CrossMark

Perturbative Calculation of Quasi-Potential in Non-equilibrium Diffusions: A Mean-Field Example

Freddy Bouchet[1] · Krzysztof Gawędzki[1] ·
Cesare Nardini[1,2,3]

Abstract In stochastic systems with weak noise, the logarithm of the stationary distribution becomes proportional to a large deviation rate function called the quasi-potential. The quasi-potential, and its characterization through a variational problem, lies at the core of the Freidlin–Wentzell large deviations theory (Freidlin and Wentzell, Random perturbations of dynamical systems, 2012). In many interacting particle systems, the particle density is described by fluctuating hydrodynamics governed by Macroscopic Fluctuation Theory (Bertini et al., arXiv:1404.6466, 2014), which formally fits within Freidlin–Wentzell's framework with a weak noise proportional to $1/\sqrt{N}$, where N is the number of particles. The quasi-potential then appears as a natural generalization of the equilibrium free energy to non-equilibrium particle systems. A key physical and practical issue is to actually compute quasi-potentials from their variational characterization for non-equilibrium systems for which detailed balance does not hold. We discuss how to perform such a computation perturbatively in an external parameter λ, starting from a known quasi-potential for $\lambda = 0$. In a general setup, explicit iterative formulae for all terms of the power-series expansion of the quasi-potential are given for the first time. The key point is a proof of solvability conditions that assure the existence of the perturbation expansion to all orders. We apply the perturbative approach to diffusive particles interacting through a mean-field potential. For such systems, the variational characterization of the quasi-potential was proven by Dawson and Gartner (Stochastics 20:247–308, 1987; Stochastic differential systems, vol 96, pp 1–10, 1987). Our perturbative analysis provides new explicit results about the quasi-potential and about fluctuations of one-particle observables in a simple example of mean field diffusions: the Shinomoto–Kuramoto model of coupled rotators (Prog Theoret Phys 75:1105–1110,

✉ Cesare Nardini
  cesare.nardini@gmail.com

[1] Laboratoire de Physique de l'Ecole Normale Supérieure de Lyon, Université de Lyon and CNRS, 46, Allée d'Italie, 69007 Lyon, France

[2] SUPA, School of Physics and Astronomy, University of Edinburgh, Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK

[3] DAMTP, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK

[74]). This is one of few systems for which non-equilibrium free energies can be computed and analyzed in an effective way, at least perturbatively.

Keywords   Large deviations theory · Freidlin-Wentzell theory · Macroscopic Fluctuation Theory · Mean-field diffusions · Perturbation theory · Fluctuating Hydrodynamics · Stochastic PDEs

1 Introduction

Large deviations theory studies the exponential decay of probabilities of large fluctuations in stochastic systems. Such probabilities are important in many fields, including physics, statistics, finance or engineering, as they often yield valuable information about extreme events far from the most probable state or trajectory of the system [23,29,82].

Weak noise large deviation theory has been developed in the 1970s by Freidlin and Wentzell [32] in a mathematical framework and by Graham and collaborators [36] with physicists' perspective. It concerns the study of large fluctuations in dynamical systems subject to weak random noise. In this framework, the stationary probability to observe some state x of the system obeys a large deviation asymptotics

$$P_\infty(x) \asymp \exp\left(-\frac{F(x)}{\epsilon}\right), \tag{1.1}$$

where ϵ denotes the noise strength squared. The function F is called the quasi-potential. It generalizes the notion of (free) energy to general finite-dimensional systems where detailed balance does not hold.but noise is weak. If known, the quasi-potential permits to calculate, at leading order in ϵ, important statistical quantities such as the probability to observe an arbitrary large fluctuation of the system or the mean residence time that the system spends close a metastable state.

In the last 15 years, Jona-Lasinio and coworkers developed a framework to study large deviations of macroscopic quantities (like particle density or current) in a class of many-body systems. Their approach, known as the Macroscopic Fluctuation Theory, was mainly applied to stochastic lattice gases, see [6] for a recent review. Without stress on mathematical rigor, the MFT can be understood as a generalization of the Freidlin–Wentzell theory to the fluctuating diffusive hydrodynamics where a weak noise is added to non-linear diffusion equations. Indeed, it is easily obtained by employing a saddle-point approximation in a path integral formalism. Even if this approach is only formal because a mathematical meaning of non-linear fluctuating hydrodynamics is still lacking, the results obtained are in complete agreement with the ones obtained through rigorous probabilistic methods in all cases where a comparison is possible. In this paper, a further example of such an agreement will be discussed in Chapter 4.

The Freidlin–Wentzell theory provides a variational characterization of the quasi-potential through dynamical large deviations. However, its explicit computation is typically very difficult and this strongly restricts the practical applicability of the theory. This is true both for finite-dimensional and infinite-dimensional systems. Only in few cases the quasi-potential can be evaluated analytically, the most important one being, of course, the case where detailed balance holds.

The main focus of the present paper is to answer a very natural but still open question. Let us consider a system depending on a control parameter λ and suppose that we are able to calculate the quasi-potential for $\lambda = 0$. Can we build a perturbation theory to calculate the

quasi-potential for small but finite λ? Surprisingly, only few works in the literature discuss this question [26,33,38–41,51,62,72,80]. All of them focus on specific examples, perform only the 1st-order analysis and most of them consider only finite-dimensional dynamical systems (see however [75] for an example where an infinite dimensional system is analyzed, and the case of transition rates between two basins of attraction is studied). We obtain here a precise answer to the above question, giving an explicit iterative formula for computing each order in the power series expansion of the quasi-potential in λ.

We first analyze the perturbation theory for finite dimensional systems, where the discussion can be made quite precise. It is well known that, given the quasi-potential function, one can deduce a simple 1st-order equation for the instanton (the corresponding variational problem minimizer). The converse is also obviously true: the values of the quasi-potential (the minima) can be easily computed from the minimizers. It is however often difficult to compute either the instantons or the quasi-potential without the knowledge of the other. We show that this loop can be broken in a perturbative setting: the quasi-potential at any order may be computed just from the knowledge of the instanton dynamics of the unperturbed problem and the quasi-potential at previous orders. This gives a very natural iterative procedure. We also explain that an equivalent simple recursive scheme appears when starting from the perturbative expansion for solutions of the Hamilton–Jacobi equation. A key point is to prove that solvability conditions hold at each order, assuring the existence of the perturbative expansion to all orders. We show that such conditions are related to the behavior close to the attractor, which gives a simple proof that they are satisfied at all orders.

In the second part of the paper, we consider a particular class of non-equilibrium many-body systems described by the Macroscopic Fluctuation Theory [6], namely diffusive particles interacting through mean-field potential and driven by an external non-conservative force. Our approach is based on the fact that for such diffusions, it is possible to derive a fluctuating hydrodynamics describing the evolution of the empirical density for large but finite number of particles N. This evolution equation was first obtained by Dean in [22] and it was thereafter called the Dean equation. Although its mathematical status is uncertain, the Dean equation allows to treat the mean-field diffusions, at least at a formal level, as a dynamical system in an infinite dimensional space perturbed by a weak noise whose strength is proportional to $1/\sqrt{N}$. In the $N \to \infty$ limit, similarly to the law of large numbers, the dynamical system becomes a deterministic equation known under the name of the McKean–Vlasov or the Vlasov–Fokker–Planck one, as was proven together with the propagation of chaos in [66,67,79].

We subsequently move our attention to the large deviations around the $N \to \infty$ behavior. The Dean equation is a formal random dynamical system to which we apply the Martin-Siggia-Rose formalism [65]. Using the saddle point approximation we then end up with an infinite-dimensional generalization of the Freidlin–Wentzell theory. This result has been previously obtained rigorously in the mathematical literature by Dawson and Gartner [18,19]. The formal approach based on the Dean equation makes an explicit connection with the Macroscopic Fluctuation Theory and adds another class of systems to the ones covered by the latter theory.

Explicit results, original to the best of our knowledge, are discussed in the case of the Shinomoto–Kuramoto model, a specific stochastic system describing coupled planar rotators [74]. This model may be also viewed as a non-equilibrium version of the dynamical mean-field XY model. We discuss in detail both its $N \to \infty$ behavior and the large deviations around it. For what concern the $N \to \infty$ limit, we are able to fully describe analytically the phase diagram of the Shinomoto–Kuramoto model, deriving self-consistent equations for the stationary states that are easily solved numerically.

The perturbation theory developed in the first part of the paper is subsequently applied to the Dean equation corresponding to the Shinomoto–Kuramoto model, resulting in an explicit calculation of the quasi-potential close to the known cases: around the free particle dynamics and in the vicinity of stationary states of the $N = \infty$ theory. This turns out to be a rather simple numerical task. We show that explicit results could be obtained to any order but, for the sake of brevity, we only present, the 1st order calculations of the quasi-potential.

The paper is organized as follows. After Introduction, we begin with a review of Freidlin–Wentzell theory results in Sect. 2. Our aim is to present the known results that can be found in [32] with a physicist perspective, but nevertheless being precise on many important points that are typically overlooked in the physics literature. In particular, we discuss under which hypothesis the quasi-potential can be seen as a solution of the Hamilton–Jacobi equation that forms the basis for the perturbative treatment that is developed in Sect. 3. In the context of finite-dimensional systems, the discussion of the perturbation theory for quasi-potential is made quite precise. We also show that the Taylor expansion of the quasi-potential close to an attractor of the deterministic dynamics is a particular case of our perturbative analysis. We conclude the section by observing that close to a codimension-one bifurcation of the deterministic dynamics, fluctuations diverge with the mean-field critical exponent. Final Sect. 4 is about N-body systems and is central to this paper. It concentrates on the case of overdamped diffusions with mean-field interaction that are introduced in Sect. 4.1. In Sect. 4.2, we discuss the fluctuating hydrodynamics for such diffusions, deriving the Dean equation making the original argument [22] more precise. The McKean-Vlasov equation is described as the limit of the Dean equation when $N \to \infty$. In Sect. 4.3, we introduce the Shinomoto–Kuramoto model [74] as a particular example of the mean-field diffusions and we analyze its McKean–Vlasov limit and the corresponding phase diagram. Sect. 4.4 derives formally, starting from the Dean equation, an extension of the Freidlin–Wentzell theory to the case of mean-field diffusions, and discusses the infinite-dimensional version of the Hamilton–Jacobi equation for the quasi-potential and, briefly, the large deviations for empirical currents. In Sect. 4.5, we adapt the perturbative approach developed for finite dimensional systems to the diffusions with mean-field interaction, concentrating on the power series in the mean-field coupling constant and the Taylor expansion close to stationary solutions of the McKean–Vlasov dynamics. A discussion of explicit results for the Shinomoto–Kuramoto model obtained by combining analytical and numerical treatments is presented. Sect. 5 summarizes the results and discusses the perspectives for future work. Two Appendices contain some additional material.

2 Freidlin–Wentzell Theory: A Brief Summary

We consider in this section a finite dimensional random dynamical system defined by the Ito stochastic differential equation[1]

$$\dot{x} = K(x) + \sqrt{2\epsilon}\, g(x)\, \eta_t, \tag{2.1}$$

where $x \in \mathbb{R}^d$, g is a $d \times m$ matrix and $\eta_t \in \mathbb{R}^m$ is a vector of white in time Gaussian noises η_t^i with zero mean and covariance

$$\mathbb{E}[\eta_t^i \eta_{t'}^j] = \delta^{ij} \delta(t - t'). \tag{2.2}$$

---

[1] Here and below, we use physicists' notation for stochastic equations rather than mathematicians' one with differentials.

Throughout the paper it is supposed that K and g are smooth and the Ito stochastic convention is employed, if not stated otherwise. We use the notation $Q(x) = g(x) g^T(x)$ with the superscript T indicating the matrix transposition and we demand that $Q(x)$ be positive definite for all x. Generalizations to the case in which Q is semi-positive definite are possible but we do not consider this situation here.

We assume that the stochastic process solving Eq. (2.1) has a unique invariant measure. Then its density $P_\infty(x)$ is smooth and solves the stationary Fokker–Planck equation

$$\sum_{i=1}^{d} \frac{\partial}{\partial x^i} \left[\left(-K^i(x) + \epsilon \sum_{j=1}^{d} \frac{\partial}{\partial x^j} Q^{ij} \right) P_\infty(x) \right] = 0. \qquad (2.3)$$

We are interested in the behavior of the above stochastic dynamical system in the small noise limit $\epsilon \ll 1$. In this limit, the stationary measure obeys the large deviation principle (1.1) [32], where the symbol $\asymp$ stands for the asymptotic logarithmic equivalence:

$$F(x) = -\lim_{\epsilon \to 0} \epsilon \ln P_\infty(x). \qquad (2.4)$$

The rate function $F(x)$ is called quasi-potential associated to the random dynamical system (2.1). It is of central importance to this paper that F, under suitable hypothesis that will be specified in Sect. 2.3, is the unique solution of the Hamilton–Jacobi equation

$$\nabla F \cdot [Q \cdot \nabla F + K](x) = 0. \qquad (2.5)$$

At an informal level, this can be guessed by looking for solutions of the stationary Fokker–Planck equation (2.3) obeying Ansatz

$$P_\infty(x) \simeq \exp\left[-\frac{F(x)}{\epsilon} + Z(x) + \ldots \right] \qquad (2.6)$$

and retaining only terms of order $1/\epsilon$. The reader may consult [36] where also the lower-order equation for $Z(x)$ is derived.

The present introductory section is devoted to an informal discussion of the Freidlin–Wentzell theory, with the particular emphasis on how the quasi-potential can be described from different points of view. This knowledge will be used in the subsequent section. We refer the reader to [32] for a mathematical presentation, to Chapter 6 of [36] for a treatment more oriented towards the physics community and to [48] for a general review oriented to computational aspects. In Sect. 2.1, we discuss dynamical large deviations, that is, the probability that a solution of Eq. (2.1) is close to a given path in the limit $\epsilon \to 0$. This permits to define quasi-potential F_A relative to an attractor A of the deterministic dynamics $\dot{x} = K(x)$. In Sect. 2.2, we describe the properties of solutions of the Hamilton–Jacobi equation (2.5) that we denote F_{HJ}. The connection between F_{HJ}, F_A, and F and the conditions under which they coincide are discussed in Sect. 2.3, concentrating on the case of attractive points. A proof of the local existence and uniqueness of F_{HJ} around a stable fixed point of the deterministic dynamics $\dot{x} = K(r)$, and of its local regularity, is sketched in Sect. 2.4.

2.1 Freidlin–Wentzell Action and Quasi-Potential Relative to an Attractor

Freidlin and Wentzell considered the probability for a trajectory of the stochastic process $x(\cdot)$ defined by Eq. (2.1) to be arbitrarily close to a given continuous path $\hat{x}(\cdot)$ on the time interval $[t_i, t_f]$. They showed rigorously that

$$\lim_{\delta \downarrow 0} \lim_{\epsilon \downarrow 0} \inf \, \epsilon \ln \mathbb{P} \left[\sup_{t_i \leq t \leq t_f} |x(t) - \hat{x}(t)| < \delta \right] = \lim_{\delta \downarrow 0} \lim_{\epsilon \downarrow 0} \sup \, \epsilon \ln \mathbb{P} \left[\sup_{t_i \leq t \leq t_f} |x(t) - \hat{x}(t)| < \delta \right]$$
$$= -\mathcal{A}[\hat{x}(\cdot)], \tag{2.7}$$

where $| \cdot |$ denotes the norm of a vector in R^d and

$$\mathcal{A}[x(\cdot)] = \frac{1}{4} \int_{t_i}^{t_f} [\dot{x}(t) - K(x(t))] \cdot Q(x(t))^{-1} [\dot{x}(t) - K(x(t))] \, dt \tag{2.8}$$

is the so-called Freidlin–Wentzell action functional, see Theorem 2.3 in Chapter 3 of [32]. Observe that such a functional vanishes on solutions of the deterministic equation

$$\dot{x} = K(x). \tag{2.9}$$

It measures the difficulty for a trajectory to deviate, due to a weak noise, from the deterministic behavior.

The above result may be elucidated in the framework of the formal path-integral approach going back to Onsager and Machlup [27,54,61,65,77]. In this approach, the transition probability from a state x_i at time $t = t_i$ to a state x_f at time $t = t_f$ is written as the path integral

$$\mathbb{P}[x_i, t_i; x_f, t_f] \asymp \frac{1}{Z_A} \int \exp \left[-\frac{\mathcal{A}[x(\cdot)]}{\epsilon} \right] \mathcal{D}[x(\cdot)], \tag{2.10}$$

where the functional integration is restricted to the paths $[t_i, t_f] \ni t \mapsto x(t)$ such that $x(t_i) = x_i$ and $x(t_f) = x_f$. Z_A is the normalization factor. The exponential factor $\exp \left[-\frac{\mathcal{A}[x(\cdot)]}{\epsilon} \right]$ plays then for small ϵ the role of the probability density in the space of paths.

Let us now consider an attractor A of the deterministic dynamical system (2.9) and a point $x_0 \in A$. The quasi-potential relative to x_0 is defined as

$$F_{x_0}(x) = \min_{\{\hat{x}(\cdot) \, | \, \hat{x}(-\infty) = x_0, \, \hat{x}(0) = x\}} \mathcal{A}[\hat{x}(\cdot)], \tag{2.11}$$

where the minimum is over all absolutely continuous paths starting from x_0 at time $t = -\infty$ and ending in x at time $t = 0$. The choice of the time interval $[-\infty, 0]$ is arbitrary and any semi-infinite interval $[-\infty, t_f]$ would give the same result. It is easy to see that F_{x_0} does not depend on the choice of $x_0 \in A$ and is constant on attractors. Indeed, given two points on the attractor, one can always find trajectories starting and ending arbitrarily close to them on which the action is arbitrarily small. As an example, the reader may consider a limit cycle where any two points of the attractor can be connected by a solution of the deterministic evolution $\dot{x} = K(x)$. For an attractor A, we shall denote by F_A the quasi-potential relative to any $x_0 \in A$. For the sake of simplicity, we shall mainly consider in this paper single point attractors referring the reader to [36] and references therein for explicit calculations of quasi-potentials with respect to non-trivial attractors, for example, limit cycles or the Lorentz attractor.

The quasi-potential F can be built from the quasi-potentials relative to the attractors of $\dot{x} = K(x)$. In the case where a fixed point $\bar{x}$ is the only attractor of the deterministic dynamics $\dot{x} = K(x)$ for any initial condition, $F_{\bar{x}}$ coincides with the quasi-potential F defined through the invariant measure in (2.4). Formally, this may be understood by observing that $\mathbb{P}[x_i, t_i; x_f, t_f]$ is actually independent of x_i in the $t_f \to \infty$ limit and thus $P_\infty(x) = \mathbb{P}[x_i, t_i; x, \infty]$. Then, one obtains the stated result by applying the saddle point approximation to the right hand side of (2.10). The reader may consult Theorem 4.3 in Chapter 4 of [32] for the rigorous result.

If more attractors $\{A_i\}_{1 \leq i \leq I}$ are present then the quasi-potential F may still be constructed once the quasi-potentials relative to each attractor F_{A_i} is known. One has

$$F(x) = \min_i \left(F_{A_i}(x) + C_i \right) - \min_i C_i \qquad (2.12)$$

where constants C_i describe the "height" of each attractor A_i. More precisely, one has to consider the i-graphs $G(i)$ on the set $\{A_1, \ldots, A_I\}$ of attractors composed of arrows $j \to k$ such that $j \neq i$, from every $A_j \neq A_i$ starts exactly one arrow and there are no closed cycles. Constants C_i are then defined as

$$C_i = \min_{G(i)} \sum_{\{j \to k\}} F_{A_j}(A_k), \qquad (2.13)$$

where the minimum is over all i-graphs and the sum over all the arrows of an i-graph and $F_{A_j}(A_k) = F_{A_j}(x)$ for any $x \in A_k$. This rigorous result is discussed in detail in Chapter 6 of [32]. Equation (2.12) balances the contributions from different attractors with the use of the invariant measure of the Markov chain with transition probabilities describing the passages between different attractors. This point was discussed in [43] from physicist's perspective. In that reference, a computation of constants C_i for few explicit examples was also carried out. In the present paper, we concentrate on the calculation of quasi-potentials $F_{\bar{x}}$ relative to attractive points $\bar{x}$ and thus we do not enter into further details on how the heights C_i may be practically found.

2.2 Transverse Decomposition, Fluctuation and Relaxation Dynamics

In the previous section we saw that the quasi-potential F can be obtained by solving the variational problem given by Eqs. (2.11), (2.12) and (2.13). In this section, we discuss a different approach that will permit to obtain F as a solution of the Hamilton–Jacobi equation (2.5) of Sect. 2.3.

Consider an open set $D \subseteq \mathbb{R}^d$ and its closure $\bar{D} = D \cup \partial D$, where ∂D denotes the boundary of D assumed to be smooth. We suppose in this section that the vector field K admits a transverse decomposition in $\bar{D}$ in the following sense: there exists a smooth function $F_{HJ}(x)$ such that

$$K(x) = -(Q \nabla F_{HJ})(x) + G(x) \qquad (2.14)$$

and

$$\nabla F_{HJ}(x) \cdot G(x) = 0 \qquad (2.15)$$

for all $x \in \bar{D}$. The existence of a transverse decomposition is equivalent to demanding that F_{HJ} solves the Hamilton–Jacobi equation (2.5). Indeed, if K admits a transverse decomposition, then $\nabla F_{HJ} \cdot K = -\nabla F_{HJ} \cdot Q \nabla F_{HJ}$. Conversely, we can define $G = K + Q \nabla F_{HJ}$ and from Eq. (2.5) we obtain the transversality condition. The term $Q \nabla F_{HJ}$ may be viewed as the gradient of F_{HJ} in the Riemannian metric defined by the matrices $Q(x)^{-1}$ and then Eq. (2.15) states its orthogonality to G with respect to the corresponding scalar product of the vector fields. Such an interpretation is often employed in the mathematical literature but we shall not pursue it here.

The deterministic dynamics (2.9) with $K(x) = -(Q \nabla F_{HJ})(x) + G(x)$ is called relaxation dynamics. The reason for this name is that as the Freidlin–Wentzell action vanishes on its trajectories, they are in the small noise limit the most probable trajectories relaxing to an attractor. The presence of a transverse decomposition permits to define the so-called fluctuation or instanton dynamics

$$\dot{x} = 2(Q \nabla F_{HJ})(x) + K(x) = (Q \nabla F_{HJ})(x) + G(x) \equiv K_r(x) \qquad (2.16)$$

which plays a fundamental role in what follows. We shall also be interested in the Freidlin–Wentzell action functional corresponding to the stochastic dynamics

$$\dot{x} = K_r + \sqrt{2\epsilon}\, g(x)\, \eta_t, \tag{2.17}$$

which is

$$\mathcal{A}_r[x(\cdot)] = \frac{1}{4} \int_{t_i}^{t_f} [\dot{x}(t) - K_r(x(t))] \cdot Q(x(t))^{-1} [\dot{x}(t) - K_r(x(t))]\, dt. \tag{2.18}$$

The fluctuation dynamics is connected to time-reversal of the stochastic equation (2.1). To understand this point, let us consider the diffusion process defined by (2.1) in the time interval $[0, T]$. It was shown in [47] that its time-reversal corresponds to the stochastic dynamics

$$\dot{x} = -\widetilde{K}(x) + \sqrt{2\epsilon}\, g(x)\, \eta_t, \tag{2.19}$$

where $\widetilde{K}(x) = K(x) - 2\epsilon\, P(x, T-t)^{-1} \nabla \cdot [Q(x)P(x, T-t)]$ and $P(x, t)$ is the solution to the time-dependent Fokker–Planck equation associated to (2.1). If we consider Eq. (2.1) with initial condition distributed accordingly to $P_\infty(x)$ then it follows from Eq. (1.1) that $\lim_{\epsilon \to 0} \widetilde{K}(x) = K(x) + 2(Q\nabla F)(x)$. Under suitable hypothesis given in Sect. 2.3, F and F_{HJ} coincide and then $\lim_{\epsilon \to 0} \widetilde{K}(x) = K_r(x)$.

We now discuss some properties of F_{HJ} in connection to relaxation and fluctuation dynamics. The stationary points of F_{HJ} correspond to zeros of the vector field K. Indeed

$$(K \cdot Q^{-1} K)(x) = (\nabla F_{HJ} \cdot Q \nabla F_{HJ})(x) + (G \cdot Q^{-1} G)(x) \geq 0, \tag{2.20}$$

as it can be directly proven by inserting Eq. (2.14) into the left hand side and using the transverse decomposition. Then, if $\bar{x}$ is such that $K(\bar{x}) = 0$, the above expressions imply that $\nabla F_{HJ}(\bar{x}) = 0 = G(\bar{x})$. The converse is also true if we suppose that the Hessian matrix of F_{HJ}, denoted by $(\nabla \nabla F_{HJ})$, is invertible at $\bar{x}$. Indeed, by taking the gradient of the transversality condition and evaluating it at $\bar{x}$, we infer that

$$(\nabla \nabla F_{HJ} G)(\bar{x}) = 0. \tag{2.21}$$

By invertibility of the Hessian matrix, this implies that $G(\bar{x}) = 0$ and hence $K(\bar{x}) = 0$. Moreover, F_{HJ} is a Lyapunov function for both the relaxation dynamics and the time-reverse of the fluctuation dynamics

$$\dot{x} = -K_r(x). \tag{2.22}$$

Indeed, if $\dot{x} = K(x)$ then

$$\frac{dF_{HJ}(x)}{dt} = -(\nabla F_{HJ} \cdot Q \nabla F_{HJ})(x) \leq 0, \tag{2.23}$$

and analogously for Eq. (2.22). From the above properties, we conclude that the relaxation and fluctuation dynamics have the same stationary points but attractors are transformed into repellers and vice-versa. Thus the relaxation and the time-reversal of the fluctuation dynamics have the same attractors and F_{HJ} as a Lyapunov function.

One could be led to an incorrect conclusion that basins of attraction of those two dynamics are the same which, however, is not true because the transverse parts of dynamics can push the systems to different attractors. This can be checked, for example, in a very simple two-dimensional bistable system, see Fig. 1.
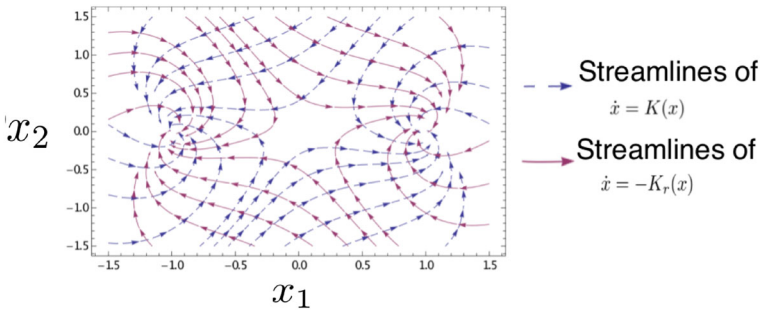
Fig. 1 Example showing a simple case where the basins of attraction of the relaxation and of the time-reversal of the fluctuation dynamics do not coincide. Here, $F_{HJ}(x_1, x_2) = -\frac{x_1^2}{2} + \frac{x_1^4}{4} + \frac{x_2^2}{2}$, $Q = \mathbb{I}$, $G = (\partial F_{HJ}/\partial x_2, -\partial F_{HJ}/\partial x_1)$, where $\mathbb{I}$ is the 2×2 unit matrix. The *dashed blue arrows* indicate the direction of the relaxation dynamics $\dot{x} = K(x) = -Q \nabla F_{HJ}(x) + G(x)$ and the continuous *red lines* the direction of the time-reversal of the fluctuation dynamics $\dot{x} = -K_r(x) = -Q \nabla F_{HJ}(x) - G(x)$ (Color figure online)

2.3 Quasi-Potential as a Solution of the Hamilton–Jacobi Equation

Let us suppose that F_{HJ} is a smooth solution of the Hamilton–Jacobi equation

$$\nabla F_{HJ} \cdot [Q \nabla F_{HJ} + K](x) = 0, \tag{2.24}$$

in $\bar{D} = D \cup \partial D \subseteq \mathbb{R}^d$ containing in the interior a fixed point $\bar{x}$ such that $F_{HJ}(\bar{x}) = 0$ and $F_{HJ}(x) > 0$, $\nabla F_{HJ}(x) \neq 0$ for $x \in \bar{D}$, $x \neq \bar{x}$. We shall assume that D is bounded and connected. We want to understand the relation between $F_{\bar{x}}$ and F_{HJ}, see Theorem 3.1 in Chapter 4 of [32] for more details.

For any path $[-\infty, 0] \ni t \mapsto \hat{x}(t) \in \bar{D}$ subjected to the boundary conditions $\hat{x}(-\infty) = \bar{x}$ and $\hat{x}(0) = x$, a simple manipulation shows that

$$\mathcal{A}[\hat{x}(\cdot)] = \mathcal{A}_r[\hat{x}(\cdot)] + \int_{-\infty}^{0} \frac{d\hat{x}(t)}{dt} \cdot \nabla F_{HJ}(\hat{x}(t)) \, dt = \mathcal{A}_r[\hat{x}(\cdot)] + F_{HJ}(x) \geq F_{HJ}(x)$$

$$\tag{2.25}$$

where $\mathcal{A}_r[\hat{x}(\cdot)]$ was defined in (2.18) and we used the transversality condition. The inequality

$$\mathcal{A}[\hat{x}(\cdot)] \geq F_{HJ}(x) \tag{2.26}$$

still holds if we allow the trajectory $\hat{x}(\cdot)$ to go out of the closed set $\bar{D}$ provided that $x \in \bar{D}_{HJ}$, the closure of the open set

$$D_{HJ} = \left\{ x \in D \mid F_{HJ}(x) < \min_{y \in \partial D} F_{HJ}(y) \right\}. \tag{2.27}$$

Indeed, writing the inequality (2.25) for the trajectory $\hat{x}(\cdot)$ restricted to $[-\infty, \tau]$, where τ is the first exit time from D, we obtain the lower bounds

$$\mathcal{A}[\hat{x}(\cdot)] \geq F_{HJ}(\hat{x}(\tau)) \geq F_{HJ}(x) \tag{2.28}$$

with the second one resulting from the condition $x \in \bar{D}_{HJ}$. Action $\mathcal{A}_r[\hat{x}(\cdot)]$ attains its minimum equal to zero on the trajectory of the fluctuation dynamics $\dot{x} = K_r(x)$. It is easy to see that for $x \in \bar{D}_{HJ}$ there exists a unique such trajectory $[-\infty, 0] \ni t \mapsto \tilde{x}(t, x)$ for which $\tilde{x}(-\infty, x) = \bar{x}$ and $\tilde{x}(0, x) = x$. Besides, such trajectory lies entirely in $\bar{D}_{HJ}$. This follows

from the fact, that on each solution of the fluctuation dynamics that ends at x, function F_{HJ} decreases backward in time and such a solution may be infinitely extended in negative time direction until it reaches $\bar{x}$ at $t = -\infty$. The resulting trajectory saturates the inequality (2.25). One infers that for $x \in \bar{D}_{HJ}$,

$$F_{HJ}(x) = \min_{\{\hat{x}(\cdot) \,|\, \hat{x}(-\infty) = \bar{x}, \, \hat{x}(0) = x\}} \mathcal{A}[\hat{x}(\cdot)] = F_{\bar{x}}(x), \qquad (2.29)$$

see (2.11). D_{HJ} is again a bounded open connected neighborhood of $\bar{x}$. Its boundary ∂D_{HJ} is composed of points of $x \in \bar{D}$ for which $F_{HJ}(x) = \min_{y \in \partial D} F_{HJ}(y)$. It is smooth since $\nabla F_{HJ} \neq 0$ at such points.

We have just obtained two results. First, a smooth solution of the Hamilton–Jacobi equation (2.5) on $\bar{D}$, with the properties stated at the beginning of the section, coincides with $F_{\bar{x}}$ on $\bar{D}_{HJ}$ where D_{HJ} is the sub-domain of D containing $\bar{x}$ defined by (2.27). Second, on $\bar{D}_{HJ}$,

$$F_{HJ}(x) = F_{\bar{x}}(x) = \mathcal{A}[\tilde{x}(\cdot, x)], \qquad (2.30)$$

where $[-\infty, 0] \ni t \mapsto \tilde{x}(t, x) \in \bar{D}_{HJ}$ is the solution to the instanton dynamics, i.e. the unique trajectory joining $\bar{x}$ to x and satisfying Eq. (2.16).

2.4 Hamiltonian Picture

Above we have assumed the local existence of a smooth solution of the Hamilton–Jacobi equation around an attractive point $\bar{x}$ of the relaxation dynamics (2.9). The linearization of such dynamics around $\bar{x}$ has the form

$$\dot{x} = Ax, \qquad (2.31)$$

where $A = (\nabla K(\bar{x}))^T$ is the matrix with entries $A^i_{\ j} = \nabla_j K^i(\bar{x})$. We shall assume that all eigenvalues of A have negative real parts, which ensures the exponential convergence to $\bar{x}$ in the vicinity of the attractor. Such attractive points will be called non-degenerate. We shall sketch below a proof of the fact that around non-degenerate attractive point $\bar{x}$ there exists a unique local smooth solution of the Hamilton–Jacobi equation, The argument we present here based on the analysis of the dynamics of extremal trajectories of the Freidlin–Wentzell action functional, see [21] for a more global discussion.

Let us start by considering the Hamiltonian $H(x, p)$ related by the Legendre transform to the Lagrangian

$$L(x, \dot{x}) = \tfrac{1}{4}(\dot{x} - K(x)) \cdot Q(x)^{-1}(\dot{x} - K(x)) \qquad (2.32)$$

appearing in the Freidlin–Wentzell action (2.8). One has

$$H(x, p) = p \cdot \dot{x} - L(x, \dot{x}) = p \cdot Q(x)p + p \cdot K(x) \qquad (2.33)$$

for $p = \tfrac{1}{2} Q(x)^{-1}(\dot{x} - K(x))$. The Euler-Lagrange equations for the extremal trajectories of the Freidlin–Wentzell action correspond to the Hamilton equations

$$\dot{x} = \nabla_p H(x, p) = 2Q(x)p + K(x), \qquad \dot{p} = -\nabla_x H(x, p) = -\nabla_x(p \cdot Q(x)p + p \cdot K(x)). \qquad (2.34)$$

The dynamical system (2.34) in the phase space $\mathbb{R}^{2d}$ possesses a hyperbolic fixed point with $(x, p) = (\bar{x}, 0)$. Indeed, the right hand sides of Eq. (2.34) vanish at this point and the linearization of (2.34) around $(\bar{x}, 0)$ has the form

$$\dot{x} = 2Q(\bar{x})p + Ax, \qquad \dot{p} = -A^T p. \qquad (2.35)$$

Fig. 2 Hamiltonian flow around $(\bar{x}, 0)$ with the stable and unstable manifolds
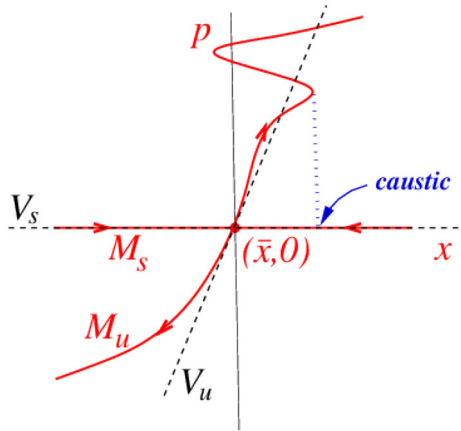
Phase space $\mathbb{R}^{2d}$ may be split in a unique way into as a direct sum $V_s \oplus V_u$ of the stable and unstable invariant subspaces of the linearized flow (2.35),

$$V_s = \{(x, 0) \,|\, x \in \mathbb{R}^d\}, \qquad V_u = \{(x, Bx) \,|\, x \in \mathbb{R}^d\}, \tag{2.36}$$

where

$$B^{-1} = 2 \int_0^\infty e^{tA} \, Q(\bar{x}) \, e^{tA^T} \, dt. \tag{2.37}$$

is the unique positive definite matrix satisfying the relation

$$BA + A^T B = -2B \, Q(\bar{x})B. \tag{2.38}$$

On V_s and V_u the linearized flow reduces to

$$(\dot{x}, 0) = (Ax, 0), \qquad (\dot{x}, B\dot{x}) = (-B^{-1}A^T Bx, -A^T Bx) \tag{2.39}$$

and is, respectively, exponentially contracting and exponentially expanding (the eigenvalues of $-B^{-1}AB$ are the negatives of those of A and have positive real parts). The subspaces V_s and V_u are Lagrangian, i.e. the symplectic form $\omega = dp \cdot dx$ vanishes when restricted to each of them. It follows from a general theory of hyperbolic fixed points, see e.g. [50,81], that in the vicinity of $(\bar{x}, 0)$ there exist unique stable and unstable smooth d-dimensional submanifolds M_s and M_u composed of close points tending to $(\bar{x}, 0)$ under, respectively, flow (2.34) and its time-reversal. M_s is a local piece of V_s around $\bar{x}$ and M_u has V_u as the tangent space at $(\bar{x}, 0)$, see Fig. 2. Besides, if K and Q are (real) analytic then the submanifolds M_s and M_u are also analytic, see Theorem 7.1 in [49].

Both M_s and M_u are Lagrangian submanifolds of the phase space (i.e. 2-form ω vanishes when restricted to their tangent subspaces). This follows from the Hamiltonian nature of the flow (2.34) which preserves ω. Besides the Hamiltonian H (conserved by the flow (2.34)) has to vanish both on M_s and on M_u since it vanishes at $(\bar{x}, 0)$. We may now define for x in a small ball D around $\bar{x}$

$$F(x) = \int_{\bar{x}}^x p \cdot dx \tag{2.40}$$

with the result independent of the integration path $(x(t), p(t))$ in M_u such that $x(t)$ lies in D. Then M_u is given locally by the equation $p(x) = \nabla F(x)$, where $(x, p(x)) \in M_u$. Clearly, F is a smooth function on D (which is analytic if K and Q are). The vanishing of the Hamiltonian H on M_u implies now that F satisfies the Hamilton–Jacobi equation (2.5) so that we may set $F_{HJ} = F$. Note that for such F_{HJ}, the Hamiltonian dynamics on the stable manifold M_s projects to the position space to the relaxation dynamics and the one on the unstable manifold M_u to the fluctuation dynamics. F_{HJ} and ∇F_{HJ} vanish at $\bar{x}$ and the Hessian $\nabla\nabla F_{HJ}(\bar{x}) = B$ is positive definite.

Conversely, if F_{HJ} is a local solution of the Hamilton–Jacobi equation around $\bar{x}$ with the latter properties then, for a small ball D around $\bar{x}$, the sets $\{(x, 0) \,|\, x \in D\}$ and $\{(x, \nabla F_{HJ}(x)) \,|\, x \in D\}$ form, respectively, the local stable and unstable manifolds of the fixed point $(\bar{x}, 0)$ of the Hamiltonian flow (2.34) so that $F_{HJ} = F$.

The above argument shows also the local existence of a transverse decomposition (2.14) and (2.15) of the vector field K around its non-degenerate stable zeros $\bar{x}$. The decomposition has positive definite Hessian $\nabla\nabla F_{HJ}(\bar{x})$ and is uniquely determined by this property.

The results discussed here, together with the local equality $F_{HJ} = F_{\bar{x}}$, show that the quasi-potential relative to $\bar{x}$ is smooth (or analytic for K and Q analytic) in the vicinity of a non-degenerate attractive fixed point $\bar{x}$ of the deterministic dynamics (2.9). It is however well known that $F_{\bar{x}}$ does not have to be smooth everywhere, see e.g. [37,42,55,63] and references therein. Such non-smoothness occurs if the unstable manifold of the fixed point $(\bar{x}, 0)$ of the Hamiltonian dynamics (2.34) has tangent vectors perpendicular to V_s leading to the caustics in its projection on V_s, see Fig. 2. In recent literature, such situations were connected to the so called Lagrangian phase transitions [6]. Non-smooth quasi-potentials may be treated using viscous solutions of Hamilton–Jacobi equations [12,31]. We leave such situations to a future investigation.

3 Perturbative Expansions of Quasi-Potentials

Let us consider a finite-dimensional stochastic dynamics depending smoothly on a real external parameter λ,

$$\dot{x} = K^\lambda(x) + \sqrt{2\epsilon}\, g^\lambda(x)\, \eta_t. \tag{3.1}$$

We shall use the notation $Q^\lambda = g^\lambda(g^\lambda)^T$ for the noise covariance and F^λ for the quasi-potential relative to the attractor $\bar{x}^\lambda$ of the dynamics $\dot{x} = K^\lambda(x)$ (dropping the subscript indicating the attractor). For simplicity, we only consider fixed points as attractors, even if all the results in this section can be generalized to other kinds of attractors.

An explicit calculation of the quasi-potentials F^λ for the dynamics of the form (3.1) is usually an impossible task. However, it is feasible in particular situations, as for example in the cases when (3.1) respects detailed balance. It is thus very natural to ask the following question. Supposing that we are able to calculate the quasi-potential for a given value of λ, say $\lambda = 0$, can we perturbatively calculate F^λ for small λ? Some works are available in the literature containing the first-order analysis [33,38–41,51,62] and concentrating mainly on specific examples. In [80], the first-order theory is presented in a general fashion and, recently, a rigorous 1st order analysis has been obtained in [72]. We extend here the approach of [80] to any order.

Before discussing in details how the perturbative approach is built in the rest of the section, we first summarize the main ideas. The strategy is to consider the quasi-potential as a solution to the Hamilton–Jacobi equation

$$\nabla F^\lambda \cdot \left[Q^\lambda \nabla F^\lambda + K^\lambda \right] (x) = 0, \tag{3.2}$$

which holds under the hypotheses discussed in Sect. 2.3. In particular, we shall assume that the vector field K^0 has a non-degenerate stable zero $\bar{x}^0$. From the Implicit Function Theorem, it follows that for sufficiently small $|\lambda|$ there exists a smooth family $\bar{x}^\lambda$ of non-degenerate zeros of K^λ. By a slight modification of the arguments in Secs. 2.3 and 2.4, invoking the dependence on a parameter of the unstable manifold of a hyperbolic fixed point, we infer that there exists a smooth family of solutions $F_{HJ}^\lambda(x)$ of the Hamilton–Jacobi equation (3.2) defined in a neighborhood of $\lambda = 0$ and $\bar{x}^0$ and that those solutions coincide with the quasi-potentials F^λ for the stochastic dynamics (3.1) relative to the attractors $\bar{x}^\lambda$ on the set D_{HJ}^0 for a sufficiently small neighborhood D of $\bar{x}^0$. D_{HJ}^0 is given by Eq. (2.27) for $\lambda = 0$. Besides $F^\lambda(\bar{x}^\lambda) = 0$, $\nabla F^\lambda(\bar{x}^\lambda) = 0$ and the Hessians $\nabla\nabla F^\lambda(\bar{x}^\lambda)$ are positive definite. For $K^\lambda(x)$ and $Q^\lambda(x)$ analytic in λ and x, both $\bar{x}^\lambda$ and $F_{HJ}^\lambda(x)$ will be analytic for sufficiently small $|\lambda|$.

For $\bar{x}^0 + y \in D_{HJ}^0$, we shall expand the function $\lambda \mapsto F^\lambda(\bar{x}^\lambda + y)$, well defined and smooth for sufficiently small $|\lambda|$, into the infinite Taylor series around $\lambda = 0$, writing

$$F^\lambda(\bar{x}^\lambda + y) = \sum_{n=0}^{\infty} \lambda^n F^{(n)}(y). \tag{3.3}$$

Such a Taylor expansion is asymptotic in the smooth case but has a finite radius of convergence in the analytic case. An analogous notation will be used for the expansions of K^λ and Q^λ centered at $\bar{x}^\lambda$. We suppose that $F^{(0)}$ is explicitly known and attempt to calculate the perturbative coefficients $F^{(n)}(y)$. Inserting the above expansion into the Hamilton–Jacobi equation (3.2), one obtains a hierarchy of equations expressing $F^{(n)}$ in terms of functions $F^{(k)}$ with $k < n$ so that they may be solved iteratively. We stress that in Eq. (3.3) we have moved attractors $\bar{x}^\lambda$ to the origin. This may seem just a detail but it is important in order to get a simple proof that the equations for $F^{(n)}$ admit a unique solution for $\bar{x}^0 + y \in D_{HJ}^0$. Recall from Sect. 2.3 that $\bar{x}^0$ may be connected to each $x \in D_{HJ}^0$ by a unique trajectory of the $\lambda = 0$ fluctuation dynamics which, after the shift by $-\bar{x}^0$, takes the form

$$\dot{y} = K_r^{(0)}(y) = 2 \left(Q^{(0)}(y) \nabla F^{(0)}(y) \right) + K^{(0)}(y). \tag{3.4}$$

The solution for each $F^{(n)}$ can be obtained with the method of characteristics by integrating along the trajectories of (3.4), see Eq. (3.16) below for the explicit expression for $F^{(n)}$. Such observation is of practical importance: to compute $F^{(n)}$, the only difficulty is to compute solutions of Eq. (3.4). This might not be doable analytically, but it is a simple problem for a numerical treatment.

From the practical point of view it may be more convenient to implement a modified version of perturbative expansion where we replace Eq. (3.3) with a Taylor expansion of the function $\lambda \mapsto F^\lambda(x)$ at $\lambda = 0$ for $x \in D_{HJ}^0$:

$$F^\lambda(x) = \sum_{n=0}^{\infty} \lambda^n \hat{F}^{(n)}(x). \tag{3.5}$$

We shall see an example where expansion (3.5), that we shall call direct, is simpler to implement than Eq. (3.3). The difference between the two expansions comes from the fact that here we do not move the λ-attractors to the origin. In this case, however, the simplest route to prove that functions $\hat{F}^{(n)}$ may be again found iteratively with the method of characteristics, is to reduce expansion (3.5) to (3.4), see Sect. 3.2.

Finally we shall show in Sect. 3.3 that, for a suitable choice of K^λ and Q^λ, the expansion of Eq. (3.3) reduces to the Taylor expansion of the quasi-potential F associated to Eq. (2.1) around the attractor $\bar{x}$.

3.1 Expansion Centered on Attractors of the Perturbed Dynamics

This section discusses the perturbative solution of the Hamilton–Jacobi equation (3.2) in the form (3.3). We iteratively construct functions $F^{(n)}$ obtaining, in such a way, two results: first, we prove that the equations obeyed by $F^{(n)}$ admit a unique solution and, second, we give an explicit recursive formula that may be used to practically calculate $F^{(n)}$.

As already mentioned, we assume that $\bar{x}^0$ is a non-degenerate fixed point of the relaxation dynamics at $\lambda = 0$. This means that $K^0(\bar{x}^0) = 0$ and the eigenvalues of $\nabla K^0(\bar{x}^0)$ have negative real parts. The assumption ensures that quasi-potential F^0 relative to $\bar{x}^0$ (recall that we have dropped the subscript indicating the attractor) is given by the solution to the Hamilton–Jacobi equation in the neighborhood D^0_{HJ} of $\bar{x}^0$, that $F^0(\bar{x}^0)$ and $\nabla F^0(\bar{x}^0)$ vanish, and that the Hessian $\nabla\nabla F^0(\bar{x}^0)$ is positive definite and $F^0(x) > 0$ for $x \in D^0_{HJ}$, $x \neq \bar{x}^0$, see Sect. 2.3.

Let us expand the drift and the noise covariance in powers of λ after shifting the attractor $\bar{x}^\lambda$ of the perturbed deterministic dynamics to the origin:

$$K^\lambda(\bar{x}^\lambda + y) = \sum_{n=0}^{\infty} \lambda^n K^{(n)}(y), \qquad Q^\lambda(\bar{x}^\lambda + y) = \sum_{n=0}^{\infty} \lambda^n Q^{(n)}(y). \tag{3.6}$$

Several properties that will be used below follow simply. Because $\bar{x}^\lambda$ is a fixed point for the deterministic dynamics, $K^{(n)}(0) = 0$. Similarly, since $F^\lambda(\bar{x}^\lambda) = 0$ and $\nabla F^\lambda(\bar{x}^\lambda) = 0$, the relations $F^{(n)}(0) = 0$ and $\nabla F^{(n)}(0) = 0$ must hold for every n. Moreover, the eigenvalues of $\nabla K^{(0)}(0)$ have negative real parts and $\nabla\nabla F^{(0)}(0)$ and $Q^{(0)}$ are positive definite.

Inserting expansions (3.3) and (3.6) into the Hamilton–Jacobi equation (3.2), we obtain the power-series identity

$$\sum_{n=0}^{\infty} \lambda^n \sum_{k=0}^{n} \left[\sum_{l=0}^{n-k} \nabla F^{(n-k-l)} \cdot Q^{(k)} \nabla F^{(l)} + \nabla F^{(n-k)} \cdot K^{(k)} \right](y) = 0. \tag{3.7}$$

Upon equating to zero order by order, this gives a hierarchy of relations

$$\nabla F^{(0)} \cdot \left[Q^{(0)} \nabla F^{(0)} + K^{(0)} \right] = 0 \qquad \text{for} \quad n = 0, \tag{3.8}$$

$$\nabla F^{(n)} \cdot K_r^{(0)} = S^{(n)}[F^{(0)}, \ldots, F^{(n-1)}] \qquad \text{for} \quad n > 0, \tag{3.9}$$

where

$$K_r^{(0)} = 2 Q^{(0)} \nabla F^{(0)} + K^{(0)} \tag{3.10}$$

and $S^{(n)}$ is a functional of $F^{(0)}, \ldots, F^{(n-1)}$ given by

$$S^{(n)}[F^{(0)}, \ldots, F^{(n-1)}] = -\sum_{k=1}^{n-1} \left[\nabla F^{(n-k)} \cdot \left(Q^{(0)} \nabla F^{(k)} + K^{(k)} \right) \right.$$

$$+ \sum_{l=0}^{n-k} \nabla F^{(n-k-l)} \cdot Q^{(k)} \nabla F^{(l)} \right]$$

$$- \nabla F^{(0)} \cdot \left[Q^{(n)} \nabla F^{(0)} + K^{(n)} \right]. \tag{3.11}$$

We have arranged Eq. (3.9) in such a way that $F^{(n)}$ appears only on the left hand side. Equation (3.8) is nothing else but the Hamilton–Jacobi equation corresponding to the dynamics (3.1) with $\lambda = 0$, once we have moved the attractor to the origin. We assumed that its solution $F^{(0)}(y)$ is known. In the following, we prove that solutions to Eq. (3.9) for $n > 0$ exist and are unique. An explicit formula will permit to obtain $F^{(n)}$ given $F^{(k)}$ for $k < n$. We start by looking at the properties of the 0th order fluctuation dynamics $\dot{y} = K_r^{(0)}(y)$ when the norm $|y|$ is small. These results will be useful to prove the existence of solutions to Eq. (3.9).

3.1.1 Fluctuation Dynamics: Exponential Escape from Attractor

Let us consider Eq. (3.4) describing the 0th order fluctuation dynamics after the shift of attractor to the origin. From the results of Sect. 2.3 about the trajectories of the fluctuation dynamics it follows that for $y \in D_{HJ}^0 - \bar{x}^0 \equiv D_0$ there exists a unique solution $[-\infty, 0] \ni t \mapsto \tilde{y}(t, y)$ of Eq. (3.4) such that $\tilde{y}(-\infty, y) = 0$ and $\tilde{y}(0, y) = y$. Besides, $\tilde{y}(t, y)$ belongs to D_0 for all t.

We now show that $\tilde{y}(t, y)$ escapes from the attractor $y = 0$ exponentially fast. Indeed, we can write for $|y|$ small

$$\tilde{y}(t, y) = e^{t(2Q^{(0)}B + A)} y + o\left(\left| e^{t(2Q^{(0)}B + A)} y \right| \right). \tag{3.12}$$

In the above expressions, A and B are defined through the small $|y|$ expansion of $K^{(0)}$ and $F^{(0)}$:

$$K^{(0)}(y) = Ay + o(|y|), \qquad F^{(0)}(y) = \tfrac{1}{2} y \cdot By + o\left(|y|^2 \right), \tag{3.13}$$

i.e. $A = (\nabla K^{(0)}(0))^T$ and $B = \nabla \nabla F^{(0)}(0)$. One should recall from the properties listed above for $K^{(0)}$ and $F^{(0)}$ that the eigenvalues of A have negative real parts and $B = B^T$ is positive definite. We have encountered matrices A and B already before when studying the Hamiltonian dynamics in Sect. 2.4. They are related by the identity

$$BA + A^T B + 2B Q^{(0)} B = 0 \tag{3.14}$$

imposed by the 2nd order in y contribution to the 0th order Hamilton–Jacobi equation (3.8). Note that identity (3.14), which implies that

$$2Q^{(0)}B + A = -B^{-1}A^T B, \tag{3.15}$$

coincides with Eq. (2.38) solved by (2.37) if we replace in the latter $Q(\bar{x})$ by $Q^{(0)}(0)$. Behavior (3.12) is dictated by the spectrum of $(2Q^{(0)}B + A)$. Indeed, as already noticed before, the eigenvalues of $-BA^T B$ have positive real parts. Thus the 0th-order fluctuation dynamics escapes exponentially fast from the attractor. This agrees with the analysis of Sect. 2.4 where we showed that the fluctuation dynamics on the position space corresponds to the Hamiltonian dynamics on the unstable manifold of the hyperbolic fixed point.

3.1.2 Iterative Solution

We shall prove that the unique solution of Eq. (3.9) on $D_0 = D_{HJ}^0 - \bar{x}^0$ that satisfies $F^{(n)}(0) = 0$ is

$$F^{(n)}(y) = \int_{-\infty}^{0} S^{(n)}[F^{(0)}, \ldots, F^{(n-1)}](\tilde{y}(t, y)) \, dt, \tag{3.16}$$

where $\tilde{y}(t, y)$ is the trajectory of the 0th-order fluctuation dynamics (3.4) joining the origin to y that was discussed above. Moreover, the expression (3.16) is well-defined, as we show that the integral in this expression is convergent whenever y belongs to D_0. Equation (3.16) gives an iterative solution of (3.9).

Let us start by proving that if a solution of Eq. (3.9) such that $F^{(n)}(0) = 0$ exists, it has to have the form (3.16). This is easily seen by taking the total time derivative of $F^{(n)}(\tilde{y}(t, y))$,

$$\frac{d}{dt} F^{(n)}(\tilde{y}(t, y)) = \nabla F^{(n)}(\tilde{y}(t, y)) \cdot \frac{d\tilde{y}(t, y)}{dt} = \left(\nabla F^{(n)} \cdot K_r^{(0)} \right)(\tilde{y}(t, y))$$
$$= S^{(n)}[F^{(0)}, \ldots, F^{(n-1)}](\tilde{y}(t, y)), \tag{3.17}$$

and by integrating over time.

In the next step, we shall show that, assuming that the integral on the right hand side of (3.16) converges, the latter equation gives a function $F^{(n)}(y)$ that solves Eq. (3.9). Indeed, on the fluctuating dynamics trajectory,

$$F^{(n)}(\tilde{y}(s, y)) = \int_{-\infty}^{0} S^{(n)}[\cdots](\tilde{y}(t, \tilde{y}(s, y))) \, dt = \int_{-\infty}^{0} S^{(n)}[\cdots](\tilde{y}(t + s, y)) \, dt$$
$$= \int_{-\infty}^{s} S^{(n)}[\cdots](\tilde{y}(t, y)) \, dt, \tag{3.18}$$

where we used the relation $\bar{y}(t, \bar{y}(s, y)) = \bar{y}(t + s, y)$ that holds because of the uniqueness of the solutions of Eq. (3.4). Now, deriving the previous expression with respect to s and evaluating at $s = 0$, we obtain Eq. (3.9). This means that expression (3.16) solves Eq. (3.9) provided that it is well defined.

To complete the proof, we need to show that the integral appearing in Eq. (3.16) is convergent and vanishes at $y = 0$. To show this, we have to analyze the behavior of the integrand when $t \to -\infty$. This corresponds to studying the behavior of the integrand for small $|y|$. One completes the prove combining Eq. (3.12) with the observation that the expression $S^{(n)}$ defined in Eq. (3.11) is at least quadratic in y for $|y|$ small if $F^{(k)}$ for $k < n$ have the same property. Under this assumption, that is true for $F^{(0)}$, the integrand on the right hand side of Eq. (3.16) converges exponentially to zero when $t \to -\infty$ so that the time integral converges. Besides, it determines function $F^{(n)}(y)$ that is at least quadratic in y for small $|y|$ so that it may be shown inductively that $F^{(n)}$ starts at worst quadratically.

Let us summarize our results: we described an iterative scheme to calculate perturbatively the quasi-potential F^λ of Eq. (3.1) as a power series (3.3). Once we know the terms $F^{(k)}$ of that expansion for $k < n$ then $F^{(n)}$ may be obtained using Eq. (3.16). In that formula, $S^{(n)}$ is defined by Eq. (3.11) and $\tilde{y}(t, y)$ is the solution to the 0th-order fluctuation dynamics (3.4) that starts at the origin at $t = -\infty$ and arrives at y at time zero. Finally, we proved that this procedure is well defined ($F^{(n)}$ are finite quantities and depend smoothly, or analyticaly in the analytic case, on $y \in D_0$). More generally, one could define this way $F^{(n)}(y)$ for y belonging to the basin of attraction of the origin for the time-reversal of the fluctuation dynamics (3.4).

Let us conclude this section observing that the procedure described here may be very easily implemented numerically. Indeed, to calculate $F^{(n)}$ at all the orders, one only needs to compute the solution to the 0th-order fluctuation dynamics (3.4). The scheme just described gives a powerful practical tool to compute quasi-potentials perturbatively.

3.2 Direct Expansion

In Sect. 3.1 we have presented an expansion in power of λ centered on the attractors of the perturbed dynamics. We discuss in this section a direct expansion, which does not depend on the knowledge of the attractors of the perturbed dynamics. The main point we want to stress in that case is the appearance of non-trivial solvability conditions.

There are several reason why, in some cases, this new expansion may be simpler or more relevant than the one centered on the attractors. The main one is that the attractors of the perturbed dynamics may not be known explicitly and should then be computed themselves by a perturbative expansion. In that case, as the definition of $K^{(n)}$ and $Q^{(n)}$ given in Eq. (3.6) involves the attractor $\bar{x}^\lambda$ of the perturbed dynamics $\dot{x} = K^\lambda(x)$, then matrices $Q^{(n)}$ may be non-zero at all orders even if the covariance Q^λ is independent of λ. An example of this kind will be encountered in Sect. 4.5.

We assume power series expansions for K^λ and Q^λ

$$K^\lambda(x) = \sum_{n=0}^{\infty} \lambda^n \hat{K}^{(n)}(x), \qquad Q^\lambda(x) = \sum_{n=0}^{\infty} \lambda^n \hat{Q}^{(n)}(x). \qquad (3.19)$$

Observe that $\hat{K}^{(0)}(\bar{x}^0) = 0$ and $\nabla \hat{F}^{(0)}(\bar{x}^0) = 0$. By assumption all the eigenvalues of $\nabla \hat{K}^{(0)}(\bar{x}^0)$ have negative real part, and $Q^{(0)}(\bar{x}^0)$ is a positive definite matrix. This implies that $\nabla \nabla F^{(0)}(\bar{x}^0)$ is a positive definite matrix. As $\nabla \hat{K}^{(0)}(\bar{x}^0)$ is invertible, the identity $K^\lambda(\bar{x}^\lambda) = 0$ permits to solve iteratively for the coefficients of the Taylor expansion

$$\bar{x}^\lambda = \sum_{n=0}^{\infty} \lambda^n \hat{x}^{(n)}. \qquad (3.20)$$

for the perturbed attractor, where $\hat{x}^{(0)} = \bar{x}^0$.

Inserting Eq. (3.5) and (3.19) into the Hamilton–Jacobi equation (3.2), we obtain a hierarchy identical to (3.8) and (3.9). The only differences are, of course, that y has to be replaced with x and $F^{(n)}$, $K^{(n)}$ and $Q^{(n)}$ by $\hat{F}^{(n)}$, $\hat{K}^{(n)}$ and $\hat{Q}^{(n)}$. Explicitly, we have

$$\nabla \hat{F}^{(0)}(x) \cdot \left[\hat{Q}^{(0)}(x)\nabla \hat{F}^{(0)}(x) + \hat{K}^{(0)}(x) \right] = 0 \qquad \text{for} \quad n = 0, \quad (3.21)$$

$$\nabla \hat{F}^{(n)}(x) \cdot \hat{K}_r^{(0)}(x) = \hat{S}^{(n)}[\hat{F}^{(0)}, \ldots, \hat{F}^{(n-1)}](x) \qquad \text{for} \quad n \neq 0, \quad (3.22)$$

where

$$\hat{K}_r^{(0)} = 2\,\hat{Q}^{(0)}\nabla \hat{F}^{(0)} + \hat{K}^{(0)} \qquad (3.23)$$

and $\hat{S}^{(n)}$ is the functional of $\hat{F}^{(0)}, \ldots, \hat{F}^{(n-1)}$ given by Eq. (3.11) with $F^{(n)}$, $K^{(n)}$ and $Q^{(n)}$ replaced by the hatted quantities.

We first remark that, as $\hat{K}_r^{(0)}(x^0) = 0$, Eq. (3.22) implies

$$\hat{S}^{(n)}[\hat{F}^{(0)}, \ldots, \hat{F}^{(n-1)}]\Big|_{x=\bar{x}^0} = 0, \qquad (3.24)$$

which appears as a solvability condition for Eq. (3.22). It is possible to prove directly this solvability condition by induction, however this involves subtle cancellations that are tedious to prove to all orders. In order to bypass this proof, we rather argue that we know *a priori* that the series expansion exists. Then Eq. (3.22) is a consequence of the existence of the series expansion and this implies that the solvability condition (3.24) is satisfied. The existence of the series expansion follows from the existence of the series expansion around the attractors of the perturbed dynamics, discussed in the previous section. Indeed, $\hat{F}^{(n)}(x)$ may be directly found using the expansion from the previous section by comparing order by order both sides of the identity

$$\sum_{n=0}^{\infty} \lambda^n \hat{F}^{(n)}(x) = F^\lambda(x) = \sum_{n=0}^{\infty} \lambda^n F^{(n)}(x - \bar{x}^\lambda), \tag{3.25}$$

where on the right hand side one inserts the Taylor expansion (3.20).

We now explain how to solve Eq. (3.22) using (3.24). Let us consider the 0th-order fluctuation dynamics

$$\dot{x} = \hat{K}_r^{(0)}(x), \tag{3.26}$$

and its trajectory $\tilde{x}(t, x)$ lying in D_{HJ}^0 such that $\tilde{x}(-\infty, x) = \bar{x}^0$ and $\tilde{x}(0, x) = x$. With the same argument as in Sect. 3.1.2, we can show that

$$\hat{F}^{(n)}(x) = C^{(n)} + \int_{-\infty}^{0} \hat{S}^{(n)}[\hat{F}^{(0)}, \ldots, \hat{F}^{(n-1)}](\tilde{x}(t, x)) \, dt. \tag{3.27}$$

has to hold for a solution of Eq. (3.22), where $C^{(n)}$ are (for the moment arbitrary) constants. This is analogous to Eq. (3.16). To obtain $\hat{F}^{(n)}(x)$ from this equation, we have to prove the convergence of the integral on the right hand side of (3.27) and to fix the constants $C^{(n)}$. The first task requires the control of the behavior of $\hat{S}^{(n)}[\hat{F}^{(0)}, \ldots, \hat{F}^{(n-1)}](x)$ around $x = \bar{x}^0$. It may be achieved by induction using (3.24) and the exponential relaxations of $\tilde{x}(t, x)$ to x^0 when t goes to $-\infty$. The second task is easier and may be accomplished iteratively since the normalization $F^\lambda(\bar{x}^\lambda) = 0$ leads upon Taylor expending to the relations that allow to express $C^{(n)} = \hat{F}^{(n)}(\bar{x}^0)$ by the values at $\bar{x}^0$ of functions $\hat{F}^{(k)}$ and their derivatives and by the coefficients $\hat{x}^{(k)}$ of the Taylor expansion (3.20), all for $k < n$. Note, however, that the choice of $C^{(k)}$ for $k < n$ in not relevant for the calculation of $\hat{F}^{(n)}$, except when it comes to the choice of $C^{(n)}$. Indeed, these are the gradients of $\hat{F}^{(k)}$ for $k < n$ that enter $\hat{S}^{(n)}$.

3.3 Taylor Expansion of the Quasi-Potential Around the Attractor of the Unperturbed Dynamics

Consider now the stochastic evolution (2.1) without an external parameter. Let $\bar{x}$ be a non-degenerate attractive point of the deterministic dynamics (2.9). We shall be interested here in the Taylor expansion of the quasi-potential in a neighborhood of $\bar{x}$:

$$F(\bar{x} + y) = \sum_{n=2}^{\infty} \left(\nabla^{(n)} F \right) y^{(n)}, \tag{3.28}$$

where $\nabla^{(n)} F$ and $y^{(n)}$ are rank-n tensors with the components

$$\left(\nabla^{(n)} F \right)_{i_1 \cdots i_n} = \frac{1}{n!} \frac{\partial^{(n)} F(\bar{x})}{\partial x^{i_1} \cdots \partial x^{i_n}}, \qquad (y^{(n)})^{i_1 \cdots i_n} = y^{i_1} \cdots y^{i_n} \tag{3.29}$$

and on the right hand side of (3.28) the contraction of all indices is implied. We shall show here that expansion (3.28) can be viewed as a particular case of the perturbative expansion considered in Sect. 3.1. Then, all the results obtained there may be applied to (3.28) providing a method to calculate tensors $\nabla^{(n)} F$.

With the same notation as in (3.28), we introduce the expansions of K and Q around $\bar{x}$:

$$K(\bar{x} + y) = \sum_{n=1}^{\infty} \left(\nabla^{(n)} K \right) y^{(n)}, \qquad Q(\bar{x} + y) = \sum_{n=0}^{\infty} \left(\nabla^{(n)} Q \right) y^{(n)}. \tag{3.30}$$

Since $\bar{x}$ is a non-degenerate attractive fixed point of the relaxation dynamics, $\nabla^{(1)} K = \nabla K(\bar{x})$ has eigenvalues with negative real parts and $\nabla^{(2)} F = \frac{1}{2} \nabla \nabla F(\bar{x})$ is a positive definite matrix.

Let us introduce a new λ-dependent system defined by

$$K^{\lambda}(\bar{x} + y) \equiv \frac{1}{\lambda} K (\bar{x} + \lambda y), \qquad Q^{\lambda}(\bar{x} + y) \equiv Q (\bar{x} + \lambda y) \tag{3.31}$$

that reduces for $\lambda = 1$ to the previous one and depends smoothly on real λ. Note that point $\bar{x}$ is a non-degenerate stable attractive zero of K^{λ} for all λ and that

$$F^{\lambda}(\bar{x} + y) \equiv \frac{1}{\lambda^2} F (\bar{x} + \lambda y) \tag{3.32}$$

satisfies the Hamilton–Jacobi equation (3.2) for all λ including $\lambda = 0$. The scaling with λ was introduced in such a way that the $\lambda = 0$ case gives a non-trivial contribution. It allows to align the notations to those of Sect. 3.1. With our choice, we indeed have

$$K^0(\bar{x} + y) = (y \cdot \nabla) K(\bar{x}) \equiv Ay, \qquad Q^0(\bar{x} + y) = Q(\bar{x}), \tag{3.33}$$

$$F^0(\bar{x} + y) = \frac{1}{2}(y \cdot \nabla)^2 F(\bar{x}) \equiv \frac{1}{2} y \cdot By. \tag{3.34}$$

Moreover

$$K^{(n)}(y) = \left(\nabla^{(n+1)} K \right) y^{(n+1)}, \qquad Q^{(n)}(y) = \left(\nabla^{(n)} Q \right) y^{(n)}, \qquad F^{(n)}(y)$$

$$= \left(\nabla^{(n+2)} F \right) y^{(n+2)} \tag{3.35}$$

in the notation of Eq. (3.6) and (3.3). It is worth stressing that, in this context, the role of the unperturbed $\lambda = 0$ stochastic dynamics is played by the linear dynamic

$$\dot{y} = Ay + \sqrt{2\epsilon} \, g(\bar{x}) \, \eta_t \tag{3.36}$$

with matrix A given by (3.33), for which the unperturbed quasi-potential is the quadratic approximation of F around the attractor. This is not surprising since the stochastic equation (3.36) defines a Gaussian process of the Orstein-Uhlenbeck type whose invariant measure is Gaussian with the covariance equal to ϵB^{-1}, where B is the Hessian matrix of F at $\bar{x}$, see (3.34). The Hamilton–Jacobi equation reduces for $\lambda = 0$ to the identity (2.38) with the solution given by (2.37). The relaxation and the fluctuation dynamics are (after the shift of the attractor to the origin), respectively,

$$\dot{y} = Ay \quad \text{and} \quad \dot{y} = -B^{-1} A^T By. \tag{3.37}$$

In particular, for $\lambda = 0$ the trajectories of the fluctuation dynamics that start at $t = -\infty$ from the origin have a particularly simple form:

$$\tilde{y}(t, y) = e^{-t B^{-1} A^T B} y \tag{3.38}$$

and Eq. (3.16) reduces to the iterative solution

$$F^{(n)}(y) = \int_{-\infty}^{0} S^{(n)}[F^{(0)}, \ldots, F^{(n-1)}](e^{-tB^{-1}A^T B} y) \, dt \tag{3.39}$$

for the Taylor coefficients of F at $\bar{x}$ with $S^{(n)}$ given Eq. (3.11).

3.4 Codimension-One Bifurcations: The Critical Exponent

Let us return to a family of dynamical systems parameterized by λ as in Eq. (3.1) and let us suppose that the deterministic dynamics $\dot{x} = K^\lambda(x)$ has an attractive fixed point $\bar{x}^\lambda$ for $\lambda \leq \lambda_c$ that is non-degenerate for $\lambda < \lambda_c$ and undergoes a codimension-one bifurcation at $\lambda = \lambda_c$. Large deviations for normal forms corresponding to codimension-one and codimension-two bifurcations were discussed in a series of papers in the 1970s and 1980s. We refer to [27,28,44,57,64] and to the review [80] for a detailed analysis. Here we want only to remark that the critical exponent in this framework is equal to that of the mean-field theory.

The assumed scenario implies that $\nabla K^\lambda(x^\lambda)$ has a simple real eigenvalue α^λ that is negative for $\lambda < \lambda_c$ such that

$$\alpha^{\lambda_c} = 0 \quad \text{and} \quad \frac{d\alpha^\lambda}{d\lambda}\Big|_{\lambda=\lambda_c} \neq 0 \tag{3.40}$$

(the saddle-node bifurcation) or a complex eigenvalue α^λ with negative real part for $\lambda < \lambda_c$, and its conjugate, such that

$$\operatorname{Re} \alpha^{\lambda_c} = 0, \quad \operatorname{Im} \alpha^{\lambda_c} \neq 0 \quad \text{and} \quad \frac{d\alpha^\lambda}{d\lambda}\Big|_{\lambda=\lambda_c} \neq 0 \tag{3.41}$$

(the Hopf bifurcation). Moreover all the other eigenvalues of $\nabla K^\lambda(\bar{x}^\lambda)$ have strictly negative real parts for $\lambda \leq \lambda_c$.

Let us consider the covariance

$$(C^\lambda)^{ij} = \int x^i x^j P_\infty^\lambda(x) \, dx - \left(\int x^i P_\infty^\lambda(x) \, dx \right) \left(\int x^j P_\infty^\lambda(x) \, dx \right) \tag{3.42}$$

of the invariant measure of the stochastic dynamics (3.1). If the minimum of the quasi-potential F^λ describing the behavior (1.1) of P_∞^λ in the limit of small noise is attained at $\bar{x}^\lambda$ then the saddle-point analysis of (3.42) implies that

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} C^\lambda = (B^\lambda)^{-1} = 2 \int_0^\infty e^{t A^\lambda} Q^\lambda(\bar{x}^\lambda) e^{t(A^\lambda)^T} dt \tag{3.43}$$

where B^λ is the Hessian matrix of F^λ at $\bar{x}^\lambda$ and the last equality with $A^\lambda = (\nabla K^\lambda(\bar{x}^\lambda))^T$ follows from Eq. (2.37). Hence $(B^\lambda)^{-1}$ may be viewed as the small-noise limit of the stationary equal-time connected 2-point function of the stochastic process solving Eq. (3.1). Let us examine the behavior of the latter limiting quantity when $\lambda \nearrow \lambda_c$. Any vector $y \in \mathbb{R}^d$ may be decomposed as

$$y = \beta^\lambda v^\lambda + \overline{\beta^\lambda v^\lambda} + y^\lambda \tag{3.44}$$

where $\beta^\lambda \in \mathbb{C}$, v^λ is the eigenvector of $(A^\lambda)^T$ with the eigenvalue α^λ and $y^\lambda \in \mathbb{R}^d$ belongs to the invariant subspace of $(A^\lambda)^T$ corresponding to the other eigenvalues so that

$$|e^{t(A^\lambda)^T} y^\lambda| \leq e^{-\delta t} \tag{3.45}$$

in the vicinity of λ_c for some $\delta > 0$. Then the 2-point function

$$
\begin{aligned}
y \cdot (B^\lambda)^{-1} y = &-2|\beta^\lambda|^2 (\operatorname{Re} \alpha^\lambda)^{-1} \overline{v^\lambda} \cdot Q^\lambda(\bar{x}^\lambda) v^\lambda \\
&- 2\operatorname{Re}\left[(\beta^\lambda)^2 (\alpha^\lambda)^{-1} v^\lambda \cdot Q^\lambda(\bar{x}^\lambda) v^\lambda\right] + O(1)
\end{aligned}
\tag{3.46}
$$

where $O(1)$ term is regular when $\lambda \nearrow \lambda_c$. Assume that $\beta^{\lambda_c} \neq 0$ which holds for generic y. Then, when $\lambda \nearrow \lambda_c$, the first term on the right hand side diverges like $(\lambda_c - \lambda)^{-1}$ and the second term is regular if the purely imaginary eigenvalue $\alpha_{\lambda_c} \neq 0$, and both terms have the same divergence when $\alpha_{\lambda_c} = 0$.

This completes the simple proof that, in the framework of a dynamical systems in the low noise limit undergoing a codimension-one bifurcation, critical exponent for two-points correlator is equal to 1, the mean field theory value of the susceptibility exponent γ.

4 Mean-Field Systems and the Shinomoto–Kuramoto Model

We consider in this section systems composed of many diffusive particles, interacting through a mean-field type two-body potential and driven out of equilibrium by external forces. Most of our analysis is valid for any model in this class (the two-body interaction could even be non-potential). However, in order to go beyond formal results, we discuss in detail the example of Shinomoto–Kuramoto system, a simple 1-D model first introduced in [74], which has recently attracted some attention in the mathematical literature [6,34,35,60] as well as in the physical one [70,71,83].

It is known since the 60s that the evolution of the empirical measure of N diffusions with mean-field interaction is described, for $N \to \infty$, by a non-linear Fokker–Planck equation known in the mathematics literature as the McKean–Vlasov equation [66], see also [67,79] for more modern presentations. We refer to this limit as the mean-field behavior of the N particle system. For the Shinomoto–Kuramoto model, the mean-field behavior exhibits a rich phase diagram with stationary and periodic phases separated by bifurcation lines, as first observed in [73].

In [18,19], Dawson and Gartner studied the large deviations of the empirical measure for the mean-field diffusions through a generalization of the Freidlin–Wentzell theory to such questions. Extensions to cases where quenched disorder is present, as for example in the Kuramoto model [16,24] or in mean-field spin glasses [4,5,45], were considered too. Also some large deviation results on a mean-field model for active matter were obtained in the physics literature [3].

In this chapter, we first discuss how the mathematical results [18,19] can be obtained formally (i.e., without mathematical rigor) by writing an effective evolution for the empirical measure in the form of a stochastic PDE, called the Dean equation [22]. The noise term in this equation is proportional to $1/\sqrt{N}$ so that it vanishes as $N \to \infty$. In the latter limit, one recovers the deterministic McKean–Vlasov equation.

The evolution of the empirical measure is thus formally given by a stochastic partial differential equation with weak noise which, at variance with the cases considered in the previous sections, is infinite-dimensional. By applying functional integral techniques from field theory (the Martin-Siggia-Rose formalism, see [65]), we can write the infinite-dimensional analogue of the Freidlin–Wentzell theory. In the physics literature, this extension goes under the name of Macroscopic Fluctuation Theory and it has attracted much attention in the statistical mechanics community over last years, see [6] for a review. In the present case, we

re-obtain formally the rigorous results by Dawson and Gartner [18,19]. We also discuss a more general large deviation result covering current fluctuations.

We then apply the infinite-dimensional analogue of the perturbative scheme discussed in Chapter 3 to obtain explicit results about the quasi-potential for the mean-field diffusions. In particular, we calculate the quasi-potential perturbatively close to the free particle dynamics with no interactions between particles present. This permits also to compute perturbatively the rate function for the fluctuations of some macroscopic observables. For the Shinomoto–Kuramoto model, explicit results obtained by implementing a numerical algorithm to compute the quasi-potential to the 1st order in the coupling are presented, together with an analysis of the fluctuations of magnetization.

We also discuss how explicit results may be obtained for the Taylor expansion of the quasi-potential around a stationary solution of the McKean–Vlasov equation. Within this analysis, it is clear that the variance of density fluctuations diverges close to bifurcations when external parameters are changed, a result rigorously obtained in [2,20]. A numerical algorithm to evaluate explicitly the Taylor expansion is also discussed but we have not implemented it.

The structure of this part of the article is as follows. In Sect. 4.1, we introduce the class of systems that will be considered. Sect. 4.2 derives the Dean equation, formally describing the evolution of the empirical measure for large but finite N. Then, in Sect. 4.3, the particular case of the Shinomoto–Kuramoto model is discussed. We describe the long-time behavior of solutions of the associated McKean–Vlasov equation. Despite its simplicity, the model, which is a kinetic and non-equilibrium version of a mean-field ferromagnet, exhibits a rather complex mean-field behavior that we describe concentrating on the results that can be obtained analytically or semi-analytically. In Sect. 4.4, we discuss how a generalization of the Freidlin–Wentzell theory may be formally obtained by applying the Martin-Siggia-Rose formalism to the Dean equation. Finally, in Sect. 4.5, the perturbative calculation of the quasi-potential is performed and some explicit results for the Shinomoto–Kuramoto model are described.

4.1 Mean-Field Diffusions

Let us consider a system composed of N particles undergoing an over-damped diffusion in $\mathbb{R}^d$ or in a torus $\mathbb{T}^d$ and coupled through a mean-field 2-body potential $V(x) = V(-x)$. The equations of motions defining the stochastic evolution are

$$\dot{x}_n = b(x_n) - \frac{J}{N} \sum_{m=1}^{N} (\nabla V)(x_n - x_m) + \sqrt{2k_B T}\, \eta_n, \tag{4.1}$$

where $\eta_n(t)$ are independent white noises with zero average and covariance $\mathbb{E}\, \eta_n^i(t)\, \eta_m^j(s) = \delta^{ij}\delta_{n,m}\delta(t-s)$, $T > 0$ is the temperature, and k_B is the Boltzmann constant.

The quantities which are of central interest for us are the empirical density and the empirical current, defined as

$$\rho_N(t, x) = \frac{1}{N} \sum_{n=1}^{N} \delta(x - x_n(t)) \tag{4.2}$$

$$j_N(t, x) = \frac{1}{N} \sum_{n=1}^{N} \delta(x - x_n(t)) \circ \dot{x}_n(t) \tag{4.3}$$

where "$\circ$" stands for the product in the Stratonovich convention. It is straightforward to show that the following continuity equation holds

$$\partial_t \rho_N + \nabla \cdot j_N = 0 \tag{4.4}$$

(this uses the chain rule which imposes the Stratonovich convention in the definition of the empirical current). Moreover, substituting the equation of motion (4.1) into the definition of j_N and returning to the Ito convention, we obtain

$$j_N(t, x) = \rho_N(t, x)\big[b(x) - J\,\nabla(V * \rho_N)(t, x)\big] - k_B T \nabla \rho_N(t, x)$$
$$+ \frac{\sqrt{2k_B T}}{N} \sum_{n=1}^{N} \delta(x - x_n(t))\,\eta_n(t), \tag{4.5}$$

where $(V * \rho_N)$ is the convolution between V and $\rho_N(t, \cdot)$ and the term with $\nabla \rho_N$ was produced by the change of the stochastic convention. Observe that Eq. (4.4) with j_N given by (4.5) is not a closed equation for ρ_N because the noise term explicitly depends on the particle positions. What we would like to do, instead, is to write a closed evolution equation for ρ_N into which the particle positions enter only through ρ_N.

4.2 Evolution of the Empirical Density: The Dean Equation

A closed equation for the evolution of the empirical density was obtained from Eqs. (4.4) and (4.5) by Dean [22]. Dean's argument (somewhat brief in the original paper) can be reformulated in the following way. The last term on the right hand side of (4.5) may be viewed as a white noise in time with values in vector fields on $\mathbb{R}^d$,

$$\frac{\sqrt{2k_B T}}{N} \sum_{n=1}^{N} \delta(x - x_n)\,\eta_n^i(t), \tag{4.6}$$

parameterized by the particle positions $(x_n)_{n=1}^{N}$. White noise (4.6) has mean zero and covariance

$$\frac{2k_B T}{N^2} \delta^{ij}\, \delta(t - t') \sum_n \delta(x - x_n)\, \delta(y - x_n) = \frac{2k_B T}{N} \delta^{ij}\, \delta(t - t')\, \delta(x - y)\, \rho(x), \tag{4.7}$$

where the last expression followed by using the distributional identity $\delta(x - x_n)\, \delta(y - x_n) = \delta(x - y)\, \delta(x - x_n)$ and introducing the particle density $\rho(x) = \frac{1}{N} \sum_n \delta(x - x_n)$. Consider now another noise,

$$\sqrt{\frac{2k_B T}{N}}\, \sqrt{\rho(x)}\, \xi^i(t, x), \tag{4.8}$$

where $\xi(t, x)$ stands for the vector-valued white noise in space and time satisfying

$$\mathbb{E}\,\xi^i(t, x) = 0, \qquad \mathbb{E}\,\xi^i(t, x)\,\xi^j(s, y) = \delta^{ij}\, \delta(t - s)\, \delta(x - y). \tag{4.9}$$

Random process (4.8) may again be viewed as a white noise in time with values in vector fields on $\mathbb{R}^d$, but now parameterized by densities $\rho(x)$. It has a zero mean and covariance

$$\frac{2k_B T}{N} \delta^{ij}\, \delta(t - t')\, \delta(x - y)\, \rho(x) \tag{4.10}$$

that coincides with the one of noise (4.6) if ρ is related to particle positions as above. Dean proceeded identifying the two white noises by writing

$$j_N(x, t) = j_{\rho_N}(t, x) + \sqrt{\tfrac{2k_B T}{N}} \sqrt{\rho_N(t, x)} \, \xi(t, x), \qquad (4.11)$$

upon which the continuity equation (4.4) became a closed stochastic PDE in the space of densities for $\rho(t, x) = \rho_N(t, x)$,

$$\partial_t \rho(t, x) + \nabla \cdot j_\rho(t, x) + \sqrt{\tfrac{2k_B T}{N}} \, \nabla \cdot \left(\sqrt{\rho(t, x)} \, \xi(t, x) \right) = 0, \qquad (4.12)$$

where j_ρ is the nonlinear functional of ρ given by

$$j_\rho(t, x) = \rho(t, x) \left(b(x) - J \int (\nabla V)(x - y) \, \rho(t, y) \, dy \right) - k_B T \, \nabla \rho(t, x). \quad (4.13)$$

We shall call (4.12) the Dean equation for the empirical density.

Dean's substitution is formal in the infinite-dimensional situation involving the space of densities but would be legitimate in a finite-dimensional setup. To explain what we mean, let us consider the backward Kolmogorov equation describing the evolution of averages for functionals of the empirical density $\mathcal{S}[\rho_N(t, \cdot)]$. Denoting by $\mathbb{E}$ the average with respect to the noises η_n and applying the Ito calculus, we infer that[2]

$$
\begin{aligned}
\frac{d}{dt} \mathbb{E} \, \mathcal{S}[\rho_N(t, \cdot)] &= \int dx \, \frac{\delta \mathcal{S}}{\delta \rho(x)} [\rho_N(t, \cdot)] \nabla \cdot \left(\rho_N(t, x) \big[- b(x) + J \, \nabla(V * \rho_N)(t, x) \big] \right. \\
&\quad \left. + k_B T \nabla \rho_N(t, x) \right) + \frac{k_B T}{N^2} \sum_{n=1}^{N} \int dx \, dy \, \frac{\delta^2 \mathcal{S}}{\delta \rho(x) \delta \rho(y)} [\rho_N(t, \cdot)] \nabla_x \cdot \nabla_y \\
&\quad \times \left(\delta(x - x_n(t)) \, \delta(y - x_n(t)) \right) \\
&= \int dx \, \frac{\delta \mathcal{S}}{\delta \rho(x)} [\rho_N(t, \cdot)] \nabla \cdot \left(\rho_N(t, x) \big[- b(x) + J \, \nabla(V * \rho_N)(t, x) \big] \right. \\
&\quad \left. + k_B T \nabla \rho_N(t, x) \right) + \frac{k_B T}{N} \int dx \int dy \, \frac{\delta^2 \mathcal{S}}{\delta \rho(x) \delta \rho(y)} [\rho_N(t, \cdot)] \nabla_x \cdot \nabla_y \\
&\quad \times \left(\rho_N(x, t) \, \delta(x - y) \right),
\end{aligned}
\tag{4.14}
$$

where the last expression was obtained proceeding as in (4.7). We thus obtain for $\mathbb{E}\,[\mathcal{S}[\rho_N(t, \cdot)]]$ the evolution equation

$$\frac{d}{dt} \mathbb{E} \, \mathcal{S}[\rho_N(t, \cdot)] = \mathbb{E} \, \mathcal{L} \mathcal{S}[\rho_N(t, \cdot)], \qquad (4.15)$$

where $\mathcal{L}$ is the generator given by

$$
\begin{aligned}
\mathcal{L} \mathcal{S}[\rho] &= \int dx \, \frac{\delta \mathcal{S}[\rho]}{\delta \rho(x)} \nabla \cdot \left(\rho_N(t, x) \big[- b(x) + J \, \nabla(V * \rho_N)(t, x) \big] + k_B T \nabla \rho_N(t, x) \right) \\
&\quad + \frac{k_B T}{N} \int dx \int dy \, \frac{\delta^2 \mathcal{S}[\rho]}{\delta \rho(x) \delta \rho(y)} (\nabla \cdot \nabla_y) \left(\rho(x, t) \, \delta(x - y) \right).
\end{aligned}
\tag{4.16}
$$

---

[2] Since the densities are normalized, functional derivatives $\delta \mathcal{S}/\delta \rho(x)$ are defined only up to a constant, but such ambiguities drop out in all expressions below where the functional derivatives are integrated against functions with vanishing integral.

Observe now that the above evolution for $\mathbb{E}\left[\mathcal{S}[\rho_N]\right]$ is given by the same backward Kolmogorov equation that the one obtained assuming that ρ_N solves the Dean stochastic PDE (4.12) anticipated at the beginning of the subsection. In finite dimension, it is not enough to know that $\mathbb{E}\,f(X(t)) = \mathbb{E}\,Lf(X(t))$ for a generator of a diffusion process L and each f to deduce that $X(t)$ has the law of the diffusion process satisfying the corresponding SDE, but this would follow if one showed that $f(X(t)) - f(X(0)) - \int_0^t Lf(X(s))\,ds$ are martingales [78]. A slight extension of the previous calculation shows that the process

$$\mathcal{S}[\rho_N(t, \cdot)] - \mathcal{S}[\rho_N(0, \cdot)]] - \int_0^t \mathcal{L}\mathcal{S}[\rho_N(s, \cdot)]\,ds \qquad (4.17)$$

possess this property. Nevertheless, since we are in infinite dimension, the derivation of the Dean equation remains formal.

To complicate things further, the mathematical status of the Dean equation is unclear: it is difficult to give sense to different terms of the equation, including the noisy one, in a function space that would contain the (distributional) empirical densities of coupled diffusions (4.1), not even speaking about a theory of solutions of such a stochastic PDE. Although there is a considerable mathematical literature about non-linear stochastic PDE's with the noise that is delta-correlated in time and space, see [17,30,46], it does not cover the case of the Dean equation. One may hope, however, that it is possible to give a meaning to this equation at least in the large deviation regime for large N. Although in what follows we completely avoid such rigorous issues and proceed formally, the fact that our results agree with those rigorously obtained by Dawson and Gartner [18,19] gives support to this assumption.

Before proceeding further, let us observe that the Dean equation can be recast in a form similar to that of the finite-dimensional SDE considered in the first part of the paper by rewriting it as

$$\partial_t \rho = \mathcal{K}[\rho] + \sqrt{\frac{2}{N}}\,\eta[\rho], \qquad (4.18)$$

where

$$\mathcal{K}[\rho](t, x) = -\nabla \cdot j_\rho(t, x) \qquad (4.19)$$

is the drift and

$$\eta[\rho](t, x) = -\sqrt{k_B T}\,\nabla \cdot \left(\sqrt{\rho(x)}\,\xi(t, x)\right) \qquad (4.20)$$

is the white noise in time, parameterized by density ρ, with zero mean and covariance

$$\mathbb{E}\,\eta[\rho](t, x)\,\eta[\rho](s, y) = \delta(t - s)\,\mathcal{Q}[\rho](x, y) \qquad (4.21)$$

for

$$\mathcal{Q}[\rho](x, y) = k_B T\,\nabla_x \cdot \nabla_y \Big(\rho(x)\delta(x - y)\Big). \qquad (4.22)$$

This form will be useful in the following to formally extend the large deviations results that we described in Sect. 2 to the present case.

In the Dean equation, the noise term becomes small when N becomes large. In particular, in the $N \to \infty$ limit, one obtains the deterministic non-linear Fokker–Planck equation

$$\partial_t \rho(t, x) + \nabla \cdot j_\rho(t, x) = 0 \qquad (4.23)$$

known in the mathematical literature as McKean–Vlasov equation [66]. Some authors refer to the above equation as the Vlasov-Fokker–Planck one. In the following, we call the evolution described by this equation the mean-field dynamics.

The above reasoning showed that the McKean–Vlasov equation should describe the evolution of the empirical density in the $N \to \infty$ limit. This is actually corroborated by a number of rigorous results, since the original paper of McKean [66]. More precisely, the property of propagation of chaos was proved under mild hypothesis on the smoothness of the b and V, see [79] and references therein. Moreover, a bound on a proper distance between the solution to the McKean–Vlasov equation and the empirical measure at time t, depending on the distance at $t = 0$ is known. Those results are the analogues in the present context of the more famous ones due to Braun-Hepp [11] and Dobrushin [25] for deterministic particles with mean-field interactions, see also [76].

The Dean equation suggests that the evolution of the empirical measure for finite but large N is described by a weak random perturbation (of order $1/\sqrt{N}$) of the McKean–Vlasov equation. We are then in a similar context to that of the Freidlin–Wentzell theory discussed in Sect. 2, except for the fact that the stochastic dynamical system is now an infinite dimensional one. We may, nevertheless, hope to obtain large deviation estimates in a similar manner by working at the formal level. This will be done in Sect. 4.4. Before, however, let us further investigate the McKean–Vlasov equation by considering a simple model system.

4.3 Shinomoto–Kuramoto Model: The Mean-Field Behavior

Models in the class of Eq. (4.1) can display a very rich mean-field dynamics. As a simple example, we consider in this section the Shinomoto–Kuramoto model introduced in [74], describing its mean-field behavior by focusing on results that can be obtained analytically or semi-analytically.

The Shinomoto–Kuramoto model is a one-dimensional model where N particles move on a circle of unit radius and are thus identified by their angular coordinate $x_n = \theta_n$ defined modulo 2π, for $n = 1, \ldots, N$. The model is obtained from Eq. (4.1) by setting $b(\theta) = F - h \sin(\theta)$ and $V(\theta) = (1 - \cos \theta)$, where F, h, J are real constants that we shall take non-negative. The equations of motion are

$$\dot{\theta}_n = (F - h \sin \theta_n) - \frac{J}{N} \sum_{m=1}^{N} \sin(\theta_n - \theta_m) + \sqrt{2k_B T}\, \eta_n, \qquad (4.24)$$

where $\eta_n(t)$ are independent standard scalar white noises. Observe that negative F is related to positive F and negative h to positive h by the changes of variables $\theta_n \mapsto -\theta_n$ and $\theta_n \mapsto \theta_n + \pi$, respectively. One of the parameters among F, h, J, T is redundant, as it can be fixed by rescaling the other three and time. We could, for example, fix the coupling strength J. We prefer, however, to leave all the parameters because we shall be interested in limiting cases where one of them vanishes. The Shinomoto–Kuramoto system is related to the more famous Kuramoto model [58] of frequency synchronization phenomena in coupled rotators, from which it is obtained by setting all the natural frequencies to the same value F and adding white noises acting on each rotator, see [1] for a review on Kuramoto and related models.

For general value of the parameters, the Shinomoto–Kuramoto stochastic dynamics (4.24) breaks the detailed balance. This is always true except for $F = 0$, where the model defines an equilibrium dynamics and reduces to a kinetic version of the mean-field ferromagnetic XY model, with θ_n describing the angles of planar spins and h the external magnetic field h in the X direction. For $F > 0$, the model may be still interpreted within ferromagnetism,

except that the planar magnetic field should be taken rotating with angular velocity F and the spin angles described in the frame rotating with it. The case $h = 0$ is also special since its dynamics only trivially breaks the detailed balance that may be restored by returning to the original frame. Thus, the Shinomoto–Kuramoto model can be seen as a non-equilibrium version of mean-field ferromagnets.

As discussed in Sect. 4.2, for $N \to \infty$, the evolution of the empirical density is described by the McKean–Vlasov equation (4.23) which, in the present case, reads:

$$\partial_t \rho + \partial_\theta j_\rho = 0, \tag{4.25}$$

$$j_\rho(t, \theta) = \rho(t, \theta) \left(F - h \sin \theta - J \int \sin(\theta - \vartheta) \, \rho(t, \vartheta) \, d\vartheta \right) - k_B T \, \partial_\theta \rho(t, \theta). \tag{4.26}$$

The mean-field behavior of the Shinomoto–Kuramoto model was first studied in [74], focusing on long-time behavior. In this work, the authors expanded in Fourier modes the stationary McKean–Vlasov equation (4.25) and numerically integrated the resulting coupled ordinary differential equations.

For $F = 0$, the system relaxes to a stationary solution that is unique for $h > 0$ and for $h = 0$ and $k_B T \geq J/2$, with ρ flat in θ (unmagnetized state) in the latter case. For $h = 0$ and $k_B T < J/2$ there is a one-parameter family of stationary solutions differing by rotation, with ρ bumped around some value of θ, a well known picture for equilibrium ferromagnets of a magnetized state spontaneously breaking the planar rotation symmetry. This equilibrium case, also known under the name of Brownian mean field model, is studied in detail in [13–15].

For $F > 0$, a more complicated phase diagram for the McKean–Vlasov equation emerges, however, see the left part of Fig. 3. For sufficiently high h or T, the system relaxes to a stationary solution of Eq. (4.25), which for $h = 0$ and $k_B T > J/2$ has ρ flat in θ, as before. For sufficiently low values of h and T, the long-time behavior is, instead, periodic. For $h = 0$, the periodic phase sets in for $k_B T < J/2$, as follows from the relation to the $F = 0$, $h = 0$ system mentioned before. The periodic and stationary regions are separated by bifurcations. The rightmost line (blue dots) is a Hopf bifurcation, while the upper line, a saddle-node one. The two bifurcations meet forming a Takens-Bogdanov bifurcation. The careful analysis of bifurcations occurring in this model has been performed in [73] and is confirmed here. In particular, a tiny region where the system is bistable is found around the Takens-Bogdanov bifurcation, see the right part of Fig. 3. Here the McKean–Vlasov equation admits two stable stationary solutions.

A number of analytical results can be obtained. In particular, we show below that all the stationary solutions (stable and unstable) can be studied analytically. Moreover, analyzing their stability we can trace the bifurcation curves. On the other hand, we were not able to find a closed form for the periodic solutions except for the trivial cases where $h = 0$ or $T = 0$.

In Sect. 4.3.1 below, we consider the case where the dynamics respects the detailed balance ($F = 0$) or trivially breaks it ($h = 0$). Then, in 4.3.2, we list all the stationary solutions for generic values of the parameters and describe how their stability may be analyzed. Finally, we discuss the special case where the particles do not interact ($J = 0$), and the singular situation of zero temperature ($T = 0$) in Secs. 4.3.3 and 4.3.4, respectively.

4.3.1 Equilibrium Dynamics ($F = 0$ or $h = 0$)

We analyze in this paragraph the long-time behavior of the Shinomoto–Kuramoto model for $F = 0$ or $h = 0$. We refer the reader to [14] for a more detailed discussion on the case $F = 0$.
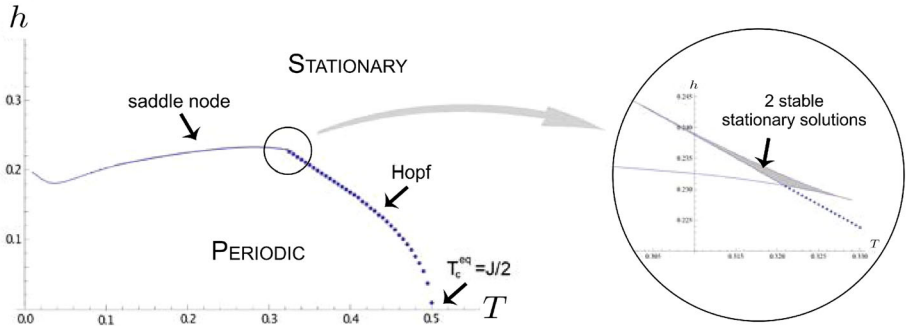
Fig. 3 Phase diagram obtained with our semi-analytical results describing the long-time behavior of the Shinomoto–Kuramoto model for $J = 1$ and $F = 0.2$. We have checked that the phase diagram for other values of the parameters is similar. Our results are in agreement with those first obtained numerically in [73,74]. Qualitatively, the long-time behavior of the Shinomoto–Kuramoto model is very different for high T and/or h and for low T and h. In the first case, the system is stationary at long times and the empirical density converges to the unique stationary stable solution of the McKean–Vlasov equation. This is true except for a very small region where two stationary stable solutions of the McKean–Vlasov equation are present, represented by the shaded region in the *inset*. For low T and h, no stationary stable solutions of the McKean–Vlasov equation exists and the empirical density converges to a periodic solution. The two regions are separated by two lines corresponding to the Hopf (*dots*) and saddle-node (*blue line*) bifurcations. These two lines merge in a Taken–Bogdanov bifurcation. Other stationary but unstable states of the McKean–Vlasov equation exist, and will be fully analyzed in the following (Color figure online)

Let us start with the case $F = 0$, $J \geq 0$ and $T > 0$. Here, the N-body system (4.24) with finite N respects the detailed balance with respect to the invariant measure given by the Boltzmann-Gibbs distribution

$$f_N(\theta_1, \ldots, \theta_N) = \frac{1}{Z} \exp\left[\frac{1}{k_B T} \left(h \sum_{n=1}^N \cos\theta_n - \frac{J}{2N} \sum_{m,n=1}^N (1 - \cos(\theta_n - \theta_m)) \right) \right]$$

(4.27)

where Z is the canonical partition function. The system is ergodic and the mean of the empirical density converges to the one in the Gibbs measure that may be easily calculated. Indeed, applying the the Hubbard-Stratonovich transformation to Eq. (4.27), we get the identity

$$f_N(\theta_1, \ldots, \theta_N) = \frac{N e^{-\frac{NJ}{2k_B T}}}{2\pi J k_B T Z} \int e^{\frac{1}{k_B T} \left((m_x + h) \sum_{n=1}^N \cos\theta_n + m_y \sum_{n=1}^N \sin\theta_n - \frac{N}{2J}(m_x^2 + m_y^2) \right)} dm_x \, dm_y.$$

(4.28)

The expectation of the empirical density $\rho_N(\theta)$ in the Gibbs stationary state is equal to the integral of $f_N(\theta, \theta_2, \ldots, \theta_N)$ over $\theta_2, \ldots, \theta_N$ that gives

$$\mathbb{E} \, \rho_N(x) = \frac{(2\pi)^{N-2} N e^{-\frac{NJ}{2k_B T}}}{J k_B T Z}$$

$$\times \int e^{\frac{1}{k_B T} \left((m_x + h) \cos\theta + m_y \sin\theta - \frac{N}{2J}(m_x^2 + m_y^2) \right)} \left(I_0\left(\frac{m(h)}{k_B T}\right) \right)^{N-1} dm_x \, dm_y,$$

(4.29)

where $m(h) = \sqrt{(m_x + h)^2 + m_y^2}$ and

$$I_0\left(\sqrt{z_1^2 + z_2^2}\right) = \frac{1}{2\pi} \int_0^{2\pi} e^{z_1 \cos\theta + z_2 \sin\theta} \, d\theta \tag{4.30}$$

is the Bessel functions of the first kind. From Eq. (4.29), we can obtain the equilibrium stationary density in the $N \to \infty$ limit by the saddle point calculation:

$$\rho_{inv}^{eq}(\theta) = \lim_{N \to \infty} \mathbb{E} \, \rho_N(\theta) = \frac{e^{\frac{1}{k_B T}\left((m_x + h)\cos\theta + m_y \sin\theta\right)}}{2\pi I_0\left(\frac{m(h)}{k_B T}\right)}, \tag{4.31}$$

where (m_x, m_y) minimizes

$$f(m_x, m_y) = \frac{m_x^2 + m_y^2}{2 J k_B T} - \ln I_0\left(\frac{m(h)}{k_B T}\right). \tag{4.32}$$

The corresponding stationarity equations are

$$\frac{m_x}{J} = \frac{m_x + h}{m(h)} \frac{I_1\left(\frac{m(h)}{k_B T}\right)}{I_0\left(\frac{m(h)}{k_B T}\right)}, \qquad \frac{m_y}{J} = \frac{m_y}{m(h)} \frac{I_1\left(\frac{m(h)}{k_B T}\right)}{I_0\left(\frac{m(h)}{k_B T}\right)} \tag{4.33}$$

with $I_1 = I_0'$. They may be rewritten as the self-consistency equations

$$m_x = J \int_0^{2\pi} \cos\theta \, \rho_{inv}^{eq}(\theta) \, d\theta, \qquad m_y = J \int_0^{2\pi} \sin\theta \, \rho_{inv}^{eq}(\theta) \, d\theta, \tag{4.34}$$

so that (m_x, m_y) has the interpretation of the magnetization vector (in units of J).

For $h > 0$, the stationarity equations (4.33) imply that $m_y = 0$ and $m_x > 0$ solves the equation

$$\frac{m_x}{J} = \frac{I_1\left(\frac{m_x + h}{k_B T}\right)}{I_0\left(\frac{m_x + h}{k_B T}\right)} \tag{4.35}$$

which has a unique solution. In the limit $h \to 0$, one recovers the 2nd-order phase transition (a pitchfork bifurcation) located at $k_B T_c = J/2$. For higher values of T, one has $\rho_{inv}^{eq} = \frac{1}{2\pi}$ and the magnetization vanishes (i.e. $m_x = 0, m_y = 0$) while for $T < T_c$ the stationary state is spontaneously magnetized ($m_x > 0, m_y = 0$). Taking $h = 0$ directly, any rotation of the low temperature solution in θ, with the corresponding rotation of the magnetization vector (m_x, m_y), provides a solution of the saddle point equations which minimizes (4.32), whereas $m_x = 0, m_y = 0$ corresponding to the flat density gives the maximum of (4.32) instead of the minimum. The same limiting densities will be obtained as the stationary solutions of the McKean–Vlasov equation for $F = 0$, see below.

Let us now consider the second case, where $F, J, T > 0$ and $h = 0$. Strictly speaking, the dynamics does not respect here the detailed balance. However, it can be recast as an equilibrium dynamics performing the change of variables $\theta_n' = \theta_n + Ft$, which corresponds to sitting on the comoving frame with angular velocity $-F$. The empirical density in this case can thus be obtained from the equilibrium results. Indicating by ρ_N^{eq} the empirical density for $F = 0$ and given values of T, J and by ρ_N^F the empirical density for the same values of

T and J but with $F \neq 0$, we have $\rho_N^F(t, \theta) = \rho_N^{eq}(t, \theta - Ft)$. We conclude that for $T > T_c$, in the limit of long times and large N, the expectation of $\rho_N^F(t, \theta)$ becomes stationary and flat, while for $T < T_c$ it becomes periodic. The pitchfork bifurcation at equilibrium is thus modified to a Hopf bifurcation.

4.3.2 Stationary States for Generic Parameters

We now consider generic values of the parameter h, T, J, F with $T > 0$. To find stationary densities ρ_{inv} at $N = \infty$, we consider the McKean–Vlasov equation (4.25) in the stationary form, imposing $\partial_t \rho_{inv} = 0$. As the Shinomoto–Kuramoto model is one-dimensional, this is equivalent to ask that the mean-field current j_ρ given by Eq. (4.26) be a constant, that we shall denote c, for $\rho = \rho_{inv}$. We infer that for the stationary densities,

$$j_{\rho_{inv}}(\theta) = \left[F - (m_x + h) \sin\theta + m_y \cos\theta \right] \rho_{inv}(\theta) - k_B T \partial_\theta \rho_{inv}(\theta) = c, \quad (4.36)$$

where m_x and m_y are given by the equations

$$m_x = J \int_0^{2\pi} \cos\theta \, \rho_{inv}(\theta) \, d\theta, \qquad m_y = J \int_0^{2\pi} \sin\theta \, \rho_{inv}(\theta) \, d\theta, \quad (4.37)$$

so that (m_x, m_y) is again the magnetization vector. As the current is constant, $j_{\rho_{inv}}(\theta) = j_{\rho_{inv}}(\theta + \theta_0)$ for any θ_0 implying that

$$
\begin{aligned}
j_{\rho_{inv}}(\theta) = & \left\{ F - \left[(m_x + h) \cos\theta_0 + m_y \sin\theta_0 \right] \sin\theta \right. \\
& \left. + \left[m_y \cos\theta_0 - (m_x + h) \sin\theta_0 \right] \cos\theta \right\} \rho_{inv}(\theta + \theta_0) \\
& - k_B T \, \partial_\theta \rho_{inv}(\theta + \theta_0) = c.
\end{aligned}
\quad (4.38)
$$

We can now choose θ_0 in such a way that

$$m_y \cos\theta_0 - (m_x + h) \sin\theta_0 = 0, \qquad (m_x + h) \cos\theta_0 + m_y \sin\theta_0 = y \geq 0, \quad (4.39)$$

so that Eq. (4.38) becomes

$$j_{\rho_{inv}}(\theta) = (F - y \sin\theta) \rho_{inv}(\theta + \theta_0) - k_B T \, \partial_\theta \rho_{inv}(\theta + \theta_0) = c. \quad (4.40)$$

For any y, the solution to this equation can be written as

$$\rho_{inv}(\theta + \theta_0) = Z^{-1} e^{\frac{1}{k_B T}(F\theta + y \cos\theta)} \int_\theta^{\theta + 2\pi} e^{-\frac{1}{k_B T}(F\vartheta + y \cos\vartheta)} \, d\vartheta \quad (4.41)$$

with Z the normalization factor. Clearly, for Eq. (4.41) to be a stationary solution, the self-consistency conditions (4.37) have to be satisfied together with Eq. (4.39). Such self-consistency condition can be made more explicit by introducing

$$
\begin{aligned}
f_x(y) &\equiv \int_0^{2\pi} \cos\theta \, \rho_{inv}(\theta + \theta_0) \, d\theta = \frac{y}{J} - \frac{h}{J} \cos\theta_0, \\
f_y(y) &\equiv \int_0^{2\pi} \sin\theta \, \rho_{inv}(\theta + \theta_0) \, d\theta = \frac{h}{J} \sin\theta_0.
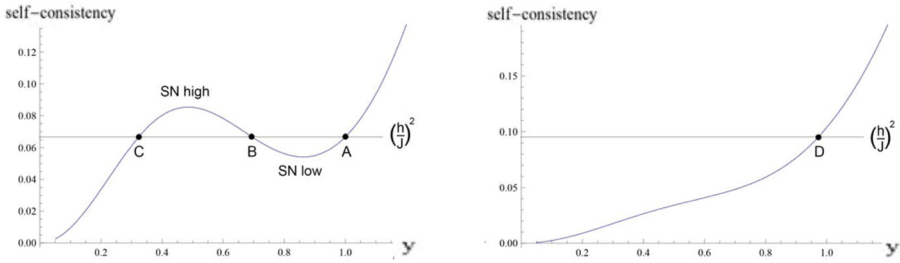\end{aligned}
\quad (4.42)
$$

Fig. 4 We report here the typical curves that are obtained for the *left* hand side of the self-consistency condition (4.43) that does not depend on h. Two cases are observed. If $T > T_{tr}$ (on the *right*), whatever the value of h, only one value of y satisfies the self-consistency condition and the McKean–Vlasov equation admits only one stationary solution. For $T < T_{tr}$ (on the *left*), depending on the value of h, one obtains one or three values of y for which the self-consistency is satisfied. Thus, the McKean–Vlasov equation admits one or three stationary solutions. Moreover, the maximum and the minimum of the curve defines two values of y for which a pair of stationary solutions is created or destroyed corresponding to the values of h where saddle-node bifurcations occur

from which we can eliminate θ_0 to obtain an equation for y

$$\left(f_x(y) - \frac{y}{J}\right)^2 + f_y(y)^2 = \left(\frac{h}{J}\right)^2. \tag{4.43}$$

We have thus showed that, for any $T > 0$, stationary solutions of the McKean–Vlasov equation associated to the Shinomoto–Kuramoto model (4.25) are given, up to rotation by angle θ_0, by Eq. (4.41), where y is chosen so that the self-consistency condition (4.43) is satisfied. At the end, angle θ_0 may be found from Eq. (4.42). The above self-consistency problem can be easily solved numerically. With such procedure, we obtained the curves reported in Fig. 4 for the left hand side of Eq. (4.43) as a function of y. Note that such a function does not depend on h. Two cases are observed. If $T > T_{tr}$, for any value of h only one value of y satisfies the self-consistency condition (4.43). In this case, whatever the value of h, the McKean–Vlasov equation admits only one stationary solution. For $T < T_{tr}$, depending on the value of h, one obtains one or three values of y for which the self-consistency is satisfied. Thus, the McKean–Vlasov equation admits one or three stationary solutions. Observe moreover that the maximum and the minimum of the curve defines two values of y where a pair of stationary solutions is created or destroyed. They thus indicate the values of h where saddle-node bifurcations occur. We leave to the reader checking how the self-consistent solutions reduce to the ones obtained in the previous section for the equilibrium case $F = 0$. Let us just note that in that case, $f_x(y) = I_1(\frac{y}{k_B T})/I_0(\frac{y}{k_B T})$ and $f_y(y) = 0$.

To finally recover the phase diagram, we must also analyze the stability of the stationary solutions just described. For this, it is enough to look at the linearization of the McKean–Vlasov dynamics around a given ρ_{inv},

$$\partial_t \delta\rho = R_{\rho_{inv}} \delta\rho(t, \cdot), \tag{4.44}$$

$$R_{\rho_{inv}} \delta\rho = -\partial_\theta \left[\delta\rho(\theta)\left(F - h\sin\theta - J\int \sin(\theta - \vartheta)\,\rho_{inv}(\vartheta)\,d\vartheta\right) \right. \tag{4.45}$$

$$\left. -\rho_{inv}(\theta)\,J\int \sin(\theta - \vartheta)\,\delta\rho(\vartheta)\,d\vartheta - k_B T\,\partial_\theta \delta\rho(\theta) \right].$$
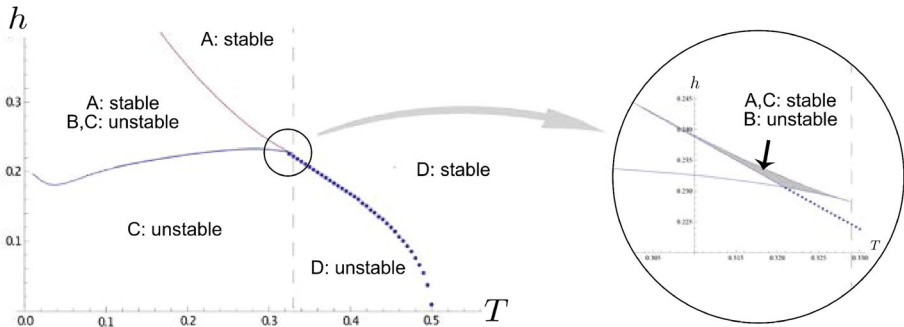
Fig. 5 Stationary states in the mean-field limit of the Shinomoto–Kuramoto model along with their stability. In each region, we report which stationary states are present and whether they are stable or unstable, where the letters A, B, C, D indicate from which self-consistent solution the stationary state is obtained, see Fig. 4

One should then study the spectra of the linearized Fokker–Planck operator $R_{\rho_{inv}}$. This is possible analytically only in very special case when ρ_{inv} does not depend on θ, which holds for $h = 0$.

We checked the stability of the stationary solutions by passing to the Fourier transformed picture and truncating the resulting system to modes $k \leq K$, in which way, $R_{\rho_{inv}}$ is reduced to a $K \times K$ matrix. It is then a simple numerical task to find the eigenvalues of such a matrix. Only few modes ($K \sim 7$) are enough to already obtain very accurate results. The stationary states and their stability are summarized in Fig. 5, from which we reconstruct the phase diagram anticipated in Fig. 3.

We conclude by observing that it is possible to calculate analytically where the shown bifurcation lines end. Indeed, from the equilibrium solution, we know that the Hopf bifurcation crosses the $h = 0$ axis in $T = J/2$. Moreover, it is simple to show that for $T = 0$ the bifurcation occurs at $F = h$.

4.3.3 Free Particles ($J = 0$)

The stationary solutions for the free particle case ($J = 0$) can be obtained very easily directly,

$$\rho_{inv}(\theta) = Z^{-1} e^{\frac{1}{k_B T}(F\theta + h\cos\theta)} \int_{\theta}^{\theta + 2\pi} e^{-\frac{1}{k_B T}(F\vartheta + h\cos\vartheta)} \, d\vartheta, \qquad (4.46)$$

or from Eq. (4.41) since in this case $m_x = m_y = 0 = \theta_0$ and thus $y = h$.

4.3.4 Zero Temperature

There are two kinds of stationary solutions for $T = 0$: delta-function-like and smooth solutions. Smooth solutions are physically less relevant because they are linearly unstable in the ferromagnetic model ($J > 0$) that we consider here. We discuss smooth solutions in Appendix 1 and consider here only the delta-function-like ones.

The stationary McKean–Vlasov equation in the form (4.40) for $T = 0$ is solved by

$$\rho_{inv}(\theta + \theta_0) = \delta(\theta - \theta_1) \qquad (4.47)$$

if θ_1 satisfies $F - y \sin \theta_1 = 0$. Clearly, such solutions exist only for $y > F$. The self-consistency condition (4.42) reduces in this case to the relation

$$\left(\pm \sqrt{1 - \left(\frac{F}{y} \right)^2} - \frac{y}{J} \right)^2 + \left(\frac{F}{y} \right)^2 = \left(\frac{h}{J} \right)^2 \tag{4.48}$$

that has no solutions for y if $h < F$, one solution if $h = F$ and two solutions for $h > F$. They are straightforward to understand. Indeed, one may show that at such solutions, $F - h \sin(\theta_0 + \theta_1) = 0$, so that the delta-function stationary states describe a situation where all particles sit in the zero of their drift $F - h \sin \theta$, which clearly provides a solution of the equations of motion (4.24) when $T = 0$. Such solutions exist for all N if and only if $h \geq F$. For large N, their stability depends on the sign of $h \cos(\theta_0 + \theta_1)$, with the positive one corresponding to stable solutions.

4.4 Large Deviations for Large But Finite N

For large but finite N, the dynamics of the empirical density $\rho_N(t, x)$ deviates from the mean-field evolution given by the McKean–Vlasov equation (4.23). Formally, ρ_N solves the Dean equation (4.12) which is an infinite-dimensional dynamical system perturbed by a weak noise. Below, starting from this equation and employing a functional integral argument, we derive the results obtained rigorously by Dawson and Gartner [18,19] that describe the dynamical large deviations of the empirical density for mean-field diffusions. We also formulate the functional Hamilton–Jacobi equation for the quasi-potential describing the asymptotic form of stationary distribution for the empirical densities. Another functional integral argument gives the large deviations asymptotics for the distribution of dynamical fluctuations of the empirical current. A generalization of the Freidlin–Wentzell theory to infinite dimensional dynamical systems whose deterministic part is given by the Fokker–Planck operator with, possibly, non-linear diffusion and drift coefficients is known in the physics literature as the Macroscopic Fluctuation Theory. It has attracted much attention in last years in non-equilibrium statistical mechanics mainly because of its applications to stochastic lattice gases, see [6] for a review. Our results may be viewed as an application of that theory to diffusions with mean-field interactions of the general form (4.1). It will be also straightforward to specify the results that follow to the case of the Shinomoto–Kuramoto model.

4.4.1 Dynamical Large Deviations for the Empirical Density

Let us fix two times $t_i < t_f$ and consider the probability that the empirical density $\rho_N(t, x)$ is arbitrarily close to a given trajectory $\hat{\rho}(t, x)$ for $t_i \leq t \leq t_f$. In fact we shall be interested only in the large deviations where such probability is described by the rate function $\mathcal{A}[\rho(\cdot, \cdot)]$ as in Eq. (2.7) where trajectories $x(t)$ become those in the infinite-dimensional space of densities equipped with an appropriate norm and the limit $\epsilon \to 0$ is replaced by $N \to \infty$. Informally,

$$\mathcal{A}[\rho(\cdot, \cdot)] = - \lim_{N \to \infty} \frac{1}{N} \ln P[\rho(\cdot, \cdot)], \tag{4.49}$$

where $P[\rho(\cdot, \cdot)]$ is the probability distribution function(al) describing the distribution of empirical densities in the time interval $[t_i, t_f]$. This rate function $\mathcal{A}$ is a functional of the trajectory $\rho(\cdot, \cdot)$ and will provide a generalization of the Freidlin–Wentzell action of Eq. (2.8) to the present infinite-dimensional setup.

We may find the form of $\mathcal{A}[\rho]$ by the Martin-Siggia-Rose functional integral argument [65]. A similar derivation to the one presented below may be found in [53]. Assuming the Dean equation, the distribution function $P[\rho]$ is that of its random solutions and may be written in the form

$$P[\rho] = \mathbb{E}\left[\delta\left(\partial_t\rho + \nabla\cdot j_\rho + \sqrt{\frac{2k_BT}{N}}\,\nabla\cdot(\sqrt{\rho}\,\xi)\right)\det\left(\frac{\delta\left(\partial_t\rho + \nabla\cdot j_\rho + \sqrt{\frac{2k_BT}{N}}\,\nabla\cdot(\sqrt{\rho}\,\xi)\right)}{\delta\rho}\right)\right]$$
(4.50)

if we fix the initial value $\rho(t_i, x) \equiv \rho_i(x)$. The Jacobian factor can be dropped, as it will not contribute to the large deviations. Expressing now the delta functional as an oscillatory functional integral in the imaginary directions over fields $u(t, x)$ such that $\int u(t, x)dx = 0$, we obtain

$$P[\rho] \asymp \mathbb{E}\left[\int \exp\left[\frac{N}{2k_BT}\int_{t_i}^{t_f} dt \int \left(\partial_t\rho + \nabla\cdot j_\rho + \sqrt{\frac{2k_BT}{N}}\,\nabla\cdot(\sqrt{\rho}\,\xi)\right)u\,dx\right]\mathcal{D}u\right]$$
(4.51)

$$= \int \exp\left[\frac{N}{2k_BT}\int_{t_i}^{t_f} dt \int \left((\partial_t\rho + \nabla\cdot j_\rho)\,u + \frac{1}{2}\rho\,(\nabla u)^2\right)dx\right]\mathcal{D}u,$$
(4.52)

where the last equality was obtained by calculating the Gaussian expectation over the white noise $\xi(t, x)$. The rate function $\mathcal{A}[\rho]$ can now be extracted by the saddle point argument:

$$\mathcal{A}[\rho] = -\min_u \frac{1}{2k_BT}\int_{t_i}^{t_f} dt \int \left((\partial_t\rho + \nabla\cdot j_\rho)\,u + \frac{1}{2}\rho\,(\nabla u)^2\right)dx,$$
(4.53)

where the minimum is taken over all functions $u(t, x)$ with $t_i \leq t \leq t_f$ and spatial mean zero. The minimum is easy to calculate resulting in the formula

$$\mathcal{A}[\rho] = \frac{1}{4k_BT}\int_{t_i}^{t_f} dt \int\int (\partial_t\rho + \nabla\cdot j_\rho)\,(t, x)\,(-\nabla\cdot\rho(t, \cdot)\nabla)^{-1}(x, y)$$
$$\times (\partial_t\rho + \nabla\cdot j_\rho)\,(t, y)\,dx\,dy,$$
(4.54)

where $(-\nabla\cdot\rho(t, \cdot)\nabla)^{-1}(x, y)$ is the kernel of the inverse of the operator $-\nabla\cdot\rho(t, \cdot)\nabla$ in the action on functions of zero mean. Eq. (4.53) may be also rewritten in the form

$$\mathcal{A}[\rho] = \min_u \frac{\left(\int_{t_i}^{t_f} dt \int (\partial_t\rho + \nabla\cdot j_\rho)\,u\,dx\right)^2}{4k_BT \int_{t_i}^{t_f} dt \int \rho(\nabla u)^2\,dx}$$
(4.55)

which was rigorously derived for the large deviations rate function $\mathcal{A}[\rho]$ by Dawson and Gartner in [18,19]. Indeed, Eq. (4.55) results from (4.53) upon substituting $u \mapsto \lambda u$ and minimizing over $\lambda \in \mathbb{R}$. On the other hand, expression (4.54) is completely analogous to the Freidlin–Wentzell action (2.8) as it may be rewritten as

$$\mathcal{A}[\rho] = \frac{1}{4}\int_{t_i}^{t_f} dt \int\int (\partial_t\rho - \mathcal{K}[\rho])\,(t, x)\,\mathcal{Q}^{-1}[\rho(t, \cdot)](x, y)$$
$$\times (\partial_t\rho - \mathcal{K}[\rho])\,(t, y)\,dx\,dy,$$
(4.56)

where $\mathcal{K}[\rho]$ is the drift and $\mathcal{Q}[\rho]$ is the noise covariance of the Dean equation that are given by Eqs. (4.19) and (4.22), respectively.

In our arguments, we ignored the fact that the empirical density $\rho_N(t, x)$ defines a genuine stochastic process only after integrating it against a spatial test function. The large deviations for such stochastic processes will, however, be governed by rate functions that are contractions of $\mathcal{A}[\rho]$ obtained by minimizing it with imposed integrals against the test functions.

4.4.2 Quasi-Potential and Hamilton–Jacobi Equation

The functional Fokker–Planck equation associated to the Dean equation (4.12) describes the evolution of the probability distribution function $P_t[\rho]$ to observe a density profile $\rho(x)$ at time t and takes the form

$$\partial_t P = \mathcal{L}^\dagger P, \tag{4.57}$$

where $\mathcal{L}^+$ is the adjoint of the generator (4.16) calculated with the rule $\left(\frac{\delta}{\delta\rho}\right)^\dagger = -\frac{\delta}{\delta\rho}$. Analogously to the case of finite-dimensional systems, we define the quasi-potential as the functional

$$\mathcal{F}[\rho] = -\lim_{N\to\infty} \frac{1}{N} \ln P_\infty[\rho], \tag{4.58}$$

where P_∞ denotes the stationary solution of the above Fokker–Planck equation. Proceeding formally, all the structure of Freidlin–Wentzell theory generalizes to the present case. We are thus very brief in what follows, as the derivation of the statements below is very similar to the one for finite dimensional systems, albeit only formal here.

As in the case of finite dimensional systems, we only consider cases where the attractor of the deterministic evolution $\partial_t \rho = \mathcal{K}[\rho]$ is a fixed point ρ_{inv}. The quasi-potential with respect to ρ_{inv} can be obtained from the minimization problem

$$\mathcal{F}[\rho] = \min_{\{\hat{\rho}(t,x)\,|\,\hat{\rho}(-\infty,x)=\hat{\rho}_{inv}(x),\,\hat{\rho}(0,x)=\rho(x)\}} \mathcal{A}[\hat{\rho}]. \tag{4.59}$$

Alternatively, one can try to solve the Hamilton–Jacobi equation

$$\int \frac{\delta\mathcal{F}}{\delta\rho(x)} \mathcal{K}[\rho](x)\,dx + \int\int \frac{\delta\mathcal{F}}{\delta\rho(x)} \mathcal{Q}[\rho](x, y) \frac{\delta\mathcal{F}}{\delta\rho(y)}\,dx\,dy = 0 \tag{4.60}$$

that can be obtained by inserting Ansatz $P_\infty[\rho] \sim e^{-N\mathcal{F}[\rho]}$ into the functional Fokker–Planck equation (4.57). Equivalently, we can obtain it from the Hamilton–Jacobi equation for finite-dimensional systems (2.24).

Finally, with the same argument as in Sect. 2.3, it is simple to show that the minimization in Eq. (4.59) is achieved over the solution of the fluctuation or instanton dynamics

$$\partial_t \rho = \mathcal{K}_r[\rho], \qquad \mathcal{K}_r[\rho](x) = \mathcal{K}[\rho](x) + 2 \int \mathcal{Q}[\rho](x, y)\,dy\,\frac{\delta\mathcal{F}}{\delta\rho(y)}$$

$$= \mathcal{K}[\rho](x) + 2k_B T\,(-\nabla\rho(x)\,\nabla)\frac{\delta\mathcal{F}}{\delta\rho(x)}. \tag{4.61}$$

4.4.3 Dynamical Large Deviations for the Empirical Current

Empirical current (4.3) is a more general quantity than empirical density (4.2). Indeed, ρ_N and j_N satisfy the continuity equation (4.4). Given the empirical current j_N in a time interval $[t_i, t_f]$ and the initial condition for $\rho_N(t_i, x) = \rho_i(x)$, we can reconstruct the empirical density in the same time interval as $\rho_N(x, t) = \rho_i(x) - \int_{t_i}^t \nabla \cdot j_N(x, s)\,ds$. The converse is

however not true, as two empirical currents that differ by a divergent-free quantity give rise to the same empirical density.

Let us consider the probability distribution function $P[j(\cdot, \cdot)]$ of empirical current $j_N(t, x)$ in the time interval $t_i \leq t \leq t_f$ for fixed initial condition for empirical density. The corresponding large deviations rate function will be denoted by $\mathcal{A}^c$:

$$\mathcal{A}^c[j(\cdot, \cdot)] = -\lim_{N \to \infty} \frac{1}{N} \ln \mathbb{P}[j(\cdot, \cdot)]. \tag{4.62}$$

It may be found using again the Martin-Siggia-Rose functional integral formalism:

$$\mathbb{P}[j] \asymp \mathbb{E}\left[\delta\left(j - j_\rho - \sqrt{\frac{2k_B T}{N}} \rho \xi\right)\right] \tag{4.63}$$

$$= \mathbb{E}\left[\int \exp\left[-\frac{N}{2k_B T} \int_{t_i}^{t_f} dt \int \left(j - j_\rho - \sqrt{\frac{2k_B T}{N}} \rho \xi\right) \cdot A\right] \mathcal{D}A\right] \tag{4.64}$$

$$= \int \exp\left[\frac{N}{2k_B T} \int_{t_i}^{t_f} dt \int \left(-(j - j_\rho) \cdot A + \frac{1}{2} \rho A^2\right) dx\right] \mathcal{D}A, \tag{4.65}$$

where A is vector valued, and evaluating the last functional integral with the saddle point method. This gives:

$$\mathcal{A}^c[j] = \frac{1}{4k_B T} \int_{t_i}^{t_f} dt \int \frac{(j - j_\rho)^2}{\rho} \, dx, \tag{4.66}$$

where $\rho(t, x) = \rho_i(x) - \int_0^t \nabla \cdot j(x, s) \, ds$.

Equation (4.66) is known in the literature as the fundamental formula of Macroscopic Fluctuation Theory. It is often viewed as giving the rate function for the joint probability of observing a trajectory of the empirical density and of the empirical current arbitrarily close to $\rho(x, t)$ and to $j(x, t)$, provided that the latter satisfy the continuity equation, and equal to infinity otherwise. This just expresses the fact that density trajectories are completely determined once we chose the current trajectory and the initial condition for the density.

Again, we ignored the fact that these are the integrals of the empirical currents against test functions, now both over time and space, that make sense as random variables. The large deviations for such random variables will, however, be governed by the rate functions obtained by contractions of $\mathcal{A}^c[j]$ that minimize it with constraints imposed for the space-time integrals of j.

Let us conclude by observing that, to the best of our knowledge, large deviations for currents in diffusions with mean-field interactions have not been discussed in the mathematical literature. We will devote a future publication to the investigation of current fluctuations in the Shinomoto–Kuramoto model.

4.5 Perturbative Calculation of the Quasi-Potential

In this final section, we discuss the perturbative calculation of the quasi-potential for the diffusions with mean-field interaction. In Sect. 3, we developed a perturbative scheme to calculate the quasi-potential for finite dimensional systems. The first objective is then to translate that perturbative scheme to the infinite-dimensional setting described by the Dean equation (4.18), which we do in Sect. 4.5.1.

It is possible to obtain explicit results in two cases: close to the free-particle dynamics $J = 0$, see Sect. 4.5.2, and for the Taylor expansion of the quasi-potential around stationary solutions of the McKean–Vlasov equation, see Sect. 4.5.5. We shall see that in these cases the

perturbative expansion of the quasi-potential reduces to solving partial differential equations instead of functional differential equations.

For the Shinomoto–Kuramoto model, in the case of perturbations close to the free-particle dynamics $J = 0$, we present explicit results at the 1st order in J, see Sect. 4.5.3. Interestingly, the quasi-potential that is a local functional of ρ for the unperturbed $J = 0$ case, is shown to become non-local already at the 1st order in J. The resulting analytical expression is evaluated with a simple numerical scheme and we also compute the rate function for the fluctuations of single-particle observables such as the magnetization. Moreover, we discuss how the numerical scheme can be generalized to higher orders stressing that this does not increase the computational complexity.

We also discuss the Taylor expansion of the quasi-potential around the fixed points of the McKean–Vlasov dynamics. It would be possible to obtain explicit results also in this case. We detail how an algorithm can be designed for this purpose in Appendix 2. We did not, however, perform the corresponding calculations, leaving them to future investigations.

Although the perturbative techniques developed in Sect. 3 in the context of finite-dimensional systems are easily transposable to the infinite-dimensional setup discussed here, we are unable to say anything about the nature of the resulting perturbative expansions (are they convergent? asymptotic?) as, to the best of our knowledges, the techniques used in finite dimensions to settle such questions do not extend in a Banach space setup.

4.5.1 Perturbative Schemes in Infinite Dimensional Setting

We generalize here the perturbative schemes for calculating the quasi-potential, developed in Sect. 3, to the case of infinite dimensional systems at hand. This can be done very easily either by starting again from the Hamilton–Jacobi equation, now in the infinite dimensional version (4.60), or by formally taking the limit of infinite dimensions in the expressions of Sect. 3. Such a formal limit transforms derivatives into functional derivatives and finite-dimensional scalar products into L^2 ones.

Both the expansion centered on the attractor ρ_{inv}^{λ} of the perturbed system (a stable fixed point of the McKean–Vlasov dynamics) and the direct one can be easily obtained. We only consider here the second one, as it will be used in the following. We denote the direct expansion of the quasi-potential as

$$\mathcal{F}^{\lambda}[\rho] = \sum_{n} \lambda^n \mathcal{F}^{(n)}[\rho], \tag{4.67}$$

dropping the hats of Sect. 3.2, and analogously for $\mathcal{K}^{\lambda}[\rho]$ and $\mathcal{Q}^{\lambda}[\rho]$.

With the procedures outlined above, the hierarchy (3.21) and (3.22) is now replaced in the infinite dimensional setting by

$$\int \frac{\delta \mathcal{F}^{(0)}}{\delta \rho(x)} \left[\mathcal{K}^{(0)}[\rho](x) + \int \mathcal{Q}^{(0)}[\rho](x, y) \frac{\delta \mathcal{F}^{(0)}}{\delta \rho(y)} \, dy \right] dx = 0 \qquad \text{for} \quad n = 0, \tag{4.68}$$

$$\int \frac{\delta \mathcal{F}^{(n)}}{\delta \rho(x)} \mathcal{K}_r^{(0)}[\rho](x) \, dx = \mathcal{S}^{(n)}[\mathcal{F}^{(0)}, \ldots, \mathcal{F}^{(n-1)}] \qquad \text{for} \quad n \neq 0, \tag{4.69}$$

where $\mathcal{K}_r^{(0)}$ is the drift of the fluctuation dynamics for the unperturbed problem:

$$\mathcal{K}_r^{(0)}[\rho] = \mathcal{K}^{(0)}[\rho] + 2\mathcal{Q}^{(0)}[\rho] \frac{\delta \mathcal{F}^{(0)}}{\delta \rho} \tag{4.70}$$

and $\mathcal{S}^{(n)}$ is given by

$$
\begin{aligned}
\mathcal{S}^{(n)}[\mathcal{F}^{(0)}, \ldots, \mathcal{F}^{(n-1)}] = & - \sum_{k=1}^{n-1} \left[\int \frac{\delta \mathcal{F}^{(n-k)}}{\delta \rho(x)} \left[\mathcal{K}^{(k)}[\rho](x) + \int \mathcal{Q}^{(0)}[\rho](x, y) \frac{\delta \mathcal{F}^{(k)}}{\delta \rho(y)} \, dy \right] dx \right. \\
& + \left. \sum_{l=0}^{n-k} \int \int \frac{\delta \mathcal{F}^{(n-k-l)}}{\delta \rho(x)} \mathcal{Q}^{(k)}[\rho](x, y) \frac{\delta \mathcal{F}^{(l)}}{\delta \rho(y)} \, dx \, dy \right] \\
& - \int \frac{\delta \mathcal{F}^{(0)}}{\delta \rho(x)} \left[\mathcal{K}^{(n)}[\rho](x) + \int \mathcal{Q}^{(n)}[\rho](x, y) \frac{\delta \mathcal{F}^{(0)}}{\delta \rho(y)} \, dy \right] dx.
\end{aligned}
\tag{4.71}
$$

As it will be clear, there is a full analogy with finite-dimensional systems and we refer the reader to Sect. 3 for the discussion on the perturbative schemes. In particular, we recall that the solution of these equations obtained with the methods of characteristics is unique and an explicit expression may be written in terms of the solution for the 0th-order fluctuation dynamics. In the actual infinite dimensional setting, this solution is given by

$$
\mathcal{F}^{(n)}[\rho] = C^{(n)} + \int_{-\infty}^{0} \mathcal{S}^{(n)}[\mathcal{F}^{(0)}, \ldots, \mathcal{F}^{(n-1)}][\widetilde{\rho}(t, \cdot)] \, dt,
\tag{4.72}
$$

where $\widetilde{\rho}(t, x)$ is the solution of the fluctuation dynamics

$$
\partial_t \rho = \mathcal{K}_r^{(0)}[\rho]
\tag{4.73}
$$

starting at $t = -\infty$ at the stable fixed point $\rho_{inv}^0 \equiv \rho_{inv}$ of the unperturbed McKean–Vlasov dynamics and satisfying the final condition $\bar{\rho}(0, x) = \rho(x)$, compare to Eq. (3.27). It should be clear from the discussion of Sect. 3 that the convergence of the above integral in this infinite-dimensional setting is assured once the unperturbed fluctuation dynamics (4.73) linearized around ρ_{inv} escapes fast enough from ρ_{inv}. This happens, for example, in the typical case in which its generator has a spectral gap. We finally recall from Sect. 3.2 that the constants $C^{(n)}$ may be iteratively fixed but do not enter the expression for $\mathcal{S}^{(n)}[\mathcal{F}^{(0)}, \ldots, \mathcal{F}^{(n-1)}]$ so are largely irrelevant and may be adjusted conveniently.

4.5.2 Expansion Around the Free Particles Dynamics ($J = 0$)

We now consider the perturbative expansion of the quasi-potential $\mathcal{F}$ around the independent particle dynamics, choosing the perturbative parameter $\lambda = J$. In this case,

$$
\mathcal{K}^J[\rho](x) = -\nabla \cdot \left[\rho(x) \left(b(x) - J \int (\nabla V)(x - y) \rho(y) \, dy \right) - k_B T \, \nabla \rho(x) \right]
\tag{4.74}
$$

and

$$
\mathcal{Q}^J[\rho](x, y) = k_B T \, \nabla_x \cdot \nabla_y \big(\rho(x) \, \delta(x - y) \big).
\tag{4.75}
$$

For $J = 0$, particles do not interact. Even if for $J = 0$ the system breaks the detailed balance, the quasi-potential $\mathcal{F}^{(0)}$ is known here as its explicit form follows from Sanov's theorem:

$$
\mathcal{F}^{(0)}[\rho] = \int \left[\rho(x) \ln \frac{\rho(x)}{\rho_{inv}(x)} - \rho(x) + \rho_{inv}(x) \right] dx,
\tag{4.76}
$$

where ρ_{inv} is the stable stationary solution of McKean–Vlasov's equation which reduces here to the linear Fokker–Planck equation for a single-particle diffusion. Hence ρ_{inv} exists

and is unique under simple assumptions about the single-particle drift $b(x)$. In the case of the Shinomoto–Kuramoto model, we have explicitly calculated ρ_{inv} for $J = 0$, see Eq. (4.46). We observe that, despite the fact that the dynamics breaks in general the detailed balance, the quasi-potential for $J = 0$ is a local functional of ρ.

As anticipated, it is convenient to perform the direct perturbative expansion, for which we have

$$\mathcal{K}^{(0)}[\rho](x) = \nabla \cdot [-\rho(x)\, b(x) + k_B T \,\nabla \rho(x)], \tag{4.77}$$

$$\mathcal{K}^{(1)}[\rho](x) = \nabla \cdot \left[\rho(x) \int (\nabla V)(x - y)\, \rho(y)\, dy\right], \tag{4.78}$$

$$\mathcal{K}^{(n)}[\rho](x) = 0 \qquad \text{for} \qquad n \geq 2 \tag{4.79}$$

and

$$\mathcal{Q}^{(0)}[\rho](x, y) = k_B T \,\nabla_x \cdot \nabla_y \big(\rho(x)\, \delta(x - y)\big), \tag{4.80}$$

$$\mathcal{Q}^{(n)}[\rho](x, y) = 0 \qquad \text{for} \qquad n \geq 1. \tag{4.81}$$

It should be clear now why, from a practical point of view, it is simpler to deal here with the direct expansion than with the expansion centered on the attractor of the perturbed dynamics. Indeed, in the former case, almost all $\mathcal{K}^{(n)}$ and $\mathcal{Q}^{(n)}$ are zero. This would not be true in the attractor-centered expansion because the attractor of the perturbed dynamics depends on J.

The quasi-potential at the 0th order $\mathcal{F}^{(0)}$ is given by Eq. (4.76). We thus have

$$\frac{\delta \mathcal{F}^{(0)}}{\delta \rho(x)} = \ln \frac{\rho(x)}{\rho_{inv}(x)}. \tag{4.82}$$

The 0th order fluctuation dynamics takes the form $\partial_t \rho = \mathcal{K}_r^{(0)}[\rho]$, where

$$\mathcal{K}_r^{(0)}[\rho] = -\nabla \cdot (\rho\, b) + 2k_B T \,\nabla \cdot [\rho \,\nabla \ln \rho_{inv}] - k_B T \,\nabla^2 \rho \equiv K_r^{(0)} \rho. \tag{4.83}$$

Note that $K_r^{(0)}$ is a linear operator. When detailed balance is respected (i.e. $b = -\nabla U$ for some potential U), one can easily check that

$$\mathcal{K}_r^{(0)}[\rho] = -\nabla \cdot (\rho \nabla U) - k_B T \,\nabla^2 \rho = -\mathcal{K}^{(0)}[\rho], \tag{4.84}$$

where we have used that $\rho_{inv} = (1/Z) \exp(-U/k_B T)$ in this case. The above relation just reflects the fact that the fluctuation dynamics is the time-reversal of the relaxation dynamics for equilibrium problems.

Let us specify the hierarchy (4.69) to the present situation. The left hand side is clearly unchanged, except that $\mathcal{K}_r^{(0)}[\rho]$ takes the particular form (4.84), while the right hand side becomes equal to

$$\mathcal{S}^{(1)}[\mathcal{F}^{(0)}] = -\int \frac{\delta \mathcal{F}^{(0)}}{\delta \rho(x)}\, \mathcal{K}^{(1)}[\rho](x)\, dx \tag{4.85}$$

for $n = 1$, and to

$$\mathcal{S}^{(n)}[\mathcal{F}^{(0)}, \ldots, \mathcal{F}^{(n-1)}] = -\int \frac{\delta \mathcal{F}^{(n-1)}}{\delta \rho(x)}\, \mathcal{K}^{(1)}[\rho](x)\, dx$$

$$- \sum_{k=1}^{n-1} \int \int \frac{\delta \mathcal{F}^{(n-k)}}{\delta \rho(x)}\, \mathcal{Q}^{(0)}[\rho](x, y)\, \frac{\delta \mathcal{F}^{(k)}}{\delta \rho(y)}\, dx\, dy \tag{4.86}$$

for $n > 1$. A very important point is that, in this case, the solutions $\mathcal{F}^{(n)}[\rho]$ of the hierarchy are for $n \geq 1$, homogeneous polynomials of degree $(n + 1)$ (up to constants),

$$\mathcal{F}^{(n)}[\rho] = C^{(n)} + \frac{1}{(n + 1)!} \int \phi^{(n)}(x_0, \ldots, x_n)\, \rho(x_0) \cdots \rho(x_n)\, dx_0 \cdots dx_n, \quad (4.87)$$

where the kernels $\phi^{(n)}$ are symmetric in $(n + 1)$ variables and the choice of constants $C^{(n)}$ is essentially irrelevant. Using Eq. (4.87), we infer that the left hand side of the hierarchy (4.69) can be written as

$$\begin{aligned}
\int \frac{\delta \mathcal{F}^{(n)}}{\delta \rho(x)} &\, \mathcal{K}_r^{(0)}[\rho](x)\, dx \\
&= \frac{1}{n!} \int \left(K_{r,x_0}^{(0)\dagger}\, \phi^{(n)}(x_0, \ldots, x_n) \right) \rho(x_0) \ldots \rho(x_n)\, dx_0 \ldots dx_n \\
&= \frac{1}{(n + 1)!} \int \left(K_{r,n+1}^{(0)\dagger}\phi^{(n)})(x_0, \ldots, x_n) \right) \rho(x_0) \ldots \rho(x_n)\, dx_0 \ldots dx_n. \quad (4.88)
\end{aligned}$$

In the above expressions, $K_{r,x_0}^{(0)\dagger}$ denotes, in a slightly abusive notation, the adjoint operator to $K_r^{(0)}$,

$$K_r^{(0)\dagger} = \left(b - 2k_B T\, (\nabla \ln \rho_{inv}) \right) \cdot \nabla_x - k_B T \nabla_x^2, \quad (4.89)$$

see Eq. (4.83), in the action on a function of variable x_0. Similarly, $K_{r,n+1}^{(0)\dagger}$ denotes the operator

$$K_{r,n+1}^{(0)\dagger} = \sum_{m=0}^{n} K_{r,x_m}^{(0)\dagger} \quad (4.90)$$

acting on functions of n-tuples $(x_0, \ldots, x_n)$. Note that the differential operator $K_r^{(0)\dagger}(x)$ is the generator of a diffusion process with drift $-b(x) + 2k_B T (\nabla \ln \rho_{inv})$ which is the time-reversal of the original single particle diffusion. This operator has a one-dimensional kernel composed of constants and a one-dimensional cokernel composed of functions proportional to ρ_{inv}. For the same reasons, the operator $K_{r,n+1}^{(0)\dagger}$ has also a one dimensional kernel composed of constants and a one dimensional cokernel composed of functions proportional to $\rho_{inv}(x_0) \cdots \rho_{inv}(x_n)$. We shall use those properties below.

In order to compute the right hand side of the hierarchy, we shall treat separately the cases with $n = 1$ and with $n > 1$. For $n = 1$, we have

$$\begin{aligned}
\mathcal{S}^{(1)}[\mathcal{F}^{(0)}] = -\int \Big[&(\nabla^2 V)(x - y) + \tfrac{1}{2}(\nabla V)(x - y) \cdot \left(\nabla \ln \rho_{inv}(x) - \nabla \ln \rho_{inv}(y) \right) \Big] \\
&\times \rho(x)\, \rho(y)\, dx\, dy, \quad (4.91)
\end{aligned}$$

where we have made the integrand symmetric under the exchange of x and y. Comparing Eqs. (4.88) and (4.91), we obtain a differential equation for $\phi^{(1)}$,

$$\left(K_{r,2}^{(0)\dagger}\phi^{(1)} \right)(x, y) = -2(\nabla^2 V)(x - y) - (\nabla V)(x - y) \cdot \left(\nabla \ln \rho_{inv}(x) - \nabla \ln \rho_{inv}(y) \right). \quad (4.92)$$

For $n > 1$, the right hand side of the hierarchy can be written in terms of the kernels $\phi^{(k)}(x_0, \ldots, x_k)$ as

$$
\begin{aligned}
&\mathcal{S}^{(n)}[\mathcal{F}^{(1)}, \ldots, \mathcal{F}^{(n-1)}] \\
&= \frac{1}{n!} \int (\nabla_{x_0} V)(x_0 - x_n) \cdot (\nabla_{x_0}\phi^{(n-1)})(x_0, \ldots, x_{n-1})\, \rho(x_0) \cdots \rho(x_n)\, dx_0 \cdots dx_n \\
&\quad - \sum_{k=1}^{n-1} \frac{k_B T}{k!(n-k)!} \int (\nabla_{x_k}\phi^{(k)})(x_0, \ldots, x_k) \cdot (\nabla_{x_k}\phi^{(n-k)})(x_k, \ldots, x_n) \\
&\qquad \times \rho(x_0) \cdots \rho(x_n)\, dx_0 \cdots dx_n.
\end{aligned}
\tag{4.93}
$$

Then, for any $n > 1$, we obtain the differential equations

$$
\begin{aligned}
&\left(K_{r,n+1}^{(0)\dagger}\phi^{(n)}\right)(x_0, \ldots, x_n) \\
&= (n+1)!\, \mathrm{Sym}\Bigg[\frac{1}{n!} (\nabla_{x_0} V)(x_0 - x_n) \cdot \nabla_{x_0}\phi^{(n-1)}(x_0, \ldots, x_{n-1}) \\
&\qquad - k_B T \sum_{k=1}^{n-1} \frac{1}{k!(n-k)!} (\nabla_{x_k}\phi^{(k)})(x_0, \ldots, x_k) \cdot (\nabla_{x_k}\phi^{(n-k)})(x_k, \ldots, x_n) \Bigg],
\end{aligned}
\tag{4.94}
$$

where $\mathrm{Sym}\,[\cdot]$ stands for the symmetrization of the argument with respect to variables $x_0, \ldots, x_n$. Eqs. (4.92) and (4.94) may be solved iteratively for the kernels $\phi^{(n)}(x_0, \ldots, x_n)$ provided that their right hand sides are orthogonal to the one-dimensional cokernels of operators $K_{r,n+1}^{(0)\dagger}$ spanned by the product functions $\rho_{inv}(x_0) \cdots \rho_{inv}(x_n)$. These are the solvability conditions $\mathcal{S}^{(n)}[\rho_{inv}] = 0$ that were discussed in Sect. 3.2 in the finite-dimensional context, see Eq. (3.24). Such conditions are easy to check in the lower orders but are difficult to prove order by order. They are, however, implied by the existence of the perturbative expansion centered at the attractors, as discussed in Sect. 3.2. Observe finally that $\phi^{(n)}$ are defined by the above equations only up to constants that are in the kernel of $K_{r,n+1}^{(0)\dagger}$. This ambiguity just leads to shifts in constants $C^{(n)}$ appearing in Eq. (4.87) which are not relevant (remember that densities ρ are normalized).

We have greatly simplified the problem of calculating the quasi-potential perturbatively around the free particles dynamics. Instead of solving the functional differential equations (4.69) we have to solve the partial differential equations for the kernels $\phi^{(n)}$ defined by Eq. (4.87). These differential equations are (4.92) for the 1st order contribution and (4.94) for the higher orders. Solving them is a rather simple numerical problem. In the next section, we discuss the results obtained in the case of the Shinomoto–Kuramoto model.

4.5.3 Expansion Around $J = 0$: Results for the Shinomoto–Kuramoto Model

Let us now specify the discussion to the case of the Shinomoto–Kuramoto model introduced in Sect. 4.3 and describe some explicit results. In this case, Eq. (4.92) for the $\phi^{(1)}$ reduces to the relation

$$
\left(K_{r,2}^{(0)\dagger}\phi^{(1)}\right)(\theta, \vartheta) = S^{(1)}(\theta, \vartheta),
\tag{4.95}
$$

where

$$
S^{(1)}(\theta, \vartheta) = -2\cos(\theta - \vartheta) - \sin(\theta - \vartheta) \cdot [\partial_\theta \ln \rho_{inv}(\theta) - \partial_\vartheta \ln \rho_{inv}(\vartheta)],
\tag{4.96}
$$

and

$$K_r^{(0)\dagger} = [F - h \sin\theta - 2k_B T \, (\partial_\theta \ln \rho_{inv})] \, \partial_\theta - k_B T \partial_\theta^2. \tag{4.97}$$

We observe that identity (4.95) is a Lyapunov equation and several techniques can be employed to solve it, see for example [9,56,69]. The one we have chosen is to perform an expansion on the eigenfunctions of operators $K_r^{(0)}$ and $K_r^{(0)\dagger}$.

Let us first set the notations. For $k = 1, \ldots,$ we denote by v_k the eigenfunctions of $K_r^{(0)}$ and by u_k those of $K_r^{(0)\dagger}$. As $K_r^{(0)}$ is not self-adjoint with respect to the L^2 scalar product, these eigenfunctions are not connected by complex conjugation but we may assume that the corresponding eigenvalues are:

$$K_r^{(0)\dagger} u_k = \alpha_k u_k \qquad \text{and} \qquad \mathcal{K}_r^{(0)} v_k = \overline{\alpha}_k v_k \tag{4.98}$$

where $\overline{\alpha}_k$ denotes the complex conjugate of α_k. $K_r^{(0)}$ is a Fokker–Planck operator. Zero is its simple eigenvalue and ρ_{inv} is the corresponding eigenfunction. We choose it to be equal to v_1. Other eigenvalues of $K_r^{(0)}$ have strictly positive real parts. Similarly, we choose $u_1 = 1$. It corresponds to the unique zero mode of $K_r^{(0)\dagger}$. We can also assume that the eigenfunctions are mutually orthonormal[3] in the following sense

$$\int \overline{u_k(\theta)} \, v_l(\theta) \, d\theta = \delta_{kl}. \tag{4.99}$$

It is then simple to show that

$$\phi^{(1)}(\theta, \vartheta) = \sum_{(k,l)\neq(1,1)} \frac{S_{kl}^{(1)}}{\alpha_k + \alpha_l} \, u_k(\theta) \, u_l(\vartheta), \tag{4.100}$$

where

$$S_{kl}^{(1)} = \int \overline{v_k(\theta)} \, \overline{v_l(\vartheta)} \, S^{(1)}(\theta, \vartheta) \, d\theta \, d\vartheta \tag{4.101}$$

(note that $\alpha_k + \alpha_l \neq 0$ for $(k, l) \neq (1, 1)$). One has $S_{11} = 0$ and we have chosen $\phi^{(1)}$ to be orthogonal to the kernel of $K_r^{(0)}$ composed of $v_1 = \rho_{inv}$. The calculation of constant $C^{(1)}$ of Eq. (4.87) will be discussed afterwards.

Equation (4.100) permits to obtain numerically the quasi-potential at the 1st order in J. One first looks for an approximation of the eigenfunctions and eigenvalues of $K_r^{(0)}$ and of $K_r^{(0)\dagger}$ by expanding Eq. (4.98) in the Fourier modes and truncating the hierarchy in order to deal with matrices. The calculation of $\phi^{(1)}$ through Eq. (4.100) is then easily performed. We checked that very few Fourier modes have to be retained ($\gtrsim 20$) to obtain an excellent approximation of eigenfunctions and eigenvalues. Even less eigenfunctions ($\gtrsim 5$) need to be employed in the calculation in order to obtain quite accurate results.

Although the scheme proposed here is a bit more involved than other methods, such as a direct Fourier expansion of Eq. (4.95), it is much more powerful than the latter. Indeed, in the method proposed here, we only need to diagonalize the matrices that approximate operators $K_r^{(0)}$ and $K_r^{(0)\dagger}$ that act on the space of functions of one periodic variable. Instead, attempting to directly solve Eq. (4.95) by expanding on Fourier modes results in diagonalizing the operator $K_{r,2}^{(0)\dagger}$ acting on the space of functions of two variables. Even more important is the

---

[3] This is possible whenever the eigenvalues α_k are all distinct. We have checked numerically that this is indeed the case.
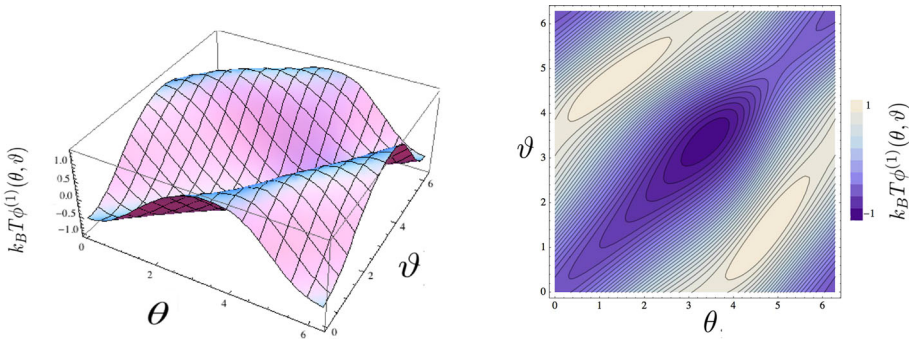
Fig. 6 On the *left* the kernel $k_B T \phi^{(1)}(\theta, \vartheta)$ that permits to obtain the first-order correction to the quasi-potential of the Shinomoto–Kuramoto model in the perturbative expansion around the free particle dynamics with $J = 0$. The values of parameters are $F = 0.2, T = 0.3, h = 0.5$. This figure was obtained by numerically solving Eq. (4.95) as detailed in the text. On the *right* the same function but in a contour plot

fact that the method proposed here is fully generalizable, without increasing the computational complexity to obtain higher perturbative orders (encoded in kernels $\phi^{(n)}$) for the quasi-potential. In contrast, in the direct Fourier expansion, one needs to diagonalize matrices of dimension $(k_{max})^n$ where k_{max} is the number of retained Fourier modes, making the problem practically intractable already for small n.

We report the results for the 1st order correction $\phi^{(1)}$ to the quasi-potential in Fig. 6 for some typical choice of parameters for which the McKean–Vlasov equation has a single stationary stable solution (namely $F = 0.2, T = 0.3, h = 0.5$). A rather non-trivial result emerges. As a check of the accuracy of the perturbative expansion, we also compared the exact stationary state of the system ρ^J_{inv} for $J > 0$ to the one, $\rho_{inv,1}$, obtained by imposing that the functional derivative of

$$\mathcal{F}^{(0)} + J\mathcal{F}^{(1)} + \gamma \left(\int \rho(\theta) \, d\theta - 1 \right) \tag{4.102}$$

vanishes. This condition implies, with the use of (4.82) and (4.87), that

$$\rho_{inv,1}(\theta) = \frac{1}{Z} \rho_{inv}(\theta) \exp\left[-J \int \rho_{inv,1}(\vartheta) \, \phi^{(1)}(\theta, \vartheta) \, d\vartheta \right], \tag{4.103}$$

where $Z = e^\gamma$ fixes the normalization of $\rho_{inv,1}$. Recall that the true mean-field stationary density ρ^J_{inv} minimizes the quasi-potential $\mathcal{F}^J[\rho]$ with the value at the minimum equal to zero. Equation (4.103) may be very easily solved numerically with an iterative scheme, once $\phi^{(1)}$ is known; we have verified that this procedure converges after few iterations. In Fig. 7, we also compare $\rho_{inv,1}$ with the exact result obtained from Eq. (4.41), which shows that already at the 1st order the stationary state for the system with $J > 0$ is much better approximated by $\rho_{inv,1}$ than by the stationary state at $J = 0$. As for the constant $C^{(1)}$ from Eq. (4.87) with $n = 1$, it will be convenient to fix it so that the quasi-potential calculated to the 1st order $\mathcal{F}^{(0)} + J\mathcal{F}^{(1)}$ has the minimal value equal to zero, as does the complete quasi-potential. This is achieved by setting

$$C^{(1)} = -J^{-1}\mathcal{F}^{(0)}[\rho_{inv,1}] - \mathcal{F}^{(1)}[\rho_{inv,1}], \tag{4.104}$$

where on the right hand side $\mathcal{F}^{(1)}$ is taken homogeneous in ρ (i.e. with $C^{(1)}$ set to zero).

Once the quasi-potential at order J is known, we can also compute the large deviations rate functions for the one-particle observables $\frac{1}{N} \sum_{n=1}^{N} g(\theta_n)$ where g is a given function.
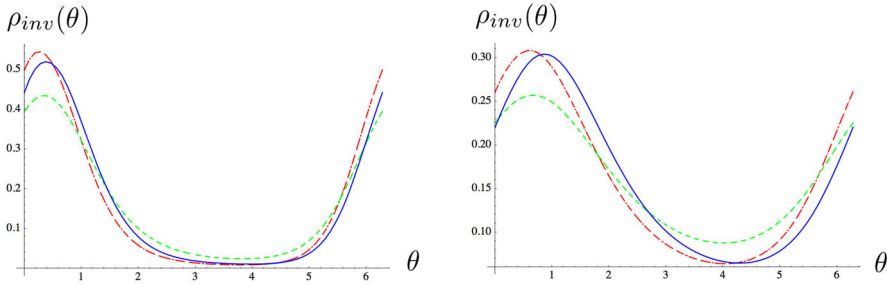
Fig. 7 Stationary state of the Shinomoto–Kuramoto model for $J = 0.3$ obtained from the perturbative expansion of the quasi-potential (*blue continuous curve*) compared to the exact result (*red dashed dotted curve*) given by the Eq. (4.41) and to the stationary state for the system with $J = 0$ (*green dashed curve*). The values of parameters are $F = 0.2, T = 0.3, h = 0.5$ on the *left* and $F = 0.2, T = 0.2, h = 0.15$ on the *right*. The plots show that the 1st order correction to the stationary state strongly improves the prediction (Color figure online)

In the following, we consider the particularly relevant examples of the magnetization along x and y axis corresponding, respectively, to $g = \cos$ and $g = \sin$. What follows may be straightforwardly generalized to any choice of g. We denote by $\mathcal{I}_x(\sigma)$ (resp. $\mathcal{I}_y(\sigma)$) the rate function for the probability of observing a value of magnetization $m_x \sim \sigma$ (resp. $m_y \sim \sigma$). Let us consider $\mathcal{I}_x$ (the case of $\mathcal{I}_y$ may be analyzed similarly). To the 1st order in J, we have

$$\mathcal{I}_x(\sigma) = \min_{\rho} \left\{ \mathcal{F}^{(0)}[\rho] + J\mathcal{F}^{(1)}[\rho] \right\}, \tag{4.105}$$

where the minimum is taken over all positive functions ρ with unit integral respecting the constraint $\int \cos(\theta) \, \rho(\theta) \, d\theta = \sigma$. The profile realizing the minimum will be called ρ_{opt} in the following.

The minimization problem in Eq. (4.105) can be solved introducing two Lagrange multipliers γ_1 and γ_2 associated, respectively, to the total mass and to the magnetization constraints. A simple calculation shows that the minimizing profile satisfies the relation

$$\rho_{opt}(\theta) = \frac{1}{Z} \rho_{inv}(\theta) \exp\left[-\gamma_2 \cos\theta - J \int \phi^{(1)}(\theta, \vartheta) \, \rho_{opt}(\vartheta) \, d\vartheta \right], \tag{4.106}$$

where Z normalizes ρ_{opt}. As for the case of the most probable state, the above self-consistent equation can be solved iteratively. We first fix γ_2 and iteratively solve Eq. (4.106) normalizing the iterated solution to unity at each step. Once the optimal profile has been obtained, we then calculate the corresponding value of m_x. As the initial condition we used the optimal profile for $J = 0$. We have checked that the iterative scheme converges in few steps. The results obtained for $\mathcal{I}_x$ and $\mathcal{I}_y$ are plotted in Fig. 8 for the same choice of parameters as those employed in Fig. 7. In the plots, the results obtained at the 1st order in J are compared with those obtained for $J = 0$ in the whole range of possible magnetizations. We observe that the 1st order corrections are significant. Fox example, the probability of fluctuations of m_x smaller than the typical value is diminished with respect to the $J = 0$ case but the probability of very rare fluctuations is instead increased.

We conclude by noting that, as already discussed, it would be simple to extend the calculations of the quasi-potential and of rate functions of observables to higher orders by computing with the method employed in this section the higher kernels $\phi^{(n)}$. We do not pursue, however, this direction here, as our scope was rather to show that the perturbative approach can lead to explicit results than to analyze in details the Shinomoto–Kuramoto model.
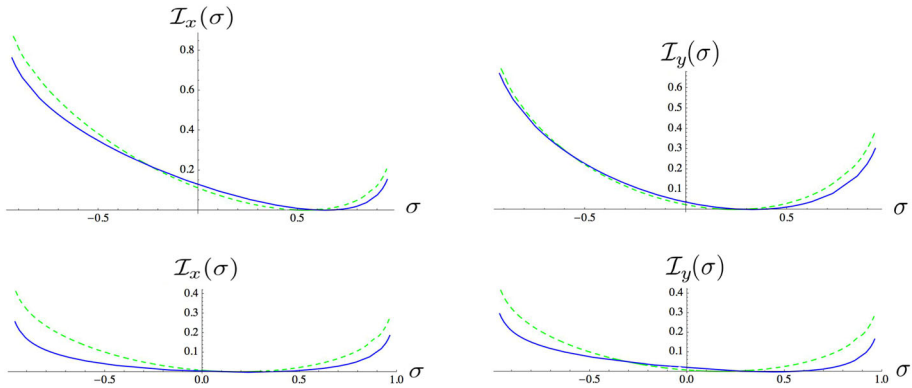
Fig. 8 Rate functions for the probability of fluctuations of the x (on the *left*) and y (on the *right*) component of magnetization. The results of our perturbative analysis to the 1st order in J (*blue continuous lines*) are compared with those of the unperturbed model with $J = 0$ (*green dashed lines*). The value of parameters used are the same as in Fig. 6 in the *upper* plots while they are set to $F = 0.2$, $T = 0.2$, $h = 0.15$ in the *lower* plots (Color figure online)

4.5.4 Expansion Around the Equilibrium Dynamics ($F = 0$ or $h = 0$)

When the drift $b(x)$ may be written as $b = -\nabla U$ for some U, the quasi-potential can be given explicitly because detailed balance holds. We note that the Dean equation (4.18) corresponding to this case has also been derived in [13,15], where it is called stochastic Smoluchowski equation. One has in this case

$$\mathcal{F}[\rho] = \frac{1}{k_B T} \int \rho(x) \left(U(x) + \frac{J}{2} \int V(x - y)\,\rho(y)\,dy \right) dx \; + \; \int \rho(x) \ln \rho(x)\, dx. \tag{4.107}$$

We did not develop the perturbative expansion close to the detailed balance because obtaining explicit results would be quite complex. Indeed, the equations for the quasi-potential at order n are functional differential equations and not, as in the case of a perturbation close to $J = 0$, ordinary differential equations. Even though results may be explicitly obtained by a discretization, the problem is not computationally straightforward.

4.5.5 Taylor Expansion Close to a Stationary State

Let us discuss now the Taylor expansion of the quasi-potential $\mathcal{F}$ around a stable stationary solution ρ_{inv} of the McKean–Vlasov equation. As we have seen in Sect. 3.3 for finite dimensional systems, this expansion may be viewed as a particular case of the perturbative expansion of the quasi-potential for a parameter-dependent system. In this setting, the role of the unperturbed quasi-potential is played by the quadratic approximation to $\mathcal{F}$ that, consistently, will be denoted in the following by $\mathcal{F}^{(0)}$. Then, once $\mathcal{F}^{(0)}$ is known, any higher order correction may, in principle, be calculated. After extending here the discussion of Sect. 3.3 to a general diffusion with mean-field interaction (4.1), we discuss in Appendix 2 how explicit results could be obtained in a similar way as for the perturbative expansion around $J = 0$.

We would like to calculate the quasi-potential $\mathcal{F}[\tilde{\rho}]$ for $\tilde{\rho} = (\rho_{inv} + \rho)$ as a power series in ρ. Note ρ has then to have the vanishing integral. In order to use the same notations as in the finite dimensional setting, see Sect. 3.3, we introduce the λ-dependent system defined by

$$\mathcal{K}^{\lambda}[\tilde{\rho}] = \frac{1}{\lambda} \mathcal{K}[\rho_{inv} + \lambda\rho] \tag{4.108}$$

$$\mathcal{Q}^{\lambda}[\tilde{\rho}] = \mathcal{Q}[\rho_{inv} + \lambda\rho] \tag{4.109}$$

$$\mathcal{F}^{\lambda}[\tilde{\rho}] = \frac{1}{\lambda^2} \mathcal{F}[\rho_{inv} + \lambda\rho]. \tag{4.110}$$

For any n we thus have

$$\mathcal{K}^{(0)}[\rho] = -\nabla \cdot [\rho\, b - J\rho_{inv}(\nabla V * \rho) - J\rho(\nabla V * \rho_{inv})] + k_B T \nabla^2 \rho, \tag{4.111}$$

$$\mathcal{K}^{(1)}[\rho] = J\nabla \cdot [\rho(\nabla V * \rho)], \tag{4.112}$$

$$\mathcal{K}^{(n)}[\rho] = 0 \qquad \text{for} \qquad n \geq 2. \tag{4.113}$$

Clearly, $\mathcal{K}^{(0)}[\rho] = R_{\rho_{inv}}\rho$, where $R_{\rho_{inv}}$ denotes the linearized Fokker–Planck operator. Moreover,

$$\mathcal{Q}^{(0)}[\rho] = -k_B T \, \nabla \cdot \rho_{inv}\nabla, \tag{4.114}$$

$$\mathcal{Q}^{(1)}[\rho] = -k_B T \, \nabla \cdot \rho\nabla, \tag{4.115}$$

$$\mathcal{Q}^{(n)}[\rho] = 0 \qquad \text{for} \qquad n \geq 2. \tag{4.116}$$

Recall that the computation of the quadratic order of $\mathcal{F}$ corresponds, in the above λ-dependent system, to the calculation of $\mathcal{F}^{(0)}$. Moreover, with these notations the hierarchy of equations to solve in order to obtain the Taylor expansion of the quasi-potential close to ρ_{inv} is given precisely by Eqs. (4.68) and (4.69) with $\mathcal{K}^{(n)}$ and $\mathcal{Q}^{(n)}$ expressed by the above formulae.

We shall search for the solution of this hierarchy in the form of homogeneous polynomials

$$\mathcal{F}^{(n)}[\rho] = \frac{1}{(n+2)!} \int \varphi^n(x_0, \ldots, x_{n+1})\, \rho(x_0) \cdots \rho(x_{n+1})\, dx_0 \cdots dx_{n+1}, \tag{4.117}$$

where $\varphi^n(x_0, \ldots, x_{n+1})$ are symmetric kernels on which we shall impose the relations

$$\int \varphi^n(x_0, \ldots, x_{n+1})\, dx_0 = 0. \tag{4.118}$$

As in the other cases, once the quadratic order $\mathcal{F}^{(0)}$ is known, the higher orders can be obtained by using the general solution to the hierarchy, see Eq. (4.72). We thus concentrate here only on $\mathcal{F}^{(0)}$. We introduce the operator Φ

$$\Phi[\rho](x) = \int \varphi^0(x, y)\, \rho(y)\, dy \tag{4.119}$$

that will be considered as acting in the space H_0 of functions with vanishing integral and L^2 scalar product. Note that Φ is a symmetric operator. From Eq. (4.68), we obtain the operator identity

$$\Phi R_{\rho_{inv}} + R_{\rho_{inv}}^{\dagger} \Phi = 2k_B T\, \Phi(\nabla \cdot \rho_{inv}\nabla)\Phi, \tag{4.120}$$

where R_{inv} is also viewed as an operator in H_0. Assuming that Φ is invertible, as it must be from the fact that $\mathcal{F}^{(0)}$ is positive definite, we infer that

$$R_{\rho_{inv}}\Phi^{-1} + \Phi^{-1}R_{\rho_{inv}}^{\dagger} = 2k_B T(\nabla \cdot \rho_{inv}\nabla), \tag{4.121}$$

which is the infinite dimensional analogue of Eq. (2.38). Its solution is given by the relation

$$\Phi^{-1} = -2k_B T \int_0^{\infty} e^{t R_{\rho_{inv}}} (\nabla \cdot \rho_{inv}\nabla)\, e^{t R_{\rho_{inv}}^{+}}\, dt, \tag{4.122}$$

see Eq. (2.37). Similarly to the finite-dimensional case, convergence of the integral is assured if ρ_{inv} a non-degenerate stable stationary solution of the McKean–Vlasov equation, i.e. if the spectrum of $R_{\rho_{inv}}$ on the space of functions with vanishing integral in contained in the complex half-plane with negative real part. Once we know the invertible operator Φ^{-1}, the function $\varphi^0(x_0, x_1)$ is extracted as the kernel of its inverse.

With an argument analogous to the one used in Sect. 3.4 for the finite dimensional systems, one can show that Eq. (4.122) implies that the covariance of density fluctuations diverges with the mean-field exponent.when approaching a codimension-one bifurcations of the McKean–Vlasov dynamics. We note that the latter result was obtained rigorously in [2].

Let us now consider the 0th-order fluctuation dynamics. We have

$$\mathcal{K}_r^{(0)}[\rho] = R_{inv}\rho - 2k_B T \, \nabla \cdot \rho_{inv} \nabla \Phi \rho \,=\, -\Phi^{-1} R_{inv}^\dagger \Phi \, \rho \,\equiv\, K_r^{(0)}\rho. \qquad (4.123)$$

$K_r^{(0)}$ is again an operator in H_0 that is invertible and has spectrum in the half-plane with positive imaginary part (it should not be confused with $K_r^{(0)}$ considered in Sect. 4.5.2). With Ansatz (4.117), the left hand side of Eq. (4.69) takes now the form

$$\int \frac{\delta \mathcal{F}^{(n)}}{\delta \rho(x)} \, \mathcal{K}_r^{(0)}[\rho](x) \, dx$$
$$= \frac{1}{(n+2)!} \int \left(K_{r,n+2}^{(0)\dagger} \, \varphi^n \right)(x_0, \ldots, x_{n+1}) \Big) \, \rho(x_0) \, \cdots \, \rho(x_{n+1}) \, dx_0 \cdots dx_{n+1} \qquad (4.124)$$

in the notation of (4.90). As for their right hand side, it follows from Eq. (4.71) that

$$\mathcal{S}^{(n)}[\mathcal{F}^{(0)}, \ldots, \mathcal{F}^{(n-1)}]$$
$$= \frac{1}{(n+2)!} \int s^n[\varphi^0, \ldots, \varphi^{n-1}](x_0, \ldots, x_{n+1}) \, \rho(x_0) \, \cdots \, \rho(x_{n+1}) \, dx_0 \cdots dx_{n+1},$$
$$\qquad (4.125)$$

where the symmetric kernels $s^{(n)}[\varphi^0, \ldots, \varphi^{n-1}](x_0, \ldots, x_{n+1})$ are expressed in terms of φ^k with $k < n$ and again satisfy the constraint (4.118). Thus in terms of the kernels, Eq. (4.69) become the identities

$$\left(K_{r,n+2}^{(0)\dagger} \, \varphi^n \right)(x_0, \ldots, x_{n+1}) = s^n[\varphi^0, \ldots, \varphi^{n-1}](x_0, \ldots, x_{n+1})(x_0, \ldots, x_{n+1})$$
$$\qquad (4.126)$$

which may be solved iteratively since the operators $\mathcal{K}_{r,n+2}^{(0)\dagger}$ are invertible on $H_0^{\otimes(n+2)}$ (note that the constraint (4.118) characterizes the symmetric functions in $H_0^{\otimes(n+2)}$). Further details are left to the reader, but in Appendix 2, we briefly discuss how φ^0, as well as higher order kernels, could be explicitly calculated numerically, leaving out of this work the explicit numerical implementation of the proposed algorithm.

5 Conclusions

The main aim of this paper was to discuss a perturbative approach to the calculation of the quasi-potential in parameter dependent stochastic dynamical systems. As we have seen, the Taylor expansion of the quasi-potential around an attractor of the deterministic dynamics was also covered by our theory.

The approach developed in the paper may be summarized as it was already done in the introduction. The perturbative expansion breaks the loop connecting the quasi-potential and

the instanton dynamics, with one needed to obtain the other. At any perturbative order the quasi-potential can be computed just from the knowledge of the instanton dynamics for the unperturbed problem and the lower order results. Indeed, our approach gives explicit formulae that permit to iteratively calculate any order of the power series expansion of the quasi-potential in the perturbative parameter. Explicit results can be obtained once the instanton dynamics for the unperturbed problem may be analytically or numerically solved.

We have first developed the theory for finite dimensional systems, where the mathematical details can be easily handled precisely. Our strategy was to perturbatively solve the Hamilton–Jacobi equation which, as it is well known [32], has the quasi-potential as the unique non-trivial solution under the hypothesis that the later is smooth enough. It is also known [21] that, at least in a neighborhood of the attractor, this is indeed the case, while further away singularities may occur, see for example [21,37,42,52,55,63]. An investigation on how to deal practically with singularities of the quasi-potential is left for the future. We also note that the perturbative study of transition rates between basin of attraction requires a specific approach, as was done for instance in the reference [75], using Melnikov's method for perturbations of Hamiltonian systems.

In the second part of the paper, we moved our attention to many body systems. In particular, we considered N particles that undergo an overdamped diffusion, interact with mean field conservative forces, and are driven out of equilibrium by an external drift, see Eq. (4.1). In this case it was possible to (formally) derive a fluctuating hydrodynamics, with the noise term proportional to $1/\sqrt{N}$, that describes in a closed way the evolution of the empirical density of particles. This fluctuating hydrodynamics is known in the literature under the name of the Dean equation [22]. In the limit when $N \to \infty$, the empirical density obeys then a deterministic PDE, the McKean–Vlasov equation [66]. This provided a formal but quick way to recover the results on propagation of chaos that appeared long ago in the mathematical literature.

In the case of a specific mean-field system, the Shinomoto–Kuramoto model of coupled rotators, we showed that such an approach can be employed to obtain explicit results. We first discussed MacKean-Vlasov's dynamics describing the behavior of the system in the $N \to \infty$ limit, characterizing analytically or semi-analytically all the stationary solutions of the system and their stability. Already this very simple out-of-equilibrium mean-field system displays quite a complex phase diagram with bifurcation lines and several attractors (stationary states as well as limit cycles). Albeit the Shinomoto–Kuramoto model has been studied in literature by numerical techniques, our analytical results characterizing all its stationary states are original.

After discussing the $N = \infty$ limit, we concentrated on large deviations around it. This was done by studying the Dean equation for N large but finite. Applying formal techniques from field theory (the Martin-Siggia-Rose representation and the WKB approximation), we showed that fluctuations of the empirical density are described by a generalization of the Freidlin–Wentzell theory to infinite dimensional systems. This permitted to recover formally rigorous results that already appeared in mathematical literature [18,19] and to make a connection with the Macroscopic Fluctuation Theory [6].

We subsequently generalized our perturbative approach to the calculation of quasi-potential to the case of fluctuating mean-field hydrodynamics. For the Shinomoto–Kuramoto model, we explicitly computed the quasi-potential close to the uncoupled particle dynamics to the 1st order in the mean-field coupling strength J and found the corresponding rate function for the probability of fluctuations of single-particle observables such as the magnetization. Our results seem to provide the first explicit calculation of the quasi-potential for diffusions with mean field interaction driven out of equilibrium. Higher orders in J could be

also computed without much difficulty. As our aim here was mainly to develop the general theory and to illustrate that it can be employed to obtain explicit results, we have not pursed such calculations further but discussed instead how one could similarly obtain explicit results for the Taylor expansion of the quasi-potential around the stationary states.

The perturbative approach developed in the present paper seems quite general and we expect it to be useful in several other problems. For example, it may be employed in the future to study other long-range interacting systems which are driven out of equilibrium by different mechanisms, such as stochastic forces which are spatially correlated. This setting is natural for systems whose constituents exhibit long-range interactions (such as plasmas or gravitational systems) but are exposed to stochastic external fields (electric for example) mainly acting on a given spatial scale. The typical behavior of this class of system was analyzed with kinetic theory in [68,69] and numerical simulations have shown how the phase transitions are modified when departing from the detailed balance [69]. The perturbative approach exposed here may give an analytical insight to such problems. This direction is presently pursued by one of the authors.

Similarly, applications of such perturbative methods could be found in the study of large deviations for $2d$ turbulence problems, which are relevant for climate modelling [7–10,59]. Even though these are quite academic questions [8] because realistic models of $2d$ turbulence and those arising from climate models are far from the perturbative regime, this approach can give valuable physical insight as it is difficult to handle the rare events in realistic systems.

Finally, while large deviations for the empirical density of mean-field interacting diffusions appeared in literature long ago, mathematical literature does not cover fluctuations of the empirical current. The formal approach through the Dean equation presented in this paper clearly permits to go in this direction. Some results on the current fluctuations that we have obtained already, as well as further explicit results on the calculation of the quasi-potential for the Shinomoto–Kuramoto model and their comparison with results obtained from direct numerical simulations are left for a future publication.

Appendix 1: Smooth Stationary Solutions of the Shinomoto–Kuramoto Model at $T = 0$

In Sect. 4.3.4, we have discussed simple delta-like stationary solutions of the McKean–Vlasov dynamics for the Shinomoto–Kuramoto model at $T = 0$. They described the particles accumulated at the fixed point of single particle drift. These are however not the only stationary solutions for vanishing temperature and also smooth solutions are present, which, however, are linearly unstable in the ferromagnetic model considered in this paper ($J > 0$). We briefly discuss such smooth stationary solutions at $T = 0$ and their stability in this appendix.

For $T = 0$, Eq. (4.36) implies that

$$\rho_{inv}(\theta + \theta_0) = \frac{\sqrt{1 - \left(\frac{y}{F}\right)^2}}{2\pi \left(1 - \frac{y}{F} \sin \theta\right)} \tag{6.1}$$

provided that $0 \le y < F$ and the self-consistency relation (4.43) holds taking now the form

$$\left(\frac{y}{J}\right)^2 + \left(\frac{1-\sqrt{1-\left(\frac{y}{F}\right)^2}}{\frac{y}{F}}\right)^2 = \left(\frac{h}{J}\right)^2. \tag{6.2}$$

The left hand side of the last equation grows monotonically from 0 at $y = 0$ to $(F/J)^2 + 1$ at $y = F$. Hence there is a unique solution for $0 \le y < F$ if and only if $F^2 + J^2 > h^2$. We infer that there is a unique stationary smooth solution at $T = 0$ when the latter constraint holds and none otherwise.

To examine the stability of such solutions, we examine the eigenfunctions $\delta\rho_\mu(\theta + \theta_0)$ with vanishing integral of the linearized Fokker–Planck operator $R_{\rho_{inv}}$ of (4.44) at $T = 0$. We may write $\delta\rho_\mu(\theta + \theta_0) = \partial_\theta \delta f_\mu(\theta)$, where δf_μ is periodic on $[0, 2\pi]$. The eigenequation takes the form

$$\begin{aligned}
&\left(R_{\rho_{inv}} \partial_\theta \delta f_\mu\right)(\theta) \\
&= -\partial_\theta \Big((F - y\sin\theta)\, \partial_\theta \delta f_\mu(\theta) - Y_1\, \rho_{inv}(\theta + \theta_0)\sin\theta + Y_2\, \rho_{inv}(\theta + \theta_0)\cos\theta\Big) \\
&= \mu\, \partial_\theta \delta f_\mu,
\end{aligned} \tag{6.3}$$

where

$$Y_1 = J \int_0^{2\pi} \sin\theta\ \delta f_\mu(\theta)\, d\theta, \qquad Y_2 = -J \int_0^{2\pi} \cos\theta\ \delta f_\mu(\theta)\, d\theta. \tag{6.4}$$

We may now define a new operator S such that

$$\partial_\theta \left(S\,\delta f\right) = R_{\rho_{inv}} \partial_\theta \delta f_\lambda. \tag{6.5}$$

Explicitly,

$$\left(S\,\delta f\right)(\theta) = -(F - y\sin\theta)\, \partial_\theta \delta f(\theta) - Y_1 \rho_{inv}(\theta + \theta_0)\,\sin\theta + Y_2 \rho_{inv}(\theta + \theta_0)\,\cos\theta. \tag{6.6}$$

Clearly, the diagonalization of $R_{\rho_{inv}}$ on the subspace of functions δg with vanishing integral is equivalent to the diagonalization of S.

A direct calculation shows that the 3-dimensional subspace V_3 spanned by functions

$$f_1(\theta) = \rho_{inv}(\theta + \theta_0)\,\sin\theta, \qquad f_2(\theta) = \rho_{inv}(\theta + \theta_0)\,\cos\theta, \qquad f_3(\theta) = \rho_{inv}(\theta + \theta_0) \tag{6.7}$$

is invariant for S. In particular, the representation of S on such subspace is given by the matrix

$$\begin{pmatrix}
J\dfrac{1-\sqrt{1-\left(\frac{y}{F}\right)^2}}{\left(\frac{y}{F}\right)^2} & F & J\dfrac{1-\sqrt{1-\left(\frac{y}{F}\right)^2}}{\frac{y}{F}} \\[2em]
-F & J\left(1 - \dfrac{1-\sqrt{1-\left(\frac{y}{F}\right)^2}}{\left(\frac{y}{F}\right)^2}\right) & -y \\[2em]
0 & -y & 0
\end{pmatrix}. \tag{6.8}$$

A little algebra shows that the above matrix has one zero eigenvalue which corresponds to the eigenvector $(-y, 0, F)$, i.e. to the constant function

$$\delta f(\theta) = \rho_{inv}(\theta + \theta_0)\,(F - y\sin\theta) = \frac{\sqrt{F^2 - y^2}}{2\pi}. \tag{6.9}$$

In this case $\delta g = \partial_\theta \delta f = 0$. Consequently, the zero eigenvalue does not occur for the operator $R_{\rho_{inv}}$. The other two eigenvalues of the matrix (6.7) are given by

$$\lambda_\pm = \frac{J}{2} \pm \sqrt{\left(\frac{J}{2}\right)^2 - J^2 \frac{\sqrt{1-\left(\frac{y}{F}\right)^2}}{\left(1+\sqrt{1-\left(\frac{y}{F}\right)^2}\right)^2} - F^2 + y^2}. \tag{6.10}$$

of which at least one has positive real part if $J > 0$. It follows that for $J > 0$, the stationary solution (6.1) is linearly unstable.

Appendix 2: Algorithm for the Taylor Expansion of the Quasi-Potential Close to a Stationary State

We discuss briefly in this Appendix how the Taylor expansion of the quasi-potential of the Shinomoto–Kuramoto model may be numerically calculated. This will be possible, in principle, at all the orders of the Taylor expansion.

Let us first concentrate on the quadratic term $\mathcal{F}^{(0)}$. In order to explicitly calculate kernel $\varphi^0(\theta, \vartheta)$, we proceed similarly as in Sect. 4.5.3 and consider the eigenfunctions of operators $R^\dagger_{\rho_{inv}}$ and $R_{\rho_{inv}}$ acting in space H_0:

$$R^\dagger_{\rho_{inv}} u_k = \alpha_k \, u_k \qquad \text{and} \qquad R_{\rho_{inv}} v_k = \overline{\alpha}_k \, v_k. \tag{7.1}$$

Assuming α_k to be different, we impose the orthogonality relations (4.99). Eigenfunctions u_k form a basis of H_0 but they are not orthogonal. Similarly for v_k. The kernel of the quadratic term of the quasi-potential may be represented as

$$\varphi^0(\theta, \vartheta) = \sum_{k,l} \Phi_{kl} \, v_k(\theta) \, \overline{v_l(\vartheta)} \tag{7.2}$$

where Φ_{kl} is defined by

$$\Phi_{kl} = \int_0^{2\pi} \overline{u_k(\theta)} \, d\theta \int_0^{2\pi} \varphi^0(\theta, \vartheta) \, u_l(\vartheta) \, d\vartheta = \langle u_k | \Phi u_l \rangle. \tag{7.3}$$

The problem is thus reduced to the calculation of Φ_{kl}. Now, from Eq. (4.122), we can easily compute the matrix elements of Φ^{-1} in the basis u_k. Indeed, from Eq. (4.122),

$$(\Phi^{-1})_{kl} \equiv \langle u_k | \Phi^{-1} u_l \rangle = -\frac{2k_B T}{\alpha_k^* + \alpha_l} \int d\theta \, \overline{(\partial_\theta u_k)(\theta)} \, \rho_{inv}(\theta) \, (\partial_\theta u_l)(\theta). \tag{7.4}$$

Note, however, that the matrix $(\Phi^{-1})_{kl}$ is not the inverse of Φ_{kl} because the basis formed by u_k is not orthonormal. This problem can be handled with simple linear algebra. We introduce the matrix of scalar products

$$P_{kl} = \langle u_k | u_l \rangle \tag{7.5}$$

whose inverse will be denoted by $(P^{-1})_{kl}$. Then, we construct another matrix

$$(B^{-1})_{kl} \equiv \sum_i (P^{-1})_{ki} \, (\Phi^{-1})_{il} \tag{7.6}$$

with inverse B_{kl}. Finally,

$$\Phi_{kl} = \sum_i P_{ki} \, B_{il}. \tag{7.7}$$

Once Φ_{kl} is known, it is straightforward to write the kernel φ^0 in the real space using Eq. (4.126).

We conclude by observing that, once φ^0 is known, one could also numerically evaluate the higher order kernels φ^n of the Taylor expansion by solving Eq. (4.126) using the fact that the basis Φv_k is composed of eigenstates of operators $K_r^{(0)\dagger} = -\Phi \, R_{\rho_{inv}} \Phi^{-1}$, see Eqs. (4.123) and (7.1).

References

1. Acebrón, J.A., Bonilla, L.L., Vicente, C.J.P., Ritort, F., Spigler, R.: The Kuramoto model: a simple paradigm for synchronization phenomena. Rev. Mod. Phys. **77**, 137–185 (2005)
2. Ben Arous, G., Brunaud, M.: Méthode de Laplace: Étude variationnelle des fluctuations de diffusions de type "champ moyen". Stoch. Stoch. Rep. **31**, 79–144 (1990)
3. Barré, J., Chetrite, R., Muratori, M., Peruani, F.: Motility-induced phase separation of active particles in the presence of velocity alignment. J. Stat. Phys. **158**, 589–600 (2015)
4. Ben Arous, G., Guionnet, A.: Large deviations for langevin spin glass dynamics. Probab. Theory Relat. Fields **102**, 455–509 (1995)
5. Ben Arous, G., Guionnet, A.: Symmetric Langevin spin glass dynamics. Ann. Probab. **25**, 1367–1422 (1997)
6. Bertini, L., De Sole, A., Gabrielli, D., Jona-Lasinio, G., Landim, C.: Macroscopic fluctuation theory. Preprint arXiv:1404.6466 (2014)
7. Bouchet, F., Laurie, J., Laurie, J., Zaboronski, O.: Control and instanton trajectories for random transitions in turbulent flows. J. Phys. Conf. Ser. **318**(022041), 1–7 (2011)
8. Bouchet, F., Laurie, J., Zaboronski, O.: Langevin dynamics, large deviations and instantons for the quasi-geostrophic model and two-dimensional Euler equations. J. Stat. Phys. **156**, 1066–1092 (2014)
9. Bouchet, F., Nardini, C., Tangarife, T.: Kinetic theory of jet dynamics in the stochastic barotropic and 2D Navier–Stokes equations. J. Stat. Phys. **153**, 572–625 (2013)
10. Bouchet, F., Nardini, C., Tangarife, T.: Stochastic averaging, large deviations and random transitions for the dynamics of 2d and geostrophic turbulent vortices. Fluid Dyn. Res **46**(061416), 1–11 (2014)
11. Braun, W., Hepp, K.: The Vlasov dynamics and its fluctuations in the 1/n limit of interacting classical particles. Commun. Math. Phys. **56**, 101–113 (1977)
12. Cameron, M.K.: Finding the quasipotential for nongradient SDEs. Physica D **241**, 1532–1550 (2012)
13. Chavanis, P.H.: The Brownian mean field model. Eur. Phys. J. B **87**, 1–33 (2014)
14. Chavanis, P.H.: Thermodynamics of the HMF model with a magnetic field. Eur. Phys. J. B **80**, 275–306 (2011)
15. Chavanis, P.H.: Hamiltonian and Brownian systems with long-range interactions: V. Stochastic kinetic equations and theory of fluctuations. Physica A **387**, 5716–5740 (2008)
16. Dai Pra, P., den Hollander, F.: Mckean–Vlasov limit for interacting random processes in random media. J. Stat. Phys. **84**, 735–772 (1996)
17. Da Prato, G., Zabczyk, J.: Stochastic Equations in Infinite Dimensions. Encyclopedia of Mathematics and Its Applications, vol. 152. Cambridge University Press, Cambridge (2014)
18. Dawson, D.A., Gärtner, J.: Large deviations from the Mckean–Vlasov limit for weakly interacting diffusions. Stochastics **20**, 247–308 (1987)
19. Dawson, D.A., Gärtner, J.: Long-time fluctuations of weakly interacting diffusions. In: Engelbert, H., Schmidt, W. (eds.) Stochastic Differential Systems. Lecture Notes in Control and Information Sciences, vol. 96, pp. 1–10. Springer, Berlin (1987)
20. Dawson, D.A.: Critical dynamics and fluctuations for a mean-field model of cooperative behavior. J. Stat. Phys. **31**, 29–85 (1983)
21. Day, M.V., Darden, T.A.: Some regularity results on the Ventcel–Freidlin quasi-potential function. Appl. Math. Optim. **13**, 259–282 (1985)
22. Dean, D.S.: Langevin equation for the density of a system of interacting Langevin processes. J. Phys. A **29**, L613–L617 (1996)

23. Dembo, A., Zeitouni, O.: Large Deviations Techniques and Applications, 2nd edn. Springer, New York (1998)
24. den Hollander, F.: Large Deviations. Fields Institute Monographs, vol. 14. American Mathematical Society, Providence, RI (2008)
25. Dobrushin, R.L.: Vlasov equations. Funct. Anal. Appl. **13**, 115–123 (1979)
26. Dykman, M.I.: Large fluctuations and fluctuational transitions in systems driven by colored Gaussian noise: a high-frequency noise. Phys. Rev. A **42**, 2020–2029 (1990)
27. Dykman, M.I., Krivoglaz, M.A.: Theory of fluctuational transitions between stable states of nonlinear oscillators. Sov. Phys. JETP **50**, 30–37 (1979)
28. Dykman, M.I., Krivoglaz, M.A.: Fluctuations in nonlinear systems near bifurcations corresponding to the appearance of new stable states. Physica A **104**, 480–494 (1980)
29. Ellis, R.S.: Entropy, Large Deviations, and Statistical Mechanics. Springer, New York (1985)
30. Faris, W.G., Jona-Lasinio, G.: Large fluctuations for a nonlinear heat equation with noise. J. Phys. A **15**, 3025–3055 (1982)
31. Fleming, W.H., Soner, H.M.: Controlled Markov Processes and Viscosity Solutions. Stochastic Modelling and Applied Probability, vol. 25. Springer, New York (2006)
32. Freidlin, M.I., Wentzell, A.D.: Random Perturbations of Dynamical Systems. Grundlehren der Mathematischen Wissenschaften, vol. 260, 3rd edn. Springer, Berlin (2012)
33. Gang, H., Haken, H.: Polynomial expansion of the potential of Fokker–Planck equations with a noninvertible diffusion matrix. Phys. Rev. A **40**, 5966–5978 (1989)
34. Giacomin, G., Pakdaman, K., Pellegrin, X., Poquet, C.: Transitions in active rotator systems: invariant hyperbolic manifold approach. SIAM J. Math. Anal. **44**, 4165–4194 (2012)
35. Giacomin, G., Poquet, C.: Noise, interaction, nonlinear dynamics and the origin of rhythmic behaviors. Braz. J. Probab. Stat. **29**, 460–493 (2015)
36. Graham, R.: Macroscopic potentials, bifurcations and noise in dissipative systems. In: Moss, F., McClintock, P.V.E. (eds.) Noise in Nonlinear Dynamical Systems, vol. 1, pp. 225–278. Cambridge University Press, Cambridge (1989)
37. Graham, R.: Fluctuations in the steady state. In: Brey, J.J., Marro, J., Rubí, J.M., San Miguel, M. (eds.) 25 Years of Non-Equilibrium Statistical Mechanics, pp. 125–134. Springer, New York (1995)
38. Graham, R., Roekaerts, D., Tél, T.: Integrability of Hamiltonians associated with Fokker–Planck equations. Phys. Rev. A **31**, 3364–3375 (1985)
39. Graham, R., Schenzle, A.: Non-equilibrium potentials and stationary probability distributions of some dissipative models without manifest detailed balance. Zeit. Phys. B **52**, 61–68 (1983)
40. Graham, R., Tél, T.: Existence of a potential for dissipative dynamical systems. Phys. Rev. Lett. **52**, 9–12 (1984)
41. Graham, R., Tél, T.: On the weak-noise limit of Fokker–Planck models. J. Stat. Phys. **35**, 729–748 (1984)
42. Graham, R., Tél, T.: Weak-noise limit of Fokker–Planck models and nondifferentiable potentials for dissipative dynamical systems. Phys. Rev. A **31**, 1109–1122 (1985)
43. Graham, R., Tél, T.: Nonequilibrium potential for coexisting attractors. Phys. Rev. A **32**, 1322–1337 (1986)
44. Graham, R., Tél, T.: Nonequilibrium potentials for local codimension-2 bifurcations of dissipative flows. Phys. Rev. A **35**, 1328–1349 (1987)
45. Grunwald, M.: Sanov results for Glauber spin-glass dynamics. Probab. Theory Relat. Fields **106**, 187–232 (1996)
46. Hairer, M.: Solving the KPZ equation. Ann. Math. **178**, 559–664 (2013)
47. Haussmann, U.G., Pardoux, E.: Time reversal of diffusions. Ann. Probab. **14**, 1188–1205 (1986)
48. Heymann, M., Vanden-Eijnden, E.: The geometric minimum action method: a least action principle on the space of curves. Commun. Pure Appl. Math. **61**, 1052–1117 (2008)
49. Ilyashenko, Yu., Yakovenko, S.: Lectures on Analytic Differential Equations. Graduate Studies in Mathematics, vol. 86. AMS, Providence, RI (2008)
50. Irvin, M.C.: Smooth Dynamical Systems. Advanced Series in Nonlinear Dynamics. World Scientific, Singapore (2001)
51. Jauslin, H.R.: Melnikov's criterion for nondifferentiable weak-noise potentials. J. Stat. Phys. **42**, 573–585 (1986)
52. Jauslin, H.R.: Nondifferentiable potentials for nonequilibrium steady states. Physica A **144**, 179–191 (1987)
53. Jouvet, B., Phythian, R.: Quantum aspects of classical and statistical fields. Phys. Rev. A **19**, 1350–1355 (1979)
54. Zinn-Justin, J.: Quantum Field Theory and Critical Phenomena. International Series of Monographs on Physics, vol. 113, 4th edn. Oxford University Press, Oxford (2002)

55. Kamenev, A.: Field Theory of Non-equilibrium Systems. Cambridge University Press, Cambridge (2011)
56. Kitagawa, G.: An algorithm for solving the matrix equation $X = F X F^T + S$. Int. J. Control **25**, 745–753 (1977)
57. Knobloch, E., Wiesenfeld, K.A.: Bifurcations in fluctuating systems: the center-manifold approach. J. Stat. Phys. **33**, 611–637 (1983)
58. Kuramoto, Y.: Self-entrainment of a population of coupled non-linear oscillators. In: Araki, H. (ed.) International Symposium on Mathematical Problems in Theoretical Physics. Lecture Notes in Physics, pp. 420–422. Springer, Berlin (1975)
59. Laurie, J., Bouchet, F.: Computation of rare transitions in the barotropic quasi-geostrophic equations. New J. Phys. **17**(015009), 1–25 (2015)
60. Luçon, E., Poquet, C.: Long time dynamics and disorder-induced traveling waves in the stochastic Kuramoto model. Preprint arXiv:1505.00497 (2015)
61. Machlup, S., Onsager, L.: Fluctuations and irreversible process. II. Systems with kinetic energy. Phys. Rev. **91**, 1512–1515 (1953)
62. Maier, R.S., Stein, D.L.: Escape problem for irreversible systems. Phys. Rev. E **48**, 931–938 (1993)
63. Maier, R.S., Stein, D.L.: A scaling theory of bifurcations in the symmetric weak-noise escape problem. J. Stat. Phys. **83**, 291–357 (1996)
64. Mangel, M.: Uniform treatment of fluctuations at critical points. Physica A **97**, 597–615 (1979)
65. Martin, P.C., Siggia, E.D., Rose, H.A.: Statistical dynamics of classical systems. Phys. Rev. A **8**, 423–437 (1973)
66. McKean Jr., H.P.: A class of Markov processes associated with nonlinear parabolic equations. Proc. Natl Acad. Sci. USA **56**, 1811–1907 (1966)
67. Méléard, S.: Asymptotic behaviour of some interacting particle systems: McKean–Vlasov and Boltzmann models. Probabilistic Models for Nonlinear PDE, pp. 42–92. Springer, Berlin (1996)
68. Nardini, C., Gupta, S., Ruffo, S., Dauxois, T., Bouchet, F.: Kinetic theory for non-equilibrium stationary states in long-range interacting systems. J. Stat. Mech. **L01002**, 1–11 (2012)
69. Nardini, C., Gupta, S., Ruffo, S., Dauxois, T., Bouchet, F.: Kinetic theory of nonequilibrium stochastic long-range systems: phase transition and bistability. J. Stat. Mech. **P12010**, 1–28 (2012)
70. Ohta, H., Sasa, S.: Critical phenomena in globally coupled excitable elements. Phys. Rev. E **78**(065101(R)), 1–4 (2008)
71. Pikovsky, A., Rosenblum, M.: Dynamics of heterogeneous oscillator ensembles in terms of collective variables. Physica D **240**, 872–881 (2011)
72. Poquet, C.: Phase reduction in the noise induced escape problem for systems close to reversibility. Stoch. Process. Appl. **124**, 3312–3341 (2014)
73. Sakaguchi, H., Shinomoto, S., Kuramoto, Y.: Phase transitions and their bifurcation analysis in a large population of active rotators with mean-field coupling. Prog. Theoret. Phys. **79**, 600–607 (1988)
74. Shinomoto, S., Kuramoto, Y.: Phase transitions in active rotator systems. Prog. Theoret. Phys. **75**, 1105–1110 (1986)
75. Smelyanskiy, V.N., Dykman, M.I., Rabitz, H., Vugmeister, B.E.: Fluctuations, escape, and nucleation in driven systems: logarithmic susceptibility. Phys. Rev. Lett. **79**, 3113–3116 (1997)
76. Spohn, H.: Large Scale Dynamics of Interacting Particles. Springer, Heidelberg (1991)
77. Stratonovič, R.L.: On the probability functional of diffusion processes (in Russian). In: Proc. Sixth All-Unian Conf. Theory Prob. and Math. Statist., pp. 471–483, Gosudarstv. Izdat. Politič. i Naučn. Lit. Litovsk. SSR, Vilnius (1962). English translation. In: Selected Transl. in Math. Statist. and Propbability, vol. 10, pp. 273–286, AMS (1971)
78. Stroock, D.W., Varadhan, S.R.S.: Diffusion processes. In: Proceedings of Sixth Berkeley Symposium on Mathematical Statistics and Probability, vol. 3, pp. 361–368, University of California Press (1972)
79. Sznitman, A.: Topics in propagation of chaos. In: Ecole d'Eté de Probabilités de Saint-Flour XIX 1989, pp. 165–251. Springer, Berlin (1991)
80. Tél, T., Graham, R., Hu, G.: Nonequilibrium potentials and their power-series expansions. Phys. Rev. A **40**, 4065–4071 (1989)
81. Teschl, G.: Ordinary Differential Equations and Dynamical Systems. Graduate Studies in Mathematics, vol. 140. American Mathematical Society, Providence, RI (2012)
82. Touchette, H.: The large deviation approach to statistical mechanics. Phys. Rep. **478**, 1–69 (2009)
83. Zaks, M.A., Neiman, A.B., Feistel, S., Schimansky-Geier, L.: Noise-controlled oscillations and their bifurcations in coupled phase oscillators. Phys. Rev. E **68**(066206), 1–9 (2003)