HAL
open science

Large deviations and entropy production in viscous fluid flows

Vojkan Jaksic, Vahagn Nersesyan, Claude-Alain Pillet, Armen Shirikyan

To cite this version:

Vojkan Jaksic, Vahagn Nersesyan, Claude-Alain Pillet, Armen Shirikyan. Large deviations and entropy production in viscous fluid flows. Archive for Rational Mechanics and Analysis, 2021, 240 (3), pp.1675-1725. 10.1007/s00205-021-01646-3 . hal-02012709

HAL Id: hal-02012709

https://hal.science/hal-02012709

Submitted on 9 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Large deviations and entropy production in viscous fluid flows

V. Jakšić[1] V. Nersesyan[2,5] C.-A. Pillet[3] A. Shirikyan[4,6]

Abstract

We study the motion of a particle in a random time-dependent vector field defined by the 2D Navier–Stokes system with a noise. Under suitable non-degeneracy hypotheses we prove that the empirical measures of the trajectories of the pair (velocity field, particle) satisfy the LDP with a good rate function. Moreover, we show that the law of a unique stationary solution restricted to the particle component possesses a positive smooth density with respect to the Lebesgue measure in any finite time. This allows one to define a natural concept of the entropy production, and to show that its time average is a bounded function of the trajectory. The proofs are based on a new criterion for the validity of the level-3 LDP for Markov processes and an application of a general result on the image of probability measures under smooth maps to the laws associated with the motion of the particle.

AMS subject classifications: 35Q30, 35R60, 60B12, 60F10, 76D05, 93B05

Keywords: Large deviations, entropy production, Navier–Stokes system, Lagrangian trajectories, regular densities

[1]Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Street West, Montreal, QC, H3A 2K6 Canada; e-mail: Jaksic@math.mcgill.ca

[2]Laboratoire de Mathématiques, UMR CNRS 8100, UVSQ, Université Paris-Saclay, 45, av. des Etats-Unis, F-78035 Versailles, France; e-mail: Vahagn.Nersesyan@math.uvsq.fr

[3]Aix Marseille Univ, Université de Toulon, CNRS, CPT, Marseille, France; e-mail: Pillet@univ-tln.fr

[4]Department of Mathematics, University of Cergy–Pontoise, CNRS UMR 8088, 2 avenue Adolphe Chauvin, 95302 Cergy–Pontoise, France; e-mail: Armen.Shirikyan@u-cergy.fr

[5]Centre de Recherches Mathématiques, CNRS UMI 3457, Université de Montréal, Montréal, QC, H3C 3J7, Canada

[6]Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Street West, Montreal, QC, H3A 2K6, Canada

Contents

0 Introduction **2**

1 Main results **8**
1.1 Formulations 8
1.2 Schemes of the proofs 13

2 Large deviations via controllability **16**
2.1 Formulation of the result 16
2.2 General scheme of the proof of Theorem 2.1 20
2.3 Proof of Proposition 2.3 24
2.4 Proof of Proposition 2.4 28
2.5 Proof of Proposition 2.5 29

3 Image of measures under non-degenerate maps **32**

4 Application to the 2D Navier–Stokes system with a particle **33**
4.1 Large Deviation Principle 33
4.2 Regularity of laws and convergence 36
4.3 Strict positivity of densities 39

5 Appendix **41**
5.1 Exponential mixing and coupling operators 42
5.2 Kifer's criterion 45
5.3 Asymptotics of Feynman–Kac semigroups 45
5.4 Proofs of Theorems 3.1 and 3.2 46
5.5 Agrachev–Sarychev theorem 48

Bibliography **49**

0 Introduction

The theory of entropic fluctuations in deterministic and stochastic systems of mathematical physics underwent a spectacular development in the last thirty years. It was initiated in the middle of nineties of the last century in physics literature (see [ECM93, ES94, GC95b, GC95a]), and was developed rapidly by various research groups. We refer the reader to the papers [Gal95, Kur98, LS99, Mae99, Rue99, ES02, Gas05, RM07, CG08, JPR11, CJPS17] and the references therein for a detailed account of major achievements in the field. The viewpoints and the frameworks adopted in these papers are not necessarily the same, and we start by briefly describing the approach to the modern theory of entropic fluctuations that we will adopt here, confining ourselves to the discrete-time setting. For additional information, see the paper [CJPS17] and the forthcoming review articles [CJN$^+$, CJPS].

The starting point of the theory of entropic fluctuations is the *Large Deviation Principle* (LDP) for the empirical measures associated with trajectories.[1] Namely, denoting by $\mathscr{X}$ the phase space of the system in question and by $\{u_k\}_{k\geq 0}$ a random trajectory, we introduce the *empirical measures* by

$$\boldsymbol{\nu}_t = t^{-1}\sum_{k=0}^{t-1}\delta_{\boldsymbol{u}_k}, \quad t\geq 1, \tag{0.1}$$

where $\boldsymbol{u}_k = (u_l, l\geq k)$. Thus, $\boldsymbol{\nu}_t$ is a *random probability measure* on the product space $\boldsymbol{X} = \mathscr{X}^{\mathbb{Z}_+}$, where $\mathbb{Z}_+$ is the set of non-negative integers. If the LDP holds for the sequence $\{\boldsymbol{\nu}_t\}_{t\geq 1}$, then we get an object—the rate function $\boldsymbol{I}$—giving a detailed information on its large-time asymptotics. Very often $\boldsymbol{I}$ does not depend on a choice of trajectory, which makes it an important characteristic of the system.

Suppose now that the system under study possesses a natural time reversal operation θ that can be lifted to an involution $\boldsymbol{\theta}$ in the space of probability measures $\mathcal{P}(\boldsymbol{X})$ (on which $\boldsymbol{I}$ is defined). One can ask then how $\boldsymbol{I}$ transforms under the action of $\boldsymbol{\theta}$. It was observed in [BL08, BC15, CJPS17] that, under some additional hypotheses, there is an affine function $\mathrm{ep}:\mathcal{P}(\boldsymbol{X})\to\mathbb{R}$ such that

$$\boldsymbol{I}(\boldsymbol{\lambda}\circ\boldsymbol{\theta}) = \boldsymbol{I}(\boldsymbol{\lambda}) + \mathrm{ep}(\boldsymbol{\lambda}) \tag{0.2}$$

for a large class of measures $\boldsymbol{\lambda}\in\mathcal{P}(\boldsymbol{X})$. Identity (0.2) is called *level-3 fluctuation relation*, and the second term on its right-hand side is called the *mean entropy production with respect to* $\boldsymbol{\lambda}$. In the Markovian situation, under some regularity hypotheses, the quantity $\mathrm{ep}(\boldsymbol{\lambda})$ is the integral of a function $\sigma:\boldsymbol{X}\to\mathbb{R}$ with respect to $\boldsymbol{\lambda}$ (which will be denoted by $\langle\sigma,\boldsymbol{\lambda}\rangle$). More generally, in practically all cases of interest, the mean entropy production can be written in the form

$$\mathrm{ep}(\boldsymbol{\lambda}) = \lim_{t\to\infty} t^{-1}\langle\sigma_t,\boldsymbol{\lambda}\rangle, \tag{0.3}$$

where $\{\sigma_t\}$ is a sequence of measurable functions on $\boldsymbol{X}$. The functions σ and σ_t (called *entropy production functional* and *entropy production in time* t) may be very irregular, and their identification is often a delicate question. Furthermore, the study of the large time behaviour of the quantities $\langle\sigma,\boldsymbol{\nu}_t\rangle$ or $t^{-1}\sigma_t$, which are called the *time average of the entropy production*, is typically a difficult mathematical problem. Of particular importance are the convergence to a limit and the LDP as $t\to\infty$ because these properties are related to the emergence of the arrow of time and its quantitive description. Namely, if the sequence $\{t^{-1}\sigma_t\}$ has a non-vanishing deterministic limit $\bar\sigma$ (called *mean entropy production rate*), then the law of the process $\{u_k\}$ and its image under the time reversal $\boldsymbol{\theta}$ separate from each other as $t\to\infty$ and eventually become mutually singular. Moreover, if $\{t^{-1}\sigma_t\}$ satisfies the LDP (or even local LDP on a sufficiently large interval), then one can give a detailed description of the above-mentioned separation of measures in terms of the Hoeffding error exponents (see [JOPS12, CJPS17, CJN+]). If,

[1] All the concepts used in this introduction are defined in the main text.

in addition, the rate function I of the full LDP for $\{t^{-1}\sigma_t\}$ is obtained from $\boldsymbol{I}$ by the contraction relation

$$I(r) = \inf\{\boldsymbol{I}(\boldsymbol{\lambda}) : \mathrm{ep}(\boldsymbol{\lambda}) = r\}, \tag{0.4}$$

then I has to satisfy the celebrated *Gallavotti–Cohen symmetry relation*

$$I(-r) = I(r) + r \quad \text{for } r \in \mathbb{R}. \tag{0.5}$$

Finally, one can prove that the mean entropy production rate $\bar{\sigma}$ is always non-negative, and its strict positivity ensures the non-triviality of the error exponents and the emergence of the arrow of time. Since mathematical justification of the above program amounts to proving a fine form of the second law of thermodynamics for the system under consideration, it should not come as a surprise that for physically relevant models each step of the program is often a formidable mathematical problem.

Summarising the above discussion, we can state the following steps in the investigation of entropic fluctuations for a given system:

(a) LDP for the empirical measures (0.1), also called level-3 LDP.

(b) Level-3 fluctuation relation (0.2).

(c) Identification of σ_t, the functional of entropy production in time t, and its relation with physical transport properties.

(d) Law of large numbers for the sequence of time averages $\{t^{-1}\sigma_t\}$.

(e) Strict positivity of the mean entropy production rate $\bar{\sigma}$.

(f) Local and global LDP for the sequence of time averages $\{t^{-1}\sigma_t\}$.

We emphasise that each of these steps is essentially a separate problem, and they do not need to be studied in the stated order.

The aim of this paper is to address questions (a) and (c) for a fluid particle moving in a two-dimensional periodic box. Namely, we consider the ordinary differential equation (ODE)

$$\dot{y} = u(t, y), \quad y \in \mathbb{T}^2, \tag{0.6}$$

where $u(t, y)$ is a time-dependent vector field defined by the 2D Navier–Stokes system subject to an external random forcing. The law of u is assumed to be invariant under the time translation $t \mapsto t + 1$, while the process itself should have good mixing properties. We do not give more details on the random field u, referring the reader to Section 1.1 for the exact hypotheses. The ODE (0.6) is supplemented with the initial condition

$$y(0) = p, \tag{0.7}$$

where $p \in \mathbb{T}^2$ is a given point. The solution of (0.6), (0.7) defines a random dynamical system $\varphi_t : \mathbb{T}^2 \to \mathbb{T}^2$, $t \geq 0$, and we are interested in the large-time behaviour of the restriction of φ_t to the integer times. More precisely, let $\boldsymbol{T} := (\mathbb{T}^2)^{\mathbb{Z}_+}$ and

$$\boldsymbol{\lambda}_t^p = t^{-1} \sum_{k=0}^{t-1} \delta_{\boldsymbol{y}_k}, \quad t \geq 1, \tag{0.8}$$

where $\delta_{\boldsymbol{y}} \in \mathcal{P}(\boldsymbol{T})$ is the Dirac mass at the point $\boldsymbol{y} \in \boldsymbol{T}$, and $\boldsymbol{y}_k = (\varphi_t(p), t \geq k)$. For any $p \in \mathbb{T}^2$, $\{\boldsymbol{\lambda}_t^p\}$ is a sequence of random probability measures on $\boldsymbol{T}$. The following theorem is a concise and informal formulation of the main results of this paper. The exact statements and further details can be found in Section 1.1.

Main Theorem. *Under suitable hypotheses on the vector field $u(t, y)$, there is a $\mathbb{T}^2$-valued random process $\{z_t, t \geq 0\}$ such that its almost every trajectory satisfies (0.6) and the following assertions hold.*

Stationarity. *The laws of the processes $\{z_t\}_{t\geq 0}$ and $\{z_{1+t}\}_{t\geq 0}$ coincide, and the law of each component coincides with the normalised Lebesgue measure on $\mathbb{T}^2$.*

Convergence. *For any $s \geq 1$ and any initial point $p \in \mathbb{T}^2$, the law of the vector $(\varphi_t(p), \dots, \varphi_{t+s}(p))$ converges exponentially fast in the total variation norm, as $t \to \infty$, to that of $(z_0, \dots, z_s)$.*

Large deviations. *For any $p \in \mathbb{T}^2$, the sequence $\{\boldsymbol{\lambda}_t^p\}_{t\geq 1}$ satisfies the LDP with some good rate function $\boldsymbol{I} : \mathcal{P}(\boldsymbol{T}) \to [0, +\infty]$.*

Entropy production. *For any $t \geq 1$, the law of $(z_1, \dots, z_t)$ has a strictly positive smooth density $\rho_t(x_1, \dots, x_t)$ with respect to the Lebesgue measure on $\mathbb{T}^{2t}$. Moreover, there is a number $C > 0$ such that the entropy production in time t, defined by*

$$\sigma_t(\boldsymbol{y}^t) = \log \frac{\rho_t(y_1, \dots, y_t)}{\rho_t(y_t, \dots, y_1)}, \quad \boldsymbol{y}^t := (y_1, \dots, y_t), \tag{0.9}$$

satisfies the inequality $-C \leq t^{-1}\sigma(\boldsymbol{y}^t) \leq C$ for all $\boldsymbol{y}^t \in \mathbb{T}^{2t}$.

Let us mention that the problem of transport of particles in time-dependent or random vector fields was studied by many authors; see, for example, the papers [Kra70, FP94, Mol96, KPS13] and the references therein. However, most of these works treat questions that are different from those studied here. To the best of our knowledge, the only exception is the article [KPS13], which establishes the law of large numbers and central limit theorem for the particle position $y(t)$ considered in the whole space $\mathbb{R}^2$ (rather than $\mathbb{T}^2$). This type of results is not sufficient to get the convergence of the law of y to a limiting measure or to study the large deviations for empirical measures. We also mention the recent article [BBP18], which studies another aspect of chaotic behaviour of fluids—the strict positivity of the top Lyapunov exponent for the dynamics of the Lagrangian particle. The hypotheses imposed in [BBP18] are somewhat different from ours and require the noise to be sufficiently irregular in the space variables.

The mathematical theory of entropic fluctuations for randomly forced PDEs is in the beginning of its development. The only two cases for which the complete program (a)–(f) has been carried out are the 1D Burgers equation and a nonlinear reaction-diffusion system perturbed by a *rough* kick noise; see [JNPS15a]. However, from the physical point of view, the roughness hypothesis on the noise is not always justified, especially in the context of the fluid motion. Although the Navier–Stokes system perturbed by a *smooth* random force satisfies the level-3 LDP (see[2] [JNPS15b]), in this case the laws of the forward and backward evolutions are typically singular with respect to each other, and the basic object of the theory of entropic fluctuations—the entropy production in time t—is not defined. The present paper bypasses this basic obstruction in a physically and mathematically natural way by focusing on the motion of a particle immersed in the fluid for which we show that all the objects of the theory of entropic fluctuations are well defined. In particular, we establish the level-3 LDP and a uniform bound for the mean entropy production in time t. At the same time, the points (b), (d), (e), and (f) of the above-mentioned program are yet to be studied. Regarding this last remark, the resolution of the points (a) and[3] (c) is technically involved and relies on two general results presented in an abstract form in Sections 2 and 3. The first of them is the main novelty of the paper and concerns a new LDP criterion for randomly forced PDEs. Its proof builds on the results of [JNPS15b] and singles out some simple controllability properties that are sufficient for the validity of LDP. This approach makes it possible to treat problems with degenerate noises and is likely to have large scope of applicability, including PDEs studied in [KNS18, Shi19]. In contrast to (a), the proof of (c) does not require development of new techniques and is based on a direct application of a particular case of the general theory presented in [Bog10]. One may anticipate that a successful resolution of the remaining points will require developments of new tools that may find applications beyond specific questions dictated by the entropic fluctuations program.

The paper is organised as follows. In Section 1, we formulate our main results and describe the scheme of their proof. Section 2 is devoted to the problem of large deviations. There we establish a general criterion for the LDP in terms of certain control properties of the system under study. Section 3 deals with the problem of existence of a density and its positivity for images of probability measures under smooth mappings. In Section 4, we study the randomly forced 2D Navier–Stokes system coupled with a Lagrangian particle. Finally, the Appendix gathers some known results used in the main text.

Acknowledgments

This research was supported by the *Agence Nationale de la Recherche* through the grant NONSTOPS (ANR-17-CE40-0006-01, ANR-17-CE40-0006-02, ANR-17-

[2]For the Navier–Stokes system perturbed by a coloured white noise, the level-2 LDP was established in [Ner19].

[3]The part of (c) concerning the relation with the physical notion of transport will be discussed elsewhere.

CE40-0006-03), the CNRS collaboration grant *Fluctuation theorems in stochastic systems*, and the *Initiative d'excellence Paris-Seine*. VJ acknowledges the support of NSERC. The work of CAP has been carried out in the framework of the Labex Archimède (ANR-11-LABX-0033) and of the A*MIDEX project (ANR-11-IDEX-0001-02), funded by the *Investissements d'Avenir* French Government programme managed by the French National Research Agency (ANR). AS acknowledges the support of the MME-DII Center of Excellence (ANR-11-LABX-0023-01) and is grateful to F. Otto for a discussion on the subject of this paper during the conference *SPDEs and Applications-X* in Trento.

Notation

We write $\mathbb{Z}^d$ for the integer lattice in $\mathbb{R}^d$, with the convention $\mathbb{Z} = \mathbb{Z}^1$, and use the notations $\mathbb{N} = \{r \in \mathbb{Z} : r \geq 1\}$, $\mathbb{Z}_\pm = \{r \in \mathbb{Z} : \pm r \geq 0\}$, $[\![a, b]\!] = [a, b] \cap \mathbb{Z}$, and $\mathbb{Z}^d_* = \mathbb{Z}^d \setminus \{0\}$. We denote by $I \subset \mathbb{R}$ a closed interval, by $\mathbb{T}^2 = \mathbb{R}^2/2\pi\mathbb{Z}^2$ the two-dimensional torus, by X a Polish space, and by $\mathcal{H}$ a separable Banach space. We shall always assume that X is endowed with the Borel σ-algebra $\mathcal{B}(X)$, and we write $\mathcal{M}(X)$ for the space of finite signed measures on X and $\mathcal{P}(X) \subset \mathcal{M}(X)$ for the simplex of probability measures. We recall the standard functional spaces of the theory of 2D Navier–Stokes equations, where $s \geq 1$ is assumed to be an integer.

H denotes the space of divergence-free vector fields on $\mathbb{T}^2$ with zero mean value. It is endowed with the usual L^2 norm $\|\cdot\|$.

H^s is the usual Sobolev space of $\mathbb{R}^2$-valued function on $\mathbb{T}^2$ and $V^s = H^s \cap H$. The corresponding norm will be denoted by $\|\cdot\|_s$.

$L^p(I, \mathcal{H})$ stands for the space of Borel-measurable functions $f : I \to \mathcal{H}$ such that

$$\|f\|_{L^p(I,\mathcal{H})} = \left(\int_I \|f(t)\|^p_{\mathcal{H}} \mathrm{d}t \right)^{1/p} < \infty.$$

$C(I, \mathcal{H})$ denotes the space of bounded continuous functions $f : I \to \mathcal{H}$, endowed with the natural norm $\|f\|_{C(I,\mathcal{H})} = \sup_{t\in I} \|f(t)\|_{\mathcal{H}}$.

$\mathcal{X}_s(I)$ is the space of functions $u \in L^2(I, V^{s+1})$ such that $\partial_t u \in L^2(I, V^{s-1})$.

Given a measure $\mu \in \mathcal{P}(X)$ and a map $F(\cdot)$ defined on X, we denote by $F_*(\mu)$ the image of μ under F. If F depends on an additional parameter u, then we shall write $F_*(u, \mu)$ to denote the image of μ for a fixed value of the parameter. For a function $f : X \to \mathbb{R}$ and a measure μ on X, we write $\langle f, \mu\rangle$ for the integral of f against μ. We shall also use the following notation for spaces of functions and measures.

$L^\infty(X)$ is the space of bounded measurable functions $f : X \to \mathbb{R}$ with the supremum norm $\|\cdot\|_\infty$.

$C_b(X)$ is the space of bounded continuous functions $f : X \to \mathbb{R}$ endowed with the norm $\|\cdot\|_\infty$. For a compact space X, we shall simply write $C(X)$.

$L_b(X)$ is the space of Lipschitz continuous functions $f \in C_b(X)$ with the norm

$$\|f\|_L = \|f\|_\infty + \sup_{u \neq v} \frac{|f(u) - f(v)|}{d_X(u,v)}.$$

$C_b(X, \mathcal{H})$ and $L_b(X, \mathcal{H})$ are defined in a similar way.

$\mathcal{M}(X)$ is endowed with the weak* topology which is generated by the functionals $\mu \mapsto \langle f, \mu \rangle$ with $f \in C_b(X)$. The restriction of this topology to $\mathcal{P}(X)$ can be metrised by the *dual-Lipschitz distance* defined as

$$\|\mu - \nu\|_L^* = \sup_{\|f\|_L \leq 1} \big| \langle f, \mu \rangle - \langle f, \nu \rangle \big|.$$

For two measures $\mu, \nu \in \mathcal{P}(X)$, we denote by $\mathrm{Ent}(\mu \,|\, \nu)$ the relative entropy of μ with respect to ν:

$$\mathrm{Ent}(\mu \,|\, \nu) = \sup_{V \in C_b(X)} \big(\langle V, \mu \rangle - \log \langle e^V, \nu \rangle \big) = \int_X \log \frac{\mathrm{d}\mu}{\mathrm{d}\nu} \mathrm{d}\mu,$$

where the second relation holds if μ is absolutely continuous with respect to ν.

1 Main results

1.1 Formulations

Setting of the problem and preliminaries

We consider the motion of a particle in a random time-dependent vector field defined by the 2D Navier–Stokes system. More precisely, we study the Cauchy problem (0.6), (0.7), in which $u = (u_1, u_2)$ is a solution of the system of equations

$$\partial_t u + \langle u, \nabla \rangle u - \nu \Delta u + \nabla \pi = \eta(t,x), \quad \operatorname{div} u = 0, \quad x \in \mathbb{T}^2, \tag{1.1}$$

supplemented with the initial condition

$$u(0,x) = u_0(x). \tag{1.2}$$

Here $\pi = \pi(t,x)$ is the pressure of the fluid, $\nu > 0$ the kinematic viscosity, u_0 is a square-integrable divergence-free vector field on the torus, and η is a random process of the form

$$\eta(t,x) = \sum_{k=1}^{\infty} \eta_k(t - k + 1, x) \mathbb{I}_k(t), \tag{1.3}$$

where $\mathbb{I}_k$ is the indicator function of the interval $[k-1, k)$, and $\{\eta_k\}$ is a sequence of i.i.d. random variables in $L^2([0,1] \times \mathbb{T}^2)$. To simplify the formulas, we assume (which can be done without loss of generality) that η_k's are divergence-free. To ensure the boundedness of the energy of solutions for $t \geq 0$, we require all the functions to have zero mean value with respect to x.

Our aim is to study the large-time asymptotics of the pair (u, y). Recall that the scale of spaces V^s is defined at the end of the Introduction. To ensure the existence of the dynamics for y, we assume that $\eta_k \in L^2(J, V^2)$ almost surely, where $J = [0, 1]$. In this case, almost every trajectory of (1.1) with an initial condition $u_0 \in V^3$ belongs to the space $C(\mathbb{R}_+, V^3)$, and it follows that the Cauchy problem (0.6), (0.7) has a unique solution $y \in C(\mathbb{R}_+, \mathbb{T}^2)$ for any initial point $p \in \mathbb{T}^2$. We shall write

$$\Upsilon(t) = \big(u(t), y(t)\big), \quad t \geq 0, \tag{1.4}$$

for the coupled trajectory and consider it as a continuous curve in $V^3 \times \mathbb{T}^2$. Under the hypotheses imposed on η, the family of trajectories $\{\Upsilon(t)\}$ corresponding to all possible initial conditions does not form a Markov process. However, their restriction to integer times does, and our goal is to study the large-time behaviour of the discrete-time process $\Upsilon_k = \Upsilon(k)$, $k \in \mathbb{Z}_+$.

We now describe the class of random forces η_k we deal with. Denote by $\{e_j\}_{j \in \mathbb{Z}^2_*}$ the L^2 normalised trigonometric basis in the space of divergence-free functions with zero mean value:

$$e_j(x) = E_j^{-1} j^\perp \begin{cases} \cos\langle j, x\rangle & \text{for } j_1 > 0 \text{ or } j_1 = 0,\ j_2 > 0, \\ \sin\langle j, x\rangle & \text{for } j_1 < 0 \text{ or } j_1 = 0,\ j_2 < 0, \end{cases} \tag{1.5}$$

where $j^\perp = (-j_2, j_1)$ and $E_j = \sqrt{2}\pi|j|$ (so that $\|e_j\| = 1$ for any $j \in \mathbb{Z}^2_*$). Note that $\{e_j\}$ is an orthogonal basis in any of the spaces V^s with respect to the inner product $(u, v)_s = (u, (-\Delta)^s v)$. Furthermore, setting $J = [0, 1]$, we fix an orthonormal basis $\{\psi_l\}_{l \geq 1}$ in the space $L^2(J)$ that satisfies the following *Poincaré property*: there are positive numbers C_r and θ such that

$$\|\mathsf{Q}_N g\|_{L^2(J)} \leq C_r N^{-\theta r} \|g\|_{H^r(J)} \quad \text{for } g \in H^r(J),\ N \geq 1, \tag{1.6}$$

where $r \geq 1$ is an arbitrary integer, and Q_N denotes the orthogonal projection in $L^2(J)$ onto the closed subspace spanned by ψ_l, $l \geq N$. For instance, the trigonometric basis $\{e^{2\pi i\, lt}\}_{l \in \mathbb{Z}}$ satisfies Poincaré property with $\theta = 1$. We now formulate our hypothesis on the noise η_k.

(N) *The random variables η_k can be written as*

$$\eta_k(t, x) = \sum_{j \in \mathbb{Z}^2_*} \sum_{l \geq 1} b_j c_l \xi^k_{lj} \psi_l(t) e_j(x), \tag{1.7}$$

where ξ^k_{lj} are independent scalar random variables. Moreover, the law of ξ^k_{lj} possesses an infinitely smooth density ρ_{lj} with support in the interval $[-1, 1]$ such that, for some $\delta > 0$ and all j, l, $\rho_{lj}(r) > 0$ for $|r| < \delta$. Finally, there are positive numbers C_m, c, and $\beta > 1/2$ such that

$$0 < |b_j| \leq C_m |j|^{-m} \quad \textit{for all } m \geq 1, \tag{1.8}$$

$$|c_l| \geq c\, l^{-\beta} \quad \textit{for all } l \geq 1, \quad \sum_{l \geq 1} c_l^2 < \infty. \tag{1.9}$$

Note that if this hypothesis is satisfied, then almost every realisation of η_k belongs to $L^2(J, V^s)$ for any $s \geq 1$. It follows that, with probability 1, the restriction to $J_k = [k-1, k]$ of the solution u for the Navier–Stokes system (1.1) with C^∞ initial condition belongs to $C(J_k, V^s)$ for any $s \geq 1$. Therefore, the resolving operator for the Cauchy problem (0.6), (0.7) can be made as smooth as we wish by choosing s sufficiently large.

Large deviations for empirical measures

Given an interval $I \subset \mathbb{R}$, we define the spaces

$$\mathcal{X}_s(I) = \big\{u \in L^2(I, V^{s+1}) : \partial_t u \in L^2(I, V^{s-1})\big\}, \quad \mathcal{Y}(I) = C(I, \mathbb{T}^2),$$

where $s \geq 1$ is an integer, and note that $\mathcal{X}_s(I)$ is continuously embedded into $C(I, V^s)$. In the case $I = [0, 1]$, we often write $\mathcal{X}_s$ and $\mathcal{Y}$, respectively. For any integer $s \geq 3$ we denote by

$$\boldsymbol{S} : V^s \times \mathbb{T}^2 \times L^2([0, 1], V^s) \to \mathcal{X}_s \times \mathcal{Y}, \quad (u_0, p, \eta) \mapsto (u, y),$$

the resolving operator of the set of equations (1.1), (0.6), (1.2), (0.7). It is well known that, if $s \geq 3$, then $\boldsymbol{S}$ is $(s-2)$-times[4] continuously differentiable in the Fréchet sense. We denote by $S(u_0, p, \eta)$ the value of $\boldsymbol{S}(u_0, p, \eta)$ at $t = 1$. Note that S is a map with range in $V^s \times \mathbb{T}^2$. We write $\boldsymbol{S} = (\boldsymbol{S}^u, \boldsymbol{S}^y)$ and $S = (S^u, S^y)$, with a natural definition of the u- and y-components.

Our first result deals with the level-3 LDP for trajectories issued from an initial point belonging to the domain of attainability from $\{0\} \times \mathbb{T}^2$ (which is also the support of the unique stationary distribution for (1.1), (0.6); see the next subsection on the regularity of laws). Namely, for a fixed $s \geq 3$, let $\mathcal{K}^s \subset L^2([0, 1], V^s)$ be the support of the law of η_k. We define the sets

$$\mathcal{A}^s_0 = \{0\}, \quad \mathcal{A}^s_k = S^u(\mathcal{A}^s_{k-1}, \mathcal{K}^s), \quad k \geq 1, \tag{1.10}$$

and denote by $\mathcal{A}^s$ the closure of the union $\cup_{k\geq 0}\mathcal{A}^s_k$ in the space V^s. The following lemma is easy to establish, and we omit its proof.

Lemma 1.1. *Let Hypothesis (N) be satisfied. Then the following properties hold for any integer $s \geq 3$.*

Compactness. *The set $\mathcal{A}^s$ is compact in V^s and contains the point 0.*

Compatibility. *If $r > s$ is another integer, then $\mathcal{A}^s$ is the closure of $\mathcal{A}^r$ in V^s.*

Invariance. *The set $\mathscr{X}^s := \mathcal{A}^s \times \mathbb{T}^2$ is invariant, that is, $S(\mathscr{X}^s, \mathcal{K}^s) \subset \mathscr{X}^s$.*

[4]The index $s-2$ comes from the fact that $u \in \mathcal{X}_s$ is a continuous function of time with range in $C^r(\mathbb{T}^2)$ for any $r < s-1$, and standard results from the theory of ODEs can ensure only the existence of $s-2$ continuous derivatives for $\boldsymbol{S}^y$.

We now introduce the *empirical measures* for (1.1), (0.6) by the formula

$$\boldsymbol{\nu}_t^{\Upsilon} = t^{-1} \sum_{n=0}^{t-1} \delta_{\boldsymbol{\Upsilon}_n}, \quad t \geq 1, \tag{1.11}$$

where $\Upsilon = (u_0, p)$ is an initial point, $\boldsymbol{\Upsilon}_n = (\Upsilon_k, k \geq n)$, and Υ_k is the value of the solution of (1.1), (0.6), (1.2), (0.7) at $t = k$. Setting $\boldsymbol{X}^s = (\mathscr{X}^s)^{\mathbb{Z}_+}$, it is straightforward to see that if $\Upsilon \in \mathscr{X}^s$, then $\boldsymbol{\Upsilon}_n \in \boldsymbol{X}^s$ for any $n \geq 0$. The following theorem uses standard notions of the theory of large deviations.[5]

Theorem 1.2. *Let Hypothesis (N) be fulfilled and let $s \geq 3$ be an integer. Then the family of empirical measures $\{\boldsymbol{\nu}_t^{\Upsilon}, \Upsilon \in \mathscr{X}^s\}_{t \geq 1}$ satisfies the uniform LDP with some good rate function $\boldsymbol{I}^s : \mathcal{P}(\boldsymbol{X}^s) \to [0, +\infty]$. Moreover, $\boldsymbol{I}^s$ is an affine function on $\mathcal{P}(\boldsymbol{X}^s)$ given by the Donsker–Varadhan entropy formula.*

Regularity of laws for the particle and convergence

We now focus on the law of the particle in more detail. Note that, for any $s \geq 3$, the compact invariant set $\mathscr{X}^s$ carries a stationary measure for the Markov process associated with (1.1), (0.6). More precisely, if (1.4) is a trajectory for (1.1), (0.6), then the vector functions $\Upsilon_k = \Upsilon(k)$ satisfy the relations

$$\Upsilon_k = S(\Upsilon_{k-1}, \eta_k), \quad k \geq 1. \tag{1.12}$$

Since $\{\eta_k\}$ is a sequence of i.i.d. random variables, Eq. (1.12) defines a discrete-time homogeneous Markov process in $V^s \times \mathbb{T}^2$ whose transition function has the form

$$P_1(\Upsilon, \cdot) = S_*(\Upsilon, \ell), \tag{1.13}$$

where ℓ stands for the law of η_k, and the right-hand side denotes the image of ℓ under the mapping $\zeta \mapsto S(\Upsilon, \zeta)$. By Lemma 1.1, the set $\mathscr{X}^s$ is invariant in the sense that $P_1(\Upsilon, \mathscr{X}^s) = 1$ for any $\Upsilon \in \mathscr{X}^s$. In what follows we consider the restriction of the Markov process defined by (1.12) to $\mathscr{X}^s$ and denote by $\mathfrak{P}_k$ and $\mathfrak{P}_k^*$ the corresponding Markov operators acting on the spaces $C(\mathscr{X}^s)$ and $\mathcal{P}(\mathscr{X}^s)$, respectively. Since $\mathscr{X}^s$ is compact, there is at least one stationary measure $\mathfrak{M} \in \mathcal{P}(\mathscr{X}^s)$. Applying Theorem 5.1, one can prove that $\mathfrak{M}$ is the unique stationary measure for (1.12). Let us note that the uniqueness of a stationary distribution was proved in [BBP18] for the coupled system (1.1), (0.6) with a coloured white noise η; however, their approach is not applicable in our situation since it is based on the strong Feller property and requires the noise to be rough in the space variables.

A simple argument based on the uniqueness of the stationary measure proves that $\mathfrak{M}$ is independent of s. Moreover, another short computation shows that[6] $\mathfrak{M} = \mu \otimes \lambda$, where μ is the unique stationary measure for (1.1) and λ is the normalised Lebesgue measure on $\mathbb{T}^2$. We shall denote by $\boldsymbol{\mathfrak{M}} \in \mathcal{P}(\boldsymbol{X}^s)$ the

[5] For their definitions we refer the reader to Section 2.1.

[6] Note, however, that this product structure is not preserved on the level of path measures.

corresponding path measure and by $\boldsymbol{\mu} \in \mathcal{P}(\boldsymbol{\mathcal{A}})$ and $\boldsymbol{\lambda} \in \mathcal{P}(\boldsymbol{T})$ its projections to the u- and y-components, where $\boldsymbol{\mathcal{A}} = \mathcal{A}^{\mathbb{Z}_+}$ and $\boldsymbol{T} = (\mathbb{T}^2)^{\mathbb{Z}_+}$. Similarly, given an initial point $\Upsilon \in \mathscr{X}^s$, we shall denote by $\mathfrak{M}^\Upsilon \in \mathcal{P}(\boldsymbol{X}^s)$ the path measure of the trajectory for (1.1), (0.6) issued from Υ, by $\mathfrak{M}_t^\Upsilon \in \mathcal{P}(\mathscr{X}^s)$ its projection the tth component, and by $\boldsymbol{\mu}^\Upsilon \in \mathcal{P}(\boldsymbol{\mathcal{A}})$ and $\boldsymbol{\lambda}^\Upsilon \in \mathcal{P}(\boldsymbol{T})$ its projections to the u- and y-components, respectively. Finally, given an integer interval $I \subset \mathbb{Z}_+$, we denote by $\boldsymbol{\lambda}_I^\Upsilon \in \mathcal{P}(\mathbb{T}^{2|I|})$ the projection of $\boldsymbol{\lambda}^\Upsilon$ to I and define $\boldsymbol{\lambda}_I$ similarly. We shall write $\boldsymbol{\lambda}_t^\Upsilon$ and $\boldsymbol{\lambda}_t$ for $I = [\![1, t]\!]$.

Theorem 1.3. *Suppose that Hypothesis (N) is satisfied. Then the following holds for any integer $t \geq 2$.*

Regularity. *For any $\Upsilon \in \mathscr{X}^3$, the measure $\boldsymbol{\lambda}_{[\![2,t]\!]}^\Upsilon$ has a density $\rho_{[\![2,t]\!]}^\Upsilon$ that belongs to $C^\infty(\mathbb{T}^{2(t-1)})$, and the function $\Upsilon \mapsto \rho_{[\![2,t]\!]}^\Upsilon$ is Lipschitz continuous from $\mathscr{X}^3$ to $C^k(\mathbb{T}^{2(t-1)})$ for any $k \geq 1$. Moreover, the measure $\boldsymbol{\lambda}_t$ has a density $\rho_t \in C^\infty(\mathbb{T}^{2t})$.*

Convergence. *There is $\gamma > 0$ such that, for any integer $k \geq 1$, we have*

$$\sup_{\Upsilon \in \mathscr{X}^3} \left\| \rho_{[\![n+1,n+t]\!]}^\Upsilon - \rho_t \right\|_{C^k(\mathbb{T}^{2t})} \leq C_{tk} e^{-\gamma n}, \quad n \geq 1, \tag{1.14}$$

where the constant $C_{tk} > 0$ does not depend on n.

Let us note that if $\Upsilon \in \mathscr{X}^s$ is not infinitely smooth, there is no reason for ρ_t^Υ to be C^∞ even for $t = 1$. Indeed, as it was mentioned in footnote 4, the map $\eta \mapsto S^y(\Upsilon, \eta)$ acting from $L^2([0,1], V^s)$ to $\mathbb{T}^2$ possesses only finite regularity, unless $\Upsilon \in \mathscr{X}^s$ is infinitely smooth. Therefore, without any regularisation mechanism, the image of a measure under the action of $S^y(\Upsilon, \cdot)$ does not need to have a smooth density. On the other hand, the following remark about finite regularity will be important in the definition of the entropy production.

Remark 1.4. The proof of Theorem 1.3 will imply that, for any integer $k \geq 0$, there is $s \geq 3$ such that, for any $t \geq 1$ and $\Upsilon \in \mathscr{X}^s$, the measure $\boldsymbol{\lambda}_t^\Upsilon$ has a density $\rho_t^\Upsilon \in C^k(\mathbb{T}^{2t})$. Moreover, the mapping $\Upsilon \mapsto \rho_t^\Upsilon$ is Lipschitz continuous from $\mathscr{X}^s$ to $C^k(\mathbb{T}^{2t})$.

Strict positivity of densities

To ensure strict positivity of the densities ρ_t and to derive a uniform bound on the mean entropy production in time t, we need to replace the random force on the right-hand side of (1.1) by $\eta^a := a\eta$, where $a > 0$ is a large parameter. We shall denote by ρ_t^a the densities corresponding to the resulting equation.

Theorem 1.5. *Suppose that Hypothesis (N) is satisfied. Then there is $a_0 > 0$ such that the following holds for any $a \geq a_0$.*

Strict positivity. *The functions ρ_t^a are bounded below by positive numbers.*

Uniform bound on the entropy production. *There is $C > 0$ such that the entropy production defined by (0.9) satisfies the inequality*

$$\left|t^{-1}\sigma_t(y_1,\dots,y_t)\right| \le C \quad \textit{for all } (y_1,\dots,y_t) \in \mathbb{T}^{2t},\ t \ge 1. \tag{1.15}$$

As we shall describe in the next section, the uniform bound on the entropy production is an easy consequence of the strict positivity of $\rho_1^{\Upsilon}(y)$. The proof given in Section 4.3 will imply that, for this theorem to be true, it suffices to have a large parameter in front of finitely many Fourier modes in x. On the other hand, the following simple observation shows that $\rho_1^{\Upsilon}(y)$ cannot be strictly positive for any $\Upsilon \in \mathscr{X}^s$ and $y \in \mathbb{T}^2$, unless the noise is sufficiently large. Indeed, suppose that $\Upsilon = (0,p)$ and $|y-p|$ is of order 1. In this case, the size of the velocity field on the interval $[0,1]$ can be bounded by the norm of the noise. If the latter is of order $\varepsilon > 0$, then the particle can travel a distance no larger than $C\varepsilon$, and so $\rho_1^{\Upsilon}(y) = 0$ for $|y-p| > C\varepsilon$.

1.2 Schemes of the proofs

Theorem 1.2

In Section 2 we shall derive a sufficient condition for the validity of LDP in the context of the Markovian RDS (1.12). Apart from the regularity of S and a decomposability hypothesis on the law of the random noise, this criterion requires two properties: approximate controllability of the nonlinear system by controls belonging to the support of the law η and the density of the image of the linearised operator; see (AC) and (ACL). The verification of these two properties is based on essentially the same idea, which we briefly outline here, leaving the details for Section 4.1. Note that some related problems on the control of a particle by the vector field appeared in the papers [Ner11, Ner15, BBP18], and our proof uses some ideas from these articles.

Suppose we wish to prove that a point $\Upsilon_0 = (0,p) \in \mathscr{X}^s$ can be exactly steered to any point $\widehat{\Upsilon} = (0,\hat{p})$ that is sufficiently close to Υ_0. Let us set

$$U_1(x) = (\cos x_2, 0), \quad U_2(x) = (0, \cos x_1), \quad \gamma(t) = \big(1-\alpha(t)\big)\,p + \alpha(t)\hat{p}, \tag{1.16}$$

where $\alpha \in C^\infty(\mathbb{R})$ is such that $\alpha(t) = 0$ for $t \le 1/3$ and $\alpha(t) = 1$ for $t \ge 2/3$. Writing

$$\dot{\gamma}(t) = \dot{\alpha}(t)(\hat{p}-p) = \big(\varphi_1(t), \varphi_2(t)\big), \tag{1.17}$$

we define the functions

$$u(t,x) = \varphi_1(t)U_1\big(x-\gamma(t)\big) + \varphi_2(t)U_2\big(x-\gamma(t)\big), \quad y(t) = \gamma(t), \tag{1.18}$$

where $t \in [0,1]$. Then the vector function $\Upsilon = (u,y)$ is infinitely smooth, coincides with $(0,p)$ and $(0,\hat{p})$ at the endpoints of the interval $[0,1]$, and satisfies Eqs. (1.1), (0.6) with

$$\eta(t) = \Pi g(t), \quad g(t) = \partial_t u + \langle u, \nabla\rangle u - \nu\Delta u, \tag{1.19}$$

where $\Pi : L^2(\mathbb{T}^2, \mathbb{R}^2) \to H$ stands for Leray's projection. It is straightforward to check that g can be written as

$$(\Pi g)(t,x) = \sum_{j\in\Lambda} \alpha_j(t) e_j(x), \tag{1.20}$$

where $\Lambda = \{j = (j_1, j_2) \in \mathbb{Z}^2_* : |j_1| + |j_2| \le 2\}$, the trigonometric basis $\{e_j\}$ is defined by (1.5), and α_j's are smooth functions of $t \in [0,1]$ whose C^r norms are proportional to $|\hat{p} - p|$ for any $r \ge 1$. We claim that Πg is in the support $\mathcal{K}^s$ of $\mathcal{D}(\eta_k)$, provided that $|\hat{p} - p| \ll 1$. Indeed, it follows from (N) that $\mathcal{K}^s$ contains any function of the form

$$h(t,x) = \sum_{j\in\Lambda} \sum_{l=1}^{\infty} h_{lj} \psi_l(t) e_j(x), \tag{1.21}$$

where the coefficients satisfy the inequality $|h_{lj}| \le \varepsilon l^{-\beta}$ with $\varepsilon \ll 1$. Since α_j's are infinitely smooth, it follows from (1.6) that the coefficients α_{jl} of the expansion of α_j in the basis $\{\psi_l\}$ decay faster than any negative degree of l. Since they are bounded by a number propositional to $|\hat{p} - p|$, we conclude that $\Pi g \in \mathcal{K}^s$, provided that $|\hat{p} - p| \ll 1$.

Theorem 1.3

In Section 3, we present a sufficient condition for the existence of a regular density for the image of a probability measure under a smooth mapping; see Theorem 3.1. Roughly speaking, it says that if a smooth map F with range in a finite-dimensional manifold is such that its derivative is surjective everywhere, then the image of a probability measure ℓ has smooth density, provided that ℓ is regular in an appropriate sense. Measures satisfying Hypothesis (N) do possess the required regularity property, and the position of the particle can be written as a smooth function F of the noise and the initial condition of the system. The fact that the derivative of F is surjective will follow from the density of the image for the linearised operator. This will establish the existence of $\rho^{\Upsilon}_{[\![2,t]\!]}$.

To prove convergence (1.14), we first note that the sequence of measures $\{\mathfrak{M}^{\Upsilon}_k\}_{k\ge1}$ converges, as $k \to \infty$, to $\mathfrak{M}$ exponentially fast in the dual-Lipschitz norm; this is established in Theorem 5.1. Let us fix any $s \ge 3$, set $\mathscr{E} = L^2(J, V^s)$, and introduce a map

$$F^t : \mathscr{X}^3 \times \underbrace{\mathscr{E} \times \cdots \times \mathscr{E}}_{t \text{ times}} \to \mathbb{T}^{2t}$$

that takes $(\Upsilon, \eta_1, \dots, \eta_t)$ to $(y_1, \dots, y_t)$, where y_k is the y-component of the trajectory Υ_k for (1.12). In this case, we can write

$$\boldsymbol{\lambda}^{\Upsilon}_{[\![n+1,n+t]\!]} = \mathbb{E}\, F^t_*(\Upsilon_n, \underbrace{\ell \otimes \cdots \otimes \ell}_{t \text{ times}}). \tag{1.22}$$

Now note that ρ_t^{Υ} is the density of $F_*^t(\Upsilon, \ell \otimes \cdots \otimes \ell)$ with respect to the Lebesgue measure on $\mathbb{T}^{2t}$. It follows that

$$\rho^{\Upsilon}_{[\![n+1,n+t]\!]}(y_1,\dots,y_t) = \int_{\mathscr{X}^s} \rho_t^{\upsilon}(y_1,\dots,y_t)\,\mathfrak{M}_n^{\Upsilon}(\mathrm{d}\upsilon). \tag{1.23}$$

Since $\rho_t^{\upsilon}(y)$ is Lipschitz continuous in υ, together with all its derivatives in y, this will imply the required convergence (1.14).

Theorem 1.5

As it was established in Theorem 1.3 and Remark 1.4, if an integer $s \geq 3$ is sufficiently large, then for any $\Upsilon \in \mathscr{X}^s$ the projection of the transition function $P_1(\Upsilon, \cdot)$ to the y-component possesses a density $\rho_1^{\Upsilon}(y)$,

$$P_1^y(\Upsilon, \mathrm{d}y) = \rho_1^{\Upsilon}(y)\,\mathrm{d}y, \tag{1.24}$$

and the mapping $(\Upsilon, y) \mapsto \rho_1^{\Upsilon}(y)$ is continuous from $\mathscr{X}^s$ to $C(\mathbb{T}^2)$. It follows that $\rho_1^{\Upsilon}(y)$ is continuous in (Υ, y) and, by the compactness of $\mathscr{X}^s \times \mathbb{T}^2$, there is $M > 0$ such that

$$\rho_1^{\Upsilon}(y) \leq M \quad \text{for all } \Upsilon \in \mathscr{X}^s,\ y \in \mathbb{T}^2. \tag{1.25}$$

By the Kolmogorov–Chapman relation, for an arbitrary non-negative function $f : \mathbb{T}^{2t} \to \mathbb{R}$, we have

$$\begin{aligned}
\langle f, \boldsymbol{\lambda}_t \rangle &= \int\limits_{\mathscr{X}^s(t+1)} f(\boldsymbol{y}^t)\,\mathfrak{M}(\mathrm{d}\Upsilon) P_1(\Upsilon, \mathrm{d}\Upsilon_1) \cdots P_1(\Upsilon_{t-1}, \mathrm{d}\Upsilon_t) \\
&= \int\limits_{\mathscr{X}^s(t)\times\mathbb{T}^2} f(\boldsymbol{y}^t) \rho_1^{\Upsilon_{t-1}}(y_t)\,\mathfrak{M}(\mathrm{d}\Upsilon) P_1(\Upsilon, \mathrm{d}\Upsilon_1) \cdots P_1(\Upsilon_{t-2}, \mathrm{d}\Upsilon_{t-1})\,\mathrm{d}y_t \\
&\leq M \int\limits_{\mathscr{X}^s(t)\times\mathbb{T}^2} f(\boldsymbol{y}^t)\,\mathfrak{M}(\mathrm{d}\Upsilon) P_1(\Upsilon, \mathrm{d}\Upsilon_1) \cdots P_1(\Upsilon_{t-2}, \mathrm{d}\Upsilon_{t-1})\,\mathrm{d}y_t,
\end{aligned}$$

where $\boldsymbol{y}^t = (y_1, \dots, y_t) \in \mathbb{T}^{2t}$, $\mathscr{X}^s(t)$ denotes the t-fold product of the space $\mathscr{X}^s$, and we used (1.24) and (1.25). Iterating this argument and using the relation $\mathfrak{M}(\mathscr{X}) = 1$, we derive

$$\langle f, \boldsymbol{\lambda}_t \rangle \leq M^t \int_{\mathbb{T}^{2t}} f(y_1, \dots, y_t)\,\mathrm{d}y_1 \dots \mathrm{d}y_t.$$

Since $f \geq 0$ was arbitrary, it follows that

$$\rho_t(y_1, \dots, y_t) \leq M^t \quad \text{for any } (y_1, \dots, y_t) \in \mathbb{T}^{2t}. \tag{1.26}$$

Note that the upper bound for the density does not require any additional hypotheses on the noise.

We now turn to the lower bound. Suppose we have proved that

$$\rho_1^{\Upsilon}(y) > 0 \quad \text{for all } \Upsilon \in \mathscr{X}^s,\ y \in \mathbb{T}^2. \tag{1.27}$$

Then, by continuity and compactness, we can find $m > 0$ such that $\rho_1^{\Upsilon}(y) \geq m$ for $\Upsilon \in \mathscr{X}^s$, $y \in \mathbb{T}^2$. Repeating the above argument, one gets that, for any non-negative function $f : \mathbb{T}^{2t} \to \mathbb{R}$,

$$\langle f, \boldsymbol{\lambda}_t \rangle \geq m^t \int_{\mathbb{T}^{2t}} f(y_1, \dots, y_t)\, \mathrm{d}y_1 \dots \mathrm{d}y_t,$$

and so it follows that

$$\rho_t(y_1, \dots, y_t) \geq m^t \quad \text{for any } (y_1, \dots, y_t) \in \mathbb{T}^{2t}. \tag{1.28}$$

Inequalities (1.26) and (1.28) allow to define the entropy production in time t by relation (0.9) and to derive the estimate (1.15) for its time-average.

The above elementary argument reduces the proof of Theorem 1.5 to the verification of (1.27). Theorem 3.2 gives a sufficient condition for the positivity of the density for the image of a probability measure ℓ under a finite-dimensional smooth map. Roughly speaking, it says that if a point $\hat{p}$ has a pre-image in the "interior" of the support of ℓ, then the density is strictly positive at $\hat{p}$. Hence, the proof further reduces to a problem of exact controllability for the Navier–Stokes system coupled to the Lagrangian particle. We shall show in Section 4.3 that this can be established by modifying the above scheme used in the proof of Theorem 1.2, provided that the noise contains a large parameter in front of finitely many Fourier modes in the space variables.

Let us also mention that the above argument cannot be applied to the full system since the transition functions corresponding to different initial points $\Upsilon = (u, p) \in \mathscr{X}^s$ are not equivalent. It is the integration with respect to $u \in \mathscr{X}^s$ that removes this singularity and allows one to prove the equivalence of the (projections of) transition probabilities. Moreover, we conjecture that the laws of the forward and backward stationary processes of the full system (1.1), (0.6) are not equivalent. Indeed, for the (linear) Stokes system perturbed by a spatially regular white noise, after integrating out the p-variable, one gets a Gaussian process for which there exist necessary and sufficient conditions (in terms of the noise) for the equivalence of forward and backward laws; cf. Theorem 7.2.1 in [DZ96]. In this case, it is not difficult to construct a noise for which the two laws are singular.

2 Large deviations via controllability

2.1 Formulation of the result

Let $\mathcal{H}$ be a separable Hilbert space, let $\mathscr{Y}$ be a compact Riemannian manifold, let $\mathscr{H} = \mathcal{H} \times \mathscr{Y}$ be the product space with natural projections $\Pi_{\mathcal{H}}$ and $\Pi_{\mathscr{Y}}$ to its components, and let $\mathscr{E}$ be a separable Banach space. We fix a continuous

mapping $S : \mathscr{H} \times \mathscr{E} \to \mathscr{H}$ and consider the random dynamical system (1.12) in which $\{\eta_k\}$ is a sequence of i.i.d. random variables in $\mathscr{E}$. We shall denote by $\mathcal{K} \subset \mathscr{E}$ the support of the law of η_k and assume that there is a compact subset $\mathcal{A} \subset \mathcal{H}$ such that $\mathscr{X} := \mathcal{A} \times \mathscr{Y}$ is invariant for (1.12) ($S(\mathscr{X} \times \mathcal{K}) \subset \mathscr{X}$). We impose the following three hypotheses on the mapping S.

(R) *There is a Banach space $\mathcal{V}$ compactly embedded into $\mathcal{H}$ such that the image of S is contained in $\mathscr{V} := \mathcal{V} \times \mathscr{Y}$, the mapping $S : \mathscr{H} \times \mathscr{E} \to \mathscr{V}$ is twice continuously differentiable, and its derivatives are bounded on bounded subsets. Moreover, there is $\overline{\Upsilon} \in \mathscr{X}$ such that $S(\overline{\Upsilon}, 0) = \overline{\Upsilon}$.*

(AC) *For any $\varepsilon > 0$, there is an integer $n \geq 1$ such that, for any initial point $\Upsilon \in \mathscr{X}$ and any target $\widehat{\Upsilon} \in \mathscr{X}$, one can find controls $\zeta_1, \dots, \zeta_n \in \mathcal{K}$ satisfying the inequality*

$$d_{\mathscr{H}}\big(S_n(\Upsilon; \zeta_1, \dots, \zeta_n), \widehat{\Upsilon}\big) \leq \varepsilon, \tag{2.1}$$

where $S_n(\Upsilon; \eta_1, \dots, \eta_n)$ stands for the vector Υ_n defined by relations (1.12) with $\Upsilon_0 = \Upsilon$.

(ACL) *For any $\Upsilon \in \mathscr{X}$ and $\eta \in \mathcal{K}$, the derivative $(D_\eta S)(\Upsilon, \eta) : \mathscr{E} \to \mathcal{H} \times T_y\mathscr{Y}$, with $y = S(\Upsilon, \eta)$, has a dense image.*

In applications to randomly forced PDEs, the mapping S is the time-1 shift along the trajectories of the system. The first part of Hypothesis (R) is a regularisation property of the flow, and the second part asserts that the unperturbed dynamics has at least one fixed point. Hypothesis (AC) is the standard property of global approximate controllability, with control functions in the support of the noise with no restriction imposed on the time of control. Hypothesis (ACL) is a similar property for the linearised equation, but it allows for a larger control space and requires the time of control to be fixed. These two properties are often satisfied if the support of the driving noise is sufficiently large.

We shall assume, in addition, that the noise has a decomposable structure in the following sense.

(D) *The support of ℓ is compact, and there are two sequences of closed subspaces $\{\mathscr{F}_n\}$ and $\{\mathscr{G}_n\}$ in $\mathscr{E}$ such that $\dim \mathscr{F}_n < \infty$ and $\mathscr{F}_n \subset \mathscr{F}_{n+1}$ for any $n \geq 1$, the union $\cup_n \mathscr{F}_n$ is dense in $\mathscr{E}$, and the following properties hold.*

- *The space $\mathscr{E}$ is the direct sum of $\mathscr{F}_n$ and $\mathscr{G}_n$, and the norms of the corresponding projections P_n and Q_n are bounded uniformly in $n \geq 1$.*
- *The measure ℓ is the product of its projections $\mathsf{P}_{n*}\ell$ and $\mathsf{Q}_{n*}\ell$ for any $n \geq 1$. Moreover, $\mathsf{P}_{n*}\ell$ has C^1-smooth density with respect to the Lebesgue measure on $\mathscr{F}_n$.*

Let us note that this condition implies, in particular, that the sequence of projections $\{\mathsf{P}_n\}$ converges to the identity operator in $\mathscr{E}$ in the strong operator

topology. In what follows, we deal with the restriction of (1.12) to the invariant set $\mathscr{X}$. We introduce the empirical measures of trajectories by the formula (1.11), in which $\boldsymbol{\Upsilon}_n = (\Upsilon_k, k \geq n)$ and $\Upsilon_k = S_k(\Upsilon; \eta_1, \dots, \eta_k)$. Thus, for each $\mathscr{X}$-valued random variable Υ, the sequence $\{\boldsymbol{\nu}_t^{\Upsilon}\}$ consists of random probability measures on the product space $\boldsymbol{X} := \mathscr{X}^{\mathbb{Z}_+}$.

To formulate the main result of this section, we first recall some definitions. The spaces $\boldsymbol{X}$ and $\mathcal{P}(\boldsymbol{X})$ are endowed with the Tikhonov and weak* topologies and the corresponding Borel σ-algebras. A mapping $\boldsymbol{I} : \mathcal{P}(\boldsymbol{X}) \to [0, +\infty]$ is called a *good rate function* if it is convex and has compact level sets. The latter property reduces to the lower semicontinuity of $\boldsymbol{I}$ since $\mathcal{P}(\boldsymbol{X})$ is a compact space. We shall say that the sequence $\{\boldsymbol{\nu}_t^{\Upsilon}\}$ satisfies the *uniform LDP* with the rate function $\boldsymbol{I}$ if

$$\begin{aligned} -\boldsymbol{I}(\dot{\Gamma}) &\leq \liminf_{t\to\infty} t^{-1} \log \inf_{\Upsilon\in\mathscr{X}} \mathbb{P}\{\boldsymbol{\nu}_t^{\Upsilon} \in \Gamma\} \\ &\leq \limsup_{t\to\infty} t^{-1} \log \sup_{\Upsilon\in\mathscr{X}} \mathbb{P}\{\boldsymbol{\nu}_t^{\Upsilon} \in \Gamma\} \leq -\boldsymbol{I}(\overline{\Gamma}) \end{aligned} \tag{2.2}$$

for any Borel subset $\Gamma \subset \mathcal{P}(\boldsymbol{X})$, where $\dot{\Gamma}$ and $\overline{\Gamma}$ stand for the interior and closure of Γ, and $\boldsymbol{I}(A)$ is the infimum of $\boldsymbol{I}$ over A. In view of the Markov property, if $\{\boldsymbol{\nu}_t^{\Upsilon}\}$ satisfies the *uniform LDP*, then inequality (2.2) remains valid if the infimum and supremum are taken over all $\mathscr{X}$-valued random variables Υ independent of the sequence $\{\eta_k\}$.

A measure $\boldsymbol{\lambda} \in \mathcal{P}(\boldsymbol{X})$ is said to be *shift-invariant* if it is invariant under the mapping $t \mapsto t+1$. The set of all shift-invariant measures is denoted by $\mathcal{P}_s(\boldsymbol{X})$. By Kolmogorov's theorem, any shift-invariant measure can be extended in a unique manner to a shift-invariant measure on $\mathscr{X}^{\mathbb{Z}}$, and we use the same notation for the extended measure. Finally, given a shift-invariant measure $\boldsymbol{\lambda} \in \mathcal{P}(\boldsymbol{X})$, we denote by $\boldsymbol{\lambda}_-$ its projection to $\boldsymbol{X}_- := \mathscr{X}^{\mathbb{Z}_-}$, and by $\{\lambda(\boldsymbol{\Upsilon}, \cdot), \boldsymbol{\Upsilon} \in \mathscr{X}^{\mathbb{Z}_-}\}$ the projection to the first component of the regular conditional probability of $\boldsymbol{\lambda}$ with respect to its projection to $\boldsymbol{X}_-$.

Theorem 2.1. *Suppose that Hypotheses (R), (AC), (ACL), and (D) hold for the random dynamical system (1.12). Then the following holds.*

Uniform LDP. *The empirical measures $\{\boldsymbol{\nu}_t^{\Upsilon}\}_{t\geq 1}$ satisfy the uniform LDP with a good rate function $\boldsymbol{I} : \mathcal{P}(\boldsymbol{X}) \to [0, +\infty]$. In particular, the LDP holds for the empirical measures of a stationary process.*

Rate function. *The rate function $\boldsymbol{I}$ is affine and is given by the Donsker–Varadhan entropy formula:*

$$\boldsymbol{I}(\boldsymbol{\lambda}) = \begin{cases} \displaystyle\int_{\boldsymbol{X}_-} \operatorname{Ent}\big(\lambda(\boldsymbol{\Upsilon}, \cdot) \,|\, P_1(\Upsilon_0, \cdot)\big)\, \boldsymbol{\lambda}_-(\mathrm{d}\boldsymbol{\Upsilon}) & \text{if } \boldsymbol{\lambda} \in \mathcal{P}_s(\boldsymbol{X}), \\ +\infty & \text{otherwise}, \end{cases} \tag{2.3}$$

where $\operatorname{Ent}(\mu \,|\, \nu)$ is the relative entropy of μ with respect to ν, and $P_1(\Upsilon, \cdot)$ is the transition function for the Markov process defined by (1.12).

The above theorem is applicable to various parabolic-type PDEs with a smooth random force. In this context, the case when all the Fourier modes are forced was studied in [JNPS15b, Ner19] (see also [Gou07, WX18] for the case of an irregular noise). The scope of applicability of Theorem 2.1 is much larger, allowing for treatment of PDEs with very degenerate noise, such as those studied in [Shi15, KNS18]. Furthermore, even though the Donsker–Varadhan formula (2.3) is by now very well known (see [DS89, Section 5.4] or [DZ00, Section 6.5]), to the best of our knowledge, all available proofs deal with the case of *strong* Feller Markov processes. Our proof presented in Section 2.5 is valid for Markov processes with Feller property in a compact metric space, and its extension to the non-compact case does not encounter any difficulties.

The proof of Theorem 2.1 is based on Kifer's criterion for LDP and a result on the asymptotoic behaviour of generalised Markov semigroups. The scheme of the proof is presented in Section 2.2, and the details are given in Sections 2.3– 2.5.

Remark 2.2. It is tempting to use the explicit formula (2.3) for the large deviations rate function to derive the level-3 fluctuation relation (0.2). Namely, for an integer $k \in \mathbb{Z}$ and a measure $\boldsymbol{\lambda} \in \mathcal{P}(\mathscr{X}^{\mathbb{Z}})$, we denote by $\mathbb{Z}_k$ the set of the integers not exceeding k and by $\boldsymbol{\lambda}_-^k$ the projection of $\boldsymbol{\lambda}$ to $\mathscr{X}^{\mathbb{Z}_k}$, so that $\mathbb{Z}_0 = \mathbb{Z}_-$ and $\boldsymbol{\lambda}_-^0 = \boldsymbol{\lambda}_-$. Using the explicit formula for the relative entropy in terms of densities and the relation $\boldsymbol{\lambda}_-^1(\mathrm{d}\boldsymbol{\Upsilon}, \mathrm{d}\Upsilon_1) = \boldsymbol{\lambda}_-(\mathrm{d}\boldsymbol{\Upsilon})\lambda(\boldsymbol{\Upsilon}, \mathrm{d}\Upsilon_1)$, for any $\boldsymbol{\lambda} \in \mathcal{P}_s(\boldsymbol{X})$ we can write[7]

$$\begin{aligned}\boldsymbol{I}(\boldsymbol{\lambda}) &= \int_{\boldsymbol{X}_-} \Big\{ \int_{\mathscr{X}} \log \frac{\lambda(\boldsymbol{\Upsilon}, \mathrm{d}\Upsilon_1)}{P_1(\Upsilon_0, \mathrm{d}\Upsilon_1)} \lambda(\boldsymbol{\Upsilon}, \mathrm{d}\Upsilon_1) \Big\} \boldsymbol{\lambda}_-(\mathrm{d}\boldsymbol{\Upsilon}) \\ &= \int_{\boldsymbol{X}_- \times \mathscr{X}} \log \frac{\boldsymbol{\lambda}_-(\mathrm{d}\boldsymbol{\Upsilon})\lambda(\boldsymbol{\Upsilon}, \mathrm{d}\Upsilon_1)}{\boldsymbol{\lambda}_-(\mathrm{d}\boldsymbol{\Upsilon})P_1(\Upsilon_0, \mathrm{d}\Upsilon_1)} \boldsymbol{\lambda}_-^1(\mathrm{d}\boldsymbol{\Upsilon}, \mathrm{d}\Upsilon_1) = \mathrm{Ent}\big(\boldsymbol{\lambda}_-^1 \,|\, \boldsymbol{\lambda}_- \otimes P_1\big),\end{aligned}$$

where $\frac{\mu(\mathrm{d}x)}{\nu(\mathrm{d}x)}$ denotes the density of μ with respect to ν, and $\boldsymbol{\lambda}_- \otimes P_1$ stands for the measure acting on a function F by the formula

$$\langle F, \boldsymbol{\lambda}_- \otimes P_1 \rangle = \int_{\boldsymbol{X}_-} \Big\{ \int_{\mathscr{X}} F(\boldsymbol{\Upsilon}, \Upsilon_1) P_1(\Upsilon_0, \mathrm{d}\Upsilon_1) \Big\} \boldsymbol{\lambda}_-(\mathrm{d}\boldsymbol{\Upsilon}).$$

Now let $\theta : \mathscr{X}^{\mathbb{Z}} \to \mathscr{X}^{\mathbb{Z}}$ be the natural time reversal taking $(\Upsilon_k, k \in \mathbb{Z})$ to $(\Upsilon_{-k}, k \in \mathbb{Z})$ and let $\boldsymbol{\theta} : \mathcal{P}(\mathscr{X}^{\mathbb{Z}}) \to \mathcal{P}(\mathscr{X}^{\mathbb{Z}})$ be the associated involution in the space of measures. Assuming that $P_1(\Upsilon_0, \mathrm{d}\Upsilon_1)$ has a positive density $\rho(\Upsilon_0, \Upsilon_1)$ with respect to a reference measure, using the above formula for $\boldsymbol{I}$, and carrying out some simple transformations, we get

$$\begin{aligned}\boldsymbol{I}(\boldsymbol{\lambda} \circ \boldsymbol{\theta}) - \boldsymbol{I}(\boldsymbol{\lambda}) &= \mathrm{Ent}\big((\boldsymbol{\lambda} \circ \boldsymbol{\theta})_-^1 \,|\, (\boldsymbol{\lambda} \circ \boldsymbol{\theta})_- \otimes P_1\big) - \mathrm{Ent}\big(\boldsymbol{\lambda}_-^1 \,|\, \boldsymbol{\lambda}_- \otimes P_1\big) \\ &= \int_{\boldsymbol{X}} \log \frac{\rho(\Upsilon_0, \Upsilon_1)}{\rho(\Upsilon_1, \Upsilon_0)} \boldsymbol{\lambda}(\mathrm{d}\boldsymbol{\Upsilon}). \qquad (2.4)\end{aligned}$$

[7] It is easy to give a rigorous meaning to the formal expressions used in the calculations below. Since these calculations do not play a role in this work, we omit the details.

Hence, denoting by $\sigma(\boldsymbol{\Upsilon})$ the integrand in (2.4), we obtain the level-3 fluctuation relation (0.2), in which $\mathrm{ep}(\boldsymbol{\lambda})$ is the mean value of σ with respect to $\boldsymbol{\lambda}$.

Unfortunately, the above argument is purely formal since the logarithmic ratio in (2.4) may not be well defined, as is expected in the case of the Navier–Stokes system with a smooth noise. Thus, the validity of level-3 LDP is not sufficient for the fluctuation relation (0.2) to be true. On the other hand, the above argument can be justified under some additional hypotheses on the map S and the driving noise η_k; see [JNPS15a].

2.2 General scheme of the proof of Theorem 2.1

Reduction to LDP for finite segments

The first step in the proof of Theorem 2.1 consists of an application of the Dawson–Gärtner theorem, which allows one to reduce the required result to the LDP for the sequence

$$\boldsymbol{\nu}_t^{\Upsilon}(r) = \frac{1}{t}\sum_{k=0}^{t-1} \delta_{\boldsymbol{\Upsilon}_k^r}, \tag{2.5}$$

where $\boldsymbol{\Upsilon}_k^r = [\Upsilon_k, \dots, \Upsilon_{k+r-1}]$, and $\{\Upsilon_k\}$ is the trajectory defined by (1.12) with $\Upsilon_0 = \Upsilon$. Thus, $\{\boldsymbol{\nu}_t^{\Upsilon}(r)\}$ is a sequence of random probability measures on the r-fold product $\mathscr{X}(r)$ of the space $\mathscr{X}$. In view of Theorem 4.6.1 in [DZ00], if for all $r \geq 1$ the sequence $\{\boldsymbol{\nu}_t^{\Upsilon}(r)\}$ satisfies a uniform LDP with a good rate function $\boldsymbol{I}_r : \mathcal{P}(\mathscr{X}(r)) \to [0, +\infty]$, then so does the sequence $\{\boldsymbol{\nu}_t^{\Upsilon}\}$, with the rate function

$$\boldsymbol{I}(\boldsymbol{\lambda}) = \sup_{r\geq 1} \boldsymbol{I}^r\big(\Pi_*^r(\boldsymbol{\lambda})\big), \tag{2.6}$$

where $\Pi^r : \boldsymbol{X} \to \mathscr{X}(r)$ stands for the natural projection to the first r components. We shall prove the uniform LDP for $\{\boldsymbol{\nu}_t^{\Upsilon}(r)\}$ with an arbitrary $r \geq 1$, establish a variational formula for the corresponding rate function $\boldsymbol{I}^r$, and use relation (2.6) to obtain the Donsker–Varadhan entropy formula (2.3).

Application of Kifer's theorem

To prove the uniform LDP for the sequence $\{\boldsymbol{\nu}_t^{\Upsilon}(r)\}_{t\geq 1}$ for a fixed $r \geq 1$, we shall apply Kifer's theorem [Kif90], which is recalled in Section 5.2. To this end, we define the set $\Theta = \{\theta = (t, \Upsilon), t \in \mathbb{N}, \Upsilon \in \mathscr{X}\}$ and endow it with a partial order $\prec$ defined by the following rule:

$$(t_1, \Upsilon_1) \prec (t_2, \Upsilon_2) \quad \text{if and only if} \quad t_1 \leq t_2.$$

The sequence $\{\boldsymbol{\nu}_t^{\Upsilon}(r)\}$ will be regarded as a directed family indexed by $\theta \in \Theta$, and in what follows we shall often write $\{\boldsymbol{\nu}_\theta\}$, dropping the fixed integer r from the notation. Let us suppose that, for any $V \in C(\mathscr{X}(r))$, the limit

$$Q^r(V) = \lim_{\theta\in\Theta} t^{-1} \log \mathbb{E} \exp\big(t\langle V, \boldsymbol{\nu}_\theta\rangle\big) \tag{2.7}$$

exists, and let $\boldsymbol{I}^r : \mathcal{M}(\mathscr{X}(r)) \to [0, +\infty]$ be its Legendre transform; see relation (5.13) for a definition. If, in addition to the existence of limit (2.7), there exists a dense subspace $\mathcal{V} \subset C(\mathscr{X}(r))$ such that, for any $V \in \mathcal{V}$, the equation[8]

$$\langle V, \boldsymbol{\sigma} \rangle - \boldsymbol{I}^r(\boldsymbol{\sigma}) = Q^r(V) \tag{2.8}$$

has a unique solution $\boldsymbol{\sigma} \in \mathcal{P}(\mathscr{X}(r))$, then the validity of the LDP follows immediately from Theorem 5.5. We show in the next step how to reduce the above two properties (existence of limit (2.7) and uniqueness of a solution of Eq. (2.8) for V in a dense subspace) to a study of the large-time asymptotics of a Feynman–Kac semigroup.

Reduction to a study of Feynman–Kac semigroups

Let us consider the following random dynamical system in $\mathscr{X}(r)$:

$$\boldsymbol{\Upsilon}_k(r) = \boldsymbol{S}\big(\boldsymbol{\Upsilon}_{k-1}(r), \eta_k\big), \quad k \geq 1, \tag{2.9}$$

where $\boldsymbol{\Upsilon}_k(r) = [\Upsilon_k^1, \dots, \Upsilon_k^r]$, $\{\eta_k\}$ is the sequence of i.i.d. random variables in $\mathscr{E}$ entering (1.12), and the mapping $\boldsymbol{S}^r : \mathscr{X}(r) \times \mathscr{E} \to \mathscr{X}(r)$ is given by

$$\boldsymbol{S}^r(\Upsilon^1, \dots, \Upsilon^r, \eta) = \big[\Upsilon^2, \dots, \Upsilon^r, S(\Upsilon^r, \eta)\big]. \tag{2.10}$$

Equation (2.9) is supplemented with the initial condition

$$\boldsymbol{\Upsilon}_0(r) = \boldsymbol{\Upsilon}(r) \in \mathscr{X}(r). \tag{2.11}$$

Given a function $V \in C(\mathscr{X}(r))$, we consider the operator

$$\big(\mathfrak{P}_k^V(r) f\big)\big(\boldsymbol{\Upsilon}(r)\big) = \mathbb{E}\big(\exp\big\{V(\boldsymbol{\Upsilon}_1(r)) + \dots + V(\boldsymbol{\Upsilon}_k(r))\big\} f(\boldsymbol{\Upsilon}_k(r))\big), \tag{2.12}$$

acting in the space $C(\mathscr{X}(r))$. The Markov property implies that the sequence $\{\mathfrak{P}_k^V(r)\}$ is a semigroup in $C(\mathscr{X}(r))$. A key observation is that, for $V \in C(\mathscr{X}(r))$ and $\boldsymbol{\Upsilon}(r) \in \mathscr{X}(r)$,

$$Q^r(V) = \lim_{k \to \infty} k^{-1} \log\big(\mathfrak{P}_k^V(r)\mathbf{1}\big)(\boldsymbol{\Upsilon}(r)), \tag{2.13}$$

provided that the limit on the right-hand side exists uniformly with respect to the initial point $\boldsymbol{\Upsilon}(r)$ and does not depend on it. The latter property is a consequence of the following proposition, which is established in Section 2.3 with the help of Theorem 5.6.

Proposition 2.3. *Under the Hypotheses of Theorem 2.1, for any integer $r \geq 1$ and any function $V \in L_b(\mathscr{X}(r))$, there is a number $\lambda_V > 0$, a positive function $h_V \in C(\mathscr{X}(r))$, and a measure $\boldsymbol{\mu}_V \in \mathcal{P}(\mathscr{X}(r))$ such that*

$$\langle h_V, \boldsymbol{\mu}_V \rangle = 1, \quad \mathfrak{P}_1^V(r) h_V = \lambda_V h_V, \quad \mathfrak{P}_1^V(r)^* \boldsymbol{\mu}_V = \lambda_V \boldsymbol{\mu}_V, \tag{2.14}$$

$$\big\| \lambda_V^{-k} \mathfrak{P}_k^V(r) f - \langle f, \boldsymbol{\mu}_V \rangle h_V \big\|_{L^\infty(\mathscr{X}(r))} \to 0 \quad \text{as } k \to \infty, \tag{2.15}$$

where $f \in C(\mathscr{X}(r))$ is an arbitrary function.

[8] The lower semicontinuity of $\boldsymbol{I}^r$ and the inversion formula for the Legendre transform imply that Eq. (2.8) has at least one solution.

Convergence (2.15), combined with (2.13) and the inequality

$$t^{-1}\big|\log \mathbb{E}\exp\big(t\langle W,\boldsymbol{\nu}_\theta\rangle\big)-\log \mathbb{E}\exp\big(t\langle V,\boldsymbol{\nu}_\theta\rangle\big)\big|\le \|W-V\|_\infty,$$

implies that limit (2.7) exists for any $V\in C(\mathscr{X}(r))$. Let us briefly outline the well-known argument proving that Proposition 2.3 also implies the uniqueness of a solution $\boldsymbol{\sigma}\in\mathcal{P}(\mathscr{X}(r))$ for Eq. (2.8) with an arbitrary V in the space $L_b(\mathscr{X}(r))$, which is dense in $C(\mathscr{X}(r))$; cf. [Kif90, Section 4] and [JNPS15b, Section 4].

Let us fix any $V\in C(\mathscr{X}(r))$. For any $W\in C(\mathscr{X}(r))$, we consider a semigroup $\mathfrak{Q}_k^W(r):C(\mathscr{X}(r))\to C(\mathscr{X}(r))$ with the generator given by

$$\mathfrak{Q}_1^W(r)f=\lambda_V^{-1}h_V^{-1}\mathfrak{P}_1^V(e^W h_V f)=\lambda_V^{-1}h_V^{-1}\mathfrak{P}_1^{V+W}(h_V f).$$

In the case $W=0$, we shall write $\mathfrak{Q}_k(r)$. A straightforward calculation shows that $\mathfrak{Q}_k$ is Markovian (that is, $\mathfrak{Q}_k\mathbf{1}=\mathbf{1}$) and

$$\mathfrak{Q}_k^W(r)f=\lambda_V^{-k}h_V^{-1}\mathfrak{P}_k^{V+W}(h_V f).$$

It follows from Proposition 2.3 that, for any $W\in L_b(\mathscr{X}(r))$, we have

$$Q_V^r(W):=\lim_{k\to\infty}k^{-1}\log\mathfrak{Q}_k^W(r)\mathbf{1}=\log\lambda_{V+W}-\log\lambda_V=Q^r(V+W)-Q^r(V).$$

By the Lipschitz continuity of Q^r and Q_V^r, the left-most and right-most terms coincide for any $W\in C(\mathscr{X}(r))$. Denoting by $I_V^r:\mathcal{P}(\mathscr{X}(r))\to[0,+\infty]$ the Legendre transform of Q_V^r, we see that

$$I_V^r(\boldsymbol{\sigma})=I^r(\boldsymbol{\sigma})+Q^r(V)-\langle V,\boldsymbol{\sigma}\rangle\quad\text{for any }\boldsymbol{\sigma}\in\mathcal{P}(\mathscr{X}(r)).\tag{2.16}$$

Thus, a measure $\boldsymbol{\sigma}\in\mathcal{P}(\mathscr{X}(r))$ is a solution for (2.8) if and only if $I_V^r(\boldsymbol{\sigma})=0$. Now note that, by Proposition 2.3, the dual semigroup $\mathfrak{Q}_1^V(r)^*$ has a unique stationary measure, which is given by $\boldsymbol{\sigma}_V=h_V\boldsymbol{\mu}_V$. Hence, the required uniqueness of solution of (2.8) will be established if we prove that any $\boldsymbol{\sigma}\in\mathcal{P}(\mathscr{X}(r))$ satisfying $I_V^r(\boldsymbol{\sigma})=0$ is a stationary measure for $\mathfrak{Q}_1^V(r)^*$.

To this end, we repeat the argument used in the proof of Lemma 2.5 in [DV75]. Namely, as will be established in Proposition 2.4, we have

$$I_V^r(\boldsymbol{\sigma})=\sup_{g>0}\int_{\mathscr{X}(r)}\log\frac{g}{\mathfrak{Q}_1^V(r)g}\,\mathrm{d}\boldsymbol{\sigma},\tag{2.17}$$

where the supremum is taken over all positive continuous functions $g:\mathscr{X}(r)\to\mathbb{R}$. If $I_V^r(\boldsymbol{\sigma})=0$, then the supremum on the right-hand side of (2.17) is attained at the function $g\equiv 1$. It follows that, for any $d\in C(\mathscr{X}(r))$, the function

$$F(\varepsilon)=\int_{\mathscr{X}(r)}\log\frac{1+\varepsilon d}{\mathfrak{Q}_1^V(r)(1+\varepsilon d)}\,\mathrm{d}\boldsymbol{\sigma}$$

is well defined for $|\varepsilon|\ll 1$ and has a local minimum at $\varepsilon=0$. Calculating its derivative at zero, we obtain $\langle d,\boldsymbol{\sigma}\rangle-\langle\mathfrak{Q}_1^V(r)d,\boldsymbol{\sigma}\rangle=0$. Recalling that $d\in C(\mathscr{X}(r))$ was arbitrary, we see that $\boldsymbol{\sigma}$ is a stationary measure for $\mathfrak{Q}_1^V(r)^*$.

We have thus established the first part of Theorem 2.1, and we turn to the explicit expression for the rate function. Relation (2.3) is proved in [DV83] in the case when the process is strong Feller. We present here a different argument applicable to our setting. To emphasise its universal character, we do it in a more general setting, under minimal hypotheses.

Donsker–Varadhan entropy formula

The first step is the derivation of a variational formula for the level-2 rate function; cf. [DV75, Section 2]. Let $\mathfrak{X}$ be a compact metric space and let $P_1(u,\Gamma)$ be a Feller transition function. Given $V \in C(\mathfrak{X})$, we denote by $\{\mathfrak{P}_k^V\}$ a semigroup in $C(\mathfrak{X})$ whose generator is given by

$$(\mathfrak{P}_1^V f)(x) = \int_{\mathfrak{X}} e^{V(y)} f(y) P_1(x, \mathrm{d}y), \quad f \in C(\mathfrak{X}). \tag{2.18}$$

In the case $V \equiv 0$, we shall write $\mathfrak{P}_k$.

Proposition 2.4. *Suppose that, for any $V \in C(\mathfrak{X})$, the limit*

$$Q(V) = \lim_{k\to\infty} \frac{1}{k} \log(\mathfrak{P}_k^V \mathbf{1})(x)$$

exists uniformly in $x \in \mathfrak{X}$ and does not depend on x. Then Q is a 1-Lipschitz convex function such that

$$Q(V + C) = Q(V) + C \quad \textit{for any } V \in C(\mathfrak{X}) \textit{ and } C \in \mathbb{R}, \tag{2.19}$$

and its Legendre transform $I : \mathcal{M}(\mathfrak{X}) \to [0, +\infty]$ has the form

$$I(\lambda) = \begin{cases} \sup\limits_{g>0} \displaystyle\int_{\mathfrak{X}} \log \frac{g}{\mathfrak{P}_1 g} \, \mathrm{d}\lambda & \textit{for } \lambda \in \mathcal{P}(\mathfrak{X}), \\ +\infty & \textit{otherwise}, \end{cases} \tag{2.20}$$

where the supremum is taken over all positive functions $g \in C(\mathfrak{X})$.

We now denote by $\mathfrak{X}(r)$ the r-fold product of the space $\mathfrak{X}$ and, given a function $V \in C(\mathfrak{X}(r))$, consider a semigroup $\mathfrak{P}_k^V(r)$ on $C(\mathfrak{X}(r))$ with the generator [9]

$$\big(\mathfrak{P}_1^V(r) f\big)(\boldsymbol{x}^r) = \int_{\mathfrak{X}} e^{V(x_2,\dots,x_r,y)} f(x_2,\dots,x_r,y) P_1(x_r, \mathrm{d}y), \tag{2.21}$$

where $\boldsymbol{x}^r = [x_1,\dots,x_r] \in \mathfrak{X}(r)$. In the case $V \equiv 0$, we shall write $\mathfrak{P}_k(r)$. Finally, let us denote $\boldsymbol{\mathfrak{X}} = \mathfrak{X}^{\mathbb{N}}$ and $\boldsymbol{\mathfrak{X}}_- = \mathfrak{X}^{\mathbb{Z}_-}$.

Proposition 2.5. *Suppose that, for any integer $r \geq 1$ and any $V \in C(\mathfrak{X}(r))$, there is a uniform limit*

$$\boldsymbol{Q}^r(V) = \lim_{k\to\infty} \frac{1}{k} \log\big(\mathfrak{P}_k^V(r)\mathbf{1}\big)(\boldsymbol{x}^r),$$

[9] In the language Markov processes, this means that $\mathfrak{P}_k^V(r)$ is the Feynman–Kac semigroup associated with the evolution of words of length r.

independent of $\boldsymbol{x}^r \in \mathfrak{X}(r)$. *Let* $\boldsymbol{I}^r : \mathfrak{P}(\mathfrak{X}(r)) \to [0, +\infty]$ *be the Legendre transform of* $\boldsymbol{Q}^r$ *and let* $\boldsymbol{I} : \mathcal{P}(\mathfrak{X}) \to [0, +\infty)$ *be defined by* (2.6). *Then, for any shift-invariant measure* $\boldsymbol{\lambda} \in \mathcal{P}(\mathfrak{X})$, *we have*

$$\boldsymbol{I}(\boldsymbol{\lambda}) = \int_{\mathfrak{X}_-} \mathrm{Ent}\big(\lambda(\boldsymbol{x}, \cdot) \,|\, P_1(x_0, \cdot)\big)\, \boldsymbol{\lambda}_-(\mathrm{d}\boldsymbol{x}), \tag{2.22}$$

where we use the same conventions as in (2.3).

Propositions 2.4 and 2.5 are established in Sections 2.4 and 2.5, respectively. Going back to the proof of Theorem 2.1, we note that Proposition 2.5 implies (2.3) for $\boldsymbol{\lambda} \in \mathcal{P}_s(\boldsymbol{X})$. The fact that $\boldsymbol{I}(\boldsymbol{\lambda})$ is infinite when $\boldsymbol{\lambda}$ is not shift-invariant follows from the observation that $\boldsymbol{\nu}_t^{\Upsilon}$ is *exponentially equivalent*[10] to a sequence of random probability measures concentrated on shift-invariant measures on $\boldsymbol{X}$; see [DV83, Section 1]. Namely, together with $\boldsymbol{\nu}_t^{\Upsilon}$, let us consider the sequence

$$\tilde{\boldsymbol{\nu}}_t^{\Upsilon} = t^{-1} \sum_{n=0}^{t-1} \delta_{\widetilde{\boldsymbol{\Upsilon}}_n(t)},$$

where $\{\widetilde{\boldsymbol{\Upsilon}}_0(t)\}$ is a t-periodic sequence whose first t components coincide with those of $\boldsymbol{\Upsilon}_0$, and $\widetilde{\boldsymbol{\Upsilon}}_n(t)$ is obtained from $\widetilde{\boldsymbol{\Upsilon}}_0(t)$ by deleting the first t components. It is straightforward to check that $\tilde{\boldsymbol{\nu}}_t^{\Upsilon}$ and $\boldsymbol{\nu}_t^{\Upsilon}$ are exponentially equivalent and that

$$\mathbb{P}\big\{\tilde{\boldsymbol{\nu}}_t^{\Upsilon} \in \mathcal{P}_s(\boldsymbol{X})\big\} = 1 \quad \text{for any } t \geq 1,\, \Upsilon \in \mathscr{X}.$$

Since exponentially equivalent sequences satisfy the same LDP, we conclude that the rate function $\boldsymbol{I}$ is infinite on $\mathcal{M}(\boldsymbol{X}) \setminus \mathcal{P}(\boldsymbol{X})$. Finally, the proof of the affine property of $\boldsymbol{I}$ given in [DV83, Theorem 3.5] uses only relation (2.3) and therefore remains valid in our setting. This completes the proof of Theorem 2.1.

2.3 Proof of Proposition 2.3

We first outline the main idea of the proof, which is based on an application of Theorem 5.6. According to that result, to prove the required claims, we need to check the uniform Feller and uniform irreducibility properties (UF) and (UI). The first of them will be established with the help of a coupling technique; see Proposition 5.3. On the other hand, the uniform irreducibility is not valid in $\mathscr{X}(r)$, and we have to restrict ourselves to the domain of attainability $\mathcal{A}(r)$, for which the validity of (UI) follows easily from the approximate controllability (AC). Thus, we can apply Theorem 5.6 with $X = \mathcal{A}(r)$. Finally, to establish convergence (2.15), we shall prove that, for any $\boldsymbol{\Upsilon} \in \mathscr{X}(r)$, there is $\widetilde{\boldsymbol{\Upsilon}} \in \mathcal{A}(r)$ such that

$$\big|\log(\mathfrak{P}_k^V(r)f)(\boldsymbol{\Upsilon}) - \log(\mathfrak{P}_k^V(r)f)(\widetilde{\boldsymbol{\Upsilon}})\big| \leq C\|f\|_\infty \quad \text{for all } k \geq 1, \tag{2.23}$$

[10] See Section 4.2.10 in [DZ00] for a definition.

where $f \in C(\mathscr{X}(r))$ is an arbitrary function, and the constant $C > 0$ does not depend on $\boldsymbol{\Upsilon}$, $\widetilde{\boldsymbol{\Upsilon}}$, and k. The details are split into three steps.

Step 1: Reduction to the domain of attainability. Let us recall that $\mathcal{K} \subset \mathscr{E}$ stands for the support of the law ℓ. Setting $\overline{\boldsymbol{\Upsilon}} = [\overline{\Upsilon}, \dots, \overline{\Upsilon}]$, we define a sequence $\{\mathcal{A}_k(r)\}_{k\geq 0}$ of compact subsets of $\mathscr{X}(r)$ by the following rule:

$$\mathcal{A}_0(r) = \{\overline{\boldsymbol{\Upsilon}}\}, \quad \mathcal{A}_k(r) = \overline{\boldsymbol{S}^r(\mathcal{A}_{k-1}(r), \mathcal{K})} \quad \text{for } k \geq 1,$$

where $\boldsymbol{S}^r$ is defined by (2.10), and $\overline{B}$ stands for the closure of $B \subset \mathscr{X}(r)$. Since $S(\overline{\Upsilon}, 0) = \overline{\Upsilon}$, the sequence $\{\mathcal{A}_k(r)\}_{k\geq 0}$ is increasing. We denote by $\mathcal{A}(r) \subset \mathscr{X}(r)$ the closure of the union $\cup_{k\geq 1}\mathcal{A}_k(r)$. A simple compactness argument yields that, for any $\delta > 0$, there is an integer $m \geq 1$ such that $\mathcal{A}(r)$ is a subset of the δ-neighbourhood of $\mathcal{A}_m(r)$.

Denoting by $\mathsf{P} : \mathscr{X}(r) \to \mathscr{X}$ the projection taking $[\Upsilon^1, \dots, \Upsilon^r]$ to Υ^r, let us show that $\mathsf{P}(\mathcal{A}(r)) = \mathscr{X}$. Indeed, since $\mathcal{A}(r)$ is compact and P is continuous, we see that the projection $\mathsf{P}(\mathcal{A}(r))$ is closed, and so it suffices to prove that it is dense in $\mathscr{X}$. Fix $\Upsilon \in \mathscr{X}$ and $\varepsilon > 0$. By (AC), there is an integer $k \geq r$ and vectors $\eta_1, \dots, \eta_k \in \mathcal{K}$ such that

$$d_{\mathscr{H}}\big(S_k(\overline{\Upsilon}; \eta_1, \dots, \eta_k), \Upsilon\big) < \varepsilon. \tag{2.24}$$

Let us denote by $\boldsymbol{S}^r_k(\boldsymbol{\Upsilon}; \eta_1, \dots, \eta_k)$ the trajectory of (2.9) issued from $\boldsymbol{\Upsilon}$. Inequality (2.24) and the definition of $\mathcal{A}_k(r)$ imply that the vector $\mathsf{P}\boldsymbol{S}^r_k(\boldsymbol{\Upsilon}; \eta_1, \dots, \eta_k)$ belongs to the ε-neighbourhood of Υ. This proves the required density.

We now prove (2.23). Without loss of generality we may assume that $f \in C(\mathscr{X}(r))$ is non-negative. Let $\boldsymbol{\Upsilon} \in \mathscr{X}(r)$. Since $\mathsf{P}(\mathcal{A}(r)) = \mathscr{X}$, we can find $\widetilde{\boldsymbol{\Upsilon}} \in \mathcal{A}(r)$ such that $\mathsf{P}(\boldsymbol{\Upsilon}) = \mathsf{P}(\widetilde{\boldsymbol{\Upsilon}})$. Since $\boldsymbol{S}^r_k(\boldsymbol{\Upsilon}; \eta_1, \dots, \eta_k)$ depends only on the rth component of $\boldsymbol{\Upsilon}$ for $k \geq r$, we have

$$\boldsymbol{S}^r_k(\boldsymbol{\Upsilon}; \eta_1, \dots, \eta_k) = \boldsymbol{S}^r_k(\widetilde{\boldsymbol{\Upsilon}}; \eta_1, \dots, \eta_k) \quad \text{for } k \geq r. \tag{2.25}$$

Denoting by $\boldsymbol{\Upsilon}_k$ and $\widetilde{\boldsymbol{\Upsilon}}_k$ the left- and right-hand terms in (2.25), we see that $\boldsymbol{\Upsilon}_k = \widetilde{\boldsymbol{\Upsilon}}_k$ for $k \geq r$. It follows from (2.12) that

$$\begin{aligned}
\big(\mathfrak{P}^V_k(r)f\big)(\boldsymbol{\Upsilon}) &= \mathbb{E}\left(\exp\big\{V(\boldsymbol{\Upsilon}_1) + \dots + V(\boldsymbol{\Upsilon}_k)\big\} f(\boldsymbol{\Upsilon}_k)\right) \\
&\leq \exp(2r\|V\|_\infty)\, \mathbb{E}\left(\exp\big\{V(\widetilde{\boldsymbol{\Upsilon}}_1) + \dots + V(\widetilde{\boldsymbol{\Upsilon}}_k)\big\} f(\boldsymbol{\Upsilon}_k)\right) \\
&= \exp(2r\|V\|_\infty)\big(\mathfrak{P}^V_k(r)f\big)(\widetilde{\boldsymbol{\Upsilon}}).
\end{aligned}$$

By symmetry, we can exchange the roles of $\boldsymbol{\Upsilon}$ and $\widetilde{\boldsymbol{\Upsilon}}$, and the resulting inequalities imply (2.23) with $C = 2r\|V\|_\infty$. Thus, we need to construct $h_V \in C(\mathscr{X}(r))$, $\boldsymbol{\mu}_V \in \mathcal{P}(\mathscr{X}(r))$, and $\lambda_V > 0$ satisfying relations (2.14) and to establish (2.15) with the L^∞-norm on $\mathcal{A}(r)$. To this end, we shall prove that the Hypotheses of Theorem 5.6 are satisfied with $X = \mathcal{A}(r)$ and $\mathcal{C} = L_b(\mathcal{A}(r))$.

Step 2: Uniform irreducibility. Our goal is to find an integer $n \geq 1$ and a number $p > 0$ such that

$$\mathbb{P}_{\boldsymbol{\Upsilon}}\big\{\boldsymbol{\Upsilon}^n \in B_{\mathscr{X}(r)}(\widehat{\boldsymbol{\Upsilon}}, \varepsilon)\big\} \geq p \quad \text{for any } \boldsymbol{\Upsilon}, \widehat{\boldsymbol{\Upsilon}} \in \mathcal{A}(r), \tag{2.26}$$

where the subscript $\boldsymbol{\Upsilon}$ on the left-hand side means that we consider the trajectory of (2.9) issued from $\boldsymbol{\Upsilon}$. Simple arguments based on the concepts of the support of a measure and of compactness show that (2.26) follows if for any $\varepsilon > 0$ we can find an integer $n \geq 1$ such that, for arbitrary $\boldsymbol{\Upsilon}, \widehat{\boldsymbol{\Upsilon}} \in \mathcal{A}(r)$ and some suitable $\eta_1, \dots, \eta_n \in \mathcal{K}$,

$$d_r\big(\boldsymbol{\Upsilon}_n, \widehat{\boldsymbol{\Upsilon}}\big) < \varepsilon, \tag{2.27}$$

where d_r stands for the distance in $\mathscr{X}(r)$ defined as the maximum of the distances between the components, and $\boldsymbol{\Upsilon}_k = \boldsymbol{S}^r_k(\boldsymbol{\Upsilon}, \eta_1, \dots, \eta_k)$ is the trajectory of (2.9) issued from $\boldsymbol{\Upsilon}$; see Section 3.3.2 in [KS12] and Section 4 in [JNPS15b].

The construction of the controls $\eta_1, \dots, \eta_n$ is carried out in two steps: we first steer the trajectory to a point close to $\overline{\boldsymbol{\Upsilon}}$ and then use the definition of $\mathcal{A}(r)$ to steer it further to the neighbourhood of $\widehat{\boldsymbol{\Upsilon}}$. More precisely, as it was mentioned in Step 1, we can find $m \geq 1$ such that $\mathcal{A}(r)$ is included in the $\varepsilon/2$-neighbourhood of $\mathcal{A}_m(r)$. Hence, there is $\widehat{\boldsymbol{\Upsilon}}_1 \in \mathcal{A}_m(r)$ such that $d_r(\widehat{\boldsymbol{\Upsilon}}, \widehat{\boldsymbol{\Upsilon}}_1) < \varepsilon/2$. Furthermore, by the definition of $\mathcal{A}_m(r)$, we can find vectors $\eta_1, \dots, \eta_m \in \mathcal{K}$ such that $\boldsymbol{S}^r_m(\overline{\boldsymbol{\Upsilon}}; \eta_1, \dots, \eta_m) = \widehat{\boldsymbol{\Upsilon}}_1$. By continuity, there is a number $\delta > 0$ such that, for any $\boldsymbol{\Upsilon}' \in \mathscr{X}(r)$ satisfying $d_r(\boldsymbol{\Upsilon}', \overline{\boldsymbol{\Upsilon}}) \leq \delta$, we have $d_r(\boldsymbol{S}^(_m \boldsymbol{\Upsilon}'; \eta_1, \dots, \eta_m), \widehat{\boldsymbol{\Upsilon}}_1) < \varepsilon/2$, so that

$$d_r\big(\boldsymbol{S}^r_m(\boldsymbol{\Upsilon}'; \eta_1, \dots, \eta_m), \widehat{\boldsymbol{\Upsilon}}\big) < \varepsilon. \tag{2.28}$$

By (AC), there is an integer $l \geq 1$ and controls $\zeta_1, \dots, \zeta_l \in \mathcal{K}$ such that $d_r(S_l(\Pi(\boldsymbol{\Upsilon}); \zeta_1, \dots, \zeta_l), \overline{\Upsilon}) < \delta$. This observation and the relation $S(\overline{\Upsilon}, 0) = \overline{\Upsilon}$ yield that

$$d_r\big(\boldsymbol{S}^r_{l+r-1}(\boldsymbol{\Upsilon}; \zeta_1, \dots, \zeta_l, \underbrace{0, \dots, 0}_{r-1 \text{ times}}), \widehat{\boldsymbol{\Upsilon}}\big) < \delta.$$

Combining this with (2.28), we derive

$$d_r\big(\boldsymbol{S}^r_{l+m+r-1}(\boldsymbol{\Upsilon}; \zeta_1, \dots, \zeta_l, 0, \dots, 0, \eta_1, \dots, \eta_m), \widehat{\boldsymbol{\Upsilon}}\big) < \varepsilon.$$

This proves the required inequality (2.27) with the integer $n = m + l + r - 1$ not depending on $\boldsymbol{\Upsilon}$ and $\widehat{\boldsymbol{\Upsilon}}$.

Step 3: Uniform Feller property. We shall show that, for any $\Upsilon, \Upsilon' \in \mathscr{X}(r)$, $k \geq r$, and non-negative functions $V, f \in L_b(\mathscr{X}(r))$,

$$\big|\big(\mathfrak{P}^V_k(r)f\big)(\boldsymbol{\Upsilon}) - \big(\mathfrak{P}^V_k(r)f\big)(\boldsymbol{\Upsilon}')\big| \leq C\|f\|_L \big\|\mathfrak{P}^V_k(r)\mathbf{1}\big\|_\infty d_r(\boldsymbol{\Upsilon}, \boldsymbol{\Upsilon}'), \tag{2.29}$$

where $C > 0$ is a number not depending on k and f, and both L^∞ and L_b norms on the right-hand side are taken over $\mathscr{X}(r)$. This will obviously imply the validity of (UF) with $\mathcal{C} = \{f \in L_b(\mathscr{X}(r)) : f \geq 1\}$.

Fix the initial points $\boldsymbol{\Upsilon}, \boldsymbol{\Upsilon}' \in \mathscr{X}(r)$ and denote by $\{\boldsymbol{\Upsilon}_k\}$ and $\{\boldsymbol{\Upsilon}'_k\}$ the trajectories of (2.9) issued from them. Let $\{\widetilde{\Upsilon}_k\}$ and $\{\widetilde{\Upsilon}'_k\}$ be the trajectories constructed in Corollary 5.4 for the initial points $\mathsf{P}(\boldsymbol{\Upsilon}) = \Upsilon_r$ and $\mathsf{P}(\boldsymbol{\Upsilon}') = \Upsilon'_r$, respectively. Note that they depend on the choice of the parameter $q \in (0, 1)$ that will be specified below. We set

$$\widetilde{\boldsymbol{\Upsilon}}_k = [\widetilde{\Upsilon}_{k-r+1}, \dots, \widetilde{\Upsilon}_k], \quad \widetilde{\boldsymbol{\Upsilon}}'_k = [\widetilde{\Upsilon}'_{k-r+1}, \dots, \widetilde{\Upsilon}'_k],$$

where $\widetilde{\Upsilon}_{j-r} = \Upsilon_j$ and $\widetilde{\Upsilon}'_{j-r} = \Upsilon'_j$ for $2 \le j \le r$. We set $D(j) = d_r(\widetilde{\boldsymbol{\Upsilon}}_j, \widetilde{\boldsymbol{\Upsilon}}'_j)$, and introduce the events

$$G_l(q) = \big\{D(j) \le q^{j-r+1} d_r\big(\boldsymbol{\Upsilon}, \boldsymbol{\Upsilon}'\big) \text{ for } 0 \le j < l, D(l) > q^{l-r+1} d_r\big(\boldsymbol{\Upsilon}, \boldsymbol{\Upsilon}'\big)\big\},$$
$$G'_l(q) = \big\{D(j) \le q^{j-r+1} d_r\big(\boldsymbol{\Upsilon}, \boldsymbol{\Upsilon}'\big) \text{ for } 0 \le j \le l\big\},$$

where $l \ge 1$. It follows from (5.11) that

$$\mathbb{P}\big(G_l(q)\big) \le C_1 q^l d_r(\boldsymbol{\Upsilon}, \boldsymbol{\Upsilon}') \quad \text{for all } l \ge 0, \tag{2.30}$$

where $C_1 > 0$ does not depend on l. Since the laws of the trajectories $\{\boldsymbol{\Upsilon}_k\}$ and $\{\boldsymbol{\Upsilon}'_k\}$ coincide with those of $\{\widetilde{\boldsymbol{\Upsilon}}_k\}$ and $\{\widetilde{\boldsymbol{\Upsilon}}'_k\}$ respectively, we have

$$\big(\mathfrak{P}_k^V(r)f\big)(\boldsymbol{\Upsilon}) = \mathbb{E}\left(\exp\{V(\widetilde{\boldsymbol{\Upsilon}}_1) + \cdots + V(\widetilde{\boldsymbol{\Upsilon}}_k)\} f(\widetilde{\boldsymbol{\Upsilon}}_k)\right) = \mathbb{E}\left(\Xi_k(\boldsymbol{\Upsilon}) f(\widetilde{\boldsymbol{\Upsilon}}_k)\right),$$

where $\Xi_k(\boldsymbol{\Upsilon}) = \exp\{V(\widetilde{\boldsymbol{\Upsilon}}_1) + \cdots + V(\widetilde{\boldsymbol{\Upsilon}}_k)\}$, and a similar representation holds for $(\mathfrak{P}_k^V(r)f)(\boldsymbol{\Upsilon}')$. Setting

$$I_k^l(\boldsymbol{\Upsilon}, \boldsymbol{\Upsilon}') = \mathbb{E}\big\{\mathbb{I}_{G_l(q)}\big(\Xi_k(\boldsymbol{\Upsilon}) f(\widetilde{\boldsymbol{\Upsilon}}_k) - \Xi_k(\boldsymbol{\Upsilon}') f(\widetilde{\boldsymbol{\Upsilon}}'_k)\big)\big\},$$
$$J_k(\boldsymbol{\Upsilon}, \boldsymbol{\Upsilon}') = \mathbb{E}\big\{\mathbb{I}_{G'_l(q)}\big(\Xi_k(\boldsymbol{\Upsilon}) f(\widetilde{\boldsymbol{\Upsilon}}_k) - \Xi_k(\boldsymbol{\Upsilon}') f(\widetilde{\boldsymbol{\Upsilon}}'_k)\big)\big\},$$

where $\mathbb{I}_G$ stands for the indicator function of G, we can write

$$\begin{aligned}\Delta_k(\boldsymbol{\Upsilon}, \boldsymbol{\Upsilon}') :&= \big(\mathfrak{P}_k^V(r)f\big)(\boldsymbol{\Upsilon}) - \big(\mathfrak{P}_k^V(r)f\big)(\boldsymbol{\Upsilon}') \\ &= \sum_{l=1}^k I_k^l(\boldsymbol{\Upsilon}, \boldsymbol{\Upsilon}') + J_k(\boldsymbol{\Upsilon}, \boldsymbol{\Upsilon}').\end{aligned} \tag{2.31}$$

The Markov property and inequality (2.30) imply that

$$\begin{aligned}I_k^l(\boldsymbol{\Upsilon}, \boldsymbol{\Upsilon}') &\le \mathbb{E}\big\{\mathbb{I}_{G_l(q)} \Xi_k(\boldsymbol{\Upsilon}) f(\widetilde{\boldsymbol{\Upsilon}}_k)\big\} = \mathbb{E}\big\{\mathbb{I}_{G_l(q)}\, \mathbb{E}\big(\Xi_k(\boldsymbol{\Upsilon}) f(\widetilde{\boldsymbol{\Upsilon}}_k) \,|\, \mathcal{F}_l\big)\big\} \\ &\le \|f\|_\infty \exp\big(l\, \|V\|_\infty\big)\, \mathbb{E}\big\{\mathbb{I}_{G_l(q)}\, \big(\mathfrak{P}_{k-l}^V(r)\mathbf{1}\big)(\widetilde{\boldsymbol{\Upsilon}}_l)\big\} \\ &\le \|f\|_\infty \exp\big(l\, \|V\|_\infty\big)\, \big\|\mathfrak{P}_k^V(r)\mathbf{1}\big\|_\infty\, \mathbb{P}\big(G_l(q)\big) \\ &\le C_1 \|f\|_\infty \exp\big(l\, \|V\|_\infty - l \log q^{-1}\big)\, \big\|\mathfrak{P}_k^V(r)\mathbf{1}\big\|_\infty d_r(\boldsymbol{\Upsilon}, \boldsymbol{\Upsilon}').\end{aligned} \tag{2.32}$$

To estimate $J_k = J_k(\boldsymbol{\Upsilon}, \boldsymbol{\Upsilon}')$, we write

$$\begin{aligned}J_k &= \mathbb{E}\big\{\mathbb{I}_{G'_k(q)}\, \Xi_k(\boldsymbol{\Upsilon})(f(\boldsymbol{\Upsilon}'_k) - f(\boldsymbol{\Upsilon}_k))\big\} + \mathbb{E}\big\{\mathbb{I}_{G'_k(q)}(\Xi_k(\boldsymbol{\Upsilon}) - \Xi_k(\boldsymbol{\Upsilon}'))f(\boldsymbol{\Upsilon}_k)\big\} \\ &=: J_k^1(\boldsymbol{\Upsilon}, \boldsymbol{\Upsilon}') + J_k^2(\boldsymbol{\Upsilon}, \boldsymbol{\Upsilon}').\end{aligned} \tag{2.33}$$

Using the Lipschitz continuity of f, we derive

$$J_k^1(\boldsymbol{\Upsilon}, \boldsymbol{\Upsilon}') \le C_2\, q^k \|f\|_L \big\|\mathfrak{P}_k^V(r)\mathbf{1}\big\|_\infty d_r(\boldsymbol{\Upsilon}, \boldsymbol{\Upsilon}'). \tag{2.34}$$

Furthermore, the Lipschitz continuity of V implies that, for $q \le 1/2$,

$$\begin{aligned}\big|\Xi_k(\boldsymbol{\Upsilon}) - \Xi_k(\boldsymbol{\Upsilon}')\big| &= \Xi_k(\boldsymbol{\Upsilon})\Big\{\exp\Big(\sum_{j=1}^{k}\big|V(\widetilde{\boldsymbol{\Upsilon}}_j) - V(\widetilde{\boldsymbol{\Upsilon}}'_j\big|\Big) - 1\Big\}\\ &\le \Xi_k(\boldsymbol{\Upsilon})\Big\{\exp\big(2q\,\|V\|_L\, d_r(\boldsymbol{\Upsilon},\boldsymbol{\Upsilon}')\big) - 1\Big\}\\ &\le C_3(V)\, d_r(\boldsymbol{\Upsilon},\boldsymbol{\Upsilon}')\,\Xi_k(\boldsymbol{\Upsilon}),\end{aligned}$$

on the set $G'_k(q)$. It follows that

$$J_k^2(\boldsymbol{\Upsilon},\boldsymbol{\Upsilon}') \le C_3(V)\,\|f\|_\infty \big\|\mathfrak{P}_k^V(r)\mathbf{1}\big\|_\infty d_r(\boldsymbol{\Upsilon},\boldsymbol{\Upsilon}'). \tag{2.35}$$

Combining this with (2.31)–(2.35), we derive

$$\big|\Delta_k(\boldsymbol{\Upsilon},\boldsymbol{\Upsilon}')\big| \le C_4(V)\|f\|_L\big\|\mathfrak{P}_k^V(r)\mathbf{1}\big\|_\infty d_r(\boldsymbol{\Upsilon},\boldsymbol{\Upsilon}')\sum_{l=0}^{k}\exp\big(l\,\|V\|_\infty - l\log q^{-1}\big).$$

Taking $q < \exp(-\|V\|_\infty)$, we arrive at (2.29).

2.4 Proof of Proposition 2.4

The fact the Q is a 1-Lipschitz convex function satisfying (2.19) is well known, as is the relation $I(\lambda) = +\infty$ for $\lambda \in \mathcal{M}(\mathfrak{X}) \setminus \mathcal{P}(\mathfrak{X})$. We thus confine ourselves to the proof of (2.20) for $\lambda \in \mathcal{P}(\mathfrak{X})$.

Step 1. Let us denote by $J(\lambda)$ the supremum on the right-hand side of (2.20). We first prove that

$$I(\lambda) \ge J(\lambda) \quad \text{for any } \lambda \in \mathcal{P}(\mathfrak{X}). \tag{2.36}$$

To this end, fix $\lambda \in \mathcal{P}(\mathfrak{X})$ and $\varepsilon > 0$. Let $g \in C(\mathfrak{X})$ be such that $g \ge 1$ and

$$J(\lambda) < \int_{\mathfrak{X}} \log\frac{g}{\mathfrak{P}_1 g}\,\mathrm{d}\lambda + \varepsilon. \tag{2.37}$$

Set $V = \log\frac{g}{\mathfrak{P}_1 g}$. A simple calculation based on the semigroup property and the inequality $1 \le \mathfrak{P}_1 g \le \|g\|_\infty$ shows that

$$\|g\|_\infty^{-1} \le \mathfrak{P}_k^V\mathbf{1} \le \|g\|_\infty,$$

and so $Q(V) = 0$. Inequality (2.37) now implies that

$$J(\lambda) < \langle V,\lambda\rangle + \varepsilon \le I(\lambda) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we arrive at (2.36).

Step 2. To establish the opposite inequality in (2.36), we again fix $\varepsilon > 0$. Let $V \in C(\mathfrak{X})$ be such that

$$I(\lambda) < \langle V,\lambda\rangle + \varepsilon, \quad Q(V) = 0. \tag{2.38}$$

The existence of such a function follows from the definition of I and the relation (2.19). We now set

$$g_\varepsilon = e^V \sum_{k=0}^{\infty} e^{-\varepsilon k}\, \mathfrak{P}_k^V \mathbf{1}.$$

The second relation in (2.38) implies that the series converges uniformly in $u \in \mathfrak{X}$ and defines a continuous function on $\mathfrak{X}$. It is straightforward to check that

$$\mathfrak{P}_1 g_\varepsilon = \sum_{k=0}^{\infty} e^{-\varepsilon k}\, \mathfrak{P}_{k+1}^V \mathbf{1} = e^\varepsilon (e^{-V} g_\varepsilon - 1).$$

It follows that

$$\log \frac{g_\varepsilon}{\mathfrak{P}_1 g_\varepsilon} \geq V - \varepsilon - \log\bigl(1 - e^V g_\varepsilon^{-1}\bigr) \geq V - \varepsilon. \tag{2.39}$$

Integrating (2.39) with respect to λ and using (2.38), we derive

$$\int_{\mathfrak{X}} \log \frac{g_\varepsilon}{\mathfrak{P}_1 g_\varepsilon} \mathrm{d}\lambda \geq \langle V, \lambda \rangle - \varepsilon \geq I(\lambda) - 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we arrive at the required inequality.

2.5 Proof of Proposition 2.5

Step 1: A formula for $\boldsymbol{I}$. We first note that $\mathfrak{P}_k^V(r)$ falls into the framework of Proposition 2.4 if we define the transition function by

$$P_1^r(\boldsymbol{x}^r, \mathrm{d}\boldsymbol{y}^r) = \delta_{[x_2,\dots,x_r]}(\mathrm{d}y_1, \dots, \mathrm{d}y_{r-1}) P_1(x_r, \mathrm{d}y_r).$$

Therefore, replacing g by e^V in (2.20) and using approximation of a bounded measurable function by continuous functions, we can write

$$\boldsymbol{I}^r(\lambda) = \sup_{V \geq 0} \bigl\langle V - \log\bigl(\mathfrak{P}_1(r) e^V\bigr), \lambda \bigr\rangle, \quad \lambda \in \mathcal{P}(\mathfrak{X}(r)),$$

where the supremum is taken over all non-negative bounded measurable functions $V : \mathfrak{X}(r) \to \mathbb{R}$. Combining this with (2.6), we derive

$$\boldsymbol{I}(\boldsymbol{\lambda}) = \sup_{r \geq 1} \sup_{V \geq 0} \int_{\mathfrak{X}(r)} \Bigl(V(\boldsymbol{x}^r) - \log \int_{\mathfrak{X}} e^{V(x_2,\dots,x_r,y)} P_1(x_r, \mathrm{d}y) \Bigr) \boldsymbol{\lambda}^r(\mathrm{d}\boldsymbol{x}^r), \tag{2.40}$$

where $\boldsymbol{\lambda}^r$ stands for the image of $\boldsymbol{\lambda}$ under the projection Π^r to the first r components. Since $\boldsymbol{\lambda}$ is shift-invariant, we can replace $[x_2, \dots, x_r, y]$ by $[x_1, \dots, x_{r-1}, y]$ in the integral over $\mathfrak{X}$. Let us denote by $\boldsymbol{\lambda}^r(\boldsymbol{x}^{r-1}; \cdot)$ the regular conditional probability of $\boldsymbol{\lambda}^r$ given the first $r-1$ coordinates and let

$$F_V(\boldsymbol{x}^{r-1}) = \int_{\mathfrak{X}} V(\boldsymbol{x}^{r-1}, y) \boldsymbol{\lambda}^r(\boldsymbol{x}^{r-1}; \mathrm{d}y) - \log \int_{\mathfrak{X}} e^{V(\boldsymbol{x}^{r-1}, y)} P_1(x_{r-1}, \mathrm{d}y).$$

We can rewrite (2.40) as

$$\boldsymbol{I}(\boldsymbol{\lambda}) = \sup_{r\geq 1}\sup_{V\geq 0}\int_{\mathfrak{X}(r-1)} F_V(\boldsymbol{x}^{r-1})\,\boldsymbol{\lambda}^{r-1}(\mathrm{d}\boldsymbol{x}^{r-1}). \tag{2.41}$$

Denoting by $\boldsymbol{J}(\boldsymbol{\lambda})$ the expression on the right-hand side of (2.22), we now prove that $\boldsymbol{I}$ is bounded from above and from below by $\boldsymbol{J}$.

Step 2: Upper bound. We recall the convention that any shift-invariant measure $\boldsymbol{\lambda}\in\mathcal{P}(\boldsymbol{X})$ can be extended (in a unique manner) to $\mathfrak{X}^{\mathbb{Z}}$. For any integer $r\geq 1$, we write $\mathbb{Z}_r=\mathbb{Z}\cap(-\infty,r]$ and, given a measure $\boldsymbol{\lambda}\in\mathcal{P}_s(\boldsymbol{\mathfrak{X}})$, denote by $\boldsymbol{\lambda}(\boldsymbol{z},\boldsymbol{x}^{r-1};\,\cdot\,)$ the regular conditional probability of the projection of $\boldsymbol{\lambda}$ to $\mathfrak{X}^{\mathbb{Z}_r}$ given $[\boldsymbol{z},\boldsymbol{x}^{r-1}]\in\mathfrak{X}^{\mathbb{Z}_{r-1}}$. Let

$$\widetilde{F}_V(\boldsymbol{z},\boldsymbol{x}^{r-1}) = \int_{\mathfrak{X}} V(\boldsymbol{x}^{r-1},y)\boldsymbol{\lambda}(\boldsymbol{z},\boldsymbol{x}^{r-1};\mathrm{d}y) - \log\int_{\mathfrak{X}} e^{V(\boldsymbol{x}^{r-1},y)}P_1(x_{r-1},\mathrm{d}y).$$

It is straightforward to check that

$$\int_{\mathfrak{X}(r-1)} F_V(\boldsymbol{x}^{r-1})\,\boldsymbol{\lambda}^{r-1}(\mathrm{d}\boldsymbol{x}^{r-1}) = \int_{\boldsymbol{\mathfrak{X}}_-}\int_{\mathfrak{X}(r-1)} \widetilde{F}_V(\boldsymbol{z},\boldsymbol{x}^{r-1})\boldsymbol{\lambda}(\mathrm{d}\boldsymbol{z},\mathrm{d}\boldsymbol{x}^{r-1}). \tag{2.42}$$

The definition of the relative entropy implies that

$$\widetilde{F}_V(\boldsymbol{z},\boldsymbol{x}^{r-1}) \leq \mathrm{Ent}\big(\boldsymbol{\lambda}(\boldsymbol{z},\boldsymbol{x}^{r-1};\,\cdot\,)\,|\,P_1(x_{r-1},\cdot)\big) \quad \text{for any } V\in C_b(\mathfrak{X}(r)).$$

Substituting this and (2.42) into (2.41), we obtain

$$\boldsymbol{I}(\boldsymbol{\lambda}) \leq \sup_{r\geq 1}\int_{\boldsymbol{\mathfrak{X}}_-}\int_{\mathfrak{X}(r-1)} \mathrm{Ent}\big(\boldsymbol{\lambda}(\boldsymbol{z},\boldsymbol{x}^{r-1};\cdot)\,|\,P_1(x_{r-1},\cdot)\big)\boldsymbol{\lambda}(\mathrm{d}\boldsymbol{z},\mathrm{d}\boldsymbol{x}^{r-1}).$$

In view of the stationarity of $\boldsymbol{\lambda}$, the expression under the supremum on the right-hand side of this inequality does not depend on r and coincides with $\boldsymbol{J}(\boldsymbol{\lambda})$.

Step 3: Lower bound. For any integer $r\geq 0$, we define the space $\boldsymbol{\mathfrak{X}}_r=\mathfrak{X}^{[\![-r,0]\!]}$ and denote by $\boldsymbol{x}_r=[x_{-r},\dots,x_0]$ its points. To prove that $\boldsymbol{I}\geq\boldsymbol{J}$, we first rewrite (2.41) in the form

$$\boldsymbol{I}(\boldsymbol{\lambda}) = \sup_{r\geq 0}\sup_{V\geq 0}\int_{\boldsymbol{\mathfrak{X}}_r} F_V(\boldsymbol{x}_r)\boldsymbol{\lambda}_r(\mathrm{d}\boldsymbol{x}_r). \tag{2.43}$$

Here, bounded measurable functions V depend on $r+2$ variables, and, with a slight abuse of notation, we write

$$F_V(\boldsymbol{x}_r) = \int_{\mathfrak{X}} V(\boldsymbol{x}_r,y)\boldsymbol{\lambda}^{[\![-r,1]\!]}(\boldsymbol{x}_r;\mathrm{d}y) - \log\int_{\mathfrak{X}} e^{V(\boldsymbol{x}_r,y)}P_1(x_0,\mathrm{d}y), \tag{2.44}$$

where $\boldsymbol{\lambda}^{[\![-r,1]\!]}(\boldsymbol{x}_r;\,\cdot\,)$ denotes the regular conditional probability for the projection of $\boldsymbol{\lambda}$ to $\mathfrak{X}^{[\![-r,1]\!]}$ given $\boldsymbol{x}_r\in\boldsymbol{\mathfrak{X}}_r$. In what follows, it will be convenient to consider $\boldsymbol{\lambda}^{[\![-r,1]\!]}(\boldsymbol{x}_r;\,\cdot\,)$ as a function of the entire trajectory $\boldsymbol{x}\in\boldsymbol{\mathfrak{X}}_-$ depending only

on $\boldsymbol{x}_r$, and accordingly we shall write $\boldsymbol{\lambda}^{[\![-r,1]\!]}(\boldsymbol{x};\,\cdot\,)$. Let $\Pi_r : \mathfrak{X}_- \to \mathfrak{X}_r$ be the projection taking $\boldsymbol{x}$ to $\boldsymbol{x}_r$ and let $\{\mathcal{F}_r\}_{r\geq 0}$ be the filtration on $\mathfrak{X}_-$ generated by the projections Π_s, $0 \leq s \leq r$. We claim that, for any bounded measurable function $f : \mathfrak{X} \to \mathbb{R}$, the sequence $\{\langle f, \boldsymbol{\lambda}^{[\![-r,1]\!]}(\boldsymbol{x},\cdot)\rangle\}_{r\geq 0}$ considered on the probability space $(\mathfrak{X}_-, \boldsymbol{\lambda}_-)$ is a martingale with respect to the filtration $\{\mathcal{F}_r\}$. To see this, let us consider the probability space $(\mathfrak{X}^{\mathbb{Z}}, \boldsymbol{\lambda})$ and the bounded random variable $\xi([x_j]_{j\in\mathbb{Z}}) = f(x_1)$ on it. By the definition of the regular conditional probability, we have

$$\langle f, \boldsymbol{\lambda}^{[\![-r,1]\!]}(\boldsymbol{x},\cdot)\rangle = \mathbb{E}^{\boldsymbol{\lambda}_-}(\xi \,|\, \mathcal{F}_r),$$

where $\mathbb{E}^{\boldsymbol{\lambda}_-}(\cdot\,|\,\cdot)$ denotes the conditional expectation with respect to $\boldsymbol{\lambda}_-$. This relation immediately implies the required martingale property.

Applying Doob's martingale convergence theorem, we see that, for any bounded measurable function $f : \mathfrak{X} \to \mathbb{R}$, the sequence $\langle f, \boldsymbol{\lambda}^{[\![-r,1]\!]}(\boldsymbol{x},\cdot)\rangle$ converges for $\boldsymbol{\lambda}_-$-almost every $\boldsymbol{x} \in \mathfrak{X}_-$. By Theorem A.5.2 in [KS12], there is a random probability measure $\mu(\boldsymbol{x},\,\cdot\,)$ such that

$$\boldsymbol{\lambda}^{[\![-r,1]\!]}(\boldsymbol{x},\,\cdot\,) \rightharpoonup \mu(\boldsymbol{x},\,\cdot\,) \quad \text{for } \boldsymbol{\lambda}_-\text{-a.e. } \boldsymbol{x} \in \mathfrak{X}_-. \tag{2.45}$$

It is straightforward to check that $\mu(\boldsymbol{x},\,\cdot\,)$ is the projection to the first component of the regular conditional probability with respect to $\boldsymbol{\lambda}$ given $\boldsymbol{x} \in \mathfrak{X}_-$. By uniqueness, it must coincide with $\lambda(\boldsymbol{x},\,\cdot\,)$ for $\boldsymbol{\lambda}_-$-almost every $\boldsymbol{x} \in \mathfrak{X}_-$.

We now recall that

$$E_r(\boldsymbol{x}_r) := \mathrm{Ent}\big(\boldsymbol{\lambda}^{[\![-r,1]\!]}(\boldsymbol{x}_r,\,\cdot\,) \,|\, P_1(x_0,\,\cdot\,)\big) = \sup_{V\geq 0} F_V(\boldsymbol{x}_r), \tag{2.46}$$

where the supremum is taken over all non-negative bounded measurable functions $V : \mathfrak{X}_r \times \mathfrak{X} \to \mathbb{R}$. Moreover, the supremum in (2.46) is saturated by the sequence of functions

$$V_N(\boldsymbol{x}_r, y) = \left\{ \left(\log \frac{\mathrm{d}\boldsymbol{\lambda}^{[\![-r,1]\!]}(\boldsymbol{x}_r,\cdot)}{\mathrm{d}P_1(x_0,\,\cdot\,)} \right) \wedge N \right\} \vee (-N) + N.$$

Therefore, for any $\varepsilon > 0$ and $\boldsymbol{x}_r \in \mathfrak{X}_r$, we can find an integer $N = N_\varepsilon(\boldsymbol{x}_r) \geq 1$ such that

$$F_{V_N}(\boldsymbol{x}_r) \geq \begin{cases} E_r(\boldsymbol{x}_r) - \varepsilon & \text{if } E_r(\boldsymbol{x}_r) < \infty, \\ \varepsilon^{-1} & \text{if } E_r(\boldsymbol{x}_r) = \infty. \end{cases} \tag{2.47}$$

If $E_r(\boldsymbol{x}_r) = +\infty$ for some $r \geq 0$ on a set of positive $\boldsymbol{\lambda}_r$-measure, then both $\boldsymbol{I}(\boldsymbol{\lambda})$ and $\boldsymbol{J}(\boldsymbol{\lambda})$ are equal to $+\infty$. In the opposite case, combining (2.43), (2.46), and (2.47), we see that

$$\boldsymbol{I}(\boldsymbol{\lambda}) \geq \sup_{r\geq 0} \int_{\mathfrak{X}_-} \mathrm{Ent}\big(\boldsymbol{\lambda}^{[\![-r,1]\!]}(\boldsymbol{x},\cdot)\,|P_1(x_0,\,\cdot\,)\big)\boldsymbol{\lambda}_-(\mathrm{d}\boldsymbol{x}) - \varepsilon. \tag{2.48}$$

Since the relative entropy is lower-semicontinuous and non-negative, using (2.45) with $\mu(\boldsymbol{x},\cdot)=\lambda(\boldsymbol{x},\cdot)$ and applying Fatou's lemma, we obtain

$$\liminf_{r\to\infty}\int_{\mathfrak{X}_-}\mathrm{Ent}\big(\boldsymbol{\lambda}^{[\![-r,1]\!]}(\boldsymbol{x},\cdot)\,|P_1(x_0,\,\cdot\,)\big)\,\boldsymbol{\lambda}_-(\mathrm{d}\boldsymbol{x})$$
$$\geq\int_{\mathfrak{X}_-}\mathrm{Ent}\big(\lambda(\boldsymbol{x},\,\cdot)\,|\,P_1(x_0,\,\cdot\,)\big)\boldsymbol{\lambda}_-(\mathrm{d}\boldsymbol{x}).$$

Combining this with (2.48) and recalling that $\varepsilon>0$ was arbitrary, we derive the required inequality $\boldsymbol{I}(\boldsymbol{\lambda})\geq\boldsymbol{J}(\boldsymbol{\lambda})$. This completes the proof of Proposition 2.5.

3 Image of measures under non-degenerate maps

In this section, we discuss some general results on the absolute continuity of the image of measures under finite-dimensional smooth maps. This type of properties are well known, and a comprehensive presentation can be found in [Bog10]. Here we only need a sufficient condition for the existence, regularity, positivity and Lipschitz dependence on the parameter of the density for the image measure. For the reader's convenience, we give a proof of the results we need (Theorems 3.1 and 3.2) in Appendix 5.4.

Let $\mathscr{E}$ be a separable Hilbert space, let $\mathscr{X}$ be a compact subset in a separable Hilbert[11] manifold $\mathscr{H}$ with a metric d (see Section II.1 in [Lan85]), and let $\mathscr{Y}$ be a Riemannian manifold without boundary. We consider a function $F:\mathscr{X}\times\mathscr{E}\to\mathscr{Y}$ satisfying the following hypothesis for an integer $k\geq 0$.

(F) *For any $\Upsilon\in\mathscr{X}$, the mapping $\eta\mapsto F(\Upsilon,\eta)$ is $(k+1)$-times continuously differentiable, and the derivative $\partial_\eta^{k+1}F$ is a Lipschitz-continuous function of (Υ,η) on bounded subsets of $\mathscr{X}\times\mathscr{E}$.*

Our aim is to study the image of signed measures on $\mathscr{E}$ under maps with the above property. Namely, given $\Upsilon\in\mathscr{X}$ and $\ell\in\mathcal{M}(\mathscr{E})$, we denote by $\lambda_\Upsilon=F_*(\Upsilon,\ell)$ the image of ℓ under the map $\eta\mapsto F(\Upsilon,\eta)$. We impose the following hypothesis on ℓ.

(P) *The support of ℓ is a compact subset of $\mathscr{E}$, and there is an orthonormal basis $\{\varphi_j\}$ in $\mathscr{E}$ such that ℓ can be written as the product of its one-dimensional projections ℓ_j onto the vector spaces $\mathscr{E}_j$ spanned by φ_j. Moreover, for any $j\geq 1$, the measure ℓ_j has a density $\rho_j\in C^k(\mathscr{E}_j)$.*

Theorem 3.1. *Let us assume that a function F and a measure $\ell\in\mathcal{P}(\mathscr{E})$ satisfy Hypotheses (F) and (P) with some integer $k\geq 0$, and for any $\Upsilon\in\mathscr{X}$ and $\eta\in\operatorname{supp}\ell$ we have*

$$\mathrm{Image}\big(\partial_\eta F(\Upsilon,\eta)\big)=T_y\mathscr{Y}\,,\tag{3.1}$$

[11] The reader not willing to deal with infinite-dimensional manifolds may assume that $\mathscr{H}$ has the same structure as in Section 2.

where $T_y\mathscr{Y}$ stands for the tangent space of $\mathscr{Y}$ at $y = F(\Upsilon, \eta)$. Then, for any $\Upsilon \in \mathscr{X}$, the measure λ_Υ is absolutely continuous with respect to the volume measure on $\mathscr{Y}$, and the corresponding density $\rho(\Upsilon, y)$, defined for $\Upsilon \in \mathscr{X}$ and $y \in \mathscr{Y}$, is continuous in (Υ, y) and C^k-smooth in y. Moreover, there is $C > 0$ such that

$$\|\rho(\Upsilon_1, \cdot) - \rho(\Upsilon_2, \cdot)\|_{C^k(\mathscr{Y})} \le C\, d_{\mathscr{X}}(\Upsilon_1, \Upsilon_2) \quad \text{for all } \Upsilon_1, \Upsilon_2 \in \mathscr{X}. \tag{3.2}$$

Theorem 3.2. *Suppose that the hypotheses of Theorem 3.1 are satisfied with some integer $k \ge 0$, and let points $\widehat{\Upsilon} \in \mathscr{X}$, $\hat{\eta} \in \mathscr{E}$, $\hat{y} \in \mathscr{Y}$ be such that $F(\widehat{\Upsilon}, \hat{\eta}) = \hat{y}$ and*

$$\rho_j(\hat{\eta}_j) > 0 \quad \text{for all } j \ge 1, \tag{3.3}$$

where $\hat{\eta}_j = \langle \hat{\eta}, \varphi_j \rangle_{\mathscr{E}}$. Then $\rho(\widehat{\Upsilon}, \hat{y}) > 0$.

4 Application to the 2D Navier–Stokes system with a particle

4.1 Large Deviation Principle

In this section we prove Theorem 1.2. To this end, we shall make use of Theorem 2.1. According to that result, it suffices to check the validity of Hypotheses (R), (AC), (ACL), and (D). Let $s \ge 3$ be an integer, set $\mathcal{H} = V^s$, $\mathscr{Y} = \mathbb{T}^2$, $\mathscr{E} = L^2(J, V^s)$, where $J = [0, 1]$, and denote by $\mathcal{A}$ and $\mathscr{X}$ the sets $\mathcal{A}^s$ and $\mathcal{A}^s \times \mathbb{T}^2$, respectively. It is straightforward to see that any point $\Upsilon^0 = (0, p) \in \mathscr{X}^s$ satisfies the relation $S(\Upsilon^0, 0) = \Upsilon^0$. Moreover, in view of the regularising property of the Navier–Stokes system and infinite differentiability of its resolving operator with respect to the initial condition and the right-hand side (see Chapters I and VII in [BV92]), the regularity property (R) is satisfied with $\mathcal{V} = V^{s+1}$. Furthermore, it is easy to see that (D) follows immediately from (N). It remains to verify the validity of (AC) and (ACL).

Approximate controllability (AC)

We need to prove that any initial state $\Upsilon = (u_0, p) \in \mathscr{X}^s$ can be steered to the small neighbourhood of any target state $\widehat{\Upsilon} = (\hat{u}, \hat{p}) \in \mathscr{X}^s$ with the help of controls belonging to the support $\mathcal{K}^s$ of the law $\ell = \mathcal{D}(\eta_k)$. This will be done in three steps.

Step 1: Reduction to $\hat{u} = 0$. Suppose we can prove that, for an arbitrary $r > 0$ and $p_1 \in \mathbb{T}^2$, any point $\Upsilon = (u_0, p) \in \mathscr{X}^s$ can be steered to the r-neighbourhood of $(0, p_1)$ at some time $n_1 \ge 1$ depending only on r. Fix $\varepsilon > 0$ and points $\Upsilon, \widehat{\Upsilon} \in \mathscr{X}^s$. Recalling the definition of the sets $\mathcal{A}^s_k$ and using the fact that they form an increasing sequence, we can find an integer $n_2 \ge 1$, depending only

on $\varepsilon > 0$, an initial point $\Upsilon_1 = (0, p_1) \in \mathscr{X}^s$, and controls $\zeta_1^2, \dots, \zeta_{n_2}^2 \in \mathcal{K}^s$, such that

$$d_{\mathscr{X}^s}\big(S_{n_2}(\Upsilon_1; \zeta_1^2, \dots, \zeta_{n_2}^2), \widehat{\Upsilon}\big) \le \varepsilon/2.$$

The continuity of S implies the existence of a number $r > 0$ such that

$$d_{\mathscr{X}^s}\big(S_{n_2}(\Upsilon_1'; \zeta_1^2, \dots, \zeta_{n_2}^2), \widehat{\Upsilon}\big) \le \varepsilon, \tag{4.1}$$

for $d_{\mathscr{X}^s}(\Upsilon_1', \Upsilon_1) \le r$. By assumption, there are controls $\zeta_1^1, \dots, \zeta_{n_1}^1 \in \mathcal{K}^s$ such that

$$d_{\mathscr{X}^s}\big(S_{n_1}(\Upsilon; \zeta_1^1, \dots, \zeta_{n_1}^1), \Upsilon_1\big) \le r.$$

Combining this with inequality (4.1) in which $\Upsilon_1' = S_{n_1}(\Upsilon; \zeta_1^1, \dots, \zeta_{n_1}^1)$ and setting $n = n_1 + n_2$, we see that the controls $(\zeta_1, \dots, \zeta_n) = (\zeta_1^1, \dots, \zeta_{n_1}^1, \zeta_1^2, \dots, \zeta_{n_2}^2)$ are such that (2.1) holds.

Step 2: Reduction to $u_0 = 0$. Suppose that, given arbitrary $\delta > 0$ and $\hat{p} \in \mathbb{T}^2$, we can steer any point $\Upsilon_1 = (0, p_1) \in \mathscr{X}^s$ to the δ-neighbourhood of $(0, \hat{p})$ at some time $n_2 \ge 1$ depending only on δ. Fix $\varepsilon > 0$ and $\Upsilon = (u_0, p)$, $\widehat{\Upsilon} = (0, \hat{p})$ in $\mathscr{X}^s$. By continuity of S, we can find a number $r > 0$ such that, for any point $\Upsilon_1' = (u_1, p_1) \in \mathscr{X}^s$ satisfying the inequality $d_{\mathscr{X}^s}(\Upsilon_1', \Upsilon_1) \le r$ with $\Upsilon_1 = (0, p_1)$, there are controls $\zeta_1^2, \dots, \zeta_{n_2}^2 \in \mathcal{K}^s$ for which (4.1) holds. In view of the dissipativity of the homogeneous Navier–Stokes system in the space V^s, we can find an integer $n_1 \ge 1$ depending only on r such that

$$d_{\mathscr{X}^s}\big(S_{n_1}(\Upsilon; 0, \dots, 0), \Upsilon_1\big) \le r,$$

where $\Upsilon_1 = (0, p_1)$ with some $p_1 \in \mathbb{T}^2$. Combining this with (4.1), we see that (2.1) holds for the controls $(\zeta_1, \dots, \zeta_n) = (0, \dots, 0, \zeta_1^2, \dots, \zeta_{n_2}^2)$, where $n = n_1 + n_2$.

Step 3: Proof in the case $u_0 = \hat{u} = 0$. It suffices to prove that, for any points $p, \hat{p} \in \mathbb{T}^2$ whose distance from each other is less than a fixed number $\varkappa > 0$, we can find $\zeta \in \mathcal{K}^s$ satisfying $S(\Upsilon, \zeta) = \widehat{\Upsilon}$, where $\Upsilon = (0, p)$ and $\widehat{\Upsilon} = (0, \hat{p})$. Indeed, suppose this property is established and fix any points $\Upsilon, \widehat{\Upsilon} \in \mathscr{X}^s$ with zero u-component. We can find an integer $n \ge 1$, depending only on $\varkappa$, and points $\Upsilon_k = (0, p_k)$, $k = 1, \dots, n-1$, such that $|p_k - p_{k-1}| \le \varkappa$, where $p_0 = p$ and $p_n = \hat{p}$. Applying the above-mentioned exact controllability result, we can find $\zeta_1, \dots, \zeta_n \in \mathcal{K}^s$ such that $S(\Upsilon_{k-1}, \zeta_k) = \Upsilon_k$ for $1 \le k \le n$. This implies that $S_n(\Upsilon; \zeta_1, \dots, \zeta_n) = \widehat{\Upsilon}$.

We now use the argument described in Section 1.2 (see the scheme of the proof of Theorem 1.2) to establish the exact controllability for p and $\hat{p}$ that are close enough. Let us define an $\mathscr{X}^s$-valued curve $\Upsilon(t) = (u(t,x), y(t))$ by relations (1.18), in which $U_1, U_2, \gamma, \varphi_1, \varphi_2$ are as in (1.16) and (1.17). The endpoints of $\{\Upsilon(t), t \in J\}$ coincide with Υ and $\widehat{\Upsilon}$, and Eqs. (1.1), (0.6) hold with a right-hand side η given by (1.19). We need to prove that $\Pi g \in \mathcal{K}^s$.

To establish this, we write Πg in the form (1.20) and note that

$$\|\alpha_j\|_{C^r(J)} \le M_r |p - \hat{p}|, \tag{4.2}$$

where the numbers M_r do not depend p and $\hat{p}$. Expanding α_j in the orthonormal basis $\{\psi_l\}$,

$$\alpha_j(t) = \sum_{l\geq 1} \alpha_{jl}\psi_l(t),$$

we can rewrite (1.20) in the form

$$(\Pi g)(t,x) = \sum_{|j|_1\leq 2}\sum_{l\geq 1} b_j\alpha_{jl}\psi_l(t)e_j(x), \tag{4.3}$$

where $|j|_1 = |j_1| + |j_2|$. In view of (4.2) and Poincaré's inequality (1.6), we have

$$|\alpha_{jl}| \leq C_r M_r |p-\hat{p}|\, l^{-\theta r} \quad \text{for } |j|_1 \leq 2,\ l,r \geq 1. \tag{4.4}$$

On the other hand, it follows from (N) that if $\varepsilon > 0$ is sufficiently small, then any function h of the form (1.21) (where $\Lambda = \{|j|_1 \leq 2\}$) belongs to the support $\mathcal{K}^s$, provided that the coefficients h_{lj} satisfy the inequality

$$|h_{lj}| \leq \varepsilon l^{-\beta} \quad \text{for } |j|_1 \leq 2,\ l \geq 1. \tag{4.5}$$

Choosing $r \geq 1$ so large that $\theta r > \beta$ and assuming that $C_r M_r |p-\hat{p}| \leq \varepsilon$, we derive from (4.4) and (4.5) that $\Pi g \in \mathcal{K}^s$. This completes the proof of (AC).

Approximate controllability of the linearisation (ACL)

We need to prove the density of the image for the linear operator

$$(D_\eta S)(\Upsilon,\eta) : L^2(J,V^s) \to V^s \times T_{y(1)}\mathbb{T}^2,$$

where $y(1) = S^y(\Upsilon,\eta)$. In what follows, we identify the tangent space $T_y\mathbb{T}^2$ with $\mathbb{R}^2$.

We first note that, for any $\zeta \in \mathscr{E}$, the vector function $(D_\eta S)(\Upsilon,\eta)\zeta = (v,z)$ is a solution of the equations

$$\partial_t v + Lv + Q(u)v = \zeta, \tag{4.6}$$

$$\dot{z} - v\big(t,y(t)\big) - (D_x u)\big(t,y(t)\big)z(t) = 0, \tag{4.7}$$

where $L = -\nu\Pi\Delta$, $Q(u)v = \Pi(\langle u,\nabla\rangle v + \langle v,\nabla\rangle u)$, and $(u,y) = \boldsymbol{S}(\Upsilon,\eta)$. These equations are supplemented with the initial conditions

$$v(0) = 0, \quad z(0) = 0. \tag{4.8}$$

Since $s \geq 3$, it is easy to check the regularity of u and v is sufficient to ensure the well-posedness of (4.6), (4.7), (4.8). For given $\hat{v} \in V^s$, $\hat{q} \in \mathbb{R}^2$, and $\varepsilon > 0$, we need to find $\zeta \in L^2(J,V^s)$ such that

$$\|v(1) - \hat{v}\|_s < \varepsilon, \quad |z(1) - \hat{q}| < \varepsilon. \tag{4.9}$$

There is no loss of generality in assuming that $\hat{v} \in C^\infty(\mathbb{T}^2)$. Let $\gamma : J \to \mathbb{R}^2$ be defined by the relation

$$\gamma(t) = \alpha(t)\hat{q} + \beta(t)\big(\hat{v}(y(1)) + (D_x u)(1, y(1))\,\hat{q}\big), \tag{4.10}$$

where $\alpha, \beta \in C^\infty(J)$ are such that

$$\alpha(t) = \beta(t) = 0 \text{ for } t \le 1/3, \quad \alpha(t) = 1 \text{ for } t \ge 2/3, \quad \beta(1) = 0, \quad \dot{\beta}(1) = 1.$$

Writing $(\varphi_1(t), \varphi_2(t)) := \dot{\gamma}(t) - (D_x u)(t, y(t))\gamma(t)$ and choosing a small parameter $\delta > 0$, we set

$$v_\delta(t, x) = \theta_\delta(t)\big(\varphi_1(t)U_1(x - y(t)) + \varphi_2(t)U_2(x - y(t))\big) + \big(1 - \theta_\delta(t)\big)\hat{v},$$

where the functions U_i are defined in (1.16), and $\theta_\delta \in C^\infty(\mathbb{R})$ is such that $0 \le \theta_\delta \le 1$, $\theta_\delta(t) = 1$ for $t \le 1 - \delta$ and $\theta_\delta(t) = 0$ for $t \ge 1$. Finally, we denote by $z(t)$ the solution of (4.7) with zero initial condition. Then the vector function (v, z), defined on J, belongs to $\mathcal{X}_s \times C^1(J, \mathbb{R}^2)$, vanishes at $t = 0$, and satisfies (4.6), (4.7) on J with $\zeta \in C(J, V^s)$. Moreover, we have $v(1) = \hat{v}$. Therefore, to complete the proof, it remains to show that the second inequality in (4.9) holds for an appropriate choice of $\delta > 0$.

To establish this, we first note that there is a number R_1 not depending on δ such that

$$\sup_{(t,x)\in J\times\mathbb{T}^2} \Big(\big|v_\delta(t, x)\big| + \big|(D_x u)(t, x)\big|\Big) \le R_1. \tag{4.11}$$

Combining this with (4.7) and applying Gronwall's inequality, we can find R_2 such that

$$\sup_{t\in J} |\dot{z}_\delta(t)| \le R_2. \tag{4.12}$$

The choice of v_δ implies that the function $\gamma(t)$, vanishing at $t = 0$, is a solution of (4.7) on the interval $[0, 1 - \delta]$. By uniqueness of solutions for ODEs, we conclude that $z_\delta(1 - \delta) = \gamma(1 - \delta)$. Therefore, using (4.12), we derive

$$|z_\delta(1) - \hat{q}| \le |z_\delta(1) - z_\delta(1 - \delta)| + |z_\delta(1 - \delta) - \hat{q}| \le R_2\delta + |\gamma(1 - \delta) - \hat{q}|.$$

Since both right-most part terms of these inequalities go to zero as $\delta \to 0$, we conclude that $|z_\delta(1) - \hat{q}| < \varepsilon$ for $\delta \ll 1$. This completes the proof of (ACL) and of Theorem 1.2.

Remark 4.1. The approximate controllability property established above shows that the y-component of the linearised problem is exactly controllable. Namely, for any $\hat{q} \in \mathbb{R}^2$ there is $\zeta \in L^2(J, V^s)$ such that $(D_\eta S^y)(\Upsilon, \eta)\zeta = \hat{q}$. This follows immediately from the approximate controllability and the fact that the image of the derivative is a linear subspace in the finite-dimensional vector space $T_y\mathbb{T}^2$.

4.2 Regularity of laws and convergence

In this section we prove Theorem 1.3. The proof is divided into three steps.

Step 1: Reduction. Let us denote by $\mathscr{X}^\infty$ the intersection of the sets $\mathscr{X}^s$ with $s \ge 1$. Consider the following two statements:

(i) For any $\Upsilon \in \mathscr{X}^\infty$, the measure $\boldsymbol{\lambda}_t^\Upsilon$ has a density $\rho_t^\Upsilon \in C^\infty(\mathbb{T}^{2t})$.

(ii) For any $k \geq 1$ there is an integer $s \geq 3$ such that the mapping $\Upsilon \mapsto \rho_t^\Upsilon$ acting from $\mathscr{X}^\infty$ to $C^k(\mathbb{T}^{2t})$ is Lipschitz continuous with respect to the norm of $\mathscr{X}^s$.

Assuming that these two statements hold, the regularity part of Theorem 1.3 is deduced as follows. The regularising property of the Navier–Stokes system implies that, for any $\Upsilon \in \mathscr{X}^3$, the function $S(\Upsilon, \eta)$ belongs to $\mathscr{X}^\infty$ with probability 1. Recall the definition of the map F^t in Section 1.2 (see the scheme of the proof of Theorem 1.3), and note that $\boldsymbol{\lambda}_t^\Upsilon$ coincides with the image of the t-fold product of ℓ under the map $F^t(\Upsilon, \cdot)$. The independence of the random variables $\eta_1, \dots, \eta_t$ implies that

$$\boldsymbol{\lambda}_{[\![2,t]\!]}^\Upsilon = \int_{\mathscr{X}^\infty} F_*^{t-1}(\upsilon, \underbrace{\ell \otimes \cdots \otimes \ell}_{t-1 \text{ times}})\, \mathfrak{M}_1^\Upsilon(\mathrm{d}\upsilon) = \int_{\mathscr{X}^\infty} \boldsymbol{\lambda}_{t-1}^\upsilon\, \mathfrak{M}_1^\Upsilon(\mathrm{d}\upsilon). \tag{4.13}$$

Combining this with the property (i), we see that $\boldsymbol{\lambda}_{[\![2,t]\!]}^\Upsilon$ has a density given by

$$\rho_{[\![2,t]\!]}^\Upsilon(y_1, \dots, y_{t-1}) = \int_{\mathscr{X}^\infty} \rho_{t-1}^\upsilon(y_1, \dots, y_{t-1})\, \mathfrak{M}_1^\Upsilon(\mathrm{d}\upsilon). \tag{4.14}$$

Furthermore, for any integer $s \geq 3$, the map $S(\Upsilon, \eta)$ is Lipschitz continuous on $\mathscr{X}^3 \times \mathscr{E}$ with range in $\mathscr{X}^s$, where $\mathscr{E} = L^2(J, V^s)$. It follows that the same holds for the map $\Upsilon \mapsto \mathfrak{M}_1^\Upsilon$ acting from $\mathscr{X}^3$ to the space $\mathcal{P}(\mathscr{X}^s)$ endowed with the dual-Lipschitz metric. Denoting by C_s the corresponding Lipschitz constant and using the property (ii) and relation (4.14), for any $\Upsilon_1, \Upsilon_2 \in \mathscr{X}^3$ we can write

$$\begin{aligned}\big\|\rho_{[\![2,t]\!]}^{\Upsilon_1} - \rho_{[\![2,t]\!]}^{\Upsilon_2}\big\|_{C^k} &= \sup_{\alpha, y} \bigg| \int_{\mathscr{X}^s} (\partial^\alpha \rho_{t-1}^\upsilon)(y) \big(\mathfrak{M}_1^{\Upsilon_1} - \mathfrak{M}_1^{\Upsilon_2}\big)(\mathrm{d}\upsilon) \bigg| \\ &\leq C_s \big\|\dot{\rho}_{t-1}\big\|_{L_b(\mathscr{X}^s, C^k)} d_{\mathscr{X}^3}(\Upsilon_1, \Upsilon_2),\end{aligned}$$

where the supremum is taken over all $y \in \mathbb{T}^{2(t-1)}$ and $\alpha \in \mathbb{Z}_+^2$ with $|\alpha| \leq k$. We have thus established the Lipschitz continuity of the function $\Upsilon \mapsto \rho_{[\![2,t]\!]}^\Upsilon$ from $\mathscr{X}^3$ to C^k for any $k \geq 1$.

To prove that $\boldsymbol{\lambda}_t$ possesses a regular density, we note that the stationary measure $\mathfrak{M}$ is concentrated on $\mathscr{X}^\infty$. It follows that (cf. (4.14))

$$\boldsymbol{\lambda}_t = \int_{\mathscr{X}^\infty} \boldsymbol{\lambda}_t^\upsilon\, \mathfrak{M}(\mathrm{d}\upsilon).$$

In view of (i), this implies the existence of a density given by

$$\rho_t(y_1, \dots, y_t) = \int_{\mathscr{X}^\infty} \rho_t^\upsilon(y_1, \dots, y_t)\, \mathfrak{M}(\mathrm{d}\upsilon). \tag{4.15}$$

The infinite smoothness of ρ_t follows now from the property (ii).

Step 2: Proof of (i) *and* (ii). Fix an integer $k \geq 1$, set $s = k+3$, and recall that $\boldsymbol{\lambda}_t^{\Upsilon}$ is the image of the t-fold product $\ell^t = \ell \otimes \cdots \otimes \ell$ under the map

$$F^t : \mathscr{X}^s \times \underbrace{L^2(J, V^s) \times \cdots \times L^2(J, V^s)}_{t \text{ times}} \to \mathbb{T}^{2t}, \quad (\Upsilon, \eta) \mapsto (y_1, \ldots, y_t).$$

We claim that the hypotheses of Theorem 3.1 are satisfied for F^t. If this holds, then the statements (i) and (ii) follow immediately from the conclusions of Theorem 3.1.

In our setting, the manifold $\mathscr{Y}$ is the t-fold product $\mathbb{T}^{2t}$ of the two-dimensional torus, $\mathscr{X} = \mathscr{X}^s$, and $\mathscr{E}$ is the t-fold product of $L^2(J, V^s)$. Hence, to prove the existence of a C^k-smooth density and its Lipschitz continuity in $\Upsilon \in \mathscr{X}^s$, we need to check that the law ℓ^t satisfies (P), the map F^t is $(k+1)$-times continuously differentiable, and that the derivative of F^t is surjective.

It is straightforward to see that if Hypothesis (N) holds for ℓ, then so does (P) with the orthonormal basis $\{\varphi_j\}_{j \geq 1}$ obtained from $\{\psi_l e_j, l \geq 1, j \in \mathbb{Z}_*^2\}$ by normalisation. It follows that the product measure ℓ^t also satisfies (P). To prove that F^t is C^{k+1}, we recall that the resolving operator for the Navier–Stokes system is an infinitely smooth map from $V^r \times L^2(J, V^r)$ to $\mathcal{X}_r(J)$ for any $r \geq 2$. Moreover, standard results in the theory of ODEs imply that the solution of (0.6) is C^{r-2} function of the initial condition and the vector field $u \in \mathcal{X}_r(J)$, and the C^{k+1} regularity of $F^1(\Upsilon, \eta) = S^y(\Upsilon, \eta)$ follows. The C^{k+1} regularity of F^t follows now by recurrence.

Finally, we prove the surjectivity of the derivative $(D_{\boldsymbol{\eta}^t} F^t)(\Upsilon, \boldsymbol{\eta}^t)$, where $\boldsymbol{\eta}^t = (\eta_1, \ldots, \eta_t) \in \mathscr{E}$. We argue by induction. For $t = 1$, this is the claim of Remark 4.1. Assume that the statement holds for t. Then

$$F^{t+1}(\Upsilon, \boldsymbol{\eta}^{t+1}) = \big(F^t(\Upsilon, \boldsymbol{\eta}^t), S^y(\Upsilon_t, \eta_{t+1})\big), \quad \Upsilon_t = S_t(\Upsilon, \boldsymbol{\eta}^t).$$

It follows that $(D_{\boldsymbol{\eta}^{t+1}} F^{t+1})(\Upsilon, \boldsymbol{\eta}^{t+1})$ can be written as

$$\big((D_{\boldsymbol{\eta}^t} F^t)(\Upsilon, \boldsymbol{\eta}^t), (D_\Upsilon S^y)(\Upsilon_t, \eta_{t+1}) \circ (D_{\boldsymbol{\eta}^t} S_t)(\Upsilon, \boldsymbol{\eta}^t) + (D_\eta S^y)(\Upsilon_t, \eta_{t+1})\big).$$

Using the induction hypothesis and Remark 4.1, we conclude that this mapping is surjective.

Step 3: Convergence. It remains to prove the convergence part of Theorem 1.3. Note that, by Theorem 5.1,

$$\sup_{\Upsilon \in \mathscr{X}^3} \big\|\mathfrak{M}_n^{\Upsilon} - \mathfrak{M}\big\|_L^* \leq C\, e^{-\gamma n} \quad \text{for } n \geq 1, \tag{4.16}$$

where C and γ are positive numbers, and the dual-Lipschitz norm can be taken over $\mathscr{X}^s$ for any integer $s \geq 3$. It now follows from relations (4.15) and (1.23) that

$$\begin{aligned}
\|\rho_{[\![n+1,n+t]\!]}^{\Upsilon} - \rho_t\|_{C^k(\mathbb{T}^{2t})} &\leq \sup_{\alpha, y} \left| \int_{\mathscr{X}^s} (\partial^\alpha \rho_t^{\upsilon})(y) \big(\mathfrak{M}_n^{\Upsilon} - \mathfrak{M}\big)(\mathrm{d}\upsilon) \right| \\
&\leq \big\|\rho_t^{\cdot}\big\|_{L_b(\mathscr{X}^s, C^k)} \big\|\mathfrak{M}_n^{\Upsilon} - \mathfrak{M}\big\|_L^* \\
&\leq C e^{-\gamma n} \big\|\rho_t^{\cdot}\big\|_{L_b(\mathscr{X}^s, C^k)}.
\end{aligned}$$

This completes the proof of Theorem 1.3.

4.3 Strict positivity of densities

This section is devoted to the proof of Theorem 1.5, which was reduced to deriving inequality (1.27). To this end, we shall apply Theorem 3.2. As it was established in the proof of Theorem 1.3, if we take $s = 3$, then the map $S^y(u_0, \eta)$ acting from $\mathscr{E} := L^2(J, V^3)$ to $\mathbb{T}^2$ satisfies the hypotheses of Theorem 3.1 with $k = 1$, so that the measure $S^y_*(\Upsilon, \ell) = P_1(\Upsilon, \cdot)$ has a C^1-smooth density $\rho^\Upsilon_1(y)$ with respect to the Lebesgue measure on $\mathbb{T}^2$. Let us denote by $\{\varphi_{lj}(t,x)\}$ the orthonormal basis in $\mathscr{E}$ formed of the normalised orthogonal functions $\psi_l(t)e_j(x)$, where $l \geq 1$ and $j \in \mathbb{Z}^2_*$. To establish the strict positivity of ρ^Υ_1, it suffices to prove that, for any $\Upsilon = (u_0, p_0) \in \mathscr{X}^s$ and $\hat{p} \in \mathbb{T}^2$, there is $\hat{\eta} \in \mathcal{K}$ such that

$$S^y(\Upsilon, \hat{\eta}) = \hat{p}, \tag{4.17}$$

$$\rho_{lj}\big(\langle \hat{\eta}, \varphi_{lj}\rangle_{\mathscr{E}}\big) > 0 \quad \text{for all } l \geq 1,\ j \in \mathbb{Z}^2_*. \tag{4.18}$$

This is done in three steps. We first describe explicitly a subset $\mathcal{K}_0$ in the support $\mathcal{K}$ of the measure ℓ such that (4.18) holds for any $\hat{\eta} \in \mathcal{K}_0$. We next combine a result from the control theory of the Navier–Stokes equation (see Theorem 5.7) with the argument used in the verification of (AC) to construct a time-regular finite-dimensional control for which (4.17) holds. And, finally, we prove that the resulting control belongs to $\mathcal{K}_0$.

Step 1: Description of $\mathcal{K}_0$. Recall that the noise in (1.1) is now replaced by η^a, so that its restriction to the interval $J_k = [k-1, k]$ has the form (if we replace t by $t + k - 1$)

$$\eta^a_k = a \sum_{l=1}^{\infty} \sum_{j \in \mathbb{Z}^2_*} b_j d_j c_l \xi^k_{jl} \varphi_{lj}(t,x), \tag{4.19}$$

where $d_j = \|e_j\|_s^{-1}$. Let us denote by ℓ^a the law of η^a_k and by ρ^a_{lj} the density of its projection to the one-dimensional space spanned by φ_{lj}. Since the density of ξ^k_{lj} is positive on the interval $(-\delta, \delta)$, we have

$$\rho^a_{lj}(r) > 0 \quad \text{for } |r| < \delta_{lj} := a b_j d_j c_l \delta. \tag{4.20}$$

Thus, defining $\mathcal{K}_0$ as the set of functions $\psi \in \mathscr{E}$ such that $|\langle \psi, \varphi_{lj}\rangle_{\mathscr{E}}| < \delta_{lj}$, we see that any element of $\mathcal{K}_0$ satisfies (4.18).

Step 2: Construction of the control. We first reduce the problem to a result on approximate controllability of the Lagrangian component. Namely, let us fix a number $\tau > 0$ such that, for any $\hat{p} \in \mathbb{T}^2$, the ball $B_{\mathbb{T}^2}(\hat{p}, \tau)$ is homeomorphic to the unit disc in $\mathbb{R}^2$, and suppose that we have proved the following property for any $\Upsilon = (u_0, p_0) \in \mathscr{X}^s$:

(C) *There is a continuous map* $\Phi^\Upsilon : B_{\mathbb{T}^2}(\hat{p}, \tau) \to \mathscr{E}$ *whose image is contained in* $\mathcal{K}_0$ *such that*

$$\sup_{p \in B_{\mathbb{T}^2}(\hat{p}, \tau)} \big| S^y(\Upsilon, \Phi^\Upsilon(p)) - p \big| \leq \tau. \tag{4.21}$$

In this case, for any $\hat{p} \in \mathbb{T}^2$, the map

$$\Phi_1 : B_{\mathbb{T}^2}(\hat{p}, \tau) \to B_{\mathbb{T}^2}(\hat{p}, \tau), \quad \Phi_1(p) = p - S^y(\Upsilon, \Phi^\Upsilon(p)) + \hat{p},$$

is well defined and continuous. By Brouwer's theorem, there is a fixed point $\bar{p} \in B_{\mathbb{T}^2}(\hat{p}, \tau)$ for Φ_1, and it is easy to see that the function $\eta = \Phi^\Upsilon(\bar{p})$ belongs to $\mathcal{K}_0$ and satisfies (4.17).

The proof of Property (C), with an arbitrary $\tau > 0$, is based on the Agrachev–Sarychev theorem (see Theorem 5.7 in Section 5.5) and a modification of the construction used in Section 4.1 to establish (AC). Let us fix a small parameter $\varkappa > 0$ that will be chosen below and define the space $\mathcal{H}_1 = \operatorname{span}\{e_j : |j|_1 \le 2\}$. By the Agrachev–Sarychev theorem, there is a number $\delta_1 > 0$ and a continuous mapping $\Psi_1 : \mathscr{X}^s \to C^\infty([0, 1/2], \mathcal{H}_1)$ such that

$$\operatorname{supp} \Psi_1(\Upsilon) \subset [\delta_1, 1/2 - \delta_1], \tag{4.22}$$

$$\sup_{\Upsilon \in \mathscr{X}^s} \|S^u_{1/2}\big(\Upsilon, \Psi_1(\Upsilon)\big)\|_s \le \varkappa. \tag{4.23}$$

We now set $p_1 = S^y_{1/2}(\Upsilon, \Psi_1(\Upsilon))$ and note that $p_1 = p_1(\Upsilon)$ is a continuous $\mathbb{T}^2$-valued function of $\Upsilon \in \mathscr{X}^s$. We define a map $\gamma : [1/2, 1] \to \mathbb{T}^2$ by

$$\gamma(t) = \gamma(t; \Upsilon, p) = \big(1 - \alpha(t)\big)p_1(\Upsilon) + \alpha(t)p, \tag{4.24}$$

where $\alpha \in C^\infty(\mathbb{R})$, $\alpha(t) = 0$ for $t \le 2/3$ and $\alpha(t) = 1$ for $t \ge 5/6$, so that $\gamma(1/2) = p_1(\Upsilon)$ and $\gamma(1) = p$. Recalling that U_1 and U_2 were introduced in (1.16) and defining $\varphi = (\varphi_1, \varphi_2)$ as the time-derivative of γ, for $1/2 \le t \le 1$ we set (cf. (1.17) and (1.18))

$$\bar{u}(t, x) = \varphi_1(t)U_1\big(x - \gamma(t)\big) + \varphi_2(t)U_2\big(x - \gamma(t)\big), \quad \bar{y}(t) = \gamma(t; \Upsilon, p).$$

It is straightforward to check that the pair $(\bar{u}(t, x), \bar{y}(t))$ defined on the time interval $[1/2, 1]$ is a solution of (1.1), (0.6) with the right-hand side

$$\eta(t) = \Psi_2(\Upsilon, p)(t) := \Pi g(t), \quad g(t) = \partial_t \bar{u} + \langle \bar{u}, \nabla \rangle \bar{u} - \nu \Delta \bar{u}. \tag{4.25}$$

Moreover, the construction implies that Ψ_2 is continuous in $(\Upsilon, p) \in \mathscr{X}^s \times \mathbb{T}^2$ with range in $C^\infty([1/2, 1])$, and that

$$\bar{u}(t) = 0 \quad \text{for } 1/2 \le t \le 2/3, \qquad \bar{y}(1/2) = p_1(\Upsilon), \qquad \bar{y}(1) = p. \tag{4.26}$$

By continuity, we can find $\varkappa > 0$ such that, if $\Upsilon_1 = (u_1, p_1(\Upsilon)) \in \mathscr{X}^s$ satisfies the inequality $\|u_1\|_s \le \varkappa$, then the solution (u, y) of (1.1), (0.6) issued from Υ_1 at time $t = 1/2$ satisfies the inequality $|y(1) - p| \le \tau$. Thus, defining $\Phi^\Upsilon : \mathscr{X}^s \to \mathscr{E}$ by the relation

$$\Phi^\Upsilon(p) = \begin{cases} \Psi_1(\Upsilon)(t) & \text{for } 0 \le t \le 1/2, \\ \Psi_2(\Upsilon, p)(t) & \text{for } 1/2 \le t \le 1, \end{cases}$$

we see that (4.21) holds. Moreover, the construction implies that the image of $\Phi^{\Upsilon}(p)$ is an $\mathcal{H}_1$-valued function of t that vanishes in the neighbourhood of $t = 1/2$ and whose restrictions to the intervals $[0, 1/2]$ and $[1/2, 1]$ are infinitely smooth, so that $\Phi^{\Upsilon}(p) \in C^\infty(J, V^s)$. Finally, the continuity properties of Ψ_1 and Ψ_2 imply that $\Phi^{\Upsilon}(p)$ is a continuous function of $(\Upsilon, p) \in \mathscr{X}^s \times \mathbb{T}^2$ with range in $C^\infty(J, V^s)$. Thus, it remains to prove that its image is contained in $\mathcal{K}_0$.

Step 3: Description of the image. Let us take an arbitrary $(\Upsilon, p) \in \mathscr{X}^s \times \mathbb{T}^2$ and consider the function $\psi = \Phi^{\Upsilon}(p)$. Writing it in the form

$$\psi(t, x) = \sum_{l=1}^{\infty} \sum_{j \in \mathbb{Z}^2_*} \psi_{lj} \varphi_{lj}(t, x), \tag{4.27}$$

where $\psi_{lj} = \psi_{lj}(\Upsilon, p) := \langle \psi, \varphi_{lj} \rangle_{\mathscr{E}}$, and recalling the definition of $\mathcal{K}_0$, we see that it suffices to prove the inequalities

$$|\psi_{lj}| < \delta_{lj} \quad \text{for all } l \geq 1 \text{ and } j \in \mathbb{Z}^2_*, \tag{4.28}$$

where δ_{lj} is defined in (4.20). Note that the function ψ takes values in $\mathcal{H}_1$, so that $\psi_{lj} = 0$ for $|j|_1 > 2$. On the other hand, in view of Poincaré's inequality (1.6), for $|j|_2 \leq 2$ we have

$$|\psi_{lj}| \leq C_r l^{-\theta r}, \quad l, r \geq 1,$$

where the numbers C_r do not depend on $(\Upsilon, p) \in \mathscr{X}^s \times \mathbb{T}^2$. Thus, the coefficient ψ_{lj} will satisfy (4.28) if

$$C_r l^{-\theta r} < a b_j d_j c_l \delta \quad \text{for } |j|_2 \leq 2,\ l \geq 1. \tag{4.29}$$

Since $r \geq 1$ can be chosen arbitrarily large, recalling (1.9), we see that (4.29) is certainly satisfied if $a > 0$ is sufficiently large. This completes the proof of Theorem 1.5.

Remark 4.2. The above proof gives that the conclusions of Theorem 1.5 remain valid if only the Fourier components e_j with $|j|_1 \leq 2$ are multiplied by a large parameter.

Remark 4.3. Since the control function η entering (1.1) has full range, we could have avoided use of the Agrachev–Sarychev theorem by connecting a given initial function u_0 with zero and defining the corresponding control by relation (4.25). However, with this construction, the control $\Psi_1(\Upsilon)$ on the interval $[0, 1/2]$ would have much larger dimension, and Remark 4.2 would not be valid.

5 Appendix

In this section, we recall various results used in the main text. Section 5.1 deals with the property of exponential mixing and construction of coupling operators. In Section 5.2, we formulate Kifer's criterion for the validity of LDP. In Section 5.3, we discuss the large-time asymptotics of generalised Markov semigroups. Section 5.4 is the devoted to the proof of the theorems on the image of probability measures under smooth maps. Finally, in Section 5.5, we recall the Agrachev–Sarychev theorem.

5.1 Exponential mixing and coupling operators

In this section, we discuss the problem of uniqueness of stationary measure and exponential mixing for the RDS (1.12) in the phase space $\mathscr{X} = \mathcal{A} \times \mathscr{Y}$, where $\mathcal{A}$ is a compact subset of a separable Hilbert space and $\mathscr{Y}$ is a compact manifold (without boundary). We denote by $P_1(\Upsilon, \Gamma)$ the transition function for (1.12), and by $\mathfrak{P}_k$ and $\mathfrak{P}_k^*$ the corresponding Markov operators acting in the spaces $C(\mathscr{X})$ and $\mathcal{P}(\mathscr{X})$, respectively. The following result has been essentially established in [KNS18, Shi19]; however, since our setting here is slightly different, we give a short proof.

Theorem 5.1. *Assume that Hypotheses (R), (AC), (ACL), and (D) are satisfied. Then the discrete-time Markov process associated with (1.12) has a unique stationary measure $\mu \in \mathcal{P}(\mathscr{X})$. Moreover, there are positive numbers γ and C such that*

$$\|\mathfrak{P}_k^* \nu - \mu\|_L^* \le C e^{-\gamma k} \quad \text{for } \nu \in \mathcal{P}(\mathscr{X}),\ k \ge 0. \tag{5.1}$$

Proof. Set $\boldsymbol{D}_\delta = \{(\Upsilon, \Upsilon') \in \mathscr{X} \times \mathscr{X} : d_{\mathscr{X}}(\Upsilon, \Upsilon') \le \delta\}$. In view of Theorem 1.1 in [Shi19], it suffices to prove that the following local stabilisation property holds:

(LS) *For any $R > 0$ and any compact set $\mathscr{K} \subset \mathscr{E}$ there is a finite-dimensional subspace $\mathcal{E} \subset \mathscr{E}$, positive numbers C, δ, and $q < 1$, and a continuous mapping*

$$\varPhi : \boldsymbol{D}_\delta \times B_{\mathscr{E}}(R) \to \mathcal{E}, \quad (\Upsilon, \Upsilon', \eta) \mapsto \eta',$$

which is continuously differentiable in η and satisfies the following inequalities for any $(\Upsilon, \Upsilon') \in \boldsymbol{D}_\delta$:

$$\sup_{\eta \in B_{\mathscr{E}}(R)} \big(\|\varPhi(\Upsilon, \Upsilon', \eta)\|_{\mathscr{E}} + \|D_\eta \varPhi(\Upsilon, \Upsilon', \eta)\|_{\mathcal{L}(\mathscr{E})} \big) \le C\, d_{\mathscr{X}}(\Upsilon, \Upsilon'), \tag{5.2}$$

$$\sup_{\eta \in \mathscr{K}} d_{\mathscr{X}}\big(S(\Upsilon, \eta), S(\Upsilon', \eta + \varPhi(\Upsilon, \Upsilon', \eta))\big) \le q\, d_{\mathscr{X}}(\Upsilon, \Upsilon'). \tag{5.3}$$

We shall show that Hypotheses (R) and (ACL) imply (LS). Let us recall that, given a separable Hilbert space H and a continuous linear operator $A : \mathscr{E} \to H$ with a dense range, we can construct an approximate inverse for A in the following way. Setting $G = AA^*$, it is easy to check that

$$G(G + \gamma I)^{-1} f \to f \quad \text{as } \gamma \to 0^+ \text{ for any } f \in H,$$

so that $A^*(G + \gamma I)^{-1}$ is an approximate right inverse of A. Moreover, if P_n are finite-dimensional projections converging to the identity in $\mathscr{E}$ for the strong operator topology (see (D)), then considering the operators $\mathsf{P}_M A^*(G + \gamma I)^{-1}$, we obtain a family of continuous finite-dimensional operators $R_\varepsilon : H \to \mathscr{E}$ such that $AR_\varepsilon \to I$ in the strong operator topology.

We now apply this procedure to the derivatives $A(\Upsilon, \eta) := (D_\eta S)(\Upsilon, \eta)$ acting continuously from $\mathscr{E}$ to $\mathcal{H} \times T_y \mathscr{Y}$, where $y = \Pi_{\mathscr{Y}} S(\Upsilon, \eta)$. Set

$$G(\Upsilon, \eta) = A(\Upsilon, \eta)\, A(\Upsilon, \eta)^*, \quad R_{M,\gamma}(\Upsilon, \eta) = \mathsf{P}_M A(\Upsilon, \eta)^* (G(\Upsilon, \eta) + \gamma I)^{-1}.$$

Repeating the compactness argument used in the proof of Proposition 2.3 in [KNS18], it is not difficult to show that, for any $\varepsilon > 0$, there are $M_\varepsilon \geq 1$ and $\gamma_\varepsilon > 0$ such that the operator $R_\varepsilon(\Upsilon, \eta) = R_{M_\varepsilon, \gamma_\varepsilon}(\Upsilon, \eta)$ (which acts continuously from $\mathscr{H}_y := \mathcal{H} \times T_y\mathscr{Y}$ to $\mathscr{E}$ and has a finite-dimensional range) satisfies the following inequalities for any $\Upsilon \in \mathscr{X}$, $\eta \in \mathcal{K}$, and $f \in \mathcal{V} \times T_y\mathscr{Y}$:

$$\big\|(D_\eta S)(\Upsilon, \eta) R_\varepsilon(\Upsilon, \eta) f - f\big\|_{\mathscr{H}_y} \leq \varepsilon \, \|f\|_{\mathcal{V} \times T_y\mathscr{Y}}, \tag{5.4}$$

$$\big\|R_\varepsilon(\Upsilon, \eta)\big\|_{\mathcal{L}(\mathscr{H}_y, \mathscr{E})} + \big\|(D_\eta R_\varepsilon)(\Upsilon, \eta)\big\|_{\mathcal{L}(\mathscr{H}_y \times \mathscr{E}, \mathscr{E})} \leq C_1(\varepsilon), \tag{5.5}$$

where $C_1(\varepsilon)$ does not depend on (Υ, η), and the tangent space $T_y\mathscr{Y}$ is endowed with the norm induced by the Riemannian metric of $\mathscr{Y}$. We now fix points $\Upsilon = (u, y) \in \mathscr{X}$ and $\eta \in B_{\mathscr{E}}(R)$, together with some local charts around y and $\Pi_{\mathscr{Y}} S(\Upsilon, \eta)$, and use Taylor's formula to write

$$S(\Upsilon', \eta') - S(\Upsilon, \eta) = (D_\Upsilon S)(\Upsilon, \eta)(\Upsilon' - \Upsilon) + (D_\eta S)(\Upsilon, \eta)(\eta' - \eta) + r(\Upsilon, \Upsilon', \eta, \eta'). \tag{5.6}$$

Here Υ' and η' are sufficiently close to Υ and η, respectively, so that $\Pi_{\mathscr{Y}}\Upsilon$ and $\Pi_{\mathscr{Y}} S(\Upsilon', \eta')$ belong to the above-mentioned local charts[12], and the remainder term r satisfies the inequality

$$\big\|r(\Upsilon, \Upsilon', \eta, \eta')\big\|_{\mathscr{H}} \leq C_2(R) \left(\big\|\Upsilon - \Upsilon'\big\|^2_{\mathscr{H}} + \big\|\eta - \eta'\big\|^2_{\mathscr{E}} \right), \tag{5.7}$$

where $\Upsilon \in \mathscr{X}$ and $\eta \in B_{\mathscr{E}}(R)$. Define

$$\Phi(\Upsilon, \Upsilon', \eta) = -R_\varepsilon(\Upsilon, \eta)(D_\Upsilon S)(\Upsilon, \eta)(\Upsilon' - \Upsilon), \tag{5.8}$$

and note that

$$\big\|(D_\Upsilon S)(\Upsilon, \eta)(\Upsilon' - \Upsilon)\big\|_{\mathcal{V} \times T_y\mathscr{Y}} \leq C_3(R) \big\|\Upsilon' - \Upsilon\big\|_{\mathcal{H} \times T_y\mathscr{Y}}. \tag{5.9}$$

Combining this inequality with (5.5) and Hypothesis (R), we see that (5.2) holds. Furthermore, it follows from (5.4)–(5.9) that

$$\begin{aligned}
&\big\|S(\Upsilon, \eta) - S(\Upsilon', \eta + \Phi(\Upsilon, \Upsilon', \eta))\big\|_{\mathscr{H}} \\
&\quad \leq \varepsilon \, \big\|(D_\Upsilon S)(\Upsilon, \eta)(\Upsilon' - \Upsilon)\big\|_{\mathcal{V} \times T_y\mathscr{Y}} + \big\|r(\Upsilon, \Upsilon', \eta, \eta + \Phi(\Upsilon, \Upsilon', \eta))\big\|_{\mathscr{H}} \\
&\quad \leq \big(C_3(R)\varepsilon + C_4(R) \, \|\Upsilon' - \Upsilon\|_{\mathscr{H}}\big) \big\|\Upsilon' - \Upsilon\big\|_{\mathscr{H}} \leq q \, \big\|\Upsilon' - \Upsilon\big\|_{\mathscr{H}},
\end{aligned}$$

where $\Upsilon, \Upsilon' \in \boldsymbol{D}_\delta$ and $\eta \in B_{\mathscr{E}}(R)$, the number $q > 0$ is arbitrary, and the positive numbers ε and δ are sufficiently small. This gives inequality (5.3) and proves the local stabilisability. □

Remark 5.2. The above proof gives that Hypothesis (AC) in Theorem 5.1 can be replaced by the following weaker variant which requires approximate controllability to some fixed point $\widehat{\Upsilon} \in \mathscr{X}$:

[12] This enables one to write the differences $\Upsilon' - \Upsilon$ and $S(\Upsilon', \eta') - S(\Upsilon, \eta)$ and to consider them as elements of $\mathcal{H} \times \mathbb{R}^d$. In particular, we shall write $\|\Upsilon' - \Upsilon\|_{\mathscr{H}}$ for the distance between Υ and Υ'.

(ACP) For any $\varepsilon > 0$, there is an integer $n \geq 1$ such that, for any initial point $\Upsilon \in \mathscr{X}$, one can find controls $\zeta_1, \dots, \zeta_n \in \mathcal{K}$ satisfying inequality (2.1).

In the proof of the LDP, we also needed the existence of coupled trajectories that converge to each other exponentially fast. Their existence is established with the help of the following coupling construction.

Proposition 5.3. *Under the hypotheses of Theorem 5.1, for any $q \in (0,1)$, there is a number $C > 0$, a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and measurable mappings*

$$\mathcal{R}, \mathcal{R}' : \mathscr{X} \times \mathscr{X} \times \Omega \to \mathscr{X},$$

such that, for any $\Upsilon, \Upsilon' \in \mathscr{X}$, the pair $(\mathcal{R}(\Upsilon, \Upsilon', \cdot), \mathcal{R}'(\Upsilon, \Upsilon', \cdot))$ is a coupling for $(P_1(\Upsilon, \cdot), P_1(\Upsilon', \cdot))$, and

$$\mathbb{P}\big\{\|\mathcal{R}(\Upsilon, \Upsilon', \cdot) - \mathcal{R}'(\Upsilon, \Upsilon', \cdot)\|_{\mathscr{H}} > q\,\|\Upsilon - \Upsilon'\|_{\mathscr{H}}\big\} \leq C\,\|\Upsilon - \Upsilon'\|_{\mathscr{H}}. \tag{5.10}$$

This result has been proved in [Shi19] (see Step 3 of the proof of Theorem 1.1), provided that Hypotheses (R), (LS), and (D) are satisfied. As was established in the proof of Theorem 5.1, under Hypotheses (R) and (ACL) property (LS) holds with any $q < 1$, so that Proposition 5.3 also holds.

We now describe the construction of coupled trajectories. Let $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ be the product of countably many copies of the probability space constructed in Proposition 5.3. For $\Upsilon, \Upsilon' \in \mathscr{X}$ and $\omega = (\omega^1, \omega^2, \dots) \in \widetilde{\Omega}$, we set

$$\begin{aligned} \widetilde{\Upsilon}_0 &= \Upsilon, & \widetilde{\Upsilon}'_0 &= \Upsilon', \\ \widetilde{\Upsilon}_k &= \mathcal{R}(\widetilde{\Upsilon}_{k-1}, \widetilde{\Upsilon}'_{k-1}, \omega^k), & \widetilde{\Upsilon}'_k &= \mathcal{R}'(\widetilde{\Upsilon}_{k-1}, \widetilde{\Upsilon}'_{k-1}, \omega^k), \end{aligned}$$

where $k \geq 1$. For any $q \in (0,1)$ and $k \geq 1$, define the event

$$G_k(q, \Upsilon, \Upsilon') := \big\{d_{\mathscr{X}}(\Upsilon_k, \Upsilon'_k) > q\, d_{\mathscr{X}}(\Upsilon_{k-1}, \Upsilon'_{k-1})\big\}.$$

The following result is a straightforward consequence of Proposition 5.3, and its proof, based on an application of the Markov property, can be carried out by repeating the argument in [KS12, Section 3.2.2] (see also [Shi15, Section 4.4]).

Corollary 5.4. *Under the hypotheses of Theorem 5.1, the trajectories $\{\widetilde{\Upsilon}_k\}$ and $\{\widetilde{\Upsilon}'_k\}$ constructed above have the following properties.*

Coupling. *The laws of the processes $\{\widetilde{\Upsilon}_k\}_{k\geq 0}$ and $\{\widetilde{\Upsilon}'_k\}_{k\geq 0}$ regarded as random variables with range in $\boldsymbol{X} = \mathscr{X}^{\mathbb{Z}_+}$ coincide with those of the trajectories for (1.12) issued from the initial points Υ and Υ', respectively.*

Estimate. *There is a number $C > 0$ depending on $q \in (0,1)$ such that, for any integer $k \geq 1$ and any points $\Upsilon, \Upsilon' \in \mathscr{X}$, we have*

$$\widetilde{\mathbb{P}}\big(G_k(q, \Upsilon, \Upsilon')\big) \leq C\,\mathbb{E}\, d_{\mathscr{X}}(\Upsilon_{k-1}, \Upsilon'_{k-1}). \tag{5.11}$$

5.2 Kifer's criterion

Let X be a compact metric space, let Θ be a directed set, let $\{r_\theta\}_{\theta\in\Theta}$ be a non-decreasing net of positive numbers converging to $+\infty$, and let $\{\mu_\theta\}_{\theta\in\Theta}$ be a net of random probability measures on X with an underlying space $(\Omega, \mathcal{F}, \mathbb{P})$. We assume that, for any $V \in C(X)$, the following limit exists:

$$Q(V) = \lim_{\theta\in\Theta} r_\theta^{-1} \log \mathbb{E} \exp\bigl(r_\theta \langle V, \mu_\theta\rangle\bigr). \tag{5.12}$$

This is a 1-Lipschitz function on $C(X)$ such that $Q(V + C) = Q(V) + C$ for any $V \in C(X)$ and $C \in \mathbb{R}$. Let $I : \mathcal{M}(X) \to [0, +\infty]$ be the Legendre transform of Q:

$$I(\sigma) = \begin{cases} \sup\limits_{V\in C(X)} \bigl(\langle V, \mu\rangle - Q(V)\bigr) & \text{for } \sigma \in \mathcal{P}(X), \\ +\infty & \text{otherwise.} \end{cases} \tag{5.13}$$

It is easy to see that I is a good rate function (see Section 2.1 for a definition). Recall that $\sigma \in \mathcal{P}(X)$ is called an *equilibrium state* for $V \in C(X)$ if

$$Q(V) = \langle V, \sigma\rangle - I(\sigma).$$

The following theorem is due to Kifer [Kif90].

Theorem 5.5. *In addition to the existence of limit (5.12), suppose that there exists a dense vector space $\mathcal{V} \subset C(X)$ such that for any $V \in \mathcal{V}$ there is a unique equilibrium state $\sigma \in \mathcal{P}(X)$. Then the LDP holds for $\{\mu_\theta\}$ with the speed $\{r_\theta\}$ and the rate function I.*

5.3 Asymptotics of Feynman–Kac semigroups

Let X be a compact metric space, let $\mathcal{M}_+(X)$ be the set of non-negative Borel measures on X, and let $\{P(x, \cdot), x \in X\} \subset \mathcal{M}_+(X)$ be a family such that $P(x, X) > 0$ for any $x \in X$, and the mapping $x \mapsto P(x, \cdot)$ is continuous from X to the space $\mathcal{M}_+(X)$ endowed with the weak* topology. We denote by $P_k(x, \cdot)$ the k-fold iteration of $P(x, \cdot)$, and by

$$\mathfrak{P}_k : C(X) \to C(X), \quad \mathfrak{P}_k^* : \mathcal{M}(X) \to \mathcal{M}(X)$$

the semigroups with the generators $\mathfrak{P}$ and $\mathfrak{P}^*$ defined by

$$(\mathfrak{P}f)(x) = \int_X P(x, \mathrm{d}y) f(y), \quad (\mathfrak{P}^*\sigma)(\Gamma) = \int_X P(x, \Gamma)\sigma(\mathrm{d}x),$$

where $f \in C(X)$, $\sigma \in \mathcal{M}(X)$, and $\Gamma \in \mathcal{B}(X)$. Note that $P_k(x, \cdot)$ is the kernel of the operator $\mathfrak{P}_k$. Recall that a family $\mathcal{C} \subset C(X)$ is said to be *determining* if, for any two measures $\mu, \nu \in \mathcal{M}_+(X)$, the validity of the relation $\langle f, \mu\rangle = \langle f, \nu\rangle$ for all $f \in \mathcal{C}$ implies that $\mu = \nu$. In addition to the above hypotheses, let us assume that the following two properties hold.

(UI) Uniform irreducibility. For any $\varepsilon > 0$ there is an integer $n \geq 1$ and a number $p > 0$ such that

$$P_n\big(x, B_X(\hat{x}, \varepsilon)\big) \geq p \quad \text{for all } x, \hat{x} \in X. \tag{5.14}$$

(UF) Uniform Feller property. *There is a determining family $\mathcal{C} \subset C(X)$ such that, for any $f \in \mathcal{C}$, the sequence $\{\|\mathfrak{P}_k \mathbf{1}\|_\infty^{-1} \mathfrak{P}_k f, k \geq 0\}$ is uniformly equicontinuous.*

The following theorem is established in [JNPS15b, Section 2].

Theorem 5.6. *Under the above hypotheses, there is a number $\lambda > 0$, a measure $\mu \in \mathcal{P}(X)$, and a positive function $h \in C(X)$ such that $\langle h, \mu \rangle = 1$ and*

$$\mathfrak{P}_1 h = \lambda h, \quad \mathfrak{P}_1^* \mu = \lambda \mu, \tag{5.15}$$

$$\lambda^{-k} \mathfrak{P}_k f \to \langle f, \mu \rangle h \quad \text{in } C(X) \text{ as } k \to \infty, \tag{5.16}$$

$$\lambda^{-k} \mathfrak{P}_k^* \sigma \to \langle h, \sigma \rangle \mu \quad \text{in } \mathcal{M}_+(X) \text{ as } k \to \infty, \tag{5.17}$$

where $f \in C(X)$ and $\sigma \in \mathcal{M}_+(X)$ are arbitrary.

5.4 Proofs of Theorems 3.1 and 3.2

Proof of Theorem 3.1. We follow the arguments in [AKSS07] and [Bog10, Section 9.6]. The proof in our case is simpler since the derivative of F has full rank everywhere.

Step 1: Localisation in Υ. We first prove that it suffices to establish the result in the neighbourhood of each point $\Upsilon \in \mathscr{X}$. Namely, suppose that for any $\Upsilon \in \mathscr{X}$ there is $\delta = \delta_\Upsilon > 0$ such that, for any $\Upsilon' \in B_{\mathscr{H}}(\Upsilon, \delta)$, the measure $F_*(\Upsilon', \ell)$ has a C^k-smooth density $\rho(\Upsilon', y)$ that satisfies the inequality in (3.2) for $\Upsilon_1, \Upsilon_2 \in B_{\mathscr{H}}(\Upsilon, \delta)$. In this case, the existence and regularity of the density for $F_*(\Upsilon, \ell)$ with $\Upsilon \in \mathscr{X}$ is trivial. To prove inequality (3.2), we consider the open cover $\{\dot{B}_{\mathscr{H}}(\Upsilon, \delta_\Upsilon/2)\}_{\Upsilon \in \mathscr{X}}$ of the compact set $\mathscr{X}$ and select a finite sub-cover $\{\mathcal{O}^m, 1 \leq m \leq M\}$, where $\mathcal{O}^m = \dot{B}_{\mathscr{H}}(\Upsilon^m, \delta_m/2)$ for some $\Upsilon^m \in \mathscr{X}$ and $\delta_m > 0$. Denoting by δ the minimum of the numbers δ_m, $1 \leq m \leq M$, we note that it suffices to establish (3.2) for points Υ_1, Υ_2 satisfying $d(\Upsilon_1, \Upsilon_2) < \delta$. For any such pair, we can find $m \in [\![1, M]\!]$ such that $\Upsilon_1, \Upsilon_2 \in B_{\mathscr{H}}(\Upsilon^m, \delta_m)$, so that (3.2) holds by assumption.

Step 2: Localisation in η. Let us fix any $\Upsilon \in \mathscr{X}$ and assume that for any point $\eta \in \mathcal{K}$ we can find positive numbers δ_η and γ_η such that, if a measure $\ell \in \mathcal{M}(\mathscr{E})$ satisfies (P), and $\psi : \mathscr{E} \to \mathbb{R}$ is a C^∞-function with a support contained in $B_{\mathscr{E}}(\eta, \gamma_\eta)$, then $F_*(\Upsilon, \psi\ell)$ has a C^k-smooth density $\rho_\psi(\Upsilon, \cdot)$ for $\Upsilon \in B_{\mathscr{H}}(\Upsilon, \delta_\eta)$, and inequality (3.2) holds for $\Upsilon_1, \Upsilon_2 \in B_{\mathscr{H}}(\Upsilon, \delta_\eta)$. In this case, we consider the open cover $\{\dot{B}_{\mathscr{E}}(\eta, \gamma_\eta)\}_{\eta \in \mathcal{K}}$ of the compact set $\mathcal{K}$ and choose a finite sub-cover $\{\mathcal{U}^m, 1 \leq m \leq M\}$, where $\mathcal{U}^m = \dot{B}_{\mathscr{E}}(\eta^m, \gamma_m)$ for some $\eta^m \in \mathcal{K}$ and $\gamma_m > 0$.

Let $\{\psi^m\}$ be an infinitely smooth partition of unity on $\mathcal{K}$ subordinate to $\{\mathcal{U}^m\}$; see [Lan85, Section II.3]. Then we can write

$$\lambda_{\Upsilon'} = F_*(\Upsilon', \ell) = \sum_{m=1}^{M} F_*(\Upsilon', \psi^m \ell). \tag{5.18}$$

Setting $\delta = \min\{\delta_m, 1 \le m \le M\}$, where $\delta_m > 0$ is the number corresponding to η^m, we see that, for any $\Upsilon' \in B_{\mathscr{H}}(\Upsilon, \delta)$, each term of the sum in (5.18) possesses a C^k-smooth density $\rho_m(\Upsilon', y)$ that satisfies (3.2) for $\Upsilon_1, \Upsilon_2 \in B_{\mathscr{H}}(\Upsilon, \delta)$. This proves the required property.

Step 3: Proof in the localised case. It remains to establish the property described in the beginning of Step 2. Let us fix any $\widehat{\Upsilon} \in \mathscr{X}$ and $\hat{\eta} \in \mathcal{K}$ and choose a local chart in the neighbourhood of $\hat{y} = F(\widehat{\Upsilon}, \hat{\eta})$, so that F can be written as $(F_1, \dots, F_d)$. By (3.1), we can find vectors $g_1, \dots, g_d \in \{\varphi_j\}_{j \ge 1}$ such that

$$D(\widehat{\Upsilon}, \hat{\eta}) := \big|\det\big(D_{g_i} F_j(\widehat{\Upsilon}, \hat{\eta})\big)\big| > 0. \tag{5.19}$$

Let us denote by $\mathscr{E}_1$ the vector span of $g_1, \dots, g_d$ and by $\mathscr{E}_2$ its orthogonal complement, so that we can write $\eta = \eta^1 + \eta^2$ with $\eta^i \in \mathscr{E}_i$. By the implicit function theorem, there are $\delta, \gamma_1, \gamma_2 > 0$ such that, for any vectors $\Upsilon \in B_{\mathscr{H}}(\widehat{\Upsilon}, \delta)$ and $\eta^2 \in B_{\mathscr{E}_2}(\hat{\eta}^2, \gamma_2)$, the mapping $\eta^1 \mapsto F(\Upsilon, \eta^1 + \eta^2)$ is a C^{k+1}-smooth diffeomorphism of the open ball $\dot{B}_{\mathscr{E}_1}(\hat{\eta}^1, 2\gamma_1)$ onto its image $W(\Upsilon, \eta^2, 2\gamma_1)$, and the determinant $D(\Upsilon, \eta^1 + \eta^2)$ is separated from zero by a number $c > 0$. Let us denote by $G(\Upsilon, \eta^2; \cdot) : W(\Upsilon, \eta^2, 2\gamma_1) \to \dot{B}_{\mathscr{E}^1}(\eta^1, 2\gamma_1)$ the inverse function. Decreasing, if necessary, the numbers δ and γ_2, we can assume that the closure of $W(\Upsilon, \eta^2, \gamma_1)$ is included in $B := W(\widehat{\Upsilon}, \hat{\eta}^2, 2\gamma_1)$, so that the map $G(\Upsilon, \eta^2; y)$ is well defined on the product set $B_{\mathscr{H}}(\widehat{\Upsilon}, \delta) \times B_{\mathscr{E}_2}(\hat{\eta}^2, \gamma_2) \times B$ whose last component is independent of Υ and η^2.

Now let $\gamma = \min(\gamma_1, \gamma_2)$ and let $\psi : \mathscr{E} \to \mathbb{R}$ be a smooth function with support in $B_{\mathscr{E}}(\hat{\eta}, \gamma)$. Then the support of $F_*(\Upsilon, \psi\ell)$ is contained in B. Set $B^1 = B_{\mathscr{E}}(\hat{\eta}^1, \gamma)$ and $B^2 = B_{\mathscr{E}}(\hat{\eta}^2, \gamma)$, and denote by ℓ^i the projection of ℓ to $\mathscr{E}^i$. Hypothesis (P) implies that ℓ^1 has a smooth density $\rho^1(\eta^1)$ with respect to the Lebesgue measure. Hence, for any function $f \in C(\mathscr{Y})$ supported in B, Fubini's theorem gives

$$\begin{aligned}
\langle f, F_*(\Upsilon, \psi\ell)\rangle &= \int_{\mathscr{E}} f\big(F(\Upsilon, \eta)\big)\psi(\eta)\ell(\mathrm{d}\eta) \\
&= \int_{B^2} \bigg\{ \int_{B^1} f\big(F(\Upsilon, \eta^1 + \eta^2)\big)\psi(\eta^1 + \eta^2)\rho^1(\eta^1)\mathrm{d}\eta^1 \bigg\} \ell^2(\mathrm{d}\eta^2) \\
&= \int_{B^2} \bigg\{ \int_{B} f(y) \frac{\psi(\eta^1 + \eta^2)\rho^1(\eta^1)}{D(\Upsilon, \eta^1 + \eta^2)}\, \mathrm{d}y \bigg\} \ell^2(\mathrm{d}\eta^2),
\end{aligned}$$

where we performed the change of variable $y = F(\Upsilon, \eta^1 + \eta^2)$ to derive the last line, in which $\eta^1 = G(\Upsilon, \eta^2; y)$. This relation implies that, for $\Upsilon \in B_{\mathscr{H}}(\widehat{\Upsilon}, \delta)$, the measure $F_*(\Upsilon, \psi\ell)$ (which is supported in B) has a density given by

$$\rho_\psi(\Upsilon, y) = \int_{B^2} \frac{\psi(G(\Upsilon, \eta^2; y) + \eta^2)\rho^1(G(\Upsilon, \eta^2; y))}{D(\Upsilon, G(\Upsilon, \eta^2; y) + \eta^2)}\, \ell^2(\mathrm{d}\eta^2), \tag{5.20}$$

where the denominator satisfies the following inequality in the support of the numerator:

$$D(\Upsilon, G(\Upsilon, \eta^2; y) + \eta^2) \geq c. \tag{5.21}$$

Relations (5.20) and (5.21) show that ρ_ψ is C^k-smooth and satisfies (3.2) for $\Upsilon_1, \Upsilon_2 \in B_{\mathscr{H}}(\widehat{\Upsilon}, \delta)$. This completes the proof of the theorem. □

Proof of Theorem 3.2. Let $\psi : \mathscr{E} \to \mathbb{R}$ be any C^∞-function with support in a small ball $B_{\mathscr{E}}(\hat{\eta}, \gamma)$ such that $0 \leq \psi \leq 1$ and $\psi(\hat{\eta}) > 0$. In this case, $F_*(\Upsilon, \ell)$ is minorised by the measure $F_*(\Upsilon, \psi\ell)$. As was established in the proof of Theorem 3.1, the latter has a density $\rho_\psi(\Upsilon, y)$ given by (5.20). Taking $\Upsilon = \widehat{\Upsilon}$ and noting that $G(\widehat{\Upsilon}, \hat{\eta}^2, \hat{y}) = \hat{\eta}^1$, we can write

$$\rho(\widehat{\Upsilon}, \hat{y}) \geq \rho_\psi(\widehat{\Upsilon}, \hat{y}) = \int_{B^2} \left(\frac{\psi(\hat{\eta})\rho^1(\hat{\eta}^1)}{D(\widehat{\Upsilon}, \hat{\eta})} + g(\eta^2) \right) \ell^2(\mathrm{d}\eta^2), \tag{5.22}$$

where $g(\eta^2)$ is a continuous function vanishing at $\hat{\eta}^2$. Now note that, in view of (3.3), the first term in the brackets under the integral is positive, so that the integrand (which is a non-negative function) is strictly positive in the neighbourhood of $\hat{\eta}^2$. Moreover, it follows from (3.3) that $\hat{\eta}^2$ is in the support of ℓ^2. We thus conclude that the integral in (5.22) is positive. □

5.5 Agrachev–Sarychev theorem

Let us consider the Navier–Stokes system on $\mathbb{T}^2$ controlled by a finite-dimensional external force. After projecting to the space H of square-integrable divergence-free vector fields with zero mean value, we write it in the form

$$\partial_t u + \nu L u + B(u) = \eta(t, x), \tag{5.23}$$

where $L = -\Pi\Delta$, $B(u) = \Pi(\langle u, \nabla\rangle u)$, and $\Pi : L^2(\mathbb{T}^2, \mathbb{R}^2) \to H$ stands for Leray's projection. We consider the problem on some interval $J_T = [0, T]$ and denote by $S_T^u(u_0, \eta)$ the map that takes functions $u_0 \in H$ and $\eta \in L^2(J_T, H)$ to $u(T)$, where $u(t, x)$ is the solution of (5.23) issued from u_0. The control force η is assumed to have the form

$$\eta(t, x) = \sum_{j \in \Lambda} \eta_j(t) e_j(x), \tag{5.24}$$

where $\Lambda = \{(1, 0), (1, 1), (-1, 0), (-1, -1)\}$, $\{e_j\}$ is the trigonometric basis in H (see (1.5)), and η_j are smooth real-valued functions on J_T. We denote by $\mathcal{H}_1$ the four-dimensional vector space spanned by $\{e_j, j \in \Lambda\}$ and endow the space $C^\infty(J_T, \mathcal{H}_1)$ with the usual Fréchet topology. The following result is established in [AS06, Sections 4,6] (see also [Shi07, Theorem 2.5] for a more explicit statement in the 3D case).[13]

[13] The papers [AS06, Shi07] deal with the case when the initial point is fixed and the target varies in a compact subset. However, exactly the same arguments enable one to handle the situation in which both the initial and target states vary in compact subsets.

Theorem 5.7. *Let $s \geq 0$ be an integer and let $\mathcal{C}_0, \mathcal{C} \subset V^s$ be compact subsets. Then, for any $\varepsilon > 0$, there is a number $\delta \in (0, T/2)$ and a continuous function $\Psi_T : \mathcal{C}_0 \times \mathcal{C} \to C^\infty(J_T, \mathcal{H}_1)$ such that the following properties hold.*

Support. *For any $u_0 \in \mathcal{C}_0$ and $\hat{u} \in \mathcal{C}$, the support of $\Psi_T(u_0, \hat{u})$ is contained in the interval $[\delta, T - \delta]$.*

Approximation. *We have the inequality*

$$\sup_{u_0 \in \mathcal{C}_0, \hat{u} \in \mathcal{C}} \left\| S_T^u\big(u_0, \Psi_T(u_0, \hat{u})\big) - \hat{u} \right\|_s \leq \varepsilon. \tag{5.25}$$

References

[AKSS07] A. Agrachev, S. Kuksin, A. Sarychev, and A. Shirikyan, *On finite-dimensional projections of distributions for solutions of randomly forced 2D Navier-Stokes equations*, Ann. Inst. H. Poincaré Probab. Statist. **43** (2007), no. 4, 399–415.

[AS06] A. A. Agrachev and A. V. Sarychev, *Controllability of 2D Euler and Navier–Stokes equations by degenerate forcing*, Comm. Math. Phys. **265** (2006), no. 3, 673–697.

[BBP18] J. Bedrossian, A. Blumenthal, and S. Punshon-Smith, *Lagrangian chaos and scalar advection in stochastic fluid mechanics*, Preprint (2018).

[BC15] A. C. Barato and R. Chetrite, *A formal view on level 2.5 large deviations and fluctuation relations*, J. Stat. Phys. **160** (2015), no. 5, 1154–1172.

[BL08] T. Bodineau and R. Lefevere, *Large deviations of lattice Hamiltonian dynamics coupled to stochastic thermostats*, J. Stat. Phys. **133** (2008), 1–27.

[Bog10] V. I. Bogachev, *Differentiable Measures and the Malliavin Calculus*, Mathematical Surveys and Monographs, vol. 164, American Mathematical Society, Providence, RI, 2010.

[BV92] A. V. Babin and M. I. Vishik, *Attractors of Evolution Equations*, North-Holland Publishing, Amsterdam, 1992.

[CG08] R. Chetrite and K. Gawędzki, *Fluctuation relations for diffusion processes*, Comm. Math. Phys. **282** (2008), no. 2, 469–518.

[CJN+] N. Cuneo, V. Jakšić, V. Nersesyan, C.-A. Pillet, and A. Shirikyan, *Mathematical theory of the fluctuation theorem*, in preparation.

[CJPS] N. Cuneo, V. Jakšić, C.-A. Pillet, and A. Shirikyan, *What is a fluctuation theorem?*, in preparation.

[CJPS17] ———, *Fluctuation theorem and thermodynamic formalism*, Preprint (2017).

[DS89] J.-D. Deuschel and D. W. Stroock, *Large Deviations*, Academic Press, Boston, 1989.

[DV75] M. D. Donsker and S. R. S. Varadhan, *Asymptotic evaluation of certain Markov process expectations for large time, I*, Comm. Pure Appl. Math. **28** (1975), 1–47.

[DV83] M. D. Donsker and S. R. S. Varadhan, *Asymptotic evaluation of certain Markov process expectations for large time. IV*, Comm. Pure Appl. Math. **36** (1983), no. 2, 183–212.

[DZ96] G. Da Prato and J. Zabczyk, *Ergodicity for Infinite Dimensional Systems*, Cambridge University Press, Cambridge, 1996.

[DZ00] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Springer–Verlag, Berlin, 2000.

[ECM93] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, *Probability of second law violations in shearing steady states*, Phys. Rev. Lett. **71** (1993), 2401–2404, 3616.

[ES94] D. J. Evans and D. J. Searles, *Equilibrium microstates which generate second law violating steady states*, Phys. Rev. E **50** (1994), 1645–1648.

[ES02] ———, *The fluctuation theorem*, Adv. Phys. **51** (2002), 1529–1585.

[FP94] A. Fannjiang and G. Papanicolaou, *Convection enhanced diffusion for periodic flows*, SIAM J. Appl. Math. **54** (1994), no. 2, 333–408.

[Gal95] G. Gallavotti, *Reversible Anosov diffeomorphisms and large deviations*, Math. Phys. Electron. J. **1** (1995), Paper 1, 1–12.

[Gas05] P. Gaspard, *Dynamical systems theory of irreversibility*, Chaotic dynamics and transport in classical and quantum systems, NATO Sci. Ser. II Math. Phys. Chem., vol. 182, Kluwer Acad. Publ., Dordrecht, 2005, pp. 107–157.

[GC95a] G. Gallavotti and E. G. D. Cohen, *Dynamical ensembles in nonequilibrium statistical mechanics*, Phys. Rev. Lett. **74** (1995), 2694–2697.

[GC95b] ———, *Dynamical ensembles in stationary states*, J. Statist. Phys. **80** (1995), no. 5-6, 931–970.

[Gou07] M. Gourcy, *A large deviation principle for 2D stochastic Navier–Stokes equation*, Stochastic Process. Appl. **117** (2007), no. 7, 904–927.

[JNPS15a] V. Jakšić, V. Nersesyan, C.-A. Pillet, and A. Shirikyan, *Large deviations and Gallavotti–Cohen principle for dissipative PDE's with rough noise*, Comm. Math. Phys. **336** (2015), no. 1, 131–170.

[JNPS15b] ———, *Large deviations from a stationary measure for a class of dissipative PDE's with random kicks*, Comm. Pure Appl. Math. **68** (2015), no. 12, 2108–2143.

[JOPS12] V. Jakšić, Y. Ogata, C.-A. Pillet, and R. Seiringer, *Quantum hypothesis testing and non-equilibrium statistical mechanics*, Rev. Math. Phys. **24** (2012), no. 6, 1230002, 67.

[JPR11] V. Jakšić, C.-A. Pillet, and L. Rey-Bellet, *Entropic fluctuations in statistical mechanics: I. Classical dynamical systems*, Nonlinearity **24** (2011), no. 3, 699–763.

[Kif90] Y. Kifer, *Large deviations in dynamical systems and stochastic processes*, Trans. Amer. Math. Soc. **321** (1990), no. 2, 505–524.

[KNS18] S. Kuksin, V. Nersesyan, and A. Shirikyan, *Exponential mixing for a class of dissipative PDEs with bounded degenerate noise*, Preprint (2018).

[KPS13] T. Komorowski, S. Peszat, and T. Szarek, *Passive tracer in a flow corresponding to two-dimensional stochastic Navier–Stokes equations*, Nonlinearity **26** (2013), no. 7, 1999–2026.

[Kra70] R. H. Kraichnan, *Diffusion by a random velocity field*, Physics of Fluids **13** (1970), no. 1, 22–31.

[KS12] S. Kuksin and A. Shirikyan, *Mathematics of Two-Dimensional Turbulence*, Cambridge University Press, Cambridge, 2012.

[Kur98] J. Kurchan, *Fluctuation theorem for stochastic dynamics*, J. Phys. A **31** (1998), no. 16, 3719–3729.

[Lan85] S. Lang, *Differential Manifolds*, Springer-Verlag, New York, 1985.

[LS99] J. L. Lebowitz and H. Spohn, *A Gallavotti-Cohen-type symmetry in the large deviation functional for stochastic dynamics*, J. Statist. Phys. **95** (1999), no. 1-2, 333–365.

[Mae99] C. Maes, *The fluctuation theorem as a Gibbs property*, J. Stat. Phys. **95** (1999), no. 1-2, 367–392.

[Mol96] S. Molchanov, *Topics in statistical oceanography*, Stochastic Modelling in Physical Oceanography, Birkhäuser Boston, Boston, MA, 1996, pp. 343–380.

[Ner11] H. Nersisyan, *Controllability of the 3D compressible Euler system*, Comm. Partial Differential Equations **36** (2011), no. 9, 1544–1564.

[Ner15] V. Nersesyan, *Approximate controllability of Lagrangian trajectories of the 3D Navier–Stokes system by a finite-dimensional force*, Nonlinearity **28** (2015), no. 3, 825–848.

[Ner19] ———, *Large deviations for the Navier–Stokes equations driven by a white-in-time noise*, Annales Henri Lebesgue (2019), accepted for publication.

[RM07] L. Rondoni and C. Mejía-Monasterio, *Fluctuations in nonequilibrium statistical mechanics: models, mathematical theory, physical mechanisms*, Nonlinearity **20** (2007), no. 10, R1–R37.

[Rue99] D. Ruelle, *Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics*, J. Stat. Phys. **95** (1999), no. 1-2, 393–468.

[Shi07] A. Shirikyan, *Exact controllability in projections for three-dimensional Navier–Stokes equations*, Ann. Inst. H. Poincaré Anal. Non Linéaire **24** (2007), no. 4, 521–537.

[Shi15] ———, *Control and mixing for 2D Navier–Stokes equations with space-time localised noise*, Ann. Sci. Éc. Norm. Supér. (4) **48** (2015), no. 2, 253–280.

[Shi19] ———, *Controllability implies mixing II. Convergence in the dual-Lipschitz metric*, J. Eur. Math. Soc. (2019), accepted for publication.

[WX18] R. Wang and L. Xu, *Asymptotics for stochastic reaction-diffusion equation driven by subordinate Brownian motion*, Stochastic Process. Appl. **128** (2018), no. 5, 1772–1796.