

# Multiscale expansion of invariant measures for SPDEs

September 30, 2003

Dirk Blömker<sup>1,2</sup> and Martin Hairer<sup>1</sup>

<sup>1</sup> Mathematics Research Centre, University of Warwick

<sup>2</sup> Institut für Mathematik, RWTH Aachen

Email: bloemker@instmath.rwth-aachen.de

Email: hairer@maths.warwick.ac.uk

## Abstract

We derive the first two terms in an  $\varepsilon$ -expansion for the invariant measure of a class of semilinear parabolic SPDEs near a change of stability, when the noise strength and the linear instability are of comparable order  $\varepsilon^2$ . This result gives insight into the stochastic bifurcation and allows to rigorously approximate correlation functions. The error between the approximate and the true invariant measure is bounded in both the Wasserstein and the total variation distance.

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Formal Derivation</b>	<b>4</b>
<b>3</b>	<b>Main Assumptions and the Approximation Result</b>	<b>6</b>
<b>4</b>	<b>Proof of the Approximation Result</b>	<b>10</b>
<b>5</b>	<b>Structure of the Invariant Measures</b>	<b>23</b>
<b>6</b>	<b>Total Variation Bounds</b>	<b>32</b>
<b>A</b>	<b>Decay Properties of the Invariant Measure</b>	<b>42</b>

## 1 Introduction

The concept of stochastic bifurcation is still a topic of ongoing research, and there are several open questions on the effect of small additive noise on a deterministic bifurcation. The only fairly complete picture available so far applies to pitchfork bifurcations in one-dimensional stochastic ordinary equations (SODEs), see [CIS99] for a classification of bifurcations. There exist several concepts of bifurcation within the framework of random dynamical systems, and we are not going to discuss them in detail here (see *e.g.* [Arn98] for an overview). One such concept

is a phenomenological bifurcation, where the density of a unique invariant measure changes its structure (*e.g.* number of maxima).

For one-dimensional SODEs there is the result of [CF98] showing that only a phenomenological bifurcation can occur. Their result mainly rules out changes in the structure of the random attractor, which would be a dynamical bifurcation, in the sense of [Arn98].

Nevertheless for SPDEs there are no analogs of these results. Only for monotone SPDEs results are established ruling out the possibility of a dynamical bifurcation (see [CCLR]). The question of defining the analog of a phenomenological bifurcation in this setting is completely open, one problem being the lack of reference measure in infinite dimensional spaces.

Our result approximates the fine structure of the invariant measure. This can be used to describe phenomenological bifurcation for SPDEs. The main tool is a multi-scale approximation of the dynamics using amplitude equations. This is a fairly natural approach, as the separation of time-scales is naturally present in a neighbourhood of a deterministic bifurcation.

The amplitude equation is a stochastic ordinary differential equation describing the dynamics of the dominating modes of the SPDE living on a much slower time-scale. A rigorous justification of the approximation of solutions of SPDEs by the solutions of the corresponding amplitude equation was obtained in [BMPS01, Blö03a]. Unfortunately, these results were only capable of describing the transient dynamics of the SPDE, *i.e.* the dynamics on time scales of the order of the characteristic time of the linear instability. However, we show in this paper that the amplitude equation is also able to describe the long-time behaviour of SPDEs. Our main results show that we can approximate the invariant measure of the original SPDE at first order by the invariant measure of the amplitude equation, which is supported on a finite dimensional subspace. Furthermore, we show that the second order correction is given by a Gaussian measure supported on the orthogonal complement of that space. Note that in the deterministic case, the approximation of solutions to PDEs by the solutions of a suitable amplitude equation is well understood. In particular, the case of unbounded (or very large) domains, which is beyond the scope of this paper, can be justified rigorously, see *e.g.* [KSM92, Sch96].

Let us now make the considerations above more precise. We consider in this paper SPDEs of the type

$$\partial_t u = Lu + \varepsilon^2 Au + \mathcal{F}(u) + \varepsilon^2 \xi, \quad (1.1)$$

where  $L$  is a dissipative operator with finite dimensional kernel,  $\varepsilon^2 Au$  is a small (linear) deterministic perturbation,  $\mathcal{F}$  is a stable cubic nonlinearity, and  $\xi$  is a Gaussian noise which will be taken to be white in time and can be either white or coloured in space. We will furthermore denote by  $X$  the Hilbert space in which we consider initial conditions and solutions to (1.1). There are many examples of equations of that type in the literature. For instance, the well-known Ginzburg-Landau equation

$$\partial_t u = \Delta u + \nu u - u^3 + \sigma \xi$$

and the Swift-Hohenberg equation

$$\partial_t u = -(\Delta + 1)^2 u + \nu u - u^3 + \sigma \xi ,$$

which was first used as a toy model for the convective instability in a Rayleigh-Bénard problem (see [HS92]), fall into the scope of our work when the parameters  $\nu$  and  $\sigma$  are small and of comparable order of magnitude. Both equations are considered on bounded domains with suitable boundary conditions (*e.g.* periodic, Dirichlet, Neumann, etc.). The boundedness of the domain is essential, since all of our proofs rely on the existence of a spectral gap of order one in the linear part.

Another example arises in the theory of surface growth in the model of Lai and Das-Sarma [LDS91]. Here  $h$  is the height-profile of a growing surface given by

$$\partial_t h = -\Delta^2 h - \nu \Delta h + \nabla \cdot (|\nabla h|^2 \nabla h) + \sigma \xi ,$$

subject to periodic boundary conditions on a square and vanishing spatial mean for  $h$ . In this case  $\nu \Delta$  splits into two parts, one belonging to  $L$  and one to  $\varepsilon^2 A$ .

Our first main result (cf. Theorem 5.2), which is based on the transient approximation result (cf. Theorem 3.8), is the following. Let  $u_\star$  be distributed according to an invariant Markov measure of (1.1). Then

$$u_\star = \varepsilon a_\star + \varepsilon^2 \psi_\star + \mathcal{O}(\varepsilon^{3^-}) ,$$

where  $a_\star$  is distributed according to the invariant measure for the amplitude equation (cf. section 2) on the finite dimensional space  $P_c X$ , with  $P_c$  being the projection onto the kernel  $\mathcal{N}$  of  $L$ . Moreover,  $\psi_\star$  is a stationary Ornstein-Uhlenbeck process on  $P_s X$ , where  $P_s = I - P_c$  is the projection onto the stable (or fast) modes. In contrast to that, the projection  $P_c$  maps onto the slow modes, which live on a much slower time-scale  $T = \varepsilon^2 t$ .

The  $\mathcal{O}(\varepsilon^{3^-})$ -term is interpreted as the Wasserstein distance of the corresponding measures, where  $\mathcal{O}(\varepsilon^{3^-})$  means  $\mathcal{O}(\varepsilon^{3-\kappa})$  for any  $\kappa > 0$ . Note that Theorem 5.2 is proved in a slightly different scaling.

The second result (cf. Theorem 5.3) establishes that the law of  $a_\star + \varepsilon \psi_\star$  is approximately given by the product measure, *i.e.*

$$\mathbf{P}^{a_\star + \varepsilon \psi_\star} = \mathbf{P}^{a_\star} \otimes \mathbf{P}^{\varepsilon \psi_\star} + \mathcal{O}(\varepsilon^{2^-}) .$$

This is a remarkable fact as, in general,  $a_\star$  and  $\psi_\star$  are not independent. This approximate independence of the “slow modes” and the “fast modes” in the invariant distribution results mainly from a separation of timescales in their approach to equilibrium.

Another surprising fact is that, due to the cubic nature of the nonlinearity, no second order correction in  $P_c X$  is necessary. This changes dramatically when one tries to go to higher order corrections. Then, the cubic nonlinearity couples strongly the dynamics in  $P_c X$  and  $P_s X$ . Nevertheless we will give formal arguments in section 2 that indicate what the third order correction should be. This

formal approach also indicates that it is not trivial to extend our approach to third order corrections.

Under stronger assumptions on the non-degeneracy of the noise process, we show in section 6 that similar approximation results hold in the much stronger total variation distance. However, since the total variation distance between two measures does not depend on the distance with which the underlying space is endowed, the error between the invariant measure of the SPDE and its approximation scales in a different way with  $\varepsilon$ . Our result is then

$$\|P_c^* \mathbf{P}^{u_*} - \mathbf{P}^{\varepsilon a_*}\|_{\text{TV}} = \mathcal{O}(\varepsilon^{\frac{3}{2}-}), \quad \|\mathbf{P}^{u_*} - \mathbf{P}^{\varepsilon a_*} \otimes \mathbf{P}^{\varepsilon^2 \psi_*}\|_{\text{TV}} = \mathcal{O}(\varepsilon^{1-}). \quad (1.2)$$

Recall the following interpretation of the total variation distance. Given two probability measures  $\mu$  and  $\nu$ , one can build two random variables  $x$  and  $y$  with respective laws  $\mu$  and  $\nu$ , such that  $\mathbf{P}(\{x = y\}) = 1 - \frac{1}{2} \|\mu - \nu\|_{\text{TV}}$ . In particular, the second estimate in (1.2) implies that the small scale spatial structure of stationary solutions of the original SPDE are indistinguishable from those of the Ornstein-Uhlenbeck process  $\varepsilon^2 \psi_*$ .

The paper is organised as follows. In section 2 we give a formal derivation of the amplitude equation and of the higher order corrections. In section 3, we give a precise statement of the assumptions that will be used throughout the paper, and we give a rigorous formulation of the approximation result for the transient dynamics of (1.1) (cf. Theorem 3.8).

The proof of the approximation result is split into several steps, which are all contained in section 4. Subsection 4.1 provides essential *a priori* bounds on the solutions and the invariant measure of the SPDE (1.1), while subsection 4.2 gives essential error estimates for the residual. Finally the proof of the approximation result is contained in subsection 4.3.

The main results on the structure of invariant measures of (1.1) are discussed in section 5, while the proofs are given in subsections 5.1 and 5.2. Finally, the total variation estimates (1.2) are discussed in section 6.

## 2 Formal Derivation

Let us first discuss a formal calculation giving approximations of our original equation (1.1). We make the ansatz

$$u(t) = \varepsilon a(\varepsilon^2 t) + \varepsilon^2 b(\varepsilon^2 t) + \varepsilon^3 c(\varepsilon^2 t) + \varepsilon^2 \psi(t) + \varepsilon^3 \varphi(t) + \mathcal{O}(\varepsilon^4), \quad (2.1)$$

with  $a, b, c \in \mathcal{N} = P_c X$  and  $\varphi, \psi \in \mathcal{S} = P_s X$ . Due to the linear damping of order one in  $\mathcal{S}$ , we expect the modes in  $\mathcal{S}$  to evolve on time scales of order one. In  $\mathcal{N}$ , on the other hand, we expect the modes to evolve on the much slower time-scale  $T = \varepsilon^2 t$ , since the linear operator acting on  $\mathcal{N}$  is of order  $\varepsilon^2$ , so its characteristic time is of order  $\varepsilon^{-2}$ .

Plugging the ansatz (2.1) back into (1.1) and expanding in orders of  $\varepsilon$ , we obtain first, by collecting all terms of order  $\varepsilon^3$  in  $\mathcal{N}$ ,

$$\partial_T a(T) = A_c a(T) + \mathcal{F}_c(a(T)) + \partial_T \beta(T). \quad (2.2)$$

Here,  $\beta(T) = \varepsilon P_c W(\varepsilon^{-2}T)$  is a Wiener process in  $\mathcal{N}$  with distribution independent of  $\varepsilon$ , and we write  $A_c = P_c A$  and  $\mathcal{F}_c = P_c \mathcal{F}$  for short. This approximation is called *amplitude equation*, and it is well-known for many examples in the physics literature. It was rigorously verified in [Blö03a] for a large class of equations.

Let us now turn to higher order corrections. Collecting terms of order  $\varepsilon^2$  in  $\mathcal{S}$  yields

$$\partial_t \psi(t) = L_s \psi(t) + P_s \xi(t) . \quad (2.3)$$

A second order correction was already used in [Blö03b] to deal with quadratic nonlinearities. However, the result there was only an approximation in lowest order, as opposed to the higher order approximation derived here.

The next step is to compute the other second order term involving  $b$ . First we can solve (2.3) for  $\psi$  using the variation of constant formula. Then we rescale  $\psi$  to the slow time-scale to obtain formally that

$$\psi(T\varepsilon^{-2}) \stackrel{\text{law}}{=} \varepsilon L_s^{-1} P_s \xi(T) + \text{“higher order terms”} .$$

As a consequence, the term  $A_c \psi(T\varepsilon^{-2})$ , when viewed on the slow time-scale, gives a contribution of order  $\varepsilon^5$  instead of  $\varepsilon^4$ . Using this, we obtain for terms of order  $\varepsilon^4$  in  $\mathcal{N}$

$$\partial_T b(T) = A_c b(T) + 3\mathcal{F}_c(a(T), a(T), b(T)) . \quad (2.4)$$

Since typical initial conditions for (1.1) can be written as  $u(0) = \varepsilon a(0) + \varepsilon^2 \psi(0)$ , one can choose  $b(0) = 0$ , and therefore  $b$  vanishes identically.

Turning again to  $\mathcal{S}$  and gathering terms of order  $\varepsilon^3$  we obtain

$$\partial_t \varphi(t) = L_s \varphi(t) + A_s a(\varepsilon^2 t) + \mathcal{F}_s(a(\varepsilon^2 t)) .$$

Hence,  $\varphi$  lives basically on the slow time-scale. Defining  $\hat{\varphi}(T) = \varphi(\varepsilon^{-2}T)$  we obtain

$$\hat{\varphi}(T) = L_s^{-1}(A_s a(T) + \mathcal{F}_s(a(T))) + \text{“higher order terms”} . \quad (2.5)$$

Finally, terms of order  $\varepsilon^5$  in  $\mathcal{N}$  (and using the fact that  $b \equiv 0$ ) give rise to

$$\begin{aligned} \partial_T c(T) &= A_c c + A_c \hat{\varphi} + A_c L_s^{-1} P_s \xi(T) + 3\mathcal{F}_c(a, a, c) + 3\mathcal{F}_c(a, a, \hat{\varphi}) \\ &\quad + 3\mathcal{F}_c(a, a, L_s^{-1} P_s \xi(T)) + 3\mathcal{F}_c(a, \psi(\varepsilon^{-2}T), \psi(\varepsilon^{-2}T)) . \end{aligned} \quad (2.6)$$

At this point, the ansatz (2.1) starts to show its limitations. Since  $\hat{\varphi}$  lives (up to higher order terms) on a finite-dimensional space and  $c$  lives on  $\mathcal{N}$  by construction, (2.5) and (2.6) seem to indicate that the solution of (1.1) is approximated at order  $\varepsilon^4$  by  $\varepsilon^2 \psi(t)$  on the orthogonal complement of a finite-dimensional subspace of  $X$ . On the other hand, (2.6) seems to be of rather limited use for studying the structure of the invariant measure of (1.1) since, because of the presence of the instability  $A_c$  in the linear equation for  $c$ , the very existence of an invariant measure for  $c$  is not clear.

Notice furthermore that expression (2.6) still depends on  $\varepsilon$ , as it is not clear *a priori* what would be the limit of  $\mathcal{F}_c(a, \psi(\varepsilon^{-2}T), \psi(\varepsilon^{-2}T))$  as  $\varepsilon \rightarrow 0$ . If we use the previous formal argument, we would obtain a term of order  $\mathcal{O}(\varepsilon^2)$ , which however does not make sense, as it involves squares of Delta-distributions. Instead another formal argument indicates that one should have

$$\mathcal{F}_c(a, \psi(\varepsilon^{-2}T), \psi(\varepsilon^{-2}T)) = \Psi[a] + \text{“higher order terms”},$$

for some linear operator  $\Psi : \mathcal{N} \rightarrow \mathcal{N}$ .

For these reasons, we limit ourselves to the study of the expansion (2.1) up to terms of order  $\varepsilon^2$ . Our main approximation result can then be summarised as follows. Given an initial condition of the form  $u(0) = \varepsilon a(0) + \varepsilon^2 \psi(0)$  with  $a(0)$  and  $\psi(0)$  of order one, the representation

$$u(t) = \varepsilon a(\varepsilon^2 t) + \varepsilon^2 \psi(t) + \mathcal{O}(\varepsilon^3),$$

where  $a \in \mathcal{N}$  is a solution of (2.2) and  $\psi \in \mathcal{S}$  is an Ornstein-Uhlenbeck process given by (2.3), holds up to times  $t = \mathcal{O}(\varepsilon^{-2})$ . The rigorous statement and verification of these results can be found in Theorem 3.8.

### 3 Main Assumptions and the Approximation Result

In this section we formulate the main assumptions needed for our proofs, and state the approximation result, which verifies the formal calculation of the previous section, at least on large transient time-scales.

Let  $X$  be some fixed separable Hilbert space  $X$ . Throughout this paper, we will always write  $\|\cdot\|$  for the norm and  $\langle \cdot, \cdot \rangle$  for the scalar product in  $X$ , unless otherwise stated. We denote by  $\mathcal{L}(X, Y)$  the space of continuous linear operators from  $X$  into  $Y$  endowed with the usual operator norm.

**Assumption 3.1** *The operator  $L$  on  $X$  has compact resolvent and generates an analytic semigroup  $\{e^{tL}\}_{t \geq 0}$  in  $X$ . Denote  $\mathcal{N} = \ker(L)$  and assume that  $\dim(\mathcal{N}) = n$ . We define  $P_c$  as the orthogonal projection onto  $\mathcal{N}$ , we set  $P_s = I - P_c$ , and we assume that  $e^{tL}$  and the projection  $P_c$  commute. We also denote the range of  $P_s$  by  $\mathcal{S}$ , so that  $X = \mathcal{N} \oplus \mathcal{S}$ . We assume*

$$\|e^{tL} P_s\|_{\mathcal{L}(X, X)} \leq e^{-t\omega} \quad \text{for all } t \geq 0. \quad (3.1)$$

We define the fractional space  $X^\alpha$  for  $\alpha \geq 0$  as usual by  $X^\alpha = D((1 - L)^\alpha)$  with norm  $\|\cdot\|_\alpha = \|(1 - L)^\alpha \cdot\|$  and scalar product  $\langle u, v \rangle_\alpha = \langle (1 - L)^\alpha u, (1 - L)^\alpha v \rangle$ . The space  $X^{-\alpha} = (X^\alpha)'$  is the dual of  $X^\alpha$  with canonical norm. See for example [Lun95] or [Paz83]. It is well-known that  $e^{tL}$  extends to an analytic semigroup on all  $X^\alpha$ ,  $\alpha \in \mathbf{R}$ . Note furthermore that obviously  $\mathcal{N} \subset X^\alpha$  for any  $\alpha \geq 0$ , as  $(1 - L)^\alpha \mathcal{N} = \mathcal{N}$ .

The conditions on the semigroup readily imply (see Lumer-Phillips theorem in [Paz83]) that  $e^{tL}$  is a contraction semigroup on  $X$  and hence  $L$  a dissipative operator. To be more precise we have  $\operatorname{Re} \langle Lv, v \rangle \leq 0$  for all  $v \in D(L)$  and moreover

$\operatorname{Re}\langle P_s L v, v \rangle \leq -\omega \|v\|^2$ , as  $L + \omega$  generates a contraction semigroup on  $\mathcal{S}$ . Since  $LP_c = P_c L = 0$ , this also implies that 0 is the only point in the spectrum of  $L$  right of the line  $\operatorname{Re}(\lambda) = -\omega$ . It is easy to verify that this result remains true in  $X^\alpha$ .

It is a well-known fact that for  $\alpha \in [0, 1)$  there is a constant  $M \geq 1$  such that

$$\|e^{tL}\|_{\mathcal{L}(X^{-\alpha}, X)} \leq M(1 + t^{-\alpha}) \quad (3.2)$$

for all  $t > 0$ .

We can combine this with (3.1) to obtain for some  $0 < \tilde{\omega} < \omega$  and  $\tilde{M} \geq 0$  that

$$\|P_s e^{tL}\|_{\mathcal{L}(X^{-\alpha}, X)} \leq \tilde{M}(1 + t^{-\alpha})e^{-t\tilde{\omega}} \quad \text{for all } t > 0. \quad (3.3)$$

All we have to do is using  $P_s e^{tL} = e^{ptL} P_s e^{qtL}$  for  $p, q \in (0, 1)$  with  $p + q = 1$ . Nevertheless, we will assume without loss of generality that  $\tilde{\omega} = \omega$  and  $\tilde{M} = M$ .

Moreover, we immediately get that under Assumption 3.1 we can extend  $P_c$  to an orthogonal projection in  $X^{-\alpha}$  onto  $\mathcal{N}$ , which commutes with the semigroup  $e^{tL}$  on  $X^{-\alpha}$ . This can be easily seen for example by using the representation of the resolvent with the semigroup in addition with Dunford-integrals.

**Assumption 3.2** *The linear operator  $A$  belongs to  $\mathcal{L}(D(L), X)$ . Moreover, there exists a constant  $C_A > 0$  such that*

$$\langle Av, v \rangle \leq C_A(\|v\|^2 + \operatorname{Re}\langle -Lv, v \rangle) \quad \text{for all } v \in D(L).$$

Finally, we assume that one can extend  $A$  to a bounded linear operator from  $X \rightarrow X^{-\alpha}$  for some  $\alpha \in [0, 1)$ .

The assumption that  $A \in \mathcal{L}(D(L), X)$  implies that  $A$  is relatively bounded with respect to  $L$ , i.e. there is a constant  $C > 0$  such that  $\|Av\| \leq C(\|v\| + \|Lv\|)$ .

**Assumption 3.3** *The function  $\mathcal{F} : X^3 \rightarrow X^{-\alpha}$  for some  $\alpha \in [0, 1)$  is continuous, trilinear, and symmetric. Furthermore, its restriction to  $D(L)$  is continuous from  $D(L)^3$  into  $X$ .*

Using the shorthand notations  $\mathcal{F}(u) = \mathcal{F}(u, u, u)$  and  $\mathcal{F}_c = P_c \mathcal{F}$ , we assume that

$$\langle \mathcal{F}_c(v_c, v_c, w_c), w_c \rangle < 0 \quad (3.4)$$

for all  $v_c, w_c \in \mathcal{N} \setminus \{0\}$ . We finally assume that there exist constants  $K$  and  $\gamma_L \in [0, 1)$  such that, for all  $\delta > 0$  sufficiently small,

$$\langle \mathcal{F}(v + \varphi), v \rangle \leq K\delta^2 + K\|\varphi\|^4 - \delta\|v\|^2 - \gamma_L \operatorname{Re}\langle Lv, v \rangle, \quad (3.5)$$

for any  $\varphi, v \in D(L)$ .

The assumption on the restriction  $\mathcal{F} : D(L)^3 \rightarrow X$  being continuous is mainly for convenience. It simplifies the proof of the *a priori* estimate (Theorem 4.1) significantly.

Moreover the assumption that  $\mathcal{F}$  is cubic could be relaxed by adding terms of higher order. Adding terms of degree larger or equal to five does not change any of the main results (under suitable stability assumptions). However, fourth order terms induce a nonlinearity in equation (2.4) and therefore complicate the analysis considerably. In particular, they induce second-order corrections in  $\mathcal{N}$ .

**Assumption 3.4** *The noise process is formally given by  $\xi = Q\partial_t W$ , where  $W$  is a standard cylindrical Wiener process in  $X$  with the identity as a covariance operator and  $Q \in \mathcal{L}(X, X)$  is symmetric. Furthermore, there exists a constant  $\tilde{\alpha} < \frac{1}{2}$  such that*

$$\|(1 - L)^{-\tilde{\alpha}} Q\|_{\text{HS}(X)} < \infty,$$

where  $\|\cdot\|_{\text{HS}(X)}$  denotes the Hilbert-Schmidt norm of an operator from  $X$  to  $X$ .

**Remark 3.5** Straightforward computations, combined with the properties of analytic semigroups allow to check that Assumption 3.4 implies the following:

- The stochastic convolution  $W_L(t) = \int_0^t e^{L(t-s)} Q dW(s)$  is an  $X$ -valued process with Hölder continuous sample paths.
- There exist positive constants  $C$  and  $\gamma$  such that

$$\|P_s e^{Lt} Q\|_{\text{HS}} \leq C(1 + t^{-\tilde{\alpha}}) e^{-\omega t}, \quad (3.6)$$

holds for every  $t > 0$ .

**Remark 3.6** Note that we do *not* assume that  $Q$  and  $L$  commute. Hence, it is in general not true that  $Q$  and  $P_c$  commute. Therefore, the noise processes  $P_c QW$  and  $P_s QW$  will not necessarily be independent.

Using Assumptions 3.2, 3.3, and 3.4 it is straightforward to verify for any (random) initial condition  $u(0) \in X$  the existence of a unique global mild solution  $u \in C^0([0, \infty), X)$  of (1.1), *i.e.* a solution of

$$u(t) = e^{tL} u(0) + \int_0^t e^{(t-\tau)L} (\varepsilon^2 A u(\tau) + \mathcal{F}(u(\tau))) d\tau + W_L(t), \quad (3.7)$$

where the stochastic convolution  $W_L$  was defined in Remark 3.5.

It is well-known under our assumptions that (1.1) or (3.7), respectively, exhibits a not necessarily unique invariant measure. This is quite standard to check, using for instance the methods of [DPZ96]. The main ingredients are an a priori bound in  $X$  (this will be proved in Theorem 4.1), and a bound in some  $X^\gamma$  with small  $\gamma < 1 - \alpha$ , using the methods of Lemma 6.8. The latter bound yields the compactness required by the usual Krylov–Bogoloubov argument.

Depending on the degeneracy of the noise, it is usually possible to check that the invariant measure is also unique, using for example the results from [DPZ96, KS00, EH01, Mat02, Hai02]. The next assumption is only required for the approximation results concerning the invariant measure, not for the approximation of the transient dynamics.



**Assumption 3.7** *The Wiener process  $P_c Q W$  on  $\mathcal{N}$  is not degenerate, i.e.  $P_c Q^2 P_c$  has full rank.*

Note that Assumption 3.7 together with Assumption 3.3 is sufficient for the amplitude equation to have a unique invariant measure that is exponentially attracting, see e.g. [MT94].

We consider solutions  $u$  of equation (1.1), with initial condition  $u(0) = u_0$  satisfying for some family of positive constants  $\{C_p, p \geq 1\}$

$$\mathbf{E}\|u_0\|^p \leq C_p \varepsilon^p \quad \text{and} \quad \mathbf{E}\|P_s u_0\|^p \leq C_p \varepsilon^{2p}. \quad (3.8)$$

By Theorem 4.1 we will see that any initial condition  $u_*$  distributed according to an invariant measure of (1.1) is admissible.

We define the stochastic process  $a \in C^0([0, T], \mathcal{N})$  as the solution to

$$\partial_T a(T) = A_c a(T) + \mathcal{F}_c(a(T)) + \partial_T \beta(T), \quad (3.9)$$

with initial condition  $a(0) = \varepsilon^{-1} P_c u_0$ , where we defined  $\beta(T) = \varepsilon P_c Q W(\varepsilon^{-2} T)$  as in section 2. The process  $\beta$  is a Brownian motion in  $\mathcal{N}$  with non-degenerate covariance matrix  $P_c Q^2 \in \mathcal{L}(\mathcal{N}, \mathcal{N})$ . Furthermore we set

$$\psi_c(t) := a(\varepsilon^2 t).$$

We also define  $\psi_s$  by  $\psi_s(0) = \varepsilon^{-2} P_s u_0$  and

$$\psi_s(t) = e^{tL} \psi_s(0) + P_s W_L(t). \quad (3.10)$$

Our approximation to the solution of (1.1) is then given by

$$\psi(t) := \varepsilon \psi_c(t) + \varepsilon^2 \psi_s(t). \quad (3.11)$$

Note that we do not add any second order term in  $\mathcal{N}$ , which was already justified by our formal calculation. Moreover, due to Remark 3.6 *a* and  $\psi_s$  are in general *not* independent. They are independent, for example, if  $P_c$  commutes with  $Q$  and  $\psi_s(0)$  is independent of  $a(0)$ .

The main result showing that transient dynamics of (1.1) is well approximated by  $\psi$  is the following theorem and corollary. The proof will be given in subsection 4.3.

**Theorem 3.8** *Let Assumptions 3.1, 3.2, 3.3, and 3.4 hold and let  $u$  be the mild solution of (1.1) with initial value  $u_0$  satisfying (3.8). Define  $\psi$  by (3.11) with  $a$  given by (3.9).*

*Then for all  $p > 0$ ,  $\kappa > 0$  and  $T_0 > 0$  there is a constant  $C_{\text{app}}$  explicitly depending on  $p$  and growing exponentially with  $T_0$  such that the estimate*

$$\mathbf{P}\left(\sup_{t \in [0, T_0 \varepsilon^{-2}]} \|u(t) - \psi(t)\| \leq \varepsilon^{3-\kappa}\right) \geq 1 - C_{\text{app}} \varepsilon^p$$

*holds for  $\varepsilon > 0$  sufficiently small.*

Combining this approximation result with the a-priori bound (cf. Theorem 4.1) on the moments, we easily obtain:

**Corollary 3.9** *With the same notations and assumptions as in Theorem 3.8, one has*

$$\mathbf{E} \left( \sup_{t \in [0, T_0 \varepsilon^{-2}]} \|u(t) - \psi(t)\|^p \right) \leq C \varepsilon^{3p-\kappa} .$$

This result will be proved in the following section. Note that it contains only information on transient time scales. The approximation result for the invariant measure (and thus for an infinite time) is stated and proved in section 5.

## 4 Proof of the Approximation Result

### 4.1 A-priori bounds on the invariant measure and solutions

In this subsection, we show that the invariant measure for (1.1) and the law of solutions from a certain class are concentrated on functions that are of order  $\varepsilon$  in  $\mathcal{N}$  and of order  $\varepsilon^2$  in  $\mathcal{S}$ .

**Theorem 4.1** *Let Assumptions 3.1, 3.2, 3.3, and 3.4 be satisfied. Let  $u_\star$  be an  $X$ -valued random variable distributed according to an invariant measure for equation (1.1). Then for all  $p \geq 1$  there are constants  $C_p > 0$  explicitly depending on  $p$  such that*

$$\mathbf{E} \|u_\star\|^p \leq C_p \varepsilon^p \quad \text{and} \quad \mathbf{E} \|P_s u_\star\|^p \leq C_p \varepsilon^{2p} \quad (4.1)$$

for any  $\varepsilon \in (0, 1)$ . Moreover, there are constants  $C'_p$  depending on  $C_p$ , but not on  $\varepsilon$ , such that for any mild solution of (3.7) with initial condition  $u(0)$  fulfilling (4.1) we have

$$\sup_{t>0} \mathbf{E} \|u(t)\|^p \leq C'_p \varepsilon^p \quad \text{and} \quad \sup_{t>0} \mathbf{E} \|P_s u(t)\|^p \leq C'_p \varepsilon^{2p} .$$

Chebychev inequality immediately implies:

**Corollary 4.2** *Under the conditions of the previous theorem. For all  $\kappa > 0$  and  $p \geq 1$  there is a constant  $C > 0$  such that for all  $\varepsilon \in (0, 1)$ ,*

$$\mathbf{P}(\|P_c u_\star\| \leq \varepsilon^{1-\kappa} \text{ and } \|P_s u_\star\| \leq \varepsilon^{2-\kappa}) \geq 1 - C \varepsilon^p .$$

*Proof of Theorem 4.1.* Define the norm in  $\mathcal{L}^p(\Omega, X)$  by

$$\|u\|_{\mathcal{L}^p} := (\mathbf{E} \|u\|^p)^{1/p} .$$

Consider now a mild solution  $u(t)$  of (1.1) with  $u(0) = u_\star$ . Obviously,

$$\|u(t)\|_{\mathcal{L}^p} = \|u_\star\|_{\mathcal{L}^p} , \quad \text{for all } t > 0 .$$

Define now  $L_\varepsilon := L + \varepsilon^2 A - \gamma \varepsilon^2$  for  $\gamma > 0$  to be fixed later. Since  $A$  is a relatively bounded with respect to  $L$  by Assumption 3.2, we immediately get from [Kat80,

Thm IX.2.4] for sufficiently small  $\varepsilon > 0$  that  $L + \varepsilon^2 A$  generates an analytic semigroup. Moreover, following the proof [Kat80, Thm IX.2.4] to reveal the precise values of various constants, it is easy to verify that there exists a sufficiently large (but independent of  $\varepsilon$ ) constant  $\gamma > 0$  such that  $L_\varepsilon$  generates a bounded semigroup for  $\varepsilon < 1$ .

Define now

$$v(t) := u(t) - \varepsilon^2 W_{L_\varepsilon}(t) .$$

Due to the stability of the semigroup and the fact that  $W_{L_\varepsilon}$  is Gaussian, the stochastic convolution satisfies

$$\|P_s W_{L_\varepsilon}(t)\|_{\mathcal{L}^p} \leq \tilde{C}_p \|P_s W_{L_\varepsilon}(t)\|_{\mathcal{L}^2} \leq C_p \quad \text{for any } t \geq 0 , \quad (4.2)$$

and

$$\|W_{L_\varepsilon}(t)\|_{\mathcal{L}^p} \leq \tilde{C}_p \|W_{L_\varepsilon}(t)\|_{\mathcal{L}^2} \leq C_p \varepsilon^{-1} \quad \text{for any } t \geq 0 ,$$

where the constants depend on  $p$ , but can be chosen to be independent of  $\varepsilon$ . This yields immediately

$$\|v\|_{\mathcal{L}^p} \geq \|u_\star\|_{\mathcal{L}^p} - C_p \varepsilon . \quad (4.3)$$

Moreover, a formal computation starting from (3.7) shows that we can differentiate  $v$  with respect to  $t$ , yielding

$$\partial_t v = Lv + \varepsilon^2 Av + \mathcal{F}(v + \varepsilon^2 W_{L_\varepsilon}) + \gamma \varepsilon^4 W_{L_\varepsilon} . \quad (4.4)$$

This equation is only valid in the weak formulation since  $v \notin D(L)$  in general. Therefore, the previous and the following argument are only formal, but can be justified by using smooth approximations of  $W$ , given by  $W^{(k)}$  in  $X$ . Hence, we have  $W_{L_\varepsilon}^{(k)}(t)$  in  $D(L)$ . Moreover, by Assumptions 3.3 and 3.2 we immediately get that the corresponding solution  $v^{(k)}$  is in  $D(L)$ . Our arguments are valid for the approximations  $v^{(k)}$  because they are strong solutions of (4.4) in the classical sense. It suffices then to pass to the limit.

Taking  $\langle \cdot, v \rangle$  on both sides of (4.4) and using (3.5) for  $\delta = 2\varepsilon^2$  together with Assumption 3.2, we derive for sufficiently small  $\varepsilon > 0$

$$\begin{aligned} \frac{1}{2} \partial_t \|v\|^2 &\leq K \delta^2 - \delta \|v\|^2 + C \varepsilon^8 \|W_{L_\varepsilon}\|^4 + \varepsilon^2 \|v\|^2 + C \varepsilon^6 \|W_{L_\varepsilon}\|^2 \\ &\quad - (\gamma_L - \varepsilon^2 C_A) \cdot \operatorname{Re} \langle -Lv, v \rangle \\ &\leq -\varepsilon^2 \|v\|^2 + C(\varepsilon^4 \|W_{L_\varepsilon}\|^2 + \varepsilon^2)^2 . \end{aligned}$$

A comparison argument for ODEs immediately gives

$$\|v\|^2 \leq e^{-2\varepsilon^2 t} \|u_\star\|^2 + C \int_0^t e^{-2\varepsilon^2(t-\tau)} (\varepsilon^4 \|W_{L_\varepsilon}(\tau)\|^2 + \varepsilon^2)^2 d\tau .$$

Taking the  $\mathcal{L}^{p/2}$ -norm on both sides yields

$$\|v\|_{\mathcal{L}^p}^2 \leq e^{-2\varepsilon^2 t} \|u_\star\|_{\mathcal{L}^p}^2 + C \int_0^t e^{-2\varepsilon^2 \tau} d\tau \cdot \varepsilon^2 .$$

Hence,

$$\|v\|_{\mathcal{L}^p} \leq e^{-\varepsilon^2 t} \|u_\star\|_{\mathcal{L}^p} + C(1 - e^{-2\varepsilon^2 t})^{1/2} \cdot \varepsilon .$$

Using (4.3), we get

$$\|u_\star\|_{\mathcal{L}^p} (1 - e^{-\varepsilon^2 t}) \leq C(1 - e^{-2\varepsilon^2 t})^{1/2} \varepsilon + C\varepsilon , \quad (4.5)$$

and the first assertion follows by choosing  $t = \varepsilon^{-2}$ .

For the second claim we get from the mild formulation (3.7)

$$\begin{aligned} \|P_s u(t)\| &\leq M e^{-t\omega} \|P_s u_\star\| + \varepsilon^2 \|P_s W_L(t)\| \\ &\quad + C \int_0^t (1 + (t - \tau)^{-\alpha}) e^{-(t-\tau)\omega} (\varepsilon^2 \|u(\tau)\| + \|u(\tau)\|^3) d\tau . \end{aligned}$$

Taking the  $\mathcal{L}^p$ -norm and using (4.2) yields

$$\begin{aligned} \|P_s u_\star\|_{\mathcal{L}^p} &\leq M e^{-t\omega} \|P_s u_\star\|_{\mathcal{L}^p} + C\varepsilon^2 \\ &\quad + C(\varepsilon^2 \|u_\star\|_{\mathcal{L}^p} + \|u_\star\|_{\mathcal{L}^{3p}}^3) \int_0^t (1 + \tau^{-\alpha}) e^{-\tau\omega} d\tau . \end{aligned}$$

Taking the limit  $t \rightarrow \infty$  and using the first part we derive  $\|P_s u_\star\|_{\mathcal{L}^p} \leq C\varepsilon^2$ .

For the second part of the theorem, we can proceed similarly to the first part. First, instead of (4.5) we derive

$$\|u(t)\|_{\mathcal{L}^p} \leq e^{-\varepsilon^2 t} \|u(0)\|_{\mathcal{L}^p} + C(1 - e^{-2\varepsilon^2 t})^{1/2} \varepsilon + C\varepsilon^2 .$$

Hence, by (4.1) there are constants  $\tilde{C}_p > C_p$  such that  $\|u(t)\|_{\mathcal{L}^p} \leq \tilde{C}_p$  for all  $t > 0$ . Concerning the estimates in  $\mathcal{S}$ , we easily derive

$$\|P_s u(t)\|_{\mathcal{L}^p} \leq M C_p \varepsilon^2 + C\varepsilon^2 + C \int_0^\infty (1 + \tau^{-\alpha}) e^{-\tau\omega} d\tau \cdot \varepsilon^3 (\tilde{C}_p + \tilde{C}_{3p}) .$$

Hence, we easily find a family of constants  $C'_p$  depending (among other constants) explicitly on  $C_p$ ,  $\tilde{C}_p$ , and  $\tilde{C}_{3p}$ , such that  $\|P_s u(t)\|_{\mathcal{L}^p} \leq C'_p$  for all  $t > 0$  and for all  $p > 1$ .  $\square$

## 4.2 Bounds on the residual

In this subsection, we consider solutions  $u$  of equation (1.1), with initial condition  $u(0) = u_0$  satisfying condition (3.8). We prove a sequence of lemmas establishing bounds on the approximation  $\psi(t) := \varepsilon \psi_c(t) + \varepsilon^2 \psi_s(t)$  given in (3.11).

In order to prove that  $\psi$  is a good approximation of  $u$ , we have to control the residual of  $\psi$ , which is defined by

$$\text{Res}(\psi(t)) := -\psi(t) + e^{tL} \psi(0) + \int_0^t e^{(t-\tau)L} (\varepsilon^2 A \psi(\tau) + \mathcal{F}(\psi(\tau))) d\tau + \varepsilon^2 W_L(t) . \quad (4.6)$$

First we establish some bounds on our approximation  $\psi_s$  from (3.11).

**Lemma 4.3** *Let Assumptions 3.1, 3.2, 3.3, and 3.4 be satisfied and fix some time  $T_0 > 0$ . Moreover, consider an initial condition  $u_0$  fulfilling (3.8). Then for every  $p \geq 1$  there exists a constant  $C_p$  such that*

$$\mathbf{E}\left(\sup_{T \in [0, T_0]} \|a(T)\|^p\right) = \mathbf{E}\left(\sup_{t \in [0, T_0 \varepsilon^{-2}]} \|\psi_c(t)\|^p\right) \leq C_p .$$

Moreover, for every  $p \geq 1$  and  $\kappa > 0$ , there exists a constant  $C$  depending on  $p$  and  $\kappa$  such that

$$\mathbf{E}\left(\sup_{t \in [0, T_0 \varepsilon^{-2}]} \|\psi_s(t)\|^p\right) \leq C \varepsilon^{-\kappa} .$$

Again Chebychev inequality immediately yields the following corollary.

**Corollary 4.4** *Under the conditions of the previous theorem. For all  $\kappa > 0$ ,  $T_0 > 0$  and  $p \geq 1$  there is a constant  $C > 0$  such that*

$$\mathbf{P}\left(\|\psi_c(t)\| \leq \varepsilon^{-\kappa} \text{ and } \|\psi_s(t)\| \leq \varepsilon^{-\kappa} \text{ for all } t \in [0, T_0 \varepsilon^{-2}]\right) \geq 1 - C \varepsilon^p .$$

*Proof.* Define the norm in  $\mathcal{L}^p(\Omega, L^\infty([0, T_0 \varepsilon^{-2}], X))$  by

$$\|\psi\|_{\mathcal{L}^p(L^\infty)} := \left(\mathbf{E} \sup_{t \in [0, T_0 \varepsilon^{-2}]} \|\psi(t)\|^p\right)^{1/p} , \quad (4.7)$$

and similarly for functions depending on the slow time-scale, where the supremum is taken over  $T \in [0, T_0]$ . From the boundedness of the semigroup generated by  $L$  we get

$$\|\psi_s\|_{\mathcal{L}^p(L^\infty)} \leq M \varepsilon^{-2} \|P_s u_0\|_{\mathcal{L}^p} + \|P_s W_L\|_{\mathcal{L}^p(L^\infty)} ,$$

where, by assumption (3.8), one has  $\|P_s u_0\|_{\mathcal{L}^p} \leq C \varepsilon^2$ . Moreover, the bound  $\|P_s W_L\|_{\mathcal{L}^p(L^\infty)} \leq C \varepsilon^{-\kappa/p}$  is given by standard arguments using for instance the factorisation method of Da Prato and Zabczyk [DPZ92]. This implies the second assertion.

For the first assertion define  $b = a - \beta_{-1}$  depending on the slow time-scale, where  $\beta_{-1}(T) = \int_0^T e^{-(T-s)} d\beta(s)$ . It is easy to check that  $b$  is differentiable with

$$\partial_T b = A_c(b + \beta_{-1}) + \mathcal{F}_c(b + \beta_{-1}) + \beta_{-1} .$$

By standard arguments,  $\|\beta_{-1}\|_{\mathcal{L}^p(L^\infty)} \leq C$  where the constant depends on  $T_0$  and  $p$ . Using (3.5) and  $A_c \in \mathcal{L}(\mathcal{N})$

$$\begin{aligned} \partial_T \|b\|^2 &\leq C \|b\|^2 + 2\langle \mathcal{F}_c(b + \beta_{-1}), b \rangle + C \|\beta_{-1}\|^2 \\ &\leq C \|b\|^2 + C(1 + \|\beta_{-1}\|^2)^2 . \end{aligned} \quad (4.8)$$

Gronwall's inequality yields for  $t \in [0, T_0]$

$$\|b(T)\|^2 \leq e^{CT} \|a_0\|^2 + C \int_0^T e^{C(T-\tau)} (1 + \|\beta(\tau)\|^2)^2 d\tau .$$

Taking the  $\mathcal{L}^{p/2}(L^\infty)$ -norm and using (3.8) implies

$$\|b\|_{\mathcal{L}^p(L^\infty)}^2 \leq e^{CT_0} \|a_0\|_{\mathcal{L}^p}^2 + C \leq C .$$

Hence,

$$\|a\|_{\mathcal{L}^p(L^\infty)} \leq \|b\|_{\mathcal{L}^p(L^\infty)} + \|\beta_{-1}\|_{\mathcal{L}^p(L^\infty)} \leq C ,$$

which concludes the proof of Lemma 4.3.  $\square$

Note that the constants in the previous proof grow exponentially in  $T_0$ . To avoid this we could use the method of the following lemmas to obtain a logarithmic growth rate instead.

If we do not consider the supremum under the expectation, then the bound is uniform in  $T$  or  $t$ , as shown in the following lemma. Moreover, for  $a$ , the strong dissipativity properties of the cubic nonlinearity allows one to prove a slightly stronger result, where the bound is independent of the initial condition.

**Lemma 4.5** *Let Assumptions 3.1, 3.2, 3.3, and 3.4 hold and fix some time  $T_0 > 0$  and some  $p \geq 1$ . Then there is a constant  $C > 0$  such that for any solution  $a$  of (3.9)*

$$\sup_{T \geq T_0} \mathbf{E} \|a(T)\|^p \leq C ,$$

*uniformly in the initial condition  $a(0)$ , provided  $\|a(0)\| < \infty$  almost surely.*

*Moreover, for any initial condition  $u_0$  fulfilling (3.8), we obtain for any  $p \geq 1$*

$$\sup_{t \geq 0} \mathbf{E} \|\psi_s(t)\|^p + \sup_{T > 0} \mathbf{E} \|a(T)\|^p < C'_p ,$$

*with constants depending on the family  $\{C_p\}$ , but independent of  $\varepsilon$  and  $u_0$ .*

*Proof.* Using (3.4) we obtain for any  $b \in X$  with  $\|b\| = 1$  that

$$\langle \mathcal{F}_c(b), b \rangle \leq K\delta^2 - \delta \leq -\delta/2$$

for sufficiently small  $\delta > 0$ . Hence, as  $\mathcal{F}$  is trilinear  $\langle \mathcal{F}_c(b), b \rangle \leq -\frac{\delta}{2} \|b\|^4$ . Using (4.8) and defining as before  $b = a - \beta_{-1}$ , we thus get the differential inequality

$$\partial_T \|b\|^2 \leq -\frac{\delta}{2} \|b\|^4 + C_\delta (1 + \|\beta_{-1}\|^2)^2 , \quad (4.9)$$

for some constant depending explicitly on  $\delta$ .

Define now  $K_\beta^2(T_0) = \sup_{T \in [0, T_0]} C_\delta (1 + \|\beta_{-1}(T)\|^2)^2$ . Hence, for  $T \in [0, T_0]$  we get  $\partial_T \|b\|^2 \leq -\frac{\delta}{2} \|b\|^4 + K_\beta^2(T_0)$ , and we either obtain  $\|b\|^2 \leq 4K_\beta^2/\delta$ , or  $\|b\|^2$  is strictly decreasing with  $\partial_T \|b\|^2 \leq -\frac{\delta}{4} \|b\|^4$ . Thus  $\|b\|^2 \leq 4/(\delta T + (4/\|b(0)\|^2)) \leq 4/(T\delta)$ . Finally, for any  $T \in [0, T_0]$

$$\|a(T)\| \leq \max \left\{ \frac{4}{T\delta}; \frac{4K_\beta^2(T_0)}{\delta} \right\}^{1/2} + \|\beta_{-1}(T)\| .$$

Hence,  $\|a(T_0)\|_{\mathcal{L}^p} \leq C(T_0, p)$ , with constant independent of  $a(0)$ . This immediately implies the claim.

For the second part, the bound on  $a$  follows in a similar way than in the first part. The bound on  $\psi_s$  is straightforward, since  $\|P_s W_L(t)\|_{\mathcal{L}^p} \leq C$  uniformly in  $t \geq 0$ .  $\square$

The following lemma now gives a bound on the residual.

**Lemma 4.6** *Let Assumptions 3.1, 3.2, 3.3, and 3.4 be satisfied. Moreover, consider an initial condition  $u_0$  fulfilling (3.8). Then for all  $T_0 > 0$ ,  $p > 1$ , and  $\kappa > 0$  there is a constant  $C > 0$  depending explicitly on  $p$  such that*

$$\mathbf{E} \left( \sup_{t \in [0, T_0 \varepsilon^{-2}]} \|P_s \text{Res}(\psi(t))\|^p \right) \leq C \varepsilon^{2p(1-\kappa)},$$

holds for all  $\varepsilon < 1$ .

*Proof.* From (4.6) and the definition of  $\psi_s$ ,

$$P_s \text{Res}(\psi(t)) = \int_0^t e^{(t-\tau)L} P_s (\varepsilon^2 A \psi(\tau) + \mathcal{F}(\psi(\tau))) d\tau.$$

Since  $A$  and  $F$  are bounded from  $X$  to  $Y$ , we obtain

$$\|P_s \text{Res}(\psi(t))\| \leq C \int_0^t (1 + (t-\tau)^{-\alpha}) e^{-(t-\tau)\omega} (\varepsilon^2 \|\psi(\tau)\| + \|\psi(\tau)\|^3) d\tau.$$

Therefore

$$\|P_s \text{Res}(\psi(t))\|_{\mathcal{L}^p(L^\infty)} \leq C \int_0^\infty (1 + \tau^{-\alpha}) e^{-\tau\omega} d\tau \left( \varepsilon^2 \|\psi\|_{\mathcal{L}^p(L^\infty)} + \|\psi\|_{\mathcal{L}^{3p}(L^\infty)}^3 \right),$$

and the claim follows from Theorem 4.3.  $\square$

**Lemma 4.7** *Suppose Assumptions 3.1, 3.2, 3.3, and 3.4 are satisfied. Moreover, consider initial condition  $u_0$  as in (3.8). For every  $T_0 > 0$ , every  $p > 0$ , and every  $\kappa > 0$ , there exists a constant  $C_{\text{Res}} > 0$  depending explicitly on  $p$  such that, for every  $\varepsilon \in (0, 1)$ , the residual  $P_c \text{Res}(\psi(t))$  satisfies the bound*

$$\mathbf{E} \left( \sup_{t \in [0, T_0 \varepsilon^{-2}]} \|P_c \text{Res}(\psi(t))\|^p \right) \leq C_{\text{Res}} \varepsilon^{3p(1-\kappa)}. \quad (4.10)$$

Furthermore, the residual is differentiable in  $t$ , i.e.  $P_c \text{Res}(\psi) \in C^1([0, T_0 \varepsilon^{-2}], \mathcal{N})$   $\mathbf{P}$ -almost surely.

*Proof.* Throughout this proof, we will say that a stochastic process  $\{X(t)\}_{t \geq 0}$  is  $\mathcal{O}(\varepsilon^{n-})$  (or of order  $\varepsilon^{n-}$ ) if for all  $p \geq 1$  and  $\kappa > 0$  there is a constant such that  $\|X\|_{\mathcal{L}^p(L^\infty)} \leq C \varepsilon^{n-\kappa}$ , and similarly for functions on the slow time-scale  $T$ . With

this notation we know by Lemma 4.3 that  $\psi_s$  is of order  $\varepsilon^{0-}$ , but  $\psi_c$  is of order 1 or  $\varepsilon^0$ .

Now (4.6) and (3.9) implies

$$P_c \text{Res}(\psi(t)) = \varepsilon^4 \int_0^t (A_c \psi_s + 3\mathcal{F}_c(\psi_c, \psi_c, \psi_s) + 3\varepsilon \mathcal{F}_c(\psi_c, \psi_s, \psi_s) + \varepsilon^2 \mathcal{F}_c(\psi_s)) d\tau. \quad (4.11)$$

Note first that by definition  $\psi_c$  and  $\psi_s$  are at least continuous in time. Moreover, since  $A_c$  and  $\mathcal{F}_c$  are continuous, it is obvious that the integrand is in  $C^0([0, T_0\varepsilon^{-2}], \mathcal{N})$ . Hence,  $P_c \text{Res}(\psi)$  is differentiable.

To bound (4.11) note first that by Assumption 3.3 and Lemma 4.3, the two last terms are of order  $\varepsilon^{3-}$  and  $\varepsilon^{4-}$ , respectively.

Using the definition of  $\psi_s$  from (3.10), we obtain for the first term

$$\begin{aligned} \int_0^t A_c \psi_s d\tau &= \int_0^t A_c e^{\tau L} \psi_s(0) d\tau + \int_0^t A_c \int_0^\tau e^{(\tau-\eta)L} P_s Q dW(\eta) d\tau \\ &= A_c L_s^{-1} (I - e^{tL}) \psi_s(0) + A_c L_s^{-1} P_s Q (W(t) - W_L(t)), \end{aligned}$$

by using a stochastic Fubini theorem. Since any finite-dimensional projection of the cylindrical Wiener process  $W$  (thus in particular  $A_c L_s^{-1} P_s Q W$ ) is of order  $\varepsilon^{-1-}$ , all of the above terms are of order  $\varepsilon^{3-}$  or even smaller. Note that, by Theorem 4.1,  $\psi_s(0)$  is of order 1.

We expand the remaining term in (4.11) as

$$\begin{aligned} \varepsilon^4 \int_0^t \mathcal{F}_c(a(\varepsilon^2 \tau), a(\varepsilon^2 \tau), \psi_s(\tau)) d\tau &= \varepsilon^2 \int_0^T \mathcal{F}_c(a(\tau), a(\tau), e^{\tau \varepsilon^{-2} L} \psi_s(0)) d\tau \\ &+ \varepsilon^2 \int_0^T \mathcal{F}_c(a(\tau), a(\tau), W_L(\tau \varepsilon^{-2})) d\tau =: I_1(T) + I_2(T). \end{aligned}$$

Now for the first term

$$\|I_1(T)\| \leq C \varepsilon^2 \|\psi_s(0)\| \cdot \left( \sup_{\tau \in [0, T]} \|a(\tau)\|^2 \right) \cdot \int_0^T e^{-\tau \varepsilon^{-2} \omega} d\tau = \mathcal{O}(\varepsilon^{4-}).$$

Hence,  $\|I_1\|_{\mathcal{L}^p(L^\infty)} \leq C \varepsilon^{4-\kappa}$ .

In order to bound the second term, let us define the linear operator

$$B_a(\tau) : X \rightarrow \mathcal{N} \quad \text{by} \quad B_a(\tau)u = \mathcal{F}_c(a(\tau), a(\tau), u). \quad (4.12)$$

With this notation

$$I_2(T) \stackrel{\text{law}}{=} \varepsilon \int_0^T B_a(\tau) \int_0^\tau e^{-L\varepsilon^{-2}(\tau-r)} P_s Q dW(r) d\tau.$$

Let us first consider the case where  $Q$  and  $P_c$  commute. In this case, the process  $B_a(\cdot)$  is independent of the process  $P_s Q W(\cdot)$  and one can decompose  $I_2$  in the following way:

$$I_2 \stackrel{\text{law}}{=} \varepsilon \int_0^T \int_r^{T_0} B_a(\tau) e^{-L\varepsilon^{-2}(\tau-r)} P_s Q d\tau dW(r)$$



$$- \varepsilon \int_T^{T_0} B_a(\tau) \int_0^T e^{-L\varepsilon^{-2}(\tau-r)} P_s Q dW(r) d\tau =: I_3(T) - I_4(T).$$

Since  $I_3$  is a martingale, the Burkholder–Davis–Gundy inequality (see *e.g.* section IV.4 of [RY99]) yields

$$\mathbf{E} \left( \sup_{T \in [0, T_0]} \|I_3(T)\|^p \right) \leq C_p \mathbf{E} \left( \int_0^{T_0} \|\tilde{B}(r)\|_{L_2^0}^2 dr \right)^{p/2}, \quad (4.13)$$

where we defined the process  $\tilde{B} \in C^0([0, T_0], \mathcal{L}(X, N))$  by

$$\tilde{B}(r) = \varepsilon \int_r^{T_0} B_a(\tau) e^{-L\varepsilon^{-2}(\tau-r)} P_s Q d\tau.$$

Moreover, we denote by  $L_2^0$  the space of Hilbert-Schmidt operators with norm  $\|\tilde{B}\|_{L_2^0}^2 = \text{tr}(\tilde{B}\tilde{B}^*)$ , where as usual  $\tilde{B}^*$  denotes the adjoint operator of  $\tilde{B}$ .

By Assumption 3.1,  $\|e^{-L\varepsilon^{-2}(\tau-r)} P_s\| \leq C e^{-w\varepsilon^{-2}(\tau-r)}$ , and therefore

$$\sup_{T \in [0, T_0]} \|\tilde{B}(r)\|_{L_2^0} \leq C\varepsilon^3 \cdot \sup_{r \in [0, T_0]} \|B_a(r)\|_{\mathcal{L}(X, N)},$$

where we used that  $\text{tr}(\tilde{B}\tilde{B}^*) \leq C\|\tilde{B}\|_{\mathcal{L}(X, N)}$  due to the fact that  $\tilde{B}\tilde{B}^* \in \mathcal{L}(N, N)$  is just a matrix.

Combining this with Lemma 4.3, (4.12) and (4.13), we eventually get

$$\mathbf{E} \left( \sup_{T \in [0, T_0]} |I_3(T)|^p \right) \leq C_p \varepsilon^{3p}. \quad (4.14)$$

In order to bound  $I_4$ , we write it as

$$I_4(T) = \varepsilon^2 \int_T^{T_0} B_a(\tau) e^{-L\varepsilon^{-2}(\tau-T)} P_s W_L(\varepsilon^{-2}\tau) d\tau.$$

Since  $\|P_s W_L\|_{\mathcal{L}(L^\infty)} \leq C\varepsilon^{-\kappa}$  (see the proof of Lemma 4.3), it follows immediately from Assumption 3.1 that  $I_4$  is  $\mathcal{O}(\varepsilon^{4-\kappa})$ . This finishes the proof of the lemma for this case.

It remains to bound  $I_2$  in the case when  $Q$  and  $P_c$  do not commute. Since  $B_a(\tau)$  is a semimartingale with respect to  $P_c Q W(\tau)$ , we define  $P_q$  as the orthogonal projection on the cokernel of  $P_c Q$ , which is the orthogonal complement of  $\ker P_c Q$  in  $X$ . This projector is of rank at most  $n = \dim N$  and is such that  $B_a(\cdot)$  is independent of the  $\sigma$ -field generated by the increments of  $Q(1 - P_q)W(\cdot)$ . We can therefore apply the previous calculation to this part and it remains to bound

$$\tilde{I}_2(T) = \varepsilon \int_0^T B_a(\tau) \int_0^\tau e^{-L\varepsilon^{-2}(\tau-r)} P_s Q P_q dW(r) d\tau.$$

Since  $B_a$  is a bilinear map applied to  $(a, a)$  and  $P_q$  has finite rank, we can rewrite  $\tilde{I}_2$  as

$$\tilde{I}_2(T) = \varepsilon \mathcal{B} \int_0^T a(\tau) \otimes a(\tau) \otimes \left( \int_0^\tau e^{-M\varepsilon^{-2}(\tau-r)} \otimes dw(r) \right) d\tau, \quad (4.15)$$

where  $M$  is now a matrix,  $w$  is a Wiener process of finite dimension, say  $m$ , and  $\mathcal{B}$  is a finite-dimensional linear map. Moreover, we identify  $\mathcal{N}$  with  $\mathbf{R}^n$ . We use this notation in order to avoid using multiple indices.

Since  $\mathcal{B}$  is constant and of bounded norm, we make a slight abuse of notation and omit it in the sequel. It is possible to choose  $w$  in such a way that  $a$  satisfies the equation

$$da(\tau) = (A_c a(\tau) + \mathcal{F}_c(a(\tau))) d\tau + \tilde{Q} dw(\tau) = \mathcal{G}(a) d\tau + \tilde{Q} dw(\tau), \quad (4.16)$$

for some matrix  $\tilde{Q} \in \mathbf{R}^{n \times m}$ . In particular, the process  $a$  is adapted to the filtration generated by  $w$ . In order to simplify the subsequent expressions further, we write  $\mathcal{S}_t^{(\varepsilon)}$  for  $e^{-M\varepsilon^{-2}t}$ , and we define the process

$$w_M(T) = \int_0^T \mathcal{S}_{T-s}^{(\varepsilon)} \otimes dw(s), \quad (4.17)$$

which is easily seen to be of order  $\mathcal{O}(\varepsilon^{1-})$ . This is for example done by rescaling to the fast time-scale, and using the analog for the bound on  $P_s W_L(t)$ .

In order to bound (4.15), we would like to exchange the order of integration. The problem is that the integrand for the stochastic integral is then no longer adapted. We will therefore interpret all the stochastic integrals appearing until the end of this proof as Skorokhod integrals (see [Nua95] for example). We will mainly make use of the following property of the Skorokhod integral, where  $\mathcal{D}_t$  denotes the Malliavin derivative of a random variable. We formulate only the one-dimensional version, the finite-dimensional generalisation being obvious.

**Lemma 4.8** *Let  $\beta$  be a Brownian motion,  $u$  a Skorokhod integrable process, and let  $a$  be a random variable with square integrable Malliavin derivative. Then*

$$a \int_0^T u(t) d\beta(t) = \int_0^T a u(t) d\beta(t) + \int_0^T \mathcal{D}_t a u(t) dt \quad (4.18)$$

holds.

*Proof.* The proof of this lemma can be found in [Nua95, p. 40].  $\square$

For shortness of presentation, we do not give a detailed definition of Malliavin derivatives. The only Malliavin derivative required in this paper is that of the process  $a$ , which is denoted by  $\mathcal{D}_t a(\tau)$ . Since  $a$  takes values in  $\mathcal{N}$  and  $w$  takes values in  $\mathbf{R}^m$ ,  $\mathcal{D}_t a(\tau)$  will take values in  $\mathbf{R}^m \otimes \mathcal{N}$ , which we also identify with  $\mathcal{L}(\mathbf{R}^m, \mathcal{N})$ . With this identification made,  $\mathcal{D}_t a(\tau)v$  is given, for each  $v \in \mathbf{R}^m$ , as the solution to the random differential equation

$$\frac{d\mathcal{D}_t a(\tau)v}{d\tau} = D\mathcal{G}(a(\tau)) \mathcal{D}_t a(\tau)v, \quad \mathcal{D}_t a(t)v = \tilde{Q}v. \quad (4.19)$$

(If  $t \geq \tau$ , then  $\mathcal{D}_t a(\tau) = 0$ .) Formally, (4.19) is obtained by making the substitution  $dw(\tau) \mapsto dw + h \delta(\tau - t) d\tau$  in (4.16) and then differentiating the resulting

process with respect to  $h$  and evaluating it at  $h = 0$ . For a general and rigorous definition of the Malliavin derivative, we refer to [Mal97, Nua95].

Obviously  $D\mathcal{G}(a)v = A_c v + 3\mathcal{F}_c(a, a, v)$ . Hence, using the bound on  $A_c$  and condition (3.4), we derive using a standard Gronwall argument, that one has for  $t \leq \tau$  the bound

$$\|\mathcal{D}_t a(\tau)\| \leq \|\tilde{Q}\| e^{\|A_c\|(\tau-t)}, \quad (4.20)$$

where all norms are the corresponding matrix norms *e.g.* in  $\mathbf{R}^{n \times m}$ .

Another ingredient for bounding  $\tilde{I}_2$  is the following modification of the Burkholder–Davis–Gundy inequality:

**Lemma 4.9** *Let  $w(t)$  and  $\mathcal{S}_t^{(\varepsilon)}$  be as above, and let  $Z(s)$  be a real-valued continuous adapted process whose  $L^2$ -norm has moments of all orders. Then, the following bound holds for every  $\varepsilon > 0$  and for every  $p \geq 1$ :*

$$\mathbf{E} \sup_{t \in [0, T]} \left\| \int_0^t Z(s) \mathcal{S}_{t-s}^{(\varepsilon)} \otimes dw(s) \right\|^p \leq C_p \mathbf{E} \left( \int_0^T |Z(s)|^2 ds \right)^{p/2}, \quad (4.21)$$

where the constant  $C_p > 0$  is independent of  $\varepsilon$ .

*Proof.* Using integration by parts, we get

$$\begin{aligned} \int_0^t Z(s) \mathcal{S}_{t-s}^{(\varepsilon)} \otimes dw(s) &= \text{Id} \otimes \int_0^t Z(s) dw(s) \\ &\quad + \varepsilon^{-2} M \int_0^t \mathcal{S}_{t-s}^{(\varepsilon)} \otimes \int_0^s Z(r) dw(r) ds. \end{aligned}$$

It now suffices to apply Burkholder–Davis–Gundy to  $\int_0^t Z(s) dw(s)$  and to use the fact that  $\mathcal{S}_t^{(\varepsilon)}$  is a contraction semigroup, *i.e.* a bound like (3.1) holds. Hence

$$\int_0^t \|\mathcal{S}_{t-s}^{(\varepsilon)}\| ds \leq M \int_0^t e^{-\omega \varepsilon^{-2}(t-s)} ds \leq C \varepsilon^2, \quad (4.22)$$

and the estimate (4.21) follows.  $\square$

Let us now finally turn to the bound on  $\tilde{I}_2(T)$ . Using (4.18) and the stochastic Fubini theorem (see *e.g.* [Leó93] for a quite general version that also applies to Hilbert spaces) we can rewrite it as

$$\tilde{I}_2(T) = \varepsilon \int_0^T \int_0^\tau [a(\tau) \otimes a(\tau) - a(s) \otimes a(s)] \otimes \mathcal{S}_{\tau-s}^{(\varepsilon)} \otimes dw(s) d\tau \quad (4.23a)$$

$$+ \varepsilon \int_0^T a(s) \otimes a(s) \otimes \int_s^T \mathcal{S}_{\tau-s}^{(\varepsilon)} d\tau \otimes dw(s) \quad (4.23b)$$

$$+ 2\varepsilon \int_0^T \int_0^\tau a(\tau) \otimes \mathcal{D}_s a(\tau) \otimes \mathcal{S}_{\tau-s}^{(\varepsilon)} ds d\tau. \quad (4.23c)$$

Term (4.23c) is seen to be of order  $\varepsilon^3$  by using (4.20) and Lemma 4.3. Term (4.23b) can be bounded by first computing the inner integral explicitly, which gives an

additional  $\varepsilon^2$  from integrating  $\mathcal{S}_t^{(\varepsilon)}$ , and then applying Lemma 4.9 to the remaining terms. So it remains to bound the first term (4.23a). By Itô's formula, we have for some constant matrix  $C$ , which depends only on the covariance matrix  $\tilde{Q}$ , that

$$\begin{aligned} a(\tau) \otimes a(\tau) - a(s) \otimes a(s) &= 2 \int_s^\tau a(r) \otimes \mathcal{G}(a(r)) dr \\ &\quad + (\tau - s)C + 2 \int_s^\tau a(r) \otimes \tilde{Q} dw(r). \end{aligned}$$

The term induced by  $(\tau - s)C$  can easily be bounded by  $\mathcal{O}(\varepsilon^3)$ , using (4.22). So we focus on the two remaining terms which we denote by  $J_1$  and  $J_2$ . For the first one, we get, by applying again the stochastic Fubini theorem and (4.18):

$$\begin{aligned} J_1(T) &= 2\varepsilon \int_0^T \int_0^\tau \int_s^\tau a(r) \otimes \mathcal{G}(a(r)) dr \otimes \mathcal{S}_{\tau-r}^{(\varepsilon)} \otimes dw(s) d\tau \\ &= 2\varepsilon \int_0^T \int_0^\tau a(r) \otimes \mathcal{G}(a(r)) \otimes \mathcal{S}_{\tau-r}^{(\varepsilon)} w_M(r) dr d\tau \\ &\quad + \varepsilon \int_0^T \int_0^\tau \int_0^r \mathcal{D}_s a(r) \otimes \mathcal{G}(a(r)) \otimes \mathcal{S}_{\tau-s}^{(\varepsilon)} ds dr d\tau \\ &\quad + \varepsilon \int_0^T \int_0^\tau \int_0^r a(r) \otimes D\mathcal{G}(a(r)) \mathcal{D}_s a(r) \otimes \mathcal{S}_{\tau-s}^{(\varepsilon)} ds dr d\tau. \end{aligned}$$

All these terms are easily bounded by  $\mathcal{O}(\varepsilon^3)$ , using Lemma 4.3 for  $a$ , Assumptions 3.2 and 3.3 for  $\mathcal{G}$ , (4.22) for producing  $\varepsilon^2$ , and the discussion after (4.17) showing that  $w_M = \mathcal{O}(\varepsilon^{1-})$ .

The term  $J_2$  is defined as

$$J_2(T) = 2\varepsilon \int_0^T \int_0^\tau \int_s^\tau a(r) \otimes \tilde{Q} dw(r) \otimes \mathcal{S}_{\tau-s}^{(\varepsilon)} \otimes dw(s) d\tau.$$

Changing the order of the integration and computing explicitly the integral over  $d\tau$  yields

$$J_2(T) = 2\varepsilon^3 \int_0^T \int_0^r a(r) \otimes M^{-1} \mathcal{S}_{r-s}^{(\varepsilon)} (\mathcal{S}_{T-r}^{(\varepsilon)} - \text{Id}) \otimes dw(s) \otimes \tilde{Q} dw(r).$$

At this point, we again use (4.18) in order to “pull”  $a$  out of the inner integral. This yields

$$\begin{aligned} J_2(T) &= 2\varepsilon^3 \int_0^T a(r) \otimes M^{-1} (\mathcal{S}_{T-r}^{(\varepsilon)} - \text{Id}) w_M(r) \otimes \tilde{Q} dw(r) \\ &\quad + 2\varepsilon^3 \int_0^T \int_0^r \mathcal{D}_s a(r) \otimes M^{-1} \mathcal{S}_{r-s}^{(\varepsilon)} ds (\mathcal{S}_{T-r}^{(\varepsilon)} - \text{Id}) \otimes \tilde{Q} dw(r). \end{aligned}$$

Both terms are at most of order  $\mathcal{O}(\varepsilon^3)$  by Lemma 4.9 and the fact that  $a$ ,  $\mathcal{D}_t a$ ,  $\mathcal{S}_t^{(\varepsilon)}$ ,  $w_M$  are all of order 1 or better, therefore concluding the proof of Lemma 4.7.  $\square$

### 4.3 Approximation

In this subsection, we use the bounds previously obtained for the residual to give the proofs of Theorem 3.8 and Corollary 3.9. Recall that this theorem states that  $\psi(t)$  does indeed approximate  $u(t)$  up to errors of order  $\mathcal{O}(\varepsilon^{3-})$ , where  $u$  is a solution of the original SPDE (3.7) with initial conditions satisfying the *a priori* bounds (3.8).

*Proof of Theorem 3.8.* For shorthand notation, we define  $R(t)$  by  $u(t) = \varepsilon\psi_c(t) + \varepsilon^2\psi_s(t) + \varepsilon^3R(t)$ . Note that by definition (cf. (3.11)) we have  $R(0) = 0$ .

Define for some arbitrary  $\kappa \in (0, \frac{1}{4})$  and  $T_0 > 0$  the event  $\mathcal{A}$  by

$$\mathcal{A} := \left\{ \sup_{t \in [0, T_0 \varepsilon^{-2}]} \left( \|\psi_s(t)\| + \|\psi_c(t)\| + \varepsilon^{-3} \|\mathbf{Res}_c(\psi(t))\| + \varepsilon^{-2} \|\mathbf{Res}_s(\psi(t))\| \right) \leq \varepsilon^{-\kappa} \right\}. \quad (4.24)$$

We know by Lemmas 4.3, 4.6, and 4.7 that  $\mathbf{P}(\mathcal{A}) \geq 1 - C\varepsilon^p$ . Now all we have to verify is that  $\sup_{t \in [0, T_0 \varepsilon^{-2}]} \|R(t)\| \leq C\varepsilon^{-\kappa}$  on  $\mathcal{A}$ .

Let  $R_c = P_c R$ ,  $R_s = P_s R$  and define the stopping time  $\tau_e$  by

$$\tau_e = \inf\{t > 0 : \|R(t)\| \geq \varepsilon^{-1/2}\}.$$

One obtains from (3.7) and (4.6)

$$\begin{aligned} R(t) &= \varepsilon^{-3} \mathbf{Res}(\psi(t)) + \varepsilon^2 \int_0^t e^{(t-\tau)L} A R(\tau) d\tau \\ &\quad + \varepsilon^{-3} \int_0^t e^{(t-\tau)L} (\mathcal{F}(u(\tau)) - \mathcal{F}(\psi(\tau))) d\tau. \end{aligned} \quad (4.25)$$

Since  $\mathcal{F}$  is trilinear and  $u = \psi + \varepsilon^3 R$ , we have by Assumption 3.3

$$\|\mathcal{F}(u) - \mathcal{F}(\psi)\|_{X^{-\alpha}} \leq C_F \varepsilon^5 (3 \|\varepsilon^{-1} \psi\|^2 \|R\| + 3 \varepsilon^2 \|\varepsilon^{-1} \psi\| \|R\|^2 + \varepsilon^4 \|R\|^3).$$

For  $t < \tau_e$ , we thus obtain on  $\mathcal{A}$

$$\begin{aligned} \|R_s(t)\| &\leq C_{\mathbf{Res}} \varepsilon^{-\kappa} + C \int_0^{t_\varepsilon} (1 + (t_\varepsilon - \tau)^{-\alpha}) e^{-(t_\varepsilon - \tau)\omega} (C_A \varepsilon^{3/2} + C_F C_\varepsilon) d\tau \\ &\leq C \varepsilon^{-\kappa}. \end{aligned}$$

Since, by Lemma 4.7,  $\mathbf{Res}_c(\psi(t))$  is differentiable, we get from (4.25)

$$\partial_t R_c = \varepsilon^2 A_c R_c + \varepsilon^{-3} \mathbf{Res}_c(\psi(t)) + \mathcal{F}_c(\psi_c + \varepsilon\psi_s + \varepsilon^2 R) - \mathcal{F}_c(\psi_c + \varepsilon\psi_s).$$

In order to bound  $R_c$ , we define as in (4.12) the  $\mathcal{L}(\mathcal{N}, \mathcal{N})$ -valued process  $B_a(T)$  by  $B_a(T)v = 3\mathcal{F}_c(a(T), a(T), v)$ , and we set  $\tilde{R}_c(T) = R_c(\varepsilon^2 T)$ , where  $\tilde{R}_c$  lives on the slow time-scale. With these notations we get

$$\partial_T \tilde{R}_c(T) = (A_c + B_a(T)) \tilde{R}_c(T) + \varepsilon^{-3} \partial_T (\mathbf{Res}_c(\psi(\varepsilon^{-2} T))) + K(T), \quad (4.26)$$

where we defined

$$\varepsilon^2 K(T) = \mathcal{F}_c(\psi_c + \varepsilon\psi_s + \varepsilon^2 R) - \mathcal{F}_c(\psi_c + \varepsilon\psi_s) - 3\mathcal{F}_c(\psi_c, \psi_c, \varepsilon^2 R_c).$$

Now an elementary calculation shows that  $K$  contains all terms from  $\mathcal{F}(u)$  that are of higher order. Hence, we obtain on  $\mathcal{A}$  that

$$\|K(T)\| \leq C_{T_0} \varepsilon^{1/2-2\kappa} \quad \text{for } T \in [0, T_0] \cap [0, \tau_e \varepsilon^2]. \quad (4.27)$$

To bound  $R_c$  or  $\tilde{R}_c$  we will discuss (4.26) further. First define  $U(T, S) \in \mathcal{L}(\mathcal{N}, \mathcal{N})$  as the flow generated by  $A_c + B_a(T)$ , *i.e.* the solution of

$$\partial_T U(T, S) = (A_c + B_a(T))U(T, S), \quad U(S, S) = I.$$

To bound  $U$  note first that obviously  $A_c \in \mathcal{L}(\mathcal{N}, \mathcal{N})$  and  $\langle w, B_a(T)w \rangle < 0$  by (3.4). Hence  $\langle w, (A_c + B_a(T))w \rangle \leq \|A_c\|_{\mathcal{L}(\mathcal{N})} \|w\|^2$  for all  $w \in \mathcal{N}$ . Using a standard Gronwall argument, this immediately implies  $\|U(T, S)\|_{\mathcal{L}(\mathcal{N})} \leq \exp(\|A_c\|_{\mathcal{L}(\mathcal{N})}(T - S))$ . Moreover, solving (4.26) with the help of the variation of constants formula yields

$$\tilde{R}_c(T) = \varepsilon^{-3} \int_0^T U(T, S) \partial_S (\text{Res}_c(\psi(\varepsilon^{-2} S))) dS + \int_0^T U(T, S) K(S) dS. \quad (4.28)$$

By (4.27) the second term on the right-hand side in (4.28) is bounded by

$$\left\| \int_0^T U(T, S) K(S) dS \right\| \leq C_{T_0} \varepsilon^{1/2-2\kappa} \leq C_{T_0}. \quad (4.29)$$

In order to bound the first term in (4.28), we integrate by parts:

$$\begin{aligned} \int_0^T U(T, S) \partial_S (\text{Res}_c(\psi(\varepsilon^{-2} S))) dS &= \text{Res}_c(\psi(\varepsilon^{-2} T)) - U(T, 0) \text{Res}_c(\psi(0)) \\ &\quad + \int_0^T U(T, S) (A_c + B_a(S)) \text{Res}_c(\psi(\varepsilon^{-2} S)) dS. \end{aligned}$$

Combining Lemma 4.7 with our bounds for  $A$ ,  $B_a$ , and  $U$ , we bound the latter by  $C_T \varepsilon^{-\kappa}$ .

We have shown that  $\|R(t)\| \leq C\varepsilon^{-\kappa}$  for  $t < \min\{\tau_e, \varepsilon^{-2} T_0\}$ . By the continuity of the process  $\|R(t)\|$  and the definition of  $\tau_e$ , this implies that  $\tau_e \geq \varepsilon^{-2} T_0$  on  $\mathcal{A}$  for  $\varepsilon > 0$  sufficiently small, so the proof of Theorem 3.8 is complete.  $\square$

*Proof of Corollary 3.9.* Recall the notation  $\|\cdot\|_{\mathcal{L}^p(L^\infty)}$  introduced in (4.7). Using (3.8) and Lemma 4.3 we readily obtain the following *a priori* bound.

$$\|R\|_{\mathcal{L}^p(L^\infty)} \leq \varepsilon^{-3} \|u\|_{\mathcal{L}^p(L^\infty)} + \varepsilon^{-3} \|\psi\|_{\mathcal{L}^p(L^\infty)} \leq C\varepsilon^{-2}.$$

Define  $\mathcal{R}_p = \sup_{t \in [0, T_0 \varepsilon^{-2}]} \|R(t)\|^p$ . Using Theorem 3.8 with  $6p + 4$  instead of  $p$ , we easily derive

$$\mathbf{E} \mathcal{R}_p = \int_{\{\mathcal{R}_p \geq C\varepsilon^{3p-\kappa}\}} \mathcal{R}_p d\mathbf{P} + C\varepsilon^{3p-\kappa} \int_{\{\mathcal{R}_p < C\varepsilon^{3p-\kappa}\}} d\mathbf{P}$$

$$\begin{aligned} &\leq \mathbf{P}(\mathcal{R}_p \geq C\varepsilon^{3p-\kappa})^{1/2} \cdot (\mathbf{E}\mathcal{R}_p^2)^{1/2} + C\varepsilon^{3p-\kappa} \\ &\leq C\varepsilon^{3p+2}\varepsilon^{-2} + C\varepsilon^{3p-\kappa}, \end{aligned}$$

which concludes the proof of Corollary 3.9.  $\square$

## 5 Structure of the Invariant Measures

In this section, we consider the approximation to the original equation on the slow time-scale and we rescale its amplitude by a factor  $\varepsilon^{-1}$ . In other words, we consider for the approximation the solutions of

$$dv_c = A_c v_c dT + \mathcal{F}_c(v_c) dT + P_c Q d\tilde{W}(T), \quad (5.1a)$$

$$dv_s = \varepsilon^{-2} L v_s dT + P_s Q d\tilde{W}(T), \quad (5.1b)$$

where  $\tilde{W}$  is the rescaled Wiener process, with the same distribution as  $W$ . We rewrite the original equation (1.1) in a similar way as

$$d\tilde{u} = \varepsilon^{-2} L \tilde{u} dT + A \tilde{u} dT + \mathcal{F}(\tilde{u}) dT + Q d\tilde{W}(T), \quad (5.2)$$

where  $\tilde{u}(T) = \varepsilon^{-1} u(\varepsilon^{-2}T)$  is the rescaled solution  $u$ .

We denote by  $\mathcal{P}_T$  the Markov semigroup generated by the solutions of (5.1) and by  $\mathcal{Q}_T$  the one generated by (5.2). (We make a slight abuse of notation by using the same symbol for the semigroup acting on measures and its dual group acting on functions.) Since (5.1a) and (5.1b) are only coupled through the noise, they can also be considered separately and we denote by  $\mathcal{P}_T^c$  and  $\mathcal{P}_T^s$  the corresponding semigroups.

Since we rescaled the equations in such a way that solutions are of order 1 and no longer of order  $\varepsilon$ , it will follow that the Wasserstein distance (5.6) between the invariant measures for  $\mathcal{Q}_T$  and for  $\mathcal{P}_T$  is of order  $\mathcal{O}(\varepsilon^{2-})$ . A straightforward modification of the arguments presented here also allows to show that this distance is of order  $\mathcal{O}(\varepsilon^{3-})$  in the original scaling. We however prefer to prove the result in this scaling for simplicity of presentation, as now a lot of terms are actually independent of  $\varepsilon$ .

The main results are the following. The first result basically says that  $\mathcal{P}_T$  is (up to small errors) a contraction uniform in  $\varepsilon$ . Precise definitions and properties of the Wasserstein ( $\|\cdot\|_L$ ) and other distances of probability measures will be given after the theorems.

**Theorem 5.1** *Let the assumptions of section 3 hold. There exists a  $T_0 > 0$  such that*

$$\|\mathcal{P}_{T_0}\mu - \mathcal{P}_{T_0}\nu\|_L \leq \frac{1}{2}\|\mu - \nu\|_L + \varepsilon^2 \int_X (1 + \|P_s x\|)(\mu + \nu)(dx),$$

for every pair  $\mu, \nu$  of probability measures on  $X$  and for every  $\varepsilon \in (0, 1)$ .

The next result gives the expansion of the invariant measure. Important is the following bound on moments of some measure  $\mu$ , which is necessary to apply the results of the previous sections.

Let  $C_p$  be a sequence of positive constants indexed by  $p \geq 1$  and let  $\mu$  be a measure on  $X$  satisfying

$$\int_X \|x\|^p \mu(dx) < C_p \quad \text{and} \quad \int_X \|P_s x\|^p \mu(dx) < C_p \varepsilon^p \quad (5.3)$$

for all  $p \geq 1$ .

**Theorem 5.2** *Suppose all assumptions of section 3 are satisfied. Let  $\mu$  and  $\nu$  be two measures on  $X$  satisfying (5.3). Then, there exist constants  $C$ ,  $C_*$ , and  $\gamma > 0$  depending only on the family  $\{C_p\}$  such that*

$$\|\mathcal{P}_T \mu - \mathcal{Q}_T \nu\|_{\mathbb{L}} \leq C e^{-\gamma T} \|\mu - \nu\|_{\mathbb{L}} + C_* \varepsilon^{2-\kappa} \quad (5.4)$$

*holds for every  $T > 0$  and every  $\varepsilon \in (0, 1)$ . In particular, if  $\mu_*$  and  $\nu_*$  denote invariant measures for  $\mathcal{P}_T$  and  $\mathcal{Q}_T$ , respectively, one has  $\|\mu_* - \nu_*\|_{\mathbb{L}} \leq C_* \varepsilon^{2-\kappa}$ .*

It is clear that, in the case where  $P_c$  and  $L$  commute, we have independence of the solutions of (5.1a) and (5.1b). Hence,

$$\mu_* = \mu_*^c \otimes \mu_*^s, \quad (5.5)$$

where  $\mu_*^c$  is the invariant measure for (5.1a) and  $\mu_*^s$  is the invariant measure for (5.1b). Due to the structure of the equation, it is furthermore obvious that one always has  $P_c^* \mu_* = \mu_*^c$  and  $P_s^* \mu_* = \mu_*^s$ . However, there is in general no reason for (5.5) to hold if  $P_c$  and  $L$  do not commute.

In subsection 5.2, we show that, even if  $P_c$  and  $L$  do not commute, the equality (5.5) holds up to an error term of order  $\varepsilon^{2-\kappa}$  for arbitrarily small  $\kappa$ . This is a consequence of the fact that (5.1a) and (5.1b) live on different time-scales. It is nevertheless rather surprising that the error is roughly of order  $\varepsilon^2$ , since (5.1b) needs a time of the order  $\varepsilon^2$  to reach equilibrium, but in this time interval (5.1a) moves by an amount of order  $\varepsilon$ , due to the presence of the diffusion term. The rigorous statement of this result is:

**Theorem 5.3** *Let the assumptions of section 3 hold. Then, for every  $\kappa > 0$ , there exists a constant  $C$  such that*

$$\|\mu_* - \mu_*^c \otimes \mu_*^s\|_{\mathbb{L}} \leq C \varepsilon^{2-\kappa}$$

*for all  $\varepsilon \in (0, 1)$ , where  $\mu_*$  is the invariant measure for the approximating equation (5.1). Moreover,  $P_c^* \mu_* = \mu_*^c$  and  $P_s^* \mu_* = \mu_*^s$  are the marginals of  $\mu_*$  on  $\mathcal{N}$  and  $\mathcal{S}$ .*

Before we proceed, we first recall definitions and properties of the norms on the space of signed measures on  $X$ , which are used in the statement above, and the proofs below.



Given a measurable function  $\varphi : X \rightarrow \mathbf{R}$ , we define its Lipschitz norm  $\|\varphi\|_L$  by

$$\|\varphi\|_L = \sup_{x,y \in X} \left\{ |\varphi(x)|, \frac{|\varphi(x) - \varphi(y)|}{\|x - y\|} \right\}.$$

We will also be led to consider function with bounded Lipschitz norm in the stable directions. We therefore define in a similar way

$$\|\varphi\|_{L,s} = \sup_{x,y \in X \mid P_c x = P_c y} \left\{ |\varphi(x)|, \frac{|\varphi(x) - \varphi(y)|}{\|P_s x - P_s y\|} \right\}.$$

These two norms induce corresponding norms on the space of signed measures on  $X$ . The Wasserstein distance is defined by

$$\|\mu - \nu\|_L = \sup_{\|\varphi\|_L \leq 1} \left| \int_X \varphi(x) \mu(dx) - \int_X \varphi(x) \nu(dx) \right|, \quad (5.6)$$

and we define  $\|\mu - \nu\|_{L,s}$  in a similar way. Recall also that the total variation distance between two measures on  $X$  is given by

$$\|\mu - \nu\|_{TV} = \sup_{\|\varphi\|_\infty \leq 1} \left| \int_X \varphi(x) \mu(dx) - \int_X \varphi(x) \nu(dx) \right|,$$

where  $\|\varphi\|_\infty = \sup_{x \in X} |\varphi(x)|$ . Note that these definitions imply

$$\|\mu - \nu\|_L \leq \|\mu - \nu\|_{L,s} \leq \|\mu - \nu\|_{TV}.$$

We will sometimes use the following equivalent definition of the total variation distance. Denote by  $\mathcal{C}(\mu, \nu)$  the set of all couplings of  $\mu$  and  $\nu$ , *i.e.* the set of all probability measures  $\mathbf{P}$  on the product space  $X \times X$  such that the first marginal of  $\mathbf{P}$  is equal to  $\mu$  and its second marginal is equal to  $\nu$ . Then, one has

$$\frac{1}{2} \|\mu - \nu\|_{TV} = \inf_{\mathbf{P} \in \mathcal{C}(\mu, \nu)} \mathbf{P}\{x \neq y\}.$$

Furthermore, there exists one coupling that realises the infimum.

The Monge–Kantorovitch theorem (see *e.g.* [RR00]) yields a similar representation for the Wasserstein distance:

$$\frac{1}{2} \|\mu - \nu\|_L = \inf_{\mathbf{P} \in \mathcal{C}(\mu, \nu)} \int_X \int_X \inf\{\|x - y\|, 1\} \mathbf{P}(dx, dy). \quad (5.7)$$

In particular, an inequality holds if the integrand on the right hand side is taken to be  $\|x - y\|$ .

### 5.1 Proofs

We now proceed to the proofs of Theorem 5.1 and Theorem 5.2, which will be broken into several steps. First, we show the following ‘‘smoothing property’’ of  $\mathcal{P}_T$ .

**Lemma 5.4** *There exists a constant  $K$  independent of  $\varepsilon$  such that, for every  $T > 0$ , one has*

$$\|\mathcal{P}_T\mu - \mathcal{P}_T\nu\|_{\mathbb{L},s} \leq K(T^{-\frac{1}{2}} + 1)\|\mu - \nu\|_{\mathbb{L}},$$

for every pair  $\mu, \nu$  of probability measures on  $X$ .

*Proof.* Since we assumed that  $P_c Q$  has maximal rank, we can rewrite (5.1) using the projection  $P_q$  that was defined in the proof of Lemma 4.7. We obtain:

$$\begin{aligned} dv_c &= A_c v_c dT + \mathcal{F}_c(v_c) dT + Q_c dw(T), \\ dv_s &= \varepsilon^{-2} L v_s dT + Q_s dw(T) + \tilde{Q}_s dw_\infty(T), \end{aligned} \tag{5.8}$$

where  $w$  is a standard  $n$ -dimensional Wiener process obtained from  $W$  by  $w = P_q W$ ,  $w_\infty$  is an independent cylindrical Wiener process in  $X$ , and the various covariance operators are given by

$$Q_c = P_c Q P_q, \quad Q_s = P_s Q P_q, \quad \tilde{Q}_s = P_s Q (1 - P_q).$$

The operator,  $Q_c$  can be identified with an invertible  $n \times n$  matrix by Assumption 3.7.

The proof of Lemma 5.4 follows easily from the definition of the norms, if we verify that for all measurable  $\varphi$  with  $\|\varphi\|_{\mathbb{L},s} \leq 1$  we have  $\|\mathcal{P}_T\varphi\|_{\mathbb{L}} \leq C(T^{-1/2} + 1)$  for some universal constant  $C > 0$ . In the following we fix  $\varphi$ .

First, we establish a uniform bound on the Fréchet derivative  $D_c \mathcal{P}_T \varphi$ . Denote by  $\Phi^T$  the stochastic flow generated by (5.8), i.e.  $\Phi^T(v_c(0), v_s(0)) = (v_c(T), v_s(T))$  for a solution of (5.8). Let  $\Phi_s^T$  and  $\Phi_c^T$  be its components in  $\mathcal{S}$  and  $\mathcal{N}$ , respectively and denote by  $D_s$  and  $D_c$  the Fréchet derivatives along the subspaces  $\mathcal{S}$  and  $\mathcal{N}$ .

Then, a trivial modification of the Bismut–Elworthy–Li formula [EL94, Thm. 2.1] yields

$$(D_c \mathcal{P}_T \varphi)(v_0)h = \frac{1}{T} \mathbf{E} \left( (\varphi \circ \Phi^T)(v_0) \int_0^T \langle Q_c^{-1}(D_c \Phi_c^S)(v_0)h, dw(S) \rangle \right),$$

with  $v_0 = (v_c(0), v_s(0))$  for short. Thus

$$\|(D_c \mathcal{P}_T \varphi)(v_0)\|^2 \leq \frac{C}{T^2} \int_0^T \mathbf{E} \|(D_c \Phi_c^S)(v_0)\|^2 dS \leq C(T^{-1} + 1),$$

where we used the uniform bound  $\|(D_c \Phi_c^T)(x)\| \leq e^{\|A_c\|T}$ , that can be obtained as in (4.20).

The contraction property of  $e^{Lt}$  implies  $\|(D\Phi_s^T)(v_0)\| \leq 1$  for all  $v_0 \in X$ . Since  $\|\varphi\|_{\mathbb{L},s} \leq 1$  by assumption, it is straightforward to verify that this implies that  $\|\mathcal{P}_T\varphi\|_{\mathbb{L},s} \leq C$ . Together with the estimate on  $D_c \mathcal{P}_T \varphi$  obtained above, this concludes the proof of Lemma 5.4.  $\square$

It follows immediately from the definitions that  $\|P_c^* \mu - P_c^* \nu\|_{\text{TV}} \leq \|\mu - \nu\|_{\text{L},s}$ , where  $P_c^* \mu$  is the marginal of  $\mu$  on  $P_c X$ . Therefore, the previous lemma says that if two measures are close in the Wasserstein distance  $\|\cdot\|_{\text{L}}$ , then their marginals on  $\mathcal{N}$  will be close in the total variation distance after a short time interval. The next lemma says that they will get even closer as time goes by.

**Lemma 5.5** *There exists a constant  $\kappa < 1$  such that*

$$\|\mathcal{P}_1^c \mu - \mathcal{P}_1^c \nu\|_{\text{TV}} \leq \kappa \|\mu - \nu\|_{\text{TV}}$$

for every pair  $\mu, \nu$  of probability measures on  $\mathcal{N}$ .

*Proof.* Denote by  $P_t^c(x, \cdot)$  the transition probabilities for (5.1a) and by  $B_r$  the ball of radius  $r$  centred around the origin. Using Lemma 4.5 we immediately get

$$P_T^c(x, B_r) > 1 - C(p, T)/r^p, \quad (5.9)$$

where the constant depends only on  $p$  and  $T$ , i.e. the estimate is uniform in  $x$ .

On the other hand, the following Lemma is well-known [MT94]:

**Lemma 5.6** *Let  $P_t(x, \cdot)$  a Markov family of transition probabilities on a locally compact state space  $X$  which is strong Feller and short-time irreducible. Then, for every compact set  $K \in X$  and every  $t > 0$ , there exists a  $\delta > 0$  and a probability measure  $\nu$  on  $X$  such that  $P_t(x, \cdot) \geq \delta \nu$  for every  $x \in K$ .*

Since the noise acts in a non-degenerate way on  $\mathcal{N}$  by Assumption 3.7 it is easy to check that the family  $P_t^c(x, \cdot)$  satisfies the assumptions of Lemma 5.6. Combining this with the estimate (5.9) concludes the proof of Lemma 5.5.  $\square$

The last ingredient we need for the proof of Theorem 5.1 is some control on what happens on the stable space  $\mathcal{S}$ . This is given by

**Lemma 5.7** *The estimate*

$$\|\mathcal{P}_T \mu - \mathcal{P}_T \nu\|_{\text{L}} \leq \|P_c^* \mu - P_c^* \nu\|_{\text{TV}} + M e^{-\varepsilon^{-2} \omega T} \int_X \|P_s x\| (\mu + \nu)(dx)$$

holds for every pair  $\mu, \nu$  of probability measures on  $X$ .

*Proof.* Let  $v^{(1)}(0)$  and  $v^{(2)}(0)$  be  $X$ -valued random variables with distributions  $\mu$  and  $\nu$ , respectively, and denote by  $v^{(1)}(T)$  and  $v^{(2)}(T)$  the corresponding solutions to (5.1). With this notation, we can write  $\|\mathcal{P}_T \mu - \mathcal{P}_T \nu\|_{\text{L}}$  as

$$\begin{aligned} \|\mathcal{P}_T \mu - \mathcal{P}_T \nu\|_{\text{L}} &= \sup_{\|\varphi\|_{\text{L}} \leq 1} |\mathbf{E} \varphi(v^{(1)}(T)) - \mathbf{E} \varphi(v^{(2)}(T))| \\ &\leq \sup_{\|\varphi\|_{\text{L}} \leq 1} \mathbf{E} |\varphi(v^{(1)}(T)) - \varphi(v^{(2)}(T))|. \end{aligned} \quad (5.10)$$

For  $\varphi$  with  $\|\varphi\|_L \leq 1$  rewrite as before  $\varphi(v)$  as  $\varphi(v_c, v_s)$ . We derive

$$\begin{aligned} |\varphi(v^{(1)}(T)) - \varphi(v^{(2)}(T))| &\leq |\varphi(v_c^{(1)}(T), v_s^{(1)}(T)) - \varphi(v_c^{(1)}(T), v_s^{(2)}(T))| \\ &\quad + |\varphi(v_c^{(1)}(T), v_s^{(2)}(T)) - \varphi(v_c^{(2)}(T), v_s^{(2)}(T))| \\ &\leq M e^{-\varepsilon^{-2}\omega t} \|v_s^{(1)}(0) - v_s^{(2)}(0)\| \\ &\quad + |\varphi(v_c^{(1)}(T), v_s^{(2)}(T)) - \varphi(v_c^{(2)}(T), v_s^{(2)}(T))|, \end{aligned}$$

where we used (3.1). By the definition of the total variation distance, it is possible to find a coupling between  $v^{(1)}(0)$  and  $v^{(2)}(0)$  such that  $\mathbf{P}\{v_c^{(1)}(0) \neq v_c^{(2)}(0)\} = \frac{1}{2}\|P_c^*\mu - P_c^*\nu\|_{\text{TV}}$ , and therefore also  $\mathbf{P}\{v_c^{(1)}(T) \neq v_c^{(2)}(T)\} = \frac{1}{2}\|P_c^*\mu - P_c^*\nu\|_{\text{TV}}$ . The claim follows immediately.  $\square$

We can now turn to the

*Proof of Theorem 5.1.* Using the fact that  $\mathbf{E}\|P_s W_L(t)\| \leq C$  for any  $t > 0$  and the contraction property (3.1), we immediately obtain that there exists a constant  $C_1$  such that

$$\int_X \|P_s x\| (\mathcal{P}_T \mu)(dx) \leq C_1 \int_X (1 + \|P_s x\|) \mu(dx),$$

for all  $t \geq 0$  and all probability measures  $\mu$ . One can then find a constant  $\tau > 1$  independent of  $\varepsilon \in (0, 1]$  such that

$$2C_1 M e^{-\varepsilon^{-2}\tau} \leq \varepsilon^2. \quad (5.11)$$

Furthermore, since  $\kappa < 1$ , one can find an integer  $N > 1$  such that  $2K\kappa^{N-1} < \frac{1}{2}$ , where  $K$  is the constant from Lemma 5.4. Choosing  $T_0 = N + \tau$  and combining the three previous lemmas, we obtain

$$\begin{aligned} \|\mathcal{P}_{T_0}\mu - \mathcal{P}_{T_0}\nu\|_L &\leq \|P_c^* \mathcal{P}_N \mu - P_c^* \mathcal{P}_N \nu\|_{\text{TV}} + \frac{\varepsilon^2}{C_1} \int_X \|P_s x\| \mathcal{P}_N(\mu + \nu)(dx) \\ &\leq \kappa^{N-1} \|P_c^* \mathcal{P}_1 \mu - P_c^* \mathcal{P}_1 \nu\|_{\text{TV}} + \varepsilon^2 \int_X (1 + \|P_s x\|) (\mu + \nu)(dx) \\ &\leq 2K\kappa^{N-1} \|\mu - \nu\|_L + \varepsilon^2 \int_X (1 + \|P_s x\|) (\mu + \nu)(dx), \end{aligned}$$

where we also used that obviously  $P_c^* \mathcal{P}_T = \mathcal{P}_T P_c^*$ .  $\square$

**Remark 5.8** It is clear from (5.11) that a far better asymptotic for the rest term could be achieved, but it is sufficient to establish a bound of order  $\varepsilon^2$  since our other error terms will only be of that order.

The remainder of this section is devoted to combining the contraction result from Theorem 5.1 with the results of the previous section in order to obtain information on the invariant measure for (1.1) (or equivalently on the invariant measure for the Markov semigroup  $\mathcal{Q}_T$ ). We first reformulate these results in the context of this section.

**Lemma 5.9** *Let  $\mu$  be a measure on  $X$  satisfying the bounds on the moments given by (5.3). Then, there exists a family of constants  $\{C'_p\}$  depending on  $\{C_p\}$  but not on  $\varepsilon$  or  $\mu$ , such that*

$$\begin{aligned} \int_X \|x\|^p (\mathcal{P}_T\mu)(dx) &< C'_p, & \int_X \|P_s x\|^p (\mathcal{P}_T\mu)(dx) &< C'_p \varepsilon^p, \\ \int_X \|x\|^p (\mathcal{Q}_T\mu)(dx) &< C'_p, & \int_X \|P_s x\|^p (\mathcal{Q}_T\mu)(dx) &< C'_p \varepsilon^p, \end{aligned}$$

for every  $p \geq 1$  and every  $T > 0$ .

*Proof.* One part of the proof is just a rescaled version of Theorem 4.1, and the part for  $\mathcal{P}_T$  is given in Lemma 4.5, where we just have to rescale in one case.  $\square$

**Lemma 5.10** *Let  $\mu$  be a measure on  $X$  satisfying (5.3). Then, for every  $T > 0$ , there exists a constant  $C_a$  depending on  $T_0$  and on the family of constants  $\{C_p\}$ , such that for all  $T \in [0, T_0]$*

$$\|\mathcal{P}_T\mu - \mathcal{Q}_T\mu\|_{\mathbb{L}} \leq C_a \varepsilon^{2-\kappa},$$

holds for every  $\varepsilon \in (0, 1)$  and every  $T \leq T_0$ .

*Proof.* This is just a restatement of Corollary 3.9.  $\square$

We are now ready to turn to the

*Proof of Theorem 5.2.* We fix once and for all the value  $T_0 > 0$  given by Theorem 5.1 and we choose two measures  $\mu$  and  $\nu$  satisfying (5.3). Combining Theorem 5.1 with Lemmas 5.9 and 5.10 obtained above, we get for an arbitrary integer  $n$ :

$$\begin{aligned} \|\mathcal{P}_{T_0}^n \mu - \mathcal{Q}_{T_0}^n \nu\|_{\mathbb{L}} &\leq \|\mathcal{P}_{T_0}(\mathcal{P}_{T_0}^{n-1} \mu - \mathcal{Q}_{T_0}^{n-1} \nu)\|_{\mathbb{L}} + \|(\mathcal{P}_{T_0} - \mathcal{Q}_{T_0})\mathcal{Q}_{T_0}^{n-1} \nu\|_{\mathbb{L}} \\ &\leq \frac{1}{2} \|\mathcal{P}_{T_0}^{n-1} \mu - \mathcal{Q}_{T_0}^{n-1} \nu\|_{\mathbb{L}} + C_b \varepsilon^2 + C_a \varepsilon^{2-\kappa}. \end{aligned}$$

Iterating this estimate, one readily obtains by Lemma 5.4 that, for any  $\tau \in [0, T_0]$

$$\begin{aligned} \|\mathcal{P}_{nT_0+\tau} \mu - \mathcal{Q}_{nT_0+\tau} \nu\|_{\mathbb{L}} &\leq \frac{1}{2^{n-1}} \|\mathcal{P}_{\tau+T_0} \mu - \mathcal{Q}_{\tau+T_0} \nu\|_{\mathbb{L}} + 2C_b \varepsilon^2 + 2C_a \varepsilon^{2-\kappa} \\ &\leq \frac{1}{2^{n-1}} K(T_0^{-1/2} + 1) \|\mu - \nu\|_{\mathbb{L}} + C_* \varepsilon^{2-\kappa}, \end{aligned}$$

with  $C_* = 2(C_a + C_b)$ . Choosing  $\gamma = \frac{\log 2}{T_0}$ , (5.4) follows easily.  $\square$

## 5.2 The non-diagonal case

This section is dedicated to the

*Proof of Theorem 5.3.* Denote by  $\mathcal{P}_T(u, \cdot)$  the transition probabilities for (5.1). The idea of the proof is to use the fact that, on small time-scales, the drift term for (5.1a) is not important and therefore  $\mathcal{P}_T(u, \cdot)$  is close to a Gaussian measure with mean  $e^{\varepsilon^{-2}LT}u$  and with covariance operator

$$\tilde{C}_T = \int_0^T e^{\varepsilon^{-2}Lr} Q^2 e^{\varepsilon^{-2}L^*r} dr, \quad (5.12)$$

where  $L^*$  denotes the adjoint of  $L$  in  $X$ . Denote by  $\tilde{\mathcal{P}}_T(u, \cdot)$  the Gaussian measure on  $X$  with mean  $P_c u$  and covariance operator  $\tilde{C}_T$ . We then have

**Lemma 5.11** *Fix an arbitrary  $\kappa > 0$  and fix  $T_\varepsilon = \varepsilon^{2-\kappa}$ . Then, there exists a constant  $C > 0$  independent of  $\varepsilon$  such that the measure  $\mu_T$ , defined by*

$$\mu_T(\cdot) = \int_{\mathcal{N}} \tilde{\mathcal{P}}_T(u, \cdot) \mu_\star^c(du) = \int_X \tilde{\mathcal{P}}_T(u, \cdot) \mu_\star(du), \quad (5.13)$$

satisfies  $\|\mu_{T_\varepsilon} - \mu_\star\|_L \leq C\varepsilon^{2-\kappa}$ .

*Proof.* As  $\mu_\star(\cdot) = \int_X \mathcal{P}_T(u, \cdot) \mu_\star(du)$ , it suffices to show that

$$\int_X \|\tilde{\mathcal{P}}_T(u, \cdot) - \mathcal{P}_T(u, \cdot)\|_L \mu_\star(du) \leq C\varepsilon^{2-\kappa}.$$

We first show that there exists a constant  $C > 0$  such that, for all  $T > 0$ , we have

$$\int_X \|\hat{\mathcal{P}}_T(u, \cdot) - \mathcal{P}_T(u, \cdot)\|_L \mu_\star(du) \leq CT, \quad (5.14)$$

with  $\hat{\mathcal{P}}_T(u, \cdot)$  the Gaussian measure of mean  $e^{\varepsilon^{-2}LT}u$  and covariance  $\tilde{C}_T$  given by (5.12).

Let  $(v_c, v_s)$  be a stationary solution of the approximating equations (5.1) with initial condition  $v$ . Then  $\mathcal{P}_T(v, \cdot)$  is given by the law of this solution. Since  $\hat{\mathcal{P}}_T(v, \cdot)$  corresponds to the dynamics without drift in  $\mathcal{N}$ , we can choose the process  $(v_c(0) + P_c Q \tilde{W}, v_s)$  to realise  $\hat{\mathcal{P}}_T(v, \cdot)$ . Now one has by the definition of  $\|\cdot\|_L$  the bound

$$\int_X \|\hat{\mathcal{P}}_T(v, \cdot) - \mathcal{P}_T(v, \cdot)\|_L \mu_\star(dv) \leq \mathbf{E} \int_0^T C(1 + \|v_c(t)\|^3) dt,$$

where we used (5.1a) and straightforward estimates. Since the invariant measure  $\mu_\star$  has all moments bounded of order  $\mathcal{O}(1)$  by the rescaled version of Lemma 4.1, the bound (5.14) follows.

Furthermore, we immediately get the bound

$$\|\hat{\mathcal{P}}_T(u, \cdot) - \tilde{\mathcal{P}}_T(u, \cdot)\|_L \leq \|e^{\varepsilon^{-2}LT} P_s u\| \leq e^{-\varepsilon^{-2}wT} \|P_s u\|,$$

as the Gaussian measures  $\mathcal{P}_T(u, \cdot)$  and  $\tilde{\mathcal{P}}_T(u, \cdot)$  differ only by the term  $e^{\varepsilon^{-2}LT} P_s u$  in the mean.

Using both parts of the proof and again the *a priori* bounds on the moments of  $\mu_\star$ , Lemma 5.11 follows at once.  $\square$

Using Lemma 5.11, Theorem 5.3 will follow, if we can show that

$$\|\mu_{T_\varepsilon} - \mu_\star^c \otimes \mu_\star^s\|_{\mathbf{L}} \leq C\varepsilon^{2-\kappa} \quad (5.15)$$

holds for  $T_\varepsilon = \varepsilon^{2-\kappa}$ . Let us write  $\tilde{\mathcal{P}}_T^s$  for the marginal of  $\tilde{\mathcal{P}}_T(u, \cdot)$  onto  $\mathcal{S} = P_s X$ . (Note that this projection is independent of  $u$ .) Recall that the covariance of  $\tilde{\mathcal{P}}_T^s$  is

$$\tilde{C}_T^s = \varepsilon^2 \int_0^{\varepsilon^{-2}T} e^{L_s r} P_s Q^2 P_s e^{L_s^* r} dr, \quad (5.16)$$

whereas the covariance of  $\mu_\star^s$  is

$$\tilde{C}_\star^s = \varepsilon^2 \int_0^\infty e^{L_s r} P_s Q^2 P_s e^{L_s^* r} dr.$$

Since the integrand is positive definite, it is obvious that  $\tilde{C}_\star^s - \tilde{C}_T^s$  is positive definite. One can thus write a random variable with law  $\mu_\star^s$  as the sum of one random variable with law  $\tilde{\mathcal{P}}_T^s$  and an independent centred Gaussian random variable with covariance  $\tilde{C}_\star^s - \tilde{C}_T^s$ . The representation (5.7) then immediately yields the bound

$$\|\tilde{\mathcal{P}}_T^s - \mu_\star^s\|_{\mathbf{L}} \leq \sqrt{\text{tr}(\tilde{C}_\star^s - \tilde{C}_T^s)},$$

which in turn is bounded by  $C\varepsilon e^{-\omega\varepsilon^{-2}T}$  from (3.6). For  $T = T_\varepsilon$  this expression is smaller than  $\varepsilon^2$ , so it suffices to show

$$\|\mu_{T_\varepsilon} - \mu_\star^c \otimes \tilde{\mathcal{P}}_{T_\varepsilon}^s\|_{\mathbf{L}} \leq C\varepsilon^{2-\kappa}. \quad (5.17)$$

Until the end of this proof, we will denote elements of  $\mathcal{N}$  by  $x$  and elements of  $\mathcal{S}$  by  $y$ ,  $X = \mathcal{N} \oplus \mathcal{S}$  (cf. Assumption 3.1). We will also denote the measure  $\tilde{\mathcal{P}}_T(0, \cdot)$  by  $\tilde{\mathcal{P}}_T(dx, dy)$ . Since (5.1a) is a finite-dimensional non-degenerate SDE with smooth coefficients, its invariant measure  $\mu_\star^c$  has a smooth density  $\varrho(x)$ . Furthermore, as shown in Theorem A.1 of the appendix,  $\varrho$  actually belongs to the Schwartz space. In particular, we use  $\varrho \in W^{2,1}(\mathcal{N})$ , *i.e.* there exists a constant  $c_\varrho$  such that

$$\int_{\mathcal{N}} |\nabla \varrho(x)| dx \leq c_\varrho, \quad \int_{\mathcal{N}} |\nabla^2 \varrho(x)| dx \leq c_\varrho, \quad (5.18)$$

where  $\nabla$  denotes the gradient and  $\nabla^2$  the Hessian.

In order to obtain the required estimate, we fix a test function  $\varphi : X \rightarrow \mathbf{R}$  with  $\|\varphi\|_{\mathbf{L}} \leq 1$  and, using Taylor series, we write

$$\int_X \varphi(x, y) \mu_T(dx, dy) = \int_X \int_{\mathcal{N}} \varphi(x, y') \varrho(x - x') dx \tilde{\mathcal{P}}_T(dx', dy')$$

$$\begin{aligned}
&= \int_X \int_{\mathcal{N}} \varphi(x, y') \varrho(x) dx \tilde{\mathcal{P}}_T(dx', dy') \\
&\quad - \int_X \int_{\mathcal{N}} \varphi(x, y') \langle x', \nabla \varrho(x) \rangle dx \tilde{\mathcal{P}}_T(dx', dy') \\
&\quad + \int_X \int_{\mathcal{N}} \int_0^1 \varphi(x, y') \langle x', \nabla^2 \varrho(x - rx') x' \rangle dr dx \tilde{\mathcal{P}}_T(dx', dy') \\
&=: I_1(\varphi) + I_2(\varphi) + I_3(\varphi) .
\end{aligned}$$

By definition,  $\int_{\mathcal{N}} \tilde{\mathcal{P}}_T(dx', \cdot) = \tilde{\mathcal{P}}_T^s$ , hence

$$I_1(\varphi) = \int_S \int_{\mathcal{N}} \varphi(x, y) \mu_\star^c(dx) \tilde{\mathcal{P}}_T^s(dy) = \int_S \int_{\mathcal{N}} \varphi(x, y) \mu_\star^c \otimes \tilde{\mathcal{P}}_T^s(dx, dy) ,$$

and therefore,

$$\|\mu_T - \mu_\star^c \otimes \tilde{\mathcal{P}}_T^s\|_{\mathbb{L}} = \sup_{\varphi \in \mathbb{L}} \{|I_2(\varphi) + I_3(\varphi)|\} .$$

Using (5.18),  $|\varphi(x, y)| \leq 1$ , and the definition of  $\tilde{\mathcal{P}}_T$ , it is straightforward to bound  $I_3(\varphi)$  by the second moment of  $P_c^* \tilde{\mathcal{P}}_T$ . Since this is a centred Gaussian with covariance matrix  $P_c C_T^2 P_c = T \cdot P_c Q^2 P_c$ ,

$$I_3(\varphi) \leq c_\varrho \operatorname{tr}(P_c Q^2 P_c) T ,$$

which in turn is smaller than  $C\varepsilon^{2-\kappa}$  for  $T = T_\varepsilon$ .

Define  $\psi(y) = \int_{\mathcal{N}} \varphi(x, y) \nabla \varrho(x) dx$ . If  $\|\varphi\|_{\mathbb{L}} \leq 1$ , the function  $\psi$  is obviously Lipschitz continuous with constant smaller than  $c_\varrho$ , and we have for  $I_2(\varphi)$

$$\begin{aligned}
|I_2(\varphi)| &= \left| \int_X \langle \psi(y), x \rangle \tilde{\mathcal{P}}_T(dx, dy) \right| = \left| \int_X \langle (\psi(y) - \psi(0)), x \rangle \tilde{\mathcal{P}}_T(dx, dy) \right| \\
&\leq c_\varrho \int_X \|x\| \cdot \|y\| \tilde{\mathcal{P}}_T(dx, dy) \leq c_\varrho \sqrt{\operatorname{tr}(\tilde{C}_T^s) T \operatorname{tr}(P_c Q^2 P_c)} ,
\end{aligned}$$

where the last step uses Cauchy–Schwarz inequality and (5.12). Using (5.16) and (3.6), we verify  $|I_2(\varphi)| \leq C\varepsilon\sqrt{T}$ , and the proof of (5.17) and hence Theorem 5.3 is complete.  $\square$

## 6 Total Variation Bounds

So far, most estimates obtained in this paper were formulated using the Wasserstein distance between measures. This distance is strongly linked to the topology of the underlying space. For example, the Wasserstein distance between Dirac measures located at two points  $x$  and  $y$  is given by two times the distance between  $x$  and  $y$ , as long as this distance does not exceed 1. On the other hand, the total variation distance between two such Dirac measures is always 2, unless  $x = y$ . Also, if one scales the distance of the underlying space by a factor  $\lambda$ , the corresponding Wasserstein distance between two given measures will also scale (approximately



for small distances) by  $\lambda$ , whereas the total variation distance between the two measures will always remain the same.

However, if we take  $\mu$  to be the centred Gaussian measure on  $\mathbf{R}$  with covariance 1 and  $\nu$  to be its translate by a distance  $\varepsilon$ , one has  $\|\mu - \nu\|_{\text{TV}} \approx C\varepsilon$  and  $\|\mu - \nu\|_{\text{L}} \approx C\varepsilon$  for  $\varepsilon$  small. One might thus expect that the total variation distance between two measures  $\mu$  and  $\nu$  is in general comparable to their Wasserstein distance, as long as  $\mu$  and  $\nu$  are scaled in such a way that they are localised in a region of about unity size and have smooth densities with derivatives of roughly order one. This is in general not true, as shown by the following example:

$$\mu(dx) = e^{-x^2} dx, \quad \nu(dx) = (e^{-x^2} + xe^{-(x/\varepsilon)^2}) dx. \quad (6.1)$$

In this case, a straightforward computation, using that the total variation distance is the  $L^1$ -distance for densities, shows that one has  $\|\mu - \nu\|_{\text{TV}} = \varepsilon^2$ . On the other hand it is easy to verify that  $c\varepsilon^3 \leq \|\mu - \nu\|_{\text{L}} \leq \varepsilon^3$  for some constant  $c \in (0, 1)$ . Actually, one can show that if both  $\mu$  and  $\nu$  have smooth densities with integrable first and second derivatives of order one, one has an estimate of the type

$$\|\mu - \nu\|_{\text{TV}} \leq C\|\mu - \nu\|_{\text{L}}^{2/3}.$$

To prove this, it suffices indeed to bound  $\|\mu - \nu\|_{\text{TV}}$  by

$$\|\mu - \nu\|_{\text{TV}} \leq \|\mu - \mu \star G_t\|_{\text{TV}} + \|\mu \star G_t - \nu \star G_t\|_{\text{TV}} + \|\nu - \nu \star G_t\|_{\text{TV}},$$

where  $G_t$  is the Gaussian measure with covariance  $t$ . Using the smoothness of the densities, the first and the last term are bounded by  $Ct$ . The middle term is bounded by  $C\|\mu - \nu\|_{\text{L}}/\sqrt{t}$ , using a computation very similar to the one that yields the bound (6.11) below. The claim follows by optimising  $t$ . The above example (6.1) shows furthermore that this estimate is sharp.

These considerations show that, by combining the results of section 5 with smoothness properties of the measures, one may expect to obtain bounds of the type

$$\|P_c^* \mu_\star - P_c^* \nu_\star\|_{\text{TV}} \approx C\varepsilon^{4/3}, \quad \|\mu_\star - \nu_\star\|_{\text{TV}} \approx C\varepsilon^{2/3}.$$

Here, we again denote by  $\mu_\star$  the invariant measure for (5.1) and by  $\nu_\star$  the invariant measure for (5.2). Our aim in this section is to show that one can even get slightly better estimates than that, namely we will show in Theorems 6.1 and 6.9 that, for  $\kappa$  arbitrarily small, one has

$$\|P_c^* \mu_\star - P_c^* \nu_\star\|_{\text{TV}} \leq C\varepsilon^{3/2-\kappa}, \quad \|\mu_\star - \nu_\star\|_{\text{TV}} \leq C\varepsilon^{1-\kappa}.$$

These estimates will however require us to put rather strict lower bounds on the covariance of the noise, as stated in Assumption 6.3 below. The main reason why, in the total variation distance, we do not achieve an accuracy of  $\varepsilon^2$  in  $\mathcal{N}$  is that, in our formal derivation of (5.1), we approximated  $v_s(T\varepsilon^{-2})$  by a white noise process with intensity  $\varepsilon$ . This approximation is justified in the weak topology, but not in the total variation topology where the distinction on a ‘‘microscopic’’ level (the regularity properties of both processes are of course completely different) becomes apparent.

### 6.1 Total variation distance in $\mathcal{N}$

In this subsection, we show the following result:

**Theorem 6.1** *Suppose all assumptions of section 3 are satisfied, and let  $\mu_\star$  and  $\nu_\star$  denote invariant measures for (5.1) and (5.2), respectively. Then for every  $\kappa > 0$ , there exists a constant  $C$  such that*

$$\|P_c^* \mu_\star - P_c^* \nu_\star\|_{\text{TV}} \leq C\varepsilon^{3/2-\kappa}$$

holds for every  $\varepsilon$  small enough.

The following lemma is an extension of Girsanov's theorem and will be useful for the proofs.

**Lemma 6.2** *Let  $\nu$  denote the Wiener measure on the interval  $[0, T]$  and let  $h$  be an adapted process satisfying  $h_\infty = \text{ess sup}_{w \in \Omega} \sup_{t \in [0, T]} |h(t, w)| < \infty$ . Let  $\mu$  be the measure associated to the Girsanov transformation*

$$w \mapsto \tilde{w} = w + \int_0^\cdot h(s, w) ds .$$

(i.e.  $\tilde{w}$  is again a Wiener process under  $\mu$ .) *There exists a universal constant  $C$  such that one has the bound*

$$\|\mu - \nu\|_{\text{TV}} \leq Ch_\infty \sqrt{T} .$$

This lemma is only formulated and proved in  $\mathbf{R}$ , but it is easy to see that it holds for any Hilbert space valued Wiener process, provided we use the norm of the corresponding Cameron–Martin space in the definition of  $h_\infty$ .

*Proof.* By Girsanov's theorem,  $\mu$  and  $\nu$  are mutually absolutely continuous and one has

$$\mathcal{D}(w) = \frac{d\mu}{d\nu}(w) = \exp\left(\int_0^T h(s, w) dw(s) - \frac{1}{2} \int_0^T |h(s, w)|^2 ds\right) .$$

The total variation distance is then bounded by

$$\begin{aligned} \|\mu - \nu\|_{\text{TV}}^2 &= \left(\int |1 - \mathcal{D}(w)| \nu(dw)\right)^2 \leq \int (1 - \mathcal{D}(w))^2 \nu(dw) \\ &= \int \mathcal{D}^2(w) \nu(dw) - 1 \leq e^{h_\infty^2 T} - 1 , \end{aligned}$$

where we used in the last step, that  $\exp\{\int_0^T 2h(s)dw(s) - \frac{1}{2} \int_0^T |2h(s)|^2 ds\}$  is a martingale. Since on the other hand one has by definition  $\|\mu - \nu\|_{\text{TV}}^2 \leq 4$ , one gets  $\|\mu - \nu\|_{\text{TV}}^2 \leq Ch_\infty^2 T$ , which concludes the proof of the lemma.  $\square$

We now turn to the

*Proof of Theorem 6.1.* Let us denote as previously by  $\mathcal{P}_T$  the semigroup generated by (5.1) and by  $\mathcal{Q}_T$  the semigroup generated by the rescaled SPDE (5.2). We start by showing that, for every  $\kappa > 0$  there exists a constant  $C$  such that, for every  $T \in (0, 1)$ ,

$$\|P_c^* \mathcal{P}_T \mu - P_c^* \mathcal{Q}_T \mu\|_{\text{TV}} \leq C \varepsilon^{1-\kappa} \sqrt{T} + C \varepsilon^2 \quad (6.2)$$

holds for every measure  $\mu$  with the property that

$$\int_X \|u_c\|^p \mu(du) < C_p \quad \text{and} \quad \int_X \|u_s\|^p \mu(du) < C_p \varepsilon^p \quad (6.3)$$

for every  $\varepsilon \in (0, 1)$ . (The constant  $C$  does of course depend on the family of constants  $C_p$ .)

We fix a smooth non-decreasing cut-off function  $\bar{\chi} : \mathbf{R}_+ \rightarrow [0, 1]$  with the properties that  $\bar{\chi}(x) = 0$  if  $x < 1$ ,  $\bar{\chi}(x) = 1$  if  $x > 2$ , and  $|\bar{\chi}'(x)| \leq 2$ . We then define  $\chi : \mathcal{N} \rightarrow \mathcal{N}$  by  $\chi(x) = x \bar{\chi}(\|x\|)$ . We also define  $\chi_\delta(x) = \delta \bar{\chi}(x/\delta)$ . Given some value  $\delta \in (0, 1)$ , we denote by  $\mathcal{Q}_T^\delta$  the semigroup generated by the solutions to the equation

$$\begin{aligned} du &= \varepsilon^{-2} Lu dT + Au dT + \mathcal{F}(u) dT + Q dW(T) \\ &\quad - \chi_\delta(A_c u_s + \mathcal{F}_c(u) - \mathcal{F}_c(u_c)) dT . \end{aligned} \quad (6.4)$$

Notice that there exists a constant  $C_\star$  such that, as long as  $\|u_s\|(1 + \|u_c\|^2) \leq C_\star \delta$ , the dynamics of (6.4) and of the original equation (5.2) coincide. Choose  $p > 1$  to be fixed later. By rescaled versions of Lemma 4.3 and Theorem 3.8, there exists a constant  $C$  depending on  $p$  and on  $\kappa$  such that

$$\begin{aligned} \mathbf{E} \left( \sup_{T \in [0, 1]} (\|u_s(T)\|(1 + \|u_c(T)\|^2))^p \right) \\ \leq \frac{1}{2} \mathbf{E} \sup_{T \in [0, 1]} \left( \frac{\|u_s(T)\|^{2p}}{\varepsilon^p} + \varepsilon^p (1 + \|u_c(T)\|^2)^{2p} \right) \\ \leq C \varepsilon^{p(1-\kappa/2)} , \end{aligned}$$

where  $(u_c(T), u_s(T))$  denotes the solution of (5.2).

If we choose now  $\delta_\varepsilon = \varepsilon^{1-\kappa}$  and  $p = 4/\kappa$ , we get the estimate

$$\mathbf{P} \left( \sup_{T \in [0, 1]} \|u_s(T)\|(1 + \|u_c(T)\|^2) > C_\star \delta_\varepsilon \right) \leq C \frac{\varepsilon^{p(1-\kappa/2)}}{\delta_\varepsilon^p} \leq C \varepsilon^2 .$$

Hence (for a suitable coupling between the two driving noise processes), solutions of (6.4) and of (5.2) coincide for  $T \in [0, 1]$  with probability larger than  $1 - C \varepsilon^2$ . This in turn yields the total variation estimate

$$\|P_c^* \mathcal{Q}_T^{\delta_\varepsilon} \mu - P_c^* \mathcal{Q}_T \mu\|_{\text{TV}} \leq C \varepsilon^2 , \quad (6.5)$$

holding uniformly for  $T \in (0, 1)$ . On the other hand, we can apply Lemma 6.2 to estimate the difference between  $\mathcal{Q}_T^\delta$  and  $\mathcal{P}_T$  by simply taking

$$Qh = (I - \chi_\delta)(A_c u_s + \mathcal{F}_c(u) - \mathcal{F}_c(u_c)) .$$

(Here  $I$  denotes the identity.) Since the norm of the function  $I - \chi_\delta$ , and thus the norm of  $Qh$ , is bounded by  $2\delta$ , we can apply the generalised version of Lemma 6.2, bearing in mind that by using a bounded pseudo-inverse of  $Q$ , which exists by Assumption 3.7 we can find some process  $h$  that is uniformly bounded by  $C\delta$ . Now

$$\|P_c^* \mathcal{P}_T \mu - P_c^* \mathcal{Q}_T^\delta \mu\|_{\text{TV}} \leq C\delta\sqrt{T}. \quad (6.6)$$

For  $\delta_\varepsilon = \varepsilon^{1-\kappa}$  the inequalities (6.5) and (6.6) immediately imply (6.2).

We can now estimate  $\|P_c^* \mu_\star - P_c^* \nu_\star\|_{\text{TV}}$  by

$$\begin{aligned} \|P_c^* \mu_\star - P_c^* \nu_\star\|_{\text{TV}} &\leq \|\mathcal{P}_T \mu_\star - \mathcal{P}_T \nu_\star\|_{\text{L},s} + \|P_c^* \mathcal{P}_T \nu_\star - P_c^* \mathcal{Q}_T \nu_\star\|_{\text{TV}} \\ &\leq C \frac{\varepsilon^{2-\kappa}}{\sqrt{T}} + C\varepsilon^{1-\kappa}\sqrt{T} + C\varepsilon^2, \end{aligned}$$

where we made use of Lemma 5.4 and Theorem 5.2 to bound the first term and of (6.2) to bound the second term. Theorem 4.1 ensures that (6.3) is fulfilled. Taking  $T = \varepsilon$  concludes the proof of Theorem 6.1.  $\square$

## 6.2 Total variation in $X$

In order to get bounds on  $\|\mu_\star - \nu_\star\|_{\text{TV}}$ , we have to put additional hypotheses on  $Q$  and  $\mathcal{F}$ .

**Assumption 6.3** *Let  $\alpha$  be as in Assumptions 3.2 and 3.3. There exists a constant  $\gamma_0 > 0$  such that, for all  $\gamma \in [0, \gamma_0]$ ,  $\mathcal{F} : (X^\gamma)^3 \rightarrow X^{\gamma-\alpha}$  and  $A : X^\gamma \rightarrow X^{\gamma-\alpha}$  are continuous. Furthermore, the operator  $Q^{-1}$  is continuous from  $X^{\gamma_0-\alpha}$  to  $X$  and for some  $\tilde{\alpha} \in [0, \frac{1}{2}]$  we have  $\|(1-L)^{\gamma_0-\tilde{\alpha}}Q\|_{\text{HS}(X)} < \infty$ .*

**Remark 6.4** Basically, we need that  $Q^{-1}\mathcal{F}(u(T))$  and  $Q^{-1}Au(T)$  are bounded in  $X$  for a Girsanov argument. Hence, we need bounds on  $u(T)$  in  $X^\gamma$ , which will be established in Lemma 6.8 by a bootstrapping technique. Moreover, we need further assumptions on the inverse  $Q^{-1}$ . This was not necessary in the previous result for the centre part, as we could use a pseudo-inverse, which existed by Assumption 3.7.

The *main result* of this section (cf. Theorem 6.9) shows that, if the noise is sufficiently “rough”, one has the *total variation* estimate

$$\|\nu_\star - \mu_\star^s \otimes \mu_\star^c\|_{\text{TV}} \leq C\varepsilon^{1-\kappa}. \quad (6.7)$$

As discussed at the beginning of this section, we are able to approximate  $\nu_\star$  only up to an error of order  $\varepsilon$  in the total variation distance, instead of the error of order  $\varepsilon^2$  that we achieved in the Wasserstein distance.

### 6.3 Preliminary estimates

In order to prepare the proof of (6.7), we formulate several technical results. First, we estimate the speed at which solutions to the linear equation

$$dv_s = \varepsilon^{-2} L_s v_s dT + Q_s d\mathcal{W} \quad (6.8)$$

converge toward each other in the total variation distance. Here, we consider solutions to (6.8) with values in  $\mathcal{S}$ , we defined  $Q_s^2 = P_s Q^2 P_s$  and  $L_s = P_s L$  from Assumption 3.1. Moreover  $\mathcal{W}$  is a cylindrical Wiener process on  $\mathcal{S}$  such that  $Q_s \mathcal{W} = P_s Q W$ , with  $W$  from Assumption 3.4. We denote by  $\mathcal{P}_T^s$  the semigroup generated by (6.8), which is the same semigroup than the one generated by (5.1b).

**Lemma 6.5** *Let Assumptions 3.1 and 3.4 be satisfied and assume that  $Q_s^{-1} L_s^{-\delta} \in \mathcal{L}(\mathcal{S}, \mathcal{S})$  for some  $\delta > 0$  (this is ensured, e.g., by Assumption 6.3).*

*Then, there exists a constant  $C$  such that, for every  $T \geq \varepsilon^2$  and for every pair of probability measures  $\mu$  and  $\nu$  on  $\mathcal{S}$  with bounded first order moments, one has the estimate*

$$\|\mathcal{P}_T^s \mu - \mathcal{P}_T^s \nu\|_{\text{TV}} \leq C \varepsilon^{-1} e^{-\omega \varepsilon^{-2} T} \int_{\mathcal{S}} \|u\| (\mu + \nu)(du), \quad (6.9)$$

where  $\omega$  is the constant appearing in (3.1).

*Proof.* Take  $\varphi : \mathcal{S} \rightarrow \mathbf{R}$  to be a bounded Borel-measurable function. The Bismut–Elworthy–Li formula [EL94] yields (after a simple substitution) for the Fréchet derivative of  $\mathcal{P}_{\varepsilon^2}^s \varphi$  in the direction  $h$ :

$$(D_h \mathcal{P}_{\varepsilon^2}^s \varphi)(v_s(0)) = \frac{2}{\varepsilon} \mathbf{E} \left( \varphi(v_s(\varepsilon^2)) \int_{1/4}^{3/4} \langle Q_s^{-1} e^{L_s(1-\tau)} h, d\tilde{\mathcal{W}}(\tau) \rangle \right), \quad (6.10)$$

where  $\tilde{\mathcal{W}}(t) = \varepsilon^{-1} \mathcal{W}(\varepsilon^2 t)$  is equal in law to  $\mathcal{W}$ . By assumption,  $Q_s^{-1} L_s^{-\delta}$  is bounded, so  $\|Q_s^{-1} e^{L_s(1-\tau)}\| \leq C(1-\tau)^{-\delta}$ . Applying Cauchy–Schwarz to (6.10) yields

$$\|(D \mathcal{P}_{\varepsilon^2}^s \varphi)(u)\| \leq C \varepsilon^{-1} \sqrt{\int_{1/4}^{3/4} (1-\tau)^{-2\delta} d\tau} \leq C \varepsilon^{-1}.$$

Let us denote by  $\hat{\mathcal{P}}_T^s(u, \cdot)$  the transition probabilities corresponding to (6.8). With this notation, we just proved that

$$\|\mathcal{P}_{\varepsilon^2}^s(u, \cdot) - \mathcal{P}_{\varepsilon^2}^s(v, \cdot)\|_{\text{TV}} \leq C \varepsilon^{-1} \|u - v\| \quad (6.11)$$

holds for every  $u$  and  $v$  in  $\mathcal{S}$ . Denote now by  $v_s(T, u_0)$  the solution at time  $T$  of (6.8) with initial condition  $u_0$  distributed according to the measure  $\mu$ . Let  $v_0$  be another initial condition independent of  $u_0$  and distributed according to  $\nu$ . Using first (6.11) and then (3.1), we derive the bound

$$\|\mathcal{P}_T^s \mu - \mathcal{P}_T^s \nu\|_{\text{TV}} \leq C \varepsilon^{-1} \mathbf{E} \|v_s(T - \varepsilon^2, u_0) - v_s(T - \varepsilon^2, v_0)\|$$

$$\begin{aligned} &\leq C\varepsilon^{-1}e^{-\omega\varepsilon^{-2}T}\mathbf{E}\|u_0 - v_0\| \\ &\leq C\varepsilon^{-1}e^{-\omega\varepsilon^{-2}T}\mathbf{E}(\|u_0\| + \|v_0\|), \end{aligned}$$

which is nothing but the estimate (6.9).  $\square$

We will also need the following elementary property of Gaussian measures.

**Lemma 6.6** *Let  $\mu$  be a centred Gaussian measure on a Hilbert space  $\mathcal{H} = \mathcal{H}_x \oplus \mathcal{H}_y$ . Denote by  $\mu_x$  the marginal of  $\mu$  on  $\mathcal{H}_x$  and similarly for  $\mu_y$ . Then, there exists a  $\mu_y$ -measurable linear map  $\mathcal{A} : \mathcal{H}_y \rightarrow \mathcal{H}_x$  and a centred Gaussian measure  $\mu_x^\circ$  on  $\mathcal{H}_x$  such that, for every measurable  $f : \mathcal{H} \rightarrow \mathbf{R}$ , one has*

$$\int_{\mathcal{H}} f(x, y) \mu(dx, dy) = \int_{\mathcal{H}_y} \int_{\mathcal{H}_x} f(x + \mathcal{A}y, y) \mu_x^\circ(dx) \mu_y(dy).$$

Furthermore, one has the inequalities

$$\begin{aligned} \int_{\mathcal{H}_y} \|\mathcal{A}y\|^2 \mu_y(dy) &\leq \int_{\mathcal{H}_x} \|x\|^2 \mu_x(dx), \\ \int_{\mathcal{H}_x} \|x\|^2 \mu_x^\circ(dx) &\leq \int_{\mathcal{H}_x} \|x\|^2 \mu_x(dx). \end{aligned}$$

*Proof.* The existence of  $\mathcal{A}$  and of  $\mu_x^\circ$  is standard. All we need is that the distribution of a Gaussian, given a marginal, is still a Gaussian, and that  $\mathcal{A}y = \mathbf{E}_\mu(x|y)$  is linear in  $y$ .

The first inequality follows immediately by taking expectations with respect to  $y$  on both sides of the following Cauchy–Schwarz inequality:

$$\|\mathbf{E}_\mu(x|y)\|^2 \leq \mathbf{E}_\mu(\|x\|^2|y).$$

To prove the second inequality, observe that

$$\int_{\mathcal{H}_x} \|x\|^2 \tilde{\mu}_x(dx) = \mathbf{E}_\mu(\|x - \mathcal{A}y\|^2|y) = \mathbf{E}_\mu(\|x\|^2|y) - \|\mathcal{A}y\|^2,$$

and take expectations over  $y$  on both sides of this expression.  $\square$

Finally, we show the following estimate on the total variation distance between a measure with a smooth density and its translates.

**Lemma 6.7** *Let  $\mu$  be a measure on  $\mathbf{R}^n$  with a smooth density  $\varrho$  with respect to the Lebesgue measure and such that  $\nabla\varrho$  is integrable. Let  $a \in \mathbf{R}^n$  and let  $\tau_a$  be the shift map defined by  $\tau_a(x) = x + a$ . Then, one has the estimate*

$$\|\mu - \tau_a^*\mu\|_{\text{TV}} \leq |a| \int_{\mathbf{R}^n} |\nabla\varrho(x)| dx,$$

where the absolute values in the right-hand side denote the Euclidean norm on  $\mathbf{R}^n$ .

*Proof.* Let  $f : \mathbf{R}^n \rightarrow \mathbf{R}$  be a measurable test function bounded by 1. We have

$$\begin{aligned} \left| \int_{\mathbf{R}^n} f(x+a)\varrho(x) dx - \int_{\mathbf{R}^n} f(x)\varrho(x) dx \right| &= \left| \int_{\mathbf{R}^n} f(x)(\varrho(x-a) - \varrho(x)) dx \right| \\ &= \left| \int_{\mathbf{R}^n} f(x) \int_0^1 \langle \nabla \varrho(x-at), a \rangle dt dx \right| \\ &\leq \int_{\mathbf{R}^n} |\nabla \varrho| \cdot |a| dx . \end{aligned}$$

The claim follows by taking the supremum over  $f$ .  $\square$

Another ingredient for our main result will be the following improvement of the *a priori* estimate from Theorem 4.1, which relies on a bootstrapping technique and on Assumption 6.3.

**Lemma 6.8** *Let all assumptions of section 3, as well as Assumption 6.3 be satisfied. Moreover, choose an initial condition  $u_0$  such that for any  $\kappa > 0$  and  $p > 0$ , one has*

$$\mathbf{E} \|u_0\|_{\gamma_0}^p \leq C\varepsilon^{-\kappa} \quad \text{and} \quad \mathbf{E} \|P_s u_0\|_{\gamma_0}^p \leq C\varepsilon^{p-\kappa}, \quad (6.12)$$

where  $C$  depends on  $\kappa$  and  $p$ , but not on  $\varepsilon$ .

Then, we have the following bound on the solution  $u$  of (5.2) with initial condition  $u_0$ :

$$\mathbf{E} \sup_{T \in [0, T_0]} \|u(T)\|_{\gamma_0}^p \leq C\varepsilon^{-\kappa} \quad \text{and} \quad \mathbf{E} \sup_{T \in [0, T_0]} \|P_s u(T)\|_{\gamma_0}^p \leq C\varepsilon^{p-\kappa},$$

where the constant depends on  $T_0$ ,  $\kappa$ , and  $p$ .

Moreover, every invariant measure for (5.2) is admissible, i.e.  $u_0$  distributed according to an invariant measure fulfils (6.12).

*Proof.* The proof is straightforward, we just give a brief sketch of the main arguments. Any mild solution of (5.2) fulfils

$$u(T) = e^{TL\varepsilon^{-2}} u_0 + \int_0^T e^{(T-\tau)L\varepsilon^{-2}} (Au(\tau) + \mathcal{F}(u(\tau))) d\tau + W_{\varepsilon^{-2}L}(T). \quad (6.13)$$

Note that  $W$  is, as usual, not the original  $W$  of Assumption 3.4, but a rescaled version with the same distribution.

First of all, using *e.g.* the factorisation method (cf. [DPZ92]), it is well-known that we can bound the stochastic convolution as follows, using the assumption on the covariance  $Q$  of Assumption 3.4.

$$\mathbf{E} \sup_{T \in [0, T_0]} \|W_{\varepsilon^{-2}L}(T)\|_{\gamma_0}^p \leq C\varepsilon^{-\kappa} \quad \text{and} \quad \mathbf{E} \sup_{T \in [0, T_0]} \|P_s W_{\varepsilon^{-2}L}(T)\|_{\gamma_0}^p \leq C\varepsilon^{p-\kappa}, \quad (6.14)$$

with constants  $C$  depending on  $p > 0$ ,  $T_0 > 0$  and  $\kappa > 0$ . Obviously (6.14) holds also for any  $\gamma \in [0, \gamma_0]$  instead of  $\gamma_0$ . Using this for  $\gamma = 0$ , we can easily improve Theorem 4.1 slightly to obtain

$$\mathbf{E} \sup_{T \in [0, T_0]} \|u(T)\|^p \leq C\varepsilon^{-\kappa} \quad \text{and} \quad \mathbf{E} \sup_{T \in [0, T_0]} \|P_s u(T)\|^p \leq C\varepsilon^{p-\kappa}, \quad (6.15)$$

with constants  $C$  depending on  $p > 0$ ,  $T_0 > 0$ , and  $\kappa > 0$ .

From (6.13), using Assumption 6.3 to bound  $A$  and  $\mathcal{F}$  together with 3.2, we easily derive for  $\gamma \in [0, 1 - \alpha)$ ,  $\gamma \leq \gamma_0$ ,

$$\begin{aligned} \|u(T)\|_\gamma &\leq \|u_0\|_\gamma + \|W_{\varepsilon^{-2}L}(T)\|_\gamma \\ &\quad + C \int_0^T [1 + (\varepsilon^2/(T - \tau))^{\alpha+\gamma}] \cdot (\|u(\tau)\| + \|u(\tau)\|^3) d\tau. \end{aligned}$$

Hence, using (6.14) and (6.15), we derive

$$\mathbf{E} \sup_{T \in [0, T_0]} \|u(T)\|_\gamma^p \leq \mathbf{E} \sup_{T \in [0, T_0]} \|u_0\|_\gamma^p + C\varepsilon^{-\kappa} \leq C\varepsilon^{-\kappa}.$$

The estimate for  $\|P_s u(T)\|_\gamma$  is completely analogous using (3.3). The proof for  $\gamma = \gamma_0$  (in case  $1 - \alpha < \gamma_0$ ) follows by iterating the previous argument.

Showing that the invariant measure is admissible relies on the same iteration technique. One starts from Theorem 4.1 and then gets successive bounds by following similar steps as above, removing the supremum and using the bounds obtained from the previous iteration in the right hand side.  $\square$

#### 6.4 Proof of the main estimate

**Theorem 6.9** *Assume all assumptions of section 3 together with Assumption 6.3 hold, and let  $\nu_\star$  and  $\mu_\star$  be the invariant measures introduced in Theorem 5.2.*

*Then, for every  $\kappa > 0$ , there exists a constant  $C > 0$  such that*

$$\|\nu_\star - \mu_\star^c \otimes \mu_\star^s\|_{\text{TV}} \leq C\varepsilon^{1-\kappa}.$$

*Proof.* Define  $\mathcal{Q}_T$ ,  $\mathcal{P}_T$ ,  $\hat{\mathcal{P}}_T$ , and  $\tilde{\mathcal{P}}_T$  as in section 5.2. Recall that the transition probabilities  $\tilde{\mathcal{P}}_T(u, \cdot)$  and  $\hat{\mathcal{P}}_T(u, \cdot)$  are Gaussian with means  $P_c u$  and  $e^{T\varepsilon^{-2}L}u$ , respectively, and covariance  $\int_0^T e^{r\varepsilon^{-2}L}Q^2e^{r\varepsilon^{-2}L^*}dr$ . Recall furthermore that obviously  $P_s^* \hat{\mathcal{P}}_T = \mathcal{P}_T^s$ .

Similarly to the proof of Theorem 6.1, we start by showing that the estimate

$$\|\hat{\mathcal{P}}_T \mu - \mathcal{Q}_T \mu\|_{\text{TV}} \leq C\varepsilon^{-\kappa/2} \sqrt{T} + C\varepsilon \quad (6.16)$$

holds for every  $T \in (0, 1)$  and for every measure  $\mu$  with the property that

$$\int_X \|u_c\|_\gamma^p \mu(du) < C_p \quad \text{and} \quad \int_X \|u_s\|_\gamma^p \mu(du) < C_p \varepsilon^p \quad (6.17)$$

for every  $\varepsilon \in (0, 1)$ . (The constant  $C$  does of course depend on the family of constants  $C_p$ .)



To prove (6.16) we can follow the ideas of the proof of Theorem 6.1. First, by analogy with (6.4), we define the cut-off  $\chi : X^{\gamma_0-\alpha} \rightarrow X^{\gamma_0-\alpha}$  for  $\gamma_0$  and  $\alpha$  as in Assumption 6.3 by

$$\chi(u) = u \bar{\chi}(\|u\|_{\gamma_0-\alpha}), \quad \text{and} \quad \chi_\delta(u) = \delta \chi(u/\delta).$$

The semigroup  $\mathcal{Q}_T^\delta$  is then defined by solving

$$du = \varepsilon^{-2} Lu dT + (I - \chi_\delta)(Au + \mathcal{F}(u)) dT + Q dW(T). \quad (6.18)$$

Again, there exists a constant  $C_\star$  such that, as long as  $\|u\|_{\gamma_0}(1 + \|u\|_{\gamma_0}^2) \leq C_\star \delta$ , the solutions of (6.18) and of (5.2) coincide. The estimate (6.16) follows by retracing step by step the proof of Theorem 6.1, using Lemma 6.8 to get the required *a priori* bounds on  $\|u\|_\gamma$ . This ensures that  $Q^{-1}\mathcal{F}$  and  $Q^{-1}A$  stay bounded in  $X$ , so one can apply Lemma 6.2. This finishes the proof of (6.16).

Notice furthermore that an immediate consequence of Lemma 6.5 is that

$$\|\hat{\mathcal{P}}_T(u, \cdot) - \tilde{\mathcal{P}}_T(u, \cdot)\|_{\text{TV}} \leq C\varepsilon^{-1} e^{-\omega\varepsilon^{-2}T} \|P_s u\|, \quad (6.19)$$

and thus for  $T_\varepsilon = \varepsilon^{2-\kappa}$

$$\begin{aligned} \|\nu_\star - \tilde{\mathcal{P}}_{T_\varepsilon} \mu_\star^c\|_{\text{TV}} &\leq \|\mathcal{Q}_{T_\varepsilon} \nu_\star - \hat{\mathcal{P}}_{T_\varepsilon} \nu_\star\|_{\text{TV}} + \|\hat{\mathcal{P}}_{T_\varepsilon} \nu_\star - \tilde{\mathcal{P}}_{T_\varepsilon} \nu_\star\|_{\text{TV}} \\ &\quad + \|\tilde{\mathcal{P}}_{T_\varepsilon} \nu_\star^c - \tilde{\mathcal{P}}_{T_\varepsilon} \mu_\star^c\|_{\text{TV}} \leq C\varepsilon, \end{aligned}$$

where, we used (6.16) to bound the first term, (6.19) to bound the second term, and Theorem 6.1 to bound the last term.

In order to conclude the proof of Theorem 6.9 it remains to show that

$$\|\tilde{\mathcal{P}}_{T_\varepsilon} \mu_\star^c - \mu_\star^c \otimes \mu_\star^s\|_{\text{TV}} \leq C\varepsilon^{1-\kappa}.$$

Hence, for every Borel measurable test function  $\varphi : X \rightarrow \mathbf{R}$  bounded by 1, we have to verify the estimate

$$\left| \int_{\mathcal{N}} \int_X \varphi(v) \tilde{\mathcal{P}}_{T_\varepsilon}(u_c, dv) \mu_\star^c(du_c) - \int_{\mathcal{N}} \int_S \varphi(u_c, u_s) \mu_\star^s(du_s) \mu_\star^c(du_c) \right| \leq C\varepsilon^{1-\kappa}. \quad (6.20)$$

Here and until the end of the proof, we fix  $T_\varepsilon = \varepsilon^{2-\kappa}$ .

We now apply Lemma 6.6 to the centred Gaussian measure  $\tilde{\mathcal{P}}_T(\cdot) = \tilde{\mathcal{P}}_T(0, \cdot)$  with  $\mathcal{H}_x = \mathcal{N}$  and  $\mathcal{H}_y = \mathcal{S}$ . We denote its marginal on  $\mathcal{S}$  by  $P_s^* \tilde{\mathcal{P}}_T(\cdot)$ , we denote by  $\mathcal{A}_T : \mathcal{S} \rightarrow \mathcal{N}$  the corresponding  $P_s^* \tilde{\mathcal{P}}_T$ -measurable linear map, and by  $\tilde{\mathcal{P}}_T^\circ(\cdot)$  the measure on  $\mathcal{N}$  associated to it. With these notations, we have

$$\begin{aligned} &\int_{\mathcal{N}} \int_X \varphi(v) \tilde{\mathcal{P}}_T(u_c, dv) \mu_\star^c(du_c) \\ &= \int_{\mathcal{N}} \int_S \int_{\mathcal{N}} \varphi(v_c + \mathcal{A}_T v_s + u_c, v_s) \tilde{\mathcal{P}}_T^\circ(dv_c) P_s^* \tilde{\mathcal{P}}_T(dv_s) \mu_\star^c(du_c) \\ &= \int_S \int_{\mathcal{N}} \int_{\mathcal{N}} \varphi(u_c, v_s) \tau_{v_c + \mathcal{A}_T v_s}^* \mu_\star^c(du_c) \tilde{\mathcal{P}}_T^\circ(dv_c) P_s^* \tilde{\mathcal{P}}_T(dv_s). \end{aligned}$$

From the proof of Theorem 5.3, we know that  $\mu_\star^c$  has a density with respect to the Lebesgue measure that satisfies the assumptions of Lemma 6.7, which immediately implies that there exists a constant  $C$  such that

$$\begin{aligned} & \left\| \int_{\mathcal{N}} \tilde{\mathcal{P}}_{T_\varepsilon}(u_c, \cdot) \mu_\star^c(du_c) - \mu_\star^c \otimes P_s^* \tilde{\mathcal{P}}_{T_\varepsilon} \right\|_{\text{TV}} \\ & \leq C \int_{\mathcal{S}} \int_{\mathcal{N}} \|v_c + \mathcal{A}_{T_\varepsilon} v_s\| \tilde{\mathcal{P}}_{T_\varepsilon}^\circ(dv_c) P_s^* \tilde{\mathcal{P}}_{T_\varepsilon}(dv_s) \\ & \leq 2C \sqrt{\int_{\mathcal{N}} \|v_c\|^2 P_c^* \tilde{\mathcal{P}}_{T_\varepsilon}(dv_c)} \leq C\varepsilon^{1-\kappa/2}. \end{aligned}$$

Here, we used Lemma 6.6 to go from the second to the third line, and we used the fact that the covariance of  $P_c^* \tilde{\mathcal{P}}_{T_\varepsilon}$  is  $T_\varepsilon P_c Q^2 P_c$  for the last estimate.

Using  $\tilde{\mathcal{P}}_{T_\varepsilon}(0, \cdot) = \tilde{\mathcal{P}}_{T_\varepsilon}^\circ(0, \cdot)$  together with Lemma 6.5, we get immediately

$$\|P_s^* \tilde{\mathcal{P}}_{T_\varepsilon} - \mu_\star^s\|_{\text{TV}} \leq C\varepsilon^{-1} e^{-\omega\varepsilon^{-\kappa}} \leq C\varepsilon.$$

This concludes the proof of (6.20) and thus of Theorem 6.9.  $\square$

## Appendix A Decay Properties of the Invariant Measure

In this appendix we show that the invariant measure for a non-degenerate diffusion with polynomial coefficients is smooth and all of its derivatives decay exponentially. Throughout this appendix,  $F$  denotes a fixed polynomial from  $\mathbf{R}^n$  to  $\mathbf{R}^n$  of degree  $2p - 1$  ( $p \geq 1$  is an integer which is equal to 2 in the application we have in mind) and such that there exist strictly positive constants  $C$  and  $\delta$  such that

$$\langle x, F(x) \rangle \leq C\|x\|^2 - \delta\|x\|^{2p}, \quad \forall x \in \mathbf{R}^n. \quad (\text{A.1})$$

We will consider  $x$  to be the solution of the non-degenerate SDE

$$dx(t) = F(x) dt + dW(t), \quad (\text{A.2})$$

where  $W$  consists of  $n$  independent standard Wiener processes. It is easy to show that (A.2) possesses a unique invariant measure with a smooth density  $\varrho_\star$  with respect to the Lebesgue measure on  $\mathbf{R}^n$ . The result on the decay properties of  $\varrho_\star$  is the following.

**Theorem A.1** *Let  $\varrho_\star$  be as above. Then, there exists a constant  $\varepsilon > 0$  such that  $\varrho_\star$  can be written as*

$$\varrho_\star(x) = e^{-\varepsilon\|x\|^{2p}} g_\star(x), \quad (\text{A.3})$$

where the function  $g_\star$  belongs to the Schwartz space  $\mathcal{S}$ .

*Proof.* Denote by  $\varphi_t$  the (random) flow generated by the solutions to (A.2) and by  $\mathcal{P}_t$  the semigroup defined on finite measures by

$$(\mathcal{P}_t \mu)(A) = \mathbf{E}(\mu \circ \varphi_t^{-1})(A).$$

Since  $\mathcal{P}_t$  maps every measure into a measure with a smooth density, it can be restricted to a positivity preserving contraction semigroup on  $L^1(\mathbf{R}^n, dx)$ . The generator  $\tilde{K}$  of  $\mathcal{P}_t$  is given by

$$(\tilde{K}\varrho)(x) = \frac{1}{2}\Delta\varrho(x) - F(x)\nabla\varrho(x) - (\operatorname{div} F)(x)\varrho(x).$$

We now define an operator  $K$  on  $L^2(\mathbf{R}^n, dx)$  by closing the operator defined on  $C_0^\infty$  by

$$K = e^{\varepsilon\|x\|^{2p}} \tilde{K} e^{-\varepsilon\|x\|^{2p}}. \quad (\text{A.4})$$

An easy computation shows that, if  $\varepsilon$  is small enough, one has

$$K + K^T = \Delta + G(x), \quad (\text{A.5})$$

$K^T$  is the formal adjoint (in  $L^2$ ) of  $K$  and  $G : \mathbf{R}^n \rightarrow \mathbf{R}$  is a polynomial of degree  $4p - 2$  satisfying

$$G(x) \leq C - \varepsilon^2\|x\|^{4p-2},$$

for some positive constant  $C$ . A standard argument (see *e.g.* the proof of Prop. B.3 in [EH00] or the proof of Prop. 5.5 in [HN03]) shows that  $K$  is maximal accretive and is therefore the generator of a quasi-bounded  $C_0$ -semigroup on  $L^2(\mathbf{R}^n, dx)$ . Furthermore, (A.5) implies that  $K$  has compact resolvent. Since  $e^{-\varepsilon\|x\|^{2p}}$  is an eigenfunction of  $K^*$  with eigenvalue 0,  $K$  must also have an eigenvalue 0. Denote by  $g_\star$  the corresponding eigenfunction. By the definition of  $K$ , the function  $\varrho_\star$  given by  $\varrho_\star(x) = e^{-\varepsilon\|x\|^{2p}} g_\star(x)$  must be an eigenfunction with eigenvalue 0 of  $\tilde{K}$ . An argument given for example in [EPR99] shows that  $\varrho_\star$  must be positive, and therefore it is the invariant measure of (A.2). It thus remains to show that  $g_\star$  belongs to the Schwartz space.

For this, we use some of the results of [EH03]. Following the notation of that paper, we introduce a family of weighted Sobolev spaces  $S^{\alpha,\beta}$  with  $\alpha, \beta \in \mathbf{R}$  as the following subset of tempered distributions  $\mathcal{S}'$  on  $\mathbf{R}^n$ :

$$S^{\alpha,\beta} = \{u \in \mathcal{S}' \mid \Lambda^\alpha \bar{\Lambda}^\beta u \in L^2(\mathbf{R}^n)\},$$

where the operator  $\Lambda$  is defined as the positive square root of  $\Lambda^2 = 1 - \sum_{i=1}^n \partial_i^2 = 1 - \Delta$ , and  $\bar{\Lambda}$  is the multiplication operator  $\bar{\Lambda}^2 = 1 + \|x\|^2$ .

We equip the space  $S^{\alpha,\beta}$  with the scalar product

$$\langle f, g \rangle_{\alpha,\beta} = \langle \Lambda^\alpha \bar{\Lambda}^\beta f, \Lambda^\alpha \bar{\Lambda}^\beta g \rangle_{L^2}. \quad (\text{A.6})$$

We also use the corresponding norms  $\|\cdot\|_{\alpha,\beta}$ . Note that these spaces are actually a particular case of the more general class of Sobolev spaces introduced in [BC94].

Retracing the proof of [EH03, Lem. 3.2] we see that for every pair  $\alpha, \beta$ , there exists a constant  $C$  such that

$$|\langle f, Kg \rangle_{\alpha,\beta} - \langle K^T f, g \rangle_{\alpha,\beta}| \leq C \|f\|_{\alpha+\frac{1}{2}, \beta+2p-\frac{1}{2}} \|g\|_{\alpha+\frac{1}{2}, \beta+2p-\frac{1}{2}}$$

holds for every  $f$  and  $g$  in  $\mathcal{S}$ . Combining this with (A.5) shows that, for every pair  $\alpha, \beta$ , there exists a constant  $C$  such that

$$\begin{aligned} \|u\|_{\alpha+1, \beta+2p-1} &\leq C(\langle u, (\Delta + G)u \rangle_{\alpha, \beta} + \|u\|_{\alpha+\frac{1}{2}, \beta+2p-\frac{3}{2}}) \\ &= C(\langle u, (K + K^T)u \rangle_{\alpha, \beta} + \|u\|_{\alpha+\frac{1}{2}, \beta+2p-\frac{3}{2}}) \quad (\text{A.7}) \\ &\leq C(\|Ku\|_{\alpha, \beta} + \|u\|_{\alpha+\frac{1}{2}, \beta+2p-\frac{3}{2}}) \end{aligned}$$

holds for every  $u \in \mathcal{S}$ .

Taking formally  $u = g_*$  in (A.7) and using the fact that  $Kg_* = 0$ , we get

$$\|g_*\|_{\alpha, \beta} \leq C\|g_*\|_{\alpha-\frac{1}{2}, \beta-\frac{1}{2}}.$$

Since the  $L^2$ -norm of  $g_*$  finite, this shows that  $g_*$  belongs to the intersection of all the spaces  $S^{\alpha, \beta}$  and thus to  $\mathcal{S}$ . Since we didn't know *a priori* that  $g_*$  belongs to  $\mathcal{S}$  (this is what we want to prove after all!), this argument appears not to be very rigorous. However, it can easily be made rigorous by approximating  $g_*$  by a sequence of functions in  $\mathcal{S}$  and estimating the errors carefully. This concludes the proof of Theorem A.1.  $\square$

### Acknowledgements

The work of D.B. was supported by DFG-Forschungsstipendium BL535/5-1. The work of M.H. was supported by the Fonds National Suisse. Both authors would like to thank the MRC at the University of Warwick and especially David Elworthy for their warm hospitality.

### References

- [Arn98] L. ARNOLD. *Random Dynamical Systems*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 1998.
- [BC94] J.-M. BONY and J.-Y. CHEMIN. Espaces fonctionnels associés au calcul de Weyl-Hörmander. *Bull. Soc. Math. France* **122**, no. 1, (1994), 77–118.
- [Blö03a] D. BLÖMKER. Amplitude equations for locally cubic non-autonomous nonlinearities. *SIAM Journal on Applied Dynamical Systems* To appear.
- [Blö03b] D. BLÖMKER. Approximation of the stochastic Rayleigh-Bénard problem near the onset of instability and related problems, 2003. Preprint.
- [BMPS01] D. BLÖMKER, S. MAIER-PAAPE, and G. SCHNEIDER. The stochastic Landau equation as an amplitude equation. *Discrete and Continuous Dynamical Systems, Series B* **1**, no. 4, (2001), 527–541.
- [CCLR] T. CARABALLO, H. CRAUEL, J. A. LANGA, and J. C. ROBINSON. Stabilization by additive noise. In preparation.
- [CF98] H. CRAUEL and F. FLANDOLI. Additive noise destroys a pitchfork bifurcation. *J. Dynam. Differential Equations* **10**, no. 2, (1998), 259–274.

- [CIS99] H. CRAUEL, P. IMKELLER, and M. STEINKAMP. Bifurcations of one-dimensional stochastic differential equations. In *Stochastic dynamics (Bremen, 1997)*, 145–154. Springer, New York, 1999.
- [DPZ92] G. DA PRATO and J. ZABCZYK. *Stochastic Equations in Infinite Dimensions*. University Press, Cambridge, 1992.
- [DPZ96] G. DA PRATO and J. ZABCZYK. *Ergodicity for Infinite Dimensional Systems*, vol. 229 of *London Mathematical Society Lecture Note Series*. University Press, Cambridge, 1996.
- [EH00] J.-P. ECKMANN and M. HAIRER. Non-equilibrium statistical mechanics of strongly anharmonic chains of oscillators. *Commun. Math. Phys.* **212**, no. 1, (2000), 105–164.
- [EH01] J.-P. ECKMANN and M. HAIRER. Uniqueness of the invariant measure for a stochastic PDE driven by degenerate noise. *Commun. Math. Phys.* **219**, no. 3, (2001), 523–565.
- [EH03] J.-P. ECKMANN and M. HAIRER. Spectral properties of hypoelliptic operators. *Comm. Math. Phys.* **235**, no. 2, (2003), 233–253.
- [EL94] K. D. ELWORTHY and X.-M. LI. Formulae for the derivatives of heat semigroups. *J. Funct. Anal.* **125**, no. 1, (1994), 252–286.
- [EPR99] J.-P. ECKMANN, C.-A. PILLET, and L. REY-BELLET. Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures. *Commun. Math. Phys.* **201**, (1999), 657–697.
- [Hai02] M. HAIRER. Exponential mixing properties of stochastic PDEs through asymptotic coupling. *Probab. Theory Related Fields* **124**, no. 3, (2002), 345–380.
- [HN03] B. HELFFER and F. NIER. Hypoellipticity and spectral theory for Fokker-Planck operators and Witten Laplacians, 2003. Prépublication 03-25 de l’IRMAR Université de Rennes.
- [HS92] P. HOHENBERG and J. SWIFT. Effects of additive noise at the onset of Rayleigh-Bénard convection. *Physical Review A* **46**, (1992), 4773–4785.
- [Kat80] T. KATO. *Perturbation Theory for Linear Operators*. Springer, New York, 1980.
- [KS00] S. B. KUKSIN and A. SHIRIKYAN. Stochastic dissipative PDE’s and Gibbs measures. *Commun. Math. Phys.* **213**, (2000), 291–330.
- [KSM92] P. KIRRMANN, G. SCHNEIDER, and A. MIELKE. The validity of modulation equations for extended systems with cubic nonlinearities. *Proc. Roy. Soc. Edinburgh Sect. A* **122**, no. 1-2, (1992), 85–91.
- [LDS91] Z. LAI and S. DAS SARMA. Kinetic growth with surface relaxation: Continuum versus atomistic models. *Phys. Rev. Lett.* **66**, no. 18, (1991), 2348–2351.
- [Le693] J. A. LEÓN. Fubini theorem for anticipating stochastic integrals in Hilbert space. *Appl. Math. Optim.* **27**, no. 3, (1993), 313–327.
- [Lun95] A. LUNARDI. *Analytic Semigroups and Optimal Regularity in Parabolic Problems*. Birkhäuser, Basel, 1995.

- [Mal97] P. MALLIAVIN. *Stochastic Analysis*, vol. 313 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1997.
- [Mat02] J. C. MATTINGLY. Exponential convergence for the stochastically forced Navier-Stokes equations and other partially dissipative dynamics. *Commun. Math. Phys.* **230**, no. 3, (2002), 421–462.
- [MT94] S. P. MEYN and R. L. TWEEDIE. *Markov Chains and Stochastic Stability*. Springer, New York, 1994.
- [Nua95] D. NUALART. *The Malliavin Calculus and Related Topics*. Probability and its Applications. Springer-Verlag, New York, 1995.
- [Paz83] A. PAZY. *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Springer, New York, 1983.
- [RR00] D. RAMACHANDRAN and L. RÜSCHENDORF. On the Monge-Kantorovich duality theorem. *Teor. Veroyatnost. i Primenen.* **45**, no. 2, (2000), 403–409.
- [RY99] D. REVUZ and M. YOR. *Continuous Martingales and Brownian Motion*, vol. 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third ed., 1999.
- [Sch96] G. SCHNEIDER. The validity of generalized Ginzburg-Landau equations. *Math. Methods Appl. Sci.* **19**, no. 9, (1996), 717–736.