DIMENSIONAL REDUCTION IN NONLINEAR FILTERING: A HOMOGENIZATION APPROACH

PETER IMKELLER[1], N. SRI NAMACHCHIVAYA[2], NICOLAS PERKOWSKI[1], AND HOONG C. YEONG[2]

Abstract. We propose a homogenized filter for multiscale signals, which allows to reduce the dimension of the system. We prove that the nonlinear filter converges to our homogenized filter with rate $\sqrt{\varepsilon}$. This is achieved by a suitable asymptotic expansion of the dual of the Zakai equation, and by probabilistically representing the correction terms with the help of BDSDEs.

1. Introduction

Filtering theory is an established field in applied probability and decision and control systems, which is important in many practical applications from inertial guidance of aircrafts and spacecrafts to weather and climate prediction. It provides a recursive algorithm for estimating a signal or state of a random dynamical system based on noisy measurements. More precisely, filtering problems consist of an unobservable signal process $X \stackrel{\text{def}}{=} \{X_t : t \geq 0\}$ and an observation process $Y \stackrel{\text{def}}{=} \{Y_t : t \geq 0\}$ that is a function of X corrupted by noise. The main objective of filtering theory is to get the best estimate of X_t based on the information $\mathcal{Y}_t \stackrel{\text{def}}{=} \sigma\{Y_s : 0 \leq s \leq t\}$. This is given by the conditional distribution π_t of X_t given $\mathcal{Y}_t$ or equivalently, the conditional expectations $\mathbb{E}[f(X_t)|\mathcal{Y}_t]$ for a rich enough class of functions. Since this estimate minimizes the mean square error loss, we call π_t the optimal filter. The goal of filtering theory is to characterize this conditional distribution effectively. In simplified problems where the signal and the observation models are linear and Gaussian, the filtering equation is finite-dimensional, and the solution is the well-known Kalman-Bucy filter. In more realistic problems, nonlinearities in the models lead to more complicated equations for π_t, defined by Zakai (1969) and Fujisaki et al. (1972), which describe the evolution of the conditional distribution in the space of probability measures (see, for example, Bain and Crisan (2009), Kallianpur (1980), Liptser and Shiryaev (2001)).

It is impractical to implement a numerical solution to such infinite dimensional stochastic evolution equations of the general nonlinear filtering problem by finite difference or finite element approximations. Therefore, extended Kalman filter algorithms, which use linear approximations to the signal dynamics and observation, have been used extensively in several applications. These provide essentially a first order approximation to an infinite dimensional problem and can perform quite poorly in problems with strong nonlinearities. Particle filters have been well established for the implementation of nonlinear filtering in science and engineering applications. Doucet et al. (2001) and Arulampalam et al. (2002) provide comprehensive insight into particle filtering. However, due to dimensionality issues (see, for example, Snyder et al. (2008)) and computational complexities that arise in representing the signal density using a high number of particles, the problem of particle filtering in high dimensions is still not completely resolved. As a result of these difficulties, we have established a novel particle filtering method Park et al. (2011) for multiscale signal and observation processes that combines the homogenization with filtering techniques. The theoretical basis for this new capability is presented in this paper.

2010 *Mathematics Subject Classification.* 60G35, 35B27, 60H15, 60H35.

Key words and phrases. nonlinear filtering; dimensional reduction; homogenization; particle filtering; asymptotic expansion; SPDE; BDSDE.

The results presented here are set within the context of slow-fast dynamical systems, where the rates of change of different variables differ by orders of magnitude. Multiple time scales occur in models throughout the science and engineering field. For example, climate evolution is governed by fast atmospheric and slow oceanic dynamics and state dynamics in electric power systems consists of fast- and slowly-varying elements. This paper addresses the effects of the multiscale signal and observation processes via the study of the Zakai equation. We construct a lower dimensional Zakai equation in a canonical way. This problem has also been studied in Park et al. (2010) using a different approach from what is presented here. In moderate dimensional problems, particle filters are an attractive alternative to numerical approximation of the stochastic partial differential equations (SPDEs) by finite difference or finite element methods. For the reduced nonlinear model an appropriate form of particle filter can be a viable and useful scheme. Hence, Lingala et al. (2012) presents the numerical solution of the lower dimensional stochastic partial differential equation derived here, as it is applied to a chaotic high-dimensional multiscale system.

In general, this paper provides rigorous mathematical results that support the numerical algorithms based on the idea that stochastically averaged models provide qualitatively useful results which are potentially helpful in developing inexpensive lower-dimensional filtering as demonstrated by Park et al. (2011) in the context of homogenized particle filters and by Harlim and Kang (2012) in the context of averaged ensemble Kalman filters. The convergence of the optimal filter to the homogenized filter is shown using backward stochastic differential equations (BSDEs) and asymptotic techniques.

Let us describe the main result. We assume the signal is given as solution of the two time scale stochastic differential equation (SDE)

$$\begin{aligned} dX_t^\varepsilon &= b(X_t^\varepsilon, Z_t^\varepsilon)dt + \sigma(X_t^\varepsilon, Z_t^\varepsilon)dV_t \\ dZ_t^\varepsilon &= \frac{1}{\varepsilon} f(X_t^\varepsilon, Z_t^\varepsilon)dt + \frac{1}{\sqrt{\varepsilon}} g(X_t^\varepsilon, Z_t^\varepsilon)dW_t. \end{aligned}$$

Here X^ε is the slow component and Z^ε is the fast component. We assume that for every fixed x, the solution Z^x of

$$dZ_t^x = f(x, Z_t^x)dt + g(x, Z_t^x)dW_t$$

is ergodic and converges rapidly to its unique stationary distribution. In this case it is well known that X^ε converges in distribution to a diffusion X^0 which is governed by an SDE

$$dX_t^0 = \bar{b}(X_t^0)dt + \bar{\sigma}(X_t^0)dV_t.$$

This X^0 is used to construct an averaged filter π^0. We denote the optimal filter for the full system by π^ε. Define the x-marginal of π^ε as $\pi^{\varepsilon,x}$, i.e.

$$\int \varphi(x)\pi_t^{\varepsilon,x}(dx) = \int \varphi(x)\pi_t^\varepsilon(dx, dz).$$

Our main result is then

Theorem. *Under the assumptions stated in Theorem 3.1, for every* $p \geq 1$ *and* $T \geq 0$ *there exists* $C > 0$*, such that for every* $\varphi \in C_b^4$

$$\left(\mathbb{E}_{\mathbb{Q}}\left[\left|\pi_T^{\varepsilon,x}(\varphi) - \pi_T^0(\varphi)\right|^p\right]\right)^{1/p} \leq \sqrt{\varepsilon}C||\varphi||_{4,\infty}.$$

In particular, there exists a metric d *on the space of probability measures, such that* d *generates the topology of weak convergence, and such that for every* $T \geq 0$ *there exists* $C > 0$ *such that*

$$\mathbb{E}_{\mathbb{Q}}\left[d(\pi_T^{\varepsilon,x}, \pi_T^0)\right] \leq \sqrt{\varepsilon}C.$$

We begin in Section 2 by presenting the general formulation of the multiscale nonlinear filtering problem. Here we describe the measure-valued Zakai equation and introduce the homogenized equations that we seek to derive for the reduced dimension unnormalized filter. Section 3 presents the formal asymptotic expansion of the multi scale Zakai equation that results in several SPDEs. We also present the main results of this paper in this section. Section 4 provides the probabilistic representation of the SPDEs, that is, we describe the solutions of the infinite dimensional SPDEs by finite dimensional backward doubly stochastic differential equations (BDSDEs). We restate some of the results in this context due to Rozovskii (1990) and Pardoux and Peng (1994) at the end of this section. We present some of the preliminary results of Pardoux and Veretennikov (2003) on convergence of the transition function of Z^x in section 5. These estimates are used in the proof of the main results presented in section 6.

2. Formulation of multiscale nonlinear filtering problems

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{Q})$ be a filtered probability space that supports a $(k+l+d)$-dimensional standard Brownian motion (V, W, B). Let the signal $(X^\varepsilon, Z^\varepsilon)$ be a two time scale diffusion process with a fast component Z^ε and a slow component X^ε:

$$\begin{aligned} dX_t^\varepsilon &= b(X_t^\varepsilon, Z_t^\varepsilon)dt + \sigma(X_t^\varepsilon, Z_t^\varepsilon)dV_t \\ dZ_t^\varepsilon &= \frac{1}{\varepsilon} f(X_t^\varepsilon, Z_t^\varepsilon)dt + \frac{1}{\sqrt{\varepsilon}} g(X_t^\varepsilon, Z_t^\varepsilon)dW_t, \end{aligned} \tag{1}$$

where $X_t^\varepsilon \in \mathbb{R}^m$, $Z_t^\varepsilon \in \mathbb{R}^n$, $W_t \in \mathbb{R}^l$ and $V_t \in \mathbb{R}^k$ are independent standard Brownian motions, $b : \mathbb{R}^{m+n} \to \mathbb{R}^m$, $\sigma : \mathbb{R}^{m+n} \to \mathbb{R}^{m\times k}$, $f : \mathbb{R}^{m+n} \to \mathbb{R}^n$, $g : \mathbb{R}^{m+n} \to \mathbb{R}^{n\times l}$. All the functions above are assumed to be Borel-measurable. For fixed $x \in \mathbb{R}^m$, define

$$dZ_t^x = f(x, Z_t^x)dt + g(x, Z_t^x)dW_t. \tag{2}$$

Assume that for all $x \in \mathbb{R}^m$, Z^x is ergodic and converges rapidly towards its stationary measure $\mu(x, \cdot)$. We will make this precise later.

The d-dimensional observation Y^ε is given by

$$Y_t^\varepsilon = \int_0^t h(X_s^\varepsilon, Z_s^\varepsilon)ds + B_t$$

with Borel-measurable $h : \mathbb{R}^{m+n} \to \mathbb{R}^d$. B is assumed to be a d-dimensional standard Brownian motion that is independent of W and V.

Define $\mathcal{Y}_t^\varepsilon = \sigma(Y_s^\varepsilon : 0 \le s \le t) \vee \mathcal{N}$, where $\mathcal{N}$ are the $\mathbb{Q}$-negligible sets. For a finite measure π on $\mathbb{R}^{m+n}$ and for a bounded measurable function φ on $\mathbb{R}^{m+n}$ denote $\pi(\varphi) = \int \varphi(x,z)\pi(dx,dz)$. Then our aim is to calculate the measure-valued process $(\pi_t^\varepsilon, t \ge 0)$ determined by

$$\pi_t^\varepsilon(\varphi) = \mathbb{E}[\varphi(X_t^\varepsilon, Z_t^\varepsilon)|\mathcal{Y}_t^\varepsilon].$$

Define the Girsanov transform

$$\left.\frac{d\mathbb{P}^\varepsilon}{d\mathbb{Q}}\right|_{\mathcal{F}_t} = D_t^\varepsilon = \exp\left(-\int_0^t h(X_s^\varepsilon, Z_s^\varepsilon)^* dB_s - \frac{1}{2}\int_0^t |h(X_s^\varepsilon, Z_s^\varepsilon)|^2 ds\right).$$

Under $\mathbb{P}^\varepsilon$, the observation process, Y^ε, is a Brownian motion and independent of $(X^\varepsilon, Z^\varepsilon)$. By the Kallianpur-Striebel formula,

$$\mathbb{E}_{\mathbb{Q}}[\varphi(X_t^\varepsilon, Z_t^\varepsilon)|\mathcal{Y}_t^\varepsilon] = \frac{\mathbb{E}_{\mathbb{P}^\varepsilon}\left[\varphi(X_t^\varepsilon, Z_t^\varepsilon)\left.\frac{d\mathbb{Q}}{d\mathbb{P}^\varepsilon}\right|_{\mathcal{F}_t}\middle|\mathcal{Y}_t^\varepsilon\right]}{\mathbb{E}_{\mathbb{P}^\varepsilon}\left[\left.\frac{d\mathbb{Q}}{d\mathbb{P}^\varepsilon}\right|_{\mathcal{F}_t}\middle|\mathcal{Y}_t^\varepsilon\right]}$$

with

$$\left.\frac{d\mathbb{Q}}{d\mathbb{P}^\varepsilon}\right|_{\mathcal{F}_t} = \tilde{D}_t^\varepsilon = \exp\left(\int_0^t h(X_s^\varepsilon, Z_s^\varepsilon)^* dY_s^\varepsilon - \frac{1}{2}\int_0^t |h(X_s^\varepsilon, Z_s^\varepsilon)|^2 ds\right).$$

So if we define

$$\rho_t^\varepsilon(\varphi) = \mathbb{E}_{\mathbb{P}^\varepsilon}\left[\varphi(X_t^\varepsilon, Z_t^\varepsilon)\exp\left(\int_0^t h(X_s^\varepsilon, Z_s^\varepsilon)^* dY_s^\varepsilon - \frac{1}{2}\int_0^t |h(X_s^\varepsilon, Z_s^\varepsilon)|^2 ds\right)\middle| \mathcal{Y}_t^\varepsilon\right],$$

then

$$\pi_t^\varepsilon(\varphi) = \frac{\rho_t^\varepsilon(\varphi)}{\rho_t^\varepsilon(1)}.$$

Denote by $\mathcal{L}^\varepsilon = \frac{1}{\varepsilon}\mathcal{L}_F + \mathcal{L}_S$ the differential operator associated to $(X^\varepsilon, Z^\varepsilon)$. That is,

$$\mathcal{L}_F = \sum_{i=1}^n f_i(x,z)\frac{\partial}{\partial z_i} + \frac{1}{2}\sum_{i,j=1}^n (gg^*)_{ij}(x,z)\frac{\partial^2}{\partial z_i \partial z_j}$$
$$\mathcal{L}_S = \sum_{i=1}^m b_i(x,z)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^m (\sigma\sigma^*)_{ij}(x,z)\frac{\partial^2}{\partial x_i \partial x_j}$$

where $\cdot^*$ denotes the transpose of a matrix or a vector.

Then the unnormalized measure-valued process, ρ^ε, satisfies the Zakai equation:

$$\begin{aligned} d\rho_t^\varepsilon(\varphi) &= \rho_t^\varepsilon(\mathcal{L}^\varepsilon\varphi)dt + \rho_t^\varepsilon(h\varphi)dY_t^\varepsilon \\ \rho_0^\varepsilon(\varphi) &= \mathbb{E}_{\mathbb{Q}}[\varphi(X_0^\varepsilon, Z_0^\varepsilon)] \end{aligned} \tag{3}$$

for every $\varphi \in C_b^2(\mathbb{R}^{m+n}, \mathbb{R})$ (see, for example, Bain and Crisan (2009)). For $k \geq 0$, C_b^k is the space of k times continuously differentiable functions f, such that f and all its partial derivatives up to order k are bounded.

The theory of stochastic averaging (see, for example, Papanicolaou et al. (1977)) tells us that under suitable conditions, X^ε converges in law to X^0 as $\varepsilon \to 0$, where X^0 is the solution of an SDE

$$dX_t^0 = \bar{b}(X_t^0)dt + \bar{\sigma}(X_t^0)dW_t$$

for suitably averaged $\bar{b}$ and $\bar{\sigma}$. Denote the generator of X^0 by $\bar{\mathcal{L}}$.

We want to show that as long as we are only interested in *estimating the slow component*, we can take advantage of this fact. More precisely, we want to find a homogenized (unnnormalized) filter ρ^0, such that for small ε, $\rho^{\varepsilon,x}$ which is the x-marginal of ρ_t^ε, is close to ρ^0. The x-marginal of ρ_t^ε is defined as

$$\rho_t^{\varepsilon,x}(\varphi) = \int_{\mathbb{R}^{m+n}} \varphi(x)\rho_t^\varepsilon(dx, dz)$$

for every measurable bounded $\varphi : \mathbb{R}^m \to \mathbb{R}$, and ρ^0 is the solution of

$$\begin{aligned} d\rho_t^0(\varphi) &= \rho_t^0(\bar{\mathcal{L}}\varphi)dt + \rho_t^0(\bar{h}\varphi)dY_t^\varepsilon \\ \rho_0^0(\varphi) &= \mathbb{E}_{\mathbb{Q}}[\varphi(X_0^0)], \end{aligned} \tag{4}$$

where $\bar{h}$ is a suitably averaged version of h. The measure-valued processes π^0 and $\pi^{\varepsilon,x}$ are then defined in terms of ρ^0 and $\rho^{\varepsilon,x}$ as π^ε was defined in terms of ρ^ε:

$$\pi_t^0(\varphi) = \frac{\rho_t^0(\varphi)}{\rho_t^0(1)} \qquad \text{and} \qquad \pi_t^{\varepsilon,x}(\varphi) = \frac{\rho_t^{\varepsilon,x}(\varphi)}{\rho_t^{\varepsilon,x}(\varphi)}.$$

Note that the homogenized filter is still driven by the real observation Y^ε and not by a "homogenized observation", which is practical for implementation of the homogenized filter in applications since such homogenized observation is usually not available. However, should such homogenized observation be available, using it would lead to loss of information for estimating the signal compared to using the actual observation.

In this paper, we will prove L^1-convergence of the actual filter to the homogenized filter, i.e. we will show that for any $T > 0$,

$$\lim_{\varepsilon\to 0} \mathbb{E}\left[d(\pi_T^{\varepsilon,x}, \pi_T^0)\right] = 0,$$

where d denotes a suitable distance on the space of probability measures that generates the topology of weak convergence. This convergence result is shown in Park et al. (2010) for a two-dimensional multiscale signal process with no drift in the fast component SDE. Here, we extend the result to an $\mathbb{R}^{m+n}$-dimensional signal process with drift and diffusion coefficients of the fast and slow components dependent on both components. The proof of Park et al. (2010) is based on representing the slow component as a time-changed Brownian motion under a suitable measure, which cannot be extended easily to the multidimensional setting we assume here.

Based on (3) and (4), the filter convergence problem is a problem of homogenization of a SPDE. In Papanicolaou et al. (1977), homogenization of diffusion processes with periodic structures is done using the martingale problem approach. In Papanicolaou and Kohler (1975) and Chapter 2 of Bensoussan et al. (1978), limit behavior of stochastic processes is studied using asymptotic analysis. Bensoussan et al. (1978) studies linear SPDEs with periodic coefficients and also used a probabilistic approach in Chapter 3. Homogenization in the nonlinear filtering problem framework has been studied in Bensoussan and Blankenship (1986) and Ichihara (2004) via asymptotic analysis on a dual representation of the nonlinear filtering equation. As far as we are aware, Ichihara (2004) has used BSDEs for studying homogenization of Zakai-type SPDEs for the first time. Our convergence proof applies BSDE techniques by invoking the dual representation of the filtering equation and using asymptotic analysis to determine the limit behavior of the solution of the backward equation. Pardoux and Veretennikov (2003) give precise estimates for the transition function of an ergodic SDE of the type (2), and these results are used in our proof. To our knowledge, such method of homogenization for SPDEs combining BSDE and asymptotic methods has not been done before.

To our knowledge, a result presented in Chapter 6 of Kushner (1990) is the closest to the results presented in this paper. In Theorem 6.3.1 of Kushner (1990) it is shown that for a fixed test function, the difference of the unnormalized actual and homogenized filters for multiscale jump-diffusion processes converges to zero in distribution. Standard results then give convergence in probability of the fixed time marginals. Kushner (1990)'s method of proof is by averaging the coefficients of the SDEs for the unnormalized filters and showing that the limits of both filters satisfy the same SDE that possesses a unique solution. We obtain L^p convergence of the measure valued process, not just for fixed test functions, and we are able to quantify the rate of convergence, which, to the best of our knowledge, has not been achieved before in homogenization of nonlinear filters..

In Kleptsina et al. (1997), convergence of the nonlinear filter is shown in a very general setting, based on convergence in total variation distance of the law of $(X^\varepsilon, Y^\varepsilon)$. This is then applied to two examples. Since the diffusion matrix of our slow component is allowed to depend on the fast component, our results are not a special case. In the examples of Kleptsina et al. (1997), X^ε converges to $\bar{X}$ in probability, which is no longer the case in our setting. However it might be possible to apply the total variation techniques developed in Kleptsina et al. (1997) to obtain convergence in our setting. Only the rate of convergence cannot be determined with these techniques.

For a given bounded test function φ and terminal time T, we follow Pardoux (1979) in introducing the associated dual process $v_t^{\varepsilon,T,\varphi}(x,z)$, which is a dynamic version of $\mathbb{E}_{\mathbb{P}^\varepsilon}[\varphi(X_T^\varepsilon)\tilde{D}_T^\varepsilon|\mathcal{Y}_T^\varepsilon]$:

$$v_t^{\varepsilon,T,\varphi}(x,z) = \mathbb{E}_{\mathbb{P}^\varepsilon_{t,x,z}}[\varphi(X_T^\varepsilon)\tilde{D}_{t,T}^\varepsilon|\mathcal{Y}_{t,T}^\varepsilon]$$

where $\mathbb{P}^\varepsilon_{t,x,z}$ is the measure under which X^ε and Z^ε are governed by the same dynamics as under $\mathbb{P}^\varepsilon$, but $(X^\varepsilon, Z^\varepsilon)$ stays in (x,z) until time t, then it starts to follow the SDE dynamics. $\tilde{D}_{t,T}^\varepsilon = \tilde{D}_T^\varepsilon(\tilde{D}_t^\varepsilon)^{-1}$; and $\mathcal{Y}_{t,T}^\varepsilon = \sigma(Y_r^\varepsilon - Y_t^\varepsilon : t \le r \le T) \vee \mathcal{N}$ (recall that $\mathcal{N}$ denotes the

$\mathbb{Q}$-negligible sets). From the Markov property of $(X^\varepsilon, Z^\varepsilon)$ it follows that for any $t \in [0,T]$: $\rho^\varepsilon_t(v^{\varepsilon,T,\varphi}_t) = \rho^{\varepsilon,x}_T(\varphi)$. In particular (because at time 0, ρ^ε is just the starting distribution of $(X^\varepsilon, Z^\varepsilon)$):

$$\rho^{\varepsilon,x}_T(\varphi) = \int v^{\varepsilon,T,\varphi}_0(x,z)\mathbb{Q}_{(X^\varepsilon_0, Z^\varepsilon_0)}(dx,dz).$$

Similarly introduce

$$v^{0,T,\varphi}_t(x) = \mathbb{E}_{\mathbb{P}^\varepsilon_{t,x}}[\varphi(X^0_T)\tilde{D}^0_{t,T}|\mathcal{Y}^\varepsilon_{t,T}],$$

where

$$\tilde{D}^0_{t,T} = \exp\left(\int_t^T \bar{h}(X^0_r)^* dY^\varepsilon_r - \frac{1}{2}\int_t^T |\bar{h}(X^0_r)|^2 dr\right)$$

and $\mathbb{P}^\varepsilon_{t,x}$ is the measure under which X^0 is governed by the same dynamics as under $\mathbb{P}^\varepsilon$, but stays in x until time t. We can also show that for any $t \in [0,T]$: $\rho^0_t(v^{0,T,\varphi}_t) = \rho^0_T(\varphi)$, so that

$$\rho^0_T(\varphi) = \int v^{0,T,\varphi}_0(x)\mathbb{Q}_{X^0_0}(dx).$$

Note that $\mathbb{Q}_{X^0_0} = \mathbb{Q}_{X^\varepsilon_0}$, because the homogenized process has the same starting distribution as the unhomogenized one.

Now fix T and $\varphi \in C^2_b(\mathbb{R}^m, \mathbb{R})$ and write $v^\varepsilon_t = v^{\varepsilon,T,\varphi}_t$ and $v^0_t = v^{0,T,\varphi}_t$.

Our aim is to show that for nice test functions φ, and for the dual processes v^ε and v^0 defined above, $\mathbb{E}[|v^\varepsilon_0(x,z) - v^0_0(x)|^p]$ is small (in a way that will depend on x and z). Then

$$\begin{aligned}\mathbb{E}[|\rho^{\varepsilon,x}_T(\varphi) - \rho^0_T(\varphi)|^p] &= \mathbb{E}\left[\left|\int (v^\varepsilon_0(x,z) - v^0_0(x))\mathbb{Q}_{(X^\varepsilon_0,Z^\varepsilon_0)}(dx,dz)\right|^p\right] \\ &\le \mathbb{E}\left[\int |v^\varepsilon_0(x,z) - v^0_0(x)|^p \mathbb{Q}_{(X^\varepsilon_0,Z^\varepsilon_0)}(dx,dz)\right] \\ &= \int \mathbb{E}[|v^\varepsilon_0(x,z) - v^0_0(x)|^p]\mathbb{Q}_{(X^\varepsilon_0,Z^\varepsilon_0)}(dx,dz)\end{aligned}$$

will also be small as long as $\mathbb{Q}_{(X^\varepsilon_0,Z^\varepsilon_0)}$ is well behaved.

3. Formal expansions of the filtering equations and the main results

Before we continue, let us change notation: For large parts of this article we will only work under $\mathbb{P}^\varepsilon$, and the process Y^ε is a Brownian motion under $\mathbb{P}^\varepsilon$ which is independent of $(X^\varepsilon, Z^\varepsilon, X^0)$. Therefore from now on we write $\mathbb{P}$ instead of $\mathbb{P}^\varepsilon$ and B instead of Y^ε to facilitate the reading. The distribution and notation for the Markov processes $(X^\varepsilon, Z^\varepsilon, X^0)$ do not change.

The key point is now that v^ε and v^0 solve backward SPDEs:

$$\begin{aligned}-dv^\varepsilon_t(x,z) &= \mathcal{L}^\varepsilon v^\varepsilon_t(x,z)dt + h(x,z)^* v^\varepsilon_t(x,z)d\overleftarrow{B}_t \\ v^\varepsilon_T(x,z) &= \varphi(x)\end{aligned} \tag{5}$$

and

$$\begin{aligned}-dv^0_t(x) &= \bar{\mathcal{L}} v^0_t(x,z)dt + \bar{h}(x)^* v^0_t(x)d\overleftarrow{B}_t \\ v^0_T(x) &= \varphi(x).\end{aligned} \tag{6}$$

Here and everywhere in this article, $d\overleftarrow{B}$ denotes Itô's backward integral.

We formally expand v^ε as

$$v^\varepsilon_t(x,z) = u^0_t(x,z) + \varepsilon u^1_{t/\varepsilon}(x,z) + \varepsilon^2 u^2_{t/\varepsilon}(x,z).$$

Note that rigorously this does not make any sense, because:

- We work with equations with terminal conditions. But when we send $\varepsilon \to 0$, then t/ε converges to infinity. So for which time should the terminal condition of e.g. u^1 be defined?
- The terms in this expansion will all be stochastic. Then if u^1 is adapted to $\mathcal{F}^B$, the stochastic integral $\int_t^T u^1_{s/\varepsilon}(x,z)d\overleftarrow{B}_s$ a priori does not make any sense for $\varepsilon < 1$.

However if we do such a formal asymptotic expansion, and then call

$$v^0(t,x) = u^0(t,x), \qquad \psi^1(t,x,z) = \varepsilon u^1_{t/\varepsilon}(x,z), \qquad R(t,x,z) = \varepsilon^2 u^2_{t/\varepsilon}(x,z)$$

(of course all terms except v^0 depend on ε, which we omit in the notation to facilitate the reading), then these terms have to solve the following equations:

$$-dv^0_t(x) = \bar{\mathcal{L}} v^0_t(x,z)dt + \bar{h}(x)^* v^0_t(x)d\overleftarrow{B}_t$$

$$-d\psi^1_t(x,z) = \frac{1}{\varepsilon}\mathcal{L}_F \psi^1_t(x,z)dt + (\mathcal{L}_S - \bar{\mathcal{L}})v^0_t(x)dt + \left(h(x,z) - \bar{h}(x)\right)^* v^0_t(x)d\overleftarrow{B}_t \tag{7}$$

$$-dR_t(x,z) = \mathcal{L}^\varepsilon R_t(x,z)dt + \mathcal{L}_S \psi^1_t(x,z)dt + h(x,z)^*\left(\psi^1_t(x,z) + R_t(x,z)\right)d\overleftarrow{B}_t \tag{8}$$

with terminal conditions

$$v^0(T,x) = \varphi(x), \quad \psi^1(T,x,z) = R(T,x,z) = 0.$$

Note that the equation for v^0 is exactly the desired equation (6). By existence and uniqueness of the solutions to these *linear* equations, we can apply superposition to obtain that then indeed

$$v^\varepsilon_t(x,z) = v^0_t(x) + \psi^1_t(x,z) + R_t(x,z).$$

Therefore the problem of showing L^p-convergence of v^ε to v^0 reduces to showing L^p-convergence of $\psi^1 + R$ to 0. To achieve this, we will give probabilistic representations of ψ^1 and R in terms of backward doubly stochastic differential equations. This will allow us to apply the existing estimates for the transition function of Z^x from Pardoux and Veretennikov (2003).

It will be convenient for us to work with functions that are smoother in their x-component than they are in their z-component or vice versa. To do so, introduce the function spaces $C^{k,l}(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^d)$: For $\theta : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^d$, $\theta = \theta(x,z)$, write $\theta \in C^{k,l}(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^d)$, if θ is k times continuously differentiable in its x-components and l times continuously differentiable in its z-components. If θ as well as its partial derivatives up to order (k,l) are bounded, write $\theta \in C^{k,l}_b(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^d)$.

Introduce the following assumptions:

($\mathrm{H}_{\mathrm{stat}}$) For the existence of a stationary distribution $\mu(x,dz)$ for Z^x, we suppose that there exist $M_0 > 0, \alpha > 0$, such that for all $|z| \geq M_0$

$$\sup_x \langle f(x,z), z\rangle \leq -C|z|^\alpha.$$

For the uniqueness of the stationary distribution $\mu(x,dz)$ of Z^x, we suppose uniform ellipticity, i.e. that there are $0 < \lambda \leq \Lambda < \infty$, such that

$$\lambda I \leq gg^*(x,y) \leq \Lambda I$$

in the sense of positive semi-definite matrices (I is the unit matrix).

($\mathrm{HF}_{k,l}$) The coefficients of the fast diffusion satisfy $f \in C^{k,l}_b(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^n)$ and $g \in C^{k,l}_b(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^{n\times k})$.

($\mathrm{HS}_{k,l}$) The coefficients of the slow diffusion satisfy $b \in C^{k,l}_b(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^m)$ and $\sigma \in C^{k,l}_b(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^{m\times k})$.

($\mathrm{HO}_{k,l}$) The observation function h satisfies $h \in C^{k,l}_b(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^d)$.

We will usually write $p_\infty(x, dz)$ instead of $\mu(x, dz)$. Also introduce the notation

$$p_t(z, \theta; x) := \int_{\mathbb{R}^n} \theta(x, z') p_t(z, z'; x) dz' := \mathbb{E}_z[\theta(Z^x_t)]$$

where z denotes the starting point of Z^x, and $z' \mapsto p_t(z, z'; x)$ is the density of Z^x_t if at time 0 it is started in z. Note that the density exists for all $t > 0$ under the condition ($\mathrm{H}_{\mathrm{stat}}$), because of the uniform ellipticity of gg^*. Similarly

$$p_\infty(\theta; x) = \int_{\mathbb{R}^n} \theta(x, z) p_\infty(x, dz).$$

Let the differential operator $\bar{\mathcal{L}}$ be defined as

$$\bar{\mathcal{L}} = \sum_{i=1}^m \bar{b}_i(x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^m \bar{a}_{ij}(x, z) \frac{\partial^2}{\partial x_i \partial x_j}$$

where $\bar{b}(x) = p_\infty(b; x)$ and $\bar{a} = p_\infty(\sigma\sigma^*; x)$. Also define $\bar{h}(x) = p_\infty(h; x)$.

We introduce the following notation: A multiindex $\alpha = (\alpha_1, \dots, \alpha_m) \in \mathbb{N}^n_0$ is of order

$$|\alpha| = \alpha_1 + \dots + \alpha_m.$$

Given such a multiindex, define the differential operator

$$D^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \dots x_m^{\alpha_m}}.$$

Finally introduce the following norms for $f \in C^k_b(\mathbb{R}^m, \mathbb{R}^n)$:

$$||f||_{k,\infty} = \sum_{|\alpha| \le k} ||D^\alpha f||_\infty$$

where $|| \cdot ||_\infty$ is the usual supremum norm.

Our main result is

Theorem 3.1. *Assume (H_{stat}), ($HF_{8,4}$), ($HS_{7,4}$), ($HO_{8,4}$), and that the initial distribution $\mathbb{Q}_{(X^\varepsilon_0, Z^\varepsilon_0)}$ has finite moments of every order. Then for every $p \ge 1$ and $T \ge 0$ there exists $C > 0$, such that for every $\varphi \in C^4_b$*

$$\left(\mathbb{E}_{\mathbb{Q}}\left[\left|\pi^{\varepsilon,x}_T(\varphi) - \pi^0_T(\varphi)\right|^p\right]\right)^{1/p} \le \sqrt{\varepsilon} C ||\varphi||_{4,\infty}.$$

In particular, there exists a metric d on the space of probability measures, such that d generates the topology of weak convergence, and such that for every $T \ge 0$ there exists $C > 0$, such that

$$\mathbb{E}_{\mathbb{Q}}\left[d(\pi^{\varepsilon,x}_T, \pi^0_T)\right] \le \sqrt{\varepsilon} C.$$

This result will be proven in Section 6.

In particular we can use Borel-Cantelli to conclude that if (ε_n) converges quickly enough to 0, then π^{ε_n} will a.s. converge weakly to π^0.

The ideas are rather simple: We represent the backward SPDEs by finite-dimensional stochastic equations (this will be BDSDEs). The diffusion operators get replaced by the associated diffusions. We are able to solve those finite-dimensional equations explicitly, or at least give explicit estimates up to an application of Gronwall. This allows us to estimate ψ^1 and R in terms of the transition function of the fast diffusion. But Pardoux and Veretennikov (2003) proved very precise estimates for this transition function. These estimates allow us to obtain the convergence.

While the ideas are simple, the precise formulation and the actual proofs are quite technical. We start by describing the probabilistic representation.

4. Probabilistic representation of SPDEs

In this section, we derive probabilistic representations for SPDEs of the form

$$
\begin{aligned}
-d\psi(\omega,t,x) &= \mathcal{L}\psi(\omega,t,x)dt + f(\omega,t,x)dt \\
&\qquad + (g(\omega,t,x) + G(\omega,t,x)\psi(\omega,t,x))d\overleftarrow{B}_t, \\
\psi(T,x) &= \varphi(\omega,x),
\end{aligned} \tag{9}
$$

where $\psi : \Omega \times [0,T] \times \mathbb{R}^m \to \mathbb{R}$, $f : \Omega \times [0,T] \times \mathbb{R}^m \to \mathbb{R}$, $g : \Omega \times [0,T] \times \mathbb{R}^m \to \mathbb{R}^{1\times d}$, and $G : \Omega \times [0,T] \times \mathbb{R}^m \to \mathbb{R}^{1\times d}$, $\varphi : \Omega \times \mathbb{R}^m \to \mathbb{R}$ are all jointly measurable, and $(B_t : t \in [0,T])$ is a d-dimensional standard Brownian motion under the measure $\mathbb{P}$. Equation (9) represents the general form of the equations (7) and (8) for the corrector $\psi^1_t(x,z)$ and error $R_t(x,z)$, respectively. The differential operator $\mathcal{L}$ is given by

$$
\mathcal{L} = \sum_{i=1}^m b_i(x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^m a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j}
$$

for measurable $b : \mathbb{R}^m \to \mathbb{R}^m$ and $a : \mathbb{R}^m \to \mathbb{S}^{m\times m}$ ($\mathbb{S}^{m\times m}$ denotes positive semidefinite symmetric matrices). We will represent these equations in terms of BDSDEs as introduced by Pardoux and Peng (1994). Note that for these linear equations it is possible to give a Feynman-Kac type representation without using BDSDEs. This is done, for example, in Rozovskii (1990) ("The Method of Stochastic Characteristics"). However the BDSDE-representation has the advantage that it permits us to apply Gronwall's lemma. This would not be possible with the method of stochastic characteristics.

A BDSDE is an integral equation of the form

$$
Y_t = \xi + \int_t^T f(s,Y_s,Z_s)ds + \int_t^T g(s,Y_s,Z_s)d\overleftarrow{B}_s - \int_t^T Z_s dW_s
$$

where B and W are independent Brownian motions. The solution (Y_t, Z_t) will be $\mathcal{F}^B_{t,T} \vee \mathcal{F}^W_t$-measurable. Starting from the notion of BDSDEs, we can define forward-backward doubly stochastic differential equations. Let $\sigma = a^{1/2}$ and

$$
\begin{aligned}
X^{t,x}_s &= x + \int_t^s b(X^{t,x}_s)ds + \int_t^s \sigma(X^{t,x}_s)dW_s \qquad \text{for } s \geq t \\
X^{t,x}_s &= x \qquad \text{for } s \leq t
\end{aligned}
$$

We then define the following BDSDE

$$
\begin{aligned}
-dY^{t,x}_s &= f(s,X^{t,x}_s)ds + (g(s,X^{t,x}_s)ds + G(s,X^{t,x}_s)Y^{t,x}_s)d\overleftarrow{B}_s - Z^{t,x}_s dW_s \\
Y^{t,x}_T &= \varphi(X^{t,x}_T)
\end{aligned}
$$

It turns out that Y gives a finite-dimensional probabilistic representation for equation (9), more precisely we have $Y^{t,x}_t = \psi(t,x)$. This is not completely covered by Pardoux and Peng (1994), because we have random unbounded coefficients, and because we do not assume the diffusion matrix a to have a smooth square root. On the other side, the equation is of a particularly simple linear type. In the remainder of this section, we give the precise statement and proof for this representation. This can be skipped at first reading.

We will not be able to get an existence result for classical solutions of the above SPDE from the theory of BDSDEs: This is due to the fact that for this we would need smoothness properties of a square root of a. But even when a is smooth, in the degenerate elliptic case it does not need to have a smooth square root (see, for example, Stroock (2008), Chapter 2.3). Instead we will use the existence result of Rozovskii (1990) and only reprove the uniqueness result of Pardoux and Peng (1994) in our setting. This will work under Lipschitz continuity of $a^{1/2}$.

Define for $0 \leq t \leq s \leq T$

$$\mathcal{F}^{0,B}_{t,s} = \sigma(B_u - B_t : t \leq u \leq s)$$

and $\mathcal{F}^{B}_{t,s}$ as the completion of $\mathcal{F}^{0,B}_{t,s}$ under $\mathbb{P}$. Introduce the space of adapted random fields of polynomial growth:

Definition. *$\mathcal{P}_T(\mathbb{R}^m, \mathbb{R}^n)$ is the space of random fields*

$$H : \Omega \times [0,T] \times \mathbb{R}^m \to \mathbb{R}^n$$

that are jointly measurable in (ω, t, x), and for fixed (t,x), $\omega \mapsto H(\omega, t, x)$ is $\mathcal{F}^B_{t,T}$-measurable. Further for fixed ω outside a null set, H has to be jointly continuous in (t,x), and it has to satisfy the following inequality: For every $p \geq 1$ there is $C_p > 0$, $q > 0$, such that for all $x \in \mathbb{R}^m$

$$\mathbb{E}\left[\sup_{0\leq t\leq T} |H(t,x)|^p\right] \leq C_p(1 + |x|^q)$$

We make the following assumptions on the coefficients of the SPDE:

(S_k) f and g are k times continuously differentiable and the partial derivatives up to order k are all in $\mathcal{P}_T$. G is $(k+1)$ times continuously differentiable and the partial derivatives up to order $(k+1)$ are all uniformly bounded in (ω, t, x). φ is k times continuously differentiable, and all partial derivates of order 0 to k grow at most polynomially.

We make the following assumptions on the coefficients of the differential operator $\mathcal{L}$:

(D_k) $b \in C^k_b(\mathbb{R}^m, \mathbb{R}^m)$, $a \in C^k_b(\mathbb{R}^m, \mathbb{S}^{m\times m})$, and a is degenerate elliptic: For every $\xi \in \mathbb{R}^m$ and every $x \in \mathbb{R}^m$,

$$\langle a(x)\xi, \xi\rangle = \sum_{i,j=1}^{m} a_{ij}(x)\xi_i\xi_j \geq 0.$$

Then we have the following result:

Proposition 4.1. *Assume (S_k) and (D_k) for some $k \geq 3$. Then the equation (9) has a unique classical solution ψ in the sense that for every fixed ω outside a null set, $\psi(\omega, \cdot, \cdot) \in C^{0,k-1}([0,T] \times \mathbb{R}^d, \mathbb{R})$, ψ and its partial derivatives are in $\mathcal{P}_T(\mathbb{R}^m, \mathbb{R})$, and ψ solves the integral equation. If $\tilde{\psi}$ is any other solution of the integral equation, then ψ and $\tilde{\psi}$ are indistinguishable. If further f, g and φ as well as their derivatives up to order k are uniformly bounded in (ω, t, x), then for any $p > 0$ there exist $C_p, q > 0$ (only depending on p, the dimensions involved, the bounds on a, b and G, and on T), such that for all $|\alpha| \leq k-1$ and $x \in \mathbb{R}^m$:*

$$\mathbb{E}\left[\sup_{t\leq T} |D^\alpha \psi(t,x)|^p\right] \leq C(1+|x|^q)\mathbb{E}\left[||\varphi||^p_{k,\infty} + \sup_{t\leq T} ||f(t,\cdot)||^p_{k,\infty} + \sup_{t\leq T} ||g(t,\cdot)||^p_{k,\infty}\right].$$

Proof. This is a combination of Theorem 4.3.2 and Corollary 4.3.2 of Rozovskii (1990) (The claimed bound is only given for the equation in unweighted Sobolev spaces, in Corollary 4.2.2. But from that we can deduce the result for the weighted Sobolev case). The only thing we need to verify is that our polynomial growth assumption on the coefficients is compatible with the Sobolev norm condition there. But if $\theta \in \mathcal{P}_T(\mathbb{R}^m, \mathbb{R}^n)$, then for any $p \geq 1$ there certainly is an

$r < 0$ such that θ takes its values in the weighted L^p-space with weight $(1+|x|^2)^{r/2}$:

$$\begin{aligned}\mathbb{E}\left[\sup_{0\le t\le T}\int|\theta(t,x)|^p(1+|x|^2)^{\frac{r}{2}}dx\right] &\le \mathbb{E}\left[\int\sup_{0\le t\le T}|\theta(t,x)|^p(1+|x|^2)^{\frac{r}{2}}dx\right]\\ &= \int\mathbb{E}\left[\sup_{0\le t\le T}|\theta(t,x)|^p\right](1+|x|^2)^{\frac{r}{2}}dx\\ &\le \int C_p(1+|x|^q)(1+|x|^2)^{\frac{r}{2}}dx<\infty\end{aligned}$$

for small enough r. □

Now we combine this result with the theory of BDSDEs:

Let $(W_t : t \in [0,T])$ be an n-dimensional standard Brownian motion that is independent of B. For $0 \le t \le s$, $\mathcal{F}^W_{t,s}$ is defined analogously to $\mathcal{F}^B_{t,s}$. For $0 \le t \le T$ we set

$$\mathcal{F}_t = \mathcal{F}^B_{t,T} \vee \mathcal{F}^W_t.$$

Note that this is not a filtration, as it is neither decreasing nor increasing in t. Introduce the following notation:

- $H^2_T(\mathbb{R}^m)$ is the space of measurable $\mathbb{R}^m$-valued processes Y s.t. Y_t is $\mathcal{F}_t$-measurable and
$$\mathbb{E}\left[\int_0^T|Y_t|^2dt\right]<\infty.$$
- $S^2_T(\mathbb{R}^m)$ is the space of continuous adapted $\mathbb{R}^m$-valued processes Y s.t. $Y_t \in \mathcal{F}_t$ and
$$\mathbb{E}\left[\sup_{0\le t\le T}|Y_t|^2\right]<\infty.$$

A BDSDE is an integral equation of the form

$$Y_t=\xi+\int_t^T f(s,\cdot,Y_s,Z_s)ds+\int_t^T g(s,\cdot,Y_s,Z_s)d\overleftarrow{B}_s-\int_t^T Z_sdW_s, \tag{10}$$

where $f : [0,T]\times\Omega\times\mathbb{R}\times\mathbb{R}^{1\times n}\to\mathbb{R}$, $g : [0,T]\times\Omega\times\mathbb{R}\times\mathbb{R}^{1\times n}\to\mathbb{R}^{1\times l}$, and for fixed $y\in\mathbb{R}$, $z\in\mathbb{R}^{1\times n}$ the processes $(\omega,t)\mapsto f(t,\omega,x,z)$ and $(\omega,t)\mapsto g(t,\omega,x,z)$ are $(\mathcal{F}^B_{0,T}\vee\mathcal{F}^W_T)\otimes\mathcal{B}(\mathbb{R})$-measurable, and for every t, $f(t,\cdot,x,z)$ and $g(t,\cdot,x,z)$ are $\mathcal{F}_t$-measurable.

(Y,Z) will be called solution of (10) if $(Y,Z)\in S^2_T(\mathbb{R})\times H^2_T(\mathbb{R}^{1\times n})$ and if the couple solves the integral equation.

We will also write the equation in differential form:

$$-dY_t = f(t,Y_t,Z_t)dt + g(t,Y_t,Z_t)d\overleftarrow{B}_t - Z_tdW_t.$$

Observe that with suitable adaptations, all of the following results also hold in the multidimensional case, i.e. for $Y\in\mathbb{R}^m$. We restrict to one-dimensional Y for simplicity and because ultimately we are only interested in that case.

Pardoux and Peng (1994) show that under the following conditions, equation (10) has a unique solution:

- $\xi\in L^2(\Omega,\mathcal{F}_T,\mathbb{P};\mathbb{R})$
- for any $(y,z)\in\mathbb{R}\times\mathbb{R}^{1\times n}$: $f(\cdot,\cdot,y,z)\in H^2_T(\mathbb{R})$ and $g(\cdot,\cdot,y,z)\in H^2_T(\mathbb{R}^{1\times k})$
- f and g satisfy Lipschitz conditions and g is a contraction in z: there exist constants $L>0$ and $0<\alpha<1$ s.t. for any (ω,t) and y_1,y_2,z_1,z_2:
$$\begin{aligned}&|f(t,\omega,y_1,z_1)-f(t,\omega,y_2,z_2)|^2\le L(|y_1-y_2|^2+|z_1-z_2|^2)\quad\text{and}\\ &|g(t,\omega,y_1,z_1)-g(t,\omega,y_2,z_2)|^2\le L|y_1-y_2|^2+\alpha|z_1-z_2|^2.\end{aligned}$$

Now we want to associate a diffusion X to the differential operator $\mathcal{L}$. To do so, assume that (D_k) is satisfied for some $k \geq 2$. Then $\sigma := a^{1/2}$ is Lipschitz continuous by Lemma 2.3.3 of Stroock (2008). Hence for every $(t,x) \in [0,T] \times \mathbb{R}^m$, there exists a strong solution of the SDE

$$\begin{aligned} X_s^{t,x} &= x + \int_t^s b(X_s^{t,x})ds + \int_t^s \sigma(X_s^{t,x})dW_s \qquad \text{for } s \geq t, \\ X_s^{t,x} &= x \qquad \text{for } s \leq t. \end{aligned}$$

Associate the following BDSDE to (9):

$$\begin{aligned} -dY_s^{t,x} &= f(s, X_s^{t,x})ds + (g(s, X_s^{t,x}) + G(s, X_s^{t,x})Y_s^{t,x})d\overleftarrow{B}_s - Z_s^{t,x}dW_s, \\ Y_T^{t,x} &= \varphi(X_T^{t,x}). \end{aligned} \tag{11}$$

Under the assumptions (S_k) and (D_k) for $k \geq 2$, this equation has a unique solution.

Proposition 4.2. *Assume (S_k) and (D_k) for some $k \geq 3$. Then the unique classical solution ψ of the SPDE (9) is given by $\psi(t,x) = Y_t^{t,x}$, where $(Y^{t,x}, Z^{t,x})$ is the unique solution of the BDSDE (11).*

We can give exactly the same proof as in Pardoux and Peng (1994), Theorem 3.1, taking advantage of the independence of B and W. For the reader's convenience, we include it here.

Proof. Let ψ be a classical solution of (9). It suffices to show that

$$(\psi(s, X_s^{t,x}), D\psi(s, X_s^{t,x})\sigma(X_s^{t,x}) : t \leq s \leq T)$$

solves the BDSDE (11). Here $D\psi$ is the gradient of ψ. For this purpose, consider a partition $t = t_0 < t_1 < \cdots < t_n = T$ of $[t,T]$. Then

$$\begin{aligned} \psi(t, X_t^{t,x}) &= \psi(T, X_T^{t,x}) + \sum_{i=0}^{n-1}(\psi(t_i, X_{t_i}^{t,x}) - \psi(t_{i+1}, X_{t_{i+1}}^{t,x})) \\ &= \varphi(X_T^{t,x}) + \sum_{i=0}^{n-1}(\psi(t_i, X_{t_i}^{t,x}) - \psi(t_{i+1}, X_{t_{i+1}}^{t,x})) \end{aligned}$$

and

$$\begin{aligned} \psi(t_i, &X_{t_i}^{t,x}) - \psi(t_{i+1}, X_{t_{i+1}}^{t,x}) \\ &= (\psi(t_i, X_{t_i}^{t,x}) - \psi(t_i, X_{t_{i+1}}^{t,x})) + (\psi(t_i, X_{t_{i+1}}^{t,x}) - \psi(t_{i+1}, X_{t_{i+1}}^{t,x})) \\ &= -\left(\int_{t_i}^{t_{i+1}} \mathcal{L}\psi(t_i, X_s^{t,x})ds + \int_{t_i}^{t_{i+1}} D\psi(t_i, X_s^{t,x})\sigma(X_s^{t,x})dW_s\right) \\ &\quad + \int_{t_i}^{t_{i+1}} (\mathcal{L}\psi(s, X_{t_{i+1}}^{t,x}) + f(s, X_{t_{i+1}}^{t,x}))ds \\ &\quad + \int_{t_i}^{t_{i+1}} (g(s, X_{t_{i+1}}^{t,x}) + G(X_{t_{i+1}}^{t,x})\psi(s, X_{t_{i+1}}^{t,x}))d\overleftarrow{B}_s. \end{aligned}$$

This is justified because $X^{t,x}$ and ψ are independent and because ψ grows polynomially, hence we can apply Itô's formula. We also used the fact that ψ is a classical solution to (9). If we let the mesh size tend to 0, then by continuity of $X^{t,x}$ and ψ, the result follows. □

5. Preliminary estimates

The notation D_x^α indicates that the differential operator D^α is only acting on the x-variables.

The following result will help us to justify the BDSDE-representations on the deeper levels. Recall that $p_t(z, \theta; x) = \mathbb{E}[\theta(x, Z_t^x)|Z_0^x = z]$.

Proposition 5.1. *Assume ($HF_{k,l}$). Let $\theta \in C^{k,l}(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R})$ satisfy for some $C, p > 0$*

$$\sum_{|\alpha|\le k}\sum_{|\beta|\le l}\left|D_x^\alpha D_z^\beta \theta(x,z)\right| \le C(1+|x|^p+|z|^p).$$

Then

$$(t,x,z)\mapsto p_t(z,\theta;x)\in C^{0,k,l}(\mathbb{R}_+\times\mathbb{R}^m\times\mathbb{R}^n,\mathbb{R})$$

and there exist $C_1, p_1 > 0$, such that for all $(t,x,z)\in[0,\infty)\times\mathbb{R}^m\times\mathbb{R}^n$

$$\sum_{|\alpha|\le k}\sum_{|\beta|\le l}\left|D_x^\alpha D_z^\beta p_t(z,\theta;x)\right| \le C_1 e^{C_1 t}(1+|x|^{p_1}+|z|^{p_1}).$$

If the bound on the derivatives of θ can be chosen uniformly in x, i.e.

$$\sum_{|\alpha|\le k}\sum_{|\beta|\le l}\sup_x\left|D_x^\alpha D_z^\beta \theta(x,z)\right| \le C(1+|z|^p),$$

then the bound on the derivatives of $p_t(z,\theta;x)$ is also uniform in x:

$$\sum_{|\alpha|\le k}\sum_{|\beta|\le l}\sup_x\left|D_x^\alpha D_z^\beta p_t(z,\theta;x)\right| \le C_1 e^{C_1 t}(1+|z|^{p_1}).$$

Proof. Note that

$$p_t(z,\theta;x)=\mathbb{E}[\theta(x,Z_t^x)|Z_0^x=z]=\mathbb{E}(\theta(X_t,Z_t)|(X_0,Z_0)=(x,z)]$$

is the solution of Kolmogorov's backward equation associated to (X,Z), where

$$\begin{aligned} X_t &= X_0, \\ Z_t &= Z_0+\int_0^t f(X_s,Z_s)ds+\int_0^t g(X_s,Z_s)dW_s. \end{aligned}$$

In this formulation, the first result is standard. Cf. e.g. Stroock (2008), Corollary 2.2.8.

The second statement can be proven in the same way as Stroock (2008), Corollary 2.2.8. ☐

Some results from Pardoux and Veretennikov (2003) are collected in the following Proposition:

Proposition 5.2. *Assume (H_{stat}) and ($HF_{k,3}$). Let $\theta \in C^{k,0}(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R})$ satisfy for some $C, p > 0$:*

$$\sum_{|\alpha|\le k}\sup_x|D_x^\alpha\theta(x,z)|\le C(1+|z|^p).$$

Then

(1) $x\mapsto p_\infty(\theta;x)\in C_b^k(\mathbb{R}^m,\mathbb{R})$.
(2) *Assume additionally that θ satisfies the centering condition*

$$\int_{\mathbb{R}^n}\theta(x,z)p_\infty(x,dz)=0$$

for all x, and that $\theta\in C^{k,1}(\mathbb{R}^m\times\mathbb{R}^n,\mathbb{R})$ and

$$\sum_{|\alpha|\le k}\sum_{|\beta|\le 1}\sup_x\left|D_z^\beta D_x^\alpha\theta(x,z)\right|\le C(1+|z|^p).$$

Then

$$(x,z)\mapsto\int_0^\infty p_t(z,\theta;x)dt\in C^{k,1}(\mathbb{R}^m\times\mathbb{R}^n,\mathbb{R}),$$

and for every $q > 0$ there exist $C_1, q_1 > 0$, such that for every $z \in \mathbb{R}^n$

$$\sum_{|\alpha|\le k}\sum_{|\beta|\le 1}\int_0^\infty \sup_x \left|D_z^\beta D_x^\alpha p_t(z,\theta;x)\right|^q dt \le C_1(1+|z|^{q_1}).$$

Proof. The statements in the Proposition are taken from Theorem 1, Theorem 2 and Proposition 1 of Pardoux and Veretennikov (2003):

(1) We get from Theorem 1 of Pardoux and Veretennikov (2003), that for any $q > 0$ there exists $C_q > 0$, such that for any $(x, z, z') \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^n$:

$$\sum_{|\alpha|\le k} \sup_x \left|D_x^\alpha p_\infty(z';x)\right| \le \frac{C_q}{1+|z'|^q}.$$

So if we choose q large enough and differentiate $p_\infty(\theta;x)$ under the integral sign, then we obtain the first claim. (Of course here we have to use the growth constraint on θ and its derivatives).

(2) This follows from the bounds on the derivatives of $p_t(z,\theta;x)$ that are given in Pardoux and Veretennikov (2003), Theorem 2, formulae (14) and (15): For any $k > 0$ there exist $C_k, m_k > 0$, such that for any $(t, x, z) \in [1,\infty) \times \mathbb{R}^m \times \mathbb{R}^n$

$$\sum_{|\alpha|\le k}\sum_{|\beta|\le 1}\left|D_z^\beta D_x^\alpha p_t(z,\theta;x)\right| \le C_k \frac{1+|z|^{m_k}}{(1+t)^k}.$$

We combine this estimate with Proposition 5.1, from where we obtain for $(t, x, z) \in \mathbb{R}_+ \times \mathbb{R}^m \times \mathbb{R}^n$

$$\sum_{|\alpha|\le k}\sum_{|\beta|\le l}\sup_x \left|D_x^\alpha D_z^\beta p_t(z,\theta;x)\right| \le C_1 e^{C_1 t}(1+|z|^{p_1}).$$

We choose k such that $qk > 1$ and use the first estimate on $[1,\infty)$ and the second estimate on $[0, 1)$. The result follows.

□

We will also need some moment bounds for the diffusions X^ε and Z^ε.

Proposition 5.3. *Assume (H_{stat}) and that the coefficients b and σ and f and g of the fast and slow motion are bounded and globally Lipschitz continuous. Then for any $p \ge 1$ there exists $C_p > 0$, such that*

$$\sup_{(t,\varepsilon,x)\in[0,\infty)\times[0,1]\times\mathbb{R}^m} \mathbb{E}[|Z_t^\varepsilon|^p | (X_0^\varepsilon, Z_0^\varepsilon) = (x,z)] \le C_p(1+|z|^p).$$

Also, for every $T > 0$ and every $p \ge 1$ there exist $C(p,T), q > 0$, such that

$$\sup_{(t,\varepsilon)\in[0,T]\times[0,1]} \mathbb{E}[|X_t^\varepsilon|^p | (X_0^\varepsilon, Z_0^\varepsilon) = (x,z)] \le C(p,T)(1+|x|^p).$$

Proof. The first claim can be proven exactly as in Veretennikov (1997): First write $\bar{Z}_t^\varepsilon := Z_{t\varepsilon^2}^\varepsilon$. Then

$$d\bar{Z}_t^\varepsilon = f(X_{\varepsilon^2 t}^\varepsilon, \bar{Z}_t^\varepsilon)dt + g(X_{\varepsilon^2 t}^\varepsilon, \bar{Z}_t^\varepsilon)d\bar{W}_t^\varepsilon$$

where $\bar{W}_t^\varepsilon := 1/\varepsilon W_{\varepsilon^2 t}$ is a Wiener process. Next, introduce the same time change as in Pardoux and Veretennikov, page 1063:

$$\kappa(x,z) := |g(x,z)^* z|/|z|, \quad \gamma^\varepsilon(t) := \int_0^t \kappa^2(X_{\varepsilon^2 s}^\varepsilon, \bar{Z}_s^\varepsilon)ds, \quad \tau^\varepsilon(t) := (\gamma^\varepsilon)^{-1}(t).$$

Define $\tilde{Z}_t^\varepsilon := \bar{Z}_{\tau^\varepsilon(t)}^\varepsilon$. Then,

$$d\tilde{Z}_t^\varepsilon = \kappa^{-2}(X_{\varepsilon^2 t}^\varepsilon, \tilde{Z}_t^\varepsilon)f(X_{\varepsilon^2 t}^\varepsilon, \tilde{Z}_t^\varepsilon)dt + \kappa^{-1}(X_{\varepsilon^2 t}^\varepsilon, \tilde{Z}_t^\varepsilon)g(X_{\varepsilon^2 t}^\varepsilon, \tilde{Z}_t^\varepsilon)d\tilde{W}_t^\varepsilon$$

with a new standard Brownian motion $\tilde{W}^\varepsilon$. Now we are in a position to just copy the proof of Lemma 1 in Veretennikov (1997) (which we do not do here) to get the first result.

The second claim is obvious, because the coefficients of X^ε are bounded. □

Now we we are able to impose conditions on the coefficients of the diffusions that guarantee smoothness of the coefficients of $\bar{\mathcal{L}}$. Recall that $\bar{\mathcal{L}}$ was defined as

$$\bar{\mathcal{L}} = \sum_{i=1}^m \bar{b}_i(x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^m \bar{a}_{ij}(x,z)\frac{\partial^2}{\partial x_i \partial x_j}$$

where $\bar{b} = p_\infty(b;x)$ and $\bar{a} = p_\infty(\sigma\sigma^*;x)$.

Proposition 5.4. *Assume ($HF_{k,3}$), ($HS_{k,0}$), and ($HO_{k,0}$). Then*

$$\bar{b} \in C_b^k(\mathbb{R}^m,\mathbb{R}^m), \bar{a} \in C_b^k(\mathbb{R}^m,\mathbb{S}^{m\times m}), \bar{h} \in C_b^k(\mathbb{R}^m,\mathbb{R}^k)$$

Proof. All the terms of $\bar{b}$, $\bar{a}$ and $\bar{h}$ are of the form $p_\infty(\theta;x)$. So by Proposition 5.2, we only need to verify that the respective θ are in $C^{k,0}$ and satisfy the polynomial bound

$$\sum_{|\alpha|\le k} \sup_x |D_x^\alpha \theta(x,z)| \le C(1+|z|^p)$$

for some $C, p > 0$. But we even assumed them to be in $C_b^{k,0}$, so the result follows. □

6. Proof of the main result

We will find convergence rates for the corrector and remainder terms that are expressed in terms of v^0 and its derivatives. So now we give bounds on v^0 and its derivatives in terms of the test function φ. This is necessary, because we do not only want to show convergence of the filter integrating fixed test functions, but with respect to a suitable distance on the space of probability measures.

Lemma 6.1. *Let $k \ge 2$ and assume $\bar{b}, \bar{a}, \varphi \in C_b^{k+1}$, and $\bar{h} \in C_b^{k+2}$. Then $v^0 \in C^{0,k}([0,T]\times\mathbb{R}^m,\mathbb{R})$, and for any $p \ge 1$ there exist $C_p, q > 0$, independent of φ, such that for all $x \in \mathbb{R}^m$:*

$$\sum_{|\alpha|\le k} \mathbb{E}\left[\sup_{0\le t\le T} |D^\alpha v_t^0(x)|^p\right] \le C_p(1+|x|^q)||\varphi||_{k,\infty}^p.$$

In particular, v^0 and all its partial derivatives up to order $(0,k)$ are in $\mathcal{P}_T(\mathbb{R}^m,\mathbb{R})$.

Proof. This is a simple application of Proposition 4.1, noting that the equation (6) for v^0 is of the type (9) with $f = 0$, $g = 0$, and $G = \bar{h}^*$. □

We will prove L^p-convergence of ψ^1 and R separately:

Lemma 6.2. *Let $k, l \ge 2$. Assume (H_{stat}), ($HF_{k+1,l+1}$), ($HS_{k+1,l+1}$), and ($HO_{k+1,l+1}$). Also assume $v^0 \in C^{0,k+1}([0,T]\times\mathbb{R}^m,\mathbb{R})$, and that all its partial derivatives in x up to order $k+1$ are in $\mathcal{P}_T(\mathbb{R}^m,\mathbb{R})$. Finally assume $\bar{a}, \bar{b}, \bar{h} \in C_b^k$. Then $\psi^1 \in C^{0,k,l}([0,T]\times\mathbb{R}^m\times\mathbb{R}^n,\mathbb{R})$, and ψ^1 as well as its partial derivatives up to order $(0,k,l)$ are in $\mathcal{P}_T(\mathbb{R}^m\times\mathbb{R}^n,\mathbb{R})$. For any $p \ge 1$ there exist $C_p, q > 0$, independent of φ, such that for any $(x,z) \in \mathbb{R}^{m+n}$ and any $\varepsilon \in (0,1)$*

$$\sum_{|\alpha|\le k-1} \sup_{0\le t\le T} \mathbb{E}\left[|D_x^\alpha \psi_t^1(x,z)|^p\right] \le \varepsilon^{\frac{p}{2}} C_p(1+|z|^q) \sum_{0\le|\alpha|\le k+1} \mathbb{E}\left[\sup_{0\le t\le T} \left|D_x^\alpha v_t^0(x)\right|^p\right].$$

Proof. $\psi_t^1(x,z)$ solves the BSPDE

$$\begin{aligned} -d\psi_t^1(x,z) &= \left[\frac{1}{\varepsilon}\mathcal{L}_F\psi_t^1(x,z) + (\mathcal{L}_S - \bar{\mathcal{L}})v_t^0(x)\right]dt \\ &\quad + \left[h(x,z) - \bar{h}(x)\right]^* v_t^0(x)d\overleftarrow{B}_t, \\ \psi_T^1(x,z) &= 0. \end{aligned} \tag{12}$$

Existence of the solution ψ^1 and its derivatives as well as the polynomial growth all follow from Proposition 4.1. Write $Z^{\varepsilon,x,(t,z)}$ for the solution of the SDE

$$\begin{aligned} dZ_s^{\varepsilon,x,(t,z)} &= \frac{1}{\varepsilon}f(x,Z_s^{\varepsilon,x,(t,z)})ds + \frac{1}{\sqrt{\varepsilon}}g(x,Z_s^{\varepsilon,x,(t,z)})dW_t, \qquad s\geq t \\ Z_s^{\varepsilon,x,(t,z)} &= z, \qquad s\leq t. \end{aligned}$$

We consider $(x, Z^{\varepsilon,x,(t,z)})$ as a joint diffusion, just as in the proof of Proposition 5.1 (x has generator 0). By Proposition 4.2, the solution of (12) is given by $\theta_t^{(t,x,z)(1)}$, the unique solution to the BDSDE

$$\begin{aligned} -d\theta_s^{(t,x,z)(1)} &= (\mathcal{L}_S(\cdot, Z_s^{\varepsilon,x,(t,z)}) - \bar{\mathcal{L}})v_s^0(x)ds \\ &\quad + \left(h(x,Z_s^{\varepsilon,x,(t,z)}) - \bar{h}(x)\right)^* v_s^0(x)d\overleftarrow{B}_s + \gamma_s^{t,x,z}dW_s, \\ \theta_T^{(t,x,z)(1)} &= 0. \end{aligned}$$

We will drop superscripts (t,x,z) for $\theta_s^{(t,x,z)(1)}$ and write θ_s^1 instead. Similarly, we write $Z_s^{\varepsilon,x}$ instead of $Z_s^{\varepsilon,x,(t,z)}$. $\psi_t^1(x,z)$ is $\mathcal{F}_{t,T}^B$-measurable, hence, so is θ_t^1. We can then write $\theta_t^1 = \mathbb{E}\left[\theta_t^1|\mathcal{F}_{t,T}^B\right]$, where

$$\begin{aligned} &\mathbb{E}\left[\theta_t^1|\mathcal{F}_{t,T}^B\right] \\ &= \mathbb{E}\left[\int_t^T (\mathcal{L}_S - \bar{\mathcal{L}})v_s^0(x)ds|\mathcal{F}_{t,T}^B\right] \\ &\quad + \mathbb{E}\left[\int_t^T \left[h(x,Z_s^{\varepsilon,x}) - \bar{h}(x)\right]^* v_s^0(x)d\overleftarrow{B}_s|\mathcal{F}_{t,T}^B\right] - \mathbb{E}\left[\int_t^T \gamma_s^{t,x,z}dW_s|\mathcal{F}_{t,T}^B\right]. \end{aligned}$$

W and B are independent, therefore W is a Brownian motion in the large filtration $(F_s^W \vee F_{t,T}^B : s\in[0,T])$, hence $\mathbb{E}\left[\int_t^T \gamma_s^{t,x,z}dW_s|\mathcal{F}_t^W \vee \mathcal{F}_{t,T}^B\right] = 0$, and by the tower property

$$\mathbb{E}\left[\int_t^T \gamma_s^{t,x,z}dW_s|\mathcal{F}_{t,T}^B\right] = 0.$$

v_s^0 is $\mathcal{F}_{s,T}^B$-measurable and $\bar{\mathcal{L}}$ has deterministic coefficients. Thus

$$\begin{aligned} &\mathbb{E}\left[\int_t^T \bar{\mathcal{L}}v_s^0(x)ds|\mathcal{F}_{t,T}^B\right] \\ &= \int_t^T \mathbb{E}\left[\bar{\mathcal{L}}v_s^0(x)|\mathcal{F}_{s,T}^B\right]ds \\ &= \int_t^T \left\{\sum_{i=1}^m p_\infty(b_i;x)\frac{\partial}{\partial x_i}v_s^0(x) + \sum_{i,j=1}^m p_\infty((\sigma\sigma^*)_{ij};x)\frac{\partial^2}{\partial x_i x_j}v_s^0(x)\right\}ds. \end{aligned}$$

Since $Z^{\varepsilon,x}$ is independent of B,

$$\begin{aligned}
&\mathbb{E}\left[\int_t^T \mathcal{L}_S(\cdot, Z_s^{\varepsilon,x})v_s^0(x)ds|\mathcal{F}_{t,T}^B\right] = \int_t^T \mathbb{E}\left[\mathcal{L}_S(\cdot, Z_s^{\varepsilon,x})v_s^0(x)|\mathcal{F}_{s,T}^B\right] ds \\
&= \int_t^T \left\{\sum_{i=1}^m \mathbb{E}\left[b_i(x, Z_s^{\varepsilon,x})\right] \frac{\partial}{\partial x_i} v_s^0(x)\right. \\
&\qquad \left. + \frac{1}{2}\sum_{i,j=1}^m \mathbb{E}\left[(\sigma\sigma^*)_{ij}(x, Z_s^{\varepsilon,x})\right] \frac{\partial^2}{\partial x_i x_j} v_s^0(x)\right\} ds \\
&= \int_t^T \left\{\sum_{i=1}^m p_{\frac{s-t}{\varepsilon}}(z, b_i; x) \frac{\partial}{\partial x_i} v_s^0(x)\right. \\
&\qquad \left. + \frac{1}{2}\sum_{i,j=1}^m p_{\frac{s-t}{\varepsilon}}(z, (\sigma\sigma^*)_{ij}; x) \frac{\partial^2}{\partial x_i x_j} v_s^0(x)\right\} ds,
\end{aligned}$$

so

$$\begin{aligned}
&\left|\mathbb{E}\left[\int_t^T (\mathcal{L}_S - \bar{\mathcal{L}})v_s^0(x)ds|\mathcal{F}_{t,T}^B\right]\right| \\
&= \left|\int_t^T \left\{\sum_{i=1}^m p_{\frac{s-t}{\varepsilon}}(z, b_i - p_\infty(b_i;x);x) \frac{\partial}{\partial x_i} v_s^0(x)\right.\right. \\
&\quad \left.\left. + \frac{1}{2}\sum_{i,j=1}^m p_{\frac{s-t}{\varepsilon}}(z, (\sigma\sigma^*)_{ij} - p_\infty((\sigma\sigma^*)_{ij};x);x) \frac{\partial^2}{\partial x_i x_j} v_s^0(x)\right\} ds\right|
\end{aligned}$$

(the $p_\infty(.;x)$ terms have been brought inside the integral $p_{\frac{s-t}{\varepsilon}}(z,\cdot;x)$ since they not depend on z)

$$\begin{aligned}
&\leq \varepsilon \left|\sum_{i=1}^m \int_0^{\frac{T-t}{\varepsilon}} p_u(z, b_i - p_\infty(b_i;x);x) \frac{\partial}{\partial x_i} v_{\varepsilon u+t}^0(x) du\right| \\
&\quad + \frac{\varepsilon}{2}\left|\sum_{i,j=1}^m \int_0^{\frac{T-t}{\varepsilon}} p_u(z, (\sigma\sigma^*)_{ij} - p_\infty((\sigma\sigma^*)_{ij};x);x) \frac{\partial^2}{\partial x_i x_j} v_{\varepsilon u+t}^0(x) du\right| \\
&\leq \varepsilon \sum_{i=1}^m \int_0^\infty |p_u(z, b_i - p_\infty(b_i;x);x)|\, du \sup_{t\leq s\leq T}\left|\frac{\partial}{\partial x_i} v_s^0(x)\right| \\
&\quad + \frac{\varepsilon}{2}\sum_{i,j=1}^m \int_0^\infty |p_u(z, (\sigma\sigma^*)_{ij} - p_\infty((\sigma\sigma^*)_{ij};x);x)|\, du \sup_{t\leq s\leq T}\left|\frac{\partial^2}{\partial x_i x_j} v_s^0(x)\right|
\end{aligned}$$

($f - p_\infty(f;x)$ is centered, so by Proposition 5.2, (2):)

$$\leq \varepsilon C_1(1+|z|^{q_1})\left\{\sum_{i=1}^m \sup_{t\leq s\leq T}\left|\frac{\partial}{\partial x_i} v_s^0(x)\right| + \sum_{i,j=1}^m \sup_{t\leq s\leq T}\left|\frac{\partial^2}{\partial x_i x_j} v_s^0(x)\right|\right\}$$

and therefore finally

$$\mathbb{E}\left[\left|\mathbb{E}\left[\int_t^T (\mathcal{L}_S - \bar{\mathcal{L}})v_s^0(x)ds|\mathcal{F}_{t,T}^B\right]\right|^p\right] \tag{13}$$
$$\leq \varepsilon^p C_2(1+|z|^{q_2})\mathbb{E}\left[\sum_{i=1}^m \sup_{t\leq s\leq T}\left|\frac{\partial}{\partial x_i}v_s^0(x)\right|^p + \sum_{i,j=1}^m \sup_{t\leq s\leq T}\left|\frac{\partial^2}{\partial x_i x_j}v_s^0(x)\right|^p\right].$$

Next, using again $v_s^0 \in \mathcal{F}_{s,T}^B$ and that $Z^{\varepsilon,x}$ is independent of B,

$$\mathbb{E}\left[\int_t^T \left[h(x,Z_s^{\varepsilon,x}) - \bar{h}(x)\right]^* v_s^0(x)d\overleftarrow{B}_s|\mathcal{F}_{t,T}^B\right]$$
$$= \int_t^T \mathbb{E}\left[\left[h(x,Z_s^{\varepsilon,x}) - \bar{h}(x)\right]^* v_s^0(x)|\mathcal{F}_{s,T}^B\right] d\overleftarrow{B}_s$$
$$= \int_t^T p_{\frac{s-t}{\varepsilon}}(z,h-\bar{h};x)^* v_s^0(x)d\overleftarrow{B}_s.$$

For $t \leq r \leq T$, $r \mapsto \int_r^T p_{\frac{s-t}{\varepsilon}}(z,h-\bar{h};x)^* v_s^0(x)d\overleftarrow{B}_s$, is a martingale w.r.t. $(\mathcal{F}_{r,T}^B : r \in [t,T])$ if time is run backwards. Hence by the Burkholder-Davis-Gundy inequality,

$$\mathbb{E}\left[\left|\int_t^T p_{\frac{s-t}{\varepsilon}}(z,h-\bar{h};x)^* v_s^0(x)d\overleftarrow{B}_s\right|^p\right]$$
$$\leq C_p\mathbb{E}\left[\langle\int_t^T p_{\frac{s-t}{\varepsilon}}(z,h-\bar{h};x)^* v_s^0(x)d\overleftarrow{B}_s\rangle^{\frac{p}{2}}\right],$$

where

$$\langle\int_t^T p_{\frac{s-t}{\varepsilon}}(z,h-\bar{h};x)^* v_s^0(x)d\overleftarrow{B}_s\rangle = \int_t^T \left|p_{\frac{s-t}{\varepsilon}}(z,h-\bar{h};x)^* v_s^0(x)\right|^2 ds$$
$$\leq \varepsilon \int_0^\infty \left|p_u(z,h-\bar{h};x)\right|^2 du \sup_{t\leq s\leq T}\left|v_s^0(x)\right|^2$$
$$\leq \varepsilon C_3(1+|z|^{q_3}) \sup_{t\leq s\leq T}\left|v_s^0(x)\right|^2,$$

where the last inequality is by Proposition 5.2, (2), since $h-\bar{h}$ is centered. Therefore,

$$\mathbb{E}\left[\left|\int_t^T p_{\frac{s-t}{\varepsilon}}(z,h-\bar{h};x)^* v_s^0(x)d\overleftarrow{B}_s\right|^p\right] \leq \varepsilon^{\frac{p}{2}}C_4(1+|z|^{q_4})\mathbb{E}\left[\sup_{t\leq s\leq T}\left|v_s^0(x)\right|^p\right]. \tag{14}$$

Combining (13) and (14),

$$\mathbb{E}\left[\left|\theta_t^1\right|^p\right] \leq \varepsilon^p C_4(1+|z|^{q_4})\sum_{|\alpha|\leq 2}\mathbb{E}\left[\sup_{t\leq s\leq T}|D_x^\alpha v_s^0(x)|^p\right].$$

Next, consider a first order x-derivative of θ_t^1:

$$\frac{\partial}{\partial x_k}\theta_t^1 = \frac{\partial}{\partial x_k}\int_t^T \mathbb{E}\left[\mathcal{L}_S - \bar{\mathcal{L}}\right] v_s^0(x)ds$$
$$+ \frac{\partial}{\partial x_k}\int_t^T \mathbb{E}\left[h(x,Z_s^{\varepsilon,x}) - \bar{h}(x)\right]^* v_s^0(x)d\overleftarrow{B}_s.$$

As before, the forward Itó integral term vanished after taking the (conditional) expectation.

Interchanging order of differentiation and integration,

$$
\begin{aligned}
&\left|\frac{\partial}{\partial x_k}\int_t^T \mathbb{E}\left[\mathcal{L}_S-\bar{\mathcal{L}}\right] v_s^0(x)ds\right| \\
&\leq \varepsilon\sum_{i=1}^m\left|\int_0^{\frac{T-t}{\varepsilon}}\left\{\frac{\partial}{\partial x_k}p_u(z,b_i-p_\infty(b_i;x);x)\frac{\partial}{\partial x_i}v^0_{\varepsilon u+t}(x)\right.\right. \\
&\qquad\qquad \left.\left.+p_u(z,b_i-p_\infty(b_i;x);x)\frac{\partial^2}{\partial x_k x_i}v^0_{\varepsilon u+t}(x)\right\}du\right| \\
&\quad+\frac{\varepsilon}{2}\sum_{i,j=1}^m\left|\int_0^{\frac{T-t}{\varepsilon}}\left\{\frac{\partial}{\partial x_k}p_u(z,(\sigma\sigma^*)_{ij}-p_\infty((\sigma\sigma^*)_{ij};x);x)\frac{\partial^2}{\partial x_i x_j}v^0_{\varepsilon u+t}(x)\right.\right. \\
&\qquad\qquad \left.\left.+p_u(z,(\sigma\sigma^*)_{ij}-p_\infty((\sigma\sigma^*)_{ij};x);x)\frac{\partial^3}{\partial x_i x_j x_k}v^0_{\varepsilon u+t}(x)\right\}du\right| \\
&\leq \varepsilon\sum_{i=1}^m\left\{\int_0^\infty\left|\frac{\partial}{\partial x_k}p_u(z,b_i-p_\infty(b_i;x);x)\right|du\sup_{t\leq s\leq T}\left|\frac{\partial}{\partial x_i}v_s^0(x)\right|\right. \\
&\qquad\left.+\int_0^\infty|p_u(z,b_i-p_\infty(b_i;x);x)|\,du\sup_{t\leq s\leq T}\left|\frac{\partial^2}{\partial x_k x_i}v_s^0(x)\right|\right\} \\
&\quad+\frac{\varepsilon}{2}\sum_{i,j=1}^m\left\{\int_0^\infty\left|\frac{\partial}{\partial x_k}p_u(z,(\sigma\sigma^*)_{ij}-p_\infty((\sigma\sigma^*)_{ij};x);x)\right|du\right. \\
&\qquad\qquad\qquad\times\sup_{t\leq s\leq T}\left|\frac{\partial^2}{\partial x_i x_j}v_s^0(x)\right| \\
&\qquad\left.+\int_0^\infty|p_u(z,(\sigma\sigma^*)_{ij}-p_\infty((\sigma\sigma^*)_{ij};x);x)|\,du\sup_{t\leq s\leq T}\left|\frac{\partial^3}{\partial x_i x_j x_k}v_s^0(x)\right|\right\}.
\end{aligned}
$$

Then, from Proposition 5.2, (2) again,

$$
\left|\frac{\partial}{\partial x_k}\int_t^T \mathbb{E}\left[\mathcal{L}_S-\bar{\mathcal{L}}\right] v_s^0(x)ds\right|\leq \varepsilon C_5(1+|z|^{q_5})\sum_{1\leq\beta\leq 3}\sup_{t\leq s\leq T}\left|D_x^\beta v_s^0(x)\right|.
$$

since the quantities $b-\bar{b}$ and $\sigma\sigma^*-\bar{\sigma\sigma^*}$ are centered. Taking expectation,

$$
\begin{aligned}
\mathbb{E}&\left[\left|\frac{\partial}{\partial x_k}\int_t^T \mathbb{E}\left[\mathcal{L}_S-\bar{\mathcal{L}}\right] v_s^0(x)ds\right|^p\right] \\
&\leq \varepsilon^p C_6(1+|z|^{q_6})\sum_{1\leq\beta\leq 3}\mathbb{E}\left[\sup_{t\leq s\leq T}\left|D_x^\beta v_s^0(x)\right|^p\right].
\end{aligned}
\tag{15}
$$

Next, by (HO$_{k,l}$), we can interchange the order of ordinary differentiation and stochastic integration (cf. Karandikar (1983)):

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{\partial}{\partial x_k}\left(\int_t^T \mathbb{E}\left[h(x,Z_s^{\varepsilon,x})-\bar{h}(x)\right]^* v_s^0(x)d\overleftarrow{B}_s\right)\right|^p\right] \\
&\qquad = \mathbb{E}\left[\left|\int_t^T \frac{\partial}{\partial x_k}\left(\mathbb{E}\left[h(x,Z_s^{\varepsilon,x})-\bar{h}(x)\right]^* v_s^0(x)\right)d\overleftarrow{B}_s\right|^p\right] \\
&\qquad \le C_p\mathbb{E}\left[\left(\int_t^T \left|\frac{\partial}{\partial x_k}\left(\mathbb{E}\left[h(x,Z_s^{\varepsilon,x})-\bar{h}(x)\right]^* v_s^0(x)\right)\right|^2 ds\right)^{p/2}\right],
\end{aligned}$$

where

$$\begin{aligned}
&\int_t^T \left|\frac{\partial}{\partial x_k}\left(\mathbb{E}\left[h(x,Z_s^{\varepsilon,x})-\bar{h}(x)\right]^* v_s^0(x)\right)\right|^2 ds \\
&= \varepsilon\int_0^{\frac{T-t}{\varepsilon}}\left|\frac{\partial}{\partial x_k}p_u(z,h-\bar{h};x)v^0_{\varepsilon u+t}(x)+p_u(z,h-\bar{h};x)\frac{\partial}{\partial x_k}v^0_{\varepsilon u+t}(x)\right|^2 du \\
&\le 2\varepsilon\left\{\int_0^\infty \left|\frac{\partial}{\partial x_k}p_u(z,h-\bar{h};x)\right|^2\left|v^0_{\varepsilon u+t}(x)\right|^2 du\right. \\
&\qquad \left.+\int_0^\infty \left|p_u(z,h-\bar{h};x)\right|^2\left|\frac{\partial}{\partial x_k}v^0_{\varepsilon u+t}(x)\right|^2 du\right\} \\
&\le \varepsilon C_7(1+|z|^{q_7})\left\{\sup_{t\le s\le T}\left|v_s^0(x)\right|^2+\sup_{t\le s\le T}\left|\frac{\partial}{\partial x_k}v_s^0(x)\right|^2\right\}.
\end{aligned}$$

The last step follows once again from Proposition 5.2, (2). So,

$$\begin{aligned}
(16)\qquad &\mathbb{E}\left[\left|\frac{\partial}{\partial x_k}\left(\int_t^T \mathbb{E}\left[h(x,Z_s^{\varepsilon,x})-\bar{h}(x)\right]^* v_s^0(x)d\overleftarrow{B}_s\right)\right|^p\right] \\
&\le \varepsilon^{\frac{p}{2}}C_8(1+|z|^{q_8})\left\{\mathbb{E}\left[\sup_{t\le s\le T}\left|v_s^0(x)\right|^p\right]+\mathbb{E}\left[\sup_{t\le s\le T}\left|\frac{\partial}{\partial x_k}v_s^0(x)\right|^p\right]\right\}.
\end{aligned}$$

Combining (15) and (16)

$$\mathbb{E}\left[\left|\frac{\partial}{\partial x_k}\theta_t^1\right|^p\right] \le \varepsilon^{\frac{p}{2}}C_9(1+|z|^{q_9})\sum_{\alpha\le 3}\mathbb{E}\left[\sup_{t\le s\le T}\left|D_x^\alpha v_s^0(x)\right|^p\right].$$

Iterating these arguments for the higher order derivatives of θ^1,

$$\sum_{|\alpha|\le k-1}\mathbb{E}\left[\left|D_x^\alpha\theta_t^1\right|^p\right] \le \varepsilon^{\frac{p}{2}}C_{10}(1+|z|^{q_{10}})\sum_{|\alpha|\le k+1}\mathbb{E}\left[\sup_{t\le s\le T}\left|D_x^\alpha v_s^0(x)\right|^p\right].$$

□

Lemma 6.3. *Let $k,l\ge 3$. Assume (HF$_{k,l}$), (HS$_{k,l}$), and (HO$_{k+1,l+1}$). Also assume $\psi^1\in C^{0,k+2,l}([0,T]\times\mathbb{R}^m\times\mathbb{R}^n,\mathbb{R})$ and that all its partial derivatives up to order $(0,k+2,l)$ are in $\mathcal{P}_T([0,T]\times\mathbb{R}^m,\mathbb{R})$. Then for any $p\ge 1$ there exists $C_p>0$, independent of φ, such that for any $(x,z)\in\mathbb{R}^{m+n}$, any $\varepsilon\in(0,1)$, and any $t\in[0,T]$*

$$\mathbb{E}\left[|R_t(x,z)|^p\right]\le C_p\sum_{|\alpha|\le 2}\int_t^T \mathbb{E}\left[\mathbb{E}\left[\left|D_x^\alpha\psi_s^1(x',z')\right|^p\right]_{(x',z')=(X_s^{\varepsilon,(t,x)},Z_s^{\varepsilon,(t,z)})}\right]ds.$$

Proof. $R_t(x,z)$ solves the BSPDE

$$-dR_t(x,z) = \left(\mathcal{L}^\varepsilon R_t(x,z) + \mathcal{L}_S\psi^1_t(x,z)\right)dt \\ + h(x,z)^*\left(\psi^1_t(x,z)+R_t(x,z)\right)d\overleftarrow{B}_t, \\ R_T(x,z) = 0. \tag{17}$$

Existence of the solution R and its derivatives, as well as the polynomial growth all follow from Proposition 4.1. By Proposition 4.2, the solution of (17) is given by $\theta_t^{(t,x,z)(2)}$, the solution to the BDSDE

$$\begin{aligned} -d\theta_s^{(t,x,z)(2)} &= \mathcal{L}_S\psi^1_s(X_s^{\varepsilon,(t,x)}, Z_s^{\varepsilon,(t,z)})ds \\ &\quad + h(X_s^{\varepsilon,(t,x)}, Z_s^{\varepsilon,(t,z)})^*\psi^1_s(X_s^{\varepsilon,(t,x)}, Z_s^{\varepsilon,(t,z)})d\overleftarrow{B}_s \\ &\quad + h(X_s^{\varepsilon,(t,x)}, Z_s^{\varepsilon,(t,z)})^*\theta_s^{(t,x,z)(2)}d\overleftarrow{B}_s - \gamma_s^{t,x,z}dW_s - \delta_s^{t,x,z}dV_s \\ \theta_T^{(t,x,z)(2)} &= 0. \end{aligned}$$

We will drop superscripts (t,x,z) for $\theta_t^{(t,x,z)(2)}$, (t,z) for $Z^{\varepsilon,(t,z)}$, and (t,x) for $X^{\varepsilon,(t,x)}$.

$R_t(x,z)$ is $\mathcal{F}^B_{t,T}$-measurable, hence, so is θ^2_t. As before, the stochastic integrals over dV and dW vanish when we take conditional expectation with respect to $\mathcal{F}^B_{t,T}$. Thus

$$\begin{aligned} \theta^2_t &= \mathbb{E}\left[\int_t^T \mathcal{L}_S\psi^1_s(X_s^\varepsilon, Z_s^\varepsilon)ds\,\middle|\,\mathcal{F}^B_{t,T}\right] \\ &\quad + \mathbb{E}\left[\int_t^T h(X_s^\varepsilon, Z_s^\varepsilon)^*\psi^1_s(X_s^\varepsilon, Z_s^\varepsilon)d\overleftarrow{B}_s\,\middle|\,\mathcal{F}^B_{t,T}\right] \\ &\quad + \mathbb{E}\left[\int_t^T h(X_s^\varepsilon, Z_s^\varepsilon)^*\theta^2_s d\overleftarrow{B}_s\,\middle|\,\mathcal{F}^B_{t,T}\right]. \end{aligned} \tag{18}$$

Consider each term separately:

$$\begin{aligned} &\mathbb{E}\left[\left|\mathbb{E}\left[\int_t^T \mathcal{L}_S\psi^1_s(X_s^\varepsilon, Z_s^\varepsilon)ds\,\middle|\,\mathcal{F}^B_{t,T}\right]\right|^p\right] \le \mathbb{E}\left[\left|\int_t^T \mathcal{L}_S\psi^1_s(X_s^\varepsilon, Z_s^\varepsilon)ds\right|^p\right] \\ &\le (T-t)^{p-1}\int_t^T \mathbb{E}\left[\left|\left(\sum_{i=1}^m b_i(X_s^\varepsilon, Z_s^\varepsilon)\frac{\partial}{\partial x_i}\right.\right.\right. \\ &\qquad \left.\left.\left. + \frac{1}{2}\sum_{i,j=1}^m (\sigma\sigma^*)_{ij}(X_s^\varepsilon, Z_s^\varepsilon)\frac{\partial^2}{\partial x_i x_i}\right)\psi^1_s(X_s^\varepsilon, Z_s^\varepsilon)\right|^p\right]ds \\ &\le C_1\int_t^T\left(||b||_\infty\sum_{i=1}^m \mathbb{E}\left[\left|\frac{\partial}{\partial x_i}\psi^1_s(X_s^\varepsilon, Z_s^\varepsilon)\right|^p\right]\right. \\ &\qquad \left. + \frac{1}{2}||\sigma\sigma^*||_\infty\sum_{i,j=1}^m \mathbb{E}\left[\left|\frac{\partial^2}{\partial x_i x_i}\psi^1_s(X_s^\varepsilon, Z_s^\varepsilon)\right|^p\right]\right)ds \\ &\le C_2\int_t^T \sum_{1\le|\alpha|\le 2}\mathbb{E}\left[\left|D_x^\alpha\psi^1_s(X_s^\varepsilon, Z_s^\varepsilon)\right|^p\right]ds. \end{aligned}$$

Note that Z_s^ε and X_s^ε are $\mathcal{F}_s^W \vee \mathcal{F}_s^V$-measurable, ψ_s^1 is $\mathcal{F}_{s,T}^B$-measurable, and B and (V, W) are independent. Thus

$$\mathbb{E}\left[\left|D_x^\alpha \psi_s^1(X_s^\varepsilon, Z_s^\varepsilon)\right|^p\right] = \mathbb{E}\left[\mathbb{E}\left[\left|D_x^\alpha \psi_s^1(X_s^\varepsilon, Z_s^\varepsilon)\right|^p\right| \mathcal{F}_s^V \vee \mathcal{F}_s^W\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\left|D_x^\alpha \psi_s^1(x', z')\right|^p\right]_{(x',z')=(X_s^\varepsilon, Z_s^\varepsilon)}\right],$$

so that

$$\mathbb{E}\left[\left|\mathbb{E}\left[\int_t^T \mathcal{L}_S \psi_s^1(X_s^\varepsilon, Z_s^\varepsilon) ds \middle| \mathcal{F}_{t,T}^B\right]\right|^p\right] \leq C_2 \sum_{1 \leq |\alpha| \leq 2} \int_t^T \mathbb{E}\left[\mathbb{E}\left[\left|D_x^\alpha \psi_s^1(x', z')\right|^p\right]_{(x'z')=(X_s^\varepsilon, Z_s^\varepsilon)}\right] ds. \tag{19}$$

Next, by Jensen's inequality, the tower property, and the Burkholder-Davis-Gundy inequality,

$$\begin{aligned}\mathbb{E}\left[\left|\mathbb{E}\left[\int_t^T h(X_s^\varepsilon, Z_s^\varepsilon)^* \psi_s^1(X_s^\varepsilon, Z_s^\varepsilon) d\overleftarrow{B}_s \middle| \mathcal{F}_{t,T}^B\right]\right|^p\right] &\leq \mathbb{E}\left[\left|\int_t^T h(X_s^\varepsilon, Z_s^\varepsilon)^* \psi_s^1(X_s^\varepsilon, Z_s^\varepsilon) d\overleftarrow{B}_s\right|^p\right] \\ &\leq C_p \mathbb{E}\left[\langle \int_t^T h(X_s^\varepsilon, Z_s^\varepsilon)^* \psi_s^1(X_s^\varepsilon, Z_s^\varepsilon) d\overleftarrow{B}_s \rangle^{\frac{p}{2}}\right],\end{aligned}$$

where by Hölder's inequality and the Cauchy-Schwarz inequality

$$\begin{aligned}\langle \int_t^T h(X_s^\varepsilon, Z_s^{\varepsilon,x})^* \psi_s^1(X_s^\varepsilon, Z_s^\varepsilon) d\overleftarrow{B}_s \rangle^{\frac{p}{2}} &= \left(\int_t^T \left|h(X_s^\varepsilon, Z_s^{\varepsilon,x})^* \psi_s^1(X_s^\varepsilon, Z_s^\varepsilon)\right|^2 ds\right)^{\frac{p}{2}} \\ &\leq C_3 \int_t^T |h(X_s^\varepsilon, Z_s^\varepsilon)|^p |\psi_s^1(X_s^\varepsilon, Z_s^\varepsilon)|^p ds.\end{aligned}$$

So by the same arguments as for the first term,

$$\mathbb{E}\left[\left|\mathbb{E}\left[\int_t^T h(X_s^\varepsilon, Z_s^\varepsilon)^* \psi_s^1(X_s^\varepsilon, Z_s^\varepsilon) d\overleftarrow{B}_s \middle| \mathcal{F}_{t,T}^B\right]\right|^p\right] \leq C_4 \int_t^T \mathbb{E}\left[\mathbb{E}\left[\left|\psi_s^1(x', z')\right|^p\right]_{(x',z')=(X_s^\varepsilon, Z_s^\varepsilon)}\right] ds. \tag{20}$$

Finally, using Burkholder-Davis-Gundy in the second line, and Cauchy-Schwarz in the third line

$$\begin{aligned}\mathbb{E}\left[\left|\mathbb{E}\left[\int_t^T h(X_s^\varepsilon, Z_s^\varepsilon)^* \theta_s^2 d\overleftarrow{B}_s \middle| \mathcal{F}_{t,T}^B\right]\right|^p\right] &\leq \mathbb{E}\left[\left|\int_t^T [h(X_s^\varepsilon, Z_s^\varepsilon)]^* \theta_s^2 d\overleftarrow{B}_s\right|^p\right] \\ &\leq C_p \mathbb{E}\left[\left(\int_t^T |h(X_s^\varepsilon, Z_s^\varepsilon)^* \theta_s^2|^2 ds\right)^{\frac{p}{2}}\right] \\ &\leq C_p \mathbb{E}\left[\left(\int_t^T |h(X_s^\varepsilon, Z_s^\varepsilon)|^2 |\theta_s^2|^2 ds\right)^{\frac{p}{2}}\right] \\ &\leq C_5 ||h||_\infty^p \int_t^T \mathbb{E}[|\theta_s^2|^p] ds. \qquad (21)\end{aligned}$$

Combining (18) with (19), (20), and (21)

$$\mathbb{E}\left[\left|\theta_t^2\right|^p\right] \le C_6 \sum_{|\alpha|\le 2} \int_t^T \mathbb{E}\left[\mathbb{E}\left[\left|D_x^\alpha \psi_s^1(x',z')\right|^p\right]_{(x',z')=(X_s^\varepsilon,Z_s^\varepsilon)}\right] ds + C_5 ||h||_\infty^p \int_t^T \mathbb{E}[|\theta_s^2|^p] ds.$$

By Gronwall,

$$\begin{aligned}\mathbb{E}\left[\left|\theta_t^2\right|^p\right] &\le C_6 \left(\sum_{|\alpha|\le 2} \int_t^T \mathbb{E}\left[\mathbb{E}\left[\left|D_x^\alpha \psi_s^1(x',z')\right|^p\right]_{(x',z')=(X_s^\varepsilon,Z_s^\varepsilon)}\right] ds\right) e^{(T-t)C_5||h||_\infty^p} \\ &\le C_7 \left(\sum_{|\alpha|\le 2} \int_t^T \mathbb{E}\left[\mathbb{E}\left[\left|D_x^\alpha \psi_s^1(x',z')\right|^p\right]_{(x',z')=(X_s^\varepsilon,Z_s^\varepsilon)}\right] ds\right).\end{aligned}$$

where we choose C_7 so that the inequality holds for every $t \in [0,T]$ (replace $e^{(T-t)C_5||h||_\infty}$ by $e^{TC_5||h||_\infty}$). □

Now we can collect all these results, to obtain the first step towards Theorem 3.1.

Lemma 6.4. *Assume (H_{stat}), ($HF_{8,4}$), ($HS_{7,4}$), ($HO_{8,4}$), and that $\varphi \in C_b^7(\mathbb{R}^m,\mathbb{R})$. Then for every $p \ge 1$ there exists $C, q_1, q_2 > 0$, independent of φ, such that*

$$\sup_{0\le t\le T} \mathbb{E}[|v_t^\varepsilon(x,z) - v_t^0(x)|^p] \le \varepsilon^{p/2} C\left(1 + |x|^{q_1} + |z|^{q_2}\right) ||\varphi||_{4,\infty}^p.$$

Proof of Theorem 3.1. We track the necessary conditions backward from Lemma 6.3.

1. For the solution R given in Lemma 6.3 to exist and satisfy the stated bound, we need ($HF_{3,3}$), ($HS_{3,3}$), ($HO_{4,4}$), and $\psi^1 \in C^{0,5,3}([0,T]\times\mathbb{R}^m\times\mathbb{R}^n,\mathbb{R})$. The polynomial growth condition will be satisfied anyways.

2. For ψ^1 to be in $C^{0,5,3}([0,T]\times\mathbb{R}^m\times\mathbb{R}^n,\mathbb{R})$, we need ($H_{stat}$), ($HF_{6,4}$), ($HS_{6,4}$), ($HO_{6,4}$) and $\bar a, \bar b, \bar h \in C_b^5$. We also need $v^0 \in C^{0,6}([0,T]\times\mathbb{R}^m,\mathbb{R})$. Again, the polynomial growth condition will be satisfied.

3. For v^0 to be in $C^{0,6}([0,T]\times\mathbb{R}^m,\mathbb{R})$ we need $\bar a, \bar b, \varphi \in C_b^7$ and $\bar h \in C_b^8$.

4. For $\bar a, \bar b$ to be in C_b^7 we need ($HF_{7,3}$) as well as ($HS_{7,0}$) by Proposition 5.4. Similarly we need ($HF_{8,3}$) as well as ($HO_{8,0}$) for $\bar h$ to be in C_b^8.

5. So sufficient conditions are (H_{stat}), ($HF_{8,4}$), ($HS_{7,4}$), ($HO_{8,4}$). In that case we obtain from Lemma 6.1

$$\sum_{|\alpha|\le 4} \mathbb{E}\left[\sup_{0\le t\le T} |D^\alpha v_t^0(x)|^p\right] \le C_1(1+|x|^{q_1})||\varphi||_{4,\infty}^p. \tag{22}$$

From Lemma 6.2 we obtain

$$\sum_{|\alpha|\le 2} \sup_{0\le t\le T} \mathbb{E}\left[|D_x^\alpha \psi_t^1(x,z)|^p\right] \le \varepsilon^{\frac{p}{2}} C_2(1+|z|^{q_2}) \sum_{|\alpha|\le 4} \mathbb{E}\left[\sup_{0\le t\le T} \left|D_x^\alpha v_t^0(x)\right|^p\right]. \tag{23}$$

From Lemma 6.3 we get

$$\mathbb{E}\left[|R_t(x,z)|^p\right] \le C_3 \sum_{|\alpha|\le 2} \int_t^T \mathbb{E}\left[\mathbb{E}\left[\left|D_x^\alpha \psi_s^1(x',z')\right|^p\right]_{(x',z')=(X_s^{\varepsilon,(t,x)},Z_s^{\varepsilon,(t,z)})}\right] ds. \tag{24}$$

Combining (22), (24), (24), we get for any $t \in [0,T]$ (by time-homogeneity of X^ε and Z^ε)

$$\begin{aligned} &\mathbb{E}\left[|R_t(x,z)|^p\right] + \mathbb{E}\left[|\psi_t^1(x,z)|^p\right] \\ &\le \varepsilon^{p/2} C_4 \left(1 + \sup_{0\le s\le T} \mathbb{E}\left[|X_s^\varepsilon|^{q_1} + |Z_s^{\varepsilon,x}|^{q_2} | (X_0^\varepsilon, Z_0^\varepsilon) = (x,z)\right]\right) ||\varphi||_{4,\infty}^p. \end{aligned} \tag{25}$$

From Proposition 5.3 we obtain

$$\sup_{0\le s\le T} \mathbb{E}\left[|X_s^\varepsilon|^{q_1} + |Z_s^{\varepsilon,x}|^{q_2} | (X_0^\varepsilon, Z_0^\varepsilon) = (x,z)\right] \le C_5(1 + |x|^{q_3} + |z|^{q_4}).$$

Noting that the right hand side in (25) does not depend on $t \in [0,T]$,

$$\begin{aligned} &\sup_{0\le t\le T} \mathbb{E}\left[|R_t(x,z)|^p\right] + \sup_{0\le t\le T} \mathbb{E}\left[|\psi_t^1(x,z)|^p\right] \\ &\qquad \le \varepsilon^{p/2} C_6 \left(1 + |x|^{q_3} + |z|^{q_4}\right) ||\varphi||_{4,\infty}^p. \end{aligned}$$

Finally

$$\begin{aligned} &\sup_{0\le t\le T} \mathbb{E}[|v_t^\varepsilon(x,z) - v_t^0(x)|^p] \\ &\qquad \le C_7 \left(\sup_{0\le t\le T} \mathbb{E}\left[|R_t(x,z)|^p\right] + \sup_{0\le t\le T} \mathbb{E}\left[|\psi_t^1(x,z)|^p\right]\right) \\ &\qquad \le \varepsilon^{p/2} C_8 \left(1 + |x|^{q_3} + |z|^{q_4}\right) ||\varphi||_{4,\infty}^p, \end{aligned}$$

which completes the proof. □

Now we recall that all the calculations up until now were under the changed measure $\mathbb{P}^\varepsilon$. We only wrote $\mathbb{P}$ and B to facilitate the reading. So let us transfer the results to the original measure $\mathbb{Q}$.

Lemma 6.5. *Assume (H_{stat}), ($HF_{8,4}$), ($HS_{7,4}$), ($HO_{8,4}$), and that $\varphi \in C_b^7(\mathbb{R}^m, \mathbb{R})$. Then for every $p \ge 1$ there exist $C, q_1, q_2 > 0$, independent of φ, such that*

$$\sup_{0\le t\le T} \mathbb{E}_{\mathbb{Q}}[|v_t^\varepsilon(x,z) - v_t^0(x)|^p] \le \varepsilon^{p/2} C \left(1 + |x|^{q_1} + |z|^{q_2}\right) ||\varphi||_{4,\infty}^p.$$

Proof. This is a simple application of the Cauchy-Schwarz inequality in combination with Gronwall's lemma:

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}}[|v_t^\varepsilon(x,z) - v_t^0(x)|^p] &= \mathbb{E}_{\mathbb{P}^\varepsilon}\left[|v_t^\varepsilon(x,z) - v_t^0(x)|^p \frac{d\mathbb{Q}}{d\mathbb{P}^\varepsilon}\right] \\ &\le \mathbb{E}_{\mathbb{P}^\varepsilon}[|v_t^\varepsilon(x,z) - v_t^0(x)|^{2p}]^{1/2} \mathbb{E}_{\mathbb{P}^\varepsilon}\left[\left(\frac{d\mathbb{Q}}{d\mathbb{P}^\varepsilon}\right)^2\right]^{1/2}, \end{aligned}$$

so we see that the result is true by Lemma 6.4 as long as the second expectation is finite. Recall that we had defined the notation

$$\left.\frac{d\mathbb{Q}}{d\mathbb{P}^\varepsilon}\right|_{\mathcal{F}_t} = \tilde{D}_t^\varepsilon = \exp\left(\int_0^t h(X_s^\varepsilon, Z_s^\varepsilon)^* dY_s^\varepsilon - \frac{1}{2}\int_0^t |h(X_s^\varepsilon, Z_s^\varepsilon)|^2 ds\right).$$

So $\tilde{D}^\varepsilon$ satisfies the SDE

$$d\tilde{D}_t^\varepsilon = \tilde{D}_t^\varepsilon h(X_t^\varepsilon, Z_t^\varepsilon)^* dY_t^\varepsilon, \qquad \tilde{D}_0^\varepsilon = 1.$$

Since under $\mathbb{P}^\varepsilon$, Y^ε is a Brownian motion, we get by Itô-isometry

$$\mathbb{E}_{\mathbb{P}^\varepsilon}\left[(\tilde{D}_t^\varepsilon)^2\right] = \mathbb{E}_{\mathbb{P}^\varepsilon}\left[\int_0^t (\tilde{D}_s^\varepsilon)^2 |h(X_s^\varepsilon, Z_s^\varepsilon)|^2 ds\right] \le ||h||_\infty^2 \mathbb{E}_{\mathbb{P}^\varepsilon}\left[\int_0^t (\tilde{D}_s^\varepsilon)^2 ds\right],$$

so that by Gronwall $\mathbb{E}_{\mathbb{P}^\varepsilon}\left[(\tilde{D}_T^\varepsilon)^2\right] < \infty$ □

Lemma 6.6. *Assume (H_{stat}), ($HF_{8,4}$), ($HS_{7,4}$), ($HO_{8,4}$), that $\varphi \in C_b^7$, and that the initial distribution $\mathbb{Q}_{(X_0^\varepsilon, Z_0^\varepsilon)}$ has finite moments of every order. Then for every $p \geq 1$ there exists $C > 0$, independent of φ, such that*

$$\mathbb{E}_{\mathbb{Q}}[|\rho_T^{\varepsilon,x}(\varphi) - \rho_T^0(\varphi)|^p] \leq \varepsilon^{p/2} C ||\varphi||_{4,\infty}^p.$$

Proof. As we already described in the introduction, we obtain from Lemma 6.5

$$\begin{aligned}
&\mathbb{E}_{\mathbb{Q}}[|\rho_T^{\varepsilon,x}(\varphi) - \rho_T^0(\varphi)|^p] \\
&\quad = \mathbb{E}_{\mathbb{Q}}\left[\left|\int (v_0^\varepsilon(x,z) - v_0^0(x)) \mathbb{Q}_{(X_0^\varepsilon, Z_0^\varepsilon)}(dx,dz)\right|^p\right] \\
&\quad \leq \int \mathbb{E}_{\mathbb{Q}}[|v_0^\varepsilon(x,z) - v_0^0(x)|^p] \mathbb{Q}_{(X_0^\varepsilon, Z_0^\varepsilon)}(dx,dz) \\
&\quad \leq \varepsilon^{p/2} C_1 \int (1 + |x|^{q_1} + |z|^{q_2}) \, \mathbb{Q}_{(X_0^\varepsilon, Z_0^\varepsilon)}(dx,dz) ||\varphi||_{4,\infty}^p \\
&\quad \leq \varepsilon^{p/2} C_2 ||\varphi||_{4,\infty}^p.
\end{aligned}$$

□

The convergence of the actual filter, i.e. of $\pi^{\varepsilon,x}$ to π^0, now follows exactly as in Chapter 9.4 of Bain and Crisan (2009). For the sake of completeness, we include the arguments.

Lemma 6.7. *Let $p \geq 1$. Then*

$$\sup_{\varepsilon \in (0,1], t \in [0,T]} \{\mathbb{E}_{\mathbb{Q}}[|\rho_t^{\varepsilon,x}(1)|^{-p}] + \mathbb{E}_{\mathbb{Q}}[|\rho_t^0(1)|^{-p}]\} < \infty$$

as long as h is bounded.

Proof. We give the argument for $\mathbb{E}_{\mathbb{Q}}[|\rho_t^{\varepsilon,x}(1)|^{-p}]$, $\mathbb{E}_{\mathbb{Q}}[|\rho_t^0(1)|^{-p}]$ being completely analogue. We have

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}[|\rho_t^{\varepsilon,x}(1)|^{-p}] &= \mathbb{E}_{\mathbb{P}^\varepsilon}\left[|\rho_t^{\varepsilon,x}(1)|^{-p} \frac{d\mathbb{Q}}{d\mathbb{P}^\varepsilon}\right] \\
&\leq \mathbb{E}_{\mathbb{P}^\varepsilon}\left[|\rho_t^{\varepsilon,x}(1)|^{-2p}\right]^{1/2} \mathbb{E}_{\mathbb{P}^\varepsilon}\left[\left(\frac{d\mathbb{Q}}{d\mathbb{P}^\varepsilon}\right)^2\right]^{1/2}
\end{aligned}$$

We showed in the proof of Lemma 6.5 that the second expectation is finite. Note that $x \mapsto x^{-2p}$ is convex. Therefore by Jensen's inequality,

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}^\varepsilon}[|\rho_t^{\varepsilon,x}(1)|^{-2p}] \\
&= \mathbb{E}_{\mathbb{P}^\varepsilon}\left[\left|\mathbb{E}_{\mathbb{P}^\varepsilon}\left[\exp\left(\int_0^t h(X_s^\varepsilon, Z_s^\varepsilon)^* dY_s^\varepsilon - \frac{1}{2}\int_0^t |\bar{h}(X_s^\varepsilon, Z_s^\varepsilon)|^2 ds\right)\middle| \mathcal{Y}_t^\varepsilon\right]\right|^{-2p}\right] \\
&\leq \mathbb{E}_{\mathbb{P}^\varepsilon}\left[\left|\exp\left(\int_0^t h(X_s^\varepsilon, Z_s^\varepsilon)^* dY_s^\varepsilon - \frac{1}{2}\int_0^t |\bar{h}(X_s^\varepsilon, Z_s^\varepsilon)|^2 ds\right)\right|^{-2p}\right] \\
&\leq \mathbb{E}_{\mathbb{P}^\varepsilon}\left[\left|\frac{d\mathbb{Q}}{d\mathbb{P}^\varepsilon}\right|^{-2p}\right] = \mathbb{E}_{\mathbb{Q}}\left[\left|\frac{d\mathbb{P}^\varepsilon}{d\mathbb{Q}}\right|^{2p+1}\right].
\end{aligned}$$

The result now follows exactly as in the proof of Lemma 6.5, because for $D_t^\varepsilon = d\mathbb{P}^\varepsilon / d\mathbb{Q}|_{\mathcal{F}_t}$ we have

$$dD_t^\varepsilon = -h(X_t^\varepsilon, Z_t^\varepsilon)^* dB_t, \qquad D_0^\varepsilon = 1$$

and B is a Brownian motion under $\mathbb{Q}$. □

Define for any measurable and bounded test function $\varphi : \mathbb{R}^m \to \mathbb{R}$

$$\pi_t^0(\varphi) = \frac{\rho_t^0(\varphi)}{\rho_t^0(1)}.$$

Recall that $\pi_t^{\varepsilon,x}$ was defined analogously with $\rho_t^{\varepsilon,x}$ instead of ρ_t^0. We then have

Lemma 6.8. *Assume (H_{stat}), ($HF_{8,4}$), ($HS_{7,4}$), ($HO_{8,4}$), and that the initial distribution $\mathbb{Q}_{(X_0^\varepsilon, Z_0^\varepsilon)}$ has finite moments of every order. Let $p \geq 1$. Then there exists $C > 0$ such that for every $\varphi \in C_b^7$*

$$\mathbb{E}_{\mathbb{Q}}[|\pi_T^{\varepsilon,x}(\varphi) - \pi_T^0(\varphi)|^p] \leq \varepsilon^{p/2} C ||\varphi||_{4,\infty}^p.$$

Proof. In the third line we use that $\pi^{\varepsilon,x}$ is a.s. equal to a probability measure.

$$\begin{aligned}
&\mathbb{E}_{\mathbb{Q}}[|\pi_T^{\varepsilon,x}(\varphi) - \pi_T^0(\varphi)|^p] \\
&= \mathbb{E}_{\mathbb{Q}}\left[\left|\frac{\rho_T^{\varepsilon,x}(\varphi)}{\rho_T^{\varepsilon,x}(1)} - \frac{\rho_T^0(\varphi)}{\rho_T^0(1)}\right|^p\right] \\
&= \mathbb{E}_{\mathbb{Q}}\left[\left|\frac{\rho_T^{\varepsilon,x}(\varphi) - \rho_T^0(\varphi)}{\rho_T^0(1)} - \pi_T^{\varepsilon,x}(\varphi)\frac{\rho_T^{\varepsilon,x}(1) - \rho_T^0(1)}{\rho_T^0(1)}\right|^p\right] \\
&\leq C_p\left(\mathbb{E}_{\mathbb{Q}}\left[\left|\frac{\rho_T^{\varepsilon,x}(\varphi) - \rho_T^0(\varphi)}{\rho_T^0(1)}\right|^p\right] + ||\varphi||_\infty^p \mathbb{E}_{\mathbb{Q}}\left[\left|\frac{\rho_T^{\varepsilon,x}(1) - \rho_T^0(1)}{\rho_T^0(1)}\right|^p\right]\right) \\
&\leq C_p\left(\mathbb{E}_{\mathbb{Q}}[|\rho_T^0(1)|^{-2p}]\right)^{1/2}\left(\mathbb{E}_{\mathbb{Q}}\left[\left|\rho_T^{\varepsilon,x}(\varphi) - \rho_T^0(\varphi)\right|^{2p}\right]^{1/2}\right. \\
&\qquad\qquad\left. + ||\varphi||_\infty^p \mathbb{E}_{\mathbb{Q}}\left[\left|\rho_T^{\varepsilon,x}(1) - \rho_T^0(1)\right|^{2p}\right]^{1/2}\right) \\
&\leq \varepsilon^{p/2} C_1 ||\varphi||_{4,\infty},
\end{aligned}$$

where the last step follows from Lemma 6.6 and Lemma 6.7. □

Since the bound only depends on $||\varphi||_{4,\infty}$, we can replace the assumption $\varphi \in C_b^7$ by $\varphi \in C_b^4$: Just approximate $\varphi \in C_b^4$ by $\varphi^n \in C_b^7$ in the $||\cdot||_{4,\infty}$-norm, and take advantage of the fact that $\pi_T^{\varepsilon,x}$ and π_T^0 are a.s. equal to probability measures. Therefore we have

Corollary 6.9. *Assume (H_{stat}), ($HF_{8,4}$), ($HS_{7,4}$), ($HO_{8,4}$), and that the initial distribution $\mathbb{Q}_{(X_0^\varepsilon, Z_0^\varepsilon)}$ has finite moments of every order. Let $p \geq 1$. Then there exists $C > 0$ such that for every $\varphi \in C_b^4$,*

$$\mathbb{E}_{\mathbb{Q}}[|\pi_T^{\varepsilon,x}(\varphi) - \pi_T^0(\varphi)|^p] \leq \varepsilon^{p/2} C ||\varphi||_{4,\infty}^p.$$

Now note that there exists a countable algebra $(\varphi_i)_{i\in\mathbb{N}}$ of C_b^4 functions that strongly separates points in $\mathbb{R}^m$. That is, for every $x \in \mathbb{R}^m$ and $\delta > 0$, there exists $i \in \mathbb{N}$, such that $\inf_{y:|x-y|>\delta} |\varphi_i(x) - \varphi_i(y)| > 0$. Take e.g. all functions of the form

$$\exp\left(-\sum_{j=1}^n q_j (x - x_j)^2\right)$$

with $n \in \mathbb{N}$, $q_j \in \mathbb{Q}_+$, $x_j \in \mathbb{Q}^m$. By Theorem 3.4.5 of Ethier and Kurtz (1986), the sequence (φ_i) is convergence determining for the topology of weak convergence of probability measures. That is, if μ_n and μ are probability measures on $\mathbb{R}^m$, such that $\lim_{n\to\infty} \mu_n(\varphi_i) = \mu(\varphi_i)$ for every $i \in \mathbb{N}$, then μ_n converges weakly to μ.

Define the following metric on the space of probability measures on $\mathbb{R}^m$:

$$d(\nu, \mu) = d_{(\varphi_i)}(\nu, \mu) = \sum_{i=1}^\infty \frac{|\nu(\varphi_i) - \mu(\varphi_i)|}{2^i}.$$

Because (φ_i) is convergence determining, the metric d generates the topology of weak convergence. Therefore the proof of Theorem 3.1 is complete. □

7. Conclusion and future directions

This paper presented the theoretical basis for the development of a lower-dimensional particle filtering algorithm for the state estimation in complex multiscale systems. To this end, we combined stochastic homogenization with nonlinear filtering theory to construct a homogenized SPDE which is the approximation of a lower-dimesional nonlinear filter for the "coarse-grained" process. The convergence of the optimal filter of the "coarse-grained" process to the solution of the homogenized filter is shown using BSDEs and asymptotic techniques. This homogenized SPDE can be used as the basis for an efficient multi-scale particle filtering algorithm for estimating the slow dynamics of the system, without directly accounting for the fast dynamics. In Lingala et al. (2012) we present a numerical algorithm based on this scheme, that enables efficient incorporation of observation data for estimation of the coarse-grained ("slow") dynamics, and we apply the algorithm to a high-dimensional chaotic multiscale system.

Even though this paper deals with just one widely separated characteristic time scale, one can extend this work to incorporate a more realistic setting where the signal has more than one time scale separation. As before we let ε be a small parameter that measures the ratio of slow and fast time scales. Consider the signal and observation processes governed by:

$$
\begin{aligned}
dZ_t^\varepsilon &= \frac{1}{\varepsilon^2} f(Z_t^\varepsilon, X_t^\varepsilon) + \frac{1}{\varepsilon} g(Z_t^\varepsilon, X_t^\varepsilon)\, dW_t, \quad Z_0^\varepsilon = z, \\
dX_t^\varepsilon &= \frac{1}{\varepsilon} b^I(Z_t^\varepsilon, X_t^\varepsilon) + b(Z_t^\varepsilon, X_t^\varepsilon) + \sigma(Z_t^\varepsilon, X_t^\varepsilon)\, dV_t, \quad X_0^\varepsilon = x, \\
dY_t^\varepsilon &= h(Z_t^\varepsilon, X_t^\varepsilon)\, dt + dB_t, \quad Y_0^\varepsilon = 0,
\end{aligned}
\tag{26}
$$

where W, V and B are independent Wiener processes and x and z are random initial conditions which are independent of W, V and B. It is important to realize that there are several scales in (26), even the slow process X_t^ε has a fast varying component. This case is important, in particular, for applications in geophysical flows and climate dynamics. The drift term b and the diffusion σ cause fluctuations of order order 1, and the drift term f and the diffusion g cause fluctuations of order order ε^{-2}, whereas the drift term b^I causes fluctuations at an intermediate order ε^{-1}. It was found that when the average of b^I with respect to the invariant measure of the fast component Z_t^ε (for the fixed slow component) is zero, the limit distribution of the slow component (away from the initial layer) can also be obtained in terms of the solution of some auxiliary Poisson equation in the homogenization theory. However, a *unified framework* to deal with ε^{-1} term in developing a lower-dimensional nonlinear filter for the "coarse-grained" process is still not available.

Our conditions on the coefficients are very restrictive and exclude for example linear models. This is due to the fact that we are using homogenization of SPDEs to obtain convergence of the filter, and that for existence of solutions to the SPDEs, the coefficients need to be bounded and sufficiently smooth. Working with weak solutions in place of classical solutions would not improve the conditions much. Using viscosity solutions or entirely relying on probabilistic arguments might be a way to get less restrictive conditions, however with these methods we do not expect that a rate of convergence can be obtained.

While we were able to obtain the explicit rate of convergence $\sqrt{\varepsilon}$, the constant C in Theorem 3.1 depends on the terminal time T. It would be interesting to find conditions under which this can be avoided. This might be achieved by building on stability results for nonlinear filters, see e.g. Crisan and Rozovskii (2011), Chapter 4, "Stability and asymptotic analysis".

Acknowledgement. We wish to express our gratitude to the anonymous referee for their careful reading of the manuscript and for their detailed comments which improved the presentation of this paper.

N.S. Namachchivaya and H.C. Yeong are supported by the National Science Foundation under grants number CMMI 10-30144 and EFRI 10-24772 and by AFOSR under grant number FA9550-08-1-0206. N. Perkowski is supported by a Ph.D. scholarship of the Berlin Mathematical School. Part of this research was carried out while P. Imkeller and N. Perkowski were visiting the Department of Aerospace Engineering of University of Illinois at Urbana-Champaign. They are grateful for the hospitality at UIUC. The visit of P. Imkeller was funded by NSF grant number CMMI 10-30144. The visit of N. Perkowski was funded by NSF grant number EFRI 10-24772 and by the Berlin Mathematical School. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.*, 50(2):174–188, 2002.

A. Bain and D. Crisan. *Fundamentals of Stochastic Filtering.* Springer, Berlin, 2009.

A. Bensoussan and G. L. Blankenship. Nonlinear filtering with homogenization. *Stochastics*, 17: 67–90, 1986.

A. Bensoussan, J.-L. Lions, and G. Papanicolaou. *Asymptotic Analysis for Periodic Structures.* North-Holland, Amsterdam, 1978.

D. Crisan and B. L. Rozovskii. *The Oxford Handbook of Nonlinear Filtering.* Oxford University Press, Oxford, 2011.

A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice.* Springer, Berlin, 2001.

S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence.* John Wiley & Sons, Hoboken, NJ, 1986.

M. Fujisaki, G. Kallianpur, and H. Kunita. Stochastic differential equations for the non linear filtering problem. *Osaka J. Math.*, 9(1):19–40, 1972.

J. Harlim and E. L. Kang. Filtering partially observed multiscale systems with heterogeneous multiscale methods based reduced climate models. *Mon. Wea. Rev.*, 140(3):860–873, 2012.

N. Ichihara. Homogenization problem for stochastic partial differential equations of Zakai type. *Stochastics Stochastics Rep.*, 76(3):243–266, 2004.

G. Kallianpur. *Stochastic Filtering Theory.* Springer, Berlin, 1980.

R. L. Karandikar. Interchanging the order of stochastic integration and ordinary differentiation. *Sankhyā*, 45(1):120–124, 1983.

M. L. Kleptsina, R. S. Liptser, and A. Serebrovski. Nonlinear filtering problem with contamination. *Ann. Appl. Probab.*, 7(4):917–934, 1997.

H. J. Kushner. *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems.* Birkhäuser, Boston, 1990.

N. Lingala, N. S. Namachchivaya, N. Perkowski, and H. C. Yeong. Particle filtering in high-dimensional chaotic systems. Chaos, to appear; arXiv:1204.1360, 2012.

R. S. Liptser and A. N. Shiryaev. *Statistics of Random Processes: Applications.* Springer, Berlin, 2001.

G. C. Papanicolaou and W. Kohler. Asymptotic analysis of deterministic and stochastic equations with rapidly varying components. *Comm. Math. Phys.*, 45:217–232, 1975.

G. C. Papanicolaou, D. Stroock, and S. R. S. Varadhan. Martingale approach to some limit theorems. In *Papers from the Duke Turbulence Conference (Duke Univ., Durham, N.C., 1976), Paper No. 6.* Duke Univ., Durham, N.C., 1977.

E. Pardoux. Stochastic partial differential equations and filtering of diffusion processes. *Stochastics*, 3:127–167, 1979.

E. Pardoux and S. Peng. Backward doubly stochastic differential equations and systems of quasilinear spdes. *Probab. Theory Related Fields*, 98:209–227, 1994.

E. Pardoux and A. Y. Veretennikov. On the Poisson equation and diffusion approximation. *Ann. Probab.*, (3):1061–1085.

E. Pardoux and A. Y. Veretennikov. On Poisson equation and diffusion approximation 2. *Ann. Probab.*, 31(3):1166–1192, 2003.

J. Park, R. B. Sowers, and N. S. Namachchivaya. Dimensional reduction in nonlinear filtering. *Nonlinearity*, 23(2):305–324, 2010.

J. Park, N. S. Namachchivaya, and H. C. Yeong. Particle filters in a multiscale environment: Homogenized hybrid particle filter. *J. Appl. Mech.*, 78(6):1–10, 2011.

B. L. Rozovskii. *Stochastic Evolution System: Linear Theory and Aplications to Non-linear Filtering.* Kluwer Academic Publishers, Dordrecht, 1990.

C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson. Obstacles to high-dimensional particle filtering. *Mon. Weather Rev.*, 136(12):4629–4640, 2008.

D. W. Stroock. *Partial Differential Equations for Probabilists.* Cambridge University Press, Cambridge, 2008.

A. Y. Veretennikov. On polynomial mixing bounds for stochastic differential equations. *Stochastic Process. Appl.*, 70(1):115–127, 1997.

M. Zakai. On the optimal filtering of diffusion processes. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 11(3):230–243, 1969.

[1]Institut für Mathematik, Humboldt-Universität zu Berlin, Rudower Chaussee 25, 12489 Berlin, Germany

E-mail address: imkeller@math.hu-berlin.de, perkowsk@math.hu-berlin.de

[2]Department of Aerospace Engineering, University of Illinois at Urbana-Champaign, 306 Talbot Laboratory, MC-236, 104 South Wright Street, Urbana, Illinois 61801, USA

E-mail address: navam@illinois.edu, hyeong2@illinois.edu