Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia IUTAM 5 (2012) 34 – 45

Procedia IUTAM

www.elsevier.com/locate/procedia

IUTAM Symposium on 50 Years of Chaos: Applied and Theoretical

A homogenization approach to multiscale filtering

Peter Imkeller[a], N. Sri Namachchivaya[b*], Nicolas Perkowski[a], Hoong C. Yeong[b]

[a]Institut für Mathematik, Humbolt-Universität zu Berlin, Rudower Chaussee 25, 12489 Berlin, Germany
[b]Department of Aerospace Engineering, University of Illinois at Urbana-Champaign, 306 Talbot Laboratory, 104 S. Wright St., Urbana, IL 61801, USA

Abstract

We present a homogenized nonlinear filter for multi-timescale systems, which allows the reduction of the dimension of filtering equation. We prove that the actual nonlinear filter converges to our homogenized filter. This is achieved by a suitable asymptotic expansion of the dual of the Zakai equation, and probabilistically representing the correction terms with the help of backward doubly-stochastic differential equations. This homogenized filter provides a rigorous mathematical basis for the development of reduced-dimension nonlinear filters for multiscale systems. A filtering scheme, based on the homogenized filtering equation and the technique of importance sampling, is applied to a chaotic multiscale system in Lingala et al. [1].

Keywords: nonlinear filtering; dimensional reduction; homogenization; particle filter; asymptotic expansion; SPDE; BDSDE

1. Introduction

An important aspect in the study of random dynamical systems is the estimation of state variables, which are often hidden, based on available observations. The optimal estimate based on observation data, called the *filter*, is given by the conditional expectation that can be generated by a recursive equation *driven by the observation process*. Formally, the filter is a conditional expectation $\pi_t(\varphi) \overset{\text{def}}{=} \mathbb{E}[\varphi(X_t)|\mathcal{Y}_t]$ for a rich enough class of functions φ, where $\{X_t: t \geq 0\}$ is the partially observed state (or signal) and $\{\mathcal{Y}_t: t \geq 0\}$ is the (filtration of) observation process that is a function of the signal corrupted by noise. The filtering problem involves characterizing the optimal filter π_t, which is accomplished in Zakai [2] and

* Corresponding author. Tel.: +1-217-448-0683; fax: +1-217-244-0720.
E-mail address: navam@illinois.edu.

2210-9838 © 2012 Published by Elsevier Ltd. Selection and/or Peer-review under responsibility of Takashi Hikihara and Tsutomu Kambe
Open access under CC BY-NC-ND license. doi:10.1016/j.piutam.2012.06.005

Fujisaki et al. [3] through the evolution of the conditional distribution in the space of probability measures (see, for example, Bain and Crisan [4], Kallianpur [5], or Liptser and Shiryaev [6] for details).

Another aspect of dynamical systems that is of interest in this paper is multiscale (multiple timescales) dynamics that are inherent in a wide range of scientific studies and engineering applications. For example, climate evolution is governed by atmospheric (fast) and oceanic (slow) dynamics, and state dynamics in electric power systems consists of rapidly- and slowly-varying elements. Nonlinearities of the physical processes in multiscale phenomena allow energy transfer between different time scales, resulting in complex behavior. The main challenge is to recognize how information interacts within these complex structures and scales. In this paper, we are interested in the multiscale filtering problem, which we address by taking advantage of scale interaction to appropriately reduce the dimensions of the problem.

For the implementation of the optimal filter in applications, the particle filter is a well-established method for nonlinear systems (see, for example, Doucet et al. [7] and Arulampalam et al. [8] for comprehensive insight). However, for implementation in high-dimensional systems, dimensionality issues arise when trying to represent the signal density using a high number of particles (see, for example, Snyder et al. [9]). This is the motivation for addressing the multiscale (and high-dimensional) filtering problem. We have established a rigorous theoretical basis for the development of reduced-dimension nonlinear filtering methods for multiscale systems.

The main result can be summarized as follows. We assume the signal is given as solution of the two time scale stochastic differential equations (SDEs)

$$dX_t^\varepsilon = b(X_t^\varepsilon, Z_t^\varepsilon)dt + \sigma(X_t^\varepsilon, Z_t^\varepsilon)dW_t, \quad X_0^\varepsilon = \xi \in \mathbb{R}^m \tag{1}$$

$$dZ_t^\varepsilon = \frac{1}{\varepsilon} f(X_t^\varepsilon, Z_t^\varepsilon)dt + \frac{1}{\sqrt{\varepsilon}} g(X_t^\varepsilon, Z_t^\varepsilon)dV_t, \quad Z_0^\varepsilon = \eta \in \mathbb{R}^n. \tag{2}$$

Here, $\varepsilon \ll 1$ is the timescale separation parameter, so X^ε is the slow component and Z^ε is the fast component. W and V are, respectively, l- and k-dimensional independent standard Brownian motions, independent of the random initial conditions ξ and η. We assume that for every fixed *x*, the solution Z^x of (2) 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 homogenized diffusion process X^0 governed by an SDE

$$dX_t^0 = \bar{b}(X_t^0)dt + \bar{\sigma}(X_t^\varepsilon)dW_t, \quad X_0^0 = \xi \in \mathbb{R}^m, \tag{3}$$

for appropriately averaged $\bar{b}$ and $\bar{\sigma}$. In other words, a stochastically averaged model provides a qualitatively useful approximation to the actual multiscale system. It is usually the state of the slow, or "coarse-grained", process that is of concern in the study of multiscale phenomena, so it is a good idea to make use of this homogenized X^0 for estimation. Specifically, X^0 can be used to construct an averaged, or homogenized, filter π^0 to approximate the *x*-marginal, $\pi^{\varepsilon,x}$, of the optimal filter π^ε.

The main result is that under the assumptions stated in Section 4.3, there exists a metric *d* on the space of probability measures such that for every $T > 0$ there exists $C > 0$ such that

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

In other words, the *x*-marginal of the optimal filter π^ε converges to the homogenized filter π^0 in the space of probability measures as the time scale separation between the fast and slow components becomes infinitely large (separation parameter $\varepsilon \to 0$). Hence, the homogenized filter is an appropriate measure to use in place of the actual $\pi^{\varepsilon,x}$ for estimating the "coarse-grained" dynamics X^ε in a setting with wide timescale separation.

In terms of filtering applications, π^0 presents the advantage of not requiring exact knowledge of the fast dynamics for estimating the "coarse-grained" dynamics. Only knowledge of the invariant measure of Z^x is required (discussed in sections that follow) so, by applying appropriate multiscale homogenization numerical schemes, computation and information storage for the fast dynamics can be reduced. By combining the result presented here with the importance sampling framework, a particle filtering scheme for state estimation of a large scale chaotic multiscale system is formulated in Lingala et al. [1].

2. Formulation of multi-scale nonlinear filtering problems

Let $(\Omega, \mathcal{F}, \mathcal{F}^t, \mathbb{Q})$ be a filtered probability space that supports a standard Brownian motion (V, W, B). Let the signal $(X_t^\varepsilon, Z_t^\varepsilon)$ be a two time scale diffusion process governed by the SDEs (1) and (2). Functions f, g, b, and σ are assumed to be Borel-measurable. For fixed $x \in \mathbb{R}^m$, define Z^x as the dynamics of Z^ε with $X^\varepsilon = x$, i.e. the fast component dynamics with the slow component held constant. We assume that for all $x \in \mathbb{R}^m$, Z^x is ergodic and converges rapidly towards its stationary measure $p_\infty(x,\cdot)$. The d-dimensional observation is given by

$$Y_t^\varepsilon = \int_0^t h(X_t^\varepsilon, Z_t^\varepsilon)dt + B_t$$

with Borel-measurable h. B is assumed to be a d-dimensional standard Brownian motion, independent of W and V. Define the σ-algebra generated by the observation as $\mathcal{Y}_t^\varepsilon \stackrel{\text{def}}{=} \sigma\{Y_s^\varepsilon : 0 \le s \le t\} \vee \mathcal{N}$, where $\mathcal{N}$ are the $\mathbb{Q}$-negligible sets.

The aim of the filtering problem is to calculate the (normalized) filter $\pi_t^\varepsilon(\varphi) \stackrel{\text{def}}{=} \mathbb{E}_{\mathbb{Q}}[\varphi(X_t^\varepsilon, Z_t^\varepsilon)|\mathcal{Y}_t^\varepsilon]$, where π^ε is a finite measure on $\mathbb{R}^{m+n}$ and φ is a bounded measurable function on $\mathbb{R}^{m+n}$. Define $\mathbb{P}^\varepsilon$ to be a new measure related to $\mathbb{Q}$ by the Girsanov transform, which removes the observation process drift,

$$D_t^\varepsilon \stackrel{\text{def}}{=} \frac{d\mathbb{P}^\varepsilon}{d\mathbb{Q}} = exp\left(-\int_0^t h(X_s^\varepsilon, Z_s^\varepsilon)\, dB_s - \frac{1}{2}\int_0^t |h(X_s^\varepsilon, Z_s^\varepsilon)|^2 ds\right). \quad (4)$$

Under $\mathbb{P}^\varepsilon$, the observation process Y^ε is a Brownian motion independent of $(X_t^\varepsilon, Z_t^\varepsilon)$. π^ε is related to the un-normalized filter, $\rho_t^\varepsilon(\varphi) \stackrel{\text{def}}{=} \mathbb{E}_{\mathbb{P}^\varepsilon}[\varphi(X_t^\varepsilon, Z_t^\varepsilon)(D_t^\varepsilon)^{-1}|\mathcal{Y}_t^\varepsilon]$, by the Girsanov transform as follows:

$$\pi_t^\varepsilon(\varphi) = \mathbb{E}_{\mathbb{Q}}[\varphi(X_t^\varepsilon, Z_t^\varepsilon)|\mathcal{Y}_t^\varepsilon] = \frac{\mathbb{E}_{\mathbb{P}^\varepsilon}\left[\varphi(X_t^\varepsilon, Z_t^\varepsilon)(D_t^\varepsilon)^{-1}|\mathcal{Y}_t^\varepsilon\right]}{\mathbb{E}_{\mathbb{P}^\varepsilon}\left[(D_t^\varepsilon)^{-1}|\mathcal{Y}_t^\varepsilon\right]} = \frac{\rho_t^\varepsilon(\varphi)}{\rho_t^\varepsilon(1)}.$$

The un-normalized filter ρ^ε satisfies the Zakai equation (see, for example, Bain and Crisan [4]):

$$d\rho_t^\varepsilon(\varphi) = \rho_t^\varepsilon(\mathcal{L}^\varepsilon\varphi)dt + \rho_t^\varepsilon(h\varphi)dY_t^\varepsilon, \qquad \rho_0^\varepsilon(\varphi) = \mathbb{E}_{\mathbb{Q}}[\varphi(X_0^\varepsilon, Z_0^\varepsilon)].$$

Here, $\mathcal{L}^\varepsilon = \frac{1}{\varepsilon}\mathcal{L}_F + \mathcal{L}_S$ is the differential operator associated to $(X_t^\varepsilon, Z_t^\varepsilon)$, with

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

where $\cdot^*$ denotes the transpose of a matrix or vector. Denote $\rho^{\varepsilon,x}$ as the x-marginal of ρ^ε.

The theory of stochastic averaging (see, for example, Papanicolaou et al. [10]) 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 of the form (3). So, as long as we are only interested in *estimating the slow component*, i.e. the "coarse-grained" dynamics, we want to take advantage of this fact. Specifically, we want to find a homogenized (un-normalized) filter ρ^0 that satisfies

$$d\rho_t^0(\varphi) = \rho_t^0(\bar{\mathcal{L}}\varphi)dt + \rho_t^0(\bar{h}\varphi)dY_t^\varepsilon, \qquad \rho_0^0(\varphi) = \mathbb{E}_{\mathbb{Q}}[\varphi(X_0^0)].$$

such that for small ε, the x-marginal of ρ^ε, $\rho^{\varepsilon,x}$, is close to ρ^0. We let the generator $\bar{\mathcal{L}}$ of X^0 be defined as $\mathcal{L}_S$, but with coefficients $\bar{b}(x) = \int b(x,z)p_\infty(x,dz)$ and $\bar{a}(x) = \int \sigma\sigma^*(x,z)p_\infty(x,dz)$, i.e. the drift and diffusion coefficients of (1) averaged with respect to the stationary measure of Z^x. Also, define $\bar{h}$

similarly for the sensor function h. Note that the homogenized filter is still driven by the *real observation* Y^{ε}, not by a "homogenized observation" Y^0. This is practical for implementation of the homogenized filter in applications since such Y^0 is usually not available. Even if Y^0 was available, using it would lead to loss of information for estimating the signal compared to using the actual observation.

Define the homogenized and optimal x-marginal filters π^0 and $\pi^{\varepsilon,x}$ in terms of ρ^0 and $\rho^{\varepsilon,x}$ as π^{ε} was in terms of ρ^{ε} . This paper presents the convergence result of the actual filter to the averaged filter: For every $T \geq 0$, there exists $C > 0$ such that

$$\mathbb{E}_{\mathbb{Q}}[d(\pi_T^{\varepsilon,x}, \pi_T^0)] \leq \sqrt{\varepsilon} C, \quad \text{i.e.} \quad \lim_{\varepsilon\to 0} \mathbb{E}_{\mathbb{Q}}[d(\pi_T^{\varepsilon,x}, \pi_T^0)] = 0 \quad \text{for any} \quad T > 0, \tag{5}$$

where d denotes a suitable distance on the space of probability measures that generates the topology of weak convergence. Park et al. [11] shows this convergence result for a two-dimensional multiscale signal process with no drift in the fast component SDE. The results presented here are extensions to an $\mathbb{R}^{m+n}$-dimensional signal process with drift and diffusion coefficients of the fast and slow components SDEs dependent on both components. The proof of Park et al. [11] 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.

In order to show the desired convergence, we follow Pardoux [12] in introducing the dual representations of $\rho_T^{\varepsilon,x}(\varphi)$ and $\rho_T^0(\varphi)$:

$$v_t^{\varepsilon,T,\varphi}(x,z) = \mathbb{E}_{\mathbb{P}_{t,x,z}^{\varepsilon}}\left[\varphi(X_T^{\varepsilon})\left(D_{t,T}^{\varepsilon}\right)^{-1}\middle|\mathcal{Y}_{t,T}^{\varepsilon}\right], \quad v_t^{0,T,\varphi}(x) = \mathbb{E}_{\mathbb{P}_{t,x,z}^{\varepsilon}}\left[\varphi(X_T^0)\left(D_{t,T}^0\right)^{-1}\middle|\mathcal{Y}_{t,T}^{\varepsilon}\right].$$

$\mathbb{P}_{t,x,z}^{\varepsilon}$ is the measure under which $(X^{\varepsilon}, Z^{\varepsilon})$ and X^0 are governed by the same dynamics as under $\mathbb{P}^{\varepsilon}$, but $(X^{\varepsilon}, Z^{\varepsilon})$ and X^0 stays in (x, z) and x until time t. $D_{t,T}^{\varepsilon}$ is defined as the Girsanov transform (4) but with limits of integration t to T while $D_{t,T}^0$ is the Girsanov transform using the averaged sensor function, i.e.

$$D_{t,T}^0 \stackrel{\text{def}}{=} exp\left(-\int_t^T \bar{h}(X_s^0)\, dY_s^{\varepsilon} + \frac{1}{2}\int_t^T \left|\bar{h}(X_s^0)\right|^2 ds\right).$$

$D_{t,T}^{\varepsilon}$ provides the change of measure to $\mathbb{P}_{t,x,z}^{\varepsilon}$, under which Y^{ε} is a Brownian, for the diffusion process that started at (t, x, z). $\mathcal{Y}_{t,T}^{\varepsilon} \stackrel{\text{def}}{=} \sigma\{Y_s^{\varepsilon} - Y_t^{\varepsilon}: t \leq s \leq T\} \vee \mathcal{N}$ is the filtration generated by the observation over $[t, T]$, minus the observation history up to t.

From the Markov property of $(X^{\varepsilon}, Z^{\varepsilon}, X^0)$, it follows that

$$\rho_T^{\varepsilon,x}(\varphi) = \int v_0^{\varepsilon,T,\varphi}(x,z)\mathbb{Q}_{(X_0^{\varepsilon},Z_0^{\varepsilon})}(dx,dz), \quad \rho_T^0(\varphi) = \int v_0^{0,T,\varphi}(x)\mathbb{Q}_{X_0^0}(dx). \tag{6}$$

Note that the homogenized and un-homogenized processes have the same starting distribution, i.e. $\int_{\mathbb{R}^n} \mathbb{Q}_{X_0^{\varepsilon},Z_0^{\varepsilon}}(dz) = \mathbb{Q}_{X_0^0}$. Now fix T and $\varphi \in C_b^2(\mathbb{R}^m, \mathbb{R})$ and write $v_t^{\varepsilon} = v_t^{\varepsilon,T,\varphi}$. In introducing the dual process, we are representing the conditional expectation $\rho_T^{\varepsilon,x}$ at time T by a conditional expectation v_t^{ε} that is run backwards in time from $t = T$ to $t = 0$. To construct v_t^{ε}, fix a starting point (x, z) at $t = 0$ and determine all possible trajectories $(X_T^{\varepsilon}, Z_T^{\varepsilon})$. Then, the quantity v_t^{ε} is ran backwards in time from all the possible $(X_T^{\varepsilon}, Z_T^{\varepsilon})$ that started from (x, z) to obtain v_0^{ε}. Integrating v_0^{ε} over $\mathbb{Q}_{X_0^{\varepsilon},Z_0^{\varepsilon}}$ indicates averaging over all possible starting points (x, z), hence giving ρ_T^{ε}. Similarly for the representation $v_t^0 = v_t^{0,T,\varphi}$ of ρ_T^0.

Using (6), we obtain

$$\mathbb{E}[|\rho_T^{\varepsilon,x}(\varphi) - \rho_T^0(\varphi)|^p] \leq \int \mathbb{E}[|v_0^{\varepsilon}(x,z) - v_0^0(x)|^p]\mathbb{Q}_{(X_0^{\varepsilon},Z_0^{\varepsilon})}(dx,dz). \tag{7}$$

So, $\mathbb{E}[|\rho_T^{\varepsilon,x}(\varphi) - \rho_T^0(\varphi)|^p]$ will also be small as long as $\mathbb{Q}(X_0^{\varepsilon}, Z_0^{\varepsilon})$ is well behaved. Then, (7) will lead to (5). Therefore, the goal now is to show that for nice test functions φ, the quantity $\left|v_t^{\varepsilon}(x,z) - v_t^0(x)\right|^p$ is small. The reason for the introduction of the dual processes is that they solve backward stochastic partial differential equations (BSPDEs), which are function-valued rather than measure-valued and depend on the generators $\mathcal{L}$ and $\bar{\mathcal{L}}$ instead of their adjoints (see Section 3).

3. Formal expansions of the filtering equations

For large parts of the remainder of this paper, we will only work with $\mathbb{P}^\varepsilon$, and under $\mathbb{P}^\varepsilon$, Y^ε is a Brownian motion that 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.

The key point is that ν^ε and ν^0 satisfy BSPDEs ($d\overleftarrow{B}$ denotes the backward Itô integral):

$$-d\nu_t^\varepsilon(\varphi) = \mathcal{L}^\varepsilon \nu_t^\varepsilon(x,z)dt + h(x,z)\nu_t^\varepsilon(x,z)d\overleftarrow{B}_t, \qquad \nu_T^\varepsilon(x,z) = \varphi(x),$$
$$-d\nu_t^0(\varphi) = \bar{\mathcal{L}}\nu_t^0(x)dt + \bar{h}(x)\nu_t^0(x)d\overleftarrow{B}_t, \qquad \nu_T^0(x) = \varphi(x), \tag{8}$$

We formally expand ν^ε as

$$\nu_t^\varepsilon = \underbrace{u_t^0(x,z)}_{\nu_t^0(x)} + \varepsilon\, \underbrace{u_{t/\varepsilon}^1(x,z)}_{\psi_t(x,z)} + \varepsilon^2\, \underbrace{u_{t/\varepsilon}^2(x,z)}_{R(x,z)}.$$

Note that technically, this does not make sense. Firstly, if $\varepsilon \to 0$, then $t/\varepsilon \to \infty$, so we cannot definitively set terminal conditions for the equations defined so far. Secondly, the stochastic integrals $\int_t^T u_{s/\varepsilon}^i(x,z)d\overleftarrow{B}_t$, $i = 1,2$, a priori do not make sense for $\varepsilon < 1$ since u^i's are adapted to $\mathcal{Y}_{t,T}^B$ but $s/\varepsilon \to \infty$ as $\varepsilon \to 0$. However, we perform such an expansion formally. Then, ν^0 satisfies (8) and ψ and R satisfy

$$-d\psi_t(x,z) = \frac{1}{\varepsilon}\mathcal{L}_F\psi_t(x,z)dt + (\mathcal{L}_S - \bar{\mathcal{L}})\nu_t^0(x)dt + \left(h(x,z) - \bar{h}(x)\right)^* \nu_t^0(x)d\overleftarrow{B}_t, \tag{9}$$
$$-dR_t(x,z) = \mathcal{L}^\varepsilon R_t(x,z)dt + \mathcal{L}_S\psi_t(x,z)dt + h(x,z)^*(\psi_t(x,z) + R_t(x,z))d\overleftarrow{B}_t, \tag{10}$$

respectively, with terminal conditions $\psi_T(x,z) = R_T(x,z) = 0$. By the existence and uniqueness of the solutions to these *linear* equations, we can apply superposition so that indeed,

$$\nu_t^\varepsilon(x,z) = \nu_t^0(x) + \psi_t(x,z) + R_t(x,z).$$

Thus, ν^0 is the first order approximation of ν^ε, ψ is the correction due to the first order approximation, and R represents the remaining higher order terms. Therefore, the problem of showing L^p-convergence of ν^ε to ν^0 reduces to showing L^p-convergence of the corrector plus remainder, $(\psi + R)$, to 0.

4. Statement of the main result

4.1. Assumptions

We adopt some smoothness assumptions for proving our results. For a bounded function $\theta: \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^d$ with partial derivatives in (x,z) up to order (k,l) bounded, write $\theta \in C_b^{k,l}(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^d)$. For the coefficients of (1) and (2) and the sensor function h, we assume that, for $k,l \geq 0$ to be specified later,

$$f \in C_b^{k,l}(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^n),\ b \in C_b^{k,l}(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^n),$$
$$g \in C_b^{k,l}(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^{n\times k}),\ \sigma \in C_b^{k,l}(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^{m\times l}), \text{ and } h \in C_b^{k,l}(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^d).$$

For the existence and uniqueness of the stationary distribution $p_\infty(x,dz)$, we also assume that the vector field f pulls z strongly enough back to the origin and gg^* is uniformly elliptic: There exists $M_0 > 0$, $\alpha > 0$ such that $\sup_x\langle f(x,z), z\rangle \leq -C|z|^\alpha$ for all $|z| \geq M_0$, and there are $0 < \lambda \leq \Lambda < \infty$ such that $\lambda\mathbb{I}_{n\times n} \leq gg^*(x,z) \leq \Lambda\mathbb{I}_{n\times n}$.

4.2. Notations

We introduce some notations for the operations that follow:

- For averaging over $p_t(z,z';x)$, the density of Z^x that started at z at time 0, we introduce the notations $p_\infty(\theta;x) \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} \theta(x,z)p_\infty(x,dz)$ and $p_t(z,\theta;x) \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} \theta(x,z')p_t(z,z';x)dz$.
- For $(\alpha_1,\dots,\alpha_m) \in \mathbb{N}^m$, $|\alpha| = \alpha_1 + \cdots + \alpha_m$, we define the differential operator $D^\alpha \stackrel{\text{def}}{=} \frac{\partial^{|\alpha|}}{\partial x_{\alpha_1}\dots\partial x_{\alpha_m}}$.
- For $\theta \in C_b^k(\mathbb{R}^m,\mathbb{R})$, we define the norm $\|\theta\|_{k,\infty} \stackrel{\text{def}}{=} \sum_{|\alpha|\le k}\|D^\alpha\theta\|_\infty$, $\|\cdot\|_\infty$ is the usual sup norm.

4.3. Statement of main result

Theorem 4.1. Assume the conditions in Section 4.1 and that the initial distribution $\mathbb{Q}_{(X_0^\varepsilon,Z_0^\varepsilon)}$ has finite moments of every order. Then for every $p \ge 1$ and $T \ge 0$, there exists $C > 0$, such that for every $\varphi \in C_b^4$,

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

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

$$\mathbb{E}_{\mathbb{Q}}[d(\pi_t^{\varepsilon,x},\pi_t^\varepsilon)] \le \sqrt{\varepsilon}C.$$

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.

Our method of proof is as follows: We represent the BSPDEs (8), (9) and (10) by finite-dimensional stochastic equations (these will be BDSDEs). The diffusion operators get replaced by the associated diffusions and we are able to give explicit estimates of the finite dimensional equations in terms of the transition function of the fast diffusion. Pardoux and Veretennikov [13] proved very precise estimates for this transition function. Applying those estimates allows us to obtain the desired convergence of $(\psi + R)$ to 0 and the convergence of Theorem 4.1 results from the convergence of ν^ε to ν^0.

While the ideas are simple, the precise formulation and the actual proofs are quite technical. The methods and results are described concisely in the following sections. The precise statements and detailed proofs are presented in the longer, more rigorous version of the homogenization results in Imkeller et al. [14]. The study of homogenization in the nonlinear filtering problem framework is also done by Bensoussan et al. [15] and Ichihara [16] using similar approach.

5. Proof of the main result

5.1. Probabilistic representation of SPDEs

A general form of the BSPDEs (8) and (9) can be written as

$$-d\phi(\omega,t,x) = \mathcal{L}\phi(\omega,t,x)dt + f(\omega,t,x)dt + \big(g(\omega,t,x) + G(\omega,t,x)\big)\phi(\omega,t,x)d\overleftarrow{B}_t, \qquad (11)$$
$$\phi(T,x) = \varphi(x),$$

where $(B_t\colon t \in [0,T])$ is a d-dimensional standard Brownian motion under a measure $\mathbb{P}$ and $\mathcal{L}$ is a differential operator. Let us fix $t \ge 0$ and a starting point at x at t. $\phi(\omega,t,x)$ runs *backwards* in time and is driven by B_t that generates a filtration $\mathcal{F}_{t,T}^B$. Simultaneously, it is acted upon by the diffusion operator $\mathcal{L}$ with an associated *forward* diffusion process driven by a Brownian motion, say W that generates a filtration $\mathcal{F}_t^W$. Hence, roughly, the probabilistic representation of the SPDE (11) should thus involve an equation that is "doubly-stochastic". For our proof of the main result, we represent equations of the form (11) in terms of BDSDEs as introduced by Pardoux and Peng [17]. Define a BDSDE as follows:

$$\begin{aligned} -dY_s^{t,x} &= f\left(s, X_s^{t,x}\right)ds + \left(g\left(s, X_s^{t,x}\right)ds + G\left(s, X_s^{t,x}\right)Y_s^{t,x}\right)d\overleftarrow{B}_t - Z_s^{t,x}dW_s, \\ Y_T^{t,x} &= \varphi\left(X_T^{t,x}\right), \end{aligned} \tag{12}$$

where $X_s^{t,x}$ is a diffusion process, driven by Brownian motion W, associated with the generator $\mathcal{L}$. The superscript indicates that $X^{t,x}$ stays at x up to time t. The solution (Y_t, Z_t) will be $(\mathcal{F}_{t,T}^B \vee \mathcal{F}_t^W)$-measurable. This gives a finite-dimensional probabilistic representation for (11). In particular, the unique classical solution of (11) is given by $\phi(t,x) = Y^{t,x}$, where $(Y^{t,x}, Z^{t,x})$ is the unique solution of (12) (The case we consider is not completely covered by Pardoux and Peng [17] since we have unbounded, random coefficients and we do not assume a smooth diffusion coefficient. We do not present the existence and uniqueness proof of this representation here, but it is given in Section 4 of Imkeller et al. [14]). Such BDSDE representation of SPDEs allows us to apply Gronwall's lemma in the proof of our main results.

5.2. Preliminary estimate

We state a preliminary estimate from a result from Pardoux and Veretennikov [13] in Proposition 5.1 and moment bounds for X^ε and Z^ε in Proposition 5.2:

Proposition 5.1. Let $\theta \in C^{k,0}(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R})$ and assume the conditions in Section 4.1 with $l = 3$. If θ is centered, i.e. $\int_{\mathbb{R}^n} \theta(x,z)p_\infty(x;dz) = 0$ for all $x \in \mathbb{R}^m$, then for every $q > 0$, there exist C_1, $q_1 > 0$, such that

$$\sum_{|\alpha|\le k} \sum_{|\beta|\le 1} \int_0^\infty \sup_x \left|D_z^\beta D_x^\alpha p_t(z,\theta;x)\right|^q dt \le C_1(1 + |z|^{q_1}) \quad \text{for every } z \in \mathbb{R}^n. \blacksquare$$

The proof is a simple application of results from Pardoux and Veretennikov [13] (presented in Proposition 5.2 of Imkeller et al. [14]). Proposition 5.1 states that for a centered function on $\mathbb{R}^m \times \mathbb{R}^n$, its average (and derivatives of the average) with respect to the transition function of the fast process, over all time, grows at most polynomially in z. Solution of the BDSDE representation of ψ is expressed in terms of the transition function of Z^x, hence Proposition 5.1 allows us to obtain precise estimates for them.

Proposition 5.2. Assume the conditions in Section 4.1 and that the coefficients of (1) and (2) are bounded and globally Lipschitz continuous. Then for any $p \ge 1$ and $T > 0$, there exists $C_p, C(p,T), q > 0$ such that

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

Through a rescaling of time for the diffusion Z^ε, the moment bound for Z^ε is obtained as in Lemma 1 of Veretennikov [18] (presented in Proposition 5.3 of Imkeller et al. [14]). The moment bound on X^ε follows from the boundedness of coefficients of (1). Proposition 5.2 gives polynomial growth estimate on the moments of X^ε and Z^ε. The moments are with respect to $\mathbb{P}_{0,x,z}^\varepsilon$, the measure of the diffusion process that started at $(0,x,z)$, and they arise in estimating the remainder R.

5.3. Proof of the main result

Based on the formal expansion of v^ε in Section 3, the first task in proving the main result of Theorem 4.1 is to prove that the corrector, ψ, and the remainder, R, $\to 0$ as $\varepsilon \to 0$. This is achieved by representing the solutions ψ_t, R_t of BSPDEs (9), (10) in terms of BDSDEs as described in Section 5.1. The BDSDE solutions are obtained in terms of the transition function of the diffusion process, so we can apply the estimates from Section 5.3. This gives estimates of ψ and R that are appropriately bounded by ε such that $\psi, R \to 0$ as $\varepsilon \to 0$. Using these estimates, we obtain an $\mathcal{O}(\varepsilon^{p/2})$ estimate on $\left|\nu^{\varepsilon} - \nu^{0}\right|^{p}$, which can be used to track back to obtain the estimate for the main result.

From (9) and (10), we see that ψ is forced by ν^0, and R is forced by ψ, hence we first need an estimate on ν^0. We present bounds on ν^0 and its derivatives in terms of the test function φ in Lemma 5.1. The convergence rates for ψ and R are then given in terms of ν^0 and its derivatives in Lemmas 5.2 and 5.3.

Lemma 5.1. Let $k \geq 2$ and assume $\bar{b}, \bar{a}, \varphi \in C_b^{k+1}$, and $\bar{h} \in C_b^{k+2}$. Then, for any $p \geq 1$, there exist $C_p, q > 0$, such that for all $x \in \mathbb{R}^m$:

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

Proof. By the existence and uniqueness of solutions to BSPDEs of the type (11), we can also obtain polynomial growth on derivatives of the solutions. Lemma 5.1 follows by noting that equation (8) for ν^0 is of the type (11) with $f = g = 0$ and $G = \bar{h}$. □

Lemma 5.2. Let $k, l \geq 2$. Assume the conditions in Section 4.1. Also assume $\nu^0 \in C^{0,k+1}([0,T] \times \mathbb{R}^m, \mathbb{R})$ and that all its partial derivatives in x up to order $k+1$ grow at most polynomially. Finally assume that $\bar{b}, \bar{a}, \bar{h} \in C_b^{k+1}$. Then, for any $p \geq 1$, there exist $C_p, q > 0$, such that for any $(x, z) \in \mathbb{R}^{m+n}$ and any $\varepsilon \in (0,1)$:

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

Proof. $\psi_t(x,z)$ solves the BSPDE (9), which is of the form (11). By the probabilistic representation of SPDEs described in Section 5.1, the solution of (9) is given by $\psi_t(x,z) = \theta_t^{t,x,z\,(1)}$, where

$$\begin{aligned}-d\theta_s^{t,x,z\,(1)} = {} & \left(\mathcal{L}_S\left(x, Z_s^{\varepsilon,x,(t,z)}\right) - \bar{\mathcal{L}}\right)\nu_s^0(x) + \left(h\left(x, Z_s^{\varepsilon,x,(t,z)}\right) - \bar{h}(x)\right)^{*} \nu_s^0(x) d\overleftarrow{B}_s \\ & -\gamma_s^{t,x,z} dW_s, \end{aligned} \tag{13}$$

with $\theta_T^{t,x,z\,(1)} = 0$. Here, $Z_s^{\varepsilon,t,x(t,z)}$, $s \in [t,T]$, is the solution of the SDE (2) that remains at z up to time t. For brevity of notations, we will write θ_s^1 for $\theta_s^{t,x,z\,(1)}$, and $Z_s^{\varepsilon,x}$ for $Z_s^{\varepsilon,t,x(t,z)}$.

$\psi_t(x,z)$ that solves the BSPDE (9) runs backwards in time and is driven by B_t, so $\psi_t(x,z)$ is $\mathcal{F}_{t,T}^B$-measurable (i.e. given the filtration generated by B over $[t,T]$, we know the statistics of the random quantity $\psi_s(x,z)$, $s \in [t,T]$). Hence, so is θ_t^1. Then, the conditional expectation of θ_t^1 (conditioned on $\mathcal{F}_{t,T}^B$) is the quantity itself, i.e. $\theta_t^1 = \mathbb{E}\,[\theta_t^1 | \mathcal{F}_{t,T}^B]$. Taking advantage of this fact allows us to eliminate the stochastic integral term over dW. Since W and B are independent, therefore W is a standard Brownian motion in the larger filtration $(\mathcal{F}_s^W \vee \mathcal{F}_{t,T}^B : s \in [0,T])$. So, when we take the conditional expectation with respect to $\mathcal{F}_{t,T}^B$ of the solution to the BDSDE (13), the stochastic integral over dW vanishes by the tower property. Conditional expectation of the solution of the BDSDE of θ_s^1 is

$$\mathbb{E}\left[\theta_s^1 \middle| \mathcal{F}_{t,T}^B\right] = \mathbb{E}\left[\int_t^T (\mathcal{L}_S - \bar{\mathcal{L}})\nu_s^0(x) ds \middle| \mathcal{F}_{t,T}^B\right]^{(a)} + \mathbb{E}\left[\int_t^T \left(h\left(x, Z_s^{\varepsilon,x,(t,z)}\right) - \bar{h}(x)\right)^{*} \nu_s^0(x) d\overleftarrow{B}_s \middle| \mathcal{F}_{t,T}^B\right]^{(b)}. \tag{13}$$

We can first interchange the order of integration over time and averaging over the conditional density for the first term (a) in (13). ν_s^0 is the solution of SPDE (8), which is driven by B, so ν_s^0 is $\mathcal{F}_{s,T}^B$-measurable. The random quantity $Z_s^{\varepsilon,x}$ in the diffusion operator $\mathcal{L}_S$ is independent of B, so the conditional expectation becomes expectation over the density $p_{(s-t)/\varepsilon}\,(z, z'; x)$. The fast component dynamics is of order ε^{-1}, so we perform a time-shift $u = (s-t)/\varepsilon$ for the time integral. We now have the (a) in terms of the transition function of $Z_s^{\varepsilon,x}$, so the first-order derivative term in (a) of (13) becomes:

$$\varepsilon \int_0^{\frac{T-t}{\varepsilon}} \sum_{i=1}^m p_u(z, b_i - p_\infty(b_i;x);x) \frac{\partial}{\partial x_i} v^0_{\varepsilon u+t}(x)\, du \le \varepsilon C_1(1+|z|^{q_1}) \sum_{i=1}^m \sup_{t\le s\le T} \left|\frac{\partial}{\partial x_i} v_s^0(x)\right|.$$

The polynomial growth estimate is by Proposition 5.1, since $(b - p_\infty(b;x))$ is centered. Repeating the same procedure for the second derivative term, we have the following estimate for (a):

$$\mathbb{E}\left[\int_t^T (\mathcal{L}_s - \bar{\mathcal{L}}) v_s^0(x) ds \,\middle|\, \mathcal{F}^B_{t,T}\right] \le \varepsilon C_2(1+|z|^{q_2}) \mathbb{E}\left[\sum_{i=1}^m \sup_{t\le s\le T}\left|\frac{\partial}{\partial x_i} v_s^0(x)\right| + \sum_{i,j=1}^m \sup_{t\le s\le T}\left|\frac{\partial^2}{\partial x_i \partial x_j} v_s^0(x)\right|\right]. \quad (14)$$

We can again interchange order of integration for (b) and using the same arguments as for (a), arrive at an expression in terms of $p_{(s-t)/\varepsilon}(z,z';x)$. Consider the p^{th}-moment of (b):

$$\mathbb{E}\left[\left|\int_t^T p_{(s-t)/\varepsilon}\left(z, h-\bar{h};x\right)^* v_s^0(x) d\overleftarrow{B}_s\right|^p\right] \le C_p \mathbb{E}\left[\langle \int_t^T p_{(s-t)/\varepsilon}\left(z, h-\bar{h};x\right)^* v_s^0(x) d\overleftarrow{B}_s \rangle^{p/2}\right] \le \varepsilon^{p/2} C_3(1+|z|^{q_3}) \mathbb{E}[\sup_{t\le s\le T} |v_s^0(x)|^p]. \quad (15)$$

The first inequality is by the Burkholder-Davis-Gundy inequality. The second is by applying Proposition 5.1 after rewriting the quadratic variation as a time integral and performing a time-shift as for (a). Combining (14) and (15) gives the p^{th}-moment estimate for θ_s^1 :

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

The estimate for the first order derivative of θ_s^1 is obtained by first interchanging the order of ordinary differentiation and integration twice and then applying the same procedure as for θ_s^1. Note that by the assumption on h in Section 4.1, we can also interchange the order of ordinary differentiation and stochastic integration (see, for example, Karandikar [19]). We then obtain a polynomial growth estimate on the first order derivative of θ^1. Iterating the procedure for the higher order derivatives gives us

$$\sum_{|\alpha|\le k-1} \mathbb{E}[|D_x^\alpha \theta_t^1|^p] \le \varepsilon^{p/2} C_5(1-|z|^{q_5}) \sum_{|\alpha|\le k+1} \mathbb{E}[\sup_{t\le s\le T} |D_x^\alpha v_s^0(x)|^p]. \square$$

Lemma 5.3. Let $k, l \ge 3$. Assume the conditions in Section 4.1. 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 in x up to order $(0, k+2, l)$ grow at most polynomially. Then, for any $p \ge 1$, there exist $C_p, q > 0$, such that for any $(x,z) \in \mathbb{R}^{m+n}$, $\varepsilon \in (0,1)$ and $t \in [0,T]$:

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

Proof. Again, by the probabilistic representation of SPDEs, the solution $R_t(x,z)$ of (10) is given by $\theta_t^{t,x,z\,(2)}$, the solution to the BDSDE

$$-d\theta_s^{t,x,z\,(2)} = \mathcal{L}_s \psi_s^1\left(X_s^{\varepsilon,(t,x)}, Z_s^{\varepsilon,x,(t,z)}\right) ds + h\left(X_s^{\varepsilon,(t,x)}, Z_s^{\varepsilon,x,(t,z)}\right)^* \left(\psi_s^1\left(X_s^{\varepsilon,(t,x)}, Z_s^{\varepsilon,x,(t,z)}\right) + \theta_s^{t,x,z\,(2)}\right) d\overleftarrow{B}_s - \gamma_s^{t,x,z} dW_s - \delta_s^{t,x,z} dV_s,$$
$$\theta_T^{t,x,z\,(2)} = 0.$$

We will write θ_s^2 for $\theta_s^{t,x,z\,(2)}$ and $Z_s^{\varepsilon,x}$ for $Z_s^{\varepsilon,t,x(t,z)}$ as before.

By the same arguments as for $\psi_t(x,z)$, $R_t(x,z)$ is $\mathcal{F}^B_{t,T}$-measurable, hence so is θ_t^2 and the stochastic integrals over dV and dW vanish when we take conditional expectation with respect to $\mathcal{F}^B_{t,T}$ as in the proof of Lemma 5.2. Write θ_t^2 as a conditional expectation and consider the p^{th}-moment:

$$\mathbb{E}[|\theta_t^2|^p] = \mathbb{E}\left[\left|\mathbb{E}\left[\int_t^T \mathcal{L}_S\psi_s^1\left(X_s^{\varepsilon,(t,x)}, Z_s^{\varepsilon,x,(t,z)}\right)ds\,|\mathcal{F}_{t,T}^B\right]\right|^p\right]^{(a)}$$
$$+\mathbb{E}\left[\left|\mathbb{E}\left[\int_t^T h\left(X_s^{\varepsilon,(t,x)}, Z_s^{\varepsilon,x,(t,z)}\right)^* \psi_s^1\left(X_s^{\varepsilon,(t,x)}, Z_s^{\varepsilon,x,(t,z)}\right)d\overleftarrow{B}_s\,\left|\mathcal{F}_{t,T}^B\right.\right]\right|^p\right]^{(b)}$$
$$+\mathbb{E}\left[\left|\mathbb{E}\left[\int_t^T h\left(X_s^{\varepsilon,(t,x)}, Z_s^{\varepsilon,x,(t,z)}\right)^* \theta_s^2 d\overleftarrow{B}_s\,\left|\mathcal{F}_{t,T}^B\right.\right]\right|^p\right]^{(c)}. \quad (16)$$

We consider each term in (16) separately. For (a), note that the unconditioned quantity is at least equal to the conditional expectation. So we will just consider the p^{th}-moment of the time integral in (a). Application of Hölder's inequality allows us to transfer the p-exponentiation onto the integrand and then interchange the order of time integration and expectation. Coefficients b, σ of $\mathcal{L}_S$ are bounded by their respective ∞-norms by the boundedness assumptions in Section 4.1, therefore,

$$\mathbb{E}\left[\left|\mathbb{E}\left[\int_t^T \mathcal{L}_S\psi_s^1(X_s^\varepsilon, Z_s^\varepsilon)ds\,|\mathcal{F}_{t,T}^B\right]\right|^p\right] \le \mathbb{E}\left[\left|\int_t^T \mathcal{L}_S\psi_s^1(X_s^\varepsilon, Z_s^\varepsilon)\right|^p\right]$$
$$\le C_1\int_t^T\left(\|b\|_\infty \sum_{i=1}^m \mathbb{E}\left[\left|\partial_{x_i}\psi_s^1(X_s^\varepsilon, Z_s^\varepsilon)\right|^p\right] + \frac{1}{2}\|\sigma\sigma^*\|_\infty \sum_{i,j=1}^m \mathbb{E}\left[\left|\partial_{x_i}\partial_{x_j}\psi_s^1(X_s^\varepsilon, Z_s^\varepsilon)\right|^p\right]\right)ds$$
$$\le C_2\sum_{1\le|\alpha|\le 2}\int_t^T \mathbb{E}[|D_x^\alpha\psi_s^1(x',z')|^p]ds. \quad (17)$$

For (b), Jensen's inequality allows us to interchange the p-exponentiation and conditional expectation. Application of the tower property leaves us with the p^{th}-moment of the stochastic integral. We can then apply the Burkholder-Davis-Gundy inequality to get

$$\mathbb{E}\left[\left|\mathbb{E}\left[\int_t^T h(X_s^\varepsilon, Z_s^\varepsilon)^*\psi_s^1(X_s^\varepsilon, Z_s^\varepsilon)d\overleftarrow{B}_s\,\left|\mathcal{F}_{t,T}^B\right.\right]\right|^p\right] \le C_p\mathbb{E}\left[\langle\int_t^T h(X_s^\varepsilon, Z_s^\varepsilon)^*\psi_s^1(X_s^\varepsilon, Z_s^\varepsilon)\rangle^{\frac{p}{2}}\right]$$
$$\le C_3\|h\|_\infty^p\int_t^T \mathbb{E}[|\psi_s^1(X_s^\varepsilon, Z_s^\varepsilon)|^p]ds. \quad (18)$$

The second inequality results from applying Hölder's inequality to transfer the p-exponentiation into the time integral, which allows us to interchange the order of integrations, as for (a).
By the same steps, we have, for (c):

$$\mathbb{E}\left[\left|\mathbb{E}\left[\int_t^T h\left(X_s^{\varepsilon,(t,x)}, Z_s^{\varepsilon,x,(t,z)}\right)^*\theta_s^2 d\overleftarrow{B}_s\,\left|\mathcal{F}_{t,T}^B\right.\right]\right|^p\right] \le C_4\left\|h\right\|_\infty^p\int_t^T \mathbb{E}[|\theta_s^2|^p]ds. \quad (19)$$

We make a remark on the growth of the p^{th}-moments of ψ^1 and its derivatives, which emerge in (17) and (18). Note that X_s^ε, Z_s^ε are $\mathcal{F}_s^W$-measurable, ψ_s^1 is $\mathcal{F}_{s,T}^B$-measurable, and B and W are independent. Then, we can write the expected value in (17) as

$$\mathbb{E}[|D_x^\alpha\psi_s^1(x',z')|^p] = \mathbb{E}\left[\mathbb{E}[|D_x^\alpha\psi_s^1(x',z')|^p|\mathcal{F}_s^W]\right] = \mathbb{E}\left[\mathbb{E}[|D_x^\alpha\psi_s^1(x',z')|^p]_{(x',z')=(X_s^\varepsilon,Z_s^\varepsilon)}\right]. \quad (20)$$

The inner expectation in (20) is the p^{th}-moment of an $|\alpha|^{th}$-order derivative of $\psi_s^1(X_s^\varepsilon, Z_s^\varepsilon)$, which grows at most polynomially in X_s^ε and Z_s^ε, by Lemmas 5.2 and 5.1. Taking the outer expectation leads to the p^{th}-moments of X_s^ε and Z_s^ε, which also grow at most exponentially by Proposition 5.2. Therefore, by (20), we have a polynomial growth bound on (a) in x and z, and similarly for (b).

Now, combining (17), (18) and (19), we get the following estimate for θ_s^2:

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

So by Gronwall's lemma, since the first term on the right hand side is non-decreasing,

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

where we choose C_7 such that the inequality holds for every $t \in [0,T]$. By the same time-shift as in Lemma 5.3 and the moments bound argument for expected values of the form (20), we get that the p^{th}-moment of θ_t^2 grows at most $\mathcal{O}(\varepsilon)$ polynomially in x and z. □

Combining Lemmas 5.1, 5.2 and 5.3, we can obtain a convergence result for ν^ε to ν^0. Note that all the calculations so far are under the changed measure $\mathbb{P}^\varepsilon$, but we will transfer the results for ν^ε to the original measure $\mathbb{Q}_{(X_0^\varepsilon, Z_0^\varepsilon)}$. Backtracking from Lemma 5.3, we obtain that under the conditions in Section 4.1 with $(k,l) \geq (7,4)$ and with $\varphi \in C_b^7(\mathbb{R}^m, \mathbb{R})$, we have that for any $p \geq 1$, there exist $C_p, q_1, q_2 > 0$ such that

$$sup_{0\leq t\leq T}\, \mathbb{E}_{\mathbb{Q}}[|\nu_t^\varepsilon(x,z) - \nu_t^0(x)|^p] \leq \varepsilon^{\frac{p}{2}} C_p(1 + |x|^{q_1}|z|^{q_2})\|\varphi\|_{4,\infty}^p. \quad (20)$$

The estimate with respect to the original measure $\mathbb{Q}$ is obtained by an application of the Cauchy-Schwarz inequality in combination with Gronwall's lemma. Combining (20) and (7), we obtain for ρ^ε:

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

The convergence of the actual filter, i.e. of $\pi^{\varepsilon,x}$ to π^0 using the above estimate now follows exactly as in Chapter 9.4 of Bain and Crisan [3].

6. Conclusion

We have presented the theoretical basis for the development of a reduced-dimension nonlinear filtering algorithm for state estimation in multiscale systems. To this end, we combined stochastic homogenization with nonlinear filtering theory to construct a homogenized SPDE that characterizes a reduced-dimension (homogenized) nonlinear filter for the “coarse-grained” process. Convergence of the optimal filter of the slow process to the solution of the homogenized SPDE is shown using BSDEs and asymptotic techniques. The extended version of this paper that contains more detailed statements and extensive proofs is presented in Imkeller et al. [14].

The homogenized SPDE presented here can be used as the basis for an efficient multiscale nonlinear filtering algorithm for estimating the slow dynamics of the system, without directly accounting for the fast dynamics. Such filtering scheme will indirectly address the dimensionality issues mentioned in Section 1, by avoiding the need for a sample size large enough to represent both the fast and slow components. A basic homogenized particle filter based on the results shown here is presented in Park et al. [20]. A version of the homogenized filter adapted to state estimation for chaotic systems is presented in Lingala et al. [1], where the Lorenz '96 model (see Lorenz [21]), with two time-scale simplified ordinary differential equations describing advection, is considered as a nontrivial example of an atmospheric dynamics model. Lingala et al. [1] estimated the self-contained description of the coarse-grained dynamics without fully resolving the dynamics described in the fast scale. The results from several data assimilation experiments on the Lorenz '96 model are also discussed.

Acknowledgements

N. Sri Namachchivaya and Hoong C. Yeong are supported by the National Science Foundation (NSF) under grants number CMMI 10-30144 and EFRI 10-24772 and by the AFOSR under grant number FA9550-08-1-0206. Nicolas Perkowski is supported by a Ph.D. scholarship of the Berlin Mathematical School. Part of this research was carried out while Peter Imkeller and Nicolas Perkowski were visiting the Department of Aerospace Engineering of the University of Illinois at Urbana-Champaign. They are grateful for the hospitality at UIUC. The visit of Peter Imkeller was funded by NSF grant number CMMI 10-30144. The visit of Nicolas Perkowski was funded by NSF grant number EFRI 10-24772 and 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 NSF or the AFOSR.

References

[1] Lingala N, Namachchivaya N Sri, Perkowski N, Yeong HC. Particle filtering in high-dimensional chaotic systems. *Submitted* 2012.

[2] Zakai M. On Optimal filtering of Diffusion Processes. *Z. Wahrscheinlichkeitstheorie verw. Geb.* 1969; **11**(3): 230 -243.

[3] Fujisaki M, Kallianpur G, Kunita H. Stochastic Differential Equations for the Non Linear Filtering Problem. *Osaka J. Math.* 1972; **9**(1):19-40.

[4] Bain A, Crisan D. *Fundamentals of Stochastic Filtering*. New York: Springer Verlag; 2009.

[5] Kalliapur G. *Stochastic Filtering Theory*. New York: Springer-Verlag; 1980.

[6] Liptser RS, Shiryaev N. *Statistics of Random Processes: Applications*. Springer; 2001.

[7] Doucet A, de Freitas N, Gordon N. Sequential Monte Carlo Methods in Practice. New York: Springer Verlag; 2001.

[8] Arulampalam MS, Maskell S, Gordon N, Clapp T.A Tutorial on particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE Trans. Signal Process. 2002*; **50**(2): 174-188.

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

[10] Papanicolaou G, Stroock D, Varadhan SRS. Martingale Approach to Some Limit Theorems. *Statistical Mehcanics and Dynamical Systems, and papers from 1976 Duke Turbulence Conference* 1977; Duke University Mathematics Series.

[11] Park J, Sowers R, Namachcivaya N Sri. Dimensional reduction in nonlinear filtering. *Nonlinearity* 2010: **23**(2): 305-324.

[12] Pardoux E. Stochastic partial differential equations and filtering of diffusion processes. *Stochastics* 1979; **3**:127-167.

[13] Pardoux E, Veretennikov AY. On Poisson equation and diffusion approximation 2. *Ann. Probab.* 2003; **31**(3): 1166-1192.

[14] Imkeller P, Namachchivaya N Sri, Perkowski N, Yeong HC. Dimensional reduction in nonlinear filtering: A homogenization approach. *Submitted for publication* 2012.

[15] Bensoussan A, Blankenship GL. Nonlinear filtering with homogenization. *Stochastics* 1986; **17**:67-90.

[16] Ichihara N. Homogenization problem for stochastic partial differential equations of Zakai type. *Stochastics Rep.* 2004; **76**(3): 243-266.

[17] Pardoux E, Peng S. Backward Doubly Stochastic Differential Equations and Systems of Quasilinear SPDEs. *Prob. Theory Related Fields* 1994; **98**: 209 -227.

[18] Veretennikov AY. On Polynomial Mixing Bounds for Stochastic Differential Equations. *Stochastic Process. Appl. 1997;* **70**(1): 115-127.

[19] Karandikar RL. Interchanging the Order of Stochastic Integration and Ordinary Differentiation. *Sankhya* 1983; **45**(1): 120-124.

[20] Park J, Namachchivaya N Sri, Yeong HC. Particle Filters in a Multiscale Environment: Homogenized Hybrid Particle Filter. *J. Applied Mech.* 2011; **78**(6): 1-10.

[21] Lorenz EN. Predictability: A problem partly solved. *ECMWF Seminar Proceedings on Predictability* 1995.