Quantitative Convergence of the Filter Solution for Multiple Timescale Nonlinear Systems with Coarse-Grain Correlated Noise

Ryne Beeson∗, N. Sri Namachchivaya†, and Nicolas Perkowski‡

2022/01/19

Abstract

In this paper we prove a rate of convergence for the continuous time filtering solution of a multiple timescale correlated nonlinear system to a lower dimensional filtering equation in the limit of large timescale separation. Correlation is assumed to occur between the slow signal and observation processes. Convergence is almost sure in the weak topology. An asymptotic expansion of the dual process for the solution to the Zakai equation, and probabilistic representation using backward doubly stochastic differential equations is leveraged to prove the result.

1 Introduction

In this paper we prove a rate of convergence for the continuous time filtering solution of a multiple timescale and correlated nonlinear system to a lower dimensional filtering equation. The coupled system of stochastic differential equations (SDEs) that we consider is as follows,

\begin{align}
    dX^\epsilon_t &= b(X^\epsilon_t, Z^\epsilon_t)dt + \sigma(X^\epsilon_t, Z^\epsilon_t)dW_t, \\
    dZ^\epsilon_t &= \frac{1}{\epsilon^2} f(X^\epsilon_t, Z^\epsilon_t)dt + \frac{1}{\epsilon} g(X^\epsilon_t, Z^\epsilon_t)dV_t.
\end{align}

(1.1)

We denote the infinitesimal generator of \((X^\epsilon, Z^\epsilon)\) as \(G^\epsilon\). The process \((X^\epsilon, Z^\epsilon)\) is known as the signal process and \(\epsilon \in (0, 1)\) is a timescale parameter such that \(Z^\epsilon\) is a fast process and \(X^\epsilon\) is a slow process. In filtering theory, we consider the signal process to be non-observable, and instead have indirect measurements of \((X^\epsilon, Z^\epsilon)\) via the noisy observation process,

\begin{align}
    dY^\epsilon_t &= h(X^\epsilon_t, Z^\epsilon_t)dt + \alpha dW_t + \gamma dU_t.
\end{align}

We assume \(W, V, U\) are independent Brownian motions, and the presence of \(\mathbb{R}^{d \times w} \ni \alpha \neq 0\) indicates correlation between the observation and slow (coarse-grain) process. The goal in filtering theory is then to calculate the conditional distribution of \((X^\epsilon, Z^\epsilon)\) given the observation history generated from \(Y^\epsilon\), which we denote by \(\pi^\epsilon\). At each time \(t > 0\), \(\pi^\epsilon_t\) is a random probability measure on the space \(\mathbb{R}^m \times \mathbb{R}^n\) and acts on test functions \(\varphi : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}\) by integration \(\pi^\epsilon_t(\varphi) = \int \varphi(x, z) \pi^\epsilon_t(dx, dz)\).

The question of this paper is then motivated by the well known result in homogenization of stochastic differential equations that if for every fixed \(x\), the solution \(Z^\varphi\) of

\begin{align}
    dZ^\varphi_t &= f(x, Z^\varphi_t)dt + g(x, Z^\varphi_t)dV_t,
\end{align}

is ergodic with stationary distribution \(\mu_\infty(x)\), then under appropriate assumptions, the process \(X^\epsilon\) converges in distribution to a Markov process \(X^0\) with infinitesimal generator \(\overline{G}_S\) in the limit as \(\epsilon \to 0\) [PSV76; PV03; KY05]. Therefore, if we are only interested in statistics of \(X^\epsilon\) (i.e., estimation of test functions \(\varphi : \mathbb{R}^m \to \mathbb{R}\)),

∗Princeton University
†University of Waterloo
‡Freie Universität Berlin
then it would be computationally advantageous to know if $\pi^{t,x} \Rightarrow \pi^0$ converges weakly to a lower dimensional filtering equation; $\pi^0_t$ being a random probability measure for each time $t$ on $\mathbb{R}^m$ and $\pi^{t,x}$ being the $x$-marginal of $\pi^t$.

Filtering theory has widespread applications in many fields including various disciplines of engineering for decision and control systems, the geosciences, weather and climate prediction. In many of these fields, it is not uncommon to have physics based models with multiple timescales as seen in Eq. 1.1, and also have the case were estimation of the slow process is solely of interest; for example the estimation of the ocean temperature, which is necessary for climate prediction, but the ocean model may also be coupled to a fast atmospheric model. Knowing that mathematically $\pi^{t,x} \Rightarrow \pi^0$ in the limit as $\epsilon \to 0$, enables practitioners to devise more efficient methods for estimation of the slow process without great loss of accuracy (see for instance [PNY11; KH12; BH14; Yeo+20]).

There are several papers providing results for $\pi^{t,x} \rightarrow \pi^0$ (or the associated unnormalized conditional measure or density versions) on variations of the multiple timescale filtering problem. In [PSN10], $(X^\epsilon, Z^\epsilon)$ is a two dimensional process with no drift in the fast component, no intermediate scale, and no correlation. The authors made use of a representation of the slow component by a time-changed Brownian motion under a suitable measure to yield weak convergence of the filter. Homogenization of the nonlinear filter was studied in [BB86] and [Ich04] by way of asymptotic analysis on a dual representation of the nonlinear filtering equation. In these papers, the coefficients of the signal processes are assumed to be periodic. The approach in [Ich04] is novel as the first application of backward stochastic differential equations for homogenization of Zakai-type stochastic partial differential equations (SPDEs).

Convergence of the filter for a random ordinary differential equation with intermediate timescale and perturbed by a fast Markov process was investigated in [LH03]. A two timescale problem with correlation between the slow process and observation process, but where the slow dispersion coefficient does not depend on the fast process, is investigated in [Qia19]. The main result is that the filter converges in $L^1$ sense to the lower dimensional filter. An energy method approach is used in [ZR19] to show that the probability density of the reduced nonlinear filtering problem approximates the original problem when the signal process has constant diffusion coefficients, periodic drift coefficients and the observation process is only dependent on the slow process.

Convergence of the nonlinear filter is shown in a very general setting in [KLS97], based on convergence in total variation distance of the law of $(X^\epsilon, Y^\epsilon)$. In the examples of [KLS97], the diffusion coefficient is not allowed to depend on the fast component.

The work of nonlinear filter approximation given in [Kus90, Chapter 6], for a two timescale jump-diffusion process, but with no correlation between signal and observation process. The difference of the actual unnormalized conditional measure and the reduced conditional measure is shown to converge to zero in distribution. Standard results then yield convergence in probability of the fixed time marginals. The 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, which possess a unique solution. In [BNP20] a similar approach to [Kus90, Chapter 6] is used to study a broader multiple timescale correlation filtering problem where an intermediate scaling term exists and there is correlation between the slow and observation processes. The authors make use of the perturbed test function approach where the correctors are solutions of Poisson equations to manage the difficulties introduced by the intermediate timescale. The main result in [BNP20] is that $\pi^{t,x} \rightarrow \pi^0$ in probability for a metric generating the weak topology.

In contrast to other papers on the convergence of the nonlinear filter for the multiple timescale problem, Imkeller et al. [Imk+13] showed a quantitative rate of convergence of $\epsilon$ for the system in Eq. 1.1, but without intermediate timescale nor correlation of the slow process with the observation process. This is accomplished using a suitable asymptotic expansion of the dual of the Zakai equation and then harnessing a probabilistic representation of the SPDEs in terms of backward doubly stochastic differential equations. The approach of [Imk+13] is extended in this paper to cover the case of correlation between the observation process and the coarse-grain process. The analysis is therefore similar, with the exception of additional methods to handle the components of the dual of the Zakai equation due to the correlation and the final argument of the main proof.

**Theorem** (Main Result)

*Under the assumptions stated in Theorem 2.1, for every $p \geq 1$, $T \geq 0$, there exists a $C > 0$ such that for*
every $\varphi \in C_b^0(\mathbb{R}^m; \mathbb{R})$,}
\begin{equation*}
\mathbb{E}_\mathbb{Q}\left[|\pi^{\epsilon,x}_T(\varphi) - \pi^0_T(\varphi)|^p\right] \leq \epsilon^p C|\varphi|^{p,\infty}_4.
\end{equation*}

In particular, there exists a metric $d$ on the space of probability measures on $\mathbb{R}^m$, such that $d$ generates the topology of weak convergence, and such that for every $T \geq 0$, there exists $C > 0$ so that
\begin{equation*}
\mathbb{E}_\mathbb{Q}\left[d(\pi^{\epsilon,x}_T, \pi^0_T)\right] \leq \epsilon C.
\end{equation*}

To prove the main result, we first setup the full problem in Section 2, provide useful notation, and the main result. The averaged SDE, Kushner-Stratonovich and Zakai equations are provided in this section as well. Having introduced the Zakai equations, we introduce their dual process representations in Section 3 and explain how working with the dual process will allow us to prove the main result. In Section 4, a probabilistic representation of the dual processes is given. Having established the necessary tools for the analysis, we provide preliminary estimates in Section 5 and then the main analysis in Section 6.

2 Problem Statement

In this section, we provide the full problem statement, some notation and the main result. We consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{Q})$ supporting a $(w + v + u)$-dimensional $\mathcal{F}_t$-adapted Brownian motion $(W, V, U)$. We will work with the following system of SDEs,
\begin{equation}
\begin{align*}
\text{d}X^\epsilon_t &= b(X^\epsilon_t, Z^\epsilon_t)\text{d}t + \sigma(X^\epsilon_t, Z^\epsilon_t)\text{d}W_t, \\
\text{d}Z^\epsilon_t &= \frac{1}{\epsilon^2} f(X^\epsilon_t, Z^\epsilon_t)\text{d}t + \frac{1}{\epsilon} g(X^\epsilon_t, Z^\epsilon_t)\text{d}V_t, \\
\text{d}Y^\epsilon_t &= h(X^\epsilon_t, Z^\epsilon_t)\text{d}t + \alpha \text{d}W_t + \gamma \text{d}U_t, \quad Y^\epsilon_0 = 0 \in \mathbb{R}^d,
\end{align*}
\end{equation}
where $b : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m$, $\sigma : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m \times \mathbb{R}^w$, $f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$, $g : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^u$ and $h : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^d$ are Borel measurable functions. The initial distribution of $(X, Z)$ is denoted by $\mathbb{Q}_{(X_0, Z_0)}$ and is assumed independent of the $(W, V, U)$ Brownian motion. $\mathbb{Q}_{(X_0, Z_0)}$ is also assumed to have finite moments for all orders. In Eq. 2.1, $0 < \epsilon \ll 1$, is a timescale separation parameter. We consider the case where $\alpha \in \mathbb{R}^{d \times w}$, $\gamma \in \mathbb{R}^{d \times u}$, and assume the following to be true
\begin{equation}
K \equiv \alpha \alpha^\ast + \gamma \gamma^\ast > 0, \quad \gamma \gamma^\ast > 0.
\end{equation}
This implies the existence of a unique $\mathbb{R}^{d \times d} \ni \kappa > 0$ of lower triangular form, such that $K = \kappa \kappa^\ast$. Hence there exists a unique $\kappa^{-1}$, such that we can define an auxiliary observation process
\begin{equation}
Y^{\epsilon, \kappa}_t = \int_0^t \kappa^{-1} \text{d}Y^\epsilon_s = \int_0^t \kappa^{-1} h(X^\epsilon_s, Z^\epsilon_s)\text{d}s + B_t, \quad Y^{\epsilon, \kappa}_0 = 0 \in \mathbb{R}^d,
\end{equation}
where
\begin{equation*}
B_t = \kappa^{-1} (\alpha \text{d}W_t + \gamma \text{d}U_t),
\end{equation*}
is a standard $d$-dimensional Brownian motion under $\mathbb{Q}$.

We are interested in the convergence of the $x$-marginal of the normalized filter, $\pi^{\epsilon, x}$, the conditional distribution of the signal given the observation filtration, to an averaged form. In particular, for any test function $\varphi \in C_b^2(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R})$ and time $t \in [0, T]$, the normalized filter can be characterized as
\begin{equation}
\pi^\epsilon_t(\varphi) = \mathbb{E}_\mathbb{Q} [\varphi(X^\epsilon_t, Z^\epsilon_t) | Y^\epsilon_t],
\end{equation}
where $Y^\epsilon_t \equiv \sigma(\{Y^\epsilon_s | s \in [0, t]\}) \vee \mathcal{N}$, the $\sigma$-algebra generated by the observation process over the interval $[0, t]$, joined with $\mathcal{N}$, the $\mathbb{Q}$ negligible sets.
Because the filtrations generated by $Y^\epsilon$ and $Y^{\epsilon,\kappa}$ are equivalent, from the point of view of $\pi^\epsilon$ we can use either. Hence, let us redefine the sensor function $h \leftarrow \kappa^{-1}h$, the coefficients $\alpha \leftarrow \kappa^{-1}\alpha$ and $\gamma \leftarrow \kappa^{-1}\gamma$, so that the observation process can be redefined as

$$dY_i^\epsilon = h(X_i^\epsilon, Z_i^\epsilon)dt + dB_i, \quad Y_0^\epsilon = 0 \in \mathbb{R}^d,$$

(2.4)

where $B = \alpha W + \gamma U$ is a standard Brownian motion under $\mathbb{Q}$ and still correlated with $W$.

In Eq. 2.1, we identify the infinitesimal generators of the SDEs as follows,

$$\mathcal{G}_S(x, z) = \sum_{i=1}^m b_i(x, z) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^m (\sigma \sigma^*)_{ij}(x, z) \frac{\partial^2}{\partial x_i \partial x_j},$$

$$\mathcal{G}_F(x, z) = \sum_{i=1}^n f_i(x, z) \frac{\partial}{\partial z_i} + \frac{1}{2} \sum_{i,j=1}^n (gg^*)_{ij}(x, z) \frac{\partial^2}{\partial z_i \partial z_j},$$

$$\mathcal{G} = \frac{1}{\epsilon^2} \mathcal{G}_F + \mathcal{G}_S.$$

The Kushner-Stratonovich equation for the time evolution of the filter $\pi^\epsilon$, acting on a test function $\varphi \in C^2_b(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R})$, is

$$\pi^\epsilon_t(\varphi) = \pi^\epsilon_0(\varphi) + \int_0^t \pi^\epsilon_s(\mathcal{G}^\epsilon \varphi)ds + \int_0^t \{ \pi^\epsilon_s(\varphi h + \alpha \sigma^* \nabla_x \varphi) - \pi^\epsilon_s(\varphi) \pi^\epsilon_s(h) \}, dY_s^\epsilon - \pi^\epsilon_s(h)ds,$$

(2.5)

$$\pi^\epsilon_0(\varphi) = \mathbb{E}_\mathbb{Q} [ \varphi(X_0^\epsilon, Z_0^\epsilon) ].$$

When we are interested in estimating test functions of $X^\epsilon$ only, i.e., $\varphi \in C^2_b(\mathbb{R}^m; \mathbb{R})$, we consider the $x$-marginal of $\pi^\epsilon$,

$$\pi^{\epsilon,x}_t(\varphi) = \int \varphi(x) \pi^{\epsilon,x}_t(dx, dz).$$

(2.6)

### 2.1 Diffusion Approximation and the Averaged Filter

The theory of homogenization of stochastic differential equations shows that if the process $Z^{\epsilon,x}_t$,

$$dZ^{\epsilon,x}_t = \frac{1}{\epsilon^2} f(x, Z^{\epsilon,x}_t)dt + \frac{1}{\epsilon} g(x, Z^{\epsilon,x}_t)dV_t,$$

(2.7)

is ergodic with stationary distribution $\mu_\infty(x)$, then under appropriate conditions, in the limit $\epsilon \to 0$ the process $X^\epsilon$ converges in distribution to a Markov process $X^0$ with infinitesimal generator

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

(2.8)

where the averaged drift and diffusion coefficients are

$$\tilde{b}(x) \equiv \int_{\mathbb{R}^n} b(x, z) \mu_\infty(dz; x), \quad \text{and} \quad \tilde{a}(x) \equiv \int_{\mathbb{R}^n} a(x, z) \mu_\infty(dz; x).$$

Here we denote the diffusion coefficient $a = \sigma \sigma^*$. Additionally, let us define

$$\tilde{h}(x) \equiv \int_{\mathbb{R}^n} h(x, z) \mu_\infty(dz; x), \quad \text{and} \quad \tilde{\sigma}(x) \equiv \int_{\mathbb{R}^n} \sigma(x, z) \mu_\infty(dz; x).$$

The aim of this paper is to show that the $x$-marginal filter $\pi^{\epsilon,x}$ can be approximated by an averaged filter $\pi^0$. We will show the existence and uniqueness of $\pi^0$ in Section 2.3.1. This is done by defining $\pi^0$ from the Kallianpur-Striebel formula and the existence and uniqueness of an unnormalized averaged filter $\rho^0$. The averaged filter will depend on the averaged coefficients $\tilde{b}, \tilde{a}, \tilde{\sigma}$ and $\tilde{h}$. 
2.2 Notation and Main Theorem

Before stating the main result of the paper, we set a few definitions and assumptions that will be used throughout the paper. We will use $\mathbb{N}_0$ to denote $\{0,1,2,\ldots\}$ and $\mathbb{N}$ for $\{1,2,\ldots\}$. Let $H_f$ denote the assumption that there exists a constant $C > 0$, exponent $\alpha > 0$ and an $R > 0$ such that for all $|z| > R$,

$$\sup_{x \in \mathbb{R}^m} (f(x,z),z) \leq -C|z|^\alpha. \quad (H_f)$$

$H_f$ is a recurrence condition, which provides the existence of a stationary distribution, $\mu_\infty(x)$, for the process $Z^\alpha$. Let $H_g$ denote the assumption that there are $0 < \lambda \leq \Lambda < \infty$, such that for any $(x,z) \in \mathbb{R}^m \times \mathbb{R}^n$,

$$\lambda I \preceq g^*(x,z) \preceq \Lambda I, \quad (H_g)$$

where $\preceq$ is the order relation in the sense of positive semidefinite matrices. $H_g$ is a uniform ellipticity condition, which provides the uniqueness of the stationary distribution. We will say that a function $\theta : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ is centered with respect to $\mu_\infty(x)$, if for each $x$

$$\int \theta(x,z)\mu_\infty(dz;x) = 0, \quad \forall x \in \mathbb{R}^m.$$

If $\varphi(x,z) \in C^{k,l}_b(\mathbb{R}^m \times \mathbb{R}^n;\mathbb{R}^n)$, then $\varphi$ is $k$-times continuously differentiable in the $x$-component, $l$-times continuously differentiable in the $z$-component, and all partial derivatives $\partial_z^k \partial_x^l \varphi$ for $0 \leq k' \leq k$, $0 \leq l' \leq l$ are bounded. Let $HF^{k,l}$ for $k,l \in \mathbb{N}_0$ denote the following assumption:

$$f \in C^{k,l}_b(\mathbb{R}^m \times \mathbb{R}^n;\mathbb{R}^n) \quad \text{and} \quad g \in C^{k,l}_b(\mathbb{R}^m \times \mathbb{R}^n;\mathbb{R}^{n \times k}). \quad (HF^{k,l})$$

Similarly, let $HS^{k,l}$ for $k,l \in \mathbb{N}_0$ denote the assumption:

$$b \in C^{k,l}_b(\mathbb{R}^m \times \mathbb{R}^n;\mathbb{R}^n) \quad \text{and} \quad \sigma \in C^{k,l}(\mathbb{R}^m \times \mathbb{R}^n;\mathbb{R}^{m \times k}), \quad (HS^{k,l})$$

and $HO^{k,l}$ for $k,l \in \mathbb{N}_0$ denote the assumption:

$$h \in C^{k,l}_b(\mathbb{R}^m \times \mathbb{R}^n;\mathbb{R}^d). \quad (HO^{k,l})$$

We use the notation $k = (k_1, \ldots, k_m) \in \mathbb{N}_0^m$ for a multiindex with order $|k| = k_1 + \ldots + k_m$ and define the differential operator

$$D^k_x = \frac{\partial^{\lvert k \rvert}}{\partial x_1^{k_1} \ldots \partial x_m^{k_m}}.$$

Lastly, the relation $a \preceq b$ will indicate that $a \leq Cb$ for a constant $C > 0$ that is independent of $a$ and $b$, but that may depend on parameters that are not critical for the bound being computed.

Having introduced the necessary definitions and equations, we now state the main result fully.

**Theorem 2.1**

Assume $H_f$, $H_g$, $HF^{8,4}$, $b \in C^{7,4}_b$, $\sigma \in C^{8,4}_b$, and $HO^{8,4}$. Additionally, assume that the initial distribution $\mathbb{Q}(X_0,Z_0)$ has finite moments of every order. Then for any $p \geq 1, T \geq 0$ we have that for every $\varphi \in C^4_b(\mathbb{R}^m;\mathbb{R})$,

$$\mathbb{E}_{\mathbb{Q}} \left[ \lvert \pi_T^{\epsilon,x}(\varphi) - \pi_T^0(\varphi) \rvert^p \right] \leq C\|\varphi\|_{L^\infty}^p.$$

Further, there exists a metric $d$ on the space of probability measures on $\mathbb{R}^m$ that generates the topology of weak convergence, such that

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

**Proof.** The proof of the first result is given by Corollary 6.1. The proof of the second result is from Lemma 6.9. □
Before moving beyond this theorem statement, we provide some quick remarks.

**Remark.** From \( \lim_{t \to 0} \mathbb{E}_Q [d(\pi_{t,x}^n, \pi_0^T)] = 0 \), we retrieve convergence in probability,

\[
\lim_{\epsilon \to 0} \mathbb{Q} \left( d(\pi_{t,x}^\epsilon, \pi_0^T) \geq \delta \right) \leq \frac{1}{\delta} \lim_{\epsilon \to 0} \mathbb{E}_Q [d(\pi_{t,x}^\epsilon, \pi_0^T)] = 0, \quad \text{for each } \delta > 0.
\]

And by the Borel-Cantelli lemma we can choose \( (\epsilon_n) \) so that \( \pi_{t,x}^\epsilon \) will a.s. converge weakly to \( \pi_0^T \).

**Remark.** Some quick comparisons to the main result in [Imk+13]. There the scaling for the fast process was of order one, whereas in this paper we use order two. Therefore, the rate of convergence is the same in the two works. The only difference in the conditions of our Theorem 2.1 and the equivalent one in [Imk+13], is that we require \( \sigma \in C^{8,4}_b \) instead of \( C^{7,4}_b \). This extra regularity in the slow component of the function is due to the correlation between the slow process and observation process, which then appears in our backward stochastic differential equations of Section 6.

### 2.3 Change of Probability Measure and the Zakai Equation

In Section 6, we will be interested in working with the unnormalized conditional measure since it will satisfy a linear evolution equation. To define the unnormalized conditional measure requires a change of probability measure transformation, which we will perform for each \( \epsilon \). Let us denote the new collection of probability measures by \( (\mathbb{P}^\epsilon) \). For any fixed \( \epsilon, \mathbb{P}^\epsilon \) and \( \mathbb{Q} \) will be mutually absolutely continuous with Radon-Nikodym derivatives

\[
D_t^\epsilon \equiv \frac{d\mathbb{P}^\epsilon}{d\mathbb{Q}} \bigg|_{\mathcal{F}_t} = \exp \left( - \int_0^t \langle h(X_s^\epsilon, Z_s^\epsilon), dB_s \rangle - \frac{1}{2} \int_0^t |h(X_s^\epsilon, Z_s^\epsilon)|^2 ds \right),
\]

\[
\tilde{D}_t^\epsilon \equiv (D_t^\epsilon)^{-1} \frac{d\mathbb{Q}}{d\mathbb{P}^\epsilon} \bigg|_{\mathcal{F}_t} = \exp \left( \int_0^t \langle h(X_s^\epsilon, Z_s^\epsilon), dY_s^\epsilon \rangle - \frac{1}{2} \int_0^t |h(X_s^\epsilon, Z_s^\epsilon)|^2 ds \right).
\]

Then by Girsanov’s theorem, under \( \mathbb{P}^\epsilon \) the process \( Y^\epsilon \) is a Brownian motion. For a fixed test function \( \varphi \in C^2_T(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}) \) and time \( t \in [0, T] \), we characterize the unnormalized conditional measure \( \rho_t^\epsilon \) as,

\[
\rho_t^\epsilon (\varphi) = \mathbb{E}_{\mathbb{P}^\epsilon} \left[ \varphi(X_t^\epsilon, Z_t^\epsilon) \tilde{D}_t^\epsilon \mid Y_t^\epsilon \right],
\]

and its relation to \( \pi^\epsilon \) is given by the Kallianpur-Striebel formula,

\[
\pi_t^\epsilon (\varphi) = \frac{\mathbb{E}_{\mathbb{P}^\epsilon} \left[ \varphi(X_t^\epsilon, Z_t^\epsilon) \tilde{D}_t^\epsilon \mid Y_t^\epsilon \right]}{\mathbb{E}_{\mathbb{P}^\epsilon} \left[ \tilde{D}_t^\epsilon \mid Y_t^\epsilon \right]} = \frac{\rho_t^\epsilon (\varphi)}{\rho_t^\epsilon (1)}, \quad \forall t \in [0, T], \quad \mathbb{Q}, \mathbb{P}^\epsilon\text{-a.s.}
\]

The action of \( \rho^\epsilon \) on test functions \( \varphi \in C^2 (\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}) \) gives the Zakai evolution equation,

\[
\begin{aligned}
\rho_t^\epsilon (\varphi) &= \rho_0^\epsilon (\varphi) + \int_0^t \rho_s^\epsilon (G^\epsilon \varphi) \, ds + \int_0^t \langle \rho_s^\epsilon (\varphi h + \alpha \sigma^* \nabla_x \varphi), dY_s^\epsilon \rangle, \\
\rho_0^\epsilon (\varphi) &= \mathbb{E}_Q [\varphi(X_0^\epsilon, Z_0^\epsilon)].
\end{aligned}
\]  

(2.9)

When \( \varphi \in C^2 (\mathbb{R}^m; \mathbb{R}) \), we consider the \( x \)-marginal,

\[
\rho_t^{*,x} (\varphi) = \int \varphi(x) \rho_t^\epsilon (dx, dz),
\]

which is related to \( \pi^*,x \) through the Kallianpur-Striebel formula,

\[
\pi_t^{*,x} (\varphi) = \frac{\rho_t^{*,x} (\varphi)}{\rho_t^\epsilon (1)}, \quad \forall t \in [0, T], \quad \mathbb{Q}, \mathbb{P}^\epsilon\text{-a.s.}
\]
2.3.1 The Averaged Conditional Distributions

In this section, we show that there exists a probability measure-valued process \( \pi^0 \), the averaged filter, that is defined in terms of a measure-valued process \( \rho^0 \), the averaged unnormalized filter, from the Kallianpur-Striebel formula. To do so, we start by defining under \( Q \), the SDE \( X^{0,\epsilon} \) satisfying the equation,

\[
dX^{0,\epsilon}_t = \tilde{b}(X^{0,\epsilon}_t)dt + (\tilde{a}(X^{0,\epsilon}_t) - \sigma\sigma^*(X^{0,\epsilon}_t))^{1/2}dW_t + \sigma(X^{0,\epsilon}_t)\left(dW_t - \alpha^*\tilde{h}(X^{0,\epsilon}_t)dt\right),
\]

\( X^{0,\epsilon}_0 \sim Q_{X_0}^0 \),

which under the change of measure to \( P^\epsilon \) becomes

\[
dX^{0,\epsilon}_t = \tilde{b}(X^{0,\epsilon}_t)dt + (\tilde{a}(X^{0,\epsilon}_t) - \sigma\sigma^*(X^{0,\epsilon}_t))^{1/2}d\tilde{W}_t + \sigma(X^{0,\epsilon}_t)\left(d\tilde{W}_t - \alpha^*\tilde{h}(X^{0,\epsilon}_t)dt\right),
\]

\( X^{0,\epsilon}_0 \sim Q_{X_0}^0 \),

where \( \tilde{W}_t \) is a standard Brownian motion under \( P^\epsilon \). The Cholesky factor \( (\tilde{a}(X^{0,\epsilon}_t) - \sigma\sigma^*(X^{0,\epsilon}_t))^{1/2} \) exists, since from an application of Jensen’s inequality \( \tilde{a}(x) - \sigma\sigma^*(x) \geq 0 \) for every \( x \in \mathbb{R}^m \). We now define under \( P^\epsilon \), the process

\[
\tilde{D}^0_t = \exp\left(\int_0^t \langle \tilde{h}(X^{0,\epsilon}_s), dY^\epsilon_s \rangle - \frac{1}{2} \int_0^t |\tilde{h}(X^{0,\epsilon}_s)|^2 ds\right),
\]

which also satisfies the relation,

\[
\tilde{D}^0_t = 1 + \int_0^t \langle \tilde{D}^0_s \tilde{h}(X^{0,\epsilon}_s), dY^\epsilon_s \rangle.
\]

Lemma 2.1

Assume \( \tilde{h} \) is a bounded function. Then there exists a measure-valued process \( (\rho^0_t)_{t \geq 0} \) such that for all \( \varphi \in C_b \),

\[
\rho^0_0(\varphi) = \mathbb{E}_{P^\epsilon}\left[\varphi(X^{0,\epsilon}_t)\tilde{D}^0_t \mid Y^\epsilon_t\right], \quad P^\epsilon\text{-a.s.}
\]

Additionally, for every \( \varphi \in C^2_b(\mathbb{R}^n; \mathbb{R}) \), \( \rho^0(\varphi) \) satisfies the equation,

\[
\begin{align*}
\rho^0_0(\varphi) &= \rho^0_0(\varphi) + \int_0^t \rho^0_s(\nabla \varphi)ds + \int_0^t \langle \rho^0_s(\varphi \tilde{h} + \alpha \sigma^* \nabla_x \varphi), dY^\epsilon_s \rangle, \\
\rho^0_0(\varphi) &= \int \varphi(x) Q_{X^0_0}(dx),
\end{align*}
\]

(2.11)

where \( Q_{X^0_0} \) is the initial distribution of \( X^\epsilon \) and the solution is unique if the coefficients of \( \tilde{b}, \tilde{a}, \sigma, \) and \( \tilde{h} \) are in \( C^3_b \).

Proof. Because \( \tilde{h} \) is bounded, we have by the same proof as Lemma 5.3, the uniform bound

\[
\sup_{t \in [0, 1]} \mathbb{E}_{P^\epsilon} |\tilde{D}^0_t|^p < \infty,
\]

for \( p \geq 2, T > 0 \). Then \( \mathbb{E}_{P^\epsilon} \tilde{D}^0_t = 1 \) and therefore \( \mathbb{Q}^\epsilon(\cdot) = \int \tilde{D}^0_t(\omega) P^\epsilon(d\omega) \) is a new probability measure. Since \( \tilde{D}^0_t \) is \( P^\epsilon\)-a.s. strictly positive, we also have that \( \mathbb{Q}^\epsilon \) is equivalent to \( P^\epsilon \). By the Kallianpur-Striebel formula we know that the following holds,

\[
\frac{\mathbb{E}_{P^\epsilon}\left[\varphi(X^{0,\epsilon}_t)\tilde{D}^0_t \mid Y^\epsilon_t\right]}{\mathbb{E}_{P^\epsilon}\left[\tilde{D}^0_t \mid Y^\epsilon_t\right]} = \mathbb{E}_{Q^\epsilon}\left[\varphi(X^{0,\epsilon}_t) \mid Y^\epsilon_t\right].
\]

(2.12)
Indeed, for any $\mathcal{Y}_t^\tau$-measurable random variable $\xi$:

$$
E_{P^\tau} \left[ \xi \mathbb{E}_{Q^0} \left[ \varphi(X_{0,\epsilon}^t) \mid \mathcal{Y}_t^\tau \right] \mathbb{E}_{P^\tau} \left[ \tilde{D}_t^0 \mid \mathcal{Y}_t^\tau \right] \right] = E_{P^\tau} \left[ \xi \mathbb{E}_{Q^0} \left[ \varphi(X_{0,\epsilon}^t) \mid \mathcal{Y}_t^\tau \right] \tilde{D}_t^0 \right]
$$

$$
= E_{Q^0} \left[ \xi \mathbb{E}_{Q^0} \left[ \varphi(X_{0,\epsilon}^t) \mid \mathcal{Y}_t^\tau \right] \right] = E_{P^\tau} \left[ \xi \varphi(X_{0,\epsilon}^t) \tilde{D}_t^0 \right],
$$

and therefore

$$
E_{Q^0} \left[ \varphi(X_{0,\epsilon}^t) \mid \mathcal{Y}_t^\tau \right] E_{P^\tau} \left[ \tilde{D}_t^0 \mid \mathcal{Y}_t^\tau \right] = E_{P^\tau} \left[ \varphi(X_{0,\epsilon}^t) \tilde{D}_t^0 \right].
$$

As a consequence of $\tilde{D}_t^0 > 0$ $\mathbb{P}$-a.s., we have that the random variable $E_{P^\tau} \left[ \tilde{D}_t^0 \mid \mathcal{Y}_t^\tau \right]$ is $\mathbb{P}$-a.s. strictly positive, so we can divide by the variable to obtain Eq. 2.12. Therefore, there exists a regular $Q^0$-conditional probability $\pi_t^0$ such that $Q^0$-a.s. $\pi_t^0(\varphi) = E_{Q^0} \left[ \varphi(X_{0,\epsilon}^t) \mid \mathcal{Y}_t^\tau \right]$ for each fixed $t$. Since $\mathbb{P}$ and $Q^0$ are equivalent, the identity also holds $\mathbb{P}$-a.s. This only gives a random measure at one fixed time, but [BC09, Theorem 2.24, p.29] can be used to obtain $\pi^0$ as a probability measure-valued process. Note that in our setup, $X_{0,\epsilon}^t$ is not a Markov process, as is assumed in [BC09, Theorem 2.24, p.29], but this is not important for that theorem and one could always consider $X_{0,\epsilon}^t$ to be a component of the Markov process $(X_{0,\epsilon}^t, X^\epsilon, Z^\epsilon)$. We now set $\rho_t^0(\varphi) = \pi_t^0(\varphi)E_{P^\tau} \left[ \tilde{D}_t^0 \mid \mathcal{Y}_t^\tau \right]$ to get the first part of the proof. For the last part, by standard construction of the Zakai equation (see for instance [BC09]), $\rho^0$ satisfies Eq. 2.11 and uniqueness follows from [Roz91, Theorem 3.1, p.454].

Therefore based on Lemma 2.1, the averaged (normalized) filter $\pi^0$ is then related to $\rho^0$ by the Kallianpur-Striebel relation,

$$
\pi_t^0(\varphi) = \frac{\rho_t^0(\varphi)}{\rho_t^0(1)}, \quad \forall t \in [0, T], \quad \forall \varphi \in C_b(\mathbb{R}^m; \mathbb{R}). \tag{2.13}
$$

Remark. An interesting observation regarding Eq. 2.11, is that we may have $\sigma = 0$, and this implies that the SDE for the averaged filter may have no correlation.

Remark. Note that $\pi^0$ is not the filter for the averaged system, and hence $\rho^0$ is also not the unnormalized conditional measure for the averaged system, which would instead satisfy the following equation,

$$
\ddot{\rho}_t(\varphi) = \dot{\rho}_t(\varphi) + \int_0^t \ddot{\rho}_s(\mathcal{F}_s^\tau \varphi) ds + \int_0^t \langle \dot{\rho}_s(\varphi \mathcal{F} + \alpha \sqrt{\mathcal{G}} \nabla_x \varphi), d\mathcal{G}_s \rangle,
$$

$$
\ddot{\rho}_0(\varphi) = \int \varphi(x)QX_0^t(dx),
$$

where

$$
\mathcal{F}_t = \int_0^t \ddot{\mathcal{F}}_s(X_0^s) ds + B_t,
$$

and $X^0$ is a diffusion process with infinitesimal generator $\mathcal{G}_S$ under $Q$.

3 Dual Process to the Unnormalized Conditional Distribution

We now introduce an idea by [Par80] that is an important transition to the method of proof used in this paper. The idea is to define a function-valued process for any fixed $\varphi \in C_b^2(\mathbb{R}^m; \mathbb{R})$. The function-valued process will be the dual of $\rho^0$ in an appropriate sense. We first define

$$
\tilde{D}_{t,T}^0 = \exp \left( \int_t^T \langle h(X_s^\epsilon, Z_s^\epsilon), dY_s^\epsilon \rangle - \frac{1}{2} \int_t^T |h(X_s^\epsilon, Z_s^\epsilon)|^2 ds \right),
$$

where

$$
\mathcal{F}_t = \int_0^t \ddot{\mathcal{F}}_s(X_0^s) ds + B_t.
$$

and $X^0$ is a diffusion process with infinitesimal generator $\mathcal{G}_S$ under $Q$. 

8
which is \( Q \)-a.s. equal to \( \tilde{D}_t(\tilde{D}_t)^{-1} \). Fixing \( \varphi \in C^p_b(\mathbb{R}^m; \mathbb{R}) \), we then define the dual process at time \( t \in [0, T] \) as

\[
v^\varphi_t(x, z) = \mathbb{E}^\nu_{t,x,z} \left[ \varphi(X^\nu_t) \tilde{D}_t \mid \mathcal{Y}^\nu_t \right],
\]

where \( \mathbb{P}^\nu_{t,x,z} \) is the change of probability measure that results when \( (X^*_s, Z^*_s) \) takes the constant value \( (x, z) \) for \( s \in [0, t] \) and then follows the dynamics given by Eq. 1.1 for \( s > t \). \( v^\varphi_t(x, z) \) is called the dual process, because for any \( t \in [0, T] \) we have

\[
\rho^\varphi_t(\varphi) = \rho^\nu_t(v^\varphi_t \varphi), \quad \mathbb{P}^\nu\text{-a.s.}
\]

This also means that \( \rho^\varphi_0(\varphi) = \rho^\nu_0(v^0 \varphi) \) and therefore

\[
\rho^\varphi_t(\varphi) = \int_{\mathbb{R}^m \times \mathbb{R}^n} v^\varphi_t(x, z) Q(X^0_t, Z^0_t) \, dx \, dz.
\]

We can similarly define the dual process for \( \rho^0 \). Following the construction in Lemma 2.1, we would have

\[
v^0_t(x, z) = \mathbb{E}^\nu_{t,x,z} \left[ \varphi(X^0_t) \tilde{D}_t^0 \mid \mathcal{Y}^\nu_t \right],
\]

with the same property that \( \rho^0_t(\varphi) = \rho^0_0(v^0 \varphi) \). Again \( \mathbb{P}^\nu_{t,x,z} \) is the change of probability measure that results from \( X^0_t \) taking the constant value \( x \) for \( s \in [0, t] \) and then follows the dynamics given by the SDE in Eq. 2.10. The definition of \( \tilde{D}_t^0 \) in \( v^0_t \) is

\[
\tilde{D}_t^0 = \exp \left( \int_0^T (\tilde{b}(X^0_s, dY^s) - \frac{1}{2} \int_0^T |\tilde{b}(X^0_s)|^2 \, ds) \right),
\]

which is \( \mathbb{P}^\nu\text{-a.s.} \) equal to \( \tilde{D}_t = \tilde{D}_t^0(\tilde{D}_t^0)^{-1} \).

### 3.1 The Dual Process and Filter Convergence

We now show the usefulness of the dual process in showing the convergence of \( \rho^\varphi \to \rho^0 \). We again fix \( \varphi \in C^p_b(\mathbb{R}^m; \mathbb{R}) \) and \( p \geq 1 \). Then from Jensen’s inequality and Fubini’s theorem we have the following relation,

\[
\mathbb{E}^\nu \left[ |\rho^\varphi_t(\varphi) - \rho^0_t(\varphi)|^p \right] = \mathbb{E}^\nu \left[ \left| \int v^\varphi_t(x, z) - v^0_0 \varphi(x) Q(X^0_t, Z^0_t) \, dx \, dz \right|^p \right]
\leq \mathbb{E}^\nu \left[ \left| \int v^\varphi_t(x, z) - v^0_0 \varphi(x) \right|^p \, dx \, dz \right] \leq \mathbb{E}^\nu \left[ \left| \int v^\varphi_t(x, z) - v^0_0 \varphi(x) \right|^p \right] Q(X^0_t, Z^0_t) \, dx \, dz. \quad (3.1)
\]

This implies that if \( Q(X^0_t, Z^0_t) \) is well behaved (e.g., finite moments of every order) then convergence of the \( p \)-th moment of \( v^\varphi_t(x, z) - v^0_0 \varphi(x) \) to zero will imply convergence of the \( p \)-th moment of \( \rho^\varphi_t(\varphi) - \rho^0_t(\varphi) \) to zero. Without loss of generality, we assumed that \( X^0_t \) had the same initial distribution as \( X^\nu \) in Section 2.3.1, and hence why the integration is against \( Q(X^0_t, Z^0_t) \) in Eq. 3.1.

### 3.2 Evolution Equations for the Dual Process

To introduce the next step in the techniques to prove convergence of the marginalized filter to the reduced order filter, we need to state the evolution equations for the dual processes \( v^\varphi \) and \( v^0 \). Both processes satisfy backward stochastic partial differential equations (BSPDE). To facilitate the reading, we use \( v^\nu \) and \( v^0 \) instead of the more verbose \( v^\varphi \) and \( v^0 \) in most of what follows. When clarity is needed, we will use the explicit notation.
The evolution equation for \( v^\epsilon \) is given by
\[
-dv^\epsilon_t = G^*v^\epsilon_t dt + \langle v^\epsilon_t h + \alpha \sigma^* \nabla_x v^\epsilon_t, d\tilde{B}_t \rangle, \quad v^\epsilon_T = \varphi, \tag{3.2}
\]
where \( \int d\tilde{B}_t \) will denote the backward Itô integral. The process \( v^0 \) is given by
\[
-dv^0_t = G^0v^0_t dt + \langle v^0_t h + \alpha \sigma^* \nabla_x v^0_t, d\tilde{B}_t \rangle, \quad v^0_T = \varphi. \tag{3.3}
\]

3.3 Expansion of the Dual Process

Because \( v^\epsilon \) satisfies a linear equation, we consider an expansion of \( v^\epsilon \) using \( v^0 \) and a corrector \( \psi \) and remainder \( R \) term,
\[
v^\epsilon_t(x, z) = v^0_t(x) + \psi_t(x, z) + R_t(x, z).
\]

Using this expansion in Eq. 3.2 and introducing terms for Eq. 3.3, we define \( \psi \) and \( R \) to satisfy the following linear BSPDEs
\[
-d\psi_t = \left[ \frac{1}{\epsilon^2} G^0 \psi_t + (G^0_S - G_S^0) \right] dt + \langle v^0_t h - \bar{h}, d\tilde{B}_t \rangle + \langle \alpha(\sigma - \sigma)^* \nabla_x v^0_t, d\tilde{B}_t \rangle, \quad \psi_T = 0,
\]
\[
-dR_t = (G^0 R_t + G_S \psi_t) dt + \langle (\psi_t + R_t) h, d\tilde{B}_t \rangle + \langle \alpha \sigma^* \nabla_x (\psi_t + R_t), d\tilde{B}_t \rangle, \quad R_T = 0. \tag{3.4}
\]

Therefore to show convergence of the difference \( v^\epsilon - v^0 \), we can equivalently show convergence of \( \psi \) and \( R \) to zero as \( \epsilon \to 0 \):
\[
\mathbb{E}_P^\epsilon \left[ \left( v^\epsilon_t(x, z) - v^0_t(x, z) \right)^p \right] = \mathbb{E}_P^\epsilon \left[ |\psi_t(x, z) + R_t(x, z)|^p \right] \leq \mathbb{E}_P^\epsilon \left[ |\psi_t(x, z)|^p \right] + \mathbb{E}_P^\epsilon \left[ |R_t(x, z)|^p \right].
\]

This will be our strategy in Section 6.

4 Probabilistic Representation of Stochastic PDEs

We will now show that we can find a probabilistic representation of the dual processes. This representation will be given by backward doubly stochastic differential equations (BDSDEs), which are a generalization of the Feynman-Kac solution for semilinear second order parabolic SPDEs. First let us state a result for the classical solution of the dual processes, which are linear second order parabolic SPDEs of the general form:
\[
-d\psi(\omega, t, x) = \mathcal{L}\psi(\omega, t, x) dt + f(\omega, t, x) dt + \langle g(\omega, t, x) + G(\omega, t, x) \psi(\omega, t, x), d\tilde{B}_t \rangle
\]
\[
+ \langle F(\omega, t, x) \nabla_x \psi(\omega, t, x), d\tilde{B}_t \rangle, \quad \psi(T, x) = \varphi(\omega, x),
\]
where \( \psi : \Omega \times [0, T] \times \mathbb{R}^m \to \mathbb{R}, f : \Omega \times [0, T] \times \mathbb{R}^m \to \mathbb{R}, g : \Omega \times [0, T] \times \mathbb{R}^m \to \mathbb{R}^d, G : \Omega \times [0, T] \times \mathbb{R}^m \to \mathbb{R}^d, F : \Omega \times [0, T] \times \mathbb{R}^m \to \mathbb{R}^d \times \mathbb{R}^m \) and \( \varphi : \Omega \times \mathbb{R}^m \to \mathbb{R} \) are all jointly measurable, and \( \tilde{B}_t \) is a \( d \)-dimensional standard backward Brownian motion. The generator given in Eq. 4.1 has the form
\[
\mathcal{L}(x) = \sum_{i=1}^{m} b_i(x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{m} a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j},
\]
where \( b : \mathbb{R}^m \to \mathbb{R} \) and \( a : \mathbb{R}^m \to \mathbb{S}^{m \times m} \) are measurable (\( \mathbb{S}^{m \times m} \) denotes the space of symmetric positive semidefinite matrices).

Once we have stated the results on BSPDEs, we will then give the BDSDE representation in Section 4.1. We will need some definitions for the necessary conditions on the classical solution of the SPDEs. Let us state those now, starting with the definition for the filtration \( \mathcal{F}_{t,s}^{0,B} \): let \( 0 \leq t \leq s \leq T \),
\[
\mathcal{F}_{t,s}^{0,B} = \sigma(\{B_u - B_t \mid t \leq u \leq s\}),
\]
and let \( \mathcal{F}_{t,s}^B \) be the completion of \( \mathcal{F}_{t,s}^{0,B} \) under \( \mathbb{P}^\epsilon \). We next define the space of adapted random fields of polynomial growth \( \mathcal{P}_T(\mathbb{R}^m; \mathbb{R}^n) \):
Definition 4.1. \( \mathcal{P}_T(\mathbb{R}^m;\mathbb{R}^n) \) is the space of random fields of polynomial growth

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

that are jointly measurable in \((\omega, t, x)\) and for fixed \((t, x)\), \(\omega \mapsto H(\omega, t, x)\) is \(\mathcal{F}_{t,x}^B\)-measurable. Further, for fixed \(\omega\) 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 \(C_p, q > 0\), 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 denote with \(D^k\) a definition concerning conditions on the coefficients of the generator \(\mathcal{L}\) of the BSPDE:

Definition 4.2. We define the condition \(D^k\) to indicate that \(b \in C^k_b(\mathbb{R}^m; \mathbb{R}^n)\), \(a \in C^k_b(\mathbb{R}^m; \mathbb{S}^{m \times m})\), and \(a\) is degenerate elliptic: For every \(\xi \in \mathbb{R}^m\) and every \(x \in \mathbb{R}^m\),

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

or succinctly \(a \succeq 0\).

We denote with \(S^k\) a definition concerning conditions on the coefficients (not including the generator) of the BSPDE:

Definition 4.3. The condition \(S^k\) indicates that \(f\) and \(g\) are \(k\)-times continuously differentiable and the partial derivatives up to order \(k\) are all in \(\mathcal{P}_T\). \(G\) and \(F\) are \((k+1)\)-times continuously differentiable and the partial derivatives up to order \((k+1)\) are all uniformly bounded in \((\omega, t, x)\). \(\varphi\) is \(k\)-times continuously differentiable, and all partial derivatives of order \(0\) to \(k\) grow at most polynomially.

Lemma 4.1

Assume \(D^k\) and \(S^k\) for some \(3 \leq k \in \mathbb{N}\). Additionally, assume the parabolic condition \(2a - F^*F \succeq 0\) holds. Then Eq. 4.1 has a unique classical solution \(\psi\) in the sense that for every fixed \(\omega\) outside a null set, \(\psi(\omega, \cdot, \cdot) \in C^{0,k-1}(0, T] \times \mathbb{R}^d; \mathbb{R})\), \(\psi\) and its partial derivatives are in \(\mathcal{P}_T(\mathbb{R}^m; \mathbb{R})\), and \(\psi\) solves the integral equation. If \(\tilde{\psi}\) is any other solution of the integral equation, then \(\psi\) and \(\tilde{\psi}\) are indistinguishable. If further \(f, g\) and \(\varphi\) as well as their derivatives up to order \(k\) are uniformly bounded in \((\omega, t, x)\), then for any \(p > 0\) there exist \(q > 0\) and \(C > 0\) (only depending on \(p\), the dimensions involved, the bounds on \(a, b, G\) and \(F\), and on \(T\)), such that for all \(|\beta| \leq k - 1\) and \(x \in \mathbb{R}^m\),

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

Proof. The lemma is a slight generalization of [Imk+13, Proposition 4.1, p.2302], and follows the same argument as the one given there.

4.1 Backward Doubly Stochastic Differential Equations

The theory of backward doubly stochastic differential equations has its origin in the paper by [PP94]. Although it is possible to get a different representation of the solutions of Eq. 4.1 by the Method of Stochastic Characteristics [Roz90], one benefit of the BDSDE representation is that for fixed \((x, z) \in \mathbb{R}^m \times \mathbb{R}^n\), we will have a finite dimensional representation of \(\psi(x, z)\) and therefore will be able to apply Grönwall’s lemma in the final step of Lemma 6.3, as part of the main analysis.

A BDSDE is an integral equation of the form,

\[
Y_t = \xi + \int_t^T f(s, \cdot, Y_s, Z_s) ds + \int_t^T \langle g(s, \cdot, Y_s, Z_s), d\overline{B}_s \rangle - \int_t^T \langle Z_s, dW_s \rangle,
\]

(4.2)
where \( f : [0, T] \times \Omega \times \mathbb{R} \times \mathbb{R}^n \to \mathbb{R} \), \( g : [0, T] \times \Omega \times \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^d \), and for fixed \( y \in \mathbb{R} \), \( z \in \mathbb{R}^n \), the processes \( (\omega, t) \mapsto f(t, \omega, y, z) \) and \( (\omega, t) \mapsto g(t, \omega, y, z) \) are \( (\mathcal{F}^B_t \vee \mathcal{F}^W_t) \otimes \mathcal{B}(\mathbb{R}) \)-measurable, and for every \( t \), \( f(t, \cdot, y, z) \) and \( g(t, \cdot, y, z) \) are \( \mathcal{F}_t \)-measurable. Our definition of \( \mathcal{F}_t \) is,

\[
\mathcal{F}_t = \mathcal{F}^B_{t,T} \vee \mathcal{F}^W_t,
\]

where \( \mathcal{F}^W_t = \mathcal{F}^W_{0,t} \). Because of this definition, \( \mathcal{F}_t \) is not a filtration; it is neither strictly increasing nor decreasing in \( t \). Let us now introduce some additional notation for integrability and measurability conditions of the solution of the BDSDEs.

**Definition 4.4.** Let \( H^2_2(\mathbb{R}^m) \) be the space of measurable \( \mathbb{R}^m \)-valued processes \( Y \), such that \( Y_t \) is \( \mathcal{F}_t \)-measurable for almost any \( t \in [0, T] \) and

\[
\mathbb{E} \left[ \int_0^T |Y_t|^2 dt \right] < \infty.
\]

**Definition 4.5.** Let \( S^2_2(\mathbb{R}^m) \) be the space of continuous adapted \( \mathbb{R}^m \)-valued processes \( Y \), such that \( Y_t \) is \( \mathcal{F}_t \)-measurable for every \( t \in [0, T] \) and

\[
\mathbb{E} \left[ \sup_{0 \leq t \leq T} |Y_t|^2 dt \right] < \infty.
\]

The pair \((Y, Z)\) will be called a solution of Eq. 4.2 if \((Y, Z) \in S^2_2(\mathbb{R}) \times H^2_2(\mathbb{R}^n)\), and if the pair solves the integral equation. We will also write BDSDEs in differential form at times, for example Eq. 4.2 in differential form would be,

\[
-dY_t = f(t, \cdot, Y_t, Z_t)dt + g(t, \cdot, Y_t, Z_t)dw_t - \langle Z_t, dw_t \rangle.
\]

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

In [PP94], it is shown that under the following conditions, Eq. 4.2 has a unique solution:

- \( \xi \in L^2(\Omega, \mathcal{F}_T, \mathbb{P}; \mathbb{R}) \),
- for any \((y, z) \in \mathbb{R} \times \mathbb{R}^n\), we have \( f(\cdot, \cdot, y, z) \in H^2_2(\mathbb{R}) \) and \( g(\cdot, \cdot, y, z) \in H^2_2(\mathbb{R}^d) \),
- \( f \) and \( g \) satisfy Lipschitz conditions and \( g \) is a contraction in \( z \): there exists constants \( L > 0 \) and \( 0 < \beta < 1 \) such that for any \((\omega, t)\) and \( y_1, y_2, z_1, z_2 \),

\[
|f(t, \omega, y_1, z_1) - f(t, \omega, y_2, z_2)|^2 \leq L(|y_1 - y_2|^2 + |z_1 - z_2|^2) \quad \text{and} \quad |g(t, \omega, y_1, z_1) - g(t, \omega, y_2, z_2)|^2 \leq L|y_1 - y_2|^2 + \beta|z_1 - z_2|^2.
\]

Now we associate a diffusion \( X \) to the differential operator \( \mathcal{L} \) given in Eq. 4.1. To do so, assume \( D_k \) is satisfied for some \( k \geq 2 \). Then \( \sigma \equiv a^{1/2} \) is Lipschitz continuous by [Str08, Lemma 2.3.3]. Hence for every \((t, x) \in [0, T] \times \mathbb{R}^m\), there exists a strong solution of the SDE

\[
\begin{align*}
X^{t,x}_s &= x + \int_t^s b(X^{t,x}_s)ds + \int_t^s \sigma(X^{t,x}_s)dW_s, & \text{for } s \geq t, \\
X^{t,x}_s &= x & \text{for } s \leq t.
\end{align*}
\]

For the theory of BDSDEs, we must assume that \( F \) has the form \( F = \alpha \sigma^* \), and here we consider \( \alpha \in \mathbb{R}^{d \times m} \) a constant matrix. We then associate the following BDSDE to Eq. 4.1,

\[
\begin{align*}
-dY^{t,x}_s &= f(s, X^{t,x}_s)ds + \langle g(s, X^{t,x}_s) + G(s, X^{t,x}_s)Y^{t,x}_s + \alpha Z^{t,x}_s, dw_t \rangle - \langle Z^{t,x}_s, dw_s \rangle, \\
Y^{t,x}_T &= \varphi(X^{t,x}_T).
\end{align*}
\]

Under the assumptions \( S_k \) and \( D_k \) for \( k \geq 2 \), this equation has a unique solution. The tuple \((X^{t,x}, Y^{t,x}, Z^{t,x})\) constitutes a forward backward doubly stochastic differential equation (FBDSDE).
Lemma 4.2
Assume $S_k$ and $D_k$ for $k \geq 3$ and $2a - F^* F \succeq 0$. Then the unique classical solution $\psi$ of the BSPDE in Eq. 4.1 is given by $\psi(t,x) = Y^{t,x}_t$, where $(Y^{t,x}, Z^{t,x})$ is the unique solution of the BDSDE in Eq. 4.2.

Proof. See [PP94, Theorem 3.1, p.225].

A final remark before we turn to the preliminary estimates and the main analysis where we will use BDSDEs, we will not be able to get an existence result for classical solutions of the SPDEs in Section 6 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 [Str08, Lemma 2.3.3]). We will instead use the existence result of [Roz90] stated in Lemma 4.1, and use the uniqueness result of [PP94] in our setting. This works under Lipschitz continuity of $a^{1/2}$, which we used previously.

5 Preliminary Estimates

In this section, we prove several preliminary estimates to be used in Section 6. We start with results for the moments of the SDE solutions.

5.1 Estimates on SDE Solutions

Lemma 5.1
Assume that the drift coefficient $b$, and dispersion coefficient $\sigma$, of the slow motion $X^\epsilon$ are bounded. Then for any $p \geq 1$, and every $T > 0$, there exists $C_p > 0$ such that

$$\sup_{(t,\epsilon) \in [0,T] \times (0,1]} E \|X^*_t\|^p | (X^\epsilon_0, Z^\epsilon_0) = (x,z) \| \leq C_p (1 + |x|^p).$$

Proof. The result is trivial since we assume the coefficients to be bounded and consider finite $T$.

Lemma 5.2
Assume $f$ is bounded and that $f$ and $gg^*$ are Hölder continuous in $z$ uniformly in $x$ for some uniform constant. Assume that the conditions $H_f$ and $H_g$ hold. Then for any $p > 0$ there exists $C_p > 0$ such that

$$\sup_{(t,\epsilon,x) \in [0,\infty] \times (0,1] \times \mathbb{R}^m} E \|Z^*_t\|^p | (X^\epsilon_t, Z^\epsilon_t) = (x,z) \| \leq C_p (1 + |z|^p).$$

Proof. The lemma is a slight generalization of a part of [Imk+13, Proposition 5.3, p.2307], and the proof follows the same argument as given there.

Lemma 5.3
Assume $h$ is bounded, then for $p \geq 1$ and $t \in [0,T],

$$\sup_{\epsilon \in (0,1]} \sup_{t \leq T} E_{P^\epsilon} \left| \tilde{D}^t \right|^p < \infty.$$ 

Proof. See the proof of [Imk+13, Lemma 6.5].

5.2 Estimates with the Fast Semigroup

In this section we provide estimates relating to the semigroup of the fast process.

Lemma 5.4
Assume $HF^{k,l}$, with $k \in \mathbb{N}_0, l \in \mathbb{N}$, and let $\theta \in C^{k,j}(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R})$ for $j \leq 1$ satisfy for some $C, p > 0$

$$\sum_{|\alpha| \leq k} \sum_{|\beta| \leq j} |D^\alpha \tilde{D}^\beta \theta(x,z)| \leq C (1 + |x|^p + |z|^p).$$
Then
\[(t,x,z) \mapsto T^F_t(x,\cdot) (z) \in \mathcal{C}^{0,k,j}(\mathbb{R}^+ \times \mathbb{R}^m \times \mathbb{R}^n; \mathbb{R})\]
and there exist $C_1, p_1 > 0$, such that for all $(t,x,z) \in [0,\infty) \times \mathbb{R}^m \times \mathbb{R}^n$
\[
\sum_{|\alpha| \leq k} \sum_{|\beta| \leq j} |D^\alpha_x D^\beta_z T^F_t(x,\cdot) (z)| \leq C_1 e^{C_1 t}(1 + |x|^{p_1} + |z|^{p_1}).
\]

If the bound on the derivatives of $\theta$ can be chosen uniformly in $x$, that is,
\[
\sum_{|\alpha| \leq k} \sup_x \sum_{|\beta| \leq j} |D^\alpha_x D^\beta_z \theta(x,z)| \leq C(1 + |x|^p),
\]
then the bound on the derivatives of $T^F_t(x,\cdot) (z)$ is also uniform in $x$,
\[
\sum_{|\alpha| \leq k} \sup_x \sum_{|\beta| \leq j} |D^\alpha_x D^\beta_z T^F_t(x,\cdot) (z)| \leq C_1 e^{C_1 t}(1 + |z|^{p_1}).
\]

Proof: The lemma is a slight generalization of [Imk+13, Proposition 5.1]. The proof is the same as in [Imk+13, Proposition 5.1].

Lemma 5.5
Assume $H_f$, $H_g$ and $HF^{k,3}$ for $k \in \mathbb{N}_0$. Let $\theta \in \mathcal{C}^{k,0}(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R})$ satisfy for some $C, p > 0$,
\[
\sum_{|\gamma| \leq k} \sup_x |D^\gamma_x \theta(x,z)| \leq C(1 + |z|^p).
\]
Then
\[
x \mapsto \mu_\infty(\theta; x)(x') = \int_{\mathbb{R}^n} \theta(x', z) \mu_\infty(dz; x) \in \mathcal{C}^k(\mathbb{R}^m; \mathbb{R}).
\]

Proof: See [Imk+13, Proposition 5.2], which contains the same statement.

Lemma 5.6
Assume $H_f$, $H_g$ and $HF^{k,3}$ with $k \in \mathbb{N}_0$. Let $\theta \in \mathcal{C}^{k,1}(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R})$ satisfy the growth condition,
\[
\sum_{|\alpha| \leq k} \sum_{|\beta| \leq 1} \sup_x |D^\alpha_x D^\beta_z \theta(x,z)| \leq C(1 + |z|^p),
\]
for some $C, p > 0$. Assume additionally that $\theta$ satisfies the centering condition,
\[
\int_{\mathbb{R}^m} \theta(x,z) \mu_\infty(dz; x) = 0, \quad \forall x \in \mathbb{R}^m.
\]
Then
\[
(x,z) \mapsto \int_0^\infty T^F_t(x,\cdot)(z) dt \in \mathcal{C}^{k,1}(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}),
\]
and for every $q > 0$ there exists $C', q' > 0$, such that,
\[
\sum_{|\alpha| \leq k} \sum_{|\beta| \leq 1} \int_0^\infty \left( \sup_x |D^\alpha_x D^\beta_z T^F_t(x,\cdot)(z)|^q dt \right) \leq C'(1 + |z|^{q'}).
\]

Proof: The lemma is a slight generalization of a part of [Imk+13, Proposition 5.2], and the proof follows the same argument as given there.

Lemma 5.7
Assume $H_f$, $H_g$ and $HF^{k,3}$ with $k \in \mathbb{N}_0$. If $HS^{k,1}$ holds, then $\tilde{b}, \tilde{\sigma}, \tilde{a} \in C^k_h$. Similarly, if $HO^{k,1}$ holds, then $\tilde{h} \in C^k_h$.

Proof: The result follows from Lemma 5.6.
5.3 Estimates for the Corrector Term

We now introduce a few lemmas that will help to streamline the main ideas in the analysis of the corrector term given in Lemma 6.2.

Lemma 5.8

Assume that $u \in C((0, T] \times \mathbb{R}^m; \mathbb{R})$ and is an element of $\mathcal{P}_T(\mathbb{R}^m; \mathbb{R})$, that the conditions $H_f$, $H_g$, and $HF^{0.3}$ hold, and that $\psi \in C^{0,1}_b(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R})$. Assume additionally that $\psi$ satisfies the centering condition,

$$\int_{\mathbb{R}^n} \psi(x, z) \mu_\infty(\mathrm{d}z; x) = 0, \quad \forall x \in \mathbb{R}^m.$$ 

Then given $(x, z) \in \mathbb{R}^m \times \mathbb{R}^n$ and $t \in [0, T]$, there exists $q > 0$ such that

$$\mathbb{E} \left[ \mathbb{E} \left[ \int_t^T \psi(x, Z^{r, x}(t, z)) u_s(x) \mathrm{d}s \left| \mathcal{F}^{B}_{t, T} \right. \right] \right]^p \lesssim \epsilon^p (1 + |z|^q) \mathbb{E} \left[ \sup_{t \leq s \leq T} |u_s(x)|^p \right].$$

Here $Z^{r, x}$ is the diffusion process with generator $\frac{1}{2} G_F$ (in particular, the Brownian motion driving $Z^{r, x}$ is independent of the Brownian motion $\mathcal{B}_t$ that generates the backward filtration $\mathcal{B}^{B}_{t, T}$).

Proof. Because $u_s$ is measurable with respect to $\mathcal{F}_{s, T}$ and $Z^{r, x}(t, z)$ is independent of $B$, we get from the conditional expectation with respect to $\mathcal{F}_{s, T}$ and definition of the semigroup $T^{F, x}$ the following identity

$$\mathbb{E} \left[ \int_t^T \psi(x, Z^{r, x}(t, z)) u_s(x) \mathrm{d}s \left| \mathcal{F}^{B}_{t, T} \right. \right] = \int_t^T \mathbb{E} \left[ \psi(x, Z^{r, x}(t, z)) \right] u_s(x) \mathrm{d}s = \int_t^T T^{F, x}_{(s-t)/\epsilon^2} \psi(x, \cdot) (z) u_s(x) \mathrm{d}s.$$

Now taking the absolute value, using Hölder’s inequality, and removing $|u_s(x)|$ from the integral by taking the supremum over $[t, T]$ gives

$$\left| \int_t^T T^{F, x}_{(s-t)/\epsilon^2} \psi(x, \cdot) (z) u_s(x) \mathrm{d}s \right| \lesssim \sup_{t \leq s \leq T} |u_s(x)| \int_t^T T^{F, x}_{(s-t)/\epsilon^2} \psi(x, \cdot) (z) \mathrm{d}s. \quad (5.1)$$

Performing a time reparametrization and then using Lemma 5.6, we have for some $q' > 0,$

$$\sup_{t \leq s \leq T} |u_s(x)| \int_t^T T^{F, x}_{(s-t)/\epsilon^2} \psi(x, \cdot) (z) \mathrm{d}s \lesssim \epsilon^2 \sup_{t \leq s \leq T} |u_s(x)| \int_0^{(T-t)/\epsilon^2} T^{F, x}_t \psi(x, \cdot) (z) \mathrm{d}r \lesssim \epsilon^2 \sup_{t \leq s \leq T} |u_s(x)| \int_0^{\infty} T^{F, x}_t \psi(x, \cdot) (z) \mathrm{d}r \lesssim \epsilon^2 (1 + |z|^{q'}) \sup_{t \leq s \leq T} |u_s(x)|.$$ 

Lastly, taking the $p$-th power and applying the expectation gives the desired result. □

Lemma 5.9

Assume that $u \in C((0, T] \times \mathbb{R}^m; \mathbb{R}^k)$ and is an element of $\mathcal{P}_T(\mathbb{R}^m; \mathbb{R}^k)$ for $k \geq 1$. Assume $H_f$, $H_g$ and $HF^{0.3}$. Let $\psi \in C^{0,1}_b(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^{d \times k})$. Assume additionally that $\psi$ satisfies the centering condition,

$$\int_{\mathbb{R}^n} \psi(x, z) \mu_\infty(\mathrm{d}z; x) = 0, \quad \forall x \in \mathbb{R}^m.$$ 

Then given $(x, z) \in \mathbb{R}^m \times \mathbb{R}^n$ and $t \in [0, T]$, there exists $q > 0$ such that

$$\mathbb{E} \left[ \mathbb{E} \left[ \int_t^T \psi(x, Z^{r, x}(t, z)) u_s(x, d\mathcal{B}) \left| \mathcal{F}^{B}_{t, T} \right. \right] \right]^p \lesssim \epsilon^p (1 + |z|^q) \mathbb{E} \left[ \sup_{t \leq s \leq T} |u_s(x)|^p \right].$$

Here $Z^{r, x}$ is the diffusion process with generator $\frac{1}{2} G_F$ (in particular, the Brownian motion driving $Z^{r, x}$ is independent of the Brownian motion $\mathcal{B}_t$ that generates the backward filtration $\mathcal{B}^{B}_{t, T}$).
Proof. Because $u_s$ is measurable with respect to $\mathcal{F}^B_{s,T}$ and $Z^r,x;ζ(t,z)$ is independent of $B$, we get from the conditional expectation with respect to $\mathcal{F}^B_{t,T}$ and definition of the semigroup $T^{F,x}$ the following identity

$$E\left[\int_t^T \langle \psi(x, Z^r,x;ζ(t,z))u_s(x), d\widehat{B}_s \rangle \right| \mathcal{F}^B_{t,T}] = \int_t^T \langle E[\psi(x, Z^r,x;ζ(t,z))] u_s(x), d\widehat{B}_s \rangle = \int_t^T \langle T^{F,x}_{(s-t)/\epsilon^2} (\psi(x, \cdot))(z)u_s(x), d\widehat{B}_s \rangle. $$

Now by application of the Burkholder-Davis-Gundy inequality we get

$$E\left[\left(\int_t^T \langle T^{F,x}_{(s-t)/\epsilon^2} (\psi(x, \cdot))(z)u_s(x), d\widehat{B}_s \rangle \right)^p \right] \leq E \left[ \left( \int_t^T \left( T^{F,x}_{(s-t)/\epsilon^2} (\psi(x, \cdot))(z)u_s(x) \right)^2 ds \right)^{p/2} \right].$$

Computing the quadratic variation gives

$$\left\langle \int_t^T \langle T^{F,x}_{(s-t)/\epsilon^2} (\psi(x, \cdot))(z)u_s(x), d\widehat{B}_s \rangle \right\rangle = \int_t^T |T^{F,x}_{(s-t)/\epsilon^2} (\psi(x, \cdot))(z)u_s(x)|^2 ds. \quad (5.2)$$

In the case that $u$ is real-valued, the integrand for the right side of Eq. (5.2) is bounded by

$$|T^{F,x}_{(s-t)/\epsilon^2} (\psi(x, \cdot))(z)u_s(x)|^2 \leq |u_s(x)|^2 |T^{F,x}_{(s-t)/\epsilon^2} (\psi(x, \cdot))(z)|^2. \quad (5.3)$$

Similarly, in the case where $u$ is an $\mathbb{R}^k$-valued vector for some $k > 1$, then $T^{F,x}_{(s-t)/\epsilon^2} (\psi(x, \cdot))(z)$ takes values in $\mathbb{R}^{d \times k}$, and letting $A_s \equiv T^{F,x}_{(s-t)/\epsilon^2} (\psi(x, \cdot))(z)$ for brevity, we have $|A_s u_s(x)|^2 \leq |u_s(x)|^2 \text{Tr}(A^*_s A_s) = |u_s(x)|^2 |A_s|^2$. Therefore we have the same inequality for the integrand.

Therefore using Eq. (5.3) in the quadratic variation of Eq. (5.2) and then taking the function $|u_s(x)|$ outside the integral by using its supremum value over $[t, T]$, we get

$$\left\langle \int_t^T \langle T^{F,x}_{(s-t)/\epsilon^2} (\psi(x, \cdot))(z)u_s(x), d\widehat{B}_s \rangle \right\rangle \leq \int_t^T |T^{F,x}_{(s-t)/\epsilon^2} (\psi(x, \cdot))(z)|^2 |u_s(x)|^2 ds$$

$$= \sup_{t \leq s \leq T} |u_s(x)|^2 \int_t^T |T^{F,x}_{(s-t)/\epsilon^2} (\psi(x, \cdot))(z)|^2 ds.$$

We now perform a time reparametrization and use Lemma 5.6, so that for some $q' > 0$ we get

$$\sup_{t \leq s \leq T} |u_s(x)|^2 \int_t^T |T^{F,x}_{(s-t)/\epsilon^2} (\psi(x, \cdot))(z)|^2 ds = \epsilon^2 \sup_{t \leq s \leq T} |u_s(x)|^2 \int_0^{(T-t)/\epsilon^2} |T^{F,x}_r (\psi(x, \cdot))(z)|^2 dr$$

$$= \epsilon^2 \sup_{t \leq s \leq T} |u_s(x)|^2 \int_0^{\infty} |T^{F,x}_r (\psi(x, \cdot))(z)|^2 dr$$

$$\leq \epsilon^2 (1 + |z|^q) \sup_{t \leq s \leq T} |u_s(x)|^2. \quad (5.4)$$

To see that the last step still holds in the case that $\psi$ is matrix-valued, consider the following relations

$$|T^{F,x}_r (\psi(x, \cdot))(z)|^2 = \sum_{i=1}^d \left( T^{F,x}_r (\psi(x, \cdot))(z) T^{F,x}_r (\psi(x, \cdot))(z)^* \right)_{ii}$$

$$= \sum_{i,j=1}^d \left( T^{F,x}_r (\psi(x, \cdot))(z) \right)_{ij}^2 = \sum_{i,j=1}^d |T^{F,x}_r (\psi(x, \cdot)_{ij})(z)|,$$

which shows that the same analysis holds, but now for a summation over the centered entries of $\psi$. Finally, taking the $p/2$ power and applying the expectation to Eq. (5.4) gives the desired result. \qed
6 Main Analysis

6.1 Moment Estimates for Dual Processes

In this section, we compute the main estimates for \(v^0, \psi\) and \(R\) associated with an arbitrary fixed test function \(\varphi \in C^2_b\). The estimates for \(\psi\) and \(R\) are then used in Section 6.2 to prove Theorem 2.1.

Lemma 6.1
Let \(3 \leq k \in \mathbb{N}\) and assume \(b, \sigma, \varphi \in C^k_b\) and \(\tilde{h}, \tilde{\varphi} \in C^{k+1}_b\). Then \(v^0 \in C^{0,k-1}([0, T] \times \mathbb{R}^m; \mathbb{R})\), and for any \(p \geq 1\) there exist \(q > 0\), such that for all \(x \in \mathbb{R}^m\),

\[
\sum_{|\lambda| \leq k-1} \mathbb{E} \left[ \sup_{0 \leq t \leq T} |D^\lambda v^0_t(x)|^p \right] \lesssim (1 + |x|^q) |\varphi|^p_{L^\infty}.
\]

In particular, \(v^0\) and all its partial derivatives up to order \((0, k-1)\) are in \(P_T(\mathbb{R}^m; \mathbb{R})\).

Proof. The result follows from Lemma 4.1. The only condition from Lemma 4.1 that is not immediately obvious is the parabolic condition, \(2\sigma - \sigma^* \alpha^* \sigma^* \geq 0\). This condition indeed holds for the same reason as given in Section 2.3.1 and the fact that \(I - \alpha^* \alpha \geq 0\), where \(I\) is the identity matrix (recall that \(\alpha\) was redefined in Section 2 as \(\alpha \leftarrow \kappa^{-1} \alpha\), and note that \((\kappa^{-1})^* = (\kappa^*)^{-1}\)).

Lemma 6.2
Let \(3 \leq k, l \in \mathbb{N}\) and assume \(H_f, H_g, H^{k,l}, H^{S,k,l}, HO^{k,l}\), and that \(\sigma, \alpha, \beta, \tilde{h} \in C^k_b\). Let \(v^0 \in C^{0,k}([0, T] \times \mathbb{R}^m; \mathbb{R})\), and assume that all its partial derivatives in \(x\) up to order \(k\) are in \(P_T(\mathbb{R}^m; \mathbb{R})\).

Then \(\psi \in C^{0,k-1}([0, T] \times \mathbb{R}^m \times \mathbb{R}^n; \mathbb{R})\), and \(\psi\) as well as its partial derivatives up to order \((0, k-1, l-1)\) are in \(P_T(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R})\). For any \(p \geq 1\) there exists \(q > 0\), such that for any \((x, z) \in \mathbb{R}^m \times \mathbb{R}^n\) and any \(\epsilon \in (0, 1)

\[
\sum_{|\beta| \leq k-1} \sup_{0 \leq t \leq T} \mathbb{E} \left[ |D^\beta \psi_t(x, z)|^p \right] \lesssim \epsilon^p (1 + |z|^q) \sum_{|\beta| \leq k} \mathbb{E} \left[ \sup_{0 \leq t \leq T} |D^\beta v^0_t(x)|^p \right].
\]

Proof. \(\psi_t(x, z)\) solves the following BSPDE

\[-d\psi_t(x, z) = \left[ \frac{1}{\epsilon^2} G_F \psi_t(x, z) + (G_S - \overline{G_S}) v^0_t(x, z) \right] dt + \langle v^0_t(h - \tilde{h})(x, z), dB_t \rangle + \langle \sigma(x, Z^t(x, z)) \rangle d\tilde{B}_t, \]

\[\psi_T(x, z) = 0.\]

Existence of the solution \(\psi\) and its derivatives as well as the polynomial growth follow from Lemma 4.1. From Lemma 4.2, the solution, \(\psi_t(x, z)\), has a representation in terms of a FBDSDE, \(\psi_t(x, z) = \theta_t^{x,z}\). Where \(\theta_t^{x,z}\) is a component of the pair of processes \((\theta, \gamma)\) satisfying the BDSDE

\[-d\theta^{x,z}_t = \left[ G_S(x, Z^{t,x}(t, z)) - \overline{G_S(x)} \right] \theta_t^{x}(x) ds

+ \langle v^0_t(x)(h(x, Z^{t,x}(t, z)) - \tilde{h}(x)), d\tilde{B}_s \rangle

+ \langle \sigma(x, Z^{t,x}(t, z)) - \sigma(x) \rangle \nabla_x v^0_t(x) d\tilde{B}_s

- \langle \gamma^{c,t}(t, x, z), dV_s \rangle, \]

\[\theta_T^{t,x,z} = 0, \]

and \((x, Z^{t,x}(t, z))\) is a joint diffusion process with \(X^{c,t}(t, x)\) having the zero generator,

\[X^{c,t}(t, x) = x, \quad \forall s \in [t, T],\]

and \(Z^{t,x}(t, z)\) satisfying the stochastic differential equation

\[dZ^{t,x}(t, z) = \frac{1}{\epsilon^2} f(x, Z^{t,x}(t, z)) ds + \frac{1}{\epsilon} g(x, Z^{t,x}(t, z)) dV_s, \quad s \geq t,\]

\[Z^{t,x}(t, z) = z, \quad s \leq t.\]
The second component of the pair \((\theta^t_{i,x,z}, \gamma^e_{i,t,x,z})\), has a representation as
\[
\gamma^e_{i,t,x,z} = \frac{1}{c} \theta^* \nabla_z \psi_t(x, z).
\]

For brevity, let us temporarily drop from the notation, superscripts and part of superscripts that indicate initial conditions (for example, \((t, x, z)\) and \((t, z)\)).

Since \(\psi_t\) is \(\mathcal{F}^B_{t,T}\)-measurable, so is \(\theta_t\), and therefore conditioning \(\theta_t\) on \(\mathcal{F}^B_{t,T}\) gives \(\theta_t = \mathbb{E}[\theta_t | \mathcal{F}^B_{t,T}]\). We also observe that \(V\) and \(B\) are independent. Therefore, \(V\) is a Brownian motion in the larger filtration \((\mathcal{F}^V + \mathcal{F}^B_{t,T})_{s \in [0,T]}\). Hence, from an application of the tower property of conditional expectation, we have
\[
\mathbb{E} \left[ \int_t^T \langle \gamma^e_{i,s}, dV_s \rangle \Big| \mathcal{F}^B_{t,T} \right] = \mathbb{E} \left[ \int_t^T \langle \gamma^e_{i,s}, dV_s \rangle \bigg| \mathcal{F}^V + \mathcal{F}^B_{t,T} \right] \bigg| \mathcal{F}^B_{t,T} = 0,
\]
and therefore,
\[
\theta_t = \mathbb{E} \left[ \int_t^T \left[ G_S(x, Z^t_{s,x}) - \overline{G}_S(x) \right] v^0_s(x) ds \bigg| \mathcal{F}^B_{t,T} \right] + \mathbb{E} \left[ \int_t^T \langle v^0_s(x) (h(x, Z^t_{s,x}) - \overline{h}(x)), d\overline{B}_s \rangle \bigg| \mathcal{F}^B_{t,T} \right] + \mathbb{E} \left[ \int_t^T \langle \alpha(\sigma(x, Z^t_{s,x}) - \overline{\sigma}(x))^* \nabla_x v^0_s(x), d\overline{B}_s \rangle \bigg| \mathcal{F}^B_{t,T} \right].
\]

The \(p\)-th moment is therefore bounded as follows,
\[
\mathbb{E} [\theta_t|^p] \leq \mathbb{E} \left[ \left\| \int_t^T \left[ G_S(x, Z^t_{s,x}) - \overline{G}_S(x) \right] v^0_s(x) ds \bigg| \mathcal{F}^B_{t,T} \right\|^p \right] + \mathbb{E} \left[ \left\| \int_t^T \langle v^0_s(x) (h(x, Z^t_{s,x}) - \overline{h}(x)), d\overline{B}_s \rangle \bigg| \mathcal{F}^B_{t,T} \right\|^p \right] + \mathbb{E} \left[ \left\| \int_t^T \langle \alpha(\sigma(x, Z^t_{s,x}) - \overline{\sigma}(x))^* \nabla_x v^0_s(x), d\overline{B}_s \rangle \bigg| \mathcal{F}^B_{t,T} \right\|^p \right].
\]

The first term on the right side of Eq. 6.3 has an integrand that can be written as
\[
\left[ G_S(x, Z^t_{s,x}) - \overline{G}_S(x) \right] v^0_s(x) = \sum_{i=1}^m (b - \bar{b})_i \frac{\partial}{\partial x_i} v^0_s(x, Z^t_{s,x}) + \frac{1}{2} \sum_{ij=1}^m (a - \bar{a})_{ij} \frac{\partial^2}{\partial x_i \partial x_j} v^0_s(x, Z^t_{s,x}),
\]

which shows that this term is a summation of terms that fit the conditions of Lemma 5.8 (i.e., a centered function driven by \(Z^t_{s,x}\) and multiplied with a term that has the correct bounds and measurability properties) and therefore we get for some \(q_0 > 0\) the following estimate for this term
\[
\mathbb{E} \left[ \left\| \int_t^T \left[ G_S(x, Z^t_{s,x}) - \overline{G}_S(x) \right] v^0_s(x) ds \bigg| \mathcal{F}^B_{t,T} \right\|^p \right] \leq c^{2p} (1 + |z|^q_0) \sum_{1 \leq |\beta| \leq 2} \mathbb{E} \left[ sup_{t \leq s \leq T} |D^\beta_x v^0_s(x)|^p \right].
\]

The second term, Eq. 6.4, fits the assumptions of Lemma 5.9, and therefore we get for some \(q_1 > 0\) the following estimate for this term
\[
\mathbb{E} \left[ \left\| \int_t^T \langle v^0_s(x) (h(x, Z^t_{s,x}) - \overline{h}(x)), d\overline{B}_s \rangle \bigg| \mathcal{F}^B_{t,T} \right\|^p \right] \leq c^{p} (1 + |z|^q_{1}) \mathbb{E} \left[ sup_{t \leq s \leq T} |v^0_s(x)|^p \right].
\]

Unlike the time integral term, we only get \(c^p\) for this estimate because of the application of the Burkholder-Davis-Gundy inequality in the proof of Lemma 5.9.
Lemma 5.9 also covers the case where the integrand of the stochastic integral is a matrix-vector product, as occurs in Eq. 6.5. Because each entry in Eq. 6.5 is centered and $v^0$ meets the required conditions of Lemma 5.9, we get for some $q_2 > 0$ the following estimate

$$
E \left[ \left| \frac{\partial}{\partial x_k} \theta_t \right|^p \right] \leq E \left[ \left| \frac{\partial}{\partial x_k} E \left\{ \int_t^T [\mathcal{G}_S(x, Z_s^{\tau-x}) - \mathcal{G}_S(x)] v_0^0(x) ds \mid F^B_{t,T} \right\} \right|^p \right]
$$

(6.10)

Collecting the estimates from Eqs. 6.6, 6.7, and 6.8, we get for some $q_3 > 0$ the following estimate for the BDSDE solution,

$$
E [ |\theta_t|^p ] \lesssim E [ (1 + |z|^q) \sum_{|\beta| = 1} E \left[ \sup_{t \leq s \leq T} |D^\beta v^0_0(x)|^p \right] ].
$$

(6.9)

We will also need estimates of the first and second-order derivatives of $\psi$ in the $x$-component for estimating the remainder term $R$ in Lemma 6.3 (see for instance Eq. 3.4). Therefore consider taking a first-order partial derivative of Eq. 6.2, and then taking the $p$-th moment, and separating terms on the right side of the equation by Hölder’s inequality,

$$
E \left[ \left| \frac{\partial}{\partial x_k} \theta_t \right|^p \right] \lesssim \sum_{1 \leq |\beta| \leq 2} E \left[ \left| \frac{\partial}{\partial x_k} E \left\{ \int_t^T \psi^\beta(x, Z_s^{\tau-x}) D^\beta x v_0^0(x) ds \mid F^B_{t,T} \right\} \right|^p \right],
$$

where $\psi^\beta$ is either an entry of $b - \bar{b}$ or $\frac{1}{2} (a - \bar{a})$, and hence centered. Now following the same arguments as in the proof of Lemma 5.8, we are able to get for any multiindex $1 \leq |\beta| \leq 2$,

$$
E \left[ \left| \frac{\partial}{\partial x_k} E \left\{ \int_t^T \psi^\beta(x, Z_s^{\tau-x}) D^\beta x v_0^0(x) ds \mid F^B_{t,T} \right\} \right|^p \right] = E \left[ \left| \frac{\partial}{\partial x_k} \int_t^T E \left[ \psi^\beta(x, Z_s^{\tau-x}) D^\beta x v_0^0(x) ds \right] \right|^p \right]
$$

(6.11)

$$
= E \left[ \left| \frac{\partial}{\partial x_k} \int_t^T E \left[ \psi^\beta(x, Z_s^{\tau-x}) \right] D^\beta x v_0^0(x) ds \right|^p \right].
$$

Distributing the derivative inside the time integral now gives (ignoring the $p$-th power and expectation for clarity in the next argument)

$$
E \left[ \left| \frac{\partial}{\partial x_k} \int_t^T T_{s-t}^{F,x}(\psi^\beta(x, \cdot))(z) D^\beta x v_0^0(x) ds \right|^p \right] \leq E \left[ \left| \int_t^T \frac{\partial}{\partial x_k} T_{s-t}^{F,x}(\psi^\beta(x, \cdot))(z) D^\beta x v_0^0(x) ds \right|^p \right]
$$

(6.12)

where $\psi^\beta$ is either an entry of $b - \bar{b}$ or $\frac{1}{2} (a - \bar{a})$, and hence centered. Now following the same arguments as in the proof of Lemma 5.8, we are able to get for any multiindex $1 \leq |\beta| \leq 2$,
Estimates for both terms are now achieved by applying the procedure in the proof of Lemma 5.8 starting from Eq. 5.1 onwards (and using the fact that Lemma 5.6 gives bounds for the derivative of the semigroup) to get for some $q_4 > 0$

\[
E \left[ \left| \frac{\partial}{\partial x_k} E \left[ \int_t^T \langle v^0(x)(h(x, Z^x_s) - \bar{h}(x)), d\hat{B}_s \rangle \bigg| \mathcal{F}_{t,T}^B \right] \right|^p \right] 
\]

\[
\lesssim 2^{p}(1 + |z|^{q_4}) \sum_{1 \leq |\beta| \leq 3} E \left[ \sup_{t \leq s \leq T} |D^\beta_x v^0(x)|^p \right]. \tag{6.13}
\]

Turning our attention now to Eq. 6.11, we follow the procedure of Lemma 5.9 to interchange the conditional expectation and stochastic integration, and then because of \( HO^{k+1,l+1} \), we can interchange ordinary differentiation and stochastic integration \([Kar83]\), and distribute the derivative to get

\[
E \left[ \left| \frac{\partial}{\partial x_k} E \left[ \int_t^T \langle v^0(x)(h(x, Z^x_s) - \bar{h}(x)), d\hat{B}_s \rangle \bigg| \mathcal{F}_{t,T}^B \right] \right|^p \right] 
\]

\[
\lesssim E \left[ \left| \int_t^T \left( \frac{\partial}{\partial x_k} v^0(x) T_{(s-t)/\epsilon^2}(h - \bar{h})(x, \cdot)(z), d\hat{B}_s \right) \right|^p \right] 
+ E \left[ \left| \int_t^T \left( v^0(x) \frac{\partial}{\partial x_k} T_{(s-t)/\epsilon^2}(h - \bar{h})(x, \cdot)(z) \right), d\hat{B}_s \right|^p \right].
\]

Estimates for both terms on the right side of the equation now follow from the argument in the proof of Lemma 5.9, starting from the application of the Burkholder-Davis Gundy inequality (and again using the fact that Lemma 5.6 gives bounds for the derivative of the semigroup), to yield for some $q_5 > 0$

\[
E \left[ \left| \frac{\partial}{\partial x_k} E \left[ \int_t^T \langle v^0(x)(h(x, Z^x_s) - \bar{h}(x)), d\hat{B}_s \rangle \bigg| \mathcal{F}_{t,T}^B \right] \right|^p \right] 
\]

\[
\lesssim e^p(1 + |z|^{q_5}) \sum_{|\beta| \leq 1} E \left[ \sup_{t \leq s \leq T} |D^\beta_x v^0(x)|^p \right]. \tag{6.14}
\]

The last term to address is Eq. 6.12. Just as we did when handling Eq. 6.11, we follow the procedure of Lemma 5.9 to interchange the conditional expectation and stochastic integration, and then because of \( HO^{k+1,l+1} \) interchange ordinary differentiation and stochastic integration \([Kar83]\), and distribute the derivative to get

\[
E \left[ \left| \frac{\partial}{\partial x_k} E \left[ \int_t^T \langle \alpha(x, Z^x_s) - \bar{\alpha}(x) \rangle \nabla_x v^0(x), d\hat{B}_s \rangle \bigg| \mathcal{F}_{t,T}^B \right] \right|^p \right] 
\]

\[
\lesssim E \left[ \left| \int_t^T \left( \frac{\partial}{\partial x_k} T_{(s-t)/\epsilon^2}(\alpha - \bar{\alpha})(x, \cdot)(z) \nabla_x v^0(x), d\hat{B}_s \right) \right|^p \right] 
+ E \left[ \left| \int_t^T \left( T_{(s-t)/\epsilon^2}(\alpha - \bar{\alpha})(x, \cdot)(z) \frac{\partial}{\partial x_k} v^0(x), d\hat{B}_s \right) \right|^p \right].
\]

Estimates for both terms on the right side of the equation now follow from the argument in the proof of Lemma 5.9, starting from the application of the Burkholder-Davis-Gundy inequality (and again using the fact that Lemma 5.6 gives bounds for the derivative of the semigroup), to yield for some $q_6 > 0$

\[
E \left[ \left| \frac{\partial}{\partial x_k} E \left[ \int_t^T \langle \alpha(x, Z^x_s) - \bar{\alpha}(x) \rangle \nabla_x v^0(x), d\hat{B}_s \rangle \bigg| \mathcal{F}_{t,T}^B \right] \right|^p \right] 
\]

\[
\lesssim e^p(1 + |z|^{q_6}) \sum_{|\beta| \leq 2} E \left[ \sup_{t \leq s \leq T} |D^\beta_x v^0(x)|^p \right]. \tag{6.15}
\]
Collecting the estimates from Eqs. 6.13, 6.14, and 6.15 then yields for some $q > 0$
\[
E \left[ \left| \frac{\partial}{\partial x_k} \theta_t \right|^p \right] \lesssim \epsilon^p (1 + |z|^q) \sum_{|\beta| \leq 3} E \left[ \sup_{0 \leq s \leq T} |D_x^\beta \psi_s(x)|^p \right].
\]

The procedure to take higher-order derivatives is the same as that for the first-order derivatives (simply involving more terms), and therefore taking the supremum of the estimates of these derivatives over $[0, T]$ and summing the terms, we get for some $q > 0$
\[
\sum_{|\beta| \leq k-1} \sup_{0 \leq t \leq T} E \left[ |D_x^\beta \theta_t|^p \right] \lesssim \epsilon^p (1 + |z|^q) \sum_{|\beta| \leq k+1} E \left[ \sup_{0 \leq t \leq T} |D_x^\beta \psi_t(x)|^p \right].
\]

**Lemma 6.3**

Let $3 \leq k, l \in \mathbb{N}$ and assume $HF^{k, l}$, $HS^{k, l}$, $\sigma \in C^{k+1, l+1}$, and $HO^{k+1, l+1}$. Let $\psi \in C^{0, k+2, l}([0, T] \times \mathbb{R}^m \times \mathbb{R}^n; \mathbb{R})$ and assume that all its partial derivatives up to order $(0, k+2, l)$ are in $P_T(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R})$. Then for any $p > 2$, we have that for any $(x, z) \in \mathbb{R}^m \times \mathbb{R}^n$, $\epsilon \in (0, 1)$, and $t \in [0, T]$,
\[
E \left[ |R_t(x, z)|^p \right] \lesssim \sum_{|\beta| \leq 2} \int_t^T E \left[ |D_x^\beta \psi_s(x', z')|^p \right]_{(x', z') = (X_s^{c(t, x), Z_s^{c(t, z)})} ds.
\]

**Proof.** $R_t(x, z)$ follows the solving BSPDE
\[
-dR_t = \langle \mathbf{G}^* R_t + \mathbf{G} \psi_t \rangle dt + \langle \psi_t + R_t \rangle h, d\widehat{B}_t \rangle + \langle \alpha \sigma^* \nabla_x \psi_t + R_t, d\widehat{B}_t \rangle = \epsilon \langle (\mathbf{G} \sigma)^* \psi_t + \mathbf{G} \mathbf{G}^* \psi_t + \mathbf{G} \psi_t + R_t, d\widehat{B}_t \rangle,
\]
with $R_T = 0$. (6.16)

Existence of the solution $R$ and its derivatives as well as the polynomial growth all follow from Lemma 4.1. The parabolic condition of Lemma 4.1 holds because $I - \alpha^* \alpha \geq 0$, where $I$ is the identity matrix. From Lemma 4.2, the solution, $R_t(x, z)$, has a representation in terms of a FBDSDE, $R_t(x, z) = \theta_t^{x, z}$. Where $\theta_t$ is the first component of the tuple of processes $(\theta_t, \gamma_t^{x, z}, \eta_t^{x, z})$ satisfying the BDSDE
\[
-d\theta_t^{x, z} = \mathbf{G} \psi_t(X_s^{c(t, x), Z_s^{c(t, z)})} ds
\]
and $X_s^{c(t, x), Z_s^{c(t, z)}}$ is a joint diffusion process satisfying the SDEs
\[
dX_s^{c(t, x), Z_s^{c(t, z)}} = b(X_s^{c(t, x), Z_s^{c(t, z)}, Z_s^{c(t, z)}) ds + \sigma(X_s^{c(t, x), Z_s^{c(t, z)}) dW_s, \quad s \geq t,
\]
$X_s^{c(t, x), Z_s^{c(t, z)}} = x, \quad s \leq t,$
\[
dZ_s^{c(t, z)} = \frac{1}{\epsilon^c} f(X_s^{c(t, x), Z_s^{c(t, z)}, Z_s^{c(t, z)}) ds + \frac{1}{\epsilon} g(X_s^{c(t, x), Z_s^{c(t, z)}) dV_s, \quad s \geq t,
\]
$Z_s^{c(t, z)} = z, \quad s \leq t,$
where we choose $(W, V)$ and $B$ to be independent standard Brownian motions. This is necessary when working with a stochastic representation of Eq. 6.16. The second and third components of the tuple $(\theta_t^{x, z}, \gamma_t^{x, z}, \eta_t^{x, z})$, have representations as
\[
\gamma_t^{x, z} = \frac{1}{\epsilon} g^* \nabla z R_t(x, z) \quad \text{and} \quad \eta_t^{x, z} = \sigma^* \nabla x R_t(x, z).
Let $\mathcal{A}$ be the integrand for the backward stochastic integral in Eq. 6.17; it takes the following definition

$$
\mathcal{A} = \psi_\theta h(X_s^{\tau(t,x)}, Z_s^{\tau(t,z)}) + \theta_\psi^{x,z} h(X_s^{\tau(t,x)}, Z_s^{\tau(t,z)}) + \alpha \sigma^x \nabla_x \psi_\theta (X_s^{\tau(t,x)}, Z_s^{\tau(t,z)}) + \alpha \eta_s^{t,x,z},
$$
or stripping function arguments and superscripts,

$$
\mathcal{A} = \psi_\theta h + \theta_\psi h + \alpha \sigma^x \nabla_x \psi_\theta + \alpha \eta_s.
$$

We now consider the $p$-th moment of $\theta$,

$$
\mathbb{E} [\|\theta^p\|] = \int_t^T \mathbb{E} [p|\theta|^{p-2} \theta G S \psi_\theta] \, ds + \frac{p(p-1)}{2} \int_t^T \mathbb{E} [\|\theta\|^{p-2} |\mathcal{A}|^2] \, ds - \frac{p(p-1)}{2} \int_t^T \mathbb{E} [\|\theta\|^{p-2} |\gamma_\theta|^2] \, ds.
$$

(6.18)

Using the fact that $\theta, \psi$ are real-valued functions, and $b, \sigma \in C_b^{k,l}$, applying Young’s inequality to the first term on the right side of Eq. 6.18 yields,

$$
\int_t^T \mathbb{E} [p|\theta|^{p-2} \theta G S \psi_\theta] \, ds \leq \frac{p}{2} \int_t^T \mathbb{E} [\|\theta\|^p] \, ds + \frac{p}{2} \int_t^T \mathbb{E} [\|\theta\|^{p-2} |\mathcal{A}|^2] \, ds.
$$

Application of Hölder’s inequality and Young’s inequality to the last term gives,

$$
\frac{p}{2} \int_t^T \mathbb{E} [\|\theta\|^{p-2} |\mathcal{A}|^2] \, ds \leq \frac{p-2}{2} \int_t^T \mathbb{E} [\|\theta\|^p] \, ds + \int_t^T \mathbb{E} [\|\mathcal{A}^p\|^2] \, ds.
$$

(6.19)

Application of Hölder’s inequality and the use of the boundness of $b$ and $a$, then the tower property of conditional expectation, and the Markov property of $(X', Z')$ gives the following bound for the last term in Eq. 6.19,

$$
\int_t^T \mathbb{E} [\|\mathcal{A}^p\|^2] \, ds \leq \int_t^T \sum_{|j| \leq 2} \mathbb{E} \left[ |D^j_x \psi(X_s^{\tau(t,x)}, Z_s^{\tau(t,z)})|^p \right] \, ds
$$

$$
= \int_t^T \sum_{|j| \leq 2} \mathbb{E} \left[ |D^j_x \psi(X_s^{\tau(t,x)}, Z_s^{\tau(t,z)})|^p \middle| \mathcal{F}_s^{W} \cup \mathcal{F}_s^{V} \right] \, ds
$$

$$
= \int_t^T \sum_{|j| \leq 2} \mathbb{E} \left[ |D^j_x \psi(x', z')|^p \right]_{(x', z') = (X_s^{\tau(t,x)}, Z_s^{\tau(t,z)})} \, ds.
$$

Therefore the first term on the right side of Eq. 6.18 is bounded by,

$$
\int_t^T \mathbb{E} [p|\theta|^{p-2} \theta G S \psi_\theta] \, ds \leq (p-1) \int_t^T \mathbb{E} [\|\theta\|^p] \, ds + \int_t^T \sum_{|j| \leq 2} \mathbb{E} \left[ |D^j_x \psi(x', z')|^p \right]_{(x', z') = (X_s^{\tau(t,x)}, Z_s^{\tau(t,z)})} \, ds.
$$

(6.20)

Now addressing the second term on the right side of Eq. 6.18, expanding the inner product $|\mathcal{A}|^2 = \langle \mathcal{A}, \mathcal{A} \rangle$ and separating terms using Young’s inequality with values $\lambda_1, \ldots, \lambda_6 > 0$ to be chosen later, we get

$$
\langle \mathcal{A}, \mathcal{A} \rangle \leq \left( 1 + \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3} \right) |\psi_\theta h|^2 + \left( 1 + \lambda_1 + \frac{1}{\lambda_4} + \frac{1}{\lambda_5} \right) |\theta_\psi h|^2
$$

$$
+ \left( 1 + \lambda_2 + \lambda_4 + \frac{1}{\lambda_6} \right) |\alpha \sigma^x \nabla_x \psi_\theta|^2 + (1 + \lambda_3 + \lambda_5 + \lambda_6) |\alpha \eta_s|^2.
$$
Therefore the second term on the right side of Eq. 6.18 is bounded by

$$\frac{p(p-1)}{2} \int_t^T \mathbb{E} \left[ |\theta_s|^{p-2} |A|^2 \right] ds \leq \left( 1 + \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3} \right) \frac{p(p-1)}{2} \int_t^T \mathbb{E} \left[ |\theta_s|^{p-2} |\psi_s h|^2 \right] ds$$

(6.21)

$$+ \left( 1 + \frac{1}{\lambda_4} \right) \frac{p(p-1)}{2} \int_t^T \mathbb{E} \left[ |\theta_s|^{p-2} |\psi_s h|^2 \right] ds$$

(6.22)

$$+ \left( 1 + \frac{1}{\lambda_5} \right) \frac{p(p-1)}{2} \int_t^T \mathbb{E} \left[ |\theta_s|^{p-2} |\alpha \sigma^* \nabla_x \psi_s|^2 \right] ds$$

(6.23)

$$+ \left( 1 + \frac{1}{\lambda_6} \right) \frac{p(p-1)}{2} \int_t^T \mathbb{E} \left[ |\theta_s|^{p-2} |\alpha \eta_s|^2 \right] ds.$$  

(6.24)

We now consider pairing the term given by Eq. 6.24 and the third term on the right side of Eq. 6.18,

$$\left( 1 + \lambda_3 + \lambda_5 + \lambda_6 \right) \frac{p(p-1)}{2} \int_t^T \mathbb{E} \left[ |\theta_s|^{p-2} |\alpha \eta_s|^2 \right] ds - \frac{p(p-1)}{2} \int_t^T \mathbb{E} \left[ |\theta_s|^{p-2} |\eta_s|^2 \right] ds$$

$$\equiv \Lambda$$

$$= \frac{p(p-1)}{2} \int_t^T \mathbb{E} \left[ |\theta_s|^{p-2} (\Lambda \eta_s^* \alpha \alpha^* \eta_s - \eta_s^* \text{Id} \eta_s) \right] ds$$

(6.25)

$$= \frac{p(p-1)}{2} \int_t^T \mathbb{E} \left[ |\theta_s|^{p-2} (\eta_s^* (\Lambda \alpha^* \alpha - \text{Id}) \eta_s) \right] ds.$$

The constant matrix $\alpha^* \alpha - \text{Id}$ is negative definite and we can choose $\lambda_3, \lambda_5, \lambda_6 > 0$, small enough such that $\Lambda \alpha^* \alpha - \text{Id} < 0$.

Turning our attention to the three terms of Eqs. 6.21, 6.22, and 6.23, we use the same technique as in Eq. 6.19 with $|h|_\infty < \infty$ on the first term (Eq. 6.21) to get

$$\left( 1 + \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3} \right) \frac{p(p-1)}{2} \int_t^T \mathbb{E} \left[ |\theta_s|^{p-2} |\psi_s h|^2 \right] ds \lesssim \int_t^T \mathbb{E} \left[ |\theta_s|^{p-2} |\psi_s h|^2 \right] ds$$

$$\lesssim \int_t^T \mathbb{E} \left[ |\theta_s|^p \right] ds + \int_t^T \sum_{|j| \leq 0} \mathbb{E} \left[ \left| D_x^j \psi(x', z') \right|^p \right]_{(x', z') = (X^x(t, s), Z^x(t, z))} ds.$$  

(6.26)

Similarly, again using $|h|_\infty < \infty$, Eq. 6.22 is bounded by

$$\left( 1 + \frac{1}{\lambda_4} + \frac{1}{\lambda_5} \right) \frac{p(p-1)}{2} \int_t^T \mathbb{E} \left[ |\theta_s|^{p-2} |\psi_s h|^2 \right] ds \lesssim \int_t^T \mathbb{E} \left[ |\theta_s|^p \right] ds.$$  

(6.27)

And now Eq. 6.23 using $|\sigma|_\infty < \infty$,

$$\left( 1 + \lambda_2 + \lambda_4 + \frac{1}{\lambda_6} \right) \frac{p(p-1)}{2} \int_t^T \mathbb{E} \left[ |\theta_s|^{p-2} |\alpha \sigma^* \nabla_x \psi_s|^2 \right] ds$$

$$\lesssim \int_t^T \mathbb{E} \left[ |\theta_s|^p \right] ds + \int_t^T \sum_{|j| \leq 1} \mathbb{E} \left[ \left| D_x^j \psi(x', z') \right|^p \right]_{(x', z') = (X^x(t, s), Z^x(t, z))} ds.$$  

(6.28)

Collecting the bounds of Eqs. 6.20, 6.25, 6.26, 6.27, and 6.28 we get for Eq. 6.18,

$$\mathbb{E} \left[ |\theta_t|^p \right] \lesssim \int_t^T \mathbb{E} \left[ |\theta_s|^p \right] ds + \int_t^T \sum_{|j| \leq 2} \mathbb{E} \left[ \left| D_x^j \psi(x', z') \right|^p \right]_{(x', z') = (X^x(t, s), Z^x(t, z))} ds$$

$$+ \frac{p(p-1)}{2} \int_t^T \mathbb{E} \left[ |\theta_s|^{p-2} (\eta_s^* (\Lambda \alpha^* \alpha - \text{Id}) \eta_s) \right] ds$$

$$- \frac{p(p-1)}{2} \int_t^T \mathbb{E} \left[ |\theta_s|^2 |\gamma_s|^2 \right] ds.$$
Rearranging this equation gives
\[
\mathbb{E} [\|\theta_t\|^p] - \frac{p(p-1)}{2} \int_t^T \mathbb{E} [\|\theta_s\|^{p-2} (\eta_s^\ast (\Lambda \alpha - \text{Id}) \eta_s)] \, ds + \frac{p(p-1)}{2} \int_t^T \mathbb{E} [\|\theta_s\|^{p-2} |\gamma_s|^2] \, ds
\]
\[
\lesssim \int_t^T \mathbb{E} [\|\theta_s\|^p] \, ds + \int_t^T \sum_{|j| \leq 2} \mathbb{E} \left[ \mathbb{E} \left[ |D_x^j \psi(x', z')|^p \right]_{(x', z') = (X_s^c(t, z), Z_s^c(t, z))} \right] \, ds.
\]
From the fact that $\Lambda \alpha - \text{Id} \prec 0$, the subtraction of the second term on the left side of the equation is a non-negative value. The third term on the left side of the equation is also non-negative, and therefore we can drop them from the inequality to get
\[
\mathbb{E} [\|\theta_t\|^p] \lesssim \int_t^T \mathbb{E} [\|\theta_s\|^p] \, ds + \int_t^T \sum_{|j| \leq 2} \mathbb{E} \left[ \mathbb{E} \left[ |D_x^j \psi(x', z')|^p \right]_{(x', z') = (X_s^c(t, z), Z_s^c(t, z))} \right] \, ds.
\]
Now applying Grönwall’s lemma yields,
\[
\mathbb{E} [\|\theta_t\|^p] \lesssim \int_t^T \sum_{|j| \leq 2} \mathbb{E} \left[ \mathbb{E} \left[ |D_x^j \psi(x', z')|^p \right]_{(x', z') = (X_s^c(t, z), Z_s^c(t, z))} \right] \, ds.
\]
Using the fact that the solution to the BDSDE provides the classical solution to the BSPDE, $R_t(x, z) = \theta_t^{c, x, z}$, we get the desired result.

6.2 Estimates of Dual and Filter Error
We now complete the final estimates that lead to the proof of Theorem 2.1. The remaining lemmas parallel those in [Imk+13]; we provide the statements, which have slightly different assumptions, but otherwise refer to [Imk+13] for the proofs.

Lemma 6.4
Assume $H_f, H_g, HF^{8,4}, b \in C^7_b, \sigma \in C^8_b, HO^{8,4}$, and $\varphi \in C^0_b(\mathbb{R}^m; \mathbb{R})$. Then for every $p \geq 1$ there exists $q > 0$, such that
\[
\sup_{0 \leq t \leq T} \mathbb{E}_\mathbb{P} \left[ |v_t^{c, x, \varphi}(x, z) - v_t^{0, x, \varphi}(x)|^p \right] \lesssim \epsilon^p (1 + |x|^q + |z|^q)|\varphi|_{4, \infty}^p.
\]
Proof. First we collect the conditions in reverse order of our main estimates.

(i) For the solution of $R$ in Lemma 6.3, we require $HF^{3,3}, HS^{3,3}, \sigma \in C^{1,4}_b, HO^{1,4}$, and $\psi \in C^{0,5,3}$. The polynomial growth will be satisfied.

(ii) For the solution of $\psi \in C^{0,5,3}$ in Lemma 6.2, we require $HF^{6,4}, HS^{6,4}, HO^{6,4}$, and $v^0 \in C^{0,6}$. The conditions on $\sigma, \bar{a}, \bar{b}$, and $\bar{h}$ will already be covered by the stronger conditions just stated. The polynomial growth will also be satisfied. And we require $H_f$ and $H_g$.

(iii) For the solution of $v^0 \in C^{0,6}$ in Lemma 6.1, we require $\bar{b}, \bar{a} \in C^7_b$, that $\bar{h}, \sigma \in C^8_b$, and $\varphi \in C^7_b$.

(iv) Using Lemma 5.7, for $\bar{h}, \sigma \in C^8_b$ requires $HF^{8,3}, HO^{8,1}$, and $\sigma \in C^{8,1}_b$. And this also implies $\sigma \in C^7_b$. For $\bar{b} \in C^7_b$, we need $b \in C^{7,1}_b$. We also require $H_f$ and $H_g$.

(v) Therefore the sufficient conditions are $H_f, H_g, HF^{8,4}, b \in C^7_b, \sigma \in C^{8,4}_b, HO^{8,4}$, and $\varphi \in C^7_b$. For the remainder of the proof, see [Imk+13, Lemma 6.4, p.2318].

We now show that the moment estimate of the difference of $v^c$ and $v^0$ continues to hold under the original measure $\mathbb{Q}$. 

24
Lemma 6.5
Assume $H_f$, $H_g$, $HF^{8,4}$, $b \in C_b^{7,4}$, $\sigma \in C_b^{8,4}$, $HO^{8,4}$, and $\varphi \in C_b^0(\mathbb{R}^m; \mathbb{R})$. Then for every $p \geq 1$ there exists $q > 0$, such that
\[
\sup_{0 \leq t \leq T} E_Q \left[ |\nu_t^{0,T,\varphi}(x, z) - \nu_t^{0,0,\varphi}(x)|^p \right] \lesssim \epsilon^p (1 + |x|^q + |z|^q) |\varphi|_{4,\infty}^p.
\]
Proof. Using Lemma 5.3, the proof follows [Imk+13, Lemma 6.5, p.2319].

Lemma 6.6
Assume $H_f$, $H_g$, $HF^{8,4}$, $b \in C_b^{7,4}$, $\sigma \in C_b^{8,4}$, $HO^{8,4}$, and $\varphi \in C_b^0(\mathbb{R}^m; \mathbb{R})$. Additionally, assume that the initial distribution $Q_{(x_0, z_0)}$ has finite moments of every order. Then for every $p \geq 1$ there exists $q > 0$, such that
\[
E_Q \left[ |\rho_t^{0,i}(\varphi) - \rho_t^{0}(\varphi)|^p \right] \lesssim \epsilon^p |\varphi|_{4,\infty}^p.
\]
Proof. The proof is the same as [Imk+13, Lemma 6.6, p.2320].

Lemma 6.7
Let $p \geq 1$ and assume $h$ is bounded. Then
\[
\sup_{\epsilon \in (0,1]} \sup_{0 \leq t \leq T} \left( E_Q \left[ |\rho_t^{0,i}(1)|^{-p} \right] + E_Q \left[ |\rho_t^{0}(1)|^{-p} \right] \right) < \infty.
\]
Proof. For the first term, the proof is the same as [Imk+13, Lemma 6.7, p.2321]. The same is true for the second term using the definitions given in Lemma 2.1.

Lemma 6.8
Assume $H_f$, $H_g$, $HF^{8,4}$, $b \in C_b^{7,4}$, $\sigma \in C_b^{8,4}$, $HO^{8,4}$, and $\varphi \in C_b^0(\mathbb{R}^m; \mathbb{R})$. Additionally, assume that the initial distribution $Q_{(x_0, z_0)}$ has finite moments of every order. Then for every $p \geq 1$ there exists $q > 0$, such that
\[
E_Q \left[ |\pi_t^{0,i}(\varphi) - \pi_t^{0}(\varphi)|^p \right] \lesssim \epsilon^p |\varphi|_{4,\infty}^p.
\]
Proof. The proof is the same as [Imk+13, Lemma 6.8, p.2321].

Observing that the bound in the result of Lemma 6.8 only depends on $|\varphi|_{4,\infty}^p$, even though the assumption requires $\varphi \in C_b^0$, encourages us to instead approximate a fixed test function $\varphi \in C_b^0$ by a sequence $(\varphi^n \in C_b^0)$ in the $|\cdot|_{4,\infty}$-norm, and take advantage of the fact that $\pi_t^{0,i}$ and $\pi_t^{0}$ are $\mathbb{Q}$-a.s. equal to probability measures. Therefore we can relax this condition in Lemma 6.8 slightly with the following corollary.

Corollary 6.1
Assume $H_f$, $H_g$, $HF^{8,4}$, $b \in C_b^{7,4}$, $\sigma \in C_b^{8,4}$, and $HO^{8,4}$. Additionally, assume that the initial distribution $Q_{(x_0, z_0)}$ has finite moments of every order. Then for any $p \geq 1$ we have that for every $\varphi \in C_b^0(\mathbb{R}^m; \mathbb{R})$,
\[
E_Q \left[ |\pi_t^{0,i}(\varphi) - \pi_t^{0}(\varphi)|^p \right] \lesssim \epsilon^p |\varphi|_{4,\infty}^p.
\]

The next lemma shows that indeed we have weak convergence of $\pi_t^{0,i}$ to $\pi^n$.

Lemma 6.9
Assume $H_f$, $H_g$, $HF^{8,4}$, $b \in C_b^{7,4}$, $\sigma \in C_b^{8,4}$, and $HO^{8,4}$. Additionally, assume that the initial distribution $Q_{(x_0, z_0)}$ has finite moments of every order. Then there exists a metric $d$ on the space of probability measures on $\mathbb{R}^m$ that generates the topology of weak convergence, such that
\[
E_Q \left[ d(\pi_t^{0,i}, \pi_t^{0}) \right] \lesssim \epsilon.
\]
Proof. To achieve this result, we borrow the argument following [Imk+13, Corollary 6.9, p.2322].
Acknowledgement

R.B. and N.S.N. acknowledge partial support for this work from the Air Force Office of Scientific Research under grant number FA9550-17-1-0001, and N.S.N. acknowledges partial support from the National Sciences and Engineering Research Council Discovery grant 50503-10802.

References

[BB86] A. Bensoussan and G. L. Blankenship. “Nonlinear filtering with homogenization”. In: Stochastics 17 (1986), pp. 67–90. DOI: 10.1080/17442508608833383 (cit. on p. 2).

[BC09] Alan Bain and Dan Crisan. Fundamentals of Stochastic Filtering. Springer, 2009. ISBN: 978-0-387-76896-0 DOI: 10.1007/978-0-387-76896-0 (cit. on p. 8).

[BH14] Tyrus Berry and John Harlim. “Linear Theory for Filtering Nonlinear Multiscale Systems with Model Error”. In: Proceedings of the Royal Society A 470 (July 2014). DOI: 10.1098/rspa.2014.0168 (cit. on p. 2).

[BNP20] Ryne Beeson, N. Sri Namachchivaya, and Nicolas Perkowski. Approximation of the Filter Equation for Multiple Timescale, Correlated, Nonlinear Systems. Oct. 2020. arXiv: 2010.16401 [math.PR] (cit. on p. 2).

[Ich04] Naoyuki Ichihara. “Homogenization Problem for Stochastic Partial Differential Equations of Zakai Type”. In: Stochastic and Stochastics Reports 76.3 (2004), pp. 243–266. DOI: 10.1080/10451120410001714107 (cit. on p. 2).

[Imk+13] Peter Imkeller et al. “Dimensional reduction in nonlinear filtering: A homogenization approach”. In: Ann. Appl. Probab. 23.6 (Dec. 2013), pp. 2290–2326. DOI: 10.1214/12-AAP901 (cit. on pp. 2, 6, 11, 13, 14, 24, 25).

[Kar83] Rajeeva L. Karandikar. “Interchanging the Order of Stochastic Integration and Ordinary Differentiation”. In: Indian Statistical Institute 45.1 (1983), pp. 120–124. URL: http://www.jstor.org/stable/25050420 (cit. on p. 20).

[KH12] Emily L. Kang and John Harlim. “Filtering Partially Observed Multiscale Systems with Heterogeneous Multiscale Methods-Based Reduced Climate Models”. In: Monthly Weather Review 140 (Mar. 2012), pp. 860–873. DOI: 10.1175/MWR-D-10-05067.1 (cit. on p. 2).

[KLS97] M. L. Kleptsina, R. Sh. Lipster, and A. P. Serebrovski. “Nonlinear Filtering Problem with Contamination”. In: The Annals of Applied Probability 7.4 (1997), pp. 917–934. URL: http://www.jstor.org/stable/2245252 (cit. on p. 2).

[Kus90] Harold Kushner. Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems. Birkhauser Basel, 1990. DOI: 10.1007/978-1-4612-4482-0 (cit. on p. 2).

[KY05] R. Z. Khasminskii and G. Yin. “Limit behavior of two-time scale diffusions revisited”. In: Journal of Differential Equations 212 (2005), pp. 85–113. DOI: https://doi.org/10.1016/j.jde.2004.08.013 (cit. on p. 1).

[LH03] Vladimir M. Lucic and Andrew J. Heunis. “Convergence of Nonlinear Filters for Randomly Perturbed Dynamical Systems”. In: Applied Mathematics and Optimization 48.2 (2003), pp. 93–128. DOI: 10.1007/s00245-003-0772-8 (cit. on p. 2).

[Par80] E. Pardoux. “Stochastic partial differential equations and filtering of diffusion processes”. In: Stochastics 3.1-4 (1980), pp. 127–167. DOI: 10.1080/17442507908833142 (cit. on p. 8).

[PNY11] Jun Hyun Park, N. Sri Namachchivaya, and Hoong Chieh Yeong. “Particle Filters In a Multiscale Environment: Homogenized Hybrid Particle Filter”. In: Journal of Applied Mechanics 78 (Nov. 2011), pp. 061001–061001-10. DOI: doi:10.1115/1.4003167 (cit. on p. 2).

[PP94] Etienne Pardoux and Shige Peng. “Backward doubly stochastic differential equations and systems of quasilinear SPDEs”. In: Probability Theory and Related Fields 98.2 (June 1994), pp. 209–227. DOI: 10.1007/BF01192514 (cit. on pp. 11–13).
[PSN10] J H Park, R B Sowers, and N Sri Namachchivaya. “Dimensional reduction in nonlinear filtering”. In: Nonlinearity 23.2 (Jan. 2010), pp. 305–324. doi: 10.1088/0951-7715/23/2/005 (cit. on p. 2).

[PSV76] George C. Papanicolaou, Danial Stroock, and S. R. S. Varadhan. “Martingale approach to some limit theorems”. In: Papers from the Duke Turbulence Conference. Duke University, Durham, North Carolina, 1976 (cit. on p. 1).

[PV03] E. Pardoux and A. Yu. Veretennikov. “On Poisson equation and diffusion approximation 2”. In: Ann. Probab. 31.3 (July 2003), pp. 1166–1192. doi: 10.1214/aop/1055425774 (cit. on p. 1).

[Qia19] Huijie Qiao. “Convergence of Nonlinear Filterings for Multiscale Systems with Correlated Sensor Lévy Noises”. In: ArXiv e-prints (2019). url: https://arxiv.org/abs/1910.09265v1 (cit. on p. 2).

[Roz90] Boris L. Rozovskii. Stochastic Evolution Systems: Linear Theory and Applications to Non-linear Filtering. Dordrecht: Kluwer Academic Publishers, 1990, p. 315. isbn: 0-7923-0037-8 (cit. on pp. 11, 13).

[Roz91] Boris L. Rozovskii. “Stochastic Analysis”. In: ed. by Eddy Mayer-Wolf, Ely Merzbach, and Adam Shwartz. Academic Press, Inc., 1991, pp. 449–458 (cit. on p. 8).

[Str08] Daniel W. Stroock. Partial Differential Equations for Probabilists. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2008. DOI: 10.1017/CBO9780511755255 (cit. on pp. 12, 13).

[Yeo+20] Hoong C. Yeong et al. “Particle Filters with Nudging in Multiscale Chaotic Systems: With Application to the Lorenz '96 Atmospheric Model”. In: Journal of Nonlinear Science 30.4 (2020), pp. 1519–1552. DOI: 10.1007/s00332-020-09616-x (cit. on p. 2).

[ZR19] Yanjie Zhang and Jian Ren. “Data Assimilation for a Multiscale Stochastic Dynamical System with Gaussian Noise”. In: Stochastics and Dynamics 19.3 (2019). DOI: 10.1142/S0219493719500199 (cit. on p. 2).