# Singular Limit of BSDEs and Optimal Control of two Scale Stochastic Systems in Infinite Dimensional Spaces

Giuseppina Guatteri,
Dipartimento di Matematica,
Politecnico di Milano,
Piazza Leonardo da Vinci 32,
20133 Milano, Italia.
e-mail: giuseppina.guatteri@polimi.it

Gianmario Tessitore,
Dipartimento di Matematica e Applicazioni
Università Milano-Bicocca,
via R. Cozzi 53 - Edificio U5,
20125 Milano, Italia.
e-mail: gianmario.tessitore@unimib.it

**Abstract**

In this paper we study by probabilistic techniques the convergence of the value function for a two-scale, infinite-dimensional, stochastic controlled system as the ratio between the two evolution speeds diverges. The value function is represented as the solution of a *backward stochastic differential equation* (BSDE) that it is shown to converge towards a *reduced* BSDE. The noise is assumed to be additive both in the slow and the fast equations for the state. Some non degeneracy condition on the slow equation is required. The limit BSDE involves the solution of an *ergodic* BSDE and is itself interpreted as the value function of an auxiliary stochastic control problem on a reduced state space.

## 1 Introduction

The purpose of this paper is to give a representation of the limit of the value functions of a sequence of optimal control problems for a singularly perturbed infinite dimensional state equation. Namely we consider the following system of controlled stochastic differential equations:

$$\begin{cases} dX_t^{\varepsilon,\alpha} = AX_t^{\varepsilon,\alpha} + b(X_t^{\varepsilon,\alpha}, Q_t^{\varepsilon,\alpha}, \alpha_t)dt + Rd\mathcal{W}_t^1, & X_0 = x_0, \\ \varepsilon dQ_t^{\varepsilon,\alpha} = (BQ_t^{\varepsilon,\alpha} + F(X_t^{\varepsilon,\alpha}, Q_t^{\varepsilon,\alpha}))\, dt + G\rho(\alpha_t)dt + \varepsilon^{1/2} G\, d\mathcal{W}_t^2, & Q_0 = q_0, \end{cases} \tag{1.1}$$

where both state components $X^{\varepsilon,\alpha}$ and $Q^{\varepsilon,\alpha}$ take values in a Hilbert space. In the above equation $A$ and $B$ are unbounded linear operators, $\alpha$ represents the control, $(\mathcal{W}_t^1)_{t\geq 0}$, $(\mathcal{W}_t^2)_{t\geq 0}$ are infinite dimensional cylindrical Wiener processes, $b$, $F$, $\rho$ are functions and $R$ and $G$ are bounded linear operators satisfying suitable assumptions. We notice that the presence of the constant $\varepsilon$ in the second equation corresponds to the fact that $Q$ evolves with a speed which is larger by a factor $1/\varepsilon$ then the speed of evolution of the component $X$. In other words the above equation is a good model for a so called *two scale system*. The optimal control problem is then completed by a standard cost functional of the form:

$$J^{\varepsilon}(x_0, q_0, \alpha) := \mathbb{E}\left(\int_0^1 l(X_t^{\varepsilon,\alpha}, Q_t^{\varepsilon,\alpha}, \alpha_t)dt + h(X_1^{\varepsilon,\alpha})\right), \tag{1.2}$$

and the value function is defined in the usual way:

$$V^{\varepsilon}(x_0, q_0) := \inf_{\alpha} J^{\varepsilon}(x_0, q_0, \alpha), \tag{1.3}$$

where the infimum is extended over a suitable class of progressively measurable control processes $(\alpha)$.
Our purpose is to give a characterization of the limit of $V^{\varepsilon}(x_0, q_0)$ as $\varepsilon$ (that is the ratio between the speed of slow and the quick evolution) converges to 0.

Several authors have studied the convergence of singular stochastic control problems in finite dimensional spaces, see for instance [2], [3], [18], [19], [21]. In particular [2] has been an inspiration for the present work. In that paper the authors represent the value function of a singular stochastic control problem, in finite dimensions, by the solution, in viscosity sense, of an Hamilton-Jacobi-Bellman equation. Then they show, by PDE methods their convergence towards the solution, again in viscosity sense, of a *reduced* parabolic PDE with smaller state space and a new nonlinearity usually called *effective Hamiltonian*. Such analysis is performed in the case of periodic boundary conditions. Although PDE techniques perfectly fit the finite dimensional case allowing to cover general situations (including state equations with control dependent diffusions) they are not adaptable to this infinite dimensional case and consequently to the case of two scale stochastic control problems for stochastic PDEs. Namely comparison of viscosity solution would require, in infinite dimensional frameworks, additional assumptions such as a trace class noise and strong regularity of solutions (see [11]) that would hold only in special situations and, in any case, woluld not allow to cover our case where we consider cylidrical Wiener noises (see, as well, the discussion in the Introduction of [16] about use of viscosity solutions in infinite dimensional control problems).

In this paper we choose a completely different approach based on Backward Stochastic Differential Equations, BSDEs in short, (see [22], and [16] as a reference, respectively, for the finite and infinite dimensional case) that has already proved to be well adapted to infinite dimensional extensions. This choice eventually allows us to give a representation of the limit of $V^{\varepsilon}(x_0, q_0)$ (see (1.3)) in a general Hilbertian framework that constitutes, at the best of our knowledge, the first result in this direction. Moreover our assumptions are general enough to cover a pretty large class of two scale systems of controlled partial differential equations, possibly driven by cylindrical Wiener processes (see, for instance, the system of controlled reaction diffusion equations driven by space-time white noise in Example 6.6). As a counterpart we notice that we consider state equation in which the control only affects the drift and in which the noise of the slow component is assumed to be non-degenerate.

We try now to give a few more details on our method and results.
We will solve the control problem in the *weak formulation*, see [13]. This means, in particular, that the Wiener process will not be fixed (still the formulation allows to fix a priori a filtration $(\mathcal{F}_t)$ and work with a $(\mathcal{F}_t)$ Wiener process, (see Remark (6.3) and Section 6 here, in additon section 7 in [16] for further considerations).
To start with we consider, for each $\varepsilon > 0$, the following uncontrolled *forward-backward* system:

$$\begin{cases} dX_t = AX_t\, dt + R\, dW_t^1, \\ \varepsilon dQ_t^{\varepsilon} = (BQ_t^{\varepsilon} + F(X_t, Q_t^{\varepsilon}))\, dt + \varepsilon^{1/2}\, G\, dW_t^2, \\ -dY_t^{\varepsilon} = \psi(X_t, Q_t^{\varepsilon}, Z_t^{\varepsilon}, \Xi_t^{\varepsilon}/\sqrt{\varepsilon})\, dt - Z_t^{\varepsilon} d\, W_t^1 - \Xi_t^{\varepsilon} dW_t^2, \\ X_0 = x_0 \quad Q_0^{\varepsilon} = q_0, \quad Y_1^{\varepsilon} = h(X_1), \end{cases} \tag{1.4}$$

where $\psi$ will eventually be the Hamiltonian corresponding of the stochastic control problem:

$$\psi(x,p,z,\xi) = \inf_{\alpha\in U}\{l(x,q,\alpha) + z[R^{-1}b(x,q,\alpha)] + \xi\rho(\alpha)\}.$$

Then, once we have a solution $(X, Q^\varepsilon, Y^\varepsilon, Z^\varepsilon)$ to system (1.4), we exploit the well known identification between $Y_0^\varepsilon$ and $V^\varepsilon(x_0, q_0)$ (see [10] or [16]) in order to study the limit of the value functions by the limit of the sequence $Y_0^\varepsilon$ as $\varepsilon \to 0$. Our main result is indeed stated in terms of $Y^\varepsilon$, that is, see Theorem 5.4, we prove that:

$$Y_0^\varepsilon \to \bar{Y}_0, \qquad \mathbb{P} - a.s.$$

where $(X, \bar{Y}, \bar{Z})$ is the unique solution of the following decoupled forward backward system of stochastic differential equations:

$$\begin{cases} dX_t = AX_t\,dt + R\,dW_t^1, \\ -d\bar{Y}_t = \lambda(X_t, \bar{Z}_t)\,dt - \ \bar{Z}_t\,dW_t^1, \\ X_0 = x_0, \quad \bar{Y}_1 = h(X_1). \end{cases}$$

The statement of the above mentioned result is formulated and proved in Section 5 as a general result on singular limits of BSDEs since it is independent of its control theoretic interpretation and, we believe, the proving argument has some interest on its own. It is worth mentioning that the 'reduced nonlinearity' $\lambda$ is itself a component of the unique solution $(\check{Y}, \check{Z}, \lambda)$ of the parametrized version of a, so called, *Ergodic* BSDE (see (4.1) and Theorem 4.2) similar to the ones introduced in [15] (see [9] and [20] as well). Function $\lambda$ can also be interpreted as the optimal cost of an ergodic optimal control problem, see Remark 6.5. Moreover, as it happens in the finite dimensional case, the space in which the above reduced forward-backward system lives is a subspace of the original one (corresponding to the slow evolution alone).
As a by-product of our main result, using the Bismut Elworthy formula in [17] we immediately get that the solution of the reduced BSDE, and therefore the limit value function, depends on $x_0$ in a differentiable way and is linked to the unique *mild* solution of a semilinear parabolic PDE in infinite dimensional spaces:

$$\begin{cases} \dfrac{\partial v(t,x)}{\partial t} + \dfrac{1}{2}\mathrm{Tr}[RR^*\nabla_x^2 v(t,x)] + \langle Ax, \nabla v(t,x)\rangle = \lambda(x, \nabla v(t,x)R), \qquad t\in[0,1],\, x\in H, \\ v(1,x) = h(x). \end{cases}$$

Finally, in the last section, exploiting the concavity of $\lambda$ we give a representation of $\bar{Y}_t$ as the value function of an auxiliary stochastic control problem on a reduced state space.

The paper is organized in the following way. In Section 2 we set the notation and we introduce some functional spaces while Section 3 contains some estimates on the two scale state equation that will be useful in the paper. In Section 4 we introduce parametrized ergodic BSDEs and study their regularity with respect to parameters. In Section 5 we state the form of the limit equations and prove a convergence result for BSDEs that represents the main technical issue of this paper. In Section 6, we finally link our results to the stochastic singular control problem. Finally, in section 7 we interpret the solution of the reduced BSDE in terms of a stochastic optimal control problem.

## 2 Notation

Given a Banach space $E$, the norm of its elements $x$ will be denoted by $|x|_E$, or even by $|x|$ when no confusion is possible. If $F$ is another Banach space, $L(E,F)$ denotes the space of bounded linear operators from $E$ to $F$, endowed with the usual operator norm. When $F = \mathbb{R}$ the dual space $L(E,\mathbb{R})$ will be denoted by $E^*$. The letters $\Xi$, $H$ and $K$ will always be used to denote Hilbert spaces. The scalar product is denoted $\langle\cdot,\cdot\rangle$, equipped with a subscript to specify the space, if necessary. All Hilbert spaces are assumed to be real and separable and the dual of a Hilbert space will never be identified with the space itself. By $L_2(\Xi, H)$ and $L_2(\Xi, K)$ we denote the spaces of Hilbert-Schmidt operators from $\Xi$ to $H$ and to $K$, respectively. Finally $\mathcal{G}(K, H)$ is the space of all Gateaux differentiable mappings $\phi$ from $K$ to $H$ such that the map $(k, v) \to \nabla\phi(k)v$ is continuous from $K \times K$ to $H$; see [16] for details.

Let $W^1 = (W_t^1)_{t\geq 0}$ and $W^2 = (W_t^2)_{t\geq 0}$ be two independent cylindrical Wiener processes with values in $\Xi$, defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. By $\{\mathcal{F}_t,\ t \in [0,T]\}$ we will denote the natural filtration of $(W^1, W^2)$, augmented with the family $\mathcal{N}$ of $\mathbb{P}$- null sets of $\mathcal{F}$. Obviously, the filtration $(\mathcal{F}_t)$ satisfies the usual conditions of right-continuity and completeness. All the concepts of measurability for stochastic processes will refer to this filtration. By $\mathcal{P}$ we denote $\sigma$-algebra of progressive measurable sets on $\Omega \times [0,T]$ and by $\mathcal{B}(\Lambda)$ the Borel $\sigma$-algebra of any topological space $\Lambda$.
Next we define the following two classes of stochastic processes with values in a Hilbert space $V$. Given an arbitrary time horizon $T$ and constant $p \geq 1$:

- $L^p_{\mathcal{P}}(\Omega \times [0,T]; V)$ denotes the space of equivalence classes of processes $Y \in L^p(\Omega \times [0,T]; V)$ admitting a predictable version. It is endowed with the norm
$$|Y| = \left(\mathbb{E}\int_0^T |Y_s|^p\, ds\right)^{1/p}.$$
- $L^{p,loc}_{\mathcal{P}}(\Omega \times [0,+\infty[; V)$ denotes the set of processes defined on $\mathbb{R}^+$ such that their restriction to an arbitrary $[0,T]$ belongs to $L^p_{\mathcal{P}}(\Omega \times [0,T]; V)$.
- $L^p_{\mathcal{P}}(\Omega; C([0,T]; V))$ denotes the space of predictable processes $Y$ with continuous paths in $V$, such that the norm
$$\|Y\|_p = (\mathbb{E} \sup_{s\in[0,T]} |Y_s|^p)^{1/p}$$
is finite. The elements of $L^p_{\mathcal{P}}(\Omega; C([0,T]; V))$ are identified up to indistinguishability.
- $L^{p,loc}_{\mathcal{P}}(\Omega; C[0,+\infty[; V))$ denotes the set of processes defined on $\mathbb{R}^+$ such that their restriction to an arbitrary $[0,T]$ belongs to $L^p_{\mathcal{P}}(\Omega; C([0,T]; V))$.

Given $\Phi$ in $L^2_{\mathcal{P}}(\Omega \times [0,T]; L_2(\Xi, V))$, the Itô stochastic integrals $\int_0^t \Phi_s\, dW_s^1$ and $\int_0^t \Phi_s\, dW_s^2$, $t \in [0,T]$, are $V$-valued martingales belonging to $L^2_{\mathcal{P}}(\Omega; C([0,T]; V))$.

## 3 The forward system

For arbitrarily fixed $x_0 \in H$ and $q_0 \in K$ we consider the following system of stochastic differential equations in $H \times K$:
$$\begin{cases} dX_t = AX_t dt + R\, dW_t^1, & X_0 = x_0,\ t \geq 0, \\ \\ \varepsilon dQ_t^\varepsilon = (BQ_t^\varepsilon + F(X_t, Q_t^\varepsilon))\, dt + \varepsilon^{1/2} G\, dW_t^2, & Q_0^\varepsilon = q_0,\ t \geq 0, \end{cases} \tag{3.1}$$
where the "slow" variable $X$ takes its values in $H$ and the "fast" variable $Q^\varepsilon$ takes its values in $K$, $\varepsilon \in ]0,1]$ is a small parameter.
Finally $A : D(A) \subset H \to H$ and $B : D(B) \subset K \to K$ are unbounded linear operators generating $C_0$-semigroups $\{e^{tA}\}_{t\geq 0}$ and $\{e^{tB}\}_{t\geq 0}$ over $H$ and $K$, respectively, while $R$ and $G$ are linear bounded operators from $\Xi$ to $H$ (respectively to $K$).
Moreover, we make the following, standard assumptions:

**Hypothesis 3.1** *$A : D(A) \subset H \to H$ is a linear, unbounded operator that generates a $C_0$- semigroup $\{e^{tA}\}_{t\geq 0}$, such that $|e^{tA}|_{L(H,H)} \leq M_A e^{\omega_A t}, t \geq 0$ for some positive constants $M_A$ and $\omega_A$. $B : D(B) \subset K \to K$ is a linear, unbounded operator that generates a $C_0$- semigroup $\{e^{tB}\}_{t\geq 0}$ such that $|e^{tB}|_{L(K,K)} \leq M_B e^{\omega_B t}, t \geq 0$ for some $M_B, \omega_B > 0$.*
*Moreover there exist constants $L > 0$ and $\gamma \in [0, \frac{1}{2}[$ s.t.:*

$$|e^{sA}|_{L_2(\Xi,H)} + |e^{sB}|_{L_2(\Xi,K)} \leq L s^{-\gamma}, \quad \forall s \in [0,1].$$

**Hypothesis 3.2** *$F : H \times K \to K$ is bounded and there exists a constant $L_F$ for which:*

$$|F(x,y) - F(u,v)|_K \leq L_F(|x-u|_H + |y-v|_K).$$

*for every $x, u \in H$, $y, v \in K$.*
*Moreover we assume that for every $x \in H$, $F(x,\cdot)$ is Gateaux differentiable, more precisely, $F(x,\cdot) \in \mathcal{G}^1(K,K)$.*

**Hypothesis 3.3** *$B + F$ is dissipative i.e. there exists some $\mu > 0$ such that:*

$$\langle Bq + F(x,q) - (Bq' + F(x,q')), q - q' \rangle \leq -\mu |q - q'|^2,$$

*for all $x \in H, q, q' \in D(B)$.*

**Remark 3.4** *The above dissipativity assumption is needed in Section 4 to apply the techniques in [15] when dealing with an ergodic control problem for the quick evolution (e.g. the second equation in 3.1). As far as the ergodic control problems is considered such hypothesis can be relaxed when the diffusion operator $G$ is assumed to be invertible, see [9]. We choose here to stick to the stronger form because, on the one side, we think it is important to cover the case of degenerate noise in the quick evolution (that can not, for instance, be covered, even in the finite dimensional case, in [2]) and, on the other side, strong results like the uniform exponential decay in Lemma 3.10, following from the strong formulation, seem to be technically very helpful here.*

**Hypothesis 3.5** *$R \in L(\Xi; H)$, $G \in L(\Xi; K)$.*

**Remark 3.6** *When we will apply our results to a two scale control problem, see Section 6, we will have to impose invertibility of $R$ (instead of a standard and weaker 'structure condition' on the image of the control operator, see, for instance [16]). This is indeed an unpleasant technical assumption needed to simplify the structure of the slow evolution after Girsanov transform.*

Given any cylindrical Wiener process $(\beta_t)_{t\geq 0}$ with values in $\Xi$ we denote by $(\beta_t^B)_{t\geq 0}$ the stochastic convolution

$$\beta_s^B = \int_0^s e^{(s-\ell)B} G d\beta_\ell.$$

In the following we shall assume, as in [15], that:

**Hypothesis 3.7** $\sup_{s>0} \mathbb{E}|\beta_s^B|^2 < \infty$.

**Remark 3.8** *Notice that since $(\beta_t)$ is a centered Gaussian process this implies that, $\forall p \geq 1$ it holds $\sup_{s>0} \mathbb{E}|\beta_s^B|^p < \infty$. Moreover hypothesis* 3.7 *is verified whenever $B$ is a strongly dissipative operator.*

We collect here two results we will use in the sequel. We do not provide the proof of the first, that can be found for instance in [16, Proposition 3.2]. Regarding the second result, for the reader's convenience, we briefly report the argument which is a slight modification of the one in [8, Section 6.3.2.].

**Lemma 3.9** *Under Hypothesis 3.1 and 3.5 the slow equation in system (3.1) admits a unique mild solution $X_t^{x_0}$ that has continuous trajectories and for all $p \geq 1$ satisfies:*

$$\mathbb{E}(\sup_{t\in[0,1]} |X_t^{x_0}|^p) \leq c_p(1 + |x_0|^p), \qquad x_0 \in H, \tag{3.2}$$

*for some positive constant $c_p$ depending only on $p$ and on the quantities introduced in the hypotheses.*

**Lemma 3.10** *Let $(\Gamma_s)_{s\geq 0}$ be a given, $H$-valued, predictable process with $\Gamma \in L^{p,loc}_{\mathcal{P}}(\Omega \times [0, \infty[; H)$ and let $(g_s)_{s\geq 0}$ be a given, $K$-valued, process with $g \in L^{p,loc}_{\mathcal{P}}(\Omega \times [0, +\infty[; K)$ for some $p \geq 1$.*
*Then the following equation:*

$$dQ_s = (BQ_s + F(\Gamma_s, Q_s))\, ds + g_s ds + G d\beta_s, \;\; s \geq 0, \qquad Q_0 = q_0, \tag{3.3}$$

*admits a unique mild solution $Q \in L^{p,loc}_{\mathcal{P}}(\Omega; C([0, +\infty[; K))$.*
*Under hypotheses (3.1)–(3.7), there exists a constant $k_p$ (independent of $T$) such that for all $T > 0$:*

$$\sup_{s\in[0,T]} \mathbb{E}|Q_s|^p \leq k_p(1 + |q_0|^p + \sup_{s\in[0,T]} \mathbb{E}|\Gamma_s|^p + \sup_{s\in[0,T]} \mathbb{E}|\beta^B_s|^p + \sup_{s\in[0,T]} \mathbb{E}|g_s|^p). \tag{3.4}$$

*Moreover if $(\Gamma'_s)_{s\geq 0}$ is another $H$-valued, predictable processes in $L^{p,loc}_{\mathcal{P}}(\Omega \times [0, \infty[; H)$ and $Q'$ is the mild solution of equation:*

$$dQ'_s = (BQ'_s + F(\Gamma'_s, Q'_s))\, ds + g_s ds + G d\beta_s, \;\; s \geq 0, \qquad Q_0 = q_0,$$

*then, for all $T > 0$,*

$$|Q_T - Q'_T| \leq K \int_0^T e^{-\mu(T-\ell)} |\Gamma_\ell - \Gamma'_\ell|\, d\ell, \quad \mathbb{P}\text{-a.s.},$$

*where again $K$ does not depend on $T$.*

**Proof.** Let $Z_s = e^{\mu s}(Q_s - \beta^B_s)$. By Itô rule (going through Yosida approximations) we deduce that $Z$ is the mild solution of the following equation

$$dZ_s = \mu Z_s\, ds + BZ_s\, ds + e^{\mu s} F(\Gamma_s, e^{-\mu s} Z_s + \beta^B_s))ds + e^{\mu s} g_s ds.$$

Differentiating $\sqrt{|Z_s|^2 + \varepsilon}$ (going, once more, through Yosida approximations), using dissipativity of $B+F$, see hypothesis 3.3, we obtain

$$|Z_s| \leq \sqrt{|Z_s|^2 + \varepsilon} \leq \sqrt{q_0^2 + \varepsilon} + \int_0^s e^{\mu\ell} \left|F(\Gamma_\ell, \beta^B_\ell) + g_\ell\right| d\ell + \mu \int_0^s \left[\sqrt{|Z_\ell|^2 + \varepsilon} - |Z_\ell|\right] d\ell.$$

Letting $\varepsilon \to 0$, by dominated convergence we obtain:

$$|Z_s| \leq |q_0| + \int_0^s e^{\mu\ell} \left|F(\Gamma_\ell, \beta^B_\ell) + g_\ell\right| d\ell.$$

Recalling the definition of $Z$ we conclude:

$$|Q_s| \leq |\beta^B_s| + e^{-\mu t}|q_0| + \int_0^s e^{-\mu(s-\ell)} \left|F(\Gamma_\ell, \beta^B_\ell) + g_\ell\right| d\ell,$$

and by Hölder inequality (for the last term):

$$|Q_s|^p \leq 3^p|\beta^B_s|^p + 3^p e^{-p\mu t}|q_0|^p + 3^p \left(\int_0^s e^{-p^* \frac{\mu}{2}(s-\ell)} d\ell\right)^{p/p^*} \int_0^s e^{-p\frac{\mu}{2}(s-\ell)} |F(\Gamma_\ell, \beta^B_\ell) + g_\ell|^p d\ell.$$

The claim then follows from Hypothesis 3.2.
The proof of the last statement is similar (and easier) noticing that:

$$d_s(Q_s - Q'_s) = B(Q_s - Q'_s)ds + [F(\Gamma_s, Q_s) - F(\Gamma'_s, Q'_s)]ds,$$

and then arguing as before. □

If we fix $x \in H$, $q_0 \in K$, choose $g \equiv 0$ and make a change of time $s \to \varepsilon s$, then the fast equation in system (3.1) becomes

$$d\hat{Q}_s = (B\hat{Q}_s + F(x, \hat{Q}_s))\, ds + Gd\hat{W}^2_s, \; s \geq 0, \qquad \hat{Q}_0 = q_0. \tag{3.5}$$

where $\hat{W}^2_s = \varepsilon^{-1/2} W^2_{\varepsilon s}$ is a cylindrical Wiener process. So (3.5) is a special case of (3.3), and Lemma 3.10 applies.
We will denote by $\hat{Q}^{x,q_0}_s$ the unique mild solution of equation (3.5).

# 4 The ergodic BDSE parametrized

We introduce a function $\psi : H \times K \times \Xi^* \times \Xi^* \to \mathbb{R}$. We will eventually (see Section 6) choose as $\psi$ the Hamiltonian of our control problem. Here we only assume that $\psi$ satisfies the following:

**Hypothesis 4.1** *Function $\psi$ is measurable and there exist $L_q, L_x, L_z, L_\xi > 0$ such that $\forall\, q, q' \in K$, $x, x' \in H$, $\xi, \xi', z, z' \in \Xi^*$:*

$$|\psi(x, q, z, \xi) - \psi(x', q', z', \xi')| \leq L_x(1 + |z|)|x - x'| + L_z|z - z'| + L_q(1 + |z|)|q - q'| + L_\xi|\xi - \xi'|.$$

*Moreover we assume that* $\sup_{x\in H, q\in K} |\psi(x, q, 0, 0)| < +\infty$

The next result states existence of a solution to the so called *ergodic backward stochastic differential equation* (EBSDE):

$$-\mathrm{d}\check{Y}_t = [\psi(x, \hat{Q}^{x,q_0}, z, \check{\Xi}_t) - \lambda(x, z)]\, \mathrm{d}t - \check{\Xi}_t \mathrm{d}\hat{W}^2_t, \quad \forall\, t \geq s. \tag{4.1}$$

**Theorem 4.2** *Under hypotheses 3.1, 3.2, 3.3, 3.5, 3.7 and 4.1 there exist measurable functions $\check{v} : H \times K \times \Xi^* \to \mathbb{R}$, $\check{\zeta} : H \times K \times \Xi^* \to \mathbb{R}$, $\lambda : H \times \Xi^* \to \mathbb{R}$ such that:*

1. *for all fixed $x$ and $z$, $\check{v}$ is Lipschitz with respect to $q$ and verifies:*

$$|\check{v}(x, q, z)| \leq c(1 + |z|)|q|, \tag{4.2}$$

   *($c > 0$ depends only on the constants introduced in the above mentioned hypotheses) .*

2. *if we set*

$$\check{Y}^{x,q_0,z}_t = \check{v}(x, \hat{Q}^{x,q_0}_t, z), \quad \check{\Xi}^{x,q_0,z}_t = \check{\zeta}(x, \hat{Q}^{x,q_0}_t, z), \tag{4.3}$$

   *then $\check{\Xi}^{x,q_0,z} \in L^{2,loc}_{\mathcal{P}}([0, +\infty[, \Xi^*)$ and the EBSDE (4.1) is satisfied by the triplet $(\check{Y}^{x,q,z}_t, \check{\Xi}^{x,q,z}, \lambda(x, z))$ that is:*

$$\check{Y}_t = \check{Y}_T + \int_t^T [\psi(x, \hat{Q}^{x,q_0}_s, z, \check{\Xi}_s) - \lambda(x, z)]\, \mathrm{d}s - \int_t^T \check{\Xi}_s \mathrm{d}\hat{W}^2_s, \; \mathbb{P} - a.s. \text{ for all } t \leq T.$$

3. *it holds:*

$$|\lambda(x, z) - \lambda(x', z')| \leq L^1_x(1 + |z|)|x - x'| + L^1_z|z - z'|, \tag{4.4}$$

   *for some posive constants $L^1_x$ and $L^1_z$.*

**Proof.** Fix $x \in H$ and $z \in \Xi^*$. In [15, Theorem 4.4 and Corollary 5.9] authors prove existence of functions $\check{v}(x, \cdot, z)$, $\check{\zeta}(x, \cdot, z)$ and $\lambda(x, z)$ such that (4.2) holds and if $(\check{Y}^{x,q_0,z}, \check{\Xi}^{x,q_0,z})$ are defined as in (4.3), then $\check{\Xi}^{x,q_0,z}$ is in $L^{2,loc}_{\mathcal{P}}([0, +\infty[, \Xi^*)$ and the triplet $(\check{Y}^{x,q_0,z}_t, \check{\Xi}^{x,q_0,z}, \lambda(x, z))$ is a solution to equation (4.1).
Measurably of $\check{v}$, $\check{\zeta}$ and $\lambda$ with respect to all parameters follows by their construction (see again [15] Theorem 4.4).

We only need to prove (4.4). Fixed $x, x' \in H$ and $z, z' \in \Xi^*$ we set $\widetilde{\lambda} = \lambda(x, z) - \lambda(x', z')$, $\widetilde{Y} = Y^{x,0,z} - Y^{x',0,z'}$, $\widetilde{\Xi} = \check{\Xi}^{x,0,z} - \check{\Xi}^{x',0,z'}$,

$$\theta_t = \begin{cases} \dfrac{\psi(x, \hat{Q}^{x,0}_r, z, \check{\Xi}^{x,0,z}_r) - \psi(x, \hat{Q}^{x,0}_r, z, \check{\Xi}^{x',0,z'}_r)}{|\check{\Xi}^{x,0,z}_r - \check{\Xi}^{x',0,z'}_r|^2_{\Xi^*}} (\check{\Xi}^{x,0,z}_r - \check{\Xi}^{x',0,z'}_r), & \text{if } \check{\Xi}^{x,0,z}_r \neq \check{\Xi}^{x',0,z'}_r \\ 0 & \text{otherwise} \end{cases}$$

and

$$f_t = \psi(x, \hat{Q}^{x,0}_r, z, \check{\Xi}^{x',0,z'}_r) - \psi(x', \hat{Q}^{x',0}_r, z', \check{\Xi}^{x',0,z'}_r).$$

Then we have

$$\widetilde{Y}_0 + \tilde{\lambda} T = \widetilde{Y}_T + \int_t^T f_r \,\mathrm{d}r - \int_t^T \widetilde{\Xi}_r (\theta_t dt + \mathrm{d}\hat{W}^2_r), \quad \forall T \geq t \geq 0.$$

So, by Girsanov theorem (notice that $(\theta_t)$ is uniformly bounded), there exists a probability $\widetilde{\mathbb{P}}$ (mean value denoted by $\widetilde{\mathbb{E}}$) such that $\widetilde{W}_t = \int_0^t \theta_\ell d\ell + \hat{W}^2_t$, $t \geq 0$, is a cylindrical Wiener process. Consequently:

$$\tilde{\lambda} T = \widetilde{Y}_T - \widetilde{Y}_0 + \int_0^T f_r \,\mathrm{d}r - \int_0^T \widetilde{\Xi}_r \mathrm{d}\widetilde{W}_r, \quad \forall T \geq t \geq 0$$

and consequently:

$$|\tilde{\lambda}| \leq T^{-1} |\widetilde{Y}_0| + T^{-1} \widetilde{\mathbb{E}} |\widetilde{Y}_T| + T^{-1} \int_0^T \widetilde{\mathbb{E}} |f_s| \,\mathrm{d}s. \tag{4.5}$$

Thanks to hypothesis 4.1 we get that for all $t \geq 0$:

$$|f_t| \leq L_x(1 + |z|)|x - x'| + L_z|z - z'| + L_q(1 + |z|)|\hat{Q}^{x,0}_t - \hat{Q}^{x',0}_t|, \qquad \mathbb{P} - a.s.$$

We notice that with respect to $(\widetilde{W}_t)$ processes $\hat{Q}^{x,0}$ and $\hat{Q}^{x',0}$ satisfy respectively

$$d\hat{Q}^{x,0}_s = (B\hat{Q}^{x,0}_s + F(x, \hat{Q}^{x,0}_s))\, ds + \theta_s ds + G d\widetilde{W}_s, \ s \geq 0,$$
$$d\hat{Q}^{x',0}_s = (B\hat{Q}^{x',0}_s + F(x', \hat{Q}^{x',0}_s))\, ds + \theta_s ds + G d\widetilde{W}_s, \ s \geq 0,$$

and Lemma 3.10 yields $|\hat{Q}^{x,0}_s - \hat{Q}^{x',0}_s| \leq (K/\mu)|x - x'|$ thus:

$$|f_t| \leq (L_x + L_q K/\mu)(1 + |z|)|x - x'| + L_z|z - z'|, \qquad \mathbb{P} - a.s. \text{ for all } t \geq 0. \tag{4.6}$$

From Lemma 3.10 we also have that for every $T \geq 0$ and every $p \geq 1$,

$$\sup_{s \in [0,T]} \widetilde{\mathbb{E}} |\hat{Q}^{x,0}_s|^p \leq k_p(1 + |x|^p + \sup_{s \in [0,T]} \widetilde{\mathbb{E}} |\theta_s|^p), \tag{4.7}$$

and

$$\sup_{s \in [0,T]} \widetilde{\mathbb{E}} |\hat{Q}^{x',0}_s|^p \leq k_p(1 + |x|^p + \sup_{s \in [0,T]} \widetilde{\mathbb{E}} |\theta_s|^p). \tag{4.8}$$

Since $\theta$ is uniformly bounded it holds:

$$\sup_{t\in[0,\infty[} \widetilde{\mathbb{E}}(|\hat{Q}_t^{x,0}|^p + |\hat{Q}_t^{x,0}|^p) < \infty,$$

thus, by (4.2), we get that:

$$\sup_{t\in[0,\infty[} \widetilde{\mathbb{E}}(|\widetilde{Y}_t|) < \infty.$$

Consequently $T^{-1}\widetilde{\mathbb{E}}(|\widetilde{Y}_T|) \to 0$ as $T \to \infty$ and the claim follows by (4.5) and (4.6) letting $T \to \infty$.

□

**Remark 4.3** *If, fixed $x$ and $z$, one restricts the class of triples $(Y, \Xi, \lambda)$ where to find a solution to equation (4.1), asking that there must be a constant $c > 0$ (that may depend on $q_0$, $x$ and $z$) such that $|Y_t| \leq c(1+|Q_t|)$ $\mathbb{P}$-a.s. for every $t \geq 0$ then, see [15, Theorem 4.6], the third component $\lambda$ of the solution is uniquely determined.*

# 5 Limit equation and convergence of singular BSDEs

We've eventually got to the *forward-backward system* for $t \in [0, 1]$

$$\begin{cases} dX_t = AX_t\,dt + R\,dW_t^1, \\ \varepsilon dQ_t^\varepsilon = (BQ_t^\varepsilon + F(X_t, Q_t^\varepsilon))\,dt + \varepsilon^{1/2}\,G\,dW_t^2, \\ -dY_t^\varepsilon = \psi(X_t, Q_t^\varepsilon, Z_t^\varepsilon, \Xi_t^\varepsilon/\sqrt{\varepsilon})\,dt - Z_t^\varepsilon dW_t^1 - \Xi_t^\varepsilon dW_t^2, \\ X_0 = x_0 \quad Q_0^\varepsilon = q_0, \quad Y_1^\varepsilon = h(X_1), \end{cases} \tag{5.1}$$

that, as we will see in the sequel, is also associated to a controlled multiscale dynamics. Function $h : H \to \mathbb{R}$ satisfies:

**Hypothesis 5.1** *$h$ is Lipschitz continuous with constant $L > 0$.*

We have that:

**Theorem 5.2** *Assume 3.1–3.7, 4.1 and 5.1.*
*For every $\varepsilon > 0$ there exists a unique 5-tuple of processes $(X, Q^\varepsilon, Y^\varepsilon, Z^\varepsilon, \Xi^\varepsilon)$, with $X \in L^2_{\mathcal{P}}(\Omega; C([0, 1]; H))$, $Q^\varepsilon \in L^2_{\mathcal{P}}(\Omega; C([0, 1]; K))$, $Y^\varepsilon \in L^2_{\mathcal{P}}(\Omega; C([0, 1]; \mathbb{R}))$, $Z^\varepsilon \in L^2_{\mathcal{P}}(\Omega \times [0, 1]; \Xi^*)$ and $\Xi^\varepsilon \in L^2_{\mathcal{P}}(\Omega \times [0, 1]; \Xi^*)$ such that $\mathbb{P} - a.s.$ the system (5.1) is satisfied for all $t \in [0, 1]$.*

**Proof.** The proof is contained in [16, Propositions 3.2 and 5.2], we just notice that the system is decoupled, so once the forward equation is solved then it becomes a known process in the backward equation. □

The purpose of our work is to study the limit behaviour of $Y^\varepsilon$ as $\varepsilon$ tends to 0.
We introduce the candidate limit equation, that turns out to be a forward-backward system on the *finite horizon* $[0, 1]$ and on the reduced state space $H$.

$$\begin{cases} dX_t = AX_t\,dt + R\,dW_t^1, \qquad t \in [0, 1], \\ -d\bar{Y}_t = \lambda(X_t, \bar{Z}_t)\,dt - \bar{Z}_t\,dW_t^1, \\ X_0 = x_0, \quad \bar{Y}_1 = h(X_1). \end{cases} \tag{5.2}$$

where $\lambda$ is defined in Theorem 4.2.
One has that

**Theorem 5.3** *Under Hypothesis 3.1—3.7, 4.1 and 5.1, there exists a unique triplet of processes* $(X, \bar{Y}, \bar{Z})$ *with* $X \in L^p_{\mathcal{P}}(\Omega; C([0,1]; H))$, $\bar{Y} \in L^p_{\mathcal{P}}(\Omega; C([0,1]; \mathbb{R}))$, $\bar{Z} \in L^p_{\mathcal{P}}(\Omega \times [0,1]; \Xi^*)$ *that fullfils system* (5.2), $\mathbb{P}-a.s.$ *for every* $t \in [0,1]$.

**Proof.** Thank to the regularity of $\lambda$, see (4.4), the proof of existence and uniqueness of the solution to equation (5.2) is standard (see, for instance [16, Proposition 4.3]). □

We can now state our main result:

**Theorem 5.4** *Under Hypothesis 3.1—3.7, 4.1 and 5.1, the following holds for* $\bar{Y}$ *and* $Y^\varepsilon$ *found in Theorem 5.2 and Theorem 5.3 respectively:*

$$\lim_{\varepsilon \to 0} Y^\varepsilon_0 = \bar{Y}_0. \tag{5.3}$$

**Proof.** We start by noticing that if we slow down time, that is, for $s \in [0, 1/\varepsilon[$ we set: $\hat{Q}^\varepsilon_s = Q^\varepsilon_{\varepsilon s}$, $\hat{Y}^\varepsilon_s = Y^\varepsilon_{\varepsilon s}$, $\hat{\Xi}^\varepsilon_s = \varepsilon^{-1/2}\Xi^\varepsilon_{\varepsilon s}$ then the last two equations in (5.1) becomes:

$$\begin{cases} d\hat{Q}^\varepsilon_s = (B\hat{Q}^\varepsilon_s + F(X_{\varepsilon s}, \hat{Q}^\varepsilon_s)\, dt + G\, d\hat{W}^2_s, \\ -d\hat{Y}^\varepsilon_s = \psi(X_{\varepsilon s}, \hat{Q}^\varepsilon_s, Z^\varepsilon_{\varepsilon s}, \hat{\Xi}^\varepsilon_s)\, dt - \sqrt{\varepsilon} Z^\varepsilon_{\varepsilon s} d\, \hat{W}^1_s - \hat{\Xi}^\varepsilon_s d\hat{W}^2_s, \\ X_0 = x_0 \quad \hat{Q}^\varepsilon_0 = q_0, \quad \hat{Y}^\varepsilon_{1/\varepsilon} = h(X_1). \end{cases} \tag{5.4}$$

where $\hat{W}^\ell_s = \varepsilon^{-1/2} W^\ell_{\varepsilon s}$, $\ell = 1, 2$. We will often make use of this change of time in the proof.
We must compare:

$$Y^\varepsilon_0 - \bar{Y}_0 = \int_0^1 (\psi(X_t, Q^\varepsilon_t, Z^\varepsilon_t, \Xi^\varepsilon_t/\sqrt{\varepsilon}) - \lambda(X_t, \bar{Z}_t))\, dt - \int_0^1 (Z^\varepsilon_t - \bar{Z}_t)\, dW^1_t - \int_0^1 \Xi^\varepsilon_t\, dW^2_t.$$

By adding and subtracting we split the first integral on the right hand side as:

$$\int_0^1 (\psi(X_t, Q^\varepsilon_t, Z^\varepsilon_t, \Xi^\varepsilon_t/\sqrt{\varepsilon}) - \lambda(X_t, \bar{Z}_t))\, dt = \int_0^1 (\psi(X_t, Q^\varepsilon_t, Z^\varepsilon_t, \Xi^\varepsilon_t/\sqrt{\varepsilon}) - \psi(X_t, Q^\varepsilon_t, \bar{Z}_t, \Xi^\varepsilon_t/\sqrt{\varepsilon}))\, dt$$
$$+ \int_0^1 (\psi(X_t, Q^\varepsilon_t, \bar{Z}_t, \Xi^\varepsilon_t/\sqrt{\varepsilon}) - \lambda(X_t, \bar{Z}_t))\, dt. \tag{5.5}$$

We have to use a discretization argument to cope with the second member of the sum.
Let us now introduce for every $N$ positive integer, a partition of the interval $[0,1]$ of the form $t_k = k2^{-N}$, $k = 0, 1, \dots, 2^N$ and define a couple of step processes $X^N$ and $\widetilde{Z}^N$ defined as follows:

$$X^N_t = X_{t_k}, \quad t \in [t_k, t_{k+1}[, \quad k = 0, \dots 2^N - 1, \tag{5.6}$$

$$\widetilde{Z}^N_t = 2^N \int_{t_{k-1}}^{t_k} \bar{Z}_\ell\, d\ell, \text{ for } t \in [t_k, t_{k+1}[, \ k = 1, \dots 2^N - 1, \ \widetilde{Z}_t = 0 \text{ for } t \in [0, t_1[, \tag{5.7}$$

where $X, \bar{Z}$ are part of the solution of (5.2). By construction one has that:

$$\lim_{N \to \infty} \mathbb{E} \int_0^1 |\widetilde{Z}^N_t - \bar{Z}_t|^2\, dt = 0. \tag{5.8}$$

We fix $N$, then for $k = 0, 1, \dots, 2^N - 1$ we consider the following, iteratively defined, class of forward SDE:

$$d\hat{\mathcal{Q}}^{N,k}_s = (B\hat{\mathcal{Q}}^{N,k}_s + F(X_{t_k}, \hat{\mathcal{Q}}^{N,k}_s))\, ds + G d\hat{W}^2_s, \quad s \geq t_k/\varepsilon, \qquad \hat{\mathcal{Q}}^{N,k}_{t_k/\varepsilon} = \hat{\mathcal{Q}}^{N,k-1}_{t_k/\varepsilon}, \tag{5.9}$$

Moreover we define (see Theorem 4.2):

$$\check{Y}^{N,k}_s = \check{v}(X_{t_k}, \hat{\mathcal{Q}}^{N,k}_s, \widetilde{Z}^N_{t_k}), \quad \check{\Xi}^{N,k}_s = \check{\zeta}(X_{t_k}, \hat{\mathcal{Q}}^{N,k}_s, \widetilde{Z}^N_{t_k}), \quad \text{for } s \geq t_k/\varepsilon,$$

so that the triplet $((\check{Y}_s^{N,k})_{s\geq t_{k/\varepsilon}}, \lambda(X_{t_k}, \widetilde{Z}_{t_k}^N), (\check{\Xi}_s^{N,k})_{s\geq t_{k/\varepsilon}})$ verifies:

$$-d\check{Y}_s^{N,k} = [\psi(X_{t_k}, \hat{\mathcal{Q}}_s^{N,k}, \widetilde{Z}_{t_k}^N, \check{\Xi}_s^{N,k}) - \lambda(X_{t_k}, \widetilde{Z}_{t_k}^N)]\,\mathrm{d}s - \check{\Xi}_s^{N,k}\mathrm{d}\hat{W}_s^2, \quad \text{for all } s \geq t_k/\varepsilon, \tag{5.10}$$

and

$$|\check{Y}_s^{N,k}| \leq c(1+|\widetilde{Z}_{t_k}^N|)|\hat{\mathcal{Q}}_s^{N,k}|, \qquad \text{for all } s \geq t_k/\varepsilon, \tag{5.11}$$

for some positive constant $c>0$ independent of $k$ and $N$.
We also set for $s \in [0, 1/\varepsilon[$:

$$\hat{\mathcal{Q}}_s^N = \sum_{k=0}^{2^N-1} \hat{\mathcal{Q}}_s^{N,k} I_{[t_k/\varepsilon, t_{k+1}/\varepsilon[}(s), \qquad \check{\Xi}_s^N = \sum_{k=0}^{2^N-1} \check{\Xi}_s^{N,k} I_{[t_k/\varepsilon, t_{k+1}/\varepsilon[}(s), \tag{5.12}$$

so that, for all $N \in \mathbb{N}$ and $k = 0, ..., 2^N-1$ have:

$$\check{Y}_{t_k/\varepsilon}^{N,k} - \check{Y}_{t_{k+1}/\varepsilon}^{N,k} - \int_{t_k/\varepsilon}^{t_{k+1}/\varepsilon} [\psi(X_{\varepsilon s}^N, \hat{\mathcal{Q}}_s^N, \widetilde{Z}_{\varepsilon s}^N, \check{\Xi}_s^N) - \lambda(X_{\varepsilon s}^N, \widetilde{Z}_{\varepsilon s}^N)]\,ds + \int_{t_k/\varepsilon}^{t_{k+1}/\varepsilon} \check{\Xi}_s^N\, d\hat{W}_s^2 = 0. \tag{5.13}$$

The second integral in the right hand side of (5.5) can be written as:

$$\int_0^1 (\psi(X_t, Q_t^\varepsilon, \bar{Z}_t, \Xi_t^\varepsilon/\sqrt{\varepsilon}) - \lambda(X_t, \bar{Z}_t)\,dt = \varepsilon \sum_{k=0}^{2^N-1} \int_{t_k/\varepsilon}^{t_{k+1}/\varepsilon} [\psi(X_{\varepsilon s}, \hat{Q}_s^\varepsilon, \bar{Z}_{\varepsilon s}, \hat{\Xi}_s^\varepsilon) - \lambda(X_{\varepsilon s}, \bar{Z}_{\varepsilon s})]\,ds,$$

and, adding the null terms in (5.13) for $k = 1, ..., 2^N$, as:

$$\begin{aligned}
\int_0^1 (\psi(X_t, Q_t^\varepsilon, \bar{Z}_t, \Xi_t^\varepsilon/\sqrt{\varepsilon}) - \lambda(X_t, \bar{Z}_t))\,dt &= \varepsilon \sum_{k=1}^{2^N-1} (\check{Y}_{t_k/\varepsilon}^{N,k} - \check{Y}_{t_{k+1}/\varepsilon}^{N,k}) \\
+\,\varepsilon \sum_{k=0}^{2^N-1} \int_{t_k/\varepsilon}^{t_{k+1}/\varepsilon} [\psi(X_{\varepsilon s}, \hat{Q}_s^\varepsilon, \bar{Z}_{\varepsilon s}, \hat{\Xi}_s^\varepsilon) &- \psi(X_{\varepsilon s}^N, \hat{\mathcal{Q}}_s^N, \widetilde{Z}_{\varepsilon s}^N, \check{\Xi}_s^N)]\,ds \\
+\,\varepsilon \sum_{k=0}^{2^N-1} \int_{t_k/\varepsilon}^{t_{k+1}/\varepsilon} \check{\Xi}_s^N\, d\hat{W}_s^2 - \varepsilon \sum_{k=0}^{2^N-1} \int_{t_k/\varepsilon}^{t_{k+1}/\varepsilon} &[\lambda(X_{\varepsilon s}, \bar{Z}_{\varepsilon s}) - \lambda(X_{\varepsilon s}^N, \widetilde{Z}_{\varepsilon s}^N)]\,ds.
\end{aligned} \tag{5.14}$$

Therefore coming back to our original term $Y_0^\varepsilon - \bar{Y}_0$ we have, taking into account (5.5):

$$\begin{aligned}
Y_0^\varepsilon - \bar{Y}_0 = {} & \varepsilon \sum_{k=1}^{2^N-1} (\check{Y}_{t_k/\varepsilon}^{N,k} - \check{Y}_{t_{k+1}/\varepsilon}^{N,k}) + \varepsilon \int_0^{1/\varepsilon} \Big[\psi(X_{\varepsilon s}, \hat{Q}_s^\varepsilon, Z_{\varepsilon s}^\varepsilon, \hat{\Xi}_s^\varepsilon) - \psi(X_{\varepsilon s}, \hat{Q}_s^\varepsilon, \bar{Z}_{\varepsilon s}, \hat{\Xi}_s^\varepsilon)\Big]\,ds \\
& + \varepsilon \int_0^{1/\varepsilon} [\psi(X_{\varepsilon s}, \hat{Q}_s^\varepsilon, \bar{Z}_{\varepsilon s}, \hat{\Xi}_s^\varepsilon) - \psi(X_{\varepsilon s}^N, \hat{\mathcal{Q}}_s^N, \widetilde{Z}_{\varepsilon s}^N, \check{\Xi}_s^N)]\,ds - \varepsilon \int_0^{1/\varepsilon} [\lambda(X_{\varepsilon s}, \bar{Z}_{\varepsilon s}) - \lambda(X_{\varepsilon s}^N, \widetilde{Z}_{\varepsilon s}^N)]\,ds \\
& - \sqrt{\varepsilon} \int_0^{1/\varepsilon} (Z_{\varepsilon s}^\varepsilon - \bar{Z}_{\varepsilon s}) d\hat{W}_s^1 - \varepsilon \int_0^{1/\varepsilon} (\hat{\Xi}_s^\varepsilon - \check{\Xi}_s^N) d\hat{W}_s^2.
\end{aligned}$$

Notice that we can rewrite this difference as follows:

$$\begin{aligned}
Y_0^\varepsilon - \bar{Y}_0 = {} & \varepsilon \int_0^{1/\varepsilon} \mathcal{R}_s^{\varepsilon,N}\,ds + \varepsilon \sum_{k=1}^{2^N-1} (\check{Y}_{t_k/\varepsilon}^{N,k} - \check{Y}_{t_{k+1}/\varepsilon}^{N,k}) - \varepsilon \int_0^{1/\varepsilon} [\lambda(X_{\varepsilon s}, \bar{Z}_{\varepsilon s}) - \lambda(X_{\varepsilon s}^N, \widetilde{Z}_{\varepsilon s}^N)]\,ds \\
& + \varepsilon \int_0^{1/\varepsilon} [\psi(X_{\varepsilon s}, \hat{Q}_s^\varepsilon, Z_{\varepsilon s}^\varepsilon, \hat{\Xi}_s^\varepsilon) - \psi(X_{\varepsilon s}, \hat{Q}_s^\varepsilon, \bar{Z}_{\varepsilon s}, \hat{\Xi}_s^\varepsilon)]\,ds \\
& + \varepsilon \int_0^{1/\varepsilon} [\psi(X_{\varepsilon s}^N, \hat{\mathcal{Q}}_s^N, \widetilde{Z}_{\varepsilon s}^N, \hat{\Xi}_s^\varepsilon) - \psi(X_{\varepsilon s}^N, \hat{\mathcal{Q}}_s^N, \widetilde{Z}_{\varepsilon s}^N, \check{\Xi}_s^N)]\,ds \\
& - \varepsilon \int_0^{1/\varepsilon} (\check{\Xi}_s^N - \hat{\Xi}_s^\varepsilon)\,d\hat{W}_s^2 - \sqrt{\varepsilon} \int_0^{1/\varepsilon} (Z_{\varepsilon s}^\varepsilon - \bar{Z}_{\varepsilon s})\,d\hat{W}_s^1,
\end{aligned} \tag{5.15}$$

where $\mathcal{R}_s^{\varepsilon,N} := \psi(X_{\varepsilon s}, \hat{Q}_s^\varepsilon, \bar{Z}_{\varepsilon s}, \hat{\Xi}_s^\varepsilon) - \psi(X_{\varepsilon s}^N, \hat{\mathcal{Q}}_s^N, \widetilde{Z}_{\varepsilon s}^N, \hat{\Xi}_s^\varepsilon)$. Then by Hypothesis 4.1 we deduce that for a suitable constant $c$, independent from $\varepsilon$ and $N$, the following holds:

$$|\mathcal{R}_s^{\varepsilon,N}| \leq c(1+|\bar{Z}_{\varepsilon s}|)|X_{\varepsilon s} - X_{\varepsilon s}^N| + c(1+|\bar{Z}_{\varepsilon s}|)|\hat{Q}_s^\varepsilon - \hat{\mathcal{Q}}_s^N| + c|\bar{Z}_{\varepsilon s} - \widetilde{Z}_{\varepsilon s}^N|. \tag{5.16}$$

The presence of the two stochastic terms in (5.15) allows us to get rid of the third and fourth term on the right hand side by a Girsanov argument, namely we introduce:

$$\delta^{1,\varepsilon}(s) = \begin{cases} \dfrac{[\psi(X_{\varepsilon t}, \hat{Q}_s^\varepsilon, Z_{\varepsilon s}^\varepsilon, \hat{\Xi}_s^\varepsilon) - \psi(X_{\varepsilon s}, \hat{Q}_s^\varepsilon, \bar{Z}_{\varepsilon s}, \hat{\Xi}_s^\varepsilon)]}{|Z_{\varepsilon s}^\varepsilon - \bar{Z}_{\varepsilon s}|^2}(Z_{\varepsilon s}^\varepsilon - \bar{Z}_{\varepsilon s})^* & \text{if } |Z_{\varepsilon s}^\varepsilon - \bar{Z}_{\varepsilon s}| \neq 0, \\ 0 & \text{if } |Z_{\varepsilon s}^\varepsilon - \bar{Z}_{\varepsilon s}| = 0, \end{cases} \tag{5.17}$$

and

$$\delta^{2,\varepsilon,N}(s) = \begin{cases} \dfrac{\psi(X_{\varepsilon s}^N, \hat{\mathcal{Q}}_s^N, \widetilde{Z}_{\varepsilon s}^N, \hat{\Xi}_s^\varepsilon) - \psi(X_{\varepsilon s}^N, \hat{\mathcal{Q}}_s^N, \widetilde{Z}_{\varepsilon s}^N, \check{\Xi}_s^N)}{|\hat{\Xi}_s^\varepsilon - \check{\Xi}_s^N|^2}(\hat{\Xi}_s^\varepsilon - \check{\Xi}_s^N)^* & \text{if } |\hat{\Xi}_s^\varepsilon - \check{\Xi}_s^N| \neq 0, \\ 0 & \text{if } |\hat{\Xi}_s^\varepsilon - \check{\Xi}_s^N| = 0. \end{cases} \tag{5.18}$$

We notice that processes $(\delta^{1,\varepsilon}(s))_{s\in[0,1/\varepsilon]}$ and $(\delta^{2,\varepsilon,N}(s))_{s\in[0,1/\varepsilon]}$ are bounded uniformly by $L_z$ and $L_\xi$ respectively, see Hypothesis 4.1. We have:

$$\begin{aligned} Y_0^\varepsilon - \bar{Y}_0 =& \varepsilon \int_0^{1/\varepsilon} \delta^{1,\varepsilon}(s)[Z_{\varepsilon s}^\varepsilon - \bar{Z}_{\varepsilon s}]\, ds - \varepsilon \int_0^{1/\varepsilon} \delta^{2,\varepsilon,N}(s)[\check{\Xi}_s^N - \hat{\Xi}_s^\varepsilon]\, ds \\ &+ \varepsilon \int_0^{1/\varepsilon} (\check{\Xi}_s^N - \hat{\Xi}_s^\varepsilon)\, d\hat{W}_s^2 + \sqrt{\varepsilon}\int_0^{1/\varepsilon} (Z_{\varepsilon s}^\varepsilon - \bar{Z}_{\varepsilon s})\, d\hat{W}_s^1 \\ &+ \varepsilon \int_0^{1/\varepsilon} \mathcal{R}_s^{\varepsilon,N}\, ds + \varepsilon \sum_{k=1}^{2^N-1} (\check{Y}_{t_k/\varepsilon}^{N,k} - \check{Y}_{t_{k+1}/\varepsilon}^{N,k}). \end{aligned}$$

and rescaling time (speeding it up this time)

$$\begin{aligned} Y_0^\varepsilon - \bar{Y}_0 =& \int_0^1 \delta^{1,\varepsilon}(t/\varepsilon)[Z_t^\varepsilon - \bar{Z}_t]\, dt + \int_0^1 \delta^{2,\varepsilon,N}(t/\varepsilon)[\check{\Xi}_{\varepsilon^{-1}t}^N - \check{\Xi}_{\varepsilon^{-1}t}^\varepsilon]\, dt \\ &+ \sqrt{\varepsilon}\int_0^1 (\check{\Xi}_{\varepsilon^{-1}t}^N - \check{\Xi}_{\varepsilon^{-1}t}^\varepsilon)\, dW_t^2 + \int_0^1 (Z_t^\varepsilon - \bar{Z}_t)\, dW_t^1 \\ &+ \int_0^1 \mathcal{R}_{\varepsilon^{-1}t}^{\varepsilon,N}\, dt + \varepsilon \sum_{k=1}^{2^N-1} (\check{Y}_{t_k/\varepsilon}^{N,k} - \check{Y}_{t_{k+1}/\varepsilon}^{N,k}). \end{aligned}$$

We set, for $t \in [0,1]$:

$$\widetilde{W}_t^1 =: \int_0^t \delta^{1,\varepsilon}(r/\varepsilon)\, dr + W_t^1, \tag{5.19}$$

$$\widetilde{W}_t^2 =: \varepsilon^{-1/2} \int_0^t \delta^{2,\varepsilon,N}(r/\varepsilon)\, dr + W_t^2. \tag{5.20}$$

We denote by $\widetilde{\mathbb{E}}^\varepsilon$ the expectation under the new probability $\tilde{\mathbb{P}}^\varepsilon$ with respect to which $(\widetilde{W}_t^1, \widetilde{W}_t^2)_{t\in[0,1]}$ is a $H \times K$ valued cylindrical Wiener process (recall that $(W_t^1, W_t^2)_{t\in[0,1]}$ is a $H \times K$ valued cylindrical Wiener process). Since the left hand side is deterministic, we have:

$$Y_0^\varepsilon - \bar{Y}_0 = \widetilde{\mathbb{E}}^\varepsilon \int_0^1 \mathcal{R}_{t/\varepsilon}^{\varepsilon,N}\, dt + \varepsilon \widetilde{\mathbb{E}}^\varepsilon \sum_{k=1}^{2^N-1} [\check{Y}_{t_k/\varepsilon}^{N,k} - \check{Y}_{t_{k+1}/\varepsilon}^{N,k}]. \tag{5.21}$$

Moreover, taking into account (5.16), it holds:

$$\widetilde{\mathbb{E}}^{\varepsilon}\int_0^1 |\mathcal{R}^{\varepsilon,N}_{t/\varepsilon}|\,dt \le c\widetilde{\mathbb{E}}^{\varepsilon}\int_0^1 \Big((1+|\bar{Z}_t|)|X_t - X^N_t| + (1+|\bar{Z}_t|)|\hat{Q}^{\varepsilon}_{t/\varepsilon} - \hat{\mathcal{Q}}^N_{t/\varepsilon}| + |\bar{Z}_t - \widetilde{Z}^N_t|\Big)\,dt.$$

Let us start from

$$\widetilde{\mathbb{E}}^{\varepsilon}\int_0^1 (1+|\bar{Z}_t|)|X_t - X^N_t|dt.$$

We notice that, with respect to $\widetilde{W}^1$ we have:

$$\begin{cases} dX_t = AX_t\,dt - R\delta^{1,\varepsilon}(t/\varepsilon)dt + R\,d\widetilde{W}^1_t, \\ -d\bar{Y}_t = \lambda(X_t,\bar{Z}_t)\,dt - \bar{Z}_t[-\delta^{1,\varepsilon}(t/\varepsilon)dt + d\widetilde{W}^1_t], \\ \bar{Y}_1 = h(X_1), \quad X_0 = x_0. \end{cases}$$

Define:

$$\rho := \exp\left(\int_0^1 \delta^{1,\varepsilon}(s/\varepsilon)\,d\widetilde{W}^1_s - \frac{1}{2}\int_0^1 |\delta^{1,\varepsilon}(s/\varepsilon)|^2\,ds\right),$$

then, byHölder inequality, setting $\Delta_{X,N} := \sup_{t\in[0,1]}|X_t - X^N_t|$ it holds:

$$\begin{aligned}\widetilde{\mathbb{E}}^{\varepsilon}\int_0^1 (1+|\bar{Z}_t|)|X_t - X^N_t|dt &\le \widetilde{\mathbb{E}}^{\varepsilon}\left[\Delta_{X,N}\int_0^1(1+|\bar{Z}_t|)dt\right] \\ &\le \widetilde{\mathbb{E}}^{\varepsilon}\left[\rho^{-3/4}(\rho^{1/4}\Delta_{X,N})\rho^{1/2}\int_0^1(1+|\bar{Z}_t|)dt\right] \\ &\le \left[\widetilde{\mathbb{E}}^{\varepsilon}\rho^{-3}\right]^{1/4}\left[\widetilde{\mathbb{E}}^{\varepsilon}(\rho\Delta^4_{X,N})\right]^{1/4}\left[\widetilde{\mathbb{E}}^{\varepsilon}\left(\rho\int_0^1(1+|\bar{Z}_t|^2)dt\right)\right]^{1/2}.\end{aligned}$$

Again by Girsanov the process $\left(-\int_0^t \delta^1(t/\varepsilon)dt + \widetilde{W}^1_t\right)_{t\in[0,1]}$ is a cylindrical Wiener process with respect to $\rho\,d\mathbb{P}^{\varepsilon}$. By uniqueness of the solution of the forward backward system (5.2) the law of the process $(X_t)_{t\ge0}$ under $\rho d\mathbb{P}^{\varepsilon}$ coincides with its law with respect to $\mathbb{P}$. Moreover we notice that being $\bar{Z}_t = \zeta(X_t)$ where $\zeta$ is a deterministic Borel function $H \to \Xi^*$ then the law of $\bar{Z}$ and $\tilde{Z}^N$ depend only on the law of $(X)$ in a non anticipating way. So even the law of $(\bar{Z}_t)_{t\ge0}$ and $(\bar{Z}^N_t)_{t\ge0}$ under $\rho d\mathbb{P}^{\varepsilon}$ coincides with its law with respect to $\mathbb{P}$.

Recalling that $\delta^{1,\varepsilon}$ is uniformly bounded and consequently (with respect to $\varepsilon$ as well) we have $\widetilde{\mathbb{E}}^{\varepsilon}\rho^{-3} \le c$ (where $c$ does not depend on $\varepsilon$), moreover

$$\widetilde{\mathbb{E}}^{\varepsilon}\left(\rho\int_0^1|\bar{Z}_t|^2dt\right) = \mathbb{E}\left(\int_0^1|\bar{Z}_t|^2dt\right) < +\infty.$$

Thus we can conclude

$$\widetilde{\mathbb{E}}^{\varepsilon}\int_0^1(1+|\bar{Z}_t|)|X_t - X^N_t|\,dt \le C[\mathbb{E}\Delta^4_{X,N}]^{1/4}, \tag{5.22}$$

where $C$ is independent of $N$ and $\varepsilon$.

By the continuity of trajectories of $(X_t)_{t\ge0}$, having also $\mathbb{E}\sup_{t\in[0,1]}|X_t|^4 < \infty$, we get:

$$\mathbb{E}\Delta^4_{X,N} \to 0, \qquad \text{as} \qquad N \to \infty. \tag{5.23}$$

We also have that:

$$\widetilde{\mathbb{E}}^{\varepsilon}\int_0^1|\bar{Z}_t - \widetilde{Z}^N_t|dt \le C\left[\mathbb{E}\int_0^1|\bar{Z}_t - \widetilde{Z}^N_t|^2dt\right]^{1/2} = C(\mathbb{E}\Delta_{Z,N})^{1/2}, \tag{5.24}$$

where $\Delta_{Z,N} = \displaystyle\int_0^1 |\bar{Z}_t - \widetilde{Z}^N_t|^2 dt$ and by (5.8):

$$\mathbb{E}\Delta_{Z,N} \to 0, \qquad \text{as} \qquad N \to \infty. \tag{5.25}$$

Now we deal with the term:

$$\widetilde{\mathbb{E}}^\varepsilon \int_0^1 (1 + |\bar{Z}_t|)|\hat{Q}^\varepsilon_{\varepsilon^{-1}t} - \hat{\mathcal{Q}}^N_{\varepsilon^{-1}t}|\, dt.$$

Introducing the $\mathbb{P}^\varepsilon$ Wiener process $\hat{\widetilde{W}}^2{}_s := (\varepsilon)^{-1/2}\widetilde{W}^2_{\varepsilon s}$ we have that the process $(\hat{Q}^\varepsilon_s)_{s\in[0,1/\varepsilon]}$ solves:

$$d\hat{Q}^\varepsilon_s = (B\hat{Q}^\varepsilon_s + F(X_{\varepsilon s}, \hat{Q}^\varepsilon_s))\, ds - G\delta^{2,\varepsilon,N}(s)\, ds - +Gd\hat{\widetilde{W}}^2_s, \quad s \geq 0, \qquad \hat{Q}^\varepsilon_0 = q_0, \tag{5.26}$$

moreover $(\hat{\mathcal{Q}}^N_s)_{s\in[0,1/\varepsilon]}$ solves:

$$d\hat{\mathcal{Q}}^N_s = (B\hat{\mathcal{Q}}^N_s + F(X^N_{\varepsilon s}, \hat{\mathcal{Q}}^N_s))\, dt - G\delta^{2,\varepsilon,N}(s)\, ds + Gd\hat{\widetilde{W}}^2_s, \quad s \geq 0, \qquad \hat{\mathcal{Q}}^N_0 = q_0. \tag{5.27}$$

Therefore by Lemma 3.10 and hypothesis 4.1 we have for all $p \geq 1$:

$$\sup_{s\in[0,1/\varepsilon]} \widetilde{\mathbb{E}}^\varepsilon[|\hat{\mathcal{Q}}^N_t|^p] \leq c_p \left(1 + |q_0|^p + \sup_{s\in[0,1/\varepsilon]} \widetilde{\mathbb{E}}^\varepsilon|X_s|^p + \sup_{s\in[0,1/\varepsilon]} \widetilde{\mathbb{E}}^\varepsilon \left|\int_0^s e^{(s-r)B} Gd\hat{\widetilde{W}}^2_r\right|^p + L_\xi\right), \tag{5.28}$$

for a constant $c_p$ independent of $\varepsilon$ and $N$. Arguing as before, we have that

$$\widetilde{\mathbb{E}}^\varepsilon|X_s|^p = \widetilde{\mathbb{E}}^\varepsilon(\rho^{-1/2}\rho^{1/2}|X_s|^p) \leq (\widetilde{\mathbb{E}}^\varepsilon\rho^{-1})^{1/2}(\widetilde{\mathbb{E}}^\varepsilon(\rho|X_s|^{2p}))^{1/2} \leq C(\mathbb{E}|X_s|^{2p})^{1/2},$$

and

$$\begin{aligned}
&\widetilde{\mathbb{E}}^\varepsilon \left|\int_0^s e^{(s-r)B} Gd\hat{\widetilde{W}}^2_r\right|^p = \widetilde{\mathbb{E}}^\varepsilon \left(\rho^{-1/2}\rho^{1/2} \left|\int_0^s e^{(s-r)B} Gd\hat{\widetilde{W}}^2_r\right|^p\right) \\
&\leq (\widetilde{\mathbb{E}}^\varepsilon\rho^{-1})^{1/2} \left(\widetilde{\mathbb{E}}^\varepsilon \left(\rho \left|\int_0^s e^{(s-r)B} Gd\hat{\widetilde{W}}^2_r\right|^{2p}\right)\right)^{1/2} \leq C \left(\mathbb{E} \left|\int_0^s e^{(s-r)B} Gd\hat{W}^2_r\right|^{2p}\right)^{1/2},
\end{aligned}$$

for some constant $C > 0$ independent of $\varepsilon$. Therefore, bearing in mind the estimate (3.2) for the slow component $X$ and Hypothesis 3.7, we conclude that there exists a constant $c > 0$, independent of $\varepsilon$ and $N$, such that

$$\sup_{s\in[0,1/\varepsilon]} \widetilde{\mathbb{E}}^\varepsilon[|\hat{\mathcal{Q}}^N_t|^p] \leq c. \tag{5.29}$$

Again by Lemma 3.10 one has that for all $s > 0$,

$$|\hat{Q}^\varepsilon_s - \hat{\mathcal{Q}}^N_s| \leq c \int_0^s e^{-\eta(s-\ell)}|X_{\varepsilon\ell} - X^N_{\varepsilon\ell}|d\ell \leq c\Delta_{X,N},$$

thus, arguing as in (5.22),

$$\widetilde{\mathbb{E}}^\varepsilon \int_0^1 (1 + |\bar{Z}_t|)|\hat{Q}^\varepsilon_{\varepsilon^{-1}t} - \hat{\mathcal{Q}}^N_{\varepsilon^{-1}t}|dt \leq c\widetilde{\mathbb{E}}^\varepsilon \left[\Delta_{X,N} \int_0^1 (1 + |\bar{Z}_t|)dt\right] \leq C[\mathbb{E}\Delta^4_{X,N}]^{1/4}, \tag{5.30}$$

as above.

Now we come to the last term. We apply successively (5.11) and (5.29) to get the following estimates (the value of the constant $c$ below can change from line to line but never depends neither on $k$ nor on $N$ or on $\varepsilon$):

$$\begin{aligned}
&\left|\varepsilon\widetilde{\mathbb{E}}^\varepsilon \sum_{k=1}^{2^N-1} (\check{Y}^{N,k}_{t_k/\varepsilon} - \check{Y}^{N,k}_{t_{k+1}/\varepsilon})\right| \le c\varepsilon \sum_{k=1}^{2^N-1} \widetilde{\mathbb{E}}^\varepsilon \left[(1+|\widetilde{Z}^N_{t_k}|)(1+|\hat{\mathcal{Q}}^N_{t_k/\varepsilon}|+|\hat{\mathcal{Q}}^N_{t_{k+1}/\varepsilon}|)\right] \\
&\le c\varepsilon \sum_{k=1}^{2^N-1} \left[\widetilde{\mathbb{E}}^\varepsilon(1+|\widetilde{Z}^N_{t_k}|)^{4/3}\right]^{3/4} \left[\widetilde{\mathbb{E}}^\varepsilon(1+|\hat{\mathcal{Q}}^N_{t_k/\varepsilon}|+|\hat{\mathcal{Q}}^N_{t_{k+1}/\varepsilon}|)^4\right]^{1/4} \\
&\le c\varepsilon \sum_{k=1}^{2^N-1} \left[1+\left(\widetilde{\mathbb{E}}^\varepsilon|\widetilde{Z}^N_{t_k})^{4/3}\right)^{3/4}\right].
\end{aligned}$$

Proceeding as above, recalling that the law of $\widetilde{Z}^N_{t_k}$ depends only on the law of the process $(X_t)$ we have:

$$\widetilde{\mathbb{E}}^\varepsilon[(|\widetilde{Z}^N_{t_k}|)^{4/3}] \le \left[\mathbb{E}\rho^{-2}\right]^{1/3} \left[\mathbb{E}|\widetilde{Z}^N_{t_k}|^2\right]^{2/3} \le c2^{\frac{2}{3}N} \left[\mathbb{E}\int_0^t |\bar{Z}_t|^2 dt\right]^{2/3}.$$

At last we sum up the latter result, (5.22), (5.24) and (5.30) to get:

$$\begin{aligned}
|Y^\varepsilon_0 - \bar{Y}_0| &\le \widetilde{\mathbb{E}}^\varepsilon \int_0^1 |\mathcal{R}^{\varepsilon,N}_{t/\varepsilon}|\, dt + \varepsilon\widetilde{\mathbb{E}}^\varepsilon \sum_{k=1}^{2^N} |\check{Y}^{N,k}_{t_k/\varepsilon} - \check{Y}^{N,k}_{t_{k+1}/\varepsilon}| \\
&\le C[\mathbb{E}\Delta^4_{X,N}]^{1/4} + C(\mathbb{E}\Delta_{Z,N})^{1/2} + \varepsilon c2^{\frac{3}{2}N}\left(\mathbb{E}\int_0^1 |\bar{Z}_t|^2\, dt\right)^{1/2} + \varepsilon c 2^N.
\end{aligned}$$

So letting first $\varepsilon$ tend to 0 and then $N$ to $\infty$ the claim follows, by (5.23) and (5.25). □

**Remark 5.5** Consider the following class of forward backward systems with initial time $\tau \in [0,1]$

$$\begin{cases} dX^{\tau,x}_t = AX^{\tau,x}_t\, dt + RdW^1_t, & t \le 1, \\ -d\bar{Y}^{\tau,x}_t = \lambda(X^{\tau,x}_t, \bar{Z}^{\tau,x}_t)\, dt - \bar{Z}^{\tau,x}_t\, dW^1_t, & t \le 1, \\ X^{\tau,x}_\tau = x, \quad \bar{Y}^{\tau,x}_1 = h(X^{\tau,x}_1). \end{cases} \tag{5.31}$$

If we set $v(\tau,x) = \bar{Y}^{\tau,x}_\tau$ then it is shown in [17] that $v$ is a deterministic continuous function $[0,1]\times H \to \mathbb{R}$ being Gateaux differentiable with respect to the second variable. Thus it is the unique mild solution, in the sense of definition 6.1 of [16] of the nonlinear Kolmogorov equation; see [16, Th. 6.2].

$$\begin{cases} \dfrac{\partial v(t,x)}{\partial t} + \mathcal{L}v(t,x) = \lambda(x, \nabla v(t,x)R), & t\in[0,1],\, x\in H, \\ v(1,x) = h(x), \end{cases}$$

where $\mathcal{L}$ is the second order operator

$$\mathcal{L}g(x) = \frac{1}{2}\mathrm{Tr}[RR^*\nabla^2_x g(x)] + \langle Ax, \nabla v(t,x)\rangle, \quad g \in \mathcal{C}^2(H),$$

$\nabla^2 g(x) \in \mathcal{L}(H)$ being the second derivative of $g$ in $x$.
In particular the limit $\lim_{\varepsilon\to 0} Y^\varepsilon_0$ can also be represented by the solution of the above HJB equation as:

$$\lim_{\varepsilon\to 0} Y^\varepsilon_0 = \bar{Y}^{0,x_0}_0 = v(0,x_0).$$

## 6 The two scale control problem

In this section we are finally in a position to exploit the convergence of solutions of BSDEs proved in the previous section to solve our original problem of characterizing the limit of value functions $V^\varepsilon(x_0, q_0)$ as $\varepsilon \to 0$ (see (1.3) in the Introduction).

As we've stated in the introduction we formulate the control problem in a weak form, see [13]. This choice is typical in the BSDEs approach and allows to easily identify the value function with the solution of the backward equation, see also the Section 7 in [16]. Remark 6.3 below reminds the reader about the relation with the original formulation.

As we have already explained in Remark 3.6, we need to add the following hypothesis:

**Hypothesis 6.1** *$R$ admits a bounded right inverse $R^{-1} \in L(H; \Xi)$.*

Given the solution $(X, Q^\varepsilon)$ of system (3.1) and a progressively measurable process $(\alpha_t)_{t\in[0,1]}$ taking its values in a complete metric space $U$ we denote by $\Theta^{\varepsilon,\alpha}$ the density

$$\Theta^{\varepsilon,\alpha} = \exp\left(\int_0^1 R^{-1} b(X_t, Q_t^\varepsilon, \alpha_t) dW_t^1 - \frac{1}{2}\int_0^1 |R^{-1} b(X_t, Q_t^\varepsilon, \alpha_t)|^2 \, dt\right) \times \exp\left(\int_0^1 \frac{1}{\sqrt{\varepsilon}} \rho(\alpha_t) dW_t^2 - \frac{1}{2}\int_0^1 \frac{1}{\varepsilon} |\rho(\alpha_t)|^2 \, dt\right),$$

where $b : H \times K \times U \to H$ and $\rho : U \to K$ are measurable functions satisfying suitable assumptions listed below.

We also consider the following cost functional:

$$J^\varepsilon(x_0, q_0, \alpha) = \mathbb{E}\left[\Theta^{\varepsilon,\alpha}\left(\int_0^1 l(X_t, Q_t^\varepsilon, \alpha_t) dt + h(X_1)\right)\right], \tag{6.1}$$

where $l : H \times K \times U \to \mathbb{R}$ and $h : H \to \mathbb{R}$ are measurable and satisfy the assumptions below:

**Hypothesis 6.2** *There are positive constants $L$ and $M$ such that :*

$$|b(x, q, u) - b(x', q', \alpha)| \le L(|x - x'| + |q - q'|), \qquad \forall\, q, q' \in K, x, x' \in H,\ \alpha \in U,$$

$$|l(x, q, \alpha) - l(x', q', \alpha)| \le L(|x - x'| + |q - q'|), \qquad \forall\, q, q' \in K, x, x' \in H,\ \alpha \in U,$$

$$|h(x) - h(x')| \le L|x - x'|, \qquad \forall x, x' \in H,$$

$$|b(x, q, \alpha)|, |l(x, q, \alpha)|, |\rho(\alpha)|, |h(x)| \le M, \qquad \forall q \in K, x \in H,\ \alpha \in U.$$

**Remark 6.3** *We recall that if $d\mathbb{P}^{\varepsilon,\alpha} := \Theta^{\varepsilon,\alpha} d\mathbb{P}$ then under probability $\mathbb{P}^{\varepsilon,\alpha}$ the process:*

$$(\mathcal{W}_t^1, \mathcal{W}_t^2) = (-\int_0^t R^{-1} b(X_r, Q_r^\varepsilon, \alpha_r) dr + W_t^1, -\frac{1}{\sqrt{\varepsilon}}\int_0^t \rho(\alpha_r) dr + W_t^2),$$

*is a cylindrical $\mathcal{F}_t$-Wiener process in $\Xi \times \Xi$, where $\mathcal{F}_t$ is the filtration introduced in Section 2. Moreover with respect to $(\mathcal{W}_t^1, \mathcal{W}_t^2)$ the couple of processes $(X_t, Q_t^\varepsilon)$ satisfies the controlled system:*

$$\begin{cases} dX_t = AX_t \, dt + b(X_t, Q_t^\varepsilon, \alpha_t) dt + R d\mathcal{W}_t^1, & X_0 = x_0, \\ \\ \varepsilon dQ_t^\varepsilon = (BQ_t^\varepsilon + F(X_t^\varepsilon, Q_t^\varepsilon) \, dt + G\rho(\alpha_t) dt + \varepsilon^{1/2} G \, d\mathcal{W}_t^2, & Q_0^\varepsilon = q_0. \end{cases} \tag{6.2}$$

*Moreover:*

$$J^{\varepsilon}(x_0, q_0, \alpha) = \mathbb{E}^{\mathbb{P}^{\varepsilon,\alpha}} \left( \int_0^1 l(X_t, Q_t^{\varepsilon}, \alpha_t) dt + h(X_1) \right),$$

*thus the one introduced here is a correct formulation of our original problem* (1.1) *and* (1.2).

We define, for $x \in H$, $q \in K$ and $z, \xi \in \Xi^*$ :

$$\psi(x, q, z, \xi) = \inf_{\alpha \in U} \{ l(x, q, \alpha) + z[R^{-1} b(x, q, \alpha)] + \xi \rho(\alpha) \}, \tag{6.3}$$

and notice that, by straight forward considerations, under Hypotheses 6.1 and 6.2, the Hamiltonian $\psi$ verifies hypothesis 4.1.
The main result of this paper is now just an immediate consequence of Theorem 5.4.

**Theorem 6.4** *Denote by* $V^{\varepsilon}$ *the value function of our control problem that is:*

$$V^{\varepsilon}(x_0, q_0) := \inf_{\alpha} J^{\varepsilon}(x_0, q_0, \alpha),$$

*where the infimum is taken over all progressive processes* $\alpha$ *with values in* $U$.
*The sequence* $V^{\varepsilon}(x_0, q_0)$ *converges to the solution* $\bar{Y}_0$ *of equation* (5.2) *evaluated at zero.*

**Proof.** In [14] it is shown that $V^{\varepsilon}(x_0, q_0) = Y_0^{\varepsilon}$ (see (5.1)). The claim then follows by Theorem 5.4. □

**Remark 6.5** The nonlinearity $\lambda$ in the limit equation (5.2) has itself a control theoretic interpretation. Namely, fixed $x \in H$ and $z \in \Xi^*$, let us consider the following ergodic control problem with *state equation*

$$d\hat{Q}_s^{\beta} = B\hat{Q}_s^{\beta} ds + F(x, \hat{Q}_s^{\beta})\, ds + G\rho(\beta_s) ds + G d\hat{W}_s^2, \tag{6.4}$$

and *ergodic cost functional*:

$$\check{J}(x, z, \beta) = \liminf_{\delta \to 0} \mathbb{E}\, \delta \int_0^{\frac{1}{\delta}} [z R^{-1} b(x, \hat{Q}_s^{\beta}, \beta_s) + l(x, \hat{Q}_s^{\beta}, \beta_s)] ds. \tag{6.5}$$

Then $\lambda(x, z)$ is the value function of the ergodic control problem that we have just described, that is:

$$\lambda(x, z) = \inf_{\beta} \check{J}(x, z, \beta),$$

where the infimum is taken over all progressive processes $\beta : [0, \infty[ \to U$.
Moreover notice that, in particular, being the infimum of linear functionals, the map $z \to \lambda(x, z)$ is concave.

Finally notice that the result was proven in [15] with lim inf replaced by lim sup in the definition (6.5) of the ergodic cost nevertheless, as it can be easily verified, this substitution is inessential in the argument reported in [15]

**Example 6.6** We provide a simple example to which our result applies. Let us consider the following two scale system of classical controlled reaction diffusion SPDEs in one space dimension driven by space time white noises see, for instance [8] Section 11.2 or [12]:

$$\begin{cases} \dfrac{\partial}{\partial t}u^\varepsilon(t,x) = \dfrac{\partial^2}{\partial x^2}u^\varepsilon(t,x) + b(u^\varepsilon(t,x), v^\varepsilon(t,x), \alpha(t,x)) + \sigma(x)\dfrac{\partial}{\partial t}\mathcal{W}^1(t,x), \\ \varepsilon\dfrac{\partial}{\partial t}v^\varepsilon(t,x) = (\dfrac{\partial^2}{\partial x^2} - m)v^\varepsilon(t,x) + f(u^\varepsilon(t,x), v^\varepsilon(t,x)) + \rho(x)r(\alpha(t,x)) \\ \qquad\qquad + \varepsilon^{1/2}\rho(x)\frac{\partial}{\partial t}\mathcal{W}^2(t,x), \\ u^\varepsilon(t,0) = u^\varepsilon(t,1) = v^\varepsilon(t,0) = v^\varepsilon(t,1) = 0, \\ u^\varepsilon(0,x) = u^0(x),\ v^\varepsilon(0,x) = v^0(x), \qquad\qquad t \in [0,1],\ x \in [0,1], \end{cases} \tag{6.6}$$

where $(\mathcal{W}^1(t,x))$ and $(\mathcal{W}^2(t,x))$ are independent space-time white noises. Here $(u^\epsilon)$ represents the slow state, $(v^\epsilon)$ the quick one and $\alpha$ is the control.
We make the following assumptions on the coefficients:

1. $m$ is a positive constant.
2. $b, f$ are continuous maps, $b$ is bounded and Lipschitz continuous w.r.t to the first two variables uniformly w.r.t. the control, moreover $f$ is Lipschitz continuous with a constant smaller then $m$.
3. $\sigma, \rho$ are measurable and bounded functions $[0,1] \to \mathbb{R}$. Moreover we ask that $|\sigma(x)| \geq c_\sigma$, for a.e. $x \in [0,1]$ and a suitable constant $c_\sigma > 0$.
4. $r : \mathbb{R} \to \mathbb{R}$ is a measurable and bounded map.
5. An admissibile control $\alpha$ is any bounded progressive measurable process $\alpha : \Omega \times [0,1] \times [0,1] \to \mathbb{R}$ and the cost functional is
$$J^\varepsilon(u_0, v_0) = \mathbb{E}\int_0^1 \int_0^1 \ell(u^\varepsilon(t,x), v^\varepsilon(t,x), \alpha(t,x))\,dx\,dt + \int_0^1 h(u^\varepsilon(1,x))\,dx,$$
with $\ell$ and $h$ Lipschitz continuous and bounded functions.

The abstract formulation in $H = K = L^2(0,1)$ and $U = L^2(0,1)$ is identical to the one in [17, section 5]. In this same place it is shown that Hypotheses 3.1—3.7, 4.1, 5.1 and 6.2 hold. Notice that thus Theorem 5.4 and Theorem 6.4 apply.

# 7 Control interpretation of the limit forward-backward system

Since we were able to interpret the limit value function as the solution of a *reduced* forward-backward system we can now hope to see it as the value function of a correspondingly *reduced* control problem.

Most of our analysis in this section is based on the fact that $\lambda$ is concave with respect to $z$. In particular, by Fenchel-Moreau theorem (translated in the obvious way for concave functions instead than for convex ones), we can write $\lambda = \lambda_{**}$ where for all $x \in H$:

$$\lambda_*(x,p) = \inf_{z\in\Xi^*}\big(-zp - \lambda(x,z)\big), \quad p \in \Xi$$

and the map $\lambda_*(x,\cdot)$ is an upper semicontinuous concave function with non empty domain in $\Xi$. Thus for all $x \in H$, $z \in \Xi^*$:

$$\lambda(x,z) = \inf_{p\in\Xi}\big(-zp - \lambda_*(x,p)\big).$$

Recalling that $\lambda$ is Lipschitz continuous with respect to $z$ uniformly in $x$ and denoting by $L$ the Lipschitz constant we have:

$$\lambda_*(x,p) = -\infty, \text{ whenever } |p| > L.$$

and consequently:

$$\lambda(x,z) = \inf_{p\in\Xi,\, |p|\le L} \big(-zp - \lambda_*(x,p)\big). \tag{7.1}$$

Moreover, $\lambda_*(x,p) \le -\lambda(x,0) \le c(1+|x|)$, thus, for any process $(\mathfrak{p}_t)_{0\le t\le 1}$ with values in $\Xi$, the process

$$\left(\int_0^t \lambda_*(X_s,\mathfrak{p}_s)ds\right)_{0\le t\le 1}$$

is well-defined and takes values in $[-\infty,\infty)$.
We have the following characterization:

**Theorem 7.1** *Assume 3.1—3.7, 4.1 and 5.1 then it holds:*

$$\bar{Y}_0 = \inf_{\mathfrak{p}} \mathbb{E}^{\mathfrak{p}}\left(h(X_1) - \int_0^1 \lambda_*(X_t,\mathfrak{p}_t)dt \middle| \mathcal{F}_t\right),$$

*where:*

1. *$(X_t)_{t\ge 0}$ is, as before, the mild solution of the following stochastic differential equation:*

$$dX_t = AX_t dt + RdW_t^1, \quad X_0 = x_0;$$

2. *the infimum is extended to all $\Xi$-valued, progressively measurable processes $(\mathfrak{p}_s)_{0\le s\le 1}$ that are bounded by $L$;*

3. *$\mathbb{E}^{\mathfrak{p}}$ denotes the mean value with respect the probability $\mathbb{P}^{\mathfrak{p}}$ under which*

$$W_t^{\mathfrak{p}} := \int_0^t \mathfrak{p}_s ds + W_t^1$$

*is a Wiener process.*

*Notice that, with respect to $(W^{\mathfrak{p}})$ process $(X)$ solves the controlled stochastic differential equation:*

$$dX_t = AX_t dt - R\,\mathfrak{p}_t dt + RdW_t^{\mathfrak{p}}, \quad X_0 = x_0.$$

**Proof.** Given any $\Xi$ valued progressively measurable process $(\mathfrak{p}_t)_{t\ge 0}$ with $|\mathfrak{p}_t| \le L$ by (7.1) we get:

$$\begin{aligned}\bar{Y}_t &= h(X_1) + \int_t^1 \lambda(X_s,\bar{Z}_s)ds - \int_t^1 \bar{Z}_s dW_s^1 \\ &\le h(X_1) - \int_t^1 \big(\bar{Z}_s\mathfrak{p}_s + \lambda_*(X_s,\mathfrak{p}_s)\big)\, ds - \int_t^1 \bar{Z}_s dW_s^1.\end{aligned}$$

and by the definition of $(W^{\mathfrak{p}})$:

$$\bar{Y}_t \le h(X_1) - \int_t^1 \lambda_*(X_s,\mathfrak{p}_s)ds - \int_t^1 \bar{Z}_s dW_s^{\mathfrak{p}},$$

which shows that:

$$\bar{Y}_t \le \mathbb{E}^{\mathfrak{p}}\left(h(X_1) - \int_t^1 \lambda_*(X_s,\mathfrak{p}_s)ds \middle| \mathcal{F}_t\right).$$

Conversely, we may choose, for any $n \geq 1$, $(\mathfrak{p}^n_t)_{0\leq t\leq 1}$ such that $|\mathfrak{p}^n_t| \leq L$ and

$$-\bar{Z}_t \mathfrak{p}^n_t - \lambda_*(X_t, \mathfrak{p}^n_t) - 1/n \leq \lambda(X_t, \bar{Z}_t).$$

By a measurable selection theorem, see for instance Theorem 6.9.13 in [5], we can also assume the process $\mathfrak{p}^n$ to progressive measurable.
We have

$$\bar{Y}_t \geq h(X_1) - \int_t^1 \left( \bar{Z}_s \mathfrak{p}^n_s + \lambda_*(X_s, \mathfrak{p}^n_s) + \frac{1}{n} \right) ds - \int_t^1 \bar{Z}_s dW^1_s,$$

and rewriting the above in terms of $W^{\mathfrak{p}^n}$:

$$\bar{Y}_t + \frac{1-t}{n} \geq h(X_1) - \int_t^1 \lambda_*(X_t, \mathfrak{p}^n_t) dt - \int_t^1 \bar{Z}_t dW^{\mathfrak{p}^n}_t.$$

Therefore we can conclude that:

$$\bar{Y}_t \geq \lim_{n\to\infty} \mathbb{E}^{\mathfrak{p}^n} \left( h(X_1) - \int_t^1 \lambda_*(X_s, \mathfrak{p}^n_s) ds \middle| \mathcal{F}_t \right).$$

and the claim is proved. □

## Acknowledgements

The authors wish to thank Francois Delarue for several enlightening discussions on the content of the present paper.
Moreover would like to thank both referees for their careful reading of our paper and for their remarks.

## References

[1] S. Cerrai and M. Freidlin. On the Smoluchowski-Kramers approximation for a system with an infinite number of degrees of freedom. *to appear*

[2] O. Alvarez and M. Bardi. Viscosity solutions methods for singular perturbations in deterministic and stochastic control. *SIAM J. Control Optim.* 40 No.4 (2001), pp. 1159–1188

[3] A. Bensoussan. *Perturbation Methods in Optimal Control.* Wyley/Gauthiers-Villars, U.K., 1988.

[4] P. Briand and Y. Hu Stability of BSDEs with random terminal time and homogenization of semilinear elliptic PDEs. *J.Func. Anal.,* 155 (1998), pp. 455-494.

[5] V. I. Bogachev. *Measure Theory.* Springer-Verlag, Berlin, Heidelberg, 2007

[6] G. Da Prato, S. Kwapien and J. Zabczyk. Regularity of solutions of linear stochatsic equations in Hilbert spaces. *Stochastics*, 23, 1-23.

[7] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions.* Cambridge University Press, Cambridge, 1992.

[8] G. Da Prato and J. Zabczyk. *Ergodicity for infinite-dimensional systems.* London Mathematical Society Lecture Note Series, 229. Cambridge University Press, Cambridge, 1996.

[9] A. Debussche, Y. Hu and G. Tessitore. Ergodic BSDEs under weak dissipative assumptions. *Stoc. Pro. Appl.* 121 (2011), 407-426.

[10] N. El Karoui, S. Peng and M.C. Quenez: Backward Stochastic Differential Equations in Finance, *Mathematical Finance* 7 (1997), 1-71.

[11] G. Fabbri, F. Gozzi and A. Święch. *Stochastic optimal control in infinite dimension.* Probability Theory and Stochastic Modelling, 82. Springer, Cham, 2017

[12] M.I. Freidlin. Random perturbations of reaction-diffusion equations: the quasi-deterministic approximation, *Trans. Amer. Math. Soc.*, 305 (1988), 665-697

[13] W.H Fleming and H.M. Soner. *Controlled Markov processes and viscosity solutions.* Applications of Mathematics (New York), 25. Springer-Verlag, New York, 1993, xvi+428.

[14] M. Fuhrman, G. Tessitore. Generalized directional gradients, backward stochastic differential equations and mild solutions of semilinear parabolic equations. Applied Matematics and Optimization, 51(3) (2005), 279-332.

[15] M. Fuhrman, Y. Hu and G. Tessitore. Ergodic BSDES and optimal ergodic control in Banach spaces. Siam Journal On Control and Optimization, 48(3) (2009), 1542-1566.

[16] M. Fuhrman and G. Tessitore. Nonlinear Kolmogorov equations in infinite dimensional spaces: The backward stochastic differential equations approach and applications to optimal control. *Ann. Probab.* 30 (2002), 1397–1465.

[17] M. Fuhrman and G. Tessitore. The Bismut-Elworthy formula for backward SDE's and applications to nonlinear Kolmogorov equations and control in infinite dimensional spaces. *Stochastics and Stochastic Reports* 74 (2002), 429–464.

[18] Y. Kabanov and S. Pergamenshchikov. *Two-scale Stochastic Systems.* Applications of Mathematics, 49. Springer-Verlag, Berlin, Heidelberg, New York, 2003.

[19] Y. Kabanov and W. Runggaldier. On Control of Two-scale Stochastic Systems with Linear Dynamics in the Fast Variables. *Math. Control Signals Systems* 9 (1996), 107–122.

[20] Y. Hu, P.-Y. Madec and A. Richou. A probabilistic approach to large time behavior of mild solutions of HJB equations in infinite dimension. *SIAM J. Control Optim.* 53 (2015), 378–398.

[21] H. Kushner. *Weak convergence Methods and Singularly Perturbed Stochastic Control And Filtering Problem.* Birkauser, Boston, 1990. Backward stochastic differential equations and quasilinear parabolic partial differential equations

[22] É. Pardoux and S. Peng, Backward stochastic differential equations and quasilinear parabolic partial differential equations In *Stochastic partial differential equations and their applications (Charlotte, NC) 1991*, 176 Springer, Berlin, 200–217

[23] S. Cerrai *Second order PDE's in finite and infinite dimension* Lecture Notes in Mathematics, 1762 Springer-Verlag, Berlin, (2001), x+330.

[24] A. Święch. "Unbounded" second order partial differential equations in infinite-dimensional Hilbert spaces. *Comm. Partial Differential Equations.* 19(11-12), (1994), 1999–2036.