ESAIM: COCV 29 (2023) 52
https://doi.org/10.1051/cocv/2023020

ESAIM: Control, Optimisation and Calculus of Variations
www.esaim-cocv.org

SINGULAR PERTURBATIONS IN STOCHASTIC OPTIMAL CONTROL WITH UNBOUNDED DATA

Martino Bardi[1,*] and Hicham Kouhkouh[2]

Abstract. We study singular perturbations of a class of two-scale stochastic control systems with unbounded data. The assumptions are designed to cover some relaxation problems for deep neural networks. We construct effective Hamiltonian and initial data and prove the convergence of the value function to the solution of a limit (effective) Cauchy problem for a parabolic equation of HJB type. We use methods of probability, viscosity solutions and homogenization.

Mathematics Subject Classification. 35B25, 93E20, 93C70, 49L25.

Received July 27, 2022. Accepted March 23, 2023.

1. Introduction

In this paper, we study the asymptotic behavior as $\varepsilon \to 0$ of a system of controlled two-scale stochastic differential equations

$$\begin{aligned} \mathrm{d}X_t &= f(X_t, Y_t, u_t)\,\mathrm{d}t + \sqrt{2}\,\sigma^\varepsilon(X_t, Y_t, u_t)\,\mathrm{d}W_t, \\ \mathrm{d}Y_t &= \frac{1}{\varepsilon}\, b(X_t, Y_t)\,\mathrm{d}t + \sqrt{\frac{2}{\varepsilon}}\,\varrho(X_t, Y_t)\,\mathrm{d}W_t, \end{aligned} \qquad (SDE\left(\tfrac{1}{\varepsilon}\right))$$

where $X_t \in \mathbb{R}^n$ is the *slow* dynamics, $Y_t \in \mathbb{R}^m$ is the *fast* dynamics, u_t is the control taking values in a given compact set U and W_t is a multidimensional Brownian motion. We will allow the components of the drift and the diffusion of the slow dynamics to be unbounded and with at most linear growth in the fast variables Y. While the diffusion coefficient of the process X can be degenerate (*i.e.* $\sigma^\varepsilon = 0$ is allowed), the diffusion coefficient of the process Y is required to be nondegenerate, in particular we will assume for our main result that $\varrho\varrho^\top$ is the identity matrix times a positive constant, in addition to other structural assumptions on the data that we shall make precise later. We carry our analysis in the context of stochastic optimal control problems with payoff

Keywords and phrases: Singular perturbations, two-scale systems, stochastic optimal control, homogenization, viscosity solutions, Hamilton-Jacobi-Bellman equations, invariant measures.

[1] Department of Mathematics "T. Levi-Civita", University of Padova, via Trieste, 63, 35121 Padova, Italy.
[2] RWTH Aachen University, Institut für Mathematik, RTG Energy, Entropy, and Dissipative Dynamics, Templergraben 55 (111810), 52062 Aachen, Germany.

[*] Corresponding author: bardi@math.unipd.it

functional

$$J(t,x,y,u) := \mathbb{E}\left[e^{\lambda(t-T)} g(X_T,Y_T) + \int_t^T \ell(s,X_s,Y_s,u_s) e^{\lambda(s-T)} \mathrm{d}s \,\middle|\, X_t = x, Y_t = y \right],$$

and exploit that the value function $V^\varepsilon(t,x,y) := \sup_u J(t,x,y,u)$ solves in the viscosity sense a fully nonlinear parabolic degenerate Hamilton-Jacobi-Bellman PDE in $(0,T)\times\mathbb{R}^n\times\mathbb{R}^m$.

Our motivation comes from the Stochastic Gradient Descent algorithm in the context of Deep Learning and Big Data analysis. The following special case of equation ($SDE\left(\frac{1}{\varepsilon}\right)$), without control, was proposed in [13] to justify an algorithm of Stochastic Gradient Descent called Deep Relaxation (see also [36]). Given a *loss function* $\phi:\mathbb{R}^n\to\mathbb{R}$ to be minimized, consider its quadratic perturbation in double variables

$$\Phi(y,x) := \phi(y) + \frac{1}{2\gamma}|x-y|^2$$

and the partial stochastic gradient descent associated to it

$$\begin{aligned} \mathrm{d}X_s &= -\nabla_x \Phi(Y_s,X_s)\,\mathrm{d}s, \quad X_0 = x\in\mathbb{R}^n \\ \mathrm{d}Y_s &= -\frac{1}{\varepsilon}\nabla_y\Phi(Y_s,X_s)\,\mathrm{d}s + \sqrt{\frac{2}{\varepsilon}}\beta^{-1/2}\,\mathrm{d}W_s, \quad Y_0 = y\in\mathbb{R}^n. \end{aligned} \tag{1.1}$$

The calculations in [13] show that one should expect the limit as $\varepsilon\to 0$ in the above system of SDEs to be

$$\mathrm{d}\hat{X}_s = \int_{\mathbb{R}^n} -\frac{1}{\gamma}(X_s - y)\rho_\beta^\infty(\mathrm{d}y;X_s)\,\mathrm{d}s, \quad \hat{X}_0 = x\in\mathbb{R}^n,$$

where $\rho_\beta^\infty(y;x)$ is the invariant (Gibbs) measure associated to the process $Y.$ (with $\varepsilon = 1$ and frozen $X_s = x$). The latter can be written as

$$\mathrm{d}\hat{X}_s = -\nabla\phi_\gamma(\hat{X}_s)\mathrm{d}s, \quad \hat{X}_0 = x\in\mathbb{R}^n,$$

that is the deterministic gradient descent of the regularized loss function

$$\phi_\gamma(x) := -\frac{1}{\beta}\log\left(G_{\beta^{-1}\gamma} * \exp(-\beta\phi(x))\right)$$

where

$$G_{\beta^{-1}\gamma}(x) := (2\pi\gamma)^{-n/2}\exp\left(-\frac{\beta}{2\gamma}|x|^2\right)$$

is the heat kernel, and $\beta,\gamma>0$ are parameters used to tune the algorithm. The function ϕ_γ above is called *local entropy*, and it is useful in the search of robust minima, because it measures both the depth and the flatness of the valleys in the landscape of the graph of ϕ, see [12]. Note also that in equation (1.1) the drifts are $f = (y-x)/\gamma$ and $b = -\nabla\phi + (x-y)/\gamma$, which are unbounded in x and y.

In the present paper, under rather general assumptions, we prove the convergence as $\varepsilon\to 0$ of the value functions V^ε associated with the the singularly perturbed control system ($SDE\left(\frac{1}{\varepsilon}\right)$) to a function $V(t,x)$ independent of y, and V is characterized as the unique viscosity solution of a Cauchy problem for a limiting HJB equation. The *effective Hamiltonian* $\bar{H}$ driving such equation and the *effective initial data* $\bar{g}$ are explicitly computed by suitable averages. In particular, the result applies to the model problem (1.1) if $\phi\in C^1(\mathbb{R}^n)$ is bounded from below and such that $\nabla\phi$ is Lipschitz continuous with constant L, and γ is small enough ($\gamma < \frac{1}{L}$). This holds also if the equation for $X_\cdot$ in equation (1.1) involves a control u_s, *e.g.*, it is of the form

$$\mathrm{d}X_s = -u_s \nabla_x \Phi(Y_s, X_s) \tag{1.2}$$

where $u_s \in [0,1]$ is the *learning rate* of the SGD algorithm. This variant is used in the companion paper [8] to prove that by equation (1.1) modified with equation (1.2) one reaches in expectation a value of ϕ lower than the one got by classical stochastic gradient descent. In [8] we also characterize explicitly the limiting system of controlled SDEs in $\mathbb{R}^n$ and prove results on the convergence of the trajectories of equation ($SDE\left(\frac{1}{\varepsilon}\right)$) to the trajectories of such effective system as $\varepsilon \to 0$. These results can be found also in the second author's thesis [26].

Our convergence theorem includes the previous results in [5, 7], where the coefficients in the slow variable were assumed to be bounded with respect to the fast variables and mostly viscosity method for the HJB PDE were employed. However, some important parts of the proofs in [5, 7] do not work in the current setting. Here we use first a truncation to big balls of the cell problem, an HJB equation of ergodic type that formally gives the effective Hamiltonian, and then use probabilistic estimates on the exit time τ_n^Y of the process

$$\mathrm{d}Y_t = b(x, Y_t)\,\mathrm{d}t + \sqrt{2}\varrho(x, Y_t)\,\mathrm{d}W_t \tag{1.3}$$

from balls of radius n as $n \to \infty$. This approach is new to our knowledge in the present context. Here some ideas are borrowed from [23].

Our results also allow to generalise several applications of singular perturbations to finance, *e.g.*, models of pricing and trading derivative securities in financial markets with stochastic volatility, as in [7, 18], applications in economics and advertising theory, as in [5], and connections to large deviations as in [17, 19, 39].

There is a wide literature on singular perturbations for ODEs and control systems that goes back to the late 60's, see [25] and the references therein, and for diffusion processes, with and without control, see [5, 7, 27] and their bibliographies. We mention also the series of papers [33–35] by Pardoux and Veretennikov on the approximation of diffusions without control from the point of view of Poisson equation, and the contributions by Borkar and Gaitsgory [10, 11] on singularly perturbed stochastic differential equations with control both in the slow and in the fast variables, relying on the Limit Occupational Measure Set. More recent results for uncontrolled SDEs were obtained in [29, 38] under weaker regularity assumptions and in [15] for nonautonomus systems with an application in finance. LQ problems with multiplicative noise were treated in [21]. Some extensions to infinite dimensional control systems were studied in [41] and [22]. Other results were obtained using different techniques from nonlinear filtering theory in [3]. The very recent paper [20] studies the rate of convergence in an unbounded setting.

The paper is organized as follows. In Section 2 we present the two scale stochastic control problem and the assumptions that will hold throughout the paper, together with the associated Hamilton-Jacobi-Bellman equation. Section 3 is devoted to the study of ergodicity properties of the process (1.3), the estimates on τ_n^Y, and the construction of the effective Hamiltonian and initial data and of suitable approximate correctors for the singularly perturbed HJB equation. This is a crucial step for the convergence result of the value function that we next show in Section 4. In this last section, we rely on viscosity methods, with an adaptation of the Evans' perturbed test function method [16] to fit our unbounded context.

2. The two scale stochastic control problem

2.1. The stochastic system

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ be a complete filtered probability space and let $(W_t)_t$ be an $\mathcal{F}_t$-adapted standard r-dimensional Brownian motion. We consider the following stochastic control system

$$\begin{cases} \mathrm{d}X_t = f(X_t, Y_t, u_t)\,\mathrm{d}t + \sqrt{2}\,\sigma^\varepsilon(X_t, Y_t, u_t)\,\mathrm{d}W_t, \quad X_0 = x \in \mathbb{R}^n \\ \mathrm{d}Y_t = \dfrac{1}{\varepsilon}\, b(X_t, Y_t)\,\mathrm{d}t + \sqrt{\dfrac{2}{\varepsilon}}\,\varrho(X_t, Y_t)\,\mathrm{d}W_t, \quad Y_0 = y \in \mathbb{R}^m \end{cases} \tag{2.1}$$

Throughout the paper, we shall make different sets of assumptions that we will present when needed. We start with the following

*Assumptions **(A)***

(A1) For a given compact set U, $f : \mathbb{R}^n \times \mathbb{R}^m \times U \to \mathbb{R}^n$, $\sigma^\varepsilon : \mathbb{R}^n \times \mathbb{R}^m \times U \to \mathbb{M}^{n,r}$, $b : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ and $\varrho : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{M}^{m,r}$ are continuous functions, Lipschitz continuous in (x, y) uniformly with respect to $u \in U$ and $\varepsilon > 0$, and with linear growth in both x and y, that is,

$$|f(x, y, u)|, \|\sigma^\varepsilon(x, y, u)\| \le C(1 + |x| + |y|), \quad \forall\, x, y, \ \forall \varepsilon > 0 \tag{2.2}$$

$$|b(x, y|, \|\varrho(x, y)\| \le C(1 + |x| + |y|), \quad \forall\, x, y, \tag{2.3}$$

for some positive constant C.

(A2) The diffusion σ^ε driving the slow variables X_t satisfies

$$\lim_{\varepsilon \to 0} \sigma^\varepsilon(x, y, u) = \sigma(x, y, u) \quad \text{locally uniformly,}$$

where $\sigma : \mathbb{R}^n \times \mathbb{R}^m \times U \to \mathbb{M}^{n,r}$ satisfies the same conditions as σ^ε. We will not make any nondegeneracy assumption on the matrices $\sigma^\varepsilon, \sigma$, so the cases $\sigma^\varepsilon, \sigma \equiv 0$ are allowed.

(A3) The diffusion ϱ driving the fast variables Y_t is such that $\varrho\varrho^\top$ is uniformly bounded and non degenerate, *i.e.* $\exists\ \underline{\Lambda},\ \overline{\Lambda}\ > 0$, such that $\forall\, x, y, \xi$.

$$\underline{\Lambda}|\xi|^2 \le \varrho(x, y)\varrho^\top(x, y)\xi \cdot \xi = |\varrho(x, y)^\top \xi|^2 \le \overline{\Lambda}|\xi|^2. \tag{2.4}$$

(A4) The following *recurrence condition* holds for the fast variables Y.

$$\forall\, x \in \mathbb{R}^n, \exists\, A_x, R_x > 0 \ \text{ s.t.} \quad b(x, y) \cdot y < -A_x|y|, \quad \forall\, |y| \ge R_x. \tag{2.5}$$

We will simply denote $A = A_x, R = R_x$ when there is no confusion. Note that condition (2.5) is related to the one introduced by Pardoux and Veretennikov in [33] namely, $\lim_{|y|\to\infty} \sup_{x \in \mathbb{R}^n} b(x, y) \cdot y = -\infty$ uniformly in x, and usually called Khasminskii's assumption. It will be strengthened later into assumption (C2) to get better properties of the invariant measure and effective Hamiltonian.

2.2. The optimal control problem

We define the following pay off functional for a finite horizon optimal control problem associated to system (2.1) for $t \in [0, T]$

$$J(t, x, y, u) := \mathbb{E}\left[e^{\lambda(t-T)} g(X_T, Y_T) + \int_t^T \ell(s, X_s, Y_s, u_s) e^{\lambda(s-T)} \mathrm{d}s \,\middle|\, X_t = x,\ Y_t = y \right]. \tag{2.6}$$

The associated value function is

$$V^\varepsilon(t,x,y) := \sup_{u\in\mathcal{U}} J(t,x,y,u), \quad \text{subject to } (2.1). \tag{$OCP(\varepsilon)$}$$

The set of admissible control functions $\mathcal{U}$ is the standard one in stochastic control problems, *i.e.*, the set of $\mathcal{F}_t$-progressively measurable processes taking values in U. We will make the following

Assumptions ***(B)***

(B1) The discount factor is $\lambda \geq 0$.

(B2) The utility function $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and the running cost $\ell : [0,T] \times \mathbb{R}^n \times \mathbb{R}^m \times U \to \mathbb{R}$ are continuous functions and satisfy

$$\exists\, K > 0 \ \text{ s.t. } \ |g(x,y)|, |\ell(s,x,y,u)| \leq K(1+|x|^2+|y|^2), \ \forall s \in [0,T], x, y, u. \tag{2.7}$$

(B3) The running cost ℓ is locally Hölder continuous in y uniformly in u, *i.e.*, for any $R > 0$ and s, x there are constants $\gamma, C > 0$ such that

$$|\ell(s,x,y,u) - \ell(s,x,\tilde{y},u)| \leq C|y-\tilde{y}|^\gamma \quad \forall\, |y| \leq R, u \in U. \tag{2.8}$$

2.3. The HJB equation

The HJB equation associated *via* Dynamic Programming to the value function V^ε is

$$-V_t^\varepsilon + F^\varepsilon\left(t,x,y,V^\varepsilon, D_xV^\varepsilon, \frac{D_yV^\varepsilon}{\varepsilon}, D^2_{xx}V^\varepsilon, \frac{D^2_{yy}V^\varepsilon}{\varepsilon}, \frac{D^2_{x,y}V^\varepsilon}{\sqrt{\varepsilon}}\right) = 0, \ \text{ in } \ (0,T)\times\mathbb{R}^n\times\mathbb{R}^m, \tag{2.9}$$

complemented with the obvious terminal condition

$$V^\varepsilon(T,x,y) = g(x,y). \tag{2.10}$$

This is a fully nonlinear degenerate parabolic equation (strictly parabolic in the y variables by the assumption (2.4)). We denote by $\mathbb{M}^{n,m}$ (respec. $\mathbb{S}^n$) the set of matrices of n rows and m columns (respec. the subset of n-dimensional squared symmetric matrices). The Hamiltonian $F^\varepsilon : [0,T]\times\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}\times\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{S}^n\times\mathbb{S}^m\times\mathbb{M}^{n,m}\to\mathbb{R}$ is defined as

$$F^\varepsilon(t,x,y,r,p,q,M,N,Z) := H^\varepsilon(t,x,y,p,M,Z) - \mathcal{L}(x,y,q,N) + \lambda r, \tag{2.11}$$

where

$$H^\varepsilon(t,x,y,p,M,Z) := \min_{u\in U}\left\{-\text{trace}(\sigma^\varepsilon\sigma^{\varepsilon\top}M) - f\cdot p - 2\text{trace}(\sigma^\varepsilon\varrho^\top Z^\top) - \ell\right\}, \tag{2.12}$$

with σ^ε, f computed at (x,y,u), $\ell = \ell(t,x,y,u)$, and $\varrho = \varrho(x,y)$, and

$$\mathcal{L}(x,y,q,N) := b(x,y)\cdot q + \text{trace}(\varrho(x,y)\varrho^\top(x,y)N) \tag{2.13}$$

We define also the Hamiltonian H as H^ε with σ^ε is replaced by σ

$$H(t,x,y,p,M,Z) := \min_{u\in U}\left\{-\text{trace}(\sigma\sigma^\top M) - f\cdot p - 2\text{trace}(\sigma\varrho^\top Z^\top) - \ell\right\}. \tag{2.14}$$

The next result is standard, see, *e.g.*, [7], Proposition 3.1 or [5], Proposition 2.1.

Proposition 2.1. *Under assumptions (A) and (B1,B2), for any $\varepsilon > 0$, the function V^ε in equation (OCP(ε)) is the unique continuous viscosity solution to the Cauchy problem (2.9)–(2.10) with at most quadratic growth in x and y, i.e., $\exists\, K > 0$ independent by ε such that*

$$|V^\varepsilon(t,x,y)| \le K(1+|x|^2+|y|^2), \quad \forall\, t \in [0,T],\ x \in \mathbb{R}^n,\ y \in \mathbb{R}^m. \tag{2.15}$$

Note that the functions V^ε are locally equibounded but can be unbounded. The unboundedness in y was not allowed in the previous literature on singular perturbations and it is the main difficulty and novelty in this paper.

3. Ergodicity of the fast variables and the effective limit problem

3.1. The invariant measure

Consider the diffusion processes in $\mathbb{R}^m$ obtained by setting $\varepsilon = 1$ in equation (2.1) and freezing $x \in \mathbb{R}^n$

$$\mathrm{d}Y_t = b(x, Y_t)\,\mathrm{d}t + \sqrt{2}\varrho(x, Y_t)\,\mathrm{d}W_t, \quad Y_0 = y \in \mathbb{R}^m \tag{3.1}$$

called *fast subsystem.* If we want to recall the dependence on the parameter x, we denote the process in equation (3.1) as $Y_\cdot^x$. Observe that its infinitesimal generator is $\mathcal{L}_x w := \mathcal{L}(x, y, D_y w, D^2_{yy} w)$ with $\mathcal{L}$ defined by equation (2.13).

Let us recall that a probability measure μ_x on $\mathbb{R}^m$ is an *invariant measure* for the process $Y_\cdot^x$ in (3.1) if

$$\int_{\mathbb{R}^m} \mathbb{E}[f(Y_t)\,|Y_0 = y]\,\mathrm{d}\mu_x(y) = \int_{\mathbb{R}^m} f(y)\,\mathrm{d}\mu_x(y), \quad \forall\, t > 0, \tag{3.2}$$

for all bounded Borel functions f in $\mathbb{R}^m$ (see for example [30]). We recall that an invariant measure is a stationary solution of the Fokker-Planck equation $\mathcal{L}_x^* \mu_x = 0$, where $\mathcal{L}_x^*$ is the adjoint operator to $\mathcal{L}_x$. When it exists and is unique we say that the process $Y_\cdot^x$ is ergodic.

It is well known that the assumption (2.5) on the drift ensures the existence of an invariant measure for equation (3.1), and its uniqueness follows from the non-degeneracy assumption (2.4) on the diffusion ϱ. This is proven for instance in [42] (see also [33–35]). Another proof of existence and uniqueness of the invariant measure is in [7] assuming the existence of a Lyapunov-type function, which is related to the recurrence condition [5].

In this section, we drop the explicit dependence on the frozen x. Instead, we stress the dependence of $Y_\cdot$ on its initial position y by writing

$$\mathrm{d}Y_y(t) = b(Y_y(t))\,\mathrm{d}t + \sqrt{2}\varrho(Y_y(t))\,\mathrm{d}W_t, \quad Y_y(0) = y \in \mathbb{R}^m. \tag{3.3}$$

3.2. Auxiliary results

The first result we need is the following lemma which gives a stronger form of ergodicity of the fast subsystem, that is, the convergence of the probability law of $Y_y(\cdot)$ towards its unique invariant probability measure. We use $\|\mu - \nu\|_{TV}$ for the total variation distance between two probability measures μ, ν defined by

$$\|\mu - \nu\|_{TV} = \sup_{A \in \mathcal{B}} |\mu(A) - \nu(A)|$$

where $\mathcal{B}$ is the class of Borel sets. In particular, $\|\mu\|_{TV} = \int_{\mathbb{R}^m} \mathrm{d}\mu = 1$.

Lemma 3.1. *Under assumptions (A), there exists $C, d, k > 0$ such that*

$$\|\mathbb{P}_{Y_y(t)}(\cdot) - \mu(\cdot)\|_{TV} \leq C(1+|y|^d)(1+t)^{-(1+k)}. \tag{3.4}$$

Moreover, the invariant measure μ has finite moments of any order.

Proof. This is a particular case of the more general result in [42], Theorem 6. Indeed, the main assumption in [42] is

$$\exists\, M_0 \geq 0,\ r \geq 0 \ \text{ s.t.} \quad b(y)\cdot y \leq -r, \quad \forall\, |y| \geq M_0. \tag{3.5}$$

Then, for the constants

$$\tilde{\Lambda} := \sup_y \operatorname{trace}(\varrho\varrho^\top(y))/m, \quad r_0 := [r - (m\tilde{\Lambda} - \underline{\Lambda})/2]\overline{\Lambda}^{-1},$$

Theorem 6 in [42] states that equation (3.4) holds $\forall\, k \in (0, r_0 - \frac{3}{2})$, $\forall\, d \in (2k+2, 2r_0 - 1)$ if $r_0 > \frac{3}{2}$. In our case, assumption (2.5) guarantees a constant r, and therefore r_0, as large as we want.

For the finite moments, see [42], Equation (28) in Section 6, where it is shown that the invariant measure has finite moments of order $d \in (2k+2, 2r_0-1)$ if $k \in (0, r_0 - \frac{3}{2})$. It is enough to use Hölder inequality together with the fact that $\mu(\mathbb{R}^m) = 1$ to prove finite moments of any order $d \geq 1$. □

The following result gives an estimate on the first exit time of $Y_y(\cdot)$ from the ball centered in 0 with radius n

$$\tau_n^Y := \inf\{t \geq 0 \mid \|Y_y(t)\| \geq n\}.$$

It will be needed together with the previous Lemma for constructing the limit PDE in the next section.

Lemma 3.2. *Under assumptions (A), for any compact set $\mathcal{K}$, there exist η, C positive constants and ℓ a positive function such that, for any $\delta \in (0,1)$ and for n large enough,*

$$\mathbb{E}\left[e^{-\delta\tau_n^Y}\right] \leq C\frac{\ell(\delta)}{\delta}e^{-n\eta}, \quad \forall\, y \in \mathcal{K}, \tag{3.6}$$

where $\ell(\delta) = 1 + O(\delta)$ when $\delta \to 0^+$. In particular for any $\alpha \geq 0$ and $\beta > 0$, one has

$$\mathbb{E}\left[n^\alpha e^{-\frac{1}{n^\beta}\tau_n^Y}\right] \leq Cn^{\alpha+\beta}e^{-n\eta} \longrightarrow 0 \text{ as } n \to +\infty. \tag{3.7}$$

Proof. The idea of the proof is to build a process $Z_t \in \mathbb{R}$ such that $\|Y(t)\| \leq Z_t$ a.s.. Then one has $\tau_n^Y \geq \tau_n^Z$ a.s., where $\tau_n^Z := \inf\{t \geq 0 \mid |Z(t)| \geq n\}$, and hence

$$\mathbb{E}\left[e^{-\delta\tau_n^Y}\right] \leq \mathbb{E}\left[e^{-\delta\tau_n^Z}\right], \quad \forall\, \delta > 0. \tag{3.8}$$

Once we will have such a process Z, we'll give an upper bound of the right hand side in equation (3.8).

The construction of Z such that $\|Y_t\| \leq Z_t$ a.s. is inspired by the proof of [23], Proposition 1.4. Let A and R be the positive constants in the recurrence condition (2.5) and note that R can be chosen as large as we want. Define $h : \mathbb{R}^m \to \mathbb{R}$ as a C^2 function such that $h(y) = \|y\|$ when $\|y\| \geq R$, and $h(y) < R$ otherwise. Next define

$$Z_t := R \vee \|y_o\| + \sqrt{2}M_t - \eta\xi_t + L_t, \tag{3.9}$$

where $Y_0 = y_o$, η is a positive constant to be made precise,

$$M_t := \int_0^t \nabla h(Y_s)^\top \varrho(Y_s) dW_s, \quad t \geq 0,$$

$$\xi_t := \int_0^t \|\nabla h(Y_s)^\top \varrho(Y_s)\|^2 ds$$

is the quadratic variation of the continuous local martingale M_t, and L_t is an increasing process (of finite variation) which increases only at times t for which $Z_t = R$, and is of zero value when $Z > R$. Such pair (Z, L) is the unique pair of continuous adapted process given by Skorokhod's lemma (see *e.g.* [37], Chap. VI, Sect. 2): Z is a process reflected out of the interval $]-R, R[$ and L its compensator. Note that when $\|y\| \geq R$, $\nabla h(y) = \frac{y}{\|y\|}$ so $\left\|\nabla h(y)^\top \varrho(y)\right\|^2 = \frac{1}{\|y\|^2} y^\top \varrho(y) \varrho(y)^\top y \leq \overline{\Lambda}$ by equation (2.4), and hence $d\xi_t \leq \overline{\Lambda}\, dt$ on $\{\|Y_t\| \geq R\}$. On the other hand, define $\tilde{K} := \sup_{\|y\| \leq R} \|\nabla h(y)\|^2$. Then we have $\xi_t \leq (1 \vee \tilde{K})\, \overline{\Lambda}\, t$ for all $t \geq 0$. We set $K := (1 \vee \tilde{K}) \overline{\Lambda}$, and get

$$0 \leq \xi_t \leq Kt, \quad \forall\, t \geq 0. \tag{3.10}$$

Now we choose $f \in C^2(\mathbb{R})$ such that

$$\begin{aligned} &f(x) > 0 \quad \text{and} \quad f'(x) > 0, && \forall\, x > 0 \\ &f(x) = 0, && \forall\, x \leq 0 \end{aligned}$$

We set $a(y) := \varrho(y) \varrho(y)^\top$. According to Itô's formula, for $t \geq 0$,

$$\begin{aligned} \mathrm{d}h(Y_s) &= \left(\nabla h(Y_s)^\top b(Y_s) + \operatorname{trace}\left(a(Y_s) D^2 h(Y_s)\right)\right) \mathrm{d}s + \sqrt{2} \nabla h(Y_s)^\top \varrho(Y_s) \mathrm{d}W_s \\ \mathrm{d}Z_s &= -\eta d\xi_s + \mathrm{d}L_s + \sqrt{2} \mathrm{d}M_s \\ &= -\eta \left\|\nabla h(Y_s)^\top \varrho(Y_s)\right\|^2 \mathrm{d}s + \mathrm{d}L_s + \sqrt{2} \nabla h(Y_s)^\top \varrho(Y_s) \mathrm{d}W_s \end{aligned}$$

so that

$$\mathrm{d}\left(h(Y) - Z\right)_s = \left(\nabla h(Y_s)^\top b(Y_s) + \operatorname{trace}\left(a(Y_s) D^2 h(Y_s)\right) + \eta \left\|\nabla h(Y_s)^\top \varrho(Y_s)\right\|^2\right) \mathrm{d}s - \mathrm{d}L_s.$$

Again by Itô's formula we obtain

$$\begin{aligned} f(h(Y_t) - Z_t) =& f(h(y_o) - R \vee \|y_o\|) + \int_0^t f'(h(Y_s) - Z_s) \mathrm{d}(h(Y) - Z)_s \\ &+ \frac{1}{2} \int_0^t f''(h(Y_s) - Z_s) \mathrm{d}\langle h(Y) - Z \rangle_s, \end{aligned}$$

where $\langle \zeta \rangle_t = \int_0^t \sigma(\zeta_s) \sigma^\top(\zeta_s) \mathrm{d}s$ denotes the quadratic variation of a process defined by $\mathrm{d}\zeta_t = f(\zeta_t)\mathrm{d}t + \sigma(\zeta_t)\mathrm{d}W_t$.

Note that $f(h(y_o) - R \vee \|y_o\|) = 0$ by definition of h and f. Moreover, $h(Y_\cdot) - Z_\cdot$ is a continuous process with no Wiener process term, and hence it has zero quadratic variation, *i.e.*, $\mathrm{d}\langle h(Y) - Z \rangle_s = 0$. Now, again by definition of h and Z, we have $h(Y_t) \leq Z_t$ on $\{\|Y_t\| \leq R\}$, so $\{h(Y_t) > Z_t\} = \{\|Y_t\| > Z_t\}$ is a subset of

$\{\|Y_t\| > R\}$. When $\|y\| \geq R$, we have $\nabla h(y) = \frac{y}{\|y\|}$, $D^2 h(y) = \frac{1}{\|y\|}\left(\mathbb{I}_m - \frac{y \otimes y}{\|y\|^2}\right)$, and we compute, by equation (2.4),

$$\operatorname{trace}\big(a(y) D^2 h(y)\big) = \frac{1}{\|y\|}\left(\operatorname{trace} a(y) - \sum_{i,j=1}^m a_{ij}(y)\frac{y_i y_j}{\|y\|^2}\right) \leq \frac{m\overline{\Lambda}}{\|y\|}. \tag{3.11}$$

Hence the expression

$$\int_0^t f'(\|Y_s\| - Z_s)\left\{\frac{1}{\|Y_s\|}Y_s \cdot b(Y_s) + \frac{m}{\|Y_s\|}\overline{\Lambda} + \eta\overline{\Lambda}\right\}\mathrm{d}s - \int_0^t f'(\|Y_s\| - Z_s)\mathrm{d}L_s,$$

which is valid for $\|Y_s\| > R$, is an upper bound of $f(h(Y_t) - Z_t)$. Furthermore, $\mathrm{d}L_s = 0$ for $\|Y_s\| > R$, and therefore one has

$$f(h(Y_t) - Z_t) \leq \int_0^t f'(\|Y_s\| - Z_s)\left\{\frac{1}{\|Y_s\|}Y_s \cdot b(Y_s) + \frac{m}{\|Y_s\|}\overline{\Lambda} + \eta\overline{\Lambda}\right\}\mathrm{d}s \tag{3.12}$$

By the recurrence condition (2.5) the quantity in brackets $\{...\}$ is bounded from above by

$$-A + \frac{m}{\|Y_s\|}\overline{\Lambda} + \eta\overline{\Lambda} \leq -A + \frac{m}{R}\overline{\Lambda} + \eta\overline{\Lambda}$$

because this upper bound is obtained for $\|Y_s\| > R$. Now we choose R large enough so that $A/\overline{\Lambda} > m/R$. Then for

$$0 < \eta < \frac{A}{\overline{\Lambda}} - \frac{m}{R}$$

the r.h.s. of equation (3.12) is negative, which ensures $f(h(Y_t) - Z_t) \leq 0$ and implies $\|Y_t\| \leq Z_t$ a.s. by definition of f.

Next we look for an upper bound to $\mathbb{E}\left[e^{-\delta \tau_n^Z}\right]$. For simplicity of notation, in this step we shall write $\tau_n := \tau_n^Z$, dropping the dependence on Z. Fix $\delta \in (0,1)$ and set $\gamma := \frac{\delta}{K}$ where K is the constant in equation (3.10). By Itô's formula, for any $\Phi \in C^2(\mathbb{R})$,

$$\begin{aligned}\mathrm{d}(\Phi(Z_t)e^{-\gamma\xi_t}) = \sqrt{2}\Phi'(Z_t)e^{-\gamma\xi_t}\mathrm{d}M_t + \Phi'(Z_t)e^{-\gamma\xi_t}\mathrm{d}L_t \\ + e^{-\gamma\xi_t}\left\{\Phi''(Z_t) - \eta\Phi'(Z_t) - \gamma\Phi(Z_t)\right\}\mathrm{d}\xi_t.\end{aligned}$$

Since we are interested in the limit as $n \to \infty$, we can assume without loss of generality that $n > R$. We choose Φ such that

$$\begin{cases} \Phi''(z) - \eta\Phi'(z) - \gamma\Phi(z) = 0, & \text{for } z \in [R, n] \\ \Phi'(R) = 0 \quad \text{and} \quad \Phi(n) = 1 \end{cases} \tag{3.13}$$

then $\Phi(Z_t)e^{-\gamma\xi_t}$ is a local martingale which is bounded up to time τ_n. Hence we are allowed to apply Doob's stopping theorem to obtain

$$\Phi(R \vee \|y_o\|) = \mathbb{E}\left[\Phi(Z_{\tau_n})e^{-\gamma\xi_{\tau_n}}\right] \tag{3.14}$$

and since $Z_{\tau_n} = n$, $\Phi(n) = 1$, and $\xi_t \le Kt$ for all $t \ge 0$, we have

$$\mathbb{E}\left[e^{-\gamma K \tau_n}\right] \le \mathbb{E}\left[e^{-\gamma \xi_{\tau_n}}\right] = \Phi(R \vee \|y_o\|) \tag{3.15}$$

which yields

$$\mathbb{E}\left[e^{-\delta \tau_n}\right] \le \Phi(R \vee \|y_o\|) \tag{3.16}$$

Now solving the differential equation (3.13) yields

$$\Phi(z) = \frac{-\lambda_2 e^{\lambda_1 (z-R)} + \lambda_1 e^{\lambda_2 (z-R)}}{-\lambda_2 e^{\lambda_1 (n-R)} + \lambda_1 e^{\lambda_2 (n-R)}}, \tag{3.17}$$

where $\lambda_2 < 0 < \lambda_1$ are given by

$$\lambda_1 = \frac{1}{2}\left(\eta + \sqrt{\eta^2 + 4\gamma}\right), \qquad \lambda_2 = \frac{1}{2}\left(\eta - \sqrt{\eta^2 + 4\gamma}\right).$$

Hence,

$$\begin{aligned}\Phi(z) &\le \frac{(\lambda_1 - \lambda_2) e^{\lambda_1 (z-R)}}{-\lambda_2 e^{\lambda_1 (n-R)}} \\ &\le 2\frac{\sqrt{1+\frac{4\gamma}{\eta^2}}}{\sqrt{1+\frac{4\gamma}{\eta^2}} - 1} \exp\left[z\,\frac{\eta}{2}\left(1 + \sqrt{1+\frac{4\gamma}{\eta^2}}\right)\right] \exp\left[-n\frac{\eta}{2}\left(1 + \sqrt{1+\frac{4\gamma}{\eta^2}}\right)\right]\end{aligned}$$

By Taylor expansion, for γ small,

$$1 + 2\frac{\gamma}{\eta^2} - 2\frac{\gamma^2}{\eta^4} \le \sqrt{1 + \frac{4\gamma}{\eta^2}} \le 1 + 2\frac{\gamma}{\eta^2},$$

which yields

$$\Phi(z) \le \frac{1 + 2\frac{\gamma}{\eta^2}}{\frac{\gamma}{\eta^2} - \frac{\gamma^2}{\eta^4}} \exp\left[z\,\eta\left(1 + \frac{\gamma}{\eta^2}\right)\right] e^{-n\frac{\eta}{2}}.$$

Now recall that $\gamma := \frac{\delta}{K}$ and define

$$\ell(\delta) := \frac{1 + \frac{2\delta}{K\eta^2}}{1 - \frac{\delta}{K\eta^2}} \exp\left[z\frac{\delta}{K\eta}\right] \quad \text{and} \quad C := \eta^2 K e^{z\eta}.$$

Then the right-hand side in the last inequality equals $C\frac{\ell(\delta)}{\delta}e^{-n\frac{\eta}{2}}$ and $\ell(\delta) = 1 + O(\delta)$ when $\delta \to 0^+$. Together with equation (3.16), for $z := R \vee \|y_o\|$ this yields

$$\mathbb{E}\left[e^{-\delta \tau_n^z}\right] \le C\frac{\ell(\delta)}{\delta} e^{-n\frac{\eta}{2}}.$$

By combining this inequality with equation (3.8) we finally get the desired estimate (3.6) and conclude the proof of the first statement.

The second statement of the lemma is immediately obtained by multiplying the inequality (3.6) by n^α for $\alpha \geq 0$ and choosing $\delta = n^{-\beta}$ for $\beta > 0$. □

Remark 3.3. Under assumptions (A), we can also prove that, for suitable C_1, C_2 and $\kappa > 0$,

$$C_2\,(n^2 - |y|^2) \leq \mathbb{E}[\,\tau_n^Y\,] \leq C_1\, e^{\kappa\, n^2} \quad \text{locally uniformly in } y. \tag{3.18}$$

3.3. The effective Hamiltonian and approximate correctors

We expect that the effective Hamiltonian in the limit HJB equation of the singular perturbation problem is

$$\overline{H}(t,x,p,P) := \int_{\mathbb{R}^m} H(t,x,y,p,P,0)d\mu_x(y) \tag{3.19}$$

where μ_x is the invariant measure of the process (3.1) introduced in Section 3.1 and studied in Section 3.2. In classical periodic homogenization theory one proves the convergence by means of a corrector (see [1, 16, 28]), namely, a (periodic) solution χ (for fixed (t,x,p,P)) of the *cell problem*

$$-\mathcal{L}(x,y,D\chi,D^2\chi) + H(t,x,y,p,P,0) = \overline{H} \quad \text{in } \mathbb{R}^m,$$

where $\mathcal{L}$ and H are defined in equations (2.13) and (2.14). In many cases, however, the cell problem may be hard or impossible to solve, and then one resorts to *approximate correctors*, *i.e.*, a sequence χ_n such that

$$-\mathcal{L}(x,y,D\chi_n,D^2\chi_n) + H(t,x,y,p,P,0) \to \overline{H} \quad \text{locally uniformly},$$

see, *e.g.*, [2, 5]. Here the unboundedness of both the domain and the Hamiltonian does not allow us to build the approximate correctors globally. We overcome the problem by introducing a suitable *truncated δ-cell problem* that we now describe. Fix (t,x,p,P), and let us denote for simplicity

$$\mathcal{L}\omega(y) := \mathcal{L}(x,y,D\omega,D^2\omega), \tag{3.20}$$

$$h(y) := H(t,x,y,p,P,0) \quad \text{in } \mathbb{R}^m. \tag{3.21}$$

Note that h is locally Hölder continuous by the assumptions (A) and (B).

Take a sequence of bounded and open domains D_n such that $\overline{D}_n \subset D_{n+1}$ and $\cup_n D_n = \mathbb{R}^m$. Assume in addition that ∂D_n is C^2 and $D_n \subseteq B(0,n) := \{y \in \mathbb{R}^m \mid \|y\| < n\}$, the open ball centered in 0 with radius n (*e.g.*, $D_n = B(0,n)$). Consider the Dirichlet-Poisson problem

$$\begin{cases} \delta u(y) - \mathcal{L}u(y) = -h(y), \text{ in } D_n, \\ u(y) = 0, \text{ on } \partial D_n. \end{cases} \tag{3.22}$$

It has a unique solution $u^{\delta,n}(\cdot)$ (see, *e.g.*, [32], Thm. 8.1, p. 79) given by

$$u^{\delta,n}(y) = \mathbb{E}\left[-\int_0^{\tau_n} h(Y_y(t))e^{-\delta t}\mathrm{d}t\right] \tag{3.23}$$

where τ_n is the first exist time of $Y_y(\cdot)$ from D_n. In the next result we study the limit as $\delta \to 0$ and $n \to \infty$. We will use that, under the assumptions (2.2) and (2.7), the Hamiltonian has at most a quadratic growth in y, *i.e.*

$$\exists\, K_h > 0, \quad |h(y)| \leq K_h(1+|y|^2), \quad \forall\, y \in \mathbb{R}^m \tag{3.24}$$

where K_h is a constant that depends on the slow dynamics data (f,σ) and the running cost ℓ.

Proposition 3.4. *Let $u^{\delta,n}(\cdot)$ be the solution to equation (3.22). Under assumptions (A) and (B), for any $\alpha>0$ and $\delta=\delta(n)=O\left(\frac{1}{n^{4+\alpha}}\right)$ as $n\to\infty$,*

$$\lim_{n\to\infty}\left|\delta(n)u^{\delta(n),n}(y)+\mu(h)\right|=0,\qquad \textit{locally uniformly in } y, \tag{3.25}$$

where $\mu(h)=\int_{\mathbb{R}^m}h(y)d\mu(y)=\overline{H}(t,x,p,P)$ and μ is the unique invariant probability measure for the process (3.1).

Proof. From equation (3.23), for $D_n^c:=\mathbb{R}^m\setminus D_n$, we have

$$\begin{aligned}
&u^{\delta,n}(y)+\frac{\mu(h)}{\delta}\\
&=\mathbb{E}\left[-\int_0^{\tau_n}h(Y_y(t))e^{-\delta t}\mathrm{d}t\right]+\int_0^\infty\int_{\mathbb{R}^m}h(y)e^{-\delta t}d\mu(y)\mathrm{d}t\\
&=\mathbb{E}\left[-\int_0^{\infty}\mathbb{1}_{D_n}(Y_y(t))h(Y_y(t))e^{-\delta t}\mathrm{d}t\right]+\int_0^\infty\int_{\mathbb{R}^m}h(y)e^{-\delta t}d\mu(y)\mathrm{d}t+\mathbb{E}\left[\int_{\tau_n}^{\infty}\mathbb{1}_{D_n}(Y_y(t))h(Y_y(t))e^{-\delta t}\mathrm{d}t\right]\\
&=\int_0^\infty\int_{D_n}h(y)d\Big(\mu(y)-\mathbb{P}_{Y_y(t)}(y)\Big)e^{-\delta t}\mathrm{d}t+\frac{1}{\delta}\int_{D_n^c}h(y)d\mu(y)+\mathbb{E}\left[\int_{\tau_n}^{\infty}\mathbb{1}_{D_n}(Y_y(t))h(Y_y(t))e^{-\delta t}\mathrm{d}t\right].
\end{aligned}$$

To estimate the first term we apply first Hölder inequality to get

$$\begin{aligned}
\left|\int_0^\infty\int_{D_n}h(y)\mathrm{d}\Big(\mu(y)-\mathbb{P}_{Y_y(t)}(y)\Big)e^{-\delta t}\mathrm{d}t\right|&\le\left(\int_0^\infty\left(\int_{D_n}h(y)\mathrm{d}(\mu(y)-\mathbb{P}_{Y_y(t)}(y))\right)^2\mathrm{d}t\right)^{1/2}\left(\int_0^\infty e^{-2\delta t}\mathrm{d}t\right)^{1/2}\\
&=\frac{1}{\sqrt{2\delta}}\left(\int_0^\infty\left(\int_{D_n}h(y)\mathrm{d}(\mu(y)-\mathbb{P}_{Y_y(t)}(y))\right)^2\mathrm{d}t\right)^{1/2}
\end{aligned}$$

Now we can bound the term in the r.h.s. by Lemma 3.1 and equation (3.24) as follows

$$\begin{aligned}
\int_0^\infty\left(\int_{D_n}h(y)\mathrm{d}(\mu(y)-\mathbb{P}_{Y_y(t)}(y))\right)^2\mathrm{d}t&\le\int_0^\infty\left(\sup_{D_n}|h|\int_{D_n}\mathrm{d}(\mu(y)-\mathbb{P}_{Y_y(t)}(y))\right)^2\mathrm{d}t\\
&\le\sup_{D_n}|h|^2\int_0^\infty\left\|\mathbb{P}_{Y_y(t)}(\cdot)-\mu(\cdot)\right\|_{TV}^2\mathrm{d}t\\
&\le\frac{C^2(1+|y|^d)^2}{1+2k}\sup_{D_n}|h|^2\le\frac{C^2(1+|y|^d)^2}{1+2k}K_h^2(1+n^2)^2.
\end{aligned}$$

Finally, we have the following upper bound

$$\left|\int_0^\infty\int_{D_n}h(y)\mathrm{d}\big(\mu(y)-\mathbb{P}_{Y_y(t)}(y)\big)\,e^{-\delta t}\mathrm{d}t\right|\le K_h\frac{C(1+|y|^d)}{1+2k}\frac{(1+n^2)}{\sqrt{2\delta}}. \tag{3.26}$$

We rewrite the second term as

$$\frac{1}{\delta}\int_{D_n^c}h(y)\mathrm{d}\mu(y)=\frac{1}{\delta}\left(\mu(h)-\int_{D_n}h(y)\mathrm{d}\mu(y)\right) \tag{3.27}$$

We bound the third term using the definition of D_n and equation (3.24)

$$\begin{aligned}\left|\mathbb{E}\left[\int_{\tau_n}^{\infty}\mathbb{1}_{D_n}(Y_y(t))h(Y_y(t))e^{-\delta t}\mathrm{d}t\right]\right| &\le K_h\mathbb{E}\left[\int_{\tau_n}^{\infty}\mathbb{1}_{D_n}(Y_y(t))(1+|Y_y(t)|^2)e^{-\delta t}\mathrm{d}t\right]\\ &\le K_h\mathbb{E}\left[\int_{\tau_n}^{\infty}(1+n^2)e^{-\delta t}\mathrm{d}t\right]\le K_h\frac{1+n^2}{\delta}\mathbb{E}\left[e^{-\delta\tau_n}\right]\end{aligned} \tag{3.28}$$

Now we add up equations (3.26), (3.27), and (3.28), and multiply by δ to get

$$\left|\delta u^{\delta,n}(y)+\mu(h)\right|\le K_h\sqrt{\delta}\,\frac{C(1+|y|^d)(1+n^2)}{(1+2k)\sqrt{2}}+\left|\mu(h)-\int_{D_n}h(y)\mathrm{d}\mu(y)\right|+K_h(1+n^2)\mathbb{E}\left[e^{-\delta\tau_n}\right].$$

If we set $\delta=\delta(n)=O(\frac{1}{n^{4+\alpha}})$ with $\alpha>0$, the last term converges to zero as $n\to\infty$ by Lemma 3.2, and then

$$\lim_{n\to\infty}\left|\delta(n)u^{\delta(n),n}(y)+\mu(h)\right|=0.$$

□

Remark 3.5. This result still holds true if we relax the growth condition (3.24) on h to

$$\exists\, K_h>0,\quad |h(y)|\le K_h(1+|y|^{\gamma}),\quad \forall\, y\in\mathbb{R}^m,$$

with any $\gamma\ge 0$, provided we set $\delta=O\left(\frac{1}{n^{2\gamma+\alpha}}\right)$ in Proposition 3.4. This means that the slow dynamics is allowed to have a polynomial growth w.r.t. the fast variables. The same result holds also if $u^{\delta,n}(\cdot)$ satisfies a inhomogeneous boundary condition $u(y)=\phi(y)$ on ∂D_n in the Dirichlet problem (3.22), if ϕ has a polynomial growth, that is, $\exists\, K_\phi>0$ and $\kappa\ge 0$ such that $|\phi(y)|\le K_\phi(1+|y|^{\kappa})$. The proof requires only minor modifications, see [26].

The next result is an *exchange property* which allows the effective Hamiltonian $\overline{H}$ to be of Bellman type. Such representation will be useful for applying a comparison theorem in the conclusion of our main result.

Proposition 3.6. *Under assumptions (A) and (B2), the effective Hamiltonian equation* (3.19) *can be written as*

$$\overline{H}(t,x,p,P)=\min_{\nu\in L^{\infty}(\mathbb{R}^m,U)}\int_{\mathbb{R}^m}\left[-trace(\sigma\sigma^{\top}P)-f\cdot p-\ell\right]\,d\mu_x(y) \tag{3.29}$$

where σ,f are computed in $(x,y,\nu(y))$ and ℓ in $(t,x,y,\nu(y))$.

Note that $L^{\infty}(\mathbb{R}^m,U)=L^1((\mathbb{R}^m,\mu_x),U)$ because U is bounded and μ_x is a finite measure.

Proof. Let t,x,p,P be fixed and define

$$F(y,u):=-\mathrm{trace}(\sigma(x,y,u)\sigma(x,y,u)^{\top}P)-f(x,y,u)\cdot p-\ell(t,x,y,u),$$

so that $H(t,x,y,p,P,0)=\min_{u\in U}F(y,u)$. To prove the inequality "$\le$", it suffices to observe that for any $\varepsilon>0$, there exists $\nu^{\varepsilon}\in L^{\infty}(\mathbb{R}^m,U)$ such that

$$\begin{aligned}\inf_{\nu\in L^{\infty}(\mathbb{R}^m,U)}\int_{\mathbb{R}^m}F(y,\nu(y))\,\mathrm{d}\mu_x(y)+\varepsilon\ &\ge\int_{\mathbb{R}^m}F(y,\nu^{\varepsilon}(y))\,\mathrm{d}\mu_x(y)\\ &\ge\int_{\mathbb{R}^m}\min_{u\in U}F(x,u)\,\mathrm{d}\mu_x(y)=\overline{H}\end{aligned} \tag{3.30}$$

and hence the result by the arbitrariness of ε.

To prove the inequality "$\geq$", we consider the minimization problem

$$\mathfrak{F}(y) := \min_{u \in U} F(y,u)$$

where $y \in \mathbb{R}^m$. Since F is continuous, U is compact, $\mathfrak{F}(y) \in F(\{y\} \times U)$, and $\mathfrak{F}$ is continuous, a classical selection theorem (see [24], Thm. 7.1, p. 66) implies the existence of a measurable selector $\overline{\nu}$ for which the minimization is achieved, *i.e.*,

$$\exists\, \overline{\nu} \in L^\infty(\mathbb{R}^m, U), \text{ s.t. } \forall\, y \in \mathbb{R}^m,\ \mathfrak{F}(y) = \min_{u \in U} F(y,u) = F(y, \overline{\nu}(y)).$$

Therefore one has

$$\begin{aligned} \overline{H} = \int_{\mathbb{R}^m} \min_{u \in U} F(y,u)\,\mathrm{d}\mu_x(y) &= \int_{\mathbb{R}^m} F(y, \overline{\nu}(y))\,\mathrm{d}\mu_x(y)) \\ &\geq \inf_{\nu \in L^\infty(\mathbb{R}^m,U)} \int_{\mathbb{R}^m} F(y, \nu(\cdot))\,\mathrm{d}\mu_x(y). \end{aligned}$$

This inequality together with equation (3.30) proves that the inf is a min, attained at $\nu = \overline{\nu}$, and the equality (3.29) holds. □

3.4. The effective initial data

In this section, we construct the effective terminal cost $\overline{g}(x)$ for the limit of the singular perturbations problem (2.9)-(2.10). We expect that it is

$$\overline{g}(x) := \int_{\mathbb{R}^m} g(x,y) d\mu_x(y) \tag{3.31}$$

where μ_x is the invariant measure of the process (3.1). In classical homogenization theory one uses that

$$\overline{g}(x) = \lim_{t \to +\infty} \omega(t, y; x)$$

where ω solves, for fixed x, the initial value problem:

$$\begin{cases} \omega_t - \mathcal{L}(x, y, D\omega, D^2\omega) = 0 & \text{in } (0, +\infty) \times \mathbb{R}^m, \\ \omega(0, y) = g(x, y), & \text{in } \mathbb{R}^m, \end{cases} \tag{3.32}$$

with $\mathcal{L}$ defined in equation (2.13), see, *e.g.*, [2, 5]. In our context of unbounded data we use a truncation to bounded domains of such a problem, similar to the previous section. We consider an increasing sequence of bounded and open domains D_n with C^2 boundaries invading $\mathbb{R}^m$ and such that $D_n \subseteq B(0,n)$, as in Section 3.3 (for example $D_n = B(0,n)$). Now instead of equation (3.32), we consider the Cauchy-Dirichlet problem

$$\begin{cases} \dfrac{\partial}{\partial t}\omega^{T,n} - \mathcal{L}(x, y, D\omega^{T,n}, D^2\omega^{T,n}) = 0, & \text{in } (0,T] \times D_n, \\ \omega^{T,n}(0, y) = g(x, y), & \text{in } D_n, \\ \omega^{T,n}(t, y) = 0, & \text{in } [0,T] \times \partial D_n, \end{cases} \tag{3.33}$$

where x is again a fixed parameter, and if we set $u^{T,n}(t,y) = \omega^{T,n}(T-t,y)$, then $u^{T,n}(\cdot,\cdot)$ solves the terminal-boundary value problem

$$\begin{cases} \dfrac{\partial}{\partial t} u^{T,n} + \mathcal{L}(x,y,Du^{T,n},D^2u^{T,n}) = 0, & \text{in } [0,T)\times D_n, \\ u^{T,n}(T,y) = g(x,y), & \text{in } D_n, \\ u^{t,n}(t,y) = 0, & \text{in } [0,T]\times \partial D_n, \end{cases} \tag{3.34}$$

It is known [32], Theorem 8.2, p. 81, that the problem (3.34) admits a unique solution given by

$$u^{T,n}(t,y) = \mathbb{E}[\,\mathbb{1}_{\{\tau_n\wedge T=T\}}\, g(x,Y_{y,t}(T))\,] \tag{3.35}$$

where $Y_{y,t}(\cdot)$ is the fast process defined by equation (3.1) and such that $Y_{y,t}(t) = y \in \mathbb{R}^m$, and $\tau_n = \inf\{s \in [t,T] : Y_{y,t}(s) \notin D_n\}$ is the first exit time from D_n. The next result gives an approximation of the effective initial data $\overline{g}(x)$ by $\omega^{T,n}(T,y) = u^{T,n}(0,y)$ as $T = T(n) \to +\infty$ for $n \to +\infty$.

Proposition 3.7. *Let $u^{T,n}(\cdot,\cdot)$ be as defined in equation (3.35). Under assumptions (A) and (B), for any increasing sequence $\{T(n)\}_{n>0}$ such that $T(n) \geq n^2$, we have the following*

$$\lim_{n\to+\infty} \left| u^{T(n),n}(0,y) \,-\, \overline{g} \right| = 0, \quad \textit{locally uniformly in } y, \tag{3.36}$$

where $\overline{g} = \overline{g}(x) = \int_{\mathbb{R}^m} g(x,y)\,d\mu_x(y)$ and μ_x is the unique invariant probability measure of the process (3.1). In particular $\lim\limits_{n\to+\infty} \omega^{T(n),n}(T(n),y) = \overline{g}$ locally uniformly in y and $\overline{g}$ has at most quadratic growth in x.

Proof. Since the slow variable x is frozen we drop it in the notations and write in particular $g(x,\cdot) = g(\cdot)$ and $\mu_x(\cdot) = \mu(\cdot)$. Also, the fast process $Y_{y,0}(\cdot)$ will be simply denoted by $Y_y(\cdot)$. We have the following

$$u^{T(n),n}(0,y) = \int_{D_n} \mathbb{1}_{\{\tau_n\wedge T(n)=T(n)\}} g(z)\,\mathrm{d}\mathbb{P}_{Y_y(T(n))}(z), \quad \text{from equation (3.35)}$$
$$\overline{g} = \int_{\mathbb{R}^m} g(z)\,\mathrm{d}\mu(z) = \int_{D_n} g(z)\,\mathrm{d}\mu(z) + \int_{D_n^c} g(z)\,\mathrm{d}\mu(z).$$

Hence

$$\begin{aligned} \left| u^{T(n),n}(0,y) \,-\, \overline{g} \right| &\leq \left| \int_{D_n} \mathbb{1}_{\{\tau_n\wedge T(n)=T(n)\}} g(z)\,\mathrm{d}\big(\mathbb{P}_{Y_y(T(n))} - \mu\big)\,(z) \right| + \left| \int_{D_n^c} g(z)\,\mathrm{d}\mu(z) \right| \\ &\leq C(1+n^2)\|\mathbb{P}_{Y_y(T(n))}(\cdot) - \mu(\cdot)\|_{TV} + \sqrt{\mu(g^2)}\sqrt{1-\mu(D_n)} \end{aligned}$$

where, for the first integral we used the quadratic growth of g from equation (2.7), and for the second, Hölder inequality together with the fact that the probability measure μ has finite fourth moment by Lemma 3.1. Now, again by Lemma 3.1, there exist $C, d, k > 0$ such that

$$\|\mathbb{P}_{Y_y(T(n))}(\cdot) - \mu(\cdot)\|_{TV} \leq C(1+|y|^d)(1+T(n))^{-(1+k)}$$

Therefore, by choosing $T(n) \geq n^2$ we obtain, as $n \to \infty$,

$$\left| u^{T(n),n}(0,y) \,-\, \mu(g) \right| \leq C(1+n^2)(1+|y|^d)\frac{1}{(1+n^2)^{1+k}} \,+\, \sqrt{\mu(g^2)(1-\mu(D_n))} \,\to\, 0.$$

Finally, the growth condition on $\overline{g}$ follows from equation (2.7) and the fact that μ has a finite second order moment (Lem. 3.1). □

Remark 3.8. This result still holds true if we consider, instead of the growth assumption (2.7), g such that

$$\exists\, K_g > 0, \quad |g(x,y)| \leq K_g(1+|x|^2+|y|^\gamma), \quad \forall\, y \in \mathbb{R}^m$$

where $\gamma \geq 0$ is as large as we want, provided we choose $T(n) \geq n^\gamma$.

4. The convergence theorem for the value function

We can now state and prove the main result of the paper, namely the convergence as $\varepsilon \to 0$ of the value function $V^\varepsilon(t,x,y)$, solution to equations (2.9)–(2.10), to a function $V(t,x)$ characterised as the unique solution of the Cauchy problem

$$\begin{cases} -V_t + \overline{H}(t,x,D_xV,D^2_{xx}V) + \lambda V(x) = 0, & \text{in } (0,T)\times\mathbb{R}^n, \\ V(T,x) = \overline{g}(x), & \text{in } \mathbb{R}^n, \end{cases} \tag{4.1}$$

where the effective Hamiltonian $\overline{H}$ and the effective initial data $\overline{g}(x)$ are defined by equations (3.19) and (3.31), respectively.

Before we go further, we need to check smoothness in the x variables of the data in the effective (limit) Cauchy problem. Indeed, the construction of $\overline{H}, \overline{g}$ in the previous section involves the invariant measure of the fast process Y which depends on x.

4.1. On the effective Cauchy problem

This subsection is devoted to the continuity of $\overline{H}, \overline{g}$. Under the assumptions (A) and (B), the proof of this property reduces to proving continuity of the invariant measure μ_x of the process $Y^x_\cdot$ in equation (3.1). To do so, we need the following

*Assumptions **(C)***

- **(C1)** The diffusion ϱ is constant such that $\varrho\varrho^\top = \bar{\varrho}\,\mathbb{I}_m$ where $\bar{\varrho} > 0$ is a constant and $\mathbb{I}_m$ is the identity matrix.
- **(C2)** The drift b satisfies the following *strong recurrence condition*

$$\exists\, \kappa > 0 \text{ s.t. } (b(x,y_1) - b(x,y_2))\cdot(y_1-y_2) \leq -\kappa\,|y_1-y_2|^2, \quad \forall\, x, y_1, y_2. \tag{4.2}$$

- **(C3)** The utility function g and running cost ℓ are Lipschitz continuous in y uniformly in their other arguments.

It is clear that (C3) implies (B3), (C2) implies (A4), while (C1) is a particular case of (A3).
We recall the weighted norm (when it exists) $\|\varphi\|^p_{L^p(\nu)} = \int_{\mathbb{R}^m} |\varphi(y)|^p \mathrm{d}\nu(y)$ for $p \geq 1$, ν a positive measure, and the Wasserstein distance

$$\mathcal{W}_p(\mu,\nu) = \left(\inf_\pi \iint_{\mathbb{R}^m} |y-y'|^p \,\mathrm{d}\pi(y,y')\right)^{1/p}, \quad \text{for } p \geq 1 \tag{4.3}$$

where the minimization is performed over the collection of all measures π on $\mathbb{R}^m \times \mathbb{R}^m$ having marginals μ, ν. We now state a result in [9].

Lemma 4.1. *Under assumptions (A) and (C)*

$$\mathcal{W}_2(\mu_{x_1},\mu_{x_2}) \leq \bar{\varrho}\,\kappa^{-1}\,\|b(x_1,\cdot) - b(x_2,\cdot)\|_{L^2(\mu_{x_2})}$$

where μ_{x_i} is the unique invariant probability measure associated to equation (3.1) with $x = x_i, i = 1, 2$, respectively.

Proof. The inequality with $\overline{\varrho} = 1$ is [9], Corollary 2 where it is assumed that $b(x_1, \cdot), b(x_2, \cdot)$ satisfy (C2) and such that $|b(x_1, \cdot) - b(x_2, \cdot)| \in L^2(\mu_{x_2} + \mu_{x_2})$. This last condition is satisfied as a consequence of Lemma 3.1 which guarantees existence of all moments of the invariant measure in our setting and hence the desired integrability conditions. □

Proposition 4.2. *Under assumptions (A), (B) and (C), the effective Hamiltonian $\overline{H} : [0, T] \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{S}^n \to \mathbb{R}$ and initial data $\overline{g} : \mathbb{R}^n \to \mathbb{R}$ are continuous.*

Proof. We write the proof for $\overline{H}$ only, $\overline{g}$ being completely analogous. Recall the definition of the effective Hamiltonian $\overline{H}$

$$\overline{H}(t, x, p, P) = \int_{\mathbb{R}^m} H(t, x, y, p, P, 0)\mathrm{d}\mu_x(y),$$

where μ_x is the unique invariant probability measure associated to the fast subsystem (3.1). The Hamiltonian H inherits all the regularity properties of f, σ, ℓ as easily seen from its definition (2.14). Let $(t_1, x_1, p_1, P_1), (t_2, x_2, p_2, P_2) \in [0, T] \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{S}^n$,

$$\begin{aligned}\overline{H}(t_1, x_1, p_1, P_1) - \overline{H}(t_2, x_2, p_2, P_2) = \overline{H}(t_1, x_1, p_1, P_1) - \overline{H}(t_1, x_2, p_1, P_1)\\ + \overline{H}(t_1, x_2, p_1, P_1) - \overline{H}(t_2, x_2, p_2, P_2).\end{aligned} \tag{4.4}$$

On one hand we have

$$\overline{H}(t_1, x_2, p_1, P_1) - \overline{H}(t_2, x_2, p_2, P_2) = \int_{\mathbb{R}^m} H(t_1, x_2, y, p_1, P_1, 0) - H(t_2, x_2, y, p_2, P_2, 0)\ \mathrm{d}\mu_{x_2}(y). \tag{4.5}$$

The variable x_2 being fixed here, and from continuity of H in (t, p, P), one easily deduces continuity of $\overline{H}(\cdot, x_2, \cdot, \cdot)$. On the other hand, we have

$$\overline{H}(t_1, x_1, p_1, P_1) - \overline{H}(t_1, x_2, p_1, P_1) = \int_{\mathbb{R}^m} H(t_1, x_1, y, p_1, P_1, 0)\mathrm{d}\mu_{x_1}(y) - \int_{\mathbb{R}^m} H(t_1, x_2, y, p_1, P_1, 0)\mathrm{d}\mu_{x_2}(y)$$

The variables (t_1, p_1, P_1) being fixed, we introduce $\phi(x, y) := H(t_1, x, y, p_1, P_1, 0)$. We need then to estimate the quantity

$$\begin{aligned}&\int_{\mathbb{R}^m} \phi(x_1, y)\mathrm{d}\mu_{x_1}(y) - \int_{\mathbb{R}^m} \phi(x_2, y)\mathrm{d}\mu_{x_2}(y)\\ &\qquad = \int_{\mathbb{R}^m} \big(\phi(x_1, y) - \phi(x_2, y)\big)\mathrm{d}\mu_{x_1}(y) + \int_{\mathbb{R}^m} \phi(x_2, y)\mathrm{d}(\mu_{x_1} - \mu_{x_2})(y).\end{aligned} \tag{4.6}$$

The first term in the r.h.s. is continuous thanks to continuity of ϕ in x. We are then left with the second term

$$\begin{aligned}\int_{\mathbb{R}^m} \phi(x_2, y)\mathrm{d}(\mu_{x_1} - \mu_{x_2})(y) &= \iint_{\mathbb{R}^m} \phi(x_2, y) - \phi(x_2, y')\ \mathrm{d}\pi(y, y')\\ &\leq C \iint_{\mathbb{R}^m} |y - y'|\ \mathrm{d}\pi(y, y')\end{aligned} \tag{4.7}$$

for any $\pi(\cdot,\cdot)$ a probability measure on $\mathbb{R}^m \times \mathbb{R}^m$ with marginals μ_{x_1} and μ_{x_2}. Therefore, we have

$$\iint_{\mathbb{R}^m} \phi(x_2, y) - \phi(x_2, y')\, \mathrm{d}\pi(y, y') \le C\, \mathcal{W}_1(\mu_{x_1}, \mu_{x_2}) \le C\, \mathcal{W}_2(\mu_{x_1}, \mu_{x_2}) \tag{4.8}$$

Using Lemma 4.1, we have the following

$$\mathcal{W}_2(\mu_{x_1}, \mu_{x_2}) \le\, \overline{\varrho}\, \kappa^{-1} \left(\int_{\mathbb{R}^m} |b(x_1, y) - b(x_2, y)|^2 \,\mathrm{d}\mu_{x_2}(y) \right)^{1/2}$$

and hence

$$\mathcal{W}_2(\mu_{x_1}, \mu_{x_2}) \le\, \overline{\varrho}\, \kappa^{-1}\, C\, |x_1 - x_2| \tag{4.9}$$

where $C > 0$ is now the Lipschitz constant of b. Finally, using equations (4.7), (4.8) and (4.9) we can upperbound the r.h.s. of equation (4.6) with

$$\int_{\mathbb{R}^m} \big(\phi(x_1, y) - \phi(x_2, y)\big) \mathrm{d}\mu_{x_1}(y) + \overline{\varrho}\, \kappa^{-1}\, C\, |x_1 - x_2|. \tag{4.10}$$

Finally, exchanging the roles of x_1, x_2, and using equations (4.4), (4.5) and (4.10), we get the joint continuity of $\overline{H}$ in all its arguments. □

Remark 4.3. Note that the upper bound (4.10) yields Lipschitz continuity of $x \mapsto \overline{H}(t, x, p, Y)$ provided H is Lipschitz in (x, y); the Lipschitz continuity in y being needed in equation (4.7). This observation will be useful in Step 5 of the proof of the main result.

4.2. The main result

We are now ready to state and prove our main convergence result. The last assumption we need is the following

Assumption ***(D)***

- **(D)** The matrix $\Sigma = \sigma\sigma^\top(x, y, u)$ has bounded second derivatives in x, uniformly in (y, u) and at least one of the two conditions is satisfied:
  - (a) Σ is independent of y and u, *i.e.* $\sigma = \sigma(x)$;
  - (b) the drift of the fast process is independent of x, *i.e.* $b = b(y)$.

Assumption (D) ensures that the square root of Σ is Lipschitz in x (see [40], Thm. 5.2.3, p. 132) and will be needed in Step 5 of the proof of our next result. For our main motivation as described in the introduction, assumption (D.a) is satisfied because $\sigma = 0$. Assumption (D.b) on the other hand is relevant for applications in finance, see [7, 18] and the references therein.

Theorem 4.4. *Under assumptions (A), (B), (C) and (D), the solution V^ε to equation* (2.9) *converges uniformly on compact subsets of $(0, T) \times \mathbb{R}^n \times \mathbb{R}^m$ to the unique continuous viscosity solution of the limit problem* (4.1) *satisfying a quadratic growth condition in x, i.e.*

$$\exists\, K > 0 \text{ such that } |V(t, x)| \le K(1 + |x|^2), \quad \forall\, (t, x) \in [0, T] \times \mathbb{R}^n \tag{4.11}$$

Remark 4.5. Without assumption (D) we prove that the weak semilimits $\underline{V}$ and $\overline{V}$ are independent of y and are, respectively, a super- and a subsolution of equation (4.1), as in [2], Theorem 1. If, in addition, we assume (D), we prove that the Comparison Principle holds for equation (4.1), which implies the uniform convergence of V^ε.

The last auxiliary result we need is a Liouville property for semi-solutions of the PDE

$$-\mathcal{L}V(y) = -b(x,y)\cdot\nabla V(y) - \operatorname{trace}(\varrho\varrho^\top(x,y)D^2V(y)) = 0, \quad \text{in } \mathbb{R}^m, \tag{4.12}$$

where $x\in\mathbb{R}^n$ is frozen, taken from [6], Theorems 2.1 and 2.2 or [31], Proposition 3.1.

Lemma 4.6. *Assume there exist a function $\omega\in C^\infty(\mathbb{R}^m)$ and $R_0>0$ such that*

$$-\mathcal{L}\omega \geq 0 \quad \text{in } \overline{B(0,R_0)}^C, \quad \omega(y)\to+\infty \ \text{ as } \ |y|\to+\infty. \tag{4.13}$$

Then every viscosity subsolution $V\in USC(\mathbb{R}^m)$ to equation (4.12) such that $\limsup_{|y|\to\infty}\frac{V}{\omega}\leq 0$ and every viscosity supersolution $U\in LSC(\mathbb{R}^m)$ to equation (4.12) such that $\liminf_{|y|\to\infty}\frac{U}{\omega}\geq 0$ are constant.

Proof. (Thm. 4.4) The proof follows the one of [7], Theorem 5.1 (see also [5], Thm. 3.2).

Step 1. We define the half-relaxed semilimits

$$\underline{V}(t,x,y) = \liminf_{\substack{\varepsilon\to0\\ t'\to t, x'\to x, y'\to y}} V^\varepsilon(t',x',y'), \quad \overline{V}(t,x,y) = \limsup_{\substack{\varepsilon\to0\\ t'\to t, x'\to x, y'\to y}} V^\varepsilon(t',x',y')$$

for $t<T, x\in\mathbb{R}^n, y\in\mathbb{R}^m$, and

$$\underline{V}(T,x,y) = \liminf_{\substack{\varepsilon\to0\\ t'\to T^-, x'\to x, y'\to y}} V^\varepsilon(t',x',y'),\ \overline{V}(T,x,y) = \limsup_{\substack{\varepsilon\to0\\ t'\to T^-, x'\to x, y'\to y}} V^\varepsilon(t',x',y').$$

By equation (2.15) they also have quadratic growth, that is,

$$|\underline{V}(t,x,y)|,\ |\overline{V}(t,x,y)| \ \leq K(1+|x|^2+|y|^2), \quad \forall\, t\in[0,T],\ x\in\mathbb{R}^n,\ y\in\mathbb{R}^m. \tag{4.14}$$

Step 2. *(We show that $\underline{V}(t,x,y),\overline{V}(t,x,y)$ do not depend on y for every $t\in[0,T)$ and $x\in\mathbb{R}^n$.)* Arguing as in Step 2 of the proof of [7], Theorem 5.1, we get that $\overline{V}(t,x,y)$ (resp., $\underline{V}(t,x,y)$) is, for every $t\in(0,T)$ and $x\in\mathbb{R}^n$, a viscosity subsolution (resp., supersolution) to

$$-\mathcal{L}(x,y,D_yV,D^2_{yy}V) = 0 \quad \text{in } \mathbb{R}^m \tag{4.15}$$

where $\mathcal{L}$ is the differential operator defined in equation (2.13). Consider now the function ω defined on $\mathbb{R}^m\setminus\{0\}$ such that

$$\omega(y) = \frac{1}{2}|y|^2\log|y| \tag{4.16}$$

and such that $\omega(0)=0$. It is easy to check that

$$\nabla\omega(y) = \left(\frac{1}{2}+\log(|y|)\right)y \quad \text{and} \quad D^2\omega(y) = \left(\frac{1}{2}+\log(|y|)\right)\mathbb{I}_m + \frac{y\otimes y}{|y|^2}.$$

Therefore, recalling $a = \varrho\varrho^\top$, one has

$$\begin{aligned}-\mathcal{L}\omega &= -\left(\frac{1}{2}+\log(|y|)\right)(b(y)\cdot y) - \left(\frac{1}{2}+\log(|y|)\right)\operatorname{trace}(a(y)) - \frac{1}{|y|^2}\operatorname{trace}((y\otimes y)a(y)) \\ &\geq -\left(\frac{1}{2}+\log(|y|)\right)\left((b(y)\cdot y) + m\overline{\Lambda}\right) - \overline{\Lambda} \xrightarrow[|y|\to\infty]{} +\infty\end{aligned} \tag{4.17}$$

thanks to assumptions (2.5) and (2.4). Then one can find $R > 0$ such that

$$-\mathcal{L}\omega(y) \geq 0 \quad \text{in } \overline{B(0,R)}^C, \qquad \text{and } \omega(y) \xrightarrow[|y|\to\infty]{} +\infty. \tag{4.18}$$

We can now use Lemma 4.6 with such a Lyapunov function ω, since $\overline{V}, \underline{V}$ have at most a quadratic growth in y, to conclude that the functions $y \mapsto \overline{V}(t,x,y)$, $y \mapsto \underline{V}(t,x,y)$ are constants for every $(t,x) \in (0,T)\times\mathbb{R}^n$. Finally, using the definition it is immediate to see that this implies that also $\overline{V}(T,x,y)$ and $\underline{V}(T,x,y)$ do not depend on y.

Step 3. *(We show that $\overline{V}$ and $\underline{V}$ are sub and supersolutions to the PDE in equation (4.1) in $(0,T)\times\mathbb{R}^n$.)* The proof adapts the perturbed test function method [2, 16]. To show that $\overline{V}$ is a viscosity subsolution we fix $(\bar{t},\bar{x}) \in (0,T)\times\mathbb{R}^n$ and a smooth function ψ such that $\psi(\bar{t},\bar{x}) = \overline{V}(\bar{t},\bar{x})$ and $\overline{V} - \psi$ has a strict maximum at $(\bar{t},\bar{x})$. We must prove that

$$-\psi_t(\bar{t},\bar{x}) + \overline{H}(\bar{t},\bar{x},D_x\psi(\bar{t},\bar{x}),D^2_{xx}\psi(\bar{t},\bar{x})) + \lambda\overline{V}(\bar{t},\bar{x}) \leq 0$$

Set $\bar{p} = D_x\psi(\bar{t},\bar{x})$, $\overline{P} = D^2_{xx}\psi(\bar{t},\bar{x})$, and assume by contradiction that for some $\eta > 0$

$$-\psi_t(\bar{t},\bar{x}) + \overline{H}(\bar{t},\bar{x},\bar{p},\overline{P}) + \lambda\psi(\bar{t},\bar{x}) \geq 5\eta.$$

By the continuity of $\overline{H}$ given by Proposition 4.2, we can choose $r > 0$ such that

$$-\psi_t(t,x) + \overline{H}(t,x,\bar{p},\overline{P}) + \lambda\psi(t,x) \geq 4\eta \tag{4.19}$$

for all $(x,t) \in B((\bar{t},\bar{x}),r)$, and $\varepsilon_o > 0$ such that

$$|H^\varepsilon(t,x,y,D_x\psi(t,x),D^2_{xx}\psi(t,x),0) - H(\bar{t},\bar{x},y,\bar{p},\overline{P},0)| < \eta \tag{4.20}$$

for all $(x,t) \in B((\bar{t},\bar{x}),r)$, $y \in \overline{B}(0,R)$ (R to be chosen soon), and $\varepsilon \leq \varepsilon_o$. Now consider, as in equation (3.22), the $\delta(n)$-cell problem

$$\begin{cases}\delta\chi_\delta(y) - \mathcal{L}(\bar{x},y,D\chi_\delta,D^2\chi_\delta) + H(\bar{t},\bar{x},y,\bar{p},\overline{P},0) = 0, & \text{in } D_n,\\ \chi_\delta(y) = 0, & \text{in } \partial D_n,\end{cases} \tag{4.21}$$

where $\delta := \delta(n) = O\left(\frac{1}{n^{4+\alpha}}\right)$ and $D_n = B(0,n)$. By Proposition 3.4 there exists $n_o > 0$ such that, for every $n \geq n_o$, $R < n_o$,

$$|\delta\chi_\delta(y) + \overline{H}(\bar{t},\bar{x},\bar{p},\overline{P})| \leq \eta, \quad \forall y \in B(0,R). \tag{4.22}$$

Moreover

$$|\mathcal{L}(\bar{x},y,D\chi_\delta,D^2\chi_\delta) - \mathcal{L}(x,y,D\chi_\delta,D^2\chi_\delta)| < \eta \tag{4.23}$$

for $|x-\overline{x}|<r$, by decreasing r if necessary, and we set $C_n := \max_{\overline{B(0,R)}} |\chi_\delta(y)|$. We define the perturbed test function

$$\psi^\varepsilon(t,x,y) := \psi(t,x) + \varepsilon\chi_\delta(y), \tag{4.24}$$

which is in $C^2(\overline{\Omega})$ for $\Omega := B((\overline{t},\overline{x}),r)\times B(0,R)$. We claim that ψ^ε is a strict supersolution of the PDE (2.9) in Ω for $\varepsilon\le\varepsilon_o$ and $\varepsilon\lambda C_n<\eta$. In fact

$$\begin{aligned}
&-\psi^\varepsilon_t(t,x) + H^\varepsilon(t,x,y,D_x\psi(t,x),D^2_{xx}\psi(t,x),0) - \mathcal{L}(x,y,D\chi_\delta,D^2\chi_\delta) + \lambda\psi^\varepsilon_t(t,x)\\
&\qquad \ge -\psi_t(t,x) + H^\varepsilon(t,x,y,D_x\psi(t,x),D^2_{xx}\psi(t,x),0) - \delta\chi_\delta(y) - H(\overline{t},\overline{x},y,\overline{p},\overline{P},0) + \lambda\psi^\varepsilon_t(t,x)\\
&\qquad \ge -\psi_t(t,x) - \eta + \overline{H}(\overline{t},\overline{x},\overline{p},\overline{P}) - \eta + \lambda\psi_t(t,x) + \lambda\varepsilon\chi_\delta(y)\\
&\qquad \ge 4\eta - 2\eta - \lambda\varepsilon C_n \ge \eta > 0
\end{aligned} \tag{4.25}$$

where in the first inequality we used equations (4.23) and (4.21), in the second inequality we used equations (4.20) and (4.22), and in the third inequality we used equation (4.19).

Since the maximum of $\overline{V}-\psi$ at $(\overline{t},\overline{x})$ is strict, we can decrease r so that $\overline{V}-\psi\le -2\eta$ on $\partial\Omega$. Moreover

$$\limsup_{\substack{\varepsilon\to 0\\ t'\to t, x'\to x, y'\to y}} V^\varepsilon(t',x',y') - \psi^\varepsilon(t',x',y') = \overline{V}(t,x) - \psi(t,x) \tag{4.26}$$

and the compactness of $\partial\Omega$ imply that $V^\varepsilon - \psi^\varepsilon \le -\eta$ on $\partial\Omega$ for ε small enough. We claim that, for such ε,

$$V^\varepsilon - \psi^\varepsilon \le -\eta \qquad \text{in } \Omega. \tag{4.27}$$

In fact, if this is not the case, $V^\varepsilon-\psi^\varepsilon$ has a maximum point in Ω, a contradiction to the fact that V^ε is a viscosity subsolution of equation (2.9) in Ω and ψ^ε satisfies equation (4.25). Now equations (4.26) and (4.27) imply $\overline{V}(\overline{t},\overline{x})<\psi(\overline{t},\overline{x})$, which is a contradiction and completes the proof that $\overline{V}$ is a subsolution to equation (4.1). The proof that $\underline{V}$ is a supersolution is completely analogous.

Step 4. *(Behavior of $\overline{V}$ and $\underline{V}$ at time T)* In this step, we adapt the *Step 4* in the proof of [7], Theorem 5.1 or in [5], Theorem 3.2 using our result in Proposition 3.7. The main difference relies in the use of the sequence of Cauchy problems with bounded domains (3.33) instead of the Cauchy problem (3.32) that was used in [5, 7]. We repeat the proof for the sake of consistency and clarity.
We prove only the statement for subsolution, since the proof for the supersolution is completely analogous.

We fix $\overline{x}\in\mathbb{R}^n$ and $t_0>0$, and we consider, for some $n>0$ to be later made precise, the unique bounded solution $\omega^{r,n}$ to the Cauchy problem in $[0,T(n)]\times D_n$ where $T(n) := n^2 t_0$ and D_n is the ball of radius n in $\mathbb{R}^m$

$$\begin{cases}
\omega_t - \mathcal{L}(\overline{x},y,D\omega,D^2\omega) = 0, & \text{in } (0,T(n)]\times D_n,\\
\omega(0,y) = \sup\limits_{\{|x-\overline{x}|\le r\}} g(x,y), & \text{in } D_n,\\
\qquad\qquad \omega(t,y) = 0, & \text{in } [0,T(n)]\times\partial D_n.
\end{cases} \tag{4.28}$$

Using stability properties of viscosity solutions it is not hard to see that $\omega^{r,n}$ converges, as $r\to 0$, to the solution ω^n of equation (3.33) set in $[0,T(n)]\times D_n$. We recall that

$$\overline{g}(\overline{x}) := \mu_{\overline{x}}(g(\overline{x},\cdot)) = \int_{\mathbb{R}^m} g(\overline{x},y)\,\mathrm{d}\mu_{\overline{x}}(y).$$

Using the convergence result in Proposition 3.7 and the uniform convergence of $\omega^{r,n}$ to ω^n, it is easy to see that for every $\eta > 0$ there exist r_0 and $n_0 > 0$ such that

$$\forall\, n \geq n_0 : \quad |\omega^{r,n}(T(n), y) - \overline{g}(\overline{x})| \leq \eta, \quad \forall\, r < r_0,\, y \in D_n \supseteq \overline{D}_{n_0}. \tag{4.29}$$

We now fix $r < r_0$ and a constant M_r such that $V^\varepsilon(t,x,y) \leq M_r$ and $|g(x,y)| \leq M_r/2$ for every $\varepsilon > 0$, $x \in \overline{B} := \overline{B(\overline{x}, r)}$ and $y \in \overline{D} := \overline{D}_{n_0}$. This is possible by Proposition 2.1 and assumption (2.7). Moreover we fix a smooth nonnegative function ψ such that $\psi(\overline{x}) = 0$ and $\psi(x) + \inf_{y\in\overline{D}} g(x,y) \geq 2M_r$ for every $x \in \partial B$ (which is easy to build because $\inf_{x\in\partial B}\inf_{y\in\overline{D}} g(x,y) \geq -M_r/2$). Let C_r be a positive constant such that

$$|H^\varepsilon(t,x,y,D\psi(x),D^2\psi(x),0)| \leq C_r \quad \text{for } x \in \overline{B},\ y \in \overline{D} \text{ and } \varepsilon > 0$$

where H^ε is defined in equation (2.12). Note that such a constant exists thanks to assumptions (2.2) and (2.7). We define the function

$$\psi_r^\varepsilon(t,x,y) = \omega^{r,n}\left(\frac{T-t}{\varepsilon}, y\right) + \psi(x) + C_r(T-t),$$

for some fixed $n > n_0$, and we claim that it is a supersolution to the parabolic problem

$$\begin{cases} -V_t + F^\varepsilon\left(t,x,y,V,D_xV,\dfrac{D_yV}{\varepsilon}, D^2_{xx}V^\varepsilon, \dfrac{D^2_{yy}V}{\varepsilon}, \dfrac{D^2_{xy}V}{\sqrt{\varepsilon}}\right) = 0, & \text{in } (0,T)\times B\times\overline{D} \\ V(t,x,y) = M_r, & \text{in } (0,T)\times\partial B\times\overline{D} \\ V(T,x,y) = g(x,y), & \text{in } \overline{B}\times\overline{D} \end{cases} \tag{4.30}$$

where F^ε is defined in equation (2.11). Indeed

$$\begin{aligned} &-(\psi_r^\varepsilon)_t + F^\varepsilon\left(t,x,y,D_x\psi_r^\varepsilon, \frac{D_y\psi_r^\varepsilon}{\varepsilon}, D^2_{xx}\psi_r^\varepsilon, \frac{D^2_{yy}\psi_r^\varepsilon}{\varepsilon}, \frac{D^2_{xy}\psi_r^\varepsilon}{\sqrt{\varepsilon}}\right) \\ &\qquad = \frac{1}{\varepsilon}\left[(\omega^{r,n})_t - \mathcal{L}(y, D\omega^{r,n}, D^2\omega^{r,n})\right] + C_r + H^\varepsilon(t,x,y,D\psi(x),D^2\psi(x),0) \geq 0. \end{aligned}$$

Moreover $\psi_r^\varepsilon(T,x,y) = \sup_{\{|x-\overline{x}|\leq r\}} g(x,y) + \psi(x) \geq g(x,y)$.

Finally, observe that the constant function $\min\{\,0\,;\ \inf_{y\in\overline{D}}\ \sup_{\{|x-\overline{x}|\leq r\}} g(x,y)\,\}$ is always a subsolution to equation (4.28) and then by a standard comparison principle we obtain

$$\omega^{r,n}(t,y) \geq \min\{\,0\,;\ \inf_{y\in\overline{D}}\ \sup_{\{|x-\overline{x}|\leq r\}} g(x,y)\,\}.$$

This implies, for all $x \in \partial B$,

$$\psi_r^\varepsilon(t,x,y) \geq \min\{\,0\,;\ \inf_{y\in\overline{D}}\ \sup_{\{|x-\overline{x}|\leq r\}} g(x,y)\,\} + 2M_r - \inf_{y\in\overline{D}} g(x,y) + C_r(T-t) \geq M_r\,,$$

where we have used either the fact that $|g(x,y)| \leq M_r/2$, and hence $-\inf_{y\in\overline{D}} g(x,y) \geq -M_r/2$, when we have $\min\{\,0\,;\ \inf_{y\in\overline{D}}\ \sup_{\{|x-\overline{x}|\leq r\}} g(x,y)\,\} = 0$, or otherwise, we have used the fact that $\inf_{y\in\overline{D}}\ \sup_{\{|x-\overline{x}|\leq r\}} g(x,y) - \inf_{y\in\overline{D}} g(x,y) \geq$

0. In the first case, we get $\psi_r^\varepsilon(t,x,y) \geq 3M_r/2$ and in the second case we have $\psi_r^\varepsilon(t,x,y) \geq 2M_r$. Then ψ_r^ε is a supersolution to equation (4.30). For our choice of M_r we get that V^ε is a subsolution to equation (4.30). Moreover both V^ε and ψ_r^ε are bounded in $[0,T]\times\overline{B}\times\overline{D}$, because of the estimate (2.15), of the boundedness of $\omega^{r,n}$ and of the regularity of ψ. So, a standard comparison principle for viscosity solutions gives

$$V^\varepsilon(t,x,y) \leq \psi_r^\varepsilon(t,x,y) = \omega^{r,n}\left(\frac{T-t}{\varepsilon}, y\right) + \psi(x) + C_r(T-t)$$

for every $0<r<r_0$, $n>n_0$ $\varepsilon>0$, $(t,x,y)\in[0,T]\times\overline{B}\times\overline{D}$. We compute the upper limit of both sides of the previous inequality as $(\varepsilon,t,x,y)\to(0,t',x',y')$ for $t'\in(0,T)$, $x'\in B$, $y'\in D$ and $\varepsilon \coloneqq \varepsilon(n) = \frac{T-t}{T(n)}$ (recalling $T(n)=n^2t_0$) and get, using equation (4.29),

$$\overline{V}(t',x') \leq \overline{g}(\overline{x}) + \eta + \psi(x') + C_r(T-t').$$

Then taking the upper limit for $(t',x')\to(T,\overline{x})$, we obtain $\overline{V}(T,\overline{x}) \leq \overline{g}(\overline{x})+\eta$ which permits us to conclude recalling that η is arbitrary.

The proof for $\underline{V}$ is completely analogous, once we replace the Cauchy problem (4.28) with

$$\begin{cases} \omega_t - \mathcal{L}(y, D\omega, D^2\omega) = 0, & \text{in } (0,T(n)]\times D_n, \\ \omega(0,y) = \inf\limits_{\{|x-\overline{x}|\leq r\}} g(x,y), & \text{in } D_n, \\ \omega(t,y) = 0, & \text{in } [0,T(n)]\times\partial D_n. \end{cases}$$

Step 5. *(Uniform convergence)*. We observe that by definition $\overline{V}\geq\underline{V}$ and that both $\overline{V}$ and $\underline{V}$ satisfy the same quadratic growth condition (4.11). Moreover the Hamiltonian $\overline{H}$ defined in equation (3.19) can be written thanks to Proposition 3.6 as a Bellman Hamiltonian of the form

$$\overline{H}(t,x,p,P) = \min_{\nu\in L^\infty(\mathbb{R}^m,U)} \left\{ -\operatorname{trace}(\overline{\sigma}\,\overline{\sigma}^\top P) - \overline{f}\cdot p - \overline{\ell} \right\}$$

where

$$\overline{\sigma} = \overline{\sigma}(x,\nu) = \sqrt{\int_{\mathbb{R}^m} \sigma\sigma^\top(x,y,\nu(y))\,\mathrm{d}\mu_x(y)}\,, \qquad \overline{f} = \overline{f}(x,\nu) = \int_{\mathbb{R}^m} f(x,y,\nu(y))\,\mathrm{d}\mu_x(y)$$
$$\overline{\ell} = \overline{\ell}(t,x,\nu) = \int_{\mathbb{R}^m} \ell(t,x,y,\nu(y))\,\mathrm{d}\mu_x(y).$$

Under assumptions (D), we actually have in the case (D.a)

$$\overline{\sigma} = \sqrt{\sigma\sigma^\top(x)},$$

and in the case (D.b), the invariant measure of the fast process Y does not depend on x. Therefore, $\overline{\sigma},\overline{f},\overline{\ell}$ inherit regularity and growth conditions of σ,f,ℓ thanks to assumptions (A), (B), (C), (D) and Remark 4.3, and fall in the framework of [14]. Hence we can use the comparison result between sub- and supersolutions to parabolic problems satisfying a quadratic growth condition, given in [14], Theorem 2.1, to deduce $\underline{V}\geq\overline{V}$. Therefore $\underline{V}=\overline{V}=: V$. In particular V is continuous, and by definition of half-relaxed semilimits, this implies that V^ε converges locally uniformly to V (see [4], Lem. V.1.9). □

Acknowledgements. The authors wish to thank Markus Fischer for useful conversations and for pointing out [23], Proposition 1.4 used in the proof of Lemma 3.2. The authors also thank the two anonymous referees for the many useful comments.

Authors contributions. The first author is member of the Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA) of the Istituto Nazionale di Alta Matematica (INdAM). He also participates in the King Abdullah University of Science and Technology (KAUST) project CRG2021-4674 "Mean-Field Games: models, theory, and computational aspects". The second author is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 320021702/GRK2326 – Energy, Entropy, and Dissipative Dynamics (EDDy). The results of this paper are part of his Ph.D. thesis [26] which was conducted when he was a Ph.D. student at the University of Padova.

References

[1] O. Alvarez and M. Bardi, Viscosity solutions methods for singular perturbations in deterministic and stochastic control. *SIAM J. Control Optim.* **40** (2002) 1159–1188.

[2] O. Alvarez and M. Bardi, Singular perturbations of nonlinear degenerate parabolic PDEs: a general convergence result. *Arch. Rational Mech. Anal.* **170** (2003) 17–61.

[3] S.R. Athreya, V.S. Borkar, K.S. Kumar and R. Sundaresan, Simultaneous small noise limit for singularly perturbed slow-fast coupled diffusions. *Appl. Math. Optim.* **83** (2021) 2327–2374.

[4] M. Bardi and I. Capuzzo-Dolcetta, Optimal Control and Viscosity Solutions of Hamilton–Jacobi–Bellman Equations. Springer Science & Business Media (2008).

[5] M. Bardi and A. Cesaroni, Optimal control with random parameters: a multiscale approach. *Eur. J. Control* **17** (2011) 30–45.

[6] M. Bardi and A. Cesaroni, Liouville properties and critical value of fully nonlinear elliptic operators. *J. Diff. Equ.* **261** (2016) 3775–3799.

[7] M. Bardi, A. Cesaroni and L. Manca, Convergence by viscosity methods in multiscale financial models with stochastic volatility. *SIAM J. Finan. Math.* **1** (2010) 230–265.

[8] M. Bardi and H. Kouhkouh, Deep relaxation of controlled stochastic gradient descent via singular perturbations, arXiv preprint arXiv:2209.05564 (2022).

[9] V.I. Bogachev, A.I. Kirillov and S.V. Shaposhnikov, The Kantorovich and variation distances between invariant measures of diffusions and nonlinear stationary Fokker-Planck-Kolmogorov equations. *Math. Notes* **96** (2014) 855–863.

[10] V.S. Borkar and V. Gaitsgory, Averaging of singularly perturbed controlled stochastic differential equations. *Appl. Math. Optim.* **56** (2007) 169–209.

[11] V.S. Borkar and V. Gaitsgory, Singular perturbations in ergodic control of diffusions. *SIAM J. Control Optim.* **46** (2007)1562–1577.

[12] P. Chaudhari, A. Choromanska, S. Soatto, Y. LeCun, C. Baldassi, C. Borgs, J. Chayes, L. Sagun and R. Zecchina, Entropy-SGD: Biasing gradient descent into wide valleys. *J. Stat. Mech.: Theory Exp.* **2019** (2019) 124018.

[13] P. Chaudhari, A. Oberman, S. Osher, S. Soatto and G. Carlier, Deep relaxation: partial differential equations for optimizing deep neural networks. *Res. Math. Sci.* **5** (2018) 1–30.

[14] F. Da Lio and O. Ley, Uniqueness results for second-order Bellman–Isaacs equations under quadratic growth assumptions and Applications. *SIAM J. Control Optim.* **45** (2006) 74–106.

[15] F. de Feo, The averaging principle for non-autonomous slow-fast stochastic differential equations and an application to a local stochastic volatility model. *J. Diff. Equ.* **302** (2021) 406–443.

[16] L.C. Evans, The perturbed test function method for viscosity solutions of nonlinear PDE. *Proc. Roy. Soc. Edinb. A: Math.* **111** (1989) 359–375.

[17] J. Feng, J.-P. Fouque and R. Kumar, Small-time asymptotics for fast mean-reverting stochastic volatility models. *Ann. Appl. Probab.* **22** (2012) 1541–1575.

[18] J.-P. Fouque, G. Papanicolaou, R. Sircar and K. Sølna, Multiscale Stochastic Volatility for Equity, Interest Rate, and credit Derivatives. Cambridge University Press (2011).

[19] D. Ghilli, Viscosity methods for large deviations estimates of multiscale stochastic processes. *ESAIM: Control Optim. Calc. Variat.* **24** (2018) 605–637.

[20] D. Ghilli and C. Marchi, Rate of convergence for singular perturbations of Hamilton–Jacobi equations in unbounded spaces, arXiv preprint arXiv:2201.04592 (2022).

[21] B. Goldys, G. Tessitore, J. Yang and Z. Zhou, Multiscale linear-quadratic stochastic optimal control with multiplicative noise, arXiv preprint arXiv:2011.09238 (2020).

[22] G. Guatteri and G. Tessitore, Singular limit of BSDEs and optimal control of two scale stochastic systems in infinite dimensional Spaces. *Appl. Math. Optim.* **83** (2021) 1025–1051.

[23] S. Herrmann, P. Imkeller and D. Peithmann, Transition times and stochastic resonance for multidimensional diffusions with time periodic drift: a large deviations approach. *Ann. Appl. Probab.* **16** (2006) 1851–1892.

[24] C. Himmelberg, Measurable relations. *Fundamenta Mathematicae* **87** (1975) 53–72.

[25] P. Kokotović, H.K. Khalil and J. O'Reilly, Singular perturbation methods in control: analysis and design. *SIAM*, 1999.

[26] H. Kouhkouh, Some asymptotic problems for Hamilton-Jacobi-Bellman equations and applications to global optimization. PhD thesis, University of Padova (2022).

[27] H. Kushner, Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems. Springer Science & Business Media (1990).

[28] P.-L. Lions, G. Papanicolaou and S.R.S. Varadhan, Homogenization of Hamilton–Jacobi equations, Unpublished preprint (1987).

[29] W. Liu, M. Röckner, X. Sun and Y. Xie, Averaging principle for slow-fast stochastic differential equations with time dependent locally Lipschitz coefficients. *J. Diff. Equ.* **268** (2020) 2910–2948.

[30] L. Lorenzi and M. Bertoldi, Analytical Methods for Markov Semigroups. CRC Press (2006).

[31] P. Mannucci, C. Marchi and N. Tchou, The ergodic problem for some subelliptic operators with unbounded coefficients. *Nonlinear Diff. Equ. Appl. NoDEA* **23** (2016) 1–26.

[32] X. Mao, Stochastic Differential Equations and Applications. Elsevier (2007).

[33] E. Pardoux and A.Yu. Veretennikov, On the Poisson equation and diffusion approximation. I *Ann. Prob.* (2001) 1061–1085.

[34] E. Pardoux and A.Yu. Veretennikov, On Poisson equation and diffusion approximation 2. *Ann. Prob.* **31** (2003) 1166–1192.

[35] E. Pardoux and A.Yu. Veretennikov, On the Poisson equation and diffusion approximation 3. *Ann. Prob.* **33** (2005) 1111–1133.

[36] M. Pavon, On local entropy, stochastic control and deep neural networks. *IEEE Control Syst. Lett.* **7** (2023) 437–441.

[37] D. Revuz and M. Yor, Continuous Martingales and Brownian Motion, Vol. 293. Springer Science & Business Media (2013).

[38] M. Röckner and L. Xie, Averaging principle and normal deviations for multiscale stochastic systems. *Commun. Math. Phys.* **383** (2021) 1889–1937.

[39] K. Spiliopoulos, Large deviations and importance sampling for systems of slow-fast motion. *Appl. Math. Optim.* **67** (2013) 123–161.

[40] D.W. Stroock and S.R.S. Varadhan, Multidimensional Diffusion Processes, Vol. 233. Springer Science & Business Media (1997).

[41] A. Świech, Singular perturbations and optimal control of stochastic systems in infinite dimension: HJB equations and viscosity Solutions. *ESAIM: Control Optim. Calc. Variat.* **27** (2021) 6.

[42] A.Yu. Veretennikov, On polynomial mixing bounds for stochastic differential equations. *Stochastic Processes Appl.* **70** (1997) 115–127.