The Annals of Probability
2001, Vol. 29, No. 3, 1061–1085

ON THE POISSON EQUATION AND DIFFUSION APPROXIMATION. I[1]

BY E. PARDOUX AND A. YU. VERETENNIKOV[2]

Université de Provence and Institute of Information Transmission Problems

Dedicated to N. V. Krylov on his sixtieth birthday

A Poisson equation in $\mathbb{R}^d$ for the elliptic operator corresponding to an ergodic diffusion process is considered. Existence and uniqueness of its solution in Sobolev classes of functions is established along with the bounds for its growth. This result is used to study a diffusion approximation for two-scaled diffusion processes using the method of corrector; the solution of a Poisson equation serves as a corrector.

1. Introduction. The first topic of this paper is the investigation of the Poisson equation in $\mathbb{R}^d$,

$$Lu = -f, \tag{1}$$

where L is an elliptic differential operator of second order,

$$L = \sum a_{ij}(x)\partial_{x_i}\partial_{x_j} + \sum b_i(x)\partial_{x_i},$$

which may be regarded as the infinitesimal generator of a positive recurrent diffusion process X solution of the stochastic differential equation

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t, \qquad X_0 = x \in \mathbb{R}^d,$$

with $a = \sigma\sigma^*/2$. We assume that f is "centered," that is, $\int f(x)\mu(dx) = 0$, where μ is the invariant probability measure of our diffusion process. Let

$$u(x) = \int_0^\infty E_x f(X_t)\,dt.$$

We will show that under some assumptions this function is well-defined, locally bounded, continuous, belongs locally to the Sobolev class $W^2_p(\mathbb{R}^d)$ and satisfies equation (1) in the Sobolev sense. Moreover, under certain assumptions on b, σ and f, this solution is either bounded or slowly increasing. These properties are important in limit theorems of diffusion approximation type, which is our second subject.

What we mean here by diffusion approximation is the convergence of singularly perturbed ordinary differential equations with random inputs towards

Received May 1998; revised October 2000.
[1]Supported by CNRS project PICS-99 and INTAS 99-00559.
[2]Supported by RFBR project 98-01-00062.
AMS 2000 *subject classifications*. 60H30, 60J45, 60J60, 35J15.
Key words and phrases. Poisson equation, polynomial recurrence, diffusion approximation.

stochastic differential equations. Typically, one wants to study the asymptotic behavior, as $\varepsilon \to 0$, of the solution Y^ε of

$$\frac{dY_t^\varepsilon}{dt} = F(X_t^\varepsilon, Y_t^\varepsilon) + \varepsilon^{-1} G(X_t^\varepsilon, Y_t^\varepsilon), \tag{2}$$

where $X_t^\varepsilon = X_{t/\varepsilon^2}$, X is an ergodic Markov process and for all y,

$$\int G(x, y)\mu(dx) = 0,$$

with again μ the unique invariant measure of X.

The first results in that direction seem to be due to Stratonovich (1963, 1967) and Khasminski (1966). We refer the reader to Papanicolaou, Stroock and Varadhan (1977) and Chapter 12 of Ethier and Kurtz (1986) for an account of the theory and more complete references. It turns out that the limiting coefficients are expressed in terms of solutions of Poisson equations with $G(\cdot, y)$ as the right-hand side. Consequently, bounds on the limiting coefficients depend on bounds on the solution of some Poisson equations. This is not a problem when the driving noise X takes values in a compact state space, but is a difficulty for a noncompact state space. Exploiting our results on Poisson equations, we shall give conditions on the coefficients of the singularly perturbed ordinary differential equation (2) (in fact a stochastic differential equation in greater generality) driven by the diffusion process X^ε which takes values in Euclidean space, under which we shall prove the diffusion approximation result. Our assumptions are more explicit and weaker than other results that we know of. In a subsequent publication, we intend to treat the case where the process Y^ε feeds back into the stochastic differential equation which defines the process X^ε.

The paper is organized as follows. In Section 2 we state our basic assumptions on the coefficients of the diffusion process X, which imply in particular that it is positive recurrent and ergodic, recall and prove some results on moment bounds and convergence to the invariant measure. Section 3 is devoted to the results on Poisson equations and Section 4 to diffusion approximation. Finally the Appendix contains a detailed proof of a version of the Itô–Krylov formula which is used in Section 4.

2. Moment bounds and convergence to the invariant measure for SDEs.

Consider the stochastic Itô equation

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t, \qquad X_0 = x \in \mathbb{R}^d, \tag{3}$$

where B_t is a d_1-dimensional Brownian motion, b is a locally bounded Borel vector function of dimension d, σ is a $d \times d_1$ matrix-valued uniformly continuous function, $d_1 \geq d$. We assume that $\sigma\sigma^*$ is bounded and nondegenerate. Let λ_-, λ_+ and Λ be the best constants such that for any $x \in \mathbb{R}^d \backslash \{0\}$,

$$(A_\sigma) \qquad 0 < \lambda_- \leq (\sigma\sigma^*(x)x/|x|, x/|x|) \leq \lambda_+, \qquad \operatorname{Tr} \sigma\sigma^*(x)/d \leq \Lambda.$$

Notice that $\lambda_\pm$ are not exactly the constants of the upper and lower bounds from the nondegeneracy condition and that we will use all these constants as well as the nondegeneracy condition for σ.

Let us introduce the following family of recurrence conditions :

$$(A_b) \qquad (b(x), x/|x|) \le -r|x|^\alpha, \qquad |x| \ge M_0,$$

with $M_0 \ge 0$, $\alpha \ge -1$ and $r > 0$ such that in the case $\alpha = -1$, $r > (3\lambda_+ - \lambda_- + \Lambda d)/2$. Note that this condition prevents the solution of the SDE (3) from exploding, so that the process $\{X_t\}$ is well-defined for all $t > 0$. The case $\alpha = 1$ includes the case of the Ornstein-Uhlenbeck process. Assumption (A_b) in the case $\alpha = 0$ is usually called Has'minski's assumption. Define $r_0 = [r - (\Lambda d - \lambda_-)/2]\lambda_+^{-1}$ in the case $\alpha = -1$, and $r_0 = \infty$ if $\alpha > -1$. The value r_0 plays an important role in the case $\alpha = -1$; namely, this constant must be greater than 3/2, which is implied by condition (A_b). Let $R > 0$ and $\tau = \tau^R = \inf(t \ge 0\colon |X_t| \le R)$. Let also $\kappa(x) = |\sigma^*(x)x|/|x|$, $h(t) = \int_0^t \kappa^2(X_s)\,ds$, $z(t) = h^{-1}(t)$, $\widetilde{X}_t = X_{z(t)}$. Then

$$d\widetilde{X}_t = \kappa^{-2}(\widetilde{X}_t)b(\widetilde{X}_t)\,dt + \kappa^{-1}(\widetilde{X}_t)\sigma(\widetilde{X}_t)\,d\widetilde{B}_t,$$

with a new d_1-dimensional Brownian motion $\widetilde{B}$; compare Ikeda and Watanabe (1981), page 102.

PROPOSITION 1. *Under the assumptions* (A_σ) *and* (A_b), *for any* $0 < m < 2r_0 - 1$, $t \ge 0$,

$$E_x|\widetilde{X}_t|^m \le C(1 + |x|^m), \tag{4}$$

$$E_\mu|X_t|^m = C < \infty, \tag{5}$$

and for any $2k + 2 < m < 2r_0 - 1$, $k > 0$,

$$\operatorname{var}(\mu_t^x - \mu) \le C(1 + |x|^m)(1 + t)^{-(k+1)}, \tag{6}$$

where "var" *denotes the total variation norm of a signed measure over the Borel sigma-field*, μ_t^x *is the law of* X_t *when* $X_0 = x$, μ *is the unique invariant measure of* X *and* E_μ *means the expectation w.r.t.* μ. *Moreover*, $\widetilde{X}$ *also possesses an invariant measure* $\tilde{\mu}$ *which satisfies the equation* $\tilde{\mu}(dx) = c^{-1}\,\kappa^2(x)\mu(dx)$ *with* $c = \int \kappa^2(x)\mu(dx)$. *If* $\alpha \ge 0$ *then assertions* (4)–(6) *may be strengthened to the following exponential inequalities*: *there exist constants* $C > 0$, $\nu > 0$ *and* $\lambda > 0$ *such that for any* $t \ge 0$,

$$E_x \exp(\nu|X_t|) \le C\exp(\nu|x|), \tag{7}$$

$$E_\mu \exp(\nu|X_t|) < \infty \tag{8}$$

and

$$\operatorname{var}(\mu_t^x - \mu) \le C\exp(\nu|x|)\exp(-\lambda t). \tag{9}$$

This statement has been proved in case $\alpha = -1$ in Veretennikov (1997), where continuity of σ only is used instead of uniform continuity which is assumed in this paper and is required for easier references to the PDE literature. The case $\alpha > -1$ follows from the statement for $\alpha = -1$. The case $\alpha \geq 0$ [i.e., assertions (7)–(9)] may be found in Veretennikov (1987). The relationship between the invariant measures for X and $\widetilde{X}$ is well known.

REMARK. In the case $d_1 = d$, $\sigma \equiv I$ (the unit $d \times d$ matrix) the lower bound for r_0 is 3/2. It is only for $r_0 \geq 3/2$ that the bounds for the convergence rate to the stationary regime are known, while in the case $r_0 \leq 1/2$ the process X may not possess an invariant probability distribution. The above proposition states that in general 3/2 is an upper bound for the critical value of r_0.

PROPOSITION 2. *Let the assumptions* (A_σ) *and* (A_b) *be satisfied. Then for any* $0 < p < \frac{r_0}{2} + \frac{1}{4}$,

$$E_x\Big(\sup_{0\leq t'\leq t} |X_{t'}|^p\Big) = \circ(\sqrt{t}) \quad \text{as } t \to \infty,$$

$$E_x\Big(\sup_{0\leq t'\leq t} |\widetilde{X}_{t'}|^p\Big) = \circ(\sqrt{t}) \quad \text{as } t \to \infty.$$

PROOF. The two assertions are clearly equivalent. Also we shall later use the second inequality; we prove now the first one. Let $1 \leq p < \frac{r_0}{2} + \frac{1}{4}$. From Itô's formula, we deduce

$$\begin{aligned} d|X_t|^{2p} &= 2p|X_t|^{2p-2}\sum_{i,k} X_t^k \sigma_{k,i}(X_t)\,dB_t^i + 2p|X_t|^{2p-2}\langle X_t, b(X_t)\rangle\,dt \\ &\quad + p|X_t|^{2p-2}\Bigg[\sum_{i,k}\sigma_{k,i}^2(X_t) + 2(p-1)\sum_i\Bigg(\sum_k |X_t|^{-1}X_t^k\sigma_{k,i}(X_t)\Bigg)^2\Bigg]dt. \end{aligned}$$

From the Burkholder–Davis–Gundy inequality, there exists C such that

$$E_x\Big(\sup_{0\leq t'\leq t}|X_{t'}|^{2p}\Big) \leq C|x|^{2p} + C\Big(E_x\int_0^t |X_s|^{4p-2}\,ds\Big)^{1/2} + CE_x\int_0^t |X_s|^{2p-2}\,ds.$$

Since $2p-2 < 4p-2 < 2r_0-1$, we can use estimate (4) (for the time–changed process). Namely, for any $q < 2r_0 - 1$,

$$\begin{aligned} E_x\int_0^t |X_s|^q\,ds &\leq CE_x\int_0^{Ct} |\widetilde{X}_s|^q\,ds \\ &\leq C't(1+|x|^q). \end{aligned}$$

Hence, we get

$$E_x\Big(\sup_{t'\leq t}|X_{t'}|^{2p}\Big) \leq C|x|^{2p} + C(t+t^{1/2})(1+|x|^{2p-1}),$$

and moreover, as $t \geq 1$,

$$E_x\Big(\sup_{t'\leq t}|X_{t'}|^p\Big) \leq C|x|^p + C(1+|x|^{p-1/2})\sqrt{t}.$$

Hence for any $p' < p$, $t \geq 1$,

$$E_x\Big(\sup_{t'\leq t}|X_{t'}|^{p'}\Big) \leq C|x|^{p'} + C(1+|x|^{p-1/2})t^{p'/2p}.$$

The result follows. □

One deduces the following corollary.

COROLLARY 1. *Under the same assumptions, for any* $T > 0$, $0 < p < \frac{r_0}{2} + \frac{1}{4}$,

$$\varepsilon E_x\Big(\sup_{0\leq t\leq T}|X_{t/\varepsilon^2}|^p\Big) \to 0 \quad \textit{as } \varepsilon \to 0.$$

3. The Poisson equation in $\mathbb{R}^d$. We consider the Poisson equation in $\mathbb{R}^d$,

$$Lu(x) = -f(x), \tag{10}$$

where

$$L = \sum a_{ij}(x)\partial_{x_i}\partial_{x_j} + \sum b_i(x)\partial_{x_i},$$

with

$$a(x) = \sigma\sigma^*(x)/2.$$

The problem under consideration is to describe the class of functions f such that (10) has a solution either bounded or slowly increasing at infinity. Results of this type were previously obtained for the compact space case [cf. Revuz (1984)]. For the case $\alpha = 0$ under some additional assumptions (for the dual process), see Bouc and Pardoux (1984). For results concerning some functional properties of the operator L, see Ethier and Kurtz (1986).

Now we are going to show that even rather weak recurrence assumptions with $\alpha = -1$ for the process X, plus certain assumptions on f, imply the existence of a bounded or slowly increasing solution u of (10).

It is well-known [cf. Dynkin (1965)] that the solution of (10) in a *bounded domain* D with a smooth boundary and a zero boundary condition (Dirichlet problem) has the representation

$$u(x) = E_x\int_0^\gamma f(X_s)\,ds, \qquad \gamma = \inf(t > 0 : X_t \notin D).$$

Similarly, one expects that the solution of (10) in $\mathbb{R}^d$ has the stochastic representation

$$u(x) = \int_0^\infty E_x f(X_s)\,ds, \tag{11}$$

provided that this solution exists and does not increase too rapidly at infinity. Conversely, if the function $u(x)$ given by (11) is continuous then it possesses a certain additional smoothness, belongs to some Sobolev class and is a solution of (10). In other words, the main problem is to show that the function u in (11) is continuous and does not increase too fast at infinity.

Concerning the function f, we will assume that

$$|f(x)| \le C(1+|x|^{\beta}) \quad \text{for some } \beta < 2r_0 - 3,$$

so that due to Proposition 1, f is integrable with respect to the invariant measure μ. We assume moreover that

$$(A_f) \qquad \int f(x)\mu(dx) = 0;$$

otherwise one cannot hope to get a finite value in (11) for a positive recurrent process X. Define

$$\tilde{u}(x) = \int_0^{\infty} |E_x f(X_t)|\, dt.$$

THEOREM 1. *Let the assumptions* (A_σ), (A_b) *be satisfied. We assume that there exists* $0 \le \beta < 2r_0 - 3$ *such that* $|f(x)| \le C_1 + C_2|x|^{\beta}$ *with* $C_1, C_2 \ge 1$ *and that* (A_f) *holds true. Then* (11) *defines a continuous function* u *which belongs to the Sobolev class* $W^2_{p,\text{loc}}$ *for any* $p > 1$, *is a solution of* (10) *and satisfies the following properties. For any* $m > \beta + 2$, *there exists* C_m *which depends only on* m, β, r_0, *the ellipticity constants, the modulus of continuity of the matrix function* $(a_{ij}(\cdot))$, *the value* $\sup_{i,x} |b_i(x)|$ *and on the constants* C *in* (4) *and* (5), *such that*

$$(12) \qquad |u(x)| \le \tilde{u}(x) \le C_m(C_1 + C_2 + C_2|x|^m), \qquad x \in \mathbb{R}^d,$$

so that in particular u *is* μ*–integrable. Moreover, again for any* $m > \beta + 2$,

$$(13) \qquad \sup_x (1+|x|^m)^{-1} \left| u(x) - \int_0^N E_x f(X_t)\, dt \right| \to 0, \qquad \text{as } N \to \infty.$$

In addition, u *is centered in the sense that*

$$(14) \qquad \int u(x)\mu(dx) = 0.$$

The solution is unique in the class of u *which belong to* $W^2_{p,\text{loc}}$ *for any* $p > 1$ *and satisfy properties* (12) *and* (14).

Finally, for any $m > \beta + 2$, *there exists* C_m *as above such that*

$$(15) \qquad |\nabla u(x)| \le C_m(C_1 + C_2 + C_2|x|^m), \qquad x \in \mathbb{R}^d.$$

THEOREM 2. *Let the assumptions of Theorem* 1 *be in force*:

(i) *If there exists* C *such that*

$$(16) \qquad |f(x)| \le C(1+|x|)^{\beta+\alpha-1}$$

for some $\beta < 0$, *then* u *is bounded*. *Moreover*,

$$\sup_x |u(x)| \le C \sup_x [|f(x)|(1+|x|)^{-\beta-\alpha+1}], \tag{17}$$

where the constant C *depends only on the constants* C, m, k *from* (4)–(6) *in Proposition* 1, *and*

$$|\nabla u(x)| \le C\big(1+|x|^{(\beta+\alpha-1)^+}\big). \tag{18}$$

(ii) *If there exists* C *such that for some* $\beta > 0$,

$$|f(x)| \le C(1+|x|)^{\beta+\alpha-1} \tag{19}$$

and also, *whenever* $\alpha = -1$ *and* $\beta > 4$, *the constant* r *in* (A_b) *satisfies* $2r > \Lambda d + (\beta-2)\lambda_+$, *then there exists* C' *such that*

$$|u(x)| \le C'(1+|x|)^{\beta}.$$

Moreover,

$$\sup_x \frac{|u(x)|}{1+|x|^{\beta}} \le C'' \sup_x \frac{|f(x)|}{1+|x|^{\beta+\alpha-1}}, \tag{20}$$

where the constant C'' *only depends on the constants* C, m, k *from* (4)–(6) *in Proposition* 1. *Finally there exists* C *such that*

$$|\nabla u(x)| \le C\big(1+|x|^{(\beta+\alpha-1)^+}+|x|^{\beta}\big). \tag{21}$$

Theorem 1 and Theorem 2 are not comparable. The assertion of Theorem 1 is used in Theorem 2 which means that the latter theorem gives additional information under additional assumptions. Theorem 2 gives a criterion for u and ∇u to be bounded.

PROOF OF THEOREM 1. (a) u is well defined and satisfies (12). This follows from Veretennikov (1997); see Proposition 1. Indeed,

$$\tilde{u}(x) = \int_0^\infty |E_x f(X_t)|\,dt = \int_0^\infty \left|\int f(y)\mu_t^x(dy)\right| dt$$
$$= \int_0^\infty \left|\int f(y)[\mu_t^x(dy)-\mu(dy)]\right| dt.$$

Without loss of generality, we assume that $\beta+2 < m < 2r_0 - 1$. Due to the inequalities in Proposition 1, one can choose $p > 1$, $q > 1$ with $p^{-1}+q^{-1} = 1$, such that $p\beta \le m$ and $(k+1)/q > 1$.

Indeed, if $\beta = 0$, then it is evident. Consider the case $\beta > 0$. Let $p = m/\beta$. Then $q^{-1} = 1-\beta/m$, and $(k+1)/q > 1$ is equivalent to $(k+1)(1-\beta/m) > 1$. Since $k+1$ is an arbitrary number less than $m/2$ then the last inequality can be satisfied if $(m/2)(1-\beta/m) > 1$ which is equivalent to $m > \beta+2$ and this is our assumption.

Now, using Hölder's inequality, and denoting all new constants which do not depend on C_1 and C_2 by C_0 (they may be different on each line), one has

$$\begin{aligned}
&\int_0^\infty \left|\int f(y)[\mu_t^x(dy)-\mu(dy)]\right| dt\\
&\quad\le \int_0^\infty \Big(\int |f(y)|^p[\mu_t^x(dy)+\mu(dy)]\Big)^{1/p}\Big(\int |\mu_t^x-\mu|(dy)\Big)^{1/q}\,dt\\
&\quad\le C_0\int_0^\infty \Big(\int (C_1+C_2|y|^m)[\mu_t^x(dy)+\mu(dy)]\Big)^{1/p}(\operatorname{var}(\mu_t^x-\mu))^{1/q}\,dt\\
&\quad\le C_0\int_0^\infty (2C_1+C_2E_x|X_t|^m+C_2E_\mu|X_t|^m)^{1/p}\\
&\qquad\qquad\times((1+|x|^m)(1+t)^{-(k+1)})^{1/q}\,dt\\
&\quad\le C_0(1+|x|^m)^{1/q}\int_0^\infty (C_2E_x|X_t|^m+2C_1+C_0C_2)^{1/p}(1+t)^{-(k+1)/q}\,dt\\
&\quad\le C_0(1+|x|^m)^{1/q}\int_0^\infty (C_2E_x|\widetilde{X}_t|^m)^{1/p}(1+t)^{-(k+1)/q}\,dt\\
&\qquad +C_0(1+|x|^m)^{1/q}(2C_1+C_0C_2)^{1/p}\\
&\quad\le C_0(1+|x|^m)^{1/q}\int_0^\infty (C_2(1+|x|^m))^{1/p}(1+t)^{-(k+1)/q}\,dt\\
&\qquad +C_0(1+|x|^m)^{1/q}(C_1+C_2)^{1/p}\\
&\quad\le C_0(1+|x|^m)^{1/q}[(C_1+C_2)^{1/p}+(C_2(1+|x|^m))^{1/p}]\\
&\quad\le C_0[(1+|x|^m)C_2^{1/p}+(1+|x|^m)^{1/q}(C_1+C_2)^{1/p}]\\
&\quad\le C_0[C_1+C_2+(1+C_2^{1/p})(1+|x|^m)]\\
&\quad\le C_0[C_1+C_2+C_2(1+|x|^m)].
\end{aligned}$$

Thus, u is locally bounded and, moreover, (12) holds true with any $m>\beta+2$. The assertion (13) follows from the same calculations with $\int_N^\infty$ instead of $\int_0^\infty$.

(b) u satisfies (14). Notice that if some function g is integrable w.r.t. the invariant measure μ then for any $s>0$,

$$\int E_x[g(X_s)]\mu(dx)=\int g(x)\mu(dx).$$

Due to (12), the function $\tilde{u}$ is μ-integrable. So, by virtue of Fubini's theorem,

$$\int\int_0^\infty E_xf(X_s)\,ds\,\mu(dx)=\int_0^\infty\int E_xf(X_s)\mu(dx)\,ds.$$

But clearly,

$$\int E_xf(X_s)\mu(dx)=\int f(x)\mu(dx)=0.$$

(c) u is continuous. For each fixed $t > 0$, $\{z^t(s, x) := E_x \int_s^t f(X_r)\, dr;\ 0 \le s \le t,\ x \in \mathbb{R}^d\}$ is a generalized solution of the parabolic equation

$$\frac{\partial z}{\partial s}(s, x) + Lz(s, x) = -f(x), \qquad 0 \le s \le t,\ x \in \mathbb{R}^d,$$

$$z(t, x) = 0,\ x \in \mathbb{R}^d.$$

Here z is locally uniform continuous. This is deduced from the $W^{1,2}_{p,\mathrm{loc}}$-regularity of the solution for all $p > 1$ [see Ladyzenskaja, Solonnikov and Ural'ceva (1968), Chapter 4, Veretennikov (1982)]. Moreover $z^t(0, x)$ converges to $u(x)$ as $t \to \infty$, locally uniformly in x, due to (13).

(d) u belongs to Sobolev classes (locally). Consider any ball D and the Dirichlet problem

$$L\hat{u}(x) = -f(x), \qquad x \in D,\ \hat{u}|_{\partial D} = u.$$

This equation has a unique solution $\hat{u} \in W^2_{p,\mathrm{loc}}(D) \cap C(\overline{D})$ for any $p > 1$; see Gilbarg and Trudinger (1983), Corollary 9.18. We can then apply Itô–Krylov's formula to $\hat{u}(X_t)$ on the random interval $[0, \gamma]$, where $\gamma := \inf(s \ge 0\colon X_s \notin D)$ [see Krylov (1980), Theorem 2.10.1]. One deduces that

$$\hat{u}(x) = E_x u(X_\gamma) + E_x \int_0^\gamma f(X_s)\, ds.$$

On the other hand, the function u satisfies the same representation inside D because of the strong Markov property of X_t. Indeed, let $x \in D$. One has

$$u(x) = \int_0^\infty E_x f(X_s) I(s \le \gamma)\, ds + \int_0^\infty E_x f(X_s) I(s > \gamma)\, ds,$$

where both integrals are well-defined. Indeed since f is bounded on $\overline{D}$ and $\sup_x E_x \gamma < \infty$,

$$E_x \int_0^\infty |f(X_s)| I(s \le \gamma)\, ds < \infty.$$

Moreover,

$$\begin{aligned}\int_0^\infty E_x f(X_s) I(s > \gamma)\, ds &= \int_0^\infty E_x E_{X_\gamma} f(X_s)\, ds \\ &= \lim_{N\to\infty} \int_0^N E_x E_{X_\gamma} f(X_s)\, ds \\ &= \lim_{N\to\infty} E_x \int_0^N E_{X_\gamma} f(X_s)\, ds \\ &\equiv \lim_{N\to\infty} E_x u^N(X_\gamma) \\ &= E_x u(X_\gamma),\end{aligned}$$

where $u^N(x) := \int_0^N E_x f(X_t)\, dt$. Hence, we get the desired representation, $u(x) = \hat{u}(x)$, $x \in D$.

(e) u satisfies (15). We shall write $B_{x,R}$ for the ball in $\mathbb{R}^d$ centered at x with radius R. Using successively the Sobolev embedding theorem [see Ladyzenskaja, Solonnikov and Ural'ceva (1968), Theorem 2.2.1], the a priori estimate (9.40) from Gilbarg and Trudinger (1983), the assumption on f and the inequality (12), for any $p > d$, we obtain that for some $C, C', C_m > 0$ and all $x \in \mathbb{R}^d$,

$$|\nabla u(x)| \le C\|u\|_{W_p^2(B_{x,1})} \le C'(\|u\|_{L_p(B_{x,2})} + \|Lu\|_{L_p(B_{x,2})})$$
$$\le C_m(C_1 + C_2 + C_2|x|^m + C_1 + C_2|x|^\beta).$$

(f) Uniqueness. For the difference of two solutions, $v = u - u'$, we have $Lv = 0$. So, due to Itô–Krylov's formula for functions in $W^2_{p,\mathrm{loc}}$, $\forall\ p > d$,

$$v(x) = E_x v(X_t) \to E_\mu(v(X_t)) = 0, \qquad t \to \infty.$$

Hence, $v(x) \equiv 0$. □

PROOF OF THEOREM 2. We shall prove the boundedness of u and its moderate growth. (17) and (20) follow from our proof. Inequalities (18) and (21) follow from the argument in part (e) of the proof of Theorem 1.

(a) u is bounded. We will use the representation

$$u(x) = E_x u(X_{\tau^R}) + E_x \int_0^{\tau^R} f(X_t)\, dt$$

which follows from the strong Markov property of the process. In view of (12), the first term in the above right-hand side is bounded. We now prove that the second term is bounded. We assume that for some $\beta < 0$, $R > 0$ there exists $C > 0$ such that

$$|f(x)| \le C|x|^{\beta+\alpha-1}, \qquad |x| \ge R.$$

Consider $\{X_t^x,\ 0 \le t \le \tau^R\}$, for $|x| > R$, where $R \ge M_0$ [see assumption (A_b)],

$$E_x|X_{t\wedge\tau^R}|^\beta = |x|^\beta + \beta E_x \int_0^{t\wedge\tau^R} |X_s|^{\beta-2}$$
$$\times\left(\langle X_s, b(X_s)\rangle + \frac{\beta-2}{2}\frac{\langle a(X_s)X_s, X_s\rangle}{|X_s|^2} + \frac{\mathrm{Tr}\, a(X_s)}{2}\right) ds$$
$$\ge \beta E_x \int_0^{t\wedge\tau^R} |X_s|^{\beta-2}(-r|X_s|^{\alpha+1} + \Lambda d/2)\, ds.$$

In the case $\alpha > -1$, the last expression is

$$\ge |\beta| r' E_x \int_0^{t\wedge\tau^R} |X_s|^{\beta+\alpha-1}\, ds,$$

provided $0 < r' < r$, R is large enough s.t. $(r - r')R^{\alpha+1} \ge \Lambda d/2$. Then one gets, as $t \to \infty$, from monotone and bounded convergence

$$E_x \int_0^{\tau^R} |X_s|^{\beta+\alpha-1}\, ds \le c E_x |X_{\tau^R}|^\beta = cR^\beta. \qquad □$$

In the case $\alpha = -1$, we have

$$E_x|X_{t\wedge\tau^R}|^\beta \geq |\beta|\left(r - \frac{\Lambda d + (\beta-2)\lambda_-}{2}\right)E_x\int_0^{t\wedge\tau^R}|X_s|^{\beta-2}\,ds$$

and in the limit as $t \to \infty$,

$$E_x\int_0^{\tau^R}|X_s|^{\beta-2}\,ds \leq CR^\beta$$

provided $r > (\Lambda d + (\beta-2)\lambda_-)/2$, which is implied by our standing assumption $r_0 > 3/2$. □

REMARK 1. In the case $\alpha = 1$, we have obtained the following result: if there exists $\beta < 0$ s.t. $|f(x)| \leq C(1+|x|)^\beta$, $|x| > R$ then u is bounded. This result is optimal in the sense that f bounded does not imply u bounded. Indeed, in the case $dX_t = -X_t\,dt + \sqrt{2}\,dB_t$, $f(x) = \operatorname{sign}(x)$, one has $|u(x)| \geq c\log\sqrt{1+x^2}$.

(b) u grows moderately. Let $\rho = \Lambda d/2 + (\beta-2)\lambda_+ I(\beta > 2)/2 + (\beta-2)\lambda_- I \times (\beta < 2)/2$. We start with the same computation, but this time with $\beta > 0$,

$$E_x|X_{t\wedge\tau^R}|^\beta = |x|^\beta + \beta E_x\int_0^{t\wedge\tau^R}|X_s|^{\beta-2} \times\left(\langle X_s, b(X_s)\rangle + \frac{\beta-2}{2}\frac{\langle a(X_s)X_s, X_s\rangle}{|X_s|^2} + \operatorname{Tr}a(X_s)/2\right)ds.$$

Hence,

$$0 \leq |x|^\beta + \beta E_x\int_0^{t\wedge\tau^R}|X_s|^{\beta-2}(-r|X_s|^{\alpha+1} + \rho)\,ds$$

or

$$E_x\int_0^{t\wedge\tau^R}|X_s|^{\beta-2}(r|X_s|^{\alpha+1} - \rho)\,ds \leq \beta^{-1}|x|^\beta.$$

Note that unless $d=1$, clearly $\rho > 0$, since $\Lambda \geq \lambda_\pm$, $\beta > 0$.

In the case $\alpha > -1$, choose R large enough such that $R^{1+\alpha} \geq 2\rho^+/r$, so that the above inequality yields

$$\frac{r}{2}E_x\int_0^{\tau^R}|X_s|^{\beta-1+\alpha}\,ds \leq \beta^{-1}|x|^\beta.$$

With the case $\alpha = -1$, we conclude, since $r > \rho$, which is the case due to the additional assumption if $\beta > 4$, and follows from the standing assumption $r_0 > 3/2$ if $\beta \leq 4$.

4. Diffusion approximation. Now we are going to apply Theorem 1 to the singularly perturbed SDE,

$$
\begin{aligned}
dX_t^\varepsilon &= \varepsilon^{-2} b(X_t^\varepsilon)\,dt + \varepsilon^{-1}\sigma(X_t^\varepsilon)\,dB_t^\varepsilon, \qquad X_0^\varepsilon = x,\\
dY_t^\varepsilon &= F(X_t^\varepsilon, Y_t^\varepsilon)\,dt + \varepsilon^{-1} G(X_t^\varepsilon, Y_t^\varepsilon)\,dt \\
&\quad + H(X_t^\varepsilon, Y_t^\varepsilon)\,dB_t^\varepsilon, \qquad Y_0^\varepsilon = y, \qquad 0 \le t \le T.
\end{aligned} \tag{22}
$$

Here ε is a small parameter, X_t^ε may be regarded as X_{t/ε^2}, where $X \equiv X^1$ with some Brownian motion depending on ε. The process X is the same as that of the previous sections, and we will again assume the same nondegeneracy and recurrence conditions (A_b) and (A_σ) with some $\alpha \ge -1$. F, G and H are Borel, locally bounded vector-functions. The dimension of X is again d, the dimension of Y is l. Notice that under our assumptions there exists at least a weak solution $(X^\varepsilon, B^\varepsilon)$ of the first equation in (22). That first equation may be solved independently of the second one. As above, we denote by L the generator of the process X. The problem we are interested in is the weak convergence of the slow component Y^ε as $\varepsilon \to 0$. Concerning the second equation in (22), we require the Lipschitz condition with respect to the variable y, with a constant depending on x:

$$
\begin{aligned}
&|F(x, y) - F(x, y')| + |G(x, y) - G(x, y')| \\
&\qquad + \|H(x, y) - H(x, y')\| \le C(x)|y - y'|.
\end{aligned} \tag{A_L}
$$

Note that (A_b), (A_σ) and (A_L) insure that the system (22) of SDEs is well-posed. We now assume that for all $x \in \mathbb{R}^d$, $G(x, \cdot) \in C^2(\mathbb{R}^l; \mathbb{R}^l)$, that $\partial_y^2 G \in C(\mathbb{R}^{d+l}; \mathbb{R}^{l^3})$ and the functions F, G, H satisfy the following polynomial growth conditions:

$$
\begin{aligned}
|F(x, y)| &\le K(1 + |y|)(1 + |x|^{q_1});\\
\|H(x, y)\| &\le K(1 + |y|^{1/2})(1 + |x|^{q_2});\\
|G(x, y)| &\le K(1 + |y|)(1 + |x|^{q_3});\\
\|\nabla_y G(x, y)\| &\le K(1 + |x|^{q_4});\\
\|\partial_y^2 G(x, y)\| &\le K(1 + |x|^{q_5}).
\end{aligned} \tag{A_P}
$$

We assume moreover that for all $y \in \mathbb{R}^l$ and $j = 1, 2, \dots, l$,

$$
\int G_j(x, y)\mu(dx) = 0, \tag{A_G}
$$

where $\mu(dx)$ again denotes the (unique) invariant measure of X. It then follows from Theorem 1 that the Poisson equations

$$
L\overline{G}_j(x, y) = -G_j(x, y), \qquad j = 1, \dots, l,
$$

have unique centered solutions

$$
\overline{G}_j(x, y) = \int_0^\infty E_x G_j(X_t^1, y)\,dt.
$$

Moreover, for some K and q_3', q_4', q_5', the following holds:

$$|\overline{G}(x,y)| \le K(1+|y|)(1+|x|^{q_3'});$$
$$\|\nabla_y \overline{G}(x,y)\| \le K(1+|x|^{q_4'});$$
$$\|\partial_y^2 \overline{G}(x,y)\| \le K(1+|x|^{q_5'}),$$
$$|\partial_x \overline{G}(x,y)| \le K(1+|y|)(1+|x|^{q_3'}),$$
$$|\partial_x \partial_y \overline{G}(x,y)| \le K(1+|x|^{q_4'}),$$
$$\|\partial_x \partial_y^2 \overline{G}(x,y)\| \le K(1+|x|^{q_5'}).$$

The values of q_3', q_4' and q_5' can be deduced from those of q_3, q_4 and q_5 by using Theorem 1 or Theorem 2. We assume that the q_i's are nonnegative and such that, with $r_1 := 2r_0 - 1$,

$$(A_q) \qquad \max[q_1, 2q_2, 2q_3', q_2 + q_3', q_3 + q_3', q_3 + q_4'] < r_1.$$

The above conditions would take a slightly different form, in case certain q_i's were negative.

THEOREM 3. *Let (A_b), (A_σ), (A_L), (A_P), (A_G) and (A_q) be satisfied. Then for any $T > 0$, the family of processes $\{Y_t^\varepsilon, 0 \le t \le T\}_{0<\varepsilon\le 1}$ is weakly relatively compact in $C([0,T];\mathbb{R}^l)$. Any accumulation point Y is a solution of the martingale problem associated to the operator*

$$\mathscr{L} = \sum \bar{a}_{ij}(y)\partial_{y_i}\partial_{y_j} + \sum \bar{b}_i(y)\partial_{y_i},$$

where

$$\bar{b}(y) = \overline{F}(y) + \sum_i \int G_i(x,y)\partial_{y_i}\overline{G}(x,y)\mu(dx)$$
$$+ \sum_{i,k} \int (H\sigma^*)_{ik}(x,y)\partial_{x_k}\partial_{y_i}\overline{G}(x,y)\mu(dx)$$

and

$$\bar{a}(y) = 2(\overline{\mathscr{H}} + \overline{\mathscr{G}} + \overline{\mathscr{K}})(y),$$

with

$$\overline{F}(y) = \int F(x,y)\mu(dx),$$
$$\overline{\mathscr{H}}(y) = \int HH^*(x,y)\mu(dx),$$
$$\overline{\mathscr{G}}(y) = \int [G(x,y)\overline{G}^*(x,y) + \overline{G}(x,y)G^*(x,y)]\mu(dx),$$
$$\bar{\mathscr{K}}(y) = \int [(H\sigma^*)_{ik}(x,y)\partial_{x_k}\overline{G}_j(x,y)$$
$$+ (H\sigma^*)_{jk}(x,y)\partial_{x_k}\overline{G}_i(x,y)]\mu(dx).$$

If moreover the martingale problem associated to $\mathscr{L}$ is well-posed (it is easy to state sufficient conditions for that), then $Y^\varepsilon \Rightarrow Y$, and Y is the unique (in law) diffusion process with generator $\mathscr{L}$.

Notice that all integrals in the definition of $\mathscr{L}$ are well-defined, as follows from (A_q) and Proposition 1.

PROOF. Let $f \in C_p^3(\mathbb{R}^l)$ (the set of functions of class C^3 which, together with their partial derivatives of order 1, 2 and 3, have at most polynomial growth of some order) and define

$$f^\varepsilon(x, y) = f(y) + \varepsilon u(x, y),$$

where $\varepsilon u(x, y)$ is a corrector to f; that is, u is the solution of the Poisson equation

$$Lu(x, y) = -\langle \nabla_y f(y), G(x, y)\rangle,$$

or in other words,

$$u(x, y) = \langle \nabla_y f(y), \overline{G}(x, y)\rangle, \tag{23}$$

where $\overline{G}$: $\mathbb{R}^d \times \mathbb{R}^l \to \mathbb{R}^l$ solves

$$L\overline{G}(x, y) = -G(x, y).$$

Note that

$$\int \partial_y G(x, y)\mu(dx) = 0, \qquad y \in \mathbb{R}^l,$$

and

$$\partial_y \overline{G}(x, y) = \overline{(\partial_y G(x, y))}.$$

From Itô–Krylov's formula for functions with Sobolev derivatives,

$$\begin{aligned} f^\varepsilon(X_t^\varepsilon, Y_t^\varepsilon) - f^\varepsilon(x, y) &= \int_0^t \langle \nabla_y f(Y_s^\varepsilon), F(X_s^\varepsilon, Y_s^\varepsilon) + \varepsilon^{-1} G(X_s^\varepsilon, Y_s^\varepsilon)\rangle\, ds \\ &\quad + \int_0^t \langle \nabla_y f(Y_s^\varepsilon), H(X_s^\varepsilon, Y_s^\varepsilon)\, dB_s^\varepsilon\rangle \\ &\quad + \int_0^t (1/2)\mathrm{Tr}\, \partial_y^2 f(Y_s^\varepsilon)(HH^*)(X_s^\varepsilon, Y_s^\varepsilon)\, ds \\ &\quad + \int_0^t \varepsilon\langle \nabla_y u(X_s^\varepsilon, Y_s^\varepsilon), F(X_s^\varepsilon, Y_s^\varepsilon) + \varepsilon^{-1} G(X_s^\varepsilon, Y_s^\varepsilon)\rangle\, ds \\ &\quad + \varepsilon \int_0^t \langle \nabla_y u(X_s^\varepsilon, Y_s^\varepsilon), H(X_s^\varepsilon, Y_s^\varepsilon)\, dB_s^\varepsilon\rangle \\ &\quad + \varepsilon\,(1/2)\mathrm{Tr} \int_0^t \partial_y^2 u(X_s^\varepsilon, Y_s^\varepsilon)(HH^*)(X_s^\varepsilon, Y_s^\varepsilon)\, ds \end{aligned} \tag{24}$$

$$+\varepsilon^{-1}\int_0^t Lu(X_s^\varepsilon, Y_s^\varepsilon)\,ds$$
$$+\int_0^t \langle \nabla_x u(X_s^\varepsilon, Y_s^\varepsilon), \sigma(X_s^\varepsilon)\,dB_s^\varepsilon\rangle$$
$$+\sum_{i,k}\int_0^t \partial_{y_i}\partial_{x_k} u(X_s^\varepsilon, Y_s^\varepsilon)(H\sigma^*)_{ik}(X_s^\varepsilon, Y_s^\varepsilon)\,ds.$$

If u were more regular, the above would just be the usual Itô formula. Formula (24) will be justified in the Appendix below. Loosely speaking, we use the fact that u, $\partial_{x_i}u$ and $\partial_{x_i}\partial_{x_j}u$ are continuous in y.

From the definition of u, the sum of the terms of order ε^{-1} vanishes. We then obtain

$$\begin{aligned}
f(Y_t^\varepsilon) = f(Y_{t_0}^\varepsilon) &+ \int_{t_0}^t \Big\langle \nabla_y f(Y_s^\varepsilon), F(X_s^\varepsilon, Y_s^\varepsilon) + \sum_i G_i(X_s^\varepsilon, Y_s^\varepsilon)\partial_{y_i}\overline{G}(X_s^\varepsilon, Y_s^\varepsilon) \\
&\qquad + \sum_{i,k}(H\sigma^*)_{ik}(X_s^\varepsilon, Y_s^\varepsilon)\partial_{y_i}\partial_{x_k}\overline{G}(X_s^\varepsilon, Y_s^\varepsilon)\Big\rangle\,ds \\
&+ \frac{1}{2}\int_{t_0}^t \sum_{i,j}\partial_{y_i}\partial_{y_j} f(Y_s^\varepsilon)\Big[(HH^*)_{ij}(X_s^\varepsilon, Y_s^\varepsilon) + 2(G_i\overline{G}_j)(X_s^\varepsilon, Y_s^\varepsilon) \\
&\qquad + 2\sum_k (H\sigma^*)_{ik}(X_s^\varepsilon, Y_s^\varepsilon)\partial_{x_k}\overline{G}_j(X_s^\varepsilon, Y_s^\varepsilon)\Big]\,ds \\
&+ \int_{t_0}^t \langle \nabla_y f(Y_s^\varepsilon), [H(X_s^\varepsilon, Y_s^\varepsilon) + \nabla_x\overline{G}(X_s^\varepsilon, Y_s^\varepsilon)\sigma(X_s^\varepsilon)]\,dB_s^\varepsilon\rangle \\
&+ \varepsilon R_f^\varepsilon(t_0, t),
\end{aligned}\tag{25}$$

where

$$\begin{aligned}
R_f^\varepsilon(t_0, t) = u(X_{t_0}^\varepsilon, Y_{t_0}^\varepsilon) &- u(X_t^\varepsilon, Y_t^\varepsilon) \\
&+ \frac{1}{2}\operatorname{Tr}\int_{t_0}^t \partial_y^2 u(X_s^\varepsilon, Y_s^\varepsilon)(HH^*)(X_s^\varepsilon, Y_s^\varepsilon)\,ds \\
&+ \int_{t_0}^t \langle \nabla_y u(X_s^\varepsilon, Y_s^\varepsilon), F(X_s^\varepsilon, Y_s^\varepsilon)\rangle\,ds \\
&+ \int_{t_0}^t \langle \nabla_y u(X_s^\varepsilon, Y_s^\varepsilon), H(X_s^\varepsilon, Y_s^\varepsilon)\,dB_s^\varepsilon\rangle.
\end{aligned}\tag{26}$$

We shall exploit these last formulas, both in order to establish tightness of the sequence $\{Y^\varepsilon\}$, and to identify the limit.

Tightness. We will show the relative compactness of $(Y_\cdot^\varepsilon)$ in the metric space $C([0, T]; \mathbb{R}^d)$. We will use the following slight modification of Theorem 8.2 from Billingsley (1968).

PROPOSITION 3. *The collection* $\{Y_t^\varepsilon,\ 0 \le t \le T\}_{\{0<\varepsilon\le 1\}}$ *is relatively compact if it satisfies the two conditions*:

(i) *For all* $\delta > 0$, *there exists* $M > 0$, *such that*

$$P\left(\sup_{0\le t\le T}|Y_t^\varepsilon| > M\right) \le \delta, \qquad 0 < \varepsilon \le 1.$$

(ii) *For any* $\delta > 0$, $M > 0$ *there exist* ε_0 *and* γ, *such that*

$$\sup_{0<\varepsilon\le\varepsilon_0} P\left(\sup_{0\le t_0\le T}\ \sup_{t\in[t_0,t_0+\gamma]}|Y_t^\varepsilon - Y_{t_0}^\varepsilon| \ge \delta;\ \sup_{0\le s\le T}|Y_s^\varepsilon| \le M\right) \le \delta.$$

We first prove that the sequence $(Y_\cdot^\varepsilon)$ satisfies (i). For the sake of that, we will use (25) and (26), with $t_0 = 0$ and the function $f(y) = \log(1+|y|^2)$. Remember that the function u depends on f. Notice that for this choice of f one has

$$(1+|y|)|\partial_y f(y)| + (1+|y|)^2\|\partial_y^2 f(y)\| + (1+|y|)^3\|\partial_y^3 f(y)\| \le C,$$

and then in particular,

$$|u(x, y)| \le K(1+|x|^{q_3'}).$$

From the assumption (A_q) we have that the absolute values of all Lebesgue integrands in (25) and (26) do not exceed $C(1+|X_s^\varepsilon|^q)$ with some $q < 2r_0 - 1$. So, for any Lebesgue integral $\int K(X_s^\varepsilon, Y_s^\varepsilon)\,ds$ in these two formulas we have

$$\begin{aligned} E_{x,y}\sup_{0\le t\le T}\int_0^t |K(X_s^\varepsilon, Y_s^\varepsilon)|\,ds &\le \int_0^T CE_{x,y}(1+|X_s^\varepsilon|^q)\,ds \\ &\le C\int_0^{CT} E_{x,y}(1+|\widetilde{X}_s^\varepsilon|^q)\,ds \\ &\le CT(1+|x|^q). \end{aligned}$$

Each Itô integral $\int K'\,(X_s^\varepsilon,\ Y_s^\varepsilon)\,dB_s$ has an integrand which satisfies the inequality $|K'(x, y)|^2 \le C(1+|x|^q)$ again with $q < 2r_0 - 1$. From Doob's inequality,

$$\begin{aligned} E_{x,y}\sup_{0\le t\le T}\left|\int_0^t K'(X_s^\varepsilon, Y_s^\varepsilon)\,dB_s\right|^2 &\le 4E_{x,y}\int_0^T |K'(X_s^\varepsilon, Y_s^\varepsilon)|^2\,ds \\ &\le CT(1+|x|^q). \end{aligned}$$

The term $\varepsilon u(x, y)$ is bounded (it even tends to zero). Finally, by virtue of Corollary 1,

$$\varepsilon E_{x,y}\sup_{0\le t\le T}|u(X_t^\varepsilon, Y_t^\varepsilon)| \le C\varepsilon E_{x,y}\sup_{0\le t\le T}(1+|X_t^\varepsilon|)^{|q_3'}) \to 0, \qquad \varepsilon\to 0.$$

Combining the above estimates, we deduce that

$$\sup_{0<\varepsilon\le 1} E_{x,y}\sup_{0\le t\le T}\log(1+|Y_t^\varepsilon|^2) < \infty,$$

from which (i) follows. We now prove (ii). For the sake of that, let us first write (25) and (26) in the particular case of the vector function $f(y) = y$. We obtain

$$
\begin{aligned}
Y_t^\varepsilon = Y_{t_o}^\varepsilon + \int_{t_0}^t \Big[& F(X_s^\varepsilon, Y_s^\varepsilon) + (G_i \partial_{y_i} \overline{G})(X_s^\varepsilon, Y_s^\varepsilon) \\
& + ((H\sigma^*)_{ik} \partial_{x_k} \partial_{y_i} \overline{G})(X_s^\varepsilon, Y_s^\varepsilon) \Big] \, ds \\
& + \int_{t_0}^t [H(X_s^\varepsilon, Y_s^\varepsilon) + \nabla_x \overline{G}(X_s^\varepsilon, Y_s^\varepsilon)\sigma(X_s^\varepsilon)] \, dB_s^\varepsilon \\
& + \varepsilon \Big[\overline{G}(X_{t_0}^\varepsilon, Y_{t_0}^\varepsilon) - \overline{G}(X_t^\varepsilon, Y_t^\varepsilon) \\
& \quad + \tfrac{1}{2} \operatorname{Tr} \int_{t_0}^t (\partial_y^2 \overline{G} H H^*)(X_s^\varepsilon, Y_s^\varepsilon) \, ds \\
& \quad + \int_{t_0}^t \langle \nabla_y \overline{G}(X_s^\varepsilon, Y_s^\varepsilon), F(X_s^\varepsilon, Y_s^\varepsilon) \rangle \, ds \\
& \quad + \int_{t_0}^t \langle \nabla_y \overline{G}(X_s^\varepsilon, Y_s^\varepsilon), H(X_s^\varepsilon, Y_s^\varepsilon) \, dB_s^\varepsilon \rangle \Big].
\end{aligned} \tag{27}
$$

We rewrite the above as

$$
Y_t^\varepsilon - Y_{t_0}^\varepsilon = \overline{Y}_t^\varepsilon - \overline{Y}_{t_0}^\varepsilon + \widehat{Y}_t^\varepsilon - \widehat{Y}_{t_0}^\varepsilon,
$$

where

$$
\begin{aligned}
\overline{Y}_t^\varepsilon - \overline{Y}_{t_o}^\varepsilon &= \varepsilon [\overline{G}(X_{t_0}^\varepsilon, Y_{t_0}^\varepsilon) - \overline{G}(X_t^\varepsilon, Y_t^\varepsilon)], \\
\widehat{Y}_t^\varepsilon - \widehat{Y}_{t_0}^\varepsilon &= \int_{t_0}^t J_\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) \, ds + \int_{t_0}^t K_\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) \, dB_s^\varepsilon,
\end{aligned}
$$

and

$$
\begin{aligned}
J_\varepsilon &= F + G_i \partial_{y_i} \overline{G} + (H\sigma^*)_{ik} \partial_{x_k} \partial_{y_i} \overline{G} + \frac{\varepsilon}{2} \operatorname{Tr} \partial_y^2 \overline{G} H H^* + \varepsilon \langle \nabla_y \overline{G}, F \rangle, \\
K_\varepsilon &= H + \nabla_x \overline{G} \sigma + \varepsilon \nabla_y \overline{G} H.
\end{aligned}
$$

The two processes $\overline{Y}^\varepsilon$ and $\widehat{Y}^\varepsilon$ will be treated differently; (ii) will follow from:

(ii′) For any $\delta > 0$, there exist ε_0 s.t.,

$$
\sup_{0<\varepsilon\le\varepsilon_0} P\left(\sup_{0\le t\le T} |\overline{Y}_t^\varepsilon| \ge \delta \right) \le \delta
$$

and, following Theorem 8.3 in Billingsley (1968):

(ii″) For any $\delta > 0$, $M > 0$, there exist ε_0 and $\gamma > 0$ such that

$$
\gamma^{-1} \sup_{0<\varepsilon\le\varepsilon_0} \sup_{0\le t_0\le T} P\left(\sup_{t\in[t_0,\, t_0+\gamma]} |\widehat{Y}_t^\varepsilon - \widehat{Y}_{t_0}^\varepsilon| \ge \delta;\ \sup_{0\le s\le T} |Y_s^\varepsilon| \le M \right) \le \delta.
$$

The estimate (ii′) follows from the estimate for $\overline{G}$, (i), Corollary 1 and condition (A_q). To estimate the probability in (ii″), we shall use the change of time which was described in Section 2, in order to use the estimate (4). Let

$$\widetilde{X}^{\varepsilon}_t = X^{\varepsilon}_{z_\varepsilon(t)}, \qquad \widetilde{Y}^{\varepsilon}_t = Y^{\varepsilon}_{z_\varepsilon(t)},$$
$$z_\varepsilon(t) = h_\varepsilon^{-1}(t), \qquad h_\varepsilon(t) = \int_0^t \kappa^2(X^\varepsilon_s)\,ds.$$

Let us define the stopping time $\tau^M_\varepsilon = \inf\{r \geq 0,\ |\widetilde{Y}^\varepsilon(r)| \geq M\}$. It suffices to show that for any $\delta > 0$, $M > 0$, there exist ε_0 and $\gamma > 0$ such that

$$\gamma^{-1} \sup_{0<\varepsilon\leq\varepsilon_0} \sup_{0\leq t_0\leq T} P\biggl(\sup_{t\in[t_0,\,t_0+\gamma],\ t\leq\tau^M_\varepsilon} |\widehat{Y}^\varepsilon_{z_\varepsilon(t)} - \widehat{Y}^\varepsilon_{z_\varepsilon(t_0)}| \geq \delta\biggr) \leq \delta. \tag{28}$$

We have [see Ikeda, Watanabe (1981), page 102]

$$\widehat{Y}^\varepsilon_{z_\varepsilon(t)} - \widehat{Y}^\varepsilon_{z_{\varepsilon(t_0)}} = \int_{t_0}^t J_\varepsilon(\widetilde{X}^\varepsilon_s, \widetilde{Y}^\varepsilon_s)\kappa^{-2}(\widetilde{X}^\varepsilon_s)\,ds + \int_{t_0}^t K_\varepsilon(\widetilde{X}^\varepsilon_s, \widetilde{Y}^\varepsilon_s)\kappa^{-1}(\widetilde{X}^\varepsilon_s)\,d\widetilde{B}^\varepsilon_s,$$

where $\widetilde{B}^\varepsilon_s$ is a new Brownian motion and $t > t_0$.

We derive (28) from the two following estimates, with $\nu > 0$ small enough such that condition (A_q) implies the bounds below. Note that, in particular, (4) is used in their derivation:

$$\begin{aligned}
E_{xy}&\biggl(\sup_{t_0\leq t\leq t_0+\gamma}\biggl|\int_{t_0\wedge\tau^M_\varepsilon}^{t\wedge\tau^M_\varepsilon} J_\varepsilon(\widetilde{X}^\varepsilon_s, \widetilde{Y}^\varepsilon_s)\kappa^{-2}(\widetilde{X}^\varepsilon_s)\,ds\biggr|^{1+\nu}\biggr)\\
&\leq C\gamma^\nu E_{xy}\int_{t_0\wedge\tau^M_\varepsilon}^{(t_0+\gamma)\wedge\tau^M_\varepsilon} |J_\varepsilon(\widetilde{X}^\varepsilon_s, \widetilde{Y}^\varepsilon_s)|^{1+\nu}\,ds\\
&\leq C_M\gamma^\nu \int_{t_0}^{t_0+\gamma} E_x(1+|\widetilde{X}_s|^{q''})\,ds\\
&\leq C_M\gamma^{1+\nu}(1+|x|^{q''}),\\
E_{xy}&\biggl(\sup_{t_0\leq t\leq t_0+\gamma}\biggl|\int_{t_0\wedge\tau^M_\varepsilon}^{t\wedge\tau^M_\varepsilon} K_\varepsilon(\widetilde{X}^\varepsilon_s, \widetilde{Y}^\varepsilon_s)\kappa^{-1}(\widetilde{X}^\varepsilon_s)\,d\widetilde{B}^\varepsilon_s\biggr|^{2+2\nu}\biggr)\\
&\leq CE_{xy}\biggl(\sup_{t_0\leq t\leq t_0+\gamma}\biggl|\int_{t_0\wedge\tau^M_\varepsilon}^{t\wedge\tau^M_\varepsilon} K^2_\varepsilon(\widetilde{X}^\varepsilon_s, \widetilde{Y}^\varepsilon_s)\kappa^{-2}(\widetilde{X}^\varepsilon_s)\,ds\biggr|^{1+\nu}\biggr),
\end{aligned}$$

which is estimated exactly as above.

So we get (28) which implies (ii″) for the process $\widetilde{Y}$ instead of Y. Therefore, the process $\widetilde{Y}$ satisfies condition (ii). Notice that the change of time z_ε has a derivative which is bounded and bounded away from zero. So (ii) for the process Y follows from the same assertion with $\widetilde{Y}$.

Identification of the limit. Let Y be a limiting point for (Y^ε) and let $\Phi_{t_0}(\cdot)$ be a bounded continuous functional on $C([0,T];\mathbb{R}^l)$ which is measurable with respect to the sigma-field $\sigma(\varphi_t,\ \varphi \in C([0,T];\mathbb{R}^l),\ t \le t_0)$. We are to show that for any $t_0 \ge 0$, any such Φ and any function $f \in C_0^\infty(\mathbb{R}^l)$ (infinitely differentiable functions with compact support) the following assertion holds:

$$E\left[\left(f(Y_t) - f(Y_{t_0}\right) - \int_{t_0}^t \mathscr{L} f(Y_s)\,ds)\Phi_{t_0}(Y)\right] = 0, \qquad t \ge t_0.$$

We first deduce from (25) and (26) that (in the sequel always $t \ge t_0$)

$$\begin{aligned}
&E_{x,y}\bigg[\Big(f(Y_t^\varepsilon) - f(Y_{t_0}^\varepsilon)\\
&\quad - \int_{t_0}^t \Big\langle \nabla_y f(Y_s^\varepsilon), F(X_s^\varepsilon, Y_s^\varepsilon) + G_i(X_s^\varepsilon, Y_s^\varepsilon)\partial_{y_i}\overline{G}(X_s^\varepsilon, Y_s^\varepsilon)\\
&\qquad\qquad + (H\sigma^*)_{ik}(X_s^\varepsilon, Y_s^\varepsilon)\partial_{y_i}\partial_{x_k}\overline{G}(X_s^\varepsilon, Y_s^\varepsilon)\Big\rangle\,ds\\
&\quad - \tfrac{1}{2}E_{x,y}\int_{t_0}^t \partial_{y_i}\partial_{y_j} f(Y_s^\varepsilon)\big[(HH^*)_{ij}(X_s^\varepsilon, Y_s^\varepsilon) + 2(G_i\overline{G}_j)(X_s^\varepsilon, Y_s^\varepsilon)\\
&\qquad\qquad + 2(H\sigma^*)_{ik}(X_s^\varepsilon, Y_s^\varepsilon)\partial_{x_k}\overline{G}_j(X_s^\varepsilon, Y_s^\varepsilon)\big]\,ds\\
&\qquad\qquad - \varepsilon R_f^\varepsilon(t_0, t)\Big)\Phi_{t_0}(Y^\varepsilon)\bigg] = 0.
\end{aligned} \tag{29}$$

It follows from the arguments used in the proof of compactness that

$$\varepsilon E_{x,y}[R_f^\varepsilon(t_0, t)\Phi_{t_0}(Y^\varepsilon)] \to 0,$$

as $\varepsilon \to 0$. Due to the tightness of the sequence $\{Y^\varepsilon\}$, there exists a sequence $\varepsilon_n \to 0$ and a continuous process Y, such that $Y^{\varepsilon_n} \Longrightarrow Y$, as $n \to \infty$. Consequently $E_{x,y}[\Gamma_n] \to 0$, as $n \to \infty$, where

$$\begin{aligned}
\Gamma_n := &\bigg[f(Y_t^{\varepsilon_n}) - f(Y_{t_0}^{\varepsilon_n})\\
&- \int_{t_0}^t \Big\langle \nabla_y f(Y_s^{\varepsilon_n}), F(X_s^{\varepsilon_n}, Y_s^{\varepsilon_n}) + G_i(X_s^{\varepsilon_n}, Y_s^{\varepsilon_n})\partial_{y_i}\overline{G}(X_s^{\varepsilon_n}, Y_s^{\varepsilon_n})\\
&\qquad + (H\sigma^*)_{ik}(X_s^{\varepsilon_n}, Y_s^{\varepsilon_n})\partial_{y_i}\partial_{x_k}\overline{G}(X_s^{\varepsilon_n}, Y_s^{\varepsilon_n})\Big\rangle\,ds\\
&- \tfrac{1}{2}E_{x,y}\int_{t_0}^t \partial_{y_i}\partial_{y_j} f(Y_s^{\varepsilon_n})\Big[(HH^*)_{ij}(X_s^{\varepsilon_n}, Y_s^{\varepsilon_n}) + 2(G_i\overline{G}_j)(X_s^{\varepsilon_n}, Y_s^{\varepsilon_n})\\
&\qquad + 2(H\sigma^*)_{ik}(X_s^{\varepsilon_n}, Y_s^{\varepsilon_n})\partial_{x_k}\overline{G}_j(X_s^{\varepsilon_n}, Y_s^{\varepsilon_n})\Big]\,ds\bigg]\Phi_{t_0}(Y^{\varepsilon_n}).
\end{aligned}$$

All we need to show is that $E(\Gamma_n) \to E(\Gamma)$, where

$$\Gamma := \left[f(Y_t) - f(Y_{t_0}) - \int_{t_0}^t \mathscr{L} f(Y_s)\,ds\right]\Phi_{t_0}(Y).$$

We use the standard idea of freezing the slow component; see, for example, Lemmas 3 and 5 in Pardoux and Veretennikov (1997). To this end, we establish the following lemma.

LEMMA 1. *For any* $\delta > 0$ *there exist* $N \in \mathbb{N}$ *and* $\mathbb{R}^l$*–valued step functions* $y^1, \dots, y^N$ *s.t.*

$$P\left(\bigcap_{k=1}^N \left\{\sup_{0\le t\le T} |Y_t^{\varepsilon_n} - y_t^k| > \delta\right\}\right) < \delta \qquad \forall\, n \in \mathbb{N},$$

$$P\left(\bigcap_{k=1}^N \left\{\sup_{0\le t\le T} |Y_t - y_t^k| > \delta\right\}\right) < \delta.$$

PROOF. The result follows from the tightness of the set $\{Y;\, Y^{\varepsilon_n},\, n \in \mathbb{N}\}$, the separability of $C([0, T]; \mathbb{R}^d)$ and the fact that to any continuous function we can associate a step function which is arbitrarily close to the former in sup norm.

To each $y \in C([0, T]; \mathbb{R}^l)$, and $k = 1, 2, \dots, N$, we associate the number

$$\beta_k(y) := \sup_{0\le t\le T} |y(t) - y_t^k|.$$

Let

$$\psi, \varphi_1, \dots, \varphi_N \colon C([0, T]; \mathbb{R}^l) \to [0, 1]$$

be continuous mappings such that:

(i) $\psi(y) + \sum_{k=1}^N \varphi_k(y) = 1 \ \ \forall\, y \in C([0, T]; \mathbb{R}^l)$;
(ii) $\operatorname{supp} \psi \subset \bigcap_{k=1}^N \{y; \beta_k(y) > \delta\}$;
(iii) $\operatorname{supp} \varphi_k \subset \{y;\, \beta_k(y) < 2\delta\},\ 1 \le k \le N$.

We define moreover the random variables

$$\xi_n := \psi(Y^{\varepsilon_n}),\ \ \xi = \psi(Y);\ \ \eta_n^k = \varphi_k(Y^{\varepsilon_n}),\ \ \eta^k = \varphi_k(Y), \qquad n \in \mathbb{N},\ 1 \le k \le N.$$

Note that

$$\operatorname{supp} \xi_n \subset A_n = \bigcap_{k=1}^N \left\{\sup_{0\le t\le T} |Y_t^{\varepsilon_n} - y_t^k| > \delta\right\},$$

and similarly,

$$\operatorname{supp} \xi \subset A = \bigcap_{k=1}^N \left\{\sup_{0\le t\le T} |Y_t - y_t^k| > \delta\right\}.$$

Define Γ_n^k as the random variable Γ_n, where Y^{ε_n} is replaced by y^k, and Γ^k as the quantity obtained by replacing Y by y^k in the expression for Γ. The above considerations yield the fact that

$$E_{x,y}(\Gamma_n \xi_n) + \sum_{k=1}^N E_{x,y}(\Gamma_n \eta_n^k) \to 0,$$

as $n \to \infty$. We also use the decomposition

$$E(\Gamma) = E(\Gamma\xi) + \sum_{k=1}^N E(\Gamma\eta^k).$$

(Recall that f here has a compact support.) Now, let $p, q > 1$, $p^{-1} + q^{-1} = 1$, and p is close to 1, namely,

$$p \max[q_1, 2q_2, 2q_3', q_2 + q_3', q_3 + q_3', q_3 + q_4'] < r_1.$$

Then by the Hölder inequality,

$$|E(\Gamma_n \xi_n)| \le \big(E(\Gamma_n)^p\big)^{1/p} \big(P(A_n)\big)^{1/q} \le C\delta^{1/q},$$
$$|E(\Gamma \xi)| \le \big(E(\Gamma)^p\big)^{1/p} \big(P(A)\big)^{1/q} \le C\delta^{1/q}.$$

Indeed, hypothesis (A_q) allows estimating from above all Lebesgue and Itô integrals in the expression $E(\Gamma_n)^p$ in a standard manner. The value $E(\Gamma)^p$ is bounded just because $f \in C_0^\infty$. Also,

$$\sum_{k=1}^N E(\Gamma_n \eta_n^k) = \sum_{k=1}^N E[(\Gamma_n - \Gamma_n^k)\eta_n^k] + \sum_{k=1}^N E(\Gamma_n^k \eta_n^k),$$
$$\sum_{k=1}^N E(\Gamma \eta^k) = \sum_{k=1}^N E[(\Gamma - \Gamma^k)\eta^k] + \sum_{k=1}^N E(\Gamma^k \eta^k).$$

From the Lipschitz property of the coefficients of Γ_n and Γ with respect to y,

$$\left| \sum_{k=1}^N E[(\Gamma_n - \Gamma_n^k)\eta_n^k] \right| + \left| \sum_{k=1}^N E[(\Gamma - \Gamma^k)\eta^k] \right| \le \rho(\delta),$$

where $\rho(\delta) \to 0$, as $\delta \to 0$. It finally remains to show that

$$E[\Gamma_n^k \eta_n^k] \to E[\Gamma^k \eta^k],$$

as $n \to \infty$. Since $\eta_n^k \Rightarrow \eta^k$ and $|\eta_n^k| \le 1$, it suffices to show that $\Gamma_n^k \to \Gamma^k$ in $L^1(\Omega)$ (note that Γ^k is nonrandom). Indeed,

$$\begin{aligned} \left| E(\Gamma_n^k \eta_n^k - \Gamma^k \eta^k) \right| &\le E\left| (\Gamma_n^k - \Gamma^k)\eta_n^k \right| + |\Gamma^k| \left| E(\eta_n^k - \eta^k) \right| \\ &\le E|\Gamma_n^k - \Gamma^k| + |\Gamma^k| \times |E(\eta_n^k - \eta^k)|. \end{aligned}$$

Finally, the $L^1(\Omega)$–convergence of Γ_n^k toward Γ^k follows from the following lemma.

LEMMA 2. *Let $K(X_s^{\varepsilon_n}, y_s^k)$ denote any of the functions under the integral sign in the expression for Γ_n^k. Denote $\overline{K}(y) = \int K(x, y)\tilde{\mu}(dx)$. Then for any $t < T$,*

$$E\left| \int_0^t (K(X_s^{\varepsilon_n}, y_s^k) - \overline{K}(y_s^k))\, ds \right| \to 0, \qquad \varepsilon_n \to 0.$$

PROOF. Let $(a_k, b_k) \subset [0, T]$ be an interval on which y^k is constant, equal to z^k. The a.s. convergence

$$\int_{a_k}^{b_k} K(X_s^{\varepsilon_n}, z^k)\, ds \to (b_k - a_k)\overline{K}(z^k)$$

follows from the ergodic theorem. The $L^1(\Omega)$-convergence follows by uniform integrability, which is deduced from Proposition 1 and condition (A_q). □

APPENDIX

PROPOSITION 4. *Under the assumptions of Theorem* 3, (24) *holds true*.

The proof will follow from four lemmas and a localization procedure. All lemmas are established under the assumptions of Theorem 3.

LEMMA 3. *For any* $R > 0$, $p > 1$ *and bounded* D,

$$\sup_{|y|\le R} \|\partial_y \overline{G}(\cdot, y)\|_{W^2_p(D)} < \infty, \qquad \partial_y \overline{G} \in C(\mathbb{R}^d \times \mathbb{R}^l)$$

and

$$\partial_{y_i} \overline{G}(x, y) = \int_0^\infty E_x \partial_{y_i} G(X_t, y)\, dt.$$

PROOF.

$$v^i(x, y) := \int_0^\infty E_x(\partial/\partial y_i) G(X_t, y)\, dt.$$

Due to Theorem 1 this function is well-defined because

$$\langle \partial_{y_i} G(\cdot, y), \mu\rangle = \partial_{y_i} \int G(x, y)\mu(dx) = 0.$$

The first equality here follows from the bounds

$$|G(x, y)| \le C(1 + |y| + |x|^{q_3}),$$
$$\|\nabla_y G(x, y)\| \le C(1 + |x|^{q_4}), \qquad q_3, q_4 < 2r_0 - 1$$

and from the inequality $\int |x|^q \mu(dx) < \infty$ for any $q < 2r_0 - 1$.

Due to Theorem 1, $v^i(\cdot, y) \in \bigcap_{p>1} W^2_{p,\,\mathrm{loc}}$ locally uniformly w.r.t. y, moreover, for any $R > 0$,

$$\sup_{|x|\le R} \sup_{y} \left| \int_0^N E_x(\partial/\partial y_i) G(X_t, y)\, dt - v^i(x, y)\right| \to 0, \qquad N \to \infty$$

due to the assumptions of Theorem 3 on $\partial_y G$. This implies that $v^i \in C(\mathbb{R}^d \times \mathbb{R}^l)$.

The equality $v^i(x, y) = \partial_{y_i} \overline{G}(\cdot, y)$ follows from the uniform convergence of the integral $\int_0^\infty |E_x \partial_{y_i} G(X_t, y)|\, dt$, see theorem 1. Lemma 3 is proved. □

LEMMA 4. *For any* $R > 0$, $p > 1$ *and bounded* D,

$$\sup_{|y|\le R} \|\partial^2_y \overline{G}(\cdot, y)\|_{W^2_p(D)} < \infty, \qquad \partial^2_y \overline{G} \in C(\mathbb{R}^d \times \mathbb{R}^l)$$

and

$$\partial_{y_i}\partial_{y_j} \overline{G}(x, y) = \int_0^\infty E_x \partial_{y_i}\partial_{y_j} G(X_t, y)\, dt.$$

The proof is similar to the proof of Lemma 3.

LEMMA 5. *For any $R > 0$, $p > 1$ and bounded D,*

$$\sup_{|y|\le R} \|\partial_{x_i}\partial_{y_j}\overline{G}(\cdot, y)\|_{W^1_p(D)} < \infty, \qquad \partial_{x_i}\partial_{y_j}\overline{G} \in C(\mathbb{R}^d \times \mathbb{R}^l).$$

PROOF. The first assertion follows from Lemma 3. Moreover, we get from the embedding theorem that for any $R > 0$ there exist $C, \lambda > 0$ such that

$$\sup_{|y|\le R}\sup_{|x|\le R} |\partial_x\partial_y\overline{G}(x, y) - \partial_x\partial_y\overline{G}(x', y)| \le C|x - x'|^\lambda.$$

Now the second assertion follows from this and the first assertion in Lemma 4. Lemma 5 is proved. □

LEMMA 6. *For any $p > 1$, $R > 0$,*

$$\lim_{\delta\to 0}\sup_{|y|\le R}\int_{|x|\le R}\sup_{|z|\le\delta}\Big|\partial_{x_i}\partial_{x_j}\overline{G}(x, y) - \partial_{x_i}\partial_{x_j}\overline{G}(x, y + z)\Big|^p\,dx = 0.$$

PROOF. Let $R > 0$, $p > 1$. It follows from the inequality

$$\sup_{|y|\le R+1}\|\partial_y\overline{G}(\cdot, y)\|_{W^2_p(B_R)} =: C_R < \infty$$

that

$$\int_{|y|\le R+1}\int_{|x|\le R} |\partial_x^2(\partial_y\overline{G}(x, y))|^p\,dx\,dy < \infty.$$

By virtue of the Fubini theorem, we get for a.s. $|x| \le R$,

$$\int_{|y|\le R+1}\|\partial_x^2(\partial_y\overline{G}(x, y))\|^p\,dy =: C(x) < \infty,$$

where $\int_{|x|\le R} C(x)\,dx < \infty$. So, due to the embedding theorem we get that if $|y|, |y'| \le R$,

$$\|\partial_x^2\overline{G}(x, y) - \partial_x^2\overline{G}(x, y')\| \le CC(x)^{1/p}|y - y'|^\lambda.$$

Here $\lambda > 0$ only depends on d and p which is, indeed, arbitrary large, C depends on R and d. Hence

$$\sup_{|y|\le R}\int_{|x|\le R}\sup_{|z|\le\delta}\|\partial_x^2\overline{G}(x, y) - \partial_x^2\overline{G}(x, y + z)\|^p\,dx$$

$$\le C\delta^{p\lambda}\int_{|x|\le R} C(x)\,dx \to 0$$

as $\delta \to 0$. Lemma 6 is proved. □

COROLLARY 2.

$$\sup_{|y|\le R}\|\partial_x^2\overline{G}(\cdot, y)\| \in \bigcap_{p>1} L_p(B_R).$$

PROOF OF PROPOSITION 4. We will omit the index ε. Let $\tau^R = \inf(t \geq 0\colon (X_t, Y_t) \notin B_R)$. Here $B_R = ((x, y)\colon |(x, y)| \leq R)$. It is sufficient to prove formula (24) with $\min(t, \tau^R)$ for any R, instead of t. Let us consider a convolution $\overline{G}_n(x, y) = \overline{G}(\cdot, y) * \psi_n(x)$ with the kernel $\psi_n(x) = n^d \psi(x/n)$ where $\psi \in C_0^\infty$, $\psi \geq 0$, $\int \psi(x)\,dx = 1$. Then we have

$$\sup_{B_R} \Big(|\overline{G}_n(x, y) - \overline{G}(x, y)| + \|\partial_x \overline{G}_n(x, y) - \partial_x \overline{G}(x, y)\| + \|\partial_y \overline{G}_n(x, y) - \partial_y \overline{G}(x, y)\| + \|\partial_x\partial_y \overline{G}_n(x, y) - \partial_x\partial_y \overline{G}(x, y)\| \Big) \to 0, \qquad n \to \infty.$$

Moreover, due to Lemma 6,

$$\sup_{|y|\leq R} \|\partial_x^2 \overline{G}_n(\cdot, y) - \partial_x^2 \overline{G}(\cdot, y)\|_{L_p(B_R)} \to 0, \qquad n \to \infty.$$

So, one can pass to the limit in the Itô formula for $\overline{G}_n$, by the uniform convergence in all terms but the last one, $\int_0^{\min(t,\tau)} \partial_x^2 \overline{G}_n(X_s, Y_s)\,ds$ where one can do it by Krylov's estimate and due to Lemma 6. Indeed,

$$\begin{aligned} &E \int_0^{\min(t,\tau)} |\partial_x^2 (\overline{G}_n - \overline{G})(X_s, Y_s)|\,ds \\ &\quad\leq E \int_0^{\min(t,\tau)} \sup_{|y|\leq R} \|\partial_x^2 (\overline{G}_n - \overline{G})(X_s, y)\|\,ds \\ &\quad\leq C \Big\| \sup_{|y|\leq R} \|(\partial_x^2 \overline{G}_n - \partial_x^2 \overline{G})(\cdot, y)\| \Big\|_{L_p(B_R)} \to 0, \qquad n \to \infty. \end{aligned}$$

Proposition 4 is proved. □

Acknowledgment. The authors are grateful to the anonymous referee for useful remarks.

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BOUC, R. and PARDOUX, E. (1984). Asymptotic analysis of PDEs with wide-band noise disturbance expansion of the moments. *Stochastic Anal. Appl.* **2** 369–422.

DYNKIN, E. B. (1965). *Markov Processes* **2**. Springer, Berlin.

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes. Characterization and Convergence*, Wiley, New York.

GILBARG, D. and TRUDINGER, N. S. (1983). *Elliptic Partial Differential Equations of Second Order*, 2nd ed. Springer, Berlin.

IKEDA, N. and WATANABE, S. (1981). *Stochastic Differential Equations and Diffusion Processes*. North-Holland, Amsterdam.

KHASMINSKI, R. Z. (1966). A limit theorem for solutions of differential equations with random right–hand sides. *Theory Probab. Appl.* **11** 390–406.

KHASMINSKI, R. Z. (1980). *Stochastic Stability of Differential Equations*. Sijthoff and Nordhoff, The Netherlands.

KRYLOV, N. V. (1980). *Controlled Diffusion Processes* (trans. by A. B. Aries). Springer, Berlin.

LADYZENSKAJA, O., SOLONNIKOV, V. and URAL'CEVA, N. (1968). *Linear and Quasilinear Equations of Parabolic Type*, Amer. Math. Soc., Providence, RI.

PAPANICOLAOU, G. C., STROOCK, D. W. and VARADHAN, S. R. S. (1977). Martingale approach to some limit theorems. In *Conference on Statistical Mechanics, Dynamical Systems and Turbulence* (M. Reed ed.) Duke Univ. Press.

PARDOUX, E. and VERETENNIKOV, A. YU. (1997). Averaging of backward stochastic differential equations, with application to semi–linear PDEs. *Stochastics Stochastic. Rep.* **60** 255–270.

REVUZ, D. (1984). *Markov Chains*, 2nd rev. ed. North-Holland, Amsterdam.

STRATONOVICH, R. L. (1963, 1967). *Topics in the Theory of Random Noise* **1, 2**. Gordon and Breach, New York.

STROOCK, D. W. and VARADHAN, S. R. S. (1979). *Multidimensional Diffusion Processes*. Springer, Berlin.

VERETENNIKOV, A. YU. (1982). Parabolic equations and Itô's stochastic equations with coefficients discontinuous in the time variable. *J. Math. Notes* **31** 278–283 (trans. from *Mat. Zametki* **31** 549–557).

VERETENNIKOV, A. YU. (1987). Bounds for the mixing rates in the theory of stochastic equations. *Theory Probab. Appl.* **32** 273–281.

VERETENNIKOV, A. YU. (1997). On polynomial mixing bounds for stochastic differential equations. *Stochastic Process. Appl.* **70** 115–127.

LATP, UMR–CNRS 6632
CENTRE DE MATHÉMATIQUES
ET D'INFORMATIQUE
UNIVERSITÉ DE PROVENCE
39, RUE F. JOLIOT CURIE
13453 MARSEILLE CEDEX 13
FRANCE
E-MAIL: pardoux@cmi.univ-mrs.fr

INSTITUTE OF INFORMATION
TRANSMISSION PROBLEMS
19, BOLSHOY KARETNII
101447 MOSCOW
RUSSIA
E-MAIL: veretenn@iitp.ru