

## DIFFUSION APPROXIMATION OF SYSTEMS WITH WEAKLY ERGODIC MARKOV PERTURBATIONS. I

UDC 519.21

A. YU. VERETENNIKOV AND A. M. KULIK

ABSTRACT. Diffusion approximation type results are obtained for a system perturbed by a Markov process whose transition probabilities converge to the invariant distribution nonuniformly with respect to the initial value. In general, the mode of convergence is weaker than the total variation convergence.

### 1. INTRODUCTION

We investigate the limit behavior as  $\varepsilon \rightarrow 0$  of the family  $Y^\varepsilon$ ,  $\varepsilon > 0$ , of solutions of stochastic differential equations whose coefficients depend on a “fast” random component

$$X^\varepsilon(t) = X(t\varepsilon^{-1}), \quad t \in \mathbb{R}^+,$$

where  $X$  is a homogeneous Markov process.

Since the models that include both “fast” and “slow” components are widely used in the theory as well as in applications, the studies of asymptotic problems of the type mentioned above are of obvious interest. A survey of results in this direction can be found, for example, in Sections 1 and 2 of the paper [1].

The aim of this paper is to obtain a diffusion approximation result (this term means a certain version of the functional central limit theorem) for processes  $Y^\varepsilon$  under as weak as possible restrictions on the ergodic properties of the process  $X$ .

The classical results on the diffusion approximation ([2]–[6]) assume that either a bound for the *uniform mixing coefficient* ( $\phi$ -mixing coefficient) or the *uniform ergodicity* of the process  $X$  holds. Instead we impose an essentially weaker assumption on the convergence of transition probabilities of the process  $X$  to its invariant distribution; see (2.4). This assumption implies the nonuniform (with respect to the initial value) convergence in the metric that, in general, is weaker than the convergence in variation. This allows us to sufficiently extend the class of the corresponding models. In particular, our approach is applied to processes  $X$  defined by stochastic differential equations (possibly, infinite dimensional stochastic differential equations) with the Lévy noise. Note that an attempt to apply other approaches known in the literature to the latter processes faces inextricable difficulties.

Our approach in outline follows the scheme developed in the paper [7] for the model where the “fast” random component  $X$  is a diffusion process. The latter model is related to a solution of the *Poisson equation* for the corresponding diffusion generator. The subsequent analysis is based on a standard method of using the *corrector*. The latter idea requires information about the properties of semimartingale components of the corrector. Such properties are obtained in the papers [7] and [8] with the help of an appropriate

---

2010 *Mathematics Subject Classification*. Primary 60H10, 60J10.

*Key words and phrases*. Diffusion approximation, corrector, extended potential.

variant of the Itô formula. In doing so, one needs to apply some results on the local properties of weak solutions of second order partial equations of the parabolic and elliptic types. Such results are known only under the assumption that the process  $X$  is of a diffusion type and therefore we need to essentially extend the method in the general case. In [9], a result is obtained that can be viewed as an extended version of the Dynkin formula and that can be used for describing the predictable component in the semimartingale decomposition of the corrector. This description is sufficient for the proof of the weak compactness of the family  $\{Y^\varepsilon\}$  as well as for determining its limit points. Since the properties of the martingale component in the semimartingale decomposition of the corrector are not known, we essentially modify the reasoning; see the discussion at the beginning of Section 4 below.

Due to requirements of the editorial board, the paper is split into two parts. In the current part, we introduce the main objects and state and prove the main result. For the sake of completeness and having further applications in mind, we consider two versions of this result. The first version is a limit theorem stated in terms of the convergence of finite dimensional distributions, while the second one obtains a stronger functional form of this limit theorem under somewhat stronger assumptions imposed on the process  $X$ , namely we assume condition  $\mathbf{S}(\psi_i, Q_i)$ ,  $i = 1, 2$ , instead of  $\mathbf{W}(\psi_i, Q_i)$ ,  $i = 1, 2$  (see below).

In the forthcoming paper [10] we discuss some corollaries of the main result for particular cases being of their own interest and provide an example of an application of Theorems 3.1 and 3.2. The proofs of auxiliary results of the current paper (Lemmas 4.2 and 4.4) are also given in [10].

## 2. PRELIMINARIES

**2.1. Main objects and notation.** Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space and let

$$X = \{X(t), t \in \mathbb{R}^+\}$$

be a homogeneous Markov process of the Feller type defined on  $(\Omega, \mathcal{F}, \mathbb{P})$ . Let  $X$  assume its values in a complete separable metric space  $\mathbb{X}$ . We suppose that the transition probability  $P_t(x, dy)$  of the process  $X$  is such that, whatever the distribution of the initial value  $X(0)$  is, there exists a càdlàg version of the process  $X$ . In what follows we assume that the trajectories of  $X$  itself are càdlàg.

Let  $\mathbb{P}_x$  be the distribution in  $\mathbb{D}(\mathbb{R}^+, \mathbb{X})$  of the càdlàg version of the process  $X$  with  $X(0) = x$  and let  $\mathbb{E}_x$  denote the corresponding mathematical expectation. Put

$$X^\varepsilon(t) = X(t\varepsilon^{-1}), \quad t \in \mathbb{R}^+, \quad \varepsilon \in (0, 1].$$

The assumptions above can be weakened in some cases. For example, one can assume that the phase space  $\mathbb{X}$  is a measurable space instead of assuming that  $\mathbb{X}$  is a metric space; see Remark 3.3 in [10].

Further, we assume that a Wiener process  $W$  is defined on the initial probability space and that  $W$  assumes values in  $\mathbb{R}^l$  and is independent of  $X$ . We denote by  $\mathbb{F}^X$  the natural filtration generated by the process  $X$  and by  $\mathbb{F}^\varepsilon = \{\mathcal{F}_t^\varepsilon, t \in \mathbb{R}^+\}$  the filtration generated by the pair of processes  $X^\varepsilon$  and  $W$ .

Consider the family of stochastic differential equations in  $\mathbb{R}^m$  given by

$$(2.1) \quad \begin{aligned} Y^\varepsilon(t) = y_0 + \int_0^t \left[ a(X^\varepsilon(s), Y^\varepsilon(s)) + \frac{1}{\sqrt{\varepsilon}} A(X^\varepsilon(s), Y^\varepsilon(s)) \right] ds \\ + \int_0^t \sigma(X^\varepsilon(s), Y^\varepsilon(s)) dW_s, \quad t \geq 0. \end{aligned}$$

We further assume that there exists a function  $L: \mathbb{X} \rightarrow \mathbb{R}^+$  being bounded on each compact set of  $\mathbb{X}$  and for which

$$(2.2) \quad |a(x, y)| + |A(x, y)| + |\sigma(x, y)| \leq L(x)(1 + |y|), \quad x \in \mathbb{X}, y \in \mathbb{R}^m.$$

Moreover, we assume that, for an arbitrary  $R > 0$ , there exists a function  $L_R: \mathbb{X} \rightarrow \mathbb{R}^+$  being bounded on each compact set of  $\mathbb{X}$  and such that

$$(2.3) \quad \begin{aligned} &|a(x, y_1) - a(x, y_2)| + |A(x, y_1) - A(x, y_2)| + |\sigma(x, y_1) - \sigma(x, y_2)| \\ &\leq L_R(x)|y_1 - y_2|, \quad x \in \mathbb{X}, y_1, y_2 \in \{y: |y| \leq R\}. \end{aligned}$$

Here and in what follows we use the same symbol  $|\cdot|$  for the absolute value of a real number, the norm of a vector, or for the norm of a matrix. Then equation (2.1) possesses a unique strong solution  $Y^\varepsilon$  wherever  $\varepsilon \in (0, 1]$  is (see [9], Proposition 2.1 for details).

Throughout the paper,  $b = \{b_{ij}\}_{i,j=1}^m$  denotes the matrix of diffusion coefficients for the process defined by (2.1):

$$b_{ij} = \sum_{k=1}^l \sigma_{ik}\sigma_{jk}.$$

**2.2. Conditions imposed on the process  $X$ .** In what follows we assume that we are given a *premetric*  $d$ ; that is,  $d$  is a nonnegative symmetric semicontinuous from below function on  $\mathbb{X} \times \mathbb{X}$  vanishing in the “diagonal”  $\{(x, x), x \in \mathbb{X}\} \subset \mathbb{X} \times \mathbb{X}$  and assuming positive values beyond it.

For all pairs of probability measures  $\mu$  and  $\nu$  defined on the  $\sigma$ -algebra of Borel subsets in  $\mathbb{X}$ , put

$$d(\mu, \nu) = \inf_{(\xi, \eta) \in C(\mu, \nu)} \mathbb{E} d(\xi, \eta),$$

where  $C(\mu, \nu)$  is a family of pairs  $(\xi, \eta)$  of random elements whose values belong to  $\mathbb{X}$  and such that  $\xi \sim \mu$  and  $\eta \sim \nu$ . The above definition is an appropriate version of the definition of the *minimal* (or *coupling*) distance between the probability distributions; see, for example, [11] or [12], Section 11.8.

The following is the main assumption about the ergodic properties of the process  $X$ ; in other words, this is an assumption on the limit behavior of its transition probabilities as  $t \rightarrow +\infty$ .

**$\mathbf{E}(\mathbf{d}, \mathbf{r}, \psi)$ .** There exists a unique invariant probability  $\pi$  of the process  $X$  such that

$$(2.4) \quad d(P_t(x, \cdot), \pi) \leq r(t)\psi(x), \quad x \in \mathbb{X}, t \geq 0,$$

where  $d$  is a premetric, the measurable function  $\psi$  assumes values in the interval  $[1, +\infty)$ , and the function  $r: \mathbb{R}^+ \rightarrow \mathbb{R}^+$  is measurable, bounded, and such that  $r(t) \rightarrow 0$  as  $t \rightarrow \infty$ .

We introduce yet another moment condition imposed on the transition probabilities.

**$\mathbf{M}_p(\phi, \psi)$ .** For two functions  $\phi, \psi: \mathbb{X} \rightarrow [1, \infty)$  such that  $\psi$  belongs to  $L_1(\pi)$  and for a number  $p \in [1, \infty)$ ,

$$(2.5) \quad \int_{\mathbb{X}} \phi^p(y) P_t(x, dy) \leq \psi(x), \quad x \in \mathbb{X}, t \geq 0$$

(we agree that  $1^\infty = 1$  and  $a^\infty = \infty$  for  $a > 1$ ).

The function  $\psi$  involved in conditions (2.4) and (2.5) can be viewed as a kind of “penalty” depending on the initial value of the process  $X$ . The following assumptions control (in various forms) the amount of such a penalty. As shown in the forthcoming paper [10], all these assumptions can be checked directly.

**$\mathbf{P}(\psi)$ .** For an arbitrary  $x \in \mathbb{X}$ ,

$$\mathbb{E}_x \int_0^\infty e^{-\lambda t} \psi(X(t)) dt < \infty \quad \text{for all } \lambda > 0.$$

**W( $\psi, \mathbf{Q}$ ).** For a given nondecreasing function  $Q: [1, +\infty) \rightarrow [1, +\infty)$  and for an arbitrary  $T \in \mathbb{R}^+$ ,

$$(2.6) \quad \varepsilon Q(\varepsilon^{-1}) \sup_{t \in [0, T]} \mathbf{E} \psi(X^\varepsilon(t)) \rightarrow 0, \quad \varepsilon \rightarrow 0.$$

Denote by  $\mathcal{S}^\varepsilon(T)$  the family of discrete  $\mathbb{F}^\varepsilon$ -stopping times that do not exceed  $T$ ; a number  $T > 0$  here is arbitrary.

**S( $\psi, \mathbf{Q}$ ).** For a given nondecreasing function  $Q: [1, +\infty) \rightarrow [1, +\infty)$  and for an arbitrary  $T \in \mathbb{R}^+$ ,

$$(2.7) \quad \varepsilon Q(\varepsilon^{-1}) \sup_{\tau \in \mathcal{S}^\varepsilon(T)} \mathbf{E} \psi(X^\varepsilon(\tau)) \rightarrow 0, \quad \varepsilon \rightarrow 0.$$

**2.3. Function classes and seminorms.** For a number  $p \in [1, \infty)$  and for a function  $\phi: \mathbb{X} \rightarrow [1, +\infty)$ , denote by  $H_{\phi, d, p}$  the class of functions  $f: \mathbb{X} \rightarrow \mathbb{R}$  such that

$$(2.8) \quad \|f\|_{\phi, d, p} := \sup_{x_1 \neq x_2} \frac{|f(x_1) - f(x_2)|}{d^{1/p}(x_1, x_2)(\phi(x_1) + \phi(x_2))} < +\infty.$$

It is natural to view  $H_{\phi, d, p}$  as a “weight Hölder space” with respect to the premetric  $d$  with index  $1/p$  and weight  $\phi$ .

We denote by  $\|\cdot\|_\pi$  the norm in  $L_1(\pi)$ ,

$$\|f\|_\pi = \int_{\mathbb{X}} |f| d\pi.$$

Here and in what follows  $\pi$  is a unique invariant measure of the process  $X$ . A function  $f \in L_1(\pi)$  is called centered if  $\int_{\mathbb{X}} f d\pi = 0$ . Denote by  $H_{\phi, p}(d, \pi)$  the class of measurable functions  $f: \mathbb{X} \rightarrow \mathbb{R}$  such that

$$\|f\|_{\phi, d, p, \pi} := \|f\|_{\phi, d, p} + \|f\|_\pi < +\infty.$$

For  $\kappa \in \mathbb{R}$  and for a measurable function  $f: \mathbb{X} \times \mathbb{R}^m \rightarrow \mathbb{R}$  being continuous with respect to the second argument, put

$$(2.9) \quad \|f\|_{\phi, d, p, \kappa} = \sup_{y \in \mathbb{R}^m} \frac{\|f(\cdot, y)\|_{\phi, d, p}}{(1 + |y|)^\kappa}, \quad \|f\|_{\pi, \kappa} = \int_{\mathbb{X}} \sup_{y \in \mathbb{R}^m} \frac{|f(x, y)|}{(1 + |y|)^\kappa} \pi(dx).$$

Let  $\mathbf{H}_{\phi, p, \kappa}(d, \pi)$  be the class of functions  $f: \mathbb{X} \rightarrow \mathbb{R}$  such that

$$(2.10) \quad \|f\|_{\phi, d, p, \pi, \kappa} := \|f\|_{\phi, d, p, \kappa} + \|f\|_{\pi, \kappa} < +\infty.$$

We consider a subclass of the class  $\mathbf{H}_{\phi, p, \kappa}(d, \pi)$  whose defining property is closely related to the uniform integrability condition. Consider the functions

$$\Xi_N: \mathbb{R} \ni z \mapsto |z| - |z| \wedge N, \quad N \geq 1,$$

and let  $\hat{\mathbf{H}}_{\phi, p, \kappa}(d, \pi)$  be the set of functions  $f \in \mathbf{H}_{\phi, p, \kappa}(d, \pi)$  such that

$$(2.11) \quad \int_{\mathbb{X}} \sup_{y \in \mathbb{R}^m} \Xi_N \left( \frac{f(x, y)}{(1 + |y|)^\kappa} \right) \pi(dx) \rightarrow 0, \quad N \rightarrow \infty.$$

We say that a measurable function  $f: \mathbb{X} \rightarrow \mathbb{R}$  belongs to the domain of definition of an *extended potential*  $\mathcal{R}$  of the process  $X$  if

$$\int_0^\infty |\mathbf{E}_x f(X(t))| dt < +\infty, \quad x \in \mathbb{X}$$

(see [9], Definition 2.2). The action of the extended potential  $\mathcal{R}$  at a function  $f$  is given by

$$(2.12) \quad \mathcal{R}f(x) = \int_0^\infty \mathbf{E}_x f(X(t)) dt, \quad x \in \mathbb{X}.$$

For a function of two arguments  $f: \mathbb{X} \times \mathbb{R}^m \rightarrow \mathbb{R}$ , the argument  $y \in \mathbb{R}^m$  is constant in the definition of the extended potential  $\mathcal{R}f(x, y)$ , that is,

$$(2.13) \quad \mathcal{R}f(x, y) = \int_0^\infty \mathbf{E}_x f(X(t), y) dt, \quad x \in \mathbb{X}.$$

Note that the combinations of assumptions imposed on the process  $X$  in Section 2.2 and conditions that a function  $f$  belong to one of the classes defined above ( $H_{\phi,p}(d, \pi)$  or  $\mathbf{H}_{\phi,p,0}(d, \pi)$ , etc.) allow one to prove the existence of the corresponding generalized extended potential and to derive its properties; see Theorem 3.1 in [9].

### 3. MAIN RESULTS

The main results of the paper are stated in this section. The corresponding proofs are given in Section 4.

**Theorem 3.1.** *Assume that the following conditions hold for some premetrics  $d_1$  and  $d_2$ , some functions  $r_1, r_2, \phi_1, \phi_2, \psi_1$ , and  $\psi_2$ , and some numbers  $p_1, p_2 \in [1, +\infty)$  and  $q_1, q_2 \in (1, +\infty]$  such that*

$$\frac{1}{p_i} + \frac{1}{q_i} = 1, \quad i = 1, 2.$$

- (1) *The process  $X$  satisfies conditions  $\mathbf{E}(\mathbf{d}_i, \mathbf{r}_i, \psi_i)$  and  $\mathbf{M}_{\mathbf{q}_i}(\phi_i, \psi_i)$ ,  $i = 1, 2$ , as well as  $\mathbf{P}(\psi_1)$ .*
- (2) *The process  $X$  satisfies condition  $\mathbf{W}(\psi_i, \mathbf{Q}_i)$ ,  $i = 1, 2$ , with*

$$(3.1) \quad Q_1(t) = \sqrt{t}, \quad Q_2(t) = \int_0^t r_2^{1/p_2}(s) ds, \quad t \geq 1.$$

- (3) *The function  $r_1$  is such that*

$$(3.2) \quad \int_0^\infty r_1^{1/p_1}(t) dt < +\infty.$$

- (4) *The coefficient  $A: \mathbb{X} \times \mathbb{R}^m \rightarrow \mathbb{R}^m$  as a function of the argument  $y \in \mathbb{R}^m$  belongs to the class  $C^2$  for an arbitrary fixed  $x \in \mathbb{R}^m$ . Moreover,*

$$(3.3) \quad \text{the function } A_i(\cdot, y) \text{ is centered, } \quad y \in \mathbb{R}^m, \quad i = 1, \dots, m,$$

and

$$(3.4) \quad A_i \in \mathbf{H}_{\phi_1, p_1, 1}(d_1, \pi), \quad \partial_{y_i} A_j \in \mathbf{H}_{\phi_1, p_1, 1}(d_1, \pi), \quad \partial_{y_i y_j}^2 A_k \in \mathbf{H}_{\phi_1, p_1, 1}(d_1, \pi), \\ i, j, k = 1, \dots, m.$$

- (5) *The following inclusions hold:*

$$(3.5) \quad a_i \in \hat{\mathbf{H}}_{\phi_2, p_2, 1}(d_2, \pi), \quad A_i \partial_{y_i} \mathcal{R}A_j \in \hat{\mathbf{H}}_{\phi_2, p_2, 1}(d_2, \pi), \\ a_i \partial_{y_i} \mathcal{R}A_j \in \mathbf{H}_{\phi_2, p_2, 1}(d_2, \pi), \quad b_{ij} \partial_{y_i y_j}^2 \mathcal{R}A_k \in \mathbf{H}_{\phi_2, p_2, 1}(d_2, \pi), \\ i, j, k = 1, \dots, m,$$

and

$$(3.6) \quad b_{ij} \in \hat{\mathbf{H}}_{\phi_2, p_2, 2}(d_2, \pi), \quad A_i \mathcal{R}A_j \in \hat{\mathbf{H}}_{\phi_2, p_2, 2}(d_2, \pi), \quad a_i \mathcal{R}A_j \in \mathbf{H}_{\phi_2, p_2, 2}(d_2, \pi), \\ b_{ij} \partial_{y_j} \mathcal{R}A_k \in \mathbf{H}_{\phi_2, p_2, 2}(d_2, \pi), \quad i, j, k = 1, \dots, m,$$

$$(3.7) \quad b_{ij} \mathcal{R}A_k \in \mathbf{H}_{\phi_2, p_2, 3}(d_2, \pi), \quad i, j, k = 1, \dots, m.$$

(6) There exists a function  $\varrho: \mathbb{X} \rightarrow \mathbb{R}$  such that

$$|a(x, y)| + |A(x, y) + |b(x, y)| \leq \rho(x), \quad x \in \mathbb{X}, y \in \mathbb{R}^m,$$

and

$$(3.8) \quad \mathbb{E} \int_0^\infty e^{-\lambda t} \psi_1(X(t)) \varrho(X(t)) dt < \infty, \quad \lambda > 0.$$

Then

I. The family  $\{Y^\varepsilon, \varepsilon \in (0, 1]\}$  of solutions of equations (2.1) is weakly relatively compact in the sense of the convergence of finite dimensional distributions. That is, given an arbitrary sequence  $\{\varepsilon_n\}$ , there exists a subsequence  $\{\varepsilon_{n_k}\}$  such that the subsequence of random vectors

$$(Y^{\varepsilon_{n_k}}(t_1), \dots, Y^{\varepsilon_{n_k}}(t_M))$$

weakly converges in  $(\mathbb{R}^m)^M$  for all  $M \in \mathbb{N}$  and  $t_1, \dots, t_M \geq 0$ .

II. Every limit point  $Y$  of  $Y^\varepsilon, \varepsilon \rightarrow 0$ , is stochastically continuous and thus has a measurable modification. This modification is a solution of the martingale problem associated with the operator

$$(3.9) \quad \mathcal{L}f(y) = \sum_{i=1}^m \mathbf{A}_i(y) \frac{\partial}{\partial y_i} f(y) + \frac{1}{2} \sum_{i,j=1}^m \mathbf{B}_{ij}(y) \frac{\partial^2}{\partial y_i \partial y_j} f(y), \quad f \in C_0^\infty(\mathbb{R}^m),$$

where

$$\mathbf{A}_i(y) = \int_{\mathbb{X}} \left( a_i(x, y) + \sum_{k=1}^m A_k(x, y) \partial_{y_k} \mathcal{R}A_i(x, y) \right) \pi(dx)$$

and

$$\mathbf{B}_{ij}(y) = \int_{\mathbb{X}} (b_{ij}(x, y) + A_i(x, y) \mathcal{R}A_j(x, y) + A_j(x, y) \mathcal{R}A_i(x, y)) \pi(dx).$$

III. If the martingale problem associated with operator (3.9) is well posed (see [5], Chapter 4, §3), then  $Y^\varepsilon$  weakly converges as  $\varepsilon \rightarrow 0$  in the sense of the convergence of finite dimensional distributions to the unique (in the sense of the distributions) diffusion process  $Y$  with generator (3.9).

*Remark 3.1.* By Theorem 3.1 of [9], conditions (1), (3), and (4) imply the existence of the extended potentials  $\mathcal{R}A_i, i = 1, \dots, m$ , appearing in condition (5). Moreover, Theorem 3.2 of [9] implies that these potentials are twice continuously differentiable with respect to the argument  $y$ ; see the discussion in Section 4.1 for details.

*Remark 3.2.* An extra assumption in statement III that the martingale problem associated with operator (3.9) is well posed is not very restrictive (it can be checked directly by using Theorem 4.4.1 of [5], for example).

**Theorem 3.2.** Assume that conditions (1)–(5) of Theorem 3.1 hold with stronger versions  $\mathbf{S}(\psi_i, \mathbf{Q}_i), i = 1, 2$ , instead of  $\mathbf{W}(\psi_i, \mathbf{Q}_i), i = 1, 2$ , in (2).

We further assume that either assumption (6) of Theorem 3.1 holds or the family of random variables  $\psi(X(\tau)), \tau \in \mathcal{S}(T)$ , is uniformly integrable for all  $T \in \mathbb{R}^+$ .

Then the family  $\{Y^\varepsilon, \varepsilon \in (0, 1]\}$  is weakly relatively compact in  $\mathbb{D}(\mathbb{R}^+, \mathbb{R}^m)$ , and each limit point  $Y$  of  $Y^\varepsilon, \varepsilon \rightarrow 0$ , is a solution of the martingale problem associated with operator (3.9).

Moreover, if the martingale problem associated with operator (3.9) is well defined, then  $Y^\varepsilon$  weakly converges in  $C(\mathbb{R}^+, \mathbb{R}^d)$  as  $\varepsilon \rightarrow 0$  to the unique (in the sense of the distribution) diffusion process  $Y$  with generator (3.9).

4. PROOF

We briefly describe the main ideas of the proof. First we introduce the corrector  $u_f$  for an appropriate test function  $f: \mathbb{R}^m \rightarrow \mathbb{R}$  in order to obtain a representation of the form (4.4) written below. We use the results of paper [9] on the existence and properties of solutions of the *extended Poisson equation* to construct the corrector. The representation of the form (4.4) follows from the extended version of the Dynkin formula proved in Theorems 3.2 and 3.3 of [9].

Representation (4.4) is “insufficient” in the sense that it does not contain the martingale component  $M_f^\varepsilon$ . Nevertheless, this representation is sufficient for the proof of the weak compactness in the sense of the convergence of either finite dimensional distributions (under the assumptions of Theorem 3.1) or in the space  $\mathbb{D}(\mathbb{R}^+, \mathbb{R}^m)$  (under the assumptions of Theorem 3.2). Note that the direct proof of the weak compactness in the space  $C(\mathbb{R}^+, \mathbb{R}^m)$  is rather difficult in view of the absence of the martingale component  $M_f^\varepsilon$  in representation (4.4). This part of the proof differs essentially from the corresponding part of paper [7] where an analog of representation (4.4) is obtained with the help of Itô–Krylov’s formula which allows one to describe  $M_f^\varepsilon$ .

The last step in the proof, the description of limit points, also uses representation (4.4). Since we consider a weaker mode of convergence than the convergence in  $C(\mathbb{R}^+, \mathbb{R}^m)$ , the standard tools (such as “freezing the slow component”, for example; see [7]) are not applicable. Thus we propose a new trick that can be called “delaying the slow component” (see Remark 4.2 below).

Note that our method lies in the framework of the “martingale approach” proposed in [13] but does not rely on a priori uniform bounds and smoothness conditions imposed on solutions of the Poisson equation.

**4.1. Corrector: Construction and estimates.** Consider a function  $f \in C^3(\mathbb{R}^m)$  such that

$$(4.1) \quad |\nabla^r f(y)| \leq C(1 + |y|)^{-r}, \quad r = 1, 2, 3.$$

**Lemma 4.1.** *Let all the assumptions of Theorem 3.1 hold. Then equality (2.13) well defines the extended potentials*

$$(4.2) \quad \mathcal{R}(A_i \partial_{y_i} f)(x, y), \quad i = 1, \dots, m,$$

*being twice continuously differentiable functions with respect to the argument  $y$ . Moreover, putting*

$$(4.3) \quad u_f(x, y) = \sum_{i=1}^m \mathcal{R}(A_i \partial_{y_i} f)(x, y),$$

*we obtain the following representation:*

$$(4.4) \quad f(Y^\varepsilon(t)) + \varepsilon^{1/2} u_f(X^\varepsilon(t), Y^\varepsilon(t)) = \int_0^t K_f^\varepsilon(X^\varepsilon(s), Y^\varepsilon(s)) ds + M_f^\varepsilon(t),$$

*where  $M_f^\varepsilon$  is a martingale with respect to  $\mathbb{F}$ ,*

$$(4.5) \quad K_f^\varepsilon = H_f + \varepsilon^{1/2} J_f,$$

$$(4.6) \quad H_f \in \hat{\mathbf{H}}_{\phi_2, p_2, 0}(d_2, \pi), \quad J_f \in \mathbf{H}_{\phi_2, p_2, 0}(d_2, \pi).$$

*The same statements hold if all the assumptions of Theorem 3.2 hold.*

*Remark 4.1.* It is clear that, for a given  $i$  and for any given open ball, one can choose a test function  $f \in C_0^\infty(\mathbb{R}^m)$  such that  $\partial_{y_i} f \equiv 1$  in this ball. It is also clear that (4.1) holds for such a function  $f$ , and thus Lemma 4.1 implies that equality (2.13) well defines the extended potential  $\mathcal{R}A_i(x, y)$  being a twice continuously differentiable function with respect to the argument  $y$ . Moreover, equality (2.13) implies that

$$(4.7) \quad u_f(x, y) = \sum_i \mathcal{R}A_i(x, y) \partial_{y_i} f(y)$$

for an arbitrary function  $f \in C^3(\mathbb{R}^m)$  for which (4.1) holds.

*Proof.* Let all the assumptions of Theorem 3.1 hold. We show that one can apply Theorem 3.2 of [9] with

$$d = d_1, \quad \phi = \phi_1, \quad \psi = \psi_1, \quad r = r_1, \quad p = p_1, \quad q = q_1,$$

and with the function  $\tilde{f}(x, y) = A_i(x, y) \partial_{y_i} f(y)$ . Indeed, the assumptions of Theorem 3.1 imposed on the process  $X$  and on coefficients of the equation follow from conditions (1) and (3) of Theorem 3.1 and from assumptions (2.2) and (2.3). Further, the function  $\tilde{f}(\cdot, y)$  is centered for each  $y$  in view of condition (3.3). Finally, assumptions (3.7) and (3.8) of [9, Theorem 3.2] imposed on the function  $\tilde{f}$  follow from (4.1) and (3.4). Therefore, the first statement of [9, Theorem 3.2] yields the existence of twice differentiable (with respect to  $y$ ) potentials (4.2).

Now condition  $\mathbf{W}(\psi_1, \mathbf{Q}_1)$  with the function  $Q_1$  defined by (3.1) implies conditions (3.9) of [9, Theorem 3.2] and (3.15) of [9, Theorem 3.3]. Applying condition (6) of Theorem 3.1 we prove that all the assumptions of [9, Theorem 3.3] hold.

Now [9, Theorem 3.3] implies that

$$(4.8) \quad \begin{aligned} & u_f(X^\varepsilon(t), Y^\varepsilon(t)) \\ &= -\varepsilon^{-1} \int_0^t \sum_i A_i(X^\varepsilon(s), Y^\varepsilon(s)) \partial_{y_i} f(X^\varepsilon(s)) ds \\ &+ \int_0^t \sum_i a_i(X^\varepsilon(s), Y^\varepsilon(s)) \partial_{y_i} u_f(X^\varepsilon(s), Y^\varepsilon(s)) ds \\ &+ \varepsilon^{-1/2} \int_0^t \sum_i A_i(X^\varepsilon(s), Y^\varepsilon(s)) \partial_{y_i} u_f(X^\varepsilon(s), Y^\varepsilon(s)) ds \\ &+ \frac{1}{2} \int_0^t \sum_{i,j} b_{ij}(X^\varepsilon(s), Y^\varepsilon(s)) \partial_{y_i y_j}^2 u_f(X^\varepsilon(s), Y^\varepsilon(s)) ds + M_f^{1,\varepsilon}(t), \end{aligned}$$

where  $M_f^{1,\varepsilon}$  is a martingale. On the other hand, we deduce from Itô's formula that

$$(4.9) \quad \begin{aligned} f(Y^\varepsilon(t)) &= \int_0^t \sum_i a_i(X^\varepsilon(s), Y^\varepsilon(s)) \partial_{y_i} f(Y^\varepsilon(s)) ds \\ &+ \varepsilon^{-1/2} \int_0^t \sum_i A_i(X^\varepsilon(s), Y^\varepsilon(s)) \partial_{y_i} f(Y^\varepsilon(s)) ds \\ &+ \frac{1}{2} \int_0^t \sum_{i,j} b_{ij}(X^\varepsilon(s), Y^\varepsilon(s)) \partial_{y_i y_j}^2 f(Y^\varepsilon(s)) ds + M_f^{2,\varepsilon}(t), \end{aligned}$$

where  $M_f^{2,\varepsilon}$  is a local martingale.

Further we take into account (4.1) and use the restrictions imposed on the coefficients  $a$ ,  $A$ , and  $b$ , in (3.4)–(3.6) to prove that  $M_f^{2,\varepsilon}$  is a martingale in view of the properties of  $X$ . The reasoning leading to this result is similar to that given in Section 4.2 of



the paper [9], thus we omit the details. Now we multiply (4.8) by  $\varepsilon^{1/2}$  and add the result to (4.9); this proves equality (4.4). The explicit expression for the function  $K_f^\varepsilon$  follows by adding the corresponding integrands in (4.8) and (4.9) and by differentiating in (4.7) (as we have already shown, each function  $\mathcal{R}A_i(x, y)$  is twice continuously differentiable with respect to  $y$ ). After some algebra, we obtain  $K_f^\varepsilon$  defined by equality (4.5) with

$$\begin{aligned}
 H_f(x, y) &= \sum_i \left[ a_i(x, y) + \sum_j A_j(x, y) \partial_{y_j} \mathcal{R}A_i(x, y) \right] \partial_{y_i} f(y) \\
 &\quad + \sum_{i,j} \left[ \frac{1}{2} b_{ij}(x, y) + A_i(x, y) \mathcal{R}A_j(x, y) \right] \partial_{y_i y_j}^2 f(y), \\
 J_f(x, y) &= \sum_i \left[ \sum_j a_j(x, y) \partial_{y_j} \mathcal{R}A_i(x, y) + \frac{1}{2} \sum_{j,k} b_{jk}(x, y) \partial_{y_j y_k}^2 \mathcal{R}A_i(x, y) \right] \partial_{y_i} f(y) \\
 &\quad + \sum_{i,j} \left[ a_i(x, y) \mathcal{R}A_j(x, y) + \sum_k b_{jk}(x, y) \partial_{y_k} \mathcal{R}A_i(x, y) \right] \partial_{y_i y_j}^2 f(y) \\
 &\quad + \frac{1}{2} \sum_{i,j,k} [b_{ij}(x, y) \mathcal{R}A_k(x, y)] \partial_{y_i y_j y_k}^3 f(y).
 \end{aligned}$$

The explicit expressions for  $H_f$  and  $J_f$  imply (4.6) by conditions (3.5)–(3.7) and (4.1). This completes the proof.

In the case of Theorem 3.2, we use the last statement of [9, Theorem 3.2] instead of [9, Theorem 3.3] to check equality (4.8). Assumption (3.10) of [9, Theorem 3.2] follows from condition  $\mathbf{S}(\psi_1, \mathbf{Q}_1)$ , whence we derive equality (4.8), where  $M_f^{1,\varepsilon}$  is a local martingale. This equality can be rewritten as follows:

$$u_f(X^\varepsilon(t), Y^\varepsilon(t)) = \int_0^t \Upsilon_f^\varepsilon(X^\varepsilon(s), Y^\varepsilon(s)) ds + M_f^{1,\varepsilon}(t).$$

To prove that  $M_f^{1,\varepsilon}$  is an (ordinary) martingale, we check that

$$(4.10) \quad \mathbb{E} \int_0^t |\Upsilon_f^\varepsilon(X^\varepsilon(s), Y^\varepsilon(s))| ds < \infty, \quad t \geq 0$$

(see [9, Theorem 3.2]). Similarly to the above reasoning for  $K_f^\varepsilon$ , we obtain  $\Upsilon_f^\varepsilon \in \mathbf{H}_{\phi_2, p_2, 0}(d_2, \pi)$ . Then, by (A.1),

$$\mathbb{E} |\Upsilon_f^\varepsilon(X^\varepsilon(s), Y^\varepsilon(s))| \leq C_\varepsilon \left( 1 + r^{1/p_2}(s) \mathbb{E} \psi(X(0)) \right),$$

which implies relation (4.10). This completes the proof of (4.8); the rest of the proof is unchanged.  $\square$

**4.2. The probability of sojourn in a compact set and estimates for the increments of the process  $Y^\varepsilon$ .** Using representation (4.4), we obtain the following assertions. The proofs of Lemmas 4.2–4.4 are given in the forthcoming paper [10].

**Lemma 4.2.** *Let assumptions (1)–(5) of Theorem 3.1 hold. Assume that (4.4) is satisfied for an arbitrary function  $f \in C^3(\mathbb{R}^m)$  for which inequality (4.1) holds.*

*Then, for an arbitrary  $T \in \mathbb{R}^+$ ,*

$$(4.11) \quad \sup_{\varepsilon \in (0,1]} \sup_{t \in [0,T]} \mathbb{P}(|Y^\varepsilon(t)| > R) \rightarrow 0, \quad R \rightarrow +\infty.$$

If additionally the process  $X$  satisfies condition  $\mathbf{S}(\psi_1, \mathbf{Q}_1)$ , then

$$(4.12) \quad \sup_{\varepsilon \in (0,1]} \mathbf{P} \left( \sup_{t \in [0,T]} |Y^\varepsilon(t)| > R \right) \rightarrow 0, \quad R \rightarrow +\infty.$$

**Lemma 4.3.** *Let all the assumptions of Lemma 4.2 hold. Then, given an arbitrary  $r$  and  $R > 0$ , there exists a constant  $C_{r,R} \in \mathbb{R}^+$  such that the inequality*

$$(4.13) \quad \mathbf{P} [|Y^\varepsilon(\tau + u) - Y^\varepsilon(\tau)| > r \mid \mathcal{F}_\tau^\varepsilon] \leq C_{r,R}(u + \varepsilon\psi_2(X^\varepsilon(\tau)))$$

holds almost surely in the set  $\{|Y^\varepsilon(\tau)| \leq R\}$  for all

$$u \in (0, 1], \quad \varepsilon \in (0, 1], \quad \tau \in \bigcup_{T \in \mathbb{R}^+} \mathcal{S}^\varepsilon(T).$$

**Lemma 4.4.** *Under the assumptions of Lemma 4.2,*

$$(4.14) \quad \limsup_{\varepsilon \rightarrow 0} \sup_{t \in [0,T]} \sup_{0 \leq u \leq \delta} \mathbf{E} [|Y^\varepsilon(t + u) - Y^\varepsilon(t)| \wedge 1] \rightarrow 0, \quad \delta \rightarrow 0,$$

for an arbitrary  $T \in \mathbb{R}^+$ .

If additionally the process  $X$  satisfies condition  $\mathbf{S}(\psi_i, \mathbf{Q}_i)$ ,  $i = 1, 2$ , then

$$(4.15) \quad \limsup_{\varepsilon \rightarrow 0} \sup_{\tau \in \mathcal{S}^\varepsilon(T)} \sup_{0 \leq u \leq \delta} \mathbf{E} [|Y^\varepsilon(\tau + u) - Y^\varepsilon(\tau)| \wedge 1] \rightarrow 0, \quad \delta \rightarrow 0.$$

**4.3. Weak relative compactness.** The family of distributions of the processes  $Y^\varepsilon$  is continuous in  $C(\mathbb{R}^+, \mathbb{R}^m)$  with respect to the parameter  $\varepsilon \in (0, 1]$ . Thus the weak relative compactness of the family  $\{Y^\varepsilon\}_{\varepsilon \in (0,1]}$  follows from that of an arbitrary sequence  $\{Y^{\varepsilon_n}\}$ ,  $\varepsilon_n \rightarrow 0$ .

The relative compactness in the sense of the convergence of finite dimensional distributions follows from relations (4.12) and (4.14) (the reasoning is standard and we omit it; see §1 of Chapter 5 in [14] for details). Note that each limit point  $Y$  is a stochastic process being continuous in probability by (4.14).

We use the following sufficient condition given in [5, Chapter 3] to obtain the relative compactness in  $\mathbb{D}(\mathbb{R}^+, \mathbb{R}^m)$ . Condition (c) of Theorem 8.6 in [5, Chapter 3] combined with Remark 8.7 (b) therein can be rewritten as follows: given an arbitrary  $T > 0$ , there exists a number  $\beta > 0$  such that

$$(4.16) \quad \limsup_{n \rightarrow \infty} \sup_{\tau \in \mathcal{S}^{\varepsilon_n}(T)} \sup_{0 \leq u \leq \delta} \mathbf{E} \left[ \sup_{0 \leq v \leq \tau \wedge \delta} (|Y^{\varepsilon_n}(\tau) - Y^{\varepsilon_n}(\tau - v)|^\beta \wedge 1) \right. \\ \left. \times (|Y^{\varepsilon_n}(\tau + u) - Y^{\varepsilon_n}(\tau)|^\beta \wedge 1) \right] \rightarrow 0, \quad \delta \rightarrow 0.$$

Relation (4.16) is equivalent to the relative compactness in  $\mathbb{D}(\mathbb{R}^+, \mathbb{R}^m)$  of the sequence  $\{Y^{\varepsilon_n}\}$  under an additional assumption that condition (a) of Theorem 7.2 in [5, Chapter 3] holds. In the case under consideration, condition (a) of Theorem 7.2 in [5, Chapter 3] coincides with (4.11). It is also clear that (4.15) implies (4.16) with  $\beta = 1$ , and this is what had to be proved.

**4.4. Description of limit points.** Let  $Y$  be a limit point of a sequence  $Y^{\varepsilon_n}$ ,  $\varepsilon_n \rightarrow 0$ , in the sense of the weak convergence of finite dimensional distributions. As mentioned above, the process  $Y$  is continuous in probability by (4.14), that is, there is a measurable version of  $Y$ , and thus one can assume that  $Y$  is measurable from the beginning.

To prove that  $Y$  is a solution of the martingale problem for  $\mathcal{L}$ , we check that

$$(4.17) \quad \mathbf{E} \left[ \left( f(Y(t_2)) - f(Y(t_1)) - \int_{t_1}^{t_2} \mathcal{L}f(Y(s)) ds \right) \Phi_{t_1}(Y) \right] = 0$$

for all  $f \in C_0^\infty(\mathbb{R}^m)$  and  $t_2 > t_1 > 0$ , where  $\Phi_{t_1}(Y)$  denotes a random variable such that

$$\Phi_{t_1}(Y) = \prod_{j=1}^m h_j(Y(s_j))$$

with  $m \in \mathbb{N}$ ,  $s_1, \dots, s_m < t_1$ , and with some bounded continuous functions

$$h_j, \quad j = 1, \dots, m$$

(see a discussion related to inclusions (3.4) in [5, Chapter 4]).

Now we prove equality (4.17). To have shorter notation, we write

$$X_n = X^{\varepsilon_n} \quad \text{and} \quad Y_n = Y^{\varepsilon_n}.$$

**Lemma 4.5.** *The function  $\mathcal{L}f$  is continuous for an arbitrary  $f \in C_0^\infty(\mathbb{R}^m)$ .*

*Proof.* First we write the representation

$$(4.18) \quad \mathcal{L}f(y) = \int_{\mathbb{X}} H_f(x, y) \pi(dx)$$

(see the proof of Lemma 4.1 where a similar result is given for  $H_f$ ). The function  $H_f(x, \cdot)$  is continuous for  $\pi$ -almost sure  $x \in \mathbb{X}$ . The latter property follows from the Lipschitz property of the functions  $a$ ,  $A$ , and  $\sigma$  with respect to the argument  $y$  (see (2.3)), since the functions  $\mathcal{R}A_i$ ,  $i = 1, \dots, m$ , are continuously differentiable with respect to the same argument (see Remark 4.1). Now the Lebesgue dominated convergence theorem proves that the function

$$\mathcal{L}^N f(y) := \int_{\mathbb{X}} (|H_f(x, y)| \wedge N) \text{sign } H_f(x, y) \pi(dx)$$

is continuous with respect to  $y$  for an arbitrary  $N > 0$ . On the other hand,

$$H_f \in \hat{\mathbf{H}}_{\phi_2, p_2, 0}(d_2, \pi)$$

in view of Lemma 4.1. Applying relation (2.11), we deduce from the definition of the latter class that  $\mathcal{L}^N f$  uniformly converges to  $\mathcal{L}f$  as  $N \rightarrow \infty$ . This means that  $\mathcal{L}f$ , as a uniform limit of continuous functions, is continuous.  $\square$

It is also clear that  $\mathcal{L}f$  has a compact support if  $f \in C_0^\infty(\mathbb{R}^m)$ . This result combined with the latter lemma yields the boundedness and uniform continuity of  $\mathcal{L}f$ . Taking into account relation (4.14), we obtain the following assertion whose proof is standard and thus is omitted.

**Lemma 4.6.** *If*

$$(4.19) \quad \mathbb{E} \left[ \left( f(Y_n(t_2)) - f(Y_n(t_1)) - \int_{t_1}^{t_2} \mathcal{L}f(Y_n(s)) ds \right) \Phi_{t_1}(Y_n) \right] \rightarrow 0, \quad n \rightarrow \infty,$$

*then relation (4.17) holds.*

Put

$$\Delta_f(x, y) = H_f(x, y) - \mathcal{L}f(y).$$

It is clear that (4.1) holds for the function  $f \in C_0^\infty(\mathbb{R}^m)$ , and thus such a function admits representation (4.4). Using (4.4) we obtain

$$\begin{aligned}
 & \mathbb{E} \left[ \left( f(Y_n(t_2)) - f(Y_n(t_1)) - \int_{t_1}^{t_2} \mathcal{L}f(Y_n(s)) ds \right) \Phi_{t_1}(Y_n) \right] \\
 &= \varepsilon_n^{1/2} \mathbb{E} \left[ (u_f(X_n(t_1), Y_n(t_1)) - u_f(X_n(t_2), Y_n(t_2))) \Phi_{t_1}(Y_n) \right] \\
 (4.20) \quad &+ \varepsilon_n^{1/2} \mathbb{E} \left[ \left( \int_{t_1}^{t_2} J_f(X_n(s), Y_n(s)) ds \right) \Phi_{t_1}(Y_n) \right] \\
 &+ \mathbb{E} \left[ \left( \int_{t_1}^{t_2} \Delta_f(X_n(s), Y_n(s)) ds \right) \Phi_{t_1}(Y_n) \right].
 \end{aligned}$$

This, in particular, means that the asymptotic smallness of the three terms on the right hand side of (4.20) completes the proof.

The bound

$$|u_f(x, y)| \leq C\psi_1(x)$$

holds by statement I of [9, Theorem 3.1] applied to the functions  $\tilde{f}_i = A\partial_{y_i}f$  with  $\phi = \phi_1$  and  $\psi = \psi_1$ ,

Since  $|\Phi_{t_1}(Y_n)| \leq C$  by definition, the convergence to zero as  $n \rightarrow \infty$  of the first term on the right hand side of (4.20) follows in view of condition  $\mathbf{W}(\psi_1, \mathbf{Q}_1)$ .

Next, we apply (A.1) with  $\phi = \phi_2$  and  $\psi = \psi_2$  and use the second relation in (4.6). As a result,

$$(4.21) \quad \mathbb{E} \int_{t_1}^{t_2} |J_f^\varepsilon(X^\varepsilon(s), Y^\varepsilon(s))| ds \leq C(t_2 - t_1)(1 + \mathbb{E} \psi_2(X(0))).$$

Since  $\Phi_{t_1}$  is bounded, the condition  $\mathbf{W}(\psi_2, \mathbf{Q}_2)$  yields the convergence to zero as  $n \rightarrow \infty$  of the second term on the right hand side of (4.20).

The result on the asymptotic smallness of the third term on the right hand side of (4.20) is stated as a separate assertion.

**Lemma 4.7.** *Under the above assumptions,*

$$(4.22) \quad \mathbb{E} \left[ \left( \int_{t_1}^{t_2} \Delta_f(X_n(s), Y_n(s)) ds \right) \Phi_{t_1}(Y_n) \right] \rightarrow 0, \quad n \rightarrow \infty.$$

*Proof.* The expression on the left hand side of (4.22) with  $\gamma \in (0, t_2 - t_1)$  can be rewritten as follows:

$$\begin{aligned}
 I_{n,f} &= I_{n,f,\gamma}^1 + I_{n,f,\gamma}^2 + I_{n,f,\gamma}^3 \\
 &= \mathbb{E} \left[ \left( \int_{t_1}^{t_1+\gamma} \Delta_f(X_n(s), Y_n(s)) ds \right) \Phi_{t_1}(Y_n) \right] \\
 &+ \mathbb{E} \left[ \left( \int_{t_1+\gamma}^{t_2} \Delta_f(X_n(s), Y_n(s - \gamma)) ds \right) \Phi_{t_1}(Y_n) \right] \\
 &+ \mathbb{E} \left[ \left( \int_{t_1+\gamma}^{t_2} (\Delta_f(X_n(s), Y_n(s)) - \Delta_f(X_n(s), Y_n(s - \gamma))) ds \right) \Phi_{t_1}(Y_n) \right].
 \end{aligned}$$

The integrals  $I_{n,f,\gamma}^i$ ,  $i = 1, 2, 3$ , are estimated separately.

*Remark 4.2.* We briefly describe the main idea for obtaining such a bound. First we change  $Y_n(s)$  by  $Y_n(s - \gamma)$  and get an expression in the second term where the arguments of the processes  $X_n$  and  $Y_n$  differ by  $\gamma$ . This allows one to estimate this expression by using the ergodic properties of the process  $X$ ; see inequality (4.25) below. This trick is called the “delay of the slow component”.

By the first relation in (4.6),

$$(4.23) \quad \Delta_f \in \hat{\mathbf{H}}_{\phi_2, p_2, 0}(d_2, \pi).$$

Then the reasoning similar to that used in the proof of (4.21) leads to the bound

$$(4.24) \quad |I_{n, f, \gamma}^1| \leq C\gamma.$$

Further, the function  $\Delta_f(\cdot, y)$  is centered for all  $y \in \mathbb{R}^m$  by construction. Using (4.23), the Markov property of the process  $X_n$ , and bound (4.1) of paper [9] with  $\phi = \phi_2$  and  $\psi = \psi_2$ , we get

$$\mathbb{E} [\Delta_f(X_n(s), Y_n(s - \gamma)) \mid \mathcal{F}_{s-\gamma}^{\varepsilon_n}] \leq Cr_2^{1/p_2}(\gamma/\varepsilon_n)\psi_2(X_n(s - \gamma)).$$

Since the bounded function  $\Phi_{t_1}(Y_n)$  is measurable with respect to  $\mathcal{F}_{s-\gamma}^{\varepsilon_n}$  for an arbitrary  $s \geq t_1 + \gamma$ , we obtain the following bound:

$$(4.25) \quad |I_{n, f, \gamma}^2| \leq Cr_2^{1/p_2}(\gamma/\varepsilon_n) \mathbb{E} \int_{t_1+\gamma}^{t_2} \psi_2(X_n(s - \gamma)) ds.$$

Next, since  $\Phi_{t_1}$  is bounded,

$$(4.26) \quad \begin{aligned} |I_{n, f, \gamma}^3| &\leq C \mathbb{E} \int_{t_1+\gamma}^{t_2} (|\Delta_f(X_n(s), Y_n(s)) - \Delta_f(X_n(s), Y_n(s - \gamma))| \wedge N) ds \\ &\quad + C \mathbb{E} \int_{t_1+\gamma}^{t_2} (\Xi_N(\Delta_f(X_n(s), Y_n(s))) + \Xi_N(\Delta_f(X_n(s), Y_n(s - \gamma)))) ds \end{aligned}$$

for an arbitrary  $N \geq 1$ . Recall that the function  $\Xi_N$  is introduced in Section 2.3. Since  $\Xi_N$  is a Lipschitz function with the Lipschitz constant 1, we get

$$\|\Xi_N \circ \Delta_f\|_{\phi_2, d_2, p_2, 0} \leq \|\Delta_f\|_{\phi_2, d_2, p_2, 0}.$$

This allows us to use Proposition A.1 with the function  $\Xi_N \circ \Delta_f$  in place of  $f$  and with  $Y_n(s)$  and then with  $Y_n(s - \gamma)$  in place of the random variable  $\zeta$ . As a result,

$$(4.27) \quad \begin{aligned} &\mathbb{E} (\Xi_N(\Delta_f(X_n(s), Y_n(s))) + \Xi_N(\Delta_f(X_n(s), Y_n(s - \gamma)))) \\ &\leq \varkappa_{f, N} + Cr_2^{1/p_2}(s/\varepsilon_n) \mathbb{E} \psi_2(X(0)), \end{aligned}$$

where

$$\varkappa_{f, N} = \int_{\mathbb{X}} \sup_{y \in \mathbb{R}^m} \Xi_N(\Delta_f(x, y)) \pi(dx).$$

Moreover,

$$(4.28) \quad \varkappa_{f, N} \rightarrow 0, \quad N \rightarrow \infty,$$

since (4.23) holds for  $\Delta_f$ .

The function  $\Delta_f(x, \cdot)$  is continuous for each  $x \in \mathbb{X}$  by construction, which means that this function is uniformly continuous in the ball  $\{y: |y| \leq R\}$  for an arbitrary  $R$ . Therefore

$$\Theta_R(x, \beta) := \sup_{|y_1| \leq R, |y_2 - y_1| \leq \beta} |\Delta_f(x, y_2) - \Delta_f(x, y_1)| \rightarrow 0, \quad \beta \rightarrow 0,$$

for all  $x \in \mathbb{X}$ . Hence, given an arbitrary  $\delta > 0$ , there exists  $\beta_{\delta, R} > 0$  such that

$$(4.29) \quad \pi(\{x: \Theta_R(x, \beta_{\delta, R}) > \delta\}) < \delta.$$

By Lemma 4.2, there exists  $R_\delta \in \mathbb{R}^+$  such that

$$(4.30) \quad \mathbb{P}(|Y_n(s)| > R_\delta) < \delta, \quad s \in [t_1, t_2].$$

Put  $\beta_\delta = \beta_{\delta, R_\delta}$  and

$$A_{s, \delta} = \{|Y_n(s)| \leq R_\delta, |Y_n(s) - Y_n(s - \delta)| \leq \beta_\delta\}.$$

Applying Proposition A.2 with  $\zeta_1 = Y_n(s)$ ,  $\zeta_2 = Y_n(s - \gamma)$ , and  $A = A_{s,\delta}$ , we obtain

$$\begin{aligned}
 & \mathbf{E} \left( \left| \Delta_f(X_n(s), Y_n(s)) - \Delta_f(X_n(s), Y_n(s - \gamma)) \right| \wedge N \right) \mathbb{1}_{A_{s,\delta}} \\
 & \leq \int_{\mathbb{X}} \sup_{|y_1| \leq R_\delta, |y_1 - y_2| \leq \beta_\delta} \left( \left| \Delta_f(x, y_1) - \Delta_f(x, y_2) \right| \wedge N \right) \pi(dx) \\
 (4.31) \quad & + Cr_2^{1/p_2}(s/\varepsilon_n) \mathbf{E} \psi_2(X(0)) \\
 & \leq \delta + 2N\pi(\{x: \Theta_{R_\delta}(x, \beta_\delta) > \delta\}) + Cr_2^{1/p_2}(s/\varepsilon_n) \mathbf{E} \psi_2(X(0)) \\
 & \leq (2N + 1)\delta + Cr_2^{1/p_2}(s/\varepsilon_n) \mathbf{E} \psi_2(X(0)).
 \end{aligned}$$

Note that the latter inequality holds in view of (4.29).

By Lemma 4.3, there exists  $C_\delta \in \mathbb{R}^+$  such that

$$\mathbf{P} \left[ |Y_n(s) - Y_n(s - \gamma)| > \beta_\delta \mid \mathcal{F}_{s-\gamma}^{\varepsilon_n} \right] \leq C_\delta(\gamma + \varepsilon_n \psi_2(X_n(s - \gamma)))$$

almost surely in the set  $\{|Y_n(s - \gamma)| \leq R_\delta\}$  for all  $s \in [t_1 + \gamma, t_2]$ . This inequality together with (4.30) and (4.31) implies that

$$\begin{aligned}
 (4.32) \quad & \mathbf{E} \left( \left| \Delta_f(X_n(s), Y_n(s)) - \Delta_f(X_n(s), Y_n(s - \gamma)) \right| \wedge N \right) \\
 & \leq (2N + 2)\delta + Cr_2^{1/p_2}(s/\varepsilon_n) \mathbf{E} \psi_2(X(0)) + C_\delta(\gamma + \varepsilon_n \mathbf{E} \psi_2(X_n(s))).
 \end{aligned}$$

Now we fix  $\gamma_* \in (0, t_2 - t_1)$  and consider the averages of the bounds (4.24)–(4.27) and (4.32) with respect to  $\gamma \in [0, \gamma_*]$ . Hence

$$\begin{aligned}
 |I_{n,f}| & \leq C\gamma_* + C_\delta\gamma_* + (2N + 2)\delta + \varkappa_{f,N} + C \mathbf{E} \psi_2(X(0)) \sup_{s \geq t_1} r_2^{1/p_2}(s/\varepsilon_n) \\
 & + C_\delta\varepsilon_n \int_{t_1}^{t_2} \mathbf{E} \psi_2(X_n(s)) ds \\
 (4.33) \quad & + \frac{C}{\gamma_*} \left( \int_0^{\gamma_*} r_2^{1/p_2}(\gamma/\varepsilon_n) d\gamma \right) \left( \int_{t_1}^{t_2} \mathbf{E} \psi_2(X_n(s)) ds \right) \\
 & \leq (C + C_\delta)\gamma_* + (2N + 2)\delta + \varkappa_{f,N} + C \mathbf{E} \psi_2(X(0)) \sup_{t \geq t_1/\varepsilon_n} r_2^{1/p_2}(t) \\
 & + \left( C_\delta + \frac{C}{\gamma_*} \int_0^\infty r_2^{1/p_2}(s) ds \right) \left( \varepsilon_n \int_{t_1}^{t_2} \mathbf{E} \psi_2(X_n(s)) ds \right).
 \end{aligned}$$

Then

$$\sup_{t \geq t_1/\varepsilon} r_2^{1/p_2}(t) \rightarrow 0, \quad \varepsilon \rightarrow 0,$$

since  $t_1 > 0$  and  $r_2(t) \rightarrow 0$  as  $t \rightarrow \infty$ . Thus

$$\limsup_{n \rightarrow \infty} |I_{n,f}| \leq (C + C_\delta)\gamma_* + (2N + 2)\delta + \varkappa_{f,N}$$

by (4.33) and in view of condition  $\mathbf{W}_2(\psi_2, \mathbf{Q}_2)$ .

We obtain the result desired by passing to the limit as  $\gamma_* \rightarrow 0$ , then as  $\delta \rightarrow 0$ , and finally as  $N \rightarrow \infty$ . When passing to the limit as  $N \rightarrow \infty$ , we use relation (4.28).  $\square$

**4.5. The completion of the proof.** Statements I and II of Theorem 3.1 are already proved in Sections 4.3 and 4.4, and this implies statement III of the same theorem.

We also proved that, under the assumptions of Theorem 3.2, the family  $\{Y^\varepsilon\}$  is relatively compact in  $\mathbb{D}(\mathbb{R}^+, \mathbb{R}^m)$ . This implies that if the martingale problem associated with (3.9) is well posed, then  $Y^\varepsilon$  weakly converges in  $\mathbb{D}(\mathbb{R}^+, \mathbb{R}^m)$  as  $\varepsilon \rightarrow 0$ . Since  $Y^\varepsilon$ ,  $\varepsilon \in (0, 1]$ , as well as the limit diffusion process  $Y$  have continuous trajectories, the family  $Y^\varepsilon$  weakly converges in  $C(\mathbb{R}^+, \mathbb{R}^m)$  as  $\varepsilon \rightarrow 0$ .  $\square$

APPENDIX A. AUXILIARY ESTIMATES

**Proposition A.1.** *Let  $1/p + 1/q = 1$  and let conditions  $\mathbf{E}(\mathbf{d}, \mathbf{r}, \psi)$  and  $\mathbf{M}_q(\phi, \psi)$  hold for  $X$ . Also let  $f \in \mathbf{H}_{\phi, \mathbf{p}, \mathbf{0}}(\mathbf{d}, \pi)$ . If  $\zeta$  is an arbitrary random vector assuming values in  $\mathbb{R}^m$  and being defined on the same probability space as the process  $X$ , then*

$$(A.1) \quad \mathbf{E} |f(X(t), \zeta)| \leq \|f\|_{\pi, 0} + 2\|f\|_{\phi, d, p, 0} r^{1/p}(t) \left( \int_{\mathbb{X}} \psi \, d\pi \right)^{1/q} \mathbf{E} \psi(X(0)), \quad t \in \mathbb{R}^+.$$

*Proof.* Condition  $\mathbf{E}(\mathbf{d}, \mathbf{r}, \psi)$  implies the following bound for the distribution  $\mu_t$  of the random variable  $X(t)$ :

$$(A.2) \quad d(\mu_t, \pi) \leq r(t) \mathbf{E} \psi(X(0)).$$

Consider an arbitrary pair  $(\xi, \eta) \in C(\mu_t, \pi)$ . By the definition of the class  $C(\mu_t, \pi)$ , the first element of this pair has the same distribution as the first element of the pair  $(X(t), \zeta)$ . Then, on a certain probability space, there exists a triple  $(\hat{\xi}, \hat{\eta}, \hat{\zeta})$  such that the pair  $(\hat{\xi}, \hat{\eta})$  is identically distributed with  $(\xi, \eta)$ , while the pair  $(\hat{\xi}, \hat{\zeta})$  is identically distributed with  $(X(t), \zeta)$  (see [12, Section 11.8, Problem 8]).

Denote by  $\hat{\mathbf{E}}$  the mathematical expectation in this new probability space. Then

$$(A.3) \quad \begin{aligned} \mathbf{E} |f(X(t), \zeta)| &= \hat{\mathbf{E}} |f(\hat{\xi}, \hat{\zeta})| \leq \hat{\mathbf{E}} |f(\hat{\eta}, \hat{\zeta})| + \hat{\mathbf{E}} |f(\hat{\xi}, \hat{\zeta}) - f(\hat{\eta}, \hat{\zeta})| \\ &\leq \hat{\mathbf{E}} \sup_y |f(\hat{\eta}, y)| + \hat{\mathbf{E}} |f(\hat{\xi}, \hat{\zeta}) - f(\hat{\eta}, \hat{\zeta})| \\ &= \|f\|_{\pi, 0} + \hat{\mathbf{E}} |f(\hat{\xi}, \hat{\zeta}) - f(\hat{\eta}, \hat{\zeta})|. \end{aligned}$$

Note that we used the property that  $\hat{\eta}$  has the same distribution as  $\pi$ . Thus

$$|f(x, y) - f(x', y)| \leq \|f\|_{\phi, p, d, 0} d^{1/p}(x, x')(\phi(x) + \phi(x')), \quad x, x' \in \mathbb{X}, y \in \mathbb{R}^m.$$

Substituting  $x = \hat{\xi}$ ,  $x' = \hat{\eta}$ , and  $y = \hat{\zeta}$  and applying (A.3), we conclude that

$$(A.4) \quad \begin{aligned} \mathbf{E} |f(X(t), \zeta)| &\leq \|f\|_{\pi, 0} + \|f\|_{\phi, p, d, 0} \hat{\mathbf{E}} d^{1/p}(\hat{\xi}, \hat{\eta})(\phi(\hat{\xi}) + \phi(\hat{\eta})) \\ &\leq \|f\|_{\pi, 0} + \|f\|_{\phi, d, p, 0} (\mathbf{E} d(\xi, \eta))^{1/p} (\mathbf{E}(\phi(\xi) + \phi(\eta))^q)^{1/q}, \end{aligned}$$

since  $(\hat{\xi}, \hat{\eta})$  has the same distribution as  $(\xi, \eta)$ . Note also that here we used the Hölder inequality. Since the function  $z \mapsto z^q$  is convex,

$$\begin{aligned} \mathbf{E} (\phi(\xi) + \phi(\eta))^q &\leq 2^{q-1} (\mathbf{E} \phi^q(\xi) + \mathbf{E} \phi^q(\eta)) \\ &= 2^{q-1} \left( \mathbf{E} \int_{\mathbb{X}} \phi^q(y) P_t(X(0), dy) + \int_{\mathbb{X}} \left( \int_{\mathbb{X}} \phi^q(y) P_t(z, dy) \right) \pi(dz) \right) \\ &= 2^{q-1} \left( \mathbf{E} \psi(X(0)) + \int_{\mathbb{X}} \psi \, d\pi \right) \leq 2^q \left( \int_{\mathbb{X}} \psi \, d\pi \right) \mathbf{E} \psi(X(0)) \end{aligned}$$

by condition  $\mathbf{M}_q(\phi, \psi)$ . The latter inequality is valid in view of  $\psi \geq 1$ . Then inequality (A.4) can be written as follows:

$$(A.5) \quad \mathbf{E} |f(X(t), \zeta)| \leq \|f\|_{\pi} + 2\|f\|_{\phi, p} (\mathbf{E} d(\xi, \eta))^{1/p} \left( \int_{\mathbb{X}} \psi \, d\pi \right)^{1/q} (\mathbf{E} \psi(X(0)))^{1/q}.$$

Since  $(\xi, \eta) \in C(\mu_t, \pi)$  is arbitrary, relations (A.2) and (A.5) complete the proof.  $\square$

**Proposition A.2.** *Let all the assumptions of Proposition A.1 hold. Also let  $\zeta_1$  and  $\zeta_2$  be two random vectors assuming values in the space  $\mathbb{R}^m$  and that they be defined on the same probability space  $(\Omega, \mathcal{F}, \mathbf{P})$  as the process  $X$ . We further assume that the set  $A \in \mathcal{F}$  is such that*

$$|\zeta_1| \leq R \quad \text{and} \quad |\zeta_1 - \zeta_2| \leq \beta$$

almost surely in the set  $A$  for some  $\beta, R > 0$ . Then

$$\begin{aligned}
 (A.6) \quad & \mathbf{E}(|f(X(t), \zeta_1) - f(X(t), \zeta_2)| \wedge N) \mathbb{1}_A \\
 & \leq \int_{\mathbb{X}} \sup_{|y_1| \leq R, |y_1 - y_2| \leq \beta} (|f(x, y_1) - f(x, y_2)| \wedge N) \pi(dx) \\
 & + 4\|f\|_{\phi, d, p, 0} r^{1/p}(t) \left( \int_{\mathbb{X}} \psi d\pi \right)^{1/q} \mathbf{E} \psi(X(0)), \quad t \in \mathbb{R}^+,
 \end{aligned}$$

for an arbitrary  $N \geq 1$ .

*Proof.* Similarly to the preceding proof we consider an arbitrary pair  $(\xi, \eta) \in C(\mu_t, \pi)$  and construct a family  $(\hat{\xi}, \hat{\eta}, \hat{\zeta}_1, \hat{\zeta}_2, \hat{A})$  such that the pair  $(\hat{\xi}, \hat{\eta})$  is identically distributed with  $(\xi, \eta)$  and the family  $(\hat{\xi}, \hat{\zeta}_1, \hat{\zeta}_2, \mathbb{1}_{\hat{A}})$  is identically distributed with  $(X(t), \zeta_1, \zeta_2, \mathbb{1}_A)$ . Then

$$\begin{aligned}
 (A.7) \quad & \mathbf{E}|f(X(t), \zeta_1) - f(X(t), \zeta_2)| \mathbb{1}_A = \mathbf{E}|f(\hat{\xi}, \hat{\zeta}_1) - f(\hat{\xi}, \hat{\zeta}_2)| \mathbb{1}_{\hat{A}} \\
 & \leq \mathbf{E}|f(\hat{\eta}, \hat{\zeta}_1) - f(\hat{\eta}, \hat{\zeta}_2)| \mathbb{1}_{\hat{A}} + 2\hat{\mathbf{E}}|f(\hat{\xi}, \hat{\zeta}) - f(\hat{\eta}, \hat{\zeta})|.
 \end{aligned}$$

The distribution of  $(\hat{\zeta}_1, \hat{\zeta}_2, \mathbb{1}_{\hat{A}})$  coincides with the distribution of  $(\zeta_1, \zeta_2, \mathbb{1}_A)$ , whence we conclude that  $|\hat{\zeta}_1| \leq R$  and  $|\hat{\zeta}_1 - \hat{\zeta}_2| \leq \beta$  almost surely on  $\hat{A}$ . Hence

$$\begin{aligned}
 \mathbf{E}|f(\hat{\eta}, \hat{\zeta}_1) - f(\hat{\eta}, \hat{\zeta}_2)| \mathbb{1}_{\hat{A}} & \leq \mathbf{E} \sup_{|y_1| \leq R, |y_1 - y_2| \leq \beta} |f(\hat{\eta}, y_1) - f(\hat{\eta}, y_2)| \\
 & = \int_{\mathbb{X}} \sup_{|y_1| \leq R, |y_1 - y_2| \leq \beta} |f(x, y_1) - f(x, y_2)| \pi(dx),
 \end{aligned}$$

since  $\pi$  is the distribution of  $\hat{\eta}$ . A bound for the second term in (A.7) is obtained in the preceding proof. Therefore

$$\begin{aligned}
 (A.8) \quad & \mathbf{E}|f(X(t), \zeta_1) - f(X(t), \zeta_2)| \mathbb{1}_A \\
 & \leq \int_{\mathbb{X}} \sup_{|y_1| \leq R, |y_1 - y_2| \leq \beta} |f(x, y_1) - f(x, y_2)| \pi(dx) \\
 & + 4\|f\|_{\phi, p, d, 0} (\mathbf{E} d(\xi, \eta))^{1/p} \left( \int_{\mathbb{X}} \psi d\pi \right)^{1/q} (\mathbf{E} \psi(X(0)))^{1/q}.
 \end{aligned}$$

Since  $(\xi, \eta) \in C(\mu_t, \pi)$  is arbitrary, relations (A.2) and (A.8) complete the proof.  $\square$

## BIBLIOGRAPHY

1. N. Abourashchi and A. Yu. Veretennikov, *On stochastic averaging and mixing*, Theory Stoch. Process. **16(32)** (2010), no. 1, 111–129. MR2779833 (2012a:60188)
2. R. Z. Khas'minskiĭ, *A limit theorem for the solutions of differential equations with random right-hand sides*, Teor. Veroyatnost. i Primenen. **11** (1966), no. 3, 444–462; English transl. in Theor. Probability Appl. **11** (1966), no. 3, 390–406. MR0203789 (34:3637)
3. A. N. Borodin, *A limit theorem for the solutions of differential equations with a random right-hand side*, Teor. Veroyatnost. i Primenen. **22** (1977), no. 3, 498–512; English transl. in Theor. Probability Appl. **22** (1977), no. 3, 482–497. MR0517995 (58:24546)
4. V. V. Sarafyan and A. V. Skorokhod, *Dynamical systems with fast switchings*, Teor. Veroyatnost. i Primenen. **32** (1987), no. 4, 658–669; English transl. in Theory Probab. Appl. **32** (1987), no. 4, 595–607. MR927247 (89f:60064)
5. S. N. Ethier and T. G. Kurtz, *Markov Processes. Characterization and Convergence*, Wiley, New York, 1986. MR838085 (88a:60130)
6. V. S. Koroliuk and N. Limnios, *Stochastic Systems in Merging Phase Space*, World Scientific, New Jersey, 2005. MR2205562 (2007a:60004)
7. E. Pardoux and A. Yu. Veretennikov, *On the Poisson equation and diffusion approximation. I*, Ann. Probab. **29** (2001), 1061–1085. MR1872736 (2002j:60120)



8. E. Pardoux and A. Yu. Veretennikov, *On Poisson equation and diffusion approximation. II*, *Ann. Probab.* **31** (2003), 1166–1192. MR1988467 (2004d:60156)
9. A. Yu. Veretennikov and A. M. Kulik, *Extended Poisson equation for weakly ergodic Markov processes*, *Teor. Imovir. Matem. Statyst.* **85** (2011), 22–38; English transl. in *Theor. Probability and Math. Statist.* **85** (2012), 23–39. MR2933700 (2012m:60128)
10. A. Yu. Veretennikov and A. M. Kulik, *Diffusion approximation of systems with weakly ergodic Markov perturbations. II*, *Teor. Imovir. Mat. Stat. No. 88* (2013), 1–16. MR3112631
11. V. M. Zolotarev, *Modern theory of summation of random variables*, “Nauka”, Moscow, 1986; English transl., VSP, Utrecht, 1997. MR1640024 (99m:60002)
12. R. M. Dudley, *Real Analysis and Probability*, Cambridge Univ. Press, 2002. MR1932358 (2003h:60001)
13. G. C. Papanicolaou, D. W. Stroock, and R. S. Varadhan, *Martingale approach to some limit theorems* (M. Reed, ed.), Conference on Statistical Mechanics, Dynamical Systems and Turbulence, Duke Univ. Press, 1977. MR0461684 (57:1669)
14. I. I. Gikhman and A. V. Skorokhod, *Stochastic Differential Equations and their Applications*, “Naukova dumka”, Kiev, 1982. (Russian) MR678374 (84j:60003)

SCHOOL OF MATHEMATICS, UNIVERSITY OF LEEDS, UNITED KINGDOM; INSTITUTE FOR INFORMATION TRANSMISSION PROBLEMS, MOSCOW, RUSSIA

*E-mail address:* [A.Veretennikov@leeds.ac.uk](mailto:A.Veretennikov@leeds.ac.uk)

INSTITUTE FOR MATHEMATICS, NATIONAL ACADEMY OF SCIENCE OF UKRAINE, KYIV, UKRAINE

*E-mail address:* [kulik@imath.kiev.ua](mailto:kulik@imath.kiev.ua)

Received 15/MAY/2012

Translated by N. SEMENOV