



# On diffusion approximation with discontinuous coefficients

N.V. Krylov<sup>a,\*</sup>, R. Liptser<sup>b, 2</sup>

<sup>a</sup>*School of Mathematics, University of Minnesota, Minneapolis, MN, 55455, USA*

<sup>b</sup>*Department of Electrical Engineering-Systems, Tel Aviv University, 69978 Tel Aviv, Israel*

Received 22 May 2001; received in revised form 25 November 2001; accepted 26 May 2002

---

## Abstract

Convergence of stochastic processes with jumps to diffusion processes is investigated in the case when the limit process has discontinuous coefficients. An example is given in which the diffusion approximation of a queueing model yields a diffusion process with discontinuous diffusion and drift coefficients.

© 2002 Elsevier Science B.V. All rights reserved.

*MSC:* 60B10; 60K25

*Keywords:* Diffusion approximation; Stochastic differential equations; Weak convergence

---

## 1. Introduction

Suppose that we are given a sequence of semimartingales  $(x_t^n)_{t \geq 0}$ ,  $n = 1, 2, \dots$ , with paths in the Skorokhod space  $\mathcal{D} = \mathcal{D}([0, \infty), \mathbb{R}^d)$  of  $\mathbb{R}^d$ -valued right-continuous functions on  $[0, \infty)$  having left limits on  $(0, \infty)$ . If one can prove that the sequence of distributions  $\mathbb{Q}^n$  of  $x^n$  on  $\mathcal{D}$  weakly converges to the distribution  $\mathbb{Q}$  of a diffusion process  $(x_t)_{t \geq 0}$ , then one says that the sequence of  $(x_t^n)_{t \geq 0}$  admits a diffusion approximation. In this article by diffusion processes we mean solutions of Itô equations of the form

$$x_t = x_0 + \int_0^t b(s, x_s) ds + \int_0^t \sqrt{a(s, x_s)} dw_s$$

---

\* Corresponding author.

*E-mail addresses:* krylov@math.umn.edu (N.V. Krylov), liptser@eng.tau.ac.il (R. Liptser).

<sup>1</sup> The work was partially supported by NSF grant DMS-0140405.

<sup>2</sup> The work was partially supported by Orway Research Grant.

with  $w_t$  being a vector-valued Wiener process. Usually, to investigate the question if in a particular situation there is a diffusion approximation, one uses the general framework of convergence of semimartingales as developed for instance by [Liptser and Shiryaev \(1986, Section 3, Chapter 8\)](#) (also see the references in this book).

The problem of diffusion approximation attracted attention of many researchers who obtained many deep and important results. The reason for this is that diffusion approximation is a quite efficient tool in stochastic systems theory (see [Kushner, 1984, 1990](#)), in asymptotic analysis of queueing models under heavy traffic and bottleneck regimes (see [Kogan and Liptser, 1993](#)), in finding asymptotically optimal filters (see [Kushner and Runggaldier, 1987b; Liptser and Runggaldier, 1995](#)), in asymptotical optimization in stochastic control problems (see [Kushner and Runggaldier, 1987a; Liptser et al., 1999](#)), and in many other issues.

In all the above-mentioned references the coefficients  $a(t, x)$  and  $b(t, x)$  of the limit diffusion process are continuous in  $x$ . In part, this is dictated by the approach developed in by [Liptser and Shiryaev \(1986, Section 3, Chapter 8\)](#). On the other hand, there are quite a few situations in which the limit process should have discontinuous coefficients. One of such situations is presented by [Fleming and Simon \(1999\)](#) where a queueing model is considered. It was not possible to apply standard results and the authors only conjectured that the diffusion approximation should be a process with natural coefficients. Later this conjecture was rigorously proved by [Chao \(1999\)](#). [Chao \(1999\)](#) and [Fleming and Simon \(1999\)](#) considered the case in which only the drift term is discontinuous. Another example of the limit diffusion with discontinuities in both drift and diffusion coefficients is given by [Khasminskii and Krylov \(2001\)](#) on averaging principle for diffusion processes with null-recurrent fast component.

The idea to circumvent the discontinuity of  $a$  and  $b$  is to try to show that the time spent by  $(t, x_t)$  in the set  $G$  of their discontinuity in  $x$  is zero. This turns out to be enough if outside of  $G$  the “coefficients” of  $x_t^n$  converge “uniformly” to the coefficients of  $x_t$ . By the way, even if all these hold, still the functionals

$$\int_0^t a(t, y_t) dt, \quad \int_0^t b(t, y_t) dt, \quad y_t \in \mathcal{D}$$

need not be continuous on the support of  $\mathbb{Q}$ . This closes the route of “trivial” generalizing the result by [Liptser and Shiryaev \(1986, Section 3, Chapter 8\)](#).

To estimate the time spent by  $x_t$  we use an inequality similar to the following one:

$$E \int_0^T f(t, x_t) dt \leq N \left( \int_0^T \int_{\mathbb{R}^d} f^{d+1}(t, x) dx dt \right)^{1/(d+1)}, \tag{1.1}$$

which is obtained by [Krylov \(1974\)](#) for nonnegative Borel  $f$ . Then upon assuming that  $G \subset (0, \infty) \times \mathbb{R}^d$  has  $d + 1$ -dimensional Lebesgue measure zero and substituting  $I_G$  in place of  $f$  in (1.1) we get that indeed the time spent by  $(t, x_t)$  in  $G$  is zero. However, for (1.1) to hold we need the process  $x_t$  to be uniformly nondegenerate which may not be convenient in some applications. Therefore, in [Section 5](#) we prove a version of (1.1), which allows us to get the conclusion about the time spent in  $G$  assuming that the process is nondegenerate only on  $G$ . In essence, our approach to diffusion approximation with discontinuous coefficients is close to that of [Chao](#)

(1999). However, details are quite different and we get more general results under less restrictive assumptions. In particular, we do not impose the linear growth condition. Neither do we assume that the second moments of  $x_0^n$  are bounded.

The referees of the paper insisted on comparing our results also to Theorem 5.3 in Chapter 10 of Kushner and Dupuis (2001). The fact is that the weak limits of processes with jumps appear in many other settings, in particular, in Markov chain approximations in the theory of controlled diffusion processes, where, generally, the coefficients of  $x_t^n$  are not supposed to converge to anything in any sense and yet the processes converge weakly to a process of diffusion type. The first author of the present article is currently working on the general problem of describing all weak limit points of a given sequence of semimartingales. Theorem 5.3 in Chapter 10 of Kushner and Dupuis (2001) also bears on this matter in the particular case of Markov chain approximations in the theory of controlled diffusion processes. Clearly, there is no way to specify precisely the coefficients of all limit points in the general problem. Still one can obtain some nontrivial information and one may wonder if one can get anything from general results when we are additionally given that the coefficients do converge on the major part of the space. In Remarks 2.6 and 2.7 we show that this is not the case in what concerns Theorem 5.3 in Chapter 10 of Kushner and Dupuis (2001).

Above we alluded to the “coefficients” of  $x_t^n$ . By them we actually mean the local drift and the matrix of quadratic variation. We do not use any additional structure of  $x_t^n$ . In particular, the quadratic variation is just the sum of two terms: one coming from diffusion and another from jumps. Therefore, unlike Kurtz and Protter (1996) we do not use any stochastic equations for  $x_t^n$ . This allows us to neither introduce nor use any assumptions on the martingales driving these equations and their (usual) coefficients thus making the presentation simpler and more general. On the other hand, it is worth noting that the methods of Kurtz and Protter (1996) may be more useful in other problems. Our intention was not to cover all aspects of diffusion approximation but rather give a new method allowing us to treat discontinuous coefficients. In particular, we do not discuss uniqueness of solutions to the limit equation. This is a separate issue belonging to the theory of diffusion processes and we only mention the article by Khasminskii and Krylov (2001), where the reader can find a discussion of it.

The paper is organized as follows. In Section 2 we prove our main results, Theorems 2.1 and 2.5, about diffusion approximation. Their proofs rely on the estimate proved in Section 5 we have been talking about above. But even if the set  $G$  is empty, the results which we prove are the first ones of the kind.

In Theorems 2.1 and 2.5 there is no assumption about any control of  $\sqrt{a(t,x)}$  and  $b(t,x)$  as  $|x| \rightarrow \infty$ , but instead we assume that  $\mathbb{Q}^n$  converge weakly to  $\mathbb{Q}$ . Therefore, in Section 3 we give a sufficient condition for precompactness of a sequence of distributions on Skorokhod space. Interestingly enough, this condition is different from that of Jacod and Shiriyayev (1987) and Liptser and Shiriyayev (1986), and again does not involve usual growth conditions. Section 4 contains an example of application of our results to a queueing model close to that of Chao (1999) and Fleming and Simon (1999). We slightly modify the model by Chao (1999) and Fleming and Simon (1999) and get the diffusion approximation with discontinuous drift and

diffusion coefficients. To the best of our knowledge this is the first example when the diffusion approximation leads to discontinuous diffusion coefficients.

### 2. Main results

We use notions and notation given by Liptser and Shirayev (1986). For each  $n = 1, 2, \dots$ , let

$$(\Omega^n, \mathcal{F}^n, \mathcal{F}_t^n, t \geq 0, P^n)$$

be a stochastic basis satisfying the “usual” assumptions. Let  $\mathcal{D}$  be the Skorokhod space of right-continuous  $\mathbb{R}^d$ -valued functions  $x_t$  given on  $[0, \infty)$  and having left limits on  $(0, \infty)$ . As usual we endow  $\mathcal{D}$  with Skorokhod–Lindvall metric in which  $\mathcal{D}$  becomes a Polish space (see Liptser and Shirayev, 1986, Theorem 2, Section 1, Chapter 6).

Suppose that for each  $n$  on  $\Omega^n$  we are given an  $\mathcal{F}_t^n$ -semimartingale  $x_t^n, t \geq 0$ , with trajectories in  $\mathcal{D}$ . Let  $(B^n, C^n, \nu^n)$  be the triple of predictable characteristics of  $(x_t^n, \mathcal{F}_t^n)$  and  $\mu^n$  be its jump measure (see Liptser and Shirayev, 1986, Section 1, Chapter 4). Then

$$x_t^n = x_0^n + B_t^n + x_t^{nc} + \int_0^t \int_{|x| \leq 1} x(\mu^n - \nu^n)(ds dx) + \int_0^t \int_{|x| > 1} x \mu^n(ds dx),$$

where  $B_t^n$  is a predictable process of locally bounded variation with  $B_0^n = 0$ ,  $x_t^{nc}$  is a continuous local martingale with  $\langle x^{nc} \rangle_t = C_t^n$ ,  $\nu^n$  is the compensator of  $\mu^n$ . Define

$$m_t^n = x_t^{nc} + \int_0^t \int_{|x| \leq 1} x(\mu^n - \nu^n)(ds dx), \quad j_t^n = \int_0^t \int_{|x| > 1} x \mu^n(ds dx)$$

so that  $m_t^n$  is a locally square-integrable martingale and

$$x_t^n = x_0^n + B_t^n + m_t^n + j_t^n. \tag{2.1}$$

**Assumption 2.1.** (i) For each  $n$  on  $(0, \infty) \times \mathcal{D}$  we are given an  $\mathbb{R}^d$ -valued function  $b^n = b^n(t, y)$  and a  $d \times d$  matrix-valued function  $a^n = a^n(t, y)$  which is nonnegative and symmetric for any  $t$  and  $y \in \mathcal{D}$ . The functions  $b^n$  and  $a^n$  are Borel measurable.

(ii) For each  $r \in [0, \infty)$  there exists a locally integrable function  $L(r, t)$  given on  $[0, \infty)$  such that  $L(r, t)$  increases in  $r$  and

$$|b^n(t, y)| + \text{trace } a^n(t, y) \leq L(r, t) \tag{2.2}$$

whenever  $t > 0, y \in \mathcal{D}$ , and  $|y_t| \leq r$ . (iii) We have

$$B_t^n = \int_0^t b^n(s, x_s^n) ds, \quad \langle m^n \rangle_t = 2 \int_0^t a^n(s, x_s^n) ds.$$

**Remark 2.1.** We have

$$\langle m^n \rangle_t^{ij} = \langle x^{nc} \rangle_t^{ij} + \int_0^t \int_{|x| \leq 1} x^i x^j \nu^n(ds dx)$$

and it follows from Assumption 2.1 that both summands on the right are absolutely continuous in  $t$ . In particular, they are continuous, which along with the continuity

of  $B_t^n$  implies that  $x_t^n$  is quasi-left continuous (see [Liptser and Shiryaev, 1986](#), Theorem 1, Section 1, Chapter 4).

**Assumption 2.2.** (i) On  $(0, \infty) \times \mathbb{R}^d$  we are given an  $\mathbb{R}^d$ -valued function  $b = b(t, x)$  and a  $d \times d$  matrix-valued function  $a = a(t, x)$  which is nonnegative and symmetric for any  $t$  and  $x$ . The functions  $b$  and  $a$  are Borel measurable.

(ii) There exists a Borel set  $G \subset (0, \infty) \times \mathbb{R}^d$  (perhaps empty) such that, for almost every  $t \in (0, \infty)$ , for every  $x$  lying outside of the  $t$ -section  $G_t := \{x \in \mathbb{R}^d : (t, x) \in G\}$  of  $G$  and any sequence  $y^n \in \mathcal{D}$ , which converges to a continuous function  $y$ . satisfying  $y_t = x$ , it holds that

$$b^n(t, y^n) \rightarrow b(t, x), \quad a^n(t, y^n) \rightarrow a(t, x).$$

**Remark 2.2.** It is easy to see that Assumption 2.2 implies that for almost any  $t$ , the functions  $a(t, x)$  and  $b(t, x)$  are continuous on the set  $\mathbb{R}^d \setminus G_t$  in the relative topology of this set.

Also, Assumptions 2.1 and 2.2 obviously imply that

$$|b(t, x)| + \text{trace } a(t, x) \leq L(r, t)$$

for almost every  $t \in (0, \infty)$  and all  $x$  satisfying  $|x| \leq r$ ,  $x \notin G_t$ .

**Assumption 2.3.** If  $G \neq \emptyset$ , then for almost each  $t$

- (i) the set  $G_t$  has Lebesgue measure zero,
- (ii) for every  $x \in G_t$  and each sequence  $y^n \in \mathcal{D}$ , which converges to a continuous function  $y$ . satisfying  $y_t = x$ , we have

$$\liminf_{n \rightarrow \infty} \det a^n(t, y^n) \geq \delta(t, x) > 0, \tag{2.3}$$

where  $\delta$  is a Borel function.

**Remark 2.3.** Condition (2.3) is satisfied if, for instance, the processes  $x_t^n$  are uniformly nondegenerate in a neighborhood of  $G_t$ .

**Assumption 2.4.** For any  $T, \varepsilon \in (0, \infty)$ , and any  $\alpha \in (0, 1]$ , it holds that

$$\lim_{n \rightarrow \infty} P^n(v^n((0, T] \times B_\alpha^c)) \geq \varepsilon = 0,$$

where  $B_\alpha = \{x \in \mathbb{R}^d : |x| < \alpha\}$ ,  $B_\alpha^c = \{x \in \mathbb{R}^d : |x| \geq \alpha\}$ .

**Remark 2.4.** Notice that for each  $\alpha \in (0, 1]$  and  $r, T \in [0, \infty)$

$$\begin{aligned} \theta_{rT}^n &:= \int_0^T \int_{|x| \leq 1} |x|^3 I_{|x_s| \leq r} v^n(ds dx) \leq \int_0^T \int_{|x| < \alpha} + \int_0^T \int_{|x| \geq \alpha} \\ &\leq \alpha \int_0^T \int_{|x| \leq 1} |x|^2 I_{|x_s| \leq r} v^n(ds dx) + v^n((0, T] \times B_\alpha^c), \end{aligned}$$

where according to Assumption 2.1 the first term on the right is less than

$$2\alpha \int_0^T I_{|x_s| \leq r} \text{trace } a^n(s, x^n) \, ds \leq 2\alpha \int_0^T L(r, s) \, ds.$$

It follows easily that, owing to Assumptions 2.1 and 2.4, for each  $\varepsilon > 0$  and  $r, T \in [0, \infty)$ , we have

$$\lim_{n \rightarrow \infty} P^n(\theta_{rT}^n \geq \varepsilon) = 0$$

and since  $\theta_{rT}^n \leq 2 \int_0^T L(r, s) \, ds$ , we also have  $E^n \theta_{rT}^n \rightarrow 0$  as  $n \rightarrow \infty$ , where  $E^n$  is the expectation sign relative to  $P^n$ .

**Remark 2.5.** Define

$$\gamma^n = \inf\{t \geq 0: |j_t^n| > 1\}. \tag{2.4}$$

Then  $\gamma^n$  is an  $\mathcal{F}_t^n$ -stopping time, and obviously  $j_t^n = 0$  for  $0 \leq t < \gamma^n$ . Furthermore, by Lemma VI.4.22 of Jacod and Shiriyayev (1987), Assumption 2.4 implies that

$$P^n(\gamma^n \leq T) \rightarrow 0$$

for each  $T \in [0, \infty)$ .

**Theorem 2.1.** *In addition to Assumptions 2.1–2.4, suppose that the sequence of distributions  $(\mathbb{Q}^n)_{n \geq 1}$  of  $x^n$  converges weakly on the Polish space  $\mathcal{D}$  to a measure  $\mathbb{Q}$ . Then  $\mathbb{Q}$  is the distribution of a solution of the Itô equation*

$$x_t = x_0 + \int_0^t \sqrt{2a(s, x_s)} \, dw_s + \int_0^t b(s, x_s) \, ds \tag{2.5}$$

defined on a probability space with  $w_t$  being a  $d$ -dimensional Wiener process.

**Remark 2.6.** Notice that there are *no conditions* on the values of  $a(t, x)$  and  $b(t, x)$  on the set  $G$ . Hence Theorem 2.1 holds if we replace  $a, b$  with any other Borel functions, which coincide with the original ones on the complement  $\Gamma$  of  $G$ . Of course, this can only happen if

$$\int_0^t I_G(s, x_s) \, ds = 0 \quad (\text{a.s.}).$$

This equality is proved in Lemma 2.4. In particular,  $x_t$  satisfies

$$x_t = x_0 + \int_0^t I_\Gamma(s, x_s) \sqrt{2a(s, x_s)} \, dw_s + \int_0^t I_\Gamma(s, x_s) b(s, x_s) \, ds. \tag{2.6}$$

Thus, the limit process satisfies (2.6). A particular feature of this equation is that generally its solutions are not unique. Indeed, let  $x'_t$  be a one-dimensional Wiener process  $w_t$  and  $x''_t$  the process identically equal to zero. They both satisfy  $dx_t = \sqrt{2a(t, x_t)} \, dw_t$ , where  $a(t, x) = 1/2$  for  $(t, x) \notin G$ ,  $a(t, x) = 0$  for  $(t, x) \in G$ , and  $G = [0, \infty) \times \{0\}$ . Of course, there are many more different solutions which spend some time at zero, then

follow the trajectories of  $w_t$  for a while and then again stay at zero. Therefore, the statement that  $x_t$  has the form

$$x_t = x_0 + \int_0^t \sqrt{2a_s} dw_s + \int_0^t b_s ds,$$

where  $a_s = a(s, x_s)$  and  $b_s = b(s, x_s)$  whenever  $(s, x_s) \notin G$  and  $a$  and  $b$  are not specified otherwise (cf. the first part of Theorem 5.3 in Chapter 10 of Kushner and Dupuis, 2001), contains very little information on the process: in the above example both  $x'_t$  and  $x''_t$  have this form. In contrast with this always in the above example, the fact that without changing  $x_t$  one can change  $a, b$  on  $G$  in any way, and thus take  $a \equiv 1/2$ , leaves only one possibility:  $x_t = w_t$ .

**Remark 2.7.** From Remark 2.6 we also see that the assumption that (2.5) has a unique (weak or strong) solution makes no sense unless the values of  $a(t, x)$  and  $b(t, x)$  are specified everywhere. In Theorem 5.3 in Chapter 10 of Kushner and Dupuis (2001) an attempt is presented to specify  $a(t, x)$  and  $b(t, x)$  on  $G$  consisting of requiring that they belong to the set of all possible diffusion and drift coefficients of  $x_t$  when  $x_t \in G_t$ . Generally, the set  $x_t \in G_t$  has zero probability (say, for the Wiener process) and the requirement seems to have little sense. Nevertheless, it is natural to assume that, if  $x_t = w_t$  in the example from Remark 2.6, then the only possibility for  $a(t, 0)$  is  $1/2$ , the same value as for all other  $x$ .

In that case, the equation  $dx_t = \sqrt{2a(t, x_t)} dw_t$  ( $=dw_t$ ) with zero initial condition has a unique solution, the distribution of which (by Theorem 2.1) is the weak limit of the distributions of solutions to  $dx_t^n = \sqrt{2a^n(x_t)} dw_t$  with zero initial condition, where  $a^n(x) = 1/2$  for  $|x| \geq 1/n$  and  $a^n(x) = 1/3$  for  $|x| < 1/n$ .

However, this fact does not imply that the distributions of any other processes  $z^n$  converge to the Wiener measure, provided only that  $z^n_0 = 0$  and  $dz^n_t = \sqrt{2c^n(z^n_t)} dw_t$  with  $c^n(x) = a^n(x)$  for  $|x| \geq 1/n$ ,  $c^n \geq 0$ , and  $\sup_{n,x} c^n(x) < \infty$ . To show this, it suffices to define  $c^n(x) = n^2 x^2$  for  $|x| \leq 1/n$  and notice that  $z^n_t \equiv 0$  for all  $n$ .

This somewhat contradicts the second part of Theorem 5.3 in Chapter 10 of Kushner and Dupuis (2001).

The proof of Theorem 2.1 consists of several steps throughout which we assume that the conditions of this theorem are satisfied.

The idea is to rewrite (2.5) in terms of the martingale problem of Stroock–Varadhan. Then naturally we also want to write the information about  $x^n_t$  in a martingale form not involving stochastic bases and that is convenient to pass to the limit. This is done in Lemma 2.2. After that we pass to the limit and in Lemma 2.3 derive our theorem upon additionally assuming that the time spent by the limit process  $(t, x_t)$  in the set  $G$  of possible discontinuities of coefficients is zero. This additional assumption holds, for instance, if  $G = \emptyset$ . Lemma 2.4 concludes the proof of the theorem.

After that in Theorem 2.5 we extend Theorem 2.1 to cases in which uniform non-degeneracy on  $G_t$  of diffusion is not required. We show the usefulness of Theorem 2.5 in Remark 4.3.

As any probability measure on  $\mathcal{D}$ , the measure  $\mathbb{Q}$  is the distribution on  $\mathcal{D}$  of a process  $x$ . having trajectories in  $\mathcal{D}$  and defined on a probability space. By  $E$  we denote the

expectation sign associated with that probability space. We will see that the theorem holds for this  $x$ . up to a possible enlargement of the probability space on which  $x$  lives. In the following lemma, Assumptions 2.2 and 2.3 are not used.

By  $C_0^\infty(\mathbb{R}^{d+1})$  we denote the set of all infinitely differentiable real-valued functions  $u = u(t, x)$  on  $\mathbb{R}^{d+1}$  with compact support.

**Lemma 2.2.** *For any  $0 \leq t_1 \leq \dots \leq t_q \leq s \leq t < \infty$ , continuous bounded function  $f$  on  $\mathbb{R}^{qd}$ , and  $u \in C_0^\infty(\mathbb{R}^{d+1})$ , we have*

$$\begin{aligned} & Ef(x_{t_1}, \dots, x_{t_q})[u(t, x_t) - u(s, x_s)] \\ &= \lim_{n \rightarrow \infty} E^n f(x_{t_1}^n, \dots, x_{t_q}^n) \int_s^t [u_p(p, x_p^n) + a^{nij}(p, x^n)u_{x^i x^j}(p, x_p^n) \\ &\quad + b^{ni}(p, x^n)u_{x^i}(p, x_p^n)] dp. \end{aligned} \tag{2.7}$$

Furthermore, the integrand with respect to  $p$  is less than  $NL(r, p)$ , where the constants  $N$  and  $r$  depend only on  $u$  but not on  $\omega$  and  $n$ .

**Proof.** Denote

$$z_t^n = x_t^n - j_t^n,$$

and for any process  $z_t$  on  $\Omega^n$  denote (whenever it makes sense)

$$\begin{aligned} M_t^n(z.) &:= u(t, z_t) - u(0, z_0) - \int_0^t u_t(s, z_s) ds - \int_0^t u_{x^i}(s, z_s) dB_s^{ni} \\ &\quad - (1/2) \int_0^t u_{x^i x^j}(s, z_s) d\langle m^n \rangle_s^{ij}, \end{aligned} \tag{2.8}$$

$$\rho_s^n(z., x) = u(s, z_s + x) - u(s, z_s) - x^i u_{x^i}(s, z_s) - (1/2)x^i x^j u_{x^i x^j}(s, z_s),$$

$$R_t^n(z.) = \int_0^t \int_{|x| \leq 1} \rho_s^n(z., x) v^n(ds dx). \tag{2.9}$$

Notice that by Itô’s formula (see Liptser and Shirayev, 1986, Theorem 1, Section 3, Chapter 2) the process  $M_t^n(z^n) - R_t^n(z^n)$  is a local  $\mathcal{F}_t^n$ -martingale. To be more precise Theorem 1, Section 3, Chapter 2 of Liptser and Shirayev (1986) says that

$$\begin{aligned} M_t^n(z^n) - R_t^n(z^n) &= \sum_{0 < s \leq t} [u(s, z_s^n) - u(s, z_{s-}^n) - u_{x^i}(s, z_{s-}^n)\Delta z_s^{ni}] \\ &\quad - \int_0^t \int_{|x| \leq 1} [u(s, z_s^n + x) - u(s, z_s^n) - x^i u_{x^i}(s, z_s^n)] v^n(ds dx) \\ &\quad + \int_0^t u_{x^i}(s, z_{s-}^n) dm_s^{ni}. \end{aligned}$$

Here the last term is a local martingale as is any stochastic integral with respect to a local martingale and the sum of the remaining terms equals

$$\int_0^t \int_{|x| \leq 1} [u(s, z_{s-}^n + x) - u(s, z_{s-}^n) - x^i u_{x^i}(s, z_{s-}^n)] \bar{\mu}(ds dx)$$

which is the stochastic integral with respect to the martingale measure  $\bar{\mu} = \mu - \nu$  and thus also is a local martingale.

Take the  $\mathcal{F}_t^n$ -stopping time  $\gamma^n$  introduced in (2.4). Then

$$M_{t \wedge \gamma^n}^n(z^n) - R_{t \wedge \gamma^n}^n(z^n)$$

is again a local martingale. It turns out that, for each  $T \in [0, \infty)$ , the trajectories of  $M_{t \wedge \gamma^n}^n(z^n)$ ,  $t \in [0, T]$ , are bounded and even uniformly in  $n$ . Indeed, let  $r$  be such that  $u(t, x) = 0$  for  $|x| \geq r$ . Notice that  $z_t^n = x_t^n$  for  $0 \leq t < \gamma^n$ . Then we find

$$\int_0^{t \wedge \gamma^n} u_{x^i}(s, z_s^n) dB_s^{ni} = \int_0^{t \wedge \gamma^n} u_{x^i}(s, x_s^n) b^{ni}(s, x_s^n) ds,$$

where

$$|u_{x^i}(s, x_s^n) b^{ni}(s, x_s^n)| = 0$$

if  $|x_s| \geq r$  (since  $u(t, x) = 0$  for  $|x| \geq r$ ) and

$$|u_{x^i}(s, x_s^n) b^{ni}(s, x_s^n)| \leq L(r, s) \sup_{s,x} |u_x(s, x)|$$

if  $|x_s| \leq r$  (see Assumption 2.1). Therefore,

$$\left| \int_0^{t \wedge \gamma^n} u_{x^i}(s, z_s^n) dB_s^{ni} \right| \leq \sup_{s,x} |u_x(s, x)| \int_0^t L(r, s) ds.$$

Similarly one treats the integrals with respect to  $\langle m^n \rangle_s^{ij}$ . As far as  $R_t^n(z^n)$  is concerned we notice that, for  $|x| \leq 1$  and  $0 \leq t < \gamma^n$ , we have

$$|\rho_s^n(z^n, x)| \leq N|x|^3 I_{|z_s^n| \leq r+1} = N|x|^3 I_{|x_s^n| \leq r+1},$$

where the constant  $N$  can be expressed in terms of the third-order derivatives of  $u$  only. Therefore,

$$|R_{t \wedge \gamma^n}^n(z^n)| \leq N\theta_{r+1, T}^n,$$

where  $\theta_{r, T}^n$  is introduced in Remark 2.4. By this remark for any  $t$  we have  $E|R_{t \wedge \gamma^n}^n(z^n)| \rightarrow 0$ . It follows that  $E^n|R_{t \wedge \gamma^n}^n(z^n)| < \infty$ , so that the local martingale  $M_{t \wedge \gamma^n}^n(z^n) - R_{t \wedge \gamma^n}^n(z^n)$  is in fact a martingale.

Hence,

$$E^n f(x_{t_1}^n, \dots, x_{t_m}^n) [M_{t \wedge \gamma^n}^n(z^n) - R_{t \wedge \gamma^n}^n(z^n) - (M_{s \wedge \gamma^n}^n(z^n) - R_{s \wedge \gamma^n}^n(z^n))] = 0.$$

Since  $E^n|R_{t \wedge \gamma^n}^n(z^n)| \rightarrow 0$ , we also have

$$\lim_{n \rightarrow \infty} E^n f(x_{t_1}^n, \dots, x_{t_q}^n) [M_{t \wedge \gamma^n}^n(z^n) - M_{s \wedge \gamma^n}^n(z^n)] = 0.$$

Furthermore, due to Remark 2.5,  $P(\gamma^n \leq T) \rightarrow 0$  as  $n \rightarrow \infty$  for each  $T \in [0, \infty)$ . In the light of this fact and by virtue of the uniform boundedness of  $M_{s \wedge \gamma^n}^n(z^n)$ , we obtain

$$\lim_{n \rightarrow \infty} E^n |M_{t \wedge \gamma^n}^n(z^n) - M_{s \wedge \gamma^n}^n(z^n)| I_{\gamma^n \leq t} = 0, \tag{2.10}$$

so that

$$\lim_{n \rightarrow \infty} E^n f(x_{t_1}^n, \dots, x_{t_q}^n) [M_t^n(z^n) - M_s^n(z^n)] I_{t < \gamma^n} = 0.$$

In addition, obviously,  $M_t^n(z^n) = M_t^n(x^n)$  for  $t < \gamma^n$  and in the same way as above one can prove that the trajectories of  $M_t^n(x^n)$ ,  $t \in [0, T]$ , are uniformly bounded in  $n$  for each  $T$ . It follows that (2.10) holds with  $t, s, x^n$  in the place of  $t \wedge \gamma^n, s \wedge \gamma^n, z^n$ , respectively. Thus,

$$\lim_{n \rightarrow \infty} E^n f(x_{t_1}^n, \dots, x_{t_q}^n) [M_t^n(x^n) - M_s^n(x^n)] = 0$$

which is rewritten as (2.7). The asserted boundedness of the integrand in (2.7) follows easily from the above argument. The lemma is proved.  $\square$

After we have exploited the stochastic bases  $(\Omega^n, \mathcal{F}^n, \mathcal{F}_t^n, t \geq 0, P^n)$ , we will pass to processes defined on the same probability space. We are going to rely upon two facts. First we know from Liptser and Shirayev (1986, Theorem 1, Section 5, Chapter 6) that, owing to Assumption 2.4,  $\mathbb{Q}$  is concentrated on the space of *continuous*  $\mathbb{R}^d$ -valued functions defined on  $[0, \infty)$ . Second, remember that if  $y^n \rightarrow y$  in  $\mathcal{D}$  and  $y$  is continuous, then  $|y^n - y|_t^* \rightarrow 0$  for any  $t < \infty$ , where

$$y_t^* := \sup_{r \leq t} |y_r|.$$

Owing to these facts and Skorokhod’s embedding theorem (see Skorokhod, 1961, Section 6, Chapter 1), we may assume that all the processes  $x^n$ ,  $n = 1, 2, \dots$ , are given on the same probability space and there is a continuous process  $x_t$  such that (a.s.)

$$\lim_{n \rightarrow \infty} \sup_{t \leq T} |x_t^n - x_t| = 0 \quad \forall T \in [0, \infty). \tag{2.11}$$

**Lemma 2.3.** *Assume that for any  $T$*

$$E \int_0^T I_G(t, x_t) dt = 0, \tag{2.12}$$

*which is certainly true if  $G = \emptyset$ . Then the assertion of Theorem 2.1 holds.*

**Proof.** As explained before the lemma we can write  $E$  in place of  $E^n$  in (2.7). Then we insert  $I_{x_p \notin G_p}$ , which is harmless due to (2.12), in the integral in (2.7) (notice  $x_p$  and not  $x_p^n$ ). Furthermore, we remember the last assertion of Lemma 2.2 and use Assumption 2.2, (2.11), and the dominated convergence theorem to conclude that the limit in (2.7) equals

$$\begin{aligned} & E f(x_{t_1}, \dots, x_{t_q}) \int_s^t I_{x_p \notin G_p} [a^{ij}(p, x_p) u_{x^i x^j}(p, x_p) \\ & + b^i(p, x_p) u_{x^i}(p, x_p) + u_p(p, x_p)] dp. \end{aligned} \tag{2.13}$$

By using (2.12) again, we obtain that

$$\begin{aligned} & Ef(x_{t_1}, \dots, x_{t_q})[u(t, x_t) - u(s, x_s)] \\ &= Ef(x_{t_1}, \dots, x_{t_q}) \int_s^t [u_p(p, x_p) + a^{ij}(p, x_p)u_{x^i x^j}(p, x_p) \\ &\quad + b^i(p, x_p)u_{x^i}(p, x_p)] dp \end{aligned}$$

for any bounded continuous  $f$  and  $t_i \leq s \leq t$ . The latter just amounts to saying that the process

$$u(t, x_t) - \int_0^t [u_s(s, x_s) + a^{ij}(s, x_s)u_{x^i x^j}(s, x_s) + b^i(s, x_s)u_{x^i}(s, x_s)] ds$$

is an  $\mathcal{F}_t^x$ -martingale, where  $\mathcal{F}_t^x$  is the  $\sigma$ -field generated by  $x_s$ ,  $s \leq t$ . It only remains to remember the Lévy–Doob–Stroock–Varadhan characterization theorem (see, for instance, Stroock and Varadhan (1979, Section 4.5) or Ikeda and Watanabe (1981, Sections 2.6 and 2.7). The lemma is proved.  $\square$

**Remark 2.8.** In the general case the above proof and Fatou’s theorem show that, if  $f$  is nonnegative, then

$$\begin{aligned} & Ef(x_{t_1}, \dots, x_{t_q})[u(t, x_t) - u(s, x_s)] \\ &\leq Ef(x_{t_1}, \dots, x_{t_q}) \int_s^t I_{x_p \notin G_p} [u_p(p, x_p) + a^{ij}(p, x_p)u_{x^i x^j}(p, x_p) \\ &\quad + b^i(p, x_p)u_{x^i}(p, x_p)] dp + I, \end{aligned} \tag{2.14}$$

where

$$\begin{aligned} I &= Ef(x_{t_1}, \dots, x_{t_q}) \int_s^t I_{x_p \in G_p} \overline{\lim}_{n \rightarrow \infty} [a^{nij}(p, x^n)u_{x^i x^j}(p, x_p^n) \\ &\quad + b^{ni}(p, x^n)u_{x^i}(p, x_p^n) + u_p(p, x_p^n)] dp. \end{aligned} \tag{2.15}$$

In the following lemma we complete proving Theorem 2.1. At this moment we take Theorem 5.1 for granted.

**Lemma 2.4.** *Eq. (2.12) holds and hence, by Lemma 2.3, Theorem 2.1 holds true as well.*

**Proof.** First, we estimate the  $\overline{\lim}$  in (2.15). Fix  $\omega$  and almost any  $p$  for which (2.3) holds with  $p$  in the place of  $t$  and  $x_p(\omega) \in G_p$ . Then we can replace  $\overline{\lim}_{n \rightarrow \infty}$  with  $\lim_{n' \rightarrow \infty}$ , where  $n'$  is an appropriate sequence tending to infinity. By extracting further subsequences when necessary we may assume that  $a^{n'}(p, x^{n'})$  and  $b^{n'}(p, x^{n'})$  converge to some  $\bar{a}$  and  $\bar{b}$ . Since  $x_p \in G_p$  and  $|x^n - x_p|_p^* \rightarrow 0$ , (2.3) implies that  $\det \bar{a} \geq \delta(p, x_p)$ . In addition,

$$|\bar{b}| + \text{trace } \bar{a} \leq L(|x_p| + 1, p)$$

due to Assumption 2.1. Combined with  $\det \bar{a} \geq \delta(p, x_p)$  this yields

$$\bar{a}^{ij} \lambda^i \lambda^j \geq \delta(p, x_p) L^{-(d-1)}(|x_p| + 1, p) |\lambda|^2 =: \bar{\delta}(p, x_p) |\lambda|^2 \geq \tilde{\delta}(p, x_p) |\lambda|^2$$

for all  $\lambda \in \mathbb{R}^d$ , where  $\tilde{\delta} = I_G \bar{\delta}$ . Now by replacing  $\delta$  with  $\tilde{\delta}$  and both  $K(r, t)$  and  $L(r, t)$  with  $L(r + 1, t)$  in Section 5, we conclude that

$$\begin{aligned} & \overline{\lim}_{n \rightarrow \infty} [a^{mij}(p, x^n) u_{x^i x^j}(p, x_p^n) + b^m(p, x^n) u_{x^i}(p, x_p^n) + u_p(p, x_p^n)] \\ & \leq u_p(p, x_p) + F(p, x_p, u_{xx}(p, x_p)) + L(|x_p| + 1, p) |u_x(p, x_p)|. \end{aligned}$$

Furthermore, Remark 2.2 shows that the same estimate holds for the expression in brackets in (2.13), so that according to (2.14)

$$\begin{aligned} & Ef(x_{t_1}, \dots, x_{t_q}) [u(t, x_t) - u(s, x_s)] \\ & \leq Ef(x_{t_1}, \dots, x_{t_q}) \int_s^t [u_p(p, x_p) \\ & \quad + F(p, x_p, u_{xx}(p, x_p)) + L(|x_p| + 1, p) |u_x(p, x_p)|] dp \end{aligned}$$

if  $f \geq 0$ . Hence the process

$$u(t, x_t) - \int_0^t [u_s(s, x_s) + F(s, x_s, u_{xx}(s, x_s)) + L(|x_s| + 1, s) |u_x(s, x_s)|] ds$$

is a supermartingale and by Theorem 5.1 estimate (5.2) holds. If we take there  $f = I_G$  and remember that the Lebesgue measure of  $G$  is zero and  $\tilde{\delta}(t, x) > 0$  on  $G_t$  for almost all  $t$ , then we come to (2.12) with  $T \wedge \tau_r$  in place of  $T$ . Upon letting  $r \rightarrow \infty$  we finally obtain (2.12) as is. The lemma is proved.  $\square$

The following theorem is used in Remark 4.3. Its proof is obtained by changing variables. We introduce an assumption different from Assumption 2.3.

**Assumption 2.5.** If  $G \neq \emptyset$ , then  $G = \bigcup_{m=1}^\infty G^m$ , where  $G^m$  are Borel sets. For each  $m$ , we are given an integer  $d_m \geq 1$ , a nonnegative Borel function  $\delta_m$  defined on  $(0, \infty) \times \mathbb{R}^{d_m}$ , and a continuous  $\mathbb{R}^{d_m}$ -valued function  $v^m(t, x) = (v^{m1}(t, x), \dots, v^{md_m}(t, x))$  defined on  $[0, \infty) \times \mathbb{R}^d$  and having there continuous in  $(t, x)$  derivatives  $v_t^{mi}, v_x^{mi}, v_{xx}^{mi}$ . For each  $m$  and almost every  $t \in (0, \infty)$ ,

- (i) the set  $v^m(t, G_t^m)$  has  $d_m$ -dimensional Lebesgue measure zero,
- (ii) for every  $x \in v^m(t, G_t^m)$  and each sequence  $y^n \in \mathcal{D}$ , which converges to a continuous function  $y$ . satisfying  $v^m(t, y_t) = x$ , we have

$$\overline{\lim}_{n \rightarrow \infty} \det V^{mn}(t, y^n) \geq \delta_m(t, x) > 0, \tag{2.16}$$

where the matrix  $V^{mn}(t, y.)$  is defined according to

$$V_{ij}^{mn}(t, y.) = v_{x^k}^{mi}(t, y_t) v_{x^r}^{mj}(t, y_t) a^{nkr}(t, y.), \quad i, j = 1, \dots, d_m.$$

**Remark 2.9.** Assumption 2.3 is stronger than Assumption 2.5. Indeed, if the former is satisfied, one can take  $G^m = G$ ,  $\delta_m(t, x) = \delta(t, x)$ ,  $d_m = d$ , and  $v^{mi} = x^i$ ,  $i = 1, \dots, d$ , in which case  $\det V^{mn} = \det a^n$ .

**Remark 2.10.** Another case is when again everything is independent of  $m$ , but  $d_m = 1$  and  $v(t, x) = x^1$ . Then condition (2.16) becomes

$$\lim_{n \rightarrow \infty} a^{n1}(t, y^n) \geq \delta(t, x) > 0,$$

which is much weaker than (2.3). However, typically in this case in order to satisfy requirement (i) of Assumption 2.5 we need to assume that  $G_t$  lies in a hyperplane orthogonal to the first coordinate axis.

**Remark 2.11.** Assume that  $G = \bigcup_{m=1}^{\infty} G^m$ , where  $G_t^m$  are independent of  $t$  and are hyperplanes  $G_t^m = \{x: (x, \alpha_m) = \beta_m\}$  with certain  $\alpha_m \in \mathbb{R}^d$  and  $\beta_m \in \mathbb{R}$  satisfying  $|\alpha_m| = 1$ . Assume that we have a Borel nonnegative function  $\delta(t, x)$ . Finally, assume that for every  $m \geq 1$ ,  $t > 0$ ,  $x \in \mathbb{R}^d$  such that

$$(x, \alpha_m) = \beta_m,$$

and each sequence  $y^n \in \mathcal{D}$ , which converges to a continuous function  $y$  satisfying  $y_t = x$ , we have

$$\lim_{n \rightarrow \infty} a^{nij}(t, y^n) \alpha_m^i \alpha_m^j \geq \delta(t, \beta_m) > 0.$$

Then it turns out that Assumption 2.5 is satisfied. To show this, it suffices to take  $d_m = 1$  and  $v^m(t, x) = (x, \alpha_m)$  and notice that the image of  $G_t^m$  under the mapping  $v^m(t, \cdot): G_t^m \rightarrow \mathbb{R}$  is just one point  $\beta_m$ . We will use this fact in Section 4.

**Remark 2.12.** Generally, condition (2.16) is aimed at situations in which  $x_t^n$  in the limit may degenerate in some directions but not along all those which are transversal to  $G_t$ .

**Theorem 2.5.** Suppose that Assumptions 2.1, 2.2, 2.4, and 2.5 are satisfied and the sequence of distributions  $(\mathbb{Q}^n)_{n \geq 1}$  of  $x^n$  converges weakly on  $\mathcal{D}$  to a measure  $\mathbb{Q}$ . Then the assertion of Theorem 2.1 holds true again.

**Proof.** We mimic the argument from the proof of Lemma 2.4 to show that (2.12) holds if Assumption 2.5 rather than Assumption 2.3 is satisfied. The main idea is to change variables according to the mappings  $v^m$ .

It suffices to prove that, for each  $m$ , Eq. (2.12) holds with  $G^m$  in place of  $G$ . Furthermore, without losing generality we may assume that each set  $G^m$  is bounded; otherwise we could split each of them into the union of bounded sets and consider them as new  $G^m$ 's. We fix  $m$ ,  $T$ , and  $R$  and assume that  $G^m \subset [0, T] \times B_R$ . Then the behavior of  $v^m(t, x)$  for large  $|x|$  becomes irrelevant and, changing  $v^m$  outside of  $[0, T] \times B_R$  if necessary, we assume that

$$v^m(t, x) = e_1 |x| \tag{2.17}$$

for  $(t, x) \notin [0, 2T] \times B_{2R}$ , where  $e_1$  is the first basis vector in  $\mathbb{R}^{d_m}$ . It follows that there is a constant  $N_0 < \infty$  such that

$$|v_x^m(t, x)| + |v_{x^i x^j}^m(t, x)| + |v_t^m(t, x)| \leq N_0 \quad \forall t, x. \tag{2.18}$$

It also follows that, for any  $r \geq 0$ ,

$$|v^m(t, x)| \leq r \Rightarrow |x| \leq 2R + r. \tag{2.19}$$

After that we go back to Lemma 2.2 and take there

$$u(t, x) = w(t, v^m(t, x)),$$

with  $w$  being a function of class  $C_0^\infty(\mathbb{R}^{d_m+1})$ . By the way, our stipulation (2.17) about the behavior of  $v^m$  for large  $|x|$  yields that  $u \in C_0^\infty(\mathbb{R}^{d_m+1})$ . We also take the function  $f$  in the form

$$f(y_1, \dots, y_q) = g(v^m(t_1, y_1), \dots, v^m(t_q, y_q)),$$

where  $y_i \in \mathbb{R}^d$  and  $g$  is a continuous bounded *nonnegative* function on  $\mathbb{R}^{qd_m}$ . Finally, we define

$$\tilde{x}_t^n = v^m(t, x_t^n), \quad \tilde{x}_t = v^m(t, x_t).$$

Notice that

$$\begin{aligned} & a^{nij}(p, x^n)u_{x^i x^j}(p, x_p^n) + b^{ni}(p, x^n)u_{x^i}(p, x_p^n) + u_p(p, x_p^n) \\ &= \tilde{a}^{nkr}(p, x^n)w_{x^k x^r}(p, \tilde{x}_p^n) + \tilde{b}^{nk}(p, x^n)w_{x^k}(p, \tilde{x}_p^n) + w_p(p, \tilde{x}_p^n), \end{aligned}$$

where, for  $y \in \mathcal{D}$ ,

$$\begin{aligned} \tilde{a}^{nkr}(p, y) &= a^{nij}(p, y)v_{x^i}^{mk}(p, y_p)v_{x^j}^{mr}(p, y_p), \\ \tilde{b}^{nk}(p, y) &= a^{nij}(p, y)v_{x^i x^j}^{mk}(p, y_p) + b^{ni}(p, y)v_{x^i}^{mk}(p, y_p) + v_p^{mk}(p, y_p). \end{aligned}$$

Then on the basis of Fatou’s theorem and Lemma 2.2 we get

$$\begin{aligned} & Ef(x_{t_1}, \dots, x_{t_q})[u(t, x_t) - u(s, x_s)] \\ & \leq Ef(x_{t_1}, \dots, x_{t_q}) \int_s^t \overline{\lim}_{n \rightarrow \infty} [u_p(p, x_p) + a^{nij}(p, x^n)u_{x^i x^j}(p, x_p^n) \\ & \quad + b^{ni}(p, x^n)u_{x^i}(p, x_p^n)] dp \\ & = Eg(\tilde{x}_{t_1}, \dots, \tilde{x}_{t_q}) \int_s^t \overline{\lim}_{n \rightarrow \infty} [w_p(p, \tilde{x}_p) + \tilde{a}^{nkr}(p, x^n)w_{x^k x^r}(p, \tilde{x}_p^n) \\ & \quad + \tilde{b}^{nk}(p, x^n)w_{x^k}(p, \tilde{x}_p^n)] dp. \end{aligned}$$

Also notice that owing to (2.18),  $\tilde{a}$  and  $\tilde{b}$  satisfy (2.2) with  $L(r, t)$  replaced with  $N_0L(r, t)$ . In the light of (2.19) this implies

$$|\tilde{b}^n(t, x^n)| + \text{trace } \tilde{a}^n(t, x^n) \leq N_0L(2R + |\tilde{x}_t^n|, t).$$

In addition, according to (2.16), for almost any  $t$ , for every  $\tilde{x} \in v^m(t, G_t^m)$  and each sequence  $y^n \in \mathcal{D}$ , which converges to a continuous function  $y$  satisfying  $v^m(t, y_t) = \tilde{x}$ , we have

$$\begin{aligned} \liminf_{n \rightarrow \infty} \det \tilde{a}^n(t, y^n) &\geq \delta_m(t, \tilde{x}) > 0, \\ \liminf_{n \rightarrow \infty} \tilde{a}^{nkr}(t, y^n) \lambda^k \lambda^r &\geq \tilde{\delta}_m(t, \tilde{x}) |\lambda|^2 \end{aligned}$$

for all  $\lambda \in R^{d_m}$ , where

$$\tilde{\delta}_m(t, \tilde{x}) = \delta_m(t, \tilde{x}) L^{-(d_m-1)}(2R + |\tilde{x}| + 1, t) I_{v^m(G^m)}(t, \tilde{x}).$$

Then as in the proof of Lemma 2.4 we find that

$$\begin{aligned} &Eg(\tilde{x}_{t_1}, \dots, \tilde{x}_{t_q}) [w(t, \tilde{x}_t) - w(s, \tilde{x}_s)] \\ &\leq Eg(\tilde{x}_{t_1}, \dots, \tilde{x}_{t_q}) \int_s^t w_p(p, \tilde{x}_p) + F(p, \tilde{x}_p, w_{xx}(p, \tilde{x}_p)) \\ &\quad + L(2R + |\tilde{x}_p| + 1, p) |w_x(p, \tilde{x}_p)|] dp, \end{aligned}$$

where the operator  $F$  is constructed on the basis of  $\tilde{\delta}_m$  and  $N_0 L(2R + r, t)$  in place of  $\delta$  and both  $L, K$  from Section 5, respectively, on the space of functions on  $\mathbb{R}^{d_m}$  in place of  $\mathbb{R}^d$ . Again as in the proof of Lemma 2.4 we conclude that, for any  $S$  we have

$$E \int_0^S I_{v^m(G^m)}(t, \tilde{x}_t) dt = 0.$$

Since, obviously,  $I_{G^m}(t, x) \leq I_{v^m(G^m)}(t, v^m(t, x))$  we get that (2.12) holds with  $G^m$  in place of  $G$ . As we have pointed out in the beginning of the proof, this is exactly what we need. The theorem is proved.  $\square$

### 3. A sufficient condition for precompactness

One of the conditions of Theorem 2.1 is that the sequence of distributions  $(\mathbb{Q}^n)_{n \geq 1}$  of  $x^n$  on  $\mathcal{D}$  converge. One can always extract a convergent subsequence from a sequence which is precompact and here we want to give a simple sufficient condition for precompactness to hold. The assumptions of this section are somewhat different from those of Section 2 and this was the reason to treat the issue in a separate section. We take the objects introduced in Section 2 before Assumption 2.1 and instead of that assumption we require the following.

**Assumption 3.1.** Assumption 2.1 is satisfied with condition (ii) replaced by the following weaker condition: For each  $r \in [0, \infty)$  there exists a locally integrable function  $L(r, t)$  given on  $[0, \infty)$  such that  $L(r, t)$  increases in  $r$  and

$$|b^n(t, y.)| + \text{trace } a^n(t, y.) \leq L(r, t)$$

whenever  $t > 0$ ,  $y. \in \mathcal{D}$ , and  $\sup_{s \leq t} |y_s| \leq r$ .

**Lemma 3.1.** *Under Assumptions 2.4 and 3.1 suppose that we are given  $\mathcal{F}_t^n$  stopping times  $\tau_r^n, n = 1, 2, \dots, r > 0$ , and a finite function  $\alpha(r)$  defined on  $(0, \infty)$  such that we have (i) for all  $n$  and  $r$ ,*

$$|x_t^n| \leq \alpha(r) \quad \text{if } 0 \leq t < \tau_r^n \tag{3.1}$$

and (ii)

$$\lim_{r \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} P^n(\tau_r^n \leq T) = 0 \quad \forall T \in [0, \infty). \tag{3.2}$$

Then the sequence  $(\mathbb{Q}^n)_{n \geq 1}$  is precompact.

**Proof.** Define

$$G_t^n = \int_0^t [|b^n(s, x^n)| + \text{trace } a^n(s, x^n)] ds,$$

$$F_t^n = G_t^n + \int_0^t \int_{|x| > 1} v^n(ds dx).$$

Owing to Assumption 2.4, by Theorem VI.4.18 and Remark VI.4.20 of Jacod and Shiriyayev (1987) to prove the theorem it suffices to check that the sequence of distributions on  $\mathcal{D}$  of  $F^n$  is  $C$ -tight, that is precompact and each limit point of this sequence is the distribution of a continuous process. In turn, due to Theorem VI.4.5 and Remark VI.4.6 (3) of Jacod and Shiriyayev (1987), to prove the  $C$ -tightness it suffices to show that, for any  $T \in [0, \infty)$  and  $\varepsilon > 0$ ,

$$\lim_{N \rightarrow \infty} \overline{\lim}_n P^n \left( \sup_{t \leq T} |F_t^n| \geq N \right) = 0,$$

$$\lim_{\delta \downarrow 0} \overline{\lim}_n P^n \left( \sup_{t+s \leq T, 0 \leq s \leq \delta} |F_{t+s}^n - F_t^n| \geq \varepsilon \right) = 0. \tag{3.3}$$

In view of Assumption 2.4 we need only prove (3.3) for  $G^n$  in place of  $F^n$ . We do this replacement and after that notice that, for any  $r$ , the left-hand side of the first equation in (3.3) is less than

$$\lim_{N \rightarrow \infty} \overline{\lim}_n P^n \left( \sup_{t \leq T \wedge \tau_r^n} |G_t^n| \geq N \right) + \overline{\lim}_{n \rightarrow \infty} P^n(\tau_r^n \leq T).$$

Here the first term is zero for each  $r$  since  $G_t^n$  is continuous in  $t$  and

$$|G_t^n| \leq \int_0^t L(\alpha(r), u) du$$

for  $t < \tau_r^n$  when by our assumptions  $|x_t^n| \leq r$ . In addition, the second term can be made as small as we wish by choosing a sufficiently large  $r$ . This proves the first equation in (3.3).

Similarly, the left-hand side of the second equation in (3.3) with  $G^n$  in place of  $F^n$  is less than

$$\lim_{\delta \downarrow 0} \overline{\lim}_n P^n \left( \sup_{t+s \leq T \wedge \tau_r^n, 0 \leq s \leq \delta} |G_{t+s}^n - G_t^n| \geq \varepsilon \right) + \overline{\lim}_{n \rightarrow \infty} P^n(\tau_r^n \leq T),$$

where again the first term vanishes since

$$|G_{t+s}^n - G_t^n| \leq \int_t^{t+s} L(\alpha(r), u) du.$$

The lemma is proved.  $\square$

**Remark 3.1.** It may be worth noticing that the combination of assumptions (i) and (ii) of Lemma 3.1 is equivalent to the following: for any  $T \in (0, \infty)$ , the sequence of distributions of  $\sup_{t \leq T} |x_t^n|$  is tight or put otherwise

$$\lim_{r \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} P^n \left( \sup_{t \leq T} |x_t^n| \geq r \right) = 0.$$

Lemma 3.1 reduces the investigation of precompactness to estimating  $|x_t^n|^*$ . Here the following coercivity assumption turns out to be useful.

**Assumption 3.2.** For any  $n$ , there exists a nonnegative  $\mathcal{F}_t^n$ -predictable function  $L_n(t)$  such that

$$b^{ni}(t, x^n)x_t^{ni} + \text{trace } a^n(t, x^n) \leq L_n(t)(1 + |x_t^n|^2) \tag{3.4}$$

for almost all  $(\omega, t)$ . Furthermore, for any  $T \in [0, \infty)$ ,

$$\lim_{c \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} P^n \left( \int_0^T L_n(t) dt > c \right) = 0.$$

**Remark 3.2.** Quite often one imposes a linear growth assumption on the coefficients  $a^n$  and  $b^n$ , which of course implies (3.4). However, say in one dimension, if  $a^n \equiv 0$  and  $b^{ni}(t, y) = b^n(t, y_t)$  and  $b^n(t, y_t) \geq 0$  for  $y_t < 0$  and  $b^n(t, y_t) \leq 0$  for  $y_t > 0$ , then (3.4) is satisfied with  $L \equiv 0$ . Therefore generally (3.4) does not provide any control on the behavior of  $|b^n(t, y_t)|$  for large  $|y_t|$ .

For that reason, Theorem 3.2 does not follow from the results of Jacod and Shiriyayev (1987) and Liptser and Shiriyayev (1986).

**Theorem 3.2.** *Let*

$$\lim_{N \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} P^n(|x_0^n| \geq N) = 0 \tag{3.5}$$

and let Assumptions 2.4, 3.1, and 3.2 be satisfied. Then the sequence  $(\mathbb{Q}^n)_{n \geq 1}$  is precompact. Furthermore, let  $k$  be an integer and  $f^n(t, x)$  be Borel  $\mathbb{R}^k$ -valued functions defined on  $(0, \infty) \times \mathbb{R}^d$  such that  $|f^n(t, x)| \leq L(|x|, t)$  for all  $t, x, n$ . Define

$$y_t^n = \int_0^t f^n(s, x_s^n) ds.$$

Then the sequence of distributions of  $(x^n, y^n)$  on  $\mathcal{D}([0, \infty), \mathbb{R}^{d+k})$  is precompact as well.

**Proof.** We are going to use a method introduced by Krylov and Rozovsky (1979, Section 4, Chapter II). Define

$$z_t^n = x_t^n - j_t^n, \quad \phi_n(t) = \exp\left(-2 \int_0^t L_n(s) ds\right), \quad u_n(t, x) = (1 + |x|^2)\phi_n(t).$$

Also as in the proof of Lemma 2.2, use notation (2.8) and (2.9) and notice that due to special choice of  $u$ , we have  $R_t^n(z.) \equiv 0$ .

Then by using Itô's formula, we get that the process

$$M_t^n := (1 + |z_t^n|^2)\phi_n(t) - (1 + |x_0^n|^2) - \int_0^t [2z_s^{ni} b^{ni}(s, x_s^n) + 2 \text{trace } a^n(s, x_s^n) - 2L_n(s)(1 + |z_s^n|^2)]\phi_n(s) ds \quad (3.6)$$

is a local martingale.

Now take  $\gamma^n$  again from (2.4) and remember that  $z_s^n = x_s^n$  for  $s < \gamma^n$ , so that the expression in brackets in (3.6) is negative due to Assumption 3.2. Then we see that

$$H_t^n := (1 + |z_{t \wedge \gamma^n}^n|^2)\phi_n(t \wedge \gamma^n) - (1 + |x_0^n|^2)$$

is a local supermartingale. For any constant  $N > 0$ , the process  $H_t^n I_{|x_0^n| \leq N}$  also is a local supermartingale and, since it is bounded from below by the constant  $-(1 + N^2)$ , it is a supermartingale. Therefore, upon defining

$$\kappa_r^n = \inf \left\{ t \geq 0: \sup_{s \leq t} |x_s^n| > r \right\}, \quad \tau_r^n = \gamma^n \wedge \kappa_r^n,$$

we get that, for any  $T \in [0, \infty)$ ,

$$E^n(1 + |z_{T \wedge \tau_r^n}^n|^2)\phi_n(T \wedge \tau_r^n) I_{|x_0^n| \leq N} \leq 1 + N^2,$$

$$E^n(1 + |z_{\tau_r^n}^n|^2)\phi_n(\tau_r^n) I_{|x_0^n| \leq N, \tau_r^n \leq T < \gamma^n} \leq 1 + N^2.$$

Then we notice that on the interval  $[0, \gamma^n)$  the process  $j_t^n$  is identically zero. Hence, for  $\tau_r^n \leq T < \gamma^n$  we have

$$|z_{\tau_r^n}^n| = |x_{\tau_r^n}^n| = |x_{\kappa_r^n}^n| \geq r$$

and we obtain

$$e^{-c}(1 + r^2)P^n \left( \int_0^T L_n(t) dt \leq c, |x_0^n| \leq N, \tau_r^n \leq T < \gamma^n \right) \leq 1 + N^2,$$

$$\lim_{r \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} P^n(|x_0^n| \leq N, \tau_r^n \leq T < \gamma^n) = 0.$$

This holds for any  $N$  and along with assumption (3.5) and Remark 2.5 leads first to

$$\lim_{r \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} P^n(\tau_r^n \leq T < \gamma^n) = 0$$

and then to (3.2).

Finally, observe that (3.1) is obviously satisfied even if  $0 \leq t < \kappa_r^n$  rather than  $0 \leq t < \tau_r^n$ . Hence, by referring to Lemma 3.1 we finish proving the assertion of our theorem regarding the distributions of  $x^n$ .

Lemma 3.1 yields the result for  $(x^n, y^n)$  as well since, obviously, for  $0 \leq t < r \wedge \tau_r^n$ , we have

$$|y_t^n| \leq \int_0^r L(r, s) ds.$$

The theorem is proved.  $\square$

#### 4. An example of queueing model

We consider a particular queueing system with  $d$  service stations and  $d + 1$  incoming streams of customers. We refer the reader to [Fleming and Simon \(1999\)](#) for relations of this system to practical problems. The first  $d$  streams are composed of customers “having appointments”, meaning that the customers from the  $i$ th stream only go to the  $i$ th service station. The last stream, to which we assign number 0, is that of “free” customers who, upon “checking in”, are routed to the service stations according to a certain rule to be described later. We assume that each service station consists of infinitely many servers, so that infinitely many customers can be served at each station simultaneously. Denote by  $Q_t^i$  the number of customers being served at the  $i$ th station at time  $t$ .

With station  $i$ ,  $i = 1, \dots, d$ , we associate a “cost”  $\alpha_i > 0$  and suppose that a “free” customer arriving at time  $t$  is directed to the  $i$ th station if  $i$  is the smallest integer satisfying

$$\alpha_i Q_{t-}^i \leq \alpha_j Q_{t-}^j \quad \text{for all } j \neq i.$$

Such a routing policy is called load balancing ([Fleming and Simon, 1999](#)). Here and below in this section the summation convention over repeated indices is *not enforced*.

We take some numbers  $\lambda_0, \dots, \lambda_d > 0$  and assume that the  $i$ th stream of customers forms a Poisson process with parameter  $\lambda_i$ . To describe the service times we fix some “thresholds”  $N^1, \dots, N^d$ , which are positive integers, and assume that, given  $0 < Q_t^i < N^i$ , each of  $Q_t^i$  customers at the  $i$ th station

- (i) has its own server,
- (ii) spends with its server a random time having exponential distribution with parameter 1,
- (iii) after having been served leaves the system.

However, given  $Q_t^i \geq N^i$ , the service is organized differently. All  $Q_t^i$  customers are divided into disjoint groups each consisting of two persons apart from at most one group having only one member. Then each of those groups is supposed to get service according to rules (i)–(iii) above. By the way, it is not hard to understand that on average both disciplines of servicing yield the same number of customers having been served during one unit of time.

Finally, we assume that all service times and arrival processes are as independent as they can possibly be.

Now we describe the model in rigorous terms. For any numbers  $y^1, \dots, y^d$  define

$$\operatorname{argmin}_{k=1, \dots, d} y^k = i$$

if  $i$  is the least of  $1, \dots, d$  such that  $y^i \leq y^k$  for  $k \neq i$ . For  $x \in \mathbb{R}^d$  and  $i = 1, \dots, d$ , let

$$\delta^i(x) = \begin{cases} 1 & \text{if } i = \operatorname{argmin}_{k=1, \dots, d} \alpha_k x^k, \\ 0 & \text{otherwise.} \end{cases}$$

Also take independent Poisson processes  $\Pi_t^0, \dots, \Pi_t^d$  with parameters  $\lambda_0, \dots, \lambda_d$ , respectively. Then we think of the number of arrivals at the  $i$ th station as given by

$$A_t^i = \int_0^t \delta^i(Q_{s-}) d\Pi_s^0 + \Pi_t^i,$$

where  $Q_s = (Q_s^1, \dots, Q_s^d)$  and  $Q_t^i$  are some integer-valued right continuous processes having left limits. To model the number of departures  $D_t^i$  from the  $i$ th station up to time  $t$  we take Poisson processes  $\Pi_t^{ij}$  and  $\Lambda_t^{ij}$ ,  $i = 1, \dots, d$ ,  $j = 1, 2, \dots$ , having parameter 1 and mutually independent and independent of  $(\Pi_t^0, \dots, \Pi_t^d)$ . Then we define

$$D^i(t) = \int_0^t I_{N^i > Q_{s-}^i} \sum_{j \geq 1} I_{Q_{s-}^i \geq j} d\Pi_s^{ij} + \int_0^t I_{N^i \leq Q_{s-}^i} \sum_{j \geq 1} (I_{Q_{s-}^i \geq 2j} + I_{Q_{s-}^i + 1 \geq 2j}) d\Lambda_s^{ij}.$$

To be consistent with the description,  $Q_t$  should satisfy the balance equations  $Q_t^i = Q_0^i + A_t^i - D_t^i$ . Thus, we are going to investigate the system of equations

$$dQ_t^i = \delta^i(Q_{t-}) d\Pi_t^0 + d\Pi_t^i - I_{N^i > Q_{t-}^i} \sum_{j \geq 1} I_{Q_{t-}^i \geq j} d\Pi_t^{ij} - I_{N^i \leq Q_{t-}^i} \sum_{j \geq 1} (I_{Q_{t-}^i \geq 2j} + I_{Q_{t-}^i + 1 \geq 2j}) d\Lambda_t^{ij}, \quad i = 1, \dots, d. \tag{4.1}$$

Needless to say that we assume that all the Poisson processes we are dealing with are given on a probability basis satisfying the “usual” assumptions. We also assume that the initial condition  $Q_0$  is independent of the Poisson processes.

Notice that for any initial condition  $Q_0$  there is a unique solution of (4.1). Indeed obviously, for any solution we have  $Q_t^i \leq Q_0^i + \Pi_t^0 + \Pi_t^i$ , so that, while solving (4.1) for  $t \in [0, T]$ , one can safely replace the infinite sums in (4.1) with the sums over  $j \leq Q_0^i + \Pi_T^0 + \Pi_T^i$ . After that one solves (4.1) on each  $\omega$  noticing that between the jumps of the Poisson processes  $Q_t$  is constant and the jumps of  $Q_t$  themselves are given by (4.1).

For obvious reasons we rewrite (4.1) in terms of representation (2.1). First, for  $k = 0, \dots, d$ ,  $i = 1, \dots, d$ ,  $j \geq 1$ , we define

$$\bar{\Pi}_t^k = \Pi_t^k - \lambda_k t, \quad \bar{\Pi}_t^{ij} = \Pi_t^{ij} - t, \quad \bar{\Lambda}_t^{ij} = \Lambda_t^{ij} - t.$$

These processes are square integrable martingales with

$$\langle \bar{\Pi}^k \rangle_t = \lambda_k t, \quad \langle \bar{\Pi}^{ij} \rangle_t = t, \quad \langle \bar{\Lambda}^{ij} \rangle_t = t.$$

Next, for  $i = 1, \dots, d$ , define

$$M_t^i = \int_0^t \delta^i(Q_{s-}) d\bar{\Pi}_s^0 + \bar{\Pi}_t^i - \int_0^t I_{N^i > Q_{s-}^i} \sum_{j \geq 1} I_{Q_{s-}^i \geq j} d\bar{\Pi}_s^{ij} \\ - \int_0^t I_{N^i \leq Q_{s-}^i} \sum_{j \geq 1} (I_{Q_{s-}^i \geq 2j} + I_{Q_{s-}^i + 1 \geq 2j}) d\bar{\Lambda}_s^{ij},$$

which are at least locally square integrable martingales. Then after observing that, for any integer  $q \geq 0$ ,

$$\sum_{j \geq 1} I_{q \geq j} = q, \quad \sum_{j \geq 1} (I_{q \geq 2j} + I_{q+1 \geq 2j}) = q,$$

we turn Eq. (4.1) into the equation

$$dQ_t^i = (\lambda_0 \delta^i(Q_t) + \lambda_i - Q_t^i) dt + dM_t^i. \tag{4.2}$$

In order to explain what follows (in no way is this explanation used in the proof of Theorem 4.1 below), notice that (4.2) seems to imply that

$$(EQ_t^i)' = \lambda_0 E\delta^i(Q_t) + \lambda_i - EQ_t^i. \tag{4.3}$$

We are interested in the behavior of  $Q_t$  when  $\lambda_i$ 's are large but  $\lambda_0$  is much smaller than  $\lambda_1, \dots, \lambda_d$ . Then, on the one hand,  $EQ_t^i$  should be large for moderate  $t$  and, on the other hand, the first term on the right in (4.3) can be neglected. In that situation Eq. (4.3) turns out to have a stable point  $EQ_t^i \equiv \lambda_i$ . This means that, if for the initial condition we have  $EQ_0^i = \lambda_i$ , then  $EQ_t^i = \lambda_i$  for all  $t$ . Notice that since  $\lambda_i$ 's are large, so should  $EQ_0^i$  be.

Therefore, we write  $\lambda_i = \bar{\lambda}_i + \Delta\lambda_i$ , where  $\Delta\lambda_i$  will be assumed to have order of  $\lambda_0$ , denote

$$\bar{Q}_t^i = Q_t^i - \bar{\lambda}_i$$

and rewrite (4.2) in terms of  $\bar{Q}_t$ . At this moment we introduce the assumption that

$$\bar{\lambda}_i \alpha_i = n, \quad i = 1, \dots, d, \tag{4.4}$$

with  $n$  being an integer (independent of  $i$ ) to be sent to infinity. This is convenient due to the simple fact that then

$$\delta^i(x) = \delta^i(x - \bar{\lambda}_i).$$

In this notation (4.2) becomes

$$d\bar{Q}_t^i = (\lambda_0 \delta^i(\bar{Q}_t) + \Delta\lambda_i - \bar{Q}_t^i) dt + dM_t^i.$$

To understand what kind of normalization is natural we compute the quadratic characteristics of  $M_t^i$ . Notice that, for any integer  $q \geq 0$ , we have

$$\begin{aligned} \sum_{j \geq 1} (I_{q \geq 2j} + I_{q+1 \geq 2j})^2 &= \sum_{j \geq 1} (I_{q \geq 2j} + 2I_{q \geq 2j} + I_{q+1 \geq 2j}) \\ &= 3[q/2] + [(q + 1)/2] = :qf(q), \end{aligned}$$

where  $[a]$  is the integer part of  $a$ . By the way, we can only define  $f(q)$  by the above formula for all real  $q > 0$ . If  $q \leq 0$ , we let  $f(q) = 0$ . Then

$$0 \leq f \leq 2, \quad \lim_{q \rightarrow \infty} f(q) = 2. \tag{4.5}$$

It follows that

$$d\langle M \rangle_t^{ii} = [\lambda_0 \delta^i(\bar{Q}_t) + \lambda_i + Q_t^i I_{Q_t^i < N^i} + Q_t^i f(Q_t^i) I_{Q_t^i \geq N^i}] dt.$$

Also due to independence of our Poisson processes and the fact that  $\delta^i \delta^j = 0$  for  $i \neq j$ , we get

$$\langle M \rangle_t^{ij} = 0 \quad \text{for } i \neq j.$$

If we believe that, in a sense,  $Q_t^i \sim \lambda_i$ , then  $Q_t^i/\lambda_i$  should converge as well as  $M_t^i/\sqrt{\lambda_i}$ , and we see that it is natural to expect  $\bar{Q}_t^i/\sqrt{\lambda_i}$  to converge to a certain limit. To make the model more meaningful we also assume that the thresholds  $N^i$ 's are large and roughly speaking proportional to  $\lambda_i$ . In this way we convince ourselves that the following result seems natural.

**Theorem 4.1.** *Let  $\alpha_1, \dots, \alpha_d > 0$  and  $\mu_0, \dots, \mu_d \geq 0$  and  $v_1, \dots, v_d \in \mathbb{R}$  be fixed parameters. For  $n = 1, 2, \dots$  define*

$$\lambda_i = n\alpha_i^{-1} + \mu_i\sqrt{n}, \quad i = 1, \dots, d, \quad \lambda_0 = \mu_0\sqrt{n},$$

$$N^i = n\alpha_i^{-1} + v_i\sqrt{n}, \quad i = 1, \dots, d.$$

Let  $Q_t = Q_t^n$  be the solution of (4.1) with certain initial condition independent of the Poisson processes and introduce

$$x_t^n = n^{-1/2}(Q_t^{n1} - n\alpha_1^{-1}, \dots, Q_t^{nd} - n\alpha_d^{-1}).$$

Let  $\mathbb{Q}^n$  be the distribution of  $x^n$  on  $\mathcal{D}$ . Finally, assume that the distribution of  $x_0^n$  weakly converges to a distribution  $F_0$  as  $n \rightarrow \infty$ .

Then, as  $n \rightarrow \infty$ ,  $\mathbb{Q}^n$  converges weakly to the distribution of a solution of the following system:

$$dx_t^i = (\mu_0 \delta^i(x_t) + \mu_i - x_t^i) dt + \alpha_i^{-1/2} (2 + I_{x_t^i \geq v_i})^{1/2} dw_t^i, \quad i = 1, \dots, d \tag{4.6}$$

considered on some probability space with  $w_t$  being a  $d$ -dimensional Wiener process and  $x_0$  distributed according to  $F_0$ .

**Proof.** First of all notice that (4.6) has solutions on appropriate probability spaces and any solution has the same distribution on the space of  $\mathbb{R}^d$ -valued continuous functions.

This follows from the fact that an obvious change of probability measure allows us to consider the case with no drift terms in (4.6). In that case, (4.6) becomes just a collection of unrelated one-dimensional equations with uniformly nondegenerate and bounded diffusion. Weak unique solvability of such equations is a very well-known fact (see, for instance, Krylov, 1969, Theorems 2 and 3).

In the proof of convergence we will be using Theorems 3.2 and 2.1. Observe that Assumption 2.4 is satisfied since  $x_t^{ni}$  has no jumps bigger than  $2n^{-1/2}$  and  $v^n((0, \infty) \times B_a^c) = 0$  if  $n > 4d/a^2$ . Furthermore, if in the argument before the theorem we take  $\bar{\lambda}_i = n\alpha_i^{-1}$  so that (4.4) holds, and let  $\Delta\lambda_i = \mu_i\sqrt{n}$ , then after noticing that, by definition,

$$Q^{ni} = n^{1/2}x_t^{ni} + n\alpha_i^{-1},$$

we easily obtain

$$dx_t^n = b^n(x_t^n) dt + dm_t^n, \quad \langle m^n \rangle_t = \int_0^t a^n(x_s^n) ds, \tag{4.7}$$

where

$$\begin{aligned} b^{ni}(x) &= \mu_0\delta^i(x) + \mu_i - x^i, \\ a^{nij}(x) &= \delta^{ij} (n^{-1/2}\mu_0\delta^i(x) + \alpha_i^{-1} + \mu_in^{-1/2} \\ &\quad + (x^i n^{-1/2} + \alpha_i^{-1})_+ [I_{x^i < v^i} + f(n^{1/2}x^i + n\alpha_i^{-1})I_{x^i \geq v^i}]). \end{aligned}$$

Upon remembering (4.5) we see that, for a constant  $N$  and all  $n$  and  $x$ , we have  $|b^n(x)| + \text{trace } a^n(x) \leq N(1 + |x|)$ , which shows that Assumptions 2.1 and 3.1, equivalent in our present situation, and Assumption 3.2 are satisfied. By Theorem 3.2 the sequence  $(Q^n)$  is precompact.

Next, obviously Assumption 2.2 is satisfied if we take

$$\begin{aligned} G &= \left\{ (t, x): t > 0, \prod_{i,j=1}^d (\alpha_i x^i - \alpha_j x^j)(x^i - v_i) = 0 \right\}, \\ b^i(x) &= \mu_0\delta^i(x) + \mu_i - x^i, \quad a^{ij}(x) = \delta^{ij}\alpha_i^{-1}(1 + I_{x^i < v^i} + 2I_{x^i \geq v^i}). \end{aligned}$$

Finally, Assumption 2.3 is satisfied since  $\det a^n(x) \geq \alpha_1^{-1} \cdots \alpha_d^{-1}$  everywhere.

By Theorem 2.1, every convergent subsequence of  $(Q^n)$  converges to the distribution of a solution of (4.6) with the above specified initial distribution. Since all such solutions have the same distribution, the whole sequence  $(Q^n)$  converges to the distribution of any solution of (4.6). The theorem is proved.  $\square$

**Remark 4.1.** In Theorem 4.1 we assume that  $Q_0^{ni}$  goes to infinity with certain rate, namely  $Q_0^{ni} \sim n\alpha_i^{-1}$ . Interestingly enough, if we change the rate, the diffusion approximation changes. Indeed, keep all the assumption of Theorem 4.1 apart from the assumption that  $x_0^n$  converges in distribution and instead assume that, for a  $\gamma \in [0, \infty)$  say for  $\gamma = 0$ ,

$$n^{-1/2}(Q_0^{n1} - n\gamma\alpha_1^{-1}, \dots, Q_0^{nd} - n\gamma\alpha_d^{-1})$$

converges in law to a random vector. Notice that the case  $\gamma = 1$  is covered by Theorem 4.1. We claim that, for  $\gamma > 1$ , the processes

$$y_t^n = n^{-1/2}(Q_t^{n1} - nq_t\alpha_1^{-1}, \dots, Q_t^{nd} - nq_t\alpha_d^{-1}),$$

where  $q_t = 1 + (\gamma - 1)e^{-t}$ , weakly converge to a solution of the system

$$dy_t^i = (\mu_0\delta^i(y_t) + \mu_i - y_t^i) dt + \alpha_i^{-1/2}(1 + q_t)^{1/2} dw_t^i, \quad i = 1, \dots, d,$$

and for  $\gamma \in [0, 1)$  weakly converge to a solution of

$$dy_t^i = (\mu_0\delta^i(y_t) + \mu_i - y_t^i) dt + \alpha_i^{-1/2}(1 + 2q_t)^{1/2} dw_t^i, \quad i = 1, \dots, d.$$

Indeed, we have

$$Q_t^{ni} = n^{1/2}y_t^{ni} + nq_t\alpha_i^{-1}, \quad dq_t = (1 - q_t) dt,$$

$$dy_t^n = b^n(y_t^n) dt + dm_t^n, \quad \langle m^n \rangle_t = \int_0^t a^n(y_s^n) ds,$$

where

$$\begin{aligned} b^{ni}(x) &= \mu_0\delta^i(x) + \mu_i - x^i, \\ a^{nij}(x) &= \delta^{ij} \left( n^{-1/2}\mu_0\delta^i(x) + \alpha_i^{-1} + \mu_i n^{-1/2} \right. \\ &\quad \left. + (x^i n^{-1/2} + q_t\alpha_i^{-1})_+ [I_{(\gamma-1)e^{-t} < \alpha_i(y^i - x^i)n^{-1/2}} \right. \\ &\quad \left. + f(n^{1/2}x^i + nq_t\alpha_i^{-1})I_{(\gamma-1)e^{-t} \geq \alpha_i(y^i - x^i)n^{-1/2}}] \right). \end{aligned}$$

As in the proof of Theorem 4.1 one checks that the sequence of distributions of  $y_t^n$  is precompact. Furthermore, obviously, for any  $x$

$$a^{nij}(x) \rightarrow \begin{cases} \delta^{ij}\alpha_i^{-1}(1 + q_t) & \text{if } \gamma < 1, \\ \delta^{ij}\alpha_i^{-1}(1 + 2q_t) & \text{if } \gamma > 1, \end{cases}$$

and this yields our claim in the same way as in the proof of Theorem 4.1.

**Remark 4.2.** We tried to explain before the proof of Theorem 4.1 why its statement looks natural. Now we can also explain how the function  $q_t$  from Remark 4.1 was found. The explanation is based on a kind of law of large numbers which in queueing theory is associated with the so-called “fluid approximations”. Generally, “fluid approximations” can also be derived from Theorems 3.2 and 2.1. For instance, if  $\lambda_k = \lambda_k(n)$  and  $\lambda_k(n)/n \rightarrow \beta_k$  as  $n \rightarrow \infty$ , and  $\beta_0 = 0$ , then under the condition that  $Q_0^n/n$  converges in probability to a constant vector, the processes  $Q_t^n/n$  converge in probability uniformly on each finite time interval to the deterministic solution of the system

$$dq_t^i = (\beta_i - q_t^i) dt, \quad i = 1, \dots, d.$$

This fact obviously follows from Theorems 3.2 and 2.1 applied to (4.2) written in terms of  $z_t^n := Q_t^n/n$ :

$$dz_t^{ni} = b^{ni}(z_t^n) dt + dM_t^{ni},$$

with  $d\langle M^n \rangle_t^{ij} = a_t^n(z_t^n) dt$ ,

$$b^{ni}(x) = \delta^i(x)\lambda_0/n + \lambda_i/n - x, \quad |a_t^{nij}(x)| \leq Nn^{-1}(1 + |x|),$$

where the constant  $N$  is independent of  $x, n, t$ .

The following observation can be generalized so as to be used in various control problems in which optimal controls are discontinuous with respect to space variables.

**Remark 4.3.** It turns out that many discontinuous functionals of  $x^n$  converge in law to corresponding functionals of  $x$ . For instance take a Borel vector-valued function  $f(x)$  on  $\mathbb{R}^d$  such that the set of its discontinuities lies in a closed set  $J \subset \mathbb{R}^d$  having Lebesgue measure zero. Also assume that  $f$  is locally bounded, that is bounded on any ball in  $\mathbb{R}^d$  but may behave in any way at infinity. As an example, one can take  $f(x) = (\delta^1(x), \dots, \delta^d(x))$ . Then, for

$$y_t^n := \int_0^t f(x_s^n) ds, \quad y_t := \int_0^t f(x_s) ds$$

we have that the distributions of  $(x^n, y^n)$  converge weakly to the distribution of  $(x, y)$ .

Indeed, append (4.7) with one more equation:  $dy_t^n = f(x_t^n) dt$  and consider the couple  $z^n = (x^n, y^n)$  as a process in  $\mathbb{R}^{d+1}$ . Obviously, Assumptions 2.1 and 2.4 are satisfied for thus obtained couple.

Furthermore, define

$$H = \left\{ (t, x, y): t > 0, y \in \mathbb{R}, x \in J \text{ or } \prod_{i,j=1}^d (\alpha_i x^i - \alpha_j x^j)(x^i - v_i) = 0 \right\}.$$

Since  $J$  is closed, for any  $t > 0$  and  $(x, y) \notin H_t$ , the function  $f$  (independent of  $y$ ) is continuous in a neighborhood of  $x$ , which along with the argument in the proof of Theorem 4.1 shows that Assumption 2.2 is satisfied for  $z_t^n$ . Finally, for

$$H^m \equiv H, \quad d_m = d, \quad v^{mi}(t, x, y) = x^i, \quad i = 1, \dots, d,$$

we have

$$v^m(H_t) = \left\{ x: x \in J \text{ or } \prod_{i,j=1}^d (\alpha_i x^i - \alpha_j x^j)(x^i - v_i) = 0 \right\}$$

which has  $d$ -dimensional Lebesgue measure zero and

$$\det V^{nm}(t, x^n, y^n) = \det a^n(x_t^n) \geq \alpha_1^{-1} \dots \alpha_d^{-1} > 0.$$

Hence Assumption 2.5 is satisfied as well. This along with precompactness of distributions of  $(x^n, y^n)$  guaranteed by Theorem 3.2 and along with Theorem 2.5 shows that any convergent subsequence of distributions of  $(x^n, y^n)$  converges to the distribution of a process  $(x, y)$ , whose first component satisfies (4.6) and the second one obeys  $dy_t = f(x_t) dt$ .

Thus, we get our assertion for a subsequence instead of the whole sequence. However, as we have noticed above, solutions of (4.6) are weakly unique and this obviously

implies that solutions of the system (4.6) appended with  $dy_t = f(x_t) dt$  are also weakly unique. Therefore, the whole sequence of distributions of  $(x^n, y^n)$  converges.

### 5. An $L_p$ estimate

Let  $d \geq 1$  be an integer,  $(\Omega, \mathcal{F}, P)$  be a complete probability space, and  $(\mathcal{F}_t, t \geq 0)$  be an increasing filtration of  $\sigma$ -fields  $\mathcal{F}_t \subset \mathcal{F}$  with  $\mathcal{F}_0$  being complete with respect to  $P, \mathcal{F}$ . Let  $K(r, t)$  and  $L(r, t)$  be two nonnegative deterministic functions defined for  $r, t > 0$ . Assume that they increase in  $r$  and are locally integrable in  $t$ , so that

$$\int_0^T (K(r, t) + L(r, t)) dt < \infty \quad \forall r, T \in (0, \infty).$$

Let  $\delta(t, x)$  be a nonnegative deterministic function defined for  $t \geq 0$  and  $x \in \mathbb{R}^d$  and satisfying  $\delta(t, x) \leq K(|x|, t)$ . Define  $A(t, x)$  as the set of all symmetric nonnegative  $d \times d$ -matrices  $a$  such that

$$\delta(t, x)|\lambda|^2 \leq a^{ij} \lambda^i \lambda^j \leq K(|x|, t)|\lambda|^2 \quad \forall \lambda \in \mathbb{R}^d.$$

Here, as well as everywhere in the article apart from Section 4, we use the summation convention. For any symmetric  $d \times d$ -matrix  $v = (v_{ij})$  define

$$F(t, x, v) = \sup_{a \in A(t, x)} a^{ij} v_{ij}.$$

As is easy to see, if  $\lambda_i(v)$ ,  $i = 1, \dots, d$ , are eigenvalues of  $v$  numbered in any order, then

$$F(t, x, v) = \sum_{i=1}^d \chi(t, x, \lambda_i(v)),$$

where  $\chi(t, x, \lambda) = K(|x|, t)\lambda$  for  $\lambda \geq 0$  and  $\chi(t, x, \lambda) = \delta(t, x)\lambda$  for  $\lambda \leq 0$ .

Remember that  $C_0^\infty(\mathbb{R}^{d+1})$  is the set of all infinitely differentiable real-valued function  $u = u(t, x)$  on  $\mathbb{R}^{d+1}$  with compact support.

**Theorem 5.1.** *Let  $x_t$ ,  $t \geq 0$ , be an  $\mathbb{R}^d$ -valued  $\mathcal{F}_t$ -adapted continuous process such that, for any  $u \in C_0^\infty(\mathbb{R}^{d+1})$ , the following process is a local  $\mathcal{F}_t$ -supermartingale:*

$$u(t, x_t) - \int_0^t [u_s(s, x_s) + F(s, x_s, u_{xx}(s, x_s)) + L(x_t^*, s)|u_x(s, x_s)|] ds, \tag{5.1}$$

where  $u_x$  is the gradient of  $u$  with respect to  $x$ ,  $u_{xx}$  is the matrix of second-order derivatives  $u_{x^i x^j}$  of  $u$ ,

$$u_s = \partial u / \partial s, \quad u_{x^i x^j} = \partial^2 u / \partial x^i \partial x^j.$$

Then for any  $r, T \in (0, \infty)$  there exists a constant  $N < \infty$ , depending only on  $r, L(r, T)$ , and  $d$  (but not on  $K(r, t)$ ), such that, for any nonnegative Borel  $f(t, x)$ , we have

$$E \int_0^{T \wedge \tau_r} \delta^{d/(d+1)}(t, x_t) f(t, x_t) dt \leq N \|f\|_{L_{d+1}([0, T] \times B_r)}, \tag{5.2}$$

where

$$\|f\|_{L_{d+1}([0,T] \times B_r)} = \left( \int_0^T \int_{|x| \leq r} f^{d+1}(t,x) dx dt \right)^{1/(d+1)}.$$

$B_r$  is the open ball in  $\mathbb{R}^d$  of radius  $r$  centered at the origin, and  $\tau_r$  is the first exit time of  $x_t$  from  $B_r$ .

**Proof.** First of all notice that for any  $u \in C_0^\infty(\mathbb{R}^{d+1})$  expression (5.1) makes sense. Indeed, if  $r$  is such that  $u(t,x) = 0$  for  $|x| \geq r$  and all  $t$ , then the integral is bounded by a constant times

$$\int_0^t [1 + K(r,s) + L(x_t^* + r,s)] ds,$$

which is finite since each trajectory of  $x_s$  is bounded on  $[0,t]$ . Also observe that the usual approximation techniques allow us to only concentrate on the case of infinitely differentiable functions  $f \geq 0$  vanishing for  $|x| \geq r$  for some  $r$ . We fix  $r$ , such a function  $f$ , and a nonnegative function  $\zeta \in C_0^\infty(\mathbb{R}^{d+1})$  with unit integral and support in the unit ball of  $\mathbb{R}^{d+1}$  centered at the origin. Below, for any locally bounded Borel function  $g(t,y)$  and  $\varepsilon > 0$  we use the notation

$$g^{(\varepsilon)} = g * \zeta_\varepsilon, \quad \text{where } \zeta_\varepsilon(t,x) = \varepsilon^{-d-1} \zeta(t/\varepsilon, x/\varepsilon).$$

Next, we need Theorem 2 of Krylov (1976), which states the following. There exist constants  $\alpha = \alpha(d) > 0$  and  $N_r = N(r,d) < \infty$  and there exists a bounded Borel non-positive function  $z$  on  $\mathbb{R}^{d+1}$  which is convex on  $B_{2r}$  for each fixed  $t$  and is such that, for each nonnegative symmetric matrix  $a$ ,

$$\alpha(\det a)^{1/(d+1)} f^{(\varepsilon)} \leq z_t^{(\varepsilon)} + a^{ij} z_{x_i x_j}^{(\varepsilon)} \quad \text{for } \varepsilon \leq r, t \in \mathbb{R}, |x| \leq r, \tag{5.3}$$

$$|z_x^{(\varepsilon)}| \leq 2r^{-1} |z^{(\varepsilon)}| \quad \text{for } \varepsilon \leq r/2, t \in \mathbb{R}, |x| \leq r, \tag{5.4}$$

$$|z| \leq N_r \|f\|_{L_{d+1}(\mathbb{R} \times B_r)} \quad \text{in } \mathbb{R} \times B_{2r}. \tag{5.5}$$

Notice that in Theorem 2 of Krylov (1976) there is a minus sign in front of  $z_t^{(\varepsilon)}$ . However, (5.3) is true as is, since one can replace  $t$  with  $-t$  and this does not affect any other term. Observe that (5.5) obviously implies that for  $\varepsilon \leq r$ , we have

$$|z^{(\varepsilon)}| \leq N_r \|f\|_{L_{d+1}(\mathbb{R} \times B_r)} \quad \text{in } \mathbb{R} \times B_r. \tag{5.6}$$

Fix an  $\varepsilon > 0$ . We claim that the process

$$\begin{aligned} \xi_t := & -z^{(\varepsilon)}(t \wedge \tau_r, x_{t \wedge \tau_r}) - \int_0^{t \wedge \tau_r} [-z_s^{(\varepsilon)}(s, x_s) \\ & + F(s, x_s, -z_{xx}^{(\varepsilon)}(s, x_s)) + L(r, s) |z_x^{(\varepsilon)}(s, x_s)|] ds \end{aligned} \tag{5.7}$$

is a local supermartingale. To prove the claim it suffices to prove that (5.7) is a local supermartingale on  $[0, T]$  for every  $T \in [0, \infty)$ . Fix a  $T \in [0, \infty)$  and concentrate on  $t \in [0, T]$ . Change  $-z^{(\varepsilon)}$  outside of  $[0, T] \times B_r$  in any way with the only requirement

that the new function, say  $u$ , belongs to  $C_0^\infty(\mathbb{R}^{d+1})$ . Then process (5.1) is a local supermartingale. Replacing  $t$  with  $t \wedge \tau_r$  yields a local supermartingale again. Also observe that subtracting an increasing continuous process from a local supermartingale preserves the property of being a local supermartingale. After noticing that for  $0 < s \leq t \wedge \tau_r \leq T$ , we have  $|x_s| \leq r$  and  $L(x_s^*, s) \leq L(r, s)$ , we conclude that

$$\eta_t := u(t \wedge \tau_r, x_{t \wedge \tau_r}) - \int_0^{t \wedge \tau_r} [-z_s^{(\varepsilon)}(s, x_s) + F(s, x_s, -z_{xx}^{(\varepsilon)}(s, x_s)) + L(r, s)|z_x^{(\varepsilon)}(s, x_s)|] ds$$

is a local supermartingale on  $[0, T]$ . Since

$$\eta_t - \zeta_t = [u(0, x_0) - z^{(\varepsilon)}(0, x_0)]I_{\tau_r=0}$$

is a bounded martingale, (5.7) is a local supermartingale indeed.

After having proved our claim we notice that for each  $T \in [0, \infty)$ , the process (5.7) is obviously bounded on  $[0, T]$ . Therefore (5.7) is a supermartingale and

$$E \zeta_T I_{\tau_r > 0} \leq E \zeta_0 I_{\tau_r > 0} \leq \sup_{|x| \leq r} |z^{(\varepsilon)}(0, x)|,$$

which along with (5.6), (5.4), and the fact that  $z \leq 0$ , yields that for any  $\varepsilon \leq r/2$

$$E \int_0^{T \wedge \tau_r} [z_s^{(\varepsilon)}(s, x_s) - F(s, x_s, -z_{xx}^{(\varepsilon)}(s, x_s))] ds \leq N_r \|f\|_{L_{d+1}([0, T] \times B_r)} \left( 1 + 2r^{-1} E \int_0^{T \wedge \tau_r} L(r, s) ds \right).$$

Here, owing to (5.3)

$$\begin{aligned} z_s^{(\varepsilon)} - F(s, x, -z_{xx}^{(\varepsilon)}) &= \inf_{a \in A(s, x)} [z_s^{(\varepsilon)} + a^{ij} z_{x^i x^j}^{(\varepsilon)}] \\ &\geq f^{(\varepsilon)} \alpha \inf_{a \in A(s, x)} (\det a)^{1/(d+1)} = f^{(\varepsilon)} \alpha \delta^{d/(d+1)}. \end{aligned}$$

Hence

$$E \int_0^{T \wedge \tau_r} \delta^{d/(d+1)} f^{(\varepsilon)}(s, x_s) ds \leq N \|f\|_{L_{d+1}([0, T] \times B_r)}$$

with

$$N = N_r \alpha^{-1} \left( 1 + 2r^{-1} \int_0^T L(r, s) ds \right).$$

Finally we let  $\varepsilon \downarrow 0$  and use the continuity of  $f$  which guarantees that  $f^{(\varepsilon)} \rightarrow f$ . Then upon remembering that  $f \geq 0$  and using Fatou’s theorem, we arrive at (5.2) with the above specified  $N$ . The theorem is proved.  $\square$

**Remark 5.1.** Actually, we did not use the continuity of  $x_t$ . We could have only assumed that  $x_t$  is a separable measurable process. However, then it turns out that the

assumption about processes (5.1) implies that  $x_t$  is continuous anyway and moreover that  $x_t$  is an Itô process (Krylov, 1989).

## Acknowledgements

The authors are sincerely grateful to the referees for many useful suggestions.

## References

- Chao, Yi-Ju, 1999. Diffusion approximation of a sequence of semimartingales and its application in exploring the asymptotic behavior of some queueing networks. Ph.D. Thesis, University of Minnesota, June.
- Fleming, P.J., Simon, B., 1999. Heavy traffic approximations for a system of infinite servers with load balancing. *Probab. Eng. Inform. Sci.* 13, 251–273.
- Ikeda, N., Watanabe, S., 1981. *Stochastic Differential Equations and Diffusion Processes*. North-Holland, Amsterdam, Oxford, New York.
- Jacod, J., Shiriyayev, A., 1987. *Limit theorems for stochastic processes*. Grundlehren der Mathematischen Wissenschaften, A Series of Comprehensive Studies in Mathematics. Springer, New York, Berlin, Heidelberg.
- Kogan, Ya., Liptser, R.S., 1993. Limit non-stationary behavior of large closed queueing network with bottleneck. *Queueing Systems* 14, 33–55.
- Khasminskii, R., Krylov, N., 2001. On averaging principle for diffusion processes with null-recurrent fast component. *Stochastic Process. Appl.* 93 (2), 229–240.
- Krylov, N.V., 1969. On Itô's stochastic integral equations. *Teor. Veroyatnost. i Primenen.* 14(2) 340–348 (in Russian); English translation *Theor. Probab. Appl.* 14 (1969) 330–336.
- Krylov, N.V., 1974. Some estimates of the probability density of a stochastic integral. *Izv. Akad. Nauk SSSR Ser. Mat.* 38(1), 228–248 (in Russian); English translation: *Math. USSR Izv.* 8 (1974) 233–254.
- Krylov, N.V., 1976. Sequences of convex functions and estimates of the maximum of the solution of a parabolic equation. *Sibirskii Mat. Zh.* 17, 290–303 (in Russian); English translation: *Siberian Math. J.* 17 (1976) 226–236.
- Krylov, N.V., 1989. A supermartingale characterization of a set of stochastic integrals. *Ukrain. Mat. Zh.* 41(6), 757–762 (in Russian); English translation: *Ukrain. Math. J.* 41 (1990) 650–654.
- Krylov, N.V., Rozovsky, B.L., 1979. Stochastic evolution equations. In: *Itogy nauki i tekhniki*, Vol. 14. VINITI, Moscow, pp. 71–146 (in Russian); English translation *J. Soviet Math.* 16 (1981) 1233–1277.
- Kurtz, T.G., Protter, P.E., 1996. Weak convergence of stochastic integrals and differential equations. II. Infinite-dimensional case. *Probabilistic Models for Nonlinear Partial Differential Equations* (Montecatini Terme, 1995). pp. 197–285. *Lecture Notes in Mathematics*, Vol. 1627. Springer, Berlin.
- Kushner, H.J., 1984. *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory*. MIT Press, Cambridge.
- Kushner, H.J., 1990. *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*. Birkhäuser, Basel.
- Kushner, H.J., Dupuis, P.G., 2001. *Numerical Methods for Stochastic Control Problems in Continuous Time*, 2nd Edition. Springer, Berlin.
- Kushner, H.J., Runggaldier, W., 1987a. Nearly optimal state feedback controls for stochastic systems with wideband noise disturbances. *SIAM J. Control Optim.* 25, 289–315.
- Kushner, H.J., Runggaldier, W.J., 1987b. Filtering and control for wide bandwidth noise driven systems. *IEEE Trans. Automat. Control* AC-32, 123–133.
- Liptser, R.S., Runggaldier, W.J., 1995. Non-linear filters for linear models (a robust approach). *IEEE Trans. Inform. Theory* 41, 1001–1009.

- Liptser, R.Sh., Shirayayev, A.N., 1986. Theory of Martingales. Nauka, Moscow (in Russian); English translation: Kluwer Academic Publishers, Dordrecht, 1989.
- Liptser, R.S., Runggaldier, W.J., Taksar, M., 1999. Diffusion approximation and optimal stochastic control (Russian). *Teor. Veroyatnost. i Primenen.* 44, 705–737.
- Skorokhod, A.V., 1961. *Issledovaniya po teorii sluchainykh processov.* Izd-vo Kievskogo Universiteta, Kiev (in Russian); English translation: Studies in the Theory of Random Processes. Dover, New York, 1982.
- Stroock, D.W., Varadhan, S.R.S., 1979. *Multidimensional Diffusion Processes.* Springer, Berlin, New York.