Helsinki University of Technology Institute of Mathematics Reports

Teknillisen korkeakoulun matematiikan laitoksen raporttisarja

Espoo 2004 C17

ON APPROXIMATIONS OF DIFFUSIONS WITH EQUILIBRIUM

Alexander Yu. Veretennikov

TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY
TECHNISCHE UNIVERSITÄT HELSINKI
UNIVERSITE DE TECHNOLOGIE D'HELSINKI

Helsinki University of Technology Institute of Mathematics Reports
Teknillisen korkeakoulun matematiikan laitoksen raporttisarja
Espoo 2004 C17

ON APPROXIMATIONS OF DIFFUSIONS WITH EQUILIBRIUM

Alexander Yu. Veretennikov

Helsinki University of Technology
Department of Engineering Physics and Mathematics
Institute of Mathematics

Alexander Yu. Veretennikov: *On approximations of diffusions with equilibrium*; Helsinki University of Technology Institute of Mathematics Reports C17 (2004).

Abstract: *These are the lecture notes of the summer (by definition) graduate minicourse at Helsinki University of Technology, 6-8.09.2004. The topics of the course are: "simple" Markov diffusions in $\mathbb{R}^d$ with equilibrium or stationary measures; recurrence properties and convergence to a stationary measure; "strong" and "weak" discretisation problems; convergence to equilibrium for approximations; invariant measures depending on a parameter, their smoothness. The author thanks Helsinki University of Technology for hospitality which made it possible to prepare this minicourse as a part of the programme "New Techniques in Applied Stochastics", 2004-2005, supported by Finnish Mathematical Society and the Finnish Academy*

AMS subject classifications: 60-01, 60H10, 60J05, 60F10, 60H35, 60J05

Keywords: stochastic differential equations, diffusions, discretisation, euler scheme, ergodicity

Alexander Yu. Veretennikov
University of Leeds, UK &
Institute of Information Transmission Problems, Moscow, Russia
a.veretennikov@leeds.ac.uk

ISBN 951-22-7294-6
ISSN 0784-6460

Helsinki University of Technology
Department of Engineering Physics and Mathematics
Institute of Mathematics
P.O. Box 1100, 02015 HUT, Finland
email:math@hut.fi http://www.math.hut.fi/

Introduction

Consider a stochastic differential equation in $\mathbb{R}^d$,

$$dX_t = b(X_t)\,dt + dW_t,\ X_0 = x \in \mathbb{R}^d, \tag{1}$$

with a bounded Borel drift $b : \mathbb{R}^d \to \mathbb{R}^d$, and d-dimensional Wiener process W_t. In this course we will usually assume coefficient b to be Lipschitz, unless stated directly otherwise, and the process (X_t) is a strong solution to the equation (1), – the latter is correct, in particular, for any Borel bounded b, – which is automatically homogeneous, strong Markov and unique in law. Partially, the results can be extended to more general equations

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

at least if the diffusion coefficient σ is uniformly non-degenerate and Lipschitz matrix-function $d \times d$.

The following set of questions will be addressed.

1. Strong convergence for an Euler type scheme on a finite horizon $[0, T]$; successive approximations of solutions and their Euler approximations; one case of an approximation on the infinite horizon.

2. Ergodic properties of SDEs, that is: existence of a finite invariant measure (equilibrium), convergence to this measure in total variation norm based on the coupling method; convergence of approximations.

3. Convergence of the measures, their densities and densities which depend on a parameter; smoothness on this parameter.

Exercises are provided which in particular can be used if any participant wishes to get credits for this course (the list of exercises is to be confirmed at the end of the lectures). To get the credits, one should solve several Exercises and then contact Professor E. Valkeila. One can also get points by finding mistakes in the lecture notes (the participants are encouraged to do this!), and in this case please contact also the lecturer.

Prerequisites: it is assumed that the reader is familiar with the notion of stochastic integral, its basic properties including the Itô formula, and simplest results on (strong) solutions of SDEs, of course, including the definitions. Otherwise see any text-book on SDEs; [11] is recommended.

The lecturer's intention is/was to provide in the class some details omitted in the present notes; also we do not repeat here some recent preprint publication(s) providing the link(s) instead. Two 'larger' parts here are presented in detail because of the following reasons. First, the "coupling method" sections, where an attempt has been performed by the author to give a purely analytic presentation of the method, except that measurability questions are hidden (practically, the measurability has been assumed, as in most of the modern papers on the subject). It is quite possible that similar calculus may be found in some other works where this method is used, nevertheless, in this version it looks like there is no need to really glue the two processes as it usually happen in similar texts. Secondly, the smoothness with respect to a (finite-dimensional) parameter. This question looks rather "classical", and it is strange that it cannot be found in advanced monographs on Markov processes. First short presentation by the author jointly with E. Pardoux is [27], for "polynomial setting", that is, all assumptions and conclusions are essentially polynomial. The section 3 here can be regarded as an "exponential version" of this paper, with most of the calculus presented; some minor corrections with respect to [27] have been performed without notice. On the other hand, our results for Markov chains here provide corollaries for diffusions under assumptions clearly relaxed in compare with [16] in what concerns smoothness conditions on coefficients with respect to the x variable. So, the Theorem 7 below is new indeed. On the other hand, the paper [16] provides polynomial bounds under "polynomial" assumptions. Most probably, our assumptions here could be relaxed in this direction, too, however, we are not trying to do this in order to keep the presentation as simple as possible.

There is a minimal number of references here, and most of them on the lecturer's papers, because of the nature of the text: it just suffices for the lectures. The general references on approximations for SDEs where many other problems have been investigated are [6], [14], [15].

1 Euler's scheme

1.1 Strong convergence

The Euler scheme which we will be studying has the following form,

$$X^h_{(n+1)h} = X^h_{nh} + b(X^h_{nh})h + \xi_{n+1}\sqrt{h},\ X^h_0 = x \in \mathbb{R}^d, \qquad (2)$$

where (ξ_n) is a sequence of i.i.d. random variables. Random variables ξ_k are Gaussian, although some partial results are currently established for much more general cases.

For $\sqrt{h}\xi_n := W_{nh} - W_{(n-1)h}$, it is convenient to define the process X^h_t for all $t \geq 0$ as

$$X^h_t := x + \int_0^t b(X^h_{[s/h]h})\, ds + W_t.$$

Theorem 1 *Let $b \in Lip$, and the r.v.'s ξ_n IID standard Gaussian, that is, $\sqrt{h}\xi_n := W_{nh} - W_{(n-1)h}$. Then, for any $T > 0$ there exists $C = C_T > 0$ such that*

$$E \sup_{0\leq t\leq T} |X^h_t - X_t|^2 \leq Ch.$$

Proof. For the scheme with $\sqrt{h}\xi_n = W_{nh} - W_{(n-1)h}$ we have,

$$X_t - X^h_t = \int_0^t (b(X_s) - b(X^h_{[s/h]h}))\, ds.$$

Now use the Gronwall lemma, a priory estimate $E\sup_{t\le T}(|X_t|+|X_t^h|)^2 < \infty$, plus the bound $E|X_t^h - X_s^h|^2 \le Ch$. $\square$

Exercise 1 *Show all this in detail.*

1.2 Successive approximations

This is a well known method for constructing and modelling solutions to SDEs: we set

$$X_t^0 := x,$$

and for any $n = 1, 2, \ldots$

$$X_t^{n+1} := x + W_t + \int_0^t b(X_s^n)\, ds.$$

Recall that the diffusion coefficient is now a unit matrix; however, note that the general case can be be considered similarly (instead of direct bounds for the Wiener Process one can use, e.g., the Kolmogorov-Doob inequality for continuous martingales, $E\sup_{t\le T} M_t^2 \le 4EM_T^2$). Likewise, we can define

$$X_t^{0,h} := x,$$

and for any $n = 1, 2, \ldots$

$$X_t^{n+1,h} := x + W_t + \int_0^t b(X_{[s/h]h}^{n,h})\, ds.$$

Theorem 2 *For any $T > 0$ we have,*

$$E \sup_{0\le t\le T} |X_t^n - X_t| \le \sum_{k\ge n+1} \frac{C^k T^k}{k!},$$

and uniformly w.r.t. $h \le 1$, also

$$E \sup_{0\le t\le T} |X_t^{n,h} - X_t^h| \le Ch + \sum_{k\ge n+1} \frac{C^k T^k}{k!}.$$

Notice however that the Lipschitz assumption is essential for the method, so that it is not universal.

Proof. Use the iterations,

$$\begin{aligned} E \sup_{0\le s\le t} |X_s^{n+1} - X_s^n| &\le C \int_0^t E|X_s^n - X_s^{n-1}|\, ds \qquad (3) \\ &\le C \int_0^t E \sup_{0\le s'\le s} |X_{s'}^n - X_{s'}^{n-1}|\, ds' \le \ldots \le \frac{C^n t^n}{n!}. \end{aligned}$$

Exactly the same inequalities hold true also without expectations. Then define X_t as the series,

$$X_t := X_t^0 + \sum_{k=1}^{\infty}(X_t^k - X_t^{k-1}).$$

This series converges in L_1, and in the norm $\|\cdot\|_1 = E\sup_{0\le t\le T}|\cdot_t|$, and the limit solves the SDE (1). Now, note the identity, $X_t^n = X_t^0 + \sum_{k=1}^n (X_t^k - X_t^{k-1})$. Hence, $X_t - X_t^n = \sum_{k=n+1}^{\infty}(X_t^k - X_t^{k-1})$, where we can apply the bound (3).

The second assertion can be shown likewise. $\square$

Exercise 2 *Show all this in detail.*

Exercise 3 *Consider the* $\|\cdot\|_2 = E\sup_{0\le t\le T}|\cdot_t|^2$ *norm.*

Exercise 4 *Consider the case of variable diffusion coefficient.*

1.3 Strong approximation on the infinite horizon

In rare cases the strong approximations may converge on the whole half-line, however, in a bit weaker metric.

Assumption (L): there exists $r > 0$ such that for any x, y,

$$\langle b(x) - b(y), x - y\rangle \le -r|x - y|^2. \tag{4}$$

Theorem 3 *Let (4) be satisfied in addition to the assumptions of the theorem 1. Then*

$$\sup_{0\le t<\infty} E|X_t^h - X_t|^2 \le Ch. \tag{5}$$

Proof. Denote $z_t := E|X_t^h - X_t|^2$. Using the assumption (L), we easily get for some constant $C > 0$ and any t,

$$\dot z_t \le -rz_t + Cz_t^{1/2}\sqrt{h}. \tag{6}$$

Since for any t the value z_t is finite and non-negatize, this implies (5), or, more precisely, $z_t \le C^2h/r^2$ with C from (6). $\square$

Exercise 5 *Show the latter statement in detail.*

2 Ergodic properties of SDEs

In this and next sections we are interested in weak convergence, that is, the one for the measures and densities rather than for the trajectories. A convenient method for establishing such properties is coupling method; we present a new version of the latter which perhaps may be called a function analysis version. Although all results can be extended to beta-mixing as well (see the references), we do not consider it here.

2.1 Recurrence properties

The idea of establishing convergence in total variation (as well as establishing mixing) is (1) to show exponential mixing for a "process on a certain compact", and (2) to show in addition that with "large probability" the process spends a considerable amount of the whole time on some appropriately chosen compact. So, in a sense, all convergence bounds, at least for time-homogeneous Markov processes, are exponential; however, they can be weakened if the process has not got strong enough recurrent properties. One may say that the rate of convergence to equilibrium (which is assumed, or provided by some simple sufficient conditions) is exponential, or polynomial, or whatever, depending on corresponding exponential or polynomial or likewise recurrent properties. In this course we only study the simplest exponential bounds under rather strong conditions, which can be often relaxed.

Assumption (E1):

$$\limsup_{|x|\to\infty} \frac{\langle b(x), x\rangle}{|x|^2} = -r < 0. \tag{7}$$

Consider the ball $B_R = \{x : |x| \le R\}$, and for our process X_t denote by $\tau \equiv \tau_R^x$ the first hitting time of this ball. Note that we will often drop the indices, or otherwise include them into the simbol of expectation.

Theorem 4 *Under assumption (7), for R large enough there exist a constant $C = C_R > 0$ such that*

$$E_x \exp(c\tau) \le \exp(C(|x|^2 - R^2)_+). \tag{8}$$

We also formulate another exponential bound under a weaker assumption, **Assumption (E2):**

$$\limsup_{|x|\to\infty} \frac{\langle b(x), x\rangle}{|x|} = -r < 0. \tag{9}$$

Theorem 5 *Under assumption (9), for R large enough there exist a constant $C = C_R > 0$ such that*

$$E_x \exp(c\tau) \le \exp(C(|x| - R)_+). \tag{10}$$

Proof of the Theorem 5 in the case $d = 1$. To this end, we use the method of Lyapunov functions[1] which roughly suggests that for some function of time and space variables (t, x) with X_t plugged in instead of x, the "Lebesque part" of the stochastic differential is strictly negative; this implies all desirable bounds.

Let $f(t, x) = \exp(\epsilon|x| + \delta t)$: this will be "the Lyapunov function". Both values $C > 0$ and $\epsilon > 0$ are to be fixed later. Let us apply the Itô formula to the expression $f(t, X_t)$ for $t < \tau$; in accordance to the statement of the

[1]The author of the method was A. M. Lyapunov, 1857-1918.

Theorem, if necessary for the calculus we assume that R is large enough. We can consider separately large positive values of x and large negative ones. Hence, let $x > R$, and $|x| = x$; moreover, we can drop the sign of absolute value for all $t < \tau$. We get,

$$df(t, X_t) = f(t, X_t) \left[\delta + \epsilon b(X_t) + \frac{\epsilon^2}{2}\right] dt + ...dW_t.$$

We have to choose the values ϵ and δ so that the expression in the brackets $[\ldots]$ is strictly negative. Due to our standing assumption, for any $\nu > 0$ there exists such a large R that $b(X_t) \le -(r - \nu)$. Hence, the whole expression in the brackets does not exceed the value $\delta - \epsilon(r - \nu) + \frac{\epsilon^2}{2}$. The latter can be made strictly negative and separated above from zero by an appropriate choice of the constants ϵ and δ.

Hence, applying the Fatou lemma, and possibly the localisation procedure, we get,

$$E_x f(\tau, X_\tau) \le f(0, x) - E_x \int_0^\tau f(s, X_s)\, ds.$$

This implies

$$E_x \exp(\delta\tau + \epsilon R) + E_x \int_0^\tau \exp(\delta s + \epsilon R)\, ds$$

$$\le E_x f(\tau, X_\tau) + E_x \int_0^\tau f(s, X_s)\, ds \le f(0, x) = \exp(\epsilon|x|).$$

Hence,

$$E_x \exp(\delta\tau) + \delta^{-1} E_x \exp(\delta\tau) \le \exp(\epsilon(|x| - R))$$

which means the desired bound, even more, we in some sense get two versions of this bound. Recall that we considered large ($|x| > R$) positive x, and large negative can be considered likewise. We can as well write down the uniform version of this inequality using notation $a_+ = \max(a, 0)$: for any x,

$$E_x \exp(\delta\tau) + \delta^{-1} E_x \exp(\delta\tau) \le \exp(\epsilon(|x| - R)_+).$$

This proves the Theorem. □

Exercise 6 *With the help of the Fatou Lemma, show the localisation arguments dropped in the proof.*

Exercise 7 *Show the Theorem 5 for any finite d. Assume that diffusion coefficient is a unit matrix.*

Exercise 8 *Show the Theorem 4. You may assume dimension $d = 1$ (although this is not necessary) and, of course, diffusion coefficient equal to 1. Hint: use the Lyapunov function $f(t, x) = \exp(\epsilon|x|^2 + \delta t)$.*

2.2 A priori bounds and stationary distribution

Lemma 1 *Under assumption* $(E2)$*, there exist constants* $C, c > 0$ *such that*

$$\sup_t E_x \exp(c|X_t|) \le C \exp(c|x|).$$

Proof can be derived from Comparison Theorem: compare $|X_t|$ with an appropriately chosen one-dimensional SDE on the half-line $[|x|, +\infty)$ with a non-sticky boundary condition and a constant drift $-r + \nu$; for the latter, a stationary distribution can be computed explicitly, which provides the desired a priori bound for X_t. [Note that other proofs are possible.] $\square$

Exercise 9 *Show the details. Hint: in fact, two different comparison theorems are to be exploited here, namely, one compares the Itô process* $|X_t|$ *(not a solution to a markovian SDE, generally speaking) with a solution to an appropriate SDE (a Wiener process should be chosen accurately) with a strictly inferior drift; and the other for the latter SDE with two different initial data, one constant (the boundary value, with* $R = |x|$*) and the other distributed as the stationary distribution for this SDE. The density of this distribution is computed explicitly from the Chapman - Kolmogorov equation and has the form,* $p(y) = C\exp(-cy),\ y \ge |x|$.

Lemma 2 *There exists a stationary distribution* μ_∞ *for the SDE (1).*

Proof is really standard, once the previous Lemma is established: take the Cesaro sequence of measures, then due to compactness take an appropriate subsequence which converges weakly. Then show that the (any!) limiting measure is indeed stationary, using the "Markov shift arguments". Of course, this stationary measure is unique, however, we are not going to use it. Bounded continuous functions is a right class to work with here. Certainly, the Lemma 1 provides much more than needed for compactness. $\square$

Exercise 10 *Show the details. Hint: this technique is used, e.g., in standard presentations of the Ergodic Theorem for Markov chains.*

The last auxiliary result in this subsection concerns irreducibility. Although this property can be derived for our process in a full generality in the case of variable (nondegenerate) diffusion coefficient, using other methods, here we present a very simple tool. This also concerns a condition called Doeblin type one, or local mixing. This is one of the versions of the Doeblin type condition used in all papers of the lecturer on mixing since his work [19], usually derived for diffusions from the Harnack inequality. Notice that the same kind of irreducibility condition is often useful in large deviations, cf., e.g. [3].

Lemma 3 *Given any initial data* x*, the distribution* μ_1^x *of the solution to the SDE (1) has a density with respect to Lebesgue measure. Moreover, for any* $R > 0$*,*

$$\sup_{|x-x'|\le R} \int \min\left(\frac{\mu_1^{x'}(dy)}{\mu_1^x(dy)},\, 1\right) \mu_1^x(dy) > 0. \tag{11}$$

Proof. If $b(\cdot) \equiv 0$ (standard Wiener process) then this is evident. In the case of nontrivial drift, we use the results from PDE theory (see, e.g., [2]) saying that

$$C^{-1} \exp(-c^{-1}|x - x_1|^2) \le p(x, x_1) \le C \exp(-c|x - x_1|^2), \quad C, c > 0.$$

Exercise 11 *Show the details for standard Wiener process.*

2.3 Coupling method: the idea

Now we are in a position to show convergence in total variation to (the) stationary measure for any given initial data; as we do not use uniqueness of this measure, this, in particular, will also imply its uniqueness. We will use the following trick to simplify the task: consider the process X_t at discrete times only, $t = 0, 1, 2, \dots$ The stationary measure we were talking about is automatically stationary for this process, too. If we show convergence to this measure for the distributions μ^x_k, $k = 0, 1, \dots$, $k \to \infty$, this would also provide the desired convergence of the continuous time measures, μ^x_t, $t \to \infty$.

Exercise 12 *Explain why.*

Let us formulate the result.

Theorem 6 *Under the same assumptions (E2) and Lipschitz for b, there exist constants $C, \lambda, \epsilon > 0$ such that*

$$\|\mu^x_n - \mu_\infty\|_{TV} \le C \exp(\epsilon|x| - \lambda n), \quad n = 1, 2, \dots \tag{12}$$

Corollary 1 *Under the assumptions of the Theorem 6, there exist constants $C, \lambda, \epsilon > 0$ such that*

$$\|\mu^x_t - \mu_\infty\|_{TV} \le C \exp(\epsilon|x| - \lambda t), \quad t \to \infty.$$

Note that the the bound (10) holds true (perhaps with another C) also if we redefine τ using only integer values. Moreover, if X'_t is another independent copy of the Markov process X_t with initial data X'_0 distributed according to μ_∞, and $\gamma := \inf(k : \min(|X_k|, |X'_k|) \le R)$, then the same inequality (possibly with new constants) holds true for the stopping time γ as well:

$$E_{x,\mu_\infty} \exp(\delta\gamma) \le C \exp(\epsilon|x|). \tag{13}$$

Exercise 13 *Explain why. Hint: it is natural to start with a pointwise estimate,*

$$E_{x,x'} \exp(\delta\gamma) \le C \exp(\epsilon(|x| + |x'|). \tag{14}$$

Then both parts of this inequality could be integrated with respect to $\mu_\infty(dx')$, which is possible at least for ϵ small enough, due to the estimate of the Lemma 1.

Exercise 14 *Show how the Theorem 6 implies the Corollary.*

As a next step, let us introduce a sequence of stopping times,

$$\gamma_n := \inf(k \ge \gamma_{n-1} : \min(|X_k|, |X'_k|) \le R), \quad n = 1, 2, \dots$$

Naturally, all gamma's possess similar properties, in particular, all of them are finite almost surely. Since the process (X_k, X'_k) is a strong Markov one, at each stopping time it starts as a new Markov process with the same transitions. Hence, let us consider a homogeneous Markov process $Z_n = (X_{\gamma_n}, X'_{\gamma_n})$. Due to (11) and the standard Ergodic Theorem for Markov chains, it is uniformly ergodic with an exponential rate of convergence to the (unique) equilibrium distribution in total variation metric.

Exercise 15 *Show the details. Hint: e.g., the calculus in the next section could be adjusted for this more simple case, so that the contraction property holds for the two measures providing the desired exponential convergence.*

This fact and (13), in particular, imply that the sequence $(\gamma_n, \, n = 1, 2, \dots)$ has the following law of large numbers,

$$\frac{\gamma_n}{n} \to \kappa > 0, \quad \text{a.s.},$$

the constant κ being a stationary version of the expectation $E(\gamma_{n+1} - \gamma_n)$. It is not difficult to see that $\kappa > 1$. Moreover, due to the exponential bound $E_x \exp(\delta\gamma) < C \exp(\epsilon|x|)$ for some $\delta > 0$, one can show the following also exponential bound: for any $\nu > 0$ there exist $C, \lambda' > 0$ such that

$$P_{x,\mu_\infty}\left(\frac{\gamma_n}{n} > (\kappa + \nu)\right) < C \exp(\epsilon|x| - \lambda' n), \quad n = 1, 2, \dots \tag{15}$$

This implies that for any $c > \kappa^{-1}$,

$$P_{x,\mu_\infty}\left(\gamma_{[cn]} > n\right) < C \exp(\epsilon|x| - \lambda' n), \quad n = 1, 2, \dots \tag{16}$$

Exercise 16 *Show both assertions.*

Now, the **idea** of the coupling method is the following observation. The couple (X, X') is a Markov and moreover strong Markov process which frequently visits some (actually, any, but we use the ball B_R with the appropriately chosen radius R) neighbourhood of zero. Each time when this happens, the two components are not far away from each other. At the next (discrete) time, with a positive probability bounded away from zero, both components are still in the ball B_R, and the transition measures are equivalent, e.g., because they are both equivalent to the transition measures of the Wiener process. The latter can be shown by applying Girsanov's transformation of measure, and this way is most useful because it shows, moreover, that the total variation of the difference of the two these measures is strictly less than one. This will be shown in detail on the lecture.

Once the total variation between the two measures becomes strictly less than one after each moment γ_1, γ_2, etc., then due to the Markov property after γ_{n+1} this difference in total variation is less that q^n with some $q < 1$. This shows the statement of the Theorem with an appropriate constant λ. $\square$

2.4 Coupling method: the calculus

1. Here we present a rigorous calculus justifying the intuitive ideas of the previous subsection; this may be called a formal proof of the Theorem 6. Under the Lipschitz assumption on the drift, there exists a transition density $p(x,y) = p_1(x,y)$ which is a (bounded) Borel function of all its arguments, see, e.g., [2]. In this respect, mention a recent result [17] on one-dimensional diffusion with a unit diffusion coefficient and any bounded measurable drift which states the existence of a positive bounded transition density satisfying qualitative bounds close to Gaussian ones. Recall that in fact the existence of the density is not necessary for the local Doeblin type condition.

2. For any $B \in \mathcal{B}(R^d)$ and for any $n = 1, 2, \ldots$ we are going to compare the two values,

$$
\begin{aligned}
&\underbrace{\int \ldots \int}_{n} p(x,x_1)p(x_1,x_2)\ldots p(x_{n-1},x_n)1(x_n \in B)\prod_{i=1}^{n} dx_i \\
&- \underbrace{\int \ldots \int}_{n} p(x',x'_1)p(x'_1,x'_2)\ldots p(x'_{n-1},x'_n)1(x'_n \in B)\prod_{i=1}^{n} dx'_i.
\end{aligned}
\tag{17}
$$

In fact, impotant is that x and x' are different initial data, all other x's are just integration variables, so that we can equally use idential notations for x'_i and x_i. We will exploit this remark shortly.

It was claimed, – see (11) above, – that for any R there exists $q > 0$ such that

$$
\inf_{x,x' \in B_R} \int \min(p(x,x_1), p(x',x_1))\, dx_1 > 0. \tag{18}
$$

Let us denote

$$
1 - q(x,x') := \int \min(p(x,x_1), p(x',x_1))\, dx_1;
$$

note that

$$
\sup_{x,x' \in B_R} q(x,x') \le q < 1.
$$

(currently R is fixed and, hence, we often drop it in the calculus).

3. For any x, x' such that the denominator below is positive, denote

$$
\bar{p}^{(x,x')}(x,x_1) = \frac{p(x,x_1) - \min(p(x,x_1), p(x',x_1))}{q(x,x')}. \tag{19}
$$

This is a probability density as a function of x_1. Note that if $x, x' \in B$, then the probability measures corresponding to the two densities

$\bar{p}^{(x,x')}(x, x_1)$ and $\bar{p}^{(x',x)}(x', x_1)$ are singular one with respect to each other. Also note that for any integrable function g (even bounded enough for our aims),

$$
\begin{aligned}
& \int\int p(x,x_1)p(x',x_1')(g(x_1)-g(x_1'))\,dx_1\,dx_1' \\
= \;& q(x,x')\int\int \bar{p}^{(x,x')}(x,x_1)\bar{p}^{(x',x)}(x',x_1')(g(x_1)-g(x_1'))\,dx_1\,dx_1',
\end{aligned}
\tag{20}
$$

since while integrating the initial itegral one can subtract the common part corresponding to the integration with the common part of the two densities, $\min(p(x, x_1), p(x', x_1))$.

In the case if $q(x, x') = 0$ (which means that $p(x, \cdot) \equiv p(x', \cdot)$), it is convenient to define

$$
\bar{p}^{(x,x')}(x, x_1) = p(x, x_1). \tag{21}
$$

4. The function

$$
\bar{p}^{(x,x')}(x, x_1)\bar{p}^{(x',x)}(x', x_1') \tag{22}
$$

by construction is Borel in all its variables, and may be considered as a new transition density for the couple (x, x') into (x_1, x_1'). In the other words, we have constructed a new Markov process which we denote by $(\bar{X}_n, \bar{X}_n')$ with this transition density given by (22). We do not know (and do not care) if each component of this couple is itself a Markov process.

5. Let us rewrite the difference (17) in the form,

$$
\begin{aligned}
& \underbrace{\int\ldots\int}_{2n} p(x,x_1)p(x_1,x_2)\ldots p(x_{n-1},x_n)1(x_n\in B)\prod_{i=1}^n dx_i \\
& \quad\times\;\; p(x',x_1')p(x_1',x_2')\ldots p(x_{n-1}',x_n')\prod_{i=1}^n dx_i' \\
& -\underbrace{\int\ldots\int}_{2n} p(x,x_1)p(x_1,x_2)\ldots p(x_{n-1},x_n)\prod_{i=1}^n dx_i \\
& \quad\times\;\; p(x',x_1')p(x_1',x_2')\ldots p(x_{n-1}',x_n')1(x_n'\in B)\prod_{i=1}^n dx_i' \\
= \;& \underbrace{\int\ldots\int}_{2n} p(x,x_1)p(x_1,x_2)\ldots p(x_{n-1},x_n)\prod_{i=1}^n dx_i\,\prod_{i=1}^n dx_i' \\
& \quad\times\;\; p(x',x_1')p(x_1',x_2')\ldots p(x_{n-1}',x_n')\,\left(1(x_n\in B)-1(x_n'\in B)\right).
\end{aligned}
\tag{23}
$$

6. Next, for each couple of variables (x_i, x'_i), $1 \le i \le n-1$, use the "unit decomposition",

$$1 = 1(x_i, x'_i \in B_R) + 1(x_i \notin B_R, \text{ or } x'_i \notin B_R).$$

We use notation $\alpha = (\alpha_1, \ldots, \alpha_{n-1})$ for $(n-1)$-tuples, where $\alpha_i = 1(x_i, x'_j \in B_R)$. Hence, the unit decomposition above can be represented in the form, $1 = 1(\alpha_i = 1) + 1(\alpha_i = 0)$. We will also use a "complete unit decomposition" for our integral in (24), namely,

$$1 = \sum_a 1(\alpha = a),$$

where the multi-index $a = (a_0, \ldots a_n)$ with $a_i = 1$ or 0 runs over all its possible values, that is, all possible n-tuples of ones and zeros. Hence, represent our integral in (24) as

$$\underbrace{\int \ldots \int}_{2n} p(x, x_1)p(x_1, x_2)\ldots p(x_{n-1}, x_n) \prod_{i=1}^n dx_i \prod_{i=1}^n dx'_i \sum_a 1(\alpha = a)$$
$$\times p(x', x'_1)p(x'_1, x'_2)\ldots p(x'_{n-1}, x'_n)\,(1(x_n \in B) - 1(x'_n \in B)).(24)$$

7. Next, let $\#a := \sum_{i=0}^{n-1} 1(a_i = 1)$, and consider the parts of the latter integral for which $\#a \ge [cn]$ and for which $\#a < [cn]$; we will use the identity, $\{\#a < [cn]\} = \{\gamma_{[cn]} \ge n\}$. Note that

$$\underbrace{\int \ldots \int}_{2n} p(x, x_1)p(x_1, x_2)\ldots p(x_{n-1}, x_n)$$
$$\times \prod_{i=1}^n dx_i \prod_{i=1}^n dx'_i \left(\sum_{a:\,\#a<[cn]} 1(\alpha = a)\right)$$
$$\times p(x', x'_1)p(x'_1, x'_2)\ldots p(x'_{n-1}, x'_n)\,(1(x_n \in B) - 1(x'_n \in B))$$

$$= \quad P_{x,x'}(\gamma_{[cn]} > n;\, X_n \in B) - P_{x,x'}(\gamma_{[cn]} \ge n;\, X'_n \in B).$$

We stress out that here both probabilities are written for the initial processes X_k and X'_k, with the only correction that $X'_0 = x'$ (at the end of the calculus, we can just integrate with respect to this variable x'). Both probabilities can be estimated from above by the exponential bound (16),

$$P_{x,x'}(\gamma_{[cn]} \ge n) \le C\exp(\epsilon(|x| + |x'|) - \lambda' n).$$

Therefore, the difference of the two probabilities above by absolute value does not exceed the same value $C\exp(\epsilon(|x| + |x'|) - \lambda' n)$.

8. Let us consider another part of the integral,

$$\underbrace{\int\dots\int}_{2n} p(x,x_1)p(x_1,x_2)\dots p(x_{n-1},x_n)$$
$$\times \prod_{i=1}^{n} dx_i \prod_{i=1}^{n} dx_i' \left(\sum_{a:\,\#a\geq[cn]} 1(\alpha=a)\right)$$
$$\times\ p(x',x_1')p(x_1',x_2')\dots p(x_{n-1}',x_n')\ (1(x_n\in B)-1(x_n'\in B))$$

Take any $[cn]\leq k\leq n$, and consider the value

$$\underbrace{\int\dots\int}_{2n} p(x,x_1)p(x_1,x_2)\dots p(x_{n-1},x_n)\prod_{i=1}^{n} dx_i \prod_{i=1}^{n} dx_i' \sum_{a} 1(\alpha=a)$$
$$\times\ p(x',x_1')p(x_1',x_2')\dots p(x_{n-1}',x_n')\ (1(x_n\in B)-1(x_n'\in B))\,1\left(\#\alpha=k\right).$$

By our convention, the equalities $\alpha=a$ and $\#a=k$ mean that in this expression we integrate k times over the set $B_R\times B_R$, while $(n-k)$ times over the set $(R^d\times R^d)\setminus(B_R\times B_R)$. Consider the *last* time strictly before n when the integration is over $B_R\times B_R$, e.g., suppose this is the integration over $dx_{n-1}dx_{n-1}'$. Then, according to the estimate (18) and definition (19),

$$\begin{aligned}
& \int p(x_{n-1},x_n)p(x_{n-1}',x_n')dx_n\,dx_n' \\
& \times\ (1(x_n\in B)-1(x_n'\in B))\,1\left(x_{n-1}'\in B_R, x_{n-1}\in B_R\right) \\
= \ & q(x_{n-1},x_{n-1}')\int \bar p^{x_{n-1},x_{n-1}'}(x_{n-1},x_n)\bar p^{x_{n-1}',x_{n-1}}(x_{n-1}',x_n')dx_n\,dx_n' \\
& \times\ (1(x_n\in B)-1(x_n'\in B))\,1\left(x_{n-1}'\in B_R, x_{n-1}\in B_R\right).
\end{aligned}$$

We have used here the identities, for any given x, x'

$$\begin{aligned}
q(x,x') &= q(x',x), \\
p(x,x_1) &= q(x,x')\bar p^{x,x'}(x,x_1)+\min(p(x,x_1),p(x',x_1)), \\
p(x',x_1) &= q(x',x)\bar p^{x',x}(x',x_1)+\min(p(x',x_1),p(x,x_1)),
\end{aligned}$$

and, hence,

$$p(x,x_1)-p(x',x_1)=q(x,x')(\bar p^{x,x'}(x,x_1)-\bar p^{x',x}(x',x_1)).$$

9. Denote
$$\hat q(x_i,x'_i) = \left\{ \begin{array}{ll} q(x_i,x'_i), & x_i,x'_i \in B_R, \\ 1, & \text{otherwise.} \end{array}\right.$$
Using the same trick as before and induction, the integral in question can be rewritten in the form
$$\underbrace{\int\ldots\int}_{2n} p(x,x_1)\ldots p(x_{n-1},x_n)\prod_{i=1}^n dx_i \prod_{i=1}^n dx'_i$$
$$\times\, p(x',x'_1)\ldots p(x'_{n-1},x'_n)\,\left(1(x_n\in B)-1(x'_n\in B)\right)\sum_{a:\,\#a=k} 1\,(\alpha=a)$$
$$= \underbrace{\int\ldots\int}_{2n} \bar p^{x_1,x'_1}(x,x_1)\ldots \bar p^{x_{n-1},x'_{n-1}}(x_{n-1},x_n)\prod_{i=1}^n dx_i$$
$$\times\, \bar p^{x',x}(x',x'_1)\ldots \bar p^{x'_{n-1},x_{n-1}}(x'_{n-1},x'_n)\prod_{i=1}^n dx'_i \tag{25}$$
$$\times\, \hat q(x,x')\ldots \hat q(x_{n-1},x'_{n-1})\,\left(1(x_n\in B)-1(x'_n\in B)\right)\sum_{a:\,\#a=k} 1\,(\alpha=a)\,.$$

10. In the last integral, we compute, in fact, the difference of two integrals over the transition densities ($\bar p$), certain indicator functions ($1(x_n\in B)$ or $1(x'_n\in B)$), and a product of n functions ($\hat q$) whose absolute values are bounded by 1, and (at least) k of these absolute values are less than or equal to the value q which, in turn is strictly less than 1. In terms of the process $(\bar X_k,\bar X'_k)$, this expression can be presented as follows,
$$E_{x,x'}\prod_{i=0}^{n-1}\hat q(\bar X_i,\bar X'_i)\,1\left(\left(\sum_{i=0}^{n-1}1(\bar X_i,\bar X'_i\in B_R)\right)=k\right)$$
$$\times\left(1(\bar X_n\in B_R)-1(\bar X'_n\in B_R)\right).$$
Hence, the whole expression (25) by absolute value does not exceed q^k. So, the whole integral with $[cn]\le k\le n$ is less than or equal to $\sum_{k=[cn]}^n q^k \le Cq^{[cn]}\le Cq^{cn-1}$.

11. Thence, we obtain the final estimate (12) with $\lambda=\min(-c\ln q,\lambda')$. The Theorem is proved. □

There is no Exercises for this subsection.

2.5 Final comments on convergence for X_t

1. The contents of this section is an essentially rearranged and rewritten paper [19]; the calculus for the coupling method has been written for

this course. In this paper a nondegenerate variable diffusion coefficient case is considered, the density is not used, and the key tool which leads to establishing a local Doeblin type condition is Harnack's inequality. The same inequality also helps to investigate other cases including gradient type equations.

2. The weakest known to the lecturer assumption sufficient for existence of a (unique) stationary measure for a SDE in question reads,

$$\limsup_{|x|\to\infty} \langle b(x), x\rangle < -1/2,$$

see, e.g., [21] and independently [13]; it would not be surprising to find a similar result in other papers, too.

3. Currently, the weakest sufficient assumption which allows to get any qualitative bound for convergence rate in total variation metric reads,

$$\limsup_{|x|\to\infty} \langle b(x), x\rangle < -3/2,$$

see [21] or [20]. The rate of convergence is then polynomial in time, and similar bounds have been established for so called beta-mixing.

4. It looks like an open question what could be said about qualitative ergodic properties in the region

$$-1/2 > \lim_{|x|\to\infty} \langle b(x), x\rangle \geq -3/2.$$

5. Under the assumption

$$\limsup_{|x|\to\infty} \langle b(x), x/|x|^a\rangle < 0, \quad 0 < a < 1, \tag{26}$$

intermediate convergence and mixing bounds of the kind $C(x)\exp(-ct^b)$ with certain $0 < b < 1$ are established, see [7].

6. Lower bounds have been obtained showing "epsilon-optimal" nature of upper bounds under the assumption (26) with $0 \leq a \leq 1$ (including the endpoints).

7. There are partial results about convergence and mixing for gradient type SDEs, cf. [22]. The methods for establishing recurrence properties here are different.

3 Convergence to equilibrium for X_t^h

Now, we turn to the Euler scheme (2). Although this section is intended to be the main, along with the next one, in the notes we only give some comments, because of an available reference [9] at http://www.newton.cam.ac.uk/preprints/NI03065.pdf. In the class this problem will be considered in more detail.

The first question is whether there is a (unique) stationary measure for the process X_t^h, and what is the rate of convergence to this measure, that is, if a uniform convergence rate bound independent on h exists. We notice that once we know the answer to this question, it is not hard also to show that the exact and approximate stationary measures μ_∞ and μ_∞^h are close, at least, in weak topology as h is small, e.g., due to the Theorem 1. To get to a similar conclusion in total variation metric, one would need Local Limit Theorem type results for the densities $p_t^h(x, x')$ and $p_t(x, x')$, see, e.g., [10] or [1].

Recall that the bounds for convergence in Corollary 1 and in Theorem 6 were based on two auxiliary bounds: Doeblin type condition (11) and recurrence bounds (10). It is natural to suggest, and the calculus in the previous subsection fully confirms this view, that had we established uniform bounds of both kinds which had not depended on h, say, for $h \leq h_0$ with some h_0, the result would be indeed uniform bounds for convergence rate for μ_t^h similar to Corollary 1.

Uniform – at least for small h – recurrence bounds similar to (10) can be established, indeed, under the assumptions like (E1), or (E2), or likewise. The idea is very similar to the continuous time case, however, in discrete time we cannot use such a nice tool as Itô's formula, and instead a careful analysis based on Taylor's expansion is used. Evidently, this is more involved, and this is why we do not go into details here at all. Some relevant calculus can be found in [9] (where a more general non-gaussian noise ξ_n is considered).

The second auxiliary estimate similar to (11) for Gaussian ξ's can be deduced from Harnack inequality which holds true for Itô processes, not necessarily Markov diffusions; the Euler scheme provides exactly such processes, X_t^h. Since this approach is not elementary, we present here a weaker result of the same kind which is less involved, at least, in principle, although it requires additional assumptions on coefficient b. On the other hand, the approach is based on Malliavin's calculus for approximation processes, and can be applied to much more general ξ's than Gaussian. Notice that in our case of unit diffusion matrix, one more approach is possible, different from both "Harnack's" and "Malliavin's", and it does not require any additional smoothness. Nevertheless, this short course is not a right place where it could be presented.

4 On processes with parameters

Let the coefficient b (and also σ if it were not constant) depend on a parameter, $\alpha \in R^d$ (the fact that the dimension d is the same has no importance here, although we wish this dimension to be finite). The problem under consideration here is whether the invariant density depends smoothly on this parameter. Certainly, we assume that the coefficient is itself smooth enough as a function of α. Stress out that we are looking for one or several derivatives of the limiting density, not just continuity; the latter is not difficult indeed.

The answer about smoothness is easy only in a few special cases: (1) for a compact semigroup with a spectral gap for the generator (then one can use the results of perturbation theory); (2) for an SDE with a gradient type drift (which allows an explicit representation of the invariant density): $dX_t = dW_t - \nabla U(X_t)\, dt$ with appropriate function U.

Exercise 17 *Write down an "explicit" representation for the latter case and explain how to check smoothness with respect to a parameter.*

4.1 Assumptions

We will establish the property $p(x,\cdot) \in C^1(\mathbb{R}^d)$, although in principle the same approach allows to get conditions for several derivatives, too. As earlier, we are using the assumptions which lead to the most simple calculus, not trying to establish the best possible estimates. Under weaker in some respect condition, close results can be found in [27, 16]; however, certain smoothness assumptions in this text are better than those in [16] for SDEs, this concerns assumptions with respect to state variables (x, x') (on the third side, recall that we consider the case of constant diffusion here, even though most of the result could be extended for more general case). We formulate the assumptions and give some arguments why such assumptions are reasonable for those who wished to have all conditions in terms of the coefficients of our SDE (the lecturer does share this point of view). In all assumptions below, there exist the values ϵ, δ and C.

A1 For any $\epsilon \le \epsilon_0$, there exist $C, \delta > 0$ such that

$$\mathrm{var}\,(\mu_t^x - \mu_\infty) \le C \exp(\epsilon|x| - \delta t);$$

(See the Corollary 1.)

A2 For any $\epsilon \le \epsilon_0$, there exists $C > 0$ such that

$$\exp(-\epsilon|x|) \sup_{t<\infty} \int \exp(\epsilon|x'|)\, \mu_t^x(dx') + \int \exp(\epsilon|x'|)\, \mu_\infty(dx') \le C;$$

(This condition has been justified in the Lemma 1 under the assumption (E2). One possible reasonable generalisation, – not speaking of polynomial convergence setting, – is to use different epsilons in the left hand side and right hand side of this inequality.)

A3$_i$

$$p(x, x', \cdot) \in C_b^i \text{ uniformly in } (x, x');$$

(This can be justified using the PDE technique for coefficients from C^1 in α uniformly with respect to the main variable. 'Uniformly' here means that the norms $|\cdot|_{C^i}$ are uniformly bounded.)

A4$_i$ For any $\epsilon \le \epsilon_0$, there exists $C > 0$ such that

$$p(x, x', \alpha) + \left|\nabla^i_\alpha p(x, x', \alpha)\right| \le C_i \exp(-\epsilon|x - x'|).$$

(We could, of course, use Gaussian type bounds here $C \exp(-\epsilon|x-x'|^2)$, however, this does not really provide any further simplification of the calculus. The bound for p itself is standard. The bound for $\nabla_\alpha p$ can be justified using the PDE technique.)

4.2 Smoothness of invariant density

Theorem 7 *Under assumptions $A1, A2, A3_i, A4_i$, the invariant density $p(x, \cdot) \in C^i$.*

Proof for the case $i = 1$ only.

1. From the standing assumptions the following assertions follow: for any $\epsilon \le \epsilon_0$, there exist $\delta > 0$ and $C > 0$ such that

$$\begin{aligned}
\sup_n p_n(x, x', \alpha) &\le C < \infty; && (27)\\
|p_n(x, x', \alpha) - p_\infty(x', \alpha)| &\le C \exp(\epsilon|x| - \delta n); && (28)\\
p_\infty(x', \alpha) &\le C \exp(-\epsilon|x'|/2); && (29)\\
\sup_n p_n(x, x', \alpha) &\le C \exp(\epsilon|x| - \epsilon|x'|/2); && (30)\\
|p_n(x, x', \alpha) - p_\infty(x', \alpha)| &\le C \exp\left(\epsilon|x| - \frac{\epsilon|x'|}{4} - \frac{\delta n}{2}\right). && (31)
\end{aligned}$$

We stress out that some of the inequalities will be used with different values of ϵ.

Exercise 18 *Show all. Hint: use the Chapman - Kolmogorov equation.*

2. Solutions.

   (a) (27) is indeed a straightforward consequence of the Chapman - Kolmogorov and the assumption $p \le C$.

   (b) (28): we have,

$$\begin{aligned}
&|p_n(x, x') - p_\infty(x')|\\
= &\left|\int (p_{n-1}(x, \tilde{x}) - p_\infty(\tilde{x}))\, p(\tilde{x}, x') d\tilde{x}\right|
\end{aligned}$$

$$\leq \quad C \operatorname{var}(\mu_{n-1}^x - \mu_\infty)$$

$$\leq \quad C \exp(\epsilon|x| - \delta(n-1)),$$

due to the A1 assumption.

(c) (29): we have, due to the A2 assumption,

$$p_\infty(x') = \int p_\infty(\tilde{x}) p(\tilde{x}, x')\, d\tilde{x}$$

$$\leq \quad C \int p_\infty(\tilde{x}) \exp(-\epsilon|\tilde{x} - x'|)\, d\tilde{x}$$

$$\leq \quad C \int \exp(\epsilon|\tilde{x}|) p_\infty(\tilde{x}) \exp(-\epsilon|\tilde{x} - x'| - \epsilon|\tilde{x}|)\, d\tilde{x}$$

$$= \quad C \int \exp(\epsilon|\tilde{x}|) p_\infty(\tilde{x}) \exp(-\epsilon|\tilde{x} - x'| - \epsilon|\tilde{x}|) 1(|\tilde{x}| \geq |x'|/2)\, d\tilde{x}$$

$$+ C \int \exp(\epsilon|\tilde{x}|) p_\infty(\tilde{x}) \exp(-\epsilon|\tilde{x} - x'| - \epsilon|\tilde{x}|) 1(|\tilde{x}| < |x'|/2)\, d\tilde{x}$$

$$\leq \quad 2C \int \exp(\epsilon|\tilde{x}|) p_\infty(\tilde{x}) \exp(-\epsilon|x'|/2)\, d\tilde{x}$$

$$\leq \quad 2C \exp(-\epsilon|x'|/2)\, d\tilde{x}.$$

(d) (30): using, in particular, already established (28-29) and the same hint as above, we get,

$$p_n(x, x') = \int p_{n-1}(x, \tilde{x}) p(\tilde{x}, x') d\tilde{x}$$

$$\leq \quad \int p_\infty(\tilde{x}) p(\tilde{x}, x') d\tilde{x} + |\int (p_{n-1}(x, \tilde{x}) - p_\infty(\tilde{x})) p(\tilde{x}, x') d\tilde{x}|$$

$$\leq \quad p_\infty(x') + C \int \exp(\epsilon|\tilde{x}|)(p_{n-1}(x, \tilde{x}) - p_\infty(\tilde{x}))$$

$$\times \exp(-\epsilon|\tilde{x} - x'| - \epsilon|\tilde{x}|)\left(1(|\tilde{x}| \leq |x'|/2) + 1(|\tilde{x}| < |x'|/2)\right) d\tilde{x}$$

$$\leq \quad C \exp(-\epsilon|x'|/2) + 2C\ \exp(\epsilon|x| - \epsilon|x'|/2).$$

(e) (31): follows from (28) and (30) due to the identity $a = a^{1/2} a^{1/2}$.

3. Denote

$$Lf(x) := E_x f(X_1) - f(x), \quad q_n(x, x', \alpha) = \partial_{\alpha_i} p_n(x, x', \alpha).$$

By virtue of the Chapman - Kolmogorov equation,

$$p_{n+1}(x,x') - p_n(x,x') = (Lp_n(\cdot,x'))\ (x).$$

Exercise 19 *Show.*

4. Hence, by taking derivative, we get,

$$q_{n+1}(x,x') - q_n(x,x') = (Lq_n(\cdot,x'))\ (x) + f_n^1(x,x'), \quad n \geq 0, \tag{32}$$

where

$$q_0(x,x') := 0, \ \text{ and } \ f_n^1(x,x') := \int (\partial_{\alpha_i} p(x,\tilde{x}))\, p_n(\tilde{x},x')\, d\tilde{x}.$$

Notice that f^1 is continuous in α.

5. We will need certain properties of the function f^1:

$$\int f_n^1(x,x')\, dx' = 0; \tag{33}$$

$$\left| f_n^1(x,x') \right| \leq C \exp(\epsilon|x| - \epsilon'|x'| - \delta' n), \tag{34}$$

with $\epsilon' = \epsilon/8$, $\delta' = \delta/4$.

Exercise 20 *Show both.*

6. Solutions.

   (a) (33):

$$\int f_n^1(x,x')\, dx' = \int\int (\partial_{\alpha_i} p(x,\tilde{x}))\, p_n(\tilde{x},x')\, d\tilde{x}\, dx'$$
$$= \int (\partial_{\alpha_i} p(x,\tilde{x})) \left(\int p_n(\tilde{x},x')\, dx' \right) d\tilde{x}$$
$$= \int (\partial_{\alpha_i} p(x,\tilde{x}))\, d\tilde{x} = \partial_{\alpha_i} \int p(x,\tilde{x})\, d\tilde{x} = 0,$$

   because the integral $\int |\partial_{\alpha_i} p(x,\tilde{x})|\, d\tilde{x}$ converges uniformly in α, due to the assumption A4.

   (b) (34): we have,

$$\left| f_n^1(x,x') \right| \leq \left| \int (\partial_{\alpha_i} p(x,\tilde{x}))\, p_\infty(x')\, d\tilde{x} \right|$$
$$+ \left| \int (\partial_{\alpha_i} p(x,\tilde{x}))\, (p_n(\tilde{x},x') - p_\infty(x'))\, d\tilde{x} \right|.$$

Note that

$$\int (\partial_{\alpha_i} p(x,\tilde{x}))\, p_\infty(x')\, d\tilde{x} = p_\infty(x')\partial_{\alpha_i} \int (p(x,\tilde{x}))\, d\tilde{x} = 0.$$

Hence, due to A4 and (31) which we use with $\epsilon/2$ instead of ϵ (this only changes the constants C, δ),

$$\begin{aligned}
\left|f_n^1(x,x')\right| \quad &\le \quad \left|\int (\partial_{\alpha_i} p(x,\tilde{x}))\,(p_n(\tilde{x},x') - p_\infty(x'))\, d\tilde{x}\right| \\
\le \quad & C\int \exp(-\epsilon|x-\tilde{x}|)\exp(\epsilon|\tilde{x}|/2 - \epsilon|x'|/8 - \delta n/4)\, d\tilde{x} \\
= \quad & C\exp(-\epsilon|x'|/8 - \delta n/4)\int \exp(-\epsilon|x-\tilde{x}|)\exp(\epsilon|\tilde{x}|/2)\, d\tilde{x} \\
\le \quad & C\exp(-\epsilon|x'|/8 - \delta n/4)\int \exp(\epsilon|x| - \epsilon|\tilde{x}|/2)\, d\tilde{x} \\
= \quad & C\exp(\epsilon|x| - \epsilon|x'|/8 - \delta n/4).
\end{aligned}$$

We have used the elementary inequality $|x-\tilde{x}| \ge |\tilde{x}| - |x|$.

7. The equation (32) has a solution with in some sense explicit representation,

$$q_n(x,x') = \sum_{s=0}^{n-1} \int f_s^1(x'',x')p_{n-1-s}(x,x'')\, dx'' \tag{35}$$

It is easy to see that in terms of the Markov process X_n with transition density $p(x,x')$, this formula may be rewritten as

$$q_n(x,x') = E_x \sum_{s=0}^{n-1} f_s^1(X_{n-1-s},x'). \tag{36}$$

Exercise 21 *Show the formula (35) or equivalently (36).*

8. Solution.

   (a) For $n = 0$ we have $q_0(x,x') = 0$, the assertion holds true.

   (b) Suppose we have checked the formula for n. Then let us verify it for $n+1$. We have,

$$\begin{aligned}
q_{n+1}(x,x') \quad &= \quad E_x \sum_{s=0}^{n-1} f^1(X_{n-s},x') + f_n^1(x,x') \\
&= \quad E_x \sum_{s=0}^{n} f^1(X_{n-s},x').
\end{aligned}$$

   Hence, by induction the representation (36) or equivalently (35) is proved.

9. We suspect that the function q_n has a limit as $n \to \infty$. So it is natural to define

$$q_\infty(x') := \sum_{s=0}^{\infty} \int f_s^1(x'', x') p_\infty(x'')\, dx''. \tag{37}$$

We have extended the sum to infinity, and replaced the indices $n - s$ in p_{n-s} also by ∞. Stress out that this suggestion is just a guess, so at this stage we do not pretend that this is rigorous. However, the next step is to verify whether this suggestion is right or not.

10. Let us show that the series for q_∞ does converge. Firstly, each term is finite because of the bounds (34) and (30),

$$\int \left| f_s^1(x'', x') p_\infty(x'') \right| \, dx''$$

$$\leq C \int \exp(\epsilon |x''|/3 - \epsilon' |x'| - \delta' s) \exp(-\epsilon |x''|)\, dx''$$

$$\leq C \exp(-\epsilon' |x'| - \delta' s) < \infty. \tag{38}$$

Secondly, the factor $\exp(-\delta' s)$ here makes the series converge indeed.

11. Once the sum for q_∞ converges, it defines a function continuous in α. As a next step, let us show that

$$q_n(x, x') \to q_\infty(x'). \tag{39}$$

Indeed, we can represent the difference $q_n - q_\infty$ as a sum of three terms,

$$\begin{aligned} q_n(x, x') - q_\infty(x') \;&=\; \sum_{s=0}^{[n/2]} \int f_s^1(x'', x') (p_{n-s}(x, x'') - p_\infty(x''))\, dx'' \\ &+ \sum_{s=[n/2]+1}^{n} \int f_s^1(x'', x') p_{n-s}(x, x'')\, dx'' \\ &- \sum_{s=[n/2]+1}^{\infty} \int f_s^1(x'', x') p_\infty(x'')\, dx''. \end{aligned}$$

Due to the inequality (38), the second and third term do not exceed the series $\sum_{s=[n/2]+1}^{\infty} C(x, x') \exp(-\delta' s) \leq C(x, x') \exp(-\delta' n/2)$ (with a new constant $C(x, x')$). The first term due to the bound (31), does not exceed a similar sum $\sum_{s=0}^{[n/2]} C(x, x') \exp(-\delta(n - s)/2) \leq C(x, x') \exp(-\delta n/2)$. So, (39) holds true. Moreover, in fact, we get an estimate for this convergence,

$$|q_n(x, x') - q_\infty(x')| \leq C \exp(\epsilon |x| - \epsilon' |x'| - \delta'' n). \tag{40}$$

12. Let us, finally, show that $q_\infty(x') = \partial_{\alpha_i} p_\infty(x')$.

Exercise 22 *Show. Hint: pass to the limit in the equality,*

$$p_n(x,x',\alpha) - p_n(x,x',\alpha') = \int_{\alpha'}^{\alpha} q_n(x,x',a)\,da.$$

using convergence bounds (28) and (40), to get

$$p_\infty(x,x',\alpha) - p_\infty(x,x',\alpha') = \int_{\alpha'}^{\alpha} q_\infty(x,x',a)\,da.$$

Using continuity with respect to α, explain why this means that the function $\partial_{\alpha_i} p_\infty(x,x',\cdot)$ (here $\cdot$ stands for α) is indeed a partial derivative of $p_\infty(x,x',\cdot)$. Finally, using continuity of all partial derivatives of the first order, explain why this means $p_\infty(x,x',\cdot) \in C^1$.
The Theorem 7 in the case $i = 1$ is proved. $\square$

5 Further problems

The 'mini'course unfortunately prompts that some essential related topics may remain not well represented. Here are some of them.

- Weak approximations, with indeed nongaussian noise, see [9]. (Not speaking of singular distributions, cf. [12].) Partially, this direction and the preprint were discussed on the lectures.

- Semigroup approximations, that is, in the operator topologies. This relates also to large deviations, see [23, 24, 25]. This topic was discussed briefly on the last day of the lectures.

- Equilibrium measures and approximations for nonlinear diffusions. For the McKean-Vlasov equation and 'mean field' approach, surprisingly little is known even about the existence of invariant measures, cf. [26]. This topic was very briefly announced on the lectures.

- It is interesting that certain Euler's approximations may converge to the solution of the SDE even with only Borel bounded drift, see [4], although there is no bounds available for the rate of this convergence. It turns out that a stochastic version of the following deterministic consideration works here: if any converging subsequence $a_{n'}$ ăfrom a bounded sequence a_n converges to the unique limit a, then $\lim_{n\to\infty} a_n = a$. The strong (=pathwise) uniqueness used here is provided by [18].

- "Wiener chaos" approach can be also used for approximations of solutions to SDEs, cf. [28].

References

[1] Bally, V., Talay, D. The law of the Euler scheme for stochastic differential equations. II: Convergence rate of the density. Monte Carlo Methods Appl. 2, No.2, 93-128 (1996).

[2] Friedman, A. Partial differential equations of parabolic type, Englewood Cliffs, N.J.: Prentice-Hall, Inc. (1964).

[3] Gulinsky, O. V., Veretennikov, A. Yu. Large deviations for discrete-time processes with averaging. Utrecht: VSP (1993).

[4] Gyöngy, I.; Krylov, N. Existence of strong solutions for Itô's stochastic equations via approximations. Probab. Theory Related Fields 105, no. 2, 143–158 (1996).

[5] Ibragimov, I. A., Linnik, Yu. V. Independent and stationary sequences of random variables. Wolters-Noordhoff, Groningen (1971).

[6] Kloeden, P. E., Platen, E. Numerical solution of stochastic differential equations. Springer-Verlag, Berlin (1992).

[7] Klokov, S. A.; Veretennikov, A. Yu. Subexponential mixing rate for a class of Markov diffusions. Journal of Mathematical Sciences, 123 (1), 3816-3823 (September 2004).

[8] Klokov, S. A.; Veretennikov, A. Yu. Sub-exponential mixing rate for Markov processes. Mathematical Communications 9, 9-26 (2004).

[9] Klokov, S. A.; Veretennikov, A. Yu. Mixing and convergence rates for a family of Markov processes appriximating SDEs. Preprint at http://www.newton.cam.ac.uk/preprints/NI03065.pdf (2003).

[10] Konakov, V., Mammen, E. Local limit theorems for transition densities of Markov chains converging to diffusions. Probab. Theory Relat. Fields 117, No.4, 551-587 (2000).

[11] Krylov, N. V. Introduction to the theory of random processes. AMS, Providence, RI (2002).

[12] Kushner, H. J. Probability methods for approximations in stochastic control and for elliptic equations. Academic Press (Harcourt Brace Jovanovich, Publishers), New York – London (1977).

[13] Malyshkin, M. N. Subexponential estimates of the rate of convergence to the invariant measure for stochastic differential equations. Theory Probab. Appl. 45, No.3, 466-479 (2000); translation from Teor. Veroyatn. Primen. 45, No.3, 489-504 (2000).

[14] Milstein, G. N. Numerical integration of stochastic differential equations. Kluwer Academic Publishers Group, Dordrecht (1995).

[15] Milstein, G. N., Tretyakov, M. V. Stochastic numerics for mathematical physics. Scientific Computation. Springer-Verlag, Berlin (2004).

[16] Pardoux, E.; Veretennikov, A. Yu. On Poisson equation and diffusion approximation. II. Ann. Probab. 31, No.3, 1166-1192 (2003).

[17] Qian, Z., Zheng, W. Sharp bounds for transition probability densities of a class of diffusions. C. R. Math. Acad. Sci. Paris 335, no. 11, 953–957 (2002).

[18] Veretennikov, A. Yu. On strong solutions and explicit formulas for solutions of stochastic integral equations. Mat. Sb. (N.S.) 111(153), no. 3, 434–452, 480 (1980); translation in Math. USSR, Sb. 39, 387-403 (1981).

[19] Veretennikov, A. Yu. Bounds for the mixing rate in the theory of stochastic equations. (Russian, English) Theory Probab. Appl. 32, no. 2, 273–281 (1987); translation from Teor. Veroyatn. Primen. 32, no. 2, 299–308 (1987).

[20] Veretennikov, A. Yu. On polynomial mixing bounds for stochastic differential equations. Stochastic Processes Appl. 70, No.1, 115-127 (1997).

[21] Veretennikov, A. Yu. On polynomial mixing and the rate of convergence for stochastic differential and difference equations. Teor. Veroyatnost. i Primenen. 44, 1999, no. 2, 312–327 (1999); English translation: Theory Probab. Appl. 44, no. 2, 361–374 (1999).

[22] Veretennikov, A. Yu. On polynomial mixing for SDEs with a gradient-type drift. Theory Probab. Appl. 45, No.1, 160-164 (2000); translation from Teor. Veroyatn. Primen. 45, No.1, 163-166 (2000).

[23] Veretennikov, A. Yu. On large deviations for approximations of SDEs. Probab. Theory Relat. Fields 125, No.1, 135-152 (2003).

[24] Veretennikov, A. Yu. On large deviations for approximate solutions of stochastic differential equations on a torus. Teor. Veroyatnost. i Primenen. 47, no. 4, 772–780 (2002); translation in Theory Probab. Appl. 47, no. 4, 733–741 (2003).

[25] Veretennikov, A. Yu. On approximate large deviations for 1D diffusion. Georgian Math. J. 10, No.2, 381-399 (2003).

[26] Veretennikov, A. Yu. On ergodic measures for McKean-Vlasov stochastic equations, preprints at http://www.newton.cam.ac.uk/preprints/NI03066.pdf (2003) & http://www.newton.cam.ac.uk/preprints/NI03079.pdf (2003).

[27] Veretennikov, A. Yu., Pardoux, E. On the smoothness of an invariant measure of a Markov chain with respect to a parameter. Dokl. Math. 61, No.1, 31-33 (2000); translation from Dokl. Akad. Nauk, Ross. Akad. Nauk 370, No.2, 158-160 (2000).

[28] Zvonkin, A. K., Krylov, N. V. Strong solutions of stochastic differential equations. (Russian) Proceedings of the School and Seminar on the Theory of Random Processes (Druskininkai, 1974), Part II (Russian), pp. 9–88. Inst. Fiz. i Mat. Akad. Nauk Litovsk. SSR, Vilnius, 1975.

(continued from the back cover)

HELSINKI UNIVERSITY OF TECHNOLOGY INSTITUTE OF MATHEMATICS
RESEARCH REPORTS

The list of reports is continued inside. Electronical versions of the reports are available at *http://www.math.hut.fi/reports/* .

A472 Carlo Lovadina , Rolf Stenberg
A posteriori error analysis of the linked interpolation technique for plate bending problems
September 2004

A471 Nuutti Hyvönen
Diffusive tomography methods: Special boundary conditions and characterization of inclusions
April 2004

A470 Lasse Leskelä
Stabilization of an overloaded queueing network using measurement-based admission control
March 2004

A469 Jarmo Malinen
A remark on the Hille–Yoshida generator theorem
May 2004

A468 Jarmo Malinen , Olavi Nevanlinna , Zhijian Yuan
On a tauberian condition for bounded linear operators
May 2004

ISBN 951-22-7294-6
ISSN 0784-6460