# Note on local mixing techniques for stochastic differential equations

**Alexander Veretennikov**[a,b,c]

[a]*School of Mathematics, University of Leeds, LS2 9JT, Leeds, UK*

[b]*National research university Higher School of Economics, Pokrovsky blvd. 11, 109074, Moscow, Russian Federation*

[c]*Institute for Information Transmission Problems, Russian Academy of Sciences, B. Karetny 19, 127051, Moscow, Russian Federation*

ayv@iitp.ru (A.Yu. Veretennikov)

**Abstract**   The paper discusses several techniques which may be used for applying the coupling method to solutions of stochastic differential equations (SDEs). The coupling techniques traditionally consist of two components: one is local mixing, the other is recurrence. Often in the articles they do not split. Yet, they are quite different in their nature, and this paper separates them, concentrating only on the former.

Most of the techniques discussed here work in dimension $d \geq 1$, although, in $d = 1$ there is one additional option to use intersections of trajectories, which requires nothing but the strong Markov property and nondegeneracy of the diffusion coefficient. In dimensions $d > 1$ it is possible to use embedded Markov chains either by considering discrete times $n = 0, 1, \ldots$, or by arranging special stopping time sequences and to use the local Markov–Dobrushin (MD) condition, which is one of the most efficient versions of local mixing. Further applications may be based on one or another version of the MD condition; respectively, this paper is devoted to various methods of verifying one or another form of it.

**Keywords**   SDEs, local mixing
**2010 MSC**   60H10, 37A25

VTeX

## 1    Introduction

The stochastic differential equation (SDE) in $\mathbb{R}^d$

$$X_t = x + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s, \; t \geq 0, \tag{1}$$

is considered. Here $(W_t), t \geq 0$, is a $d_1$-dimensional Wiener process, $b$ and $\sigma$ are vector and matrix-valued Borel measurable functions of dimensions $d$ and $d \times d_1$, respectively. To avoid any ambiguity, both coefficients will be assumed bounded, although, it is not always necessary in what follows.

It is assumed that Equation (1) has a (weak or strong) solution, unique in distribution, which is a strong Markov process, see [8]. Naturally, under this condition the process $X_n$ – that is, our solution $X_t$ considered at integer times $t = 0, 1, \ldots$ – is a Markov chain (MC), which is, of course, also strong Markov. The advantage of the total variation distance (although, it is not unique in this respect) for Markov processes is that once a convergence rate is established, say,

$$\|\mu_n - \mu\|_{TV} \leq \psi(n) \to 0, \quad n \to \infty,$$

where $\mu_t$ is the marginal distribution of $X_t$, $\mu$ is any probability measure (ergodic limit for $(\mu_n)$), then this rate of convergence can be nearly verbatim transferred to the continuous time:

$$\|\mu_t - \mu\|_{TV} \leq \psi([t]) \to 0, \quad t \to \infty,$$

where $[t]$ is the integer part of $t$. Consider two independent versions of our Markov process $X_t$ (in continuous time), say, $X_t^1$ and $X_t^2$, with two different initial values $x^1$ and $x^2$, respectively (or distributions). Since we allow weak solutions, the processes $X_t^2$ and $X_t^2$, generally speaking, are defined on two different probability spaces with two different Wiener processes; without loss of generality, we may assume that these two probability spaces are *independent* on their direct product. Thus, we have two trajectories $X^2$ and $X^2$ on the same probability space (do not forget that Wiener processes are also different and independent of each other). Denote by $Q(x, dx')$ the transition kernel

$$Q(x, dx') = \mathbb{P}_x(X_1 \in dx').$$

**Definition 1.** A (global) Markov–Dobrushin (MD) condition holds for the Markov process $X_n$ iff

$$\inf_{x_1, x_2} \int \left( \frac{Q(x_1, dx')}{Q(x_2, dx')} \wedge 1 \right) Q(x_2, dx') > 0. \tag{2}$$

Here $Q(x_1, dx')$ is not necessarily assumed to be absolutely continuous with respect to $Q(x_2, dx')$; the integrand here is understood as the Radon–Nikodym derivative of the absolutely continuous component of $Q(x_1, dx')$ with respect to $Q(x_2, dx')$. In what follows a localised version of this condition will be stated and this localised version will be the object of our main interest in this paper. General approaches to coupling for SDEs require a (usually positive) recurrence and some form of local mixing. For the latter, beside intersections applicable only in the case $d = 1$, the following tools can be used.

- Lower and upper bounds of the transition density (requires Hölder's continuity, at least, for the diffusion coefficient, as well as an "elliptic" or "hypoelliptic" nondegeneracy); here the *petite sets condition*, popular in the discrete time theory of ergodic Markov chains, along with recurrence properties *may* be used.

- Lower and upper bounds of the transition density *only for the equation without the drift*, including degenerate and highly degenerate cases: here Girsanov's transformation is an efficient tool; petite sets conditions, generally speaking, do not work.

- In the absence of lower and upper bounds of the density, under the nondegeneracy condition and for general measurable coefficients of SDE, Harnack inequalities in parabolic or elliptic versions may be applied; a petite sets condition apparently may be proved; however, they are less efficient than the MD condition because the latter guarantees better estimates of the convergence rate.

Hence, the goal of this paper is to attract more attention to the (local) MD condition, as this condition deserves it in the humble opinion of the author. There is also a hope that this list of available techniques may help in the future in studying ergodic properties for more general classes of processes.

One more point is that except for the method based on lower and upper bounds of the density, in all other more involved situations the popular DD condition [4], or, in its local version, the *petite set* condition is difficult to apply to SDEs, unlike the MD one; and even if it could be applied, the MD condition requires weaker assumptions and provides better estimates of the convergence rate.

Note that for discrete time stochastic models, and in certain cases for continuous time, too, one more natural approach to coupling is to use regeneration. Unfortunately, for general SDEs this method is not available. So, we do not discuss it here, although, the multidimensional coupling constructions for processes with continuous distributions are sometimes called "a generalised regeneration".

Let us warn the reader that most of the results of this paper are known, perhaps, in a slightly different form; we just collect them here together. Yet, Section 2.6 is new, and the version of the result as it is stated in Section 2.5 is new, too. For simplicity we do not touch more general equations such as SDEs with jumps. However, in principle, more general Markov processes, in particular, SDEs with jumps, may also be tackled with the help of similar techniques.

It should be also highlighted that all methods discussed in what follows (but the elliptic Harnack inequality) can be applied with minor differences to nonhomogeneous SDEs, too, except that convergence would be for the distance in total variation between marginal distributions corresponding to any two initial measures, not to the invariant measure which does not exist in this situation.

The paper consists of two sections: this introduction and the main Section 2; in its turn, Section 2 is split into six subsections, most of them related to one of the coupling tools listed above. The majority of proofs are sketchy or dropped because the results are known; the only exceptions are the parts in Subsections 2.5 and 2.6; the latter about the elliptic Harnack inequality is new to the best of author's knowledge. The proof of Lemma 2 has been provided for the reader's convenience by the suggestion

of one of the referees. We repeat that the paper presents the set of various *tools* for coupling for (homogeneous) SDEs. Neither recurrence – the necessary second ingredient in studying convergence and mixing rates – nor coupling itself (except for the basic Lemma 1 added for the reader's convenience) are not the goals of this paper.

## 2   Main results

### 2.1   Case $d = 1$, MD condition and coupling using intersections

In the 1D case for local coupling we can use **intersections** of two independent solutions of the same SDE with different initial values. Assume that $X_t$ and $X'_t$ are two independent solutions of Equation (1) with different initial values $X_0 = x$ and $X'_0 = x'$ in the one-dimensional case. The basis for applying coupling via intersections is the following result.

**Proposition 1.** *If $b$, $\sigma$, and $\sigma^{-1}$ are bounded then*

$$\inf_{-1 \le x,x' \le 1} \mathbb{P}_{x,x'}(\exists \, s \in [0,1] : \ X_s = X'_s) > 0. \tag{3}$$

*The first meeting time $\tau := (t \ge 0 : X_t = X'_t)$ is a stopping time.*

**Proof.** It follows from the following two elementary steps.

1. Change the time for both SDEs making diffusion coefficients equal to one. There is no need to make it by *the same* random time change: generally speaking, the latter is not possible unless the diffusion coefficient is a constant. Since $\sigma$ and $\sigma^{-1}$ are bounded, the interval $[0, 1]$ after this change becomes random, but for both processes it contains a nonrandom interval $[0, T] := [0, \inf_x \sigma^{-2}(x)]$. This can also be applied to nonhomogeneous SDEs with coefficients depending on time.

2. The random time change leaves the drift bounded. Hence, due to Girsanov's transformation of measure it can be seen that the probability that the process with a lower initial value will attain the level $+1$ over $[0, T]$ is positive and bounded away from zero. Similarly, the probability that the process with a higher initial value will attain the level $-1$ over $[0, T]$ is positive and bounded away from zero. Therefore, they meet on $[0, T]$ with a positive probability which is bounded away from zero, as required.                                                                                      □

### 2.2   MD condition, "case b" and "petite set" conditions

In dimensions $d > 1$ intersections do not work for the "normal" SDEs, and we now switch to the main topic of this paper – local mixing conditions. Global and local versions of the "petite set" and Markov–Dobrushin (MD) conditions will be stated. Most frequently either of them is applied in its local variant, but the global option also works in cases of a uniform ergodicity. It should be noted that, in fact, local versions may vary slightly depending on a particular setting; we only show their main appearances. The "petite set" condition is a localised version of the "case (b)" condition from [4, Chapter V, Section 5], which is, in turn, a simplification of the "condition D" (nowadays called the Doeblin–Doob one) from the same chapter in [4]. Let us highlight that the MD condition may also be in a global or local form.

**Definition 2.** The process satisfies the condition "b" (from [4, Chapter V]) iff there exists a probability measure $\nu$ on the state space $\mathcal{X}$ and constants $T, c > 0$ such that

$$\inf_{x_0 \in \mathcal{X}} \frac{\mu_T^{x_0}(dy)}{\nu(dy)} \geq c. \tag{4}$$

**Definition 3.** The process satisfies the local condition 'b', or "petite set" condition iff there exists a set $D \subset \mathcal{X}$, a probability measure $\nu$ on $D$ and constants $T, c > 0$ such that

$$\inf_{x_0 \in D} \frac{\mu_T^{x_0}(dy)}{\nu(dy)} \geq c. \tag{5}$$

See, in particular, [10] about the usage of the petite set condition in convergence studies. Let us recall that normally this local condition – as well as the local MD condition in Definition 5 in the next paragraphs – should be accomplished with certain recurrence assumptions or properties; however, as it was said earlier, recurrence is not the goal of this paper; it makes sense to study recurrence separately. Both conditions "case b" and MD in their global forms lead to an efficient exponential convergence uniform in the initial data. The next is a more general version of Definition 1.

**Definition 4.** The following is called the global Markov–Dobrushin condition: there exists $T > 0$ such that

$$\kappa(T) := \inf_{x_0, x_1} \int \left( \frac{\mu_T^{x_0}(dy)}{\mu_T^{x_1}(dy)} \wedge 1 \right) \mu_T^{x_1}(dy) > 0. \tag{6}$$

**Definition 5.** The following is called a local Markov–Dobrushin condition: there exist sets $D, D' \subset \mathcal{X}$ in the state space and a constant $T > 0$ such that

$$\kappa(D, D'; T) := \inf_{x_0, x_1 \in D} \int_{D'} \left( \frac{\mu_T^{x_0}(dy)}{\mu_T^{x_1}(dy)} \wedge 1 \right) \mu_T^{x_1}(dy) > 0, \tag{7}$$

**Remark 1.** Usually, but not necessarily $D' = D$; in this case we use the notation $\kappa(D, D'; T) =: \kappa(D; T)$. Another possibility is $D' = R^d$. A sufficient condition for (7) is as follows: there exists a dominating measure $\nu(dy)$ such that $\mu_T^x(dy) \ll \nu(dy)$ for any $x \in D$, and

$$\kappa(D, D'; T) = \inf_{x_0, x_1 \in D} \int_{D'} \left( \frac{\mu_T^{x_0}(dy)}{\nu(dy)} \wedge \frac{\mu_T^{x_1}(dy)}{\nu(dy)} \right) \nu(dy) > 0. \tag{8}$$

In general, there might be no dominating measure for all $x$ simultaneously. Yet, as we shall see, (8) may be verified in most of the cases in what follows. Note that, of course, (8) with any $(D, D')$ implies the same condition with $(D, \mathbb{R}^d)$; however, it may be more convenient by technical reasons to have a bounded set $D'$, as it was realised, for example, in [1].

Clearly, the "petite set" condition implies the MD one, both in the global ("case b") and local versions; for example, (4) implies (6): we have

$$\inf_{x_0, x_1} \int \left( \frac{\mu_T^{x_0}(dy)}{\mu_T^{x_1}(dy)} \wedge 1 \right) \mu_T^{x_1}(dy) = \int \left( \frac{\mu_T^{x_0}(dy)}{\nu(dy)} \wedge \frac{\mu_T^{x_1}(dy)}{\nu(dy)} \right) \nu(dy) \geq c > 0$$

under (4); but, generally speaking, not vice versa. The basis for applying coupling via any of them is the following *coupling lemma* (not to be confused with the *coupling inequality*). Let us add that the MD condition admits some further generalisation, see [15, 16], which provides in certain cases a slightly better bound for efficient convergence under slightly weaker assumptions. However, this note is just about tools which allow to check a local condition MD for SDEs. The following lemma clarifies why the MD condition is so useful; at the same time it serves as the basis for a further application of the MD condition to coupling technique for Markov processes.

**Lemma 1.** *Let $X^1$ and $X^2$ be two random vectors in $\mathbb{R}^d$ on their probability spaces $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$ and $(\Omega^2, \mathcal{F}^2, \mathbb{P}^2)$ (without loss of generality different, which may be made independent after we take their direct product) with densities $p^1$ and $p^2$ with respect to some reference measure $\Lambda$, correspondingly. Then, if*

$$1 - p := q = \int \left( p^1(x) \wedge p^2(x) \right) \Lambda(dx) > 0,$$

*then there exists one more probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and two random variables on it, $\tilde{X}^1$, $\tilde{X}^2$, such that*

$$\mathcal{L}(\tilde{X}^j) = \mathcal{L}(X^j), \; j = 1, 2, \; \& \; \frac{\|\mathcal{L}(X^1) - \mathcal{L}(X^2)\|_{TV}}{2} = P(\tilde{X}^1 \neq \tilde{X}^2) = p.$$

This is a well-known technical tool in the coupling method. The proof – which is simple enough – may be found, for example, in [13]. This reference should not be regarded as a claim that this lemma belongs to the author, although, who was the first inventor of this lemma is not clear to him. The way this lemma may be used for studying convergence rates for Markov chains can be seen, for example, in [13, 16]. The state space could be much more general than $\mathbb{R}^d$, but in this paper we are only concerned with SDEs in finite-dimensional Euclidean spaces.

The next lemma justifies the hint that to estimate the convergence rate for a Markov process to its invariant measure (assuming that it exists) in continuous time $(X_t, t \geq 0)$ it suffices to evaluate it for discrete times $n = 0, 1, \dots$ Its elementary proof is provided for the reader's convenience. Let $\mu_t^X$ be the marginal distribution of $X_t$, and let $\mu^X$ be its invariant measure.

**Lemma 2.** *It holds*

$$\|\mu_t^X - \mu^X\|_{TV} \leq \|\mu_n^X - \mu^X\|_{TV}, \quad t \geq n.$$

**Proof.** Due to the Markov property of $X$, by the Chapman–Kolmogorov equation using the convention $a_+ = a \vee 0$, $a_- = (-a) \vee 0$, we get

$$\frac{1}{2} \|\mu_t^X - \mu^X\|_{TV} = \sup_A (\mathbb{P}_x(X_t \in A) - \mathbb{P}_\mu(X_t \in A))$$

$$= \sup_A \iint 1(z \in A)(\mathbb{P}_x(X_n \in dy) - \mathbb{P}_\mu(X_0 \in dy))\mathbb{P}_y(X_{t-n} \in dz)$$

$$= \sup_A \left( \iint 1(z \in A)(\mathbb{P}_x(X_n \in dy) - \mathbb{P}_\mu(X_0 \in dy))_+ \mathbb{P}_y(X_{t-n} \in dz) \right.$$

$$- \iint 1(z \in A)(\mathbb{P}_x(X_n \in dy) - \mathbb{P}_\mu(X_0 \in dy))_-\mathbb{P}_y(X_{t-n} \in dz))$$

$$\leq \sup_A \iint 1(z \in A)(P_x(X_n \in dy) - \mathbb{P}_\mu(X_0 \in dy))_+\mathbb{P}_y(X_{t-n} \in dz)$$

$$= \iint \mathbb{P}_y(X_{t-n} \in dz)(\mathbb{P}_x(X_n \in dy) - \mathbb{P}_\mu(X_0 \in dy))_+$$

$$= \int (\mathbb{P}_x(X_n \in dy) - \mathbb{P}_\mu(X_0 \in dy))_+ = \frac{1}{2}\|\mu_n^X - \mu^X\|_{TV},$$

as required. □

### 2.3 MD condition using lower and upper bounds of the transition density

Assume $d \geq 1$ and that for the transition densities (fundamental solutions in the PDE language) there exist the Gaussian type upper and lower bounds

$$C_t' \exp(-c_t g(x, x')) \leq f_t(x, x') \leq C_t \exp(-c_t^{-1} g(x, x')) \tag{9}$$

with some appropriate function $g(x, x') \geq 0$ and constants $C_t, C_t', c_t, c_t' > 0$; in the nondegenerate case the function $g$ may be taken in the form $g(x, x') = |x - x'|^2$; in the hypoelliptic cases it may be chosen similarly replacing the Euclidean norm $|x-x'|$ with some other appropriate norm, which reflects the structure of the Hörmander type condition assumed. The *upper* bound in (9) can be established under the nondegeneracy of $\sigma\sigma^*$ for bounded Hölder coefficients in $x$, see the details in [5, 6, 12]. For the full *double* Inequality (9) under the nondegeneracy and some other conditions which are not specified here, see, for example, [17, Theorem A], [2, Theorem 21], [7, Theorem 5]; under certain hypoellipticity conditions, see [3, Theorem 1.1] (where, actually, $g$ itself also depends on $t$). Various similar inequalities may be also found in other sources. In particular, under the nondegeneracy condition on $\sigma\sigma^*$, one can use $C_t = Ct^{-d/2}$, $C_t' = C't^{-d/2}$, $c_t = ct$, $c_t' = c't$ with some $C, C', c, c' > 0$; under the hypoelliptic conditions such constants have some other form; let us emphasize that various versions of Inequality (9) suit our goal: the most important property must be that there are both lower and upper bounds locally finite and locally bounded away from zero for some $t > 0$.

Clearly, a local "petite set" condition is satisfied under (9) with any bounded domain $D$ (an open set by definition) and with the Lebesgue measure as $\nu$. Hence, the MD condition is also valid. To the best of author's knowledge this is the only case where the "petite set" condition can be applied to Markov SDEs in order to arrange coupling; however, this class of coefficients is wide enough, although, far from the most general.

### 2.4 MD condition using stochastic exponentials

In this section let us assume that $d \geq 1$ and that lower and upper bounds for transition densities hold true for the SDE with a "truncated drift"

$$X_t^0 = x + \int_0^t \sigma(X_s^0)dW_s + \int_0^t b_1(X_s^0)ds,$$

while the goal is to arrange a local coupling for the full SDE (1) with the more involved drift of the form

$$b = b_1 + b_2,$$

where the function $b_2$ is just Borel measurable and bounded (this boundedness may be relaxed); $\sigma^{-1}$ is assumed bounded, too. We are interested in establishing an MD condition for the full Equation (1). Note that in general, upper and lower bounds from the previous subsection for the solution of Equation (1) are not known. Denote

$$\tilde{b}_2(x) := \sigma^{-1}(x) b_2(x)$$

and let

$$\rho_T := \exp\left(-\int_0^T \tilde{b}_2(X_t)\,dW_t - \frac{1}{2}\int_0^T \left|\tilde{b}_2(X_t)\right|^2\,dt\right).$$

Recall that $\rho_T$ is a probability density for any $T > 0$. Denote by $\mu_t$ the marginal distribution of $X_t$.

**Theorem 1** (local MD condition via Girsanov). *For any $T > 0$ and $R > 0$,*

$$\kappa(R, T) := \inf_{x_0, x_1 \in B_R} \int_{B_R} \left(\frac{\mu_T^{x_0}(dy)}{\mu_T^{x_1}(dy)} \wedge 1\right) \mu_T^{x_1}(dy) > 0. \tag{10}$$

This inequality suffices for applications to coupling and convergence rates (given suitable recurrence estimates). For the *proofs* of very close statements (actually, even for degenerate SDEs), see [1, 14]. These proofs are based on Girsanov's change of measure via the stochastic exponential $\rho_T$. Some other localised versions of this result may be established: for example, the sets $B_R$ under the infimum sign and as a domain of integration may, actually, differ.

### 2.5   MD condition using parabolic Harnack inequalities
As usual in this paper, in this section we assume that $d \geq 1$, coefficients $b$ and $\sigma$ are bounded (which can be relaxed by a localisation) and Borel measurable, and $\sigma\sigma^*$ is uniformly nondegenerate. Under such conditions Krylov–Safonov's Harnack parabolic inequality holds true [9, Theorem 1.1]; stated in terms of probabilities rather than solutions of PDEs it reads:

$$\sup_{|x_1|, |x_2| \leq 1/4} \frac{\mathbb{P}(X_\tau^{0,x_1} \in d\gamma)}{\mathbb{P}(X_\tau^{\epsilon,x_2} \in d\gamma)}\Big|_{\Gamma_\epsilon} \leq N < \infty, \tag{11}$$

where $\Gamma_\epsilon$ is the parabolic boundary of the cylinder $((t, x) : |x| \leq 1; \epsilon \leq t \leq 1)$, i.e. $(\Gamma_\epsilon = \Gamma_\epsilon^{(t=1)} \cup \Gamma_\epsilon^{(t<1)})$, $\Gamma_\epsilon = ((t, x) : (|x| = 1 \,\&\, \epsilon \leq t \leq 1) \cup (|x| \leq 1 \,\&\, t = 1))$, and

$$\tau := \inf(t \geq 0 : |X_t| \geq 1), \quad \text{with a convention } \inf(\emptyset) = 1;$$

the constant $N$ depends on $d$, on the ellipticity constants of the diffusion, on the sup-norm of the drift, and on $\epsilon$. Note that in (11) the measure in the numerator is absolutely continuous with respect to the one in the denominator, that is, there is no singular component in this situation. Let

$$\mu^{x_1}(d\gamma) = \mathbb{P}(X_\tau^{0,x_1} \in d\gamma), \quad \mu^{\epsilon,x_2}(d\gamma) = \mathbb{P}(X_\tau^{\epsilon,x_2} \in d\gamma),$$

where $d\gamma$ is the element of the boundary $\Gamma_\epsilon$. Then the following local mixing bound holds true.

**Theorem 2** (local MD via parabolic Harnack). *Let $\mu^{x_1}(\Gamma_\epsilon) \geq q$ with some $q > 0$. Then a version of the Markov–Dobrushin condition holds:*

$$\inf_{|x_1|, |x_2| \leq 1/4} \int_{\Gamma_\epsilon} \left( \frac{\mu^{\epsilon, x_2}(d\gamma)}{\mu^{x_1}(d\gamma)} \wedge 1 \right) \mu^{x_1}(d\gamma) \geq \frac{q}{N} > 0. \tag{12}$$

Note that the value $q$ here may be chosen arbitrarily close to one, if $\epsilon > 0$ is small enough. However, the decrease of $\epsilon$ implies the increase of the constant $N$ in (11).

**Proof.** Indeed, due to Inequality (11) we have

$$f := \frac{d\mu^{x_1}}{d\mu^{\epsilon, x_2}}|_{\Gamma_\epsilon} \leq N \quad \& \quad \mu^{x_1} \ll \mu^{\epsilon, x_2} \text{ on } \Gamma_\epsilon.$$

Denote by $\tilde{\mu}^{\epsilon, x_2}$ the absolutely continuous part of $\mu^{\epsilon, x_2}$ with respect to $\mu^{x_1}$ (we do not know whether there exists a nontrivial singular component here, but the calculus in what follows does not depend on this). Then

$$\frac{d\tilde{\mu}^{\epsilon, x_2}}{d\mu^{x_1}}|_{\Gamma_\epsilon} = \frac{1}{f} \geq \frac{1}{N}.$$

Hence, the assumption $\mu^{x_1}(\Gamma_\epsilon) \geq q$ implies

$$\int_{\Gamma_\epsilon} \left( \frac{\mu^{\epsilon, x_2}(d\gamma)}{\mu^{x_1}(d\gamma)} \wedge 1 \right) \mu^{x_{x_1}}(d\gamma) = \int_{\Gamma_\epsilon} \left( \frac{\tilde{\mu}^{\epsilon, x_2}(d\gamma)}{\mu^{x_1}(d\gamma)} \wedge 1 \right) \mu^{x_1}(d\gamma)$$

$$\geq \int_{\Gamma_\epsilon} \frac{1}{N} \mu^{x_1}(d\gamma) \geq \frac{q}{N} > 0.$$

$\square$

Sometimes it may be more convenient to use another version of the MD condition, which follows from Theorem 2. Denote

$$\mu_1^x(A) := \mathbb{P}_x(X_1 \in A), \quad A \subset R^d.$$

**Corollary 1.** *Under the assumptions of Theorem 2 the following version of the MD condition holds: there exists $q' \in (0, q)$ such that*

$$\inf_{|x_1|, |x_2| \leq 1/8} \int_{\mathbb{R}^d} \left( \frac{\mu_1^{x_1}(dy)}{\mu_1^{x_2}(dy)} \wedge 1 \right) \mu_1^{x_2}(dy) \geq \frac{q'}{N}. \tag{13}$$

Note that here $\mathbb{R}^d$ plays the role of $D'$ in the MD condition. In some cases this may not be convenient; however, using moment bounds of the solution a reasonable version of this inequality with some bounded ball $B_R$ in place of $R^d$ is, of course, possible. We leave it till further studies where such a replacement may be required.

**Proof.** Note that due to the boundedness of $\sigma$ and $b$,

$$\inf_{|x|\leq 1/8} \mathbb{P}_x(|X_\epsilon| \leq 1/4) > 0.$$

Denote

$$q' := q \inf_{|x|\leq 1/8} \mathbb{P}_x(|X_\epsilon| \leq 1/4).$$

We have

$$\mu_1^{x_2}(dy) = \mathbb{P}_{x_2}(X_1 \in dy) = \mathbb{E}_{x_2}\mathbb{E}(X_1 \in dy|X_\epsilon) \geq$$

$$\geq \mathbb{E}_{x_2}1(|X_\epsilon| \leq 1/4)\mathbb{E}(X_1 \in dy|X_\epsilon) = \mathbb{E}_{x_2}1(|X_\epsilon| \leq 1/4)\mu_1^{\epsilon,X_\epsilon}(dy).$$

Hence, denoting $\nu_{\epsilon,x_2}(dz) := \mathbb{P}_{x_2}(X_\epsilon \in dz)$, we find

$$1 \wedge \frac{\mu_1^{x_2}(dy)}{\mu_1^{x_1}(dy)} \geq 1 \wedge \frac{\mathbb{E}_{x_2}1(|X_\epsilon| \leq 1/4)\mu_1^{\epsilon,X_\epsilon}(dy)}{\mu_1^{x_1}(dy)}$$

$$= 1 \wedge \frac{\int \nu_{\epsilon,x_2}(dz)1(|z| \leq 1/4)\mu_1^{\epsilon,z}(dy)}{\mu_1^{x_1}(dy)}$$

$$= 1 \wedge \left( \int \nu_{\epsilon,x_2}(dz)1(|z| \leq 1/4)\frac{\mu_1^{\epsilon,z}(dy)}{\mu_1^{x_1}(dy)} \right)$$

$$\geq \int \nu_{\epsilon,x_2}(dz)\left( 1 \wedge 1(|z| \leq 1/4)\frac{\mu_1^{\epsilon,z}(dy)}{\mu_1^{x_1}(dy)} \right)$$

$$\geq \int \nu_{\epsilon,x_2}(dz)1(|z| \leq 1/4)\left( 1 \wedge \frac{\mu_1^{\epsilon,z}(dy)}{\mu_1^{x_1}(dy)} \right).$$

So,

$$\int \left( 1 \wedge \frac{\mu_1^{x_2}(dy)}{\mu_1^{x_1}(dy)} \right) \mu_1^{x_1}(dy)$$

$$\geq \int \left[ \int \nu_{\epsilon,x_2}(dz)1(|z| \leq 1/4)\left( 1 \wedge \frac{\mu_1^{\epsilon,z}(dy)}{\mu_1^{x_1}(dy)} \right) \right] \mu_1^{x_1}(dy)$$

$$= \int \nu_{\epsilon,x_2}(dz)1(|z| \leq 1/4)\left[ \int \left( 1 \wedge \frac{\mu_1^{\epsilon,z}(dy)}{\mu_1^{x_1}(dy)} \right) \mu_1^{x_1}(dy) \right]. \tag{14}$$

This was the first step in the reduction of the MD characteristics in the left hand side of (14) to the Harnack inequality: now we may deal with the measures $\mu_1^{\epsilon,z}(dy)$ and $\mu_1^{x_1}(dy)$. However, these are still not the ones which show up in (11) or in (12). The next step will complete this reduction. Let

$$\tilde{\Lambda}_{x_1,z}(dy) := \mu_1^{\epsilon,z}(dy) + \mu_1^{x_1}(dy).$$

Then

$$\int \left( 1 \wedge \frac{\mu_1^{\epsilon,z}(dy)}{\mu_1^{x_1}(dy)} \right) \mu_1^{x_1}(dy) = \int \left( \frac{\mu_1^{x_1}(dy)}{\tilde{\Lambda}_{x_1,z}(dy)} \wedge \frac{\mu_1^{\epsilon,z}(dy)}{\tilde{\Lambda}_{x_1,z}(dy)} \right) \tilde{\Lambda}_{x_1,z}(dy). \tag{15}$$

We have

$$\frac{\mu_1^{\epsilon,z}(dy)}{\tilde{\Lambda}_{x_1,z}(dy)} \wedge \frac{\mu_1^{x_1}(dy)}{\tilde{\Lambda}_{x_1,z}(dy)} \geq \frac{\mathbb{P}_{x_1}(X_1 \in dy, \tau < 1)}{\tilde{\Lambda}_{x_1,z}(dy)} \wedge \frac{\mathbb{P}_z(X_{1-\epsilon} \in dy, \tau < 1-\epsilon)}{\tilde{\Lambda}_{x_1,z}(dy)}. \quad (16)$$

Therefore,

$$\int \left( \frac{\mu_1^{x_1}(dy)}{\tilde{\Lambda}_{x_1,z}(dy)} \wedge \frac{\mu_1^{\epsilon,z}(dy)}{\tilde{\Lambda}_{x_1,z}(dy)} \right) \tilde{\Lambda}_{x_1,z}(dy)$$

$$\geq \int_{R^d} \left( \frac{\mathbb{P}_{x_1}(X_1 \in dy, \tau < 1)}{\tilde{\Lambda}_{x_1,z}(dy)} \wedge \frac{\mathbb{P}_z(X_{1-\epsilon} \in dy, \tau < 1-\epsilon)}{\tilde{\Lambda}_{x_1,z}(dy)} \right) \tilde{\Lambda}_{x_1,z}(dy)$$

$$= \int_{R^d} \left( \frac{\mathbb{P}_{x_1}(X_1 \in dy, \tau < 1)}{\mathbb{P}_z(X_{1-\epsilon} \in dy, \tau < 1-\epsilon)} \wedge 1 \right) \mathbb{P}_z(X_{1-\epsilon} \in dy, \tau < 1-\epsilon).$$

Further, since $|x_1| \leq 1/4$ and $|z| \leq 1/8$, then due to the strong Markov property and by virtue of Inequality (11) we have

$$\mathbb{P}_{x_1}(X_1 \in dy, \tau < 1) = \mathbb{E}_{x_1} 1(\tau < 1)(X_1 \in dy)$$
$$= \mathbb{E}_{x_1} \mathbb{E}\left(1(\tau < 1)(X_1 \in dy)|\mathcal{F}_\tau\right) = \mathbb{E}_{x_1} 1(\tau < 1)\mathbb{E}\left((X_1 \in dy)|\mathcal{F}_\tau\right)$$
$$= \mathbb{E}_{x_1} 1(\tau < 1)\mathbb{E}\left((X_1 \in dy)|X_\tau\right)$$
$$= \mathbb{E}_{x_1} 1(\tau < 1)\mathbb{E}_{t,y}\left((X_{1-t} \in dy)\right)|_{(t,y)=(\tau,X_\tau)}$$
$$\geq \frac{q}{N} E_{x_2} 1(\tau < 1)\mathbb{E}_{t,y}\left((X_{1-t} \in dy)\right)|_{(t,y)=(\tau,X_\tau)}$$
$$= \frac{q}{N}\mathbb{P}_z(X_{1-\epsilon-t} \in dy, \tau < 1).$$

So,

$$\int_{R^d} \left( \frac{\mathbb{P}_{x_1}(X_1 \in dy, \tau < 1)}{\mathbb{P}_z(X_{1-\epsilon} \in dy, \tau < 1)} \wedge 1 \right) \mathbb{P}_z(X_{1-\epsilon} \in dy, \tau < 1)$$

$$\geq \frac{q}{N} \int_{R^d} \mathbb{P}_z(X_{1-\epsilon} \in dy, \tau < 1) = \frac{q}{N}\mathbb{P}_z(\tau < 1).$$

Recall that in (14) the integrand involves the indicator $1(|z| \leq 1/4)$. Clearly,

$$\kappa := \inf_{|z| \leq 1/4} \mathbb{P}_z(\tau < 1) > 0.$$

Hence, due to (14), (15) and (16),

$$\int \nu_{\epsilon,x_2}(dz)1(|z| \leq 1/4)\left[ \int \left( 1 \wedge \frac{\mu_1^{\epsilon,z}(dy)}{\mu_1^{x_1}(dy)} \right) \mu_1^{x_1}(dy) \right]$$

$$\geq \frac{q\kappa}{N} \int \nu_{\epsilon,x_2}(dz)1(|z| \leq 1/4) = \frac{q\kappa}{N} \int 1(|z| \leq 1/4)\mathbb{P}_{x_2}(X_\epsilon \in dz) \geq \frac{q'}{N}$$

with some $0 < q' < q\kappa$, as required. Inequality (13) follows. $\qquad \square$

### 2.6 *MD condition using elliptic Harnack inequalities*

The assumptions of this sections are the same as in the previous one: $d \geq 1$, co-
efficients $b$ and $\sigma$ are bounded (which can be relaxed) and Borel measurable, and
$\sigma\sigma^*$ is uniformly nondegenerate. We have the elliptic Harnack inequality due to [11,
Theorem 3.1], stated here in its probabilistic form (while in [11] it is offered in the
language of elliptic PDEs): there exists a constant $N > 0$ such that for any $0 < R \leq 1$
and any $A \in \partial B_R$,

$$\sup_{|x| \leq R/8} \mathbb{P}_x(X_{\tau_R} \in A) \leq N \inf_{|x| \leq R/8} \mathbb{P}_x(X_{\tau_R} \in A), \tag{17}$$

where $\tau_R = \inf(t \geq 0 : |X_t| \geq R)$, and $\partial B_R$ is the boundary of the ball $B_R$. This
inequality itself is some MD condition. In fact, it is not clear whether this version
of the Harnack inequality may be helpful for estimating the convergence rate of the
distribution of $X_t$ to its stationary regime. Nevertheless, if it can be used for such a
purpose – which is the author's hope – then it might be more convenient to apply the
following version of the MD condition based on Inequality (17). Let

$$Q_{R,T} := \{(t, x) : \ t \leq T, \ |x| \leq R\}.$$

Note that

$$\bigcup_{T>0} Q_{R,T} = R_+ \times B_R = R_+ \times (x : |x| \leq R).$$

Denote by $\Gamma_{R,T}$ the part of the parabolic boundary of $Q_{R,T}$ corresponding to $t < T$,
namely,

$$\Gamma_{R,T} = ((t, x) : 0 \leq t \leq T, \ |x| = R).$$

Denote $\tau_{R,T} := \inf(t \geq 0 : X_t \notin B_R) \wedge T$, $\tau_R := \inf(t \geq 0 : X_t \notin B_R)$, and

$$\nu_{R,T}^x(A) := \mathbb{P}_x(X_{\tau_{R,T}} \in A), \quad \nu_R^x(A) := \mathbb{P}_x(X_{\tau_R} \in A), \quad A \subset \partial B_R.$$

**Theorem 3** (local MD via elliptic Harnack). *The local MD condition*

$$\inf_{x_1, x_2 \in B_R} \int_{B_R} \left( \frac{\nu_{R,T}^{x_1}(d\gamma)}{\nu_R^0(d\gamma)} \wedge \frac{\nu_{R,T}^{x_2}(d\gamma)}{\nu_R^0(d\gamma)} \right) \nu_R^0(d\gamma) \geq \frac{C_R}{2N} \tag{18}$$

*holds true for any $T > 0$ large enough.*

In principle, it is possible to evaluate such values of $T$ for which (18) holds, and
potentially it might be useful for estimating convergence rates.

**Proof.** Clearly, $\tau_{R,T} \leq \tau_R$. Note that due to the nondegeneracy of $\sigma$ we have $\tau_R < \infty$
a.s., and

$$\lim_{T \to \infty} \inf_{|x| \leq R/8} \mathbb{P}_x(\tau_{R,T} = \tau_R) = \lim_{T \to \infty} \inf_{|x| \leq R/8} \mathbb{P}_x(\tau_R < T) = 1.$$

Equivalently,

$$\lim_{T \to \infty} \sup_{|x| \leq R/8} \mathbb{P}_x(\tau_{R,T} < \tau_R) = \lim_{T \to \infty} \inf_{|x| \leq R/8} \mathbb{P}_x(T < \tau_R) = 0.$$

Hence,

$$v_{R,T}^x(A) = \mathbb{P}_x(X_{\tau_{R,T}} \in A) = \mathbb{P}_x(X_{\tau_R} \in A, \tau_R < T) \uparrow \mathbb{P}_x(X_{\tau_R} \in A) = v_R^x(A), \tag{19}$$

as $T \uparrow \infty$, where the convergence is uniform with respect to $A$ and $|x| \leq R/8$.

Inequality (17) implies the following:

$$0 < N^{-1} \leq \inf_{|x| \leq 1/8} \frac{v_R^x(d\gamma)}{v_R^0(d\gamma)}|_{\partial B_R} \leq \sup_{|x| \leq 1/8} \frac{v_R^x(d\gamma)}{v_R^0(d\gamma)}|_{\partial B_R} \leq N < \infty.$$

As a consequence, for any $R > 0$ we get

$$\inf_{x_1, x_2 \in B_R} \int_{B_R} \left( \frac{v_R^{x_1}(d\gamma)}{v_R^0(d\gamma)} \wedge \frac{v_R^{x_2}(d\gamma)}{v_R^0(d\gamma)} \right) v_R^0(d\gamma) \geq \frac{C_R}{N} > 0. \tag{20}$$

By virtue of the monotone convergence theorem and due to (19) we have for any $x_1$, $x_2$,

$$\int_{B_R} \left( \frac{v_{R,T}^{x_1}(d\gamma)}{v_R^0(d\gamma)} \wedge \frac{v_{R,T}^{x_2}(d\gamma)}{v_R^0(d\gamma)} \right) v_R^0(d\gamma)$$

$$\rightarrow \int_{B_R} \left( \frac{v_R^{x_1}(d\gamma)}{v_R^0(d\gamma)} \wedge \frac{v_R^{x_2}(d\gamma)}{v_R^0(d\gamma)} \right) v_R^0(d\gamma), \quad T \rightarrow \infty.$$

Hence, for $T$ large enough we obtain from (20),

$$\int_{B_R} \left( \frac{v_{R,T}^{x_1}(d\gamma)}{v_R^0(d\gamma)} \wedge \frac{v_{R,T}^{x_2}(d\gamma)}{v_R^0(d\gamma)} \right) v_R^0(d\gamma) \geq \frac{C_R}{2N}.$$

However, technically this is still not sufficient because we want a similar inequality with infimum $\inf_{x_1, x_2 \in B_R}$. Using the elementary inequality $(a - b) \wedge (c - d) \geq a \wedge c - b - d$ along with the identity

$$\mathbb{P}_x(X_{\tau_R} \in d\gamma, \tau_R < T) = \mathbb{P}_x(X_{\tau_R} \in d\gamma) - \mathbb{P}_x(X_{\tau_R} \in d\gamma, \tau_R \geq T),$$

we get

$$\int_{B_R} \left( \frac{v_R^{x_1}(d\gamma)}{v_R^0(d\gamma)} \wedge \frac{v_R^{x_2}(d\gamma)}{v_R^0(d\gamma)} \right) v_R^0(d\gamma) \geq \int_{B_R} \left( \frac{v_{R,T}^{x_1}(d\gamma)}{v_R^0(d\gamma)} \wedge \frac{v_{R,T}^{x_2}(d\gamma)}{v_R^0(d\gamma)} \right) v_R^0(d\gamma)$$

$$= \int_{B_R} \left( \frac{\mathbb{P}_{x_1}(X_{\tau_R} \in d\gamma, \tau_R < T)}{v_R^0(d\gamma)} \wedge \frac{\mathbb{P}_{x_1}(X_{\tau_R} \in d\gamma, \tau_R < T)}{v_R^0(d\gamma)} \right) v_R^0(d\gamma)$$

$$\geq \int_{B_R} \left( \frac{\mathbb{P}_{x_1}(X_{\tau_R} \in d\gamma)}{v_R^0(d\gamma)} \wedge \frac{\mathbb{P}_{x_2}(X_{\tau_R} \in d\gamma)}{v_R^0(d\gamma)} \right) v_R^0(d\gamma)$$

$$- \int_{B_R} \left( \frac{\mathbb{P}_{x_1}(X_{\tau_R} \in d\gamma, \tau_R \geq T)}{v_R^0(d\gamma)} + \frac{\mathbb{P}_{x_2}(X_{\tau_R} \in d\gamma, \tau_R \geq T)}{v_R^0(d\gamma)} \right) v_R^0(d\gamma).$$

Here

$$\left(\sup_{x_1,x_2\in B_R}\right)\int_{B_R}\left(\frac{\mathbb{P}_{x_1}(X_{\tau_R}\in d\gamma,\tau_R\geq T)}{\nu_R^0(d\gamma)}+\frac{\mathbb{P}_{x_2}(X_{\tau_R}\in d\gamma,\tau_R\geq T)}{\nu_R^0(d\gamma)}\right)\nu_R^0(d\gamma)$$

$$=\left(\sup_{x_1,x_2\in B_R}\right)\int_{B_R}\left(\mathbb{P}_{x_1}(X_{\tau_R}\in d\gamma,\tau_R\geq T)+\mathbb{P}_{x_2}(X_{\tau_R}\in d\gamma,\tau_R\geq T)\right)$$

$$\leq 2\sup_{x\in B_R}\mathbb{P}_x(\tau_R\geq T)\to 0,\quad T\to\infty.$$

So,

$$\inf_{x_1,x_2\in B_R}\int_{B_R}\left(\frac{\nu_{R,T}^{x_1}(d\gamma)}{\nu_R^0(d\gamma)}\wedge\frac{\nu_{R,T}^{x_2}(d\gamma)}{\nu_R^0(d\gamma)}\right)\nu_R^0(d\gamma)$$

$$\to\inf_{x_1,x_2\in B_R}\int_{B_R}\left(\frac{\nu_R^{x_1}(d\gamma)}{\nu_R^0(d\gamma)}\wedge\frac{\nu_R^{x_2}(d\gamma)}{\nu_R^0(d\gamma)}\right)\nu_R^0(d\gamma),\quad T\to\infty.$$

On the other hand,

$$\inf_{x_1,x_2\in B_R}\int_{B_R}\left(\frac{\mathbb{P}_{x_1}(X_{\tau_R}\in d\gamma)}{\nu_R^0(d\gamma)}\wedge\frac{\mathbb{P}_{x_2}(X_{\tau_R}\in d\gamma)}{\nu_R^0(d\gamma)}\right)\nu_R^0(d\gamma)\geq\frac{C_R}{N}.$$

Therefore, there exists $T_0$ such that

$$\inf_{x_1,x_2\in B_R}\int_{B_R}\left(\frac{\nu_{R,T}^{x_1}(d\gamma)}{\nu_R^0(d\gamma)}\wedge\frac{\nu_{R,T}^{x_2}(d\gamma)}{\nu_R^0(d\gamma)}\right)\nu_R^0(d\gamma)\geq\frac{C_R}{2N},\quad T\geq T_0,$$

which completes the proof. □

## Acknowledgments

The author is grateful to two anonymous referees for useful remarks.

## Funding

## References

[1] Abourashchi, N., Veretennikov, A.Yu.: On stochastic averaging and mixing. Theory Stoch. Process. **16(32)**(1), 111–130 (2010). MR2779833

[2] Bally, V.: Lower bounds for the density of locally elliptic Itô processes. Ann. Probab. **34**(6), 2406–2440 (2006). MR2294988. https://doi.org/10.1214/009117906000000458

[3] Delarue, F., Menozzi, S.: Density estimates for a random noise propagatingthrough a chain of differential equations. J. Funct. Anal. **259**, 1577–1630 (2010). MR2659772. https://doi.org/10.1016/j.jfa.2010.05.002

[4] Doob, J.L.: Stochastic processes. J. Wiley and Sons, New York (1953). MR0058896

[5] Eidelman, S.D.: Parabolic systems. Amsterdam, North-Holland and Pub. Co., Groningen, Wolters-Noordhoff (1969). MR0252806

[6] Friedman, A.: Partial Differential Equations of Parabolic Type. Prentice-Hall Inc., Englewood Cliffs, NJ (1964). MR0181836

[7] Kohatsu-Higa, A.: Lower bounds for densities of uniformly elliptic random variables on Wiener space. Probab. Theory Relat. Fields **126**, 421–457 (2003). MR1992500. https://doi.org/10.1007/s00440-003-0272-4

[8] Krylov, N.V.: On the selection of a Markov process from a system of processes and the construction of quasi-diffusion processes. Math. USSR, Izv. **7**(3), 691–709 (1973). MR0339338. https://doi.org/10.1070/IM1973v007n03ABEH001971

[9] Krylov, N.V., Safonov, M.V.: A certain property of solutions of parabolic equations with measurable coefficients. Math. USSR, Izv. **16**(1), 151–164 (1981). MR0563790. https://doi.org/10.1070/IM1981v016n01ABEH001283

[10] Meyn, S., Tweedie, R.L.: Markov Chains and Stochastic Stability, 2nd edn. Cambridge University Press, Cambridge (2009). MR2509253. https://doi.org/10.1017/CBO9780511626630

[11] Safonov, M.V.: Harnack's inequality for elliptic equations and the Hölder property of their solutions. J. Sov. Math. **21**, 851–863 (1983). https://doi.org/10.1007/BF01094448

[12] Solonnikov, V.A.: On boundary value problems for linear parabolic systems of differential equations of general form. Proc. Steklov Inst. Math. **83**, 1–184 (1965). MR0211083

[13] Veretennikov, A.Yu.: Ergodic Markov processes and Poisson equations (lecture notes). In: Panov, V. (ed.) Modern problems of stochastic analysis and statistics – Selected contributions in honor of Valentin Konakov, pp. 457–511. Springer (2017). MR3747677. https://doi.org/10.1007/978-3-319-65313-6_18

[14] Veretennikov, A.Yu.: On Weak Solutions of Highly Degenerate SDEs. Autom. Remote Control **81**(3), 398–410 (2020). MR4100228. https://doi.org/10.1134/S0005117920030029

[15] Veretennikov, A.Yu., Veretennikova, M.A.: On convergence rates for homogeneous Markov chains. Dokl. Math. **101**(1), 12–15 (2020). https://rdcu.be/b4h3F. https://doi.org/10.1134/S1064562420010081

[16] Veretennikov, A.Yu., Veretennikova, M.A.: On improved convergence conditions and bounds for Markov chains. https://arxiv.org/abs/2006.12134 (accepted to Izvestiya Mathem. **86**(1) (2022). https://doi.org/10.4213/im9076).

[17] Zhang, Q.S.: Gaussian bounds for the fundamental solutions of $\nabla(A\nabla u) + B\nabla u - u_t = 0$. Manuscr. Math. **93**, 381–390 (1997). MR1457736. https://doi.org/10.1007/BF02677479