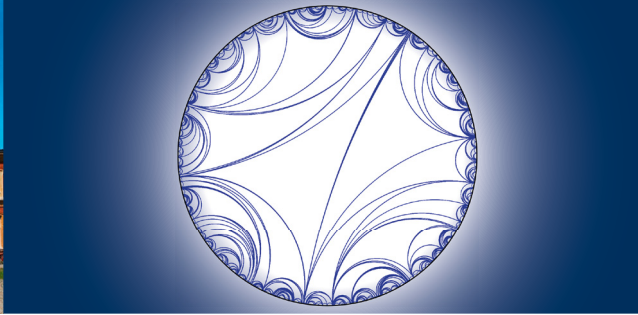




Potsdam University Press



Alexei Kulik

Introduction to Ergodic Rates for Markov Chains and Processes

With Applications to Limit Theorems

Lectures in Pure and Applied Mathematics

Lectures in Pure and Applied Mathematics | 2

Alexei Kulik

Introduction to Ergodic Rates for Markov Chains and Processes

With Applications to Limit Theorems

Potsdam University Press

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.dnb.de>.

Potsdam University Press 2015

<http://verlag.ub.uni-potsdam.de/>

Am Neuen Palais 10, 14469 Potsdam
Tel.: +49 (0)331 977 2533 / Fax: 2292
E-Mail: verlag@uni-potsdam.de

The monograph series **Lectures in Pure and Applied Mathematics** is edited by the Institute of Mathematics of the University of Potsdam.

ISSN (print) 2199-4951
ISSN (online) 2199-496X

Contact:

Institut für Mathematik
Am Neuen Palais 10
14469 Potsdam
Tel.: +49 (0)331 977 1028
WWW: <http://www.math.uni-potsdam.de>

Cover images:

1. Karla Fritze | Building of the Institute at Campus Neues Palais
 2. Nicolas Curien, Wendelin Werner | Random hyperbolic triangulation
- Published at: Journal of the European Mathematical Society, Volume 15, Issue 4, 2013, pp. 1309–1341, DOI: 10.4171/JEMS/393

This work is licensed under a Creative Commons License:

Attribution – Share Alike 4.0 International

To view a copy of this license visit

<http://creativecommons.org/licenses/by-sa/4.0/>

Print: docupoint GmbH Magdeburg

Alexei Kulik, Ukrainian National Academy of Sciences, Institute of Mathematics

ISBN 978-3-86956-338-1

Simultaneously published online at the Institutional Repository of the University of Potsdam:

URN [urn:nbn:de:kobv:517-opus4-79360](http://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-79360)

<http://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-79360>

To Masha and Katja, my beloved daughters

Preface

The general topic of this lecture course is the *ergodic behavior* of Markov processes. Calling a Markov process *ergodic* one usually means that this process has a unique invariant probability measure. For an ergodic Markov process it is very typical that its transition probabilities converge to the invariant probability measure when the time variable tends to $+\infty$. This feature is commonly called the *stochastic stability*, and the rate of such a convergence is called the *ergodic rate* for the Markov process under consideration. Explicit ergodic rates is a point of a major interest, since such rates typically may serve as a basis for a wide range of further applications.

This lecture course pursues two major goals. The first one is a detailed introduction to general methods for proving explicit upper bounds for ergodic rates. This topic is quite diverse: depending on an actual long-time behavior of a Markov process, different notions and tools appear to be most natural for the description of the respective ergodic rate. We explain the genealogy of these ideas and methods, starting from a comparatively simple and classical case of a Markov process which possesses *uniform* ergodic rates in *total variation distance*, and ending up with much less studied *weak* ergodic rates, which appear to be very efficient for Markov processes with a complicated local structure. Our second aim is to discuss one particularly important field where the ergodic rates for Markov processes find their natural application: the limit theorems for functionals of Markov processes. The two principal parts of the lecture course are closely connected; in particular, the form in which we give the ergodic rates is strongly motivated by their further applications.

The history of the topic is extremely rich, and it would be impossible to give in a short lecture course a detailed exposition of the whole diverse set of the methods and settings developed in the field so far. Thus we design this course in order to give a reader complete view of *one* possible route across this extremely interesting field, and to give hints which

would help to understand better the diversity of notions, methods, and ideas used within the field. For a more extended reading we refer to [12], Chapters V,VI, [45], [28], [43], [22]; this list is definitely far from being complete.

The lecture course is mainly oriented on graduate and post-graduate students, specialized in Probability and Statistics, and on specialists in other fields where Markov models are typically applied. Its minimal pre-requisites are standard courses of Probability and Measure Theory, but a basic knowledge in Stochastic Processes, Stochastic Calculus, and SDE's are helpful for a better understanding of the particular examples, which we discuss in details because they strongly motivate the choice of the form in which we present the theory.

This lecture course has its origin in two minicourses, given by the author at the University of Potsdam, TU Berlin and Humboldt University of Berlin (March 2013), and in Ritsumeikan University (August 2013). It has been prepared partially during the authors stay at TU Dresden (January 2014), Ritsumeikan University (January 2015), and Institute Henri Poincaré (July 2014, "Research in Paris" programme). The author is glad to express his gratitude to all these institutions. Specially, the author would like to thank Dr. Prof. Sylvie Roelly for her encouragement to compose these lecture notes and for a persistent support during their preparation. Many thanks also to Max Schneider for his help in the English formulation of a part of this text and to Dr. Mathias Rafterl for his \TeX expertise, which considerably improved the presentation.

Contents

Preface	vii
Preliminaries	1
1 Ergodic Rates in Total Variation Distance	3
1.1 Total variation distance and the Coupling Lemma	3
1.2 Uniform ergodic rate.	6
1.3 Non-uniform ergodic rates	12
1.4 Continuous-time Markov processes	26
1.5 Various forms of the irreducibility assumption.	33
1.6 Diffusions and Lévy driven SDEs	36
2 Weak Ergodic Rates	51
2.1 Markov models with intrinsic memory	51
2.2 Coupling distances for probability measures	55
2.3 Dissipative stochastic systems	64
2.4 General Harris-type theorem.	73
3 Limit Theorems	79
3.1 Preliminaries: Covariance and Mixing Coefficient	79
3.2 Covariances and the LLN, revisited	85
3.3 The corrector term and the CLT.	88
3.4 Autoregressive models with Markov regime	100
Bibliography	117

Preliminaries

In what follows, $(\mathbb{X}, \mathcal{X})$ is a Borel measurable space; that is, a measurable space which admits a bijection to $([0, 1], \mathcal{B}([0, 1]))$, measurable together with its inverse. A typical case here would be a Polish space \mathbb{X} with the Borel σ -algebra \mathcal{X} .

The main object of our interest is a time-homogeneous Markov process X with the state space \mathbb{X} . Mainly we consider the case with the discrete time set $\mathbb{T} = \mathbb{Z}_+$; that is, in other words $X = \{X_n, n \geq 0\}$ is a *Markov chain with a general state space*. The main constructions in the discrete time setting are more transparent and easier to explain. We address the case of the continuous time set $\mathbb{T} = \mathbb{R}_+$ as well, mainly to give examples, explain motivation, and expose some particular methods for the continuous-time setting.

In the discrete time setting, we denote by $P_n(x, dx')$, $n \geq 1$, the transition probability kernels for X , and by $P(x, dx') = P_1(x, dx')$ its one-step transition probability kernel. If we need to emphasize that $P(x, dx')$ relates to a chain X , we write $P^X(x, dx')$ instead. Given $P(x, dx')$ and an initial distribution μ on \mathbb{X} , the law of the whole sequence $\{X_n, n \geq 0\}$ in $(\mathbb{X}^\infty, \mathcal{X}^{\otimes \infty})$ is completely defined; we denote this law by P_μ and respective expectation by E_μ . If μ is concentrated in a point x , we write P_x, E_x instead of $P_{\delta_x}, E_{\delta_x}$.

In what follows, we denote by $\mathcal{P}(\mathbb{X})$ the set of all probability measures on \mathbb{X} . A measure $\mu \in \mathcal{P}(\mathbb{X})$ is called an *invariant probability measure* (IPM) for X , if

$$\mu(A) = \int_{\mathbb{X}} P(x, A) \mu(dx), \quad A \in \mathcal{X}.$$

Observe that $\mu \in \mathcal{P}(\mathbb{X})$ is an IPM for X if, and only if, the law of every $X_n, n \geq 1$, under P_μ equals μ ; in that case, $X = \{X_n, n \geq 0\}$ is a *strictly stationary* random sequence.

In the continuous time case the transition probability kernels are denoted by $P_t(x, dx')$,

$t \geq 0$, and $\mu \in \mathcal{P}(\mathbb{X})$ is an IPM for X if

$$\mu(A) = \int_{\mathbb{X}} P_t(x, A) \mu(dx), \quad A \in \mathcal{X}, t > 0.$$

The semigroup, generated by X in the Banach space $\mathcal{B}(\mathbb{X})$ of bounded measurable real-valued functions on \mathbb{X} with the sup-norm, is defined by

$$T_t f(x) = \int_{\mathbb{X}} f(y) P_t(x, dy) = \mathbb{E}_x f(X_t), \quad f \in \mathcal{B}(\mathbb{X}), t \geq 0.$$

Chapter 1

Ergodic Rates in Total Variation Distance

This chapter is devoted to the part of the theory which studies the ergodic rates w.r.t. the total variation distance in $\mathcal{P}(\mathbb{X})$. For the reader's convenience, at first we briefly recall the definition of total variation distance and the properties of this distance crucial for the subsequent analysis of ergodic rates (Section 1.1). Uniform and non-uniform ergodic rates first will be developed separately for discrete-time Markov processes (Section 1.2 and Section 1.3), and then will be extended for continuous-time processes (Section 1.4). In Section 1.5 we discuss separately the various forms of the principal *irreducibility* assumption used in the theory; in Section 1.6 applications to particularly important classes of diffusions and solutions to Lévy driven SDE's are given.

.....

1.1 Total variation distance and the Coupling Lemma

For two probability measures μ and ν on \mathbb{X} , the total variation distance $\|\mu - \nu\|_{\text{TV}}$ is just the total variation of the signed measure $\mu - \nu$. If λ is a σ -finite measure such that $\mu \ll \lambda$ and $\nu \ll \lambda$, then

$$\|\mu - \nu\|_{\text{TV}} = \int_{\mathbb{X}} \left| \frac{d\mu}{d\lambda} - \frac{d\nu}{d\lambda} \right| d\lambda.$$

Observe that for any $\mu, \nu \in \mathcal{P}(\mathbb{X})$ there exists a σ -finite (and even a probability) measure λ such that $\mu \ll \lambda$ and $\nu \ll \lambda$.

Exercise 1.1 Construct such a measure $\lambda \in \mathcal{P}(\mathbb{X})$.

In other words, the total variation distance between μ, ν is just the $L_1(\mathbb{X}, \lambda)$ -distance between their Radon-Nikodym derivatives w.r.t. a measure λ , where λ should be chosen in such a way that respective Radon-Nikodym derivatives exist.

An interesting and useful fact is that the total variation distance, defined in an “analytical” way, admits another characterisation in “probabilistic” terms. Consider any pair of \mathbb{X} -valued random variables ξ and η , defined on the same probability space, such that $\xi \sim \mu, \eta \sim \nu$. Let $\mathbb{X} = A_+ \cup A_-$ be the Hahn decomposition for $\mu - \nu$, then

$$\begin{aligned} \|\mu - \nu\|_{\text{TV}} &= (\mu(A_+) - \nu(A_+)) - (\mu(A_-) - \nu(A_-)) \\ &= (2\mu(A_+) - 1) - (2\nu(A_+) - 1) = 2(\mu(A_+) - \nu(A_+)). \end{aligned}$$

Because $\xi \sim \mu$ and $\eta \sim \nu$, we then have

$$\begin{aligned} \|\mu - \nu\|_{\text{TV}} &= 2(\mu(A_+) - \nu(A_+)) = 2\mathbb{E}(\mathbb{I}_{\xi \in A_+} - \mathbb{I}_{\eta \in A_+}) \\ &\leq 2\mathbb{P}(\xi \in A_+, \eta \notin A_+) \leq 2\mathbb{P}(\xi \neq \eta). \end{aligned} \tag{1.1}$$

The above inequality offers a convenient probabilistic tool to bound the total variation distance $\|\mu - \nu\|_{\text{TV}}$: one should construct properly a “couple of representatives” (ξ, η) for the measures μ, ν and then use the bound (1.1).

An important observation is that the above procedure is exact, in the sense that a proper choice of (ξ, η) turns inequality (1.1) into an identity. To formulate this result, frequently called “the Coupling Lemma” in the literature, we denote by $\mathcal{C}(\mu, \nu)$ the set of pairs of random variables (ξ, η) , defined on the same probability space such that $\xi \sim \mu$ and $\eta \sim \nu$. In what follows, we call any representative from $\mathcal{C}(\mu, \nu)$ a *coupling* for the measures μ, ν .

Theorem 1.2

$$\|\mu - \nu\|_{\text{TV}} = 2 \min_{(\xi, \eta) \in \mathcal{C}(\mu, \nu)} \mathbb{P}(\xi \neq \eta). \tag{1.2}$$

Proof. We construct explicitly the joint law κ of $(\xi, \eta) \in \mathcal{C}(\mu, \nu)$, which turns inequality (1.1) into an equality. Consider a probability measure λ such that $\mu \ll \lambda$ and $\nu \ll \lambda$

and put

$$f = \frac{d\mu}{d\lambda}, \quad g = \frac{d\nu}{d\lambda}, \quad h = f \wedge g.$$

If $p = \int h \, d\lambda = 1$, then $\mu = \nu$ and we put $\kappa(A_1 \times A_2) = \mu(A_1 \cap A_2)$; that is, we set the components ξ, η equal to one another, with the law $\mu = \nu$. Otherwise, we decompose

$$\mu = p\theta + (1-p)\sigma_1, \quad \nu = p\theta + (1-p)\sigma_2, \quad d\theta = \frac{1}{p}h \, d\lambda, \quad (1.3)$$

with the convention that $\theta = \lambda$ if $p = 0$. We note that

$$\kappa(A_1 \times A_2) = p\theta(A_1 \cap A_2) + (1-p)\sigma_1(A_1)\sigma_2(A_2)$$

and observe that κ is the distribution of some $(\xi, \eta) \in \mathcal{C}(\mu, \nu)$ and

$$\kappa(\{(x, y) : x \neq y\}) \leq 1 - p.$$

To explain the above inequality, recall that in the above construction of κ , we first represent μ and ν as mixtures (1.3) with the same first part. Then we toss a coin with probability of success p , and in the case of a success, we take both components equal to the law θ ; otherwise the components are chosen independently with the laws σ_1, σ_2 . Clearly, under such a construction the probability for the components to be different does not exceed $1 - p$. Hence

$$\begin{aligned} 2\kappa(\{(x, y) : x \neq y\}) &\leq 2(1-p) = 2 \int (1-h) \, d\lambda \\ &= \int (f+g-2(f \wedge g)) \, d\lambda = \int |f-g| \, d\lambda. \quad \square \end{aligned}$$

Let us mention some commonly used terminology related to the above proof; as a reference, see e.g. [55, Section 1.4]. Representation of a probability law μ in the form $\mu = p\theta + (1-p)\sigma$ is called the *splitting representation*; any coupling which gives the minimum in (1.2) is called a *maximal coupling*. Note also that the law of a pair $(\xi, \eta) \in \mathcal{C}(\mu, \nu)$ is frequently called a “coupling” rather than the *pair* itself.

• • • • •

1.2 Uniform ergodic rate

In this section, we establish a bound for the rate of convergence of transition probabilities $P_n(x, dx')$ to a (unique) IPM π in total variation distance. This bound is called *uniform* because it holds true uniformly for all $x \in \mathbb{X}$.

Theorem 1.3 Let there be a $\rho < 1$ such that

$$\|P_1(x_1, \cdot) - P_1(x_2, \cdot)\|_{\text{TV}} \leq 2\rho, \quad x_1, x_2 \in \mathbb{X}. \quad (1.4)$$

Then

$$\|P_n(x_1, \cdot) - P_n(x_2, \cdot)\|_{\text{TV}} \leq 2\rho^n, \quad n \geq 1, x_1, x_2 \in \mathbb{X}. \quad (1.5)$$

In addition, there exists a unique IPM π for X , and

$$\|P_n(x, \cdot) - \pi\|_{\text{TV}} \leq 2\rho^n, \quad n \geq 1, x \in \mathbb{X}. \quad (1.6)$$

Here we give a “probabilistic” proof, based on a *coupling approach*. An alternative “analytical” argument will be discussed later.

Proof. For every $y = (x_1, x_2) \in \mathbb{X} \times \mathbb{X}$, denote by $Q(y, \cdot)$ the measure κ from the above proof of the Coupling Lemma, which corresponds to the pair $\mu = P(x_1, \cdot)$, $\nu = P(x_2, \cdot)$. Then, by its construction, $\{Q(y, B), y \in \mathbb{X} \times \mathbb{X}, B \in \mathcal{X} \otimes \mathcal{X}\}$ is a transition probability; that is, $Q(y, \cdot)$ is a probability measure for any y and $Q(\cdot, B)$ is a measurable function for any B .

Exercise 1.4 Verify the fact that the function $y \mapsto Q(y, B)$ is measurable.

By the construction of Q , we have

$$Q((x_1, x_2), A \times \mathbb{X}) = P(x_1, A), \quad Q((x_1, x_2), \mathbb{X} \times A) = P(x_2, A).$$

Then, if we consider a two-component Markov process $Y = (Y^1, Y^2)$ with the transition kernel Q , we have that both its components Y^1, Y^2 are Markov processes with the same one-step transition probability, which equals the initial transition probability $P(x, A)$. To show this, say, for Y^1 , we denote $\mathcal{F}_n^Y = \sigma(Y_k, k \leq n)$ and $\mathcal{F}_n^{Y^1} = \sigma(Y_k^1, k \leq n)$, $n \geq 0$ and

write

$$\mathbb{P}(Y_{n+1}^1 \in A | \mathcal{F}_n^Y) = \mathbb{P}(Y_{n+1} \in A \times \mathbb{X} | \mathcal{F}_n^Y) = Q(Y_n, A \times \mathbb{X}) = P(Y_n^1, A).$$

Because the latter expression is clearly $\mathcal{F}_n^{Y^1}$ -measurable, we get the required Markov property of Y^1 :

$$\mathbb{P}(Y_{n+1}^1 \in A | \mathcal{F}_n^{Y^1}) = P(Y_n^1, A), \quad A \in \mathcal{X}.$$

In other words, if we consider the Markov process Y with the transition kernel Q and starting point $Y_0 = (x_1, x_2)$, we get a two-component process, such that its components have prescribed distributions P_{x_1} and P_{x_2} , respectively. In particular,

$$Y_n^1 \sim P_n(x_1, \cdot), \quad Y_n^2 \sim P_n(x_2, \cdot),$$

and therefore by (1.1)

$$\|P_n(x_1, \cdot) - P_n(x_2, \cdot)\|_{\text{TV}} \leq 2\mathbb{P}(Y_n^1 \neq Y_n^2). \quad (1.7)$$

Denote by $D = \{(x, y) : x = y\}$ the “diagonal” in $\mathbb{X} \times \mathbb{X}$, and set $q(y) = Q(y, D)$, $y \in \mathbb{X} \times \mathbb{X}$. Then $q(Y_n)$ equals the conditional probability w.r.t. $\mathcal{F}_n^Y = \sigma(Y_k, k \leq n)$ of the event $\{Y_{n+1} \in D\}$. By the construction of the kernel Q , we have

$$q(y) = 1 - \frac{1}{2} \|P(x_1, \cdot) - P(x_2, \cdot)\|_{\text{TV}}, \quad y = (x_1, x_2).$$

Hence

$$q(y) \begin{cases} = 1, & y \in D \\ \geq 1 - \rho, & \text{otherwise} \end{cases}.$$

In other words, the value Y_{n+1} :

- 1) stays on the diagonal if Y_n is on the diagonal;
- 2) has a conditional probability $\geq 1 - \rho$ to hit the diagonal if Y_n is not on the diagonal.

From 1) and 2) it follows that

$$\mathbb{P}(Y_n^1 \neq Y_n^2) \leq \rho^n, \quad (1.8)$$

which completes the proof of the first part of the theorem.

Exercise 1.5 Please give a formal proof of Equation (1.8).

The first part of the theorem implies, that for any given $x \in \mathbb{X}$, and any $m < n$

$$\begin{aligned} \|P_n(x, \cdot) - P_m(x, \cdot)\|_{\text{TV}} &= \left\| \int_{\mathbb{X}} (P_m(x', \cdot) - P_n(x, \cdot)) P_{n-m}(x, dx') \right\|_{\text{TV}} \\ &\leq \int_{\mathbb{X}} \|P_m(x', \cdot) - P_n(x, \cdot)\|_{\text{TV}} P_{n-m}(x, dx') \leq 2\rho^m. \end{aligned}$$

That is, the sequence $P_n(x, \cdot)$ is Cauchy w.r.t. the total variation distance, and therefore has a limit π in total variation distance. It is easy to check that π is an IPM for X .

Exercise 1.6 Please verify this statement.

Similar calculations lead to the required ergodic rate:

$$\begin{aligned} \|P_n(x, \cdot) - \pi\|_{\text{TV}} &= \left\| \int_{\mathbb{X}} (P_n(x, \cdot) - P_n(x', \cdot)) \pi(dx') \right\|_{\text{TV}} \\ &\leq \int_{\mathbb{X}} \|P_n(x, \cdot) - P_n(x', \cdot)\|_{\text{TV}} \pi(dx') \leq 2\rho^n. \quad \square \end{aligned}$$

For a better understanding of the condition of Theorem 1.3, let us consider its particular version in the classical case of a finite state space \mathbb{X} . Now any $\mu \in \mathcal{P}(\mathbb{X})$ is represented by a vector $\{\mu_i\}_{i \in \mathbb{X}}$ and the total variation distance between $\mu, \nu \in \mathcal{P}(\mathbb{X})$ simply equals

$$\|\mu - \nu\|_{\text{TV}} = \sum_{i \in \mathbb{X}} |\mu_i - \nu_i|.$$

Transition probabilities of X are now represented by matrices

$$\left\{ p_{ij}^n = P_n(i, \{j\}) \right\}_{i, j \in \mathbb{X}}, \quad n \geq 1.$$

Denote by $p_{ij} = p_{ij}^1$, $i, j \in \mathbb{X}$, the respective one-step transition probabilities. The following statement is now a direct consequence of Theorem 1.3 above.

Theorem 1.7 Assume that $\text{card } \mathbb{X} < \infty$ and

$$\sum_j |p_{ij} - p_{i'j}| < 2, \quad i, i' \in \mathbb{X}. \quad (1.9)$$

Then there exists a unique IPM π for X and

$$\sum_{j \in \mathbb{X}} |p_{ij}^n - \pi_j| \leq 2\rho^n, \quad n \geq 1, i \in \mathbb{X} \quad (1.10)$$

for some $\rho < 1$.

This statement definitely deserves a separate discussion. First, observe that the following simpler condition is *sufficient* for (1.9):

$$\text{there exists } j_* \in \mathbb{X} \text{ such that } p_{ij_*} > 0, \quad i \in \mathbb{X}. \quad (1.11)$$

This is actually the condition used by A.A. Markov in his seminal paper [40] in order to prove both the “stabilization of the law” property and the law of large numbers for a time-homogeneous Markov chain (in the modern terminology) with a finite state space. To better understand the sense of this condition, let us consider its following N -step version:

$$\text{there exist } j_* \in \mathbb{X}, N \geq 1 \text{ such that } p_{ij_*}^N > 0, \quad i \in \mathbb{X}. \quad (1.12)$$

Clearly, (1.12) is *necessary* for (1.10) (and hence for (1.9)) to hold true; moreover, one can verify that (1.12) is *necessary and sufficient* for the following slightly weaker version of (1.10):

$$\text{there exist } C > 1, \rho \in (0, 1) : \sum_{j \in \mathbb{X}} |p_{ij}^n - \pi_j| \leq C\rho^n, \quad n \geq 1, i \in \mathbb{X} \quad (1.13)$$

Exercise 1.8 Please verify this statement and give an example showing that (1.12) does not necessarily imply (1.10).

Condition (1.12) admits the following transparent re-arrangement in terms of the classical “classification of states” terminology. Namely, (1.12) fails if, and only if, there exist in the state space \mathbb{X} at least two (or more) non-connected classes of states (please verify this!). That is why conditions like (1.9), (1.11) and (1.12) have a natural interpretation as *irreducibility conditions* on the chain X .

Coming back to our Theorem 1.3, where the structure of the state space is not specified, we emphasize that now the “classification of states” becomes much more cumbersome and non-transparent. Hence it is very practical that condition (1.4), which of course

still has an interpretation as an irreducibility condition, is formulated straightforwardly in terms of the initial transition probability. The “probabilistic” method of proof, based on the *coupling construction*, goes back to pioneering works by W. Döblin at the end of 1930s [10, 11], although it should be mentioned that Döblin used different forms of the irreducibility condition. The statement close to the one we give in Theorem 1.3 was proved by R. Dobrushin in 1956 [9], hence in the sequel we call (1.4) *the Dobrushin condition*. Note that some authors use the name *Markov-Dobrushin condition* instead, e.g. [18].

Note that Dobrushin’s original proof in [9] was completely different from the “coupling” proof we gave before: it was based mainly on analytical ideas similar e.g. to those used already by Markov in [40]. Below we give a sketch of this “analytical” proof, which is both simple and instructive.

Another proof of Theorem 1.3. Denote by $\mathcal{M}(\mathbb{X})$ the family of finite signed measures on \mathbb{X} . It is well known that $\mathcal{M}(\mathbb{X})$ is a Banach space w.r.t. the total variation norm $\|\mu\| = \mu_+(\mathbb{X}) + \mu_-(\mathbb{X})$ (where $\mu = \mu_+ - \mu_-$ is the Hahn decomposition for μ). Clearly, $\mathcal{P}(\mathbb{X}) \subset \mathcal{M}(\mathbb{X})$ and $\|s \cdot\|$ coincides with $\|\cdot\|_{\text{TV}}$ on $\mathcal{P}(\mathbb{X})$. The Banach space $\mathcal{M}(\mathbb{X})$ is the dual one to the Banach space $\mathcal{B}(\mathbb{X})$ of bounded measurable real-valued functions on \mathbb{X} with the sup-norm:

$$\|f\| = \sup_{\|\mu\|=1} |\langle f, \mu \rangle|, \quad \|\mu\| = \sup_{\|f\|=1} |\langle f, \mu \rangle|;$$

here and below, we denote

$$\langle f, \mu \rangle = \int_{\mathbb{X}} f(x) \mu(dx).$$

Define the linear operator $P : \mathcal{B}(\mathbb{X}) \rightarrow \mathcal{B}(\mathbb{X})$ by

$$Pf(x) = \int_{\mathbb{X}} f(x') P(x, dx');$$

then the adjoint operator $P^* : \mathcal{M}(\mathbb{X}) \rightarrow \mathcal{M}(\mathbb{X})$ is given by

$$P^* \mu(A) = \int_{\mathbb{X}} P(x, A) \mu(dx), \quad A \in \mathcal{X}.$$

In particular, the n -step transition probability $P^n(x, \cdot)$ for X is just the image of the delta measure δ_x under $(P^*)^n$.

The key idea of the proof is to show that for any $\mu, \nu \in \mathcal{P}(\mathbb{X})$

$$\|P^*\mu - P^*\nu\| \leq \rho \|\mu - \nu\|. \quad (1.14)$$

This would immediately imply that (1.5) holds true, because then

$$\|P^n(x_1, \cdot) - P^n(x_2, \cdot)\| \leq \rho^n \|\delta_{x_1} - \delta_{x_2}\| = \begin{cases} 2\rho^n, & x_1 \neq x_2 \\ 0, & x_1 = x_2 \end{cases}.$$

The rest of the proof then will be the same as in the proof of Theorem 1.3. In other words, the aim is to prove that P^* is a contraction and then one can use essentially the Banach fixed point theorem. Note that P^* is *not* a contraction when considered on the entire $\mathcal{M}(\mathbb{X})$.

Exercise 1.9 Give an example of two measures $\mu, \nu \in \mathcal{M}(\mathbb{X})$ such that (1.14) for them fails.

The latter observation gives a hint that, while proving (1.14), we need to use the additional property of μ and ν being probability measures (in fact, we will need only that $\mu(\mathbb{X}) = \nu(\mathbb{X})$).

We have

$$\langle f, P^*\mu \rangle = \int_{\mathbb{X}} \int_{\mathbb{X}} f(x') P(x, dx') \mu(dx) = \langle Pf, \mu \rangle,$$

hence

$$\|P^*\mu - P^*\nu\| = \sup_{\|f\|=1} |\langle Pf, \mu - \nu \rangle|.$$

For $g \in \mathbb{B}(\mathbb{X})$, denote $g_s(x) = g(x) - s$, $x \in \mathbb{X}$, $s \in \mathbb{R}$, and observe that

$$\langle g, \mu - \nu \rangle = \langle g_s, \mu - \nu \rangle, \quad s \in \mathbb{R},$$

and therefore

$$|\langle g, \mu - \nu \rangle| \leq \|\mu - \nu\| \inf_{s \in \mathbb{R}} \|g_s\|.$$

Observe that

$$\inf_{s \in \mathbb{R}} \|g_s\| = \frac{1}{2} \sup_{x_1, x_2 \in \mathbb{X}} |g(x_1) - g(x_2)|.$$

Exercise 1.10 Prove this identity.

Summarizing the above relations, we get

$$\|P^*\mu - P^*\nu\| = \frac{1}{2} \|\mu - \nu\| \sup_{\|f\|=1} \sup_{x_1, x_2 \in \mathbb{X}} |Pf(x_1) - Pf(x_2)|.$$

But by the condition (1.4)

$$|Pf(x_1) - Pf(x_2)| = |\langle f, P(x_1, \cdot) - P(x_2, \cdot) \rangle| \leq \|f\| \|P(x_1, \cdot) - P(x_2, \cdot)\| \leq 2\rho \|f\|,$$

which completes the proof of (1.14). □

.

1.3 Non-uniform ergodic rates

The theory exposed in the previous section is very simple and self-contained but it is not applicable for many particular cases of interest. The reason is that, for a Markov process X with a non-compact state space (like \mathbb{R}, \mathbb{R}^m , etc.), it is non-typical to have a *uniform* bound similar to (1.6). To explain this, let us consider a typical example.

Example 1.11 Let $X_t, t \geq 0$, be an Ornstein-Uhlenbeck process, i.e. the solution to the linear SDE

$$dX_t = -aX_t dt + dW_t.$$

It is a Markov process and its transition probabilities are given explicitly as

$$P_t(x, \cdot) \sim \mathcal{N}\left(e^{-at}x, \frac{1}{2a}\left(1 - e^{-2at}\right)\right).$$

The latter formula follows from the explicit expression for the solution

$$X_t = e^{-at}X_0 + \int_0^t e^{-at+as} dW_s, \quad t \geq 0$$

(please use the Itô formula to verify the latter identity). Then it is an explicit calculation to show that for a given t

$$\|P_t(x, \cdot) - P_t(0, \cdot)\|_{TV} \rightarrow 2, \quad x \rightarrow \infty,$$

hence

$$\sup_{x_1, x_2} \|P_t(x_1, \cdot) - P_t(x_2, \cdot)\|_{\text{TV}} \equiv 2. \quad (1.15)$$

Relation (1.15) is quite typical and corresponds to a situation which can informally be explained in the following way: for “very distant” starting points of X , it should take a large time for the process to come back “close to the origin”; hence any bound for the total variation distance between the transition probabilities, which is uniform w.r.t. the initial value, is necessarily trivial and non-informative. This is the reason why for processes valued in non-compact state spaces, it is not typical to satisfy *uniform* irreducibility conditions, e.g. (1.4). This is the main motivation for the following definition.

Definition 1.12 A Markov chain X satisfies the (*local*) *Dobrushin condition on the set* K , if

$$\rho(K) = \sup_{x_1, x_2 \in K} \left(\frac{1}{2} \|P(x_1, \cdot) - P(x_2, \cdot)\|_{\text{TV}} \right) < 1.$$

Our goal in this section is to explain how a *local irreducibility* condition, combined with a *recurrence* condition (which will be discussed later on), produce a non-uniform ergodic rate bound of the form

$$\|P_n(x_1, \cdot) - P_n(x_2, \cdot)\|_{\text{TV}} \leq r_n(V(x_1) + V(x_2)), \quad x_1, x_2 \in \mathbb{X}, n \geq 1. \quad (1.16)$$

Here $r_n, n \geq 1$, corresponds to the rate of convergence, and V has a natural interpretation as a “penalty”, which should be large for “distant” starting points.

Before formulating exact statements, let us explain the main idea on which the whole approach is based. The “probabilistic” proof of Theorem 1.3 above was based on the relation (1.7), which follows from the Coupling Lemma and holds true for *any* two-component process $Y = (Y^1, Y^2)$ such that its components have prescribed distributions P_{x_1} and P_{x_2} , respectively. Such a process is typically called “a coupling process” or simply “a coupling” for the initial Markov process X . The key question in this approach is how to construct a coupling process Y which would admit explicit bounds for the probability that Y_n hits the diagonal (i.e., for the components Y_n^1, Y_n^2 to be “coupled”).

In Theorem 1.3, the coupling process Y was built in such a way that, in every step, a “coupling attempt” is performed with the conditional probability of success $\geq 1 - \rho$. Now this can be done with the conditional probability of a success $\geq 1 - \rho(K)$ *only* if at the given time, both components of the process Y belong to K . So instead of (1.8), we

will have a bound involving the (random) time spent by the coupling process Y at the set K and one will need to control the “recurrence” properties of the coupling process Y , e.g., the distribution of the time required for Y to hit $K \times K$.

A strategy to combine local irreducibility with recurrence conditions dates back to T. Harris [24] and gives rise to a wide range of theorems, frequently called *Harris-type theorems*. Below we give several statements of that kind. In order to explain the arguments more transparently, we separate the exposition into two parts. The first part contains a general Harris-type theorem with the recurrence condition imposed on an auxiliary two-component Markov chain Z . In the second part, we show how this condition can easily be verified in terms of the initial Markov process X .

Denote by Λ the class of continuous monotone functions $\lambda : [1, \infty) \rightarrow [1, \infty)$ with $\lambda(1) = 1$ and $\lambda(\infty) = \infty$ which are *sub-multiplicative* in the sense that

$$\lambda(t+s) \leq \lambda(t)\lambda(s), \quad s, t \in [1, \infty).$$

Next, for a given Markov chain X , we call a *Markov coupling* for X any time-homogeneous Markov chain Z on $\mathbb{X} \times \mathbb{X}$, such that for a corresponding transition probability $P^Z((x_1, x_2), \cdot)$ the marginal distributions equal $P^X(x_1, \cdot)$ and $P^X(x_2, \cdot)$, respectively. For any such Z , we denote by $P_{(x_1, x_2)}^Z$ the distribution of Z with $Z_0 = (x_1, x_2)$, by $E_{(x_1, x_2)}^Z$ the respective expectation and by $\tau_C^Z = \inf\{n \geq 0 : Z_n \in C\}$ the hitting time by Z of a measurable set $C \subset \mathbb{X} \times \mathbb{X}$.

Theorem 1.13 Let a set $K \in \mathcal{X}$ and functions $V : \mathbb{X} \rightarrow [1, \infty)$, $\lambda \in \Lambda$, be such that the following conditions hold true.

I. X satisfies the Dobrushin condition on the set K .

R.1. There exists a Markov coupling Z for X such that

$$E_{(x_1, x_2)}^Z \lambda(\tau_{K \times K}^Z) \leq V(x_1) + V(x_2), \quad x_1, x_2 \in \mathbb{X}.$$

R.2. $Q(K) := \sup_{x \in K} E_x V(X_1) < \infty$.

Then for every $p > 1$ there exist $\gamma \in (0, 1]$ and C such that (1.16) holds true with

$$r_n = \frac{C}{\lambda^{1/p}(\gamma n)}.$$

Consequently, there exists a unique IPM π for X . This measure satisfies

$$\int_{\mathbb{X}} V \, d\pi < \infty, \quad (1.17)$$

and

$$\|P_n(x, \cdot) - \pi\|_{\text{TV}} \leq r_n \left(V(x) + \int_{\mathbb{X}} V \, d\pi \right), \quad x \in \mathbb{X}, n \geq 1. \quad (1.18)$$

Before proceeding with the proof, let us explain the assumptions of the theorem. Condition **I** is just the *local irreducibility condition* we have discussed above. The idea behind the *recurrence* conditions **R.1**, **R.2** becomes visible when we recall that we need to control the total time spent in $K \times K$ up to the time instant n by the coupling process Y (which is yet to be constructed). The main recurrence condition **R.1** controls the λ -moment of the length of a waiting period of the process Y (that is, the time spent outside of $K \times K$) in terms of the penalty function V . Because we need to iterate the waiting period in case where the preceding coupling attempt was unsuccessful, we need to control the expectation of the penalty V at the end of the coupling period; this is made by means of the auxiliary recurrence condition **R.2**.

Proof. Let Q be the transition kernel constructed in the proof of Theorem 1.3. Consider a Markov process Y on $\mathbb{X} \times \mathbb{X}$ with the transition kernel

$$P^Y((x_1, x_2), \cdot) = \begin{cases} Q((x_1, x_2), \cdot), & (x_1, x_2) \in K \times K \text{ or } x_1 = x_2; \\ P^Z((x_1, x_2), \cdot), & \text{otherwise.} \end{cases}$$

Denote by S_k , $k \geq 1$, the time moment of the k -th visit of Y to $K \times K$ and by $R_k = S_k + 1$ the time moment when the k -th ‘‘coupling attempt’’ is finished. Denote

$$L = \min\{R_k : Y^1(R_k) = Y^2(R_k)\},$$

i.e. the ‘‘coupling time’’ for the process Y . Once the components are coupled, they are evolving together, hence

$$P(Y_n^1 \neq Y_n^2) \leq P(L > n) \leq \lambda^{-1/p}(\gamma n) E \lambda^{1/p}(\gamma L),$$

and we need to estimate $E \lambda^{1/p}(\gamma L)$. It follows from the bounds (1.21) and (1.22) below

that

$$P(L > R_k) \rightarrow 0, \quad k \rightarrow \infty.$$

Recall that $R_k = S_k + 1 \geq R_{k-1} + 1$ provided $R_{k-1} < \infty$, and therefore $R_k \rightarrow \infty$. Hence $P(L = \infty) = 0$ because, by the construction above and conditions **R.1** and **R.2**, there will be an arbitrarily large number of finite time moments S_k, R_k (i.e. of coupling attempts) before the coupling time L . Consequently, we have

$$E\lambda^{1/p}(\gamma L) = \sum_{k=1}^{\infty} E\lambda^{1/p}(\gamma L) \mathbb{I}_{L=R_k}. \quad (1.19)$$

Consider one term in the above sum:

$$E\lambda^{1/p}(\gamma L) \mathbb{I}_{L=R_k} = E\lambda^{1/p}(\gamma L) \mathbb{I}_{L=R_k} \mathbb{I}_{L>R_{k-1}} \leq E\mathbb{I}_{L>R_{k-1}} \lambda^{1/p}(\gamma R_k),$$

where we denote $R_0 = 0$. Then by the identity $R_k = S_k + 1$ and the sub-multiplicativity of λ , we have

$$E\mathbb{I}_{L>R_{k-1}} \lambda^{1/p}(\gamma R_k) \leq \lambda^{1/p}(\gamma) E\left[\mathbb{I}_{L>R_{k-1}} \lambda^{1/p}(\gamma R_{k-1}) \times \left(E_{R_{k-1}} \lambda^{1/p}(\gamma(S_k - R_{k-1})) \right) \right]. \quad (1.20)$$

Here by $E_{R_{k-1}}$ we denote the conditional expectation w.r.t. $\mathcal{F}_{R_{k-1}}$, where $\{\mathcal{F}_t, t \geq 0\}$ is the natural filtration of the process Y . Let us study the term

$$E_{R_{k-1}} \lambda^{1/p}(\gamma(S_k - R_{k-1}))$$

in more detail. On the time interval $[R_{k-1}, S_k]$, the process Y behaves as the process $Z = (Z^1, Z^2)$ with independent components, which starts from the point $Y(R_{k-1})$. Then by the condition **R.1**,

$$E_{R_{k-1}} \lambda(S_k - R_{k-1}) \leq V(Y^1(R_{k-1})) + V(Y^2(R_{k-1})) =: A_{k-1}.$$

Because $\gamma \leq 1$ and λ is monotone, then using the above bound, the Hölder inequality, and the Chebyshev inequality, we can write for any $p_1 \in (1, p)$

$$\begin{aligned}
\mathbb{E}_{R_{k-1}} \lambda^{1/p} (\gamma(S_k - R_{k-1})) &\leq \left(\mathbb{E}_{R_{k-1}} \lambda^{p_1/p} (\gamma(S_k - R_{k-1})) \right)^{1/p_1} \\
&\leq \left(\lambda^{p_1/p} (\gamma^{1/2}) + \mathbb{E}_{R_{k-1}} \left(\lambda^{p_1/p} (S_k - R_{k-1}) \mathbb{I}_{S_k - R_{k-1} > \gamma^{-1/2}} \right) \right)^{1/p_1} \\
&\leq \left(\lambda^{p_1/p} (\gamma^{1/2}) + \lambda^{-1+p_1/p} (\gamma^{-1/2}) A_{k-1} \right)^{1/p_1} =: B_{k-1}
\end{aligned}$$

Coming back to the bound (1.20), we can now write for $k > 1$

$$\begin{aligned}
\mathbb{E} \mathbb{I}_{L > R_{k-1}} \lambda^{1/p} (\gamma R_k) &\leq \lambda^{1/p} (\gamma) \mathbb{E} \left(\mathbb{I}_{L > R_{k-1}} \lambda^{1/p} (\gamma R_{k-1}) B_{k-1} \right) \\
&= \lambda^{1/p} (\gamma) \mathbb{E} \left(\mathbb{I}_{L > R_{k-2}} \lambda^{1/p} (\gamma R_{k-1}) \mathbb{E}_{S_{k-1}} \left(\mathbb{I}_{L > R_{k-1}} B_{k-1} \right) \right).
\end{aligned}$$

Here we have used that $R_{k-1} = S_{k-1} + 1$ is $\mathcal{F}_{S_{k-1}}$ -measurable. A coupling attempt is made on the time interval $[S_{k-1}, R_{k-1}]$ and because $Y(S_{k-1}) \in K \times K$, we have on the set $\{L > R_{k-2}\} = \{L > S_{k-1}\}$

$$\mathbb{E}_{S_{k-1}} \mathbb{I}_{L > R_{k-1}} \leq \rho(K).$$

On the other hand, recalling the formulae for B_{k-1} and A_{k-1} and using the elementary inequality $(a+b)^{p_1} \leq 2^{p_1-1}(a^{p_1} + b^{p_1})$, we get

$$\begin{aligned}
\mathbb{E}_{S_{k-1}} B_{k-1}^{p_1} &\leq \lambda^{p_1/p} (\gamma^{1/2}) + \lambda^{-1+p_1/p} (\gamma^{-1/2}) \\
&\quad \times \mathbb{E}_{S_{k-1}} \left(V(Y^1(S_{k-1} + 1)) + V(Y^2(S_{k-1} + 1)) \right) \\
&\leq \lambda^{p_1/p} (\gamma^{1/2}) + \lambda^{-1+p_1/p} (\gamma^{-1/2}) (2Q(K)).
\end{aligned}$$

In the last inequality we have used condition **R.1** and the fact that, by construction, the conditional distributions of $Y^{1,2}(S_{k-1} + 1)$ w.r.t. $\mathcal{F}_{S_{k-1}}$ equal $P(Y^{1,2}(S_{k-1}), \cdot)$.

Then by the Hölder inequality

$$\mathbb{E}_{S_{k-1}} \left(\mathbb{I}_{L > R_{k-1}} B_{k-1} \right) \leq \rho^{1-1/p_1}(K) \left(\lambda^{p_1/p} (\gamma^{1/2}) + \lambda^{-1+p_1/p} (\gamma^{-1/2}) (2Q(T, K)) \right)^{1/p_1},$$

and therefore, for every $k > 1$,

$$\mathbb{E} \mathbb{I}_{L > R_{k-1}} \lambda^{1/p} (\gamma R_k) \leq \theta(\gamma) \mathbb{E} \mathbb{I}_{L > R_{k-2}} \lambda^{1/p} (\gamma R_{k-1}) \leq \dots \leq \theta^{k-1}(\gamma) \mathbb{E} \lambda^{1/p} (\gamma R_1) \quad (1.21)$$

with

$$\theta(\gamma) = \lambda^{1/p}(\gamma) \left(\lambda^{p_1/p}(\gamma^{1/2}) + \lambda^{-1+p_1/p}(\gamma^{-1/2}) (2Q(K)) \right)^{1/p_1} \rho^{1-1/p_1}(K).$$

For $k = 1$, we have simply

$$\mathbb{E} \lambda^{1/p}(\gamma R_1) \leq \lambda^{1/p}(\gamma) \mathbb{E} \lambda^{1/p}(\gamma S_1) \leq \lambda^{1/p}(\gamma) (V(x_1) + V(x_2))^{1/p} \quad (1.22)$$

by the condition **R.1**.

Now, we proceed in the following way. Fix $p_1 \in (1, p)$; for instance, $p_1 = \sqrt{p}$. Because $\rho(K) < 1$ and

$$\lambda^{1/p}(\gamma) \left(\lambda^{p_1/p}(\gamma^{1/2}) + \lambda^{-1+p_1/p}(\gamma^{-1/2}) (2Q(K)) \right)^{1/p_1} \rightarrow 1, \quad \gamma \rightarrow 0,$$

there exists $\gamma \in (0, 1]$ such that

$$\theta(\gamma) < \rho^{1/2-1/2p_1}(K).$$

Using (1.19) – (1.22), we deduce the required bound (1.16) with r_n given in the statement of the theorem, γ chosen above and

$$C = \frac{\lambda^{1/p}(\gamma)}{1 - \rho^{1/2-1/2p_1}(K)}.$$

We can now finalize the proof. Fix a point $x_0 \in K$ and denote $\pi_n(dx) = P_n(x_0, dx)$, $n \geq 1$. Then by (1.16) and condition **R.2** we have, for any $n \geq m$,

$$\begin{aligned} \|\pi_n - \pi_m\|_{\text{TV}} &= \left\| \int_{\mathbb{X}} (P_m(x, \cdot) - P_m(x_0, \cdot)) P_{n-m}(x_0, dx) \right\|_{\text{TV}} \\ &\leq \int_{\mathbb{X}} \|P_m(x, \cdot) - P_m(x_0, \cdot)\|_{\text{TV}} P_{n-m}(x_0, dx) \\ &\leq C \rho^m \int_{\mathbb{X}} (V(x) + V(x_0)) P_{n-m}(x_0, dx) \leq C(Q(K) + V(x_0)) \rho_m. \end{aligned}$$

Hence the sequence $\{\pi_n\}$ is Cauchy in the total variation distance, and therefore there exists a limit $\pi = \lim_{n \rightarrow \infty} \pi_n$. The following properties are then easy to verify:

- ◇ π is an IPM for X ;

◇ π is the unique IPM for X ;

◇ π satisfies (1.17).

Exercise 1.14 Prove the above properties.

Then, using (1.16) once more, we get (1.18). \square

Theorem 1.13 is very general, but usually it would be more convenient to have a sufficient condition for (1.16) in a simpler form, which would only involve the distribution of the process X . One very practical way to do this is exposed in the following theorem. The principal assumption (1.23) therein is very close, both in form and spirit, to the classical Lyapunov condition in the theory of stability of ordinary differential equations; hence, it is usually called a *Lyapunov-type condition*.

Theorem 1.15 Assume that, for a given Markov chain X , functions $V : \mathbb{X} \rightarrow [1, +\infty)$ and $\phi : [1, +\infty) \rightarrow (0, \infty)$ satisfy the following:

1) there exists some $C > 0$ such that

$$E_x V(X_1) - V(x) \leq -\phi(V(x)) + C, \quad x \in \mathbb{X}; \quad (1.23)$$

2) ϕ admits a non-negative, increasing and concave extension to $[0, +\infty)$;

3) V is bounded on K , and

$$\phi \left(1 + \inf_{x \notin K} V(x) \right) > 2C \quad (1.24)$$

with C which comes from (1.23).

Then any Markov coupling Z for the chain X satisfies condition **R.1** of Theorem 1.13 with

$$\lambda(t) = \Phi^{-1}(\alpha t),$$

where

$$\alpha = 1 - \frac{2C}{\phi \left(1 + \inf_{x \notin K} V(x) \right)},$$

and Φ^{-1} denotes the inverse function to

$$\Phi(v) = \int_1^v \frac{1}{\phi(w)} dw, \quad v \in [1, \infty). \quad (1.25)$$

In addition, the function λ belongs to the class Λ , and the chain X satisfies condition **R.2** of Theorem 1.13.

Proof. Denote

$$H(t, v) = \Phi^{-1}(t + \Phi(v)), \quad t \geq 0, v \geq 1. \quad (1.26)$$

It would be a key point in the proof of **R.1** to show that, for every $n \geq 1$,

$$\mathbb{E}_{(x_1, x_2)}^Z \left[H(\alpha n, V(Z_n^1) + V(Z_n^2)) \middle| \mathcal{F}_{n-1}^Z \right] \leq H(\alpha(n-1), V(Z_{n-1}^1) + V(Z_{n-1}^2)) \quad (1.27)$$

a.s. on the set $\{Z_{n-1} \notin K \times K\}$, where $\{\mathcal{F}_n\}$ denotes the natural filtration of Z . Once we manage to do this, we would have

$$\mathbb{E}_{(x_1, x_2)}^Z H(\alpha(\tau \wedge n), V(Z_{\tau \wedge n}^1) + V(Z_{\tau \wedge n}^2)) \leq V(x_1) + V(x_2), \quad n \geq 0 \quad (1.28)$$

with $\tau = \tau_{K \times K}^Z$.

Exercise 1.16 Prove (1.28) using (1.27).

Because

$$H(\alpha t, v) \geq \Phi^{-1}(\alpha t) = \lambda(t),$$

this would imply by the Fatou's lemma the required bound

$$\mathbb{E}_{(x_1, x_2)}^Z \lambda(\tau) \leq V(x_1) + V(x_2).$$

Observe the following properties of the function $H(t, v)$:

- ◇ $H_t'(t, v) = \phi(H(t, v))$;
- ◇ $H_v'(t, v) = \phi(H(t, v))/\phi(v)$;
- ◇ $H(t, v)$ is concave w.r.t. the variable v .

The first two properties are verified straightforwardly. To prove the third one, if ϕ is additionally assumed to be smooth, we write

$$H_{vv}''(t, v) = \frac{\phi'(H(t, v))\phi(H(t, v)) - \phi(H(t, v))\phi'(v)}{\phi^2(v)} \leq 0,$$

because $H(t, v) \geq v$ and ϕ' is decreasing. For non-smooth ϕ we can approximate it by a sequence of smooth ϕ_n ; this shows us that $H(t, v)$ is concave w.r.t. the variable v as the pointwise limit of a sequence of concave functions.

Then we have

$$H(t_2, v) - H(t_1, v) \leq (t_2 - t_1)\phi(H(t_2, v)), \quad t_1 \leq t_2, \quad (1.29)$$

and

$$H(t, v_2) - H(t, v_1) \leq \frac{\phi(H(t, v_1))}{\phi(v_1)}(v_2 - v_1), \quad v_1, v_2 \geq 1. \quad (1.30)$$

Now we can proceed with the proof of (1.27) on the set $\{Z_{n-1} \notin K \times K\}$. Write

$$\begin{aligned} & H(\alpha n, V(Z_n^1) + V(Z_n^2)) - H(\alpha(n-1), V(Z_{n-1}^1) + V(Z_{n-1}^2)) \\ &= \left[H(\alpha n, V(Z_{n-1}^1) + V(Z_{n-1}^2)) - H(\alpha(n-1), V(Z_{n-1}^1) + V(Z_{n-1}^2)) \right] \\ & \quad + \left[H(\alpha n, V(Z_n^1) + V(Z_n^2)) - H(\alpha n, V(Z_{n-1}^1) + V(Z_{n-1}^2)) \right] =: \Delta_1 + \Delta_2. \end{aligned}$$

By (1.29), we have

$$\Delta_1 \leq \alpha \phi\left(H\left(\alpha n, V(Z_{n-1}^1) + V(Z_{n-1}^2)\right)\right),$$

which is \mathcal{F}_{n-1}^Z -measurable. By (1.30), we have

$$\Delta_2 \leq \frac{\phi\left(H\left(\alpha n, V(Z_{n-1}^1) + V(Z_{n-1}^2)\right)\right)}{\phi\left(V(Z_{n-1}^1) + V(Z_{n-1}^2)\right)} \left(V(Z_n^1) + V(Z_n^2) - V(Z_{n-1}^1) - V(Z_{n-1}^2) \right).$$

Because Z is a Markov coupling for X , we have

$$\mathbb{E}_{(x_1, x_2)}^Z \left[V(Z_n^i) - V(Z_{n-1}^i) \mid \mathcal{F}_{n-1}^Z \right] = \left(\mathbb{E}_x^X V(X_1) - V(x) \right) \Big|_{x=Z_{n-1}^i}, \quad i = 1, 2,$$

and therefore by the Lyapunov-type condition (1.23)

$$\mathbb{E}_{(x_1, x_2)}^Z \left[\Delta_2 \mid \mathcal{F}_{n-1}^Z \right] \leq \frac{\phi\left(H\left(\alpha n, V(Z_{n-1}^1) + V(Z_{n-1}^2)\right)\right)}{\phi\left(V(Z_{n-1}^1) + V(Z_{n-1}^2)\right)} \times$$

$$\times \left(2C - \phi \left(V(Z_{n-1}^1) \right) - \phi \left(V(Z_{n-1}^2) \right) \right).$$

Hence we have

$$\begin{aligned} \mathbb{E}_{(x_1, x_2)}^Z \left[\Delta_1 + \Delta_2 \middle| \mathcal{F}_{n-1}^Z \right] &\leq \phi \left(H \left(\alpha n, V(Z_{n-1}^1) + V(Z_{n-1}^2) \right) \right) \\ &\times \left(\alpha + \frac{2C}{\phi \left(V(Z_{n-1}^1) + V(Z_{n-1}^2) \right)} - \frac{\phi \left(V(Z_{n-1}^1) \right) + \phi \left(V(Z_{n-1}^2) \right)}{\phi \left(V(Z_{n-1}^1) + V(Z_{n-1}^2) \right)} \right). \end{aligned}$$

On the set $\{Z_{n-1} \notin K \times K\}$ at least one term in the sum $V(Z_{n-1}^1) + V(Z_{n-1}^2)$ is not less than $\inf_{x \notin K} V(x)$, while the other term is not less than 1. Hence we have

$$\frac{2C}{\phi \left(V(Z_{n-1}^1) + V(Z_{n-1}^2) \right)} \leq \frac{2C}{\phi \left(1 + \inf_{x \notin K} V(x) \right)} = 1 - \alpha.$$

Finally, observe that because ϕ admits a concave non-negative extension $\tilde{\phi}$ to $[0, \infty)$, we have

$$\phi(v_1 + v_2) - \phi(v_1) \leq \phi(v_2) - \tilde{\phi}(0) \leq \phi(v_2), \quad v_1, v_2 \geq 1;$$

hence

$$\frac{\phi(v_1) + \phi(v_2)}{\phi(v_1 + v_2)} \geq 1, \quad v_1, v_2 \geq 1.$$

Summarizing the above inequalities, we obtain

$$\mathbb{E}_{(x_1, x_2)}^Z \left[\Delta_1 + \Delta_2 \middle| \mathcal{F}_{n-1}^Z \right] \leq 0,$$

which completes the proof of (1.27).

To verify that λ belongs to Λ , observe the following.

- ◇ $\Phi(1) = 0$, hence $\lambda(0) = 1$.
- ◇ Since ϕ is concave, ϕ possesses a linear growth bound; and therefore, $\Phi(\infty) = \infty$. This implies that $\lambda(\infty) = \infty$.
- ◇ To prove the sub-multiplicativity property of λ , it is sufficient to verify that for every fixed $s \geq 0$

$$\frac{d}{dt} \left(\frac{\lambda(t+s)}{\lambda(s)} \right) \leq 0. \tag{1.31}$$

Because ϕ has a non-negative convex extension to $[0, \infty)$, one has

$$\phi(b)a \leq \phi(a)b, \quad b \geq a. \quad (1.32)$$

Calculating the derivatives straightforwardly and using (1.32), one can easily verify (1.31).

Exercise 1.17 Prove (1.32) and verify (1.31).

Finally, condition **R.2** follows trivially from (1.23) because V is assumed to be bounded on K . \square

Let us finish this section by several particularly important corollaries of Theorem 1.13 and Theorem 1.15, which give respectively *exponential*, *polynomial* and *sub-exponential* ergodic rates for a Markov process X . The form in which we formulate these corollaries is motivated by the following observation: if X satisfies (1.23) with $\phi(\infty) = \infty$, then (1.24) would hold true when one chooses K equal to a *level set* $\{x : V(x) \leq c\}$ of the function V with sufficiently large c .

Theorem 1.18 (On exponential ergodic rate) Let there exist, for a given Markov chain X , a function $V : \mathbb{X} \rightarrow [1, +\infty)$ such that X satisfies the Dobrushin condition on every level set $\{x : V(x) \leq c\}$ of the function V , and for some $a, C > 0$

$$\mathbb{E}_x V(X_1) - V(x) \leq -aV(x) + C, \quad x \in \mathbb{X}. \quad (1.33)$$

Then there exist $c_1, c_2 > 0$ such that

$$\|P_n(x_1, \cdot) - P_n(x_2, \cdot)\|_{\text{TV}} \leq c_1 e^{-c_2 n} (V(x_1) + V(x_2)), \quad x_1, x_2 \in \mathbb{X}, n \geq 1. \quad (1.34)$$

In addition, there exists a unique IPM π for X which satisfies (1.17), and

$$\|P_n(x, \cdot) - \pi\|_{\text{TV}} \leq c_1 e^{-c_2 n} \left(V(x) + \int_{\mathbb{X}} V d\pi \right), \quad x \in \mathbb{X}, n \geq 1. \quad (1.35)$$

Proof. The required statements follow directly from Theorem 1.13 and Theorem 1.15, applied to $\phi(v) = av$ and $K = \{x : V(x) \leq c\}$ with sufficiently large c . Now $\phi(v) = av$, $v \geq 0$ is a concave increasing function with $\phi(0) = 0$ and $\phi(\infty) = \infty$. Straightforward

calculations show that

$$\Phi(v) = \frac{1}{a} \log v, \quad \lambda(t) = \Phi^{-1}(\alpha t) = e^{a\alpha t},$$

where α depends on the choice of the level c in the definition of the set K . Consequently, the term $\lambda^{-1/p}(\gamma n)$, which appears in the expression for the ergodic rate in Theorem 1.13, now equals

$$\lambda^{-1/p}(\gamma n) = e^{-c_2 n}, \quad c_2 := a\alpha\gamma/p. \quad \square$$

Theorem 1.19 (On polynomial ergodic rate) For a given Markov chain X , let there exist a function $V : \mathbb{X} \rightarrow [1, +\infty)$ such that X satisfies the Dobrushin condition on every level set $\{x : V(x) \leq c\}$ of the function V , and for some $a, C > 0$ and $\sigma \in (0, 1)$

$$E_x V(X_1) - V(x) \leq -aV^\sigma(x) + C, \quad x \in \mathbb{X}. \quad (1.36)$$

Then for all $\eta < 1/(1-\sigma)$, there exist $c_1, c_2 > 0$ such that

$$\|P_n(x_1, \cdot) - P_n(x_2, \cdot)\|_{\text{TV}} \leq c_1(1 + c_2 n)^{-\eta} (V(x_1) + V(x_2)), \quad x_1, x_2 \in \mathbb{X}, n \geq 1. \quad (1.37)$$

In addition, there exists a unique IPM π for X which satisfies (1.17), and

$$\|P_n(x, \cdot) - \pi\|_{\text{TV}} \leq c_1(1 + c_2 n)^{-\eta} \left(V(x) + \int_{\mathbb{X}} V d\pi \right), \quad x \in \mathbb{X}, n \geq 1. \quad (1.38)$$

Proof. Again, we apply Theorem 1.13 and Theorem 1.15 with $\phi(v) = av^\sigma$ and $K = \{x : V(x) \leq c\}$ with sufficiently large c . Again, $\phi(x) = av^\sigma$, $v \geq 0$ is a concave increasing function with $\phi(0) = 0$, $\phi(\infty) = \infty$. Now we have

$$\Phi(v) = \frac{1}{a(1-\sigma)} (v^{1-\sigma} - 1), \quad \lambda(t) = \Phi^{-1}(\alpha t) = (1 + a(1-\sigma)\alpha t)^{\frac{1}{1-\sigma}}.$$

Consequently, the term $\lambda^{-1/p}(\gamma n)$ now equals

$$\lambda^{-1/p}(\gamma n) = (1 + c_2)n^{-1/p(1-\sigma)}, \quad c_2 := a(1-\sigma)\alpha\gamma,$$

and, choosing p close enough to 1, one can make $1/p(1-\sigma) > \eta$. \square

Theorem 1.20 (On sub-exponential ergodic rate) For a given Markov chain X , let there exist a function $V : \mathbb{X} \rightarrow [1, +\infty)$ such that X satisfies the Dobrushin condition on every level set $\{x : V(x) \leq c\}$ of the function V , and, for some $a, b, \sigma, C > 0$,

$$\mathbb{E}_x V(X_1) - V(x) \leq -aV(x) \log^{-\sigma}(V(x) + b) + C, \quad x \in \mathbb{X}. \quad (1.39)$$

Then there exist $c_1, c_2 > 0$ such that

$$\|P_n(x_1, \cdot) - P_n(x_2, \cdot)\|_{\text{TV}} \leq c_1 e^{-c_2 n^{1/(1+\sigma)}} (V(x_1) + V(x_2)), \quad x_1, x_2 \in \mathbb{X}, n \geq 1. \quad (1.40)$$

In addition, there exists a unique IPM π for X which satisfies (1.17), and

$$\|P_n(x, \cdot) - \pi\|_{\text{TV}} \leq c_1 e^{-c_2 n^{1/(1+\sigma)}} \left(V(x) + \int_{\mathbb{X}} V d\pi \right), \quad x \in \mathbb{X}, n \geq 1. \quad (1.41)$$

Proof. Now the proof is slightly more cumbersome because the function $v \mapsto av \log^{-\sigma}(v + b)$, although well-defined on $[1, \infty)$, may fail to have a non-negative concave increasing extension to $[0, \infty)$ (e.g. when $b > 0$ is small). Hence we can not apply Theorem 1.13 and Theorem 1.15 straightforwardly; instead, we tune up the relation (1.39) first.

We can check straightforwardly that, for a given σ , there exists $b_\sigma > 1$ such that the function $v \mapsto v \log^{-\sigma} v$ is concave and increasing on $[b_\sigma, \infty)$. Hence rather than using (1.39) with Theorems 1.13 and 1.15 we can apply a (weaker) inequality (1.23) with

$$\phi(v) = \tilde{a}(v + b_\sigma) \log^{-\sigma}(v + b_\sigma),$$

where

$$\tilde{a} = a \inf_{v \geq 1} \frac{v \log^{-\sigma}(v + b)}{(v + b_\sigma) \log^{-\sigma}(v + b_\sigma)} > 0.$$

Then we have

$$\begin{aligned} \Phi(v) &= \frac{1}{\tilde{a}(1 + \sigma)} \left(\log^{1+\sigma}(v + b_\sigma) - \log^{1+\sigma}(1 + b_\sigma) \right), \\ \lambda(t) = \Phi^{-1}(\alpha t) &= \exp \left\{ \left(\log^{1+\sigma}(1 + b_\sigma) + \tilde{a}(1 + \sigma) \alpha t \right)^{1/(1+\sigma)} \right\} - b_\sigma \\ &\geq C(\lambda) \exp \left\{ \left(\tilde{a}(1 + \sigma) \alpha t \right)^{1/(1+\sigma)} \right\} \end{aligned}$$

with

$$C(\lambda) = \inf_{t \geq 0} \left[\lambda(t) \exp \left\{ - \left(\bar{a}(1 + \sigma)\alpha t \right)^{1/(1+\sigma)} \right\} \right] > 0.$$

Consequently, the term $\lambda^{-1/p}(\gamma n)$ now possesses the bound

$$\lambda^{-1/p}(\gamma n) \leq (C(\lambda))^{-1/p} e^{-c_2 n^{1/(1+\sigma)}}, \quad c_2 := (\bar{a}(1+\sigma)\alpha\gamma)^{1/(1+\sigma)}/p. \quad \square$$

.....

1.4 Continuous-time Markov processes

Frequently, one is interested in statements similar to those given in Section 1.3, but for a continuous-time Markov process $X_t, t \geq 0$ rather than for a Markov chain $X_n, n \in \mathbb{Z}_+$. One straightforward way to design such a statement is to consider the initial process at the discrete set of time moments $nh, n \in \mathbb{Z}_+$, with a given $h > 0$; that is, to consider a so called “skeleton chain” $X^h = \{X_{nh}, n \in \mathbb{Z}_+\}$. Because of the inequality

$$\|P_{s+nh}(x, \cdot) - P_{s+nh}(y, \cdot)\|_{\text{TV}} \leq \|P_{nh}(x, \cdot) - P_{nh}(y, \cdot)\|_{\text{TV}},$$

a bound as in (1.16) for the skeleton chain would yield a similar bound for the initial process:

$$\|P_t(x_1, \cdot) - P_t(x_2, \cdot)\|_{\text{TV}} \leq \tilde{r}_t(V(x_1) + V(x_2)), \quad x_1, x_2 \in \mathbb{X}, t \geq 0 \quad (1.42)$$

with

$$\tilde{r}_t = r_{h\lceil t/h \rceil}.$$

However, to prove the bound (1.16) for a skeleton chain, one still should verify the irreducibility and recurrence assumptions. It would be convenient to do this in terms of the process X itself rather than the auxiliary chain X^h . Below, we explain one practical way to do this for the recurrence assumption.

Let us begin from the key calculation, which would then make clear Definition 1.21 below. Take some function V from the domain of the generator A of the semigroup $\{T_t, t \geq 0\} \subset \mathbb{B}(\mathbb{X})$ corresponding to X . Then, by the Dynkin formula (e.g. [15], Chapter

5 §1), the process

$$V(X_t) - \int_0^t AV(X_s) ds$$

is a martingale w.r.t. any law P_x , $x \in \mathbb{X}$, and hence

$$E_x V(X_t) = V(x) + \int_0^t E_x AV(X_s) ds.$$

Let us assume for now that there exist $a, C > 0$ such that

$$AV \leq -aV + C. \quad (1.43)$$

Then for the function $v(t) = E_x V(X_t)$ one has the inequality

$$v(t_2) - v(t_1) \leq \int_{t_1}^{t_2} (-av(s) + C) ds. \quad (1.44)$$

Applying Gronwall's Lemma yields

$$v(t) \leq v(0)e^{-at} + C \frac{1 - e^{-at}}{a}, \quad t \geq 0.$$

This gives finally

$$E_x V(X_1^h) - V(x) \leq -\tilde{a}V(x) + \tilde{C} \quad (1.45)$$

with

$$\tilde{a} = 1 - e^{-ah}, \quad \tilde{C} = C \frac{1 - e^{-ah}}{a}.$$

This simple calculation shows that we can verify a linear Lyapunov-type condition for the skeleton chain in terms of the linear Lyapunov-type condition (1.43) formulated from the generator of the initial process. However, the above argument is an outline of the main idea rather than a ready-made tool for proving the general bound (1.42), because there are two weak points hidden in this argument. First, recall that our aim is to generalize Lyapunov-type conditions from linear ones (and hence generalize convergence rates from the exponential). If, e.g., we assume

$$AV \leq -\phi(V) + C$$

with *concave* ϕ , we would not be able to derive, for instance, such an inequality similar

to (1.44)

$$v(t_2) - v(t_1) \leq \int_{t_1}^{t_2} (-\phi(v(s)) + C) \, ds.$$

To do that, we would need inequality $-E_x \phi(V(X_s)) \geq -\phi(E_x V(X_s))$, which is just the Jensen inequality, valid for *convex* ϕ . This explains that the linear function $\phi(v) = av$, which is both concave and convex, is a special case, and in a general setting, the argument needs to be modified.

This first difficulty is technical and in the proof of Theorem 1.23 below we show one possible way to avoid it. The second difficulty is more deep: a close inspection of the Lyapunov-type relation (1.45) for the skeleton chain shows that it is hardly applicable as a recurrence type condition. Recall that non-uniform ergodic rates (1.16) or (1.42) appear naturally in the case where the irreducibility assumption is only locally verified. However, if V is taken from the domain of the generator of the $\mathbb{B}(\mathbb{X})$ -semigroup of X , then V is bounded; therefore, the claim that “on every level set for V the Dobrushin condition holds,” used e.g. in Theorems 1.18–1.20, just means that the Dobrushin condition holds true on the entire \mathbb{X} and in this case no recurrence condition is required.

To keep the whole argument operational, we have to extend the domain of the generator in order to include some unbounded functions V therein. One very natural way to do this is to remove from the definition of the extended generator all technicalities except the main feature required above; that is, the Dynkin formula. Denote by \mathbb{F}^X the natural filtration of the process X .

Definition 1.21 A measurable function $f : \mathbb{X} \rightarrow \mathbb{R}$ belongs to the domain of the *extended generator* \mathcal{A} of the Markov process X if there exists a measurable function $g : \mathbb{X} \rightarrow \mathbb{R}$ such that the process

$$f(X_t) - \int_0^t g(X_s) \, ds, \quad t \in \mathbb{R}^+ \tag{1.46}$$

is well-defined and is an \mathbb{F}^X -martingale w.r.t. every measure $P_x, x \in \mathbb{X}$. For any such pair (f, g) , we write $f \in \text{Dom}(\mathcal{A})$ and $\mathcal{A}f = g$.

Remark 1.22 This definition, both very convenient and very useful, apparently dates back to H. Kunita [38]. It is regarded as mathematical “common knowledge,” used widely in research papers with various technical modifications, though somehow it is missing from the classical textbooks (with the important exception of [51, Chapter VII.1]).

Now we can finally formulate the main result of this section, which gives a sufficient

condition for the Lyapunov-type condition for the skeleton chain in terms of the extended generator of the initial process. For a given $\phi : [1, +\infty) \rightarrow (0, \infty)$, denote by Φ and H the corresponding functions defined by (1.25) and (1.26).

Theorem 1.23 Assume that for a given Markov process X there exist a function $V : \mathbb{X} \rightarrow [1, +\infty)$ from the domain of the extended generator, a function $\phi : [1, +\infty) \rightarrow (0, \infty)$ which admits a concave, non-negative extension to $[0, \infty)$ and a constant $C \geq 0$ such that

$$\mathcal{A}V \leq -\phi(V) + C. \quad (1.47)$$

Assume also that the process $V(X_t)$, $t \geq 0$, is continuous in probability and for any $T > 0$ the family of the random variables

$$\left(H(t_2, V(X_{t_1})) \right)^2, \quad 0 \leq t_1 \leq t_2 \leq T, \quad (1.48)$$

where the function H is given by (1.26), is uniformly integrable w.r.t. every \mathbb{P}_x , $x \in \mathbb{X}$. Then for any $h > 0$

$$\mathbb{E}_x V^h(X_1^h) - V^h(x) \leq -\phi^h(V^h(x)) + C^h \quad (1.49)$$

with

$$V^h(x) = \Phi^{-1}\left(h + \Phi(V(x))\right) - \Phi^{-1}(h) + 1, \quad x \in \mathbb{X}, \quad (1.50)$$

$$\phi^h(v) = v - 1 + \Phi^{-1}(h) - \Phi^{-1}\left(-h + \Phi(v - 1 + \Phi^{-1}(h))\right), \quad v \geq 1, \quad (1.51)$$

$$C^h = Ch \sup_{v \geq 1} \frac{H(h, v)}{v}. \quad (1.52)$$

The function $\phi^h : [1, \infty) \rightarrow (0, \infty)$ is increasing and admits a concave, non-negative extension to $[0, \infty)$.

Proof. First we prove that

$$\mathbb{E}_x H(t, V(X_t)) \leq V(x) + C \sup_{v \geq 1} \frac{H(t, v)}{v}. \quad (1.53)$$

We use essentially the same argument as in the proof of Theorem 1.15. By inequali-

ties (1.29) and (1.30) we have, for any $t_1 \leq t_2$,

$$\begin{aligned} H(t_2, V(X_{t_2})) - H(t_1, V(X_{t_1})) &\leq (t_2 - t_1) \phi \left(H(t_2, V(X_{t_2})) \right) \\ &\quad + \frac{\phi \left(H(t_2, V(X_{t_1})) \right)}{\phi(V(X_{t_1}))} [V(X_{t_2}) - V(X_{t_1})]. \end{aligned}$$

Because

$$V(X_t) - \int_0^t \mathcal{A}V(X_s) \, ds$$

is a martingale and (1.47) is assumed, we have

$$\begin{aligned} &E_x H(t_2, V(X_{t_2})) - E_x H(t_1, V(X_{t_1})) \\ &\leq (t_2 - t_1) E_x \phi \left(H(t_2, V(X_{t_2})) \right) + E_x \frac{\phi \left(H(t_2, V(X_{t_1})) \right)}{\phi(V(X_{t_1}))} \int_{t_1}^{t_2} \left(-\phi(V(X_s)) + C \right) \, ds. \end{aligned}$$

Then for every partition $0 = t_0 < t_1 < \dots < t_n = t$ of a segment $[0, t]$, we get

$$\begin{aligned} &E_x H(t, V(X_t)) - V(x) \\ &\leq \sum_{k=1}^n E_x \int_{t_{k-1}}^{t_k} \left[\phi \left(H(t_k, V(X_{t_k})) \right) - \frac{\phi \left(H(t_k, V(X_{t_{k-1}})) \right)}{\phi(V(X_{t_{k-1}}))} \phi(V(X_s)) \right] \, ds \\ &\quad + C \sum_{k=1}^n E_x \int_{t_{k-1}}^{t_k} \frac{\phi \left(H(t_k, V(X_{t_{k-1}})) \right)}{\phi(V(X_{t_{k-1}}))} \, ds. \end{aligned}$$

The functions ϕ , H are continuous, $\phi(v) \geq \phi(1) > 0$ and the process $V(X_s)$, $s \geq 0$, is continuous in probability. From this, using the auxiliary uniform integrability condition for the family (1.48), it is easy to deduce that when we consider a sequence of partitions the size of which tends to 0, the above inequality turns into

$$E_x H(t, V(X_t)) - V(x) \leq C E_x \int_0^t \frac{\phi \left(H(s, V(X_s)) \right)}{\phi(V(X_s))} \, ds. \quad (1.54)$$

Applying (1.32) with $a = V(X_s)$ and $b = H(s, V(X_s))$ we get

$$\mathbb{E}_x H(t, V(X_t)) - V(x) \leq C \mathbb{E}_x \int_0^t \frac{H(s, V(X_s))}{V(X_s)} ds,$$

which finally yields (1.53).

Relation (1.53) with $t = h$ has the form

$$\mathbb{E}_x H(h, V(X_h)) \leq V(x) + C^h, \quad (1.55)$$

which after a proper change of notation would give the required Lyapunov-type condition (1.49) for the skeleton chain. Note that the domains where the functions Φ , Φ^{-1} and H are well-defined can be naturally extended. Namely, Φ is well-defined on $(0, \infty)$ and takes values in (κ, ∞) with

$$\kappa = \int_1^0 \frac{1}{\phi(v)} dv \in [-\infty, 0).$$

Hence Φ^{-1} is well-defined on (κ, ∞) . Finally, H is well-defined on the set of the pairs (t, v) such that $t + \Phi(v) > \kappa$.

Denote $\tilde{V}^h(x) = H(h, V(x))$. Then $V(x) = H(-h, \tilde{V}^h(x))$ and (1.55) can be written in the form

$$\mathbb{E}_x \tilde{V}^h(X_h) - \tilde{V}^h(x) \leq -\tilde{\phi}^h(\tilde{V}^h(x)) + C^h$$

with

$$\tilde{\phi}^h(v) = v - H(-h, v).$$

The function $\tilde{V}^h(x)$ takes its values in $[H(h, 1), \infty)$, and $H(h, 1) = \Phi^{-1}(h + \Phi(1)) = \Phi^{-1}(h)$. Then the function

$$V^h(x) = \tilde{V}^h(x) - \Phi^{-1}(h) + 1$$

takes its values in $[1, \infty)$ and (1.49) is just (1.55) written in terms of V^h . Note that

$$\phi^h(v) = \tilde{\phi}^h(v - 1 + \Phi^{-1}(h)).$$

To finalize the proof, we need to prove the required properties for ϕ^h . Observe that

the function $H(-h, \cdot)$ is *convex* on its natural domain $\{v : h + \Phi(v) > \kappa\}$; the proof here is the same used for the concavity of $H(t, \cdot)$ in the proof of Theorem 1.15. Hence $\tilde{\phi}^h$ is concave on its natural domain. Because

$$H'_v(-h, v) = \phi(H(-h, v))/\phi(v) \leq 1,$$

$\tilde{\phi}^h$ is non-decreasing. Since ϕ^h is just $\tilde{\phi}^h$ with a shift in argument, ϕ^h is also convex and non-decreasing on its natural domain.

Finally, recall that ϕ is well-defined and increasing $[0, \infty)$. Hence, because $\Phi^{-1}(h) > 1$,

$$\begin{aligned} -h + \Phi(-1 + \Phi^{-1}(h)) &= -\Phi(\Phi^{-1}(h)) + \Phi(-1 + \Phi^{-1}(h)) \\ &= -\int_1^{\Phi^{-1}(h)} \frac{dw}{\phi(w)} + \int_1^{-1 + \Phi^{-1}(h)} \frac{dw}{\phi(w)} \\ &= -\int_{-1 + \Phi^{-1}(h)}^{\Phi^{-1}(h)} \frac{dw}{\phi(w)} > -\int_0^1 \frac{dw}{\phi(w)} = \kappa. \end{aligned}$$

Therefore the natural domain for ϕ^h contains $[0, \infty)$. Clearly, one has

$$\begin{aligned} \phi^h(0) &= -1 + \Phi^{-1}(h) - \Phi^{-1}\left(-h + \Phi(-1 + \Phi^{-1}(h))\right) \\ &> -1 + \Phi^{-1}(h) - \Phi^{-1}\left(\Phi(-1 + \Phi^{-1}(h))\right) = 0, \end{aligned}$$

and therefore ϕ^h is positive, convex and non-decreasing on $[0, \infty)$. □

Exercise 1.24 Calculate ϕ^h when

- ◇ $\phi(v) = av$,
- ◇ $\phi(v) = av^\sigma$ with $\sigma \in (0, 1)$,
- ◇ $\phi(v) = a(v+b)\log^{-\sigma}(v+b)$ with $\sigma \in (0, 1)$ and sufficiently large b ,

and verify that in any of these cases $V^h \leq cV$ with some c and $\phi^h(v) \geq \tilde{\phi}(v)$, where $\tilde{\phi}(v)$ has the same form as ϕ with possibly different constants a and b (the constant σ remains the same).

Typically, it is practical to verify Lyapunov-type condition (1.47) for particular classes of processes X by means of standard stochastic calculus tools, such as the Itô formula.

We give some examples of such a calculation in Section 1.6. Observe that Theorem 1.23 above allows various modifications, which may in some cases relax the assumptions and simplify the whole proof. For instance, one can avoid the auxiliary assumption of the uniform integrability of the family (1.48), by using the standard stochastic calculus tools to verify directly the inequality (1.54).

Another natural and much more substantial modification of the above argument is that, instead of using the time discretization and a skeleton chain, one can prove a version of Theorem 1.13 in the continuous time setting, and then use (1.47) directly to verify the analogue of the recurrence assumption therein in the continuous time setting. In general, this way might be practical because then one avoids cumbersome calculations, e.g., similar to those which arise in Exercise 1.24. One should then take care of continuous-time technicalities, such as trajectory-wise properties of X (in order to guarantee that a hitting time is a stopping time), the strong Markov property of X (which one would require to justify the argument from the proof of Theorem 1.13 in the continuous time setting), and so on. In order not to over-burden the exposition, we do not discuss this possibility here, referring the reader to [35, Theorem 2.2], for a continuous-time analogue of Theorem 1.13 in the particular case of an exponential ergodic rate.

• • • • •

1.5 Various forms of the irreducibility assumption

The choice of a particular form of the irreducibility condition for a Markov chain is far from trivial and may depend both on the structure of the particular chain and the goals of the analysis. A variety of forms of the irreducibility assumption other than the Dobrushin condition used above are available in the literature. In this section, we briefly discuss several such assumptions.

The chain X is said to satisfy the *minorization condition* if there exist a probability measure λ and $\rho \in (0, 1)$ such that

$$P(x, dy) \geq \rho \lambda(dy).$$

Clearly, the minorization condition implies the Dobrushin condition with the same ρ . The inverse implication, in general, is not true.

Exercise 1.25 Give a corresponding example.

The *Harris irreducibility condition* requires that there exist a measure λ and a positive function $f \in \mathbb{B}(\mathbb{X})$ such that

$$P(x, dy) \geq f(x)\lambda(dy).$$

In a sense, the Harris irreducibility condition is a relaxed minorization condition, perfectly adapted to the duality structure between the spaces of measures $\mathcal{M}(\mathbb{X})$ and functions $\mathbb{B}(\mathbb{X})$.

The *Döblin condition* requires that there exist a probability measure λ and $\varepsilon > 0$ such that

$$P(x, A) \leq 1 - \varepsilon \quad \text{if } \lambda(A) < \varepsilon.$$

In the case of \mathbb{X} being a compact metric space, this condition can be verified in the terms of the *Doob condition* which claims that X is strong Feller; that is, for every bounded and measurable $f : \mathbb{X} \rightarrow \mathbb{R}$, the mapping $x \mapsto E_x f(X_1) = \int_{\mathbb{X}} f(y) P(x, dy)$ is continuous.

All the conditions listed above can also be relaxed by imposing them on some N -step transition probability $P_N(x, dy)$ instead of $P(x, dy)$ or, more generally, on a “mixed” (or “sampled”, cf. [43]) kernel

$$P^Q(x, dy) = \sum_{n=1}^{\infty} Q(n) P_n(x, dy)$$

for some weight sequence $\{Q(n), n \geq 1\} \subset (0, \infty)$ with $\sum_n Q(n) < \infty$. This typically makes it possible to give conditions for the ergodicity which are sufficient and close to necessary. For instance, the Dobrushin condition for an N -step transition probability $P_N(x, dy)$ yields that the “ N -step skeleton chain” possesses a uniform exponential ergodic rate (1.5). As we have explained at the beginning of Section 1.4, this would yield a similar rate (with other constants) for the chain X itself. On the other hand, if X possesses a uniform ergodic bound

$$\sup_{x_1, x_2 \in \mathbb{X}} \|P_n(x_1, \cdot) - P_n(x_2, \cdot)\|_{\text{TV}} \rightarrow 0, \quad n \rightarrow \infty \quad (1.56)$$

with *arbitrary* rate, then for some N , the Dobrushin condition for an N -step transition probability $P_N(x, dy)$ holds.

This simple argument reveals the general feature that, on the level of *uniform* ergodic

rates, the only rate which actually occurs is an exponential one. In addition, the N -step version of the Dobrushin condition appears to be necessary and sufficient for X to possess a uniform exponential ergodic rate. This property is not exclusive to the Dobrushin condition: the N -step versions of the Harris irreducibility condition and the Döblin condition are necessary and sufficient, as well. The necessity of the N -step version of the Döblin condition is simple.

Exercise 1.26 Prove that (1.56) yields the N -step version of the Döblin condition.

Hint. Deduce from (1.56) the uniform convergence of $P_n(x, dy)$ to the (unique) IPM μ for X in the total variation distance and then take $\lambda = \mu$ in the Döblin condition.

The proof of the necessity of the Harris irreducibility condition is much more involved; see [45, Chapter 3].

All the conditions listed above are also available in local versions, i.e. on a set $K \in \mathcal{X}$, and together with proper recurrence conditions typically would lead to non-uniform ergodic rates. In the terminology explained above, the central notions of a *small set* and a *petite set* from the Meyn-Tweedie approach [43], widely used in the literature, can be formulated as follows: a set $K \in \mathcal{X}$ is *small* if X verifies the N -step version of the minorization condition, and $K \in \mathcal{X}$ is *petite* if a “mixed” (or “sampled”) chain verifies the minorization condition.

Note that, although all the irreducibility conditions discussed above are genuinely equivalent at least in their global versions, there are still good reasons to use them separately for various purposes and classes of Markov models. The Döblin condition (or the strong Feller property) is the mildest one and hence is easiest to verify, but it is then rather difficult to get ergodic bounds with *explicit* constants. The minorization condition leads to much more explicit ergodic bounds, but it is too restrictive for models which have transition probabilities with a complicated local behavior. A good example here is given by a Markov process solution to an SDE with a Lévy noise; see Section 1.6. The Dobrushin condition is more balanced in the sense that it leads to explicit ergodic bounds but it is more flexible and might be less restrictive than the minorization condition. This was the reason for us to ground our exposition mainly on the Dobrushin condition.

• • • • •

1.6 Diffusions and Lévy driven SDEs

Below we give two examples where the general methods developed above are well applicable. The Markov models from these examples have considerable interest themselves but also illustrate clearly some natural methods to verify the general assumptions from Section 1.3 and Section 1.4.

1. A diffusion process. Let $X_t, t \geq 0$, be a diffusion in \mathbb{R}^m ; that is, a Markov process solution to an SDE

$$dX_t = a(X_t) dt + b(X_t) dW_t \tag{1.57}$$

driven by a k -dimensional Wiener process $W_t, t \geq 0$. Let the coefficients $a : \mathbb{R}^m \rightarrow \mathbb{R}^m$ and $b : \mathbb{R}^m \rightarrow \mathbb{R}^{m \times k}$ satisfy usual conditions for a (weak) solution to exist uniquely (e.g. [26]); then this solution X is a time-homogeneous Markov process in \mathbb{R}^m .

Below we give explicit conditions in terms of the coefficients of the equation which are sufficient for (1.47) to hold for a particular ϕ . Denote for $f \in C^2(\mathbb{R}^m)$

$$\mathcal{L}f = \sum_{i=1}^m a_i \partial_{x_i} f + \frac{1}{2} \sum_{i,j=1}^m B_{ij} \partial_{x_i x_j}^2 f \tag{1.58}$$

with $B = bb^*$. By the Itô formula (e.g. [26, Chapter II.5]), the process

$$M_t^f = f(X_t) - \int_0^t \mathcal{L}f(X_s) ds$$

is a *local* martingale w.r.t. every $P_x, x \in \mathbb{R}^m$. A natural way to get (1.47) with some ϕ is the following:

- ◊ construct $V \in C^2(\mathbb{R}^m)$ such that

$$\mathcal{L}V \leq -\phi(V) + C; \tag{1.59}$$

- ◊ verify that the respective process M^V is not just a local martingale, but a martingale.

Let us begin with the second technical step.

Proposition 1.27 Let (1.59) hold true with some ϕ which possesses a linear bound; that is, for some $c_1, c_2 > 0$

$$|\phi(v)| \leq c_1|v| + c_2.$$

Assume that there exists a positive function $G \in C^2(\mathbb{R})$ such that $G(v)/|v| \rightarrow \infty$, $|v| \rightarrow \infty$ and that for the function $U = G(V)$, there exist constants $c_3, c_4 > 0$ such that

$$\mathcal{L}U \leq c_3U + c_4.$$

Then the process M^V is a martingale w.r.t. every P_x , $x \in \mathbb{R}^m$.

Proof. Let $x \in \mathbb{R}^m$ be fixed. Both M^V and M^U are P_x -local martingales; thus, there exists an increasing sequence of stopping times $\{\tau_n\}$ such that $\tau_n \rightarrow \infty$ P_x -a.s. and for every n the stopped processes

$$M^V(\tau_n \wedge \cdot), \quad M^U(\tau_n \wedge \cdot)$$

are P_x -martingales. Hence for every fixed $n \geq 1$ we have

$$\begin{aligned} \mathbb{E}_x U(X_{\tau_n \wedge t}) &= U(x) + \mathbb{E}_x \int_0^{\tau_n \wedge t} \mathcal{L}U(X_s) ds \leq U(x) + \mathbb{E}_x \int_0^{\tau_n \wedge t} (c_1 U(X_s) + c_2) ds \\ &\leq U(x) + \int_0^t \mathbb{E}_x (c_1 U(X_{\tau_n \wedge s}) + c_2) ds. \end{aligned}$$

Hence, by the Gronwall inequality one has

$$\mathbb{E}_x U(X_{\tau_n \wedge t}) \leq e^{c_1 t} (U(x) + c_2 t).$$

Recall that $U = G(V)$ and the growth of G at ∞ is faster than linear. Hence the above bound yields that the family

$$V(X_{\tau_n \wedge s}), \quad s \in [0, t], \quad n \geq 1$$

is uniformly integrable w.r.t. P_x . By (1.59) and the linear bound on ϕ , the family

$$\mathcal{L}V(X_{\tau_n \wedge s}), \quad s \in [0, t], \quad n \geq 1$$

is uniformly integrable as well. Because for every $t \geq 0$ and $s \in [0, t]$, we have that

$$V(X_{\tau_n \wedge t}) \rightarrow V(X_t), \quad \mathcal{L}V(X_s)1_{s \leq \tau_n} \rightarrow \mathcal{L}V(X_s), \quad n \rightarrow \infty$$

P_x -a.s., this implies the convergence in $L_1(P_x)$ of

$$M_{\tau_n \wedge t}^V = V(X_{\tau_n \wedge t}) - \int_0^t \mathcal{L}V(X_s) \mathbf{1}_{s \leq \tau_n} ds \rightarrow M_t^V, \quad n \rightarrow \infty$$

for every t . Since every $M^V(\tau_n \wedge \cdot)$ is a martingale, their L_1 -limit M^V is a martingale as well. \square

Now, let us proceed with the relation (1.59). Both the choice of respective Lyapunov-type function V and the function ϕ which arise therein would depend on the properties of the coefficients a and b . To simplify the exposition, we reduce the variety of possibilities and assume the coefficient b to be bounded. The coefficient a is assumed to be locally bounded and to satisfy

$$\limsup_{|x| \rightarrow \infty} \left(a(x), \frac{x}{|x|^{\kappa+1}} \right) = -A_\kappa \in [-\infty, 0) \quad (1.60)$$

for some $\kappa \in \mathbb{R}$.

The *drift condition* (1.60) is quite transparent: it requires that the radial part of the drift is negative far from the origin; that is, the drift pushes the diffusive point towards to the origin when this point is located far from the origin. The index κ controls the growth rate of the absolute value of the radial part at ∞ (actually its decay rate, if $\kappa < 0$), and hence indicates how powerful the drift is towards the origin.

Proposition 1.28 Let the drift condition (1.60) hold true, coefficient a be locally bounded, and coefficient b be bounded. Denote

$$\|B\| = \sup_{x \in \mathbb{R}^m, l \in \mathbb{R}^m \setminus \{0\}} |l|^{-2} |(B(x)l, l)|;$$

recall that $B = bb^*$. Then the following holds:

- 1) If $\kappa \geq 0$, then (1.59) holds true for $\phi(v) = cv$ with some $c, C > 0$ and $V \in C^2(\mathbb{R}^m)$ such that

$$V(x) = e^{\alpha|x|}, \quad |x| \geq 1.$$

The above constant $\alpha > 0$ should satisfy

$$\alpha < \frac{2A_0}{\|B\|};$$

note that if $\kappa > 0$ then $A_0 = \infty$.

2) If $\kappa \in (-1, 0)$, then (1.59) holds true for $\phi(v) = cv \log^{-\sigma} v$ with some $c, C > 0$,

$$\sigma = -\frac{2\kappa}{1+\kappa} > 0$$

and $V \in C^2(\mathbb{R}^m)$ such that $V > 1$ and

$$V(x) = e^{\alpha|x|^{1+\kappa}}, \quad |x| \geq 1$$

with the constant

$$\alpha < \frac{2A_\kappa}{(1+\kappa)\|B\|}.$$

3) If $\kappa = -1$ and in addition

$$2A_{-1} > \sup_x (\text{Trace } B(x)),$$

then (1.59) holds true for $\phi(v) = cv^{1-2/p}$ with some $c, C > 0$ and $V \in C^2(\mathbb{R}^m)$ such that

$$V(x) = |x|^p, \quad |x| \geq 1,$$

where $p > 2$ is such that

$$2A_{-1} > \sup_x (\text{Trace } B(x) + (p-2)\|B(x)\|).$$

Proof. Because both a and B are locally bounded and $V \in C^2(\mathbb{R}^m)$, the function $\mathcal{L}(V)$ is locally bounded. Hence we only have to verify (1.59) outside a large ball in \mathbb{R}^m .

Case 1) ($\kappa \geq 0$). Let $\kappa = 0$ (the case $\kappa > 0$ is similar and simpler). Then for $V(x) = e^{\alpha|x|}$ we have, on the set $\{|x| \geq 1\}$,

$$\begin{aligned} \partial_{x_i} V(x) &= \alpha e^{\alpha|x|} \frac{x_i}{|x|}, \\ \partial_{x_i x_j}^2 V(x) &= \alpha^2 e^{\alpha|x|} \frac{x_i x_j}{|x|^2} + \alpha e^{\alpha|x|} \left(\frac{\delta_{ij}}{|x|} - \frac{x_i x_j}{|x|^3} \right), \end{aligned} \quad (1.61)$$

where $\delta_{ij} = 1_{i=j}$ is the Kroenecker symbol. Then

$$\begin{aligned} \mathcal{L}V(x) &= e^{\alpha|x|} \left[\alpha \left(a(x), \frac{x}{|x|} \right) + \frac{\alpha}{2} \sum_{i,j=1}^d B_{ij}(x) \left(\alpha \frac{x_i x_j}{|x|^2} + \frac{\delta_{ij}}{|x|} - \frac{x_i x_j}{|x|^3} \right) \right] \\ &\leq e^{\alpha|x|} \left[\alpha \left(a(x), \frac{x}{|x|} \right) + \frac{\alpha}{2} \left(\alpha + \frac{d}{|x|} \right) \|B\| \right]; \end{aligned}$$

here we have used the bound

$$\left| \sum_{i,j} B_{ij} C_{ij} \right| \leq d \|B\| \|C\|,$$

valid for symmetric $d \times d$, matrices, and the fact that the norms of the matrices $\{(x_i x_j / |x|^2)\}, \{\delta_{ij} - (x_i x_j / |x|^2)\}$ equal 1.

By the drift condition and assumption on α , we have then that

$$\mathcal{L}V(x) \leq -c e^{\alpha|x|} = -cV(x)$$

outside some large ball.

Case 2) ($\kappa \in (-1, 0)$). The argument here is completely the same, so we just give the short calculation. For $V(x) = e^{\alpha|x|^{1+\kappa}}$ we have

$$\begin{aligned} \partial_{x_i} V(x) &= \alpha(1+\kappa)|x|^\kappa V(x) \frac{x_i}{|x|}, \\ \partial_{x_i x_j}^2 V(x) &= \alpha^2(1+\kappa)^2 |x|^{2\kappa} V(x) \frac{x_i x_j}{|x|^2} + \alpha(1+\kappa)\kappa |x|^{\kappa-1} V(x) \frac{x_i x_j}{|x|^2} \\ &\quad + \alpha(1+\kappa)|x|^\kappa V(x) \left(\frac{\delta_{ij}}{|x|} - \frac{x_i x_j}{|x|^3} \right) \\ &= \alpha^2(1+\kappa)^2 |x|^{2\kappa} V(x) \frac{x_i x_j}{|x|^2} + \alpha(1+\kappa)|x|^{\kappa-1} V(x) \left(\delta_{ij} + (-1+\kappa) \frac{x_i x_j}{|x|^2} \right). \end{aligned}$$

Then

$$\begin{aligned} \mathcal{L}V(x) &\leq V(x) \left[\alpha(1+\kappa)|x|^{2\kappa} \left(a(x), \frac{x}{|x|^{1+\kappa}} \right) \right. \\ &\quad \left. + \frac{\alpha(1+\kappa)}{2} |x|^{2\kappa} \left(\alpha(1+\kappa) + \frac{d}{|x|^{1+\kappa}} \right) \|B\| \right], \end{aligned}$$

and therefore

$$\mathcal{L}V(x) \leq -\tilde{c}V(x)|x|^{2\kappa}$$

outside some large ball with some $\tilde{c} > 0$. Because

$$|x| = \left(\frac{1}{\alpha} \log V(x) \right)^{1/(1+\kappa)},$$

this gives

$$\mathcal{L}V(x) \leq -cV(x) \log^{-\sigma} V(x)$$

outside some large ball.

Case 3) ($\kappa = -1$). Again, we only give the short calculation. For $V(x) = |x|^p$ we have

$$\begin{aligned} \partial_{x_i} V(x) &= p|x|^{p-1} \frac{x_i}{|x|}, \\ \partial_{x_i x_j}^2 V(x) &= p(p-1)|x|^{p-2} \frac{x_i x_j}{|x|^2} + p|x|^{p-1} \left(\frac{\delta_{ij}}{|x|} - \frac{x_i x_j}{|x|^3} \right) \\ &= p|x|^{p-2} \left(\delta_{ij} + (p-2) \frac{x_i x_j}{|x|^2} \right). \end{aligned}$$

Then

$$\begin{aligned} \mathcal{L}V(x) &= p|x|^{p-2} \left[(a(x), x) + \frac{1}{2} \left(\text{Trace } B(x) + (p-2) \left(B(x) \frac{x}{|x|}, \frac{x}{|x|} \right) \right) \right] \\ &\leq p|x|^{p-2} \left[(a(x), x) \left(\text{Trace } B(x) + (p-2) \|B(x)\| \right) \right], \end{aligned}$$

and hence

$$\mathcal{L}V(x) \leq -c|x|^{p-2} = -cV(x)^{1-2/p}$$

outside some large ball. □

Observe that in all the cases 1)–3) of Proposition 1.28 above, the assertion would still hold true if we changed V to $V^{1+\varepsilon}$ with a properly chosen small $\varepsilon > 0$. Hence by Proposition 1.27 we deduce that in each of these cases the Lyapunov-type condition (1.47) (in terms of the extended generator) holds true for corresponding V and ϕ . Then by Theorem 1.23, in each of cases 1)–3) of Proposition 1.28, the Lyapunov-type condition for a skeleton chain X^h holds true with the pair V^h, ϕ^h which may only differ from the

above V , ϕ by some constants; see Exercise 1.24. This completes the verification of the recurrence properties required in Theorem 1.13.

Remark 1.29 The careful reader should notice that we have not yet verified an auxiliary uniform integrability assumption from Theorem 1.23. One easy way to justify this assumption is to apply Proposition 1.27 to $V^{2+\varepsilon}$ with a properly chosen small $\varepsilon > 0$. This however would lead to unnecessary extra limitations on α in cases 1) and 2) or on p in case 3). A more precise way, which we have actually already mentioned in the discussion after Theorem 1.23, is to use the Itô's formula directly to verify the super-martingale property of the process $H(t, V(X_t))$, which would not bring unnecessary extra conditions on α and p . We do not expose the proof here, leaving these details for an interested reader.

The irreducibility properties are now easily understandable, e.g. by means of the analytical approach which dates back to Kolmogorov, and treats the transition probability density $p_t(x, y)$ of a diffusion process as the fundamental solution to the parabolic second order PDE

$$\partial_t - \mathcal{L} = 0.$$

Assume, for instance, that a and B are Hölder continuous and B is *uniformly elliptic*, i.e.

$$\inf_{x \in \mathbb{R}^m, l \in \mathbb{R}^m \setminus \{0\}} |l|^{-2} |B(x)l, l| > 0.$$

Then it follows from classical PDE results, e.g. [17], that $p_t(x, y)$ is a continuous function on $(0, \infty) \times \mathbb{R}^m \times \mathbb{R}^m$. The Stroock-Varadhan support theorem (e.g. [26, Chapter VI.8]) combined with the Markov property of X then implies that for a fixed $h > 0$, the function $p_h(\cdot, \cdot)$ is separated from zero on every compact set in $\mathbb{R}^m \times \mathbb{R}^m$. This immediately yields the minorization condition for the skeleton chain with the measure λ equal to the uniform distribution on a fixed ball.

Summarizing all the above and applying Theorems 1.18–1.20 in various cases of Proposition 1.28, we get the following.

Theorem 1.30 Let X be a diffusion process with Hölder continuous coefficients a and b . The drift coefficient a is assumed to verify the drift condition (1.60) and the diffusion matrix $B = bb^*$ is assumed to be bounded and uniformly elliptic.

Then the following assertions hold true, depending on the value of the index κ involved

in the drift condition.

Case 1. ($\kappa \geq 0$) For any positive $\alpha < 2A_0/\|B\|$ ($\alpha > 0$ is arbitrary in the case $\kappa > 0$), there exist $c_1, c_2 > 0$ such that, for any $x_1, x_2 \in \mathbb{R}^m$,

$$\|P_t(x_1, \cdot) - P_t(x_2, \cdot)\|_{\text{TV}} \leq c_1 e^{-c_2 t} \left(e^{\alpha|x_1|} + e^{\alpha|x_2|} \right), \quad t \geq 0. \quad (1.62)$$

In addition, there exists a unique IPM π for X which satisfies

$$\int_{\mathbb{R}^m} e^{\alpha|y|} \pi(dy) < \infty,$$

and, for every $x \in \mathbb{R}^m$,

$$\|P_t(x, \cdot) - \pi\|_{\text{TV}} \leq c_1 e^{-c_2 t} \left(e^{\alpha|x|} + \int_{\mathbb{R}^m} e^{\alpha|y|} \pi(dy) \right), \quad t \geq 0. \quad (1.63)$$

Case 2. ($\kappa \in (-1, 0)$) For any positive $\alpha < 2A_\kappa/(1+\kappa)\|B\|$, there exist $c_1, c_2 > 0$ such that, for any $x_1, x_2 \in \mathbb{R}^m$,

$$\|P_t(x_1, \cdot) - P_t(x_2, \cdot)\|_{\text{TV}} \leq c_1 e^{-c_2 t^{(1+\kappa)/(1-\kappa)}} \left(e^{\alpha|x_1|^{1+\kappa}} + e^{\alpha|x_2|^{1+\kappa}} \right), \quad t \geq 0. \quad (1.64)$$

In addition, there exists a unique IPM π for X which satisfies

$$\int_{\mathbb{R}^m} e^{\alpha|y|^{1+\kappa}} \pi(dy) < \infty,$$

and, for every $x \in \mathbb{R}^m$,

$$\|P_t(x, \cdot) - \pi\|_{\text{TV}} \leq c_1 e^{-c_2 t^{(1+\kappa)/(1-\kappa)}} \left(e^{\alpha|x|^{1+\kappa}} + \int_{\mathbb{R}^m} e^{\alpha|y|^{1+\kappa}} \pi(dy) \right), \quad t \geq 0. \quad (1.65)$$

Case 3. ($\kappa = -1$) Under the additional assumption

$$2A_{-1} > \sup_x (\text{Trace } B(x)),$$

for any

$$\eta \in \left(1, \frac{1}{2\|B\|} \left(2A_{-1} - \sup_x (\text{Trace } B(x)) \right) \right),$$

there exist $c_1, c_2 > 0$ such that, for any $x_1, x_2 \in \mathbb{R}^m$,

$$\|P_t(x_1, \cdot) - P_t(x_2, \cdot)\|_{\text{TV}} \leq c_1(1 + c_2 t)^{-\eta} (1 + |x_1|^{2\eta} + |x_2|^{2\eta}), \quad t \geq 0. \quad (1.66)$$

In addition, there exists a unique IPM π for X which satisfies

$$\int_{\mathbb{R}^m} |y|^{2\eta} \pi(dy) < \infty,$$

and, for every $x \in \mathbb{R}^m$,

$$\|P_t(x, \cdot) - \pi\|_{\text{TV}} \leq c_1(1 + c_2 t)^{-\eta} \left(1 + |x|^{2\eta} + \int_{\mathbb{R}^m} |y|^{2\eta} \pi(dy)\right), \quad t \geq 0. \quad (1.67)$$

Exponential, sub-exponential and polynomial ergodic rates for a diffusion process in terms of the drift condition were established by A. Veretennikov in the series of papers [56, 29, 57]. In our framework, the same set of results is deduced in a unified way from the Theorem 1.13, Theorem 1.15 and Theorem 1.23 where the particular form of the function ϕ in the Lyapunov-type condition is not specified.

We remark that in the design of Theorem 1.15 and Theorem 1.23 we have been strongly motivated by the paper [13], devoted to sub-exponential ergodic rates corresponding to the general Lyapunov-type condition, although both our main assumptions and the strategy of proofs have some substantial differences from [13]. One of the important points is that we avoid using the minorization condition, e.g. in the “small set” or “petite set” forms and use the Dobrushin condition instead. The second example we consider in this section contains the family of Markov models, in which the difference between the Dobrushin condition and the minorization condition becomes really substantial.

2. A Lévy driven SDE. Consider the following analogue of the SDE (1.57) in \mathbb{R}^m :

$$dX_t = a(X_t) dt + b(X_{t-}) dZ_t \quad (1.68)$$

where Z is now a Lévy process in \mathbb{R}^m which has the Itô-Lévy decomposition

$$Z_t = \int_0^t \int_{|u| \leq 1} u \tilde{\nu}(ds, du) + \int_0^t \int_{|u| > 1} u \nu(ds, du).$$

This is a particular case of a Lévy driven SDE where, in general, instead of the term $b(X_{t-}) dZ_t$, an expression of the form

$$\sigma(X_{t-}) dW_t + \int_{|u| \leq 1} c(X_{t-}, u) \tilde{\nu}(dt, du) + \int_{|u| > 1} c(X_{t-}, u) \nu(dt, du)$$

should appear. In order to make comparison with the diffusive case more transparent, we restrict our consideration. We assume that Z does not contain a diffusive term and the jump coefficient $c(x, u)$ is linear w.r.t. the variable u , which corresponds to a jump amplitude; that is, $c(x, u) = b(x)u$. The coefficients a and b are assumed to be locally Lipschitz and to satisfy the linear growth condition, hence (1.68) has a unique (strong) solution X , which defines a Markov process in \mathbb{R}^m with càdlàg trajectories.

To verify the recurrence condition for the solution to (1.68), one can essentially follow the same line developed above in the diffusive case, although some technical issues deserve separate attention and discussion. Now the role of the second order differential operator (1.58) is played by the integro-differential operator

$$\begin{aligned} \mathcal{L}^{\text{Levy}} f(x) &= \sum_{i=1}^d a_i(x) \partial_{x_i} f(x) \\ &+ \int_{\mathbb{R}^m} \left[f(x + b(x)u) - f(x) - 1_{|u| \leq 1} \sum_{i=1}^d b_i(x) \partial_{x_i} f(x) \right] \mu(du), \end{aligned} \quad (1.69)$$

where μ is the Lévy measure of the process; note that the choice of the operators \mathcal{L} and $\mathcal{L}^{\text{Levy}}$ is motivated by the particular form of the Itô formula for respective processes (e.g. [26, Chapter II.5]).

Let us explain in detail the calculations which would lead to an analogue of assertion 1) in Proposition 1.28. Just as we did in Proposition 1.28, we assume that the coefficient b is bounded and the coefficient a is locally bounded.

Observe that the operator $\mathcal{L}^{\text{Levy}}$ is not *local*; that is, the value of $\mathcal{L}^{\text{Levy}} f$ at some point x involves the values of f in all other points. Thus, we need to choose carefully the candidate for the Lyapunov function V ; see the more detailed discussion below, where the term $\mathcal{L}_3^{\text{Levy}} V$ is considered. In what follows, we deal with $V \in C^2(\mathbb{R}^m)$ such that $V(x) = e^{\alpha|x|}$ for $|x| \geq 1$ and $V(x) \leq e^{\alpha|x|}$ for $|x| \leq 1$; the latter assertion is motivated by the non-locality of $\mathcal{L}^{\text{Levy}}$.

We write

$$\begin{aligned} \mathcal{L}^{\text{Levy}}V(x) &= \sum_{i=1}^d a_i(x) \partial_{x_i} V(x) + \int_{|u| \leq 1} \left[V(x + b(x)u) - V(x) - \sum_{i=1}^d b_i(x) \partial_{x_i} V(x) \right] \mu(du) \\ &\quad + \int_{\mathbb{R}^m} [V(x + b(x)u) - V(x)] \mu(du) \\ &= \mathcal{L}_1^{\text{Levy}}V(x) + \mathcal{L}_2^{\text{Levy}}V(x) + \mathcal{L}_3^{\text{Levy}}V(x) \end{aligned}$$

and analyse these three terms separately. Just as in Proposition 1.28, we have

$$\mathcal{L}_1^{\text{Levy}}V(x) = \alpha e^{\alpha|x|} \left(a(x), \frac{x}{|x|} \right), \quad |x| \geq 1.$$

Under the drift condition (1.60) with $\kappa = 0$, for every $A < A_0$ there exists a large enough R such that

$$\mathcal{L}_1^{\text{Levy}}V(x) \leq -\alpha AV(x), \quad |x| \geq R.$$

Next, we write

$$V(x + b(x)u) - V(x) - \sum_{i=1}^d b_i(x) \partial_{x_i} V(x) = \int_0^1 (1-s) \left(V''(x + sb(x)u) b(x)u, b(x)u \right) ds.$$

The matrix V'' outside the unit ball is given by (1.61), and a straightforward calculation shows that its matrix norm possesses the bound

$$\|V''(x)\| \leq \alpha e^{\alpha|x|} \left(\alpha + \frac{1}{|x|} \right), \quad |x| \geq 1.$$

For any u with $|u| \leq 1$, we have

$$|x| - s\|b\| \leq |x + sb(x)u| \leq |x| + s\|b\|, \quad s \in [0, 1],$$

where

$$\|b\| = \sup_x \|b(x)\|.$$

Then for $|x| \geq \|b\| + 1$, any point $x + sb(x)u$ with $s \in [0, 1]$, $|u| \leq 1$, is located outside the unit ball. Then, for any $\varepsilon > 0$, there exists an R_1 large enough such that for $|x| \geq R_1$ and

$|u| \leq 1$

$$\left| V(x + b(x)u) - V(x) - \sum_{i=1}^d b_i(x) \partial_{x_i} V(x) \right| \leq \|b\|^2 |u|^2 \left(\int_0^1 \alpha^2 e^{\alpha|x| + s\alpha\|b\|} ds + \varepsilon \right).$$

Consequently,

$$\mathcal{L}_2^{\text{Levy}} V(x) \leq \alpha V(x) \left(\|b\| \left(e^{\alpha\|b\|} - 1 \right) \int_{|u| \leq 1} |u|^2 \mu(du) + \frac{\varepsilon \|b\|}{\alpha} \right), \quad |x| \geq R_1.$$

When $|u| > 1$, we can not specify the location of $x + b(x)u$ in general in terms of the position of x : this is exactly the point where the non-locality of $\mathcal{L}^{\text{Levy}}$ is most evident. In any case, we have by convention that

$$V(x + b(x)u) \leq e^{\alpha|x+b(x)u|} \leq e^{\alpha|x|} e^{\alpha\|b\||u|}.$$

Hence for any point x with $|x| \geq 1$, where $V(x) = e^{\alpha|x|}$ by convention, one has

$$\mathcal{L}_3^{\text{Levy}} V(x) \leq V(x) \int_{|u| \geq 1} \left(e^{\alpha\|b\||u|} - 1 \right) \mu(du).$$

Summarizing the above calculations we get the following: if $\alpha > 0$ is such that

$$-A_0 + \|b\| \left(e^{\alpha\|b\|} - 1 \right) \int_{|u| \leq 1} |u|^2 \mu(du) + \int_{|u| \geq 1} \left(\frac{e^{\alpha\|b\||u|} - 1}{\alpha} \right) \mu(du) < 0, \quad (1.70)$$

then, outside some large ball, the inequality

$$\mathcal{L}^{\text{Levy}} V \leq -cV$$

holds true with some positive constant c .

Clearly, for (1.70) to hold true for some $\alpha > 0$, it is *necessary* that the “tails” of μ are exponentially integrable; that is,

$$\exists \beta > 0 : \int_{|u| > 1} e^{\beta|u|} \mu(du) < \infty. \quad (1.71)$$

On the other hand, for (1.70) to hold true for some $\alpha > 0$, it is sufficient that μ satis-

fies (1.71) and

$$A_0 > \| \| b \| \| \int_{|u|>1} |u| \mu(du). \quad (1.72)$$

Exercise 1.31 Please prove this.

Finally, if μ satisfies (1.71) and $\alpha \in (0, \beta)$, then it is straightforward to verify that, for the function V chosen above, the function $\mathcal{L}^{\text{Levy}}V$ is locally bounded. This completes the proof of the following analogue of assertion 1) from Proposition 1.28.

Proposition 1.32 Let the drift condition (1.60) with $\kappa \geq 0$ hold true, coefficient a be locally bounded and coefficient b be bounded. Also assume measure μ to satisfy (1.71) and inequality (1.72) to hold true. Then there exists a positive α such that (1.70) holds, and for which (1.59) holds true for $\phi(v) = cv$ with some $c, C > 0$ where $V \in C^2(\mathbb{R}^m)$ such that

$$V(x) \begin{cases} = e^{\alpha|x|}, & |x| \geq 1, \\ \leq e^{\alpha|x|}, & |x| < 1. \end{cases}$$

In the current setting, a straightforward analogue of Proposition 1.27 is available as well, with the natural change that \mathcal{L} therein should be replaced by $\mathcal{L}^{\text{Levy}}$: this is caused by the particular form of the Itô formula for Lévy driven semimartingales. Thus, in fact, we have proved the Lyapunov-type condition (1.47) with linear $\phi(v) = cv$ and henceforth the required recurrence properties for X are established.

As seen previously in the diffusive case, irreducibility properties of X are closely related to local properties of the transition probabilities, i.e. existence and regularity of the transition probability density $p_t(x, y)$. For a solution to a Lévy driven SDE, these properties are in general more delicate than in the diffusive case; below we briefly outline several methods applicable in that concern.

Following an analytical approach, similar to the one we used for a diffusion before, one should consider $p_t(x, y)$ as the fundamental solution to the *pseudo-differential operator*

$$\partial_t - \mathcal{L}^{\text{Levy}}.$$

Within this approach, one requires an analogue of the classical *parametrix method* [17] to be developed in a Lévy noise setting. Such an analogue typically requires non-trivial “structural” assumptions on the noise; a commonly studied model here concerns the case of a Lévy process being a mixture of α -stable processes. For an overview of the topic,

details and further bibliography, we refer a reader to the monograph [19].

As an alternative to the analytical approach, a variety of “stochastic calculus of variations” methods are available, cf. [3, 49] or, for a more recent exposition [1, 27]; these are just a few references from an extensively developing field, which we can not discuss in detail here. These methods are based either on the integration by parts formula (in the Malliavin calculus case) or the duality formula (in the Picard approach) and typically provide existence and *continuity* (or, moreover, smoothness) of the transition probability density $p_t(x, y)$. The cost is that relatively strong assumptions on the Lévy measure of the noise should be required; a typical requirement here is that for some $\alpha \in (0, 2)$ and $c_1, c_2 > 0$,

$$c_1 \varepsilon^{2-\alpha} |l|^2 \leq \int_{|u| \leq \varepsilon} (u, l)^2 \mu(du) \leq c_2 \varepsilon^{2-\alpha} |l|^2, \quad l \in \mathbb{R}^m. \quad (1.73)$$

Condition (1.73) is a kind of a frequency regularity assumption on the Lévy measure and heuristically means that the intensity of small jumps is comparable with that for an α -stable noise. When this assumption fails, genuinely new effects may appear, which is illustrated by the following simple example, cf. [5, Example 2]. Consider an equation (1.68) with $d = m = 1$, $a(x) = cx$ with $c \neq 0$, $b(x) \equiv 1$ and the process Z of the form $Z_t = \sum_{k=1}^{\infty} 1/k! N_t^k$, where $\{N^k\}$ are independent copies of a Poisson process. Then the solution to (1.68) possesses the transition probability density $p_t(x, y)$, but for every t and x the function $p_t(x, \cdot) \in L_1(\mathbb{R})$ does not belong to any $L_{p, \text{loc}}(\mathbb{R})$ and therefore is not continuous. Hence, when the intensity of small jumps is low, it may happen that the transition probability density $p_t(x, y)$ exists, but is highly irregular. In this framework another kind of stochastic calculus of variations is highly appropriate, based on the Davydov’s stratification method, cf. [36] for a version of this method specially designed for Lévy driven SDEs with minimal requirements for the Lévy measure of the noise. The crucial point is that this method is well designed to give continuity of the function $x \mapsto p_t(x, \cdot)$ in the integral form; that is, as a mapping $\mathbb{R}^m \rightarrow L_1(\mathbb{R}^m)$. But this continuity is exactly the key ingredient for the proof of the local Dobrushin condition for the process X . Following this general line, it is possible to obtain the following sufficient condition (we necessarily omit the numerous technical details, referring the reader to [34], Theorem 1.3 and Section 4).

Proposition 1.33 Let coefficients a and b in (1.68) belong to $C^1(\mathbb{R}^m)$ and $C^1(\mathbb{R}^{d \times d})$

respectively, the Lévy measure μ satisfy $\int_{|u| \leq 1} |u| \mu(du) < \infty$, and at some point x_* the matrix

$$\nabla a(x_*)b(x_*) - \nabla b(x_*)a(x_*)$$

is non-degenerate. Assume also that for the measure μ the following cone condition holds: for every $l \in \mathbb{R}^m \setminus \{0\}$ and $\varepsilon > 0$, there exists a cone

$$V_{l,\rho} = \{u : (u, l) \geq \rho |u| |l|\}, \quad \rho \in (0, 1),$$

such that

$$\mu(V_{l,\rho} \cap \{|u| \leq \varepsilon\}) > 0.$$

Then the Markov process solution to (1.68) X satisfies the local Dobrushin condition on every compact set in \mathbb{R}^m .

Observe that if one intends to verify the irreducibility in the form of the minorization condition, this would require assumptions on the Lévy measure of the noise similar to (1.73), which is much more restrictive than the cone condition used in Proposition 1.33. This well-illustrates the above-mentioned point that in Markov models with comparatively complicated local structure, the Dobrushin condition is more practical than the minorization one.

Let us summarize what has been obtained so far.

Theorem 1.34 Let the assumptions of Proposition 1.33 hold true. Assume also that the drift condition (1.60) with $\kappa \geq 0$ holds true, coefficient b is bounded, the measure μ satisfies (1.71) and inequality (1.72) holds true.

Then there exists a positive α such that (1.70) holds true, and for this α , there exist $c_1, c_2 > 0$ such that, for any $x_1, x_2 \in \mathbb{R}^m$,

$$\|P_t(x_1, \cdot) - P_t(x_2, \cdot)\|_{\text{TV}} \leq c_1 e^{-c_2 t} \left(e^{\alpha|x_1|} + e^{\alpha|x_2|} \right), \quad t \geq 0. \quad (1.74)$$

In addition, there exists a unique IPM π for X which satisfies

$$\int_{\mathbb{R}^m} e^{\alpha|y|} \pi(dy) < \infty,$$

and, for every $x \in \mathbb{R}^m$,

$$\|P_t(x, \cdot) - \pi\|_{\text{TV}} \leq c_1 e^{-c_2 t} \left(e^{\alpha|x|} + \int_{\mathbb{R}^m} e^{\alpha|y|} \pi(dy) \right), \quad t \geq 0. \quad (1.75)$$

Chapter 2

Weak Ergodic Rates

In many cases of interest, the theory developed in the previous chapter is not applicable because of a lack of the irreducibility property of the process. However, it may be that the process is still ergodic, i.e. it possesses a unique IPM, and in addition its transition probabilities converge as $t \rightarrow \infty$ to this IPM, but in a sense, weaker than w.r.t. the total variation distance. This chapter is devoted to the study of such a “weak ergodicity” property. In Section 2.1, we start with motivating examples; in Section 2.2 we briefly recall the construction and basic properties of coupling (or minimal) probability distances, which will be our main tool for measuring the rates of weak convergence. In Section 2.3 exponential weak ergodic rates are established for dissipative Markov chains; in Section 2.4 a Harris-type theorem for weakly ergodic Markov chains is developed.

.....

2.1 Markov models with intrinsic memory

Example 2.1 Let $X = (X^1, X^2)$ be a process in \mathbb{R}^2 which is the solution to the system of SDEs

$$\begin{cases} dX_t^1 = -aX_t^1 dt + dW_t \\ dX_t^2 = -aX_t^2 dt + dW_t \end{cases}$$

where $a > 0$, the components X^1, X^2 have different initial values x^1, x^2 and the Wiener process W is the same for both components. The process X can be given explicitly:

$$\begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix} = e^{-at} \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} + \left(\int_0^t e^{-at+as} dW_s \right) \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

From this formula, one can easily derive two principal properties of the system. First, if we take two initial points $x = (x^1, x^2), y = (y^1, y^2)$ with

$$x^1 - x^2 \neq y^1 - y^2,$$

then, for every $t > 0$, the respective transition probabilities $P_t(x, \cdot), P_t(y, \cdot)$ are supported by the following disjoint sets of \mathbb{R}^2

$$\{z = (z^1, z^2) : z^1 - z^2 = e^{-at}(x^1 - x^2)\}, \quad \{z = (z^1, z^2) : z^1 - z^2 = e^{-at}(y^1 - y^2)\},$$

respectively. Hence, the total variation distance between them remains equal to 2. Next, because $e^{-at} \rightarrow 0, t \rightarrow \infty$, for any $x = (x^1, x^2)$ we have

$$\begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix} \Rightarrow \left(\int_0^\infty e^{-as} dW_s \right) \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

This means that for every $x \in \mathbb{R}^2$, transition probabilities $P_t(x, \cdot)$ converge weakly as $t \rightarrow \infty$ to the (unique) IPM, which is concentrated on the diagonal in \mathbb{R}^2 .

Exercise 2.2 Specify the IPM, verify its uniqueness and prove the above convergence.

Note that if $x^1 \neq x^2$, the above argument shows that $P_t(x, \cdot)$ and the IPM are mutually singular; hence, the convergence does not hold true in the sense of the total variation norm.

One can consider the SDE for $X = (X^1, X^2)$ as a particular case of the SDE (1.57). Then

$$a(x) = -a \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad b(x) = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

and the “technical” explanation of the lack of ergodicity in the total variation norm is now

that the system is not irreducible because of the degeneracy of the diffusion matrix

$$B(x) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Heuristically, because of this degeneracy the system contains some partial “intrinsic memory”: no matter how much time has passed in period t , the value X_t keeps the partial information about the initial point x ; namely, the difference $x^1 - x^2$ can be recovered completely given the value X_t :

$$x^1 - x^2 = e^{at} (X_t^1 - X_t^2).$$

The above simple example gives a natural insight to understand ergodic properties in more general and sophisticated Markov models. The process X , in fact, represents the 2-point motion for the stochastic flow which naturally corresponds to the Ornstein-Uhlenbeck process

$$dU_t = -aU_t dt + dW_t;$$

e.g. [39]. Hence one can naturally expect that the ergodic properties observed in the above example should also be present for stochastic flows generated by SDEs; in particular, because of the degeneracy of SDEs for N -point motions, the whole system typically would contain some partial intrinsic memory.

Example 2.3 This beautiful example comes from [52], see also the introduction to [23]. Consider the following real-valued *stochastic differential delay equation* (SDDE):

$$dX_t = -aX_t dt + b(X_{t-r}) dW_t \tag{2.1}$$

with a fixed $r > 0$. We will not go into the basics of the theory of SDDEs, referring the reader to [44]. We just mention that, in order to determine the values X_t , $t \geq 0$, one should initially specify the values X_s , $s \in [-h, 0]$, since otherwise the term $b(X_{t-r})$ is not well-defined for small positive t . Assuming b to be e.g. Lipschitz continuous, one has that for every function $h \in C(-r, 0)$, there exists a unique (strong) solution to (2.1) with $X_s = h_s$, $s \in [-r, 0]$, which has continuous trajectories. In general, the process X_t , $t \geq 0$, unlike the diffusion process solution to (1.57), is not Markov; instead, the *segment*

process $\mathbf{X} = \{X(t), t \geq 0\}$

$$X(t) = \{X_{t+s}, s \in [-r, 0]\} \in C(-r, 0), \quad t \geq 0$$

possesses the Markov property.

Denote by $\mathcal{F}_{=t}$ the completion of $\sigma(X(t))$. Any random variable measurable w.r.t. $\mathcal{F}_{=t}$ a.s. equals a Borel measurable functional on $C(-r, 0)$ applied to $X(t)$. We fix $t > 0$ and apply to the segment X_{t+s} , $s \in [-r, 0]$ the well known statistical procedure, which makes it possible to consistently derive the variance part from the observation of an Itô type process. Namely, we put

$$V_{t,n}(s) = \sum_{k=1}^{\lfloor 2^n(-s/r) \rfloor} (X_{t-k2^{-n}r} - X_{t-(k-1)2^{-n}r})^2, \quad s \in [-r, 0],$$

and obtain that, with probability 1,

$$V_{t,n}(s) \rightarrow V_t(s) = \int_{t+s}^t b^2(X_{v-r}) dv, \quad s \in [-r, 0].$$

Exercise 2.4 Prove this convergence.

Consequently, for every $s \in [-r, 0]$,

$$b^2(X_{t+s-r}) = \lim_{\varepsilon \rightarrow 0} \frac{V_{t,n}(s+\varepsilon) - V_{t,n}(s)}{\varepsilon}$$

belongs to $\mathcal{F}_{=t}$. If we assume that b is positive and strictly monotone, then the above argument shows that every value X_{t+s-r} , $s \in [-r, 0]$, of the segment $X(t-r)$ is $\mathcal{F}_{=t}$ -measurable. This means that $X(t-r)$ can be recovered uniquely given the value $X(t)$. Repeating this argument, we obtain that the initial value $X(0)$ of the segment process can be recovered uniquely given the value $X(t)$; hence, any two transition probabilities for the segment process \mathbf{X} with different initial values are mutually singular. This means that the Markov system described by \mathbf{X} contains the full “intrinsic memory,” which clearly prohibits this system from converging to an invariant distribution in the total variation norm.

An ergodicity for X in a weaker sense is still possible; we discuss some methods of proving such a weak ergodicity in the subsequent sections. Here we just mention that the

presence of an “intrinsic memory” is a typical feature for Markov systems with “complicated” state spaces, like $\mathbb{X} = C(-r, 0)$ in the above example. The same effect can be observed for stochastic partial differential equations (SPDEs), SDEs driven by fractional noises, etc. Example 2.1 above indicates one possible heuristic reason for such an effect: for systems with “complicated” state spaces, there are many possibilities for the noise to degenerate, in a sense, and therefore for the whole system to not be irreducible.

• • • • •

2.2 Coupling distances for probability measures

Our main aim in this chapter is to develop tools both for proving that a given Markov process is weakly ergodic and to control the *rate* of such an ergodicity. The first natural step for this aim would be to quantify the weak convergence, i.e. to specify a distance on the set $\mathcal{P}(\mathbb{X})$ of probability measures on \mathbb{X} adjusted with the weak convergence in $\mathcal{P}(\mathbb{X})$. Respective theory of *probability metrics* is well-developed, and we do not pretend to expose its constructions and ideas here, referring a reader to [14, Chapter 11], or [59, Chapter 1]. However, the part of this theory related to *coupling* (or *minimal*) probability metrics would be crucial for our subsequent constructions; hence, in this section we outline this topic.

We call a *distance-like function* any measurable function $d : \mathbb{X} \times \mathbb{X} \rightarrow [0, \infty)$, which is symmetric and satisfies

$$d(x, y) = 0 \iff x = y.$$

Denote by the same letter d the respective *coupling distance* on the class $\mathcal{P}(\mathbb{X})$, defined by

$$d(\mu, \nu) = \inf_{(\xi, \eta) \in \mathcal{C}(\mu, \nu)} \mathbb{E}d(\xi, \eta), \quad \mu, \nu \in \mathcal{P}(\mathbb{X}). \quad (2.2)$$

We use the term “coupling distance” instead of “coupling metric” because, in general, $d : \mathcal{P}(\mathbb{X}) \times \mathcal{P}(\mathbb{X}) \rightarrow [0, \infty]$ may fail to satisfy the triangle inequality. However, the following statement shows that the coupling distance may inherit the triangle inequality property from the initial distance-like function.

Proposition 2.5 Assume that the distance-like function d possesses the following extension of the triangle inequality: there exists a $c \geq 1$ such that

$$d(x, z) \leq c(d(x, y) + d(y, z)), \quad x, y, z \in \mathbb{X}.$$

Then the same property holds true for the respective coupling distance d :

$$d(\mu, \lambda) \leq c(d(\mu, \nu) + d(\nu, \lambda)), \quad \mu, \nu, \lambda \in \mathcal{P}(\mathbb{X}).$$

Proof. Fix $\varepsilon > 0$ and choose $(\xi_\varepsilon, \eta_\varepsilon) \in \mathcal{C}(\mu, \nu)$, $(\xi'_\varepsilon, \eta'_\varepsilon) \in \mathcal{C}(\nu, \lambda)$ such that

$$\text{Ed}(\xi_\varepsilon, \eta_\varepsilon) \leq d(\mu, \nu) + \varepsilon, \quad \text{Ed}(\xi'_\varepsilon, \eta'_\varepsilon) \leq d(\nu, \lambda) + \varepsilon.$$

The following useful fact is well known (e.g. [14, Problem 11.8.8]); for the reader's convenience, we sketch its proof after completing the proof of the proposition.

Lemma 2.6 Let (ξ, η) and (ξ', η') be two pairs of random elements valued in a Borel measurable space $(\mathbb{X}, \mathcal{X})$ such that η and ξ' have the same distribution. Then on a proper probability space, there exist three random elements $\zeta_1, \zeta_2, \zeta_3$ such that the law of (ζ_1, ζ_2) in $(\mathbb{X} \times \mathbb{X}, \mathcal{X} \otimes \mathcal{X})$ coincides with the law of (ξ, η) and the law of (ζ_2, ζ_3) coincides with the law of (ξ', η') .

Applying this fact, we get a triple $\zeta_1, \zeta_2, \zeta_3$, defined on the same probability space, such that

$$\text{Ed}(\zeta_1, \zeta_2) \leq d(\mu, \nu) + \varepsilon, \quad \text{Ed}(\zeta_2, \zeta_3) \leq d(\nu, \lambda) + \varepsilon.$$

In addition, the laws of ζ_1, ζ_3 are equal to μ, λ , respectively. Hence

$$d(\mu, \lambda) = \inf_{(\xi, \eta) \in \mathcal{C}(\mu, \lambda)} \text{Ed}(\xi, \eta) \leq \text{Ed}(\zeta_1, \zeta_3) \leq c(d(\mu, \nu) + d(\nu, \lambda) + 2\varepsilon).$$

Because $\varepsilon > 0$ is arbitrary, this gives the required inequality. \square

Proof of Lemma 2.6. We construct the joint law of the triple $(\zeta_1, \zeta_2, \zeta_3)$ using the representation of the laws of the pairs based on the *disintegration formula*. Since \mathbb{X} is assumed to have a measurable bijection to $[0, 1]$ with a measurable inverse, we can consider the case $\mathbb{X} = \mathbb{R}$, only. For any pair of random variables ξ, η there exists a *regular version of the conditional probability* $P_{\eta|\xi}(x, dy)$ (e.g. [14, Chapter 10.2]), which is measurable w.r.t. x , is a probability measure w.r.t. dy , and satisfies

$$P(\xi \in A, \eta \in B) = \int_A P_{\eta|\xi}(x, B) \mu(dx), \quad A, B \in \mathcal{X},$$

where μ denotes the law of ξ . Now, let us define the joint law κ for the triple $(\zeta_1, \zeta_2, \zeta_3)$

by

$$\kappa(A_1 \times A_2 \times A_3) = \int_{A_2} P_{\xi|\eta}(x, A_1) P_{\eta'|\xi'}(x, A_3) \mu(dx), \quad A_1, A_2, A_3 \in \mathcal{X},$$

where μ now denotes the same distribution for η and ξ' . This corresponds to the choice of the conditional probability $P_{(\xi_1, \xi_3)|\xi_2}$ equal to the product measure

$$P_{\xi|\eta} \otimes P_{\eta'|\xi'}.$$

It is straightforward to verify that such a triple $(\zeta_1, \zeta_2, \zeta_3)$ satisfies the required properties. \square

The definition of the coupling distance is strongly related to the classical Monge-Kantorovich *mass transportation problem* (e.g. [50]). Namely, given two mass distributions μ, ν and the transportation cost $d : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}_+$, the coupling distance $d(\mu, \nu)$ introduced above represents exactly the minimal cost to transport μ to ν . An important fact is that the “optimal transportation plan” in this problem exists under some natural topological assumptions on the model. In terms of couplings (which is just another name for transportation plans) and coupling distances, this fact can be formulated as follows.

Proposition 2.7 Let \mathbb{X} be a Polish space and the distance-like function d be lower semi-continuous w.r.t. this metric; that is, for any sequences $x_n \rightarrow x, y_n \rightarrow y$

$$d(x, y) \leq \liminf_n d(x_n, y_n).$$

Then for any $\mu, \nu \in \mathcal{P}(\mathbb{X})$, there exists a coupling $(\xi_*, \eta_*) \in \mathcal{C}(\mu, \nu)$ such that

$$d(\mu, \nu) = \mathbb{E}d(\xi_*, \eta_*). \quad (2.3)$$

In other words, “inf” in (2.2) in fact can be replaced by “min.” We call any pair $(\xi_*, \eta_*) \in \mathcal{C}(\mu, \nu)$ which satisfies (2.3) an *optimal coupling* and denote the class of optimal couplings $\mathcal{C}_{d,\text{opt}}(\mu, \nu)$. The same terminology and notation is also used when we deal with laws on $\mathbb{X} \times \mathbb{X}$ instead of pairs of random elements. By $\mathcal{C}(\mu, \nu)$, we will denote the class of measures on $\mathbb{X} \times \mathbb{X}$ such that their projections on the first and second coordinates equal μ and ν respectively and by $\mathcal{C}_{d,\text{opt}}(\mu, \nu)$ the subclass of measures $\kappa \in \mathcal{C}(\mu, \nu)$ such

that

$$d(\mu, \nu) = \int_{\mathbb{X} \times \mathbb{X}} d(x, y) \kappa(dx, dy).$$

Proof of Proposition 2.7. Observe that the family of measures $\mathcal{C}(\mu, \nu)$ is *tight* (e.g. [4]): to construct a compact set $K \subset \mathbb{X} \times \mathbb{X}$ such that, for a given ε ,

$$\kappa(K) \geq 1 - \varepsilon,$$

one can simply choose two compact sets $K_1, K_2 \subset \mathbb{X}$ such that

$$\mu(K_1) \geq 1 - \frac{\varepsilon}{2}, \quad \nu(K_2) \geq 1 - \frac{\varepsilon}{2}$$

and then set $K = K_1 \times K_2$.

Consider a sequence of pairs $\{(\xi_n, \eta_n)\} \subset \mathcal{C}(\mu, \nu)$ such that

$$Ed(\xi_n, \eta_n) \leq d(\mu, \nu) + \frac{1}{n}.$$

Then by the Prokhorov theorem, there exists a subsequence $\{(\xi_{n_k}, \eta_{n_k})\}$ which converges in law to some pair (ξ_*, η_*) . Then both sequences of components $\{\xi_{n_k}\}, \{\eta_{n_k}\}$ also converge in law to ξ_*, η_* respectively, and hence $(\xi_*, \eta_*) \in \mathcal{C}(\mu, \nu)$. Next, by Skorokhod's "common probability space" principle (e.g. [14, Theorem 11.7.2]), there exists a sequence $\{(\tilde{\xi}_k, \tilde{\eta}_k)\}$ and a pair $(\tilde{\xi}_*, \tilde{\eta}_*)$, defined on the same probability space, such that the laws of respective pairs (ξ_{n_k}, η_{n_k}) and $(\tilde{\xi}_k, \tilde{\eta}_k)$ coincide and

$$(\tilde{\xi}_k, \tilde{\eta}_k) \rightarrow (\tilde{\xi}_*, \tilde{\eta}_*)$$

with probability 1. Then by the lower semi-continuity of d , one has

$$d(\tilde{\xi}_*, \tilde{\eta}_*) \leq \liminf_k d(\tilde{\xi}_k, \tilde{\eta}_k),$$

and hence the Fatou lemma gives

$$Ed(\tilde{\xi}_*, \tilde{\eta}_*) \leq \liminf_k Ed(\tilde{\xi}_k, \tilde{\eta}_k) = \liminf_k Ed(\xi_{n_k}, \eta_{n_k}) = d(\mu, \nu).$$

Because $(\tilde{\xi}_*, \tilde{\eta}_*)$ has the same law as $(\xi_*, \eta_*) \in \mathcal{C}(\mu, \nu)$, this completes the proof. \square

Let us give several typical examples. In what follows, \mathbb{X} is a Polish space with the metric ρ .

Example 2.8 Let $d(x, y) = \rho(x, y)$; we denote the respective coupling distance with $W_{\rho,1}$ and discuss its properties. First, since ρ satisfies the triangle inequality, so does $W_{\rho,1}$. Next, $W_{\rho,1}$ is symmetric and non-negative. Finally, it possesses the identification property:

$$W_{\rho,1}(\mu, \nu) = 0 \iff \mu = \nu.$$

The part “ \Leftarrow ” of this statement is trivial; to prove the “ \Rightarrow ” part we just notice that there exists an optimal coupling (ξ_*, η_*) for μ, ν : because d is continuous, we can apply Proposition 2.7. For this coupling, we have

$$Ed(\xi_*, \eta_*) = W_{\rho,1}(\mu, \nu) = 0,$$

and because d has the identification property this means that in fact, $\xi_* = \eta_*$ a.s. Hence their laws coincide.

We have just seen that for the coupling distance $W_{\rho,1}$ all the axioms of a metric hold true; the one detail which may indeed cause $W_{\rho,1}$ to not be a metric is that $W_{\rho,1}$, in general, may take value ∞ . If ρ is *bounded*, this does not happen, and $W_{\rho,1}$ is a metric on $\mathcal{P}(\mathbb{X})$.

Example 2.9 Let $p > 1$ and $d(x, y) = \rho^p(x, y)$; we denote the respective coupling distance $W_{\rho,p}^p$ (this notation and related terminology will be discussed at the end of this section). Similarly to the case $d = 1$, this coupling distance is symmetric, non-negative and possesses the identification property. In this case, the triangle inequality does not hold in general. Instead, we have the following weaker version

$$W_{\rho,p}^p(\mu, \lambda) \leq 2^{p-1} (W_{\rho,p}^p(\mu, \nu) + W_{\rho,p}^p(\nu, \lambda)),$$

which by Proposition 2.5 is an easy consequence of the triangle inequality for ρ and the elementary inequality $(a + b)^p \leq 2^{p-1}(a^p + b^p)$, $a, b \geq 0$.

Example 2.10 Let $d(x, y) = 1_{x \neq y}$ be the *discrete metric* on \mathbb{X} . Then, by the Coupling Lemma,

$$d(\mu, \nu) = \frac{1}{2} \|\mu - \nu\|_{\text{TV}}.$$

Hence the total variation distance is a particular representative of the class of coupling

distances. Note that the function $d(x, y) = 1_{x \neq y}$ is lower semi-continuous; thus, Proposition 2.7 gives an alternative way to obtain the statement of the Coupling Lemma.

The last example makes it easier to explain the origins of the “measurable optimal selection” problem discussed below. In the “coupling” proofs of the ergodic rates given in Sections 1.2 and 1.3, it was essential to have for a given transition probability $P(x, dx')$ a probability kernel Q on $\mathbb{X} \times \mathbb{X}$ such that for any x_1, x_2 the measure $Q((x_1, x_2), \cdot)$ gives a maximal coupling for the pair $P(x_1, \cdot), P(x_2, \cdot)$. In the subsequent sections we will extend this approach to a wider class of coupling distances, but in that concern we have to return to the same question: given a transition probability $P(x, dx')$ and a distance-like function, is it possible to choose a probability kernel Q on $\mathbb{X} \times \mathbb{X}$ in such a way that for any x_1, x_2

$$Q((x_1, x_2), \cdot) \in \mathcal{C}_{d, \text{opt}}(P(x_1, \cdot), P(x_2, \cdot))?$$

Clearly, we will assume d to be lower semi-continuous in order for an optimal coupling to exist for any pair $\mu, \nu \in \mathcal{P}(\mathbb{X})$. However, the proof of such an existence in Proposition 2.7 is more implicit than the proof of the Coupling Lemma in Section 1.1, and therefore it is not easy to adapt it directly for measurability purposes, as done in Section 1.1 (cf. Exercise 1.4). Hence we consider this problem separately.

Consider the spaces $\mathbb{S} = \mathcal{P}(\mathbb{X}) \times \mathcal{P}(\mathbb{X})$, $\mathbb{S}' = \mathcal{P}(\mathbb{X} \times \mathbb{X})$ and the set-valued mapping $\psi : \mathbb{S} \ni s \mapsto \psi(s) \subset \mathbb{S}'$,

$$\psi((\mu, \nu)) = \mathcal{C}_{d, \text{opt}}(\mu, \nu). \quad (2.4)$$

Our aim is to choose a *selector* ϕ for this mapping, i.e. a function $\phi : \mathbb{S} \rightarrow \mathbb{S}'$ such that

$$\phi(s) \in \psi(s), \quad s \in \mathbb{S},$$

which is also measurable.

In general, the *measurable selection* problem just outlined is quite complicated, and in some cases it may fail to have a solution. We refer to [16, Appendix 3] for a compact but very informative exposition of the measurable selection topic and to §3 therein for a counterexample where the measurable selection does not exist. We also refer a reader deeply interested in the general measurable selection topic to an excellent survey paper [58].

Fortunately, the particular spaces \mathbb{S} , \mathbb{S}' and the set-valued mapping ψ we have in-

troduced above possess fine topological properties, which make it possible to solve the required measurable selection problem. In what follows we evaluate these properties; our goal is to apply the following theorem.

Theorem 2.11 ([54, Theorem 12.1.10, Lemma 12.1.8]) Let \mathbb{S} be a metric space, \mathbb{S}' be a Polish space, and a set-valued mapping ψ have the following properties:

- ◇ for every $s \in \mathbb{S}$ the set $\psi(s) \subset \mathbb{S}'$ is compact;
- ◇ the mapping ψ is *semi-continuous* in the following sense: as soon as $s_n \rightarrow s$ in \mathbb{S} and $s'_n \in \psi(s_n)$, $n \geq 1$, the sequence $\{s'_n\}$ has a limit point $s' \in \psi(s)$.

Then there exists a measurable map $\phi : \mathbb{S} \rightarrow \mathbb{S}'$ such that $\phi(s) \in \psi(s)$, $s \in \mathbb{S}$.

Remark 2.12 [54, Theorem 12.1.10] is a weaker version of the Kuratovskii and Ryll-Nardzewski theorem, also called the Fundamental Measurable Selection Theorem; see [58].

Evaluation of the required topological properties of the particular spaces \mathbb{S} , \mathbb{S}' and the set-valued mapping ψ is both rather simple and informative. Hence we propose them for the reader in the following series of questions; for the first three of them, see also [14, Chapter 11].

In what follows we assume that the metric ρ on \mathbb{X} is bounded: if this fails, we replace it by an equivalent metric $\tilde{\rho} = \rho \wedge 1$. We denote now by the same symbol ρ the respective coupling distance $W_{\rho,1}$ on $\mathcal{P}(\mathbb{X})$. By Example 2.8, ρ is a metric on $\mathcal{P}(\mathbb{X})$.

Exercise 2.13 Prove that convergence w.r.t. ρ is equivalent to weak convergence in $\mathcal{P}(\mathbb{X})$.

Hint: If $\mu_n \Rightarrow \mu$ then the “common probability space” principle and the Lebesgue dominated convergence theorem provide $\rho(\mu_n, \mu) \rightarrow 0$. If $\rho(\mu_n, \mu) \rightarrow 0$, then for any Lipschitz continuous function $f : \mathbb{X} \rightarrow \mathbb{R}$

$$\left| \int f \, d\mu_n - \int f \, d\mu \right| \leq \|f\|_{\text{Lip}} \rho(\mu_n, \mu) \rightarrow 0, \quad \|f\|_{\text{Lip}} = \sup_{x \neq y} \frac{|f(x) - f(y)|}{\rho(x, y)}$$

(please prove the first inequality), and therefore $\mu_n \Rightarrow \mu$; cf. [4, Theorem 2.1].

Exercise 2.14 Prove that the metric space $(\mathcal{P}(\mathbb{X}), \rho)$ is separable.

Hint: If $\{x_i\}$ is a separability set in \mathbb{X} , then a countable set of all finite sums of the form

$$\sum_k c_k \delta_{x_{i_k}}, \quad \{c_k\} \subset \mathbb{Q} \cap [0, \infty), \quad \sum_k c_k = 1$$

is dense in $(\mathcal{P}(\mathbb{X}), \rho)$.

Exercise 2.15 Prove that the metric space $(\mathcal{P}(\mathbb{X}), \rho)$ is complete.

Hint: Prove that, if a sequence $\{\mu_n\}$ is Cauchy w.r.t. ρ , it is tight. Then use the Prokhorov theorem and Exercise 2.13 to show that this sequence has a limit point in $(\mathcal{P}(\mathbb{X}), \rho)$.

Hence $(\mathcal{P}(\mathbb{X}), \rho)$ is a Polish space, and then both $\mathbb{S} = \mathcal{P}(\mathbb{X}) \times \mathcal{P}(\mathbb{X})$ and $\mathbb{S}' = \mathcal{P}(\mathbb{X} \times \mathbb{X})$ are Polish spaces with respective metrics, as well.

Exercise 2.16 Prove that for any fixed μ, ν the set $\mathcal{C}(\mu, \nu)$ is a compact subset of $\mathcal{P}(\mathbb{X} \times \mathbb{X})$.

Hint: Use the tightness argument from the proof of Proposition 2.7.

Exercise 2.17 Prove that for any fixed μ, ν the set $\mathcal{C}_{d, \text{opt}}(\mu, \nu)$ is a closed subset of $\mathcal{P}(\mathbb{X} \times \mathbb{X})$; this together with the previous exercise shows that it is compact.

Hint: Apply the “common probability space” principle and use the lower semi-continuity of d .

Exercise 2.18 Prove that if d is *continuous*, then the set-valued mapping ψ defined by (2.4) is semi-continuous.

Hint: Combine the tightness argument similar to that from the proof of Proposition 2.7 with the “common probability space” principle, and use the lower semi-continuity of d .

Summarizing all the above we get the following.

Proposition 2.19 Let a distance-like function d be *continuous*. Then there exists a measurable function

$$\Phi : \mathcal{P}(\mathbb{X}) \times \mathcal{P}(\mathbb{X}) \rightarrow \mathcal{P}(\mathbb{X} \times \mathbb{X})$$

such that for every $\mu, \nu \in \mathcal{P}(\mathbb{X})$

$$\Phi((\mu, \nu)) \in \mathcal{C}_{d, \text{opt}}(\mu, \nu).$$

Observe that any probability kernel $P(x, dx')$ can now be understood as the measurable

mapping $\mathbb{X} \rightarrow \mathcal{P}(\mathbb{X})$. Hence we can construct the required kernel Q e.g. in the form

$$Q((x_1, x_2), \cdot) = \Phi\left(\left(P(x_1, \cdot), P(x_2, \cdot)\right)\right).$$

Remark 2.20 Continuity requirement on d comes from our intent to restrict our consideration to the case of a semi-continuous set-values mapping ψ . This excludes from consideration the discrete metric d , for example, and respectively, the total variation distance. It is simple and instructive to check that in this case the semi-continuity property for ψ fails; we leave this as an example for the reader. A more elaborate analysis shows that the *measurability* property of ψ , required in [54, Theorem 12.1.10], actually still holds true when d is assumed to be only lower semi-continuous; in order to keep the exposition reasonably transparent we do not address this question here.

At the end of this section we give a short discussion of some important and closely related topics, which will not be used in the sequel. Our construction of a coupling (minimal) distance is in some aspects more restrictive than the one available in the literature. Generally, such a distance is defined as

$$\inf_{(\xi, \eta) \in \mathcal{C}(\mu, \nu)} H(\xi, \eta),$$

where H is an analogue of a distance-like function on a class of random variables, defined on a common probability space. The particular choice $H(\xi, \eta) = \mathbb{E}d(\xi, \eta)$ leads to the definition used above. Another natural choice of H is the L_p -distance $H(\xi, \eta) = \left(\mathbb{E}d^p(\xi, \eta)\right)^{1/p}$, $p > 1$. Observe that such an H possesses the triangle inequality if d does as well, and hence, the respective coupling distance inherits this property; the proof here is the same as in Proposition 2.5. The probability metric

$$W_{p,p}(\mu, \nu) = \inf_{(\xi, \eta) \in \mathcal{C}(\mu, \nu)} \left(\mathbb{E}d^p(\xi, \eta)\right)^{1/p}$$

is called the L_p -Kantorovich(-Wasserstein) distance on $\mathcal{P}(\mathbb{X})$.

One good reason to use coupling distances is that they are very convenient for estimation purposes. We have already seen this in the previous chapter: to bound the total variation distance between the laws of X_n with various starting points, we have constructed the pair of processes with the prescribed law of the components, and thus transformed the

initial problem to estimating the probability $P(X_n^1 \neq X_n^2)$. A similar argument appears to be practical in other frameworks and for other coupling distances. This is the reason why it is very useful that some natural probability metrics possess a coupling representation. We have already seen one such example: the coupling representation given by the Coupling Lemma for the total variation distance. Another example is the *Lipschitz metric*

$$d_{\text{Lip}}(\mu, \nu) = \sup_{f: \|f\|_{\text{Lip}}=1} \left| \int f \, d\mu - \int f \, d\nu \right|;$$

the coupling representation here is given by the *Kantorovich-Rubinshtein (duality) theorem*, which states that

$$d_{\text{Lip}} = W_{\rho,1}.$$

Finally, for the classical *Lévy-Prokhorov metric*

$$d_{\text{LP}}(\mu, \nu) = \inf \left\{ \varepsilon > 0 : \mu(A) \leq \nu(\{y : \rho(y, A) \leq \varepsilon\}) + \varepsilon, A \in \mathcal{X} \right\},$$

the *Strassen theorem* gives the coupling representation

$$d_{\text{LP}}(\mu, \nu) = \inf_{(\xi, \eta) \in \mathcal{C}(\mu, \nu)} H_{\text{KF}}(\xi, \eta),$$

where H_{KF} stands for the *Ky Fan metric*

$$H_{\text{KF}}(\xi, \eta) = \inf \left\{ \varepsilon > 0 : P(\rho(\xi, \eta) > \varepsilon) < \varepsilon \right\}.$$

For a detailed exposition of this topic, we refer to [14, Chapter 11].

• • • • •

2.3 Dissipative stochastic systems

Let us come back to examples we discussed in Section 2.1 and use them in order to illustrate one simple and natural argument which, for some classes of Markov models, both gives the ergodicity and controls respective weak ergodic rates in a very transparent way. The linear SDE which describes the process X in Example 2.1 is “too simple” in the sense that we can give the solution explicitly. To illustrate the main argument, it is

convenient to consider an example of a slightly more general SDE in \mathbb{R}^m of the form

$$dX_t = -aX_t dt + b(X_t) dW_t. \quad (2.5)$$

Here the Wiener process takes values in \mathbb{R}^k , and we assume the coefficient $b: \mathbb{R}^m \rightarrow \mathbb{R}^{m \times k}$ to be Lipschitz continuous. Fix two initial values $x, x' \in \mathbb{R}^m$ and consider two respective solutions X, X' to (2.5) with the *same* Wiener process W . Then for $\Delta_t = X_t - X'_t$ one has

$$d\Delta_t = -a\Delta_t dt + (b(X_t) - b(X'_t)) dW_t,$$

and, by the Itô formula applied to the function $F(x) = |x|^p$, $p \geq 2$,

$$\begin{aligned} |\Delta_t|^p &= |x - x'|^p - pa \int_0^t |\Delta_s|^p ds \\ &+ \frac{p}{2} \int_0^t |\Delta_s|^{p-2} \sum_{i,j=1}^m \sum_{l=1}^k (b_{il}(X_s) - b_{il}(X'_s)) (b_{jl}(X_s) - b_{jl}(X'_s)) \left(\delta_{ij} + (p-2) \frac{\Delta_s^i \Delta_s^j}{|\Delta_s|^2} \right) ds \\ &+ M_t, \end{aligned}$$

where M is a local martingale. Denote

$$\|b\|_{\text{Lip}} = \sup_{x \neq y} \frac{\|b(x) - b(y)\|}{|x - y|},$$

and observe that the expression under the second integral is dominated by

$$m(p-1) \|b\|_{\text{Lip}}^2 |\Delta_s|^p.$$

To see that, note that the norm of the matrix

$$\left(\delta_{ij} + (p-2) \frac{\Delta_s^i \Delta_s^j}{|\Delta_s|^2} \right)_{i,j=1}^m$$

does not exceed $p-1$, and that for any two $m \times k$ -matrices B, C ,

$$\left| \sum_{i=1}^m \sum_{l=1}^k B_{ij} C_{il} \right| \leq m \|B\| \|C\|.$$

Now the same argument as we have used in the proof of Proposition 1.27 shows that for every $p \geq 2$, the respective process M is a (usual) martingale, and

$$\mathbb{E}|\Delta_{t_2}|^p - \mathbb{E}|\Delta_{t_1}|^p \leq \int_{t_1}^{t_2} \left[-ap + \frac{mp(p-1)}{2} \|b\|_{\text{Lip}}^2 \right] \mathbb{E}|\Delta_s|^p ds.$$

Therefore the comparison principle for ODEs gives the bound

$$\mathbb{E}|\Delta_t|^p \leq |x - x'|^p e^{-\theta_p t}, \quad \theta_p = ap - \frac{mp(p-1)}{2} \|b\|_{\text{Lip}}^2.$$

We can summarize these calculations as follows. If in (2.5) the coefficient b is constant (which is the case considered in Example 2.1), then Δ_t follows a linear ODE and can be given explicitly: $\Delta_t = (x - x')e^{-at}$. When $a > 0$, this means that the whole system is contractive in the following sense: if two points perform the motion according to the same SDE, the distance between these points decreases at exponential rate. This feature appears to be quite non-sensitive w.r.t. complications in the structure of the model. What we have shown above is that, if the Lipschitz norm of the diffusion coefficient b is small when compared with the contraction index a , the system still performs a contraction, with the change that now instead of the norm of the difference $|\Delta_t|$ itself its moment of the order p should be considered. The linearity of the drift term is also not required; without any essential changes, the above argument provides the following statement.

Proposition 2.21 For two solutions X, X' of the SDE (1.57) with $X_0 = x, X'_0 = x'$ and for any $p \geq 2$,

$$\mathbb{E}|X_t - X'_t|^p \leq |x - x'|^p e^{-\theta_p t},$$

where

$$\theta_p = -p \sup_{x \neq y} \frac{1}{|x - y|^2} \left[(a(x) - a(y), x - y) + \frac{m(p-1)}{2} \|b(x) - b(y)\|^2 \right].$$

If $\theta_p > 0$ for some p , it is usually said that the model based on the SDE (1.57) is *dissipative* in the L_p sense. We remark that the statement of Proposition 2.21 still holds true for $p \in [1, 2)$: although in this case the function $F(x) = |x|^p$ does not belong to $C^2(\mathbb{R}^m)$, its only “irregular” point is 0. Once Δ_t hits 0, the solutions X_t and X'_t coincide, and therefore they coincide afterwards because of the strong Markov property and uniqueness of the solution to (1.57). This allows one to apply the Itô formula “locally”, i.e. up to the

first time for Δ to hit 0; we omit further details here.

Therefore, we can summarize that for the model (1.57) to be dissipative in the L_p sense for some $p \geq 1$, it is sufficient to assume that the coefficient b is Lipschitz continuous and the coefficient a satisfies the following “drift dissipativity condition”:

$$\sup_{x \neq y} \frac{(a(x) - a(y), x - y)}{|x - y|^2} < 0. \quad (2.6)$$

As a straightforward corollary of Proposition 2.21, for the transition probabilities of the solution to (1.57), one gets the following bound in terms of the coupling distance, which corresponds to the distance-like function $d_p(x, y) = |x - y|^p$:

$$d_p(P_t(x, \cdot), P_t(x', \cdot)) \leq e^{-\theta p t} d_p(x, x').$$

To verify this bound, one should just note that the pair (X_t, X'_t) above is a representative of the class of couplings $\mathcal{C}(P_t(x, \cdot), P_t(x', \cdot))$.

Motivated by this example, we call a Markov chain *d-dissipative* for a given distance-like function d if there exists $\theta > 0$ such that, for every $x, x' \in \mathbb{X}$

$$d(P(x, \cdot), P(x', \cdot)) \leq e^{-\theta} d(x, x'). \quad (2.7)$$

If a Markov chain is *d-dissipative*, the distance-like function d is called *contractive* for this chain.

For a dissipative chain, a natural analogue of the first part of Theorem 1.3 holds true, with the total variation distance replaced by the coupling distance d .

Proposition 2.22 Let a *continuous* distance-like function d be contracting for a given Markov chain X . Then for every $x, x' \in \mathbb{X}$,

$$d(P_n(x, \cdot), P_n(x', \cdot)) \leq e^{-\theta n} d(x, x'), \quad (2.8)$$

where θ is the same as in (2.7).

Proof. Consider a kernel Q on $\mathbb{X} \times \mathbb{X}$ such that for every $x, x' \in \mathbb{X}$ the measure $Q((x, x'), \cdot)$ belongs to the set of d -optimal couplings of $P(x, \cdot)$ and $P(x', \cdot)$. Due to Proposition 2.19, we know that such a kernel exists; note that this is exactly the

point where the continuity assumption for d is required. Consider a Markov process $Z = (Z^1, Z^2)$ in $\mathbb{X} \times \mathbb{X}$ with the transition probability Q . We have

$$\begin{aligned} \mathbb{E}\left(d(Z_n^1, Z_n^2) \middle| \mathcal{F}_{n-1}^Z\right) &= \int_{\mathbb{X} \times \mathbb{X}} d(y, y') Q((x, x'), dy dy') \Big|_{x=Z_{n-1}^1, x'=Z_{n-1}^2} \\ &= d(P(x, \cdot), P(x', \cdot)) \Big|_{x=Z_{n-1}^1, x'=Z_{n-1}^2} \leq e^{-\theta} d(Z_{n-1}^1, Z_{n-1}^2). \end{aligned}$$

Iterating this inequality, we get the bound

$$\mathbb{E}_{(x, x')}^Z d(Z_n^1, Z_n^2) \leq e^{-\theta n} d(x, x'), \quad n \geq 1.$$

This proves the required statement, because the laws of Z_n^1, Z_n^2 w.r.t. $\mathbb{P}_{(x, x')}^Z$ equal $P(x, \cdot)$ and $P(x', \cdot)$, respectively. \square

The coupling construction used in the above proof also yields that the chain X has at most one IPM. Indeed, consider two IPMs π, π' , and let the chain Z have the initial distribution $\pi \otimes \pi'$. Fix $\varepsilon > 0$, $\delta > 0$ and choose C and n large enough such that

$$\mathbb{P}\left(d(Z_0^1, Z_0^2) > C\right) < \varepsilon, \quad Ce^{-\theta n} < \varepsilon \delta.$$

Then

$$\begin{aligned} \mathbb{P}\left(d(Z_n^1, Z_n^2) > \delta\right) &\leq \mathbb{P}\left(d(Z_0^1, Z_0^2) > C\right) + \mathbb{P}\left(d(Z_n^1, Z_n^2) > \delta, d(Z_0^1, Z_0^2) \leq C\right) \\ &\leq \varepsilon + \frac{e^{-\theta n}}{\delta} \mathbb{E}d(Z_0^1, Z_0^2) \mathbb{1}_{d(Z_0^1, Z_0^2) \leq C} < 2\varepsilon. \end{aligned}$$

Because Z_n^1, Z_n^2 have the laws π, π' respectively and ε and δ are arbitrary, this means that $d(\pi, \pi') = 0$ and therefore $\pi = \pi'$.

In many cases of interest, the same construction also leads to the existence of an IPM and gives a convergence rate of transition probabilities to the IPM; that is, the full analogue of Theorem 1.3 holds true.

Proposition 2.23 Let a continuous distance-like function d be contracting for a given

Markov chain X . Assume also that for some $x_* \in \mathbb{X}$

$$C_* := \sup_n \int_{\mathbb{X}} d(x_*, y) P_n(x_*, dy) < \infty, \quad (2.9)$$

and there exists $p \geq 1$ such that $d^{1/p}$ dominates the initial metric ρ on \mathbb{X} . Then there exists a unique IPM π for \mathbb{X} , and, for every $x \in \mathbb{X}$,

$$d(P_n(x, \cdot), \pi) \leq e^{-\theta n} \int_{\mathbb{X}} d(x, y) \pi(dy), \quad n \geq 1. \quad (2.10)$$

Proof. Consider a sequence $P_n(x_*, \cdot)$, $n \geq 1$. Take a pair m, n with, say, $m < n$, and consider the chain Z with $Z_0^1 = x_*$ and Z_0^2 having the law $P_{n-m}(x_*, \cdot)$. Then

$$d(P_m(x_*, \cdot), P_n(x_*, \cdot)) \leq e^{-\theta m} \mathbb{E}d(Z_0^1, Z_0^2) = e^{-\theta m} \int_{\mathbb{X}} d(x_*, y) P_{n-m}(x_*, dy).$$

By Jensen's inequality, this yields the bound

$$\rho(P_m(x, \cdot), P_n(x, \cdot)) \leq C_* e^{-\theta m/p}, \quad m < n;$$

at this point, we use the fact that $d^{1/p}$ dominates ρ . Hence the sequence $P_n(x_*, \cdot)$, $n \geq 1$, is Cauchy w.r.t. the coupling distance ρ and therefore weakly converges to some $\pi \in \mathcal{P}(\mathbb{X})$. It is easy to verify that π is an IPM and (2.10) holds; we leave this as an exercise for the reader. \square

In the continuous time setting, we call a Markov process *d-dissipative* if there exist $\theta > 0$ and $T > 0$ such that, for every $x, x' \in \mathbb{X}$,

$$d(P_T(x, \cdot), P_T(x', \cdot)) \leq e^{-\theta T} d(x, x')$$

and, in addition, if there exists $C > 0$ such that

$$d(P_t(x, \cdot), P_t(x', \cdot)) \leq C d(x, x'), \quad t \leq T.$$

For such a process, straightforward analogues of Proposition 2.22 and Proposition 2.23 are available, with obvious changes: the time variable n should be changed to t and the additional multiplier $Ce^{\theta T}$ should appear at the right hand sides of the bounds (2.8)

and (2.10); we leave the details for the reader.

Coming back to Markov processes defined by SDEs, we note that the dissipativity feature appears also to be rather insensitive w.r.t. to the structure of the noise term. One can see this already in Proposition 2.21, where any non-degeneracy assumption on the diffusion matrix is not involved. A similar argument is applicable in a wide variety of models, e.g., infinite-dimensional SDEs including Lévy driven SDEs ([7, Chapter 11.5]; [48, Chapter 16.2]), SDDEs ([23]), etc. Below we outline a version of this argument for SDDEs, which in particular proves the weak ergodicity we have claimed in Example 2.3.

Consider an SDDE in \mathbb{R}^m of the form

$$dX_t = a(X_t) dt + b(X_{t-r}) dW_t, \quad (2.11)$$

where the delay constant $r > 0$ is fixed and functions $a : \mathbb{R}^m \rightarrow \mathbb{R}^m$, $b : \mathbb{R}^m \rightarrow \mathbb{R}^{m \times k}$ are Lipschitz continuous. Take two initial values $h, h' \in C([-r, 0], \mathbb{R}^m)$ and consider respective processes X, X' and $\Delta = X - X'$. Denote

$$\alpha = \sup_{x \neq y} \frac{(a(x) - a(y), x - y)}{|x - y|^2},$$

and assume that $\alpha < 0$; that is, the drift dissipativity condition (2.6) holds true. Assume also that, for some $p \geq 2$,

$$\frac{mp(p-1)\|b\|_{\text{Lip}}^2}{2} < -\alpha.$$

Denote

$$\theta = -\alpha p - \frac{mp(p-1)\|b\|_{\text{Lip}}^2}{2} > 0$$

and apply the Itô formula to the process Δ_t , $t \geq 0$, with the function $F(t, x) = e^{\theta t} |x|^p$:

$$\begin{aligned} e^{\theta t} |\Delta_t|^p &= |\Delta_0|^p + \int_0^t e^{\theta s} \left(\theta |\Delta_s|^p + p |\Delta_s|^{p-2} (a(X_s) - a(X'_s), \Delta_s) \right) ds \\ &\quad + \frac{p}{2} \int_0^t e^{\theta s} |\Delta_s|^{p-2} \sum_{i,j=1}^m \sum_{l=1}^k (b_{il}(X_{s-r}) - b_{il}(X'_{s-r})) \\ &\quad \times (b_{ji}(X_{s-r}) - b_{ji}(X'_{s-r})) \left(\delta_{ij} + (p-2) \frac{\Delta_s^i \Delta_s^j}{|\Delta_s|^2} \right) ds \end{aligned}$$

$$+ p \int_0^t e^{\theta s} |\Delta_s|^{p-2} (\Delta_s, b(X_{s-r}) - b(X'_{s-r})) \, dW_s.$$

The same argument from the beginning of this section leads to the following inequality:

$$\begin{aligned} \mathbb{E} e^{\theta t} |\Delta_t|^p &\leq |\Delta_0|^p + \mathbb{E} \int_0^t e^{\theta s} \left(\theta |\Delta_s|^p + p |\Delta_s|^{p-2} (a(X_s) - a(X'_s), \Delta_s) \right) \, ds \\ &\quad + \frac{mp(p-1)}{2} \mathbb{E} \int_0^t e^{\theta s} |\Delta_s|^{p-2} \|b(X_{s-r}) - b(X'_{s-r})\|^2 \, ds. \end{aligned}$$

By the Young inequality, the term under the second integral does not exceed

$$\|b\|_{\text{Lip}}^2 \left(\frac{(p-2)|\Delta_s|^p}{p} + \frac{2|\Delta_{s-r}|^p}{p} \right).$$

Denote

$$\alpha = \sup_{x \neq y} \frac{(a(x) - a(y), x - y)}{|x - y|^2},$$

and observe that

$$\int_0^t |\Delta_{s-r}|^p \, ds = \int_{-r}^{t-r} |\Delta_s|^p \, ds \leq \int_0^t |\Delta_s|^p \, ds + \int_{-r}^0 |h_s - h'_s|^p \, ds.$$

Then we can continue the above estimate and write

$$\begin{aligned} \mathbb{E} e^{\theta t} |\Delta_t|^p &\leq |\Delta_0|^p + m(p-1) \|b\|_{\text{Lip}}^2 \int_{-r}^0 |h_s - h'_s|^p \, ds \\ &\quad + \left(\alpha + \theta + \frac{mp(p-1) \|b\|_{\text{Lip}}^2}{2} \right) \mathbb{E} \int_0^t e^{\theta s} |\Delta_s|^p \, ds. \end{aligned}$$

Recall that $\Delta_0 = h_0 - h'_0$; hence, from our choice of θ , we finally get

$$\mathbb{E} |\Delta_t|^p \leq C_1 e^{-\theta t} \|h\|_{C(-r,0)}^p, \quad t \geq 0, \quad (2.12)$$

where

$$\|h\|_{C(-r,0)} = \sup_{s \in [-r,0]} |h(s)|, \quad C_1 = 1 + rm(p-1) \|b\|_{\text{Lip}}^2.$$

Now we are ready to give the final bound for the $\|\cdot\|_{C(-r,0)}$ -norm for the values of the

segment process $\{\Delta(t)\}$; that is,

$$\|\Delta(t)\|_{C(-r,0)} = \sup_{s \in [t-r,t]} |\Delta_s|.$$

Write

$$\begin{aligned} |\Delta_{t_2} - \Delta_{t_1}|^2 &= \int_{t_1}^{t_2} \left(2(a(X_s) - a(X'_s), \Delta_s) + \sum_{i=1}^m \sum_{l=1}^k (b_{il}(X_{s-r}) - b_{il}(X'_{s-r}))^2 \right) ds \\ &\quad + 2 \int_{t_1}^{t_2} (\Delta_s, b(X_{s-r}) - b(X'_{s-r})) dW_s = (I_{t_2} - I_{t_1}) + (M_{t_2} - M_{t_1}). \end{aligned}$$

Since

$$|I_{t_2} - I_{t_1}| \leq C_2 \int_{t_1}^{t_2} |\Delta_s|^2 ds,$$

it follows from the Hölder inequality that, for any $p \geq 2$ such that (2.12) holds,

$$\mathbf{E} \sup_{s \in [t-r,t]} |I_s - I_{t-r}|^{p/2} \leq C_3 e^{-\theta t}.$$

On the other hand, the quadratic characteristic of the martingale M_t has the form $d\langle M \rangle_t = f_t dt$ with

$$f_t \leq C_4 |\Delta_t| |\Delta_{t-r}|.$$

Then by the Burkholder-Davis-Gundy inequality, for any $p > 2$ such that (2.12) holds,

$$\mathbf{E} \sup_{s \in [t-r,t]} |M_s - M_{t-r}|^{p/2} \leq C_5 e^{-\theta t}.$$

Summarizing all the above, we get the following sufficient condition for the segment process $X(t)$, $t \geq 0$, to be dissipative.

Proposition 2.24 Assume

$$\frac{m(p-1)\|b\|_{\text{Lip}}^2}{2} < -\alpha = \sup_{x \neq y} \frac{(a(x) - a(y), x - y)}{|x - y|^2}. \quad (2.13)$$

Then, for every

$$p \in \left(2, 1 + \frac{2\alpha}{m\|b\|_{\text{Lip}}^2} \right),$$

there exists a positive C such that, for every two initial values $h, h' \in C(-r, 0)$, respective solutions X, X' to (2.11) satisfy

$$E\|X(t) - X'(t)\|_{C(-r,0)}^p \leq Ce^{-\theta_p t} \|h - h'\|_{C(-r,0)}^p, \quad t \geq 0,$$

where

$$\theta_p = -\alpha p - \frac{mp(p-1)\|b\|_{\text{Lip}}^2}{2} > 0.$$

Repeating the argument from the proof of Proposition 2.23, one can now easily deduce the following.

Theorem 2.25 Let coefficients of the equation (2.11) satisfy (2.13). Assume in addition that, for some $h \in C(-r, 0)$, the respective solution to (2.11) satisfies

$$\sup_{t \geq 0} E\|X(t)\|_{C(-r,0)}^p < \infty \tag{2.14}$$

for some

$$p \in \left(2, 1 + \frac{2\alpha}{m\|b\|_{\text{Lip}}^2} \right).$$

Then the segment process $X(t)$, $t \geq 1$, has a unique IPM π , and the following rate of convergence holds true in terms of the coupling distance corresponding to the distance-like function $d_r(h, h') = \|h - h'\|^p$:

$$d_p(P_t(h, \cdot), \pi) \leq Ce^{-\theta_p t} \int_{C(-r,0)} \|h - h'\|^p \pi(dh'), \quad t \geq 0.$$

Note that it is easy to give sufficient conditions for (2.14), formulated in terms of the coefficients. Calculations here are similar to those used in Section 1.6, with the modifications similar to those used in the proof of Proposition 2.24; we leave details for the reader.

.....

2.4 General Harris-type theorem

The condition for the Markov chain to be dissipative w.r.t. d is, in a sense, a condition for the chain to be “uniformly ergodic” w.r.t. the distance-like function d : this analogy

corresponds well with the analogy between Proposition 2.22 and Proposition 2.23 on one hand, and Theorem 1.3 on the other hand. In this section we extend this analogy; it appears that the global condition for the chain to be dissipative w.r.t. d (i.e., on the distance-like function d to be contracting) can be localized in a way very similar to that explained in Section 1.3.

We say that a distance-like function d is *contracting* for a Markov chain X on a set K if there exists $\theta > 0$ such that

$$d(P(x_1, \cdot), P(x_2, \cdot)) \leq e^{-\theta} d(x_1, x_2), \quad x_1, x_2 \in K.$$

We also say that a distance-like function d is *non-expanding* for a Markov chain X if

$$d(P(x_1, \cdot), P(x_2, \cdot)) \leq d(x_1, x_2), \quad x_1, x_2 \in \mathbb{X}.$$

Theorem 2.26 Let a continuous distance-like function d be non-expanding for a given Markov chain X and be contracting for this chain on some set K . Assume that, for the chain X and set K , the conditions of Theorem 1.15 hold true. Then for every $p > 1$, $\sigma > 1$ there exist $\delta > 0$, $C > 0$ such that

$$d_p(P_n(x_1, \cdot), P_n(x_2, \cdot)) \leq (\Phi^{-1}(\delta t))^{-1/q\sigma} (V(x_1) + V(x_2))^{1/q\sigma} \times d_p(x_1, x_2), \quad n \geq 0, \quad (2.15)$$

where $1/p + 1/q = 1$, function Φ^{-1} is an inverse of (1.25), and by the symbol d_p we denote both the distance-like function $d^{1/p}$ and the corresponding minimal probability distance.

In addition, there exists a unique IPM π for X which satisfies

$$\int_{\mathbb{X}} V \, d\pi < \infty, \quad (2.16)$$

and

$$d_p(P_n(x, \cdot), \pi) \leq (\Phi^{-1}(\delta t))^{1/q\sigma} \left(V(x) + \int_{\mathbb{X}} V \, d\pi \right)^{1/q\sigma} \times \left(\int_{\mathbb{X}} d(x, y) \pi(dy) \right)^{1/p}, \quad n \geq 0.$$

Before proceeding with the proof, let us engage the above in discussion. In the case of total variation distance (that is, the coupling distance corresponding to $d(x, y) = \mathbb{I}_{x \neq y}$), the condition on d to be contracting holds true for *any* Markov chain. Hence the assumptions imposed on d should be understood as a one-to-one analogue of the local Dobrushin condition from Theorem 1.13. Therefore Theorem 2.26, in the framework of weak convergence, exhibits the same effect as discussed in Section 1.3. Namely, the weak local irreducibility combined with the recurrence assumption provide weak ergodicity with explicit bounds for the rate of weak convergence to an IPM. This type of theorem was first developed in [23], where it was called a general *Harris type theorem*. In [23], a linear Lyapunov-type condition and corresponding exponential ergodic bounds are considered. More general Lyapunov-type conditions and respective ergodic bounds were treated in [6].

Proof. The strategy of the proof is very similar to that of Theorem 1.13. Consider a Markov chain $Z = (Z^1, Z^2)$ with the transition probability Q such that for every $x, x' \in \mathbb{X}$ the measure $Q((x, x'), \cdot)$ belongs to the set of d -optimal couplings of $P(x, \cdot)$ and $P(x', \cdot)$; such a kernel exists due to Proposition 2.19. For a given $x_1, x_2 \in \mathbb{X}$ consider the law $\mathbb{P}_{(x_1, x_2)}^Z$ of this process with $Z_0 = (x_1, x_2)$. Then for every $n \geq 1$, the pair Z_n^1, Z_n^2 gives a coupling for $P_n(x_1, \cdot), P_n(x_2, \cdot)$, and therefore

$$d_p(P_n(x_1, \cdot), P_n(x_2, \cdot)) \leq \mathbb{E}_{(x_1, x_2)}^Z d^{1/p}(Z_n^1, Z_n^2).$$

Denote by $S_k, k \geq 1$, the time moments of consecutive visits of Z to $K \times K$. Then we can continue the previous bound:

$$\begin{aligned} d_p(P_n(x_1, \cdot), P_n(x_2, \cdot)) &\leq \mathbb{E}_{(x_1, x_2)}^Z d^{1/p}(Z_n^1, Z_n^2) \mathbb{I}_{S_1 \geq n} \\ &\quad + \sum_{k=1}^{\infty} \mathbb{E}_{(x_1, x_2)}^Z d^{1/p}(Z_n^1, Z_n^2) \mathbb{I}_{n \in (S_k, S_{k+1}]}. \end{aligned} \quad (2.17)$$

Let us separately analyse the terms in the right hand side of the previous inequality. By Theorem 1.15, conditions **R.1** and **R.2** of Theorem 1.13 are satisfied with

$$\lambda(t) = \Phi^{-1}(\alpha t), \quad \alpha = 1 - \frac{2C}{\phi(1 + \inf_{x \notin K} V(x))}$$

for *any* Markov coupling Z and therefore for the process Z under consideration. Then the

first term in (2.17) can be estimated simply as

$$\begin{aligned} \mathbb{E}_{(x_1, x_2)}^Z d^{1/p}(Z_n^1, Z_n^2) \mathbb{1}_{S_1 \geq n} &\leq \left(\mathbb{E}_{(x_1, x_2)}^Z d(Z_n^1, Z_n^2) \right)^{1/p} (\mathbb{P}(S_1 \geq n))^{1/q} \\ &\leq \left(\mathbb{E}_{(x_1, x_2)}^Z d(Z_n^1, Z_n^2) \right)^{1/p} \left(\mathbb{E}_{(x_1, x_2)}^Z \lambda^\sigma(S_1) \right)^{1/q\sigma} (\lambda(n))^{-1/q\sigma} \\ &\leq \left(\mathbb{E}_{(x_1, x_2)}^Z d(Z_n^1, Z_n^2) \right)^{1/p} (V(x_1) + V(x_2))^{1/q\sigma} (\lambda(n))^{-1/q\sigma}. \end{aligned}$$

Because the kernel Q is optimal w.r.t. d and d is non-expanding, we have

$$\mathbb{E}_{(x_1, x_2)}^Z \left(d(Z_n^1, Z_n^2) \middle| \mathcal{F}_{n-1}^Z \right) = \int_{\mathbb{X} \times \mathbb{X}} d(y, y') Q((x, x'), dy dy') \Big|_{(x, x') = Z_{n-1}} \leq d(Z_{n-1}^1, Z_{n-1}^2),$$

and consequently

$$\mathbb{E}_{(x_1, x_2)}^Z d(Z_n^1, Z_n^2) \leq \mathbb{E}_{(x_1, x_2)}^Z d(Z_{n-1}^1, Z_{n-1}^2) \leq \dots \leq d(x_1, x_2).$$

This finally gives the bound for the first term in the right hand side of (2.17):

$$\mathbb{E}_{(x_1, x_2)}^Z d^{1/p}(Z_n^1, Z_n^2) \mathbb{1}_{S_1 \geq n} \leq (d(x_1, x_2))^{1/p} (V(x_1) + V(x_2))^{1/q\sigma} (\lambda(n))^{-1/q\sigma}.$$

The idea of how to get a bound for subsequent terms is mainly the same but some new technicalities arise. Observe that, by its construction, the chain Z has the property that once its coordinates coincide, they stay equal afterwards. Because $d(x, x) = 0$, this means that

$$\mathbb{E}_{(x_1, x_2)}^Z d^{1/p}(Z_n^1, Z_n^2) \mathbb{1}_{n \in (S_k, S_{k+1}]} = \mathbb{E}_{(x_1, x_2)}^Z d^{1/p}(Z_n^1, Z_n^2) \mathbb{1}_{n \in (S_k, S_{k+1}]} \mathbb{1}_{L > R_k},$$

where

$$L = \inf\{n : Z_n^1 = Z_n^2\}$$

is the coupling time for Z and $R_k = S_k + 1$. This gives

$$\begin{aligned} \mathbb{E}_{(x_1, x_2)}^Z d^{1/p}(Z_n^1, Z_n^2) \mathbb{1}_{n \in (S_k, S_{k+1}]} &\leq \left(\mathbb{E}_{(x_1, x_2)}^Z d(Z_n^1, Z_n^2) \mathbb{1}_{S_k < n} \right)^{1/p} \\ &\quad \times \left(\mathbb{P}_{(x_1, x_2)}^Z (L > R_k, S_{k+1} \geq n) \right)^{1/q}. \end{aligned}$$

We have

$$P_{(x_1, x_2)}^Z(L > R_k, S_{k+1} \geq n) \leq \lambda^{-1/\sigma}(\gamma n) E_{(x_1, x_2)}^Z \mathbb{I}_{L > R_k} \lambda^{1/\sigma}(\gamma R_{k+1}).$$

because $R_{k+1} = S_{k+1} + 1$ and λ is monotone. Using (1.21) and (1.22) with p replaced by σ , we get

$$E \mathbb{I}_{L > R_k} \lambda^{1/\sigma}(\gamma R_{k+1}) \leq \vartheta(\gamma)(V(x_1) + V(x_2))^{1/\sigma},$$

where $\vartheta(\gamma)$ is some function which can be given explicitly and is such that $\vartheta(\gamma) \rightarrow 1$, $\gamma \rightarrow 0$.

On the other hand, on the set $\{S_k < n\}$ the chain Z visits the set $K \times K$, where d is contracting, at least k times. Therefore

$$E_{(x_1, x_2)}^Z d(Z_n^1, Z_n^2) \mathbb{I}_{S_k < n} \leq e^{-\theta k} d(x_1, x_2),$$

where $\theta > 0$ is the constant from the condition on d to be contracting for X on the set K . This inequality can be obtained using induction in k ; we leave the details for the reader.

Exercise 2.27 Give a rigorous proof of the above inequality.

Hence, we can finally write, for any $\gamma \in (0, 1)$,

$$d_p(P_n(x_1, \cdot), P_n(x_2, \cdot)) \leq (d(x_1, x_2))^{1/p} (V(x_1) + V(x_2))^{1/q\sigma} \\ \times (\lambda(\gamma n))^{-1/q\sigma} \left[1 + \sum_{k=1}^{\infty} e^{-\theta k} \vartheta^k(\gamma) \right].$$

Taking γ small enough, we can make $e^{-\theta} \vartheta(\gamma) < 1$, which completes the proof of (2.15); now

$$\delta = \alpha\gamma \quad \text{and} \quad C = 1 + \sum_{k=1}^{\infty} e^{-\theta k} \vartheta^k(\gamma).$$

The rest of the proof is similar to the corresponding part of the proof of Theorem 1.13 and is omitted. \square

Chapter 3

Limit Theorems

In this chapter we study limit theorems for functionals of a Markov chain, with the main assumptions formulated in the terms of the ergodic rates of the chain. This field of applications of ergodic rates is both, very natural and very common, and actually dates back to the seminal A.A. Markov's papers [40], [41]. Section 3.1 is an introductory one, where we start with an outline of basic notions and tools, which apply very naturally when ergodic rates in the total variation distance are available. The main part of the chapter (Sections 3.2–3.4) is devoted to the less studied case of weakly ergodic Markov chains. To treat this case, we develop an extension of the “martingale approach” which dates back to [46]. This extension is insensitive w.r.t. the structure of the Markov process and is well applicable to Markov models with intrinsic memory.

.....

3.1 Preliminaries: Covariance and Mixing Coefficient

Throughout this chapter, we assume an ergodic Markov chain X_n , $n \geq 0$ be fixed, and denote by $P(x, dy)$ its one-step transition probability and by π its unique IPM.

Let X possess uniform ergodic rate in the total variation distance. In Section 1.5 we observed that such a rate then should be exponential; that is, there exist $C > 0$, $\rho \in (0, 1)$ such that

$$\|P_n(x, dy) - \pi(dy)\|_{TV} \leq C\rho^n, \quad n \geq 1, x \in \mathbb{X}. \quad (3.1)$$

Choose a *bounded* function $f : \mathbb{X} \rightarrow \mathbb{R}$, and consider a sequence

$$\frac{1}{n} \sum_{k=1}^n f(X_k), \quad n \geq 1.$$

Because

$$\mathbb{E}f(X_k) = \mathbb{E} \int_{\mathbb{X}} f(y) P_k(X_0, dy), \quad k \geq 1,$$

and

$$\mathbb{E}[f(X_k)|X_0, \dots, X_j] = \int_{\mathbb{X}} f(y) P_{k-j}(X_j, dy), \quad k > j,$$

it is easy to deduce from (3.1) that

$$\mathbb{E}f(X_k) \rightarrow \int_{\mathbb{X}} f d\pi, \quad k \rightarrow \infty, \quad (3.2)$$

$$\mathbb{E}f(X_j)f(X_k) \rightarrow \left(\int_{\mathbb{X}} f d\pi \right)^2, \quad j \rightarrow \infty, k - j \rightarrow \infty. \quad (3.3)$$

Hence

$$\text{Cov}(f(X_j), f(X_k)) \rightarrow 0, \quad j \rightarrow \infty, k - j \rightarrow \infty, \quad (3.4)$$

which one can use to derive easily the following version of the Law of Large Numbers (LLN),

$$\frac{1}{n} \sum_{k=1}^n f(X_k) \rightarrow \int_{\mathbb{X}} f d\pi, \quad n \rightarrow \infty, \quad (3.5)$$

where the convergence holds in the mean square sense.

Exercise 3.1 Please, prove (3.2), (3.3), and (3.5).

The above argument can be summarized as follows: if the chain X satisfies (3.1), which should be understood as a kind of a *stabilization* property for the transition probabilities of the chain, then the covariances for distant values of the sequence $f(X_n)$, $n \geq 0$ vanish, which yields the LLN for this sequence. We will see below that this simple argument is quite flexible, and remains useful in various settings where the stabilization property holds true in a weaker sense than (3.1). For instance, if instead of (3.1) we assume a non-uniform bound

$$\|P_n(x, dy) - \pi(dy)\|_{\text{TV}} \leq V(x)r_n, \quad n \geq 1, \quad (3.6)$$

with $r_n \rightarrow 0$, then the relation (3.4) and therefore the LLN (3.5) would hold true under an auxiliary assumption

$$\sup_{n \geq 0} \text{EV}(X_n) < \infty.$$

Exercise 3.2 Please, prove this.

If the function f is *centered* in the sense that

$$\int_{\mathbb{X}} f \, d\pi = 0,$$

it is natural to expect the sequence

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n f(X_k), \quad n \geq 1,$$

to converge weakly to a normal distribution; that is, the Central Limit Theorem (CLT) to hold true. One possible way to prove such an assertion under the assumption (3.1) is to establish analogues of (3.2), (3.3) for higher order mixed moments of the sequence $f(X_n)$, $n \geq 0$, and to use then the classical *method of moments*. Such an argument was used e.g. in [41]; note that the CLT proved by A.A. Markov in this paper for *dependent* random variables was conceptually new, because at that time a common point of view was that the CLT is essentially an effect which appears only for independent sequences.

Method of moments is historically the first method for proving the CLT, it is but rather cumbersome. A modification of the other classical method based on the characteristic functions for the case of *dependent* random variables was developed by S.N. Bernstein; cf. [2]. For an excellent and classical exposition of the CLT with the proof based on *Bernstein's block method* we refer to [25], Chapter 18. Below we briefly outline the main results available within this approach.

Let $\xi_n, n \in \mathbb{Z}$, be a strictly stationary sequence of random variables such that $E\xi_0 = 0$. Denote $\mathcal{F}_{\leq n} = \sigma(\xi_k, k \leq n)$, $\mathcal{F}_{\geq n} = \sigma(\xi_k, k \geq n)$, and define the *uniform mixing* (or *Ibragimov's coefficient* for the sequence $\{\xi_n\}$ by

$$\phi(n) = \sup_{A \in \mathcal{F}_{\geq n}, B \in \mathcal{F}_{\leq 0}, \text{P}(B) > 0} |\text{P}(A) - \text{P}(A|B)|, \quad n \geq 0,$$

where $\text{P}(A|B) = \text{P}(A \cap B) / \text{P}(B)$ is the conditional probability.

Theorem 3.3 ([25], Theorem 18.5.2) Let the strictly stationary sequence satisfy $E\xi_0 = 0$, $E\xi_0^2 < \infty$, and

$$\sum_n (\phi(n))^{1/2} < \infty. \quad (3.7)$$

Then the sum

$$\sigma^2 = E\xi_0^2 + 2 \sum_{k=1}^{\infty} E\xi_k \xi_0 \quad (3.8)$$

converges and

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n \xi_k \Rightarrow \mathcal{N}(0, \sigma^2), \quad n \rightarrow \infty.$$

The heuristics of this statement is clear and simple. The coefficients $\phi(n)$, in a sense, measure the rate of dependence between $\mathcal{F}_{\geq k+n}$ and $\mathcal{F}_{\leq k}$ for any k . Condition (3.7) means that such a dependence vanishes as $n \rightarrow \infty$ sufficiently fast; and this is a principal assumption in the above version of the CLT.

We will see below that the uniform mixing coefficient, in some cases, is too strong and hardly can be used as a sufficient condition. Define the *strong mixing* (or *complete regularity*, or *Rosenblatt's*) coefficient by

$$\alpha(n) = \sup_{A \in \mathcal{F}_{\geq n}, B \in \mathcal{F}_{\leq 0}} |P(A \cap B) - P(A)P(B)|, \quad n \geq 0.$$

The following theorem is of the same virtue as Theorem 3.3, but the rate of dependence therein is bounded by means of the strong mixing coefficient instead of the uniform one; this leads to stronger integrability assumptions on the sequence $\{\xi_n\}$.

Theorem 3.4 ([25], Theorem 18.5.3) Let for some $\delta > 0$ the strictly stationary sequence satisfy $E\xi_0 = 0$, $E|\xi_0|^{2+\delta} < \infty$, and

$$\sum_n (\alpha(n))^{\delta/(2+\delta)} < \infty. \quad (3.9)$$

Then the sum (3.8) converges and

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n \xi_k \Rightarrow \mathcal{N}(0, \sigma^2), \quad n \rightarrow \infty.$$

Let us return to our Markov setting. Consider a *strictly stationary version* of the chain X ; that is, a Markov chain X defined on \mathbb{Z} with the one-step transition probability $P(x, dy)$

and all one-dimensional distributions equal to the IPM π .

Exercise 3.5 Use the Kolmogorov consistency theorem to prove that, given a Markov chain with an IPM π , one can construct on a proper probability space its strictly stationary version $X_n, n \in \mathbb{Z}$.

For every $f: \mathbb{X} \rightarrow \mathbb{R}$ the real-valued sequence $\xi_n = f(X_n), n \in \mathbb{Z}$, is strictly stationary as well. To apply the above general sufficient conditions for the CLT, one should verify the principal assumptions (3.7) or (3.9). Below we denote by the same symbols $\alpha(n), \phi(n)$ the mixing coefficients for the sequence $\{X_n\}$ itself; clearly, they dominate respective mixing coefficients for the sequence $\{\xi_n\}$.

Recall that the Markov property of X is equivalent to the following: for every n and every $A \in \mathcal{F}_{\geq n}, B \in \mathcal{F}_{\leq n}$,

$$P(A \cap B | X_n) = P(A | X_n)P(B | X_n).$$

Then for any $n \geq 0$ and any $A \in \mathcal{F}_{\geq n}, B \in \mathcal{F}_{\leq 0}$ we have

$$P(A | \mathcal{F}_{\leq n}) = \psi_{\leq n, A}(X_n), \quad P(B | \mathcal{F}_{\geq 0}) = \psi_{\geq 0, B}(X_0)$$

with some measurable $\psi_{\leq n, A}, \psi_{\geq 0, B}$ which take their values in $[0, 1]$.

Exercise 3.6 Please, prove the above identities.

Then

$$\begin{aligned} P(A \cap B) &= E[\mathbb{I}_B P(A | \mathcal{F}_{\leq n})] = E[\mathbb{I}_B \psi_{\leq n, A}(X_n)] \\ &= E[\psi_{\leq n, A}(X_n) P(B | \mathcal{F}_{\geq 0})] = E[\psi_{\leq n, A}(X_n) \psi_{\geq 0, B}(X_0)] \\ &= \int_{\mathbb{X}} \int_{\mathbb{X}} \psi_{\leq n, A}(y) P_n(x, dy) \psi_{\geq 0, B}(x) \pi(dx). \end{aligned}$$

On the other hand,

$$P(A) = \int_{\mathbb{X}} \psi_{\leq n, A}(y) \pi(dy), \quad P(B) = \int_{\mathbb{X}} \psi_{\geq 0, B}(x) \pi(dx),$$

hence

$$|P(A \cap B) - P(A)P(B)| \leq \int_{\mathbb{X}} \left| \int_{\mathbb{X}} \psi_{\leq n, A}(y) (P_n(x, dy) - \pi(dy)) \right| \psi_{\geq 0, B}(x) \pi(dx)$$

$$\leq \frac{1}{2} \int_{\mathbb{X}} \|P_n(x, dy) - \pi(dy)\|_{\text{TV}} \psi_{\geq 0, B}(x) \pi(dx). \quad (3.10)$$

This, on one hand, yields the bound

$$|\mathbb{P}(A) - \mathbb{P}(A|B)| = \frac{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|}{\mathbb{P}(B)} \leq \frac{1}{2} \sup_{x \in \mathbb{X}} \|P_n(x, dy) - \pi(dy)\|_{\text{TV}},$$

and therefore

$$\phi(n) \leq \frac{1}{2} \operatorname{esssup}_{x \in \mathbb{X}} \|P_n(x, dy) - \pi(dy)\|_{\text{TV}},$$

where the essential supremum is taken w.r.t. π . On the other hand, (3.10) simply yields

$$\alpha(n) \leq \frac{1}{2} \int_{\mathbb{X}} \|P_n(x, dy) - \pi(dy)\|_{\text{TV}} \pi(dx).$$

We can summarize the above argument as follows.

- ◇ If X is uniformly ergodic in the total variation distance, then it is uniformly mixing at exponential rate, i.e.

$$\phi(n) \leq C\rho^n$$

with some $C \geq 0$, $\rho \in (0, 1)$, and then (3.7) holds true.

- ◇ If X possesses a non-uniform ergodic rate in the total variation distance (3.6) and in addition it is assumed that the function V involved therein is integrable,

$$\int_{\mathbb{X}} V d\pi < \infty,$$

then X satisfies

$$\alpha(n) \leq Cr_n$$

with some $C > 0$. Consequently, if

$$\sum_{n=1}^{\infty} r_n^{\delta/(2+\delta)} < \infty,$$

assertion (3.9) holds true.

Observe that the above bound for $\phi(n)$, in fact, is an optimal one: if $A = \{X_n \in C\}$, $B = \{X_0 \in D\}$, then $\psi_{\leq n, A} = 1_C$, $\psi_{\geq 0, B} = 1_D$, and taking the supremum over all the pairs

C, D yields the same *lower* bound for $\phi(n)$; we leave the details of this calculation as an exercise to the reader. Typically, there is no actual difference between taking the supremum over x and essential supremum over x w.r.t. π , hence one can conclude that the uniform mixing coefficient applies well, only in the case of X being uniformly ergodic in the total variation distance. Otherwise the assumptions of Theorem 3.3 typically would fail, and one should use Theorem 3.4 instead, where the main assumption (3.9) follows from a non-uniform ergodic rate (3.6), if V is integrable and $r_n \rightarrow 0$ sufficiently fast.

For Markov chains with intrinsic memory (see Section 2.1),

$$\alpha(n) \geq \alpha > 0, \quad n \geq 1,$$

is usually satisfied.

Exercise 3.7 For stationary versions of the processes in Example 2.1 and Example 2.3 consider their time discretizations $X_n, n \geq 0$ and construct $\alpha > 0$, and $C, D \in \mathcal{X}$ such that

$$|\mathbb{P}(X_n \in C, X_0 \in D) - \mathbb{P}(X_n \in C)\mathbb{P}(X_n \in D)| \geq \alpha, \quad n \geq 1.$$

Hence one can not apply directly the above classical results in order to prove the CLT for functionals of Markov chains which are only *weakly* ergodic. Below we show that this difficulty is of a technical nature, rather than of a conceptual one, and explain a practical way to prove limit theorems for functionals of weakly ergodic Markov chains.

• • • • •

3.2 Covariances and the LLN, revisited

In this section we explain a natural way to deduce the relation (3.4), and consequently to prove the LLN (3.5), for a chain X which is ergodic only in a weak sense. In what follows, we assume that the chain X satisfies

$$d(P_n(x, dy), \pi(dy)) \leq V(x)r_n, \quad n \geq 0, \quad (3.11)$$

where $V : \mathbb{X} \rightarrow [1, \infty)$, $r_n \rightarrow 0$, and d denotes the coupling distance on $\mathcal{P}(\mathbb{X})$ which corresponds to some distance-like function d on \mathbb{X} . In addition, to make the exposition most transparent, we restrict the consideration to the case of stationary X ; that is, we assume $X_0 \sim \pi$. We outline the (general) non-stationary case at the end of this section.

Consider a function $W : \mathbb{X} \rightarrow \mathbb{R}^+$, and define for $\gamma \in (0, 1]$ the *weighted Hölder class* $H_{\gamma, W}(\mathbb{X}, d)$ w.r.t. d with the index γ and the weight W as the set of functions $f : \mathbb{X} \rightarrow \mathbb{R}$ such that

$$\|f\|_{d, \gamma, W} = \sup_{x_1 \neq x_2} \frac{|f(x_1) - f(x_2)|}{d^\gamma(x_1, x_2)(W(x_1) + W(x_2))^{1-\gamma}} < \infty.$$

Here and below we use the convention $a^0 = 1$, $a \in \mathbb{R}^+$; hence for $\gamma = 1$ the weight W is inessential, and $H_{1, W}(\mathbb{X}, d) = H_1(\mathbb{X}, d)$ is just the Lipschitz class w.r.t. d .

The following statement is very simple, but it plays a key role in the sequel.

Proposition 3.8 Let a function f belong to $H_{\gamma, W}(\mathbb{X}, d)$ for some $\gamma \in (0, 1]$. Then for any $\mu, \nu \in \mathcal{P}(\mathbb{X})$,

$$\left| \int_{\mathbb{X}} f \, d\mu - \int_{\mathbb{X}} f \, d\nu \right| \leq \|f\|_{d, \gamma, W} (d(\mu, \nu))^\gamma \left(\int_{\mathbb{X}} W \, d\mu + \int_{\mathbb{X}} W \, d\nu \right)^{1-\gamma}. \quad (3.12)$$

Proof. For any coupling $(\xi, \eta) \in \mathcal{C}(\mu, \nu)$ we have

$$\begin{aligned} \left| \int_{\mathbb{X}} f \, d\mu - \int_{\mathbb{X}} f \, d\nu \right| &= |\mathbb{E}f(\xi) - \mathbb{E}f(\eta)| \leq \mathbb{E}|f(\xi) - f(\eta)| \\ &\leq \|f\|_{d, \gamma, W} \mathbb{E} \left[d^\gamma(\xi, \eta) (W(\xi) + W(\eta))^{1-\gamma} \right]. \end{aligned}$$

If $\gamma < 1$, we apply the Hölder inequality with $p = 1/\gamma$:

$$\begin{aligned} \left| \int_{\mathbb{X}} f \, d\mu - \int_{\mathbb{X}} f \, d\nu \right| &\leq \|f\|_{d, \gamma, W} (\mathbb{E}d(\xi, \eta))^\gamma (\mathbb{E}W(\xi) + \mathbb{E}W(\eta))^{1-\gamma} \\ &= \|f\|_{d, \gamma, W} (\mathbb{E}d(\xi, \eta))^\gamma \left(\int_{\mathbb{X}} W \, d\mu + \int_{\mathbb{X}} W \, d\nu \right)^{1-\gamma}. \end{aligned}$$

For $\gamma = 1$ the same bound holds true directly. Taking the infimum in this bound w.r.t. all the couplings $(\xi, \eta) \in \mathcal{C}(\mu, \nu)$, we get the required statement. \square

Now one can easily deduce the relation (3.4).

Proposition 3.9 Let $f \in H_{\gamma, W}(\mathbb{X}, d)$ and

$$\int_{\mathbb{X}} f^2 \, d\pi < \infty, \quad \int_{\mathbb{X}} V^2 \, d\pi < \infty, \quad \int_{\mathbb{X}} W^2 \, d\pi < \infty.$$

Then

$$\begin{aligned} & \left| \text{Cov}(f(X_j), f(X_k)) \right| \\ & \leq 2^{1-\gamma} r_{k-j}^\gamma \|f\|_{d,\gamma,W} \left(\int_{\mathbb{X}} f^2 \, d\pi \right)^{1/2} \left(\int_{\mathbb{X}} V^2 \, d\pi \right)^{\gamma/2} \left(\int_{\mathbb{X}} W^2 \, d\pi \right)^{(1-\gamma)/2} \end{aligned} \quad (3.13)$$

and consequently (3.4) holds true.

Proof. Using the bound (3.12) with $\mu(dy) = P_n(x, dy)$, $\nu(dy) = \pi(dy)$ as well as (3.11), we get

$$\left| \mathbb{E}_x f(X_n) - \int_{\mathbb{X}} f \, d\pi \right| \leq r_n^\gamma \|f\|_{d,\gamma,W} V^\gamma(x) \left(\mathbb{E}_x W(X_n) + \int_{\mathbb{X}} W \, d\pi \right)^{1-\gamma}. \quad (3.14)$$

Then, by the Hölder inequality,

$$\begin{aligned} \left| \text{Cov}(f(X_j), f(X_k)) \right| &= \left| \mathbb{E} f(X_j) \left(\mathbb{E}_{X_j} f(X_{k-j}) - \int_{\mathbb{X}} f \, d\pi \right) \right| \\ &\leq r_{k-j}^\gamma \|f\|_{d,\gamma,W} \mathbb{E} \left[|f(X_j)| V^\gamma(X_j) \left(\mathbb{E}_{X_j} W(X_{k-j}) + \int_{\mathbb{X}} W \, d\pi \right)^{1-\gamma} \right] \\ &\leq r_{k-j}^\gamma \|f\|_{d,\gamma,W} (\mathbb{E} f^2(X_j))^{1/2} (\mathbb{E} V^2(X_j))^{\gamma/2} \\ &\quad \times \left(\mathbb{E} \left(\mathbb{E}_{X_j} W(X_{k-j}) + \int_{\mathbb{X}} W \, d\pi \right)^2 \right)^{(1-\gamma)/2}. \end{aligned}$$

Recall that X is strictly stationary, hence

$$\begin{aligned} \mathbb{E} f^2(X_j) &= \int_{\mathbb{X}} f^2 \, d\pi, & \mathbb{E} V^2(X_j) &= \int_{\mathbb{X}} V^2 \, d\pi, \\ \mathbb{E} \left(\mathbb{E}_{X_j} W(X_{k-j}) + \int_{\mathbb{X}} W \, d\pi \right)^2 &\leq 2\mathbb{E} W^2(X_k) + 2 \int_{\mathbb{X}} W^2 \, d\pi = 4 \int_{\mathbb{X}} W^2 \, d\pi, \end{aligned}$$

which completes the proof of (3.13). \square

We can summarize the above argumentation as follows.

Theorem 3.10 Let X satisfy (3.11) and be stationary. Assume that $V \in L_2(\mathbb{X}, \pi)$ and let $f \in L_2(\mathbb{X}, \pi)$ belong to $H_{\gamma,W}(\mathbb{X}, d)$ for some $\gamma \in (0, 1]$ and some $W \in L_2(\mathbb{X}, \pi)$. Then the LLN (3.5) holds true.

The above proof of LLN is mainly based on the following feature: if the transition probabilities of chain X possess the “stabilization” property w.r.t. a coupling probability distance, then a kind of “stabilization” can be derived also for the expectations w.r.t. the laws of the chain of functionals from a properly chosen Hölder class, cf. Proposition 3.8 and (3.14). This feature remains quite useful when more sophisticated limit theorems are considered; see Section 3.4 below. Clearly, this feature does not require the process X to be stationary, and an analogue of the above LLN would hold true in a (general) non-stationary case in essentially the same manner, up to some minor technicalities related to a slightly more cumbersome analysis of the integrability issues. We propose the reader two exercises, which lead to such an extension.

Exercise 3.11 Let X satisfy (3.11), and denote by μ the distribution of X_0 . Assuming $f \in H_{\gamma,W}(\mathbb{X}, d)$ and

$$\int_{\mathbb{X}} V \, d\mu < \infty, \quad \int_{\mathbb{X}} W \, d\pi < \infty, \quad \sup_{n \geq 0} E_{\mu} W(X_n) < \infty,$$

prove Equation (3.2).

Exercise 3.12 Assume in addition to the assumptions of Exercise 3.11 that

$$\int_{\mathbb{X}} f^2 \, d\pi < \infty, \quad \int_{\mathbb{X}} W^2 \, d\pi < \infty, \quad \sup_{n \geq 0} E_{\mu} V^2(X_n) < \infty, \quad \sup_{n \geq 0} E_{\mu} W^2(X_n) < \infty.$$

Prove that (3.4) holds true, and consequently the LLN (3.5) follows.

.....

3.3 The corrector term and the CLT

In this section we explain one practical method to prove the CLT

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n f(X_j) \Rightarrow \mathcal{N}(0, \sigma_f^2), \tag{3.15}$$

where X is a weakly ergodic Markov chain. For convenience, we slightly modify the ergodicity assumption (3.11) imposed on X before. In fact, the bound (3.11) controls the rate of convergence of the one-dimensional distributions of X w.r.t. P_x to those w.r.t. the

law of the stationary version of X . The assumption below requires a similar bound for two-dimensional distributions. Namely, given a distance-like function d on \mathbb{X} , denote by the same letter d both the distance-like function on $\mathbb{X} \times \mathbb{X}$ defined by

$$d((x_1, x_2), (x'_1, x'_2)) = d(x_1, x'_1) + d(x_2, x'_2),$$

and the respective coupling distance on $\mathcal{P}(\mathbb{X} \times \mathbb{X})$. For $0 < m < n$, denote by $P_{m,n}(x, dy_1, dy_2)$ the distribution of (X_m, X_n) w.r.t. P_x , and by $\pi_m(dy_1, dy_2)$ the distribution of (X_0, X_m) w.r.t. P_π ; that is,

$$P_{m,n}(x, dy_1, dy_2) = P_m(x, dy_1)P_{n-m}(y_1, dy_2), \quad \pi_m(dy_1, dy_2) = \pi(dy_1)P_m(y_1, dy_2).$$

We assume the following: there exist a function $V : \mathbb{X} \rightarrow [1, \infty)$, a sequence $r_n \rightarrow 0$, and a sequence $C_l > 0$, $l \in \mathbb{N}$ such that for every $l \in \mathbb{N}$

$$d(P_{n,n+l}(x, dy_1, dy_2), \pi_l(dy_1, dy_2)) \leq C_l V(x) r_n, \quad n \geq 0. \quad (3.16)$$

Clearly, (3.16) yields (3.11) with $C_l V(x)$ instead of $V(x)$: to show that, one can simply integrate w.r.t. dy_2 on the entire space \mathbb{X} . Though, the difference between (3.11) and (formally, stronger) (3.16) is not really substantial: a natural way to prove (3.11), which we developed in Section 2.4, is to construct a two-component process $Z = (Z^1, Z^2)$, with the laws of the components equal to P_x and P_π , such that

$$\mathbb{E}d(Z_n^1, Z_n^2) \leq V(x) r_n.$$

One can easily see that, once such a process is constructed and r_n is decreasing, one has (3.16) with $C_l = 2$, $l \in \mathbb{N}$.

The main result of this section is presented in the following theorem.

Theorem 3.13 Let X satisfy (3.16) and be stationary. Let function $f : \mathbb{X} \rightarrow \mathbb{R}$ satisfy the following:

- 1) for some $\delta > 0$, $\int_{\mathbb{X}} |f|^{2+\delta} d\pi < \infty$;
- 2) f is centered, i.e. $\int_{\mathbb{X}} f d\pi = 0$;
- 3) $f \in H_{\gamma, W}(\mathbb{X}, d)$, $f \in H_{\bar{\gamma}, W}(\mathbb{X}, \sqrt{d})$ for some $\gamma, \bar{\gamma} \in (0, 1]$ and some W .

Assume furthermore that

$$\sum_n r_n^\gamma < \infty. \quad (3.17)$$

$$\int_{\mathbb{X}} V^2 \, d\pi < \infty, \quad \int_{\mathbb{X}} W^2 \, d\pi < \infty.$$

Then the CLT (3.15) holds true with

$$\sigma_f^2 = \int_{\mathbb{X}} f^2(x) \pi(dx) + 2 \sum_{k=1}^{\infty} \int_{\mathbb{X}} \int_{\mathbb{X}} f(x)f(y) \pi(dx) P_k(x, dy).$$

Remark 3.14 Comparing Theorem 3.10 and Theorem 3.13 with the limit theorems discussed in Section 3.1, one can see that the weak ergodicity assumptions (3.11) and (3.16) therein play the role analogous to that of the mixing conditions. In that concern, the principal assumption on a functional f is the Hölder-type condition (unlike the moment type conditions used in Theorem 3.3, Theorem 3.4), and respective Hölder index should relate to the weak ergodic rate; cf. (3.17).

Remark 3.15 We impose *two* Hölder-type conditions on f , because we will need to use the “stabilization” property both of f itself and for some quadratic expressions involving f ; see Equation (3.21) and Lemma 3.20. We formulate these two conditions with one weight function W , which does not restrict generality: if there are two different W_1, W_2 such that f satisfies Hölder-type conditions with d, \sqrt{d} and those weights, respectively, then one may choose $W = W_1 \vee W_2$. Observe that only the Hölder index γ , which corresponds to the initial distance-like function d , is involved into the condition (3.17).

Proof of Theorem 3.13. Our aim is to verify that, for every $\lambda \in \mathbb{R}$,

$$\mathbb{E} \exp \left(\frac{i\lambda}{\sqrt{n}} \sum_{j=1}^n f(X_j) \right) \rightarrow e^{-\lambda^2 \sigma_f^2 / 2}, \quad n \rightarrow \infty; \quad (3.18)$$

then (3.15) holds true by the continuity theorem for characteristic functions. Denote

$$\xi_{j,n} := \frac{f(X_j)}{\sqrt{n}}, \quad \zeta_{k,n} := \sum_{j=1}^k \xi_{j,n},$$

and for a fixed $\lambda \in \mathbb{R}$ denote $\phi(x) = e^{i\lambda x}$. Consider the family of expectations

$$m_{k,n} = \mathbb{E}\phi(\zeta_{k,n}), \quad k = 1, \dots, n;$$

we will prove that the sequence of functions

$$m_n(t) = m_{[nt],n}, \quad t \in [0, 1]$$

converge uniformly to a continuous function m which satisfies

$$\frac{d}{dt}m(t) = -\frac{\lambda^2 \sigma_f^2}{2}m(t), \quad m(0) = 1.$$

The unique solution to this equation is

$$m(t) = e^{-t\lambda^2 \sigma_f^2/2},$$

hence such a convergence would yield (3.16) immediately.

Write

$$\phi(\zeta_{k,n}) - \phi(\zeta_{k-1,n}) = \phi'(\zeta_{k-1,n})\xi_{k,n} + \frac{1}{2}\phi''(\zeta_{k-1,n})\xi_{k,n}^2 + \chi_{k,n} \quad (3.19)$$

with

$$\chi_{k,n} = \phi(\zeta_{k-1,n}) \left[e^{i\lambda \xi_{k,n}} - 1 - i\lambda \xi_{k,n} + \frac{\lambda^2}{2} \xi_{k,n}^2 \right],$$

then

$$\begin{aligned} m_{[nt],n} - 1 &= \sum_{k=1}^{[nt]} \mathbb{E}[\phi'(\zeta_{k-1,n})\xi_{k,n}] + \frac{1}{2} \sum_{k=1}^{[nt]} \mathbb{E}[\phi''(\zeta_{k-1,n})\xi_{k,n}^2] + \sum_{k=1}^{[nt]} \mathbb{E}\chi_{k,n} \\ &=: \Sigma_{[nt],n}^1 + \Sigma_{[nt],n}^2 + \Sigma_{[nt],n}^3. \end{aligned}$$

Let us analyse the terms $\Sigma_{[nt],n}^i$, $i = 1, 2, 3$ separately, beginning with the most simple $\Sigma_{[nt],n}^3$ and finishing with the most intrinsic $\Sigma_{[nt],n}^1$.

A bound for $\Sigma_{[nt],n}^3$. Let δ be as in assumption 1). There exists a constant $C_{\lambda,\delta}$, which depends only on fixed λ and δ , such that

$$|\chi_{k,n}| \leq C_{\lambda,\delta} |\xi_{k,n}|^{2+\delta}.$$

Exercise 3.16 Please, prove the above upper bound.

Recall that

$$|\xi_{k,n}|^{2+\delta} = n^{-1-\delta/2} |f(X_k)|^{2+\delta},$$

and X is stationary. Then we have

$$|\Sigma_{[nt],n}^3| \leq C_{\lambda,\delta} \sum_{k=1}^n \mathbb{E} |\chi_{k,n}| = n C_{\lambda,\delta} \mathbb{E} |\xi_{k,n}|^{2+\delta} = O(n^{-\delta/2}), \quad n \rightarrow \infty. \quad (3.20)$$

Managing $\Sigma_{[nt],n}^2$. The following *delay* trick appears to be very convenient in order to analyze the limit behavior as $n \rightarrow \infty$ of

$$\Sigma_{[nt],n}^2 = \frac{1}{2} \sum_{k=1}^{[nt]} \mathbb{E} [\phi''(\zeta_{k-1,n}) f^2(X_k)].$$

Choose a sequence $\{D_n\} \subset \mathbb{N}$ and write

$$\tilde{\Sigma}_{[nt],n}^2 := \frac{1}{2n} \sum_{k=D_n}^{[nt]} \mathbb{E} [\phi''(\zeta_{k-D_n,n}) f^2(X_k)].$$

For the difference between the present value $\zeta_{k-1,n}$ and the “delayed” value $\zeta_{k-D_n,n}$, we have

$$\mathbb{E}(\zeta_{k-1,n} - \zeta_{k-D_n,n})^2 = \frac{1}{n} \mathbb{E} \left(\sum_{j=k-D_n+1}^{k-1} f(X_j) \right)^2 = \frac{1}{n} \sum_{j_1, j_2=k-D_n+1}^{k-1} \text{Cov}(f(X_{j_1}), f(X_{j_2})),$$

because f is centered and X is stationary. Using (3.13), we get then

$$\mathbb{E}(\zeta_{k-1,n} - \zeta_{k-D_n,n})^2 \leq C \frac{D_n}{n} \sum_{l=0}^{D_n} r_l^\gamma.$$

Choose $\{D_n\}$ such that $D_n \rightarrow \infty$, but $D_n/n \rightarrow 0$. Then by condition 3)

$$\frac{D_n}{n} \sum_{l=0}^{D_n} r_l^\gamma \rightarrow 0,$$

and therefore

$$\max_{D_n \leq k \leq n} \mathbb{E}(\zeta_{k-1,n} - \zeta_{k-D_n,n})^2 \rightarrow 0, \quad n \rightarrow \infty.$$

Because ϕ'' is bounded and has a bounded derivative, this yields

$$\max_{D_n \leq k \leq n} \mathbb{E}|\phi''(\zeta_{k-1,n}) - \phi''(\zeta_{k-D_n,n})|^{2+\delta} \rightarrow 0, \quad n \rightarrow \infty.$$

Exercise 3.17 Please, prove this.

Finally, we get that the difference between $\Sigma_{[n],n}^2$ and $\tilde{\Sigma}_{[n],n}^2$ is negligible:

$$\begin{aligned} |\Sigma_{[n],n}^2 - \tilde{\Sigma}_{[n],n}^2| &\leq \frac{C}{n} \sum_{k=1}^{D_n-1} \mathbb{E}f^2(X_k) + \frac{C}{n} \sum_{k=D_n}^n |\phi''(\zeta_{k-1,n}) - \phi''(\zeta_{k-D_n,n})| f^2(X_k) \\ &\leq \frac{CD_n}{n} \int_{\mathbb{X}} f^2 d\pi + C \left(\max_{D_n \leq k \leq n} \mathbb{E}|\phi''(\zeta_{k-1,n}) - \phi''(\zeta_{k-D_n,n})|^{2+\delta} \right)^{\delta/(2+\delta)} \\ &\quad \times \left(\int_{\mathbb{X}} |f|^{2+\delta} d\pi \right)^{2/(2+\delta)} \rightarrow 0. \end{aligned}$$

On the other hand,

$$\tilde{\Sigma}_{[n],n}^2 = \frac{1}{2n} \sum_{k=D_n}^{[n]} \mathbb{E}(\phi''(\zeta_{k-D_n,n}) \mathbb{E}_{X_{k-D_n}} f^2(X_{D_n})),$$

and because $D_n \rightarrow \infty$, the conditional expectations $\mathbb{E}_{X_{k-D_n}} f^2(X_{D_n})$ enjoy the “stabilization” property. Namely, one has

$$|f^2(x) - f^2(y)| \leq \|f\|_{\sqrt{d}, \bar{\gamma}, W} d^{\bar{\gamma}/2}(x, y) (W(x) + W(y))^{1-\bar{\gamma}} (|f(x)| + |f(y)|),$$

hence similarly to the proof of Proposition 3.8 one can show that

$$\begin{aligned} \left| \mathbb{E}_x f^2(X_k) - \int_{\mathbb{X}} f^2 d\pi \right| &\leq 4r_k^{\bar{\gamma}} \|f\|_{\sqrt{d}, \bar{\gamma}, W} V^{\bar{\gamma}/2}(x) \\ &\quad \times \left(\mathbb{E}_x W^2(X_k) + \int_{\mathbb{X}} W^2 d\pi \right)^{(1-\bar{\gamma})/2} \left(\mathbb{E}_x f^2(X_k) + \int_{\mathbb{X}} f^2 d\pi \right)^{1/2}. \end{aligned} \quad (3.21)$$

Exercise 3.18 Please, prove this.

Then, by the Hölder inequality applied to p_1, p_2, p_3 with $1/p_1 = \bar{\gamma}/2, 1/p_2 = 1-\bar{\gamma}/2, 1/p_3 = 1/2$,

$$\begin{aligned} \mathbb{E} \left| \mathbb{E}_{X_{k-D_n}} f^2(X_{D_n}) - \int_{\mathbb{X}} f^2 d\pi \right| &= \int_{\mathbb{X}} \left| \mathbb{E}_x f^2(X_{D_n}) - \int_{\mathbb{X}} f^2 d\pi \right| \pi(dx) \\ &\leq 4r_{D_n}^{\bar{\gamma}} \|f\|_{\sqrt{d}, \bar{\gamma}, W} \left(\int_{\mathbb{X}} V^2 d\pi \right)^{\bar{\gamma}/2} \left(2 \int_{\mathbb{X}} W^2 d\pi \right)^{(1-\bar{\gamma})/2} \left(2 \int_{\mathbb{X}} f^2 d\pi \right)^{1/2} \rightarrow 0. \end{aligned} \quad (3.22)$$

Because ϕ'' is bounded and $\phi'' = -\lambda^2 \phi$, this finally gives

$$\left| \Sigma_{[n],n}^2 + \frac{\lambda^2}{2n} \left(\int_{\mathbb{X}} f^2 d\pi \right) \sum_{k=1}^{[n]-1} m_{k,n} \right| \rightarrow 0, \quad n \rightarrow \infty.$$

Managing $\Sigma_{[n],n}^1$ using the corrector term. To manage the “most dangerous” term $\Sigma_{[n],n}^1$, an auxiliary *corrector term* construction appears to be very useful. An important ingredient of this construction is the *potential* of the function f :

$$\mathcal{R}f(x) = \sum_{k=0}^{\infty} \mathbb{E}_x f(X_k), \quad x \in \mathbb{X}. \quad (3.23)$$

Proposition 3.19 Let X satisfy (3.11) and condition 3) of Theorem 3.13 hold true. Let a function $f \in H_{\gamma, W}(\mathbb{X}, d)$ belong to $L_2(\mathbb{X}, \pi)$ and be centered. Then the series (3.23) converges in $L_2(\mathbb{X}, \pi)$ sense, and the sequence

$$M_k = \sum_{j=0}^{k-1} f(X_j) + \mathcal{R}f(X_k), \quad k \geq 0,$$

is a martingale w.r.t. the natural filtration $\{\mathcal{F}_k\}$ of the process X . If, in addition,

$$\sup_k \mathbb{E}_x W(X_k) < \infty, \quad x \in \mathbb{X}, \quad (3.24)$$

then the series (3.23) converges for every $x \in \mathbb{X}$.

Proof. Write

$$P^k f(x) = \mathbb{E}_x f(X_k) = \int_{\mathbb{X}} f(y) P_k(x, dy).$$

Because f is centered, we get from (3.14)

$$|P^k f(x)| = \left| \mathbb{E}_x f(X_k) - \int_{\mathbb{X}} f \, d\pi \right| \leq \|f\|_{d,\gamma,W} r_k^\gamma V^\gamma(x) \left(\mathbb{E}_x W(X_k) + \int_{\mathbb{X}} W \, d\pi \right)^{1-\gamma}. \quad (3.25)$$

This immediately yields that (3.23) converges for every $x \in \mathbb{X}$ under the additional assumption (3.24). On the other hand, the same bound yields

$$\|P^k f\|_{L_2(\mathbb{X},\pi)} \leq \|f\|_{d,\gamma,W} r_k^\gamma \left(\int_{\mathbb{X}} V^2 \, d\pi \right)^\gamma \left(4 \int_{\mathbb{X}} W^2 \, d\pi \right)^{1-\gamma}$$

(see the calculation at the end of the proof of Proposition 3.9), which proves that (3.23) converges in $L_2(\mathbb{X}, \pi)$.

Because X is stationary and $f \in L_2(\mathbb{X}, \pi)$, for every $\rho \in (0, 1)$ we have that the series

$$S^{(\rho)} := \sum_{j=0}^{\infty} \rho^j f(X_j)$$

converges in $L_2(\Omega, \mathcal{F}, \mathbb{P})$. Then

$$M_k^{(\rho)} := \mathbb{E}(S^{(\rho)} | \mathcal{F}_k) = \sum_{j=0}^{\infty} \rho^j \mathbb{E}(f(X_j) | \mathcal{F}_k), \quad k \geq 0$$

is a martingale. Denote

$$\mathcal{R}^{(\rho)} f(x) = \sum_{k=0}^{\infty} \rho^k \mathbb{E}_x f(X_k) = \sum_{k=0}^{\infty} \rho^k P^k f(x),$$

then we get

$$M_k^{(\rho)} = \sum_{j=0}^{k-1} \rho^j f(X_j) + \rho^k \mathcal{R}^{(\rho)} f(X_k), \quad k \geq 0.$$

It follows from the bound (3.25) that

$$\mathcal{R}^{(\rho)} f \rightarrow \mathcal{R} f, \quad \rho \rightarrow 1-$$

in $L_2(\mathbb{X}, \pi)$. Hence, because every X_k obeys law π , we deduce

$$M_k^{(\rho)} \rightarrow M_k, \quad \rho \rightarrow 1-$$

in $L_2(\Omega, \mathcal{F}, P)$ for every $k \geq 0$. Therefore $\{M_k\}$ is a martingale. \square

Write

$$\eta_k = \mathcal{R}f(X_k) - f(X_k), \quad k \geq 0,$$

and consider the sequence

$$U_{k,n} = \phi'(\zeta_{k,n})\eta_k, \quad k = 0, \dots, n.$$

Observe that

$$\begin{aligned} \frac{U_{k,n} - U_{k-1,n}}{\sqrt{n}} &= \phi'(\zeta_{k-1,n}) \frac{\eta_k - \eta_{k-1}}{\sqrt{n}} + (\phi'(\zeta_{k,n}) - \phi'(\zeta_{k-1,n})) \frac{\eta_k}{\sqrt{n}} \\ &= \phi'(\zeta_{k-1,n}) \frac{M_k - M_{k-1}}{\sqrt{n}} - \phi'(\zeta_{k-1,n}) \xi_{k,n} + (\phi'(\zeta_{k,n}) - \phi'(\zeta_{k-1,n})) \frac{\eta_k}{\sqrt{n}}. \end{aligned}$$

Because M is a martingale, the expectation of the first term in the right hand side equals 0. The second term is exactly the one involved into $\Sigma_{[n],n}^1$. Hence, summing over k and averaging yields

$$\frac{EU_{[n],n} - EU_{0,n}}{\sqrt{n}} + \Sigma_{[n],n}^1 = \sum_{k=1}^{[n]} E \left[\phi''(\zeta_{k-1,n}) \xi_{k,n} \frac{\eta_k}{\sqrt{n}} \right] + \sum_{k=1}^{[n]} E \vartheta_{k,n} =: \Upsilon_{[n],n}^1 + \Upsilon_{[n],n}^2,$$

where

$$\begin{aligned} \vartheta_{k,n} &= (\phi'(\zeta_{k,n}) - \phi'(\zeta_{k-1,n}) - \phi''(\zeta_{k-1,n}) \xi_{k,n}) \frac{\eta_k}{\sqrt{n}} \\ &= i\lambda e^{i\lambda \zeta_{k-1,n}} \left(e^{i\lambda \xi_{k,n}} - 1 - i\lambda \xi_{k,n} \right) \frac{\eta_k}{\sqrt{n}}. \end{aligned}$$

We will see that the terms $\Upsilon_{[n],n}^1$ and $\Upsilon_{[n],n}^2$ can be treated similarly to $\Sigma_{[n],n}^2$ and $\Sigma_{[n],n}^3$, respectively. On the other hand, because ϕ' is bounded and $\{\eta_k\}$ is a stationary and a square integrable sequence,

$$\left| \frac{EU_{[n],n} - EU_{0,n}}{\sqrt{n}} \right| \leq \frac{C}{\sqrt{n}}.$$

This is the essence of the *corrector term method*, which dates back to [46]: by adding a small *corrector term* $n^{-1/2}E(U_{[n],n} - U_{0,n})$, one transforms a “dangerous” expression

$\Sigma_{[n],n}^1$, where the summands have an order $n^{-1/2}$, into a more manageable one, where the summands have an order at most n^{-1} . Clearly, such a transformation exploits substantially the semimartingale structure of the summands involved in $\Sigma_{[n],n}^1$.

A bound for $\Upsilon_{[n],n}^2$. We have

$$|\vartheta_{k,n}| \leq C |\xi_{k,n}|^{1+\delta/2} \left| \frac{\eta_k}{\sqrt{n}} \right| = C \left| \frac{f(X_k)}{\sqrt{n}} \right|^{1+\delta/2} \left| \frac{\eta_k}{\sqrt{n}} \right|$$

with C which depends only on λ, δ . Then

$$|\vartheta_{k,n}| \leq C n^{-1-\delta/2} \left(\mathbb{E} f^{2+\delta}(X_k) \right)^{1/2} \left(\mathbb{E} (\mathcal{R}f(X_k) - f(X_k))^2 \right)^{1/2},$$

and since X is stationary we get

$$|\Upsilon_{[n],n}^2| = O\left(n^{-\delta/2}\right), \quad n \rightarrow \infty.$$

Managing $\Upsilon_{[n],n}^1$. We have

$$\Upsilon_{[n],n}^1 = \sum_{k=1}^{[n]} \mathbb{E} \phi''(\zeta_{k-1,n}) \frac{g(X_k)}{n},$$

where we denote

$$g(x) = f(x)(\mathcal{R}f(x) - f(x)) = f(x) \sum_{l=1}^{\infty} \mathbb{E}_x f(X_l).$$

Hence $\Upsilon_{[n],n}^1$ has exactly the same form as $\Sigma_{[n],n}^2$, and we can use the same arguments to analyze its asymptotic behavior. In what follows we sketch the calculations, leaving the details for the reader.

Firstly, since

$$\int_{\mathbb{X}} |g|^{1+\delta/2} d\pi \leq \left(\int_{\mathbb{X}} |f|^{2+\delta} d\pi \right)^{1/2} \left(\int_{\mathbb{X}} (\mathcal{R}f - f)^2 d\pi \right)^{1/2} < \infty,$$

one shows that $\Upsilon_{[nt],n}^1$, up to a negligible term, equals

$$\tilde{\Upsilon}_{[nt],n}^1 := \sum_{k=D_n}^{[nt]} \mathbb{E} \phi''(\zeta_{k-D_n,n}) \frac{g(X_k)}{n}.$$

The limit behavior of $\tilde{\Upsilon}_{[nt],n}^1$ now is well understood because the conditional expectation $\mathbb{E}_{X_{k-D_n}} g(X_{D_n})$ has the following “stabilization” property.

Lemma 3.20 Under the conditions of Theorem 3.13,

$$\int_{\mathbb{X}} \left| \mathbb{E}_x g(X_n) - \int_{\mathbb{X}} g \, d\pi \right| \pi(\mathrm{d}x) \rightarrow 0, \quad n \rightarrow \infty.$$

Proof. Let

$$g_l(x) = f(x) \mathbb{E}_x f(X_l) = f(x) P^l f(x), \quad l \geq 1.$$

Because the series $\mathcal{R}f - f = \sum_{l=1}^{\infty} P^l f$ converges in $L_2(\mathbb{X}, \pi)$, we know that

$$\sum_{l=1}^L g_l \rightarrow g, \quad L \rightarrow \infty$$

in $L_1(\mathbb{X}, \pi)$. Then, to prove the required statement, it is sufficient to prove that

$$\int_{\mathbb{X}} \left| \mathbb{E}_x g_l(X_n) - \int_{\mathbb{X}} g_l \, d\pi \right| \pi(\mathrm{d}x) \rightarrow 0, \quad n \rightarrow \infty \quad (3.26)$$

for every $l \geq 1$. Because

$$\begin{aligned} \mathbb{E}_x g_l(X_n) &= \int_{\mathbb{X} \times \mathbb{X}} f(y_1) f(y_2) P_{n,n+l}(x, \mathrm{d}y_1, \mathrm{d}y_2), \\ \int_{\mathbb{X}} g_l \, d\pi &= \int_{\mathbb{X} \times \mathbb{X}} f(y_1) f(y_2) \pi_l(\mathrm{d}y_1, \mathrm{d}y_2), \end{aligned}$$

(3.26) follows naturally from assumption (3.16) imposed on two-dimensional distributions of X . Namely, for every $y = (y_1, y_2)$, $y' = (y'_1, y'_2)$ we have

$$\begin{aligned} &|f(y_1) f(y_2) - f(y'_1) f(y'_2)| \\ &\leq \|f\|_{\sqrt{d}, \bar{\gamma}, W} d^{\bar{\gamma}/2}(y, y') (W(y_1) + W(y_2) + W(y'_1) + W(y'_2))^{1-\bar{\gamma}} (|f(y_1)| + |f(y'_2)|). \end{aligned}$$

Hence for every pair of random vectors $\xi = (\xi_1, \xi_2)$, $\eta = (\eta_1, \eta_2)$ we obtain

$$\begin{aligned} \mathbb{E}|f(\xi_1)f(\xi_2) - f(\eta_1)f(\eta_2)| &\leq \|f\|_{\sqrt{d}, \gamma, W} (\mathbb{E}d(\xi, \eta))^{\tilde{\gamma}/2} \\ &\times \left(4\mathbb{E}(W^2(\xi_1) + W^2(\xi_2) + W^2(\eta_1) + W^2(\eta_2))\right)^{(1-\tilde{\gamma})/2} \left(2\mathbb{E}(f^2(\xi_1) + f^2(\eta_2))\right)^{1/2}. \end{aligned}$$

Taking the infimum over all pairs $(\xi, \eta) \in \mathcal{C}(P_{n,n+l}(x, \cdot), \pi_l)$, we get by means of (3.16)

$$\begin{aligned} \left| \mathbb{E}_x g_l(X_n) - \int_{\mathbb{X}} g_l \, d\pi \right| &\leq \|f\|_{\sqrt{d}, \gamma, W} (C_l r_n)^{\tilde{\gamma}/2} (V(x))^{\tilde{\gamma}/2} \\ &\times \left(4\mathbb{E}_x W^2(X_n) + 4\mathbb{E}_x W^2(X_{n+l}) + 8 \int_{\mathbb{X}} W^2 \, d\pi\right)^{(1-\tilde{\gamma})/2} \times \left(2\mathbb{E}_x f^2(X_n) + 2 \int_{\mathbb{X}} f^2 \, d\pi\right)^{1/2}. \end{aligned}$$

Now (3.26) can be deduced similarly to (3.22); we leave details to a reader. \square

As a corollary, we obtain that

$$\left| \tilde{\Upsilon}_{[n],n}^1 + \frac{\lambda^2}{n} \left(\int_{\mathbb{X}} g \, d\pi \right) \sum_{k=1}^{[n]-1} m_{k,n} \right| \rightarrow 0, \quad n \rightarrow \infty.$$

We saw that

$$\left| \Sigma_{[n],n}^3 - \tilde{\Upsilon}_{[n],n}^1 \right| \rightarrow 0,$$

hence we can summarize the above calculations, and write

$$\left| m_{[n],n} - 1 + \frac{\lambda^2}{2n} \left(\int_{\mathbb{X}} (f^2 + 2g) \, d\pi \right) \sum_{k=1}^{[n]-1} m_{k,n} \right| \rightarrow 0, \quad n \rightarrow \infty.$$

Then it is easy to prove that

$$m_{[n],n} \rightarrow e^{-t\lambda^2\sigma^2/2}, \quad t \in [0, 1]$$

with

$$\sigma^2 = \int_{\mathbb{X}} (f^2 + 2g) \, d\pi$$

(please, prove this!). Because

$$\int_{\mathbb{X}} g \, d\pi = \int_{\mathbb{X}} \left(f \sum_{k=1}^{\infty} P^k f \right) \, d\pi = \sum_{k=1}^{\infty} \iint_{\mathbb{X}^2} f(x)f(y)\pi(dx)P_k(x, dy),$$

this completes the proof of Theorem 3.13. □

We finish this section with a similar remark we made concerning the LLN: an analogue of the above CLT holds true in a non-stationary case in essentially the same manner. Minor technicalities, related to an analysis of the integrability issues, need to be then taken into account; we propose the proof of such an extension as an exercise for the interested reader.

Exercise 3.21 Let all the assumptions of Theorem 3.13, except the stationarity of X , be satisfied, and

$$\sup_{n \geq 0} E_{\mu} V^2(X_n) < \infty, \quad \sup_{n \geq 0} E_{\mu} W^2(X_n) < \infty,$$

where μ denotes the distribution of X_0 . Prove that the CLT (3.15) holds true.

.

3.4 Autoregressive models with Markov regime

Both, the LLN and the CLT studied in the previous sections, can be considered as particular cases of the following model. Let $\zeta_{k,n}$, $1 \leq k \leq n < \infty$ be a triangular array of random variables, with the n -th row being determined by the recurrence relation

$$\zeta_{k,n} = F_n(\zeta_{k-1,n}, X_k), \quad k = 1, \dots, n, \quad \zeta_{0,n} = y_0, \tag{3.27}$$

where X is a given Markov chain, and F_n , $n \geq 1$ is a given sequence of functions. What can be said then about the (weak) convergence of $\zeta_{n,n}$?

Choosing either

$$F_n(\zeta, x) = \zeta + \frac{1}{n} f(x)$$

or

$$F_n(\zeta, x) = \zeta + \frac{1}{\sqrt{n}} f(x),$$

we obtain the models studied in the two previous sections, which are “additive” in the sense that an increment $\zeta_{k,n} - \zeta_{k-1,n}$ depends only on the value X_{k-1} of the “driving process” X . In the general model (3.27), an increment may also depend on the current value $\zeta_{k-1,n}$ of the target process, and that is why the name *autoregressive* is used. Because

the driving noise is assumed to be Markov, we call (3.27) an *autoregressive model with Markov regime*.

Remark 3.22 In the available literature, when an autoregressive model with Markov regime is discussed, a typical assumption on the term on the right hand side of (3.27) is that it is of the form

$$F_n(\zeta_{k-1,n}, X_k, \varepsilon_k),$$

where $\{\varepsilon_k\}$ is a sequence of i.i.d. random variables. Introducing an additional “random seed” $\{\varepsilon_k\}$ allows to extend the model (3.27), and to consider sequences with $\zeta_{k,n}$, not being defined by $\zeta_{k-1,n}, X_k$ through a functional relation, but having its conditional distribution defined by $\zeta_{k-1,n}, X_k$. In our framework, there is no substantial difference between these settings, because, at least formally, we can consider the pair $\tilde{X}_k = (X_k, \varepsilon_k)$ as a new Markov “driving process”. To simplify the notation, in what follows we do not introduce separately a “random seed” $\{\varepsilon_k\}$, and consider the model (3.27).

We consider (3.27) with $F_n : \mathbb{R} \times \mathbb{X} \rightarrow \mathbb{R}$ of the form

$$F_n(\zeta, x) = \zeta + \frac{1}{n}f(\zeta, x) + \frac{1}{\sqrt{n}}g(\zeta, x), \quad (3.28)$$

which is a natural “non-additive” extension of the models studied in the previous two sections. We will show that the corrector term method, explained in Section (3.3), still allows one to study this much more general setting without an essential complication of neither the assumptions nor the structure of the proof.

One could notice already from the proof of Theorem 3.13 above, that it is more convenient to study the limit behavior of the whole time series $\{\zeta_{k,n}\}_{k=1,\dots,n}$ rather than of its endpoint $\zeta_{n,n}$. Namely, we will study the limit behavior of the sequence of piece-wise constant processes

$$Y_n(t) = \zeta_{[nt],n}, \quad t \in [0, 1]. \quad (3.29)$$

To do that, we will follow the standard plan of proving functional limit theorems, which consists of two principal steps:

- 1) prove that the sequence of processes $\{Y_n\}$ is weakly compact in a certain sense;
- 2) describe all the weak limit points of this sequence, and prove that, actually, such a limit point is unique.

Here, we will consider the weak convergence in the sense of finite-dimensional distributions; in that case, weak compactness is equivalent to the following: any subsequence $\{\zeta_{n_k}\}$ contains a (sub-)subsequence $\{Y_{n'_k}\}$ such that, for any $m \geq 1$, $t_1, \dots, t_m \in [0, 1]$, the vector-valued sequence $(Y_{n'_k}(t_1), \dots, Y_{n'_k}(t_m))$ converges weakly as $k \rightarrow \infty$.

To describe the limit points, we will use the concept of *martingale problems*; below we briefly recall respective results (for a detailed exposition, we refer to [54] and [20]).

Let \mathcal{L} be an operator defined on some set \mathcal{D} of functions $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$. A process $Y = \{Y(t), t \in [0, 1]\}$, taking values in \mathbb{R}^d , is called a *solution of the martingale problem* $(\mathcal{L}, \mathcal{D})$, if for any $\phi \in \mathcal{D}$ the process

$$\phi(Y(t)) - \int_0^t \mathcal{L}\phi(Y(s)) \, ds, \quad t \in [0, 1]$$

is a martingale w.r.t. the natural filtration of Y . Note that this definition includes as a natural pre-requisite the claim that the above integral is well defined; typically, it is assumed that process Y is *measurable*.

A martingale problem $(\mathcal{L}, \mathcal{D})$ is said to be *well-posed*, if for every $y \in \mathbb{R}^d$ any two solutions of $(\mathcal{L}, \mathcal{D})$ with the same initial value $Y(0) = y$, the finite-dimensional distributions agree. Actually, this means that the solution of $(\mathcal{L}, \mathcal{D})$ with $Y(0) = y$ is (weakly) unique.

Theorem 3.23 Let X satisfy (3.16) and be stationary. Consider the triangular array $\{\zeta_{k,n}\}$ defined by (3.27) with F_n having the form (3.28). Assume furthermore

- 1) There exists a derivative $\partial_\zeta g$, and $f(\zeta, \cdot)$, $g(\zeta, \cdot)$, $\partial_\zeta g(\zeta, \cdot) \in L_1(\mathbb{X}, \pi)$ for every $\zeta \in \mathbb{R}$.
- 2) There exist $\gamma, \bar{\gamma} \in (0, 1]$ and a function W , such that for every $\zeta \in \mathbb{R}$,

$$g(\zeta, \cdot), \partial_\zeta g(\zeta, \cdot) \in H_{\gamma, W}(\mathbb{X}, d),$$

$$f(\zeta, \cdot) \in H_{\bar{\gamma}, W}(\mathbb{X}, d), \quad g(\zeta, \cdot) \in H_{\bar{\gamma}, W}(\mathbb{X}, \sqrt{d}).$$

In addition,

$$\sup_{\zeta \in \mathbb{R}} (\|g(\zeta, \cdot)\|_{d, \gamma, W} + \|\partial_\zeta g(\zeta, \cdot)\|_{d, \gamma, W} + \|f(\zeta, \cdot)\|_{d, \bar{\gamma}, W} + \|g(\zeta, \cdot)\|_{\sqrt{d}, \bar{\gamma}, W}) < \infty,$$

and there exists $\delta > 0$ such that

$$\sup_{\zeta_1 \neq \zeta_2} |\zeta_1 - \zeta_2|^{-\delta} (\|g(\zeta_1, \cdot) - g(\zeta_2, \cdot)\|_{d, \gamma, W} + \|\partial_\zeta g(\zeta_1, \cdot) - \partial_\zeta g(\zeta_2, \cdot)\|_{d, \gamma, W}) < \infty,$$

$$\sup_{\zeta_1 \neq \zeta_2} |\zeta_1 - \zeta_2|^{-\delta} \|f(\zeta_1, \cdot) - f(\zeta_2, \cdot)\|_{d, \bar{\gamma}, W} < \infty.$$

3) For every $\zeta \in \mathbb{R}$, the functions $g(\zeta, \cdot)$, $\partial_\zeta g(\zeta, \cdot)$ are centered; that is,

$$\int_{\mathbb{X}} g(\zeta, x) \pi(\mathrm{d}x) = 0, \quad \int_{\mathbb{X}} \partial_\zeta g(\zeta, x) \pi(\mathrm{d}x) = 0.$$

The function f satisfies

$$\sup_{\zeta \in \mathbb{R}} \left| \int_{\mathbb{X}} f(\zeta, x) \pi(\mathrm{d}x) \right| < \infty,$$

$$\sup_{\zeta_1 \neq \zeta_2} |\zeta_1 - \zeta_2|^{-\delta} \left| \int_{\mathbb{X}} f(\zeta_1, x) \pi(\mathrm{d}x) - \int_{\mathbb{X}} f(\zeta_2, x) \pi(\mathrm{d}x) \right| < \infty.$$

$$4) \sum_n r_n^\gamma < \infty, \quad \int_{\mathbb{X}} V^{2+\delta} \mathrm{d}\pi < \infty, \quad \int_{\mathbb{X}} W^{2+\delta} \mathrm{d}\pi < \infty.$$

Then for the sequence $\{Y_n\}$ of the piece-wise constant processes (3.29) the following statements hold true:

I The sequence $\{Y_n\}$ is weakly compact in the sense of convergence of finite-dimensional distributions.

II Any weak limit point of the sequence $\{Y_n\}$ is a solution of the martingale problem $(\mathcal{L}, \mathcal{D})$ with

$$\mathcal{L}\phi(\zeta) = A(\zeta)\phi'(\zeta) + \frac{1}{2}B(\zeta)\phi''(\zeta), \quad \mathcal{D} = C_0^\infty(\mathbb{R}^d), \quad (3.30)$$

where

$$A(\zeta) = \int_{\mathbb{X}} f(\zeta, x) \pi(\mathrm{d}x) + \sum_{k=1}^{\infty} \iint_{\mathbb{X}^2} g(\zeta, x) \partial_\zeta g(\zeta, y) \pi(\mathrm{d}x) P_k(x, \mathrm{d}y),$$

$$B(\zeta) = \int_{\mathbb{X}} g^2(\zeta, x) \pi(dx) + 2 \sum_{k=1}^{\infty} \iint_{\mathbb{X}^2} g(\zeta, x) g(\zeta, y) \pi(dx) P_k(x, dy).$$

Consequently, if the martingale problem (3.30) is well-posed, then the sequence $\{Y_n\}$ converges weakly, in the sense of convergence of finite-dimensional distributions, to a diffusion process which is the unique solution of this martingale problem.

Remark 3.24 Sufficient conditions for a martingale problem to be well-posed in the classical “diffusive” case, where \mathcal{L} is a second order differential operator, are well studied; cf. [54]. Hence, under the assumptions of Theorem 3.23, it is easy to give sufficient conditions for the weak convergence of the sequence $\{Y_n\}$. In particular, the coefficients A, B are continuous, therefore for the martingale problem (3.30) to be well-posed it is sufficient to assume the diffusion coefficient to be non-degenerate.

Remark 3.25 Note that in the case $f \equiv 0$ and $g(\zeta, x) = g(x)$, the assumptions of Theorem 3.23 coincide with those of Theorem 3.13. That is, the CLT proved above actually can be considered just as a particular case of Theorem 3.23: to do that, one should recall that the martingale problem

$$\mathcal{L}\phi(\zeta) = \frac{1}{2} \sigma^2 \phi''(\zeta), \quad \mathcal{D} = C_0^\infty(\mathbb{R})$$

is well-posed, and its unique solution is σW , where W denotes the Wiener process. However, in order to explain the method clearly, we have split the exposition, and, before considering the general statement, we have to prove the important particular case of CLT separately and more explicitly, using the characteristic functions technique rather than the concept of the martingale problem.

Remark 3.26 The statement of Theorem 3.23 can be extended in various ways: instead of real-valued $\{\zeta_{k,n}\}$, one can consider sequences with values in \mathbb{R}^d , and in that case the limit process Y is a multidimensional diffusion; instead of considering uniform conditions in ζ on f and g , one can impose weaker (but more cumbersome) conditions which allow the coefficients to have at most linear growth; instead of a weak convergence in the sense of finite-dimensional distributions, convergence in the Skorokhod topology can be proved under slightly modified conditions on the weight functions V, W . Not to overburden the exposition, we do not discuss these possibilities here, referring the interested reader to [37], where such results were obtained in the continuous time setting.

Before proving Theorem 3.23, we give some analytical preliminaries and develop an auxiliary *corrector term* construction. Some of the calculations below are similar to those we developed in previous section, in that case we just sketch the argument and leave the details to the reader.

Define the potential of the function g , which now depends on an additional variable ζ , by treating this variable as a “frozen” parameter; that is,

$$\mathcal{R}g(\zeta, x) = \sum_{k=0}^{\infty} \mathbb{E}_x g(\zeta, X_k), \quad x \in \mathbb{X}. \quad (3.31)$$

One can prove the following properties similarly to Proposition 3.19; we omit the details:

- (i) for every ζ , the series (3.31) converges in $L_2(\mathbb{X}, \pi)$;
- (ii) for every $k \geq 1$, $\zeta \in \mathbb{R}$,

$$\mathbb{E}[\mathcal{R}g(\zeta, X_k) | \mathcal{F}_{k-1}] = \mathcal{R}g(\zeta, X_{k-1}) - g(\zeta, X_{k-1}),$$

where $\{\mathcal{F}_k\}$ denotes the natural filtration for the process X .

Proposition 3.27 There exists a function $H : \mathbb{X} \rightarrow \mathbb{R}^+$ such that $\int_{\mathbb{X}} H^{2+\delta} d\pi < \infty$ and

$$\begin{aligned} |f(\zeta, x)| &\leq H(x), & |f(\zeta_1, x) - f(\zeta_2, x)| &\leq |\zeta_1 - \zeta_2|^\delta H(x), \\ |g(\zeta, x)| + |\mathcal{R}g(\zeta, x)| &\leq H(x), & |g(\zeta_1, x) - g(\zeta_2, x)| &\leq |\zeta_1 - \zeta_2|^\delta H(x), \\ |\mathcal{R}g(\zeta, x)| &\leq H(x), & |\mathcal{R}g(\zeta_1, x) - \mathcal{R}g(\zeta_2, x)| &\leq |\zeta_1 - \zeta_2|^\delta H(x). \end{aligned}$$

Proof. Let us construct a function $H^{\mathcal{R}g}$ which gives the required bounds for $\mathcal{R}g$. Recall that, because $g(\zeta, \cdot)$ is centered, for any $k \geq 0$

$$\begin{aligned} |\mathbb{E}_x g(\zeta, X_k)| &\leq \|g(\zeta, \cdot)\|_{d, \gamma, W} r_k^\gamma V^\gamma(x) \left(\mathbb{E}_x W(X_k) + \int_{\mathbb{X}} W d\pi \right)^{1-\gamma}, \\ |\mathbb{E}_x g(\zeta_1, X_k) - g(\zeta_2, X_k)| &\leq \|g(\zeta_1, \cdot) - g(\zeta_2, \cdot)\|_{d, \gamma, W} r_k^\gamma V^\gamma(x) \left(\mathbb{E}_x W(X_k) + \int_{\mathbb{X}} W d\pi \right)^{1-\gamma}; \end{aligned}$$

see (3.25). Then, by the condition 2) of Theorem 3.23,

$$|\mathbb{E}_x g(\zeta, X_k)| \leq C r_k^\gamma H_k(x), \quad |\mathbb{E}_x g(\zeta_1, X_k) - g(\zeta_2, X_k)| \leq C |\zeta_1 - \zeta_2|^\delta r_k^\gamma H_k(x)$$

with

$$H_k(x) = V^\gamma(x) \left(\mathbb{E}_x W(X_k) + \int_{\mathbb{X}} W \, d\pi \right)^{1-\gamma}.$$

Then the required bounds for $\mathcal{R}g$ hold true with

$$H^{\mathcal{R}g}(x) = C \sum_{k \geq 0} r_k^\gamma H_k(x).$$

Note that $H^{\mathcal{R}g} \in L_{2+\delta}(\mathbb{X}, \pi)$ because, by the Hölder inequality and the condition 4),

$$\begin{aligned} \|H_k\|_{L_{2+\delta}(\mathbb{X}, \pi)}^{2+\delta} &= \mathbb{E}_\pi \left[V^\gamma(X_0) \left(\mathbb{E}_{X_0} W(X_k) + \int_{\mathbb{X}} W \, d\pi \right)^{1-\gamma} \right]^{2+\delta} \\ &\leq \left(\mathbb{E}_\pi V^{2+\delta}(X_0) \right)^\gamma \left[\mathbb{E}_\pi \left(W(X_k) + \int_{\mathbb{X}} W \, d\pi \right)^{2+\delta} \right]^{1-\gamma} \\ &\leq 2^{(2+\delta)(1-\gamma)} \left(\int_{\mathbb{X}} V^{2+\delta} \, d\pi \right)^\gamma \left(\int_{\mathbb{X}} W^{2+\delta} \, d\pi \right)^{1-\gamma} = C < \infty, \quad k \geq 0. \end{aligned}$$

For f, g the construction of respective functions H^f, H^g is similar and simpler: in that case one needs to consider the term with $k = 0$, only; we omit the details. By choosing

$$H = H^f + H^g + H^{\mathcal{R}g}$$

we complete the proof. □

Proposition 3.28 The function $\mathcal{R}g$, considered as a map $\mathbb{R} \ni \zeta \mapsto \mathcal{R}g(\zeta, \cdot) \in L_2(\mathbb{X}, \pi)$, has a continuous derivative, which has a representation

$$\partial_\zeta \mathcal{R}g(\zeta, x) = \sum_{k=0}^{\infty} \mathbb{E}_x \partial_\zeta g(\zeta, X_k).$$

In addition, there exists a function $\tilde{H} : \mathbb{X} \rightarrow \mathbb{R}^+$ such that

$$\int_{\mathbb{X}} \tilde{H}^{2+\delta} \, d\pi < \infty$$

and

$$|\partial_\zeta g(\zeta, x)| \leq \tilde{H}(x), \quad |\partial_\zeta g(\zeta_1, x) - \partial_\zeta g(\zeta_2, x)| \leq |\zeta_1 - \zeta_2|^\delta \tilde{H}(x),$$

$$|\partial_\zeta \mathcal{R}g(\zeta, x)| \leq \tilde{H}(x), \quad |\partial_\zeta \mathcal{R}g(\zeta_1, x) - \partial_\zeta \mathcal{R}g(\zeta_2, x)| \leq |\zeta_1 - \zeta_2|^\delta \tilde{H}(x).$$

Proof. Consider the potential of the function $\partial_\zeta g(\zeta, x)$:

$$\mathcal{R}(\partial_\zeta g)(\zeta, x) = \sum_{k=0}^{\infty} \mathbb{E}_x \partial_\zeta g(\zeta, X_k),$$

which is well defined by condition 2), because $\partial_\zeta g(\zeta, x)$ is centered. Denote also

$$\mathcal{R}^L(\partial_\zeta g)(\zeta, x) := \sum_{k=0}^L \mathbb{E}_x \partial_\zeta g(\zeta, X_k), \quad L \geq 0;$$

note that $\mathcal{R}^0(\partial_\zeta g) = \partial_\zeta g$. Similarly to the proof of the previous Proposition 3.27, one constructs a function $\tilde{H} \in L_{2+\delta}(\mathbb{X}, \pi)$ such that

$$\begin{aligned} \left| \mathcal{R}^L(\partial_\zeta g)(\zeta, x) \right| &\leq \tilde{H}(x), \\ \left| \mathcal{R}^L(\partial_\zeta g)(\zeta_1, x) - \mathcal{R}^L(\partial_\zeta g)(\zeta_2, x) \right| &\leq |\zeta_1 - \zeta_2|^\delta \tilde{H}(x), \quad L \geq 0. \end{aligned}$$

Hence, for every $L \geq 0$, the function $\mathcal{R}^L g(\partial_\zeta g)$, considered as a map $\mathbb{R} \rightarrow L_2(\mathbb{X}, \pi)$, is continuous. In addition, by the dominated convergence theorem,

$$\mathcal{R}^L g(\zeta_2, x) - \mathcal{R}^L g(\zeta_1, x) = \int_{\zeta_1}^{\zeta_2} \mathcal{R}^L(\partial_\zeta g)(\zeta, x) d\zeta$$

where the integral in the right hand side is well-defined in the mean square sense; we omit the details. Because $\mathcal{R}^L g$, $\mathcal{R}^L(\partial_\zeta g)$ converge to $\mathcal{R}g$, $\mathcal{R}(\partial_\zeta g)$ in $L_2(\mathbb{X}, \pi)$ as $L \rightarrow \infty$ for every fixed ζ , we have the same identity and the same bounds for $\mathcal{R}g$ and $\mathcal{R}(\partial_\zeta g)$. \square

Let

$$\tilde{\mathcal{R}}g = \mathcal{R}g - g.$$

For a given function $\phi \in C^3(\mathbb{R})$ with bounded derivatives, define the respective corrector term by

$$U_{k,n} = \phi'(\zeta_{k,n}) \tilde{\mathcal{R}}g(\zeta_{k,n}, X_k), \quad k \geq 0.$$

Lemma 3.29 For any $k \geq 1$,

$$\begin{aligned} \phi(\zeta_{k,n}) + \frac{U_{k,n}}{\sqrt{n}} &= \phi(\zeta_{k-1,n}) + \frac{U_{k-1,n}}{\sqrt{n}} \\ &\quad + \frac{1}{n} \phi'(\zeta_{k-1,n}) \left(f(\zeta_{k-1,n}, X_k) + g(\zeta_{k-1,n}, X_k) (\partial_\zeta \tilde{\mathcal{R}}g)(\zeta_{k-1,n}, X_k) \right) \\ &\quad + \frac{1}{2n} \phi''(\zeta_{k-1,n}) \left(g^2(\zeta_{k-1,n}, X_k) + 2g(\zeta_{k-1,n}, X_k) \tilde{\mathcal{R}}g(\zeta_{k-1,n}, X_k) \right) \\ &\quad + \phi'(\zeta_{k-1,n}) \frac{M_k - M_{k-1}}{\sqrt{n}} + \vartheta_{k,n}, \end{aligned} \quad (3.32)$$

where

$$\mathbb{E}|\vartheta_{k,n}| \leq Cn^{-1-\delta/2}, \quad (3.33)$$

and $\{M_k\}$ is a martingale w.r.t. $\{\mathcal{F}_k\}$ such that

$$\mathbb{E}(M_k - M_{k-1})^2 \leq C. \quad (3.34)$$

Proof. We have

$$\begin{aligned} \phi(\zeta_{k,n}) &= \phi(\zeta_{k-1,n}) + \phi'(\zeta_{k-1,n}) \left(\frac{1}{n} f(\zeta_{k-1,n}, X_k) + \frac{1}{\sqrt{n}} g(\zeta_{k-1,n}, X_k) \right) \\ &\quad + \frac{1}{2n} \phi''(\zeta_{k-1,n}) g^2(\zeta_{k-1,n}, X_k) + \vartheta_{k,n}^1 \end{aligned} \quad (3.35)$$

with the residue term

$$\vartheta_{k,n}^1 = \phi(\zeta_{k,n}) - \phi(\zeta_{k-1,n}) - \phi'(\zeta_{k-1,n})(\zeta_{k,n} - \zeta_{k-1,n}) - \frac{1}{2n} \phi''(\zeta_{k-1,n}) g^2(\zeta_{k-1,n}, X_k).$$

Next, recall that $\zeta_{k-1,n}$ is \mathcal{F}_{k-1} -measurable. Hence, by the property (ii) of the potential defined by (3.31), we obtain

$$\begin{aligned} \mathbb{E}[\mathcal{R}g(\zeta_{k-1,n}, X_k) | \mathcal{F}_{k-1}] &= \mathbb{E}[\mathcal{R}g(\zeta, X_k) | \mathcal{F}_{k-1}]_{\zeta=\zeta_{k-1,n}} \\ &= \tilde{\mathcal{R}}g(\zeta, X_{k-1}) \Big|_{\zeta=\zeta_{k-1,n}} = \tilde{\mathcal{R}}g(\zeta_{k-1,n}, X_{k-1}). \end{aligned}$$

Then

$$M_k = \sum_{j=1}^k (\tilde{\mathcal{R}}g(\zeta_{j-1,n}, X_j) - \mathcal{R}g(\zeta_{j-1,n}, X_{j-1})), \quad k \geq 0$$

is a martingale, and we have

$$\tilde{\mathcal{H}}g(\zeta_{k-1,n}, X_k) - \tilde{\mathcal{H}}g(\zeta_{k-1,n}, X_{k-1}) = -g(\zeta_{k-1,n}, X_k) + M_k - M_{k-1}. \quad (3.36)$$

Then

$$\begin{aligned} U_{k,n} - U_{k-1,n} &= (\phi'(\zeta_{k,n}) - \phi'(\zeta_{k-1,n}))\tilde{\mathcal{H}}g(\zeta_{k,n}, X_k) \\ &\quad + \phi'(\zeta_{k-1,n})(\tilde{\mathcal{H}}g(\zeta_{k,n}, X_k) - \tilde{\mathcal{H}}g(\zeta_{k-1,n}, X_k)) \\ &\quad + \phi'(\zeta_{k-1,n})(-g(\zeta_{k-1,n}, X_k) + M_k - M_{k-1}). \end{aligned}$$

Multiplying by $1/\sqrt{n}$ and summing with (3.35), we get the required identity (3.32) with the residue term

$$\vartheta_{k,n} = \vartheta_{k,n}^1 + \vartheta_{k,n}^2 + \vartheta_{k,n}^3 + \vartheta_{k,n}^4,$$

where $\vartheta_{k,n}^1$ is defined above, and

$$\begin{aligned} \vartheta_{k,n}^2 &= \frac{1}{\sqrt{n}}(\phi'(\zeta_{k,n}) - \phi'(\zeta_{k-1,n}))(\tilde{\mathcal{H}}g(\zeta_{k,n}, X_k) - \tilde{\mathcal{H}}g(\zeta_{k-1,n}, X_k)), \\ \vartheta_{k,n}^3 &= \frac{1}{\sqrt{n}}\left(\phi'(\zeta_{k,n}) - \phi'(\zeta_{k-1,n}) - \frac{1}{\sqrt{n}}g(\zeta_{k,n}, X_k)\phi''(\zeta_{k-1,n})\right)\tilde{\mathcal{H}}g(\zeta_{k-1,n}, X_k), \\ \vartheta_{k,n}^4 &= \frac{1}{\sqrt{n}}\phi'(\zeta_{k-1,n})\left(\tilde{\mathcal{H}}g(\zeta_{k,n}, X_k) - \tilde{\mathcal{H}}g(\zeta_{k-1,n}, X_k) - \frac{1}{\sqrt{n}}g(\zeta_{k,n}, X_k)\partial_\zeta\tilde{\mathcal{H}}g(\zeta_{k-1,n}, X_k)\right). \end{aligned}$$

Let us proceed with the bounds on the residue terms $\vartheta_{k,n}^i$, $i = 1, 2, 3, 4$. Because ϕ has bounded derivatives ϕ' , ϕ'' , ϕ''' , one has

$$\left|\phi(y) - \phi(x) - (y-x)\phi'(x) - \frac{1}{2}(y-x)^2\phi''(x)\right| \leq C|y-x|^{2+\delta}.$$

Then

$$\begin{aligned} |\vartheta_{k,n}^1| &\leq C\left|\frac{1}{n}f(\zeta_{k-1,n}, X_k) + \frac{1}{\sqrt{n}}g(\zeta_{k-1,n}, X_k)\right|^{2+\delta} \\ &\quad + \frac{1}{2n}\phi''(\zeta_{k-1,n})\left|\left(\frac{1}{\sqrt{n}}f(\zeta_{k-1,n}, X_k) + g(\zeta_{k-1,n}, X_k)\right)^2 - g^2(\zeta_{k-1,n}, X_k)\right|. \end{aligned}$$

By Proposition 3.27, we have then $|\vartheta_{k,n}^1| \leq Cn^{-1-\delta/2}H^{2+\delta}(X_k)$, which yields the bound

$$\mathbb{E}|\vartheta_{k,n}^1| \leq Cn^{-1-\delta/2}.$$

The bound for $\vartheta_{k,n}^3$ is similar, and we leave it for the reader; we only note that because $\tilde{\mathcal{H}}g = \mathcal{R}g - g$, from Proposition 3.27 we get $|\tilde{\mathcal{H}}g(\zeta, x)| \leq 2H(x)$. To bound $\vartheta_{k,n}^2$, observe that by Proposition 3.27

$$|\vartheta_{k,n}^2| \leq \frac{C}{\sqrt{n}} |\zeta_{k,n} - \zeta_{k-1,n}|^{1+\delta} H(X_k),$$

and

$$|\zeta_{k,n} - \zeta_{k-1,n}| = \left| \frac{1}{n} f(\zeta_{k-1,n}, X_k) + \frac{1}{\sqrt{n}} g(\zeta_{k-1,n}, X_k) \right| \leq \frac{C}{\sqrt{n}} H(X_k).$$

Hence $|\vartheta_{k,n}^2| \leq Cn^{-1-\delta/2} H^{2+\delta}(X_k)$, and

$$\mathbb{E}|\vartheta_{k,n}^2| \leq Cn^{-1-\delta/2}.$$

The bound for $\vartheta_{k,n}^4$ can be obtained similarly, using the integral identity

$$\tilde{\mathcal{H}}g(\zeta_{k,n}, X_k) - \tilde{\mathcal{H}}g(\zeta_{k-1,n}, X_k) = \int_{\zeta_{k-1,n}}^{\zeta_{k,n}} \partial_{\zeta} \tilde{\mathcal{H}}g(\zeta, X_k) d\zeta$$

and the Hölder continuity of $\partial_{\zeta} \tilde{\mathcal{H}}g(\zeta, x)$ w.r.t. ζ ; we leave the details for the reader. \square

Proof of Theorem 3.23. Step I: weak compactness. To prove that the sequence $\{Y_n\}$ is weakly compact in the sense of convergence of finite-dimensional distributions, it suffices to prove that

$$\limsup_{n \rightarrow \infty} \sup_{s,t \in [0,1]: |t-s| < \delta} \mathbb{E}|Y_n(t) - Y_n(s)| \rightarrow 0, \quad \delta \rightarrow 0. \quad (3.37)$$

Choose $\phi(x) = x$, then by Lemma 3.29 we have

$$Y_n(t) = Y_n^0(t) + Y_n^1(t) + Y_n^2(t) + Y_n^3(t),$$

where

$$Y_n^0(t) = y_0 + \frac{U_{0,n} - U_{[nt],n}}{\sqrt{n}},$$

$$Y_n^1(t) = \frac{1}{n} \sum_{k=1}^{[nt]} \left(f(\zeta_{k-1,n}, X_k) + g(\zeta_{k-1,n}, X_k) (\partial_{\zeta} \tilde{\mathcal{H}}g)(\zeta_{k-1,n}, X_k) \right),$$

$$Y_n^2(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} (M_k - M_{k-1}),$$

$$Y_n^3(t) = \sum_{k=1}^{[nt]} \vartheta_{k,n};$$

here we used that $\phi'(x) = 1$, $\phi''(x) = 0$. Let us show that for every $i = 0, 1, 2, 3$, (3.37) holds true with Y_n replaced by Y_n^i . By Proposition 3.27,

$$|U_{k,n}| = |\tilde{\mathcal{R}}g(\zeta_{k,n}, X_k)| \leq 2H(X_k),$$

and therefore the sequence $E|U_{k,n}|$ is bounded (recall that X is stationary and, in particular, every X_k obeys law π). Then

$$E|Y_n^0(t) - Y_n^0(s)| \leq \frac{C}{\sqrt{n}},$$

and 3.37 holds true for $Y_n^0(t)$. By Proposition 3.27 and Proposition 3.28,

$$\left| f(\zeta_{k-1,n}, X_k) + g(\zeta_{k-1,n}, X_k) (\partial_{\zeta} \tilde{\mathcal{R}}g)(\zeta_{k-1,n}, X_k) \right| \leq H(X_k) + 2H(X_k)\tilde{H}(X_k),$$

then

$$E|Y_n^1(t) - Y_n^1(s)| \leq \frac{1}{n} \sum_{[ns] < k \leq [nt]} E(H(X_k) + 2H(X_k)\tilde{H}(X_k)) \leq C \left(|t-s| + \frac{1}{n} \right),$$

which implies (3.37) for $Y_n^1(t)$. Using Proposition 3.27 once again, we get

$$|M_k - M_{k-1}| = |\tilde{\mathcal{R}}g(\zeta_{k-1,n}, X_k) - \mathcal{R}g(\zeta_{k-1,n}, X_{k-1})| \leq 3H(X_k),$$

which shows that the sequence $E(M_k - M_{k-1})^2$ is bounded. Because $\{M_k\}$ is a martingale, this yields

$$E(Y_n^2(t) - Y_n^2(s))^2 = \frac{1}{n} \sum_{[ns] < k \leq [nt]} E(M_k - M_{k-1})^2 \leq C \left(|t-s| + \frac{1}{n} \right),$$

which implies (3.37) for $Y_n^2(t)$. Finally, by (3.33),

$$\sup_{t \in [0,1]} \mathbb{E}|Y_n^3(t)| \rightarrow 0, \quad n \rightarrow \infty,$$

which implies (3.37) for $Y_n^3(t)$ and completes the proof of the weak compactness.

Step II: identification of a limit point. Let Y be a limit point for the sequence $\{Y_n\}$; that is, for some subsequence $\{Y_{n_k}\}$ all its finite-dimensional distributions converge weakly to the respective distributions of Y . By (3.37) and by Fatou's lemma, the process Y is continuous in L_1 :

$$\sup_{s,t \in [0,1]: |t-s| < \delta} \mathbb{E}|Y(t) - Y(s)| \rightarrow 0.$$

In particular, there exists its measurable modification. In what follows, we assume that Y itself is measurable; in order to simplify the notation, we assume that Y is a weak limit of the entire sequence $\{Y_n\}$.

Fix $\phi \in C_0^\infty(\mathbb{R})$. To prove that

$$\phi(Y(t)) - \int_0^t \mathcal{L}\phi(Y(s)) \, ds, \quad t \in [0, 1]$$

is a martingale w.r.t. the natural filtration of Y , it suffices to prove the following: for every $t > s$, $m \geq 1$, $s_1, \dots, s_m \in [0, s]$, $F \in C_b(\mathbb{R}^m)$

$$\mathbb{E} \left(\phi(Y(t)) - \int_s^t \mathcal{L}\phi(Y(v)) \, dv \right) G(Y(s_1), \dots, Y(s_m)) = 0. \quad (3.38)$$

Denote

$$\begin{aligned} \mathcal{Q}\phi(\zeta, x) &= \phi'(\zeta) \left(f(\zeta, x) + g(\zeta, x) (\partial_\zeta \tilde{\mathcal{H}}g)(\zeta, x) \right) \\ &\quad + \frac{1}{2} \phi''(\zeta) (g^2(\zeta, x) + 2g(\zeta, x) \tilde{\mathcal{H}}g(\zeta, x)), \end{aligned} \quad (3.39)$$

then by (3.32), for every $n \geq 1$, the process

$$\phi(Y_n(t)) + \frac{1}{\sqrt{n}} U_{[nt], n} - \int_0^{[nt]} \left(\mathcal{Q}\phi(Y_n(v), X_{[nv]}) + n \mathfrak{D}_{[nv], n} \right) \, dv$$

is a martingale w.r.t. the filtration $\{\mathcal{F}_t^n = \sigma(X_k, k \leq nt), t \in [0, 1]\}$. Recall that $\{\mathfrak{D}_{k,n}\}$

satisfies (3.33) and

$$\mathbb{E}|U_{k,n}| \leq C, \quad \mathbb{E}|\mathcal{Q}\phi(Y_n(v), X_{[nv]})| \leq C$$

by Proposition 3.27. Hence

$$\mathbb{E} \left(\phi(Y_n(t)) - \int_s^t \mathcal{Q}\phi(Y_n(v), X_{[nv]}) \, dv \right) F(Y_n(s_1), \dots, Y_n(s_m)) \rightarrow 0, \quad n \rightarrow \infty. \quad (3.40)$$

Because $(Y_n(t), Y_n(s_1), \dots, Y_n(s_m))$ converges weakly to $(Y(t), Y(s_1), \dots, Y(s_m))$ and ϕ , F are continuous and bounded,

$$\mathbb{E}\phi(Y_n(t))F(Y_n(s_1), \dots, Y_n(s_m)) \rightarrow \mathbb{E}\phi(Y(t))F(Y(s_1), \dots, Y(s_m)), \quad n \rightarrow \infty.$$

Hence, to obtain (3.38), we have to prove that

$$\begin{aligned} & \mathbb{E} \left(\int_s^t \mathcal{Q}\phi(Y_n(v), X_{[nv]}) \, dv \right) F(Y_n(s_1), \dots, Y_n(s_m)) \\ & \rightarrow \mathbb{E} \left(\int_s^t \mathcal{L}_\phi(Y(v)) \, dv \right) F(Y(s_1), \dots, Y(s_m)), \quad n \rightarrow \infty. \end{aligned} \quad (3.41)$$

We follow the same idea we used in the proof of Theorem 3.13: first, we will *delay* the time variable in the term $Y_n(v)$ under the integral; second, we will make use of the *stabilization* property and derive (3.41).

By Proposition 3.27 and Proposition 3.28,

$$|\mathcal{Q}\phi(\zeta, x)| \leq G(x), \quad |\mathcal{Q}\phi(\zeta_1, x) - \mathcal{Q}\phi(\zeta_2, x)| \leq |\zeta_1 - \zeta_2|^\delta G(x),$$

where

$$G = C(H + \tilde{H})^2 \in L_{1+\delta/2}(\mathbb{X}, \pi).$$

Denote $\delta' = \delta/(2+\delta) < \delta$, then

$$|\mathcal{Q}\phi(\zeta_1, x) - \mathcal{Q}\phi(\zeta_2, x)| \leq \max \{ |\zeta_1 - \zeta_2|^\delta G(x), 2G(x) \} \leq 2|\zeta_1 - \zeta_2|^{\delta'} G(x).$$

Hence for any $v_n \in [0, v]$,

$$\mathbb{E}|\mathcal{Q}\phi(Y_n(v), X_{[nv]}) - \mathcal{Q}\phi(Y_n(v - v_n), X_{[nv]})|$$

$$\begin{aligned}
&\leq 2\mathbb{E}|Y_n(v) - Y_n(v - v_n)|^{\delta'} G(X_k) \\
&\leq 2 \left(\mathbb{E}|Y_n(v) - Y_n(v - v_n)|^{\delta'(2+\delta)/\delta} \right)^{\delta/(2+\delta)} \left(\mathbb{E}(G(X_k))^{1+\delta/2} \right)^{2/(2+\delta)} \\
&= C(\mathbb{E}|Y_n(v) - Y_n(v - v_n)|)^{\delta/(2+\delta)}.
\end{aligned}$$

Using (3.37) and the previous inequality, we get

$$\mathbb{E} \left(\int_{s+v_n}^t \left(\mathcal{Q}\phi(Y_n(v), X_{[nv]}) - \mathcal{Q}\phi(Y_n(v - v_n), X_{[nv]}) \right) dv \right) F(Y_n(s_1), \dots, Y_n(s_m)) \rightarrow 0$$

for any deterministic sequence $\{v_n\}$ such that $v_n \rightarrow 0$. Note also that the integrals over the time interval $[s, s + v_n]$ in both sides of (3.41) are negligible because $\mathcal{Q}\phi$, $\mathcal{L}\phi$ are bounded.

This was the first (*delay*) part of the argument. The second (*stabilization*) part is almost literally similar to that in the proof of Theorem 3.13, and hence we just sketch it. Let $\mathcal{Q}^L\phi$, $L \geq 1$ be the functions defined by the formula (3.39), where $\tilde{\mathcal{R}}$, $\partial_\zeta \tilde{\mathcal{R}}$ are replaced by $\tilde{\mathcal{R}}^L$, $\partial_\zeta \tilde{\mathcal{R}}^L$, $L \geq 1$, respectively; see the notation in the proof of Proposition 3.28. Let $\{v_n\}$ be such that $nv_n \rightarrow 0$, then for any $L \geq 1$,

$$\mathbb{E} \left(\int_{s+v_n}^t \left(\mathcal{Q}^L\phi(Y_n(v - v_n), X_{[nv]}) - \mathcal{L}^L\phi(Y_n(v - v_n)) \right) dv \right) F(Y_n(s_1), \dots, Y_n(s_m)) \rightarrow 0,$$

where

$$\mathcal{L}^L\phi(\zeta) = \int_{\mathbb{X}} \mathcal{Q}^L\phi(\zeta, x) \pi(dx).$$

This relation follows from our principal assumption (3.16); the argument is very similar to that in the proof of Lemma 3.20, and we leave the detailed proof as an exercise for the reader. It follows from the estimates given in the proof of Proposition 3.27 and Proposition 3.28 that

$$\mathbb{E} \int_{s+v_n}^t \left| \mathcal{Q}^L\phi(Y_n(v - v_n), X_{[nv]}) - \mathcal{Q}\phi(Y_n(v - v_n), X_{[nv]}) \right| dv \rightarrow 0, \quad L \rightarrow \infty$$

uniformly w.r.t. s, t, n , and that

$$\mathcal{L}^L\phi(\zeta) \rightarrow \mathcal{L}\phi(\zeta), \quad L \rightarrow \infty$$

uniformly w.r.t. ζ . Every function $\mathcal{L}^L\phi$ is continuous (actually, even Hölder continuous with the index δ) and bounded, hence $\mathcal{L}\phi$ is continuous and bounded, as well. Then it follows from the weak convergence of Y_n to Y and (3.37), that

$$\begin{aligned} & \mathbb{E} \left(\int_{s+v_n}^t \mathcal{L}\phi(Y_n(v-v_n)) \, dv \right) F(Y_n(s_1), \dots, Y_n(s_m)) \\ &= \left(\int_s^{t-v_n} \mathcal{L}\phi(Y_n(v)) \, dv \right) F(Y_n(s_1), \dots, Y_n(s_m)) \\ &\rightarrow \mathbb{E} \left(\int_s^t \mathcal{L}\phi(Y(v)) \, dv \right) F(Y(s_1), \dots, Y(s_m)), \quad n \rightarrow \infty; \end{aligned}$$

again, we leave the detailed proof for the reader. This completes both the proof of (3.41) and the proof of Theorem 3.23. \square

Bibliography

- [1] Bally, V., Clément, E.: *Integration by parts formula and applications to equations with jumps*, Probab. Theory Relat. Fields **151**(3–4), 613–657 (2011).
- [2] Bernstein, S. N.: *Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes*, Math. Ann. **97**, 1–59 (1926).
- [3] Bichteler, K., Gravereaux, J.-B., and Jacod, J.: *Malliavin calculus for processes with jumps*. Gordon and Breach, New York (1987).
- [4] Billingsley, P.: *Convergence of probability measures*, 2nd ed., Wiley, New York (1999).
- [5] Bodnarchuk, S., Kulik, A.: *Conditions for existence and smoothness of the distribution density of an Ornstein-Uhlenbeck process with Lévy noise*, Prob. Theory and Math. Stat. **79**, 20–33 (2008).
- [6] Butkovsky, O.: *Subgeometric rates of convergence of Markov processes in the Wasserstein metric*, Ann. Appl. Probab. **24**(2), 526–552 (2012).
- [7] Da Prato, G., Zabczyk, Z.: *Stochastic Equations in Infinite Dimensions*, Cambridge University Press, Cambridge (1992).
- [8] Da Prato, G., Zabczyk, Z.: *Ergodicity for Infinite-Dimensional Systems*, Cambridge University Press, Cambridge (1996).
- [9] Dobrushin, R.L.: *Central limit theorem for nonstationary Markov chains*, Theor. Prob. Appl. **1**, 65–80; 329–383 (1956).

- [10] Doeblin, W.: *Sur les propriétés asymptotiques de mouvements régis par certains types de chaînes simples*, Thèse de doctorat Sciences Mathématiques, Paris (1938).
- [11] Doeblin, W.: *Eléments d'une théorie générale des chaînes simples constantes de Markoff*, Ann. Sci. E.N.S. **57**, 61–111 (1940).
- [12] Doob, J.L. *Stochastic processes*. Wiley, New York (1953).
- [13] Douc, R., Fort, G., and Guillin, A.: *Subgeometric rates of convergence of f -ergodic strong Markov processes*, Stoch. Proc. and Appl. **119**(2), 897–923 (2009).
- [14] Dudley, R.M.: *Real analysis and probability*. Cambridge Univ. Press, Cambridge (2002).
- [15] Dynkin, E.B.: *Markov processes*, vols. 1 and 2. Academic Press, New York (1965).
- [16] Dynkin, E.B., Yushkevich, A.A.: *Controlled Markov processes*, Springer, Berlin (1979).
- [17] Friedman, A.: *Partial differential equations of parabolic type*. Prentice-Hall, New York (1964).
- [18] Griffeath, D.: *Uniform coupling of non-homogeneous Markov chains*, J. Appl. Probab. **12**(4), 753–762 (1975).
- [19] Eidelman, S.D., Ivasyshen, S.D., and Kochubei, A.N.: *Analytic methods in the theory of differential and pseudo-differential equations of parabolic type*. Birkhäuser, Basel (2004).
- [20] Ethier, S.N., Kurtz, T.G.: *Markov Processes. Characterization and Convergence*. Wiley, New York (1986).
- [21] Hairer, M.: *Ergodic properties of Markov processes*. Lecture Notes (2006), <http://www.hairer.org/notes/Markov.pdf>.
- [22] Hairer, M.: *P@W course on the convergence of Markov processes*. Lecture Notes (2010), <http://www.hairer.org/notes/Convergence.pdf>.

-
- [23] Hairer, M., Mattingly, J. and Scheutzow, M.: *Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations*, Prob. Theory Rel. Fields, **149**(1–2), 223–259 (2011).
- [24] Harris, T.E.: *The existence of stationary measures for certain Markov processes*, in: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. II, 113–124. University of California Press, Berkeley (1956).
- [25] Ibragimov, I.A., Linnik, Yu.V.: *Independent and stationary sequences of random variables*, Wolters-Noordhoff, Groningen (1971).
- [26] Ikeda, N., Watanabe, S.: *Stochastic differential equations and diffusion processes*, North-Holland, Amsterdam (1981).
- [27] Ishikawa, Y. and Kunita, H.: *Existence of density for canonical differential equations with jumps*, Stoch. Process. Appl., **116**(12), 1743–1769 (2006).
- [28] Khasminskii, R.Z.: *Stochastic Stability of Differential Equations*. Sijthoff & Noordhoff, Alphen aan den Rijn (1980).
- [29] Klokov, S.A., Veretennikov, A.Yu.: *On the sub-exponential mixing rate for a class of Markov diffusions*, J. Math. Sci. (N.Y.) **123**(1), 3816–3823 (2004).
- [30] Klokov S.A., Veretennikov, A.Yu.: *Sub-exponential mixing rate for a class of Markov chains*, Math. Comm. **9**, 9–26 (2004).
- [31] Komorowski T., Walczuk, A.: *Central limit theorem for Markov processes with spectral gap in the Wasserstein metric*, Stoch. Proc. Appl. **122**, 2155–2184 (2012).
- [32] Komorowski, T., Peszat, Sz., Szarek, T.: *On ergodicity of some Markov processes*, Ann. Probab. **38**, 1401–1443 (2010).
- [33] Koroliuk, V.S., Limnios, N.: *Stochastic Systems in Merging Phase Space*. World Scientific, New Jersey (2005).
- [34] Kulik, A.M.: *Exponential ergodicity of the solutions to SDE's with a jump noise*, Stoch. Proc. and Appl. **119**(2), 602–632 (2009).

- [35] Kulik, A.M.: *Asymptotic and spectral properties of exponentially ϕ -ergodic Markov processes*, Stoch. Proc. and Appl. **121**(5), 1044–1075 (2011).
- [36] Kulik, A.M.: *Absolute continuity and convergence in variation for distributions of functionals of Poisson point measure*, J. Theor. Probab. **24**, 1–38 (2011).
- [37] Kulik, A.M., Veretennikov, A.Yu.: *Diffusion approximation for systems with weakly ergodic Markov perturbations I, II*, Theor. Probab. Math. Stat. **87**, 1–16 (2012); **88**, 1–16 (2013).
- [38] Kunita, H.: *Absolute continuity of Markov processes and generators*, Nagoya Math. J. **36**, 1–26 (1969).
- [39] Kunita, H.: *Stochastic Flows and Stochastic Differential Equations*. Cambridge University Press, Cambridge (1997).
- [40] Markov, A.A.: *Rasprostranenie zakona bolshikh chisel na velichiny, zavisyashchie drug ot druga (The Extension of the Law of Large Numbers on Mutually Dependent Variables)*. Ivestia Fiz.-mat. obsch. Kazan univ. **4**(15) 135–156 (1906). See also [42], pp. 339–361.
- [41] Markov, A.A.: *Rasprostranenie predel'nyh teorem ischisleniya veroyatnostej na summu velichin svyazannyh v cep' (Extension of the limit theorems of probability theory to a sum of variables connected in a chain)*, Zapiski Akademii Nauk po Fiziko-matematicheskomu otdeleniyu, VIII seriya **25**(3) (1908). Translated into English, in: R.A. Howard (Ed.), *Dynamic Probabilities Systems*, vol. 1. Wiley, New York (1971), 552–576. See also [42], pp. 364–397.
- [42] Markov, A.A.: *Izbrannye trudy. Teoria chisel. Teoria veroyatnostei*. In: Selected Works. Number Theory. Probability Theory, Yu.V. Linnik (ed.), Moscow-Leningrad (1951).
- [43] Meyn, S.P., Tweedie, R.L.: *Markov chains and stochastic stability*. Springer, London (1993).
- [44] Mohammed, S.E.A.: *Stochastic functional differential equations*. Research Notes in Mathematics **99**, Pitman, Boston (1984).

-
- [45] Numellin, E.: *General irreducible Markov chains and non-negative operators*, 2nd ed., Cambridge University Press, Cambridge (2004).
- [46] Papanicolaou, G.C., Stroock, D.W., Varadhan, S.R.S.: *Martingale approach to some limit theorems*. In: Conference on Statistical Mechanics, Dynamical Systems and Turbulence, M. Reed (ed.), Duke Univ. Press (1977).
- [47] Pardoux, E., Veretennikov A.Yu.: *On Poisson equation and diffusion approximation 1, 2*, Ann. Prob. **29**, 1061–1085 (2001); Ann. Prob. **31**, 1166–1192 (2003).
- [48] Peszat, S., Zabczyk, J.: *Stochastic Partial Differential Equations with Lévy Noise (an Evolution Equation Approach)*. Cambridge University Press, Cambridge (2007).
- [49] Picard, J.: *On the existence of smooth densities for jump processes*, Probab. Theory Rel. Fields **105**, 481–511 (1996).
- [50] Rachev, S.T., Rüschendorf, L.: *Mass Transportation Problems, Volume I: Theory*. Springer, Berlin (1998).
- [51] Revuz, D., Yor, M.L: *Continuous martingales and Brownian motion*, 3rd ed. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. **293**. Springer-Verlag, Berlin (1999).
- [52] Scheutzow, M.: *Exponential growth rates for stochastic delay differential equations*, Stoch. Dyn. **5**, 163–174 (2005).
- [53] Szarek, T.: *The uniqueness of invariant measures for Markov operators*. Studia Math. **189**, 225–233 (2008).
- [54] Stroock, D.W., Varadhan, S.R.S.: *Multidimensional diffusion processes*. Springer, Berlin (1979).
- [55] Thorisson, H.: *Coupling, stationarity, and regeneration*. Springer, New York (2000).
- [56] Veretennikov, A.Yu.: *Bounds for the mixing rates in the theory of stochastic equations*, Theory Probab. Appl. **32**, 273–281 (1987).
- [57] Veretennikov, A.Yu.: *On polynomial mixing bounds for stochastic differential equations*, Stoch. Proc. Appl. **70**(1), 115–127 (1997).

- [58] Wagner, D.H.: *Survey of measurable selection theorems: An update*. Lecture Notes in Mathematics, Volume **794**, 176–219 (1980).
- [59] Zolotarev, V.M.: *Sums of independent random variables: a modern approach*. Wiley, New York (1995).

The present lecture notes aim for an introduction to the ergodic behaviour of Markov processes and addresses graduate students, post-graduate students and interested readers.

Different tools and methods for the study of upper bounds on uniform and weak ergodic rates of Markov Processes are introduced. These techniques are then applied to study limit theorems for functionals of Markov processes.

This lecture course originates in two mini courses held at the University of Potsdam, Technical University of Berlin and Humboldt University in spring 2013 and Ritsumameikan University in summer 2013.

Alexei Kulik, Doctor of Sciences, is a Leading researcher at the Institute of Mathematics of Ukrainian National Academy of Sciences.

