Труды Москов. Матем. Общ.
Том 77 (2016), вып. 1

Trans. Moscow Math. Soc.
2016, Pages 1–53
http://dx.doi.org/10.1090/mosc/256
Article electronically published on November 28, 2016

ESTIMATES OF THE RATE OF CONVERGENCE IN THE VON NEUMANN AND BIRKHOFF ERGODIC THEOREMS

ALEKSANDR G. KACHUROVSKII AND IVAN V. PODVIGIN

Abstract. We present estimates (which are necessarily spectral) of the rate of convergence in the von Neumann ergodic theorem in terms of the singularity at zero of the spectral measure of the function to be averaged with respect to the corresponding dynamical system as well as in terms of the decay rate of the correlations (i.e., the Fourier coefficients of this measure). Estimates of the rate of convergence in the Birkhoff ergodic theorem are given in terms of the rate of convergence in the von Neumann ergodic theorem as well as in terms of the decay rate of the large deviation probabilities. We give estimates of the rate of convergence in both ergodic theorems for some classes of dynamical systems popular in applications, including some well-known billiards and Anosov systems.

Introduction

The problem about the rates of convergence in ergodic theorems naturally arises in physical applications of these theorems; von Neumann [1] was the first to consider it in 1932. Since the ergodic theorems themselves stem from attempts to justify the ergodic hypothesis in statistical mechanics, it is naturally of interest to consider, say, the problem on the rate of convergence in these theorems for various billiards modeling a gas, for example, when averaging the characteristic function of a subset of the phase space. (The ergodic means of this function give the mean time spent in the corresponding subset of the phase space of the billiard.) It is not clear without solving this problem how long a physicist has to wait for the convergence guaranteed by the ergodic theorems and whether he or she lives to see any noticeable manifestations of this convergence. Once we state the problem, we immediately have to ask what units should be used to measure the rates of convergence in ergodic theorems, what characteristics of dynamical systems affect these rates and in what way, and which of these characteristics can be computed for specific dynamical systems to be studied (say, for the above-mentioned billiards).

The aim of the present survey is to provide a description of new positive results obtained since 1996 (when the preceding survey [2] was published) on the rate of convergence in the von Neumann and Birkhoff ergodic theorems.

The recent years' progress aimed at possible applications of the general theory has finally permitted obtaining estimates of these rates of convergence for various specific dynamical systems that are of interest in applications, including some well-known billiards and Anosov systems. A selection of these new, yet unpublished specific estimates is also included in the survey.

2010 *Mathematics Subject Classification.* Primary 37A30; Secondary 37D20, 37D50, 60G10.

Key words and phrases. Convergence rates in ergodic theorems, correlation decay, large deviation decay, billiard, Anosov system.

The research was supported by the Program for State Support of Leading Scientific Schools of the Russian Federation (grant NSh-5998.2012.1).

0.1. **Notation.** Let $(\Omega, \mathfrak{F}, \lambda)$ be a space with a probability measure, and let T be an endomorphism of this space, that is, a map $T\colon \Omega \to \Omega$ such that $T^{-1}A \in \mathfrak{F}$ and $\lambda(A) = \lambda(T^{-1}A)$ for all $A \in \mathfrak{F}$. Further, let $\{T^t,\ t \in \mathbb{R}^+\}$ be a semiflow on $(\Omega, \mathfrak{F}, \lambda)$, that is, a one-parameter group of endomorphisms T^t of this space such that the function $f(T^t\omega)$ is measurable on the Cartesian product $\Omega \times \mathbb{R}^+$ for any measurable function $f(\omega)$ on Ω. The endomorphism T defines a discrete-time dynamical system; the semiflow T^t is called a continuous-time dynamical system.

For $f \in L_1(\Omega)$, $\omega \in \Omega$, $n \in \mathbb{N}$, and $t \in \mathbb{R}^+$, set

$$A_n f(\omega) = \frac{1}{n}\sum_{k=0}^{n-1} f(T^k\omega), \qquad \overline{A}_t f(\omega) = \frac{1}{t}\int_0^t f(T^\tau\omega)\, d\tau.$$

Birkhoff's pointwise ergodic theorem claims that the limits $f^* = \lim_{n\to\infty} A_n f$ and $\overline{f}^* = \lim_{t\to\infty} \overline{A}_t f$ exist λ-almost everywhere and satisfy

$$\int f^*\, d\lambda = \int f\, d\lambda = \int \overline{f}^*\, d\lambda.$$

Let U_T be the isometric operator acting on the (complex) Hilbert space $L_2(\Omega)$ by the formula $U_T f = f \circ T$; accordingly, $\{U^t, t \in \mathbb{R}^+\}$ is the one-parameter semigroup of isometric operators acting on the same space by the formula $U^t f = f \circ T^t$. Von Neumann's statistical ergodic theorem claims the existence of the following limits in $L_2(\Omega)$:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} U_T^k f = \lim_{n\to\infty} A_n f = f^*, \qquad \lim_{t\to\infty} \frac{1}{t}\int_0^t U^\tau f\, d\tau = \lim_{t\to\infty} \overline{A}_t f = \overline{f}^*;$$

further, it turns out that f^* is the orthogonal projection of f onto the subspace of fixed vectors of U_T and $\overline{f}^*$ is the orthogonal projection of f onto the subspace of fixed vectors of the semigroup U^t. Now let us define the correlation coefficients $b_k f$ and the spectral measure σ_f of the vector f with respect to the discrete-time dynamical system by setting $b_k f = (U_T^k f, f)$ for $k \geq 0$ and $b_k f = \overline{b_{-k} f}$ for $k < 0$. It is well known (e.g., see [3, Sec. 1.7]) that there exists a (unique) well-defined spectral measure σ_f, that is, a finite Borel measure on the unit circle such that

$$b_k f = \int_{(-\pi,\pi]} e^{ikx}\, d\sigma_f(x)$$

for all integer k (so that the correlation coefficients are its Fourier coefficients in the complex form).

The correlation function $b_t f$ and the spectral measure σ_f of a vector f with respect to a continuous-time dynamical system are defined in a similar way, $b_t f = (U^t f, f)$ for $t \geq 0$ and $b_t f = \overline{b_{-t} f}$ for $t < 0$. In this case, there exists a (unique) spectral measure σ_f (see [3, Sec. 1.7]), that is, a finite Borel measure on the real line such that

$$b_t f = \int_{-\infty}^{+\infty} e^{itx}\, d\sigma_f(x)$$

for all $t \in \mathbb{R}$.

The rate of convergence in the von Neumann ergodic theorem is the rate of convergence of the expressions $\|A_n f - f^*\|_2^2$ and $\|\overline{A}_t f - \overline{f}^*\|_2^2$ to zero as $n \to \infty$ and $t \to \infty$, respectively. The rate of convergence in the Birkhoff ergodic theorem is determined by the decay, for each $\varepsilon > 0$, of the sequences of numbers

$$\mathrm{P}_n^\varepsilon = \lambda\Big\{\sup_{k\geq n} |A_k f - f^*| \geq \varepsilon\Big\} \quad \text{as} \quad n \to \infty, \quad \overline{\mathrm{P}}_t^\varepsilon = \lambda\Big\{\sup_{s\geq t} |\overline{A}_s f - \overline{f}^*| \geq \varepsilon\Big\} \quad \text{as} \quad t \to \infty.$$

Note that the condition

$$\lim_{n\to\infty} \mathrm{P}_n^\varepsilon = 0 \quad \text{or} \quad \lim_{t\to\infty} \overline{\mathrm{P}}_t^\varepsilon = 0 \quad \text{for all } \varepsilon > 0$$

is equivalent to the a.e. convergence of $A_n f$ as $n \to \infty$ or $\overline{A}_t f$ as $t \to \infty$, respectively.

Along with these characteristics, we consider the large deviation probabilities

$$\mathrm{p}_n^\varepsilon = \lambda\{|A_n f - f^*| \geq \varepsilon\}, \qquad \overline{\mathrm{p}}_t^\varepsilon = \lambda\{|\overline{A}_t f - \overline{f}^*| \geq \varepsilon\}.$$

The property

$$\lim_{n\to\infty} \mathrm{p}_n^\varepsilon = 0 \quad \text{or} \quad \lim_{t\to\infty} \overline{\mathrm{p}}_t^\varepsilon = 0 \quad \text{for all } \varepsilon > 0$$

is, by definition, the convergence in measure of $A_n f$ as $n \to \infty$ or $\overline{A}_t f$ as $t \to \infty$, respectively.

0.2. **What affects the rates of convergence in ergodic theorems?** It is well known (see the detailed discussion of the topic and its history in [2]) that one cannot estimate the rates of convergence in these theorems solely in terms of the function f to be averaged; the estimates also depend on the choice of the dynamical system. The key role is played here by the spectral measures σ_f of f with respect to the dynamical systems T in question and by the Fourier coefficients of these measures, that is, by the correlation coefficients $b_n f$ (which depend on both f and T).

According to [3, Sec. 1.7]), "The spectral theory of dynamical systems studies the properties of dynamical systems which can be expressed via the properties of the measures σ_f." In this sense, estimates of the rate of convergence in ergodic theorems are necessarily spectral. For example, Theorem 3 in [2] (for the case of discrete time) and its exact continuous-time counterpart, Theorem 1 in [4], show that a power-law rate of convergence in the von Neumann ergodic theorem occurs if and only if the spectral measure of the function to be averaged with respect to the corresponding dynamical system has a power-law singularity with the same exponent at zero.

Further, the well-known formulas (1.1), (1.2) and (1.3), (1.4) given below show that the spectral measure, as well as the correlations, contains full information on the norms of deviations of all ergodic means from the limit. Hence, in our opinion, it is one of the most significant properties of the measures σ_f that these measures determine the rate of convergence of ergodic means in the norm; apparently, this obvious fact has not been especially mentioned in the literature yet. The converse is also true: V. V. Ryzhikov noticed and told us that if the function f is real-valued, then the exact values of all correlations can be reconstructed (and hence the corresponding spectral measure can be computed) from the exactly known norms of deviations of ergodic means from the limit; see Remarks 9 and 11.

Since the problem of computing the spectral measures for specific dynamical systems is practically unsolvable (to begin with, one has to have full information about all correlations, from which the spectral measures are essentially constructed), we see that these measures are of fundamental theoretical importance here; they show that it is natural to obtain estimates in the von Neumann ergodic theorem in terms of the decay rate of their Fourier coefficients, that is, correlation coefficients, which are already important from the viewpoint of practical applications.

The spectral measures and their Fourier coefficients not only completely determine the rates of convergence in the von Neumann ergodic theorem but also thereby permit one to obtain coarse estimates of the rate of convergence in the Birkhoff ergodic theorem. But that is all: using the norms of deviations of ergodic means from the limit, one can only make coarse assumptions about the distribution functions of these deviations; see the detailed discussion in [2]. (The converse estimates of $\|A_n f - f^*\|_2^2$ via P_n^ε are completely impossible for obvious geometric reasons.) It turns out that the rates of convergence in the Birkhoff ergodic theorem can be estimated much more efficiently via the decay rate of the large deviation probabilities for ergodic means, provided that one manages to compute these probabilities.

Needless to say, the problem of computing the decay rates, correlations, and probabilities of large deviations of ergodic means for real-world dynamical systems is hard. By estimating the rates of convergence in ergodic theorems via these characteristics, we just reduce the new problem we are interested in to an equivalently hard problem. However, the latter problem, with all of its difficulties, is well known and can be successfully solved for a number of specific systems. For example, the exponential decay of correlations is the classical, textbook chaos parameter in dynamical systems, and there is a well-developed technique for studying this parameter; e.g., see [5, Sec. 1].

The correlation coefficients (the rate of their convergence to zero), as well as the decay of large deviation probabilities, in specific dynamical systems of interest in applications have been intensively studied in recent years by various authors (e.g., see [6, 7, 8, 9, 10]), which has permitted obtaining the estimates given in Chapter 3 of the present paper for the rates of convergence in the von Neumann and Birkhoff ergodic theorems for all of these systems.

0.3. **Brief description of new results obtained after the publication of** [2]**.** The positive results obtained and discussed in [2] concerning the rates of convergence in ergodic theorems required substantial technical improvements before they could be applied to the study of specific dynamical systems. Indeed, they were only stated in terms of asymptotic relations using O-and-o notation without writing out any specific absolute constants; to obtain the corresponding inequalities, one had yet to understand on what and exactly how these constants depend. Further, these results were only obtained for the case of discrete time. Leaping ahead, note that there have been no difficulties in extending these results to the case of continuous time, even though things might go a different way: not all results about rates of convergence can be extended from discrete to continuous time; e.g., see what happened with Sinai's conjecture in [11, 12] and also cf. Remark 18.

The asymptotic result in [2] on the equivalence of a power-law rate of convergence in the von Neumann ergodic theorem and a power-law singularity at zero with the same exponent of the spectral measure of the function to be averaged was transferred from discrete to continuous time in [4]. The subsequent transition from this asymptotic result to the corresponding inequalities with specific absolute constants was made in [13, 14, 15] for the case of discrete as well as continuous time (see Theorems 5–8 in the present survey). Finally, the constants in the power-law estimates of the rate of convergence in the Birkhoff ergodic theorem via a given power-law rate of convergence in the von Neumann ergodic theorem were obtained in [14, 16, 17]; see Theorems 9 and 10 below (further generalized in [18] to a slightly wider range of rates of convergence).

It was already noted that the rates of convergence in the Birkhoff theorem can be estimated much more precisely via the large deviation probabilities rather than via the rates of convergence in the von Neumann theorem. The asymptotic result in [2] on the equivalence of a power-law decay of the probabilities of large deviations of ergodic means and the power-law rate of convergence with the same exponent in the Birkhoff theorem (for essentially bounded functions) was extended to a wider range of rates of convergence and simultaneously generalized to the case of continuous time in [19] (see Theorems 11 and 12 in our survey). The same paper made the transition from the asymptotic relations obtained there to the corresponding inequalities with specific constants for the rates of convergence popular in applications; see Theorem 13. The estimates given in the literature for the decay rate of large deviation probabilities are usually stated for the case of Hölder functions, and so it is likely that Theorem 14, which was announced in [20, 21] and is proved below, might be useful in possible applications, because it provides the transition from Hölder functions to characteristic functions of subsets of the phase space.

It was shown in [22, 23] how some results in [2] can be transferred to ergodic theorems for general transformation groups. Further, the paper [24] discussed the theoretical possibility of studying the convergence of ergodic means of orthogonal stochastic measures (e.g., see [25, Secs. VI.2–VI.3] for the definition) with the use of the corresponding stationary process, that is, of measures bearing full information on all ergodic means and hence practically incomputable and unusable for the study of any real-world dynamical systems. One should also note the curious recent papers of abstract-theoretical nature (e.g., see [26, 27, 28] and the bibliographies therein) dealing with rates of convergence in operator ergodic theorems; by the same pattern, there is no apparent possibility of applying these results to the study of specific dynamical systems yet.

Just as in [2], here we do not consider results on the central limit theorem for processes stationary in the narrow sense. The paper [29] is one of the first papers on the topic; the bibliography in the application-oriented paper [30], one of the most recent ones, offers a glimpse of the state of the art in the field. The same paper discusses the interesting conjecture that, for the central limit theorem to hold, it suffices to require that the so-called pair correlations decay rapidly (see the correlations c_τ in the introduction to Chapter 3 below).

0.4. **Applications.** The general theory of rates of convergence in ergodic theorems described in the first two chapters of the survey has been constructed having in mind its possible use in the study of specific dynamical systems that are of interest in applications. Billiards and Anosov systems, with the function to be averaged being the characteristic function of a measurable subset of the phase space, have been the main model examples from the very beginning. That is why the progress in this theory after the publication of the preceding survey [2] involves estimates via the decay rates of correlations (the estimates in the von Neumann theorem) and the large deviation probabilities (the estimates in the Birkhoff theorem) rather than, say, via spectral measures (which have nevertheless been studied thoroughly owing to their key importance in the general theory) for the von Neumann theorem and orthogonal stochastic measures for the Birkhoff theorem. This is because it is the decay rates of correlations and large deviation probabilities for specific dynamical systems of interest that have been intensively and successfully studied by various authors in the last twenty years.

This explains the multitude of applications presented in Chapter 3. In line with our interests, we pay special attention to billiards and Anosov systems; some of these new results were announced in [20, 21].

Chapter 1. Estimates of the rate of convergence in the von Neumann theorem

Throughout this chapter, we assume that $f \in L_2(\Omega)$. Note that all of the eight theorems in Chapter 1 hold not only for ergodic means (i.e., for the case of the law of large numbers for processes stationary in the narrow sense) but also for the case of the law of large numbers for processes stationary in the wide sense [25]: the spectral theory of stationary processes used here is the same for both cases, and so the proofs coincide word for word. For the same cause, all of these assertions remain valid for the case of ergodic theorems for contractions in Hilbert spaces; a relevant generalization of the spectral theory used here for the case of contraction operators can be found, say, in [31, Sec. 2.3]. Thus, all eight theorems in this chapter are essentially of operator nature; the characteristics of rates of convergence considered in these theorems are solely determined by the spectral properties of the corresponding (isometric) operators.

In this and the next chapter, we do not strive for maximum possible generality of the results; accordingly, we deal with the usual ergodic means for endomorphisms and semiflows and place obvious possible generalizations into comments.

1. Spectral measures and rates of convergence in the von Neumann theorem

The estimates given in this section can hardly be used when studying specific dynamical systems, because the spectral measures used in these estimates are practically impossible to compute. However, the results presented here (Theorems 1–4 below) convincingly show that the estimates of the rate of convergence in the von Neumann theorem are necessarily of spectral character, and hence these results explain why the estimates obtained in the next section via the decay rate of the Fourier coefficients of these measures (i.e., correlations) are natural.

1.1. General case.

Formulas (1.1) and (1.3) below permit us to exactly compute the norms of deviations of all ergodic means from their limit provided that the spectral measure is known exactly, which is virtually impossible for real-world systems. Hence it seems to be more natural to estimate these norms above and below via known behavior of the spectral measure at zero (see Theorems 1 and 2 in this section). Note that we impose no additional restrictions (such as absolute continuity, etc.) on the spectral measures; the most general case is considered.

1.1.1. *Discrete time.* Let $L_2^0(\Omega) \subset L_2(\Omega)$ be the subspace of functions with zero mean. Note that if $f \in L_2(\Omega)$, then $f - f^* \in L_2^0(\Omega)$ and hence $A_n(f - f^*) \in L_2^0(\Omega)$. Since the variance

$$\mathsf{D}A_n(f - f^*) = \|A_n(f - f^*)\|_2^2 = \|A_n f - f^*\|_2^2$$

just measures the L_2-norm of the deviation of $A_n f$ from the limit f^*, it follows that the behavior of the norm of this deviation coincides with the behavior, considered in [2], of the variance of ergodic means. Just as in [2], this behavior is completely determined by the singularity of the spectral measure σ_{f-f^*} at zero (see Theorem 3 below). The passage from σ_f to σ_{f-f^*} is done just by discarding the measure $\sigma_f\{0\} = \|f^*\|_2^2$ concentrated at zero. (This is because f^* is the orthogonal projection of f onto the subspace of eigenvectors of U_T corresponding to the eigenvalue 1; for the same reason, one has $\|f\|_2^2 = \|f - f^*\|_2^2 + \|f^*\|_2^2$.)

Set

$$\Phi_n(x) = \frac{1}{n}\frac{\sin^2(nx/2)}{\sin^2(x/2)}, \qquad 0 < |x| \le \pi.$$

Note that Φ_n only differs in the coefficient from the $(n-1)$-st Fejér kernel K_{n-1}, $\Phi_n = 2K_{n-1}$. The properties of Fejér kernels [32, Sec. I.47] imply that for all $n \in \mathbb{N}$ one has the relations

$$\int_{-\pi}^{\pi} \Phi_n(x)\,dx = 2\pi$$

and the standard inequalities

$$\frac{1}{n}\Phi_n(x) \le \frac{\pi^2}{n^2x^2} \quad (\text{for } \pi/n \le |x| \le \pi), \qquad \frac{1}{n}\Phi_n(x) \le 1 \quad (\text{for } |x| \le \pi).$$

It is well known that for each $g \in L_2^0(\Omega)$ the variance $\mathsf{D}A_ng = \|A_ng\|_2^2$ can be computed exactly via the integral of $\Phi_n(x)$ with respect to the spectral measure σ_g or expressed via the correlation coefficients b_kg. The main results of this chapter have been obtained by an analysis of the asymptotics of the following two formulas, which were essentially used as early as by von Neumann [33] (a derivation of these formulas can be found in [34, Theorem 18.2.1]):

$$\|A_ng\|_2^2 = \int_{(-\pi;\pi]} \frac{1}{n}\Phi_n(x)\,d\sigma_g(x), \tag{1.1}$$

$$\|A_ng\|_2^2 = \frac{1}{n^2}\sum_{|k|<n}(n-|k|)b_kg. \tag{1.2}$$

When thinking over the asymptotics of formula (1.1), it is useful to look at the graph of the kernel $\frac{1}{n}\Phi_n$. (The graph of the kernel K_n can be found, say, in [32, Sec. I.52].) The mass of that kernel is mainly concentrated in a small neighborhood of zero; the larger n, the smaller neighborhood can be taken. Hence it is no surprise that the asymptotics of $\|A_ng\|_2^2$ exactly correspond to the asymptotics of the singularity of the measure σ_g at zero. (See Theorem 3 in [2] and its refinement, Theorem 3 below.)

The following lemma, which was used in [14] in the proof of Theorem 1 below, is remarkable in that it contains a sharp constant, even though we do not pose the problem of determining whether constants are sharp in this paper.

Lemma 1 ([14]). *For any $a \in (0,\pi]$ and $n \in \mathbb{N}$, one has the inequality*

$$\left(\frac{\sin(a/2)}{a/2}\right)^2 \sigma_{f-f^*}\left(-\frac{a}{n}, \frac{a}{n}\right] \le \|A_nf - f^*\|_2^2,$$

which is sharp in the sense that the constant on the left-hand side cannot be increased in the general case.

Theorem 1 ([14]). *Set*

$$S_k = \sigma_{f-f^*}\left(-\frac{\pi}{k}, \frac{\pi}{k}\right], \quad \sigma_k = \sigma_{f-f^*}\left\{\left(-\frac{\pi}{k}, -\frac{\pi}{k+1}\right] \cup \left(\frac{\pi}{k+1}, \frac{\pi}{k}\right]\right\} = S_k - S_{k+1}.$$

Then

$$\frac{4}{\pi^2}S_n \le \|A_nf - f^*\|_2^2 \le S_n + \frac{1}{n^2}\sum_{k=1}^{n-1}(k+1)^2\sigma_k = \frac{1}{n^2}\left(S_1 + \sum_{k=1}^{n-1}(2k+1)S_k\right)$$

for all $n \in \mathbb{N}$.

Similar inequalities in different form and with different constants were used by Gaposhkin; e.g., see [35, Lemma 1].

1.1.2. *Continuous time.* Set

$$F_t(x) = \left(\frac{\sin(tx/2)}{tx/2}\right)^2.$$

Just as in the case of discrete time, the variance $\mathsf{D}\overline{A}_t g = \|\overline{A}_t g\|_2^2$ can be computed exactly for each $g \in L_2^0(\Omega)$ via the integral of the kernel $F_t(x)$ with respect to the spectral measure σ_g or expressed via the correlation function $b_t g$. The main results of this section have been obtained by an analysis of the asymptotics of the following formulas, which were used by von Neumann in [33] (a derivation of these formulas can also be found in [34, Theorem 18.3.1]):

$$\|\overline{A}_t g\|_2^2 = \int_{-\infty}^{+\infty} F_t(x)\, d\sigma_g(x), \tag{1.3}$$

$$\|\overline{A}_t g\|_2^2 = \frac{1}{t^2}\int_{-t}^{t} (t - |\tau|) b_\tau g\, d\tau. \tag{1.4}$$

Just as in the case of discrete time, the mass of the kernel $F_t(x)$ is mainly concentrated in a small neighborhood of zero; the larger t, the smaller neighborhood can be taken. This explains why the asymptotics of the rate of convergence is determined by the singularity of the spectral measure at zero in this case as well.

Theorem 2 ([15])**.** *Set*

$$S_k(t) = \sigma_{f-\overline{f}^*}\left(\frac{-2\pi k}{t}, \frac{2\pi k}{t}\right].$$

Then

$$\frac{4}{\pi^2} S_{1/2}(t) \le \|\overline{A}_t f - \overline{f}^*\|_2^2 \le \frac{\pi^2 - 1}{\pi^2} S_1(t) + \frac{1}{2\pi^2} S_2(t) + \frac{2}{\pi^2}\sum_{k=3}^{\infty} \frac{1}{k(k-1)(k-2)} S_k(t)$$

for all $t > 0$.

Similar inequalities in different form and with different constants were used by Gaposhkin long before [15]; e.g., see Remark 2 after Theorem 1 in [36].

1.2. **Case of power-law rate of convergence.** The proofs of both Theorems 3 and 4 in this subsection are obtained by specification of the estimates given in Theorems 1 and 2 of the preceding subsection for the power-law rates of convergence considered here.

1.2.1. *Discrete time.* The following theorem provides constants relating the (mutually equivalent) power-law convergence in the von Neumann ergodic theorem and the power-law singularity (with the same exponent) at zero of the spectral measure of the function to be averaged with respect to the corresponding dynamical system; i.e., this theorem refines the asymptotic Theorem 3 in [2].

Theorem 3 ([14])**.** *Let* $\alpha \in [0, 2)$*. The following assertions hold:*

1. If the spectral measure σ_{f-f^*} *has a power-law singularity at zero, i.e., if the inequality*

$$\sigma_{f-f^*}(-\delta, \delta] \le A\delta^\alpha$$

holds with some positive constant A for all $\delta \in (0, \pi]$, then the rate of convergence of the ergodic means $A_n f$ is power-law with the same exponent; i.e.,

$$\|A_n f - f^*\|_2^2 < A\pi^\alpha \left(\frac{2}{2-\alpha} n^{-\alpha} + \frac{1}{1-\alpha} n^{-1-\alpha}\right) \qquad \text{for } \alpha \in [0,1),$$

$$\|A_n f - f^*\|_2^2 < A\pi \left(2n^{-1} + \frac{\ln n}{n^2}\right) \qquad \text{for } \alpha = 1,$$

$$\|A_n f - f^*\|_2^2 < A\pi^\alpha \left(\frac{2}{2-\alpha} n^{-\alpha} + \left(4 - \frac{2}{2-\alpha} + \frac{1}{\alpha - 1}\right) n^{-2}\right) \qquad \text{for } \alpha \in (1,2)$$

for all $n \in \mathbb{N}$.

2. If the rate of convergence of the ergodic means $A_n f$ is power-law, i.e., if the inequality

$$\|A_n f - f^*\|_2^2 \le B n^{-\alpha}$$

holds with some positive constant B for all $n \in \mathbb{N}$, then the spectral measure σ_{f-f^*} has a power-law singularity at zero (with the same exponent); i.e.,

$$\sigma_{f-f^*}(-\delta, \delta] \le C\delta^\alpha, \qquad \text{where} \quad C = \begin{cases} \frac{\pi^{2-\alpha}}{4} B, & 0 \le \alpha < 1, \\ \frac{\pi^{2-\alpha}}{2^{3-\alpha}} B, & 1 \le \alpha < 2, \end{cases}$$

for each $\delta \in (0, \pi]$, where the constant C is sharp in the sense that it cannot be diminished.

To solve the problem considered in Theorem 3 on the equivalence of the power-law rate of convergence and the power-law singularity of the spectral measure at zero exhaustively, it remains to consider the case of $\alpha \ge 2$.

Remark 1. For $\alpha = 2$, no counterpart of Theorem 3 holds with any constants. According to [2, Remark 2 to Theorem 3], the condition $\sigma_{f-f^*}(-\delta, \delta] = O(\delta^2)$ as $\delta \to 0$ is not sufficient in general for the relation $\|A_n f - f^*\|_2^2 = O(n^{-2})$ to hold as $n \to \infty$; as was shown in [35, Corollary 2], this condition only guarantees that $\|A_n f - f^*\|_2^2 = O(n^{-2} \ln n)$. As to the rate of convergence with $\alpha > 2$, it simply does not exist for $f - f^* \not\equiv 0$; even the rate of convergence $\|A_n f - f^*\|_2^2 = o(n^{-2})$ as $n \to \infty$ is impossible (e.g., see [37, Corollary 5]).

Hence it is not surprising that the constant in the assertion of the first part of Theorem 3 tends to infinity as $\alpha \to 2$.

Remark 2 ([38]). The relation $\|A_n f - f^*\|_2^2 = O(n^{-2})$ as $n \to \infty$ is equivalent to the condition that the function $f - f^*$ is cohomological to zero, i.e., that $f - f^* = g \circ T - g$ for some $g \in L_2(\Omega)$.

It is well known (e.g., see [34, Theorem 18.2.1]) that if the measure σ_{f-f^*} is absolutely continuous with density ρ continuous at zero, then one has the asymptotic relation

$$\|A_n f - f^*\|_2^2 = \mathsf{D} A_n (f - f^*) = 2\pi\rho(0) n^{-1} + o(n^{-1}) \quad \text{as} \quad n \to \infty. \tag{1.5}$$

The following analog of this assertion in the form of an inequality refining the asymptotic Theorem 4 in [2] is an obvious corollary of Theorem 3.

Remark 3 ([14]). If the measure σ_{f-f^*} is absolutely continuous with density $\rho \in L_\infty(-\pi, \pi]$, then

$$\|A_n f - f^*\|_2^2 < 2\pi \|\rho\|_\infty \left(2n^{-1} + \frac{\ln n}{n^2}\right)$$

for every $n \in \mathbb{N}$.

Gaposhkin [35] proved that the range of rates of convergence for which an analog of the asymptotic Theorem 3 in [2] discussed here holds is much wider than the power-law range.

Remark 4 ([35]). Let $\alpha \in [0, 2)$, and let $\varphi(u)$ be a weakly oscillating function on $[1, \infty)$; i.e., the function $\varphi(u)u^{\delta}$ is monotone increasing and the function $\varphi(u)u^{-\delta}$ is monotone decreasing for each $\delta > 0$ [39, Sec. V.2]. Then the conditions

$$\|A_n f - f^*\|_2^2 = O(n^{-\alpha}\varphi(n)) \qquad \text{as} \quad n \to \infty,$$
$$\sigma_{f-f^*}(-\delta, \delta] = O(\delta^{\alpha}\varphi(\delta^{-1})) \qquad \text{as} \quad \delta \to 0$$

are equivalent.

This gives Theorem 3 in [2] if $\varphi(n) \equiv 1$ and its counterpart in [40] for the logarithmic rate of convergence if $\alpha = 0$ and $\varphi(n) = \ln^{\beta} n$, $\beta \geq 0$. The passage from asymptotic relations for rates of convergence in the range in question to specific inequalities with specific constants (as in Theorem 3 above) can be done by a simple specification of the estimates in Theorem 1 for each of these rates.

1.2.2. *Continuous time.* The following theorem, which is a full counterpart of Theorem 3 for the case of continuous time, provides constants relating the (mutually equivalent) power-law convergence in the von Neumann ergodic theorem and the power-law singularity (with the same exponent) at zero of the spectral measure of the function to be averaged with respect to the corresponding dynamical system.

Theorem 4 ([15]). *Let $\alpha \in [0, 2)$. The following assertions hold:*

1. If the spectral measure $\sigma_{f-\overline{f}^*}$ has a power-law singularity at zero, i.e., if the inequality

$$\sigma_{f-\overline{f}^*}(-\delta, \delta] \leq A\delta^{\alpha}$$

holds with some positive constant A for all $\delta > 0$, then the rate of convergence of the ergodic means $\overline{A}_t f$ is power-law with the same exponent; i.e.,

$$\|\overline{A}_t f - \overline{f}^*\|_2^2 \leq Bt^{-\alpha}$$

for all $t > 0$, where

$$B = A(2\pi)^{\alpha}\left(1 + \frac{2^{\alpha-2} - 1 + 2C}{\pi^2}\right), \qquad C = \begin{cases} \frac{3^{\alpha-1}}{2} + \frac{1}{2-\alpha}, & \alpha \in [0, 1), \\ 3^{\alpha-1} + \frac{\alpha-1}{2-\alpha}2^{\alpha-2}, & \alpha \in [1, 2). \end{cases}$$

2. If the rate of convergence of the ergodic means $\overline{A}_t f$ is power-law, i.e., if the inequality

$$\|\overline{A}_t f - \overline{f}^*\|_2^2 \leq Bt^{-\alpha}$$

holds with some positive constant B for all $t > 0$, then the spectral measure $\sigma_{f-\overline{f}^*}$ has a power-law singularity at zero (with the same exponent; i.e.,

$$\sigma_{f-\overline{f}^*}(-\delta, \delta] \leq A\delta^{\alpha}, \quad \textit{where} \quad A = \frac{\pi^{2-\alpha}}{4}B,$$

for all $\delta > 0$.

To solve the problem considered in Theorem 4 on the equivalence of the power-law rate of convergence and the power-law singularity of the spectral measure at zero exhaustively, it remains to consider the case of $\alpha \geq 2$.

Remark 5 ([15]). Just as in the case of discrete time, no counterpart of Theorem 4 holds with any constants for $\alpha = 2$. (The condition $\sigma_{f-\overline{f}^*}(-\delta, \delta] = O(\delta^2)$ as $\delta \to 0$ is not sufficient in general for the relation $\|\overline{A}_t f - \overline{f}^*\|_2^2 = O(t^{-2})$ to hold as $t \to \infty$.) As to the rate of convergence with $\alpha > 2$, it simply does not exist for $f - \overline{f}^* \not\equiv 0$; even the rate of convergence $\|\overline{A}_t f - \overline{f}^*\|_2^2 = o(t^{-2})$ is impossible.

Remark 6 ([15]). If the measure $\sigma_{f-\overline{f}^*}$ is absolutely continuous with density ρ continuous at the point 0, then one has the asymptotic relation

$$\|\overline{A}_t f - \overline{f}^*\|_2^2 = \mathsf{D}\overline{A}_t(f - \overline{f}^*) = 2\pi\rho(0)t^{-1} + o(t^{-1}) \qquad \text{as} \quad t \to \infty. \tag{1.6}$$

This coincides with the relation in Theorem 18.3.1 in [34] except for the constant multiplying $\rho(0)$; our constant in (1.6) is twice as large. Following [15], let us verify that there is indeed a misprint in [34] at that place. Since

$$\int_{-\infty}^{+\infty} \left(\frac{\sin(tx/2)}{tx/2}\right)^2 dx = \frac{2}{t}\int_{-\infty}^{+\infty} \left(\frac{\sin x}{x}\right)^2 dx = \frac{2\pi}{t},$$

it follows from the representation (1.1) for $g = f - \overline{f}^*$ that

$$\begin{aligned}\|\overline{A}_t f - \overline{f}^*\|_2^2 - \frac{2\pi}{t}\rho(0) &= \int_{-\infty}^{+\infty} \left(\frac{\sin(tx/2)}{tx/2}\right)^2 \rho(x)\,dx - \int_{-\infty}^{+\infty} \left(\frac{\sin(tx/2)}{tx/2}\right)^2 \rho(0)\,dx \\ &= \int_{-\infty}^{+\infty} \left(\frac{\sin(tx/2)}{tx/2}\right)^2 (\rho(x) - \rho(0))\,dx,\end{aligned}$$

and the subsequent proof exactly follows that of Theorem 18.2.1 in [34].

The following analog of this assertion in the form of an inequality refining the asymptotic Theorem 4 in [2] is an obvious corollary of Theorem 4.

Remark 7 ([15]). If the measure $\sigma_{f-\overline{f}^*}$ is absolutely continuous with density $\rho \in L_\infty(\mathbb{R})$, then

$$\|\overline{A}_t f - \overline{f}^*\|_2^2 \le 4\pi\left(1 + \frac{3}{2\pi^2}\right)\|\rho\|_\infty t^{-1}$$

for every $t > 0$.

2. Correlations and rates of convergence in the von Neumann theorem

It is the estimates obtained in this section (Theorems 5–8 below) that will be used in Chapter 3 when estimating the rate of convergence in the von Neumann theorem for specific dynamical systems. This proves possible because there are numerous examples in the literature where the decay rates of correlations are computed for various systems of interest in applications. These decay rates are usually power-law or exponential; that is why we pay special attention to these two cases in this section.

2.1. **Case of arbitrary decay rate of correlations.** Formulas (1.2) and (1.4) permit us to exactly compute the norms of deviations of all ergodic means from their limit, provided that one exactly knows all correlations, which is virtually impossible for real-world systems. Hence estimates of these norms via the known decay rate of correlations (Theorems 5 and 6 below) seem quite natural. Since the convergence to zero of the correlations of all $g \in L_2^0(\Omega)$ is known to be equivalent to the mixing property of the dynamical system in question (e.g., see [3, Sec. 1.7]), we, in a sense, obtain estimates of the rate of convergence in the von Neumann theorem for sufficiently rapidly mixing systems.

2.1.1. *Discrete time.* The following theorem refines Theorem 3 in [11] and Theorem 18.2.1 in [34] as well as Theorem 6 in [2].

Theorem 5 ([14]). *The following assertions hold:*

1. $\|A_n f - f^*\|_2^2 \le \frac{1}{n}\|f - f^*\|_2^2 + \frac{2}{n}\sum_{k=1}^{n-1}|b_k(f - f^*)|$.

2. If $\{b_k(f - f^*)\}_{k=0}^{\infty} \in l_p$ *for some* $p \in [1, +\infty]$, *then*

$$\|A_n f - f^*\|_2^2 \le 2\|\{b_k(f - f^*)\}\|_p n^{-1/p}.$$

3. If the series $\sum_{k=-\infty}^{+\infty} b_k(f - f^*)$ *converges absolutely, then the measure* σ_{f-f^*} *is absolutely continuous and has continuous (nonnegative density* ρ*; further,*

$$\rho(x) = \frac{1}{2\pi}\sum_{k=-\infty}^{+\infty} b_k(f - f^*)e^{ikx}$$

for all $x \in (-\pi, \pi]$, *and hence*

$$\sum_{k=-\infty}^{+\infty} b_k(f - f^*) = 2\pi\rho(0).$$

The asymptotic relation (1.5) *holds, and*

$$\|A_n f - f^*\|_2^2 - 2\pi\rho(0)n^{-1} = -\frac{1}{n^2}\sum_{|k|<n}|k|b_k(f - f^*) - \frac{1}{n}\sum_{|k|\ge n} b_k(f - f^*)$$

for each $n \in \mathbb{N}$.

4. If, in addition, the series $\sum_{k=-\infty}^{+\infty} kb_k(f - f^*)$ *converges absolutely, then the density* ρ *of the measure* σ_{f-f^*} *is continuously differentiable; its derivative satisfies*

$$\rho'(x) = \frac{1}{2\pi}\sum_{k=-\infty}^{+\infty} ikb_k(f - f^*)e^{ikx}$$

for all $x \in (-\pi, \pi]$, *and hence*

$$\sum_{k=-\infty}^{+\infty}|k|b_k(f - f^*) = 2\pi(\rho')^c(0),$$

where

$$(\rho')^c(x) = \frac{1}{2\pi}\sum_{k=-\infty}^{+\infty}|k|b_k(f - f^*)e^{ikx}$$

is the function trigonometrically conjugate to $\rho'(x)$ *(see* [32, Chap. VIII] *and* [41, vol. 2, Sec. 12.8]*). One has the asymptotic relation*

$$\|A_n f - f^*\|_2^2 = 2\pi\rho(0)n^{-1} - 2\pi(\rho')^c(0)n^{-2} + o(n^{-2}) \quad \text{as} \quad n \to \infty, \tag{2.1}$$

and

$$\|A_n f - f^*\|_2^2 - 2\pi\rho(0)n^{-1} + 2\pi(\rho')^c(0)n^{-2} = \frac{1}{n^2}\sum_{|k|\ge n}(|k| - n)b_k(f - f^*)$$

for every $n \in \mathbb{N}$.

Remark 8. Remark 3 after Theorem 3 in [11] also gives the integral representation

$$\sum_{k=-\infty}^{+\infty}|k|b_k(f - f^*) = -\int_{-\pi}^{\pi}\frac{\rho(x) + \rho(-x) - 2\rho(0)}{4\sin^2(x/2)}\,dx$$

(in the assumptions and notation of part 4 of our Theorem 5). In particular, it follows from this representation that the two coefficients $2\pi\rho(0)$ and $2\pi(\rho')^c(0)$ in the asymptotic relation in part 4 of Theorem 5 are zero simultaneously if and only if $\rho \equiv 0$, i.e., if $f - f^* \equiv 0$ almost everywhere. (Thus, the convergence of $\|A_n f - f^*\|_2^2$ at the rate $o(n^{-2})$ as $n \to \infty$ is impossible even under the assumptions of part 4 of Theorem 5; cf. Remark 1 above and Remark 1 after Theorem 3 in [2].)

The derivation of this integral representation was omitted in [11] for being "elementary but cumbersome". It was discovered in [14] that this is an integral representation of the number $2\pi(\rho')^c(0)$, which readily follows from the well-known properties of the trigonometrically conjugate function; e.g., see [39, Sec. IV.3].

Remark 9 (V. V. Ryzhikov, 2010, personal communication). Let $f \in L_2(\Omega)$ be a real-valued function. Then

$$b_0(f - f^*) = \|f - f^*\|_2^2 = \|A_1(f - f^*)\|_2^2,$$
$$b_1(f - f^*) = 2\|A_2(f - f^*)\|_2^2 - \|A_1(f - f^*)\|_2^2,$$

and

$$2b_n(f - f^*) = (n+1)^2\|A_{n+1}(f - f^*)\|_2^2 - 2n^2\|A_n(f - f^*)\|_2^2 + (n-1)^2\|A_{n-1}(f - f^*)\|_2^2$$

for all $n \geq 2$. These relations readily follow, say, from formula (1.2) (this version of proof was suggested by V. V. Sedalishchev): everything is obvious for $n < 2$, and for the other n one can use the following computations:

$$\begin{aligned}(n+1)^2\|A_{n+1}(f - f^*)\|_2^2 &= \sum_{|k|<n+1} (n + 1 - |k|)b_k(f - f^*) \\ &= b_n(f - f^*) + b_{-n}(f - f^*) + \sum_{|k|<n} (n + 1 - |k|)b_k(f - f^*) \\ &= b_n(f - f^*) + b_{-n}(f - f^*) + 2\sum_{|k|<n} (n - |k|)b_k(f - f^*) \\ &\quad - \sum_{|k|<n} (n - 1 - |k|)b_k(f - f^*) \\ &= b_n(f - f^*) + b_{-n}(f - f^*) + 2n^2\|A_n(f - f^*)\|_2^2 \\ &\quad - (n-1)^2\|A_{n-1}(f - f^*)\|_2^2.\end{aligned}$$

This remark is of fundamental importance here, because it shows that (for the case of a real-value function f) one cannot only compute the norms of deviations of ergodic means from the limit but also, conversely, reconstruct the correlations from the known norms of these deviations. This once more emphasizes that our approach to the computation of the rate of convergence in the von Neumann theorem is natural. The authors are grateful to V. V. Ryzhikov for permission to reproduce this remarkable unpublished result obtained by him.

We point out that the result is no longer true in the general case where the function f to be averaged is complex-valued; one can reconstruct the real parts of correlations but cannot always reconstruct their imaginary parts. For example, if g and h are two arbitrary real-valued functions in $L_2(\Omega)$, then elementary computations show that the respective norms of deviations of the ergodic means of the functions $f_1 = g + ih$ and $f_2 = h + ig$ from the limit coincide, while all respective correlation coefficients have coinciding real parts (which follows from the proof of Remark 9) and differ in the sign of their imaginary parts.

2.1.2. *Continuous time.* The following theorem refines Theorem 3 in [11] and Theorem 18.3.1 in [34] by extending the estimates in Theorem 5 to the case of continuous time.

Theorem 6 ([15]). *The following assertions hold:*

1. $\|\overline{A}_t f - \overline{f}^*\|_2^2 \le \frac{2}{t}\int_0^t |b_\tau(f-\overline{f}^*)|\,d\tau$ *for every* $t>0$.

2. If $b_\tau(f-\overline{f}^*) \in L_p([-t,t])$ *for some* $p\in[1,+\infty]$ *and* $t>0$, *then*

$$\|\overline{A}_t f - \overline{f}^*\|_2^2 \le 2\|b_\tau(f-\overline{f}^*)\|_p t^{-1/p}.$$

3. If $b_\tau(f-\overline{f}^*) \in L_1(\mathbb{R})$, *then the measure* $\sigma_{f-\overline{f}^*}$ *is absolutely continuous and has continuous (nonnegative) density* ρ*; further,*

$$\rho(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} b_\tau(f-\overline{f}^*)e^{-ix\tau}\,d\tau$$

for all $x\in\mathbb{R}$, *and hence*

$$\int_{-\infty}^{+\infty} b_\tau(f-\overline{f}^*)\,d\tau = 2\pi\rho(0).$$

The asymptotic relation (1.6) *holds, and*

$$\|\overline{A}_t f - \overline{f}^*\|_2^2 - 2\pi\rho(0)t^{-1} = -\frac{1}{t^2}\int_{|\tau|<t}|\tau| b_\tau(f-\overline{f}^*)\,d\tau - \frac{1}{t}\int_{|\tau|\ge t} b_\tau(f-\overline{f}^*)\,d\tau$$

for all $t>0$.

4. If, in addition, $\tau b_\tau(f-\overline{f}^*) \in L_1(\mathbb{R})$, *then the density* ρ *of the measure* $\sigma_{f-\overline{f}^*}$ *is continuously differentiable; further,*

$$\rho'(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \tau b_\tau(f-\overline{f}^*)e^{i\tau x}\,d\tau$$

for all $x\in\mathbb{R}$, *and hence*

$$\int_{-\infty}^{+\infty} |\tau| b_\tau(f-\overline{f}^*)\,d\tau = 2\pi(\rho')^c(0),$$

where

$$(\rho')^c(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |\tau| b_\tau(f-\overline{f}^*)e^{i\tau x}\,d\tau$$

is the function trigonometrically conjugate to $\rho'(x)$ *(see* [32, Chap. VIII] *and* [41, vol. 2, Sec. 12.8]*). One has the asymptotic relation*

$$\|\overline{A}_t f - \overline{f}^*\|_2^2 = 2\pi\rho(0)t^{-1} - 2\pi(\rho')^c(0)t^{-2} + o(t^{-2}) \quad \text{as} \quad t\to\infty, \tag{2.2}$$

and

$$\|\overline{A}_t f - \overline{f}^*\|_2^2 - 2\pi\rho(0)t^{-1} + 2\pi(\rho')^c(0)t^{-2} = \frac{1}{t^2}\int_{|\tau|\ge t}(|\tau|-t) b_\tau(f-\overline{f}^*)\,d\tau$$

for all $t>0$.

Remark 10. Remark 3 after Theorem 3 in [34] also gives the integral representation

$$\int_{-\infty}^{+\infty} |\tau| b_\tau(f-\overline{f}^*)\,d\tau = \int_{-\infty}^{+\infty} \frac{\rho(x)+\rho(-x)-2\rho(0)}{x^2}\,dx$$

(in the assumptions and notation of part 4 of our Theorem 6 above). In particular, it follows from this representation that the two coefficients $2\pi\rho(0)$ and $2\pi(\rho')^c(0)$ in the asymptotic relation in part 4 of Theorem 6 are zero simultaneously if and only if $\rho\equiv 0$, i.e., if $f-\overline{f}^*\equiv 0$ almost everywhere. (Thus, the convergence of $\|\overline{A}_t f - \overline{f}^*\|_2^2$ at the rate

$o(t^{-2})$ as $t \to \infty$ is impossible even under the assumptions of part 4 of Theorem 6; cf. Remark 5 above.)

Remark 11 (V. V. Ryzhikov; an analog of Remark 9 for the case of continuous time). Let $f \in L_2(\Omega)$ be a real-valued function; then the correlations $b_t(f - \overline{f}^*)$ can be reconstructed from the known norms $\|\overline{A}_t f - \overline{f}^*\|_2$ of deviations of the ergodic means from the limit,

$$2b_t g = \lim_{\varepsilon \to 0+} \frac{1}{\varepsilon^2}\big((t+\varepsilon)^2 \|\overline{A}_{t+\varepsilon} g\|_2^2 - 2t^2\|\overline{A}_t g\|_2^2 + (t-\varepsilon)^2\|\overline{A}_{t-\varepsilon} g\|_2^2\big),$$

where $g = f - \overline{f}^*$, for all $t > 0$. One can readily obtain this relation, say, by applying formula (1.3) and Lebesgue's dominated convergence theorem. (This version of the proof was suggested by Podvigin.) Namely,

$$\begin{aligned}
&\frac{1}{\varepsilon^2}\big((t+\varepsilon)^2 \|\overline{A}_{t+\varepsilon} g\|_2^2 - 2t^2\|\overline{A}_t g\|_2^2 + (t-\varepsilon)^2\|\overline{A}_{t-\varepsilon} g\|_2^2\big) \\
&\quad = \int_{-\infty}^{\infty} \frac{4}{\varepsilon^2 x^2}\left(\sin^2\frac{(t+\varepsilon)x}{2} - 2\sin^2\frac{tx}{2} + \sin^2\frac{(t-\varepsilon)x}{2}\right) d\sigma_g(x) \\
&\quad = \int_{-\infty}^{\infty} \frac{4}{\varepsilon^2 x^2}\left(2\sin^2\frac{tx}{2}\cos^2\frac{\varepsilon x}{2} - 2\sin^2\frac{tx}{2} + 2\sin^2\frac{\varepsilon x}{2}\cos^2\frac{tx}{2}\right) d\sigma_g(x) \\
&\quad = \int_{-\infty}^{\infty} \frac{4}{\varepsilon^2 x^2}\left(-2\sin^2\frac{tx}{2}\sin^2\frac{\varepsilon x}{2} + 2\sin^2\frac{\varepsilon x}{2}\cos^2\frac{tx}{2}\right) d\sigma_g(x) \\
&\quad = \int_{-\infty}^{\infty} 2\cos tx\left(\frac{\sin(\varepsilon x/2)}{\varepsilon x/2}\right)^2 d\sigma_g(x)
\end{aligned}$$

for an arbitrary $\varepsilon \in (0, t)$, and we obtain

$$\int_{-\infty}^{\infty} 2\cos tx\, d\sigma_g(x) = \int_{-\infty}^{\infty} (e^{itx} + e^{-itx})\, d\sigma_g(x) = b_t g + b_{-t} g$$

in the limit as $\varepsilon \to 0+$.

2.2. **Cases of power-law and exponential decay rates of the correlations.** The proofs of both Theorems 7 and 8 in this subsection are obtained by specification of the estimates given in Theorems 5 and 6 of the preceding subsection for the decay rates of correlations considered here and popular in applications.

2.2.1. *Discrete time.* Theorem 3 shows that the power-law rate of convergence in the von Neumann ergodic theorem is completely determined by the behavior of the spectral measure σ_{f-f^*} and hence of its Fourier coefficients, that is, correlation coefficients. The following theorem refines the asymptotic Theorem 6 in [2]. A discussion of why the inequalities in this theorem cannot be reversed, i.e., why there is no criterion here for the power-law convergence in terms of decay of the correlation coefficients, can be found in [2].

Theorem 7 ([14]). *Assume that the correlation coefficients tend to zero at a power-law rate; i.e., the inequality $|b_n(f - f^*)| \le Cn^{-\gamma}$ holds with some positive constant C for all $n \in \mathbb{N}$. Then the following assertions hold:*

1. If $0 \le \gamma < 1$, then

$$\|A_n f - f^*\|_2^2 \le \|f - f^*\|_2^2 n^{-1} + \frac{2C}{1-\gamma} n^{-\gamma}$$

for all $n \in \mathbb{N}$.

2. If $\gamma = 1$, then

$$\|A_n f - f^*\|_2^2 \le \|f - f^*\|_2^2 n^{-1} + 2C\frac{\ln n + 1}{n}$$

for all $n \in \mathbb{N}$. If $\gamma > 1$, then the measure σ_{f-f^*} is absolutely continuous and has continuous (nonnegative) density ρ; in this case,

$$\sum_{k=-\infty}^{+\infty} b_k(f-f^*) = 2\pi\rho(0) = \|f-f^*\|_2^2 + \sum_{|k|=1}^{\infty} b_k(f-f^*) \le \|f-f^*\|_2^2 + 2C\frac{\gamma}{\gamma-1}.$$

The asymptotic relation (1.5) *holds, and moreover,*

3. If $1 < \gamma < 2$, then

$$\big|\|A_n f - f^*\|_2^2 - 2\pi\rho(0)n^{-1}\big| < 2C\frac{1}{(\gamma-1)(2-\gamma)}\left(1-\frac{1}{n}\right)n^{-\gamma}$$

for all $n \ge 2$.

4. If $\gamma = 2$, then

$$\big|\|A_n f - f^*\|_2^2 - 2\pi\rho(0)n^{-1}\big| < 2C\left(\ln n + 2 + \frac{1}{n(n-1)}\right)n^{-2}$$

for all $n \ge 2$.

5. If $\gamma > 2$, then, in addition, the density ρ of the measure σ_{f-f^*} is continuously differentiable, and

$$2\pi(\rho')^c(0) = \sum_{k=-\infty}^{+\infty} |k| b_k(f-f^*), \qquad |2\pi(\rho')^c(0)| \le 2C\frac{\gamma-1}{\gamma-2}.$$

The asymptotic relation (2.1) *holds, and*

$$\big|\|A_n f - f^*\|_2^2 - 2\pi\rho(0)n^{-1} + 2\pi(\rho')^c(0)n^{-2}\big| < \frac{2C}{\gamma-2}\left(1-\frac{1}{n}\right)^2 n^{-\gamma}$$

for all $n \ge 2$.

6. If, in addition, the correlation coefficients decay exponentially, i.e., if $|b_n(f-f^*)| \le Ae^{-Bn}$ with some positive constants A and B for all $n \in \mathbb{N}$ (which is equivalent to the analyticity of the density ρ; e.g., see [32, Sec. I.25]*), then*

$$2\pi\rho(0) = \|f-f^*\|_2^2 + \sum_{|k|=1}^{\infty} b_k(f-f^*) \le \|f\|_2^2 + \frac{2A}{e^B-1},$$

$$2\pi(\rho')^c(0) = \sum_{|k|=1}^{\infty} |k| b_k(f-f^*), \quad \textit{that is,} \quad |2\pi(\rho')^c(0)| \le \frac{2A}{(e^B-1)(1-e^{-B})},$$

and

$$\big|\|A_n f - f^*\|_2^2 - 2\pi\rho(0)n^{-1} + 2\pi(\rho')^c(0)n^{-2}\big| \le 2A\frac{e^B}{e^B-1}\left(1+\frac{1}{e^B-1}\cdot\frac{1}{n}\right)\frac{e^{-Bn}}{n}$$

for all $n \in \mathbb{N}$.

2.2.2. *Continuous time.* Theorem 4 shows that the power-law rate of convergence in the von Neumann ergodic theorem is completely determined by the behavior of the spectral measure $\sigma_{f-\overline{f}^*}$ and hence of its correlation function. The following theorem is a full continuous-time counterpart of Theorem 7.

Theorem 8 ([15]). *Assume that the correlation function tends to zero at a power-law rate as $\tau \to \infty$; i.e., the inequality $|b_\tau(f - \overline{f}^*)| \le C\tau^{-\gamma}$ holds with some positive constant C for all $\tau > 0$. Then the following assertions hold:*

1. If $0 \le \gamma < 1$, then

$$\|\overline{A}_t f - \overline{f}^*\|_2^2 \le \frac{2C}{1-\gamma} t^{-\gamma}$$

for all $t > 0$.

2. If $\gamma = 1$, then

$$\|\overline{A}_t f - \overline{f}^*\|_2^2 \le 2(\|f - \overline{f}^*\|_2^2 + C \ln t) t^{-1}$$

for all $t > 1$. If $\gamma > 1$, then the measure $\sigma_{f-\overline{f}^*}$ is absolutely continuous and has continuous density ρ; in this case,

$$\rho(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} b_\tau(f - \overline{f}^*) e^{-i\tau x}\, d\tau$$

for all $x \in \mathbb{R}$ and hence

$$2\pi\rho(0) = \int_{-\infty}^{+\infty} b_\tau(f - \overline{f}^*)\, d\tau \le 2\|f - \overline{f}^*\|_2^2 + \frac{2C}{\gamma - 1}.$$

The asymptotic relation (1.6) *holds, and moreover,*

3. If $1 < \gamma < 2$, then

$$\big|\|\overline{A}_t f - \overline{f}^*\|_2^2 - 2\pi\rho(0)t^{-1}\big| \le \frac{2C}{(\gamma-1)(2-\gamma)} t^{-\gamma}$$

for all $t > 0$.

4. If $\gamma = 2$, then

$$\big|\|\overline{A}_t f - \overline{f}^*\|_2^2 - 2\pi\rho(0)t^{-1}\big| \le 2\big(\|f - \overline{f}^*\|_2^2 + C\ln t + C\big) t^{-2}$$

for all $t > 1$.

5. If $\gamma > 2$, then in addition, the density ρ of the measure $\sigma_{f-\overline{f}^*}$ is continuously differentiable, and the function

$$(\rho')^c(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\tau| b_\tau(f - \overline{f}^*) e^{i\tau x}\, d\tau$$

trigonometrically conjugate to its derivative is well defined (and continuous on $\mathbb{R}$; consequently,

$$2\pi(\rho')^c(0) = \int_{-\infty}^{+\infty} |\tau| b_\tau(f - \overline{f}^*)\, d\tau,$$

$$|2\pi(\rho')^c(0)| \le 2C \int_0^{\infty} \tau^{1-\gamma}\, d\tau \le 2\|f - \overline{f}^*\|_2^2 + \frac{2C}{\gamma - 2}.$$

One has the asymptotic relation (2.2), *and*

$$\big|\|\overline{A}_t f - \overline{f}^*\|_2^2 - 2\pi\rho(0)t^{-1} + 2\pi(\rho')^c(0)t^{-2}\big| \le \frac{2C}{\gamma-2} t^{-\gamma}$$

for all $t > 0$.

6. If, in addition, the correlation function decays exponentially, i.e., if $|b_\tau(f - \overline{f}^*)| \le Ae^{-B\tau}$ with some positive constants A and B for all $\tau > 0$ (in this case, the density ρ admits analytic continuation from the real line into a strip on the complex plane; e.g., see [42, Sec. 7.1, Exercise 3]*), then*

$$2\pi\rho(0) \le \frac{2A}{B}, \qquad |2\pi(\rho')^c(0)| \le \frac{2A}{B^2},$$

and

$$\big|\|\overline{A}_t f - \overline{f}^*\|_2^2 - 2\pi\rho(0)t^{-1} + 2\pi(\rho')^c(0)t^{-2}\big| \le \frac{2A}{B}\left(1 + \frac{1}{Bt}\right)\frac{e^{-Bt}}{t}$$

for all $t > 0$.

Chapter 2. Estimates of the rate of convergence in the Birkhoff theorem

The estimates given in Section 3 of the present chapter (Theorems 9 and 10 below) for the rate of convergence in the Birkhoff theorem are of little use in applications: although it suffices to know the decay rate of correlations to compute them, they usually fail to be sharp even asymptotically. The point is that, by using the norm of deviation of ergodic means from the limit, one can only make coarse assumptions about the distribution function of this deviation; see the discussion in the preceding survey [2, Sec. 2.0]. The estimates in Section 4 (Theorems 12 and 13) prove to be more efficient, and it is these estimates that will be applied in Chapter 3 when studying specific dynamical systems: although these estimates require knowing the decay rate of large deviation probabilities, they are much more accurate (and asymptotically sharp provided that so are the estimates used for the large deviations).

3. Estimates of power-law rates of convergence in the Birkhoff theorem via the rates of convergence in the von Neumann theorem

In this section, just as in the entire preceding chapter dealing with the von Neumann theorem, we assume that $f \in L_2(\Omega)$. Both Theorems 9 and 10 of this section, as well as all of the eight theorems in Chapter I, hold both in the case of the strong law of large numbers for processes stationary in the wide sense and in the case of ergodic theorems for contraction operators in Hilbert spaces.

Note also that the narrow range $\alpha \in (0, 2]$ of the exponent in Theorems 9 and 10 is quite reasonable, because, as was already noted in the preceding chapter (Remarks 1 and 5), the rate of convergence in the von Neumann theorem cannot be higher than quadratic except in the trivial case where $f - f^* = 0$ (or, accordingly, $f - \overline{f}^* = 0$) λ-almost everywhere.

3.1. **Discrete time.** The following Theorem 9 refines the asymptotic Theorem 11 in [2], Theorem 1 in [35], and also Theorems 1–3 in [43]. (The constants given here were extracted in [16] from Gaposhkin's proofs of the above-mentioned Theorems 1–3 in [43].) Recall that the constant B and the exponent α in the assumptions of Theorem 9 below can be computed from the behavior of the spectral measure σ_{f-f^*} in a neighborhood of zero or from the decay rate as $n \to \infty$ of the correlation coefficients $b_n(f - f^*)$ of the function $f - f^*$ (see the preceding chapter).

Theorem 9 ([16])**.** *Let* $\alpha \in (0, 2]$. *If the inequality* $\|A_n f - f^*\|_2^2 \le Bn^{-\alpha}$ *holds with some constant* $B > 0$ *for all* $n \in \mathbb{N}$, *then for each* $\varepsilon > 0$ *the following estimates hold for the quantities* $\mathrm{P}_n^\varepsilon = \lambda\Big\{\sup_{k\ge n} |A_k f - f^*| \ge \varepsilon\Big\}$ *for all* $n \ge 2$:

1. *If* $\alpha \in (0, 1)$, *then* $\mathrm{P}_n^\varepsilon < \dfrac{B}{\varepsilon^2}\dfrac{2^{1+\alpha}}{1 - 2^{-\alpha}}\left(1 + \dfrac{1}{(1 - 2^{(\alpha-1)/2})^2}\right)n^{-\alpha}$.
2. *If* $\alpha = 1$, *then* $\mathrm{P}_n^\varepsilon < \dfrac{8B}{\varepsilon^2}(\log_2^2 n + 2\log_2 n + 4)n^{-1}$.
3. *If* $\alpha \in (1, 2]$, *then* $\mathrm{P}_n^\varepsilon < \dfrac{8B}{\varepsilon^2}\left(1 + \dfrac{\pi^2}{6}\dfrac{1 + 2^{1-\alpha}}{(1 - 2^{1-\alpha})^3}\right)n^{-1}$.

Remark 12 ([16]). For $\alpha = 2$, the inequality

$$\lambda\Big\{\sup_{k\geq n}|A_k f - f^*| > \varepsilon\Big\} < \frac{B}{\varepsilon^2}n^{-1}$$

holds under the assumptions of Theorem 9.

3.2. **Continuous time.** Recall that, just as in the case of discrete time, the constant B and the exponent α in the assumptions of Theorem 10 below can be computed from the behavior of the spectral measure $\sigma_{f-\overline{f}^*}$ in a neighborhood of zero or from the decay rate as $t \to \infty$ of the correlation function $b_t(f - \overline{f}^*)$ (see the preceding chapter). The proof of this theorem in [17] was obtained by transferring Gaposhkin's construction in [43] to the case of continuous time.

Theorem 10 ([17]). *Let the inequality*

$$\|\overline{A}_t f - \overline{f}^*\|_2^2 \leq B(t+1)^{-\alpha}$$

hold for all $t > 0$, where B is a positive constant and $0 < \alpha \leq 2$. Then for each $\varepsilon > 0$ one has

$$\overline{\mathrm{P}}_t^{\varepsilon} = \lambda\Big\{\sup_{s\geq t}|\overline{A}_s f - \overline{f}^*| \geq \varepsilon\Big\} < \varphi(\alpha, t)B\varepsilon^{-2}$$

for all $t \geq 2$ and, depending on α,

1. *If $\alpha \in (0,1)$, then* $\varphi(\alpha, t) = \dfrac{3\cdot 2^{\alpha}}{1-2^{-\alpha}}\left(1 + \dfrac{1}{(1-2^{(\alpha-1)/2})^2}\right)t^{-\alpha} + 16t^{-2}.$
2. *If $\alpha = 1$, then* $\varphi(\alpha, t) = 12\dfrac{(1+\log_2 t)^2 + 3}{t} + 16t^{-2}.$
3. *If $\alpha \in (1,2]$, then* $\varphi(\alpha, t) = 12\left(1 + \dfrac{\pi^2}{6}\dfrac{1+2^{1-\alpha}}{(1-2^{1-\alpha})^3}\right)t^{-1} + 16t^{-2}.$

Remark 13. The results of both theorems in this section can be generalized to a much wider range of rates of convergence than the power-law range. This has been done in Sedalishchev's most recent paper [18] (together with all corresponding exact counterparts for processes stationary in the wide sense and for semigroups of contraction operators in $L_p(\Omega)$, $1 < p < \infty$).

4. Estimates of rates of convergence in the Birkhoff theorem via the decay rate of large deviations

This section, in contrast to all the preceding ones, only deals with functions f of the class $L_\infty(\Omega)$ to be averaged; however, this is sufficient for the numerous applications given in Chapter 3.

Theorems 11–13 given here have the specific feature that they hold not only for the time means of processes stationary in the narrow (and wide) sense or for the case of ergodic theorems for contraction operators but also in general for all essentially bounded stochastic processes with a.e. convergent time means (see Remark 14 below).

The approach considered here also has the important advantage that it applies to estimating the rates of convergence in the pointwise ergodic theorem for a fairly broad class of rates popular in applications, including not only power-law rates with an arbitrary exponent (in contrast to the preceding Section 3) but also all exponential rates.

4.1. **Coincidence of the asymptotics for the rate of convergence and for large deviations.** It was shown in [2, Theorem 12] for essentially bounded functions f that a power-law decay rate as $n \to \infty$ of the large deviation probabilities p_n^ε is equivalent to a power-law (with the same exponent) decay rate of the quantities P_n^ε. The following theorem establishes a similar asymptotic result for more general decay rates both for endomorphisms and for semiflows.

Theorem 11 ([19])**.** *Let $f \in L_\infty(\Omega)$, and let $\varphi(x)\colon \mathbb{R}^+ \to \mathbb{R}^+$ be a function monotone decreasing to zero for $x > x_0$ such that*

$$\int_x^\infty \frac{\varphi(t)}{t}\,dt = O(\varphi(x)) \quad as \quad x \to \infty. \tag{4.1}$$

Then the following two assertions are equivalent:

1. $\mathrm{p}_n^\varepsilon = O(\varphi(n))$ as $n \to \infty$ for every $\varepsilon > 0$.

2. $\mathrm{P}_n^\varepsilon = O(\varphi(n))$ as $n \to \infty$ for every $\varepsilon > 0$.

In the case of continuous time, the following two similar assertions are equivalent:

3. $\overline{\mathrm{p}}_t^\varepsilon = O(\varphi(t))$ as $t \to \infty$ for every $\varepsilon > 0$.

4. $\overline{\mathrm{P}}_t^\varepsilon = O(\varphi(t))$ as $t \to \infty$ for every $\varepsilon > 0$.

Let us introduce some notation. For each $f \in L_\infty(\Omega)$, set $\Delta = \|f - f^*\|_\infty$ in the case of a discrete dynamical system and $\Delta = \|f - \overline{f}^*\|_\infty$ in the case of a continuous dynamical system. Throughout the following, we assume that the function f is not invariant with respect to the dynamical system in question, i.e., that $\Delta \neq 0$; otherwise, there is nothing to estimate (see Remark 15 below). For each $\varepsilon > 0$, set

$$r = r(\varepsilon) = 1 + \frac{\varepsilon}{\Delta}.$$

In applications, one more often deals with a situation in which the decay rate $\varphi(n)$ of the large deviation probabilities p_n^ε is not the same for all positive ε but rather depends on ε. Hence one has to consider a one-parameter family $\{\varphi_\varepsilon(x)\}_{\varepsilon>0}$ of functions rather than a single function $\varphi(x)$. Let us refine Theorem 11 for this case in the direction of interest to us (the estimates of P_n^ε via p_n^ε).

Theorem 12 ([19])**.** *Let $f \in L_\infty(\Omega)$. Assume that, for a given $\varepsilon > 0$, $\varphi_\varepsilon(x)\colon \mathbb{R}^+ \to \mathbb{R}^+$ is a function monotone decreasing to zero for $x > x_0(\varepsilon)$ in such a way that relation* (4.1) *holds. If there exist constants $C(\varepsilon) > 0$ and $n_0(\varepsilon) \in \mathbb{Z}^+$ such that*

$$\mathrm{p}_n^\varepsilon \le C(\varepsilon)\varphi_\varepsilon(n)$$

for all $n > n_0(\varepsilon)$, then the inequality

$$\mathrm{P}_n^{2\varepsilon} \le C(\varepsilon)\left(1 + \frac{M(\varepsilon)}{\ln r(\varepsilon)}\right)\varphi_\varepsilon(n) \tag{4.2}$$

holds with some constant $M(\varepsilon) > 0$ for all $n > \max\{x_0(\varepsilon); n_0(\varepsilon)\}$.

In the case of continuous time, a similar assertion holds with n replaced by t and with p_n^ε and $\mathrm{P}_n^{2\varepsilon}$ replaced by $\overline{\mathrm{p}}_t^\varepsilon$ and $\overline{\mathrm{P}}_t^{2\varepsilon}$, respectively.

The following lemma shows how wide the class of functions satisfying the asymptotic relation (4.1) is.

Lemma 2 ([19])**.** *Let $\psi_1(x)\colon \mathbb{R}^+ \to \mathbb{R}^+$ be a function nonincreasing for $x > x_0$; then the function $\varphi_1(x) = \psi_1(x)/x^\alpha$ satisfies condition* (4.1) *for each $\alpha > 0$.*

If a function $\psi_2(x)\colon \mathbb{R}^+ \to \mathbb{R}^+$ satisfies condition (4.1)*, then so does the function $\varphi_2(x) = \psi_2(x^\delta)$ for each $\delta > 0$.*

It readily follows from this lemma that the asymptotic relation (4.1) is satisfied not only by power-law functions ($x^{-\alpha}$, $\alpha > 0$) but also by power-law functions with a logarithmic factor ($x^{-\alpha}\ln^{\beta} x$, $\alpha > 0$, $\beta \in \mathbb{R}$, $x > 1$), exponential decay functions ($e^{-\gamma x}$, $\gamma > 0$), and exponential decay functions with a power-law factor ($x^{\alpha}e^{-\gamma x}$, $\alpha \in \mathbb{R}$, $\gamma > 0$). Further, a power-law change of variables takes an exponential function to stretched exponential decay functions $e^{-\gamma x^{\delta}}$, $\gamma, \delta > 0$, which thus satisfy condition (4.1) as well.

Let us refine Theorem 12 (inequality (4.2)) for all functions described above. We introduce the incomplete gamma function $\Gamma(a,x)$ and its special case, the integral exponential $E_1(x)$,

$$\Gamma(a,x) = \int_x^{\infty} e^{-t}t^{a-1}\,dt, \quad x > 0, \quad a \in \mathbb{R}, \qquad E_1(x) = \Gamma(0,x) = \int_x^{\infty} \frac{e^{-t}}{t}\,dt, \quad x > 0.$$

Consider also the functions

$$A(a,x) = \frac{e^x}{x^a}\Gamma(a,x), \quad x > 0, \quad a \in \mathbb{R},$$

$$D(a,x) = \begin{cases} \frac{1}{a}, & x \le 0, \quad a > 0, \\ \frac{1}{a} + \frac{x}{a^2} + \frac{x(x-1)}{a^3} + \cdots + \frac{x(x-1)\cdots(x-[x])}{a^{[x]+2}}, & x > 0, \quad a > 0. \end{cases}$$

Theorem 13 ([19]). *Let $f \in L_\infty(\Omega)$; then the following assertions hold for each $\varepsilon > 0$:*

1. If $\mathrm{p}_n^{\varepsilon} \le C(\varepsilon)n^{-\alpha(\varepsilon)}$ with some constants $C(\varepsilon) > 0$ and $\alpha(\varepsilon) > 0$ for all $n > n_0(\varepsilon)$, $n_0(\varepsilon) \in \mathbb{Z}^+$, then

$$\mathrm{P}_n^{2\varepsilon} \le C(\varepsilon)\left(1 + \frac{1}{r(\varepsilon)^{\alpha(\varepsilon)} - 1}\right)n^{-\alpha(\varepsilon)} < C(\varepsilon)\left(1 + \frac{1}{\alpha(\varepsilon)\ln r(\varepsilon)}\right)n^{-\alpha(\varepsilon)}$$

for all $n > n_0(\varepsilon)$.

2. If $\mathrm{p}_n^{\varepsilon} \le C(\varepsilon)n^{-\alpha(\varepsilon)}\ln^{\beta(\varepsilon)} n$ with some constants $\alpha(\varepsilon) > 0$ and $\beta(\varepsilon) \in \mathbb{R}$ for all $n > n_0(\varepsilon)$, $n_0(\varepsilon) \in \mathbb{N}$, then

$$\mathrm{P}_n^{2\varepsilon} \le C(\varepsilon)\left(1 + \frac{D(\alpha(\varepsilon), \beta(\varepsilon))}{\ln r(\varepsilon)}\right)n^{-\alpha(\varepsilon)}\ln^{\beta(\varepsilon)}$$

for all $n > \max\{[e^{\beta(\varepsilon)/\alpha(\varepsilon)}]; n_0(\varepsilon); 2\}$.

3. If $\mathrm{p}_n^{\varepsilon} \le C(\varepsilon)e^{-\gamma(\varepsilon)n^{\delta(\varepsilon)}}$ with some constants $\gamma(\varepsilon) > 0$ and $\delta(\varepsilon) > 0$ for all $n > n_0(\varepsilon)$ and $n_0(\varepsilon) \in \mathbb{Z}^+$, then

$$\mathrm{P}_n^{2\varepsilon} \le C(\varepsilon)\frac{1}{\delta(\varepsilon)}\left(1 + \frac{e^{\gamma(\varepsilon)}E_1(\gamma(\varepsilon))}{\ln r(\varepsilon)}\right)e^{-\gamma(\varepsilon)n^{\delta(\varepsilon)}} < C(\varepsilon)\frac{1}{\delta(\varepsilon)}\left(1 + \frac{\ln(1 + 1/\gamma(\varepsilon))}{\ln r(\varepsilon)}\right)e^{-\gamma(\varepsilon)n^{\delta(\varepsilon)}}$$

for all $n > n_0(\varepsilon)$.

4. If $\mathrm{p}_n^{\varepsilon} \le C(\varepsilon)n^{\alpha(\varepsilon)}e^{-\gamma(\varepsilon)n}$ with some constants $\alpha(\varepsilon) \in \mathbb{R}$ and $\gamma(\varepsilon) > 0$ for all $n > n_0(\varepsilon)$ and $n_0(\varepsilon) \in \mathbb{Z}^+$, then

$$\mathrm{P}_n^{2\varepsilon} \le C(\varepsilon)\left(1 + \frac{A(\alpha(\varepsilon), \gamma(\varepsilon))}{\ln r(\varepsilon)}\right)n^{\alpha(\varepsilon)}e^{-\gamma(\varepsilon)n}$$

for all $n > \max\{[\alpha(\varepsilon)/\gamma(\varepsilon)]; n_0(\varepsilon)\}$. In the case of continuous time, similar assertions hold with n replaced by t and with $\mathrm{p}_n^{\varepsilon}$ and $\mathrm{P}_n^{2\varepsilon}$ replaced by $\overline{\mathrm{p}}_t^{\varepsilon}$ and $\overline{\mathrm{P}}_t^{2\varepsilon}$, respectively.

Remark 14 ([19]). All results of this section (Theorems 11–13) hold not only for ergodic means (i.e., time means of stationary processes) but also for arbitrary essentially bounded stochastic processes whose time means converge almost everywhere. Let us discuss this remark, which is important in what follows, in more detail.

Let $\{\xi_\tau \colon \tau \in \mathbb{N} \text{ or } \mathbb{R}^+\}$ be a sequence (or a family) of random variables on $(\Omega, \mathfrak{F}, \lambda)$ (i.e., a stochastic process) such that λ-almost everywhere there exists a limit

$$\xi^* = \lim_{n\to\infty} \frac{1}{n} \sum_{\tau=0}^{n-1} \xi_\tau \qquad \text{for the case of} \quad \tau \in \mathbb{N},$$

$$\overline{\xi}^* = \lim_{t\to\infty} \frac{1}{t} \int_0^t \xi_\tau \, d\tau \qquad \text{for the case of} \quad \tau \in \mathbb{R}^+,$$

of time means; let $\xi_\tau \in L_\infty(\Omega)$ for all τ, and finally, let $\sup_\tau \|\xi_\tau\|_{L_\infty(\Omega)} < \infty$. Then an analysis of the proofs of Theorems 11–13 shows that these theorems remain valid for these time means of the stochastic processes $\{\xi_\tau\}$ in question. In particular, they hold for the means of a bounded function if these means are generated by a Dunford–Schwartz operator in $L_1(\Omega)$ (by a semigroup of such operators in the continuous case), i.e., by a linear operator that is a contraction both in $L_1(\Omega)$, and in $L_\infty(\Omega)$.

Remark 15. Throughout this section, we have assumed that $\Delta \neq 0$, i.e., that the function f is not invariant with respect to the dynamical system. It is easily seen that if f is invariant, then the rate of convergence in the Birkhoff theorem is most fast: one has

$$\mathrm{p}_n^\varepsilon = \mathrm{P}_n^\varepsilon = 0 \tag{4.3}$$

for any $n \geq 1$ and $\varepsilon > 0$.

A similar situation occurs in the more general case where the function f is cohomological to some constant c. For a discrete dynamical system, this means that there exists a $g \in L_\infty(\Omega)$ such that λ-a.e. one has

$$f(\omega) = g(T\omega) - g(\omega) + c.$$

Indeed, then $f^* \equiv c$ and

$$|A_n f(\omega) - f^*(\omega)| = \frac{|g(T^n\omega) - g(\omega)|}{n} \leq \frac{2\|g\|_\infty}{n}.$$

Hence relation (4.3) holds for any $\varepsilon > 0$ and all positive integers $n > \frac{2\|g\|_\infty}{\varepsilon}$.

For the case of continuous time, the fact that f is cohomological to a constant means that there exists a function $g \in L_\infty(\Omega)$ such that λ–a.e. there exists a limit

$$g'(\omega) = \lim_{t\to 0} \frac{g(T^t\omega) - g(\omega)}{t}$$

and one has

$$f(\omega) = g'(\omega) + c.$$

Just as in the discrete case, one can readily show that a relation similar to (4.3) holds for each $\varepsilon > 0$ and all $t > \frac{2\|g\|_\infty}{\varepsilon}$.

4.2. **Large deviations for characteristic functions.** Many of the estimates given in the literature for the decay of large deviation probabilities have only been obtained for some classes of continuous functions; usually, these are functions satisfying the Hölder condition. It is natural to ask whether these estimates can be extended to other function classes, say, the class, important in applications, of characteristic functions χ_A of measurable subsets A of the phase space Ω. It turns out (see Theorem 14 below) that if the boundary of A is regular (say, piecewise smooth), then the functions χ_A admit monotone upper and lower approximations by smooth functions φ such that the large deviations for χ_A can be estimated via known large deviations for these φ. In Chapter 3, this will permit us to apply Theorems 12–13 to obtain estimates of the rates of convergence in the Birkhoff ergodic theorem for a broad variety of dynamical systems of interest in applications, the class of functions to be averaged including not only Hölder functions but also the characteristic functions of nonpathological subsets of the phase space.

Theorem 14. *Let Ω be a finite-dimensional C^r manifold, $r \geq 1$, let λ be a probability measure on its Borel σ-algebra, and let T be an endomorphism of the measure space (Ω, λ). Further, let $A \subset \Omega$ be a bounded Borel set of nonzero measure with boundary of measure zero, and let χ_A be the characteristic function of A. Then for each sufficiently small $\varepsilon > 0$ there exist C^r functions φ_1^ε and φ_2^ε such that $0 \leq \varphi_1^\varepsilon \leq \chi_A \leq \varphi_2^\varepsilon \leq 1$ and*

$$\lambda\{|A_n\chi_A - \lambda(A)| \geq 2\varepsilon\} \leq \lambda\left\{\left|A_n\varphi_1^\varepsilon - \int_\Omega \varphi_1^\varepsilon\, d\lambda\right| \geq \varepsilon\right\} + \lambda\left\{\left|A_n\varphi_2^\varepsilon - \int_\Omega \varphi_2^\varepsilon\, d\lambda\right| \geq \varepsilon\right\}$$

for all $n \geq 1$.

A similar assertion holds for the case of continuous time.

Proof. Consider the case of discrete time; for continuous time, the argument is similar. First, assume that the phase space Ω is a domain in the space $\mathbb{R}^m$, $m \in \mathbb{N}$.

It is well known (e.g., see [44, Theorem 1.4.8.]) that for any Borel set $A \subset \Omega$ and any $\varepsilon > 0$ there exists an open set U_ε and a compact set K_ε such that

$$K_\varepsilon \subset A \subset U_\varepsilon, \qquad \lambda(U_\varepsilon \backslash K_\varepsilon) < \varepsilon.$$

Since

$$\lambda(\operatorname{cl} A) - \lambda(\operatorname{int} A) = \lambda(\partial A) = 0 \quad \text{and} \quad \lambda(\operatorname{cl} A) = \lambda(A) = \lambda(\operatorname{int} A) > 0$$

under the assumptions of the theorem, we can take K_ε and U_ε such that

$$K_\varepsilon \subset \operatorname{int} A \subseteq A \subseteq \operatorname{cl} A \subset U_\varepsilon.$$

(To this end, we approximate the bounded open set $\operatorname{int} A$ by a compact set $K_\varepsilon \subset \operatorname{int} A$ such that $\lambda(\operatorname{int} A \backslash K_\varepsilon) < \varepsilon/2$ and the compact set $\operatorname{cl} A$ by an open set $U_\varepsilon \supset \operatorname{cl} A$ such that $\lambda(U_\varepsilon \backslash \operatorname{cl} A) < \varepsilon/2$.) For any compact set $K \subset \Omega$ and any open neighborhood U of K, there exists an infinitely differentiable function $\psi = \psi_{K,U} \colon \Omega \to \mathbb{R}$ such that (e.g., see [45, Part 2, Sec. 2.8])

1. $0 \leq \psi \leq 1$.
2. $K \subset \operatorname{int}\{\psi = 1\}$.
3. $\operatorname{supp} \psi \subset U$.

The desired functions φ_1^ε and φ_2^ε can be defined as follows:

$$\varphi_1^\varepsilon = \varphi_1 = \psi_{K_\varepsilon, \operatorname{int} A}, \qquad \varphi_2^\varepsilon = \varphi_2 = \psi_{\operatorname{cl} A, U_\varepsilon}.$$

Then the functions φ_1, φ_2, and χ_A satisfy the relations

$$\varphi_1 \leq \chi_A \leq \varphi_2, \qquad A_n\varphi_1 \leq A_n\chi_A \leq A_n\varphi_2, \tag{4.4}$$

for all $n \in \mathbb{N}$.

Further,

$$\int_\Omega \varphi_1\, d\lambda = \int_{\operatorname{supp}\varphi_1} \varphi_1\, d\lambda = \int_{\operatorname{int} A} \varphi_1\, d\lambda = \int_{K_\varepsilon} \varphi_1\, d\lambda + \int_{A\backslash K_\varepsilon} \varphi_1\, d\lambda$$
$$\geq \int_{K_\varepsilon} \varphi_1\, d\lambda = \lambda(K_\varepsilon) = \lambda(A) - \lambda(A\backslash K_\varepsilon) > \lambda(A) - \varepsilon.$$

In a similar way, we obtain a chain of equalities for φ_2,

$$\int_\Omega \varphi_2\,d\lambda = \int_{\operatorname{supp}\varphi_2} \varphi_2\,d\lambda = \int_{U_\varepsilon} \varphi_2\,d\lambda = \int_A \varphi_2\,d\lambda + \int_{U_\varepsilon\setminus A} \varphi_2\,d\lambda$$
$$= \lambda(A) + \int_{U_\varepsilon\setminus A} \varphi_2\,d\lambda \le \lambda(A) + \lambda(U_\varepsilon\setminus A) < \lambda(A) + \varepsilon.$$

In view of the estimates obtained above for the mean values of φ_1 and φ_2, we find that

$$\int \varphi_1\,d\lambda + \varepsilon \ge \lambda(A) \ge \int \varphi_2\,d\lambda - \varepsilon. \tag{4.5}$$

We use relations (4.4) and (4.5) and obtain

$$A_n\varphi_1 - \lambda(A) \le A_n\chi_A - \lambda(A) \le A_n\varphi_2 - \lambda(A),$$
$$A_n\varphi_1 - \int_\Omega \varphi_1\,d\lambda - \varepsilon \le A_n\chi_A - \lambda(A) \le A_n\varphi_2 - \int_\Omega \varphi_2\,d\lambda + \varepsilon,$$

for all $n \in \mathbb{N}$ and $\varepsilon > 0$, whence it follows that

$$|A_n\chi_A - \lambda(A)| \le \max\left\{\left|A_n\varphi_1 - \int_\Omega \varphi_1\,d\lambda\right|;\ \left|A_n\varphi_2 - \int_\Omega \varphi_2\,d\lambda\right|\right\} + \varepsilon.$$

Then

$$\{|A_n\chi_A - \lambda(A)| \ge 2\varepsilon\} \subseteq \left\{\left|A_n\varphi_1 - \int_\Omega \varphi_1\,d\lambda\right| \ge \varepsilon\right\} \cup \left\{\left|A_n\varphi_2 - \int_\Omega \varphi_2\,d\lambda\right| \ge \varepsilon\right\}; \tag{4.6}$$

here we apply the measure λ and find that the inequality

$$\lambda\{|A_n\chi_A - \lambda(A)| \ge 2\varepsilon\} \le \lambda\left\{\left|A_n\varphi_1 - \int_\Omega \varphi_1\,d\lambda\right| \ge \varepsilon\right\} + \lambda\left\{\left|A_n\varphi_2 - \int_\Omega \varphi_2\,d\lambda\right| \ge \varepsilon\right\}$$

holds for all $n \in \mathbb{N}$ and $\varepsilon > 0$.

One can reduce the proof of existence of C^r functions of the form $\psi_{K,U}$ for an arbitrary finite-dimensional C^r manifold Ω to the case of a domain in $\mathbb{R}^m$ considered here by using C^r coordinate neighborhoods (e.g., see [45, Part 2,Sec. 2.8]); the subsequent proof of the theorem in the general case does not differ in any way from the one given above. □

Remark 16. For ergodic dynamical systems, the inequality in Theorem 14 is equivalent to the inequality

$$\mathrm{p}_n^{2\varepsilon}(\chi_A) \le \mathrm{p}_n^{\varepsilon}(\varphi_1^\varepsilon) + \mathrm{p}_n^{\varepsilon}(\varphi_2^\varepsilon);$$

a similar assertion holds for the case of continuous time.

Remark 17. In the assumptions of Theorem 14, let m be another (in contrast to λ, not necessarily invariant) probability measure on the Borel σ-algebra of the manifold Ω. Then, by applying the measure m rather than λ to the inclusion (4.6), we additionally obtain the inequality

$$m\{|A_n\chi_A - \lambda(A)| \ge 2\varepsilon\} \le m\left\{\left|A_n\varphi_1^\varepsilon - \int_\Omega \varphi_1^\varepsilon\,d\lambda\right| \ge \varepsilon\right\} + m\left\{\left|A_n\varphi_2^\varepsilon - \int_\Omega \varphi_2^\varepsilon\,d\lambda\right| \ge \varepsilon\right\}$$

with the same φ_1^ε and φ_2^ε for all $n \in \mathbb{N}$ and $\varepsilon > 0$; a similar inequality holds for the case of continuous time.

In Section 6.2, the last remark will permit us to obtain estimates of the rate of convergence a.e. of the ergodic means $A_n\chi_A$ with respect to the (noninvariant) Riemannian volume on the manifold Ω for the case of hyperbolic dynamical systems.

Chapter 3. Applications: Estimates of rates of convergence for specific dynamical systems

In this chapter, we consider some well-known dynamical systems, including Bernoulli shifts and the Rényi map, the Gauss transformation, Anosov systems, nonuniformly hyperbolic systems modeled by a Young tower (the logistic map and the Hénon map with the Benedicks–Carleson parameters, one-dimensional expanding maps with a neutral point, and some Sinai and Bunimovich billiards), and also Teichmüller flows. For these dynamical systems, we obtain estimates of the rate of convergence in the Birkhoff theorem and/or von Neumann theorem. Our selection of dynamical systems is motivated by their popularity and accordingly the availability of estimates of large deviations and/or correlation coefficients for these systems in the literature.

In what follows, we use correlation coefficients different from the coefficients $b_n f$ and $b_t f$ considered in the first two chapters; sometimes, they are referred to as *pair* correlation coefficients. Let $f, g \in L_2(\Omega)$, and let $\tau \in \mathbb{Z}^+$ in the case of discrete time and $\tau \in \mathbb{R}^+$ in the case of continuous time. The number

$$c_\tau(f, g) = \int_\Omega f(\omega)\overline{g(T^\tau\omega)}\, d\lambda - \int_\Omega f(\omega)\, d\lambda \int_\Omega \overline{g(\omega)}\, d\lambda$$

is called the *correlation coefficient* of the functions f and $g \circ T^\tau$. For negative τ, one sets $c_\tau(f, g) = \overline{c_{-\tau}(f, g)}$; here the bar above the expression stands for complex conjugation. The number $c_\tau(f) = c_\tau(f, f)$ is called the *autocorrelation coefficient.*

The coefficient $c_\tau(f)$ is related to the correlation coefficient $b_\tau(f)$ (see the Introduction) by the simple formula

$$c_\tau(f) = c_\tau\left(f - \int_\Omega f(\omega)\, d\lambda\right) = \overline{b_\tau\left(f - \int_\Omega f(\omega)\, d\lambda\right)}, \tag{5.0}$$

which holds for all $f \in L_2(\Omega)$. Clearly, these coefficients have the same absolute value, $|c_\tau(f)| = |b_\tau(f)|$, for all $f \in L_2^0(\Omega)$ and coincide for real-valued $f \in L_2^0(\Omega)$.

By P we denote the adjoint of the Koopman operator U (see the Introduction), $P = U^*$. This is the so-called transfer operator, or the Perron–Frobenius operator. In the case of a reversible dynamical system, it coincides with the inverse of the Koopman operator (which is well known to be unitary in this case). If the action of this operator is known, then the autocorrelation coefficient for $f \in L_2^0(\Omega)$ can be found by the formula

$$c_\tau(f) = \int_\Omega P^\tau(f)\overline{f}\, d\lambda.$$

Note that the main role in the asymptotics of decay of these coefficients as $\tau \to \infty$ is played by the second-in-modulus eigenvalue of the operator P on a special Banach function space (e.g., see [46, Chap. 2]). (The first eigenvalue is always 1.)

Most of the examples given below deal with Hölder or Lipschitz real-valued functions to be averaged. Recall that if for some $\alpha \in (0, 1]$ there exists a constant $h \geq 0$ such that the inequality

$$|f(x) - f(y)| \leq h d^\alpha(x, y)$$

holds for any points x and y of a metric space (X, d), then the function f is said to be *Hölder with exponent* α if $\alpha \in (0, 1)$ and *Lipschitz* if $\alpha = 1$. The minimum constant h is denoted by $\text{Höld}_\alpha(f)$ in the first case and by $\text{Lip}(f)$ in the second case. Hölder (Lipschitz) functions ranging in an arbitrary metric space are defined in a similar way.

For some dynamical systems, we only use estimates of the correlation coefficients $c_\tau(f, g)$ for real-valued functions f and g in special Banach spaces. The passage to estimates for complex-valued functions $f = f_1 + if_2$ and $g = g_1 + ig_2$ whose real and imaginary parts lie in the same Banach space can be made with the use of the simple inequality

$$\begin{aligned}|c_\tau(f,g)| &= \big|c_\tau(f_1,g_1)+c_\tau(f_2,g_2)-ic_\tau(f_1,g_2)+ic_\tau(f_2,g_1)\big|\\ &= \big((c_\tau(f_1,g_1)+c_\tau(f_2,g_2))^2+(c_\tau(f_1,g_2)-c_\tau(f_2,g_1))^2\big)^{1/2}\\ &\le 2\sqrt{2}\max_{n,m=1,2}\{|c_\tau(f_n,g_m)|\}.\end{aligned}$$

Although we only give examples of dynamical systems with power-law and exponential estimates for the decay rate of correlations and large deviations in what follows (accordingly, the examples are divided into two groups given in Sections 5 and 6, respectively), we point out that the range of possible decay rates is by no means exhausted by these rates alone. Oppositely, it was shown in [47] that, for any ergodic aperiodic discrete-time dynamical system, the decay of large deviation probabilities at a rate less than an arbitrary given one is typical; i.e., it takes place for an everywhere dense G_δ-set of functions f in $L_p^0(\Omega)$, $1\le p<\infty$. (Consequently, the rate of convergence in the Birkhoff theorem for such functions f is arbitrarily small as well in view of the obvious inequality $\mathrm{P}_n^\varepsilon \ge \mathrm{p}_n^\varepsilon$ for all $n\in\mathbb{N}$.) Needless to say, correlations may decay arbitrarily slowly as well. It was shown, say, in [48] that, for any mixing discrete-time dynamical system, the decay rate of correlations can be less than any prescribed one; namely, for each sequence $\{a_n\}$ of numbers tending to zero and each nonzero function $g\in L_2^0(\Omega)$, there exists a function $f\in L_2(\Omega)$ such that $c_n(f,g)\ne O(a_n)$ as $n\to\infty$.

5. Dynamical systems with power-law decay of correlations and large deviations

In this section, we present estimates of rates of convergence in ergodic theorems for discrete dynamical systems modeled by a Young tower with a polynomial tail. (The correlations and large deviations have power-law decay in this case.) Of these systems, special attention is paid to the well-known Bunimovich stadium and one-dimensional expanding maps with a single neutral point.

However, first we consider an extremely simple example of a discrete-time dynamical system for which the spectral measure can be computed. This example was already presented as early as in [2].

5.1. Periodic automorphism. We have already established in Chapter 1 (Theorem 3) that, for each $f\in L_2(\Omega)$, the existence of a singularity of the spectral measure at zero, i.e., the validity of the condition

$$\sigma_{f-f^*}(-\delta,\delta]=O(\delta^\alpha)\quad\text{as}\quad\delta\to0$$

with some $\alpha\in(0,2)$, is equivalent to the power-law rate of convergence with the same exponent α in the von Neumann ergodic theorem,

$$\|A_nf-f^*\|_2^2=O(n^{-\alpha})\quad\text{as}\quad n\to\infty.$$

For a function f with such a spectral measure, we readily find from the Chebyshev inequality that

$$\mathrm{p}_n^\varepsilon\le\frac{\|A_nf-f^*\|_2^2}{\varepsilon^2}\le\frac{K}{\varepsilon^2}n^{-\alpha}\quad\text{for all}\quad n\in\mathbb{N}$$

with some constant K presented in Theorem 3 (or in Theorem 4 for the case of continuous time) for each $\varepsilon>0$. From this, by applying assertion 1 in Theorem 13, we obtain an estimate of the rate of convergence for $f\in L_\infty(\Omega)$,

$$\mathrm{P}_n^{2\varepsilon}\le\frac{K}{\varepsilon^2}\left(1+\frac{1}{(1+\varepsilon\|f-f^*\|_\infty^{-1})^\alpha-1}\right)n^{-\alpha}\quad\text{for all}\quad n\in\mathbb{N}$$

for any $\varepsilon > 0$. The following example, which has been borrowed from [2], shows that the exponent α in the asymptotics in the example above relating the singularity at zero of the spectral measure to the rate of convergence in Birkhoff theorem is sharp.

Let $\{\Omega_n\}_{n\in\mathbb{N}}$ be a partition of Ω into atoms Ω_n of measure

$$\lambda(\Omega_n) = \frac{1}{\zeta(\alpha+1)n^{\alpha+1}},$$

where

$$\zeta(\alpha+1) = \sum_{k=1}^{\infty} \frac{1}{k^{\alpha+1}}$$

is the usual zeta function. Let us split each of the atoms Ω_n into $2n$ atoms Ω_{nj} of equal measure. On each Ω_n, take an automorphism T_n of period $2n$ cyclically permuting the atoms Ω_{nj}. Set $\lambda_n = e^{i\pi/n}$. Consider the function f_{λ_n} on Ω_n equal to $2\varepsilon\lambda_n^j n \sin\frac{\pi}{2n}$ on each Ω_{nj}, where ε is a given positive number. Then, clearly, $T_n f_{\lambda_n} = \lambda_n f_{\lambda_n}$. We define the desired automorphism T and function $f \in L_\infty$ by the condition that their restrictions to the atom Ω_n coincide with T_n and f_{λ_n}, respectively, for each $n \in \mathbb{N}$.

A straightforward verification shows that $\|f\|_\infty \le \pi\varepsilon$,

$$\sigma_f(-\delta,\delta] = \sum_{-\delta<\arg\lambda_n\le\delta} \|f_{\lambda_n}\|_2^2 \le \sum_{n\ge\pi/\delta} \pi^2\varepsilon^2\lambda(\Omega_n) \le \frac{2^\alpha\pi^{2-\alpha}\varepsilon^2}{\alpha\zeta(\alpha+1)}\delta^\alpha$$

for each $\delta \in (0,\pi)$, and the rate of convergence has the lower bound

$$\mathrm{P}_n^{2\varepsilon} = \sum_{k=n}^{\infty} \lambda(\Omega_k) = \sum_{k=n}^{\infty} \frac{1}{\zeta(\alpha+1)n^{\alpha+1}} \ge \frac{1}{\alpha\zeta(\alpha+1)} n^{-\alpha}$$

for all positive integer n, which proves that the upper bound in question is asymptotically sharp.

Unfortunately, the sharp estimates discussed here of the rate of convergence via the singularity of the spectral measure σ_{f-f^*} at zero are practically of little use in the study of specific dynamical systems. As was already noted, this is due to the difficulty in the computation of these measures (for which one should first have full information about all autocorrelations $b_\tau(f-f^*)$ for all τ) and their possible asymptotic inaccuracy when estimating the rate of convergence in the Birkhoff theorem.

5.2. **Dynamical systems modeled by a Young tower with a polynomial tail.** The approach in this subsection to the derivation of power-law estimates of the decay rate of the correlation coefficients and large deviation probabilities p_n^ε as $n \to \infty$ is technically much more complicated than that adopted in the first two chapters. Nevertheless, so far there exist quite a few examples in which this approach has been applied efficiently. Its obvious advantage is that it provides highly accurate estimates and can be used when dealing with power-law estimates with arbitrarily high exponents.

This approach is based on the construction, well known in ergodic theory, of a Kakutani tower (or skyscraper) with additional properties studied by Young. Let us briefly describe this important model constructed in [7] and called a *Young tower with a polynomial tail.* The reader may well skip this description (up to and including inequality (5.1)) without any harm for understanding the estimates obtained further on the basis of this model.

Young tower. Let Δ_0 be an arbitrary set, and let $\xi = \{\Delta_{0,k}, k \in \mathbb{N}\}$ be a partition of Δ_0 into atoms. Let a function $R\colon \Delta_0 \to \mathbb{Z}^+$ (return time) be given such that $R|_{\Delta_{0,k}} = z_k$ and the greatest common denominator of all z_k is 1.

Define a dynamical system $(\Delta, \mathfrak{F}, \mu; S)$ as follows. The phase space is the set $\Delta = \{(\omega,k), \omega \in \Delta_0, 0 \le k < R(\omega)\}$, which is referred to as a tower over Δ_0; the set $\Delta_n =$

$\Delta \cap \{k = n\}$ is the nth floor of the tower, and $\Delta_{n,k} = \Delta_n \cap \{\omega \in \Delta_{0,k}\}$ is the nth floor of the tower over $\Delta_{0,k}$. All the sets $\Delta_{n,k}$ are $\mathfrak{F}$-measurable.

The map S takes each point (ω, k) to $(\omega, k+1)$ for all $k < R(\omega)$ and maps the points of the last floor $\Delta_{z_k-1,k}$ bijectively into $\Delta_0 \times \{0\}$ in such a way that the partition $\{\Delta_{n,k}\}$ is a generator for S. We identify the zero floor $\Delta_0 \times \{0\}$ with Δ_0 and define a map $S^R\colon \Delta_0 \to \Delta_0$ by the formula $S^R(\omega) = S^{R(\omega)}(\omega)$. The measure μ is a measure on $(\Delta, \mathfrak{F})$ such that $\mu(\Delta_0) < \infty$. For points $x, y \in \Omega$, we define a function (separation time) $s(x, y)$ as follows. First, for distinct points $x = (\omega_1, 0)$ and $y = (\omega_2, 0)$ in Δ_0 we define $s(x, y)$ to be the minimum nonnegative n such that the points $(S^R)^n(\omega_1)$ and $(S^R)^n(\omega_2)$ lie in distinct elements of the partition ξ and set $s(x, x) = \infty$. For all other points $x, y \in \Delta$, we set $s(x, y) = 0$ if they do not lie in the same $\Delta_{n,k}$; otherwise, $s(x, y) = s(x', y')$, where x' and y' are the corresponding points of the zero floor (the preimages under the map S). Fix some $\beta \in (0, 1)$ and set $d(x, y) = \beta^{s(x,y)}$; then (Δ, d) is a metric space.

There is an important property of S related to this metric structure. Assume that the map $S^R\colon \Delta_{0,k} \to \Delta_0$ has a nonsingular (with respect to μ) inverse for all k. Hence each $\Delta_{0,k}$ is equipped with the measure $S^R_*\mu$ absolutely continuous with respect to μ. (For a measurable $A \subset \Delta_0$, we set $S^R_*\mu(A) = \mu((S^R)^{-1}A)$.) Then we assume that the Radon–Nikodym derivative

$$\rho = \frac{dS^R_*\mu}{d\mu}$$

of this measure has the following property: there exists a constant $C_S > 0$ such that

$$\left|\frac{\rho(x)}{\rho(y)} - 1\right| \leq C_S d(S^R x, S^R y)$$

for all k and $x, y \in \Delta_{0,k}$.

Estimates of correlations. If the system written out above satisfies the condition

$$\int_{\Delta_0} R\, d\mu < \infty,$$

then there exists an S-invariant probability measure ν absolutely continuous with respect to μ such that $(\Delta, \mathfrak{F}, \nu; S)$ is a mixing dynamical system.

If, moreover, the tail distribution of the time $\overline{R}\colon \Delta \to \mathbb{Z}^+$ of return to the zero floor has the property

$$\mu(\overline{R} > n) \leq C_R n^{-\gamma} \quad \text{for some} \quad \gamma > 0 \quad \text{and} \quad C_R > 0 \quad \text{for all} \quad n \in \mathbb{N},$$

then the inequality

$$|c_n(f, g)| = \left|\int_\Delta f(\omega) g(S^n\omega)\, d\nu - \int_\Delta f(\omega)\, d\nu \int_\Delta g(\omega)\, d\nu\right| \leq C_f \|g\|_\infty n^{-\gamma}$$

holds for any real-valued functions $g \in L_\infty$ and f satisfying the Lipschitz condition and for all $n \in \mathbb{N}$, where the constant C_f depends on the dynamical system and the function f. When modeling by a Young tower, it suffices to require (as one usually does) that the observable function f on the phase space Ω (which is usually a Riemannian manifold) of the dynamical system T to be studied satisfies the Hölder condition. Then, when passing from Ω to the corresponding tower Δ (i.e., when dealing with a map $\pi\colon \Delta \to \Omega$ for which $\pi \circ S = T \circ \pi$), the function $f \circ \pi$ will be Lipschitz with respect to the metric d for an appropriate choice of β. For the base $\Delta_0 \subseteq \Omega$ one takes a uniformly hyperbolic set of system T, and the measure μ is defined to be the Lebesgue measure m^u on the unstable manifolds in Δ_0; then the measure $\lambda = \pi_*\nu$ proves to be an SRB (Sinai–Ruelle–Bowen) measure, i.e., a measure that has absolutely continuous conditional measure on the unstable manifolds and at least one positive Lyapunov exponent. (For more details on SRB measures, e.g., see Young's paper [49] dedicated to them.)

Then the inequalities on the tail distribution of the time of return to the zero floor and the estimates of correlation coefficients rewritten in terms of the dynamical system T to be studied become

$$m^u(\overline{R} > n) \le C_R n^{-\gamma} \quad \text{for some} \quad \gamma > 0 \quad \text{and} \quad C_R > 0 \quad \text{for all} \quad n \in \mathbb{N}, \tag{5.1}$$

and the inequality

$$|c_n(f,g)| = \left| \int_\Omega f(\omega) g(T^n\omega)\, d\lambda - \int_\Omega f(\omega)\, d\lambda \int_\Omega g(\omega)\, d\lambda \right| \le C_f \|g\|_\infty n^{-\gamma} \tag{5.2}$$

holds for any real-valued functions $g \in L_\infty(\Omega)$ and f satisfying the Hölder condition and for all $n \in \mathbb{N}$ with some constant C_f depending on the dynamical system and the function f.

Estimates of large deviations. The decay of large deviations for systems satisfying the estimate (5.2) was recently studied in [9, 10]. It was shown in these papers that if for some function $f \in L_\infty(\Omega)$ and $\gamma > 0$ there exists a constant $C_f > 0$ depending on f such that

$$|c_n(f,g)| \le C_f \|g\|_\infty n^{-\gamma} \quad \text{for all} \quad g \in L_\infty(\Omega), \quad n \in \mathbb{N},$$

then for each $\varepsilon > 0$ there exists a constant $C_{f,\varepsilon,\gamma} > 0$ depending on f, ε, and γ such that

$$\mathrm{p}_n^\varepsilon \le C_{f,\varepsilon,\gamma} n^{-\gamma} \tag{5.3}$$

for all $n \in \mathbb{N}$. An analysis of the proof of this estimate in [9, 10] shows that the constant $C_{f,\varepsilon,\gamma}$ there has the form

$$C_{f,\varepsilon,\gamma} = C_f \min_{\substack{q \ge 1 \\ q > \gamma}} \left\{ \frac{\|f - f^*\|_\infty^{2q-1}}{\varepsilon^{2q}} \left(\frac{36q^2 - 4q\gamma}{q - \gamma} \right)^q \right\}. \tag{5.4}$$

(We note parenthetically that if $\gamma \in (0,1)$, then the same asymptotics of decay of p_n^ε in the more general case of $f \in L_2(\Omega)$ under the weaker condition

$$|b_n(f - f^*)| \le C n^{-\gamma} \quad \text{for all} \quad n \in \mathbb{N}$$

can readily be obtained by the methods described in Chapters 1 and 2, namely, by combining the results of Theorems 7 and 9. The asymptotics given by those methods for $\gamma = 1$ is worse, and the values $\gamma > 1$ do not give any further refinement of the asymptotics at all.)

Estimates of the rate of convergence in ergodic theorems. By applying Theorem 13 to (5.3), we obtain an estimate of the rate of convergence in the Birkhoff theorem for dynamical systems modeled by a Young tower with a polynomial tail (with exponent γ) for any Hölder function f and any $\varepsilon > 0$,

$$\mathrm{P}_n^{2\varepsilon} \le C_{f,\varepsilon,\gamma} \left(1 + \frac{1}{r(\varepsilon)^\gamma - 1} \right) n^{-\gamma} \tag{5.5}$$

for all $n \in \mathbb{N}$.

By applying the large deviation estimate (5.3) for smooth functions φ_1 and φ_2 and Remark 16 to Theorem 14, we obtain an estimate of large deviations for the characteristic function χ_E of a bounded Borel set E with regular boundary: for all sufficiently small $\varepsilon > 0$ and all $n \in \mathbb{N}$, one has the inequality

$$\mathrm{p}_n^{2\varepsilon}(\chi_E) \le \mathrm{p}_n^\varepsilon(\varphi_1) + \mathrm{p}_n^\varepsilon(\varphi_2) \le C(\varepsilon) n^{-\gamma},$$

where

$$C(\varepsilon) = C_{\varphi_1,\varepsilon,\gamma} + C_{\varphi_2,\varepsilon,\gamma}.$$

We apply Theorem 13 to the resulting inequality and recall that the constant for the characteristic function χ_E has the form $r(\varepsilon) = 1 + \varepsilon\|\chi_E - \lambda(E)\|_\infty^{-1}$, thus obtaining the following estimate of the rate of convergence in the Birkhoff theorem for such functions:

$$\mathrm{P}_n^{4\varepsilon}(\chi_E) \le C(\varepsilon)\left(\frac{(\|\chi_E - \lambda(E)\|_\infty + 2\varepsilon)^\gamma}{(\|\chi_E - \lambda(E)\|_\infty + 2\varepsilon)^\gamma - \|\chi_E - \lambda(E)\|_\infty^\gamma}\right)n^{-\gamma} \tag{5.6}$$

for all sufficiently small $\varepsilon > 0$ and all $n \in \mathbb{N}$.

Theorem 7 applied to (5.2) with Hölder functions $f = g$ readily gives estimates of the rate of convergence in the von Neumann theorem for all $\gamma > 0$; the exact form of these estimates is given below in specific examples.

Let us proceed to specific dynamical systems modeled by the construction described above.

5.2.1. *Expanding maps with a neutral point.* Consider one most popular one-dimensional expanding map with a neutral point (see [7, 50, 51, 52, 53] and the references therein), namely, the map $T_\alpha\colon [0,1] \to [0,1]$, $\alpha \in (0,1)$, acting by the rule

$$T_\alpha(x) = \begin{cases} x(1 + 2^\alpha x^\alpha), & x \in [0, 1/2), \\ 2x - 1, & x \in [1/2, 1]. \end{cases}$$

The neutral point is $x = 0$, for it is only at this point that the derivative is equal to unity. An invariant measure is an absolutely continuous measure λ_α with density ϱ_α that is Lipschitz outside any neighborhood of zero.

Estimates of the rate of convergence for Hölder functions. By using the Young tower with a polynomial tail constructed in [7] for such systems, we obtain the power-law estimates (5.2) for correlations with exponent $\gamma = 1/\alpha - 1$ for any Hölder function f and bounded function g. This implies the estimates (5.5) with $\gamma = 1/\alpha - 1$ of the rate of convergence in the Birkhoff theorem for any Hölder function f and the estimates (5.6) of the rate of convergence in the Birkhoff theorem for the characteristic function χ_E of a bounded Borel set E. By applying the results of Theorem 7 with various $\alpha > 0$, we obtain estimates of the rate of convergence in the von Neumann theorem for any Hölder function f similar to the estimates presented below in detail for Lipschitz functions with nonzero mean.

Estimates of the rate of convergence for Lipschitz functions supported outside a neighborhood of zero. Surprisingly, the exponent γ, which is normally $1/\alpha - 1$, becomes $1/\alpha$ if the function f to be averaged vanishes in a neighborhood of zero and has zero mean with respect to the measure λ_α [54, 55].

Let f be a Lipschitz function and g a bounded function whose supports do not contain zero; i.e., these functions vanish in some neighborhood of zero. If

$$\int_0^1 f\,d\lambda_\alpha \int_0^1 g\,d\lambda_\alpha \ne 0,$$

then there exists a constant $C_1 > 0$ such that

$$|c_n(f,g)| \le \frac{C_2}{n^{1/\alpha - 1}}, \qquad C_2 = C_1 \varrho_\alpha\left(\frac{1}{2}\right)\frac{\alpha^{1-1/\alpha}}{1-\alpha}\left|\int_0^1 f\,d\lambda_\alpha \int_0^1 g\,d\lambda_\alpha\right|$$

for all $n \in \mathbb{N}$.

If

$$\int_0^1 f\,d\lambda_\alpha = 0,$$

then there exists a constant $C_3 > 0$ such that

$$|c_n(f,g)| \le \frac{C_3}{n^{1/\alpha}}$$

for all $n \in \mathbb{N}$.

We use these estimates of correlations, relations (5.0) between the correlation coefficient, and the result in Theorem 7 for various $\alpha > 0$ to obtain the following estimates for the rate of convergence in the von Neumann theorem for Lipschitz functions f vanishing in a neighborhood of zero:

1. If $\int_0^1 f\,d\lambda_\alpha = 0$ for all $n \ge 2$, then

$$\Bigl|\|A_nf - f^*\|_2^2 - \frac{2\pi\rho(0)}{n}\Bigr| < \frac{C_3\alpha^2}{(1-\alpha)(2\alpha-1)}\frac{1}{n^{1/\alpha}}, \qquad \alpha \in (1/2,1),$$

$$\Bigl|\|A_nf - f^*\|_2^2 - \frac{2\pi\rho(0)}{n}\Bigr| < 2C_3\biggl(\ln n + 2 + \frac{1}{n(n-1)}\biggr)\frac{1}{n^2}, \qquad \alpha = 1/2,$$

$$\Bigl|\|A_nf - f^*\|_2^2 - \frac{2\pi\rho(0)}{n} + \frac{2\pi(\rho')^c(0)}{n^2}\Bigr| < \frac{C_3\alpha}{2-4\alpha}\frac{1}{n^{1/\alpha}}, \qquad \alpha \in (0,1/2).$$

For the corresponding values of α, one has

$$0 \le 2\pi\rho(0) \le \|f - f^*\|_2^2 + \frac{2C_3}{1-\alpha}, \qquad |2\pi(\rho')^c(0)| \le 2C_3\frac{1-\alpha}{1-2\alpha}.$$

2. If $\int_0^1 f\,d\lambda_\alpha \ne 0$, then

$$\|A_nf - f^*\|_2^2 \le \frac{\|f - f^*\|_2^2}{n} + \frac{2C_2\alpha}{2\alpha-1}\frac{1}{n^{1/\alpha-1}}, \qquad \alpha \in (1/2,1), \quad n \in \mathbb{N},$$

$$\|A_nf - f^*\|_2^2 \le \frac{\|f - f^*\|_2^2}{n} + 2C_2\frac{\ln n + 1}{n}, \qquad \alpha = 1/2, \quad n \in \mathbb{N},$$

$$\Bigl|\|A_nf - f^*\|_2^2 - \frac{2\pi\rho(0)}{n}\Bigr| < \frac{C_2\alpha^2}{(1-2\alpha)(3\alpha-1)}\frac{1}{n^{1/\alpha-1}}, \qquad \alpha \in (1/3,1/2), \quad n \ge 2,$$

$$\Bigl|\|A_nf - f^*\|_2^2 - \frac{2\pi\rho(0)}{n}\Bigr| < 2C_2\biggl(\ln n + 2 + \frac{1}{n(n-1)}\biggr)\frac{1}{n^2}, \quad \alpha = 1/3, \quad n \ge 2,$$

$$\Bigl|\|A_nf - f^*\|_2^2 - \frac{2\pi\rho(0)}{n} + \frac{2\pi(\rho')^c(0)}{n^2}\Bigr| < \frac{C_2\alpha}{2-6\alpha}\frac{1}{n^{1/\alpha-1}}, \quad \alpha \in (0,1/3), \quad n \ge 2.$$

For the corresponding values of α, one has

$$0 \le 2\pi\rho(0) \le \|f - f^*\|_2^2 + 2C_2\frac{1-\alpha}{1-2\alpha}, \qquad |2\pi(\rho')^c(0)| \le 2C_2\frac{1-2\alpha}{1-3\alpha}.$$

Inequality (5.5) written out for Lipschitz functions f vanishing in a neighborhood of zero with exponent $\gamma = 1/\alpha$ if $\int_0^1 f\,d\lambda_\alpha = 0$ and with exponent $\gamma = 1/\alpha - 1$ if $\int_0^1 f\,d\lambda_\alpha \ne 0$ is common for all examples in this subsection and provides an estimate of the rate of convergence in the Birkhoff theorem.

5.2.2. *Billiards with a power-law estimate of the rate of convergence* [56, 57, 58, 59]. Billiards are dynamical systems modeling the inertial motion (with elastic reflections) of a mass point in some domain $\mathcal{M} \subset \mathbb{R}^m$, $m \ge 2$, with piecewise smooth boundary $\partial\mathcal{M}$. (In a more general setting, $\mathcal{M}$ is a Riemannian manifold, and the motion before and after each reflection is along geodesics.) The transformation group T^t is formed by the maps taking each point $x = (q,v)$, where $q \in \mathcal{M}$ and $v \in T_q\mathcal{M}$ is a unit tangent vector, to the point T^tx lying at the distance t from x along the trajectory of motion in the direction of v (with regard to reflections at the boundary). In what follows, we only deal with plane (two-dimensional) billiards. Along with this continuous system, one often considers the discrete dynamical system $(\Omega, \mathrm{T}, \lambda)$, where $\Omega = \partial\mathcal{M} \times [-\pi/2, \pi/2]$ is the collision space and $\mathrm{T}\colon \Omega \to \Omega$ is the billiard map (the Poincaré map of the billiard boundary) taking each point r on the boundary $\partial\mathcal{M}$ with a direction φ to the point r' where a collision with reflection in the direction φ' will occur; the T-invariant measure on Ω has the form $d\lambda = c_\lambda \cos\varphi\, dr\, d\varphi$, where (r, φ) are the canonical coordinates on Ω and c_λ is a normalizing factor.

The statistical properties of such systems depend on the character of the boundary. One singles out focusing billiards (where the boundary $\partial\mathcal{M}$ is convex to the outside of the domain), known as Bunimovich billiards; scattering billiards (where the boundary $\partial\mathcal{M}$ is convex to the inside of the domain), known as Sinai billiards; scattering billiards with traps (where some parts of the boundary meet at a zero angle); and semi-scattering billiards (see [3, Chap. 6] and [60] for details). Note that billiard flows and maps for convex plane domains with sufficiently smooth (C^7) boundary are not ergodic, and their statistical properties are trivial. On the other hand, Stepin and Troubetzkoy [61] showed that billiard systems for C^1-typical plane domains with C^1 boundary are weakly mixing.

Estimates of the rate of convergence in the Birkhoff theorem. It was shown in [59] for the Bunimovich stadium and for some classes of semi-scattering billiards and billiards with traps that the correlations with respect to their billiard maps exhibit power-law decay. Namely, the inequality

$$|c_n(f,g)| = \left| \int_\Omega f(r,\varphi) g(\mathrm{T}^n(r,\varphi))\, d\lambda - \int_\Omega f(r,\varphi)\, d\lambda \int_\Omega g(r,\varphi)\, d\lambda \right| \le C_f \|g\|_\infty n^{-1}$$

holds for any Hölder functions f and g with some constant $C_f > 0$ for all $n \in \mathbb{N}$.

The proof of this assertion was obtained in [59] by an application of the construction of a Young tower with a polynomial tail; hence Young's theory was actually used in [59] to prove the stronger assertion that this inequality holds for all $g \in L_\infty(\Omega)$ and Hölder functions f (as had been noted as early as in [10, Example 1.7]). This is just condition (5.3) with $\gamma = 1$. Consequently, if the function f to be averaged satisfies the Hölder condition, then such billiards also satisfy the power-law estimate (5.5) of the rate of convergence in the Birkhoff ergodic theorem with the same exponent $\gamma = 1$; namely,

$$\mathrm{P}_n^{2\varepsilon} \le C_{f,\varepsilon,1} \left(\frac{r(\varepsilon)}{r(\varepsilon) - 1} \right) n^{-1}$$

for any $\varepsilon > 0$ and all $n \in \mathbb{N}$. It is of interest to study the resulting constants. One can estimate the constant $C_{f,\varepsilon,1}$, say, by taking $q = 3/2$ in inequality (5.4),

$$C_{f,\varepsilon,1} < 1875 C_f \frac{\|f - f^*\|_\infty^2}{\varepsilon^3}.$$

Recall that $r(\varepsilon) = 1 + \frac{\varepsilon}{\|f - f^*\|_\infty}$; thus, we obtain

$$\mathrm{P}_n^{2\varepsilon} \le C_f \frac{1875}{\varepsilon} \left(\left(\frac{\|f - f^*\|_\infty}{\varepsilon} \right)^3 + \left(\frac{\|f - f^*\|_\infty}{\varepsilon} \right)^2 \right) n^{-1}.$$

Since

$$\|A_k f - f^*\|_\infty = \|A_k(f - f^*)\|_\infty \le \|f - f^*\|_\infty,$$

it follows that the inequalities $\|A_k f - f^*\|_\infty < 2\varepsilon$ hold for all k provided that $\|f - f^*\|_\infty < 2\varepsilon$, whence we see that $\mathrm{P}_n^{2\varepsilon} = 0$ for all n. Hence we can assume that $\|f - f^*\|_\infty \ge 2\varepsilon$ in our last estimate and weaken it by rewriting it in the following simplified form:

$$\mathrm{P}_n^{2\varepsilon} \le C_f \frac{2812{,}5}{\varepsilon} \left(\frac{\|f - f^*\|_\infty}{\varepsilon} \right)^3 n^{-1}$$

for any $\varepsilon > 0$ and all $n \in \mathbb{N}$.

Estimates for characteristic functions. By using the estimate obtained above for the constant $C_{f,\varepsilon,1}$ and inequality (5.6) with $\gamma = 1$, we obtain the following estimate for the rate of convergence in the Birkhoff theorem for the characteristic function χ_E of a bounded Borel set E with regular boundary for all sufficiently small $\varepsilon > 0$ and all $n \in \mathbb{N}$:

$$\mathrm{P}_n^{4\varepsilon}(\chi_E) \le C(\varepsilon)\biggl(1 + \frac{\|\chi_E - \lambda(E)\|_\infty}{2\varepsilon}\biggr)n^{-1},$$

where

$$C(\varepsilon) = C_{\varphi_1,\varepsilon,1} + C_{\varphi_2,\varepsilon,1} \le 1875\frac{C_{\varphi_1} + C_{\varphi_2}}{\varepsilon^3}.$$

Estimate of the rate of convergence in the von Neumann theorem. By applying Theorem 7 with $\gamma = 1$, we obtain the following estimate for the rate of convergence in the von Neumann theorem for all billiards considered here, for every Hölder function f, and for all $n \in \mathbb{N}$:

$$\|A_n f - f^*\|_2^2 \le \|f - f^*\|_2^2 n^{-1} + 2C_f\|f - f^*\|_\infty \frac{\ln n + 1}{n}$$
$$< (C_f + \|f - f^*\|_\infty)^2 n^{-1} + 2C_f\|f - f^*\|_\infty \frac{\ln n}{n}.$$

Remark 18. It is remarkable that the decay rate of correlations (and hence the resulting estimate of the rate of convergence in the von Neumann theorem) may have distinct asymptotics in the cases of discrete and continuous time for billiards in one and the same domain $\mathcal{M}$. For example, it was shown in [62] (for sufficiently regular observables) that the sharp asymptotics of the correlation decay for the billiard flow of the Bunimovich stadium is the same as for the billiard map, that is, $O(t^{-1})$ as $t \to \infty$. However, the correlation decay for scattering billiards with traps has distinct asymptotics for distinct types of time; namely, it admits the estimates $O(t^{-k})$ for every $k \ge 1$ in the case of continuous time and has a sharp asymptotic of the order $O(t^{-1})$ in the case of discrete time.

6. Dynamical systems with exponential decay of correlations and large deviations

Dynamical systems with hyperbolic behavior (one-dimensional expanding maps, uniformly and nonuniformly hyperbolic systems, and also partially hyperbolic systems) provide numerous examples of dynamical systems with exponential estimates of the decay of correlations and large deviations, usually for the case of Hölder observables.

In this section, we estimate the rate of convergence in the von Neumann theorem for β-adic shifts with $\beta > 2$ and for the Gauss map. We obtain estimates of the rate of convergence both in the von Neumann theorem and in the Birkhoff theorem for Anosov systems, for dynamical systems modeled by a Young tower with an exponential tail (of which main attention is paid to quadratic maps and the periodic Lorentz gas), and also for Teichmüller flows.

The decay of the remainder term in the asymptotic representations (2.1) and (2.2) of the rate of convergence in the von Neumann theorem is exponential for these systems as well.

6.1. **Piecewise expanding maps.**

In this subsection, we deal with one-dimensional maps that, almost everywhere with respect to their invariant measures, have derivatives whose absolute value is greater than unity. All the examples considered here are well-known one-dimensional chaotic maps. The correlation coefficients used to estimate the rate of convergence in the von Neumann theorem for such maps are either computed exactly for functions representable by series in Bernoulli or Euler polynomials (the Bernoulli shifts and the tent map) or estimated efficiently for some special classes of regular functions in which the action of the Perron–Frobenius operator is well studied (the Rényi and Gauss maps).

6.1.1. *Bernoulli shift and the tent map.* For an arbitrary positive integer $N \geq 2$, consider the Nth Bernoulli shift $T_N \colon [0,1) \to [0,1)$ defined by

$$T_N(x) = Nx \pmod 1 = \sum_{k=0}^{N-1} (Nx - k)\chi_{[\frac{k}{N}, \frac{k+1}{N})}.$$

By $\widetilde{T}_N$ we denote the tent map with N branches. It only differs from T_N in that the slope of the even linear branches is negative. More precisely, $\widetilde{T}_N \colon [0,1) \to [0,1)$ is given by the formula

$$\widetilde{T}_N(x) = \sum_{k=0}^{N-1} \biggl((-1)^k Nx + (-1)^{k+1} \biggl(k + \frac{1-(-1)^k}{2} \biggr) \biggr) \chi_{[\frac{k}{N}, \frac{k+1}{N})}.$$

The invariant measure for these systems is the Lebesgue measure (with invariant density $\varrho \equiv 1$).

Estimate of correlations for T_N. It is well known that the Bernoulli polynomials $B_n(x)$ are polynomial eigenfunctions of the Perron–Frobenius operator for the Bernoulli shift T_N with eigenvalues $1/N^n$ (e.g., see [63]). One can use this fact to compute the correlation coefficients exactly (and in turn use them to estimate the rate of convergence in the von Neumann theorem) for functions admitting series expansions in Bernoulli polynomials. By way of example, we do this below for the powers $x^k, k \in \mathbb{N}$. By the Euler–Maclaurin formula, we have the expansion

$$x^k = \frac{1}{k+1} + \sum_{n=1}^{k} \frac{\binom{n}{k+1}}{k+1} B_n(x),$$

where $\binom{n}{k+1} = \frac{(k+1)!}{n!(k+1-n)!}$. By using this expansion, we obtain

$$c_m(x^k) = \int_0^1 x^k P^m x^k \, dx - \biggl(\int_0^1 x^k \, dx \biggr)^2 = \sum_{n=1}^{k} \frac{\binom{n}{k+1}}{k+1} \int_0^1 x^k P^m B_n(x) \, dx$$
$$= \sum_{n=1}^{k} \frac{\binom{n}{k+1}}{N^{nm}(k+1)} \int_0^1 x^k B_n(x) \, dx$$

for all $m \in \mathbb{N}$.

By using relation (5.0) between the correlation coefficients, we obtain the estimate

$$\biggl| b_m \biggl(x^k - \frac{1}{k+1} \biggr) \biggr| \leq A_k^{(1)} e^{-B^{(1)} m} \tag{6.1}$$

for all $m \in \mathbb{N}$, where

$$A_k^{(1)} = \sum_{n=1}^{k} \frac{\binom{n}{k+1}}{k+1} \biggl| \int_0^1 x^k B_n(x) \, dx \biggr|$$

and $B^{(1)} = \ln N$ is the Lyapunov exponent of the dynamical system in question.

Estimate of correlations for $\widetilde{T}_N$. The polynomial eigenfunctions of the tent map $\widetilde{T}_N$ already depend on the parity of N. If $N = 2M$, then the eigenfunction corresponding to the eigenvalue $1/N^{2l}$ is 1 for $l = 0$ and the polynomial $B_{2l}(x) - lE_{2l-1}x$ for $l \geq 1$, where the $B_{2l}(x)$ are the Bernoulli polynomials and the $E_{2l-1}(x)$ are the Euler polynomials. The eigenfunctions corresponding to the zero eigenvalue are the Bernoulli and Euler polynomials $B_{2l-1}(x)$ and $E_{2l-1}(x)$, $l \geq 1$, with odd numbers. If $N = 2M - 1$, then the eigenvalues are again $1/N^{2l}$, $l \geq 0$; they are of multiplicity 2 for $l \geq 1$, and the corresponding eigenfunctions are the polynomials $B_{2l}(x)$ and $E_{2l-1}(x)$ [64, 63].

By applying the Euler–Maclaurin formula to the monomials x^k, $k \in \mathbb{N}$, we obtain the expansion

$$x^k = \frac{1}{k+1} + \sum_{n=1}^{[k/2]} \frac{\binom{2n}{k+1}}{k+1} B_{2n}(x) + \sum_{n=1}^{[(k+1)/2]} \frac{2^{\delta(n,(k+1)/2)}}{2} \binom{2n-1}{k} E_{2n-1}(x)$$

in Bernoulli and Euler polynomials, where $\delta(n,(k+1)/2)$ is the Kronecker delta.

We use this expansion and obtain

$$\begin{aligned} c_m(x^k) &= \sum_{n=1}^{[k/2]} \frac{\binom{2n}{k+1}}{k+1} \int_0^1 x^k P^m B_{2n}(x)\,dx \\ &= \sum_{n=1}^{[k/2]} \frac{\binom{2n}{k+1}}{k+1} \int_0^1 x^k P^m (B_{2n}(x) - nE_{2n-1}(x))\,dx \\ &= \sum_{n=1}^{[k/2]} \frac{\binom{2n}{k+1}}{N^{2nm}(k+1)} \int_0^1 x^k (B_{2n}(x) - nE_{2n-1}(x))\,dx \end{aligned}$$

for $N = 2M$ and for all $m \in \mathbb{N}$ and

$$\begin{aligned} c_m(x^k) = \sum_{n=1}^{[k/2]} \frac{\binom{2n}{k+1}}{N^{2mn}(k+1)} \int_0^1 x^k B_{2n}(x)\,dx \\ + \sum_{n=1}^{[(k+1)/2]} \frac{2^{\delta(n,(k+1)/2)}}{2} \frac{\binom{2n-1}{k}}{N^{2mn}} \int_0^1 x^k E_{2n-1}(x)\,dx \end{aligned}$$

for $N = 2M - 1$ and for all $m \in \mathbb{N}$.

Let us again use relation (5.0) between the correlation coefficients; then for all $m \in \mathbb{N}$ we obtain the estimate

$$\left| b_m\left(x^k - \frac{1}{k+1}\right)\right| \leq A_k^{(2)} e^{-B^{(2)}m}, \tag{6.2}$$

where

$$\begin{aligned} A_k^{(2)} = \max\Bigg\{ &\sum_{n=1}^{[k/2]} \frac{\binom{2n}{k+1}}{k+1} \left|\int_0^1 x^k (B_{2n}(x) - nE_{2n-1}(x))\,dx\right|; \\ &\sum_{n=1}^{[k/2]} \frac{\binom{2n}{k+1}}{k+1} \left|\int_0^1 x^k B_{2n}(x)\,dx\right| \\ &\quad + \frac{1}{2} \sum_{n=1}^{[(k+1)/2]} 2^{\delta(n,(k+1)/2)} \binom{2n-1}{k} \left|\int_0^1 x^k E_{2n-1}(x)\,dx\right| \Bigg\} \end{aligned}$$

and $B^{(2)} = 2\ln N$.

Estimates of the rate of convergence in the von Neumann theorem. Let us apply assertion 6 of Theorem 7 to the estimates (6.1) and (6.2). Then we arrive at the asymptotic relation (2.1) with an estimate of the rate of convergence with respect to both dynamical systems in question in the von Neumann theorem for the monomials x^k, $k \in \mathbb{N}$,

$$\left| \left\| A_n x^k - \frac{1}{k+1} \right\|_2^2 - 2\pi\rho(0)n^{-1} + 2\pi(\rho')^c(0)n^{-2} \right| \\ \leq \frac{2A_k^{(i)} e^{B^{(i)}}}{e^{B^{(i)}} - 1} \left(1 + \frac{1}{e^{B^{(i)}} - 1} \cdot \frac{1}{n} \right) \frac{e^{-B^{(i)}n}}{n} < 2A_k^{(i)} \left(\frac{N^i}{N^i - 1} \right)^2 \frac{1}{nN^{in}}, \quad i = 1, 2,$$

for all $n \in \mathbb{N}$; here

$$0 \leq 2\pi\rho(0) \leq \frac{1}{2k+1} + \frac{2A_k^{(i)}}{N^i - 1}, \qquad |2\pi(\rho')^c(0)| \leq \frac{2A_k^{(i)} N^i}{(N^i - 1)^2}, \quad i = 1, 2.$$

Since an arbitrary polynomial is a linear combination of monomials, it follows that it can be expanded in Bernoulli and Euler polynomials as well, and hence similar estimates of the rate of convergence in the von Neumann theorem remain valid for arbitrary polynomials (with a much more awkward estimate for the coefficients $A^{(i)}$, $i = 1, 2$).

6.1.2. *β-adic Rényi map.* It is only rarely that the correlations can be computed exactly for a fairly broad function class. Nevertheless, there exist efficient estimates of correlations in the class of Lipschitz functions, say, for dynamical systems with a finite special partition (see [46, Chap. 2] for the definition). Here we consider an application of this approach to the well-known Rényi map. (Note that this approach also applies to the Bernoulli shifts and tent maps considered above.)

For an arbitrary noninteger $\beta > 1$, consider the Rényi map $T_\beta\colon [0,1) \to [0,1)$, which generalizes the Bernoulli shift and is given by the formula

$$T_\beta(x) = \beta x \pmod 1 = \sum_{i=1}^{W} (\beta x - (i-1))\chi_{Y_i},$$

where $W = [\beta] + 1$, $Y_i = \left[\frac{i-1}{\beta}, \frac{i}{\beta}\right)$, $1 \leq i < W$, and $Y_W = \left[\frac{W-1}{\beta}, 1\right)$. Starting from the pioneering papers by Rényi and Rokhlin, there have been many papers dealing with such systems. In particular, it was established that this dynamical system has a unique absolutely continuous invariant measure, which is the measure λ_β whose density ϱ_β satisfies the inequality

$$\frac{\beta - 1}{\beta} \leq \varrho_\beta(x) \leq \frac{\beta}{\beta - 1}$$

for all $x \in [0,1]$ and can be expressed by the formula

$$\varrho_\beta(x) = \frac{1}{c_\beta} \sum_{n=0}^{\infty} \frac{1}{\beta^n} \chi_{\{z < T_\beta^n(1)\}}(x),$$

where c_β is a normalizing factor (e.g., see [65]).

Further, by applying the results in [46, Theorems 2.2.3 and 2.4.1] with $\beta > 2$, we find that the density ϱ_β of the unique absolutely continuous measure λ_β is a function of bounded variation; the correlations of Lipschitz observables decay exponentially with respect to this measure. Let us state this result in [46] more precisely.

Estimate of correlations. For each function $g\colon [0,1] \to \mathbb{R}$ of bounded variation, consider the norm

$$\|g\|_{BV} = \operatorname{var}(g) + \int_0^1 |g(x)|\, dx.$$

Let f and g be Lipschitz functions; then the estimate

$$|c_n(f,g)| \leq A(\operatorname{Lip}(f) + \|f\|_\infty)(\operatorname{Lip}(g) + \|g\|_\infty)e^{-Bn},$$

where

$$A = 5 + 2QW \max_{1\le i\le W}\Big\{(2+\lambda_\beta(Y_i))\|\varrho_\beta\|_{BV} + \frac{2}{|Y_i|}\Big\}, \qquad B = \frac{1}{2}\ln\frac{1}{q},$$

holds for every $n \in \mathbb{N}$. The constants $Q > 0$ and $q \in (0,1)$ are determined from the spectral decomposition of the Perron–Frobenius operator into the sum of the projection P_1 onto the fixed point space and an operator P_0,

$$P = P_0 + P_1, \qquad \|P_0^n\|_{BV} \le Qq^n.$$

For the map T_β, the estimate of correlations for any Lipschitz function f and $n \in \mathbb{N}$ has the form

$$|b_n(f - f^*)| \le A_f e^{-Bn}, \tag{6.3}$$

where the constant A_f is obtained from A as follows:

$$A_f = \left[2Q([\beta]+1)\left(3 + \frac{2\beta}{\{\beta\}} + 3\operatorname{var}(\varrho_\beta)\right) + 5\right]\big(\operatorname{Lip}(f-f^*) + \|f-f^*\|_\infty\big)^2.$$

Estimate of the rate of convergence in the von Neumann theorem. We apply assertion 6 of Theorem 7 to the estimate (6.3) and obtain the asymptotic relation (2.1) with an estimate of the rate of convergence for a Lipschitz function f in the von Neumann theorem for the Rényi map T_β, $\beta > 2$,

$$\big|\|A_nf - f^*\|_2^2 - 2\pi\rho(0)n^{-1} + 2\pi(\rho')^c(0)n^{-2}\big|$$
$$\le \frac{2A_fe^B}{e^B-1}\left(1 + \frac{1}{e^B-1}\cdot\frac{1}{n}\right)\frac{e^{-Bn}}{n} < \frac{2A_f}{(1-\sqrt{q})^2}\frac{q^{n/2}}{n}$$

for all $n \in \mathbb{N}$; here

$$0 \le 2\pi\rho(0) \le \|f\|_2^2 + \frac{2A_f\sqrt{q}}{1-\sqrt{q}}, \qquad |2\pi(\rho')^c(0)| \le \frac{2A_f\sqrt{q}}{(1-\sqrt{q})^2}.$$

6.1.3. *Continued fractions.* The continued fraction expansion

$$x = [a_1, a_2, a_3, \dots] = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\ddots}}}}$$

of a real number $x \in [0,1)$ is related to the well-known dynamical system (e.g., see [3]), the Gauss transformation $T_G\colon [0,1) \to [0,1)$, which acts by the rule

$$T_Gx = \frac{1}{x} \pmod 1 = [a_2, a_3, a_4, \dots], \quad x > 0, \qquad T_G0 = 0.$$

The digits a_n can be found by the formula

$$a_n = \left\lfloor \frac{1}{T_G^{n-1}x} \right\rfloor,$$

where $\lfloor\cdot\rfloor$ stands for the real part of a number.

In contrast to the piecewise linear systems considered above, this nonlinear dynamical system has countably many points of discontinuity. Nevertheless, one can still obtain efficient estimates of the decay rate of correlations (and hence estimates of the rate of convergence in the von Neumann theorem) with respect to the unique absolutely continuous invariant measure λ_G with density $\varrho_G(x) = \frac{1}{\ln 2(1+x)}$ for some special classes of regular functions (see [66, 67, 68] and also the survey [69]). Let us describe these results.

Estimates of correlations. Let $D_1 = \{z \in \mathbb{C}, \operatorname{Re} z > -1/2\}$. For a measure on the Borel subsets of D_1 we take the normalized truncated Cauchy distribution, i.e., the absolutely continuous measure ν (with respect to the plane Lebesgue measure) with density

$$\varrho_\nu(z) = \begin{cases} \frac{1}{\pi \ln 2((1+x)^2+y^2)}, & -\frac{1}{2} < \operatorname{Re} z < 0, \\ 0, & \operatorname{Re} z \geq 0. \end{cases}$$

Let $\mathscr{H}^2_\nu(D_1)$ be the Banach space of holomorphic functions $f(z)$ in D_1 such that the function $\left|\frac{f(z)}{1+z}\right|$ is bounded in the half-plane $\operatorname{Re} z > -1/2+\varepsilon$ for each $\varepsilon > 0$ and the norm

$$\|f\|_{2,\nu} = \left(\int_{D_1} |f(z)|^2 \, d\nu\right)^{1/2}$$

is finite. For $f \in \mathscr{H}^2_\nu(D_1)$, let $\widehat{f}$ be the restriction of f to the unit segment $\Omega = \{0 \leq \operatorname{Re} z < 1, \operatorname{Im} z = 0\}$. (In what follows, we use the same notation for the restriction to Ω of functions f of other classes.) Then each such restriction lies in $L_2(\Omega, \lambda_G)$, and its norm satisfies the estimate $\|\widehat{f}\|_2 \leq \|f\|_{2,\nu}$. This relation was used in [66] to obtain the following estimate for the decay of correlations. For any $f \in L_2(\Omega, \lambda_G)$ and $g \in \mathscr{H}^2_\nu(D_1)$, one has

$$|c_n(\widehat{g}, f)| \leq q^n \|f\|_2 \|g\|_{2,\nu} \tag{6.4}$$

for all $n \in \mathbb{N}$, where the Kuzmin–Levy–Wirsing constant $q \simeq 0{,}30366$ is defined as the second (in ascending order of absolute values) eigenvalue of the transfer operator (see [66]).

A similar result was obtained in [67] for another special class of regular functions. Let $D_2 = \{z \in \mathbb{C}\colon |z-1| < 3/2\}$. By $\mathscr{A}_{1,\infty}(D_2)$ we denote the Banach space of functions holomorphic in the disk D_2 and, together with the first derivative, continuous in its closure $\overline{D}_2$. The norm of a function $f \in \mathscr{A}_{1,\infty}(D_2)$ is defined by

$$\|f\|_{1,\infty} = \max\bigl\{\max_{z\in\overline{D}_2} |f(z)|;\ \max_{z\in\overline{D}_2} |f'(z)|\bigr\}.$$

Then the estimate

$$|c_n(\widehat{f}, \widehat{g})| \leq q^n \|f\|_{1,\infty} \|g\|_{1,\infty} \tag{6.5}$$

holds for any $f, g \in \mathscr{A}_{1,\infty}(D_2)$ and $n \in \mathbb{N}$, where q is the Kuzmin–Levy–Wirsing constant (the same as in inequality (6.4)).

Estimates of the form (6.4) and (6.5) can be extended to smooth functions. It was shown in [68] that for any $k \in \mathbb{N}$ and any functions $f, g \in C^k[0,1]$ there exists a constant $C = C_f \geq \|f\|_{C^k}$ such that

$$|c_n(f,g)| \leq \widetilde{q}^n C_f \|g\|_{C^k} \tag{6.6}$$

for all $n \in \mathbb{N}$, where $\widetilde{q} = q$ for $k \geq 2$; for $k = 1$, one has the estimate $(3-\sqrt{5})/2 < \tilde{q} < 1$ for $\tilde{q}$, and $C_f = \|f\|_{C^1}$.

Estimates of the rate of convergence in the von Neumann theorem. We use inequalities (6.4), (6.5), and (6.6), relations (5.0) for the correlation coefficients, and assertion 6 in Theorem 7 to obtain the asymptotic relation (2.1) with an estimate of the rate of convergence in the von Neumann theorem for the Gauss map and for any f in the special classes of regular functions considered above; namely,

$$\big|\|A_n\widehat{f} - \widehat{f}^*\|_2^2 - 2\pi\rho(0)n^{-1} + 2\pi(\rho')^c(0)n^{-2}\big| < 2A^2\left(\frac{1}{1-\tilde{q}}\right)^2\frac{\tilde{q}^n}{n}$$

for all $n \in \mathbb{N}$, where

$$A = \begin{cases} \|f\|_{2,\nu}, & f \in \mathscr{H}^2_\nu(D_1), \tilde{q} = q, \\ \|f\|_{1,\infty}, & f \in \mathscr{A}_{1,\infty}(D_2), \tilde{q} = q, \\ C_f, & f \in C^k[0,1], k \geq 2, \tilde{q} = q, \\ \|f\|_{C^1}, & f \in C^1[0,1], \frac{3-\sqrt{5}}{2} < \tilde{q} < 1. \end{cases}$$

Here

$$0 \leq 2\pi\rho(0) \leq \|\widehat{f}\|_2^2 + \frac{2A\tilde{q}}{1-\tilde{q}}, \qquad |2\pi(\rho')^c(0)| \leq \frac{2A\tilde{q}}{(1-\tilde{q})^2}.$$

Yet another example of one-dimensional dynamics (quadratic maps) will be considered in Section 6.3.1 below.

6.2. **Uniformly hyperbolic systems.** In this section, we give estimates of the rate of convergence in the von Neumann and Birkhoff theorems for Anosov diffeomorphisms and Hölder functions to be averaged.

For Anosov flows, we consider examples with exponential and subexponential decay rate of correlations and obtain the corresponding estimates of the rate of convergence in the von Neumann theorem for these examples. In the Birkhoff theorem for Anosov flows, we establish exponential estimates of the rate of convergence for Hölder functions satisfying the condition of independence from the flow. This independence condition is a technical condition needed when deriving the large deviation estimates used here; the author of these estimates claims in [70, Remark 2] that “similar estimates can be obtained without this condition”.

For Anosov diffeomorphisms, we also obtain an estimate of the rate of convergence in the Birkhoff theorem for the characteristic functions of bounded Borel subsets of the phase space with boundary of measure zero.

6.2.1. *Anosov diffeomorphisms.* Let $\mathcal{M}$ be a compact C^∞ Riemannian manifold on which a smooth diffeomorphism T is defined. This dynamical system is called an *Anosov diffeomorphism* if for each point $x \in \mathcal{M}$ the tangent space $T_x\mathcal{M}$ splits into a sum of subspaces forming continuous subbundles of the bundle $T\mathcal{M}$ invariant under the differential DT,

$$T_x\mathcal{M} = E^s_x \oplus E^u_x,$$

where

$$\|DT^n w\| \leq ce^{-\theta n}\|w\| \qquad \text{for all} \quad w \in E^s_x, \quad n \geq 1,$$
$$\|DT^n w\| \leq ce^{\theta n}\|w\| \qquad \text{for all} \quad w \in E^u_x, \quad n \leq 1,$$

with some positive constants c and θ independent of x.

For the invariant measure λ we take the SRB measure whose existence is proved in [71] even for the case of a diffeomorphism T of the class $C^{1+\alpha}$; i.e., for the case in which the differential DT belongs to the Hölder class with exponent $\alpha \in (0,1)$. We assume that the diffeomorphism T is transitive. The decay of correlations with respect to this λ for Hölder functions has long been known; it is exponential [72]. This result was first obtained for mixing topological Markov chains and then extended by methods of symbolic dynamics to transitive Anosov diffeomorphisms; it can also be obtained by an application of the construction of a Young tower with an exponential tail [6] discussed below. However, we follow the similar estimates with sharper constants obtained by the coupling method in [71].

Estimate of correlations. Let T be a $C^{1+\alpha}$ transitive Anosov diffeomorphism for some given $\alpha \in (0,1)$, and let $d_s(x,y)$ and $d_u(x,y)$ be the metrics induced on the stable and unstable manifolds by the Riemannian metric on $\mathcal{M}$. Take a $\delta > 0$. For a function f measurable on $\mathcal{M}$, consider the seminorms

$$\|f\|_s = \|f\|_\infty + \sup_{d_s(x,y)\le\delta} \frac{|f(x)-f(y)|}{d_s^\alpha(x,y)},$$
$$\|f\|_u = \|f\|_1 + \sup_{d_u(x,y)\le\delta} \frac{|f(x)-f(y)|}{d_u^\alpha(x,y)},$$

where the norm $\|f\|_1$ is taken with respect to the Riemannian volume on $\mathcal{M}$. It is clear that if f is a Hölder function with exponent $\alpha \in (0,1)$, then $\|f\|_s < \infty$ and $\|f\|_u < \infty$. It was shown in [71] that there exist constants $0 < \vartheta < 1$ and $A > 0$ such that

$$|c_n(f,g)| \le A\|f\|_s\|g\|_u\vartheta^n \tag{6.7}$$

for any Hölder functions f and g with exponent $\alpha \in (0,1)$ and all $n \in \mathbb{N}$.

Estimate of the rate of convergence in the von Neumann theorem. We use the estimate (6.7), relations (5.0) between the correlation coefficients, and assertion 6 of Theorem 7 to obtain the asymptotic relation (2.1) with an estimate of the rate of convergence in the von Neumann theorem for $C^{1+\alpha}$ transitive Anosov diffeomorphisms and for any Hölder function f with exponent $\alpha \in (0,1)$. Namely,

$$\big|\|A_nf - f^*\|_2^2 - 2\pi\rho(0)n^{-1} + 2\pi(\rho')^c(0)n^{-2}\big| < 2A\|f\|_s\|f\|_u\left(\frac{1}{1-\vartheta}\right)^2\frac{\vartheta^n}{n}$$

for all $n \in \mathbb{N}$, where

$$0 \le 2\pi\rho(0) \le \|f\|_2^2 + \frac{2A\|f\|_s\|f\|_u\vartheta}{1-\vartheta}, \qquad |2\pi(\rho')^c(0)| \le \frac{2A\|f\|_s\|f\|_u\vartheta}{(1-\vartheta)^2}.$$

Estimate of large deviations. Estimates of the decay rate of large deviation probabilities for C^2 transitive Anosov diffeomorphisms are well known for the SRB measure λ as well as for any Gibbs state λ_g generated by a Hölder function g [73, 74]. (Recall that the SRB measure λ is the Gibbs state λ_g with the Hölder function $g(x) = -\ln \operatorname{Jac}(DT\,|_{E_x^u})$.) Let us state these results following [75].

For each Hölder function f that is not cohomological to a constant (this condition is not restrictive by Remark 15), there exists an interval $[a,b]$ and a real-analytic function $I(t)\colon [a,b] \to \mathbb{R}^+$, which also depend on the function g, such that the large deviation principle

$$\lim_{n\to\infty}\frac{1}{n}\ln\lambda_g\left\{\left|A_nf - \int_\Omega f\,d\lambda_g\right| \ge \varepsilon\right\} = -\gamma_0(\varepsilon),$$
$$\gamma_0(\varepsilon) = \inf\left\{I(t)\colon t \in \left[a, \int_\Omega f\,d\lambda_g - \varepsilon\right] \cup \left[\int_\Omega f\,d\lambda_g + \varepsilon, b\right]\right\} > 0,$$

holds for any sufficiently small $\varepsilon > 0$.

This principle readily implies that for any Hölder function f not cohomological to a constant, any sufficiently small $\varepsilon > 0$, and any $\gamma \in (0, \gamma_0(\varepsilon))$ there exists a constant $C > 0$ depending on ε, γ, and f such that the inequality $\mathrm{p}_n^\varepsilon \le Ce^{-\gamma n}$ holds for all $n \ge 1$.

Estimates of the rate of convergence in the Birkhoff theorem. By applying the last inequality and the exponential estimate in Theorem 13, we obtain an estimate of the rate of convergence in the Birkhoff theorem for Anosov diffeomorphisms with respect to the measure λ_g for any Hölder function f not cohomological to a constant, for sufficiently small $\varepsilon > 0$, and for any $\gamma \in (0, \gamma_0(\varepsilon))$. Namely,

$$\mathrm{P}_n^{2\varepsilon} \le C\left(1 + \frac{\ln(1+\gamma^{-1}(\varepsilon))}{\ln\left(1+\varepsilon\left\|f - \int f\,d\lambda_g\right\|_\infty^{-1}\right)}\right)e^{-\gamma n}$$

for all $n \ge 1$.

Now consider the characteristic function of a bounded Borel set E with regular boundary (i.e., with boundary of zero λ_g-measure). By applying Theorem 14, we obtain an estimate of large deviations and hence an estimate of the rate of convergence in the Birkhoff theorem also for this function,

$$\begin{aligned} &\mathrm{p}_n^{2\varepsilon}(\chi_E) \le \mathrm{p}_n^{\varepsilon}(\varphi_1) + \mathrm{p}_n^{\varepsilon}(\varphi_2) \le C_1(\varepsilon)e^{-\widetilde{\gamma}(\varepsilon)n}, \\ &\mathrm{P}_n^{4\varepsilon} \le C_1(\varepsilon)\left(1 + \frac{\ln(1+\widetilde{\gamma}^{-1}(\varepsilon))}{\ln\left(1+2\varepsilon\|\chi_E - \lambda_g(E)\|_\infty^{-1}\right)}\right)e^{-\widetilde{\gamma}(\varepsilon)n}; \end{aligned} \tag{6.8}$$

these estimates hold for sufficiently small $\varepsilon > 0$ with some constants $C_1(\varepsilon), \widetilde{\gamma}(\varepsilon) > 0$ for all $n \ge 1$.

Remark 19. In view of Remark 14, one can also obtain an estimate of the rate of convergence for Anosov diffeomorphisms in the Bowen theorem. Recall that this theorem claims (e.g., see [72]) that for each function f continuous on $\mathcal{M}$ the ergodic means $A_n f$ converge to $f^* = \int f\,d\lambda$ as $n \to \infty$ a.e. with respect to the Riemannian volume m of the manifold $\mathcal{M}$. Here, just as above, λ is the SRB measure.

Estimates of large deviation probabilities with respect to the measure m for Anosov systems are well studied. As was shown in [73, 74, 76], the large deviation principle

$$\overline{\lim_{n\to\infty}}\, \frac{1}{n} \ln m\left\{\left|A_n f - \int f\,d\lambda\right| \ge \varepsilon\right\} = -\gamma(\varepsilon)$$

holds for each function f continuous on $\mathcal{M}$ and for all sufficiently small $\varepsilon > 0$ with some constant $\gamma(\varepsilon) > 0$ depending on the dynamical system and the function f. It readily follows from this limit relation that for each $\gamma \in (0, \gamma(\varepsilon))$ there exists a constant $C > 0$ depending on γ, ε, and f such that the inequality

$$\mathrm{p}_n^{\varepsilon}(m) = m\left\{\left|A_n f - \int f\,d\lambda\right| \ge \varepsilon\right\} \le Ce^{-\gamma n}$$

holds for all $n \ge 1$. By Theorem 13, in view of Remark 14, this implies an estimate of the rate of convergence in the Bowen theorem for any continuous function f and any sufficiently small $\varepsilon > 0$; namely,

$$\mathrm{P}_n^{2\varepsilon}(m) = m\left\{\sup_{k\ge n}\left|A_k f - \int f\,d\lambda\right| \ge 2\varepsilon\right\} \le C\left(1 + \frac{\ln(1+1/\gamma)}{\ln\left(1+\varepsilon\left\|f - \int f\,d\lambda\right\|_\infty^{-1}\right)}\right)e^{-\gamma n} \tag{6.9}$$

for all $n \ge 1$.

Note that the last relation gives a proof other than the original one of the Bowen theorem itself. Further, by applying Remark 17 to the Riemannian volume m, we obtain an estimate, similar to (6.8), of the rate of convergence of ergodic means a.e. with respect to Riemannian volume for the characteristic function of a bounded Borel set with boundary of zero λ-measure. This, in turn, proves that the Bowen theorem holds not only for continuous functions but also for the discontinuous characteristic functions in question.

6.2.2. *Anosov flows.* Let $\mathcal{M}$ be a compact C^∞ Riemannian manifold equipped with a C^2 flow T^t. This system is called an *Anosov flow* if for each point $x \in \mathcal{M}$ the tangent space $T_x\mathcal{M}$ splits into a sum of three subspaces forming continuous subbundles of the bundle $T\mathcal{M}$ invariant with respect to the differential DT^t,

$$T_x\mathcal{M} = E_x \oplus E_x^s \oplus E_x^u,$$

where E_x is the tangent to the trajectory of the system T^t through x and

$$\|DT^t w\| \le c e^{-\theta t}\|w\| \qquad \text{for all} \quad w \in E_x^s, \quad t \ge 0,$$
$$\|DT^t w\| \le c e^{\theta t}\|w\| \qquad \text{for all} \quad w \in E_x^u, \quad t \le 0,$$

with some positive constants c and θ independent of x. We also assume that the flow is topologically transitive; i.e., for any nonempty open sets $U, V \subset \mathcal{M}$ there exists a τ such that $U \cap T^\tau V \ne \emptyset$. This condition is not restrictive, because Smale's spectral decomposition theorem [77] says that otherwise $\mathcal{M}$ can be decomposed into a finite sum of sets the restrictions to which give topologically transitive flows.

Estimate of large deviations. For the invariant measure for T^t we take the SRB measure λ whose existence and ergodicity were proved, say, in [72]. It was shown in [70, Sec. 7] that the following large deviation principle (which is a corollary of a more general principle) holds for each Hölder function f on $\mathcal{M}$. Let J be a closed interval not containing the value $\int f\,d\lambda$; then there exist constants C and $\gamma > 0$ depending on J and f such that

$$\lambda\{\overline{A}_t f \in J\} \le C\frac{e^{-\gamma t}}{\sqrt{t}}$$

for all sufficiently large t. This principle was proved in [70] under the additional assumption (which should be noted to be inconvenient) that the function f and the flow T^t are independent (in the sense of Definition 2 in [70]).

Let $K = \max_{x\in\mathcal{M}} |f(x)|$; then $\overline{A}_t f(x) \in [-K, K]$ for all $t > 0$ and λ-almost all x. By using this fact and the principle described above, we obtain the inequalities

$$\begin{aligned}
\overline{\mathrm{p}}_t^\varepsilon &= \lambda\left\{\left|\overline{A}_t f - \int f\,d\lambda\right| \ge \varepsilon\right\} = \lambda\left\{\int f\,d\lambda - \varepsilon \ge \overline{A}_t f \ge \int f\,d\lambda + \varepsilon\right\}\\
&= \lambda\left\{\overline{A}_t f \in \left[-K, \int f\,d\lambda - \varepsilon\right] \cup \left[\int f\,d\lambda + \varepsilon, K\right]\right\}\\
&= \lambda\left\{\overline{A}_t f \in \left[-K, \int f\,d\lambda - \varepsilon\right]\right\} + \lambda\left\{\overline{A}_t f \in \left[\int f\,d\lambda + \varepsilon, K\right]\right\}\\
&\le C_1\frac{e^{-\gamma_1 t}}{\sqrt{t}} + C_2\frac{e^{-\gamma_2 t}}{\sqrt{t}} \le C_3\frac{e^{-\gamma(\varepsilon) t}}{\sqrt{t}}
\end{aligned}$$

for all sufficiently small $\varepsilon > 0$ and sufficiently large t, i.e., for $t > t_0$, where the constants C_1, γ_1 and C_2, γ_2 correspond to the intervals

$$J_1 = \left[-K, \int f\,d\lambda - \varepsilon\right] \quad \text{and} \quad J_2 = \left[\int f\,d\lambda + \varepsilon, K\right],$$

respectively, $C_3 = C_1 + C_2$, and $\gamma(\varepsilon) = \min\{\gamma_1, \gamma_2\}$. Then there exists a constant $C(\varepsilon) > 0$ (one can take $C(\varepsilon) = \frac{\sqrt{t_0}}{C_3}e^{\gamma(\varepsilon)t_0}$) such that the inequality

$$\overline{\mathrm{p}}_t^\varepsilon = \lambda\left\{\left|\overline{A}_t f - \int f\,d\lambda\right| \ge \varepsilon\right\} \le C(\varepsilon)\frac{e^{-\gamma(\varepsilon) t}}{\sqrt{t}}$$

holds for all $t > 0$.

Estimate of the rate of convergence in the Birkhoff theorem. By using the last inequality and the estimate 4 in Theorem 13, for sufficiently small $\varepsilon > 0$ we obtain an estimate of the rate of convergence in the Birkhoff theorem for Anosov flows for a Hölder function f under the restrictions imposed above. Namely,

$$\begin{aligned}\overline{\mathrm{P}}_t^{2\varepsilon} &\le C(\varepsilon)\left(1+\frac{A(-1/2,\gamma(\varepsilon))}{\ln\bigl(1+\varepsilon\|f-\int f\,d\lambda\|_\infty^{-1}\bigr)}\right)\frac{e^{-\gamma(\varepsilon)t}}{\sqrt{t}}\\ &\le C(\varepsilon)\left(1+\frac{2}{\ln\bigl(1+\varepsilon\|f-\int f\,d\lambda\|_\infty^{-1}\bigr)}\right)\frac{e^{-\gamma(\varepsilon)t}}{\sqrt{t}}\end{aligned}$$

for all $t>0$; the latter inequality follows by simple computations,

$$A\left(-\frac{1}{2},x\right)=\sqrt{x}e^x\int_x^\infty e^{-t}t^{-3/2}\,dt\le\sqrt{x}\,e^xe^{-x}\int_x^\infty t^{-3/2}\,dt=2.$$

A similar result with a slightly sharper estimate for $A(-1/2,\gamma(\varepsilon))$ was presented in [78],

$$A\left(-\frac{1}{2},\gamma(\varepsilon)\right)\le 2\frac{\sqrt{\gamma(\varepsilon)+2}-\sqrt{\gamma(\varepsilon)}}{\sqrt{\gamma(\varepsilon)+2}+\sqrt{\gamma(\varepsilon)}}.$$

We do not present an estimate for the rate of convergence in the Birkhoff theorem for the characteristic function of a bounded Borel set, because, to obtain this estimate, we would need estimates of the decay rates of large deviations for an arbitrary smooth function without any inconvenient restrictions such as the independence of the function from the flow.

Remark 20 ([78]). Since the Bowen theorem also holds for Anosov flows as well as diffeomorphisms (see Remark 19), and since estimates of the decay rate of large deviation probabilities with respect to the Riemannian volume m for any continuous function are known as well (see [74, 76]), we can use the same computations as in Remark 19 to obtain the estimate (6.9) of the rate of convergence in the Bowen theorem for a continuous observable function (one should only replace discrete averaging with its continuous counterpart) and also an estimate of the rate of convergence in this theorem for the characteristic function of a set with boundary of zero λ-measure.

Now let us proceed to estimates of the rate of convergence in the von Neumann theorem via the estimates given in [79, 80, 81, 82, 83] for the decay rate of correlations for smooth functions.

Estimates of the rate of convergence in the von Neumann theorem via power-law estimates of correlations. For a broad class of Anosov flows, the correlations can decay faster than at any polynomial rate [81]. For example, let a transitive Anosov flow have two periodic trajectories with periods l_1 and l_2 whose ratio is a Diophantine number; i.e., there exist constants $C,\alpha>0$ such that

$$\left|\frac{l_1}{l_2}-\frac{p}{q}\right|\ge\frac{C}{q^\alpha}$$

for all $p\in\mathbb{Z}$ and $q\in\mathbb{N}$. (For example, any irrational algebraic number is Diophantine. It is well known that Diophantine numbers form a set of full Lebesgue measure on the real line.) Then, as was proved in [81], for any functions $f,g\in C^\infty(\mathcal{M})$ and any $m\in\mathbb{N}$ there exists a constant $A=A(f,g,m)>0$ such that the correlations with respect to the SRB measure admit the estimate

$$|c_t(f,g)|\le At^{-m} \tag{6.10}$$

for all $t>0$.

By using relations (5.0) between the correlation coefficients, the estimate (6.10), and assertion 5 of Theorem 8 we obtain the asymptotic relation (2.2) with an estimate of the rate of convergence in the von Neumann theorem for the Anosov flows in question for any $f \in C^\infty(\mathcal{M})$ and any $m > 2$. Namely,

$$\big|\|\overline{A}_t f - \overline{f}^*\|_2^2 - 2\pi\rho(0)t^{-1} + 2\pi(\rho')^c(0)t^{-2}\big| \le \frac{2A}{m-2}t^{-m}$$

for all $t > 0$, where

$$0 \le 2\pi\rho(0) \le 2\|f - \overline{f}^*\|_2^2 + \frac{2A}{m-1}, \qquad |2\pi(\rho')^c(0)| \le 2\|f - \overline{f}^*\|_2^2 + \frac{2A}{m-2}.$$

We see that the decay is faster than power-law with any exponent, but we do not know whether an exponential decay takes place.

Estimates of the rate of convergence in the von Neumann theorem via exponential estimates of correlations. Now let us proceed to examples with an exponential decay of correlations.

Example 1 ([80])**.** Let $\mathcal{M}$ be a C^7 surface of negative curvature, and let T^t be the geodesic flow on its unit tangent bundle. For any Gibbs measure constructed from a Hölder function (including the SRB measure) there exist constants $A, B > 0$ such that the correlations of any C^5 functions f and g with respect to this measure satisfy the inequality

$$|c_t(f,g)| \le A\|f\|_{C^5}\|g\|_{C^5}e^{-Bt} \tag{6.11}$$

for all $t > 0$.

Example 2 ([80])**.** Now let T^t be a C^5 Anosov flow on a compact manifold $\mathcal{M}$ with C^1 stable fibers E_x^s and unstable fibers E_x^u. Then there exist constants $A, B > 0$ such that the correlations of any C^5 functions f and g with respect to the SRB measure satisfy the same estimate as in (6.11) for all $t > 0$.

Example 3 ([83])**.** Let $\mathcal{M}$ be a $(2d+1)$-dimensional connected compact Riemannian manifold, and let T^t be a C^4 contact Anosov flow on $\mathcal{M}$. The contact property of the flow T^t means that the manifold is equipped with the differential $(2d+1)$-form $\omega = \alpha \wedge (d\alpha)^d$, where α is a C^2 1-form such that

$$\omega \ne 0, \quad \alpha(D^tTv) = \alpha(v), \qquad v \in T\mathcal{M}.$$

The invariant measure is the volume induced by the form ω. It was shown in [83] that there exists a constant $\sigma > 0$ such that for any Hölder functions f and g with exponent $\alpha \in (0,1)$ there exists a constant $A = A_\alpha > 0$ for which the inequality

$$|c_t(f,g)| \le A\big(\|f\|_\infty + \text{Höld}_\alpha(f)\big)\big(\|g\|_\infty + \text{Höld}_\alpha(f)\big)e^{-Bt}, \qquad B = \frac{\alpha\sigma}{2-\alpha}, \tag{6.12}$$

holds for all $t > 0$.

By using the estimates (6.11) and (6.12), relations (5.0) between the correlation coefficients, and assertion 6 of Theorem 8, we obtain the asymptotic relation (2.2) with an estimate of the rate of convergence in the von Neumann theorem for Anosov flows in all the cases considered above. Namely,

$$\big|\|\overline{A}_t f - \overline{f}^*\|_2^2 - 2\pi\rho(0)t^{-1} + 2\pi(\rho')^c(0)t^{-2}\big| \le \frac{2A\|f\|^2}{B}\left(1 + \frac{1}{Bt}\right)\frac{e^{-Bt}}{t}$$

for each $t > 0$; here

$$0 \le 2\pi\rho(0) \le \frac{2A\|f\|^2}{B}, \qquad |2\pi(\rho')^c(0)| \le \frac{2A\|f\|^2}{B^2},$$

and we have $\|f\| = \|f\|_{C^5}$ in the first two examples for the case of $f \in C^5(\mathcal{M})$ and $\|f\| = \|f\|_\infty + \text{Höld}_\alpha(f)$ in the third example for a Hölder function f with exponent $\alpha \in (0,1)$.

6.3. **Dynamical systems modeled by a Young tower with an exponential tail** [9, 7, 6, 8]**.** Let Ω be a finite-dimensional Riemannian manifold, and let $T\colon \Omega \to \Omega$ be a dynamical system modeled by a Young tower with an exponential tail; i.e., condition (5.1) for the time of return to the base (zero floor) $\Delta_0 \subset \Omega$ is replaced with the condition

$$m^u\{y \in \Delta_0 \colon \overline{R}(y) > n\} \le C_R e^{-\alpha n} \quad \text{for some} \quad \alpha > 0 \quad \text{and} \quad C_R > 0 \quad \text{for all} \quad n \in \mathbb{N},$$

where m^u is the Lebesgue measure on the unstable manifolds of the uniformly hyperbolic set Δ_0.

Estimate of correlations. Let λ be the SRB measure for T (see Section 5.2); we denote the mean value of any function $f \in L_1(\Omega)$ by

$$\lambda(f) = \int_\Omega f\, d\lambda.$$

It was shown in [7, 6] that the decay of autocorrelations for Hölder functions f is exponential; namely, there exists a constant $C_f > 0$ and an exponent $\gamma > 0$ depending on the Hölder exponent of f such that the inequality

$$|c_n(f)| = \left| \int_\Omega f(\omega) f(T^n \omega)\, d\lambda - \left(\int_\Omega f(\omega)\, d\lambda \right)^2 \right| \le C_f \|f\|_\infty e^{-\gamma n}$$

holds for every $n \in \mathbb{N}$.

Estimate of the rate of convergence in the von Neumann theorem. By applying the last estimate to assertion 6 of Theorem 7, we obtain the asymptotic relation (2.1) with an estimate of the rate of convergence in the von Neumann theorem for nonuniformly hyperbolic dynamical systems modeled by a Young tower with an exponential tail for any Hölder function f. Namely,

$$\big| \|A_n f - \lambda(f)\|_2^2 - 2\pi\rho(0) n^{-1} + 2\pi(\rho')^c(0) n^{-2} \big| < 2C_f \|f\|_\infty \left(\frac{e^\gamma}{e^\gamma - 1} \right)^2 \frac{e^{-\gamma n}}{n} \tag{6.13}$$

for all $n \in \mathbb{N}$, where

$$0 \le 2\pi\rho(0) \le \|f\|_2^2 + \frac{2C_f \|f\|_\infty e^\gamma}{e^\gamma - 1}, \quad |2\pi(\rho')^c(0)| \le \frac{2C_f \|f\|_\infty e^\gamma}{(e^\gamma - 1)^2}.$$

Estimate of large deviations. As was proved in [8] (see also [9]), the following large deviation principle for an invariant measure λ holds for dynamical systems modeled by a Young tower with an exponential tail. For each Hölder function f, there exists a constant $\theta_0 > 0$ such that the function

$$e(\theta) = \lim_{n\to\infty} \frac{1}{n} \ln \lambda(e^{n\theta A_n f})$$

is real-analytic on the interval $(-\theta_0, \theta_0)$; further, $e'(0) = \lambda(f)$, and

$$\lim_{n\to\infty} \frac{1}{n} \ln \lambda\{|A_n f - \lambda(f)| \ge \varepsilon\} = -\inf_{t \in R_\varepsilon} I(t) \tag{6.14}$$

for each sufficiently small $\varepsilon > 0$, where

$$R_\varepsilon = (e'(-\theta_0), e'(0) - \varepsilon] \cup [e'(0) + \varepsilon, e'(\theta_0))$$

and

$$I(t) = \sup_{\theta \in [-\theta_0, \theta_0]} \theta t - e(\theta)$$

is the Legendre transform of the function $e(\theta)$. If f is cohomological to a constant $c \in \mathbb{R}$, then $e(\theta) = c\theta$ for every θ; relation (6.14) becomes trivial, because the infimum on the right-hand side is taken over the empty set and the large deviations p_n^ε on the left-hand side are zero for any $\varepsilon > 0$ for all positive integer $n > 2\|g\|_\infty/\varepsilon$. In this case, according to Remark 15, the rate of convergence in the Birkhoff theorem is trivial as well, $\mathrm{P}_n^\varepsilon = 0$ for each $\varepsilon > 0$ and for all positive integer $n > 2\|g\|_\infty/\varepsilon$. If the function f is not cohomological to a constant, then, as shown in [8], the function $e(\theta)$ is strictly convex (because $e''(\theta) > 0$ for all $\theta \in (-\theta_0, \theta_0)$); at the same time, for each $\varepsilon \in (0, \min\{e'(0)-e'(-\theta_0), e'(\theta_0)-e'(0)\})$ the function

$$\gamma_0(\varepsilon) = \inf_{t \in R_\varepsilon} I(t),$$

which depends on ε and f, is positive and does not take infinite values.

It follows from (6.14) that for each Hölder function f that is not cohomological to a constant, each sufficiently small $\varepsilon > 0$, and each $\gamma \in (0, \gamma_0(\varepsilon))$ there exists a constant $C_1 > 0$ depending on ε, γ, and f such that the inequality $\mathrm{p}_n^\varepsilon \le C_1 e^{-\gamma n}$ holds for all $n \ge 1$.

Estimates of the rate of convergence in the Birkhoff theorem. By applying Theorem 13 to the last inequality, we obtain and estimate the rate of convergence in the Birkhoff theorem for nonuniformly hyperbolic dynamical systems modeled by a Young tower with an exponential tail for any Hölder function f noncohomological to a constant, any sufficiently small $\varepsilon > 0$, and any $\gamma \in (0, \gamma_0(\varepsilon))$. Namely,

$$\mathrm{P}_n^{2\varepsilon} \le C_1\left(1 + \frac{\ln(1+\gamma^{-1})}{\ln(1+\varepsilon\|f-\lambda(f)\|_\infty^{-1})}\right)e^{-\gamma n} \tag{6.15}$$

for all $n \ge 1$.

Now consider the characteristic function χ_E of a bounded Borel set E with regular boundary (i.e., with boundary of zero λ-measure). By the same computations as in the proof of the estimate (6.8), we also obtain an estimate for the rate of convergence in the Birkhoff theorem for this function,

$$\mathrm{P}_n^{4\varepsilon}(\chi_E) \le C_1(\varepsilon)\left(1 + \frac{\ln(1+\gamma^{-1}(\varepsilon))}{\ln(1+2\varepsilon\|\chi_E-\lambda(E)\|_\infty^{-1})}\right)e^{-\gamma(\varepsilon)n}, \tag{6.16}$$

which holds for sufficiently small $\varepsilon > 0$ with some constants $C_1(\varepsilon), \gamma(\varepsilon) > 0$ for all $n \ge 1$.

Remark 21. Note that there exist estimates of correlations in nonuniformly hyperbolic systems for non-Hölder continuous functions (e.g., see [84, 85, 86]). In this case, one can also obtain estimates for such functions in the von Neumann theorem with the use of Theorem 5.

6.3.1. *Quadratic maps.* Let $a \in (0, 2]$. Consider the family of maps $T_a\colon [-1,1] \to [-1,1]$ acting by the rule $T_a(x) = 1 - ax^2$. This family, as well as the family of maps $\widetilde{T}_a\colon [0,1] \to [0,1]$ acting by the rule $\widetilde{T}_a(x) = 2ax(1-x)$, is known as the family of logistic maps. The functions T_2 and $\widetilde{T}_2$ are called the Ulam–von Neumann maps (parabolas).

It is well known (e.g., see [87]) that there exists an $\varepsilon > 0$ such that for all parameters a in the Benedicks–Carleson set $\Delta_\varepsilon \subseteq [2-\varepsilon, 2]$ of positive Lebesgue measure the dynamical system in question has an absolutely continuous invariant (SRB) measure λ_a. Further, $2 \in \Delta_\varepsilon$, and the measure λ_2 has density $1/(\pi\sqrt{x(1-x)})$ for $\widetilde{T}_2$ and density $1/(\pi\sqrt{(1-x^2)})$ for T_2. For all parameter values $a \in \Delta_\varepsilon$, the logistic map is modeled by a Yang tower with an exponential tail [6]. Hence the estimate (6.13) of the rate of convergence in the von Neumann theorem holds for each Hölder function f, and the exponential estimates (6.15) and (6.16) of the rate of convergence in the Birkhoff theorem are true for Hölder functions and for the characteristic functions of bounded Borel sets with boundary of zero measure, respectively.

Estimates of correlations for functions of bounded variation. An estimate of the rate of convergence in the von Neumann theorem for the logistic map can also be obtained for functions f of bounded variation. The decay of correlations for such functions with respect to the measure λ_a will also be exponential for $a \in \Delta_\varepsilon$ (see [87]), because for the map T_a there exists a constant $B > 0$ such that for any functions f and g of bounded variation there exists a constant $A = A_f > 0$ with the property that

$$|c_n(f, g)| \leq A_f \|g\|_\infty e^{-Bn}$$

for all $n \in \mathbb{N}$. A similar result was obtained in [88] for the maps $\widetilde{T}_a$, $a \in \Delta_\varepsilon$: there exist constants $A, B > 0$ such that, for any function f of bounded variation and any function $g \in L_{2+\delta}[0, 1]$ with some $\delta > 0$, one has

$$|c_n(f, g)| \leq A \operatorname{var}(f) \|g\|_{2+\delta} e^{-Bn}$$

for all $n \in \mathbb{N}$.

Estimate of the rate of convergence in the von Neumann theorem. By using the resulting estimates for correlations, relations (5.0) between the correlation coefficients, and assertion 6 of Theorem 7, we obtain the asymptotic relation (2.1) with an estimate of the rate of convergence in the von Neumann theorem for the logistic map with parameter $a \in \Delta_\varepsilon$ for any function f of bounded variation; namely,

$$\big| \|A_n f - f^*\|_2^2 - 2\pi\rho(0)n^{-1} + 2\pi(\rho')^c(0)n^{-2} \big| \leq 2\widetilde{A}\|f\|_\infty \left(\frac{e^B}{e^B - 1} \right)^2 \frac{e^{-Bn}}{n},$$

$$\widetilde{A} = A_f \quad \text{or} \quad \widetilde{A} = A \operatorname{var}(f),$$

for all $n \in \mathbb{N}$, where

$$0 \leq 2\pi\rho(0) \leq \|f\|_2^2 + \frac{2\widetilde{A}e^B}{e^B - 1}, \qquad |2\pi(\rho')^c(0)| \leq \frac{2\widetilde{A}e^B}{(e^B - 1)^2}.$$

Note that, for all parameter values $a \in \Delta_\varepsilon$, an exponential estimate of the rate of convergence almost everywhere with respect to the (noninvariant) Lebesgue measure holds for functions f of bounded variation in view of Remark 14 applied to the exponential estimates obtained in [88] for large deviations with respect to the Lebesgue measure for functions of bounded variation.

6.3.2. *Hénon maps.* Consider the two-parameter family of maps $T_{a,b} \colon \mathbb{R}^2 \to \mathbb{R}^2$ acting by the rule $T_{a,b} = (1 - ax^2 - y, bx)$ with parameters $(a, b) \in [2 - \varepsilon, 2) \times (0, \delta)$ for some given sufficiently small ε and $\delta > 0$. It is well known (e.g., see [89]) that there exist quite a few parameter values (a, b) of positive plane Lebesgue measure for which the maps $T_{a,b}$ have an invariant (SRB) measure $\lambda_{a,b}$. The decay of correlations and large deviations with respect to this measure will be exponential. This claim follows for the dynamical systems in question from the construction of a Young tower with an exponential tail in [89] modeling these systems. Consequently, the estimates (6.13) in the von Neumann theorem and (6.15) and (6.16) in the Birkhoff theorem for the rate of convergence, common in this section, also hold for the Hénon map with the indicated parameters.

6.3.3. *Billiards with an exponential estimate of the rate of convergence* [90, 91, 92]. Of the Sinai billiards, consider the periodic Lorentz gas on the plane [3, Secs. 6.5 and 9.1], which is popular and important in applications. This is a well-known model of an ideal electron gas in statistical physics, where the heavy ions are immovable and form a regular lattice on the plane, while the lighter electrons move, elastically reflect from the ions, and do not interact with each other (so that one may well consider the motion of a single electron).

The domain $\mathcal{M}$ in which this billiard is considered in the unit two-dimensional torus from which finitely many convex domains (modeling fixed ions) with C^3 boundary of nonzero curvature have been cut out. The billiard map (see the definition in Section 5.2.2) is modeled by a Young tower with an exponential tail (see the discussion in [91]), and so it also obeys the estimates, common in this section, of the rate of convergence in the von Neumann theorem (equation (6.13)), in the Birkhoff theorem for Hölder functions (equation (6.15)), and in the Birkhoff theorem for the characteristic functions of bounded Borel sets with boundary of zero measure (equation (6.16)).

However, we present estimates of the rate of convergence in the von Neumann theorem based on the more accurate estimates (than the estimates (6.13) resulting from modeling by the Young tower) for the decay rate of correlations obtained in [92] by the coupling method (mentioned in Section 6.2.1 above in the discussion of the estimates of decay rates of correlations for Anosov diffeomorphisms).

Estimate of correlations. It follows from [92, Theorem 4.3] that for any Hölder functions f and g on $\partial\mathcal{M} \times [-\pi/2, \pi/2]$ with exponents $\alpha \in (0,1)$ and $\beta \in (0,1)$, respectively, the inequality

$$|c_n(f,g)| \le A\big(\text{Höld}_\alpha(f)B^\alpha\|g\|_\infty + \text{Höld}_\beta(g)B^\beta\|f\|_\infty + \|f\|_\infty\|g\|_\infty\big)\theta_{f,g}^n$$

holds for all $n \in \mathbb{N}$, where $\theta_{f,g} = (\max\{\vartheta, \Lambda^{-\alpha}, \Lambda^{-\beta}\})^{1/4} < 1$ and the constants $A, B > 0$, $\Lambda > 1$, and $0 < \vartheta < 1$ depend on $\mathcal{M}$ alone. Note that estimates of the correlation coefficients were obtained in [92] for functions f and g in a class of bounded functions (referred to as dynamically Hölder functions there) wider than the Hölder class; see also the paper [30], which deals with an even wider class of observable functions satisfying an estimate for correlations (from which a central limit theorem is derived there) similar to the one written out above.

Estimate of the rate of convergence in the von Neumann theorem. By using the inequality written out for the correlations, relations (5.0) between the correlation coefficients, and assertion 6 of Theorem 7, we obtain the asymptotic relation (2.1) with an estimate of the rate of convergence in the von Neumann theorem for a periodic Lorentz gas on the plane for any Hölder function f with exponent $\alpha \in (0,1)$. Namely,

$$\big|\|A_nf - f^*\|_2^2 - 2\pi\rho(0)n^{-1} + 2\pi(\rho')^c(0)n^{-2}\big| < 2C_f\|f\|_\infty\left(\frac{1}{1-\theta_{f,f}}\right)^2\frac{\theta_{f,f}^n}{n},$$
$$C_f = A\big(2\text{Höld}_\alpha(f)B^\alpha + \|f\|_\infty\big),$$

for all $n \in \mathbb{N}$, where

$$0 \le 2\pi\rho(0) \le \|f\|_2^2 + \frac{2C_f\|f\|_\infty\theta_{f,f}}{1-\theta_{f,f}}, \qquad |2\pi(\rho')^c(0)| \le \frac{2C_f\|f\|_\infty\theta_{f,f}}{(1-\theta_{f,f})^2}.$$

6.4. **Teichmüller flows** [93, 94, 95]. Let an integer $g \ge 2$ and a vector $\vec{k} = (k_1, k_2, \dots, k_\sigma)$ of positive integers satisfy $\sum_{j=1}^\sigma k_j = 2g-2$. Let $\mathcal{M}_{\vec{k}}$ be the moduli space of abelian differentials, i.e., the set of pairs $w = (\mathcal{M}, \omega)$, where $\mathcal{M}$ is a compact orientable C^∞ Riemannian surface of genus g and ω is a holomorphic 1-form with zeros of orders $k_1, k_2, \dots, k_\sigma$ such that $\frac{1}{2i}\int_\mathcal{M} \omega \wedge \overline{\omega} = 1$. Let H be a connected component of $\mathcal{M}_{\vec{k}}$.

The Teichmüller flow T^t is defined on H by the formula

$$T^t(\mathcal{M}, \omega) = (\mathcal{M}', \omega'),$$

where the form $\omega' = e^t\,\text{Re}(\omega) + ie^{-t}\,\text{Im}(\omega)$ and the structure on $\mathcal{M}'$ are determined by the holomorphy of ω'. Let $\mu_{\vec{k}}$ be a T^t-invariant smooth Masur–Veech probability measure on H, and let d_H be the Finsler metric on H [93]. For $w = (\mathcal{M}, \omega)$, let $\mathfrak{S}(w)$ be the shortest distance (in the Riemannian metric $\mathcal{M}$) between zeros of the form ω.

For any $\alpha \in (0,1)$ and any function $f\colon H \to \mathbb{R}$, set

$$H_\alpha(f,w) = \sup_{0<d_H(w',w)\le 1}\frac{|f(w)-f(w')|}{d_H^\alpha(w,w')}.$$

For any $\alpha \in (0,1)$ and $k \in \mathbb{N}$, let $\mathfrak{D}_{k,\alpha}$ be the set of functions $f\colon H \to \mathbb{R}$ with finite norm

$$\|f\|_{k,\alpha} = \sup_{w\in H} |f(w)|\mathfrak{S}^k(w) + \sup_{w\in H} H_\alpha(f,w)\mathfrak{S}^k(w).$$

Estimates of correlations and the rate of convergence in the von Neumann theorem. It was shown in [93] that there exist constants $A, B > 0$ (depending on k, α, p, and q) such that for any functions $f \in \mathfrak{D}_{k,\alpha} \cap L_p(H, \mu_{\vec{k}})$ and $g \in \mathfrak{D}_{k,\alpha} \cap L_q(H, \mu_{\vec{k}})$, $p, q > 0$, $1/p + 1/q = 1$, one has

$$|c_t(f,g)| \le A(\|f\|_{k,\alpha} + \|f\|_p)(\|g\|_{k,\alpha} + \|g\|_q)e^{-Bt} \tag{6.17}$$

for all $t \ge 0$.

By using inequality (6.17), relations (5.0) between the correlation coefficients, and assertion 6 of Theorem 8, we obtain the asymptotic relation (2.2) with an estimate of the rate of convergence in the von Neumann theorem for Teichmüller flows for any $f \in \mathfrak{D}_{k,\alpha} \cap L_2(H, \mu_{\vec{k}})$. Namely,

$$\big|\|\overline{A}_t f - \overline{f}^*\|_2^2 - 2\pi\rho(0)t^{-1} + 2\pi(\rho')^c(0)t^{-2}\big| \le \frac{2A(\|f\|_{k,\alpha} + \|f\|_2)^2}{B}\left(1 + \frac{1}{Bt}\right)\frac{e^{-Bt}}{t}$$

for all $t > 0$, where

$$0 \le 2\pi\rho(0) \le \frac{2A(\|f\|_{k,\alpha} + \|f\|_2)^2}{B}, \qquad |2\pi(\rho')^c(0)| \le \frac{2A(\|f\|_{k,\alpha} + \|f\|_2)^2}{B^2}.$$

Estimates of large deviations and the rate of convergence in the Birkhoff theorem. Let $f\colon H \to \mathbb{R}$ be a bounded function with the following properties: f is a Hölder function in the sense of Veech (see [95]),

$$\int_H f\, d\mu_{veck} = 0,$$

and

$$\int_0^\tau f(T^t z)\, dt \ne 0$$

for some periodic point $z \in H$ of period $\tau > 0$. Then, as shown in [95], one has

$$\overline{\lim_{t\to\infty}} \frac{1}{t} \ln \mu_{\vec{k}}\{z \in H\colon |\overline{A}_t f(z)| \ge \varepsilon\} = -\gamma(\varepsilon) < 0$$

for every $\varepsilon > 0$. From this, we see that for each $\gamma \in (0, \gamma(\varepsilon))$ there exists a constant $C > 0$ depending on γ, ε, and f such that

$$\mu_{\vec{k}}\{|\overline{A}_t f| \ge \varepsilon\} \le Ce^{-\gamma t}$$

for all $t > 0$. By using inequality 3 in Theorem 13, we obtain the following estimate of the rate of convergence in the Birkhoff theorem for Teichmüller flows for the functions f considered: for any $\varepsilon > 0$ and $t > 0$, one has

$$\mu_{\vec{k}}\{\sup_{s\ge t} |\overline{A}_s f| \ge 2\varepsilon\} \le C\left(1 + \frac{\ln(1 + 1/\gamma)}{\ln(1 + \varepsilon\|f\|_\infty^{-1})}\right)e^{-\gamma t}.$$

References

[1] J. von Neumann, *Physical applications of the ergodic hypothesis*, Proc. Nat. Acad. Sci. USA **18** (1932), no. 3, 263–266.

[2] A. G. Kachurovskii, *Rates of convergence in ergodic theorems*, Uspekhi Mat. Nauk **51** (1996), no. 4(310), 73–124; English transl., Russian Math. Surveys **51** (1996), no. 4, 653–703. MR1422228

[3] I. P. Kornfeld, Ya. G. Sinai, and S. V. Fomin, *Ergodic theory*, Nauka, Moscow, 1980; English transl., Springer, Berlin–Heidelberg–New York, 1982. MR610981

[4] A. G. Kachurovskii and A. V. Reshetenko, *On the rate of convergence in von Neumann's ergodic theorem with continuous time*, Mat. Sb. **201** (2010), no. 4, 25–32; English transl., Sb. Math. **201** (2010), no. 3–4, 493–500. MR2675340
[5] Ya. G. Sinai, *Ergodic theory of smooth dynamical systems*, Chap. 6. *Stochasticity of smooth dynamical systems. Elements of KAM theory*, Current problems in mathematics, Fundamental directions, vol. 2, Dynamical systems–2, Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow, 1985, 115–122. (Russian)
[6] L.-S. Young, *Statistical properties of dynamical systems with some hyperbolicity*, Ann. of Math. (2) **147** (1998), no. 3, 585–650. MR1637655
[7] L.-S. Young, *Recurrence times and rates of mixing*, Israel J. Math. **110** (1999), 153–188. MR1750438
[8] L. Rey-Bellet and L.-S. Young, *Large deviations in non-uniformly hyperbolic dynamical systems*, Ergodic Theory Dynam. Systems **28** (2008), no. 2, 587–612. MR2408394
[9] I. Melbourne and M. Nicol, *Large deviations for nonuniformly hyperbolic systems*, Trans. Amer. Math. Soc. **360** (2008), no. 12, 6661–6676. MR2434305
[10] I. Melbourne, *Large and moderate deviations for slowly mixing dynamical systems*, Proc. Amer. Math. Soc. **137** (2009), no. 5, 1735–1741. MR2470832
[11] V. P. Leonov, *On the dispersion of time means of a stationary stochastic process*, Teor. Veroyatnost. Primenen. **6** (1961), no. 1, 93–101; English transl., Theory Probab. Appl. **6** (1961), no. 1, 87–93. MR0133173
[12] Yu. K. Belyaev, *An example of a process with mixing*, Teor. Veroyatnost. Primenen. **6** (1961), no. 1, 101–102; English transl., Theory Probab. Appl. **6** (1961), no. 1, 93–94.
[13] A. G. Kachurovskii and V. V. Sedalishchev, *On the constants in the estimates for the rate of convergence in von Neumann's ergodic theorem*, Mat. Zametki **87** (2010), no. 5, 756–763; English transl., Math. Notes **87** (2010), no. 5–6, 720–727. MR2766588
[14] A. G. Kachurovskii and V. V. Sedalishchev, *Constants of estimates for the rate of convergence in the von Neumann and Birkhoff ergodic theorems*, Mat. Sb. **202** (2011), no. 8, 21–40; English transl., Sb. Math. **202** (2011), no. 7–8, 1105–1125. MR2866197
[15] N. A. Dzhulai and A. G. Kachurovskii, *Constants in estimates for the rate of convergence in von Neumann's ergodic theorem with continuous time*, Sibirsk. Mat. Zh. **52** (2011), no. 5, 1039–1052; English transl., Sib. Math. J. **52** (2011), no. 5, 824–835. MR2908125
[16] A. G. Kachurovskii and V. V. Sedalishchev, *On the constants in the estimates of the rate of convergence in the Birkhoff ergodic theorem*, Mat. Zametki **91** (2012), no. 4, 624–628; English transl., Math. Notes **91** (2012), no. 3–4, 582–587. MR3201463
[17] V. V. Sedalishchev, *Constants in estimates for the rate of convergence in Birkhoff's ergodic theorem with continuous time*, Sibirsk. Mat. Zh. **53** (2012), no. 5, 1102–1110; English transl., Sib. Math. J. **53** (2012), no. 5, 882–888. MR3057930
[18] V. V. Sedalishchev, *The relation between rates of convergence in the von Neumann and Birkhoff ergodic theorems in L_p*, Sibirsk. Mat. Zh. **55** (2014), no. 2, 412–426; English transl., Sib. Math. J. **55** (2014), no. 2, 336–348. MR3237344
[19] A. G. Kachurovskii and I. V. Podvigin, *Large deviations and the rate of convergence in the Birkhoff ergodic theorem*, Mat. Zametki **94** (2013), no. 4, 569–577; English transl., Math. Notes **94** (2013), no. 3–4, 524–531. MR3423283
[20] A. G. Kachurovskii and I. V. Podvigin, *Convergence rates in ergodic theorems for some billiards and Anosov diffeomorphisms*, Dokl. Akad. Nauk **451** (2013), no. 1, 11–13; English transl., Dokl. Math. **88** (2013), no. 1, 385–387. MR3155153
[21] A. G. Kachurovskii and I. V. Podvigin, *Rates of convergence in ergodic theorems for the periodic Lorentz gas on the plane*, Dokl. Akad. Nauk **455** (2014), no. 1, 11–14; English transl., Dokl. Math. **89** (2014), no. 2, 139–142. MR3237613
[22] A. G. Kachurovskii, *On the convergence of averages in the ergodic theorem for the groups $\mathbb{Z}^d$*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **256** (1999), 121–128; English transl., J. Math. Sci. (New York) **107** (2001), no. 5, 4231–4236. MR1708562
[23] A. M. Vershik and A. G. Kachurovskii, *Rates of convergence in ergodic theorems for locally finite groups, and reversed martingales*, Differ. Uravn. Protsessy Upr. (1999), no. 1, 19–26. (Russian) MR1890198
[24] A. G. Kachurovskii, *On uniform convergence in the ergodic theorem*, J. Math. Sci. (New York) **95** (1999), no. 5, 2546–2551. MR1712742
[25] A. N. Shiryaev, *Probability*, Nauka, Moscow, 1989; English transl., Springer, New York, 1995. MR1024077
[26] C. Cuny and M. Lin, *Pointwise ergodic theorems with rate and application to the CLT for Markov chains*, Ann. Inst. Henri Poincaré Probab. Stat. **45** (2009), no. 3, 710–733. MR2548500

[27] A. Gomilko, M. Haase, and Yu. Tomilov, *On rates in mean ergodic theorems*, Math. Res. Lett. **18** (2011), no. 2, 201–213. MR2784667

[28] A. Gomilko, M. Haase, and Yu. Tomilov, *Bernstein functions and rates in mean ergodic theorems for operator semigroups*, J. Anal. Math. **118** (2012), no. 2, 545–576. MR3000691

[29] M. I. Gordin, *The central limit theorem for stationary processes*, Dokl. Akad. Nauk SSSR **188** (1969), no. 4, 739–741; English transl., Sov. Math. Dokl. **10** (1969), 1174–1176. MR0251785

[30] M. Stenlund, *A strong pair correlation bound implies the CLT for Sinai billiards*, J. Stat. Phys. **140** (2010), no. 1, 154–169. MR2651443

[31] U. Krengel, *Ergodic theorems*, de Gruyter Stud. in Math., vol. 6, Walter de Gruyter, Berlin, 1985.

[32] N. K. Bari, *A treatise on trigonometric series*, Fizmatgiz, Moscow, 1961; English transl., vol. I, II, Pergamon Press, Oxford–London–New York–Paris–Frankfurt, 1964. MR0171116

[33] J. von Neumann, *Proof of the quasi-ergodic hypothesis*, Proc. Nat. Acad. Sci. USA **18** (1932), no. 1, 70–82.

[34] I. A. Ibragimov and Yu. V. Linnik, *Independent and stationary sequences of random variables*, Nauka, Moscow, 1965; English transl., Wolters–Noordhoff Publishing Company, Groningen, 1971. MR0322926

[35] V. F. Gaposhkin, *On the rate of decrease of the probabilities of ε-deviations for means of stationary processes*, Mat. Zametki **64** (1998), no. 3, 366–372; English transl., Math. Notes **64** (1998), no. 3–4, 316–321 (1999). MR1680150

[36] V. F. Gaposhkin, *Estimates of means for almost all realizations of stationary processes*, Sibirsk. Mat. Zh. **20** (1979), no. 5, 978–989; English transl., Sib. Math. J. **20** (1980), no. 5, 691–699. MR559060

[37] V. F. Gaposhkin, *Convergence of series connected with stationary sequences*, Izv. Akad. Nauk SSSR Ser. Mat. **39** (1975), no. 6, 1366–1392; English transl., Math. USSR-Izv. **9** (1975), no. 6, 1297–1321. MR0402880

[38] F. Browder, *On the iterations of transformations in noncompact minimal dynamical systems*, Proc. Amer. Math. Soc. **9** (1958), no. 5, 773–780. MR0096975

[39] A. Zygmund, *Trigonometric series*, vol. 1, At the University Press, Cambridge, 1959. MR0107776

[40] I. Assani and M. Lin, *On the one-sided Hilbert transform*, Ergodic Theory and Related Fields, Contemp. Math., vol. 430, Amer. Math. Soc., Providence, RI, 2007, 21–39. MR2331323

[41] R. E. Edwards, *Fourier series. A modern introduction*, vol. 1, 2, Springer-Verlag, New York–Heidelberg–Berlin, 1979, 1982. MR545506

[42] A. Ya. Helemskii, *Lectures and exercises on functional analysis*, MCCME, Moscow, 2004; English transl., Transl. of Math. Monographs, vol. 233, Amer. Math. Soc., Providence, RI, 2006. MR2248303

[43] V. F. Gaposhkin, *Some examples concerning the problem of ε-deviations for stationary sequences*, Teor. Veroyatnost. Primenen. **46** (2001), no. 2, 370–375; English transl., Theory Probab. Appl. **46** (2003), no. 2, 341–346. MR1968692

[44] V. I. Bogachev, *Measure theory*, vol. 1, Regul. i Khaotich. Dinamika, Moscow–Izhevsk, 2003; English transl., Springer-Verlag, Berlin–Heidelberg, 2007. MR2267655

[45] B. A. Dubrovin, S. P. Novikov, and A. T. Fomenko, *Modern geometry. Methods and applications*, Nauka, Moscow, 1986; English transl., parts 1–3, Springer-Verlag, New York, 1992, 1985, 1990. MR864355

[46] M. L. Blank, *Stability and localization in chaotic dynamics*, MCCME, Moscow, 2001. (Russian)

[47] E. Lesigne and D. Volný, *Large deviations for generic stationary processes*, Colloq. Math. **84/85** (2000), part 1, 75–82. MR1778841

[48] O. Sarig, *Decay of correlations*, Handbook of Dynamical Systems **1** (2006), part B, 244–263.

[49] L.-S. Young, *What are SRB measures, and which dynamical systems have them?*, J. Stat. Phys. **108** (2002), no. 5–6, 733–754. MR1933431

[50] C. Liverani, B. Saussol, and S. Vaienty, *A probabilistic approach to intermittency*, Ergodic Theory Dynam. Systems **19** (1999), no. 3, 671–685. MR1695915

[51] H. Hu, *Decay of correlations for piecewise smooth maps with indifferent fixed points*, Ergodic Theory Dynam. Systems **24** (2004), no. 2, 495–524. MR2054191

[52] M. Pollicott, R. Sharp, and M. Yuri, *Large deviations for maps with indifferent fixed points*, Nonlinearity **11** (1998), no. 4, 1173–1184. MR1632614

[53] M. Pollicott and R. Sharp, *Large deviations for intermittent maps*, Nonlinearity *22* (2009), no. 9, 2079–2092. MR2534293

[54] O. Sarig, *Subexponential decay of correlations*, Invent. Math. **150** (2002), no. 3, 629–653. MR1946554

[55] S. Gouëzel, *Sharp polynomial estimates for the decay of correlations*, Israel J. Math. **139** (2004), 29–65. MR2041223

[56] R. Markarian, *Billiards with polynomial decay of correlations*, Ergodic Theory Dynam. Systems **24** (2004), no. 1, 177–197. MR2041267

[57] N. Chernov and H.-K. Zhang, *Billiards with polynomial mixing rates*, Nonlinearity **18** (2005), no. 4, 1527–1553. MR2150341

[58] N. Chernov and R. Markarian, *Dispersing billiards with cusps: slow decay of correlations*, Comm. Math. Phys. **270** (2007), no. 3, 727–758. MR2276463

[59] N. Chernov and H.-K. Zhang, *Improved estimates for correlations in billiards*, Comm. Math. Phys. **277** (2008), no. 2, 305–321. MR2358286

[60] N. Chernov and R. Markarian, *Chaotic billiards*, Math. Surveys and Monographs, vol. 127, Amer. Math. Soc., Providence, RI, 2006. MR2229799

[61] S. Troubetzkoy, *Approximation and billiards*, Dynamical systems and Diophantine approximation, Sémin. congr., vol. 19, Soc. Math. France, Paris, 2009, 173–185. MR2808408

[62] P. Balint and I. Melbourne, *Decay of correlations and invariance principle for dispersing billiards with cusps, and related planar billiard flows*, J. Stat. Phys. **133** (2008), no. 3, 435–447. MR2448631

[63] V. M. Anikin and A. F. Golubentsev, *Analytical models of deterministic chaos*, Fizmatlit, Moscow, 2007. (Russian)

[64] V. M. Anikin, S. S. Arkadakskii, and A. S. Remizov, *An analytical solution to the spectral problem for the Perron-Frobenius operator of piecewise linear chaotic maps*, Izv. Vyssh. Uchebn. Zaved. Prikl. Nelinein. Din. **14** (2006), no. 2, 16–34.

[65] W. Parry, *On the β-expansions of real numbers*, Acta Math. Acad. Sci. Hungar. **11** (1960), no. 3–4, 401–416. MR0142719

[66] D. Mayer and G. Roepstorff, *On the relaxation time of Gauss's continued–fraction map I. The Hilbert space approach (Koopmanism)*, J. Stat. Phys. **47** (1987), no. 1–2, 149–171. MR892927

[67] D. Mayer and G. Roepstorff, *On the relaxation time of Gauss's continued–fraction map II. The Banach space approach (Transfer operator method)*, J. Stat. Phys. **50** (1988), no. 1–2, 331–344. MR939491

[68] I. Antoniou and S. Shkarin, *Analyticity of smooth eigenfunctions and spectral analysis of the Gauss map*, J. Stat. Phys. **111** (2003), no. 1–2, 355–369. MR1964274

[69] M. Iosifescu and C. Kraaikamp, *Metrical theory and continued fractions*, Mathematics and Its Applications, vol. 547, Kluwer Acad. Publ., Dordrecht, 2002. MR1960327

[70] S. Waddington, *Large deviation asymptotics for Anosov flows*, Ann. Inst. H. Poincaré Anal. Non Linéaire **13** (1996), no. 4, 445–484. MR1404318

[71] X. Bressaud and C. Liverani, *Anosov diffeomorphisms and coupling*, Ergodic Theory Dynam. Systems **22** (2002), no. 1, 129–152. MR1889567

[72] R. Bowen, *Methods of symbolic dynamics*, Mathematics, vol. 13, Mir, Moscow, 1979. (Russian)

[73] L. Orey and S. Pelikan, *Deviation of trajectory averages and the defect in Pesin's formula for Anosov diffeomorphisms*, Trans. Amer. Math. Soc. **315** (1989), no. 2, 741–753. MR935534

[74] Y. Kifer, *Large deviations in dynamical systems and stochastic processes*, Trans. Amer. Math. Soc. **321** (1990), no. 2, 505–524. MR1025756

[75] M. Pollicott and R. Sharp, *Large deviations, fluctuations and shrinking intervals*, Comm. Math. Phys. **290** (2009), no. 1, 321–334. MR2520516

[76] L.-S. Young, *Large deviations in dynamical systems*, Trans. Amer. Math. Soc. **318** (1990), no. 2, 525–543. MR975689

[77] S. Smale, *Differentiable dynamical systems*, Bull. Amer. Math. Soc. **73** (1967), 747–817. MR0228014

[78] A. G. Kachurovskii and I. V. Podvigin, *Estimates of the rate of convergence in the Birkhoff and Bowen theorems for Anosov flows*, Vestnik Kemerov. Univ. **47** (2011), no. 3/1, 255–258. (Russian)

[79] N. I. Chernov, *Markov approximations and decay of correlations for Anosov flows*, Ann. of Math. (2) **147** (1998), no. 2, 269–324. MR1626741

[80] D. Dolgopyat, *On decay of correlations in Anosov flows*, Ann. of Math. (2) **147** (1998), no. 2, 357–390. MR1626749

[81] D. Dolgopyat, *Prevalence of rapid mixing in hyperbolic flows*, Ergodic Theory Dynam. Systems **18** (1998), no. 5, 1097–1114. MR1653299

[82] D. Dolgopyat, *Prevalence of rapid mixing. II. Topological prevalence*, Ergodic Theory Dynam. Systems **20** (2000), no. 4, 1045–1059. MR1779392

[83] C. Liverani, *On contact Anosov flows*, Ann. of Math. (2) **159** (2004), no. 3, 1275–1312. MR2113022

[84] K. Díaz-Ordaz, *Decay of correlations for non-Hölder observables for one-dimensional expanding Lorenz-like maps*, Discrete Contin. Dyn. Syst. **15** (2006), no. 1, 159–176. MR2191390

[85] V. Lynch, *Decay of correlations for non-Hölder observables*, Discrete Contin. Dyn. Syst. **16** (2006), no. 1, 19–46. MR2221734

[86] H.-K. Zhang, *Decay of correlations on non-Hölder observables*, Int. J. Nonlinear Sci. **10** (2010), no. 3, 359–369. MR2755057

[87] L.-S. Young, *Decay of correlations for certain quadratic maps*, Comm. Math. Phys. **146** (1992), no. 1, 123–138. MR1163671

[88] G. Keller and T. Nowicki, *Spectral theory, zeta functions and the distribution of periodic points for Collet–Eckmann maps*, Comm. Math. Phys. **149** (1992), no. 1, 31–69. MR1182410

[89] M. Benedicks and L.-S. Young, *Markov extensions and decay of correlations for certain Hénon maps*, Astérisque **261** (2000), 13–56. MR1755436

[90] N. Chernov, *Decay of correlations and dispersing billiards*, J. Stat. Phys. **94** (1999), no. 3–4, 513–556. MR1675363

[91] N. Chernov and L.-S. Young, *Decay of correlations for Lorentz gases and hard balls*, Encyclopaedia Math. Sci., vol. 101, Springer, Berlin, 2000, 89–120. MR1805327

[92] N. Chernov, *Advanced statistical properties of dispersing billiards*, J. Stat. Phys. **122** (2006), no. 6, 1061–1094. MR2219528

[93] A. Avila, S. Gouëzel, and J.-C. Yoccoz, *Exponential mixing for the Teichmüller flow*, Publ. Math. Inst. Hautes Études Sci. **104** (2006), 143–211. MR2264836

[94] J. S. Athreya, *Quantitative recurrence and large deviations for Teichmüller geodesic flow*, Geom. Dedicata **119** (2006), no. 1, 121–140. MR2247652

[95] V. Araujo and A. Bufetov, *A large deviations bound for the Teichmüller flow on the moduli space of abelian differentials* Ergodic Theory Dynam. Systems **31** (2011), no. 4, 1043–1071. MR2818685

Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
E-mail address: agk@math.nsc.ru

Faculty of Physics, Novosibirsk State University, Novosibirsk, Russia
E-mail address: ivan_podvigin@ngs.ru

Translated by V. E. NAZAIKINSKII
Originally published in Russian