Bernoulli **23**(4A), 2017, 2129–2180
DOI: 10.3150/15-BEJ778

Convergence rate of the powers of an operator. Applications to stochastic systems

BERNARD DELYON

IRMAR, Campus de Beaulieu, 35042 Rennes cedex, France. E-mail: bernard.delyon@univ-rennes1.fr

We extend the traditional operator theoretic approach for the study of dynamical systems in order to handle the problem of non-geometric convergence. We show that the probabilistic treatment developed and popularized under Richard Tweedie's impulsion, can be placed into an operator framework in the spirit of Yosida–Kakutani's approach. General theorems as well as specific results for Markov chains are given. Application examples to general classes of Markov chains and dynamical systems are presented.

Keywords: Markov chains

1. Introduction

This paper is mainly concerned with the asymptotic behavior of homogeneous Markov chains, that is, processes of the form

$$X_{n+1} = \varphi(X_n, U_n), \tag{1}$$

where U_n is an i.i.d. sequence and φ a certain function; the initial condition X_0 is deterministic or random. There exist schematically two different approaches for the analysis of the asymptotic behavior of such systems: the *operator theoretic* approach and the *probabilistic* approach. In simple words, the second approach considers Harris chains where, typically, total variation convergence to the invariant measure in expected to occur, while the first one is a more general approach which captures the behaviour of chains with weaker mixing properties; we can notice that, at first sight, total variation convergence is more natural in the sense that the transition operator is actually by definition a contraction for the total variation norm on measures.

The first approach is based on the study of the properties of the transition operator T defined as

$$Tf(x) = E\bigl[f(X_{n+1}) \mid X_n = x\bigr] = E\bigl[f\bigl(\varphi(x, U_n)\bigr)\bigr] = \int f\bigl(\varphi(x, u)\bigr)\mu(du), \tag{2}$$

where μ is the distribution of U_n. The second one is based on the fine study of the trajectories of X_n, especially the recurrence properties.

The most typical objective of both approaches is to understand the behaviour of

$$T^n f(x) = E\bigl[f(X_n) \mid X_0 = x\bigr] \tag{3}$$

allowing in particular to study arbitrary correlations

$$E\big[f(X_n)g(X_0)\big] = E\big[g(X_0)T^n f(X_0)\big]. \tag{4}$$

In many situations, (3) converges pointwise for a broad class of functions f and we know that if this convergence holds for any bounded continuous function, with a limit independent of x, it would imply the convergence in distribution of X_n to the invariant measure π. This is the case for the stochastic system

$$X_{n+1} = \frac{1}{2}X_n + U_n, \tag{5}$$

where U_n an i.i.d. Bernoulli sequence. For any bounded continuous function f, $T^n f(x)$ converges, because given $X_0 = x$, the variable $X_n = U_{n-1} + \frac{1}{2}U_{n-2} + \cdots + 2^{-n+1}U_0 + 2^{-n}x$ has same distribution as $U_1 + \frac{1}{2}U_2 + \cdots + 2^{-n+1}U_n + 2^{-n}x$ which converges with probability one. The sequence X_n converges in distribution. If now U_n is a Gaussian sequence, the distribution of X_n converges in total variation, which is not the case if U_n is Bernoulli.

At the other extreme, for a stochastic system like

$$X_{n+1} = \{2X_n\}, \tag{6}$$

where $\{\cdot\}$ denote the fractional part, there is no pointwise limit to (3) (because X_n is a deterministic function of X_0) whereas (4) may well converge, depending on the distribution of X_0. This is the case for many *dynamical systems* (we denote by "dynamical system" the situation where φ depend only on its first variable and the only source of randomness comes from X_0); this means that (3) converges actually in some weak sense. Notice, however, that the control of correlations, that is, equation (4), leads to elementary [24] or more sophisticated [21] arguments for proving laws of large numbers and invariance principles.

With these examples, we see that the distribution of X_n may converge in total variation, in law, or in Wassertein distance to its limit. Many other intermediate kinds of convergence may be envisaged. For any given stochastic system (1), the problem can thus be summarized as follows: In *which sense* does $T^n f$ converges, for which functions f, and at which rate?

This paper uses the operator theoretic approach, although some fruitful ideas have been borrowed from the probabilistic one, especially concerning the case where the convergence is not geometric.

A huge amount of literature is concerned with both of these points of views. In this section, we shall first give a sketch of the main ideas of each approach (operator theoretic in Section 1.1 and probabilistic in Section 1.2) with typical examples of the simplest situations, and then we shall present in Section 1.3 our plan of action.

1.1. The Yosida–Kakutani theorem and the Ionescu–Tulcea–Marinescu theorem for quasi-compactness

It is well known that in the finite case (i.e., when X_n takes values in a finite state space), T is actually a matrix, and when T^n converges, the rate is always geometric. It is given by ρ^n, where ρ

is the second largest modulus of the eigenvalues of T. The gap between 1 (first eigenvalue) and ρ, is the spectral gap. The case of more than one eigenvalues of modulus one is more complicated and treated via a first splitting of the space into irreducible classes (the normal form, Chapter XIII, equation (69) of [10]) due to the multiplicity of the eigenvalue 1, and another splitting of each irreducible class into cyclic classes via the Frobenius theorem (Chapter XIII, equation (5) of [10]) due to the complex eigenvalues.

We present now the classical operator approach in the case of a general state space, which may be seen as the infinite dimensional extension of this matrix treatment. What is expected here is that for some norm $\|\cdot\|$ and any complex valued function f with $\|f\| < \infty$

$$\left\|T^n f - \pi(f)\right\| \leq C\rho^n \|f\| \tag{7}$$

for some $C > 0$ and $0 < \rho < 1$ (there is a harmless abuse of notation in the whole paper, appearing already in Equation (7): $\pi(f)$ will stand for the complex number $\int f(x)\pi(dx)$ as well as for the constant function with value $\pi(f)$). Examples are given below, this simplest case being $\|f\| = \|f\|_\infty$.

An operator T on a Banach space $(E, \|\cdot\|)$ is said quasi-compact if some power of T can be written as

$$T^n = K + V, \tag{8}$$

where V has spectral radius < 1 (i.e. $\|V^k\| < 1$ for some k) and K is a compact operator (e.g., K is finite-rank). Quasi-compactness has been extensively studied [12].

The Yosida–Kakutani theorem [28] says that, if, in addition to (8), the sequence $\|T^k\|$ is bounded, then E splits as:

(i) $E = E_c \oplus E_0$,
(ii) E_c is the finite dimensional space generated by the eigenvectors with eigenvalues of modulus 1,
(iii) E_0 is closed with $TE_0 \subset E_0$ and the restriction of T to E_0 has spectral radius < 1.

Denoting by λ_i, $i = 1, \dots, p$ the eigenvalues of T with modulus one, by E_i the corresponding eigenspaces, by P_i the projection on E_i parallel $\oplus_{j\neq i} E_j$, one has the equivalent formulation of points (i) to (iii):

$$T = \sum_{i=1}^{p} \lambda_i P_i + Q, \qquad Q = TP_0 = P_0 T, \tag{9}$$

where

$$|\lambda_1| = \cdots = |\lambda_p| = 1 \tag{10}$$

each P_i is a $\|\cdot\|$-continuous projection, with finite rank if $i > 0$ (11)

$$\sum_{i=0}^{p} P_i = \mathrm{Id}, \tag{12}$$

$$P_i P_j = P_j P_i = 0, \qquad 0 \le i < j \le p, \tag{13}$$

$$\left| Q^n \right| \to 0. \tag{14}$$

The last equation implies of course that $|Q^n| \le C\rho^n$ for some $C > 0$, $0 < \rho < 1$ (because if we define q as the first integer such that $|Q^q| = \rho_0 < 1$ and set $n = kq + r$, $r < q$, one has $|Q^n| \le \rho_0^k \sup_{r<q} |Q^r|$); another consequence of these equations is that for any $k \ge 1$

$$T^k = \sum_{i=1}^{p} \lambda_i^k P_i + Q^k. \tag{15}$$

The simplest case is when $p = 1$ and the eigenvalue 1 is simple, which means that P_1 has rank one and the convergence of (3) to $P_1 f$ is obtained (at least in some sense depending on the norm $|\cdot|$). In the general case, the operators P_i are described in terms of the invariant subspaces and their decomposition into cyclic classes [28].

A decade later, Ionescu–Tulcea and Marinescu provided a useful tool [13–15] for checking that quasi-compactness holds when $|T^n|$ is bounded:[1] it is assumed that there exists a weaker norm $\|\cdot\|$ on E (i.e., $\|f\| \le C|f|$ for some C and all $f \in E$), for which $\{Tf : f \in E, |f| \le 1\}$ is $\|\cdot\|$-compact and in addition, for some $\gamma < 1$, $c \ge 0$ and $k > 0$, and all $f \in E$

$$\left| T^k f \right| \le \gamma |f| + c\|f\|. \tag{16}$$

Under these assumptions, (9) to (15) hold.

It turns out that conditions (8) and (16) have different natural domains of applications. For an illustrative purpose, we give below two simple but typical examples concerning Markov chains. Namely, we show that (8) is well suited for dealing with Harris chain with convergence in total variation of the distribution of the variable, whereas (16) is more adapted to non-necessarily irreducible chains where, on the other hand, the transition has, for some metric on the state space, a contraction effect on the variable (Equation (18) below).

Before giving these examples, we would like to point out the important fact that the use of two norms is particularly adequate for treating the case where the convergence is not geometric; we shall come back to this below Section 1.2. More theoretical aspects of this are given in Appendix A.

Example 1. We consider here a Markov chain on a measured space S, which satisfies a Doeblin condition in the sense that there exists a positive measure $\nu(dx)$ such that its transition kernel satisfies for all $x \in S$

$$T(x, dy) \ge \nu(dy).$$

[1]The point had been actually introduced much sooner by Doeblin and Fortet in [5], equation (2) and (3) page 143, but in a more specific context.

Under these circumstances, one can write (T stands in the whole paper for the transition probability $T(x, dy)$ as well as for the transition operator $f \mapsto TF$)

$$Tf(x) = \int f(y)\nu(dy) + \int f(y)\big(T(x, dy) - \nu(dy)\big) = Kf + Vf$$

and (8) applies with $|\!|\!| f |\!|\!| = \|f\|_\infty$, on the space E of bounded measurable functions; K is indeed compact because its rank is one; finally for any $f \in E$, $|\!|\!| Vf |\!|\!| \le (1 - \nu(S))|\!|\!| f |\!|\!|$, hence $|\!|\!| V |\!|\!| \le 1 - \nu(S) < 1$. The Yosida–Kakutani theorem applies. With some additional efforts, one can show that E_c is the one-dimensional space of constant functions. If π is the invariant measure, one gets

$$\big|\!\big|\!\big| T^n f - \pi(f) \big|\!\big|\!\big| \le C\rho^n |\!|\!| f |\!|\!|. \tag{17}$$

For any initial measure μ, we obtain

$$\big|\mu\big(T^n f\big) - \pi(f)\big| \le C\rho^n |\!|\!| f |\!|\!|$$

and this means exactly that

$$\|\mu_n - \pi\|_{\mathrm{TV}} \le C\rho^n \|\mu\|_{\mathrm{TV}},$$

where μ_n is the distribution of X_n when $X_0 \sim \mu$, and $\|\cdot\|_{\mathrm{TV}}$ is the total variation norm.

Example 2. Let us consider now a chain defined on some metric space (S, d) with the form

$$X_{n+1} = \varphi(X_n, U_n),$$

where U_n is an i.i.d. sequence with distribution μ, and X_n belongs to S. Hence,

$$Tf(x) = \int f\big(\varphi(x, u)\big)\mu(du).$$

The function φ is supposed to satisfy adequate measurability assumptions and the following uniform Lipschitz property on (S, d):

$$d\big(\varphi(x, u), \varphi(y, u)\big) \le \gamma d(x, y) \tag{18}$$

for some $\gamma < 1$ and all x, y, u. On can see $\varphi(\cdot, \cdot)$ as a family of contractions on S parametrized by u [14]. In this case, it is easy to check that (16) applies with

$$\begin{aligned}
\|f\| &= \|f\|_\infty, \\
|\!|\!| f |\!|\!| &= \|f\| + [f], \\
[f] &= \sup_{x \neq y} \frac{f(x) - f(y)}{d(x, y)}.
\end{aligned} \tag{19}$$

In order to have the $\|\cdot\|$-compactness of $B = \{Tf : f \in E, |\!|\!| f |\!|\!| \le 1\}$, we assume that the state space is compact. Application of the Yosida–Kakutani theorem leads to the geometric convergence of $T^n f$ the Lipschitz norm $f \mapsto |\!|\!| f |\!|\!|$, which means the convergence of the distribution of X_n in Wasserstein distance, and opens the way for coupling methods (cf. [27], Chapter 6, in particular Equation (6.3)). Convergence in total variation will not hold in general (e.g. the chain (5) when U_n is Bernoulli). This approach has recently received increased attention, specifically concerning subgeometric convergence rates, in cases where typically γ depends on (x, y) and may approach 1 in some ways [2,9]; we will come back to this in Section 4.

All this is based on the fact that $x \mapsto \varphi(x, u)$ is a contraction for any u, making the dependence with respect to the initial condition decrease with time. Another fruitful approach [17], which we have not yet mentioned, is based on the assumption that $x \mapsto \mu_x$, where μ_x is the distribution of $\varphi(x, U)$, is a contraction for the Wasserstein distance, assumption ensured here by (18). We would say that this theory is actually neither operator-oriented nor probabilistic but rather geometric; its advantage is probably to give more explicit bounds through more direct proofs.

1.2. The probabilistic approach

Let us consider an irreducible aperiodic Markov chain with invariant measure π. Interestingly, it appears that in many situations, geometric convergence like (17) does not occur, but nevertheless for many $f \in E$, $T^n f - \pi(f)$ converges exponentially fast to 0. In other words, the convergence is not $| \cdot |$-uniform, and sometimes this convergence does not follow an exponential rate, but is slower. This situation has been treated quite successfully with a very probabilistic approach, where the speed of convergence is related to the integrability of recurrence times. The reference [22], and more specifically [16], deals with these situations. Two key concepts are used: the ψ-irreducibility, and a drift condition for controlling moments of recurrence times. A simple illustrative example of this absence of spectral gap is the following operator on $(\mathbb{R}^{\mathbb{N}}, \|\cdot\|_\infty)$:

$$Tf(x) = \frac{1}{2}\big(f(x) + f\big((x-1)_+\big)\big), \qquad x \in \mathbb{N}$$

corresponding to the following chain on $\mathbb{N}$

$$X_{n+1} = (X_n - U_{n+1})_+, \qquad P(U_n = 0) = P(U_n = 1) = \frac{1}{2}. \tag{20}$$

The pointwise convergence $T^n f(x) \to \pi(f) = f(0)$ is very fast, but this convergence is not uniform. In particular, this makes (17) impossible to occur with $|f| = \|f\|_\infty$. A possible operator theoretical approach is that one has for some weaker norm $\|\cdot\|$

$$\big\|T^n f - \pi(f)\big\| < \rho_n |f| \tag{21}$$

for any $f \in E$, and some fixed decreasing sequence ρ_n. For example, the first equation in [7] is (21) with

$$\|f\| = \sup_x \frac{|f(x)|}{g_0(x)},$$

$$|f| = \sup_x \frac{|f(x)|}{f_0(x)}$$

for some functions $f_0, g_0 \geq 1$ (called f and g in the paper). The norm $\|\cdot\|$ introduced here has actually strong connections with the one involved in the Ionescu Tulcea–Marinescu approach. The rate of decrease of ρ_n depends on the choice of $\|\cdot\|$. Notice that if in (21) the norms were equal, the convergence of ρ_n to zero would imply the geometric convergence; however, *this is not the case any more when the norms are different*.

The probabilistic school has thus slowly shifted towards more functional theoretic arguments as illustrated by the addition of Chapter 20 in [22], or the use of Nagaev's method in [18], [19], but still restricting its work to ψ-irreducible chains and total variation convergence of measures (e.g., [7]), making, for instance, the study of (20) impossible unless the law of U_n is changed for a non a discrete distribution.

1.3. Aim of the paper

The aim of this paper is to show that these ideas can be combined successfully and that they lead to an operator theoretic approach where non-geometric convergence is considered. The main feature of this theory is to work simultaneously with two norms and to use this for measuring non-geometric rates of convergence.

Our approach has essentially two steps: we first give conditions under which (9) to (13) hold with

$$\|Q^n f\| \leq \rho_n |\!|\!| f |\!|\!|, \qquad \rho_n \to 0 \tag{22}$$

instead of (14). This is the main objective of Section 2 (see Theorem 1). Notice that in this decomposition the Banach space is $(E, |\!|\!|\cdot|\!|\!|)$, and the norm $\|\cdot\|$ only appears in (22); in particular nothing guarantees that $|\!|\!| Q^n f |\!|\!|$ tends to zero.

Section 3 is concerned with geometric convergence, that is, $\rho_n = C\rho^n$. Specifically Theorem 3 shows how the Yoshida–Kakutani and Ionescu Tulcea–Marinescu approaches can be combined into a single statement. This allows an easy treatment of chains having an irreducible component and another component behaving like Example 2 above.

Section 4 is concerned with sub-geometric convergence. Theorem 7 proposes a way to estimate the decay rate of the sequence ρ_n.

General theorems concerning Markov chains and examples are given throughout the paper in order to point out that this approach is very versatile for the study of a large class of dynamical systems, in particular for irreducible as well as for non-irreducible Markov chains.

2. General results

In the whole paper, we shall consider an operator T on a vector space $(E, |\!|\!|\cdot|\!|\!|)$ endowed with another norm $\|\cdot\|$. We shall denote by B_0, B the unit balls for these norms:

$$B_0 = \{f \in E : \|f\| \leq 1\}, \tag{23}$$

$$B = \{f \in E : |\!|\!| f |\!|\!| \leq 1\}. \tag{24}$$

We shall work under the following assumptions:

(A0) $(E, |\!|\cdot|\!|)$ is a Banach space, B is complete for the metric induced by $\|\cdot\|$, and for some C_0

$$\forall f \in E, \qquad \|f\| \le C_0 |\!|f|\!|. \tag{25}$$

(A1) The number $C_T = \sup_n |\!|T^n|\!|$ is finite.

$(E, \|\cdot\|)$ is typically not complete. For instance, one can have $E = C_b(\mathbb{R})$, the space of bounded continuous functions on $\mathbb{R}$, $|\!|f|\!| = \|f\|_\infty$ and $\|f\| = \sup_x \frac{|f(x)|}{1+|x|^2}$.

The following theorem gives a necessary and sufficient condition to have the decomposition (9) to (13) and (22):

Theorem 1. *If in addition to* (A0) *and* (A1), T *is a sum of two operators*

$$T = K + V \tag{26}$$

both $|\!|\cdot|\!|$*-continuous and* $\|\cdot\|$*-continuous, which satisfy for some* $C_K > 0$ *and for any* n *and any* $f \in E$

$$KT^nKB \text{ is } \|\cdot\|\text{-totally bounded}, \tag{27}$$

$$\|V^n f\| \le \varepsilon'_n |\!|f|\!|, \qquad \varepsilon'_n \to 0, \tag{28}$$

$$\sum_{k\ge 0} |\!|KV^k|\!| < \infty, \tag{29}$$

$$|\!|Kf|\!| \le C_K \|f\|, \tag{30}$$

then (9) *to* (13) *and* (22) *hold true.*

If T *is* $|\!|\cdot|\!|$*-continuous and* $\|\cdot\|$*-continuous, and* T^k *satisfies the assumptions* (26) *to* (30) *for some* $k > 0$, *then* (9) *to* (13) *and* (22) *hold true.*

We observe that the conditions are clearly necessary by taking $V = Q$ and $K = T - Q$ (because $(T - Q)Q = 0$ and $T - Q$ has finite rank). The proof is postponed to Appendix A. This proof utilizes the more general Theorem 11 stated in Section A.1, and is based on an extensive use of the identity:

$$T^n = \sum_{i=1}^{n} T^{n-i}(T - V)V^{i-1} + V^n = \sum_{i=1}^{n} T^{n-i}KV^{i-1} + V^n. \tag{31}$$

Very coarsely, the assumptions combined with (31), imply that for any sequence $f_k \in B$, the sequence $T^k f_k$ is $\|\cdot\|$-totally bounded. This allows us to prove that E is the direct sum of two $|\!|\cdot|\!|$-closed, T-stable spaces

$$E = \{f : \|T^n f\| \to 0\} \oplus \Big\{f : \liminf_n \|f - T^n f\| = 0\Big\} = E_0 \oplus E_c. \tag{32}$$

Next we prove that E_c is finite dimensional (by proving that its unit ball is compact) with a basis of eigenvectors. The projection P_0 of Equation (9) is then the projection on E_0 parallel to E_c.

Application to Markov chains. We shall consider a measurable space $(S, \mathcal{F})$ with a measurable weight function $v \geq 1$ and we adopt the following notation

$$\|f\|_v = \|f/v\|_\infty. \tag{33}$$

We denote by E the Banach space of bounded measurable functions on $(S, \mathcal{F})$, with the norm $|\!|\!| f |\!|\!| = \|f\|_\infty$. The conclusion of the theorem will lead directly to the total variation convergence of the distributions. We have $\|f\|_v \leq |\!|\!| f |\!|\!|$. We recall that a transition operator on $(S, \mathcal{F})$ is a function $(x, A) \mapsto T(x, A)$ such that for any $x \in S$, $A \to T(x, A)$ is a probability measure, and for any $A \in \mathcal{F}$, $x \to T(x, A)$ is measurable.

Theorem 2. *Let T be a Markov transition operator*:

$$(Tf)(x) = \int_y f(y) T(x, dy).$$

Assume that for some set K_0 and some $c_v > 0$

$$Tv(x) \leq v(x) - c_v, \qquad \forall x \notin K_0, \tag{34}$$

$$Tv \qquad \text{is bounded on } K_0 \tag{35}$$

and that there exists another kernel $K(x, dy)$ such that $0 \leq K(x, dy) \leq T(x, dy)$, and such that for some $\varepsilon > 0$, and some non-negative measure ν one has

$$K(x, S) \geq \varepsilon, \qquad \forall x \in K_0, \tag{36}$$

$$K(x, S) = 0, \qquad \forall x \notin K_0, \tag{37}$$

$$\|Kf\|_\infty \leq \nu\big(|f|\big), \qquad \forall f \in E, \tag{38}$$

$$\nu(v) < \infty. \tag{39}$$

Set

$$|\!|\!| f |\!|\!| = \|f\|_\infty, \tag{40}$$

$$\|f\| = \|f\|_v. \tag{41}$$

Then Theorem 1 applies with K and $V = T - K$. In particular Equations (9) to (13) and (22) hold true.

If in addition there is no measurable set A such that $x \mapsto T(x, A)$ is a non-trivial indicator function[2] *then there exist a probability measure π and a sequence $\rho_n \to 0$ such that for any $f \in E$*

$$\big\|T^n f - \pi(f)\big\|_v \leq \rho_n \|f\|_\infty. \tag{42}$$

[2]This would mean that $1_{X_1 \in A}$ would be a deterministic non-constant function of the initial state X_0.

We recall that $\pi(f)$ stands here for the constant function with value $\pi(f)$. The proof of this consequence of Theorem 1 is postponed to Appendix B. Estimations of ρ_n will be given later in Theorem 9.

Remark. Equation (34) is known as the "drift condition" (cf. Theorem 11.0.1 of [22] or Proposition 5.10 in [23]). Equations (36) to (39) are reminiscent of the T-chain property (cf. [22] Theorem 6.0.1), used to check the irreducibility assumption (cf. [22] page 87). However, the Feller property is not required here. Equation (37) is not a restriction, since cancelling K outside K_0 does not affect the other assumptions. The essential difficulty with the present assumptions is that the set K_0 has to be the same in (34) and in (36). Notice however that the sets K_0 satisfying assumptions (36) and (37) are stable by finite union.

Example. Consider the Markov chain on $\mathbb{R}_+$ defined by

$$X_{n+1} = X_n + 1 + X_n^\alpha W_{n+1}, \tag{43}$$

where W_n is an i.i.d. sequence of non-zero centred random variables with values in $[-1, 1]$, with a non-zero absolutely continuous component. In addition, we assume that

$$1/2 < \alpha < 1.$$

Take

$$v(x) = x^p + 1$$

for some $p \le 1$ which will be chosen later as $2(1-\alpha)$. Then

$$Tv(x) = 1 + E\big[\big(x + 1 + x^\alpha W_1\big)^p\big].$$

By the second order Taylor formula applied to the function v in the neighbourhood of $x+1$, there exist a random number $0 < \theta < 1$ such that

$$\begin{aligned} Tv(x) &= 1 + (x+1)^p - \frac{p(1-p)}{2} x^{2\alpha} E\big[\big(x + 1 + \theta x^\alpha W_1\big)^{p-2} W_1^2\big] \\ &\le 1 + (x+1)^p - \frac{p(1-p)}{2} x^{2\alpha} \big(x + 1 - x^\alpha\big)^{p-2} \sigma^2, \end{aligned} \tag{44}$$

where σ^2 is the variance of W_1. Taking $p = 2(1-\alpha)$, we have $0 < p < 1$ and

$$\begin{aligned} Tv(x) &\le 1 + (x+1)^p - \frac{p(1-p)}{2} \left(\frac{x}{x + 1 - x^\alpha}\right)^{2\alpha} \sigma^2 \\ &\le 1 + x^p - \frac{p(1-p)}{3} \sigma^2 \qquad \text{for } x \text{ large enough.} \end{aligned}$$

Equation (34) is satisfied for some interval $K_0 = [0, M]$. Equation (35) is obvious. In order to check Equations (36) to (39), notice that if the absolutely continuous component of W_1 has a density $\geq \varepsilon$ on a subset A of $[-1, 1]$ with positive measure, $K(x, dy)$ can be taken as $\varepsilon\lambda(A)$ times the distribution of $x + 1 + x^\alpha \tilde{W}_1$, where $\tilde{W}_1$ has density $1_A/\lambda(A)$, ν is some multiple of the uniform measure on $[0, M + M^\alpha + 1]$. Therefore, theorem applies. In order to get (42), it remains to prove that $T1_A = 1_B$ is impossible unless $B = \mathbb{R}_+$ or $B = \varnothing$. If B is non-trivial one can find two sequences x_n and y_n having the same limit such that $x_n \in B$ and $y_n \notin B$. The relation $T1_A = 1_B$ would mean that for each n, the distributions of $x_n + 1 + x_n^\alpha W_1$ and $y_n + 1 + y_n^\alpha W_1$ are mutually singular (supported on A and A^c), which is impossible for n large because W_1 has an absolutely continuous component. As a consequence, B is necessarily trivial and (42) holds.

Notice that nevertheless $E[X_n] = E[X_0] + n$.

3. Geometric convergence: Quasi-compactness

In this section, we give a theorem which encompasses both Yosida–Kakutani and Ionescu–Tulcea–Marinescu theorems, and present an application to Markov chains which mixes both kinds of situations presented above. A specific application to autoregressive processes with Markov switching is finally studied. We recall that B denotes the unit closed ball for the norm $|\!|\!|\cdot|\!|\!|$.

Theorem 3. *Let T be an operator on $(E, |\!|\!|\cdot|\!|\!|)$ satisfying* (A0), (A1) *and*

(A3) *T is $\|\cdot\|$-continuous. For some $\|\cdot\|$-totally bounded set K_B, $\gamma < 1$, $c > 0$ and $q > 0$*

$$T^q B \subset \gamma B + K_B, \tag{45}$$

$$|\!|\!|T^q f|\!|\!| \leq \gamma |\!|\!|f|\!|\!| + c\|f\|. \tag{46}$$

Then Equations (9) *to* (14) *hold.*

Under the assumption of the Yosida–Kakutani theorem, one can take here $\|\cdot\| = |\!|\!|\cdot|\!|\!|$, and clearly Theorem 3 applies. Under the assumptions of the Ionescu–Tulcea–Marinescu theorem, we can take $K_B = \|T^q\|B$ and Theorem 3 applies. Like Theorem 1, this theorem is a consequence of the general Theorem 11 stated in Section A.1; its proof is postponed to Appendix C.

The following theorem may seem very general and unclear for the applications. It says that if T can be lower bounded by an operator with nice properties, then quasi-compactness holds.

We should point out that we intend to bridge a continuum over two extreme cases: the convergence of the Markov chain in Wasserstein distance and the convergence in total variation. This will be exemplified below.

Let us just mention that $[\cdot]$ below is typically a Lipschitz semi-norm like in Equation (19) or simply $[\cdot] \equiv 0$, in which case we shall we get total variation convergence (cf. the following corollary).

Theorem 4. *Let (S, d) be a metric space and $\mathcal{B}$ its Borel σ-field. We assume that is given a continuous function $v(x) \geq 1$ on S such that for any $A > 0$, $\{x : v(x) \leq A\}$ is compact. Consider a vector space E of $\mathcal{B}$-measurable functions defined on S, with values on $\mathbb{C}$, containing compactly* *supported Lipschitz functions. On* E *is defined a semi-norm* $f \mapsto [f]$ *and we set for any function* f *on* S:

$$|f| = \|f\| + [f], \tag{47}$$

$$\|f\| = \sup_x \frac{|f(x)|}{v(x)}. \tag{48}$$

We assume that $(E, |\cdot|)$ *is a Banach space and that* (A0) *holds.*

Let T *be a Markov transition operator defined on* E*. We assume the existence of* $0 < \gamma_b, \gamma_v < 1$ *and* $c_v > 0$ *such that*

$$[Tf] \leq \gamma_b [f], \qquad f \in E, \tag{49}$$

$$Tv(x) \leq \gamma_v v(x) + c_v. \tag{50}$$

We assume the existence of a non-negative kernel $K(x, dz)$*, of functions* $\psi \geq 0$, $\varepsilon_d > 0$ *and* $\tau \geq 0$*, such that for any* $x, y \in S$ *and* $f \in E$,

$$K(x, dz) \leq T(x, dz), \tag{51}$$

$$K(x, S) \geq \varepsilon_d(x), \tag{52}$$

$$\left|Kf(y) - Kf(x)\right| \leq \tau(x, y)\big([f] + \psi\big(d(x, y)\big)\|f\|\big). \tag{53}$$

Moreover the function $\tau(\cdot, \cdot)$ *is assumed to be bounded on compact subsets of* $S \times S$, $\psi(x)$ *tends to* 0 *as* $x \to 0$*, and* $\varepsilon_d(x)$ *is satisfies*

$$\lim_{v(x) \to \infty} \varepsilon_d(x) v(x) = +\infty, \tag{54}$$

$$\forall A, \qquad \min_{v(x) \leq A} \varepsilon_d(x) > 0. \tag{55}$$

Then Theorem 3 *applies* (*i.e.* (A2) *holds true*) *with a pair of norms* $(|\cdot|', \|\cdot\|')$ *respectively equivalent to* $|\cdot|$ *and* $\|\cdot\|$*. In particular, if the constant functions are the only eigenvectors of* T *with an eigenvalue of modulus* 1*, there exist* $C > 0$, $0 < \rho < 1$ *and a probability measure* π *such that for any* $f \in E$,

$$\left|\pi(f) - T^n f\right| \leq C\rho^n |f| \tag{56}$$

and $\pi(v) < \infty$.

The proof is postponed to Appendix D. We use Theorem 3 with $q = 1$. The idea is to set

$$Sf(x) = \sum_{i=1}^{n} \theta_i(x) Kf(x_i),$$

where $\theta_1, \dots, \theta_n$ is a partition of the unity of a large portion of the space, each x_i being a point of the support of θ_i. Clearly $S(B)$ is compact. It remains to prove that $\|(T - S)f\| \leq \gamma \|f\|$ (which implies (45)) and that (46) holds true.

We shall consider two examples, one where $[f]$ is trivially chosen as $[f] \equiv 0$ and we get geometric convergence in $\|\cdot\|$ norm (which, by duality, corresponds to geometric weighted total variation convergence for the distribution of the Markov chain), and another case where $[\cdot]$ plays an important role.

Application to geometric total variation convergence. In the case $[f] \equiv 0$ we get the following corollary:

Corollary 5. *Let (S, d) be a metric space and $\mathcal{B}$ its Borel σ-field. We assume that is given a continuous function $v(x) \geq 1$ on S such that for any $A > 0$, $\{x : v(x) \leq A\}$ is compact. Consider the Banach space $(E, |\!|\!| \cdot |\!|\!|)$ of $\mathcal{B}$-measurable functions f defined on S such that*

$$|\!|\!| f |\!|\!| = \sup_x \frac{|f(x)|}{v(x)} \tag{57}$$

is finite.

Let T be a Markov transition operator defined on E. We assume the existence of $0 < \gamma_v < 1$ and $c_v > 0$ such that

$$Tv(x) \leq \gamma_v v(x) + c_v. \tag{58}$$

We assume the existence of of a non-negative kernel $K(x, dz)$, functions $\varepsilon_d > 0$, $\psi \geq 0$, such that for any $x, y \in S$ and $f \in E$

$$K(x, dz) \leq T(x, dz), \tag{59}$$

$$K(x, S) \geq \varepsilon_d(x), \tag{60}$$

$$\bigl|Kf(y) - Kf(x)\bigr| \leq \psi\bigl(d(x, y)\bigr)\|f\|. \tag{61}$$

Moreover, we assume that $\psi(x)$ tends to 0 as $x \to 0$, and that the function $\varepsilon_d(x)$ satisfies

$$\lim_{v(x)\to\infty} \varepsilon_d(x)v(x) = +\infty, \tag{62}$$

$$\forall A, \qquad \min_{v(x)\leq A} \varepsilon_d(x) > 0. \tag{63}$$

Then Equations (9) to (14) hold. In particular, if the constant functions are the only eigenvectors of T with an eigenvalue of modulus 1, there exist $C > 0$, $0 < \rho < 1$ and a probability measure π such that for any $f \in E$,

$$\bigl|\!\bigl|\!\bigl|\pi(f) - T^n f\bigr|\!\bigr|\!\bigr| \leq C\rho^n |\!|\!| f |\!|\!| \tag{64}$$

and $\pi(v) < \infty$. In addition, for any $x \in S$, the distribution μ_n^x of X_n when $X_0 = x$ converges exponentially fast in total variation to π.

$$\bigl\|\pi - \mu_n^x\bigr\| \leq C\rho^n v(x). \tag{65}$$

Proof. It suffices to prove the last statement. Equation (64) implies that for any bounded function $f \in E$ and any measure μ

$$\left|\pi(f) - \mu_n(f)\right| \le C\rho^n |\!|\!| f |\!|\!| v(x) \le C\rho^n \|f\|_\infty v(x),$$

where μ_n is the distribution of X_n starting from x. This means the total variation convergence of μ_n to π. □

In many cases $\varepsilon_d(x) = 1/2$ will do the job, but in the following example the situation is more complicated:

$$X_{n+1} = \begin{cases} \frac{1}{2} X_n, & \text{with probability } 1 - p(X_n), \\ V_n, & \text{with probability } p(X_n), \end{cases}$$

where V_n is an i.i.d. sequence and p is a positive function of x; V_n can be constant. We see that only the second type of transition contributes to the convergence in total variation, this is why we shall need $p(x)$ not to be too small. Let us assume that for some $0 < \alpha < 1$ and for some positive uniformly continuous function $q(x)$

$$E\left[|V_n|^\alpha\right] < \infty,$$
$$0 < q(x) \le p(x),$$
$$\lim_{x \to \infty} q(x)|x|^\alpha = +\infty.$$

Then Equations (58) to (63) are clearly satisfied with

$$v(x) = |x|^\alpha + 1,$$
$$K(x, A) = q(x) P(V_1 \in A),$$
$$\varepsilon_d(x) = q(x).$$

Indeed

$$T v(x) = \big(1 - p(x)\big) v(x/2) + p(x) E\big[v(V_1)\big] \le |x|^\alpha 2^{-\alpha} + 1 + E\big[v(V_1)\big]$$

(i.e. $\gamma_v = 2^{-\alpha}$) and Equations (59) to (63) are immediately checked.

The exponential convergence holds. If one tries to prove the same convergence by the probabilistic approach, e.g. Theorem 16.1.2 of [22], the problem is to prove the ψ-irreducibility, that is, the existence of a measure ψ such that if $\psi(A) > 0$, for any x, $P_x(X_n \in A \text{ for some } n) > 0$. This condition is implicitly checked by the assumptions.

Application to functional autoregressive processes with Markov switching. We consider the following mixed Markov process $(I_n, X_n) \in S$ where $S = \{1, \ldots, s\} \times \mathbb{R}^d$:

$$P(I_{n+1} = j \mid I_n = i) = p_{ij}, \qquad 1 \le i, j \le s, \tag{66}$$

$$X_{n+1} = \alpha(I_n)\varphi(X_n) + \beta(I_n, V_n), \tag{67}$$

where α is a matrix valued measurable function, φ and β are vector valued measurable functions, and V_n is an independent i.i.d. sequence. In other words

$$Tf(i,x) = \sum_k p_{ik} E\big[f\big(k, \alpha(i)\varphi(x) + \beta(i, V_1)\big)\big].$$

If for all i the variable $\beta(i, V_1)$ has a density, we can apply Corollary 5 at the price of extra reasonable assumptions because (59) to (61) would be satisfied for some kernel K (the continuity of φ is important here); our point is to deal with singular measures. As in [3], Theorem 1.4, we have made efforts to give conditions which allow for non-contracting values for α, as one can see in Equation (69).

Theorem 6. *Consider the Markov chain defined by* (66) *and* (67). *We assume that the chain I_n is irreducible and aperiodic with invariant measure π on its finite state space, and that for some $q > 0$*

$$\big|\varphi(y) - \varphi(z)\big| \le |y - z|, \tag{68}$$

$$\sum_i \pi_i \log\big(\|\alpha(i)\|\big) < 0, \tag{69}$$

$$\sup_i E\big[\big|\beta(i, V_1)\big|^q\big] < +\infty, \tag{70}$$

where $|\cdot|$ the euclidean norm and $\|\cdot\|$ is the usual matrix norm $\|M\| = \sup_{|x|=1} |Mx|$. Then Theorem 4 applies and (56) *holds with the norm*

$$|\!|\!| f |\!|\!|' = \sup_{i,x,x'} \frac{|f(i,x) - f(i,x')|}{|x - x'|^\eta} + \sup_{i,x} \frac{|f(i,x)|}{|x|^\eta + 1}$$

for η small enough. This implies that for any realization (I_n, X_n) of the chain at time n with an arbitrary initial distribution, one can find a coupling with a pair (I', X') having the stationary distribution, such that

$$P\big(I_n \ne I'\big) + E\big[\big|X_n - X'\big|^\eta\big] < C\rho^n\big(1 + E\big[|X_0|^\eta\big]\big).$$

Proof. We will choose

$$[f] = \sum_i \nu_i [f]_i, \qquad [f]_i = \sup_{x,y} \frac{|f(i,x) - f(i,y)|}{|x - y|^\eta},$$

$$v(i,x) = |x|^\varepsilon e^{\varepsilon\lambda(i)} + 1,$$

$$d\big((i,x),(j,y)\big) = 1_{i\ne j} + |x - y|^\eta$$

for some constants ν_i and $\lambda(i)$ which will be specified later. Concerning K we simply set:

$$K = T.$$

In that case, (51) and (52) are obvious ($\varepsilon_d = 1$), and (53) will be a consequence of (49). The technical part is to prove that (49) and (50) hold true. We now focus on (50). We first note that since

$$X_{n+1} = \alpha(I_n)\big(\varphi(X_n) - \varphi(0)\big) + \big(\alpha(I_n)\varphi(0) + \beta(I_n, V_n)\big)$$

we can assume that $\varphi(0) = 0$. Unsurprisingly, the contraction property (50) is related to the rate at which the product of $\alpha(I_k)$'s converges to zero, this one being itself controlled by the speed at which the law of large numbers acts on the sums of $\log(\|\alpha(I_k)\|)$'s. This uses classically the Poisson equation: Since the chain I_n is irreducible aperiodic on a finite state space, there exists a unique (up to a constant) solution λ to the Poisson equation

$$E\big[\lambda(I_1) \mid I_0 = i\big] = \lambda(i) - l(i) + \pi(l), \qquad l(i) = \log\big(\|\alpha(i)\|\big)$$

(it is simply $\lambda = \sum_{k=0}^{\infty}(T_0^k - \pi)l$ where $T_0 = (p_{ij})_{1\le i,j\le s}$ is the transition operator of the chain I_n). The process

$$Z_n = |X_n|^{\varepsilon} e^{\varepsilon\lambda(I_n)}$$

satisfies, thanks to (67), (68), and $\varphi(0) = 0$:

$$\begin{aligned}
Z_{n+1} &\le \big(\|\alpha(I_n)\|\,|\varphi(X_n)| + |\beta(I_n, V_n)|\big)^{\varepsilon} e^{\varepsilon\lambda(I_{n+1})} \\
&\le \|\alpha(I_n)\|^{\varepsilon}|X_n|^{\varepsilon} e^{\varepsilon\lambda(I_{n+1})} + |\beta(I_n, V_n)|^{\varepsilon} e^{\varepsilon\lambda(I_{n+1})} \\
&= Z_n e^{\varepsilon\{\log\|\alpha(I_n)\| + \lambda(I_{n+1}) - \lambda(I_n)\}} + e^{\varepsilon\lambda(I_{n+1})}|\beta(I_n, V_n)|^{\varepsilon}.
\end{aligned}$$

And since the factor of ε is bounded, we have for some c

$$\begin{aligned}
Z_{n+1} &\le Z_n\big(1 + \varepsilon\big(\lambda(I_{n+1}) - \lambda(I_n) + \log\|\alpha(I_n)\|\big) + c\varepsilon^2\big) + e^{\varepsilon\lambda(I_{n+1})}|\beta(I_n, V_n)|^{\varepsilon}, \\
E[Z_{n+1} \mid \mathcal{F}_n] &\le Z_n\big(1 + \varepsilon\pi(l) + c\varepsilon^2\big) + e^{\varepsilon \sup_i \lambda(i)} \sup_i E\big[|\beta(i, V_1)|^{\varepsilon}\big],
\end{aligned}$$

where $\mathcal{F}_n$ stand for the σ-field $\sigma(I_i, X_i, 0 \le i \le n)$. Hence, if we take $\varepsilon \le q$ such that $\varepsilon\pi(l) + c\varepsilon^2 < 0$, we obtain (50). Concerning (49):

$$\begin{aligned}
|Tf(i,y) - Tf(i,x)| &\le \sum_k p_{ik} E\big[\big|f\big(k, \alpha(i)\varphi(y) + \beta(i, V_1)\big) - f\big(k, \alpha(i)\varphi(x) + \beta(i, V_1)\big)\big|\big] \\
&\le |\varphi(y) - \varphi(x)|^{\eta}\|\alpha(i)\|^{\eta} \sum_k p_{ik}[f]_k,
\end{aligned}$$

$$[Tf]_i \le \|\alpha(i)\|^{\eta} \sum_k p_{ik}[f]_k,$$

$$\sum_i \nu_i [Tf]_i \le \sum_{i,k} \nu_i \|\alpha(i)\|^{\eta} p_{ik}[f]_k.$$

We see that if we can find ν such that

$$\forall k, \qquad \sum_i \|\alpha(i)\|^\eta \nu_i p_{ik} < \nu_k \tag{71}$$

then Equation (49) will be satisfied. To this aim, we define

$$\nu = \pi + \eta \sum_{k\geq 1} \bigl(\pi.l - \pi(l)\pi\bigr) P^k$$

with the notation

$$(\pi.l)(i) = \pi_i l(i).$$

Set $a_i = \|\alpha(i)\|^\eta$; since

$$a_i = 1 + \eta l(i) + O\bigl(\eta^2\bigr)$$

we get

$$\nu.a = \pi + \eta \sum_{k\geq 1} \bigl(\pi.l - \pi(l)\pi\bigr) P^k + \eta\pi.l + O\bigl(\eta^2\bigr)$$

hence,

$$(\nu.a)P = \pi + \eta \sum_{k\geq 1} \bigl(\pi.l - \pi(l)\pi\bigr) P^k + \eta\pi(l) + O\bigl(\eta^2\bigr) = \nu + \eta\pi(l) + O\bigl(\eta^2\bigr).$$

This equation implies that for η small enough, Equation (71) is satisfied. In particular, we shall impose $\eta \leq \varepsilon$. We have now proved (49) to (54).

As a byproduct, Equation (49) implies that any eigenfunction f, with associated eigenvalue $|\lambda| = 1$, does not depend on x, and consequently, since I_n is irreducible, f is necessarily constant.

Theorem 4 applies and (56) holds with

$$[\![f]\!] = \sup_{x,i} \frac{|f(i,x)|}{|x|^\varepsilon e^{\varepsilon\lambda(i)} + 1} + \sum_i \nu_i [f]_i.$$

Since by irreducibility, $\nu_i > 0$ for all i, this norm is equivalent to

$$N(f) = \sup_{i,x} \frac{|f(i,x)|}{|x|^\varepsilon + 1} + \sup_{i,x,y} \frac{|f(i,x) - f(i,y)|}{|x-y|^\eta}.$$

This norm is also equivalent to $[\![f]\!]'$ because, on the one hand, $\eta \leq \varepsilon$, and on the other hand

$$\begin{aligned} \sup_{i,x} \frac{|f(i,x)|}{|x|^\eta + 1} &\leq \sup_{i,x} \frac{|f(i,x) - f(i,0)| + |f(i,0)|}{|x|^\eta + 1} \\ &\leq \sup_{i,x} \frac{|f(i,x) - f(i,0)|}{|x|^\eta} + \sup_i \bigl|f(i,0)\bigr| \leq N(f). \end{aligned}$$

By the duality properties of the Wasserstein distance (cf. [27], Theorem 5.10, Equations (5.11) and (6.3))

$$\begin{aligned}
&\inf_{I_n, I', X_n, X'} P\big(I_n \neq I'\big) + E\big[\big|X_n - X'\big|^\eta\big] \\
&\quad = \sup_{f \text{ Lipschitz}} E\big[f(I_n, X_n) - f\big(I', X'\big)\big],
\end{aligned}$$

where the infimum is taken over all the pairs of random variables (I', X') and (I_n, X_n) having respectively the stationary distribution and the chain distribution at time n, and f is 1-Lipschitz w.r.t. the distance d. The expectation in the right-hand side is just $E[(Q^n f)(I_0, X_0)]$, which is smaller than $C\rho^n(1 + E[|X_0|^\eta])$. □

4. Subgeometric rates

In the rest of the paper, we shall find conditions under which the rate of convergence of V^n to 0 will give us an insight about the rate of convergence of Q^n to 0. We set for any operator S on E

$$\begin{aligned}
\|S\|_{E0} &= \sup_{|\!|\!| f |\!|\!| \leq 1} \|Sf\|, \\
\|S\|_{0E} &= \sup_{\|f\| \leq 1} |\!|\!| Sf |\!|\!|.
\end{aligned}$$

With this convention, one has

$$\begin{aligned}
\|UV\| &\leq \|U\|_{E0} \|V\|_{0E}, \\
|\!|\!| UV |\!|\!| &\leq \|U\|_{0E} \|V\|_{E0}.
\end{aligned}$$

We shall consider positive rate sequences α_n, $n \geq 1$, satisfying the conditions (R1) to (R3) below. For instance, sequences like $\alpha_n = (n+1)^{-p}$, $p > 1$, or $\alpha_n = \exp -\sqrt{n}$, or $\alpha_n = (n+1)^{-1}(\log(n+1))^{-2}$ satisfy these assumptions (notice that the first part of (R2) holds if $x \mapsto \log \alpha_x$ is convex). These conditions make it easy to solve some recursive equations (cf. Appendix F).

Theorem 7. *Let* (A0) *be satisfied and T be a $|\!|\!| \cdot |\!|\!|$- and $\| \cdot \|$-continuous operator on E satisfying* (A1), *Equations* (9) *to* (13) *and* (22). *Let α_n be a sequence satisfying*

(R1) $n \mapsto \alpha_n$ *is decreasing*,

(R2) $n \mapsto \dfrac{\alpha_{n+1}}{\alpha_n}$ *is increasing and converges to* 1,

(R3) $\displaystyle\sum_{n \geq 1} \frac{\alpha_n^2}{\alpha_{2n}} < \infty.$

We assume that T can be rewritten as $T = K + V$ with

$$\|V^k\|_{E0} \le C_1\alpha_k, \qquad k > 0, \tag{72}$$

$$|KV^k| \le C_2\alpha_k, \qquad k > 0, \tag{73}$$

$$|KQ^k| \to 0 \qquad \text{as } k \to \infty \tag{74}$$

(*Equations* (73) *and* (74) *are clearly satisfied if* (72) *and* (30) *hold true*). *Then one has for some $C > 0$ and all $n > 0$*

$$\|Q^n\|_{E0} \le C \sum_{k\ge n} \alpha_k.$$

If in addition $\sup_n \|T^n\| < \infty$, *then*

$$\|Q^n\|_{E0} \le C\alpha_n. \tag{75}$$

The proof is based on (31) and on the key result of Proposition 13. It is postponed to Appendix E.

Remarks. (1) If Theorem 1 is used for checking the assumptions, there is no need to check (29), which is automatically satisfied thanks to (28), (30) and the summability of α_n (consequence of (R1) and (R3)). (2) Condition (R2) excludes geometric rates. The theorem is indeed wrong in this case: For example, in the finite dimensional case, Theorem 1 holds with $V = 0$, and (75) only holds with some geometric rate.

Application to Markov chains. We consider here Markov chains which satisfy the following is strengthening of (34):

$$Tv(x) \le v(x) - \theta\big(v(x)\big), \qquad x \notin K_0 \tag{76}$$

for some function θ, e.g. $\theta(u) = u^q$, $0 < q < 1$. Our goal here is to use this information for bounding the sequence ρ_n in (42).

Lemma 8. *Let T be a Markov transition operator on a space S*:

$$(Tf)(x) = \int_y f(y)T(x, dy).$$

Assume that for some set $K_0 \subset S$, some $c, \varepsilon > 0$, some non-negative function v bounded below by a positive number, some function θ and some submarkovian operator V

$$Tv \le v - \theta(v) + c1_{K_0}, \tag{77}$$

$$V \le T, \tag{78}$$

$$(V1)(x) \le 1 - \varepsilon 1_{x\in K_0} \tag{79}$$

((V1)(x) is V(x, S)). We assume in addition that θ be a non-decreasing non-negative concave differentiable function on $[0, +\infty)$ *with a derivative which tends to zero at infinity. Then, for some constant* c'

$$V^n 1 \leq \frac{v + c'}{\psi^{(-1)}(n)}, \tag{80}$$

where the exponent (-1) *stands for the reciprocal function and*

$$\psi(x) = \int_0^x \frac{1}{\theta(y)}\, dy.$$

In addition, for some constant c''

$$T^n \theta(v) \leq \frac{v}{n} + c''. \tag{81}$$

The point here is that (80) implies (72) with $\alpha_n = \psi^{(-1)}(n)^{-1}$ as soon as $|\!|\!|\cdot|\!|\!| \geq \|\cdot\|_\infty$ and $\|f\| \leq \|f/v\|_\infty$.

In view of (77), a natural choice for V is $Vf(x) = (1 - 1_{x\in K_0})Tf(x)$, and we shall do this later in the proof of Theorem 10, but in the following application, we see that a more general situation is useful.

Theorem 9. *Let all the assumptions and notations of Theorem 2 hold and assume that (34) is strengthened as*

$$Tv(x) \leq v(x) - \theta\big(v(x)\big), \qquad x \notin K_0 \tag{82}$$

for some concave function θ satisfying the assumptions of Lemma 8. In addition, we assume that the sequence

$$\alpha_n = \frac{1}{\psi^{(-1)}(n)}$$

(ψ is given by (133)) satisfies the conditions (R1) to (R3) of Theorem 7. Then for some $c > 0$ *and any bounded measurable function f*

$$\sup_x \left| \frac{T^n f(x) - \pi(f)}{v(x)} \right| \leq c\rho_n \|f\|_\infty, \qquad \rho_n = \sum_{k\geq n} \alpha_k, \tag{83}$$

$$\pi\big(\big|T^n f - \pi(f)\big|\big) \leq c\alpha_n \|f\|_\infty. \tag{84}$$

The proof is a straightforward application of Lemma 9 together with Theorem 7 in the case $|\!|\!|f|\!|\!| = \|f\|_\infty$, $\|f\| = \|f\|_v$, and is postponed to Appendix H. We find the following matchings between drift function and rates (Table 1).

It is has been known for a certain time that the function $\psi^{(-1)}$ plays a key role in the estimation of the rate of convergence (e.g., [6,7] for Harris chains), and applications of this kind of result in

Table 1. Rates for various drift functions

$\theta(t)$	α_n	ρ_n
$\log(t+1)^2$	$\sim n^{-1}(\log n)^{-2}$	$\sim(\log n)^{-1}$
$t^q, 0<q<1$	$\sim n^{-1/(1-q)},$	$\sim n^{-q/(1-q)}$
$\frac{ct}{\log(t+1)}$	$\sim e^{-\sqrt{2cn}}$	$\sim e^{-\sqrt{2cn}}\sqrt{n}$

the field of Markov chains are not uncommon. For example, in [16] Jarner and Roberts give an application to Monte Carlo Markov Chains. They also consider (Example 1) the random walk on $[0,+\infty)$

$$X_{n+1}=(X_n+W_{n+1})_+,$$

where W_n is an i.i.d. sequence with $E[W_1]<0$. Under the assumption that there exists an integer $m\ge 2$ such that

$$E\bigl[|W_1|^m\bigr]<\infty$$

they prove that the drift condition (82) is satisfied with

$$v(x)=(x+1)^m,$$

$$\theta(x)=x^\alpha,\qquad \alpha=\frac{m-1}{m}.$$

In Theorem 3.6, they state that for any x, $\sup_{\|f\|_\infty\le 1}|T^nf(x)-\pi(f)|=o(n^{-\alpha/(1-\alpha)})$, which is somewhat intermediary between (83) and (84). On this example, we clearly see the interpretation of the difference of rates between (83) and (84): if the initial state $X_0=x_0$ is very large, it takes a long time to come back to the invariant measure (this time is certainly proportional to x_0), but if the initial state is drawn from π, it won't be large and the convergence rate is increased.

Similarly, in example (43), it is easily shown using (44) that

$$Tv(x)\le v(x)-\frac{p(1-p)}{3}x^{2\alpha+p-2}\sigma^2$$

for x large enough, as soon as $2-2\alpha<p<1$. This means that (82) is satisfied with $\theta(t)=t^{(2\alpha+p-2)/p}$. Hence, for any $2-2\alpha<p<1$, Theorem 9 applies with $\alpha_n=cn^{-p/(2\alpha-2)}$.

An application of Lemma 8 to weakly contractive stochastic dynamical systems. Consider a complete separable metric space (S,d). We define the Lipschitz seminorm

$$[g]=\sup_{x\ne y}\frac{|g(x)-g(y)|}{d(x,y)}.$$

We shall consider a transition operator on S, having a Lyapunov function [equation (85)] and a contraction property with is strict only in a part K of the space [equation (86), (87)]. The importance of considering such transition operators has been highlighted and exemplified by Butkovsky in [2]. He shows that, under these circumstances, Equation (88) holds (Equation (2.3) of the article, which is actually slightly weaker than (88), see below). We show in addition that Equation (89) holds true (an analogous result is more or less implicit in the proofs of [2], cf. Equation (4.8) of the article, but with a much worse rate of convergence).

Following [2], we have proved (88) only in the case $d \leq 1$. It is apparent in the proof that the general case can be treated similarly, starting from (89) again, as soon as one manages to get control of $T^n f(x)$, where f is the function $f(x) = d(x, y_0)\theta(v(x))$, y_0 being arbitrary.

While Butkovsky works on the space of measures (i.e., considering the action of the dual operator T^*), we will show this theorem by working directly on the space of Lipschitz functions and by using Lemma 8. The proof is postponed to Appendix I, and uses as another key point a theorem of Shaoyi Zhang which allows to perform a dynamical coupling of two realizations of a Markov chain, with different initial conditions, with a single Markov chain on the product space $S \times S$.

Durmus, Fort and Moulines present also an analogous result in [9] (Theorem 3) improving Equation (2.3) of [2], but there, the bound on $T^n g(x) - \pi(g)$ still appears with a third extra term (in comparison with (88)). Equation (89) is not given. They apply the result to the Metropolis algorithm.

Theorem 10. *Let (S, d) be a complete separable metric space with $d \leq 1$. Let $T(x, dy)$ be a transition operator on S such that for some function v bounded below by a positive number, some set $K \subset S$, some constant c:*

$$Tv \leq v - \theta(v) + c1_K, \tag{85}$$

where θ is a non-decreasing non-negative concave differentiable function on $[0, +\infty)$ with a derivative which tends to zero at infinity. We assume in addition that for the same set K, some $\varepsilon > 0$, some constant c, and any Lipschitz function g on S:

$$[Tg] \leq [g], \tag{86}$$

$$\left|Tg(x) - Tg(y)\right| \leq (1 - \varepsilon)\, d(x, y)[g], \qquad x, y \in K. \tag{87}$$

Then there exists a unique invariant measure π and for any Lipschitz function g, $x \in S$ and $n > 0$

$$\left|T^n g(x) - \pi(g)\right| \leq [g] \min\left(1, \frac{v(x)}{\psi^{(-1)}(n)}\right) + [g]\frac{c}{\theta(\psi^{(-1)}(n))}, \tag{88}$$

where ψ in given in Lemma 8. In addition

$$\left|T^n g(x) - T^n g(y)\right| \leq [g]\, d(x, y) \frac{v(x) + v(y) + c}{\psi^{(-1)}(n)} \tag{89}$$

which is true even without the assumption that $d \leq 1$.

Since for $0 \le x \le y$ one has $\frac{x}{y} \le \frac{\theta(x)}{\theta(y)}$ (the function $x \mapsto \frac{x}{y} - \frac{\theta(x)}{\theta(y)}$ is convex and non-positive at $x = 0$ and $x = y$), (88) leads to

$$\left|T^n g(x) - \pi(g)\right| \le [g] \frac{\theta(v(x)) + c}{\theta(\psi^{(-1)}(n))}.$$

This is Equation (2.3) obtained in [2], but with $\theta(v)$ instead of v, and without an extra exponent.

Equation (89) is interesting because it allows to estimate correlations: if the initial measure of the chain is μ, we have

$$\begin{aligned}\left|E\left[f(X_0)\left(g(X_n) - E[g(X_n)]\right)\right]\right| &= \left|\int f(x)\left(T^n g(x) - T^n g(y)\right)\mu(dy)\mu(dx)\right| \\ &\le \frac{[g]}{\psi^{(-1)}(n)} \int \left|f(x)\right| d(x,y)\left(v(x) + v(y) + c\right)\mu(dy)\mu(dx).\end{aligned}$$

Notice that the difference in convergence rate between (88) and (89) seems to shows that the forgetting of initial conditions holds at a strictly faster rate than the convergence to the invariant measure. This is due to the fact that the invariant measure may give strong weight to points with large value of v, points which are difficult for the Markov chain to reach, but are not important when comparing trajectories with close initial conditions.

Application of Theorem 7 to an expansive dynamical system. Consider the following application defined on [0, 1]

$$v(x) = \begin{cases} x\left(1 + 2^\gamma x^\gamma\right), & 0 \le x < 1/2, \\ 2x - 1, & 1/2 \le x \le 1, \end{cases} \tag{90}$$

where $0 < \gamma < 1$ is fixed, and the corresponding operator

$$Tf(x) = f\left(v(x)\right). \tag{91}$$

We are interested in the asymptotics of T^n. There exists an extensive literature on the subject [11,20] and the result we are going to present here, Equation (95), is already known [29]; our point is to give a new and direct proof of this estimate which plays a key role in the obtainment of central limit theorems (through the Gordin–Liverani theorem), and which is known to be optimal [26]. Notice that this proof does not require any explicit assumption on the invariant measure (see Equation (5.2) in [20]). We detail only here the example (90) but it will appear clearly that the following development extends to many other cases. Nevertheless, we feel that such extensions fall beyond the scope of this paper.

For any integrable function f on [0, 1], we set

$$F(x) = \int_0^x f(t)\,dt - x\bar{f}, \qquad \bar{f} = \int_0^1 f(t)\,dt.$$

We start with the following identity which we prove below:

$$\begin{aligned} Tf(x) &= \big(v'(x)^{-1}F\big(v(x)\big)\big)' + \big(\bar{f} - \big(v'(x)^{-1}\big)'F\big(v(x)\big)\big) \\ &= Vf(x) + Kf(x), \end{aligned} \tag{92}$$

where the prime denotes *in the whole present section* the density of the absolutely continuous part of the distributional derivative (which will always be a measure). We shall take $E = L_\infty([0,1])$:

$$\boldsymbol{|} f \boldsymbol{|} = \|f\|_\infty.$$

In order to prove (92), note that $v'(x)^{-1}F(v(x))$ is clearly Lipschitz because $F(v(x))$ cancels at the discontinuity point of v', implying that this function as well as its distributional derivative belongs to E with

$$\big(v'(x)^{-1}F\big(v(x)\big)\big)' = f\big(v(x)\big) - \bar{f} + \big(v'(x)^{-1}\big)'F\big(v(x)\big)$$

which proves (92). We obtain also by induction on n that

$$V^n f(x) = \big(v_n'(x)^{-1}F\big(v_n(x)\big)\big)', \tag{93}$$

where v_n is the nth iterate of v. In order to prove this, notice that $v_n'(x)^{-1}F(v_n(x))$ being Lipschitz, it is the integral of its derivative and (93) leads to

$$V^{n+1}f(x) = \big(v'(x)^{-1}\big(v_n'(\cdot)^{-1}F\big(v_n(\cdot)\big)\big)\big(v(x)\big)\big)' = \big(v_{n+1}'(x)^{-1}F\big(v_{n+1}(x)\big)\big)'.$$

On the other hand, it is proved by induction in appendix J that

$$v_n'(x) \ge c_1 n^{1/\gamma} v_n(x) \tag{94}$$

with $c_1 = (2^\gamma - 1)^{1/\gamma}$. Hence, if we consider the norm $\|f\| = \|\int_0^\cdot f(t)\,dt\|_\infty$, we are led to

$$\begin{aligned} \big\|V^n f\big\| &= \big\|v_n'(x)^{-1}F\big(v_n(x)\big)\big\|_\infty \\ &\le c_1^{-1}n^{-1/\gamma}\big\|x^{-1}F(x)\big\|_\infty \\ &\le c_1^{-1}n^{-1/\gamma}\sup_{0\le x\le 1} x^{-1}\int_0^x \big|f(y)\big|\,dy \\ &\le c_1^{-1}n^{-1/\gamma}\boldsymbol{|} f \boldsymbol{|}. \end{aligned}$$

Because $B = \{f \in E : \boldsymbol{|} f \boldsymbol{|} \le 1\}$ is $\|\cdot\|$-compact (F is 1-Lipschitz if $f \in B$), the assumptions of Theorem 1 and of Theorem 7 are all satisfied (but here $\|T^n\|$ is not bounded). Thanks to classical distortion arguments (see, for instance, [29] Theorem 1), one knows that T admits a unique absolutely continuous invariant probability measure π, which is ergodic and mixing. In particular, there is no nontrivial eigenfunction for any eigenvalue of modulus 1 and we can conclude that

$$\big\|T^n f - \pi(f)\big\| \le Cn^{1-1/\gamma}\boldsymbol{|} f \boldsymbol{|}. \tag{95}$$

Appendix A: Proof of Theorem 1

The proof of Theorem 1 requires two preliminary results which are the subject of the forthcoming section.

A.1. Asymptotically almost periodic powers of an operator

Theorem 11 below gives conditions under which, in some sense, the powers of an operator T can be rewritten

$$T^n = \sum_{i\geq 1} \lambda_i^n P_i + T^n P_0,$$

where each P_i is a projection, $P_i P_j = 0$, $i \neq j$, and $T^n P_0$ tends to zero in some sense. However, if each term of the series will be well defined (eigenspace and eigenvalue), the series may fail to converge, as in the case of almost periodic sequences; but since the set of points x for which $P_i x = 0$ except for a finite number of indices i will appear to be dense, the series $\sum_{i\geq 1} \lambda_i^n P_i x$ will converge at least on a dense subspace of E. Lemma 12 will give a condition under which there is only a finite number of non-zero λ_i's.

Let us say a few words concerning Assumptions (B1) and (B2) below, since they are the key assumptions and may appear somehow complicated; it is easily shown that under these assumptions, for any $x \in E$ the sequence $T^n x$ has $\|\cdot\|$-compact closure. These assumptions are essentially used to prove the total boundedness of the sequence $(T^n)_{n>0}$ for a certain norm (Step 1 of the proof of Theorem 11). These assumptions are reminiscent of that of the De Leeuw–Glicksberg theorem [4], but here we consider $\|\cdot\|$-total boundedness rather than $|\!|\!|\cdot|\!|\!|$-weak total boundedness (which is actually not a weaker assumption).

For the statement of this theorem, we refer to the equations (23) to (25).

Theorem 11. *Let T be a continuous operator on the Banach space $(E, |\!|\!|\cdot|\!|\!|)$ satisfying assumptions* (A0), (A1) *and*:

(B1) *The sequence T^n is uniformly $\|\cdot\|$-equicontinuous on $|\!|\!|\cdot|\!|\!|$-bounded sets in the following sense*:

$$\lim_{x\in B, \|x\|\to 0} \sup_n \|T^n x\| = 0. \tag{96}$$

(B2) *$T^n B$ is asymptotically $\|\cdot\|$-totally bounded in the following sense*: *There exist a sequence of finite sets $K_n \subset E$, and a sequence $\varepsilon_n \to 0$ such that for any $n \geq 0$*

$$T^n B \subset K_n + \varepsilon_n B_0. \tag{97}$$

Then the following facts hold true: *The space E is the direct sum of two $|\!|\!|\cdot|\!|\!|$-closed spaces*

$$E = \left\{x : \|T^n x\| \to 0\right\} \oplus \left\{x : \liminf_n \|x - T^n x\| = 0\right\} = E_0 \oplus E_c. \tag{98}$$

The projection P_c on E_c parallel to E_0 satisfies $|P_c| \leq C_T$. There exist a non-negative sequence ρ_n converging to 0 *such that*

$$\|T^n x\| \leq \rho_n |x|, \qquad x \in E_0, n \geq 0. \tag{99}$$

The space E_u of the finite linear combinations of eigenvectors with eigenvalue of modulus one is $\|\cdot\|$-dense in E_c.

The set Λ of these eigenvalues is at most countable, *and for each $\lambda \in \Lambda$ there exists a continuous projection P_λ on the corresponding eigenspace parallel to the others and to E_0. It satisfies $|P_\lambda| \leq C_T$ and*

$$\lim_{n\to\infty} \left\| P_\lambda x - \frac{1}{n}\sum_{i=0}^{n-1} \lambda^{-i} T^i x \right\| = 0, \qquad x \in E. \tag{100}$$

There exists a sequence k_i such that the projection P_c on E_c satisfies

$$\lim_{i\to\infty} \sup_{x\in B} \|P_c x - T^{k_i} x\| = 0. \tag{101}$$

The unit ball of E_c, *$B \cap E_c$*, *is $\|\cdot\|$-totally bounded.*

If the integer powers of T extend to a $\|\cdot\|$-C_0-semi-group $(T^t)_{t\geq 0}$, *that is*,

$$\forall x \in E, \qquad \lim_{t\to 0} \|T^t x - x\| = 0, \tag{102}$$

the space E_c is generated by the vectors x such that for some ω_x, *$T^t x = e^{i\omega_x t} x$ for any $t \geq 0$.*

Proof. *Step* 1: The non-negative powers of T form a totally bounded set for the distance

$$d(f,g) = \sup_{|x|\leq 1} \|f(x) - g(x)\|$$

on bounded functions on B. Any limit point of its closure is a continuous operator on $(E, |\cdot|)$, with norm $\leq C_T$.

We start with a simple modification of K_n in order to imbed it in $C_T B$. Fix $n > 0$, denote by $y_k, 1 \leq k \leq N_n$ the points of K_n, choose arbitrary N_n points $x_k \in T^n B$ such that $\|x_k - y_k\| \leq \varepsilon_n$ and define $\tilde{K}_n = \{x_k, 1 \leq k \leq N_n\}$. Assumption (B2) is still satisfied with $\tilde{K}_n$ but ε_n is now two times larger; in addition $\tilde{K}_n \subset C_T B$.

Hence there exist two functions u_n and v_n such that for $|x| \leq 1$

$$T^n x = u_n(x) + v_n(x), \qquad u_n(x) \in \tilde{K}_n,$$

and

$$\|v_n(x)\| \leq 2\varepsilon_n, \qquad |v_n(x)| \leq 2C_T. \tag{103}$$

Fix n large; for any p:

$$
\begin{aligned}
T^{2n+p}x &= C_T T^n\big(C_T^{-1}T^p u_n(x)\big) + T^{n+p}v_n(x)\\
&= C_T u_n\big(C_T^{-1}T^p u_n(x)\big) + C_T v_n\big(C_T^{-1}T^p u_n(x)\big) + T^{n+p}v_n(x)\\
&= \alpha_p(x) + \beta_p(x) + \gamma_p(x).
\end{aligned}
$$

The set of functions $\{\alpha_p(\cdot),\, p \geq 0\}$ has at most $N_n^{N_n}$ elements; clearly $\|\beta_p(x)\| \leq 2C_T\varepsilon_n$; and Assumption (B1) with Equation (103) implies that $\|\gamma_p(x)\| \leq \eta_n$, for all $p \geq 0$ and some sequence $\eta_n \to 0$. We have just proved that the set $\{T^k, k \geq 2n\}$ can be covered with $N_n^{N_n}$ d-balls of radius $2C_T\varepsilon_n + \eta_n$; hence $\{T^k, k \geq 0\}$ is totally bounded for the distance d.

For any $x \in B$, the sequence $T^n x$ belongs to $C_T B$, hence any $\|\cdot\|$-cluster point of this sequence belongs to $C_T B$ (because of (A0)), and the continuity follows.

Step 2: For any limits $d(T^{u_k}, U) \to 0$ and $d(T^{v_k}, V) \to 0$, one has $d(T^{u_j+v_k}, UV) \to 0$ if $\min(j,k) \to +\infty$. In particular $UV = VU$ and for any third similar limit operator W, $d(WU, WV) \leq C_T d(U,V)$.

One has indeed:

$$
\begin{aligned}
d\big(T^{u_j+v_k}, UV\big) &\leq d\big(T^{u_j+v_k}, T^{u_j}V\big) + d\big(T^{u_j}V, UV\big)\\
&\leq \sup\big\{\big\|T^{u_j}x\big\| : \|x\| \leq d\big(T^{v_k}, V\big), \lvert x \rvert \leq 2C_T\big\} + d\big(T^{u_j}, U\big)\lvert V \rvert.
\end{aligned}
$$

The second term obviously converges to zero, and the first one also because of Assumption (B1). For the last assertion

$$
d(WU, WV) = d(UW, VW) \leq d(U,V)\lvert W \rvert.
$$

Step 3: Proof of Equations (98) and (101).

Let n_k be a sequence such that T^{n_k} d-converges to some limit S. We can assume that $n_k - n_{k-1} \to \infty$. From the sequence $n_k - n_{k-1}$ one can extract a sequence $p_i = n_{k_i+1} - n_{k_i}$ such that T^{p_i} and T^{p_i-1} d-converge to some limit P_c and R. Set $m_i = n_{k_i}$.

$$
S = d\text{-}\lim T^{m_i+p_i} = SP_c.
$$

Since $p_i \to \infty$, there exists $q_i \to \infty$ such that $P_c = d\text{-}\lim T^{m_i+q_i}$ and we get

$$
P_c = d\text{-}\lim T^{m_i}T^{q_i} = d\text{-}\lim ST^{q_i} = d\text{-}\lim P_c S T^{q_i} = P_c^2.
$$

P_c is a projection on P_cE and Equation (101) holds. We shall prove now that P_cE is indeed E_c and that (98) holds true.

Clearly $P_cE \subset E_c$. On the other hand, for any $x \in E_c$ there exists a sequence r_k such that $\|x - T^{r_k}x\|$ converges to 0. We can assume that $r_k > p_k$ and that $d(T^{r_k-p_k}, U) \to 0$ for some U; in particular $d(T^{r_k}, P_cU) \to 0$. Hence, $x = P_cUx \in P_cE$. Finally, $P_cE = E_c$. The null space of P_c clearly contains E_0. On the other hand for any point $x \notin E_0$, there exists a sequence r_k such that $\|T^{r_k}x\| \geq \varepsilon$ and $T^{r_k-p_k}$ d-converges to some limit V; the bound $\|VP_cx\| \geq \varepsilon$ leads to $P_c x \neq 0$. This implies by contradiction that any point of the null space of P_c belongs to E_0; hence the null space of P_c is E_0 and $E = E_0 \oplus E_c$.

The bound on the norm of P_c is a consequence of the last point of Step 1.

Step 4: T is one-to-one on E_c. The powers of T on E_c generate a compact G group of operators on E_c with the distance

$$d_c(f, g) = \sup_{\|x\| \leq 1, x \in E_c} \|f(x) - g(x)\|.$$

Since $TP_c = P_c T$ and $P_c = TR = RT$ (R is defined in Step 3), E_c is T-stable and R is its inverse on E_c. The monoid generated by the powers of T

$$G = \overline{\{T^n, n \geq 0\}}$$

is a group since we have seen that $R \in G$. The continuity of the multiplication on G comes from Step 2, and the compactness from Step 1.

Step 5: E_u is $\|\cdot\|$-dense in E_c. Properties of P_λ.

Each character χ on G is uniquely determined by the value of $\chi(T)$, because of the definition of G and $\chi(T^n) = \chi(T)^n$.

For any eigenvalue λ of T with modulus 1, there exists a unique character χ such that $\chi(T) = \lambda$ which can be defined as follows: pick an eigenvector x, a $\|\cdot\|$-continuous linear form u such that $u(x) = 1$ and set $\chi(S) = u(Sx)$; χ is indeed a character since it is d_c-continuous with $\chi(T^n) = \chi(T)^n$; in particular since the set of characters of a compact group is at most countable, there is at most a countable number of eigenvalues of modulus one.

In order to show now that for any character χ, $\chi(T)$ is an eigenvalue, we proceed as follows. Let μ be the Haar measure on G, consider a character χ on G and define

$$Q_\chi = \int_G \chi(S)^{-1} S \mu(dS) \tag{104}$$

(as a continuous function on G, $f(S) = S$ is the uniform limit of simple functions (by compactness) and this integral is well defined with the usual properties, cf. [8], Section III.2). If x is a $\chi(T)$-eigenvector, then the relation $T^n x = \chi(T^n)x$ extends to G as $Sx = \chi(S)x$, and clearly $Q_\chi x = x$.

The invariance of μ implies that for $U \in G$:

$$Q_\chi = \int_G \chi(SU)^{-1} SU \mu(dS) = \chi(U)^{-1} U Q_\chi. \tag{105}$$

In particular, taking $U = T$, for any $x \in E$, $Q_\chi x$ is 0 or an eigenvector with eigenvalue $\chi(T)$. In addition integrating this expression w.r.t. $\mu(dU)$ we get that Q_χ is a projector. If Q_χ is non-zero, Q_χ is thus a projector on the $\chi(T)$-eigenspace. If $Q_\chi = 0$, for any $\|\cdot\|$-continuous linear form u on E and $y \in E$, one has

$$\int_G \chi(S)^{-1} u(Sy) \mu(dS) = 0.$$

The Fourier transform of $S \mapsto u(Sy)$ being 0, this d_c-continuous function is itself 0. Hence $u(Sy) = 0$ for any such u and y and any $S \in G$, which is impossible. Hence, Q_χ is non-zero, $\chi(T)$ is an eigenvalue, an Q_χ is a projection whose range is the $\chi(T)$-eigenspace.

In summary, there is a one-to-one correspondence between characters and eigenvalues with modulus one, defined by $\lambda = \chi(T)$, and Q_χ is a projector whose range is exactly the eigenspace.

Since $|S| \le C_T$ we have $|Q_\chi| \le C_T$, and since by (104) they commute, Q_χ is a projector parallel to the other eigenspaces.

In order to show that E_u is $\|\cdot\|$-dense in E_c, consider a $\|\cdot\|$-continuous linear form u such that $u(x) = 0$ for any eigenvector x, then for any $y \in E_c$, $S \mapsto u(Sy)$ is d_c-continuous and for any character χ one has

$$\int_G \chi(S)^{-1} u(Sy)\mu(dS) = u\big(Q_\chi(y)\big) = 0.$$

The Fourier transform of $S \mapsto u(Sy)$ being 0, this continuous function is itself 0. Hence, $u(y) = 0$. E_u is $\|\cdot\|$-dense in E_c. The projection P_λ is finally well defined on E by setting $P_\lambda x = Q_{\chi(T)}x$ if $x \in E_c$ and $P_\lambda x = 0$ if $x \in E_0$.

We now prove (100). This equation holds on E_0 and E_u. Set

$$P_{\lambda,n} = \frac{1}{n}\sum_{i=0}^{n-1} \lambda^{-i} T^i.$$

For any $x \in E_c$ we can pick out $y \in E_u$ such that $\|x - y\| \le \varepsilon$ and get

$$\begin{aligned}\|P_{\lambda,q}x - P_{\lambda,n}x\| &\le \big\|P_{\lambda,q}(x-y)\big\| + \big\|P_{\lambda,n}(x-y)\big\| + \|P_{\lambda,q}y - P_{\lambda,n}y\| \\ &\le 2\sup_k \big\|T^k(x-y)\big\| + \|P_{\lambda,q}y - P_{\lambda,n}y\|.\end{aligned}$$

Since this quantity can be made smaller than 3ε by taking n and q large, this proves that $P_{\lambda,q}x$ is a $\|\cdot\|$-Cauchy sequence, and its limit $P_\lambda x$ satisfies (100). Since for all $x \in E$, $|P_{\lambda,n}x| \le C_T|x|$ and $\|P_{\lambda,n}x - P_\lambda x\| \to 0$, Assumption (A0) implies that $|P_\lambda x| \le C_T|x|$.

Step 6: Equation (99).

Using a sequence p_k such that $d(T^{p_k}, P_c) = \alpha_k \to 0$, we obtain $\|T^{p_k}x\| \le \alpha_k$ for $x \in B \cap E_0$. For $n \ge p_k$ large, one can write $\|T^n x\| \le \|T^{p_k}(T^{n-p_k}x)\| \le C_T\alpha_k$. This implies (99).

Step 7: $B_c = E_c \cap B$ is $\|\cdot\|$-totally bounded.

Using the same sequence p_k, we get with (97)

$$B_c \subset \big(P_c - T^{p_k}\big)B_c + T^{p_k}B_c \subset \alpha_k B_0 + K_{p_k} + \varepsilon_{p_k}B_0.$$

This means that B_c is $\|\cdot\|$-totally bounded.

Step 8: Case of semi-group T^t.

We can carry on Steps 1 to 4 with $t \in \mathbb{R}_+$ instead of $n \in \mathbb{N}$. The group G is now $G = \overline{\{T^s, s \ge 0\}}$. In Equation (105) we take $U = T^t$ and we obtain that $y = P_\chi x$ is a vector such that $T^t y = \chi(T^t)y$. In particular if $y \ne 0$, we have $\chi(T^{s+t}) = \chi(T^s)\chi(T^t)$, and on the other hand assumption (102) implies that the function $t \to \|T^t y\|$ is continuous, and so is $t \to \chi(T^t)$; hence $\chi(T^t) = e^{i\omega t}$ for some $\omega \in \mathbb{R}$. □

The following lemma gives a condition for checking that E_c is finite dimensional. This could be checked specifically on examples but we shall see in Theorem 1 that this holds naturally in general situations; in addition, this finite dimensionality assumption is very important in Theorem 7.

Lemma 12. *If in addition to* (A0) *and* (A1), T *is* $\|\cdot\|$*-continuous and satisfies the following assumption*:

(B1′) *There exists two sequences* $\eta_n \to 0$ *and* $\eta'_{n,p} \to 0$ (*as* $\min(n,p) \to \infty$), *such that for any* $n, p > 0$

$$T^n\big(B \cap p^{-1}B_0\big) \subset \eta_n B_0 + \eta'_{n,p} B,$$

then (B1) *is also satisfied*. *If* (B2) *is also satisfied*, *then* (9) *to* (13) *hold true and*

$$\big\|Q^n x\big\| \le \rho_n |x|, \qquad \rho_n \to 0. \tag{106}$$

Proof. We start with (B1). We have to prove that any sequence x_p of B such that $\|x_p\| \to 0$ satisfies $\sup_{n>0} \|T^n x_p\| \to 0$. Without loss of generality, we can assume that $\|x_p\| \le 1/p$. One has

$$\big\|T^n x_p\big\| \le \eta_n + \eta'_{n,p} C_0.$$

Since on the other hand

$$\big\|T^n x_p\big\| \le \|T\|^n \|x_p\|$$

we have for any n_0

$$\sup_{n>0}\big\|T^n x_p\big\| \le \max_{n\ge n_0}\big(\eta_n + \eta'_{n,p} C_0\big) + \frac{1}{p}\max_{n<n_0}\|T\|^n$$

which can be made arbitrarily small by taking n_0 large first and then by increasing p.

Let us prove now that E_c is finite-dimensional. It suffices to prove that $B_c = E_c \cap B$ is $|\cdot|$-totally bounded; since we already know that $E_c \cap B$ is $\|\cdot\|$-totally bounded, it suffices to prove that $|\cdot|$ and $\|\cdot\|$ induce the same topology on B_c. Notice first that if $x \in E$ and $\|x - x_n\| \to 0$ then

$$|x| \le \varlimsup_n |x_n|$$

because of (A0) (the inequality is obviously true if $|x_n|$ is not bounded). Let $x \in B \cap E_c$. We want to prove that $|x|$ can be made arbitrarily small by taking $\|x\|$ small enough. Consider an integer p such that $\|x\| \leq p^{-1}$. There exists a sequence n_k such that $\|x - T^{n_k}x\|$ tends to zero. Thanks to (B1′), there exist $u_k \in B_0$ and $v_k \in B$ such that

$$T^{n_k}x = \eta_{n_k}u_k + \eta'_{n_k,p}v_k.$$

Since $\|x - T^{n_k}x + \eta_{n_k}u_k\|$ tends to zero, using the previous remark:

$$|\!|\!|x|\!|\!| \leq \overline{\lim_k} |\!|\!|T^{n_k}x - \eta_{n_k}u_k|\!|\!| = \overline{\lim_k} |\!|\!|\eta'_{n_k,p}v_k|\!|\!| \leq \overline{\lim_k} \eta'_{n_k,p}$$

which can be made arbitrarily small by taking p large. Hence, $|\!|\!|\cdot|\!|\!|$ and $\|\cdot\|$ are topologically equivalent on E_c and the compactness holds.

Now that E_c is finite dimensional, Equations (9) to (13) and (106) are an immediate rewording of the conclusion of Theorem 11 (notice that ρ_n has changed from equation (99) by a factor $|\!|\!|P_0|\!|\!|$). □

A.2. Proof of Theorem 1

Let us recall the identity (31)

$$T^n = \sum_{i=1}^{n} T^{n-i}(T-V)V^{i-1} + V^n = \sum_{i=1}^{n} T^{n-i}KV^{i-1} + V^n. \tag{107}$$

In particular, Assumption (A1) together with (29) implies that the sequence $|\!|\!|V^n|\!|\!|$ is bounded by a constant C_V, and KV^nKB is $\|\cdot\|$-totally bounded. We set $\alpha_n = |\!|\!|KV^n|\!|\!|$ and $\bar{\alpha}_k = \sum_{i=k}^{\infty} \alpha_i$. Let $x \in E$, for any $0 \leq k \leq n$:

$$\begin{aligned}
|\!|\!|(T^n - V^n)x|\!|\!| &\leq \sum_{i=1}^{n} |\!|\!|T^{n-i}KV^{i-1}x|\!|\!| \\
&\leq C_T \sum_{i=1}^{k} |\!|\!|KV^{i-1}x|\!|\!| + C_T \sum_{i=k+1}^{n} |\!|\!|KV^{i-1}x|\!|\!| \\
&\leq C_T C_K \sum_{i=1}^{k} \|V^{i-1}x\| + C_T\bar{\alpha}_k |\!|\!|x|\!|\!| \\
&\leq c_k\|x\| + C_T\bar{\alpha}_k |\!|\!|x|\!|\!|
\end{aligned}$$

for some c_k. In particular if $x \in B \cap p^{-1}B_0$ one has

$$|\!|\!|(T^n - V^n)x|\!|\!| \leq \min_{k \leq n}\left(\frac{c_k}{p} + C_T\bar{\alpha}_k\right).$$

This implies (B1′) where $\eta'_{n,p}$ is the right-hand side of the previous equation and $\eta_n = \varepsilon'_n$. We proceed now with (97):

$$\begin{aligned} T^n &= \sum_{i=1}^{n} T^{i-1} K V^{n-i} + V^n \\ &= \sum_{i=1}^{n} \left(\sum_{j=1}^{i-1} T^{j-1} K V^{i-j-1} + V^{i-1} \right) K V^{n-i} + V^n \\ &= \sum_{1 \le j < i \le n} T^{j-1} K V^{i-j-1} K V^{n-i} + \sum_{i=1}^{n} V^{i-1} K V^{n-i} + V^n \\ &= A_n + B_n + C_n. \end{aligned} \tag{108}$$

The set $A_n B$ is $\|\cdot\|$-totally bounded; on the other hand

$$C_n B + B_n B \subset \left(\varepsilon'_n + \sum_{i=1}^{n} \alpha_{n-i} \varepsilon'_{i-1} \right) B_0.$$

The sum tends to zero as n tends to infinity and this leads finally to (97).

We turn now to the last assertion. If T^k satisfies (B1) and (B2) and T is $|\cdot|$-continuous and $\|\cdot\|$-continuous, clearly T also satisfies (B1) and (B2). Theorem 11 applies to T. Since any eigenvector of T associated with an eigenvalue of modulus one is an eigenvector of T^k associated with an eigenvalue of modulus one, E_c is finite dimensional, and (9) to (13) and (22) hold.

Appendix B: Proof of Theorem 2

(A0) is clearly satisfied. In addition T is a $|\cdot|$-contraction, and (A1) holds true. Up to a replacement of v with v/c_v, we can assume that $c_v = 1$. Since $T1 = 1$, Equations (34), (35) and (36) imply

$$Vv \le Tv \le v - 1 + c 1_{K_0}, \tag{109}$$

$$V1 \le 1 - \varepsilon 1_{K_0} \tag{110}$$

for some $c > 0$. Combining these equations, we obtain that the function $\bar{v} = v + c/\varepsilon$ satisfies

$$V\bar{v} \le \bar{v} - 1. \tag{111}$$

Multiplying (111) by V^k and summing, up we obtain

$$V^n \bar{v} + \sum_{k=0}^{n-1} V^k 1 \le \bar{v}. \tag{112}$$

Equation (30) is obvious from (38) and (39) and Equation (29) is a consequence of (112) and (39) since for any f

$$\left| KV^n f \right| \le \nu\left(\left|V^n f\right|\right) \le \|f\|_\infty \nu\left(V^n 1\right) \le |f| \nu\left(V^n 1\right).$$

For (28) notice that $V^n 1$ is a decreasing sequence of functions, because $V1 \le 1$, hence:

$$V^n 1 \le \frac{1}{n}\sum_{k=0}^{n-1} V^k 1 \le \frac{\bar{v}}{n}$$

and (28) holds. It remains to prove the compactness of KT^pK. Notice first that in the assumptions we can replace ν with $\bar{\nu}$ defined as

$$\bar{\nu}(f) = \sum_{i\ge 0} 2^{-i}\nu\left(T^i f\right),$$

which makes T continuous on $L_1(\bar{\nu})$ with norm ≤ 2. Second, notice that $\bar{\nu}(\bar{v}) < \infty$, that is (39) still holds. From (38), we get that $\|Kf\|_\infty \le \nu(|f|)$ implies that the measure $\mu(g) = \int g(x,y)K(x,dy)\bar{\nu}(dx)$ is absolutely continuous w.r.t. $\bar{\nu}(dx) \otimes \bar{\nu}(dy)$, and let $p(x,y)$ be its density; if g has the form $g(x,y) = h(x)f(y)$, one has

$$\int h(x)f(y)p(x,y)\bar{\nu}(dx)\bar{\nu}(dy) = \int h(x)(Kf)(x)\bar{\nu}(dx)$$

hence one has for any bounded measurable function f and for $\bar{\nu}$-a.e. x,

$$Kf(x) = \int f(y)p(x,y)\bar{\nu}(dy).$$

The function p can be approximated in $L_1(\bar{\nu} \otimes \bar{\nu})$ as

$$p(x,y) = \sum_{i=1}^{n} q_i(x)r_i(y) + \rho(x,y), \qquad \int \left|\rho(x,y)\right|\bar{\nu}(dx)\bar{\nu}(dy) < \varepsilon.$$

This finite rank approximation implies that K is a compact operator of $\mathcal{L}(E, L_1(\nu))$, hence KB is totally bounded in $L_1(\bar{\nu})$. By continuity, the same property holds for T^pKB. Equation (38) implies now that KT^pKB is totally bounded in $(E, |\cdot|)$. The assumptions of Theorem 1 are thus satisfied.

To obtain (42), it remains to prove that the space E_c is one dimensional. For this, let n_k be a sequence such that $\lambda_i^{n_k} \to \lambda_i$ for each eigenvalue λ_i with modulus 1, and denote by P_c the projector on E_c parallel to E_0. Then $\|T^{n_k}f - TP_cf\|$ converges to 0 for any $f \in E$. Hence, TP_c is a Markov transition operator with the same one-modulus eigenvectors as T. It is compact on E and if there exists more than one eigenvector, one can find two non-trivial measurable sets A and B such that $TP_c 1_A = 1_B$ ([25] Chapter 6, Section 3, Theorem 3.7). Notice now that the function $f = P_c 1_A$ satisfies $0 \le f \le 1$ and by Jensen's inequality

$$T\left(f^n\right) \ge (Tf)^n = 1_B.$$

On the other hand, since $f^n \le f$, we have $T(f^n) \le Tf = 1_B$, and we obtain that $T(f^n) = 1_B$ for all $n > 0$; letting n tend to infinity, we get

$$T(1_{f=1}) = 1_B.$$

Appendix C: Proof of Theorem 3

We begin with the case $q = 1$.

Elementary inductions lead to

$$\begin{aligned}\left|T^n x\right| &\le \gamma^n |x| + c\left\|T^{n-1}x\right\| + \gamma c\left\|T^{n-2}x\right\| + \cdots + \gamma^{n-1}c\|x\| \\ &\le \gamma^n |x| + c_n\|x\|.\end{aligned} \tag{113}$$

This may be improved as

$$\left|T^n x\right| \le C_T \min_{k\le n}\left|T^k x\right| \le C_T \min_{k\le n}\bigl(\gamma^k |x| + c_k\|x\|\bigr).$$

This implies (B1′) of Lemma 12 with $\eta_n = 0$ and

$$\eta'_{n,p} = C_T \min_{k\le n}\left(\gamma^k + \frac{c_k}{p}\right). \tag{114}$$

We have similarly

$$T^n B \subset \gamma^n B + \gamma^{n-1}K_B + \gamma^{n-2}TK_B + \cdots + T^{n-1}K_B$$

and this implies now (B2) in Theorem 11.

It remains to prove that Q [from equation (9)] has a spectral radius < 1. Notice that for any $n > 0$, $Q^n = T^{n-1}Q$, this proves that $C_Q = \sup_n |Q^n|$ is finite. For any $x \in B$ we have from (9), (113) and (22)

$$\left|Q^{n+k}x\right| = \left|T^n Q^k x\right| \le (C_Q + 1)\left|T^n \frac{Q^k x}{|Q^k x| + 1}\right| \le (C_Q + 1)\eta'_{n,p}$$

with

$$p^{-1} = \frac{\|Q^k x\|}{|Q^k x| + 1} \le \rho_k.$$

By choosing n and k large enough, this ensures that some power of Q is a $|\cdot|$-contraction.

If now $q > 1$, the operator T^q satisfies the assumptions for the case $q = 1$, thus T^q satisfies (A1), (B1′) and (B2). Since T is $|\cdot|$ and $\|\cdot\|$-continuous, this clearly implies that T also satisfies these assumptions, by writing $T^n = T^{r+kq}$ with $0 \le r < q$.

Appendix D: Proof of Theorem 4

For the proof, we shall change $\|\cdot\|$ into

$$\|f\|' = \sup_x \frac{|f(x)|}{v'(x)}, \qquad v'(x) = \frac{v(x)+A}{1+A}$$

for some constant $A \geq 1$ which will be chosen later, and $|\!|\!| f |\!|\!|$ as

$$|\!|\!| f |\!|\!|' = \|f\|' + \alpha[f]$$

for a small constant α, and prove that the assumptions of Theorem 3 are fulfilled with $q = 1$. Notice that $\|f\| \leq \|f\|'$. For any $f \in E$, by the positivity of T and Equation (50),

$$\begin{aligned}
\left|Tf(x)\right| &\leq \|f\|' T v'(x) \\
&\leq \|f\|' \frac{\gamma_v v(x) + A + c_v}{1+A} \\
&\leq \|f\|' \left(v'(x) + \frac{c_v}{1+A} \right)
\end{aligned} \tag{115}$$

hence

$$\|Tf\|' \leq \left(1 + \frac{c_v}{A}\right)\|f\|'. \tag{116}$$

T is $\|\cdot\|'$-continuous. In addition, Equation (50) implies that for any $n > 0$

$$T^n v(x) \leq \gamma_v^n v(x) + \frac{c_v}{1-\gamma_v} \tag{117}$$

hence $\|T^n\|'$ is bounded. Equation (49) with (116) implies (46) with $\gamma = \gamma_b$, and $c = 1 + c_v/A$. With (117), it implies also that $|\!|\!| T^n |\!|\!|'$ is bounded. Thus (A0), (A1) and (46) are satisfied.

In order to prove that Theorem 3 applies, it remains to prove that (45) holds true. Consider $A_0 > 0$ which will be chosen large enough later, and η small; if $v(x) \leq A_0$ the set $O_x = \{y : d(x,y) \leq \eta\} \cap \{v < 2A_0\}$ is still an open neighbourhood of x because v is continuous. Consider a finite sequence $(x_i)_{1 \leq i \leq I}$ such that $v(x_i) \leq A_0$ and $\{v \leq A_0\} \subset \bigcup_{i=1}^{I} O_{x_i}$. This is possible thanks to the compactness of $\{v \leq A_0\}$. There exist $\theta_1(x), \theta_2(x), \ldots, \theta_{I+1}(x)$ a locally Lipschitz partition of the unity of S such that the support of each θ_i, $i \leq I$, is contained in O_{x_i}, and the support of θ_{I+1} is contained in $\{x : v(x) > A_0\}$ (see [1], Theorem 2, page 10). We define $\varphi = 1 - \theta_{I+1}$ which is 0 on $\{v \geq 2A_0\}$ and 1 on $\{v \leq A_0\}$. We split Tf as

$$\begin{aligned}
Tf(x) &= \left(\sum_{i=1}^{I} \big\{Tf(x) - \varepsilon\varphi(x) Kf(x_i)\big\}\theta_i(x) + Tf(x)\theta_{I+1}(x) \right) + \varepsilon\varphi(x)\sum_{i=1}^{I} Kf(x_i)\theta_i(x) \\
&= Vf(x) + Sf(x).
\end{aligned}$$

Clearly, for $\lvert f\rvert \le 1$, Sf belongs to a fixed $\|\cdot\|$-compact set because the sum is finite. We are going to show that

$$\lvert Vf\rvert \le \gamma_2 \lvert f\rvert \tag{118}$$

for some $\gamma_2 < 1$; this will imply (45). One has

$$\begin{aligned}
[Vf] &\le [Tf] + \varepsilon \sum_i \big|Kf(x_i)\big|[\varphi\theta_i] \\
&\le \gamma_b[f] + \varepsilon\|f\| \sum_i \big(\gamma_v v(x_i) + c_v\big)[\varphi\theta_i] \\
&\le \gamma_b[f] + \varepsilon c_0\|f\|', \qquad c_0 = (A_0 + c_v)\sum_i [\varphi\theta_i].
\end{aligned} \tag{119}$$

It is more complicated to bound $\|Vf\|'$. For $i \le I$ and $\theta_i(x) > 0$ then $d(x, x_i) \le \eta$ and Equations (115) and (53) imply that

$$\begin{aligned}
&\big|Tf(x) - \varepsilon\varphi(x)Kf(x_i)\big| \\
&\quad = \big|\big(1 - \varepsilon\varphi(x)\big)Tf(x)\big| + \big|\varepsilon\varphi(x)\big(Tf(x) - Kf(x)\big)\big| + \big|\varepsilon\varphi(x)\big(Kf(x) - Kf(x_i)\big)\big| \\
&\quad \le \big(1 - \varepsilon\varphi(x)\big)\big(\gamma_v v(x) + c_v + A\big)\frac{\|f\|'}{1+A} + \varepsilon\varphi(x)(T-K)v'(x)\|f\|' + \varepsilon c_1\big([f] + \psi(\eta)\|f\|\big),
\end{aligned}$$

where c_1 is the maximum of τ on $\{\varphi > 0\}$. Since $\varphi(x) > 0$ implies $v(x) \le 2A_0$, if we denote by γ_0 the maximum of $1 - \varepsilon_d$ on $\{v \le 2A_0\}$ the second term can be bounded as

$$(T-K)v'(x) \le \frac{Tv(x) + A - K(v+A)(x)}{1+A} \le \frac{\gamma_v v(x) + c_v + \gamma_0 A}{1+A} \le \gamma_d v'(x)$$

with $\gamma_d = \max(\gamma_v, \gamma_0 + c_v/A)$. Notice that $\gamma_d < 1$ as soon as $A > c_v/(1-\gamma_0)$. Our bound becomes

$$\begin{aligned}
&\big|Tf(x) - \varepsilon\varphi(x)Kf(x_i)\big| \\
&\quad \le \big(1 - \varepsilon\varphi(x)\big)\big(\gamma_v v(x) + c_v + A\big)\frac{\|f\|'}{1+A} + \varepsilon\gamma_d v'(x)\|f\|' + \varepsilon c_1[f] + \varepsilon c_1\psi(\eta)\|f\|.
\end{aligned}$$

If in this expression, $\varphi(x) < 1$, then $v(x) \ge A_0$ and

$$\begin{aligned}
\big(1 - \varepsilon\varphi(x)\big)\big(\gamma_v v(x) + c_v + A\big) &\le \gamma_v v(x) + c_v + A \\
&\le \left(\sup_{u \ge A_0} \frac{\gamma_v u + c_v + A}{u + A}\right)\big(v(x) + A\big) \\
&= \frac{\gamma_v A_0 + c_v + A}{A_0 + A}\big(v(x) + A\big)
\end{aligned}$$

and if $\varphi(x) = 1$:

$$\big(1-\varepsilon\varphi(x)\big)\big(\gamma_v v(x)+c_v+A\big) \le (1-\varepsilon)\left(\sup_{u\ge 0}\frac{\gamma_v u+c_v+A}{u+A}\right)\big(v(x)+A\big)$$
$$= (1-\varepsilon)\left(\frac{c_v}{A}+1\right)\big(v(x)+A\big).$$

In any case, we get

$$\begin{aligned} &\big|Tf(x)-\varepsilon\varphi(x)K_{x_i}f(x_i)\big| \\ &\le \gamma_1\|f\|'v'(x)+\varepsilon\gamma_d\|f\|'v'(x)+\varepsilon c_1[f]+\varepsilon c_1\psi(\eta)\|f\| \end{aligned} \tag{120}$$

with

$$\gamma_1 = \max\left(\frac{\gamma_v A_0+c_v+A}{A_0+A}, (1-\varepsilon)\left(1+\frac{c_v}{A}\right)\right).$$

In order to bound the factor of θ_{I+1} in the expression of Vf, we notice that in the case where $\theta_{I+1}(x) > 0$, we have $v(x) \ge A_0$ and

$$\begin{aligned} \big|Tf(x)\big| &\le \frac{\gamma_v v(x)+c_v+A}{1+A}\|f\|' \\ &= \left(\gamma_v v'(x)+\frac{c_v+(1-\gamma_v)A}{1+A}\right)\|f\|' \\ &\le \left(\gamma_v+\frac{c_v+(1-\gamma_v)A}{A_0+A}\right)\|f\|'v'(x) \\ &\le \frac{c_v+A+\gamma_v A_0}{A_0+A}\|f\|'v'(x) \\ &\le \gamma_1\|f\|'v'(x). \end{aligned} \tag{121}$$

Since (120) is true if $\theta_i(x) > 0$, and (121) holds if $\theta_{I+1}(x) > 0$, we obtain for all x

$$\big|Vf(x)\big| \le \gamma_1 v'(x)\|f\|'+\varepsilon\gamma_d\|f\|'v'(x)+\varepsilon c_1[f]+\varepsilon c_1\psi(\eta)\|f\|$$

thus

$$\|Vf\|' \le \big(\gamma_1+\varepsilon c_1\psi(\eta)+\varepsilon\gamma_d\big)\|f\|'+\varepsilon c_1[f] \tag{122}$$

and combining (122) and (119) leads to

$$\|Vf\|'+\alpha[Vf] \le \big(\gamma_1+\varepsilon c_1\psi(\eta)+\varepsilon\gamma_d+\varepsilon\alpha c_0\big)\|f\|'+(\alpha\gamma_b+\varepsilon c_1)[f]. \tag{123}$$

In order to get (118) for some $\gamma_2 < 1$, we need simultaneously:

$$\frac{\gamma_v A_0 + c_v + A}{A_0 + A} + \varepsilon c_1 \psi(\eta) + \varepsilon \gamma_d + \varepsilon \alpha c_0 < 1,$$
$$1 + \frac{c_v}{A} - \varepsilon\left(1 + \frac{c_v}{A} - c_1 \psi(\eta) - \gamma_d - c_0 \alpha\right) < 1,$$
$$\gamma_b + \varepsilon \frac{c_1}{\alpha} < 1.$$

In other words, it suffices that

$$\varepsilon\bigl(\gamma_d + c_1 \psi(\eta) + c_0 \alpha\bigr) < \frac{A_0 - \gamma_v A_0 - c_v}{A_0 + A},$$
$$c_v < \varepsilon A\bigl(1 - \gamma_d - c_1 \psi(\eta) - c_0 \alpha\bigr),$$
$$\varepsilon \frac{c_1}{\alpha} < 1 - \gamma_b.$$

Remember that c_0 and c_1 depend on A_0, and

$$1 - \gamma_d = \min\Bigl(1 - \gamma_v, \min_{v \le 2A_0} \varepsilon_d(x) - c_v/A\Bigr).$$

Assumption (62) implies that for some $B > 0$, $\varepsilon(x) > 14c_v/v(x)$ for $v(x) > B$, and if A_0 is such that $\varepsilon_d(x) > 7c_v/A_0$ for $v(x) < B$ then $\varepsilon_d(x) > 7c_v/A_0$ for $v(x) \le 2A_0$. Thus, if A_0 is large enough, and $A \ge A_0$ (A will be chosen later)

$$1 - \gamma_d \ge \frac{6c_v}{A_0}. \tag{124}$$

This makes our choice of A_0, together with the condition $A_0 - \gamma_v A_0 - c_v \ge 1$. We choose now η such that $c_1 \psi(\eta) \le c_v/A_0$ and

$$\alpha = \frac{1 - \gamma_d - 2c_1 \psi(\eta)}{2c_0}.$$

With this choice of α, our equation set becomes

$$\frac{1}{2}\varepsilon(1 + \gamma_d) < \frac{A_0 - \gamma_v A_0 - c_v}{A_0 + A},$$
$$2c_v < \varepsilon A(1 - \gamma_d),$$
$$2\varepsilon c_0 c_1 < (1 - \gamma_b)\bigl(1 - \gamma_d - 2c_1 \psi(\eta)\bigr)$$

and by (124), with $A_0 - \gamma_v A_0 - c_v \ge 1$, this is implied by

$$\varepsilon < \frac{1}{A_0 + A},$$

$$A_0 < 3\varepsilon A,$$

$$\varepsilon c_0 c_1 A_0 < 2(1-\gamma_b)c_v.$$

If we take $\varepsilon = A_0/2A$, these equation are satisfied for A large enough, as well as the condition $A \geq c_v/(1-\gamma_0)$ that has been required before.

It remains to prove the last assertion. Since 1 is the only eigenvalue of modulus one and since its multiplicity is one, there exists a linear form π on E such that (56) holds. This equation implies that for $f \in E$

$$\left\|\pi(f)1 - T^n f\right\| \leq C\rho^n \mathbf{|} f \mathbf{|}$$

hence

$$\left|\pi(f)\right|\|1\| \leq \sup_k \left\|T^k\right\|\|f\| + C\rho^n \mathbf{|} f \mathbf{|}.$$

Now we can let n tend to infinity and conclude that π is $\|\cdot\|$-continuous. This $\|\cdot\|$-continuous linear functional defined on the set of compactly supported Lipschitz functions extends to a positive functional on $C_c(S)$, the set of all compactly supported functions on S. By the Riesz theorem, there exists a Borel measure μ such that $\pi(f) = \mu(f)$ for any $f \in C_c(S)$; since v is the increasing limit of a sequence of functions of $C_c(S)$, we have $\pi(v) = \mu(v) < \infty$. Any f in E being the $\|\cdot\|$-limit of compactly supported Lipschitz functions, by $\|\cdot\|$-continuity of π we obtain that $\pi(f) = \mu(f)$, $f \in E$.

Appendix E: Proof of Theorem 7

Multiplying both sides of (31) by P_0 on the left and by Q^q on the right we get

$$Q^{n+q} = \sum_{i=1}^{n-1} Q^{n-i} K V^{i-1} Q^q + P_0 K V^{n-1} Q^q + P_0 V^n Q^q. \tag{125}$$

We consider first the simpler case when $\|T^n\|$ is bounded, say $\|T^n\| \leq c$. In this case, considering a sequence n_k such that $\lambda_i^{n_k}$ converges to 1, for $i = 1, \ldots, p$ (this can be done by considering a converging subsequence λ^{m_k} of $\lambda^m = (\lambda_i^m, \ldots, \lambda_p^m)$ and taking $n_k = m_{2k} - m_k$), Equation (15) implies that for any $x \in E$

$$\left\|\sum_{i=1}^{p} \lambda_i^{n_k} P_i x\right\| \leq \left\|T^{n_k} x\right\| + \left\|Q^{n_k} x\right\|$$

and letting k tend to infinity, thanks to (22):

$$\left\|\sum_{i=1}^{p} P_i x\right\| \leq c\|x\|.$$

Hence, $\|P_0\| \leq 1 + c$ is finite, and Equation (125) leads directly to

$$\|Q^{n+q}\|_{E0} \leq \sum_{i=1}^{n-1} \|Q^{n-i}\|_{E0} \left| KV^{i-1}Q^q \right| + \|P_0\| \left| Q^q \right| \|KV^{n-1}\|_{E0} + \|P_0\| \left| Q^q \right| \|V^n\|_{E0}.$$

We plan to apply Proposition 13 of the Appendix F with $u_n = \|Q^n\|_{E0}$ and $\beta_i = \left| KV^{i-1}Q^q \right|$ for some q large enough. We remark that (130) is satisfied since

$$\left| KV^iQ^q \right| = \left| KV^iT^qP_0 \right| \leq \alpha_i C_2 C_T \left| P_0 \right|. \tag{126}$$

Because of the summability of α_i (a consequence of (R1) and (R3)), and with the help of the Lebesgue Dominated Convergence theorem, Equation (131) will be satisfied for q large enough if we can prove that for any $i \geq 0$

$$\lim_q \left| KV^iQ^q \right| = 0. \tag{127}$$

But this is easily obtained by induction on i since it is true for $i = 0$ and for any $i, q > 0$

$$\begin{aligned} \left| KV^iQ^q \right| &= \left| KV^{i-1}(T-K)Q^q \right| \\ &\leq \left| KV^{i-1}Q^{q+1} \right| + \left| KV^{i-1}KQ^q \right| \\ &\leq \left| KV^{i-1}Q^{q+1} \right| + \left| KV^{i-1} \right| \left| KQ^q \right|. \end{aligned}$$

Hence, Proposition 13 applies and (75) holds.

If now $\|T^n\|$ is not bounded, we have to work slightly more on Equation (125). Consider

$$f(z) = \prod_{i=1}^{p} (1 - z\bar{\lambda}_i).$$

Since Equations (9) to (13) imply that $T^n = \sum \lambda_i^n P_i + P_0Q^n$, $n \geq 0$ (this differs from (15) because we have to take into account the case $n = 0$) we have $f(T) = P_0 f(Q)$. Hence, after multiplication on the left by $f(Q)$ Equation (125) becomes

$$f(Q)Q^{n+q} = \sum_{i=1}^{n-1} f(Q)Q^{n-i}KV^{i-1}Q^q + f(T)KV^{n-1}Q^q + f(T)V^nQ^q$$

thus

$$\begin{aligned} &\|f(Q)Q^{n+q}\|_{E0} \\ &\leq \sum_{i=1}^{n-1} \|f(Q)Q^{n-i}\|_{E0} \left| KV^{i-1}Q^q \right| + \|f(T)KV^{n-1}Q^q\|_{E0} + \|f(T)V^nQ^q\|_{E0}. \end{aligned} \tag{128}$$

Since $\| f(T)\| < \infty$, (126) implies that there exists a constant C such that

$$\left\| f(T)KV^{n-1}Q^q \right\|_{E0} + \left\| f(T)V^n Q^q \right\|_{E0} \le C\alpha_n$$

and we obtain, as before (because (126) and (127) still hold true) that

$$\left\| f(Q)Q^n \right\|_{E0} \le C'\alpha_n.$$

Set $g(z) = 1/f(z) = \sum_{i\ge 0} g_i z^i$. The partial fraction decomposition of g implies that $\sup_i |g_i| < \infty$. For any $n \ge 0$

$$\left\| Q^n \right\|_{E0} \le \left\| Q^n g(Q) f(Q) \right\|_{E0} \le \sum_k \left\| Q^{n+k} g_k f(Q) \right\|_{E0} \le \sup_i |g_i| \sum_k \left\| Q^{n+k} f(Q) \right\|_{E0}$$

hence

$$\left\| Q^n \right\|_{E0} \le C \sum_{k\ge n} \alpha_k.$$

Appendix F: Convolution of sequences

Proposition 13. *Let $(\alpha_n)_{n\ge 1}$ be a positive sequence satisfying Assumptions* (R1) *to* (R3) *of Theorem* 7, *and $(\beta_i)_{i\ge 1}$ be a non-negative sequence. Let q be a non-negative integer and $(u_n)_{n\ge 1}$ be a non-negative sequence such that*

$$u_{n+q} \le C_0\alpha_n + \sum_{i=1}^{n-1} u_{n-i}\beta_i, \qquad n \ge 1 \tag{129}$$

for some $C_0 > 0$. If

$$\sup_k \frac{\beta_k}{\alpha_k} < \infty, \tag{130}$$

$$\sum_{i=1}^{\infty} \beta_i < 1 \tag{131}$$

then

$$\sup_n \frac{u_n}{\alpha_n} < \infty. \tag{132}$$

Proof. Set

$$v_n = \frac{u_n}{\alpha_n},$$

$$v_n^* = \sup_{k \le n} v_k,$$

$$\theta_n = \frac{\alpha_n}{\alpha_{n+q}},$$

$$C_\beta = \sup_k \frac{\beta_k}{\alpha_k}$$

then, for any i_0 and $n > i_0$

$$\begin{aligned}
v_{n+q} &\le C_0\theta_n + \theta_n \sum_{i=1}^{n-1} v_{n-i} \frac{\alpha_{n-i}\beta_i}{\alpha_n} \\
&\le C_0\theta_n + \theta_n v_{i_0}^* \sum_{i=n-i_0}^{n-1} \frac{\alpha_{n-i}\beta_i}{\alpha_n} + \theta_n v_n^* \sum_{i=i_0}^{n-i_0} \frac{\alpha_{n-i}\beta_i}{\alpha_n} + \theta_n v_n^* \sum_{i=1}^{i_0} \frac{\alpha_{n-i}\beta_i}{\alpha_n} \\
&\le C_0\theta_n + \theta_n v_{i_0}^* C_\beta \sum_{i=n-i_0}^{n-1} \frac{\alpha_{n-i}\alpha_i}{\alpha_n} + \theta_n v_n^* C_\beta \sum_{i=i_0}^{n-i_0} \frac{\alpha_{n-i}\alpha_i}{\alpha_n} + \theta_n v_n^* \frac{\alpha_{n-i_0}}{\alpha_n} \sum_{i=1}^{i_0} \beta_i \\
&\le C_0\theta_n + \theta_n v_{i_0}^* C_\beta i_0 \frac{\alpha_1\alpha_{n-i_0}}{\alpha_n} + \theta_n' v_n^* \left(C_\beta \sum_{i=i_0}^{n-i_0} \frac{\alpha_{n-i}\alpha_i}{\alpha_n} + \sum_{i=1}^{i_0} \beta_i \right),
\end{aligned}$$

where θ_n' tends to 1 (Assumption (R2)). By assumption (R2), for any i, the sequence $j \mapsto \alpha_{j-i}/\alpha_j$, $j \ge i$ is decreasing, hence for $i \le n/2$ one has

$$\frac{\alpha_{n-i}}{\alpha_n} \le \frac{\alpha_i}{\alpha_{2i}}$$

thus for $1 \le i_0 < n$

$$\sum_{i=i_0}^{n-i_0} \frac{\alpha_{n-i}\alpha_i}{\alpha_n} \le 2 \sum_{i=i_0}^{[n/2]} \frac{\alpha_{n-i}\alpha_i}{\alpha_n} \le 2 \sum_{i=i_0}^{[n/2]} \frac{\alpha_i^2}{\alpha_{2i}} \le 2 \sum_{i=i_0}^{\infty} \frac{\alpha_i^2}{\alpha_{2i}}$$

and we get, for $n > i_0$

$$v_{n+q} \le C' + \theta_n' \rho v_n^*,$$

$$\rho = 2\left(\sum_{i=i_0}^{\infty} \frac{\alpha_i^2}{\alpha_{2i}} \right) \sup_k \frac{\beta_k}{\alpha_k} + \sum_{i=1}^{i_0} \beta_i,$$

where C' depends on everything except on n. Since $\theta_n' \to 1$ and i_0 can be chosen large enough to have $\rho < 1$, this proves that for some $n_0 > 0$ and $0 < \rho' < 1$

$$v_{n+q} \le C' + \rho' v_n^*, \qquad n \ge n_0.$$

In particular

$$v_{n+q} \le C' + \rho' v^*_{n+q}, \qquad n \ge n_0.$$

By increasing C' we even get

$$v_n \le C'' + \rho' v^*_n, \qquad n \ge 1$$

and since the r.h.s. is also an upper bound for v_k, $k \le n$ (because $v^*_k \le v^*_n$), we get

$$v^*_n \le C'' + \rho' v^*_n, \qquad n \ge 1$$

which proves that v_n is bounded. □

Appendix G: Proof of Lemma 8

We need a preparatory lemma which will be essential for working with (76); the point of this lemma is to bring out a function ζ which satisfies (135), is significantly larger than $\zeta(t) = t$ and that can be easily iterated (Equation (134) implies $\zeta^{(n)}(t) = \psi^{(-1)}(\psi(t) + n)$):

Lemma 14. *Let θ be a non-decreasing non-negative concave differentiable function on* $[0, +\infty)$ *with a derivative which tends to zero at infinity, and define for $t \ge 0$*

$$\psi(t) = \int_0^t \frac{1}{\theta(y)}\, dy, \tag{133}$$

$$\zeta(t) = \psi^{(-1)}\big(\psi(t) + 1\big). \tag{134}$$

We assume that ψ is finite and tends to infinity.

$$\theta(t) \le t.$$

Then ζ is concave and for any t such that $t \ge \theta(t)$

$$\zeta\big(t - \theta(t)\big) \le t. \tag{135}$$

For any $t \ge 0$

$$\zeta(t) \le t + \theta\big(\zeta(t)\big). \tag{136}$$

Proof. The equation

$$\psi\big(\zeta(t)\big) = \psi(t) + 1$$

implies that $\zeta(t) > t$. By differentiating this equation, we get

$$\zeta'(t) = \frac{\theta(\zeta(t))}{\theta(t)} \tag{137}$$

and

$$\zeta''(t) = \frac{\theta'(\zeta(t))\zeta'(t)\theta(t) - \theta(\zeta(t))\theta'(t)}{\theta(t)^2} = \frac{\theta(\zeta(t))}{\theta(t)^2}\big(\theta'\big(\zeta(t)\big) - \theta'(t)\big) \le 0.$$

We turn now to Equation (135); since ψ is strictly increasing, (135) is equivalent to

$$\psi\big(t - \theta(t)\big) + 1 \le \psi(t)$$

but since θ is non-decreasing

$$\psi(t) - \psi\big(t - \theta(t)\big) = \int_{t-\theta(t)}^{t} \frac{1}{\theta(y)}\, dy \ge \theta(t)\frac{1}{\theta(t)} = 1.$$

Concerning (136), notice that (134) means that

$$\int_t^{\zeta(t)} \frac{1}{\theta(y)}\, dy = 1$$

and that on the other hand

$$\int_t^{\zeta(t)} \frac{1}{\theta(y)}\, dy \ge \big(\zeta(t) - t\big)\frac{1}{\theta(\zeta(t))}. \qquad \square$$

We can now proceed to the proof of Lemma 8. Combining equations (77) to (79), we get

$$Tv \le v - \theta(v) + \lambda(1 - V1), \qquad \lambda = \frac{c}{\varepsilon}.$$

We define the functions ζ and ψ from θ as in Lemma 14 and we set for $t \ge 0$

$$\zeta_n(t) = \psi^{(-1)}\big(\psi(t) + n\big) = \zeta\big(\zeta_{n-1}(t)\big). \tag{138}$$

Differentiating (138) and using (137), we obtain

$$\zeta_n'(t) = \zeta'\big(\zeta_{n-1}(t)\big)\zeta_{n-1}'(t) = \frac{\theta(\zeta_n(t))}{\theta(\zeta_{n-1}(t))}\zeta_{n-1}'(t) \tag{139}$$

hence

$$\zeta_n'(x) = \frac{\theta(\zeta_n(x))}{\theta(x)}. \tag{140}$$

The function ζ_n is concave, as a composition of increasing concave functions. Using the Jensen inequality and the concavity of ζ_k (as a composition of increasing concave functions), we obtain

$$T\big(\zeta_k(v)\big) \le \zeta_k(Tv) \le \zeta_k\big(v - \theta(v) + \lambda - \lambda V1\big). \tag{141}$$

Set $\delta = \min_x (v(x)/2$. We proceed now by considering two cases depending on $x - \theta(x) \geq \delta$ or not (x is the implicit argument in (141)). By concavity of ζ_k, we get on the set $\{x : x - \theta(x) \geq \delta\}$

$$\begin{aligned} T\big(\zeta_k(v)\big) &\leq \zeta_k\big(v - \theta(v)\big) + \lambda\zeta_k'\big(v - \theta(v)\big)(1 - V1) \\ &\leq \zeta_{k-1}(v) + \lambda\zeta_k'(\delta)(1 - V1), \end{aligned} \tag{142}$$

the last inequality coming from the fact that ζ_k' is decreasing.

In the case where x, the implicit argument in (141), satisfies $x - \theta(x) < \delta$, we have:

$$T\big(\zeta_k(v)\big) \leq \zeta_k(\delta + \lambda - \lambda V1) \leq \zeta_k(\delta) + \lambda\zeta_k'(\delta)(1 - V1) \tag{143}$$

but since $\zeta(x) \leq x + \theta(\zeta(x))$ (Equation (136))

$$\zeta_k(\delta) \leq \zeta_{k-1}(\delta) + \theta\big(\zeta_k(\delta)\big) = \zeta_k(\delta) + \zeta_k'(\delta)\theta(\delta). \tag{144}$$

On the other hand, from $x - v(x) < \delta$, we get $2\delta \leq Tv \leq \delta + \lambda - \lambda V1$, thus $\delta \leq \lambda(1 - V1)$, and (143), (144) lead to

$$\begin{aligned} T\big(\zeta_k(v)\big) &\leq \zeta_{k-1}(\delta) + \zeta_k'(\delta)\theta(\delta) + \lambda\zeta_k'(\delta)(1 - V1) \\ &\leq \zeta_{k-1}(\delta) + \left(\frac{2\theta(\delta)}{\delta} + \lambda\right)\zeta_k'(\delta)(1 - V1). \end{aligned} \tag{145}$$

Putting together (142) and (145), we obtain that everywhere

$$T\big(\zeta_k(v)\big) \leq \zeta_{k-1}(v) + \lambda_1\zeta_k'(\delta)(1 - V1) \tag{146}$$

with $\lambda_1 = 2\delta^{-1}\theta(\delta) + \lambda$. Thus, since $v \geq \delta$,

$$\begin{aligned} V\zeta_k(v) &= V\big(\zeta_k(v) - \zeta_k(\delta)\big) + \zeta_k(\delta)V1 \\ &\leq T\big(\zeta_k(v) - \zeta_k(\delta)\big) + \zeta_k(\delta)V1 \\ &\leq \zeta_{k-1}(v) - \big(\zeta_k(\delta) - \lambda_1\zeta_k'(\delta)\big)(1 - V1). \end{aligned} \tag{147}$$

Since $\zeta_n(\delta)$ tends to infinity ($\psi(\zeta_n(t)) = \psi(t) + n$) and $\theta(x)/x$ tends to zero (θ is concave with a derivative which tends to zero), the sequence $\zeta_n'(\delta)/\zeta_n(\delta)$ tends to 0 (cf. (140)). As a consequence, there exist n_0 such that $\lambda_1\zeta_k'(\delta) - \zeta_k(\delta) \leq 0$ for $k > n_0$, hence multiplying both sides of (147) by V^{k-1} and summing up from 1 to $n > n_0$, we get

$$V^n\zeta_n(v) \leq v + c'$$

with $c' = \sum_{k=1}^{n_0} |\zeta_k(\delta) - \lambda_1\zeta_k'(\delta)|$. Since $\zeta_n(x) \geq \psi^{(-1)}(n)$ we get finally

$$V^n 1 \leq \frac{v + c'}{\psi^{(-1)}(n)}.$$

This proves (80). Concerning (81), notice that (146) implies that for any n:

$$T\big(\zeta_k(v)\big) \le \zeta_{k-1}(v) + c_1$$

for some $c_1 > 0$. Hence, multiplying both sides by T^{k-1} and summing up, we get

$$T^n\big(\zeta_n(v)\big) \le v + nc_1. \tag{148}$$

Since by definition of ζ_n, one has

$$\int_x^{\zeta_n(x)} \frac{dt}{\theta(t)} = n$$

we obtain in particular that $n \le \frac{\zeta_n(x)-x}{\theta(x)}$, and (148) becomes

$$T^n\big(v + n\theta(v)\big) \le v + nc_1.$$

This implies (81).

Appendix H: Proof of Theorem 9

We plan to apply Theorem 7 with

$$\begin{aligned} |\!|\!| f |\!|\!| &= \|f\|_\infty, \\ \|f\| &= \|f\|_v. \end{aligned}$$

Clearly, since Theorem 2 applies, Equations (9) to (13) and (22) are satisfied. As in the proof of Theorem 2 we set $V = T - K$; we recall that $K(x, S) = 0$ if $x \notin K_0$ (cf. the statement of Theorem 2). We have to estimate $\|V^n\|_{E0}$; but since Equation (36) with the fact that $K(x, S) = 0$ for $x \notin K_0$ imply that

$$V1 \le 1 - \varepsilon 1_{K_0}.$$

Lemma 8 leads to

$$V^n 1 \le \frac{v + c}{\psi^{(-1)}(n)}.$$

Theorem 7 applies and, in particular, we obtain (83). For (84), we consider

$$\|f\| = \pi\big(|f|\big).$$

Since $|\!|\!| \cdot |\!|\!|$ is unchanged, Equations (9) to (13), (73) and (74) are still satisfied, as well as (22) because $\pi(|f|) \le \|f\|_v \pi(v)$. In addition $\|T^n\| = 1$, and

$$\big\|V^n\big\|_{E0} = \pi\big(V^n 1\big) \le \frac{c'}{\psi^{(-1)}(n)}.$$

Theorem 7 still applies and we obtain (84).

Appendix I: Proof of Theorem 10

As is [2], the idea is to prove directly that that $T^n g$ is a Cauchy sequence. If we set

$$r(x, y) = 1 - \varepsilon 1_{x,y\in K},$$

Equations (86), (87) can be summarized as

$$\left|Tf(x) - Tf(y)\right| \le r(x, y)\, d(x, y)[f].$$

By the Kantorovich–Rubinstein formula ([27] equation (5.11) and (6.3)) this means that given $X_0 = x$ and $Y_0 = y$ there exists a coupling of X_1 and Y_1 such that

$$E\left[d(X_1, Y_1)\right] \le r(x, y)\, d(x, y). \tag{149}$$

Using Theorem 1.1 of [30], this coupling may be done measurably w.r.t. x and y in the sense that there exists a transition kernel $\mathbb{T}((x, y), \cdot)$ on $E \times E$ with marginal transitions given by T and such that (149) is satisfied:

$$\int f\left(x'\right)\mathbb{T}\left(x, y, dx', dy'\right) = \int f\left(y'\right)\mathbb{T}\left(y, x, dx', dy'\right) = Tf(x),$$
$$\int d\left(x', y'\right)\mathbb{T}\left(x, y, dx', dy'\right) \le r(x, y)\, d(x, y).$$

This result is very important since it gives directly the best coupling method as the realization of a Markov chain on the product space. We have thus with standard notations

$$\mathbb{T}f(x, y) = E_{x,y}\left[f(X_1, Y_1)\right],$$
$$\mathbb{T}d \le rd.$$

Set for any function f on $S \times S$

$$[\![f]\!] = \sup_{x\neq y} \frac{|f(x, y)|}{d(x, y)}.$$

For the application of Lemma 8, we define sub-Markovian transition operator

$$\mathbb{V}u(x, y) = d(x, y)^{-1}\mathbb{T}(ud)(x, y).$$

Since obviously, for any measurable positive bounded function u,

$$f(x, y) \le [\![f/u]\!]\, d(x, y)u(x, y),$$

we get

$$\mathbb{T}f \le [\![f/u]\!]\mathbb{T}(du) = [\![f/u]\!]d.\mathbb{V}u,$$

hence

$$[\![\mathbb{T}f/\mathbb{V}u]\!] \le [\![f/u]\!].$$

Replacing f with $\mathbb{T}^{n-1}f$ and u with $\mathbb{V}^{n-1}u$, we get

$$\left[\!\!\left[\mathbb{T}^n f/\mathbb{V}^n u\right]\!\!\right] \le \left[\!\!\left[\mathbb{T}^{n-1} f/\mathbb{V}^{n-1} u\right]\!\!\right],$$

and by induction

$$\mathbb{T}^n f(x,y) \le d(x,y)\mathbb{V}^n u(x,y)[\![f/u]\!].$$

In particular, taking $u=1$:

$$\mathbb{T}^n f(x,y) \le d(x,y)\big(\mathbb{V}^n 1\big)(x,y)[\![f]\!]. \tag{150}$$

In order to apply Lemma 8, we need to check that (77) is satisfied. Setting $\bar{v}(x,y)=v(x)+v(y)$, one has

$$\begin{aligned}
\mathbb{T}\bar{v}(x,y) &= Tv(x)+Tv(y) \\
&\le v(x)+v(y)-\theta\big(v(x)\big)-\theta\big(v(y)\big)-c1_K(x)-c1_K(y) \\
&\le v(x)+v(y)-\theta\big(v(x)+v(y)\big)-c1_{K\times K}(x,y)
\end{aligned}$$

since by concavity and positivity of θ, $\theta(a+b)\le\theta(a)+\theta(b)$ for $a,b\ge 0$ (differentiate w.r.t. a). Obviously (78) and (79) are satisfied. Lemma 8 applies and (80) implies

$$\mathbb{V}^n 1 \le \frac{\bar{v}+c}{\psi^{(-1)}(n)}, \tag{151}$$

where

$$\psi(x)=\int_0^x \frac{1}{\theta(y)}\,dy.$$

Since in addition $\mathbb{V}^n 1 \le 1$, Equation (150) becomes now

$$\mathbb{T}^n f(x,y) \le d(x,y)\min\left(1,\frac{\bar{v}(x,y)+c}{\psi^{(-1)}(n)}\right)[\![f]\!]. \tag{152}$$

For any function f of the form $f(x,y)=g(x)-g(y)$, this leads to

$$T^n g(x)-T^n g(y) \le d(x,y)\min\left(1,\frac{v(x)+v(y)+c}{\psi^{(-1)}(n)}\right)[g]. \tag{153}$$

This proves (89).

From (153), we get

$$T^n g(x) - T^n g(y) \le d(x,y) \min\left(1, \frac{v(x)+c}{\psi^{(-1)}(n)}\right)[g] + d(x,y)\min\left(1, \frac{v(y)}{\psi^{(-1)}(n)}\right)[g]$$
$$= A(x,y) + B(x,y). \tag{154}$$

Since for $0 \le x \le y$ one has $\frac{x}{y} \le \frac{\theta(x)}{\theta(y)}$ (the function $x \mapsto \frac{x}{y} - \frac{\theta(x)}{\theta(y)}$ is convex and non-positive at $x = 0$ and $x = y$), we have

$$B(x,y) \le d(x,y) \frac{\theta(v(y))}{\theta(\psi^{(-1)}(n))}[g]. \tag{155}$$

Since $T^p g(x) = \int g(y) T^p(x, dy)$ we get from (154):

$$\begin{aligned}\left|T^n g(x) - T^{n+p} g(x)\right| &= \left|\int \left(T^n g(x) - T^n g(y)\right) T^p(x,dy)\right| \\ &\le \int A(x,y) T^p(x,dy) + [g]\int d(x,y) \frac{\theta(v(y))}{\theta(\psi^{(-1)}(n))} T^p(x,dy) \\ &\le \min\left(1, \frac{v(x)+c}{\psi^{(-1)}(n)}\right)[g] + [g]\theta\left(\psi^{(-1)}(n)\right)^{-1}\left(T^p\theta(v)(x)\right)\end{aligned}$$

because $d \le 1$. Since, by (81), $T^p\theta(v) \le c'' + v/p$ (we just apply Lemma 8 with $Vf(x) = (1 - 1_{x\in K})Tf(x)$)

$$\left|T^n g(x) - T^{n+p} g(x)\right| \le [g]\min\left(1, \frac{v(x)+c}{\psi^{(-1)}(n)}\right) + \frac{[g]}{\theta(\psi^{(-1)}(n))}\left(v(x)/p + c''\right). \tag{156}$$

This shows that $T^n g(x)$ is a Cauchy sequence and its limit (a constant function because of (155)) is necessarily $\pi(g)$ where π is the invariant measure. Letting p tend to infinity we get

$$\left|T^n g(x) - \pi(g)\right| \le [g]\min\left(1, \frac{v(x)+c}{\psi^{(-1)}(n)}\right) + \frac{c''[g]}{\theta(\psi^{(-1)}(n))}. \tag{157}$$

Appendix J: Proof of Equation (94)

We shall prove that for $0 \le x < 1$

$$v_n'(x)^\gamma \ge 1 + a n v_n(x)^\gamma, \qquad a = 2^\gamma - 1. \tag{158}$$

We recall that

$$v(x) = \begin{cases} x\left(1 + 2^\gamma x^\gamma\right), & 0 \le x < 1/2, \\ 2x - 1, & 1/2 \le x \le 1 \end{cases} \tag{159}$$

and that the prime sign stands for the right derivative. In the case $n=0$, the inequality is obvious. In the case $n \geq 1$, we assume by induction that (158) is satisfied and since $v'_{n+1}(x) = v'_n(x)v'(v_n(x)))$, valid for $n \geq 0$, Equation (158) with $n+1$ will be implied by

$$\left(1 + anv_n(x)^\gamma\right)v'\left(v_n(x)\right)^\gamma \geq 1 + a(n+1)v_{n+1}(x)^\gamma.$$

This has to be proved for $n \geq 0$. It suffices to show that for any $0 \leq y \leq 1$

$$\left(1 + any^\gamma\right)v'(y)^\gamma \geq 1 + a(n+1)v(y)^\gamma \tag{160}$$

(i.e. $y = v_n(x)$). By linearity of both sides of (160) w.r.t. n, we only have to check this for $n=0$, and $n \to \infty$, that is

$$\begin{cases} v'(y)^\gamma \geq 1 + av(y)^\gamma, \\ yv'(y) \geq v(y) \end{cases} \tag{161}$$

(the first equation is (158) with $n=1$). In the case, $y < 1/2$ this is rewritten as

$$\begin{cases} \left(1 + (\gamma+1)2^\gamma y^\gamma\right)^\gamma \geq 1 + ay^\gamma\left(1 + 2^\gamma y^\gamma\right)^\gamma, \\ 1 + (\gamma+1)2^\gamma y^\gamma \geq 1 + 2^\gamma y^\gamma. \end{cases}$$

The second inequality is obvious. For the first one, since $2y < 1$, setting $z = 2^\gamma y^\gamma$, this holds if

$$\left(1 + (\gamma+1)z\right)^\gamma \geq 1 + az$$

for $0 \leq z \leq 1$. Since the difference of both sides is a concave function of z which vanishes at $z=0$, and is non-negative at $z=1$ (we recall that $a = 2^\gamma - 1$), the inequality is satisfied. In the case $y \geq 1/2$, (161) is

$$\begin{cases} 2^\gamma \geq 1 + a(2y-1)^\gamma, \\ 2y \geq 2y - 1 \end{cases}$$

which is obviously satisfied.

References

[1] Aubin, J.-P. and Cellina, A. (1984). *Differential Inclusions*: *Set-Valued Maps and Viability Theory*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **264**. Berlin: Springer. MR0755330

[2] Butkovsky, O. (2014). Subgeometric rates of convergence of Markov processes in the Wasserstein metric. *Ann*. *Appl*. *Probab*. **24** 526–552. MR3178490

[3] Cloez, B. and Hairer, M. (2015). Exponential ergodicity for Markov processes with random switching. *Bernoulli* **21** 505–536. MR3322329

[4] de Leeuw, K. and Glicksberg, I. (1961). Applications of almost periodic compactifications. *Acta Math*. **105** 63–97. MR0131784

[5] Doeblin, W. and Fortet, R. (1937). Sur des chaînes à liaisons complètes. *Bull. Soc. Math. France* **65** 132–148. MR1505076

[6] Douc, R., Fort, G., Moulines, E. and Soulier, P. (2004). Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab.* **14** 1353–1377. MR2071426

[7] Douc, R., Moulines, E. and Soulier, P. (2007). Computable convergence rates for sub-geometric ergodic Markov chains. *Bernoulli* **13** 831–848. MR2348753

[8] Dunford, N. and Schwartz, J.T. (1988). Linear operators. Part I. In *General Theory, with the Assistance of. Wiley Classics Library*. New York: Wiley.

[9] Durmus, A., Fort, G. and Moulines, É. (2016). Subgeometric rates of convergence in Wasserstein distance for Markov chains. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 1799–1822. MR3573296

[10] Gantmacher, F.R. (1998). *The Theory of Matrices. Vol.* 1. Providence, RI: AMS Chelsea Publishing.

[11] Gouëzel, S. (2004). Central limit theorem and stable laws for intermittent maps. *Probab. Theory Related Fields* **128** 82–122.

[12] Hennion, H. and Hervé, L. (2001). *Limit Theorems for Markov Chains and Stochastic Properties of Dynamical Systems by Quasi-Compactness. Lecture Notes in Math.* **1766**. Berlin: Springer. MR1862393

[13] Ionescu Tulcea, C.T. and Marinescu, G. (1950). Théorie ergodique pour des classes d'opérations non complètement continues. *Ann. of Math.* (2) **52** 140–147. MR0037469

[14] Iosifescu, M. (1993). A basic tool in mathematical chaos theory: Doeblin and Fortet's ergodic theorem and Ionescu Tulcea and Marinescu's generalization. In *Doeblin and Modern Probability* (*Blaubeuren*, 1991). *Contemp. Math.* **149** 111–124. Providence, RI: Amer. Math. Soc. MR1229957

[15] Iosifescu, M. and Theodorescu, R. (1969). *Random Processes and Learning*. New York: Springer. MR0293704

[16] Jarner, S.F. and Roberts, G.O. (2002). Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.* **12** 224–247. MR1890063

[17] Joulin, A. and Ollivier, Y. (2010). Curvature, concentration and error estimates for Markov chain Monte Carlo. *Ann. Probab.* **38** 2418–2442.

[18] Kontoyiannis, I. and Meyn, S.P. (2003). Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Probab.* **13** 304–362. MR1952001

[19] Kontoyiannis, I. and Meyn, S.P. (2012). Geometric ergodicity and the spectral gap of non-reversible Markov chains. *Probab. Theory Related Fields* **154** 327–339. MR2981426

[20] Liverani, C., Saussol, B. and Vaienti, S. (1999). A probabilistic approach to intermittency. *Ergodic Theory Dynam. Systems* **19** 671–685. MR1695915

[21] Merlevède, F., Peligrad, M. and Utev, S. (2006). Recent advances in invariance principles for stationary sequences. *Probab. Surv.* **3** 1–36. MR2206313

[22] Meyn, S. and Tweedie, R.L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge: Cambridge Univ. Press. MR2509253

[23] Nummelin, E. (1984). *General Irreducible Markov Chains and Nonnegative Operators. Cambridge Tracts in Mathematics* **83**. Cambridge: Cambridge Univ. Press. MR0776608

[24] Poznyak, A.S. (2001). The strong law of large numbers for dependent vector processes with decreasing correlation: "double averaging concept". *Math. Probl. Eng.* **7** 87–95.

[25] Revuz, D. (1975). *Markov Chains. North-Holland Mathematical Library* **11**. Amsterdam: North-Holland.

[26] Sarig, O. (2002). Subexponential decay of correlations. *Invent. Math.* **150** 629–653. MR1946554

[27] Villani, C. (2009). *Optimal Transport: Old and New. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Berlin: Springer.

[28] Yosida, K. and Kakutani, S. (1941). Operator-theoretical treatment of Markoff's process and mean ergodic theorem. *Ann. of Math.* (2) **42** 188–228. MR0003512

[29] Young, L.-S. (1999). Recurrence times and rates of mixing. *Israel J. Math.* **110** 153–188. MR1750438
[30] Zhang, S. (2000). Existence and application of optimal Markovian coupling with respect to non-negative lower semi-continuous functions. *Acta Math. Sin.* (*Engl. Ser.*) **16** 261–270. MR1778706

Received May 2014 and revised August 2015