Version of 6.8.98

Iterates of expanding maps

A. D. Barbour*, R. M. Gerrard and G. Reinert **

Universität Zürich; City University, London;
University of California, Los Angeles

Abstract. The iterates of expanding maps of the unit interval into itself have many of the properties of a more conventional stochastic process, when the expanding map satisfies some regularity conditions and when the starting point is suitably chosen at random. In this paper, we show that the sequence of iterates can be closely tied to an m-dependent process. This enables us to prove good bounds on the accuracy of Gaussian approximations. Our main tools are coupling and Stein's method.

Keywords: Expanding maps, functional iteration, coupling, decay of correlations, Gaussian approximation

AMS 1991 Subject Classifications: 58F08, 58F11, 60F05, 60F17, 60G10

---

* Partially supported by Schweizer Nationalfonds Grant 20-43453.95

** Partially supported by NSF grant DMS-9505075

1. Introduction

Let $h : I = [0, 1] \to I$ be piecewise monotone C^1 and uniformly expanding: that is, there is a finite set $U = U(h)$ of points

$$0 = u_0 < u_1 < \cdots < u_{m_1} < u_{m_1+1} = 1 \tag{1.1}$$

in I such that, for each interval $J_i = J_i(h) = (u_{i-1}, u_i)$, both h restricted to J_i and its continuous extension to $[u_{i-1}, u_i]$ are C^1 and monotone, satisfying

$$1 < c(h) \leq |h'(x)| \leq C(h) < \infty \tag{1.2}$$

for all x. We shall be interested in the behaviour of the sequence $\{x_r = h_r(x_0),\ r \geq 0\}$, where h_r denotes the r'th iterate of h. We need the following basic assumptions A1–A3, which are to hold throughout.

A (measurable) set A is said to have period r if $h_r(A) \doteq A$, where $A \doteq B$ means that $\lambda(A \triangle B) = 0$ and λ denotes Lebesgue measure; if A has period 1, it is called invariant. An invariant measure is a measure μ such that $\mu(h^{-1}(A)) = \mu(A)$ for all A. We assume that:

A1: There are no periodic sets A with $0 < \lambda(A) < 1$.
A2: There exists an $r = r_1 \geq 1$ for which $|h_r^{-1}(x)| \geq 4$ for all $x \notin h_r(U(h_r))$.

If h is piecewise monotone and uniformly expanding and h' is piecewise smooth, there exists at least one invariant probability measure μ which is absolutely continuous with respect to λ. The celebrated theorem of Lasota and Yorke (1973) proves this when h' is piecewise C^1, and shows also that the density f of μ is of bounded variation. In Keller (1985, Theorem 3.5), the conditions are relaxed somewhat beyond

A3: h' is piecewise Hölder continuous with exponent ζ, for some $0 < \zeta \leq 1$.

Under A3, Keller shows that any invariant density f has the following regularity property: there exists a $K_{(1.3)} < \infty$ such that, for all $0 < \varepsilon < 1$,

$$\int_I \operatorname{osc}(f, \varepsilon, x)\, dx \leq K_{(1.3)} \varepsilon^{\zeta}, \tag{1.3}$$

where

$$\operatorname{osc}(f, \varepsilon, x) = \operatorname{ess} \sup_{y,z \in S_x(\varepsilon)} |f(y) - f(z)|, \tag{1.4}$$

and $S_x(\varepsilon) = \{y : |y - x| < \varepsilon\}$. Under our additional assumption A1, there is in fact only one invariant measure that is λ–absolutely continuous, since $\operatorname{supp}(\mu)$ is an invariant set if μ is an invariant measure, and thus $\operatorname{supp}(\mu) = I$ for any such μ.

Our interest lies mainly in the extent to which the properties of the h–sequence $\{h_r(x_0), r \geq 0\}$ mimic those of a more conventional stochastic process, when x_0 is suitably chosen at random. If x_0 is exactly known, the whole future of the h–sequence is completely determined, and randomness does not enter at all. However, in practice, x_0 can never be known without error, and the small uncertainty in the value of x_0 has an enormous effect on the later values in the sequence. It also makes sense to ask for the properties of a 'typical' sequence, where 'typical' could, for instance, be interpreted as meaning that x_0 is chosen uniformly at random from I. An example of the parallels with stochastic processes is the functional central limit theorem for partial sum processes derived from an h–sequence, which was proved by Hofbauer and Keller (1982). They also derive a rate of convergence, using general theorems for mixing sequences given in Stout (1974). However, the rate they obtain is not explicitly characterized, and can be expected to be very poor. In this paper, we address the stochastic structure of h–sequences more directly, and are thus able to make much more concrete statements.

Our approach, illustrated in an elementary setting in Barbour (1995), is based on two observations. The first, which we will probably not be the last to rediscover, is that if x_0 is chosen at random according to the invariant measure μ, then the h–sequence is a stationary Markov chain taking values in I, as is its time–reversal; see, for instance, the references given in Isham (1993, Section 3.6.3). The h–sequence is Markovian because its evolution, for given x_0, is deterministic, and stationarity follows directly if $x_0 \sim \mu$. Now the invariance of μ implies that

$$f(y) = \sum_{x \in h^{-1}(y)} f(x)/|h'(x)| \tag{1.5}$$

for (almost) all y such that $h^{-1}(y) \cap U(h) = \emptyset$, so that then the quantities

$$p(y, x) = \frac{f(x)}{|h'(x)| f(y)}, \quad x \in h^{-1}(y), \tag{1.6}$$

are probabilities. Moreover, if x is such that $h_i(x) \notin U(h)$, $0 \leq i \leq m$, if X is an interval around x so small that the maps $h : h_i(X) \to h_{i+1}(X)$ are 1–1 for each $0 \leq i < m$ and if $g : I^{m+1} \to \mathbb{R}$ is measurable, then

$$\begin{aligned}
\int_X g(x, h(x), \ldots, h_m(x))\, \mu(dx) &= \int_X g(x, h(x), \ldots, h_m(x)) f(x)\, dx \\
&= \int_{h_m(X)} g(h_m^{-1}(y), \ldots, h^{-1}(y), y) f(h_m^{-1}(y)) \prod_{i=0}^{m-1} \frac{1}{|h'(h_{m-i}^{-1}(y))|}\, dy \\
&= \int_{h_m(X)} g(h_m^{-1}(y), \ldots, h^{-1}(y), y) \prod_{i=0}^{m-1} \Big\{ \frac{f(h_{m-i}^{-1}(y))}{f(h_{m-i-1}^{-1}(y)) |h'(h_{m-i}^{-1}(y))|} \Big\} f(y)\, dy \\
&= \int_{h_m(X)} g(h_m^{-1}(y), \ldots, h^{-1}(y), y) \prod_{i=0}^{m-1} p(h_{m-i-1}^{-1}(y), h_{m-i}^{-1}(y))\, \mu(dy),
\end{aligned}$$

so that the time reversal of $(X_0, X_1, \ldots)$ with $X_0 \sim \mu$ and with $X_i = h_i(X_0)$ is the Markov chain $(Y_0, Y_1, \ldots)$ with $Y_0 \sim \mu$ and with transition probabilities given by (1.6).

An advantage of considering the reversed process is that randomness enters progressively at each step, and not only when setting the initial state x_0, making the analogy with classical stochastic processes clearer. Our second observation is that the time reversal of the h–sequence of a uniformly expanding map has an induced contraction property, which enables coupling methods to be introduced. It is shown in Section 3 under A1–A3 that, if x is close enough to x', then $p(x, \phi(x))$ is close to $p(x', \phi(x'))$, where ϕ is a given branch of h^{-1}. Thus the first steps in a reversed chain starting in x_0 and in one starting in x'_0 can typically be realized in such a way that, with high probability, $x_1 = \phi(x_0)$ and $x'_1 = \phi(x'_0)$ for the same branch ϕ of h^{-1}. If this is the case, then

$$|x_1 - x'_1| = |\phi(x_0) - \phi(x'_0)| \leq |x_0 - x'_0| \sup_{y \in I}\{1/|h'(y)|\} = c(h)^{-1}|x_0 - x'_0|, \tag{1.7}$$

and the positions of the two chains after one step are closer than they were initially, at least by a geometric factor of $c(h)^{-1} < 1$. The main effort is then devoted to showing in Theorem 3.4 that, however two reversed chains $(Y_n,\ n \geq 0)$ and $(Y'_n,\ n \geq 0)$ are started, they can be realized simultaneously in such a way that $|Y_n - Y'_n| \leq Zc(h)^{-n}$ for all n, where Z is a random variable with Pareto tail; we refer to this as a 'successful' coupling. The coupling approach has already proved to be a powerful tool in many areas (Lindvall, 1992), and it plays the main part in our arguments.

A traditional approach to the limiting behaviour of the sequence of iterates $\{h_r(x_0),\ r \geq 0\}$ is first to study a sequence of labels $\{l_r,\ r \geq 0\}$ determined by the iterates; typically, $l_r = i$ if $h_r(x_0) \in J_i$. Since h is expanding, the finite vectors $(l_r, l_{r+1}, \ldots, l_{r+s})$, $s \geq 0$, determine the value of $h_r(x_0)$ ever more precisely as $s \to \infty$. Under conditions such as in Hofbauer and Keller (1982), the label process is proved to be absolutely regular (Ibragimov and Rozanov, 1978), and the 'stochastic' behaviour of the $h_r(x_0)$ sequence is then deduced from that of the l_r sequence. Our coupling argument has a somewhat similar flavour, but with an important difference. In Theorem 3.5, we are able to approximate the finite sequence $\{h_r(x_0),\ 0 \leq r \leq N\}$ by a sequence $\{x'_r,\ 0 \leq r \leq N\}$ which is m–dependent, where $m = m(N)$ need only grow logarithmically with N to achieve a uniform accuracy of approximation of order N^{-s}, for any chosen $s > 0$. Being able to work with an m–dependent process instead of with an absolutely regular mixing process brings substantial advantages.

The purpose of this paper is twofold: first, to introduce the coupling approach to the sequence of iterates, and then to illustrate how it can be used to sharpen various stochastic limit theorems. Berry–Esseen theorems with rate $N^{-1/2}$ have already been established under a variety of circumstances (Rousseau–Egele 1983, Coelho and Parry 1990, Heinrich 1996) for the distribution of the centred and normalized partial sums of the sequence. Here, we use the m–dependent approximation and Stein's method to obtain, in Theorem 4.5, a near optimal convergence rate of $N^{-1/2}\log^3 N$ for the multivariate central limit theorem. We also obtain rates of similar order with respect to appropriate Wasserstein metrics for the functional central limit theorem, and for approximation of the empirical process. Dembo and Zeitouni (1996) use the results of this paper to obtain information about moderate deviations.

The structure of the paper is as follows. In Section 2, we establish the properties of h and f which we need in order to prove that our coupling is successful. We do so assuming A1–A3, and making heavy use of the results of Keller (1985); the arguments simplify somewhat if λ itself is invariant, which requires that

$$\sum_{x\in h^{-1}(y)} 1/|h'(x)| = 1$$

for all $y \notin h(U)$. In Section 3, we demonstrate that the coupling of time reversals of h–sequences is successful. In Section 4, we use the coupling method to prove the stochastic properties of h–sequences in which we are primarily interested.

2. Properties of h and f

In order to prove that the coupling of the next section is successful, we need to show that the set $h_r^{-1}(y)$ of pre–images of a typical point y under the iterates h_r of h becomes dense as $r \to \infty$. More precisely, starting with any interval $K \subset I$, we show that there is an $r_0 = r_0(K) < \infty$ such that $h_r(K) \doteq I$ for all $r \geq r_0$: this is the substance of Theorem 2.6. We cannot in general claim that $h_r(K) = I$ for such r, since it need not (quite) be the case that $h(I) = I$, and this generates some technical complication. The following lemma describes what happens.

Lemma 2.1. *Under A1 and A3, the set $I \setminus h(I)$ is at most finite.*

Proof. Since $h(I) \subset I$, it follows that the sets $h_r(I)$ are decreasing. Define $I^* = \cap_{r\geq 0} h_r(I)$. Then $h(I^*) = I^*$, so that I^* is invariant, and by A1 we then have $\lambda(I^*) \in \{0, 1\}$.

Now, since μ is invariant, we have

$$\mu(h_r(I)) = \mu(h_r^{-1}h_r(I)) \geq \mu(I) = 1,$$

for all $r \geq 0$, and thus $\mu(I^*) = 1$. Hence, since $\mu \ll \lambda$, we have $\lambda(I^*) = 1$ also, from which it follows that $\lambda(h(I)) = 1$ also. But h is piecewise continuous, and so $h(I)$ is a finite union of intervals. Thus $I \setminus h(I)$ is at most finite.

We will now usually restrict attention to I^*, so that all the inverses h_r^{-1} are properly defined; note that, from Lemma 2.1, $I \setminus I^*$ is at most countable. We use the notation A^* to denote $A \cap I^*$.

As a first step in showing that $h_{r_0}(K) \doteq I$ for some r_0, we prove that, if $A_r = h_r^{-1}h_r(K)$, then

$$A = \lim_{r\to\infty} A_r \doteq I. \tag{2.1}$$

Note that the sets A_r are increasing, and that

$$\begin{aligned} h_r^{-1}h_r(A) &= h_r^{-1}h_r(\cup_{n>r}A_n) = \cup_{n>r}h_r^{-1}h_r(A_n) \\ &= \cup_{n>r}h_r^{-1}h_r h_n^{-1}h_n(K) = \cup_{n>r}h_r^{-1}h_{n-r}^{-1}h_n(K) \\ &= \cup_{n>r}h_n^{-1}h_n(K) = A. \end{aligned}$$

This motivates a further definition: a set B is called preinvariant if $h_r^{-1}h_r(B) \doteq B$ for all $r \geq 1$.

Lemma 2.2. *The family of preinvariant sets is closed under Boolean operations.*

Proof. If B_1 and B_2 are preinvariant and $r \geq 1$, then

$$h_r^{-1}h_r(B_1 \cap B_2) \subset h_r^{-1}h_r(B_1) \cap h_r^{-1}h_r(B_2) \doteq B_1 \cap B_2;$$

however, $B \subset h_r^{-1}h_r(B)$ for any B, so that in fact $h_r^{-1}h_r(B_1 \cap B_2) \doteq B_1 \cap B_2$ for all r.

Now, for any $y \in h_r^{-1}h_r(B^c)$, there exists an $x \in B^c$ such that $h_r(x) = h_r(y)$, or equivalently such that $x \in h_r^{-1}h_r(y) \subset h_r^{-1}h_r(B) \doteq B$, if also $y \in B$. The set of such x has to have λ–measure zero; hence the same is true of $h_r^{-1}h_r(B^c) \cap B$, since h is piecewise monotone C^1.

Lemma 2.3. *If B is preinvariant, then so is $h_r(B)$ for any $r \geq 0$, and $\mu(h_r(B)) = \mu(B)$.*

Proof. For any $s \geq 0$, observe that

$$h_s^{-1}h_s h_r(B) = h_r^{-1}h_r\{h_s^{-1}h_s h_r(B)\} = h_r\{h_{r+s}^{-1}h_{r+s}(B)\} \doteq h_r(B),$$

because B is preinvariant, proving the first part. The second part follows because μ is invariant, so that $\mu(h_r(B)) = \mu(h_r^{-1}h_r(B)) = \mu(B)$.

Theorem 2.4. *Let $K \subset I$ be a non–empty interval, and let $A = \lim_{r\to\infty} h_r^{-1}h_r(K)$. Under A1 and A3, we have $A \doteq I$.*

Proof. A is preinvariant, and so are the sets $B_r = h_r(A)$, for all $r \geq 1$, and all sets obtainable from them by finitely many Boolean operations, by Lemmas 2.2 and 2.3. In addition, $\lambda(A) \geq \lambda(K) > 0$.

If $\lambda(A) = 1$, the theorem is proved. If not, given any $\varepsilon > 0$, there can be no more than $2^{1/\varepsilon}$ distinct B_r [that is, $\lambda(B_r \triangle B_s) = 0$ if $r \neq s$] such that all the Boolean atoms derived from them have λ–measure zero or at least ε. Thus, under A1, there are two possibilities: either $\lambda(B_r) = 1$ for some (periodic) B_r, in which case $\lambda(A) = \lambda(h_r^{-1} B_r) = 1$ also, and the theorem is proved; or else all the B_r are distinct, and, for any $\varepsilon > 0$, there exists a preinvariant set C, derived by finitely many Boolean operations from the B_r, such that $0 < \lambda(C) < \varepsilon$. Indeed, since $\operatorname{supp}(\mu) = I$ and μ is λ–absolutely continuous, we can take $0 < \mu(C) < \varepsilon$ also. It remains to be shown that this latter possibility cannot in fact occur.

If it were possible, pick r_2 so large that $c(h)^{r_2} > 4$, in which case $|h'_{r_2}(x)| > 4$ for all x. Now $U(h_{r_2})$ dissects I into a finite number of intervals $J_j = J_j(h_{r_2})$; let $\eta_1 = \min_j \lambda(J_j) > 0$, $\eta_2 = \min_j \mu(J_j) > 0$ and $\eta = \min(\eta_1, \eta_2)$. Choose a preinvariant set C as above with $0 < \lambda(C), \mu(C) < \eta$. Since A is the limit of finite unions of intervals and C is derived by Boolean operations on the sets $h_r(A)$, C contains a non–empty interval I_0 which is entirely contained in some J_j. Thus $h_{r_2}(I_0)$ is also an interval, and of length at least $4\lambda(I_0)$. If it covers one of the J_j, stop and set $k = 0$. If not, it contains an interval I_1 of length at least $2\lambda(I_0)$ which is contained in some J_j. Continue the process of applying h_{r_2} and selecting a new sub–interval until, for some k, $h_{r_2}(I_k)$ covers one of the J_j: this must happen in finitely many steps, since, at each stage, $\lambda(I_j) \geq 2\lambda(I_{j-1})$. Then, if $h_{r_2}(I_k) \supset J_j$, it follows that $\mu\{h_{(k+1)r_2}(I_0)\} \geq \mu(J_j) \geq \eta$. But $h_{(k+1)r_2}(I_0) \subset h_{(k+1)r_2}(C)$, and so, by Lemma 2.3,

$$\mu\{h_{(k+1)r_2}(I_0)\} \leq \mu(h_{(k+1)r_2}(C)) = \mu(C) < \eta,$$

the desired contradiction.

Let the dissection of I induced by h from $U(h)$,

$$0 = v_0 < v_1 \ldots < v_{m_2} < v_{m_2+1} = 1,$$

be denoted by

$$V = V(h) = \{v_1, \ldots, v_{m_2}\} = \{h(u^+), h(u^-);\ u \in U(h)\} \setminus \{0, 1\}, \tag{2.2}$$

and, for $H = h_{r_1}$, let $V = V(H)$ be defined accordingly; because of A2, $|H^{-1}(y)| \geq 4$ for all $y \notin H(U(H))$. Set

$$\eta_3 = \min_{0 \leq i \leq m_2(H)} \mu\{(v_i, v_{i+1})\},$$

with $\eta_3 > 0$, because $\operatorname{supp}(\mu) = I$. Now, for any $A \subset I$, writing $A_i = A \cap J_i(H)$ (see (1.1)), we have

$$\mu(H(A)) \leq \sum_i \mu(H(A_i)) = \sum_i \int_{A_i} f(H(x))|H'(x)|\, dx \leq K\lambda(A),$$

by (1.3) and because $|H'| \leq C(h)^{r_1}$; whereas $\mu(A) = \int_A f(x)\, dx$ is bounded below by some $w(\lambda(A))$ with $w(y) > 0$ for all $y > 0$, because $\operatorname{supp}(\mu) = I$. Hence there is an $\eta_4 > 0$ such that, if $\mu(H(A)) \geq \eta_3$, then $\mu(A) \geq \eta_4$ also.

Lemma 2.5. *Let W be any union of k non–empty intervals such that $\mu(W) < \eta_4$. Define*

$$\mathcal{H} = \mathcal{H}(W) = \{J : \ J \text{ a } \textit{maximal} \text{ interval s.t. } \lambda(J) > 0 \text{ and } H^{-1}(J \setminus V(H)) \subset W\}. \quad (2.3)$$

Then, under A1–A3, $\sum_{J\in\mathcal{H}} \mu(J) \le \mu(W) < \eta_4$ and $|\mathcal{H}| \le k-1$.

Proof. For the first part, observe that the intervals $J \in \mathcal{H}$ are non–overlapping, and hence

$$\sum_{J\in\mathcal{H}} \mu(J) = \mu\Big(\bigcup_{J\in\mathcal{H}} J\Big) = \mu\Big(H^{-1}\Big(\bigcup_{J\in\mathcal{H}} J^*\Big)\Big),$$

because μ is invariant; however, from the definition of $\mathcal{H}$, $H^{-1}\Big(\bigcup_{J\in\mathcal{H}} J^*\Big) \subset W$.

For the second part, if W is such a union, then $W \backslash U(H)$ is a collection of $l \le k+m_1(H)$ intervals $I_1, \ldots, I_l$. Write

$$G_j = H(I_j) = \langle a_j, b_j \rangle, \quad 1 \le j \le l, \quad (2.4)$$

where the angle brackets indicate that the endpoints may or may not belong to the interval. We now partition the G_j into overlapping clusters. Define an equivalence relation on $\{1, 2, \ldots, l\}$ by $j \sim j'$ if there exists a t and $j_1, \ldots, j_t$ such that

$$G_j \cap G_{j_1} \ne \emptyset, \ G_{j_1} \cap G_{j_2} \ne \emptyset, \ \ldots, \ G_{j_t} \cap G_{j'} \ne \emptyset. \quad (2.5)$$

Let the equivalence classes be $P_1, \ldots, P_u$. Since $\mu(W) < \eta_4$, it follows that $\mu(H(W)) < \eta_3$, and hence that each contiguous cluster

$$\Gamma_i = \bigcup_{j\in P_i} G_j \quad (2.6)$$

contains at most one point of V. If $|\Gamma_i \cap V| = 0$, set $s_i = 0$. If $\Gamma_i \cap V = \{v_j\}$, set

$$s_i = |\{t \in P_i : \ b_t = v_j\}| + |\{t \in P_i : \ a_t = v_j\}|. \quad (2.7)$$

We now count the intervals of $\mathcal{H}$ which are contained in a cluster Γ_i.

To do this, we further define

$$w_i = |H^{-1}(y)| \quad \text{for} \quad v_{i-1} < y < v_i, \qquad 1 \le i \le m_2 + 1; \quad (2.8)$$

note that $|H^{-1}(y)|$ is indeed constant over such intervals. We then define the index sets $\mathcal{J}(y) = \{j : \ I_j \cap H^{-1}(y) \ne \emptyset\}$. If $J \in \mathcal{H}$ and $y \in J \cap (v_{i-1}, v_i)$, we must have

$$|\mathcal{J}(y)| = w_i \ge 4. \quad (2.9)$$

This is the key observation for the counting argument. We now distinguish five cases.

Case 1: $\Gamma_i \subset (v_{j-1}, v_j)$ for some j.

If $J \in \mathcal{H}$ and $J \subset \Gamma_i$, identify J with an r, $1 \le r \le |P_i|$, defined by

$$r = r(J) = \max\{R(r');\ r' \in P_i, b_{r'} = \sup J\}, \tag{2.10}$$

where $R(r')$ is the rank in *descending* order of $b_{r'}$ among $\{b_t,\ t \in P_i\}$: within ties, take any fixed order. Because of (2.9), $r \ge w_j$; also distinct intervals J, being non–overlapping, have distinct $r(J)$. Hence

$$|\{J \in \mathcal{H}:\ J \subset \Gamma_i\}| \le |P_i| - (w_j - 1) \le |P_i| - (s_i \vee 3), \tag{2.11}$$

the last inequality being true because $w_j \ge 4$ and $s_i = 0$ in this case.

Case 2: $\Gamma_i \ni v_j$ and $|\mathcal{J}(v_j^+)| < w_{j+1}$, $|\mathcal{J}(v_j^-)| < w_j$.

In this case, v_j is not covered by an interval $J \in \mathcal{H}$, though it may still happen that $H^{-1}(v_j) \in W$.

For $J \in \mathcal{H}$ such that $J \subset \Gamma_i \cap (v_j, 1]$, identify it with

$$\bar{r} = \bar{r}(J) = \max\{\overline{R}(r');\ r' \in P_i, a_{r'} = \inf J\}, \tag{2.12}$$

where $\overline{R}(r')$ is the rank of $a_{r'}$ among $\{a_t;\ t \in P_i\}$ in *ascending* order. Because of (2.9) and the assumption of Case 2, $a_{\bar{r}} > v_j$, and distinct intervals J have different $\bar{r}(J)$.

For $J \in \mathcal{H}$ such that $J \subset \Gamma_i \cap [0, v_j)$, identify J with $r(J)$ as defined in (2.10), obtaining distinct indices r with $b_r < v_j$, so that they are also distinct from the $\bar{r}$ obtained above. Hence, in this case,

$$|\{J \in \mathcal{H}:\ J \subset \Gamma_i\}| \le |P_i| - |\{t \in P_i:\ a_t = v_j\}| - |\{t \in P_i:\ b_t = v_j\}| \le |P_i| - s_i; \tag{2.13}$$

also, arguing as in Case 1, one has

$$|\{J \in \mathcal{H}:\ J \subset \Gamma_i\}| \le |P_i| - (\min(w_j, w_{j+1}) - 1) \le |P_i| - 3. \tag{2.14}$$

Case 3: $\Gamma_i \ni v_j$ and $|\mathcal{J}(v_j^+)| = w_{j+1}$, $|\mathcal{J}(v_j^-)| < w_j$.

Use the algorithm of Case 1, identifying each J with $r(J)$. As before, $r(J) \ge w_{j+1}$. Also, no J has supremum v_j, because of the assumption of Case 3. Thus

$$|\{J \in \mathcal{H}:\ J \subset \Gamma_i\}| \le |P_i| - (w_{j+1} - 1) - |\{t \in P_i:\ b_t = v_j\}| \le |P_i| - (s_i - 1) \vee 3. \tag{2.15}$$

Case 4: $\Gamma_i \ni v_j$ and $|\mathcal{J}(v_j^+)| < w_{j+1}$, $|\mathcal{J}(v_j^-)| = w_j$.

In this case, the argument is similar to that of Case 3, leading again to the estimate

$$|\{J \in \mathcal{H}: \ J \subset \Gamma_i\}| \le |P_i| - (s_i - 1) \vee 3. \tag{2.16}$$

Case 5: $\Gamma_i \ni v_j$ and $|\mathcal{J}(v_j^+)| = w_{j+1}$, $|\mathcal{J}(v_j^-)| = w_j$.

In this case, there is an interval $J \in \mathcal{H}$ which contains v_j, so no J has supremum v_j. Now argue as in Case 3, obtaining the same inequality.

Collecting the results of Cases 1–5, we find that

$$|\mathcal{H}| \le l - \sum_i (s_i - 1) \vee 3,$$

implying that

$$|\mathcal{H}| \le l - 3 = k + (l - k) - 3, \tag{2.17}$$

and that

$$|\mathcal{H}| \le l - \sum_i 3s_i/4 \le l - 3(l-k)/2 = k - (l-k)/2. \tag{2.18}$$

Taking the worst value of $l - k$ for (2.17) and (2.18) still gives $|\mathcal{H}| \le k - 1$, as claimed.

Theorem 2.6. *Under A1–A3, given any non–empty interval $K \subset I$, there exists an $r_0 = r_0(K)$ and a finite set $N = N(K) \subset I$ such that $h_{r_0}(K) = I \setminus N$.*

Proof. By Theorem 2.4, there exists an r_2 such that $\mu(h_{r_2}^{-1} h_{r_2}(K)) > 1 - \eta_4$. By the invariance of μ, $\mu(h_{r_2}(K)) > 1 - \eta_4$, and $[h_{r_2}(K)]^c$ is of the form $W_0 \cup N_0$, where W_0 is a union of some number k_0 of intervals as for Lemma 2.5, and N_0 is a finite set. Furthermore, applying $H = h_{r_1}$, $W_1 = \bigcup_{J \in \mathcal{H}(W_0)} J$ is, but for a finite set of exceptional points, just $[H h_{r_2}(K)]^c$, and, by Lemma 2.5, is a union of $k_1 \le k_0 - 1$ intervals and satisfies $\mu(W_1) < \eta_4$. Thus, recursively applying H in all $k' \le k$ times and using Lemma 2.5 at each step, we arrive at a point where $H_{k'} h_{r_2}(K) = h_{k' r_1 + r_2}(K)$ consists of all but finitely many points of I.

The remainder of the section is concerned with showing that the invariant density f has 'nice' properties outside a set of λ–measure zero. The properties of f implied by (1.3) apply only off certain null sets. We shall use some specific consequences of (1.3).

Lemma 2.7. *Under A1–A3, there exists an $N = N_{2.7} \subset I$ with $\lambda(N) = 0$ and a $K_{2.7} < \infty$ such that, for any $0 < \varepsilon < 1$,*

$$\int_0^1 M_x(f, \varepsilon; N)\, dx \le K_{2.7} \varepsilon^{\zeta},$$

where

$$M_x(f,\varepsilon;N) = \sup_{y,z\in S_x(\varepsilon)\setminus N} |f(y)-f(z)|. \tag{2.19}$$

Proof. By successively removing countably many null sets from I, whose union we call N, we can ensure that

$$\sup_{y,z\in S_{jr}\setminus N} |f(y)-f(z)| \le \mathrm{osc}\,(f;2^{-r},x_{jr}),$$

where $S_{jr} = S_{x_{jr}}(2^{-r})$ and $x_{jr} = j2^{-r}$ for $1 \le j \le 2^r-1$, $r \ge 1$. Then, given $0<\varepsilon<1/4$, pick $r = r(\varepsilon) = [-\log_2 \varepsilon]-1$; it follows that $S_x(\varepsilon) \subset S_{jr} \subset S_x(6\varepsilon)$, where x_{jr} is the nearest of the set $\{x_{ir},\ 1 \le i \le 2^r-1\}$ to x. Hence

$$\int_0^1 M_x(f,\varepsilon;N)\,dx \le 2^{-r+1}\sum_{j=1}^{2^r-1} \mathrm{osc}\,(f;2^{-r},x_{jr}) \le 2\int_0^1 \mathrm{osc}\,(f;6\varepsilon,x)\,dx \le 2K_{(1.3)}(6\varepsilon)^{\zeta},$$

and the lemma follows.

Lemma 2.8. *Under A1–A3, there exists* $N = N_{2.8} \subset I$ *with* $\lambda(N)=0$ *such that*

$$f_{max} = \sup_{x\in I^*\setminus N} f(x) < \infty; \quad f_{min} = \inf_{x\in I^*\setminus N} f(x) > 0.$$

Proof. The first claim follows directly from (1.3). Then, since μ is invariant, we can take f to satisfy

$$f(y) = \sum_{x\in h_r^{-1}(y)} f(x)/|h_r'(x)| \tag{2.20}$$

(see (1.5)) for all $y \in I^*$ and all $r \ge 1$. Take any of the intervals S_{jr} defined in Lemma 2.7, and let $r_0 = r_0(S_{jr})$ be as in Theorem 2.6, so that $N_{jr} = I \setminus h_{r_0}(S_{jr})$ is finite. If $\inf\{f(x):\ x \in I^* \setminus \{h_{r_0}(N_{2.7}) \cup N_{jr}\}\} = 0$, then (2.20) implies that S_{jr} contains a zero of f or points where f is arbitrarily small, so that then

$$\sup_{y\in S_{jr}\setminus N_{2.7}} f(y) \le \mathrm{osc}\,(f;2^{-r},x_{jr}).$$

Hence, taking

$$N = N_{2.8} = \Big\{\bigcup_{r\ge 0} h_r(N_{2.7})\Big\} \cup \Big\{\bigcup_{r\ge 1}\bigcup_{1\le j\le 2^r-1} N_{jr}\Big\},$$

it follows as in the proof of Lemma 2.7 that, if $\inf_{x\in I^*\setminus N} f(x) = 0$, then $\int_0^1 f(x)\,dx = 0$, which is impossible, since $\mu(I)=1$.

The points of $U(h)$ split I into intervals $J_i = J_i(h)$, $1 \le i \le m_1+1$. Let $B_i = h(J_i)$, and let ϕ_i denote the inverse of $h|_{J_i}$. For $y \in B_i^* \setminus N_{2.8}$, define

$$q_i(y) = f(\phi_i(y))|\phi_i'(y)|/f(y). \tag{2.21}$$

Lemma 2.9. *There exists a $K_{2.9} < \infty$ and an $N = N_{2.9} \subset I$ with $\lambda(N) = 0$ such that, for any $0 < \varepsilon < 1$ and for each i,*

$$\int_0^1 M_x(q_i, \varepsilon; N)\, dx \le K_{2.9}\varepsilon^{\zeta}.$$

Proof. Take $N = N_{2.9} = \cup_{r\ge 0} h_r(N_{2.8})$. Then the functions $f \circ \phi_i$ and $|\phi_i'|$ are bounded off N, and f is also bounded away from zero off N. Hence, for $x \in B_i$ and $y, z \in (B_i \cap S_x(\varepsilon)) \setminus N$,

$$|q_i(y) - q_i(z)| \le K_1|f(y) - f(z)| + K_2|f(\phi_i(y)) - f(\phi_i(z))| + K_3|\phi_i'(y) - \phi_i'(z)|, \quad (2.22)$$

for suitable constants K_l, $1 \le l \le 3$. Now, for such y, z, $|f(y) - f(z)| \le M_x(f, \varepsilon; N)$, and $|f(\phi_i(y)) - f(\phi_i(z))| \le M_{\phi_i(x)}(f, \varepsilon/c(h); N)$, since also $|\phi_i'| \le 1/c(h)$; and, by A3, $|\phi_i'(y) - \phi_i'(z)| \le K_{(1.3)}\varepsilon^{\zeta}$. Thus, in view of Lemma 2.7, it remains to be shown that

$$\int_0^1 M_{\phi_i(x)}(f, \varepsilon/c(h); N)\, dx \le K_4\varepsilon^{\zeta}, \quad (2.23)$$

for some $K_4 < \infty$. However, since μ is invariant for h,

$$\int_0^1 M_{\phi_i(x)}(f, \varepsilon/c(h); N)\, dx \le \frac{f_{max}}{f_{min}} \int_0^1 M_x(f, \varepsilon/c(h); N)\, dx,$$

so that Lemmas 2.7 and 2.8 conclude the proof.

The importance of the q_i comes from the fact that f is the invariant density, since then, from (1.6), for $y \in B_i^* \setminus N_{2.9}$, we find that $q_i(y) = p(y, \phi_i(y))$ is a transition probability of the time reversal of the stationary h–process. In view of Lemma 2.8, each q_i is uniformly bounded away from zero on $B_i \setminus N_{2.9}$; furthermore, if

$$F(\varepsilon, a) = \bigcup_i \{x :\ M_x(q_i, \varepsilon; N_{2.9}) > a\}, \quad (2.24)$$

it follows from Lemmas 2.8 and 2.9 that

$$\mu(F(\varepsilon, a)) \le K_{2.9} f_{max}\varepsilon^{\zeta}/a \quad (2.25)$$

for all $a > 0$ and $0 < \varepsilon < 1$.

3. Coupling

Consider a Markov chain $(Y_n,\ n \geq 0)$ on the set $T = I^* \setminus N_{2.9} \subset I$ with transition probabilities given by

$$y \mapsto \phi_i(y) \quad \text{with probability} \quad q_i(y)1_{B_i}(y); \tag{3.1}$$

note that $\lambda(T) = 1$ and that $h^{-1}(T) \subset T$, so that the chain is almost surely well defined for all $n \geq 0$ if $\mathbb{P}[Y_0 \in T] = 1$, and in particular if $Y_0 \sim \mu$. The Markov chain $(Y_n,\ n \geq 0)$ can be recursively constructed in the following standard fashion, as a function of a sequence $(U_n,\ n \geq 0)$ of independent $U[0,1]$ random variables and a starting value y_0, which may depend on U_0. Given $Y_n = y$, set

$$Y_{n+1} = \phi_i(y) \quad \text{if} \quad \sum_{r=1}^{i-1} q_r(y)1_{B_r}(y) < U_{n+1} \leq \sum_{r=1}^{i} q_r(y)1_{B_r}(y). \tag{3.2}$$

In what follows, we shall usually be interested in realizing two or more such chains simultaneously, in such a way that, as far as possible, their paths remain close together. The main result is that of Theorem 3.4, showing that a pair of realizations with almost arbitrary starting points can be coupled so as to approach each other geometrically fast. This is then used in Theorem 3.5, to show that any chain Y constructed in accordance with (3.2) can be closely approximated by an m–dependent sequence Y', with $m = m(N) = O(\log N)$.

The joint realization of two such chains, $(Y_n^1,\ n \geq 0)$ and $(Y_n^2,\ n \geq 0)$, is achieved by realizing Y^1 as in (3.2) from a sequence U^1 of independent uniform random variables, and then realizing Y^2 as a function of Y^1 and a second sequence U^2 of uniform random variables, according to the following rules. Define $I_n^j = i$ whenever $Y_n^j \in C_i$, where $(C_i,\ 1 \leq i \leq m_2(h)+1)$ are the intervals into which $V(h)$ dissects I, and set $J_{n+1}^j = r$ whenever $Y_{n+1}^j = \phi_r(Y_n^j)$, $j = 1, 2$. Then, given Y_n^2 and the whole path Y^1, determine Y_{n+1}^2 according to (3.2) with U_{n+1}^2 for U_{n+1} if $I_n^1 \neq I_n^2$; if $I_n^1 = I_n^2$, define

$$Y_{n+1}^2 = \begin{cases} \phi_r(y_2) & \text{if } U_{n+1}^2 \leq p_{r2}/p_{r1}; \\ \phi_l(y_2) & \text{if } U_{n+1}^2 > p_{r2}/p_{r1} \text{ and } R_{l-1} < (1-U_{n+1}^2)/(1-p_{r2}/p_{r1}) \leq R_l, \end{cases} \tag{3.3}$$

where

$$r = J_{n+1}^1; \quad y_1 = Y_n^1; \quad y_2 = Y_n^2; \quad p_{rj} = q_r(y_j),\ j = 1, 2, \tag{3.4}$$

and

$$R_l = \sum_{s=1}^{l} (p_{s2} - p_{s1})^+ 1_{B_s}(y_2) \Big/ \sum_{s=1}^{m_2(h)+1} (p_{s2} - p_{s1})^+ 1_{B_s}(y_2). \tag{3.5}$$

This construction makes Y^2 choose the same branch as Y^1 with as high a probability as possible when Y^1 and Y^2 are both in the same set C_i. It follows, in particular, that

$$\mathbb{P}[J^2_{n+1} \neq J^1_{n+1} \,|\, \Sigma^{12}_{\infty n} \cap \{I^1_n = I^2_n\} \cap \{J^1_{n+1} = r\}] = 1 - (q_r(Y^1_n) \wedge q_r(Y^2_n))/q_r(Y^1_n), \quad (3.6)$$

where $\Sigma^j_l = \sigma(Y^j_s,\ 0 \le s \le l)$ and $\Sigma^{12}_{ln} = \Sigma^1_l \vee \Sigma^2_n$, and hence that, for some $K_{(3.7)} < \infty$,

$$\mathbb{P}[J^2_{n+1} \neq J^1_{n+1} \,|\, \{I^1_n = I^2_n\} \cap \Sigma^{12}_{\infty n}] \le K_{(3.7)} |q_{J^1_{n+1}}(Y^1_n) - q_{J^1_{n+1}}(Y^2_n)|. \quad (3.7)$$

Now define

$$D_{sl} = \bigcap_{n=0}^{l} \{Y^1_n \notin E_{n+s}\} \in \Sigma^1_l,$$

where

$$E_r = \{y \in T : \min_{v \in V(h) \cup \{0,1\}} |y - v| < c^{-r}\} \cup F(c^{-r}, c^{-r\zeta'}), \quad (3.8)$$

now henceforth with $c = c(h)$ and with ζ' fixed, $0 < \zeta' < \zeta$: $F(\varepsilon, a)$ as in (2.24). The sets D_{sl} are those in which Y^1 does not too soon approach points where the q_i may change abruptly.

Lemma 3.1. *For any $y_1, y_2 \in T$ satisfying $|y_2 - y_1| < c^{-s}$, we have*

$$\mathbb{P}\Big[\bigcap_{n=0}^{l} \{|Y^2_n - Y^1_n| < c^{-(s+n)}\} \,|\, Y^2_0 = y_2, Y^1_0 = y_1, \Sigma^1_\infty \cap D_{sl}\Big] \ge 1 - K_{3.1} c^{-s\zeta'},$$

where $K_{3.1} = K_{(3.7)} c/(c-1)$.

Proof. If s is such that $K_{3.1} c^{-s\zeta'} \ge 1$, there is nothing to prove. Otherwise, since $|\phi'_r(y)| \le c^{-1}$ for all r and all $y \in B_r$, it follows that, for all $y, z \in C_k$ and all r such that $B_r \supset C_k$, we have $|\phi_r(y) - \phi_r(z)| \le c^{-1}|y - z|$. Hence it suffices to show that, if $|y_2 - y_1| < c^{-s}$, then

$$\mathbb{P}\Big[\bigcap_{n=0}^{l} \{J^1_n = J^2_n\} \,\Big|\, Y^2_0 = y_2, Y^1_0 = y_1, \Sigma^1_\infty \cap D_{sl}\Big] \ge 1 - K_{3.1} c^{-s\zeta'}.$$

However, from (3.7), (3.8) and (2.24), it follows that

$$\mathbb{P}\Big[J^2_{n+1} \neq J^1_{n+1} \,\Big|\, \bigcap_{r=1}^{n} \{J^1_r = J^2_r\} \cap \{|Y^2_0 - Y^1_0| < c^{-s}\} \cap D_{sl} \cap \Sigma^{12}_{\infty n}\Big] \le K_{(3.7)} c^{-(s+n)\zeta'}, \quad (3.9)$$

and the lemma follows.

Remark. By choosing $s \ge s_0$ for a suitable s_0, the lower bound in Lemma 3.1 can be made to exceed 9/10.

Now define

$$P(s,y) = \mathbb{P}[D_{s\infty} \mid Y_0^1 = y]; \qquad \pi(s,y) = \mathbb{E}\Big\{\sum_{n\ge 0} 1_{E_{n+s}}(Y_n^1) \mid Y_0^1 = y\Big\}, \tag{3.10}$$

and observe that

$$1 - P(s,y) = \mathbb{P}\Big[\sum_{n\ge 0} 1_{E_{n+s}}(Y_n^1) > 0 \,\Big|\, Y_0^1 = y\Big] \le \pi(s,y).$$

Since Y^1 is stationary if $Y_0^1 \sim \mu$, it follows that

$$\begin{aligned}\int_0^1 \pi(s,y)\,\mu(dy) &= \sum_{n\ge 0} \mu(E_{n+s}) \\ &\le f_{max} \sum_{n\ge 0} [2(m_2(h)+1)c^{-(s+n)} + \mu\{F(c^{-(s+n)}, c^{-(s+n)\zeta'})\}] \\ &\le f_{max}\{2(m_2(h)+1)c^{-s}/(c-1) + K_{2.9}c^{-s(\zeta-\zeta')}/(c^{\zeta-\zeta'}-1)\},\end{aligned} \tag{3.11}$$

where the final estimate comes from (2.25). Choose $s = s_1 \ge s_0$ large enough to make the final bound in (3.11) smaller than $1/10$. Then the set

$$S = \{y \in T:\ \pi(s_1, y) < 1/10\}$$

has $\mu(S) > 0$, and, for each $y \in S$, $P(s_1, y) > 9/10$.

We next show that, if $Y_0^1 \in S$ and Y_0^2 is close enough to Y_0^1, then there is a good chance that the paths of Y^1 and Y^2 never get far from one another.

Lemma 3.2. *For any $y_1 \in S$ and $y_2 \in T$ such that $|y_2 - y_1| < c^{-s_1}$, we have*

$$\mathbb{P}\Big[\bigcap_{n\ge 0}\{|Y_n^2 - Y_n^1| < c^{-s_1-n}\} \,\Big|\, Y_0^1 = y_1, Y_0^2 = y_2\Big] \ge 8/10.$$

Proof. From Lemma 3.1 and because $s_1 \ge s_0$,

$$\mathbb{P}\Big[\bigcap_{n\ge 0}\{|Y_n^2 - Y_n^1| < c^{-s_1-n}\} \,\Big|\, Y_0^2 = y_2, Y_0^1 = y_1, \Sigma_\infty^1 \cap D_{s_1,\infty}\Big] \ge 9/10. \tag{3.12}$$

Hence

$$\begin{aligned}&\mathbb{P}\Big[\bigcap_{n\ge 0}\{|Y_n^2 - Y_n^1| < c^{-s_1-n}\} \,\Big|\, Y_0^2 = y_2, Y_0^1 = y_1\Big] \\ &\ge \mathbb{P}[D_{s_1,\infty} \mid Y_0^1 = y_1] \\ &\qquad \times \mathbb{E}\Big\{\mathbb{P}\Big[\bigcap_{n\ge 0}\{|Y_n^2 - Y_n^1| < c^{-s_1-n}\} \,\Big|\, Y_0^2 = y_2, Y_0^1 = y_1, \Sigma_\infty^1 \cap D_{s_1,\infty}\Big]\Big\} \\ &\ge 9P(s_1,y_1)/10 > 81/100,\end{aligned}$$

which proves the lemma.

Now take any $y_0 \in S$ and set $J = S_{y_0}(c^{-s_1}) \cap T$. The next lemma shows that if Y_0^1 and Y_0^2 are both in J, then Y^1 and Y^2 stay close to one another for ever with substantial probability.

Lemma 3.3. *For any $y_1, y_2 \in J$, it is possible to couple two Markov chains Y^1 and Y^2 with transitions governed by (3.1) in such a way that*

$$\mathbb{P}\Big[\bigcap_{n\geq 0}\{c^n|Y_n^2 - Y_n^1| < 2c^{-s_1}\}\,|\,Y_0^2 = y_2, Y_0^1 = y_1\Big] \geq 6/10.$$

Proof. Realize a chain $\widetilde{Y}$ according to (3.2) with $\widetilde{Y}_0 = y_0$. Then realize two processes $\overline{Y}$ and $\widehat{Y}$ with $\overline{Y}_0 = y_1$ and $\widehat{Y}_0 = y_2$, each coupled to $\widetilde{Y}$ as in (3.3). Apply Lemma 3.2 to each pair, and then observe that, by the triangle inequality,

$$|\overline{Y}_n - \widehat{Y}_n| \leq |\overline{Y}_n - \widetilde{Y}_n| + |\widetilde{Y}_n - \widehat{Y}_n|.$$

Now set $Y^1 = \overline{Y}$ and $Y^2 = \widehat{Y}$.

Lemma 3.3 shows that Y^1 and Y^2 can be constructed in such a way that they stay close for all time with substantial probability, provided that they are both initially in the (small) set J. We can now greatly extend the scope of this result, proving that Y^1 and Y^2 can be realized in such a way that their paths are eventually close for all time with probability 1, whatever their starting states.

Theorem 3.4. *If A1–A3 hold, then there exist $K_{3.4} < \infty$ and $0 < \beta < \zeta$ such that, for a suitable joint realization of Y^1 and Y^2,*

$$\mathbb{P}[\sup_{r\geq 0} c^r|Y_r^2 - Y_r^1| \geq x\,|\,Y_0^2 = y_2, Y_0^1 = y_1] \leq K_{3.4}x^{-\beta},$$

uniformly for all $x > 0$ and $y_1, y_2 \in T \setminus N$, where $N = N_{3.4}$ is finite.

Proof. From Theorem 2.6, there exists an r_0 such that $h_{r_0}(S_{y_0}(c^{-s_1})) = I \setminus N$, with $N = N_{3.4}$ finite. Hence, for any $y \in T \setminus N$, $h_{r_0}^{-1} \cap J$ is non–empty. Thus, letting Y^1 and Y^2 evolve independently for r_0 steps, it follows that

$$\mathbb{P}[Y_{r_0}^1 \in J, Y_{r_0}^2 \in J\,|\,Y_0^2 = y_2, Y_0^1 = y_1] \geq \{\inf_{i,y} q_i(y)\}^{2r_0} = \delta > 0,$$

by Lemma 2.8, for all pairs $y_1, y_2 \in T \setminus N$. Hence, from Lemma 3.3,

$$\mathbb{P}\Big[\bigcap_{n\geq r_0}\{c^n|Y_n^2 - Y_n^1| < 2c^{r_0-s_1}\,|\,Y_0^2 = y_2, Y_0^1 = y_1\Big] \geq 6\delta/10, \tag{3.13}$$

for all such pairs.

Now, given $y_1, y_2 \in T \setminus N$, define Y^1 and Y^2 for all time as follows. Let $Y_0^1 = y_1$ and $Y_0^2 = y_2$, and let Y^1 and Y^2 evolve independently for r_0 steps. Then, if $\{y_1, y_2\} \not\subset J$, set

$N_1 = r_0$; otherwise, using the Markov property, let Y^1 and Y^2 continue to evolve according to a coupling constructed as for Lemma 3.3, and set

$$N_1 = \inf\{n \geq r_0 :\ c^n|Y_n^2 - Y_n^1| \geq 2c^{r_0 - s_1}\} \leq \infty. \tag{3.14}$$

If $N_1 = \infty$, the chains Y^1 and Y^2 are fully defined. If $N_1 < \infty$, use the strong Markov property to restart the whole construction from new initial values $Y_{N_1}^1$ and $Y_{N_1}^2$, covering all times $0 \leq n \leq N_1 + N_2 \leq \infty$, and continue to repeat until some $N_j = \infty$. Note that, in view of (3.9), for any $j, l \geq 0$,

$$\mathbb{P}[N_j I[N_j < \infty] \geq l] = \sum_{t \geq l} \mathbb{P}[N_j = r_0 + t] \leq \sum_{t \geq l} K_{(3.7)} c^{-(s_1+t)\zeta'};$$

hence, from (3.13), the sum $\sum_{j \geq 1} N_j I[N_j < \infty]$ is stochastically dominated by a sum $N^* = \sum_{j=1}^{\tau} \widetilde{N}_j$ of independent random variables $\widetilde{N}_j$, each with distribution

$$\mathbb{P}[\widetilde{N}_j = r_0 + l] = K_{(3.7)} c^{-(s_1+l)\zeta'},\ l \geq 1; \quad \mathbb{P}[\widetilde{N}_j = r_0] = 1 - K_{(3.7)} c^{-s_1\zeta'}/(c-1),$$

where τ is independent of $\{\widetilde{N}_j,\ j \geq 1\}$ and has geometric distribution

$$\mathbb{P}[\tau \geq l] = \Bigl(1 - \frac{6\delta}{10}\Bigr)^l, \quad l \geq 0.$$

Hence

$$\mathbb{P}[\sup_{r \geq 0} c^r |Y_r^2 - Y_r^1| \geq 2c^{l+r_0-s_1} \mid Y_0^2 = y_2, Y_0^1 = y_1] \leq \mathbb{P}[N^* \geq l],$$

uniformly in $y_1, y_2 \in T \setminus N$.

Now $\widetilde{N}_j$ has a probability generating function $\widetilde{\Gamma}(z)$ which converges in $|z| < c^{\zeta'}$. Hence the probability generating function of N^* also converges for some $1 < z_0 < c^{\zeta'}$, implying that $\mathbb{P}[N^* \geq l] \leq z_0^{-l}\mathbb{E}\{z_0^{N^*}\}$. Hence, given any $x > 0$, take $l = [s_1 - r_0 + \log(x/2)/\log c]$, giving

$$\mathbb{P}[\sup_{r \geq 0} c^r |Y_r^2 - Y_r^1| \geq x] \leq \mathbb{E}\{z_0^{N^*}\} e^{-l \log z_0} \leq K_{3.4} x^{-\beta},$$

for suitable $K_{3.4} < \infty$ and $\beta > 0$.

Remark. Note that the value of β indicated by the proof is smaller than ζ, and, through the probability generating function of N^*, is also small with δ. It is not in general obvious how to make good estimates of the best possible β from the properties of h, except in very simple cases. In the traditional approach, some progress has been made in the analogous problem: see Liverani (1995).

An important consequence of Theorem 3.4 is that, in a certain limited sense, the value of Y_n has little effect on that of Y_{n+m} when m is large. This can be made precise in the following theorem.

Theorem 3.5. *Suppose that the chain $(Y_n,\ n \geq 0)$ is constructed as in (3.2) from a sequence of independent $U[0,1]$ random variables $(U_n,\ n \geq 0)$, with U_0 determining the value of Y_0. Then it is possible to construct a sequence $(Y'_n,\ n \geq 0)$ of m–dependent random variables in such a way that*

$$\mathbb{P}\Big[\bigcup_{j=m}^{N+m} \{|Y'_j - Y_j| \geq N^{-2}\}\Big] \leq K_{3.4} N^{-2}, \tag{3.15}$$

if $m \geq (2+3/\beta)\log N/\log c$.

Proof. With $(Y_n,\ n \geq 0)$ constructed as in (3.2), and for any fixed m, construct random variables $(Y'_{n+m},\ n \geq 0)$ using an additional, independent sequence $(U'_n,\ n \geq 0)$ of independent $U[0,1]$ random variables; for each $n \geq 0$, start a chain $(Y^{(n)}_{n+r},\ r \geq 0)$ by letting U'_n determine $Y^{(n)}_n$, and then run $Y^{(n)}$ using (3.2) and the *original* values $(U_{n+r},\ r \geq 1)$; set $Y^{(n)}_{n+m} = Y'_{n+m}$. Because of Theorem 3.4, for each $n \geq 0$, we have

$$\mathbb{P}[|Y'_{n+m} - Y_{n+m}| \geq xc^{-m}] \leq K_{3.4} x^{-\beta}, \quad x > 0,$$

and (3.15) follows if $m \geq (2+3/\beta)\log N/\log c$.

Thus m need only be of order $\log N$ for Y' to be a uniformly small perturbation of Y throughout an index set of length N. The advantage of the sequence Y' is that longer term dependence has been entirely eliminated.

4. Rates of approximation

4.1. Decay of Correlations

The first set of implications of Theorem 3.4 concern the rate of decay of the dependence upon initial conditions in h–sequences. This rate itself depends on the form of the initial condition — if x_0 were exactly known, there would be no decay at all. The quantities in terms of which we express our rates are derived from the following measure of the average smoothness of a function. For any $g : [0,1] \to \mathbb{R}$, we define

$$\bar{m}(g,\eta) = \int_0^1 \sup_{\{z:|x-z|<\eta\}\cap T} |g(z) - g(x)|\, dx \leq \infty. \tag{4.1}$$

Note that, from Lemma 2.7,

$$\bar{m}(f,\eta) \leq K_{2.7}\eta^{\zeta}. \tag{4.2}$$

Let g_0 be the density of X_0 with respect to λ, and let g_n denote the density of $X_n = h_n(X_0)$.

Theorem 4.1. *As $n \to \infty$, $g_n \to f$ in L_1. Furthermore, if g_0 satisfies $\lim_{\eta\to 0} \bar{m}(g_0, \eta) = 0$, then*

$$\lim_{n\to\infty} \sup_{x\in T\setminus N_{3.4}} |g_n(x) - f(x)| = 0,$$

where T is as in Section 3.

Remark. The class of densities g_0 such that $\lim_{\eta\to 0} \bar{m}(g_0, \eta) = 0$ is just the class of Riemann integrable densities, and includes in particular all densities belonging to $D[0,1]$.

Proof. We start by assuming that g_0 satisfies $\lim_{\eta\to 0} \bar{m}(g_0, \eta) = 0$, proving the first part by approximation when the second is known. Let $X'_r = h_r(X'_0)$, $1 \le r \le n$, with $X'_0 \sim \mu$, and set $Y_r = X'_{n-r}$, $0 \le r \le n$. For any $A \subset I$, we have

$$\begin{aligned}\int_A g_n(x)\,dx &= \int_I 1_A(h_n(x)) g_0(x)\,dx \\ &= \mathbb{E}\{1_A(X'_n) g_0(X'_0)/f(X'_0)\} = \mathbb{E}\{1_A(Y_0) g_0(Y_n)/f(Y_n)\},\end{aligned} \tag{4.3}$$

so that $\mathbb{E}\{g_0(Y_n)/f(Y_n) \,|\, Y_0 = y\} f(y)$, $y \in T$, is a version of g_n. Hence, realizing Y^1 and Y^2 together as for Theorem 3.4, with $Y^1_0 = y \in T \setminus N_{3.4}$ and $Y^2 \sim \mu$, we have, for any $k_n > 0$,

$$\begin{aligned}|g_n(y)/f(y) - 1| &= |\mathbb{E}\{g_0(Y^1_n)/f(Y^1_n)\} - \mathbb{E}\{g_0(Y^2_n)/f(Y^2_n)\}| \\ &\le 2GK_{3.4}k_n^{-\beta} + \int_T f(x) \sup_{\{y:|y-x|<k_n c^{-n}\}\cap T} |(g_0(y)/f(y)) - (g_0(x)/f(x))|\,dx \\ &\le 2GK_{3.4}k_n^{-\beta} + G\bar{m}(f, k_n c^{-n}) + \bar{m}(g_0, k_n c^{-n}),\end{aligned} \tag{4.4}$$

where $G = \sup_{x\in T} g_0(x)/f(x) < \infty$, by Lemma 2.8 and because $\bar{m}(g_0,\eta) < \infty$ for η sufficiently small. Choosing $(k_n,\ n \ge 1)$ in such a way that $k_n \to \infty$ and $k_n c^{-n} \to 0$, and recalling (4.2), the second part of the lemma follows.

Now if g_0 does not satisfy $\lim_{\eta\to 0} \bar{m}(g_0, \eta) = 0$, it can still be approximated arbitrarily closely in L_1 by densities g_0^ε which do. Hence we can write

$$\int_I |g_n(x) - f(x)|\,dx \le \int_T |g_n(x) - g_n^\varepsilon(x)|\,dx + \int_T |g_n^\varepsilon(x) - f(x)|\,dx,$$

with the latter integral converging to zero as $n \to \infty$ by the first part; g_n^ε denotes the density of X_n when $X_0 \sim g_0^\varepsilon$. For the former integral, we have, for $y \in T$,

$$|(g_n(y) - g_n^\varepsilon(y))/f(y)| \le \mathbb{E}\{|g_0(Y^1_n) - g_0^\varepsilon(Y^1_n)|/f(Y^1_n) \,|\, Y^1_0 = y\},$$

and hence it follows that

$$\begin{aligned}\int_T |g_n(y) - g_n^\varepsilon(y)|\,dy &\le \int_T \mathbb{E}\{|g_0(Y^1_n) - g_0^\varepsilon(Y^1_n)|/f(Y^1_n) \,|\, Y^1_0 = y\} f(y)\,dy \\ &= \int_I |g_0(x) - g_0^\varepsilon(x)|\,dx,\end{aligned} \tag{4.5}$$

which can be made arbitrarily small by choice of g_0^ε.

With slightly stronger assumptions on g_0, we can prove a geometric rate for the convergence in Theorem 4.1. For $0 < \gamma \le 1$, define

$$\begin{aligned} m_\gamma(g) &= \sup_{0<\eta\le 1} \eta^{-\gamma} \bar{m}(g,\eta) \le \infty; \\ m_0(g) &= \sup_{x,y\in T} |g(x) - g(y)| = \bar{m}(g,1). \end{aligned} \tag{4.6}$$

Note that $m_\gamma(g)$ is increasing with γ, and that, if g_1 and g_2 are such that $\inf_x g_i(x) \le 0 \le \sup_x g_i(x)$, $i = 1,2$, then $\sup_x |g_i(x)| \le 2m_0(g_i)$, and hence

$$m_\gamma(g_1 g_2) \le 4 m_\gamma(g_1) m_\gamma(g_2). \tag{4.7}$$

Note also that $m_\zeta(f) < \infty$, from (4.2).

Theorem 4.2. *Suppose that $m_\gamma(g_0) < \infty$ for some $0 < \gamma \le 1$, and set $\gamma' = \min(\gamma,\zeta)$ and $\alpha = c^{-\beta\gamma'/(\beta+\gamma')}$. Then there exists a $K_{4.2} < \infty$ such that, for all $x \in T \setminus N_{3.4}$ and all $n \ge 0$,*

$$|g_n(x) - f(x)| \le K_{4.2}\alpha^n \{1 + m_{\gamma'}(g_0)\}.$$

Proof. All that is required is to make estimates of the quantities appearing in (4.4): $G \le \{1 + 2m_{\gamma'}(g_0)\}/f_{min}$, $\bar{m}(f,\eta) \le \eta^\zeta m_\zeta(f)$ and $\bar{m}(g_0,\eta) \le \eta^\gamma m_\gamma(g_0)$. Then choose $k_n = c^{-n\gamma'/(\beta+\gamma')}$.

Theorem 4.2 implies a corresponding rate of decay of correlations.

Corollary 4.3. *Suppose that u_1 and u_2 are integrable functions, and that $m_\gamma(u_1) < \infty$ for some $0 < \gamma \le 1$. Suppose also that X_0 has a density g_0 satisfying $m_\gamma(g_0) < \infty$. Then, for some $K_{4.3} < \infty$,*

$$\begin{aligned} &\left| \mathbb{E}\{u_1(X_0) u_2(h_n(X_0))\} - \int_I u_1(x) g_0(x)\,dx \int_I u_2(x) f(x)\,dx \right| \\ &\qquad \le K_{4.3}\alpha^n \int_I |u_2(x)|\,dx \Big\{ \int_I |u_1(x)| g_0(x)\,dx + m_{\gamma'}(g_0) m_{\gamma'}(|u_1|) \Big\}, \end{aligned}$$

where, as before, $\gamma' = \min(\gamma,\zeta)$ and $\alpha = c^{-\beta\gamma'/(\beta+\gamma')}$.

Proof. It is enough to prove the corollary for nonnegative u_1, since a general u_1 can be split into its negative and nonnegative parts. Note also that a constant may be added to u_2 without changing the quantity to be estimated. If $u_1(x) \ge 0$ for all x, define

$$g(x) = g_0(x) u_1(x) \Big/ \int_I g_0(y) u_1(y)\,dy,$$

and observe that

$$\begin{aligned}\mathbb{E}\{u_1(X_0)u_2(h_n(X_0))\} &= \int_I g_0(x)u_1(x)u_2(h_n(x))\,dx \\ &= \int_I g(x)u_2(h_n(x))\,dx \int_I g_0(y)u_1(y)\,dy.\end{aligned} \tag{4.8}$$

By Theorem 4.2, we have

$$\left|\int_I g(x)u_2(h_n(x))\,dx - \int_I u_2(y)f(y)\,dy\right| \le K_{4.2}\alpha^n\{1+m_{\gamma'}(g)\}\int_I |u_2(y)|\,dy, \tag{4.9}$$

and the corollary follows, since, from (4.7),

$$m_{\gamma'}(g) \le 4m_{\gamma'}(g_0)m_{\gamma'}(u_1)\Big/\int_I g_0(y)u_1(y)\,dy.$$

Remark. The quantity estimated in Corollary 4.3, although perhaps the most useful expectation estimate, is neither a correlation nor even a covariance, since, in the product of integrals, $\mathbb{E}u_2(h_n(X_0))$ is replaced by its limiting value as $n \to \infty$, $\int_I u_2(x)f(x)\,dx$. To obtain a true correlation estimate, first observe that a constant may be added to u_2 without changing the quantity to be estimated, so that u_2 can be taken to be centered at its expectation $\bar{u}_2(n) = \mathbb{E}\{u_2(h_n(X_0))\}$. Then, substituting g_0 for g in (4.9), it follows that a similar estimate holds for the covariance as well:

$$\begin{aligned}&\mathrm{Cov}\,\{u_1(X_0), u_2(h_n(X_0))\} \\ &\qquad \le K\alpha^n \int_I |u_2(x)-\bar{u}_2(n)|g_n(x)\,dx\left\{\int_I |u_1(x)|g_0(x)\,dx + m_{\gamma'}(g_0)m_{\gamma'}(|u_1|)\right\},\end{aligned}$$

with the constant K incorporating a factor $\{\inf g_n(x)\}^{-1}$, which approaches $1/f_{min}$ as $n \to \infty$, in view of Theorem 4.2. This in turn leads to an estimate of the correlation:

$$\mathrm{Corr}\,\{u_1(X_0), u_2(h_n(X_0))\} \le K\alpha^n\{1+R(u_1)\},$$

where K depends on the properties of g_0 as well as on h, and where

$$R(u_1) = m_{\gamma'}(|u_1 - \mathbb{E}u_1(X_0)|)/\sqrt{\mathrm{Var}\,u_1(X_0)}.$$

The final result of this nature concerns probabilities of sets more general than product sets.

Theorem 4.4. *If $X_0 \sim \mu_0$, where μ_0 has density g_0 such that $m_\gamma(g_0) < \infty$, then the Lévy–Prohorov distance ρ between the distribution of $(X_0, h_n(X_0))$ and $\mu_0 \times \mu$ satisfies*

$$\rho \le K_{4.4}\alpha^n\{1 + m_{\gamma'}(g_0)\}$$

for some $K_{4.4} < \infty$, where $\gamma' = \min(\gamma, \zeta)$ and $\alpha = c^{-\beta\gamma'/(\beta+\gamma')}$ are as usual.

Proof. Let A be any measurable subset of $I \times I$, and let $A_x = \{y : \ (y,x) \in A\}$ denote the corresponding section. Then

$$\begin{aligned}\mu_0\{x : \ (x, h_n(x)) \in A\} &= \mathbb{E}\{[g_0(Y_n)/f(Y_n)]I[(Y_n, Y_0) \in A] \,|\, Y_0 \sim \mu\} \\ &= \int_I \mathbb{E}\{[g_0(Y_n)/f(Y_n)]I[Y_n \in A_x] \,|\, Y_0 = x\}f(x)\,dx\end{aligned} \tag{4.10}$$

For $x \in T \setminus N_{3.4}$, let $(Y_{n,x}, \ n \ge 0)$ be a Markov chain with transitions governed by (3.1) having $Y_{0,x} = x$, and let Y' be another with $Y'_0 \sim \mu$, realized together as in (3.3) with $Y' = Y^1$ and $Y_{n,x} = Y^2$. Then since, for any $\eta \ge 0$,

$$I[Y_{n,x} \in A_x] \le I[Y'_n \in A^\eta_x]I[|Y_{n,x} - Y'_n| < \eta] + I[|Y_{n,x} - Y'_n| \ge \eta], \tag{4.11}$$

it follows from Theorem 3.4 that

$$\begin{aligned}&\mu_0\{x : (x,h_n(x)) \in A\} \\ &\quad\le \int_T \mathbb{E}\{[g_0(Y_{n,x})/f(Y_{n,x})]I[Y'_n \in A^\eta_x]I[|Y_{n,x} - Y'_n| < \eta]\}f(x)\,dx \\ &\qquad + G\int \mathbb{P}[|Y_{n,x} - Y'_n| \ge \eta]f(x)\,dx \\ &\quad\le \int_T \mathbb{E}\{[g_0(Y'_n)/f(Y'_n)]I[Y'_n \in A^\eta_x]\}f(x)\,dx \\ &\qquad + \int_T \mathbb{E}\Big\{\Big|\frac{g_0(Y_{n,x})}{f(Y_{n,x})} - \frac{g_0(Y'_n)}{f(Y'_n)}\Big|I[|Y_{n,x} - Y'_n| < \eta]\Big\}f(x)\,dx + GK_{3.4}(\eta c^n)^{-\beta} \\ &\quad\le (\mu_0 \times \mu)\{A^\eta\} + G\bar{m}(f,\eta) + \bar{m}(g_0,\eta) + GK_{3.4}(\eta c^n)^{-\beta},\end{aligned}$$

this last from (4.4): $G = \sup_x\{g_0(x)/f(x)\}$ as before. Hence, taking $\eta = c^{-n\beta/(\beta+\gamma')}$, it follows that

$$\mu_0\{x : (x, h_n(x)) \in A\} \le (\mu_0 \times \mu)\{A^\varepsilon\} + \varepsilon, \tag{4.12}$$

with $\varepsilon = K_{4.4}\alpha^n\{1 + m_{\gamma'}(g_0)\}$, for a suitable $K_{4.4}$.

4.2. Multivariate normal approximation

Let X_0 have distribution μ, where μ is the invariant measure. Then recall that, for any N, $(X_0,\dots,X_N)$ is time-reversible, and its time reversal $(Y_0,\dots,Y_N)$ is stationary, with $Y_0 \sim \mu$. Thus for any function u, in distribution,

$$\mathcal{L}(u(X_0,\dots,X_N)) = \mathcal{L}(u(Y_0,\dots,Y_N)).$$

Hence limit theorems for functions of $(X_0,\dots,X_N)$ can be obtained by deriving limit theorems for functions of $(Y_0,\dots,Y_N)$. Due to stationarity, these are equivalent to limit thoerems for functions of $(Y_m,\dots,Y_{m+N})$, for any fixed m. The latter process we can approximate by the above stationary m-dependent process $(Y'_m,\dots,Y'_{m+N})$ for which known results can easily be applied. For normal approximations, there is a vast literature about rates of convergence for stationary m-dependent sequences. However, we will want to get an explicit dependence on m, and therefore will have to rule out results such as Stein (1972), Tikhomirov (1980), where the rate of convergence is given in terms of a constant that depends on the distribution of the m-dependent sequence in an unspecified way. Moreover, there are results about Edgeworth expansions (see Heinrich (1982), Heinrich (1985), Loh (1994), Götze and Hipp (1983)), but these involve the cumulants of the distribution of the m-dependent sequence and are therefore too complex for our goals. Instead, we will apply a result by Rinott and Rotar (1996) for multivariate normal approximation.

Let $\mathbf{J} = (J_1,\dots,J_s) \in \mathbb{N}^s$ be a fixed vector with $0 = J_1 \le J_2 \le \cdots \le J_s$, and let $I = \{i+\mathbf{J}; i = 1,\dots,N\}$, where we abbreviate $i+\mathbf{J} = (i+J_1,\dots,i+J_s)$. For each i, $1 \le i \le N$, put

$$\mathbf{X}_{i+\mathbf{J}} = (X_{i+J_1},\dots,X_{i+J_s}) \in \{X_1,\dots,X_{N+J_s}\}^s;$$

let $\mathbf{u}^{(i)}$ be a Lipschitz continuous function from $[0,1]^s$ to $\mathbf{R}^d$, with Lipschitz constant $\tau^{(i)}$ and satisfying $\mathbb{E}\mathbf{u}^{(i)}(\mathbf{X}_{1+\mathbf{J}}) = (0,\dots,0)$. Here, and in what follows, the norm of any vector or matrix is understood to be the sum of the absolute values of its elements, and vectors are understood to be column vectors. We also write $\|\mathbf{u}\| = \max_{\mathbf{y}\in[0,1]^s} \sum_{j=1}^d |\mathbf{u}(\mathbf{y})_j| \le s\tau^{(j)}$, where $\mathbf{u}(\mathbf{y})_j$ denotes the jth coordinate of $\mathbf{u}(\mathbf{y})$, and we set

$$\mathbf{u}^{(i)}(\mathbf{X}_{i+\mathbf{J}}) = \mathbf{u}^{(i)}(X_{i+J_1},\dots,X_{i+J_s}); \quad \mathbf{W} = \sum_{i=1}^N \mathbf{u}^{(i)}(\mathbf{X}_{i+\mathbf{J}}).$$

Further, we define

$$\Sigma_N = (\sigma_{i,j})_{i,j=1,\dots,d} = \mathrm{Var}\,(\mathbf{W}); \quad \tau = \max_{1\le i\le N} \tau^{(i)}; \quad \varepsilon_N = s\tau|\Sigma_N^{-\frac{1}{2}}|.$$

Theorem 4.5. *With the above definitions,*

$$\sup\left\{|\mathbb{E}g(W) - \mathbb{E}g(\sigma_N^{\frac{1}{2}}\mathcal{N})| \,:\, g \in \mathcal{G}\right\} \le O\left\{(\log N + J_s)^3(N\varepsilon_N^2 + 1)\varepsilon_N\right\},$$

where $\mathcal{G}$ is the set of indicators of convex sets in $\mathbf{R}^d$ and $\mathcal{N}$ denotes a standard normal random vector in $\mathbf{R}^d$.

Remark. A detailed form of the error estimate, derived from that given in Rinott and Rotar (1996), is given below. Although it appears rather complicated, it makes the dependence of the order terms on the parameters of the problem very explicit. In fact, there exist universal constants $c = c(d)$ and $a = a(d)$ such that

$$\begin{aligned}\sup\{|\mathbb{E}g(W) - \mathbb{E}g(\Sigma_N^{\frac{1}{2}}\mathcal{N})| \,:\, g \in \mathcal{G}\} &\leq a(2+K_{3.4})\varepsilon_N N^{-1} + K_{3.4}N^{-2} \\ &+ 3c\Big\{4ab\varepsilon_N + Nb^2\varepsilon_N^2[\log N + 2|\log\varepsilon_N| + \log(4b)]\big[8a\varepsilon_N + N^{-2}\psi\big]\Big\},\end{aligned} \tag{4.13}$$

where

$$b = m + J_s + 1; \quad \psi = 8(1+K_{3.4}) + (1-\alpha)^{-1}K_{4.3}(1 + s^{-1}K_{1.4}), \tag{4.14}$$

and

$$m = \max\left\{\frac{2+\frac{3}{\beta}}{\log c}, \frac{2}{\log\frac{1}{\alpha}}\right\}\log N. \tag{4.15}$$

It is also shown in Rinott and Rotar (1996) that similar estimates are valid, with different choices of a, for other classes $\mathcal{G}$, and that the order of the bound can be slightly improved if $d = 1$. In many applications, $|\Sigma_N^{-1/2}| = O(N^{-\frac{1}{2}})$ and $J_s = O(\log N)$, giving an error in the approximation of order $O(N^{-\frac{1}{2}}\log^3 N)$.

Proof of Theorem 4.4. Let $\{Y_1,\ldots,Y_{N+J_s}\}$ be the time reversal of $\{X_1,\ldots,X_{N+J_s}\}$, so that

$$\mathbf{W} = \sum_{i=1}^N \mathbf{u}^{(N-i+1)}(\mathbf{Y}_{i+J_s-\mathbf{J}}).$$

We approximate the time-reversed process by the m-dependent process $(Y_i')_{i=1,2,\ldots}$ constructed above, with the particular choice of m given in (4.15). By stationarity, we may shift the indices in I by $m+J_s$ without changing the distribution of $\mathbf{W}$. Alternatively, we may suppose the process to have started at time $-m-J_s$, which allows us to maintain the notation $(Y_1',\ldots,Y_N')$. Put

$$\mathbf{W}' = \sum_{i=1}^N \tilde{\mathbf{u}}^{(N-i+1)}(\mathbf{Y}'_{i-\mathbf{J}}),$$

where $\tilde{\mathbf{u}}^{(j)}(Y'_{i-\mathbf{J}}) = \mathbf{u}^{(j)}(Y'_{i-\mathbf{J}}) - \mathbb{E}\mathbf{u}^{(j)}(Y'_{1+J_s-\mathbf{J}})$, and note that $\|\tilde{\mathbf{u}}^{(j)}\| \leq s\tau$; let

$$\Sigma_N' = (\sigma'_{i,j})_{i,j=1,\ldots,d} = \mathrm{Var}\,(\mathbf{W}')$$

be the covariance matrix of $\mathbf{W}'$.

First, we transform $\mathbf{W}'$; we put

$$\mathbf{W}_{\mathcal{Z}} = \Sigma_N^{-\frac{1}{2}}\Big(\sum_{i=1}^N \mathbf{u}^{(N-i+1)}(Y_{i-\mathbf{J}})\Big) \quad \text{and} \quad \mathbf{W}'_{\mathcal{Z}} = \Sigma_N^{-\frac{1}{2}}\Big(\sum_{i=1}^N \tilde{\mathbf{u}}^{(N-i+1)}(Y'_{i-\mathbf{J}})\Big).$$

Note that, as $\mathcal{G}$ is closed under affine transformations, we have

$$\begin{aligned}\sup\{|\mathbb{E}g(\mathbf{W}) - \mathbb{E}g(\Sigma_N^{\frac{1}{2}}\mathcal{N})| \,:\, g \in \mathcal{G}\} &= \sup\{|\mathbb{E}g(\mathbf{W}_{\mathcal{Z}}) - \mathbb{E}g(\mathcal{N})| \,:\, g \in \mathcal{G}\} \\ &\le \sup\{|\mathbb{E}g(\mathbf{W}_{\mathcal{Z}}) - \mathbb{E}g(\mathbf{W}'_{\mathcal{Z}})| + |\mathbb{E}g(\mathbf{W}'_{\mathcal{Z}}) - \mathbb{E}g(\mathcal{N})| \,:\, g \in \mathcal{G}\}.\end{aligned}$$

Our strategy is now to estimate the difference $\mathbb{E}g(\mathbf{W}_{\mathcal{Z}}) - \mathbb{E}g(\mathbf{W}'_{\mathcal{Z}})$, and then to apply Rinott and Rotar's result to the m-dependent sequence $\mathbf{u}(\mathbf{Y}'_{i-\mathbf{J}})$. We have

$$\begin{aligned}&|\mathbb{E}g(\mathbf{W}_{\mathcal{Z}}) - \mathbb{E}g(\mathbf{W}'_{\mathcal{Z}})| \\ &\quad\le \mathbb{E}\Big(|g(\mathbf{W}_{\mathcal{Z}}) - g(\mathbf{W}'_{\mathcal{Z}})|\Big|\bigcup_{j=1}^N\{|Y'_j - Y_j| \ge N^{-2}\}\Big)\mathbb{P}\Big[\bigcup_{j=1}^N\{|Y'_j - Y_j| \ge N^{-2}\}\Big] \\ &\quad+ \mathbb{E}\Big(|g(\mathbf{W}_{\mathcal{Z}}) - g(\mathbf{W}'_{\mathcal{Z}})|\Big|\bigcap_{j=1}^N\{|Y'_j - Y_j| < N^{-2}\}\Big)\mathbb{P}\Big[\bigcap_{j=1}^N\{|Y'_j - Y_j| < N^{-2}\}\Big].\end{aligned} \tag{4.16}$$

The first summand can easily be bounded, using $m \ge (2 + 3/\beta)\log N/\log c$ and Theorem 3.5:

$$\begin{aligned}&\mathbb{E}\Big(|g(\mathbf{W}_{\mathcal{Z}}) - g(\mathbf{W}'_{\mathcal{Z}})|\Big|\bigcup_{j=1}^N\{|Y'_j - Y_j| \ge N^{-2}\}\Big)\mathbb{P}\Big[\bigcup_{j=1}^N\{|Y'_j - Y_j| \ge N^{-2}\}\Big] \\ &\quad\le \mathbb{P}\Big[\bigcup_{j=m}^{N+m}\{|Y'_j - Y_j| \ge N^{-2}\}\Big] \le K_{3.4}N^{-2}.\end{aligned} \tag{4.17}$$

For the second summand, we have

$$\begin{aligned}&\mathbb{E}\Big(|g(\mathbf{W}_{\mathcal{Z}}) - g(\mathbf{W}'_{\mathcal{Z}})|\Big|\bigcap_{j=1}^N\{|Y'_j - Y_j| < N^{-2}\}\Big) \\ &\quad= \mathbb{E}\Big(\Big|g\Big(\Sigma_N^{-\frac{1}{2}}\sum_{i=1}^N \tilde{\mathbf{u}}^{(N-i+1)}(\mathbf{Y}'_{i-\mathbf{J}}) + \Sigma_N^{-\frac{1}{2}}\sum_{i=1}^N\big(\mathbf{u}^{(N-i+1)}(\mathbf{Y}_{i-\mathbf{J}}) - \tilde{\mathbf{u}}^{(N-i+1)}(\mathbf{Y}'_{i-\mathbf{J}})\big)\Big) \\ &\qquad\quad - g\Big(\Sigma_N^{-\frac{1}{2}}\sum_{i=1}^N \tilde{\mathbf{u}}^{(N-i+1)}(\mathbf{Y}'_{i-\mathbf{J}})\Big)\Big|\,\Big|\,\bigcap_{j=1}^N\{|Y'_j - Y_j| < N^{-2}\}\Big).\end{aligned}$$

On the set $\bigcap_{j=1}^N\{|Y_j'-Y_j|<N^{-2}\}$, we have, from the Lipschitz property of $\mathbf{u}^{(j)}$, that

$$\begin{aligned}\left|\mathbf{u}^{(N-i+1)}(\mathbf{Y}_{i-\mathbf{J}})-\tilde{\mathbf{u}}^{(N-i+1)}(\mathbf{Y}'_{i-\mathbf{J}})\right| &\le \sum_{k=1}^{s}\tau|Y'_{i-J_k}-Y_{i-J_k}|+|\mathbb{E}\mathbf{u}^{(N-i+1)}(\mathbf{Y}'_{1+J_s-\mathbf{J}})|\\ &\le s\tau N^{-2}+|\mathbb{E}\mathbf{u}^{(N-i+1)}(\mathbf{Y}'_{1+J_s-\mathbf{J}})|.\end{aligned}$$

But now $|\mathbb{E}\mathbf{u}^{(N-i+1)}(\mathbf{Y}'_{1+J_s-\mathbf{J}})|\le\mathbb{E}|\mathbf{u}^{(N-i+1)}(\mathbf{Y}'_{1+J_s-\mathbf{J}})-\mathbf{u}^{(N-i+1)}(\mathbf{Y}_{1+J_s-\mathbf{J}})|$, and splitting the expectation again as in (4.16) gives $|\mathbb{E}\mathbf{u}^{(N-i+1)}(\mathbf{Y}'_{1+J_s-\mathbf{J}})|\le s\tau(1+K_{3.4})N^{-2}$; hence

$$\left|\mathbf{u}^{(N-i+1)}(\mathbf{Y}_{i-\mathbf{J}})-\tilde{\mathbf{u}}^{(N-i+1)}(\mathbf{Y}'_{i-\mathbf{J}})\right|\le s\tau(2+K_{3.4})N^{-2} \tag{4.18}$$

and

$$\left|\Sigma_N^{-\frac12}\Big[\sum_{i=1}^N\big(\mathbf{u}^{(N-i+1)}(\mathbf{Y}_{i-\mathbf{J}})-\tilde{\mathbf{u}}^{(N-i+1)}(\mathbf{Y}'_{i-\mathbf{J}})\big)\Big]\right|\le(2+K_{3.4})\varepsilon_N N^{-1}.$$

This implies that, with $\delta_N=(2+K_{3.4})\varepsilon_N N^{-1}$,

$$\begin{aligned}&\mathbb{E}\Big(\Big|g(\mathbf{W}_{\mathcal{Z}})-g(\mathbf{W}'_{\mathcal{Z}})\Big|\Big|\bigcap_{j=1}^N\{|Y_j'-Y_j|<N^{-2}\}\Big)\\ &\qquad\le\mathbb{E}\Big(g^+_{\delta_N}(\mathbf{W}'_{\mathcal{Z}})-g^-_{\delta_N}(\mathbf{W}'_{\mathcal{Z}})\Big|\bigcap_{j=1}^N\{|Y_j'-Y_j|<N^{-2}\}\Big),\end{aligned}$$

where, following Götze (1991) and Rinott and Rotar (1996), for any $\delta>0$, we define

$$g^+_\delta(x)=\sup\{g(x+y):|y|\le\delta\};\quad g^-_\delta(x)=\inf\{g(x+y):|y|\le\delta\},$$

observing also that, for all $\delta>0$, the functions $g^+_\delta(x)$ and $g^-_\delta(x)$ are in $\mathcal{G}$, and that

$$\sup\{(\mathbb{E}g^+_\delta(\mathcal{N})-\mathbb{E}g^-_\delta(\mathcal{N})):g\in\mathcal{G}\}\le a\delta \tag{4.19}$$

for a universal constant $a=a(d)>1$. Thus we have

$$\begin{aligned}&\mathbb{E}\Big(|g(\mathbf{W}_{\mathcal{Z}})-g(\mathbf{W}'_{\mathcal{Z}})||\bigcap_{j=1}^N\{|Y_j'-Y_j|<N^{-2}\}\Big)\mathbb{P}\Big[\bigcap_{j=1}^N\{|Y_j'-Y_j|<N^{-2}\}\Big]\\ &\quad\le\mathbb{E}\Big(g^+_{\delta_N}(\mathbf{W}'_{\mathcal{Z}})-g^-_{\delta_N}(\mathbf{W}'_{\mathcal{Z}})\Big)\\ &\quad=\mathbb{E}\big(g^+_{\delta_N}(\mathbf{W}'_{\mathcal{Z}})-g^+_{\delta_N}(\mathcal{N})\big)-\mathbb{E}\big(g^-_{\delta_N}(\mathbf{W}'_{\mathcal{Z}})-g^-_{\delta_N}(\mathcal{N})\big)+\ \mathbb{E}\big(g^+_{\delta_N}(\mathcal{N})-g^-_{\delta_N}(\mathcal{N})\big)\\ &\quad\le 2\ \sup\{|\mathbb{E}g(\mathbf{W}'_{\mathcal{Z}})-\mathbb{E}g(\mathcal{N})|:g\in\mathcal{G}\}+a\delta_N\\ &\quad=2\sup\{|\mathbb{E}g(\mathbf{W}'_{\mathcal{Z}})-\mathbb{E}g(\mathcal{N})|:g\in\mathcal{G}\}+a(2+K_{3.4})\varepsilon_N N^{-1},\end{aligned}$$

which, combined with (4.17), gives

$$|\mathbb{E}g(\mathbf{W}_{\mathcal{Z}})-\mathbb{E}g(\mathbf{W}'_{\mathcal{Z}})|\le K_{3.4}N^{-2}+a(2+K_{3.4})\varepsilon_N N^{-1}+3\sup\{|\mathbb{E}g(\mathbf{W}'_{\mathcal{Z}})-\mathbb{E}g(\mathcal{N})|. \tag{4.20}$$

Thus it remains to bound $\sup\{|\mathbb{E}(g(\mathbf{W}'_{\mathcal{Z}})-\mathbb{E}g(\mathcal{N})|:g\in\mathcal{G}\}$, for which we apply Theorem 2.1 in Rinott and Rotar (1996). They consider $\mathbf{W}=\sum_{j=1}^N\mathbf{Z}_j$, where $\mathbf{Z}_j$ are bounded random vectors taking values in R^d, that is, $|\mathbf{Z}_j|\le B$, $1\le j\le N$ for some constant B. (Rinott and Rotar have $\mathbf{X}$ instead of $\mathbf{Z}$, $\mathcal{H}$ for $\mathcal{G}$, and n for N.)

Theorem (Rinott and Rotar). *For each $j = 1, \dots, N$ assume that we have two representations of* $\mathbf{W}$, $\mathbf{W} = \mathbf{U}_j + \mathbf{V}_j$ *and* $\mathbf{W} = \mathbf{R}_j + \mathbf{T}_j$, *such that* $|\mathbf{U}_j| \le A_1$, *and* $|\mathbf{R}_j| \le A_2$ *for constants satisfying* $A_1 \le A_2$. *Define*

$$\chi_1 = \sum_{j=1}^N \mathbb{E}|\mathbb{E}(\mathbf{Z}_j \mid \mathbf{V}_j)|,\ \chi_2 = \sum_{j=1}^N \mathbb{E}|\mathbb{E}(\mathbf{Z}_j\mathbf{U}_j^T) - \mathbb{E}(\mathbf{X}_j\mathbf{U}_j^T \mid \mathbf{T}_j)|,\ \chi_3 = |\mathbf{I} - \sum_{j=1}^N \mathbb{E}(\mathbf{Z}_j\mathbf{U}_j^T)|,$$

where $\mathbf{I}$ *denotes the identity matrix. Then for any $d \ge 1$, there exists a constant c depending only on the dimension d such that*

$$\begin{aligned}\sup\{|\mathbb{E}g(\mathbf{W}) - \mathbb{E}g(\mathcal{N})| \,:\, g \in \mathcal{G}\} \le c\{&aA_2 + NaA_1A_2B(|\log A_2B| + \log N) \\ &+ \chi_1 + (|\log A_1B| + \log N)(\chi_2 + \chi_3)\},\end{aligned} \tag{4.21}$$

with a as in (4.19).

To apply this theorem, for each $i = 1, \dots, N$, put

$$\mathcal{Z}_{i-\mathbf{J}} = \Sigma_N^{-\frac{1}{2}} \tilde{\mathbf{u}}^{(N-i+1)}(Y'_{i-\mathbf{J}}); \qquad \mathbf{W}'_{\mathcal{Z}} = \sum_{i=1}^N \mathcal{Z}_{i-\mathbf{J}}.$$

Clearly,

$$|\mathcal{Z}_{i-\mathbf{J}}| \le s\tau|\Sigma_N^{-\frac{1}{2}}| = \varepsilon_N, \tag{4.22}$$

showing that we can take $B = \varepsilon_N$ in (4.21). Moreover, to find the two representations of $\mathbf{W}'_{\mathcal{Z}}$ needed for (4.21), we define neighbourhoods of dependence

$$B_j = \{i = 1, \dots, N : i \le j + J_s + m \ \text{ and } \ i + J_s \ge j - m\}, \quad N_j = \bigcup_{i \in B_j} B_i$$

for $1 \le j \le n$. Because of the m-dependence, if $i \notin B_j$, then $\mathcal{Z}_{j-\mathbf{J}}$ and $\mathcal{Z}_{i-\mathbf{J}}$ are independent. If $i \notin N_j$, and if $k \in B_j$, then $\mathcal{Z}_{i-\mathbf{J}}$ and $\mathcal{Z}_{k-\mathbf{J}}$ are independent. Let

$$\mathbf{U}_j = \sum_{i \in B_j} \mathcal{Z}_{i-\mathbf{J}},\ \mathbf{V}_j = \sum_{i \notin B_j} \mathcal{Z}_{i-\mathbf{J}},\ \mathbf{R}_j = \sum_{i \in N_j} \mathcal{Z}_{i-\mathbf{J}},\ \mathbf{T}_j = \sum_{i \notin N_j} \mathcal{Z}_{i-\mathbf{J}}.$$

Then $\mathbf{W}'_{\mathcal{Z}} = \mathbf{U}_j + \mathbf{V}_j = \mathbf{R}_j + \mathbf{T}_j$. To bound $|\mathbf{U_j}|$ and $|\mathbf{R_j}|$, note that, for all $j = 1, \dots, N$,

$$|B_j| \le 2(J_s + m) + 1 \le 2b$$

and

$$|N_j| \le 4(J_s + m) + 1 \le 4b,$$

where b is as in (4.14), so that therefore

$$|\mathbf{U_j}| \le s\tau\,|B_j|\,|\Sigma_N^{-1/2}| \le 2b\varepsilon_N \quad \text{and} \quad |\mathbf{R_j}| \le s\tau\,|N_j|\,|\Sigma_N^{-1/2}| \le 4b\varepsilon_N,$$

so that we can take $2A_1 = A_2 = 4b\varepsilon_N$ in (4.21). It thus only remains to bound the χ_i. From the choice of neighbourhoods and m-dependence, the first two characteristics in the theorem vanish: $\chi_1 = 0$ and $\chi_2 = 0$. For the third characteristic, we get

$$
\begin{aligned}
\chi_3 &= |\mathbf{I} - \sum_{i=1}^{N} \mathbb{E}(\mathcal{Z}_{i-\mathbf{J}} \mathbf{U}_i^T)| = |\mathbf{I} - \mathbb{E}(\mathbf{W}'_{\mathcal{Z}}(\mathbf{W}'_{\mathcal{Z}})^T)| \\
&= |\mathbf{I} - \Sigma_N^{-\frac{1}{2}} \Sigma'_N \Sigma_N^{-\frac{1}{2}}| \leq |\Sigma_N^{-\frac{1}{2}}|^2 |\Sigma_N - \Sigma'_N|.
\end{aligned}
$$

To bound this quantity, observe that

$$
\begin{aligned}
&|\Sigma_N - \Sigma'_N| \\
&\quad = \Big| \sum_{i,j=1}^{N} \mathbb{E}\big(\mathbf{u}^{(N-i+1)}(\mathbf{Y}_{i-\mathbf{J}})\mathbf{u}^{(N-j+1)}(\mathbf{Y}_{j-\mathbf{J}})^T - \tilde{\mathbf{u}}^{(N-i+1)}(\mathbf{Y}'_{i-\mathbf{J}})\tilde{\mathbf{u}}^{(N-j+1)}(\mathbf{Y}'_{j-\mathbf{J}})^T\big)\Big| \\
&\quad \leq \Big| \sum_{j=1}^{N} \sum_{i \in B_j} \mathbb{E}\big(\mathbf{u}^{(N-i+1)}(\mathbf{Y}_{i-\mathbf{J}})\mathbf{u}^{(N-j+1)}(\mathbf{Y}_{j-\mathbf{J}})^T - \tilde{\mathbf{u}}^{(N-i+1)}(\mathbf{Y}'_{i-\mathbf{J}})\tilde{\mathbf{u}}^{(N-j+1)}(\mathbf{Y}'_{j-\mathbf{J}})^T\big)\Big| \\
&\qquad + \Big| \sum_{j=1}^{N} \sum_{i \notin B_j} \mathbb{E}\big(\mathbf{u}^{(N-i+1)}(\mathbf{Y}_{i-\mathbf{J}})\mathbf{u}^{(N-j+1)}(\mathbf{Y}_{j-\mathbf{J}})^T\big)\Big|.
\end{aligned}
\tag{4.23}
$$

Consider the first term in (4.23), for which we have

$$
\begin{aligned}
&\Big| \sum_{j=1}^{N} \sum_{i \in B_j} \mathbb{E}\big(\mathbf{u}^{(N-i+1)}(\mathbf{Y}_{i-\mathbf{J}})\mathbf{u}^{(N-j+1)}(\mathbf{Y}_{j-\mathbf{J}})^T - \tilde{\mathbf{u}}^{(N-i+1)}(\mathbf{Y}'_{i-\mathbf{J}})\tilde{\mathbf{u}}^{(N-j+1)}(\mathbf{Y}'_{j-\mathbf{J}})^T\big)\Big| \\
&\quad \leq \sum_{j=1}^{N} \sum_{i \in B_j} \mathbb{E}\Big\{ \big|\mathbf{u}^{(N-i+1)}(\mathbf{Y}_{i-\mathbf{J}})\mathbf{u}^{(N-j+1)}(\mathbf{Y}_{j-\mathbf{J}})^T - \tilde{\mathbf{u}}^{(N-i+1)}(\mathbf{Y}'_{i-\mathbf{J}})\tilde{\mathbf{u}}^{(N-j+1)}(\mathbf{Y}'_{j-\mathbf{J}})^T\big| \\
&\qquad\qquad \Big| \bigcap_{j=1}^{N} \{|Y_j - Y'_j| < N^{-2}\} \Big\} \, \mathbb{P}\Big(\bigcap_{j=1}^{N} \{|Y_j - Y'_j| < N^{-2}\} \Big) \\
&\qquad + \sum_{j=1}^{N} \sum_{i \in B_j} \mathbb{E}\Big\{ \big|\mathbf{u}^{(N-i+1)}(\mathbf{Y}_{i-\mathbf{J}})\mathbf{u}^{(N-j+1)}(\mathbf{Y}_{j-\mathbf{J}})^T - \tilde{\mathbf{u}}^{(N-i+1)}(\mathbf{Y}'_{i-\mathbf{J}})\tilde{\mathbf{u}}^{(N-j+1)}(\mathbf{Y}'_{j-\mathbf{J}})^T\big| \\
&\qquad\qquad \Big| \bigcup_{j=1}^{N} \{|Y_j - Y'_j| \geq N^{-2}\} \Big\} \, \mathbb{P}\Big(\bigcup_{j=1}^{N} \{|Y_j - Y'_j| \geq N^{-2}\} \Big).
\end{aligned}
$$

In the first sum, we bound the probability by one, giving

$$\sum_{j=1}^{N}\sum_{i\in B_j}\mathbb{E}\Big\{\big|\mathbf{u}^{(N-i+1)}(\mathbf{Y}_{i-\mathbf{J}})\mathbf{u}^{(N-j+1)}(\mathbf{Y}_{j-\mathbf{J}})^T-\tilde{\mathbf{u}}^{(N-i+1)}(\mathbf{Y}'_{i-\mathbf{J}})\tilde{\mathbf{u}}^{(N-j+1)}(\mathbf{Y}'_{j-\mathbf{J}})^T\big|$$

$$\Big|\bigcap_{j=1}^{N}\{|Y_j-Y'_j|<N^{-2}\}\Big\}\ \mathbb{P}\Big(\bigcap_{j=1}^{N}\{|Y_j-Y'_j|<N^{-2}\}\Big)$$

$$\begin{aligned}
\leq&\sum_{j=1}^{N}\sum_{i\in B_j}\mathbb{E}\Big\{\big|\mathbf{u}^{(N-i+1)}(\mathbf{Y}_{i-\mathbf{J}})\big[\mathbf{u}^{(N-j+1)}(\mathbf{Y}_{j-\mathbf{J}})^T-\tilde{\mathbf{u}}^{(N-j+1)}(\mathbf{Y}'_{j-\mathbf{J}})^T\big]\big|\\
&+\big|\big[\mathbf{u}^{(N-i+1)}(\mathbf{Y}_{i-\mathbf{J}})-\tilde{\mathbf{u}}^{(N-i+1)}(\mathbf{Y}'_{i-\mathbf{J}})\big]\tilde{\mathbf{u}}^{(N-j+1)}(\mathbf{Y}'_{j-\mathbf{J}})^T\big|\ \Big|\bigcap_{j=1}^{N}\{|Y_j-Y'_j|<N^{-2}\}\Big\}\\
\leq&4(2+K_{3.4})(s\tau)^2bN^{-1},
\end{aligned}$$

using (4.18); b as before, in (4.14). For the second sum, we use Theorem 3.5 and the fact that $m\geq(2+3/\beta)\log N/\log c$ to obtain

$$\sum_{j=1}^{N}\sum_{i\in B_j}\mathbb{E}\Big\{\big|\mathbf{u}^{(N-i+1)}(\mathbf{Y}_{i-\mathbf{J}})\mathbf{u}^{(N-j+1)}(\mathbf{Y}_{j-\mathbf{J}})^T-\tilde{\mathbf{u}}^{(N-i+1)}(\mathbf{Y}'_{i-\mathbf{J}})\tilde{\mathbf{u}}^{(N-j+1)}(\mathbf{Y}'_{j-\mathbf{J}})^T$$

$$\big|\Big|\bigcup_{j=1}^{N}\{|Y_j-Y'_j|\geq N^{-2}\}\Big\}\ \mathbb{P}\Big(\bigcup_{j=1}^{N}\{|Y_j-Y'_j|\geq N^{-2}\}\Big)$$

$$\leq 4Nb(s\tau)^2K_{3.4}N^{-2}=4bN^{-1}K_{3.4}(s\tau)^2.$$

For the second term in (4.23), we need to bound the sum over all indices i that are not in B_j. Using Corollary 4.3 with $g_0=f$ and since, by choice of m, $N\alpha^{m+1}\leq N^{-1}$, we obtain

$$\begin{aligned}
\Big|\sum_{j=1}^{N}\sum_{i\notin B_j}\mathbb{E}\big(\mathbf{u}^{(N-i+1)}(\mathbf{Y}_{i-\mathbf{J}})\mathbf{u}^{(N-j+1)}(\mathbf{Y}_{j-\mathbf{J}})^T\big)\Big|\leq&K_{4.3}\sum_{j=1}^{N}\sum_{i\notin B_j}\alpha^{|i-j|}\|\mathbf{u}\|(\|\mathbf{u}\|+\tau K_{1.4})\\
\leq&K_{4.3}N\frac{\alpha^{m+1}}{1-\alpha}s\tau(s\tau+\tau K_{1.4})\\
\leq&\frac{K_{4.3}}{1-\alpha}N^{-1}(s\tau)^2(1+s^{-1}K_{1.4}).
\end{aligned}$$

Collecting these estimates of (4.23), we get

$$\begin{aligned}
|\Sigma_N-\Sigma'_N|&\leq bN^{-1}(s\tau)^2\left\{4(2+K_{3.4})+4K_{3.4}+\frac{K_{4.3}}{1-\alpha}\left(1+s^{-1}K_{1.4}\right)\right\}\\
&=bN^{-1}(s\tau)^2\psi,
\end{aligned}\tag{4.24}$$

and thus $\chi_3 \leq bN^{-1}\varepsilon_N^2\psi$. Substituting our estimates of A_1, A_2, B and χ_3 into (4.21), and using (4.20), the theorem follows.

The freedom to choose the $\mathbf{u}^{(j)}$ to be different for each j enables one for instance to consider the joint distributions of the partial sum process $(N^{-1/2}\sum_{i=1}^{[Nt]} X_i,\ 0 \leq t \leq 1)$ at a finite number of different time points. Another natural multivariate central limit theorem involves the joint distribution of $(M_1(\mathbf{X}), \ldots, M_d(\mathbf{X}))$, where $M_k(\mathbf{X}) = \sum_{i=1}^N I[X_i \in L_k]$, for a set of d intervals $L_1, \ldots, L_d \in I$. The Lipschitz assumption on the $\mathbf{u}^{(j)}$ in Theorem 4.5 does not directly allow this example. However, by choosing m to be a possibly larger multiple of $\log N$, it is easy to construct $\mathbf{M}(\mathbf{X})$ and $\mathbf{M}(\mathbf{Y}')$ so as to be identical, except on an event of negligible probability, and m–dependent theory can be used once again for $\mathbf{M}(\mathbf{Y}')$: the details are omitted.

4.3. A functional central limit theorem

For a functional central limit theorem, we use ideas similar to those above. We approximate the process $(Y_i)_{i\in N}$ by the m-dependent process $(Y_i')_{i\in N}$ and apply a functional central limit theorem with a bound on the rate of convergence for the m-dependent process. A general functional central limit theorem for strongly mixing processes has been obtained by Doukhan, Massart and Rio (1994), but it does not give rates for the convergence. Instead, we use results by Barbour (1990); as they are derived using Stein's method, they give bounds on the rate of convergence as a function of m.

Let $\sigma_N^2 = \mathrm{Var}\,(\sum_{i=1}^N X_i)$ and $\mathbb{E}u(X_i) = 0$ be as above, u being a Lipschitz continuous function from $[0,1]$ to $\mathbf{R}$ with Lipschitz constant τ; $\|u\| = \max_{y\in[0,1]}|u(y)| \leq \tau$. For $t \in [0,1]$, put

$$Z_N(t) = \sigma_N^{-1}\sum_{i=1}^N u(X_i)J_{\frac{i}{N}}(t),$$

where

$$J_\alpha(t) = \mathbf{1}(t \geq \alpha).$$

Then Z_N takes on values in $D = D[0,1]$, the space of right-continuous functions with left-hand limits on $[0,1]$, equipped with the Skorohod topology. Once again, we can write

$$Z_N = \sigma_N^{-1}\sum_{i=1}^N u(Y_i)J_{\frac{N-i+1}{N}},$$

where $(Y_1, Y_2, \ldots, Y_N)$ is the time reversal of $(X_1, X_2, \ldots, X_N)$. As before, fix $m = \max\{(2+3/\beta)/\log c, 2/\log(1/\alpha)\}\log N$, and assume the process to start at time $-m$.

Let $(Y'_1, Y'_2, \ldots, Y'_N)$ be the m-dependent approximation of $(Y_1, Y_2, \ldots, Y_N)$ constructed above. Define $\sigma'^2_N = \mathrm{Var}\,(\sum_{i=1}^N u(Y'_i))$, and put

$$\tilde{u}(Y'_i) = u(Y'_i) - \mathbb{E}u(Y'_1); \quad s'_{i,j} = \mathbb{E}\tilde{u}(Y'_i)\tilde{u}(Y'_j); \quad s'_{i+} = \sum_{j=1}^N s'_{i,j};$$

$$S'_i = \sum_{j=1}^i s'_{j+}; \quad \kappa_N = \max_{1\le i\le N} \frac{1}{s'_{i+}} \sum_{j=1}^N |s'_{i,j}| \text{ and } t_{i,N} = \frac{S'_i}{S'_N}$$

Note that $\sigma'^2_N = S'_N$. Because of the stationarity of the (Y_i) process and the exponential decay of correlations, the s'_{i+} are almost constant in i, and nearly equal to $S' = \sum_{j=-\infty}^{\infty} \mathbb{E}u(Y_i)u(Y_j)$. If $S' > 0$, then the $t_{i,N}$ are very close to i/N for almost all values of i, and κ_N is bounded in N.

Define

$$Z'_N = \sigma'^{-1}_N \sum_{i=1}^N \tilde{u}(Y'_i) J_{\frac{N-i+1}{N}}.$$

Let Z be a standard Brownian motion on D. As for our multivariate central limit theorem, we prove weak convergence of Z_N to Z by using a class of test functions. Here, following Barbour (1990), we introduce

$$M = \{F : D \to R \text{ twice Frechet differentiable, } \|D^2 f(\omega + h) - D^2 f(\omega)\| \le k_f \|h\| \\ \text{for some constant } k_f, \text{ uniformly for all } \omega, h \in D\},$$

where $D^k f$ denotes the kth derivative of f; $\|D^k f(\omega)\| = \sup_{\{h:\|h\|=1\}} |D^k f(\omega)[h^{(k)}]|$, $h^{(k)}$ denotes the k-tuple $(h, h, \ldots, h)$, and the underlying norm $\|\cdot\|$ is the supremum norm on D. Equip M with the norm

$$\|f\|_M = \sup_{\omega\in D} \frac{|f(\omega)|}{1+\|\omega\|^3} + \sup_{\omega\in D} \frac{\|Df(\omega)\|}{1+\|\omega\|^2} + \sup_{\omega\in D} \frac{\|D^2 f(\omega)\|}{1+\|\omega\|} \\ + \sup_{\omega,h\in D} \frac{\|D^2 f(\omega+h) - D^2 f(\omega)\|}{\|h\|},$$

and then define the space

$$M_0 = \{f \in M : \|f\|_{M_0} := \|f\|_M + \sup_{\omega\in D} |f(\omega)| + \sup_{\omega\in D} \|Df(\omega)\| \\ + \sup_{\omega\in D} \|D^2 f(\omega)\| < \infty\}.$$

As in the previous section, we put

$$\psi = 8(1 + K_{3.4}) + K_{4.3}(1 + K_{1.4})/(1 - \alpha).$$

Theorem 4.6. *Suppose that $\sigma_N \geq \tau$, and that $m = \max\{(2+\frac{3}{\beta})/\log c, 2/\log\frac{1}{\alpha}\}\log N$ satisfies $m \leq (3N/4\psi) - 1$. Then, for all $g \in M_0$ satisfying the smoothness condition*

$$\sup{}_{\omega\in D}|D^2 g(\omega)[J_r, J_s - J_t]| \leq C_1\|g\|_{M_0}|t-s|^{\frac{1}{2}}, \tag{4.25}$$

for any fixed $C_1 > 0$, it follows that

$$|\mathbb{E}g(Z_N) - \mathbb{E}g(Z)| \leq \|g\|_{M_0}\rho_N,$$

where Z is standard Brownian motion and ρ_N is defined by

$$\begin{aligned}\rho_N = C\Big\{&8\varepsilon_N^3(m^2+1)N + \Big[2\varepsilon_N(m+1) + C_1\Big(\frac{m}{N}\Big)^{\frac{1}{2}}\Big]\kappa_N \\ &+\ C_1\max\Big[\max{}_i\Big(t_{i,N} - \frac{i}{N}\Big); \max{}_i\Big(\frac{i+1}{N} - t_{(i+1),N}\Big)\Big]\Big\} \\ &+ 4(1+K_{3.4})\varepsilon_N N^{-1} + \frac{4}{3}(m+1)\varepsilon_N^2\psi,\end{aligned}$$

with $\varepsilon_N = \tau\sigma_N^{-1}$ and C a universal constant. Furthermore, if $\frac{\rho_N}{(\log N)^2} \to 0$, it follows that $Z_N \Rightarrow Z$ in the Skorohod topology.

Proof. Much as in the proof of Theorem 4.4, we write

$$|\mathbb{E}g(Z_N) - \mathbb{E}g(Z)| \leq |\mathbb{E}g(Z_N) - \mathbb{E}g(Z_N')| +\ |\mathbb{E}g(Z_N') - \mathbb{E}g(Z)|$$

for each $g \in M_0$.

For the first summand, as Z_N and Z_N' are defined on the same probability space, we may use a Taylor expansion; for all $g \in M_0$,

$$\begin{aligned}&|\mathbb{E}g(Z_N) - \mathbb{E}g(Z_N')| \\ &\qquad\leq \|g\|_{M_0}\mathbb{E}\|Z_N - Z_N'\| \\ &\qquad= \|g\|_{M_0}\mathbb{E}\Big\|(\sigma_N^{-1} - \sigma_N'^{-1})\sum_{i=1}^N u(Y_i)J_{\frac{N-i+1}{N}} + \sigma_N'^{-1}\sum_{i=1}^N(u(Y_i) - \tilde{u}(Y_i'))J_{\frac{N-i+1}{N}}\Big\| \\ &\qquad\leq \|g\|_{M_0}\Big[N\tau|\sigma_N^{-1} - \sigma_N'^{-1}| + \sigma_N'^{-1}\mathbb{E}\Big\|\sum_{i=1}^N(u(Y_i) - \tilde{u}(Y_i'))J_{\frac{N-i+1}{N}}\Big\|\Big].\end{aligned}$$

Now (4.24) is still true for $s = 1$ and $J_s = 0$, and hence, because also $x - y = \frac{x^2-y^2}{x+y}$, we find that

$$\begin{aligned}|\sigma_N^{-1} - \sigma_N'^{-1}| &= \frac{1}{\sigma_N\sigma_N'(\sigma_N + \sigma_N')}|\sigma_N^2 - (\sigma_N')^2| \\ &\leq \frac{1}{\sigma_N\sigma_N'(\sigma_N + \sigma_N')}(m+1)N^{-1}\tau^2\psi \\ &= \frac{\sigma_N}{\sigma_N'(\sigma_N + \sigma_N')}\varepsilon_N^2(m+1)N^{-1}\psi.\end{aligned}$$

To bound the denominator, observe that, from (4.24) and the definition of m,

$$|\sigma_N^2 - (\sigma_N')^2| < (m+1)N^{-1}\tau^2\psi \leq \frac{3}{4}\tau^2.$$

Thus, because $\sigma_N \geq \tau$, it follows that $(\sigma_N')^2 \geq \frac{1}{4}\tau^2$, so that $(\sigma_N'/\tau) \geq \frac{1}{2}$, and thus also

$$\frac{\sigma_N^2}{(\sigma_N')^2} \leq 1 + \frac{|\sigma_N^2 - (\sigma_N')^2|}{(\sigma_N')^2} \leq 1 + \frac{3\tau^2}{4(\sigma_N')^2} \leq 4 \quad \text{and} \quad \frac{\sigma_N}{\sigma_N'} \leq 2; \tag{4.26}$$

this in turn now gives

$$|\sigma_N^{-1} - \sigma_N'^{-1}| \leq \frac{4}{3}\varepsilon_N^2(m+1)N^{-1}\psi\tau^{-1}.$$

Moreover, conditioning on the event $\bigcup_{j=1}^N \{|Y_j' - Y_j| \geq N^{-2}\}$ and on its complement yields

$$\begin{aligned}\mathbb{E}\Big\|\sum_{i=1}^N (u(Y_i) - \tilde{u}(Y_i'))J_{\frac{N-i+1}{N}}\Big\| \leq& \mathbb{E}\Big(\Big\|\sum_{i=1}^N (u(Y_i) - \tilde{u}(Y_i'))J_{\frac{N-i+1}{N}}\Big\|\Big| \bigcap_{j=1}^N \{|Y_j' - Y_j| < N^{-2}\}\Big) \\ &+ N\tau\mathbb{P}\Big[\bigcup_{j=1}^N \{|Y_j' - Y_j| \geq N^{-2}\}\Big].\end{aligned}$$

Hence, from Theorem 3.5, since $m \geq (2 + (3/\beta))\log N/\log c$, and recalling (4.18), it follows that

$$\mathbb{E}\Big\|\sum_{i=1}^N (u(Y_i) - \tilde{u}(Y_i'))J_{\frac{N-i+1}{N}}\Big\| \leq 2\tau(1 + K_{3.4})N^{-1}. \tag{4.27}$$

Summarizing the above and using (4.26), we obtain

$$|\mathbb{E}g(Z_N) - \mathbb{E}g(Z_N')| \leq \|g\|_{M_0}\left\{\frac{4}{3}(m+1)\varepsilon_N^2\psi + 4\varepsilon_N(1 + K_{3.4})N^{-1}\right\}. \tag{4.28}$$

It remains to bound $|\mathbb{E}g(Z_N') - \mathbb{E}g(Z)|$. To this purpose, we employ Theorem 4 in Barbour (1990) (note that our notation here is slightly different). Note that $\mathbb{E}|u(Y_i')|^3 \leq \tau^3$; hence, from Theorem 4 in Barbour (1990) we find that, for any $g \in M$ with

$$\sup{}_{\omega\in D}|D^2g(\omega)[J_r, J_s - J_t]| \leq C_1\|g\|_M|t-s|^{\frac{1}{2}}, \tag{4.29}$$

we have

$$\begin{aligned}&|\mathbb{E}g(Z_N') - \mathbb{E}g(Z)| \\ &\quad\leq C\|g\|_M\Big\{\tau^3(m^2+1)N(\sigma_N')^{-3} + \Big[\tau(m+1)(\sigma_N')^{-1} + C_1\Big(\frac{m}{N}\Big)^{\frac{1}{2}}\Big]\kappa_N \\ &\qquad + C_1\max\Big[\max{}_i\Big(t_{i,N} - \frac{i}{N}\Big); \max{}_i\Big(\frac{i+1}{N} - t_{i,N}\Big)\Big]\Big\},\end{aligned}$$

where C is a universal constant. Combining this with (4.28) and using $\|g\|_M \leq \|g\|_{M_0}$ yields the bound ρ_N.

To derive convergence of Z_N to Z in the Skorohod topology we employ the Corollary of Theorem 2 in Barbour (1990); note that Z_N has intervals of constancy of length $\frac{1}{N}$ and that, if $g \in M_0$ satisfies the smoothness condition (4.25), then g satisfies (4.29) because $\|g\|_M \leq \|g\|_{M_0}$. This completes the proof.

4.4 A Gaussian approximation for the empirical process

Gaussian approximations for the empirical process of stationary mixing sequences with rates have been obtained by Philipp (1982) and by Arcones and Yu (1994). Philipp (1982) obtains almost sure invariance principles with logarithmic rates. In our context, though, logarithmic rates may not suffice. Hence we will employ results from Reinert (1997) for better rates.

Let μ be the law of X_i, and let

$$\xi_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\delta_{X_i} - \mu)$$

be the empirical process of $X_i, i = 1, \ldots, N$. Here, δ_a denotes the Dirac measure at a point $a \in \mathbf{R}$. Then ξ_N is a random measure with values in the space $M_{\mathrm{f}}([0,1])$ of finite signed measures on $[0,1]$; for bounded $\phi : [0,1] \to \mathbb{R}$, we write

$$\langle \xi_N, \phi \rangle = \int \phi \, d\xi_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\phi(X_i) - \int \phi \, d\mu). \tag{4.30}$$

Equivalently, we can write

$$\xi_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\delta_{Y_i} - \mu),$$

where, as usual, $Y_1, \ldots, Y_N$ is the time reversal of $X_1, \ldots, X_N$. Put

$$\xi'_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\delta_{Y'_i} - \mu),$$

where, taking the process to have started before time $-m$, $Y'_1, \ldots, Y'_N$ is the m-dependent approximation of $Y_1, \ldots, Y_N$. Define

$$\mathcal{FL}([0,1]) = \{f : [0,1] \to \mathbf{R}; |f(x) - f(y)| \leq c_f |x - y|, x, y \in K, \\ \text{for some constant } c_f \text{ depending only on } f\},$$

and let $\mathcal{FL}^+([0,1])$ be the space of nonnegative functions in $\mathcal{FL}([0,1])$. Then $\mathcal{FL}^+([0,1])$ is convergence-determining for vague convergence in $M_{\mathrm{f}}([0,1])$ (see Rachev (1991), p. 209, Corollary 10.2.1). Let ζ_N be a random measure with values almost surely in $M_{\mathrm{f}}([0,1])$ such that, for all $l \in \mathbf{N}, \phi_1, \ldots, \phi_l \in \mathcal{FL}^+([0,1])$,

$$\mathcal{L}(\langle \zeta_N, \phi_1 \rangle, \ldots, \langle \zeta_N, \phi_l \rangle) = \mathcal{MVN}_l(0, B_N(\phi_1, \ldots, \phi_l)),$$

where

$$\begin{aligned} B_N(\phi_1, \ldots, \phi_l) &= \big(b_N(\phi_j, \phi_k)\big)_{j,k=1,\ldots,m} \\ &= \left(\frac{1}{N} \sum_{r,s=1}^{N} \operatorname{Cov}\left(\phi_j(X_r), \phi_k(X_s)\right) \right)_{j,k=1,\ldots,l}. \end{aligned}$$

Note that the laws $\mathcal{L}(\xi_N)$ are in the space $M_1(M_{\mathrm{f}}([0,1]))$ of all positive measures ν on $M_{\mathrm{f}}([0,1])$ with total mass $\nu(M_{\mathrm{f}}([0,1])) \leq 1$. To describe vague and weak convergence of laws of empirical processes, following Reinert (1997) we use as set of test functions

$$\begin{aligned} C_{\mathcal{F}}(M^f([0,1])) := \{F \in C_b(M^f([0,1])) : \; & F \text{ has the form} \\ & F(\nu) = f(\langle \nu, \phi_1 \rangle, \ldots, \langle \nu, \phi_l \rangle) \\ & \text{for an } l \in \mathbf{N}, f \in C_b^\infty(\mathbf{R}^l), \phi_1, \ldots, \phi_l \in \mathcal{FL}^+([0,1])\}. \end{aligned}$$

Then $C_{\mathcal{F}}(M^f([0,1]))$ is convergence-determining for vague convergence in $M_1(M_{\mathrm{f}}([0,1]))$. Moreover, for empirical processes, $C_{\mathcal{F}}(M^f([0,1]))$ is convergence-determining for weak convergence of the laws $\mathcal{L}(\xi_N)$. There is also a Zolotarev-type metric to quantify this convergence; convergence in this metric is slightly stronger than vague convergence. For $\phi \in \mathcal{FL}^+([0,1])$, let c_ϕ denote its Lipschitz constant, as before. Define

$$\begin{aligned} \mathcal{F}_{\mathcal{F}} = \{F \in C_{\mathcal{F}}(M^f([0,1])) : F \text{ has the form } F(\nu) = f(\langle \nu, \phi_1 \rangle, \ldots, \langle \nu, \phi_l \rangle) \\ \text{for an } l \in \mathbf{N}, f \in C_b^\infty(\mathbf{R}^l) \text{ with } \|f'\| \leq 1, \|f''\| \leq 1, \|f'''\| \leq 1; \\ \phi_1, \ldots, \phi_l \in \mathcal{FL}^+([0,1]) \text{ with } \|\phi_i\| \leq 1, c_{\phi_i} \leq 1, i = 1, \ldots, l\}. \end{aligned}$$

Here, $\|f'\| = \sum_{j=1}^{l} \|f(\mathbf{x})_{(j)}\|$, where $f(\mathbf{x})_{(j)}$ denotes the partial derivative $\frac{\partial}{\partial x_j} f(\mathbf{x})$; and $\|f\| = \sup_{\mathbf{x} \in \mathbf{R}^l} \sum_{i=1}^{l} |f(\mathbf{x})_i|$, where $f(\mathbf{x})_i$ denotes the ith coordinate of $f(\mathbf{x})$. For $\nu, \eta \in M_1(M_{\mathrm{f}}([0,1]))$, we put

$$\varsigma_{\mathcal{F}}(\nu, \eta) = \sup_{f \in \mathcal{F}_{\mathcal{F}}} \left| \int_{M_{\mathrm{f}}([0,1])} f \, d\nu - \int_{M_{\mathrm{f}}([0,1])} f \, d\eta \right|.$$

Then $\varsigma_{\mathcal{F}}$ is a metric on the space $M_1(M^f([0,1]))$, and for all $(\nu_N)_{N \in \mathbf{N}}, \nu \in M_1(M^f([0,1]))$,

$$\text{if } \varsigma_{\mathcal{F}}(\nu_N, \nu) \to 0 \quad (N \to \infty) \quad \text{then} \quad \nu_N \Rightarrow_v \nu \quad (N \to \infty).$$

Here $\Rightarrow_v$ denotes vague convergence.

Theorem 4.7. *(i) If* $m = \max\{(2+\frac{3}{\beta})/\log c, 2/\log\frac{1}{\alpha}\}\log N$, *then*

$$\begin{aligned}
\varsigma_{\mathcal{F}}(\xi_N,\zeta_N) & \\
&\leq 20m^2 N^{-\frac{1}{2}} + (K_{3.4}+1)N^{-3/2} + N^{-2}\left\{\frac{K_{4.3}}{1-\alpha}(1+K_{1.4}) + 4(m+1)(K_{3.4}+1)\right\} \\
&= O(N^{-\frac{1}{2}}\log^2 N).
\end{aligned}$$

(ii) Suppose that $b : \mathcal{FL}^+([0,1]) \times \mathcal{FL}^+([0,1]) \to \mathbf{R}$ *is such that for all* $l \in \mathbf{N}, \phi_1,\ldots,\phi_l \in \mathcal{FL}^+([0,1])$, $B(\phi_1,\ldots,\phi_l) = \big(b(\phi_i,\phi_j)\big)_{i,j=1,\ldots,l}$ *is a symmetric, positive definite matrix. Then, if* ζ *is a random measure with values in* $M_{\mathrm{f}}([0,1])$ *almost surely such that*

$$\mathcal{L}(\langle\zeta,\phi_1\rangle,\ldots,\langle\zeta,\phi_l\rangle) = \mathcal{MVN}_l(0,B(\phi_1,\ldots,\phi_l)),$$

then

$$\varsigma_{\mathcal{F}}(\xi_N,\zeta) \leq \zeta_{\mathcal{F}}(\xi_N,\zeta_N) + \max_{\phi,\psi\in C^*}|b_N(\phi,\psi) - b(\phi,\psi)|,$$

where $C^* = \{\phi \in \mathcal{FL}^+(0,1) : c_\phi \leq 1, \|\phi\| \leq 1\}$.

Proof. To assess $\varsigma_{\mathcal{F}}(\xi_N,\zeta_N)$, we once again employ the m-dependent approximation. Let ζ'_N be a random measure with values in $M_{\mathrm{f}}([0,1])$ almost surely such that, for all $l \in \mathbf{N}, \phi_1,\ldots,\phi_l \in \mathcal{FL}^+([0,1])$,

$$\mathcal{L}(\langle\zeta'_N,\phi_1\rangle,\ldots,\langle\zeta'_N,\phi_l\rangle) = \mathcal{MVN}_l(0,B'_N(\phi_1,\ldots,\phi_l)),$$

where

$$\begin{aligned}
B'_N(\phi_1,\ldots,\phi_l) &= \big(b'_N(\phi_j,\phi_k)\big)_{j,k=1,\ldots,m} \\
&= \left(\frac{1}{N}\sum_{r,s=1}^{N}\mathrm{Cov}\,(\phi_j(Y'_r),\phi_k(Y'_s))\right)_{j,k=1,\ldots,l}.
\end{aligned}$$

By the triangle inequality, we have

$$\varsigma_{\mathcal{F}}(\xi_N,\zeta_N) \leq \varsigma_{\mathcal{F}}(\xi_N,\xi'_N) + \varsigma_{\mathcal{F}}(\xi'_N,\zeta'_N) + \varsigma_{\mathcal{F}}(\zeta'_N,\zeta_N). \tag{4.31}$$

For the first summand, for all $F \in C_{\mathcal{F}}(M^f([0,1]))$ of the form $F(\nu) = f(\langle\nu,\phi_1\rangle,\ldots,\langle\nu,\phi_l\rangle)$, we find by Taylor's expansion that

$$\begin{aligned}
&|\mathbb{E}(F(\xi_N) - F(\xi'_N))| \\
&\quad = \big|\mathbb{E}\big(f(\langle\xi_N,\phi_1\rangle,\ldots,\langle\xi_N,\phi_l\rangle) - f(\langle\xi'_N,\phi_1\rangle,\ldots,\langle\xi'_N,\phi_l\rangle)\big)\big| \\
&\quad \leq \|f'\|\sum_{j=1}^{l}\mathbb{E}\big|\langle\xi_N - \xi'_N,\phi_j\rangle\big|.
\end{aligned}$$

For each $\phi \in \mathcal{FL}^+([0,1])$, using (4.30) and proceeding as for (4.27),

$$\begin{aligned}\mathbb{E}\big|\langle \xi_N - \xi_N', \phi\rangle\big| &\le c_\phi \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbb{E}\big|Y_i - Y_i'\big| \\ &\le c_\phi (K_{3.4}+1) N^{-\frac{3}{2}},\end{aligned}$$

because $m \ge (2 + (3/\beta)) \log N / \log c$ as usual. Thus

$$\varsigma_{\mathcal{F}}(\xi_N, \xi_N') \le (K_{3.4}+1) N^{-\frac{3}{2}}. \tag{4.32}$$

Moreover, from Corollary 17 in Reinert (1997), we obtain

$$\varsigma_{\mathcal{F}}(\xi_N', \zeta_N') \le \frac{20}{\sqrt{N}} m^2. \tag{4.33}$$

Finally, for $\varsigma_{\mathcal{F}}(\zeta_N', \zeta_N)$ we employ Corollary 16 in Reinert (1997). As ζ_N is Gaussian, we get

$$\varsigma_{\mathcal{F}}(\zeta_N', \zeta_N) \le \sup_{\phi,\psi \in C^*} |b_N(\phi,\psi) - b_N'(\phi,\psi)|, \tag{4.34}$$

where $C^* = \{\phi \in \mathcal{FL}^+(0,1) : c_\phi \le 1, \|\phi\| \le 1\}$. We have, for all $\phi, \psi \in C^*$,

$$\begin{aligned}&b_N(\phi_j,\phi_k) - b_N'(\phi_j,\phi_k)\\ =&\frac{1}{N}\sum_{r,s=1}^N \Big(\mathrm{Cov}\,(\phi_j(Y_r),\phi_k(Y_s)) - \mathrm{Cov}\,(\phi_j(Y_r'),\phi_k(Y_s'))\Big)\\ =&\frac{1}{N}\sum_{r=1}^N \sum_{s:|s-r|>m} \mathrm{Cov}\,(\phi_j(Y_r),\phi_k(Y_s))\\ &+\frac{1}{N}\sum_{r=1}^N \sum_{s:|s-r|\le m} \Big(\mathrm{Cov}\,(\phi_j(Y_r),\phi_k(Y_s)) - \mathrm{Cov}\,(\phi_j(Y_r'),\phi_k(Y_s'))\Big).\end{aligned}$$

For the first summand, as for (4.24), we obtain, with $\|\phi\| \le 1$ and $c_\phi \le 1$,

$$\frac{1}{N}\sum_{r=1}^N \sum_{s:|s-r|>m} \mathrm{Cov}\,(\phi_j(Y_r),\phi_k(Y_s)) \le N^{-2}\frac{K_{4.3}}{1-\alpha}(1+K_{1.4}),$$

by the choice of m. For the second summand, conditioning on whether or not all distances $|Y_j - Y_j'|$ are less than N^{-2} yields

$$\begin{aligned}&\frac{1}{N}\sum_{r=1}^N \sum_{s:|s-r|\le m} \Big(\mathrm{Cov}\,(\phi_j(Y_r),\phi_k(Y_s)) - \mathrm{Cov}\,(\phi_j(Y_r'),\phi_k(Y_s'))\Big)\\ \le&4(m+1)N^{-2}(K_{3.4}+1).\end{aligned}$$

Thus we get that

$$\sup_{\phi,\psi\in C^*}|b_N(\phi,\psi)-b'_N(\phi,\psi)|\leq N^{-2}\frac{K_{4.3}}{1-\alpha}(1+K_{1.4})+4(m+1)N^{-2}(K_{3.4}+1). \quad (4.35)$$

Combining (4.31)–(4.35), we obtain part (i) of the theorem. Part (ii) follows as for (4.34).

Acknowledgement. We would like to thank Walter Philipp and John Einmahl for fruitful discussions.

References

M. A. Arcones and B. Yu (1994) Central limit theorems for empirical and U-processes of stationary mixing sequences. *J. Theor. Probab.* **7,** 47–71.

A. D. Barbour (1990) Stein's method for diffusion approximations. *Probab. Th. Rel. Fields* **84,** 297–322.

A. D. Barbour (1996) Chance and chaos. *Elemente der Mathematik* **51,** 57–63.

L. H. Y. Chen (1986) The rate of convergence in a central limit theorem for dependent random variables with arbitrary index set. *IMA Preprint Series* **243,** Univ. Minnesota.

Z. Coelho and W. Parry (1990) Central limit asymptotics for shifts of finite type. *Israel J. Math.* **69,** 235–249.

A. Dembo and O. Zeitouni (1996) Moderate deviations for iterates of expanding maps. pp. 1–11 of *Statistics and control of Stochastic processes: the Liptser Festschrift.* Yu. M. Kabanov, B. L. Rozovskii, V. I. Shiryayev, eds., World Scientific, Singapore.

P. Doukhan, P. Massart and E. Rio (1994) The functional central limit theorem for strongly mixing processes. *Ann. Inst. Henri Poincaré* **30,** 63–82.

F. Götze (1991) On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19,** 724–739.

F. Götze and C. Hipp (1983) Asymptotic expansions for sums of weakly dependent random vectors. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **64,** 211–239.

L. Heinrich (1982) A method for the derivation of limit theorems for sums of m-dependent random variables. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **60,** 501–515.

L. Heinrich (1985) Some remarks on asymptotic expansions in the Central Limit Theorem for m-dependent random variables. *Math. Nachr.* **122**, 151–155.

L. Heinrich (1996) Mixing properties and central limit theorem for a class of non–identical piecewise monotonic C^2–transformations. *Math. Nachr.* **181,** 185–214.

F. Hofbauer and G. Keller (1982) Ergodic properties of invariant measures for piecewise monotonic transformations. *Math. Z.* **180,** 119–140.

I. A. Ibragimov and Y. A. Rozanov (1978) *Gaussian processes.* Springer, New York.

V. Isham (1992) Statistical aspects of chaos: a review. pp 124–200 of *Networks and Chaos – Statistical and Probabilistic Aspects.* O. E. Barndorff-Nielsen, J. L. Jensen, W. S. Kendall, eds., Chapman and Hall, London.

G. Keller (1985) Generalized bounded variation and applications to piecewise monotonic transformations. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **69,** 461–478.

A. Lasota and J. A. Yorke (1973) On the existence of invariant measures for piecewise monotonic transformations. *Trans. Amer. Math. Soc.* **186,** 481–488.

T. Lindvall (1992) *Lectures on the coupling method.* Wiley, New York.

C. Liverani (1995) Decay of correlations for piecewise expanding maps. *J. Stat. Phys.* **78,** 1111–1129.

W.L. Loh (1994) On m-dependence and Edgeworth expansions. *Ann. Inst. Statis. Math.* **46,** 147–164.

W. Philipp (1982) Invariance principles for sums of mixing random elements and the multivariate empirical process *Colloquia Mathematica Societatis Janos Bolyai, 36. Limit Theorems in Probability and Statistics, Veszprem (Hungary),* 843–873.

W. Philipp and W. Stout (1975) Almost sure invariance principles for partial sums of weakly dependent random variables. *Mem. Amer. Math. Soc.* **161.**

S. T. Rachev (1991) *Probability metrics and the stability of stochastic models.* Wiley, Chichester.

G. Reinert (1997) A Gaussian approximation for random measures via Stein's method. *Preprint.*

Y. Rinott and V. Rotar (1996) A multivariate CLT for local dependence with $n^{-\frac{1}{2}} logn$ rate, and applications to multivariate graph related statistics. *J. Multivariate Analysis* **56,** 333–350.

L. Rousseau–Egele (1983) Un theorème de la limite locale pour une classe de transformations dilétantes et monotones par morçeaux. *Ann. Probab.* **11,** 772–788.

V. V. Shergin (1979) On the convergence rate in the central limit theorem for m-dependent random variables. *Theory Probab. Appl.* **24,** 782–796.

V. V. Shergin (1984) On a global form of the central limit theorem for m-dependent random variables. *Theor. Probability and Math. Statist.* **29,** 149–156.

C. Stein (1972) A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2**, 583–602.

W. Stout (1974) *Almost sure convergence.* Academic Press, New York.

A. N. Tikhomirov (1980) On the convergence rate in the central limit theorem for weakly dependent random variables. *Theory Probab. Appl.* **25,** 790–809.