Can. J. Math. Vol. **49** (1), 1997 pp. 3–23

SWEEPING OUT PROPERTIES OF OPERATOR SEQUENCES

MUSTAFA A. AKCOGLU, DZUNG M. HA AND ROGER L. JONES

ABSTRACT. Let $L_p = L_p(X, \mu)$, $1 \leq p \leq \infty$, be the usual Banach Spaces of real valued functions on a complete non-atomic probability space. Let $(T_1, \ldots, T_K)$ be L_2-contractions. Let $0 < \varepsilon < \delta \leq 1$. Call a function f a δ-spanning function if $\|f\|_2 = 1$ and if $\|T_k f - Q_{k-1}T_k f\|_2 \geq \delta$ for each $k = 1, \ldots, K$, where $Q_0 = 0$ and Q_k is the orthogonal projection on the subspace spanned by $(T_1 f, \ldots, T_k f)$. Call a function h a (δ, ε) -sweeping function if $\|h\|_\infty \leq 1$, $\|h\|_1 < \varepsilon$, and if $\max_{1\leq k\leq K} |T_k h| > \delta - \varepsilon$ on a set of measure greater than $1 - \varepsilon$. The following is the main technical result, which is obtained by elementary estimates. There is an integer $K = K(\varepsilon, \delta) \geq 1$ such that if f is a δ-spanning function, and if the joint distribution of $(f, T_1 f, \ldots, T_K f)$ is normal, then $h = \Big((f \wedge M) \vee (-M)\Big)/M$ is a (δ, ε)-sweeping function, for some $M > 0$. Furthermore, if T_ks are the averages of operators induced by the iterates of a measure preserving ergodic transformation, then a similar result is true without requiring that the joint distribution is normal. This gives the following theorem on a sequence (T_i) of these averages. Assume that for each $K \geq 1$ there is a subsequence $(T_{i_1}, \ldots, T_{i_K})$ of length K, and a δ-spanning function f_K for this subsequence. Then for each $\varepsilon > 0$ there is a function h, $0 \leq h \leq 1$, $\|h\|_1 < \varepsilon$, such that $\limsup_i T_i h \geq \delta$ a.e.. Another application of the main result gives a refinement of a part of Bourgain's "Entropy Theorem", resulting in a different, self contained proof of that theorem.

1. **Introduction.** Let (X, μ) be a complete non-atomic probability space and $L_p = L_p(X, \mu)$ the usual Banach Spaces of functions on (X, μ), $1 \leq p \leq \infty$. We will consider only the real valued case; possible extensions to the complex case will be obvious. A linear operator T on L_p is called positive if $Tf \geq 0$ whenever $f \geq 0$ and a contraction, or an L_p-contraction, if $\|T\|_p \leq 1$. Let (T_i) be a sequence of operators on L_p. In most cases T_i's will be contractions of L_2. We are interested in knowing whether or not the sequence $(T_i f)$ converges pointwise (a.e.) for all $f \in L_p$. If $(T_i f)$ diverges pointwise, then we are also interested in the degree of divergence, as measured by sweeping out properties defined below. We will obtain sufficient conditions for divergence of a given degree, in terms of the L_2 behaviour of the sequence $(T_i f)$. Our methods also show how to construct functions h that will result in divergent sequences $T_i h$ of a given degree. We will now start with some definitions and remarks. The main result is stated as Theorems 1.6

The first author was partially supported by an NSERC Grant.
The second author was partially supported by an NSERC Grant.
The third author was partially supported by an NSF Grant.
Received by the editors May, 1995.
AMS subject classification: Primary: 28D99; Secondary: 60F99.
Key words and phrases: Strong and δ-sweeping out, Gaussian distributions, Bourgain's entropy theorem.

https://doi.org/10.4153/CJM-1997-001-3 Published online by Cambridge University Press

and 1.7. The relation between these theorems and Bourgain's entropy theorem [10] is discussed in the remarks following the statement of Theorem 1.7.

1.1 *Definition of δ-sweeping out.* Let (T_i) denote a sequence of positive operators and $0 < \delta \leq 1$. Then (T_i) is said to be δ*-sweeping out* if for each $\varepsilon > 0$ there is a set E such that $\mu(E) < \varepsilon$ and such that $\limsup T_i\chi_E(x) \geq \delta$ a.e..

REMARKS. If $\delta = 1$ then the sequence is said to have the *strong sweeping out* property. This was the main definition used to characterize the sequences that behave very badly. Several examples of strongly sweeping out sequences are given in [1]. In [18] Rosenblatt pointed out that there are sequences of positive L^∞ contractions for which a.e. divergence occurs, but which do not have the strong sweeping out property. He then introduced the above definition of δ-sweeping out and showed that there are sequences of operators (T_n) which are δ sweeping out, but not $\delta + \epsilon$ sweeping out for any $\epsilon > 0$. Thus it becomes interesting to establish which sequences of operators are δ-sweeping out, since it gives us a measure of divergence. The δ-sweeping out character of a sequence will follow from a certain type of behaviour of its finite segments, stated in the next definition.

1.2 *Definition of (δ, ε)-sweeping.* Let $0 < \varepsilon < \delta \leq 1$. A sequence of functions $(h_1, \ldots, h_K)$ will be called a (δ, ε)*-sweeping* sequence if $\|h_k\|_1 < \varepsilon$ for each $k = 1, \ldots, K$, and if $\max_{1 \leq k \leq K} |h_k| > \delta - \varepsilon$ on a set of measure greater than $1 - \varepsilon$. Let $(T_1, \ldots, T_K)$ be finitely many operators and h be a function. Then h is called a (δ, ε)-sweeping function (for $(T_1, \ldots, T_K)$) if $\|h\|_\infty \leq 1$, $\|h\|_1 < \varepsilon$, and if $(T_1h, \ldots, T_Kh)$ is a (δ, ε)-sweeping sequence.

REMARK 1.3. Let (T_n) be a sequence of positive L_∞ contractions and $\delta > 0$ fixed. If for each $\varepsilon > 0$ there is a (δ, ε)-sweeping function for a finite set of operators from this sequence, then it is clear that (T_n) is δ-sweeping out. (See also Rosenblatt [18].) If T_n's are not assumed to be positive, then the existence of (δ, ε)-sweeping functions for each $\varepsilon > 0$ still implies the existence of an L_∞ function h such that T_nh diverges a.e.. This, however, is not a direct observation as in the positive case, but follows from the results of Bellow and Jones in [7].

The following definition describes an L_2 behaviour for a finite sequence of functions, which implies, in certain cases, the existence of (δ, ε)-sweeping sequences, as discussed below.

1.4 *Definition of δ-spanning.* Let $0 < \delta \leq 1$. A sequence of vectors $(a_1, \ldots, a_K)$ in an inner product space will be called a δ*-spanning sequence* if $\|a_k\| \leq 1$ and $\|a_k - Q_{k-1}a_k\| \geq \delta$ for each $k = 1, \ldots, K$, where $Q_0 = 0$ and Q_k is the orthogonal projection on the subspace spanned by $(a_1, \ldots, a_k)$. In particular a sequence of functions $(f_1, \ldots, f_K)$ will be called a δ*-spanning sequence* if they form a δ-spanning sequence in L_2. Let $(T_1, \ldots, T_K)$ be finitely many operators and f be a function. Then f is called a δ*-spanning function* (for $(T_1, \ldots, T_K)$) if $\|f\|_2 = 1$ and if $(T_1f, \ldots, T_Kf)$ is a δ-spanning sequence.

The following is the main technical result of this article, which establishes a relation between the δ-spanning and (δ, ε)-sweeping functions.

THEOREM 1.5. *Let $0 < \varepsilon < \delta \le 1$. Then there is an integer $K = K(\delta, \varepsilon) \ge 1$ with the following property. Let $(T_1, \ldots, T_K)$ be K contractions in L_2. If there is a δ-spanning function f for $(T_1, \ldots, T_K)$ such that the joint distribution of $(f, T_1 f, \ldots, T_K f)$ is normal, then there is an $M > 0$ such that $h = (1/M)\big((f \wedge M) \vee (-M)\big)$ is a (δ, ε)-sweeping function for $(T_1, \ldots, T_K)$.*

This result may not be too useful by itself, because of the restrictive hypothesis that the functions involved must have a joint normal distribution. This hypothesis is not needed, however, in the following two cases described below. The first case seems to be more important, more intuitive, and easier to deal with. As the second case is more general, however, we will not give the details for the first case, and prove only the second case. The discussion of the second case starts in Section 1.9, with the definition of ergodic sequences. The corresponding result is stated as Theorem 1.10.

AVERAGES OF ERGODIC TRANSFORMATIONS. Let $\tau: X \longrightarrow X$ be a measure preserving and ergodic transformation of (X, μ). We do not assume that τ is invertible. Let $\mathcal{A}$ be the class of operators of the form $Tf = \sum_{i=0}^{n} \alpha_i f \cdot \tau^i$. Let $\mathcal{T}$ be the class of all L_2-contractions that can be approximated by operators from $\mathcal{A}$ in the sense that for each $T \in \mathcal{T}$, $f \in L_2$, and $\varepsilon > 0$, there is a $T' \in \mathcal{A}$ such that $\|Tf - T'f\|_2 < \varepsilon$. For transformations in $\mathcal{T}$ the following result is true.

THEOREM 1.6. *Let $(T_0, T_1, \ldots, T_K)$ be $K+1$ transformations in $\mathcal{T}$ and $f \in L_2$, where $T_0 = I$ is the identity transformation. Then there is a sequence f_n in L_2, such that all the inner products $(T_i f_n,\ T_j f_n)$ converge to the corresponding products $(T_i f,\ T_j f)$, $i, j = 0, 1, \ldots, K$ and such that the joint distributions of $(T_0 f_n,\ T_1 f_n, \ldots, T_K f_n)$ converge to a normal distribution in the weak topology of measures in R^{K+1}, induced by the real valued bounded continuous functions on R^{K+1}.*

The proof is constructive and uses Rokhlin's Lemma and the central limit theorem. We will omit the details. See also [3], where a part of this result has been proved under the assumption that the measure space is the unit circle and the ergodic transformation τ is an irrational rotation. As known, this example is typical and, with some additional work, implies the general case. Finally, routine approximations show that the Theorems 1.5 and 1.6 together imply the following result.

THEOREM 1.7. *Given $0 < \varepsilon < \delta \le 1$ there is an integer $K = K(\delta, \varepsilon) \ge 1$ with the following property. If $(T_1, \ldots, , T_K)$ are K transformations in $\mathcal{T}$, and if there is a δ-spanning function f for $(T_1, \ldots, , T_K)$, then there is also a $(\delta - \varepsilon,\ \varepsilon)$-sweeping function h for $(T_1, \ldots, T_K)$.*

Here the extra ε in the (δ, ε)-sweeping function is due to the fact that an approximately normal distribution must be used in the application of the first theorem. Again, we will omit the easy details. It is also clear that this result, together with the remark in 1.3, implies the following theorem.

THEOREM 1.8. *Let $\delta > 0$ and let (T_i) be a sequence of transformations in $\mathcal{T}$. If for each integer $K \geq 1$ there are K operators from this sequence admitting a δ-spanning function, then $(T_i h)$ diverges a.e. for an L_∞ function h. Further, if the operators are positive, then (T_i) is δ-sweeping out.*

1.9 *Ergodic Sequences.* We will now describe the second case referred to above. A sequence of operators (P_n) on $L_2(X, \mu)$ will be called an *ergodic sequence* if each P_n is a positive isometry of L_2 with $P_n 1 = 1$, and if $(1/n)\sum_{j=1}^{n} P_j f$ converges in L_2 to the constant function $\int_X f\,d\mu$ for each $f \in L_2$. In this case it is easy to see that each P_n is a positive contraction of all L_p spaces, $P_n f P_n g = P_n(fg)$ for each $f, g \in L_2$, and that $(1/n)\sum_{j=1}^{n} P_j f$ converges to $\int_X f\,d\mu$ in L_1, for each $f \in L_1$. We will say that a family of operators commute with an ergodic sequence, if there is an ergodic sequence (P_n) such that each P_n commutes with each member of that family.

THEOREM 1.10. *Given $0 < \varepsilon < \delta \leq 1$ there is an integer $K = K(\delta, \varepsilon) \geq 1$ with the following property. If $(T_1, \ldots, , T_K)$ are K contractions in L_2 commuting with an ergodic sequence, and if there is a δ-spanning function f for $(T_1, \ldots, , T_K)$, then there is also a (δ, ε)-sweeping function h for $(T_1, \ldots, T_K)$.*

As before, it is clear that this result, together with the remark in 1.3, implies the following theorem.

THEOREM 1.11. *Let $\delta > 0$ and let (T_i) be a sequence of L_2-contractions commuting with an ergodic sequence. If for each integer $K \geq 1$ there are K operators from this sequence admitting a δ-spanning function, then $(T_i h)$ diverges a.e. for an L_∞ function h. Further, if the operators are positive, then (T_i) is δ-sweeping out.*

RELATION TO BOURGAIN'S ENTROPY THEOREM. These results are similar to the results obtained by Bourgain in his entropy theorem [10]. This similarity will be explained further in the Appendix, Section 6. The method introduced by Bourgain in that theorem for establishing divergence has proven very useful, and resulted in the first proof of divergence for several sequences of operators. (See [10], [18], and [14] for example.) In fact, Rosenblatt [18] was able to show that for many interesting diverging sequences of operators, there is a $\delta > 0$ so that the sequence is δ-sweeping out. However, with the estimates used in Bourgain's proof, it was unclear what the relationship was between the entropy of the diverging sequence of operators, and the value of δ for δ-sweeping out. In particular, it was not clear if the entropy criteria could be used to establish strong sweeping out. In this paper we refine Bourgain's result so that an estimate of sweeping out can also be obtained, and in particular in some situations we can obtain strong sweeping out (see Theorem 1.10). In addition, we give a self contained different proof of a part of his results, concerning the divergence for bounded functions. Although the present proof depends on the same basic idea due to Bourgain, namely reducing the general case to the normally distributed case and making the estimates for the normally distributed case, it differs from Bourgain's original proof at both of these stages. In particular, the estimates for the normal case are done completely differently, in an elementary and self contained

https://doi.org/10.4153/CJM-1997-001-3 Published online by Cambridge University Press

way, without the use of Slepian's lemma or Sudakov's estimate, and the estimates are aimed at obtaining the sweeping out properties of the sequence.

REMARK. In the case of operators induced by averages of an ergodic measure preserving transformation, the existence of a commuting ergodic sequence (P_j) is automatic. However, for more general operators, the (P_j) play an important role. In particular, the results obtained here are false if we do not assume such a sequence exits. As a simple example, consider the sequence

$$T_nf(x) = \left(\sqrt{n(n+1)}\int_0^1 f(t)\,dt\right)\chi_{(\frac{1}{n+1},\frac{1}{n})}(x).$$

These operators are contractions on L_2, and it is easy to check that for every $K > 1$ the function 1 is a δ-spanning function. (In fact we see that $(T_n1, T_m1) = 0$ and $\|T_n1\|_2 = 1$ so these functions even form an orthonormal set.) However, we clearly do not have a (δ, ε) sweeping function for $(T_1, T_2, \ldots, T_K)$ for any $K > 1$.

RELATED RESULTS. Most of the related results are associated with ergodic transformations. Rosenblatt [17] considered the sequence of binomial averages (b_n) defined by $b_nf(x) = \frac{1}{2^n}\sum_{k=0}^n \binom{n}{k} f(\tau^k x)$, and showed that they can diverge, even for functions $f \in L_\infty$. Krengel [16] considered averages along subsequences of the form $C_nf(x) = \frac{1}{n}\sum_{k=1}^n f(\tau^{n_k}x)$ where (n_k) is an increasing sequence of integers. If $(n_k) = k$ we have the usual ergodic averages, which of course converge a.e.. However, Krengel [16] showed that there are subsequences for which the averages diverge. Later Bellow [4], [5] showed that these subsequence averages diverge if $(n_k) = (2^k)$, while Bourgain [10] showed they converge if $(n_k) = (k^2)$ and $f \in L_p$, $p > 1$. Finally we mention a classical paper of Rudin [19], in which pointwise divergence properties of Riemann sums were established. The present proofs for the ergodic case apply to this example. Many other special cases are known, as discussed in the given references, but understanding for which subsequences we have convergence of averages along these subsequences remains an open question.

In addition to operators that arise from questions in ergodic theory, we can also consider operators that arise in other areas. For example, in the case of singular integrals we can consider the operator defined by the Fourier multiplier $n^{i\gamma}$, and take a sequence of γ's that converge to zero. The question of a.e. convergence or divergence was open for many years, and the divergence has been shown [14] by the entropy theorem.

COROLLARIES AND EXAMPLES. In some applications it will be convenient to replace the L_2 behaviour described by the δ-spanning sequences by a different type of behaviour. The following theorems are obtained in this way.

THEOREM 1.12. *Let $0 \le \rho < 1$ and let (T_i) be a sequence of L_2-contractions commuting with an ergodic sequence. If for each integer $K \ge 1$ there are K operators $(T_{i_1}, \ldots, T_{i_K})$ from this sequence and a function f such that $\|T_{i_k}f\|_2 \ge 1$ and $(T_{i_k}f,\ T_{i_l}f) \le \rho$ for all $1 \le k \ne l \le K$, then (T_ih) diverges a.e. for an L_∞ function h. Further, if the operators are positive, then (T_i) is δ-sweeping out with $\delta = \sqrt{1-\rho}$.*

https://doi.org/10.4153/CJM-1997-001-3 Published online by Cambridge University Press

THEOREM 1.13. *Let (T_i) be a sequence of positive L_2-contractions commuting with an ergodic sequence. Assume that for each ρ, $0 < \rho < 1$, and for each integer $K \geq 1$ there are K operators $(T_{i_1}, \ldots, T_{i_K})$ from this sequence and a function f such that $\|T_{i_k} f\|_2 \geq 1$ and $(T_{i_k} f,\, T_{i_l} f) \leq \rho$ for all $1 \leq k \neq l \leq K$. Then (T_i) is strong sweeping out.*

As an application of these last two theorems we mention the following result. A special case of this result, corresponding to "arbitrarily small α" and strong sweeping out, is contained in [1], but with a different proof. The construction given in the present proof is similar to an argument of Rosenblatt in [18], pp. 237–238.

COROLLARY 1.14. *Let $0 < \alpha < (1/5)$ and let τ be an invertible measure preserving ergodic transformation. Define a sequence of averages by $T_n f(x) = \sum_{k=-\infty}^{\infty} \nu_n(k) f(\tau^k x)$, where (ν_n) is a sequence of probability measures on $\mathbb{Z}$, the set of integers. The Fourier Transform of ν_n is*

$$\hat{\nu}_n(\gamma) = \sum_{k=-\infty}^{\infty} \nu_n(k) \exp(2\pi i \gamma k).$$

Assume that for each $K \times 2^K$ matrix $\Sigma = (\sigma_{kj})$ with each $\sigma_{kj} \in \{-1, 1\}$, we can find $n_1, n_2, \ldots, n_K$ and $\gamma_1, \ldots, \gamma_{2^K}$ such that $|\hat{\nu}_{n_k}(\gamma_j) - \sigma_{kj}| < \alpha$. Then the sequence of operators T_n are δ-sweeping out with $\delta = \sqrt{(1-5\alpha)/(1-\alpha)}$. Further, if these hypotheses are satisfied with any choice of α, then (T_n) is strong sweeping out.

PROOF. By standard transfer arguments it is enough to establish this result for a single dynamical system. We will establish it for the irrational shift on $[0, 1)$ corresponding to an irrational number θ. Choose an integer $K \geq 1$ and 2^K integers ℓ_j and define

$$f(x) = \frac{1}{\sqrt{2^K}} \sum_{j=1}^{2^K} e^{2\pi i \ell_j x}.$$

Then we see that

$$T_n f(x) = \frac{1}{\sqrt{2^K}} \sum_{j=1}^{2^K} \hat{\nu}_n(\ell_j \theta) e^{2\pi i \ell_j x}.$$

Hence

$$(T_n f,\, T_m f) = \frac{1}{2^K} \sum_{j=1}^{2^K} \hat{\nu}_n(\ell_j \theta) \hat{\nu}_m(\ell_j \theta).$$

Form the matrix $(\sigma_{k,j})$ so that each row has half 1's and half -1's, and such that the rows are independent. (Note that this implies $\sum_{j=1}^{2^K} \sigma_{uj}\sigma_{vj} = 0$.) Then find γ_k's so that $|\hat{\nu}_{n_k}(\gamma_j) - \sigma_{k,j}| < \alpha$. There are integers ℓ_j such that we also have $|\hat{\nu}_{n_k}(\ell_j \theta) - \sigma_{k,j}| < \alpha$.

https://doi.org/10.4153/CJM-1997-001-3 Published online by Cambridge University Press

Hence if f is defined in terms of these integers then we see that, if $1 \le u \ne v \le K$,

$$\begin{aligned}|(T_{n_u}f,\, T_{n_v}f)| &= \left|\frac{1}{2^K}\sum_{j=1}^{2^K}\hat{\nu}_{n_u}(\ell_j\theta)\hat{\nu}_{n_v}(\ell_j\theta)\right| \\ &= \left|\frac{1}{2^K}\sum_{j=1}^{2^K}\Big(\hat{\nu}_{n_u}(\ell_j\theta)\hat{\nu}_{n_v}(\ell_j\theta) - \sigma_{uj}\sigma_{vj}\Big)\right| \\ &\le \frac{1}{2^K}\sum_{j=1}^{2^K}|\hat{\nu}_{n_u}(\ell_j\theta)\hat{\nu}_{n_v}(\ell_j\theta) - \sigma_{uj}\sigma_{vj}| \le 2\alpha.\end{aligned}$$

Also note that, since $|\sigma_{kj}| = 1$, $|\hat{\nu}_{n_k}(\ell_j\theta)| > 1-\alpha$. Hence $\|T_{n_k}f\|_2 > 1-\alpha$ for all $1 \le k \le K$. Looking at the real and imaginary parts of $T_{n_k}f$ and noticing that for at least half of the k's, one of these parts has an L_2-norm not less than $(1-\alpha)/2$, we see that the hypotheses of Theorem 1.12 are satisfied with $\rho = 4\alpha/(1-\alpha)$. Hence the proof of the first part follows. The last part follows from Theorem 1.13.

EXAMPLE 1.15. Fix a probability measure μ on Z, such that μ satisfies

$$\lim_{\gamma\to 1}\frac{|\hat{\mu}(\gamma)-1|}{1-|\hat{\mu}(\gamma)|} = \infty.$$

Then the measure ν_n defined by $\nu_n = \mu \star \mu \star \mu \star \cdots \star \mu$, the n-fold convolution of μ with itself, has the strong sweeping out property.

The estimates necessary to apply Corollary 1.15 to the above example follow as in [1]. The sequence of binomial averages considered by Rosenblatt, and defined $b_nf(x) = \frac{1}{2^n}\sum_{j=1}^n\binom{n}{j}f(\tau^jx)$, can be viewed as an n-fold convolution of $\frac{1}{2}(\delta_0 + \delta_1)$. Consequently these averages can be shown to have the strong sweeping out property by the above result.

OUTLINE OF THE PAPER. In Section 2 we review the normal distributions briefly, mainly to establish our notation and state the results we are going to assume. Section 3 contains the estimates to prove the main technical result, Theorem 1.5, which is obtained in Section 4. In Section 5 it is shown that the general case of Theorem 1.10 can be reduced to the normally distributed case of Theorem 1.5. Finally Theorem 1.12 is obtained in Section 6, together with other formulations of some of the results.

2. **Gauss Measures.**

2.1 *The Standard Gauss Measure.* A finite dimensional inner product space W has a particular measure $\Gamma = \Gamma_W$ on its Borel sets $B \in \mathcal{B}$ of W, defined by

$$\Gamma(B) = \frac{1}{K}\int_B \exp\Big(-\frac{1}{2}(w,\, w)\Big)\,\lambda(dw),$$

where λ is a Lebesgue (or Haar) measure for W, and

$$K = \int_W \exp\Big(-\frac{1}{2}(w,\, w)\Big)\,\lambda(dw)$$

is the normalizing factor to make $\Gamma(W) = 1$. Since any two Lebesgue measures for W differ only by a multiplicative constant, the standard Gauss measure Γ is independent of the choice of λ. In particular, if $W = \mathbb{R}$, with its usual inner product, then we see that

$$\Gamma_{\mathbb{R}}(B) = \frac{1}{\sqrt{2\pi}} \int_B e^{-s^2/2}\, ds$$

for any Borel set $B \subset \mathbb{R}$. For $y \in \mathbb{R}$ we will let

$$\varpi(y) = \Gamma_{\mathbb{R}}\big((y, \infty)\big) = \frac{1}{\sqrt{2\pi}} \int_y^\infty e^{-s^2/2}\, ds.$$

The following elementary estimate on $\varpi(y)$ is obtained by integration by parts. The details are given, for example, in [12], Chapter 7, Lemma 2.

LEMMA 2.2. *If $y > 0$, then*

$$\frac{1}{y\sqrt{2\pi}}\Big(1 - \frac{1}{y^2}\Big) \exp\Big(-\frac{y^2}{2}\Big) < \varpi(y) < \frac{1}{y\sqrt{2\pi}} \exp\Big(-\frac{y^2}{2}\Big).$$

We will use this estimate to obtain the following two computational lemmas.

LEMMA 2.3. *Let $0 < \varepsilon < 1$ and $2 < L$. If*

$$K > 10Le^{L^2/2} \log(1/\varepsilon)$$

then

$$[1 - 2\varpi(L)]^K < \varepsilon.$$

PROOF. Let K_0 be the solution of the equation $[1 - 2\varpi(L)]^{K_0} = \varepsilon$. Hence

$$K_0 = \frac{\log \varepsilon}{\log[1 - 2\varpi(L)]} \le [2\varpi(L)]^{-1} \log \frac{1}{\varepsilon}.$$

Using the estimate

$$\frac{1}{L\sqrt{2\pi}}\Big(1 - \frac{1}{L^2}\Big)e^{-L^2/2} < \varpi(L)$$

given above we obtain

$$K_0 \le \frac{\sqrt{2\pi}L^3}{2(L^2 - 1)} e^{L^2/2} \log(1/\varepsilon) \le 10Le^{L^2/2} \log(1/\varepsilon).$$

Hence $[1 - 2\varpi(L)]^K < \varepsilon$ whenever $K > K_0$.

The following estimate can be improved substantially. The given form is sufficient for our purpose, however.

LEMMA 2.4. *If $M > 1$ then*

$$\frac{1}{\sqrt{2\pi}}\int_M^\infty t^2 e^{-t^2/2}\,dt < Me^{-M^2/2}.$$

PROOF. Integrating by parts and using the estimate

$$\int_M^\infty e^{-t^2/2}\,dt < (1/M)e^{-M^2/2}$$

we obtain

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\int_M^\infty t^2 e^{-t^2/2}\,dt &= \frac{1}{\sqrt{2\pi}}\Big(Me^{-M^2/2} + \int_M^\infty e^{-t^2/2}\,dt\Big)\\ &\le \frac{1}{\sqrt{2\pi}}\Big(M + \frac{1}{M}\Big)e^{-M^2/2},\end{aligned}$$

which implies the estimate in the Lemma.

NOTATION 2.5. A finite sequence of functions $F = (f_1, \ldots, f_K)$ on (X, μ) defines a function

$$F = (f_1, \ldots, f_K)\colon X \longrightarrow \mathbb{R}^K$$

which transports μ to a measure Δ_F on $\mathbb{R}^K$, defined by $\Delta_F(G) = \mu(F^{-1}G)$ for measurable sets $G \subset \mathbb{R}^K$, and called the distribution measure of F. If $f\colon X \longrightarrow \mathbb{R}$ and $M > 0$, then $(f)_M = (f \wedge M) \vee (-M)$ is the truncated function at $\pm M$.

COROLLARY 2.6. *Let the distribution measure of $f\colon X \longrightarrow \mathbb{R}$ be the standard Gauss measure $\Gamma_{\mathbb{R}}$. Then $\|(f)_M\|_2^2 > 1 - 2Me^{-M^2/2}$ and $\|f - (f)_M\|_2^2 < 2Me^{-M^2/2}$ for all $M > 1$.*

PROOF. We have

$$\begin{aligned}\|(f)_M\|_2^2 &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty (t^2 \wedge M^2)e^{-t^2/2}\,dt\\ &= 1 - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty [t^2 - (t^2 \wedge M^2)]e^{-t^2/2}\,dt\\ &> 1 - 2\frac{1}{\sqrt{2\pi}}\int_M^\infty t^2 e^{-t^2/2}\,dt > 1 - 2Me^{-M^2/2},\end{aligned}$$

which is the first estimate. For the second estimate note that f and $f - (f)_M$ have the same sign at every point. Hence $\big(f, f - (f)_M\big) \ge 0$ and

$$\|f - (f)_M\|_2^2 \le \|f\|_2^2 - \|(f)_M\|_2^2.$$

2.7 *Properties of the Standard Gauss Measure.* Isometries of W leave Γ_W invariant. An orthogonal projection $P\colon W \longrightarrow V$ of W onto a subspace V transports Γ_W to Γ_V. If V and U are two orthogonal subspaces spanning W then Γ_W is equal to the Cartesian product measure $\Gamma_V \times \Gamma_U$, with the usual identification between W and $V \times U$. In particular, if

U is a one dimensional subspace spanned by a unit vector u_0 orthogonal to V, and if the pairs $(v,t) \in V \times \mathbb{R}$ are identified with the vectors $w = (v + tu_0) \in W$, then

$$\int_W f\, d\Gamma_W = \frac{1}{\sqrt{2\pi}} \int_V \int_{\mathbb{R}} f(v + tu_0) e^{-t^2/2}\, dt \Gamma_V(dv).$$

Finally,

$$0 = \int_W (a,w)\, \Gamma_W(dw), \text{ and,}$$
$$(a,b) = \int_W (a,w)(b,w)\, \Gamma_W(dw)$$

for any $a, b \in W$.

2.8 *General Gauss Measures.* A measure γ on a finite dimensional vector space V is called a (general) Gauss measure if it can be obtained by transporting the standard Gauss measure Γ_W of a finite dimensional inner product space W by a linear transformation $L\colon W \longrightarrow V$. In particular, Gauss measures on $\mathbb{R}^K$ are identified with $K \times K$ covariance matrices (that is, symmetric nonnegative definite matrices) in the usual way, as follows. Given K vectors $(a_1, \ldots, a_K)$ in W, the linear transformation $L\colon W \longrightarrow \mathbb{R}^K$ defined by

$$L(w) = \big((w, a_1), \ldots, (w, a_K)\big) \in \mathbb{R}^K$$

induces a Gauss measure γ on $\mathbb{R}^K$. Since Γ_W is invariant under the linear isometries of W, this measure γ depends only on the matrix $A = \{(a_i, a_j)\}$ formed by the inner products of $(a_1, \ldots, a_K)$. This is a covariance matrix. Conversely, any covariance matrix is of this form and induces a Gauss measure on $\mathbb{R}^K$. Sometimes we will write γ_A to denote the Gauss measure corresponding to A. Finally, for a fixed bounded and continuous function $\varphi\colon \mathbb{R}^K \longrightarrow \mathbb{R}$, the integral $\int_{\mathbb{R}^K} \varphi\, d\gamma_A$ depends only on A. We will denote this integral by $\Phi(A)$. It is clear that $\Phi(A)$ is a continuous function of A. Finally we have, with the usual coordinate functions $\xi_k\colon \mathbb{R}^K \longrightarrow \mathbb{R}$, and with the Gauss measure γ defined above in terms of the K vectors $(a_1, \ldots, a_K)$,

$$\int_{\mathbb{R}^K} \xi_i \xi_j\, d\gamma = \int_W (w, a_i)(w, a_j)\, \Gamma_W(dw) = (a_i, a_j)$$

for all $i, j = 1, \ldots, K$. Hence,

$$\int_{\mathbb{R}^K} \xi_i \xi_j\, d\gamma_A = A_{ij},$$

where $A = \{A_{ij}\}$ is a covariance matrix.

2.9 *Gauss Measures Induced by L_2-Functions.* In particular, if $(a_1, \ldots, a_K)$ is a set of K functions $F = (f_1, \ldots, f_K)$ in $L_2(X, \mu)$, considered as an inner product space, then the induced Gauss measure γ on $\mathbb{R}^K$, as defined above, will be called the Gauss measure induced by $F = (f_1, \ldots, f_K)$. It is uniquely specified by

$$\int_X f_i f_j\, d\mu = \int_{\mathbb{R}^K} \xi_i \xi_j\, d\gamma,$$

as before. If this Gauss measure γ is also the distribution measure of the mapping

$$(f_1, \ldots, f_K)\colon X \longrightarrow \mathbb{R}^K,$$

then the functions $(f_1, \ldots, f_K)$ are said to have a joint normal distribution.

https://doi.org/10.4153/CJM-1997-001-3 Published online by Cambridge University Press

3. **A Key Proposition.** Let $(T_1, T_2, \ldots, , T_K)$ be K contractions in L_2, and $f \in L^2(X)$. In this section we obtain a proposition which gives a condition on the function f and on the distribution of

$$(f,\ T_1 f, \ldots, T_K f)\colon X \longrightarrow \mathbb{R}^{K+1}$$

which is sufficient for the existence of a (δ, ε)-sweeping function for these operators. This proposition will imply the main Theorems 1.5 and 1.10 if we can show that the hypothesis are satisfied. Thus Sections 4 and 5 will be devoted to giving conditions under which the hypotheses of this proposition are satisfied.

PROPOSITION 3.1. *Let ε and δ be given, $0 < \varepsilon < \delta \leq 1$. Then there is an integer $K \geq 1$, and two numbers $R > 0$, $M > 1$ with the following property. Let $(T_1, \ldots, T_K)$ be K contractions in L_2, and f a function such that $\|f\|_2 \leq 3$, $\|f - (f)_M\|_2^2 < 10Me^{-M^2/2}$, and*

$$\mu\Big(\big\{x \mid x \in X,\ \max_{1\leq k\leq K} |T_k f(x)| \leq R\big\}\Big) < 10[1 - 2\varpi(R/\delta)]^K.$$

Then $h = (1/M)(f)_M$ is a (δ, ε)-sweeping function for $(T_1, \ldots, T_K)$.

PROOF. Given ε and δ as in the Proposition, find α and β such that $0 < \alpha < 1 < \beta$ and such that $\varepsilon > [1 - (\alpha/\beta)]\delta$. Then find a sufficiently large $R > 0$ such that

$$\frac{200\beta}{\delta^2(1-\alpha)^2} e^{-R^2(\beta^2-1)/(2\delta^2)} \log(20/\varepsilon) < \frac{\varepsilon}{2},$$
$$K_1 = 10(R/\delta) e^{R^2/(2\delta^2)} \log(20/\varepsilon) > 2,$$

and such that $\varepsilon/3 > 1/R$. Let K be any integer satisfying $K_1 < K < 2K_1$ and $M = \beta R/\delta$. With these choices for R, M, and K, assume that there is a function f and K operators $(T_1, \ldots, T_K)$ satisfying the conditions stated in the Proposition. We claim that $h = (1/M)(f)_M$ is a (δ, ε)-sweeping function for $(T_1, \ldots, T_K)$. First, clearly, $\|h\|_\infty \leq 1$. Also,

$$\|h\|_2 \leq (1/M)\|f\|_2 \leq (1/M) \leq (1/R) < \varepsilon.$$

Hence $\|h\|_1 < \varepsilon$. Now let A and B be the sets on which $\max_{1\leq k\leq K} |T_k f| \leq R$ and $\max_{1\leq k\leq K} |T_k h| \leq \delta - \varepsilon$, respectively. We would like to show that $\mu(B) < \varepsilon$. Since $(f)_M = Mh$, we see that B is also the set on which

$$\max_{1\leq k\leq K} |T_k (f)_M| \leq M(\delta - \varepsilon).$$

If C is the set on which

$$\max_{1\leq k\leq K} |T_k (f)_M| \leq R\alpha,$$

then $B \subset C$, since $M(\delta - \varepsilon) < R\alpha$, by the definitions of α and M. Hence it will be enough to show that $\mu(C) < \varepsilon$. We already know that

$$\mu(A) < 10[1 - 2\varpi(R/\delta)]^K < \varepsilon/2.$$

https://doi.org/10.4153/CJM-1997-001-3 Published online by Cambridge University Press

Here the first inequality follows from the hypothesis and the second inequality from Lemma 2.3, observing that $K > K_1$. Now $C - A$ is contained in the union of the sets D_k on which $|T_k(f)_M| \le R\alpha$ and $|T_k f| > R$, $k = 1, \ldots, K$. These sets, in turn, are contained in the sets E_k, on which $|T_k\big(f - (f)_M\big)| > R(1-\alpha)$. Since T_k's are L_2 contractions, and since $\|f - (f)_M\|_2^2 < 10Me^{-M^2/2}$ we see that

$$\mu(D_k) \le \mu(E_k) < \frac{1}{R^2(1-\alpha)^2} 10Me^{-M^2/2},$$

and, consequently,

$$\begin{aligned}
\mu(C - B) &< \frac{K}{R^2(1-\alpha)^2} 10Me^{-M^2/2} \\
&< \frac{2K_1}{R^2(1-\alpha)^2} 10Me^{-M^2/2} \\
&= \frac{200(R/\delta)e^{R^2/(2\delta^2)}}{R^2(1-\alpha)^2} \frac{\beta R}{\delta} e^{-\beta^2 R^2/(2\delta^2)} \log(20/\varepsilon) \\
&< \varepsilon/2,
\end{aligned}$$

because of the choice of R. Hence $\mu(B) < \varepsilon$ and the proof is completed.

4. **The Normally Distributed Case.**

As indicated in Section 3, we will now show, in Lemma 4.3 that the hypotheses of Proposition 3.1 are satisfied in the special case that the functions involved have a joint normal distribution. This will prove Theorem 1.5.

LEMMA 4.1. *Assume that an inner product space W is the linear span of a subspace V and a vector a not in V. Let $P\colon W \longrightarrow V$ be the orthogonal projection on V and $\delta = \|a - Pa\|$. Let $B \subset V$ be a Borel subset of the vector space V, $R > 0$ a number, and*

$$C = \{w \mid w \in W,\ Pw \in B,\ |(w, a)| \le R\}.$$

Then

$$\Gamma_W(C) \le [1 - 2\varpi(R/\delta)]\, \Gamma_V(B),$$

where Γ_W and Γ_V denote the respective standard Gauss measures of W and V.

PROOF. Let $u_0 = (a - Pa)/\delta$ be the unit vector in the direction of $(a - Pa)$. If U is the subspace spanned by u_0 then W is spanned by the orthogonal subspaces V and U. Representing the vectors $w \in W$ as $w = v + tu_0$ with $v = Pw \in V$ and $t \in \mathbb{R}$ we see that $w = v + tu_0$ is in C if and only if $v \in B$ and t belongs to the closed interval

$$I(v) = \Big[-\frac{(v, a)}{\delta} - \frac{R}{\delta}, \frac{(v, a)}{\delta} + \frac{R}{\delta}\Big].$$

Hence,

$$
\begin{aligned}
\Gamma_W(C) &= \int_W \chi_C(w)\Gamma(dw) \\
&= \int_V \int_{\mathbb{R}} \chi_C(v + tu_0)\Gamma_{\mathbb{R}}(dt)\Gamma_V(dv) \\
&= \frac{1}{\sqrt{2\pi}} \int_B \int_{I(v)} e^{-t^2/2}\, dt\Gamma_V(dv) \\
&\leq \frac{1}{\sqrt{2\pi}} \int_B \int_{-R/\delta}^{R/\delta} e^{-t^2/2}\, dt\Gamma_V(dv) \\
&= [1 - 2\varpi(R/\delta)]\Gamma_V(B),
\end{aligned}
$$

where the inequality follows from the observation that

$$\int_{\lambda-\alpha}^{\lambda+\alpha} e^{-t^2/2}\, dt \leq \int_{-\alpha}^{\alpha} e^{-t^2/2}\, dt$$

for all $\lambda \in \mathbb{R}$ and $\alpha \geq 0$.

LEMMA 4.2. *Let $(a_1, \ldots, a_K)$ be a δ-spanning sequence in W and $R > 0$. If C is the set of vectors $w \in W$ for which $|(w, a_k)| \leq R$ for each $k = 1, \ldots, K$, then*

$$\Gamma_W(C) \leq [1 - 2\varpi(R/\delta)]^K.$$

PROOF. Let V be the subspace spanned by the a_k's and U the orthogonal complement of V. Note that $C = (C \cap V) \times U$, which shows that $\Gamma_W(C) = \Gamma_V(C \cap V)\Gamma_U(U) = \Gamma_V(C \cap V)$. Hence, if $K = 1$ then, with $\alpha = \|a_1\| > \delta$,

$$\Gamma_W(C) = \frac{1}{\sqrt{2\pi}} \int_{-R/\alpha}^{R/\alpha} e^{-t^2/2}\, dt \leq [1 - 2\varpi(R/\delta)].$$

The general case follows by induction from the previous lemma.

LEMMA 4.3. *Let f be a δ-spanning function for $(T_1, \ldots, T_K)$. Assume that the functions $(f, T_1 f, \ldots, T_K f)$ have a joint normal distribution, as defined in Section 2.9. Then, for any $R > 0$ and $M > 1$, $\|f - (f)_M\|_2^2 < 2Me^{-M^2/2}$, and*

$$\mu\Big(\big\{x \mid x \in X,\ \max_{1 \leq k \leq K} |T_k f(x)| \leq R\big\}\Big) < 2[1 - 2\varpi(R/\delta)]^K.$$

PROOF. Let $\xi_k \colon \mathbb{R}^{K+1} \longrightarrow \mathbb{R}$, $k = 0, 1, \ldots, K$, be the coordinate functions in $\mathbb{R}^{K+1}$. We will write T_0 for the identity operator. Let γ be the Gauss measure on $\mathbb{R}^{K+1}$ induced by the functions $(T_0 f, T_1 f, \ldots, T_K f)$. Since these functions are assumed to have a joint normal distribution, the measure γ is also the distribution measure of the mapping

$$(T_0 f,\ T_1 f, \ldots, T_K f) \colon X \longrightarrow \mathbb{R}^{K+1}.$$

Let D be the subset of $\mathbb{R}^{K+1}$ defined by the condition that

$$\max_{1 \leq k \leq K} |\xi_k| < R.$$

https://doi.org/10.4153/CJM-1997-001-3 Published online by Cambridge University Press

Hence the second conclusion of the Lemma can be stated as

$$\gamma(D) < 2[1 - 2\varpi(R/\delta)]^K.$$

This follows from the corresponding statement in the previous lemma. In fact, let W be the finite dimensional inner product space obtained as the linear span of the vectors $a_k = T_k f$, $k = 0, 1, \ldots, K$, in the Hilbert Space $L_2(X, \mu)$. Consider W as a measure space with its standard Gauss measure Γ_W. Let C be the subset of W as defined in Lemma 4.3. Then $\Gamma_W(C) = \gamma(D)$, since γ is also the distribution measure of $L\colon W \longrightarrow \mathbb{R}^{K+1}$ that takes $w \in W$ to

$$L(w) = \big((w, a_0),\ (w, a_1), \ldots, (w,\ a_K)\big) \in \mathbb{R}^{K+1},$$

as defined in Section 2.9. To obtain the first statement, note that the norm of $\xi_0\colon \mathbb{R}^{K+1} \longrightarrow \mathbb{R}$ in $L_2(\mathbb{R}^{K+1}, \gamma)$ is equal to the norm of f in $L_2(X, \mu)$, which is one, because of the definition of a spanning function, as given in Section 1.4. Hence the distribution measure of ξ_0 is the standard Gauss measure on $\mathbb{R}$ and the result follows from Corollary 2.6.

5. **General Case.**

5.1 *Notation.* Recall that (X, μ) is a probability space. Let (P_n) be an ergodic sequence. Hence (P_n) is a sequence of positive isometries of $L_2(X, \mu)$ such that $P_n 1 = 1$ and such that $(1/N)\sum_{n=1}^{N} P_n f$ converges in L_2 to the constant function $\int_X f\, d\mu$, for each $f \in L_2$. It is easy to see that P_n's are contractions of each one of the L_p spaces, $1 \leq p \leq \infty$, and that $(1/N)\sum_{n=1}^{N} P_n f$ converges to $\int_X f\, d\mu$ in L_1, for each $f \in L_1$. Also, $P_n f P_n g = P_n(fg)$ for all f,g in L_2. Let W be a finite dimensional inner product space with its standard Gauss measure $\Gamma = \Gamma_W$. The product measure space $(X, \mu) \times (W, \Gamma)$ is denoted by (Z, ρ). Let (u_n) be a sequence of N orthonormal vectors in W. By means of these vectors we associate to each function $f\colon X \longrightarrow \mathbb{R}$ in L_2 another function $\psi f\colon X \longrightarrow W$, defined by

$$\psi f(x) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (P_n f)(x) u_n.$$

We also define $\Psi f\colon Z \longrightarrow \mathbb{R}$ by

$$\Psi f(x, w) = \big(\psi f(x), w\big) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (P_n f)(x)(u_n, w).$$

To denote the dependence of these mappings on N sometimes we also write $\psi_N f$ and $\Psi_N f$ instead of ψf and Ψf, respectively.

In what follows we will fix a set $(f_1, \ldots, f_K)$ of K functions in $L_2(X, \mu)$, with the covariance matrix $A = \{(f_k, f_l)\}$. For each $x \in X$, let $A(x) = A_N(x)$ be the covariance matrix of the K vectors $\big(\psi_N f_1(x), \ldots, \psi_N f_K(x)\big)$ in W. We will also fix a norm on the space of $K \times K$ matrices. As in Section 2.9, the Gauss measure corresponding to a covariance matrix B is denoted by γ_B and the integral of a continuous and bounded function $\varphi\colon \mathbb{R}^K \longrightarrow \mathbb{R}$ with respect to γ_B by $\Phi(B)$. Note that, if $(b_1, \ldots, b_K)$ are K vectors in W with the covariance matrix B, then

$$\Phi(B) = \int_{\mathbb{R}^K} \varphi\, d\gamma_B = \int_W \varphi\big((b_i, w)\big)\, d\Gamma_W.$$

Here the value of φ at $(\xi_1, \ldots, \xi_K) \in \mathbb{R}^K$ is denoted by $\varphi(\xi_i)$.

https://doi.org/10.4153/CJM-1997-001-3 Published online by Cambridge University Press

LEMMA 5.2. *Given any $\varepsilon > 0$, there is an N_0 such that if $N > N_0$ then*

$$\|A - A_N(x)\| < \varepsilon,$$

for all x in a set $E_N \subset X$ with $\mu(E_N) > 1 - \varepsilon$.

PROOF. Let f and g be two L_2 functions. We have

$$\begin{aligned}\big(\psi_N f(x),\ \psi_N g(x)\big) &= \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{N} P_n f(x) P_m g(x)(u_n, u_m)\\ &= \frac{1}{N}\sum_{n=1}^{N} P_n f(x) P_n g(x)\\ &= \frac{1}{N}\sum_{n=1}^{N} P_n(fg)(x).\end{aligned}$$

This last sequence converges in L_1 to the constant function $(f,g) = \int_X fg\,d\mu$. Hence given any $\varepsilon_1 > 0$ there is an N_1 such that, if $N \geq N_1$ then

$$\Big|\big(\psi_N f(x),\ \psi_N g(x)\big) - (f,g)\Big| < \varepsilon_1$$

for all $x \in F_N \subset X$, with $\mu(F_N) > 1 - \varepsilon_1$. Then the proof follows by applying this argument K^2 times, with a sufficiently small ε_1.

LEMMA 5.3. *If f and g belong to $L_2(X,\mu)$ then Ψf and Ψg belong to $L_2(Z,\rho)$ and*

$$(f,g) = (\Psi f, \Psi g),$$

where the inner products are in their respective L_2 spaces.

PROOF. We have

$$\begin{aligned}(\Psi f, \Psi g) &= \int_Z \Psi f \Psi g\, d\rho\\ &= \int_X \int_W \big(\psi f(x), w\big)\big(\psi g(x), w\big)\Gamma_W(dw)\mu(dx)\\ &= \int_X \big(\psi f(x), \psi g(x)\big)\mu(dx)\\ &= \frac{1}{N}\sum_{n=1}^{N}\int_X P_n f(x) P_n g(x)\mu(dx)\\ &= \frac{1}{N}\sum_{n=1}^{N}(P_n f, P_n g)\\ &= (f,g).\end{aligned}$$

Here, the third equality follows from a basic property of the standard Gauss measures, as stated in Section 2.7.

LEMMA 5.4. *As in Notation 5.1, let* $(f_1, \ldots, f_K)$ *be* K *functions in* $L_2(X, \mu)$ *with the covariance matrix* A. *Let* θ_N *be the distribution measure of the mapping*

$$(\Psi_N f_1, \ldots, \Psi_N f_K) \colon Z \longrightarrow \mathbb{R}^K,$$

and let $\varphi_i \colon \mathbb{R}^K \longrightarrow \mathbb{R}$, $i = 1, \ldots, I$, *be finitely many bounded and continuous functions. Then for each* $\varepsilon > 0$ *there is an* N_0 *such that if* $N \geq N_0$ *then*

$$\left| \int_{\mathbb{R}^K} \varphi_i \, d\theta_N - \int_{\mathbb{R}^K} \varphi_i \, d\gamma_A \right| < \varepsilon$$

for each $i = 1, \ldots, I$.

PROOF. We use the notation of 5.1. We will assume, without loss of generality, that $I = 1$ and write φ for φ_1. We have

$$\begin{aligned} \int_{\mathbb{R}^K} \varphi \, d\theta_N &= \int_Z \varphi\big((\psi_N f_i(x), w)\big) \rho(dx\, dw) \\ &= \int_X \int_W \varphi\big((\psi_N f_i(x), w)\big) \Gamma_W(dw) \mu(dx) \\ &= \int_X \Phi\big(A_N(x)\big) \mu(dx). \end{aligned}$$

Let $M = \|\varphi\|_\infty$. Given ε_1, $0 < \varepsilon_1 < \varepsilon/(2M+1)$, find $\delta > 0$ such that $|\Phi(B) - \Phi(A)| < \varepsilon_1$ whenever B is a covariance matrix with $\|B - A\| < \delta$. Use Lemma 5.2 to find an N_0 with the following property. For each $N \geq N_0$ there is a set $E_N \subset X$ with $\mu(E_N) > 1 - \varepsilon_1$ such that $\|A_N(x) - A\| < \delta$ for each $x \in E_N$. Then, if $N \geq N_0$,

$$\begin{aligned} \left| \int_{\mathbb{R}^K} \varphi \, d\gamma_A - \int_{\mathbb{R}^K} \varphi \, d\theta_N \right| &= \left| \Phi(A) - \int_X \Phi\big(A_N(x)\big) \mu(dx) \right| \\ &= \left| \int_X \Phi(A) \mu(dx) - \int_X \Phi\big(A_N(x)\big) \mu(dx) \right| \\ &\leq \int_X \left| \Phi(A) - \Phi\big(A_N(x)\big) \right| \mu(dx) \\ &\leq 2M\varepsilon_1 + \varepsilon_1 < \varepsilon. \end{aligned}$$

LEMMA 5.5. *Let* $(T_1, \ldots, T_K)$ *be* K *contractions in* L_2, *commuting with an ergodic sequence* (P_n). *Let* f *be a* δ-*spanning function for* $(T_1, \ldots, T_K)$. *Then, given* $R > 0$ *and* $M > 1$, *there are* N *real numbers* α_n *such that the function*

$$g = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \alpha_n P_n f$$

satisfies the conditions that $\|g\|_2 < 3$, $\|g - (g)_M\|_2^2 < 10Me^{-M^2/2}$, *and*

$$\mu\Big(\big\{x \mid x \in X,\ \max_{1 \leq k \leq K} |T_k g(x)| \leq R\big\}\Big) < 10[1 - 2\varpi(R/\delta)]^K.$$

PROOF. We will use Lemma 4.3 and the notation introduced in this Lemma and its proof. In particular, T_0 denotes the identity operator. Let A be the $(K+1) \times (K+1)$

covariance matrix of the L_2 functions $f_k = T_k f$, $0 \le k \le K$, and γ_A the Gauss measure on $\mathbb{R}^{K+1}$ corresponding to A. From Lemma 4.4 and from the fact that $\|f_0\|_2 = \|f\|_2 = 1$ (by the definition of a δ-spanning function) we obtain

$$\int_{\mathbb{R}^{K+1}} \xi_0^2 \, d\gamma_A = 1,$$
$$\int_{\mathbb{R}^{K+1}} (\xi_0^2 \wedge M^2) \, d\gamma_A > 1 - 2Me^{-M^2/2},$$

and

$$\int_{\mathbb{R}^{K+1}} \chi_D \, d\gamma_A < 2[1 - 2\varpi(R/\delta)]^K,$$

where χ_D is the characteristic function of the set $D \subset \mathbb{R}^{K+1}$ on which

$$\max_{1 \le k \le K} |\xi_k| \le R.$$

Hence we can find a bounded and continuous function $\varphi : \mathbb{R}^{K+1} \longrightarrow \mathbb{R}$ such that $\chi_D \le \varphi$ and

$$\int_{\mathbb{R}^{K+1}} \varphi \, d\gamma_A < 2[1 - 2\varpi(R/\delta)]^K.$$

Now apply Lemma 5.4 with the bounded and continuous functions $(\xi_0^2 \wedge M^2)$ and φ to find a sufficiently large N, such that if $\Psi = \Psi_N$ and if θ is the distribution measure of the mapping

$$(\Psi f_0, \, \Psi f_1, \ldots, \Psi f_K) : Z \longrightarrow \mathbb{R}^{K+1},$$

then

$$\int_{\mathbb{R}^{K+1}} (\xi_0^2 \wedge M^2) \, d\theta > 1 - 2Me^{-M^2/2},$$

and

$$\int_{\mathbb{R}^{K+1}} \chi_D \, d\theta \le \int_{\mathbb{R}^{K+1}} \varphi \, d\theta < 2[1 - 2\varpi(R/\delta)]^K.$$

Also, note that, by Lemma 5.3,

$$\int_{\mathbb{R}^{K+1}} \xi_0^2 \, d\theta = \|\Psi f_0\|_2^2 = \|f_0\|_2^2 = 1.$$

Returning to the domain Z of the mapping

$$(\Psi f_0, \, \Psi f_1, \ldots, \Psi f_K) : Z \longrightarrow \mathbb{R}^{K+1},$$

we see that, if C is the subset of Z on which

$$\max_{1 \le k \le K} |\Psi f_k| \le R,$$

then

$$\int_Z (\Psi f_0)^2 \, d\rho = 1,$$
$$\int_Z [(\Psi f_0)^2 - (\Psi f_0)^2 \wedge M^2] \, d\rho < 2Me^{-M^2/2},$$

https://doi.org/10.4153/CJM-1997-001-3 Published online by Cambridge University Press

and

$$\int_Z \chi_C \, d\rho < 2[1 - 2\varpi(R/\delta)]^K.$$

For each $w \in W$ let

$$F(w) = \int_X (\Psi f_0)^2(x, w)\mu(dx),$$
$$G(w) = \int_X [(\Psi f_0)^2 - (\Psi f_0)^2 \wedge M^2](x, w)\mu(dx),$$
$$H(w) = \int_X \chi_C(x, w)\mu(dx).$$

Note that the values of these functions depend on the functions f_k only as members of L_2, not on their representations. Since F, G, and H are nonnegative functions of w with

$$\int_W F(w)\Gamma_W(dw) = 1,$$
$$\int_W G(w)\Gamma_W(dw) < 2Me^{-M^2/2},$$
$$\int_W H(w)\Gamma_W(dw) < 2[1 - 2\varpi(R/\delta)]^K,$$

it is clear that there is a point $w_0 \in W$ such that $F(w_0) < 9$, $G(w_0) < 10Me^{-M^2/2}$, and

$$H(w_0) < 10[1 - 2\varpi(R/\delta)]^K.$$

Let

$$g = \Psi f_0(\,\cdot\,, w_0) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (u_n, w_0) P_n f.$$

It is clear that $\|g\|_2^2 = F(w_0) < 9$ and, as in the Proof of Corollary 2.6,

$$\|g - (g)_M\|_2^2 \le \|g\|_2^2 - \|(g)_M\|_2^2 = G(w_0) < 10Me^{-M^2/2}.$$

Also, since

$$T_k g = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (u_n, w_0) P_n T_k f = \Psi f_k(\,\cdot\,, w_0),$$

we have

$$\mu\Bigl(\bigl\{x \mid x \in X, \max_{1 \le k \le K} |T_k g(x)| \le R\bigr\}\Bigr) = H(w_0) < 10[1 - 2\varpi(R/\delta)]^K.$$

Hence g satisfies the requirements of the Lemma.

Proof of Theorem 1.10. This proof follows easily from Proposition 3.1 and Lemma 5.5. Given ε and δ, $0 < \varepsilon < \delta \le 1$, find K, M and R from Proposition 3.1. Let $(T_1, \ldots, T_K)$ be K contractions in L_2, commuting with an ergodic sequence. Let f be a δ-spanning function for these operators. Use Lemma 5.5 to find a function g with the properties stated there. Then this function satisfies the hypotheses of Proposition 3.1, and $h = (1/M)(g)_M$ is a (δ, ε)-sweeping function for $(T_1, \ldots, T_K)$.

6. **Appendix.** We will first show, in Lemma 6.1 that Theorem 1.12 can be reduced to the Theorem 1.10. We also discuss the relation between the entropy condition used in Bourgain's original results and δ-spanning sequences. All the arguments are completely geometrical, valid in finite dimensional inner product spaces.

LEMMA 6.1. *Let $0 \le \rho < 1$ and $0 < \delta < \sqrt{1-\rho}$. Then for each integer $K \ge 1$ there is another integer $M = M_K \ge 2$ with the following property. Let $(a_1, \ldots, a_M)$ be a sequence of vectors in an inner product space W such that $\|a_m\| \ge 1$ and $(a_n, a_m) < \rho$ for all $1 \le n \ne m \le M$. Then $(a_1, \ldots, a_M)$ contains δ-spanning subsequence of length K.*

PROOF. Proceed by induction over K. For $K = 1$ we may take $M = 1$. Assume that M_K has been obtained. Let $\eta = \sqrt{1-\delta^2} - \sqrt{\rho} (> 0)$. Find an integer $N = N(K, \eta/6)$ such that the unit ball in a K-dimensional subspace of W can be covered by N balls of radius $\eta/6$. Choose an integer $A \ge 1$ such that $\sqrt{\rho + (2/A)} < \sqrt{\rho} + (\eta/3)$. We then let $M_{K+1} = M_K + AN$. To simplify the expressions we will say that a sequence in W is a ρ-sequence if each term has a norm ≥ 1 and the inner products of different terms are $\le \rho$. Note that a ρ-sequence stays a ρ-sequence if each term a_i is replaced by the unit vector $a_i/\|a_i\|$. Consider a ρ-sequence of length M_{K+1}. By the preceding remark we will assume, without loss of generality, that this sequence consists of unit vectors. To see that it contains a δ-spanning sequence of length $K+1$, first choose a δ-spanning sequence $(g_1, \ldots, g_K)$ of length K from the first M_K terms of the ρ-sequence. Let P be the orthogonal projection on the K-dimensional subspace E spanned by the δ-spanning sequence $(g_1, \ldots, g_K)$. Apply P to the last AN terms of the ρ-sequence to obtain AN vectors in the unit ball of E. Since this unit ball can be covered by N balls of radius $\eta/6$, there will be A of these projected vectors, say $(Pf_1, \ldots, Pf_A)$, that are contained in a ball of radius $\eta/6$. Note that here $(f_1, \ldots, f_A)$ is a ρ-sequence of length A chosen from the last AN terms of the original ρ-sequence.

Assume that $\|Pf_i\|_2 \ge \sqrt{1-\delta^2}$ for all $i = 1, \ldots, A$. Let g be the center of the ball of radius $\eta/6$ that contains all Pf_i's. Then, we have, for each $i = 1, \ldots, A$,

$$\begin{aligned}\frac{1}{36}\eta^2 &\ge \|g - Pf_i\|^2 \\ &= \|g\|^2 + \|Pf_i\|^2 - 2(g, Pf_i) \\ &\ge \|g\|^2 + \Big(\|g\| - \frac{1}{6}\eta\Big)^2 - 2(g, Pf_i) \\ &= 2\|g\|^2 - 2\|g\|\frac{1}{6}\eta + \frac{1}{36}\eta^2 - 2(g, Pf_i)\end{aligned}$$

which shows that

$$(g, Pf_i) \ge \|g\|\Big(\|g\| - \frac{1}{6}\eta\Big).$$

Hence

$$\Big(g, P\frac{1}{A}\sum_{i=1}^{A} f_i\Big) \ge \|g\|\Big(\|g\| - \frac{1}{6}\eta\Big),$$

https://doi.org/10.4153/CJM-1997-001-3 Published online by Cambridge University Press

or that

$$\Big\|\frac{1}{A}\sum_{i=1}^{A} f_i\Big\| \geq \|g\| - \frac{1}{6}\eta.$$

Since

$$\|g\|_2 \geq \|Pf_i\| - \frac{1}{6}\eta \geq \sqrt{1-\delta^2} - \frac{1}{6}\eta = \sqrt{\rho} + \frac{5}{6}\eta,$$

we then have

$$\Big\|\frac{1}{A}\sum_{i=1}^{A} f_i\Big\| \geq \sqrt{\rho} + \frac{2}{3}\eta.$$

We show that this is a contradiction. In fact,

$$\begin{aligned}\Big\|\frac{1}{A}\sum_{i=1}^{A} f_i\Big\|^2 &= \frac{1}{A^2}\Big(\sum_{i=1}^{A}\sum_{j=1}^{A}(f_i, f_j)\Big)\\ &\leq \frac{1}{A^2}\big(A + (A^2 - A)\rho\big)\\ &= \rho + \frac{1}{A}(1-\rho) \leq \rho + \frac{1}{A},\end{aligned}$$

which means that

$$\Big\|\frac{1}{A}\sum_{i=1}^{A} f_i\Big\|_2 \leq \sqrt{\rho + (1/A)} < \sqrt{\rho} + (\eta/3),$$

by the choice of A. Hence we must have $\|Pf_i\| < \sqrt{1-\delta^2}$ for at least one $i = 1, \ldots, A$. In this case we have

$$\|f_i - Pf_i\|^2 = \|f_i\|^2 - \|Pf_i\|^2 > 1 - (1-\delta^2) = \delta^2.$$

Therefore f_i can be added to the initial δ-spanning sequence $(g_1, \ldots, g_K)$ to obtain a δ-spanning sequence of length $K + 1$.

REMARK. We will now discuss the relation between the following two conditions on a family $\mathcal{C}$ of L_2 contractions. The first condition is a hypothesis in Bourgain's entropy theorem. The second condition is a hypothesis in Theorem 1.11.

(A) There is a $\lambda > 0$ such that for each integer $N \geq 1$ one can find N operators $(S_1, \ldots, S_N)$ from $\mathcal{C}$ and a function f, such that $\|f\|_2 \leq 1$ and $\|S_n f - S_m f\|_2 \geq \lambda$ for all $1 \leq n \neq m \leq N$.

(B) There is a $\delta > 0$ such that for each integer $K \geq 1$ one can find a sequence $(T_1, \ldots, T_K)$ of K operators from $\mathcal{C}$ for which there is a δ-spanning function.

It is clear that if (B) is satisfied then (A) is also satisfied with $\lambda = \delta$. Conversely, if (A) is satisfied then we will show that (B) is also satisfied with $\delta = \lambda/3$. In fact, the following general result is true. We will sketch the proof briefly.

LEMMA 6.2. *Let $\delta > 0$. Then for each integer $K \geq 1$ there is another integer $N = N_K$ such that if (s_i) is a sequence of N vectors in the unit ball of an inner product space W*

with $\|s_i - s_j\| > 3\delta$ *whenever* $i \neq j$, *then* (s_i) *contains a* δ*-spanning subsequence of length* K.

PROOF. First note that there is a number $M = M_n$ such that the unit ball of an n-dimensional subspace of W can not contain more than M vectors having a distance of at least δ between any two of them. For the proof the lemma, apply an induction over K, for a fixed $\delta > 0$. Take $N_1 = 2$. Assume that N_K has been obtained and let $N_{K+1} = N_K + M_K + 1$. It is easy to see that this choice works.

REFERENCES

1. M. Akcoglu, A. Bellow, R. Jones, V. Losert, K. Reinhold and M. Wierdl, *The strong sweeping out property for lacunary sequences, Riemann sums, convolution powers, and related matters*, Ergodic Theory Dynamical Systems, to appear.

2. M. Akcoglu, A. del Junco, W. Lee, *A solution to a problem of A. Bellow*. In: Almost Everywhere Convergence II, (eds. A. Bellow and R. Jones), Academic Press, 1991, 1–7.

3. M. Akcoglu, M. D. Ha and R. L. Jones, *Divergence of Ergodic Averages*. In: Topological vector spaces, algebras and related areas, (eds. Anthony To-Ming Lau and Ian Tweddle), Pitman Research Notes in Mathematics Series **316**, Longman, 1994, 175–192.

4. A. Bellow, *On bad universal sequences in ergodic theory (II)*, Springer Verlag Lecture Notes in Math., **1033**(1983).

5. ———, *Sur la structure des suites mauvaises universelles en theorie ergodique*, Comptes Rendus Acad. Sci. Paris, **294**(1982), 55–58.

6. ———, *Two problems*, Proc. Oberwolfach Conference on Measure Theory (June 1981), Springer Lecture Notes in Math. **945**(1982).

7. A. Bellow and R. Jones, *A Banach Principle for* L_∞, Adv. Math., to appear.

8. A. Bellow, R. Jones and J. Rosenblatt, *Convergence for moving averages*, Ergodic Theory Dynamical Systems, **10**(1990), 43–62.

9. J. Bourgain, *On the maximal theorem for certain subsets of the integers*, Israel J. Math. **61**(1988), 39–72.

10. ———, *Almost sure convergence and bounded entropy*, Israel J. Math. **63**(1988), 79–97.

11. A. del Junco and J. Rosenblatt, *Counterexamples in ergodic theory and number theory*, Math. Ann. **245** (1979), 185–197.

12. W. Feller, *An introduction to Probability theory and its applications*, vol. **1**, 2nd edition, John Wiley & Sons, Inc., 1957.

13. D. Ha, *Operators with Gaussian distribution property,* L_2*-entropy, and almost everywhere convergence*, Ph.D. Thesis, University of Toronto, 1994.

14. R. Jones, *A remark on singular integrals with complex homogeneity*, Proc. Amer. Math. Soc. **114**(1992), 763–768.

15. R. Jones and M. Wierdl, *Convergence and divergence of ergodic averages*, Ergodic Theory Dynamical Systems **14**(1994), 515–535.

16. U. Krengel, *On the individual ergodic theorem for subsequences*, Ann. Math. Stat. **42**(1971), 1091–1095.

17. J. Rosenblatt, *Ergodic group actions*, Arch. Math. **47**(1986), 263–269.

18. ———, *Universally bad sequences in ergodic theory*. In: Almost Everywhere Convergence II, (eds. A. Bellow and R. Jones), Academic Press, 1991, 227–245.

19. W. Rudin, *An arithmetic property of Riemann sums*, Proc. Amer. Math. Soc. **15**(1964), 321–324.

Department of Mathematics
University of Toronto
Toronto, Ontario
M5S 1A1

Department of Mathematics and Statistics
Sultan Qaboos University
Sultanate of Oman

Department of Mathematics
DePaul University
2219 N. Kenmore
Chicago IL
USA 60614