

On the Kaczmarz algorithm of approximation in infinite-dimensional spaces

by

STANISŁAW KWAPIEŃ (Warszawa) and JAN MYCIELSKI (Boulder, CO)

Abstract. The Kaczmarz algorithm of successive projections suggests the following concept. A sequence (e_k) of unit vectors in a Hilbert space is said to be *effective* if for each vector x in the space the sequence (x_n) converges to x where (x_n) is defined inductively: $x_0 = 0$ and $x_n = x_{n-1} + \alpha_n e_n$, where $\alpha_n = \langle x - x_{n-1}, e_n \rangle$. We prove the effectivity of some sequences in Hilbert spaces. We generalize the concept of effectivity to sequences of vectors in Banach spaces and we prove some results for this more general concept.

Introduction. In 1937 S. Kaczmarz [2] proved that given nonzero vectors (e_k) in a finite-dimensional vector space and numbers (c_k) , a solution (if it exists) to the system of linear equations $\langle x, e_k \rangle = c_k$, $k = 1, \dots, N$, can be obtained as the limit of the sequence defined inductively by

$$x_0 = 0, \quad x_n = x_{n-1} + \frac{c_n - \langle x_{n-1}, e_n \rangle}{\langle e_n, e_n \rangle} e_n$$

(here the sequences $(e_n), (c_n)$ are extended to infinite sequences by making them N -periodic). Since then, the method (also known as the “row-action method”) has been extended and modified in many ways. (For a review of the method, its generalizations, outlines of applications and references see [1].)

In this paper we consider the following generalization of the method: Let X be a real or complex Banach space. Let (e_k, f_k) be a sequence such that $e_k \in X$, $f_k \in X^*$, $f_k(e_k) = 1$ for all positive integers k (X^* denotes the dual space). The *Kaczmarz algorithm* of approximation of elements of X constructs for each $x \in X$ an approximating sequence x_n defined by the iterative procedure

$$x_0 = 0, \quad x_n = x_{n-1} + f_n(x - x_{n-1})e_n.$$

2000 *Mathematics Subject Classification*: 41A65, 60G25, 60H25.

Research of S. Kwapien supported in part by Polish Grant KBN 2 PO3A 043 15.

If Y is a (not necessarily closed) subspace of X then we will say that the sequence (e_k, f_k) is *effective in Y* if the sequence (x_n) converges to x for each $x \in Y$. If (e_k, f_k) is effective in the whole space X we say simply that it is *effective*. If X is a Hilbert space then we will assume that $e_k = f_k$ for all k and we will speak about effectivity of the sequence (e_k) rather than of (e_k, e_k) .

Our motivation for considering this generalization is the following problem connected with a model in learning theory (see [3], [4]): given a T -periodic continuous function $h : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ with $h(0) = 1$ (*base function*) and a sequence of times (t_n) , the question is which continuous T -periodic functions x (*patterns*) can be uniformly approximated (*recognized*) by the sequence (x_n) defined by the iteration (*learning procedure*)

$$x_0(t) = 0, \quad x_n(t) = x_{n-1}(t) + \alpha_n h(t - t_n) \quad \text{where} \quad \alpha_n = x(t_n) - x_{n-1}(t_n).$$

If X is the Banach space of all T -periodic continuous functions on \mathbb{R}^1 , $h \in X$, e_n is the function $h(t - t_n)$ and f_n is the evaluation at t_n functional, that is, $f_n(u) = u(t_n)$, then the question is in which subspaces the sequence (e_k, f_k) is effective.

We are especially interested in the sequences $(t_n) = (n\alpha)$ with α/T irrational (*systematic learning*) or when the t_n are chosen at random independently and uniformly in T (*casual learning*). For such sequences of times, under some natural assumptions on the function h , we prove that the class of functions x which can be uniformly approximated by the procedure contains some special, everywhere dense subspace with a natural Hilbert structure. We also prove that if we modify the above procedure (as suggested by the relaxation method in row-action algorithms, for references see [1]) by putting in the above definition $\alpha_n = 0$ whenever $|x(t_n) - x_{n-1}(t_n)| < \varepsilon$ (ε is a fixed positive number) then each continuous, T -periodic function x is uniformly, up to ε , approximated by the sequence x_n which for large n is a fixed function. These results will be derived from more general theorems on effective sequences in Banach (mainly Hilbert) spaces. We think that these abstract theorems are of independent interest.

Section 1 contains Examples 1, 2 in which we reformulate the above problem in learning theory. In Section 2 we prove that if we choose the sequence (e_k) at random (more exactly if (e_k) is a sequence of independent random unit vectors in H , with distribution which has support linearly dense in H) then the sequence is effective a.s. This is a consequence of a result (Proposition 1) on products of random projections in a Hilbert space. In Section 3 we characterize the stationary sequences in Hilbert space (see Section 3) which are effective as those whose spectral measure is either singular or coincides with the Lebesgue measure. Finally in Section 4, using the results obtained for Hilbert spaces, we prove an abstract version of a

result concerning the problem of uniform approximation by means of the relaxed algorithm.

1. Effective sequences. Definitions and basic properties. Let X be a real or complex Banach space and let X^* denotes its dual space. Let (e_k, f_k) be a sequence such that $e_k \in X$, $f_k \in X^*$, $f_k(e_k) = 1$ for all positive integers k . If we define the sequence (g_k) in X^* by induction: $g_1 = f_1$ and $g_n = f_n - \sum_{k=1}^{n-1} f_n(e_k)g_k$ then the approximating sequence x_n for x , which was defined in the introduction, can be written in the form $x_n = \sum_{k=1}^n g_k(x)e_k$. Hence the last sum, which we will abbreviate by $K_n(x)$, is a linear operator on X .

Obvious necessary conditions for the effectivity of the sequence (e_k, f_k) are that (e_k) be linearly dense in X and that (f_k) separate points of X . This will be assumed throughout the paper.

We check easily that if we define $P_k = \text{Id} - e_k \otimes f_k$ (i.e. $P_k(x) = x - f_k(x)e_k$ for $x \in X$) then P_k is a projection onto the kernel of f_k and for $x \in X$ we have

$$(1.0) \quad x - K_n(x) = P_n P_{n-1} \dots P_1(x).$$

Therefore by the Banach–Steinhaus Theorem the sequence (e_k, f_k) is effective if and only if there exists a constant C such that

$$(1.1) \quad \|P_n P_{n-1} \dots P_1\| \leq C \quad \text{for all } n,$$

and

$$(1.2) \quad \lim_{n \rightarrow \infty} P_n P_{n-1} \dots P_1(x) = 0 \quad \text{for all } x$$

in a linearly dense subset of X .

If X is a Hilbert space then the assumptions $f_k(e_k) = 1$ and $f_k = e_k$ imply that (e_k) is a sequence of unit vectors and that the projections P_k are orthogonal. Then the condition (1.1) is satisfied with $C = 1$. Moreover for each n the vector $p_n = x - x_n = P_n P_{n-1} \dots P_1(x)$ is orthogonal to $p_{n-1} - p_n = g_n(x)e_n = \langle x, g_n \rangle e_n$. Hence $\|x - x_n\|^2 = \|x\|^2 - (|g_1(x)|^2 + \dots + |g_n(x)|^2)$. Therefore the sequence (e_k) is effective in the Hilbert space X if and only if the associated sequence (g_k) satisfies $\|x\|^2 = \sum_{k=1}^{\infty} |\langle x, g_k \rangle|^2$ for all $x \in X$. In wavelet theory this condition means that (g_k) is a normalized tight frame.

Not all linearly dense sequences of unit vectors in a Hilbert space are effective. Indeed, we check easily that for all $x \in X$ and each $n > 2$ we have $\|x - x_n\| \geq |\langle e_{n-1}, e_n \rangle| \cdot \|x - x_{n-1}\|$. Hence $\|x - x_n\| \geq |\langle e_1, e_2 \rangle \dots \langle e_{n-1}, e_n \rangle| \cdot \|x - x_1\|$. Therefore an obvious necessary condition for the effectivity of a sequence (e_k) of unit vectors in a Hilbert space is that

$$(1.3) \quad \prod_{k=1}^{\infty} \langle e_k, e_{k+1} \rangle = 0.$$

In the general case of Banach spaces it is hard to check the condition (1.1) which is necessary for the effectivity of a sequence (e_k, f_k) . But in a Hilbert space it suffices only to check (1.2). In some concrete examples this can still be a difficult problem. In what follows we will discuss some special cases for which we are able to prove that this condition holds. The cases we consider constitute an abstract version of the following two examples which we will keep in mind throughout the paper.

EXAMPLE 1. Let $\mathbb{T} = \{z : |z| = 1\}$ be the unit circle in the complex plane (which is a group under multiplication), and let dt be the normalized Lebesgue measure (Haar measure) on \mathbb{T} . Let $L_2(\mathbb{T})$ be the Hilbert space of all square integrable functions on \mathbb{T} . Let $b \in L_2(\mathbb{T})$ be such that $\|b\|^2 = \int_{\mathbb{T}} |b(t)|^2 dt = 1$ and all the Fourier coefficients $\widehat{b}(n)$ are nonzero. Moreover let (t_k) be a dense sequence in \mathbb{T} . For each $t \in \mathbb{T}$ the translation of b by t is denoted by b_t , i.e., $b_t(s) = b(st)$. We are interested in the question if the sequence (b_{t_k}) is effective. In the next sections we will prove that it is so for all functions b as above if $(t_k) = (z_0^k)$ for any constant $z_0 \in \mathbb{T}$ which is not a root of 1, and it is so almost surely (a.s.) if the sequence (t_k) is chosen at random such that the t_k are independent and uniformly distributed in \mathbb{T} .

Note that not all sequences (b_{t_k}) are effective. Indeed, for $t, s \in \mathbb{T}$ we have $\langle b_s, b_t \rangle = h(t\bar{s})$, where the function h is defined below in Example 2. Therefore if h has a derivative at $t = 1$ it is easy to find a sequence (t_k) with distances between consecutive times going to zero, which is dense in \mathbb{T} and such that $\prod_{k=1}^{\infty} |h(t_k \bar{t}_{k+1})| > 0$. This shows that the condition (1.3) is not satisfied and hence the sequence (b_{t_k}) is not effective. (Effective learning needs bigger breaks between lessons!)

EXAMPLE 2. Under the same notation and assumptions as in Example 1 consider the function $h = b \star b^*$, i.e., $h(t) = \int_{\mathbb{T}} b(ts) \bar{b}(s) ds$ for $t \in \mathbb{T}$. A question of interest is if in the Banach space $C(\mathbb{T})$ the sequence (h_{t_k}, δ_{t_k}) is effective (here δ_t denotes the evaluation functional at $t \in \mathbb{T}$, i.e., $\delta_t(x) = x(t)$). We do not know if such sequences exist. In Section 4 we will show that this is not so for the sequences (t_k) of the two forms in Example 1. However, in that section, for sequences of these types we will construct large subspaces $Y \subset C(\mathbb{T})$ in which the sequence (h_{t_k}, δ_{t_k}) is effective. And we will show that the relaxed algorithm gives an ε -uniform approximation for each $x \in C(\mathbb{T})$.

REMARK 1. We made the assumption that $\widehat{b}(n) \neq 0$ for all n because only in this case is the sequence (b_{t_k}) (resp. (h_{t_k})) linearly dense in $L_2(\mathbb{T})$ (resp. $C(\mathbb{T})$).

2. Random effective sequences in a Hilbert space. We will prove that in a Hilbert space “almost all” sequences (e_k) of vectors of norm 1 are effective. To make the statement precise we introduce the following notation. If μ is a Borel probability measure on a separable metric space S then μ^∞ denotes the infinite product of μ on S^∞ , which is the metric space consisting of all sequences of elements from S . The *support* of μ is defined as the smallest closed set with measure 1. $S(H)$ denotes the unit sphere of a Hilbert space H .

THEOREM 1. *If H is a separable Hilbert space and μ is a probability measure on $S(H)$ with support linearly dense in H then $\mu^\infty(\{(e_k) : (e_k) \text{ is effective}\}) = 1$.*

Theorem 1 can be reformulated as follows: If (e_k) is a sequence of independent random unit vectors in a separable Hilbert space H which are identically distributed by a law with linearly dense support in H then the sequence is effective a.s.

For such sequences the projections P_k , which were defined in Section 1, are equidistributed and moreover $\text{Id} - P_k = e_k \otimes e_k$ is an orthogonal projection of rank 1. Hence the operator $R = E(\text{Id} - P_k)$ is nuclear with trace $\text{tr}(R)$ equal to 1 (here E denotes the expectation of a random operator and the integration is taken with respect to the strong operator topology). The operator R is strictly positive (i.e., $\langle Rx, x \rangle > 0$ for $x \neq 0$). Indeed, if $\langle Rx, x \rangle = 0$ for some $x \neq 0$ then, since $\langle Rx, x \rangle = E|\langle x, e_k \rangle|^2$, we get $\langle x, e_k \rangle = 0$ a.s. This implies that the support of the distribution of e_k is contained in the space orthogonal to x and so it is not linearly dense in H . Therefore, by (1.0), we see that Theorem 1 is a direct consequence of the following

PROPOSITION 1 ⁽¹⁾. *If (P_k) is a sequence of independent identically distributed random orthogonal projections such that the operator $R = E(\text{Id} - P_1)$ is nuclear and strictly positive, then the sequence $Q_n = P_n P_{n-1} \dots P_1$ is a.s. convergent to 0 in the strong operator topology, i.e., $Q_n(x) \rightarrow 0$ for each x .*

Proof. Since each P_k is a contraction, for each $x \in H$ the sequence $(\|Q_n(x)\|)$ is nonincreasing. Therefore to prove that a.s. the sequence $(\|Q_n(x)\|)$ converges to 0 for all $x \in X$, it is enough to prove that for each $x \in H$ the sequence converges in probability to 0. Indeed, let D be any countable dense subset of H and assume that the sequence $(Q_n(x))$ converges in probability to 0 for each $x \in H$. Given $d \in D$ we can choose a sequence (m_n) of positive integers such that $\|Q_{m_n}(d)\| \rightarrow 0$ a.s. In view

⁽¹⁾ A more general result was proved by Joel Zinn and the first author. It will appear elsewhere.

of the monotonicity of the sequence $(\|Q_n(d)\|)$ this implies that the last sequence is convergent to 0 a.s. Then, since D is countable, for all $d \in D$ the sequence $(\|Q_n(d)\|)$ is convergent to 0 a.s. Since D is dense in H and the Q_n are contractions we have $\|Q_n(x)\| \rightarrow 0$ for all $x \in H$ a.s. Convergence in probability follows from convergence in the second moment. Thus to prove Proposition 1 it is enough to show that $E\|Q_n(x)\|^2 \rightarrow 0$ for each $x \in H$. For each x we have $E\|Q_n(x)\|^2 = E\langle x, Q_n^* Q_n x \rangle = \langle x, E Q_n^* Q_n x \rangle = \langle x, A_n x \rangle$, where

$$A_n = E Q_n^* Q_n = E(P_1 P_2 \dots P_n P_n \dots P_2 P_1).$$

Thus it suffices to prove that $A_n \rightarrow 0$ in the strong operator topology. Since the P_n are independent and equidistributed we obtain

$$A_n = E P_1 (E P_2 \dots P_n P_n \dots P_2) P_1 = E P_1 A_{n-1} P_1.$$

The last equality and an easy induction imply that $A_{n-1} - A_n$ are positive operators. Indeed, if we put $A_0 = \text{Id}$ then for $n = 1$ we have $0 \leq E P_1 = A_1 \leq \text{Id} = A_0$. For $n > 1$ if $0 \leq A_n \leq A_{n-1} \leq \text{Id}$ then

$$0 \leq A_{n+1} = E P_1 A_n P_1 \leq E P_1 A_{n-1} P_1 = A_n \leq \text{Id}.$$

This proves that the sequence of operators A_n is monotone and bounded. Such a sequence is convergent in the strong operator topology. Denoting the limit by A_∞ we obtain $0 \leq A_\infty \leq \text{Id}$ and $E P_1 A_\infty P_1 = A_\infty$. By the last equality,

$$E(\text{Id} - P_1) A_\infty + E A_\infty (\text{Id} - P_1) = E(\text{Id} - P_1) A_\infty (\text{Id} - P_1),$$

or equivalently, $R A_\infty + A_\infty R = E(\text{Id} - P_1) A_\infty (\text{Id} - P_1)$. Taking the trace on both sides we get

$$\begin{aligned} 2 \text{tr}(R A_\infty) &= \text{tr}(E(\text{Id} - P_1) A_\infty (\text{Id} - P_1)) = E \text{tr}((\text{Id} - P_1) A_\infty (\text{Id} - P_1)) \\ &= E \text{tr}((\text{Id} - P_1)^2 A_\infty) = E \text{tr}((\text{Id} - P_1) A_\infty) \\ &= \text{tr}(E(\text{Id} - P_1) A_\infty) = \text{tr}(R A_\infty). \end{aligned}$$

Thus $\text{tr}(R A_\infty) = 0$. Since R is strictly positive and $A_\infty \geq 0$, this yields $A_\infty = 0$, which concludes the proof of Proposition 1.

COROLLARY 1. *If (t_k) is a sequence of independent, equidistributed random variables in \mathbb{T} and the support of their common distribution is \mathbb{T} then the sequence (b_{t_k}) , defined as in Example 1, is effective a.s.*

Proof. The random vectors b_{t_k} satisfy the assumptions of Theorem 1.

COROLLARY 2. *If (e_k) is a sequence of unit vectors which is linearly dense in a Hilbert space H then there exists a sequence (m_k) of positive integers such that the sequence (e_{m_k}) is effective.*

This corollary also follows from Theorem 1. It is enough to pick the m_k at random independently according to any law which assigns positive

probability to each positive integer. Of course in such sequences each positive integer appears infinitely many times. (Repetitions are needed for an effective learning!)

3. Stationary effective sequences in Hilbert space. A sequence (e_k) in a Hilbert space H is said to be *stationary* if $\langle e_{k+m}, e_{l+m} \rangle = \langle e_k, e_l \rangle$ for any positive integers k, l, m . It follows by Bochner's Theorem that if we define a_m for $m \in \mathbb{Z}$ (the set of all integers) by $a_m = \langle e_k, e_{k+m} \rangle$ where k is any positive integer with $k > -m$ then there exists a unique positive measure σ on \mathbb{T} (called the *spectral measure* of the stationary sequence) such that

$$a_m = \int_{\mathbb{T}} \bar{z}^m \sigma(dz) \quad \text{for each } m \in \mathbb{Z}.$$

A stationary sequence consists of unit vectors if and only if its spectral measure is a probability measure. And it is orthonormal if and only if its spectral measure coincides with the normalized Lebesgue measure on \mathbb{T} .

THEOREM 2. *A stationary sequence of unit vectors which is linearly dense in a Hilbert space is effective if and only if its spectral measure either coincides with the normalized Lebesgue measure or is singular with respect to Lebesgue measure.*

Proof. Let (e_k) be a stationary sequence of unit vectors in a Hilbert space H . Let σ be its spectral probability measure on \mathbb{T} and let (a_m) be defined as above. If we define the complex function F by

$$F(z) = \int_{\mathbb{T}} \frac{\sigma(dw)}{1 - z\bar{w}}$$

then F is well defined in $\mathbb{D} = \{z : |z| < 1\}$ and $F(z) = \sum_{n=0}^{\infty} a_n z^n$. Since the map $z \mapsto 1/(1-z)$ transforms \mathbb{D} onto $\{z : \operatorname{Re}(z) > 1/2\}$ and since σ is a probability measure we get $\operatorname{Re}(F(z)) > 1/2$ for all $z \in \mathbb{D}$. Taking the inverse map we see that for the function $G(z) = 1 - 1/F(z)$ we have $|G(z)| < 1$ for $z \in \mathbb{D}$. Therefore, since $F(0) = 1$, the Taylor series of $1/F(z) = \sum_{n=0}^{\infty} c_n z^n$ is convergent in \mathbb{D} and $\sum_{n=0}^{\infty} |c_n|^2 \leq 2$. It is clear that

$$(2.1) \quad c_0 = a_0 = 1 \quad \text{and} \quad \sum_{k=0}^n c_k a_{n-k} = 0 \quad \text{for } n \geq 1.$$

The above equalities easily imply that if (g_k) is the sequence associated to (e_k) as in Section 1, then $g_k = \sum_{i=1}^k c_{k-i} e_i$. Hence

$$K_n(x) = \sum_{k=1}^n \left(\sum_{i=1}^k c_{k-i} \langle x, e_i \rangle \right) e_k.$$

Since (e_k) is linearly dense the sequence is effective if and only if $\lim_{n \rightarrow \infty} K_n(e_i) = e_i$ for $i = 1, 2, \dots$

Using (2.1) we check that $K_n(e_1) = e_1$ and, for $i = 2, 3, \dots$,

$$K_n(e_i) - e_i = \sum_{j=1}^{i-1} \bar{a}_j \left(\sum_{l=0}^{n+j-i} c_l e_{l+i-j} \right).$$

Therefore the condition (1.2) is satisfied if and only if for any positive integers i, j such that $a_j \neq 0$ and $i > j$ the series $\sum_{l=0}^{\infty} c_l e_{l+i-j}$ converges to 0 in H . Since the sequence (e_k) is stationary, for fixed positive integers $i > j$ the map $e_l \mapsto e_{l+i-j}$, $l = 1, 2, \dots$, extends to an isometry of H into itself. Thus the last series converges to 0 if and only if $\sum_{l=0}^{\infty} c_l e_{l+1}$ converges to 0. Hence if $a_j \neq 0$ for some $j > 0$, then the sequence is effective if and only if the series $\sum_{l=0}^{\infty} c_l e_{l+1}$ is convergent to 0 in H , or equivalently, the sequence (r_n) of numbers defined by $r_n^2 = \|\sum_{l=0}^n c_l e_{l+1}\|^2$ converges to zero. However, if $a_j = 0$ for all $j = 1, 2, \dots$, then the sequence (e_k) is orthonormal, the spectral measure coincides with the normalized Lebesgue measure on \mathbb{T} and Theorem 1 is trivial in this case.

We will prove by induction that (r_n) satisfies

$$r_n^2 = 1 - |c_1|^2 - |c_2|^2 - \dots - |c_n|^2.$$

Indeed, for $n = 0$ both sides are equal to 1. For each $n > 1$ we have

$$\begin{aligned} r_n^2 &= r_{n-1}^2 + 2 \operatorname{Re} \left(\left\langle \sum_{l=0}^{n-1} c_l e_{l+1}, c_n e_{n+1} \right\rangle \right) + |c_n|^2 \\ &= r_{n-1}^2 + 2 \operatorname{Re} \left(\bar{c}_n \sum_{l=0}^{n-1} c_l a_{n-l} \right) + |c_n|^2 = r_{n-1}^2 - |c_n|^2 \end{aligned}$$

(because by (2.1), $\sum_{l=0}^{n-1} c_l a_{n-l} = -c_n$). This concludes the inductive step.

Since $\sum_{n=1}^{\infty} |c_n|^2 = \int_{\mathbb{T}} |G(z)|^2 dz$, we see that the sequence (e_k) is effective if and only if $\int_{\mathbb{T}} |G(z)|^2 dz = 1$. Also $|G(z)| \leq 1$, therefore this is so if and only if $G(z) = 1$ a.e. on \mathbb{T} . The last condition means that G is an inner function. This is equivalent to the fact that σ is a singular measure with respect to the Lebesgue measure on \mathbb{T} (see [5] where the correspondence between singular measures and inner functions is studied in detail). This concludes the proof of Theorem 2.

COROLLARY 3. *If $(t_k) = (z_0^k)$ for some $z_0 \in \mathbb{T}$ which is not a root of 1 and b is as in Example 1 then the sequence (b_{t_k}) is effective in $L_2(\mathbb{T})$.*

Proof. With $(t_k) = (z_0^k)$ the sequence (b_{t_k}) is stationary and its spectral measure is $\sigma = \sum_{m=-\infty}^{\infty} |\widehat{b}(m)|^2 \delta_{z_0^m}$. This measure is discrete and so it is singular with respect to the Lebesgue measure. Under the given assumption

on b and z_0 the sequence (b_{t_k}) is linearly dense in $L_2(\mathbb{T})$. Thus by Theorem 2 the sequence is effective.

REMARK 2. From the point of view of prediction theory of stationary sequences in Hilbert space, it is interesting to note that the proof of Theorem 2 yields that if a stationary sequence (e_m) , $m \in \mathbb{Z}$, has a singular probability spectral measure then $e_{m+1} = -\sum_{l=-\infty}^m \bar{c}_{m+1-l} e_l$. This is an explicit prediction formula. The formula is valid only for stationary sequences with singular spectral measures. As far as we know, no such formula is known for all predictable stationary sequences.

4. Effective sequences in Banach spaces. Let us prove the following claims made in Example 2:

PROPOSITION 2. *If b and h are as in Examples 1 and 2 and $(t_k) = (z_0^k)$ then the sequence (h_{t_k}, δ_{t_k}) is not effective in $C(\mathbb{T})$. If (t_k) is chosen at random, as in Corollary 1, then a.s. the sequence (h_{t_k}, δ_{t_k}) is not effective in $C(\mathbb{T})$.*

Proof. Let (t_k) be an arbitrary sequence of different points in \mathbb{T} and assume that $K_n(x)$ converges uniformly to x for each $x \in C(\mathbb{T})$. Then by the Banach–Steinhaus Theorem, $\|K_n\| \leq C$ for some constant C and all n . Since $g_n(x) = (x - K_{n-1}(x))(t_n)$ we have $\|g_n\| \leq C + 1$. On the other hand $g_n(x) = \sum_{i=1}^n c_{n,i} x(t_i)$, where the triangular matrix $(c_{n,i})_{n \geq i}$ is the inverse of the triangular matrix $(h(t_n \bar{t}_i))_{n \geq i}$ which we denote by $(a_{n,i})$. Therefore if the sequence (h_{t_k}, δ_{t_k}) is effective in $C(\mathbb{T})$ then $\sum_{i=1}^n |c_{n,i}| = \|g_n\| \leq C + 1$.

Now let $(t_k) = (z_0^k)$ and assume that (h_{t_k}, δ_{t_k}) is effective. Let (c_k) and (a_k) be sequences and F the function defined as in Section 2 for the stationary sequence (b_{z^k}) in $L_2(\mathbb{T})$. We check easily that in this case $c_{n,i} = c_{n-i}$ and $a_{n,i} = a_{n-i} = h(z_0^{n-i})$. Thus $\sum_{n=0}^{\infty} |c_n| \leq C + 1$ for some C . Therefore $1/F$ is continuous on $\mathbb{D} \cup \mathbb{T}$ and hence the same is true about F . But then $\sum_{n=0}^{\infty} |a_n|^2 < \infty$ and this is a contradiction since $(a_k) = (h(t_k))$ does not converge to 0.

The above also proves that if (t_k) is chosen at random, as in Corollary 1, then a.s. the sequence (h_{t_k}, δ_{t_k}) is not effective in $C(\mathbb{T})$. Indeed, assume to the contrary that the probability that it is effective is positive. Because the matrix $(a_{n,i})$ is triangular, for each interval J of positive integers the entries of the inverse matrix $(c_{n,i})$ with indices $n, i \in J$ depend only on the entries $a_{n,i}$ with $n, i \in J$. Fix $z_0 \in \mathbb{T}$ which is not a root of 1. If (t_k) is random, as in Corollary 1, then a.s. we can find arbitrarily large intervals of integers k for which t_k is as close to z_0^k as we wish and thus for n, i in such intervals the terms of the inverse matrices for these two sequences are close. Therefore the sums of absolute values of the rows of the matrix $(c_{n,i})$ are arbitrarily

large because this was shown for the sequence (z_0^k) in the first case. This is a contradiction.

Thus in the above cases we can only hope for “large” subspaces $Y \subset C(\mathbb{T})$ in which the sequence (h_{t_k}, δ_{t_k}) is effective. A construction of such subspaces is provided by the following abstract proposition.

PROPOSITION 3. *Let $i : H \rightarrow X$ be a continuous operator from a Hilbert space into a Banach space with a dense image. Assume that (d_k) is an effective sequence in H and (e_k, f_k) is a sequence in $X \times X^*$ such that $i(d_k) = e_k$ and $i^*(f_k) = d_k$ ($i^* : X^* \rightarrow H$ is the adjoint operator). Then in the Banach space X the sequence (e_k, f_k) is effective in $i(H)$.*

Proof. This is obvious since if (x_n) is an approximating sequence for x corresponding to the sequence (d_k) in H , then $(i(x_n))$ is the approximating sequence for $i(x)$ corresponding to (e_k, f_k) in X .

COROLLARY 4 (cf. Example C of [4]). *Let b, h and (t_k) be as in Examples 1, 2. If the sequence (b_{t_k}) is effective in $L_2(\mathbb{T})$ then in $C(\mathbb{T})$ the sequence (h_{t_k}, δ_{t_k}) is effective in the space $Y = \{x \in C(\mathbb{T}) : \sum_{n=-\infty}^{\infty} |\hat{x}(n)|^2 / \hat{h}(n) < \infty\}$.*

Proof. Let $i : L_2(\mathbb{T}) \rightarrow C(\mathbb{T})$ be the operator defined by

$$i(x)(z) = \sum_{n=-\infty}^{\infty} \hat{x}(n) \bar{\hat{b}}(n) z^n.$$

We check that $i(b_t) = h_t$, $i^*(\delta_t) = b_t$ for each $t \in \mathbb{T}$ and $i(L_2(\mathbb{T})) = Y$. Thus the assertion follows from Proposition 3.

The last and the preceding corollaries can be combined to prove the effectivity of sequences (h_{t_k}, δ_{t_k}) in subspaces $Y \subset C(\mathbb{T})$.

We will conclude the paper with a nonlinear modification of the Kaczmarz algorithm.

Let ε be a fixed positive number and let (e_k, f_k) be a sequence in a Banach space X as at the beginning of Section 1. Given $x \in X$ we define by induction sequences (x_n^ε) of vectors and $(g_k^\varepsilon(x))$ of numbers as follows:

$$x_0 = 0, \quad g_0^\varepsilon = 0 \quad \text{and} \quad x_n^\varepsilon = x_{n-1}^\varepsilon + g_n^\varepsilon(x) e_n$$

where $g_n^\varepsilon(x) = f_n(x - x_{n-1}^\varepsilon)$ if $|f_n(x - x_{n-1}^\varepsilon)| > \varepsilon$ and otherwise $g_n^\varepsilon(x) = 0$.

The functionals g_k^ε and the operators K_n^ε defined by $K_n^\varepsilon(x) = x_n^\varepsilon = \sum_{k=1}^n g_k^\varepsilon(x) e_k$ are nonlinear in X . To formulate our next result we need the following definition. A sequence (f_k) in X^* is said to be *norming* if $\|f_k\| \leq 1$ for all k and $\limsup_{k \rightarrow \infty} |f_k(x)| = \|x\|$ for each $x \in X$. An example of such a sequence is (δ_{t_k}) in $C(\mathbb{T})$ whenever (t_k) is dense in \mathbb{T} . $\#A$ is the number of elements in A .

THEOREM 3. *Let (e_k, f_k) be a sequence in $X \times X^*$ such that (e_k) is linearly dense in X and (f_k) is norming. Assume that for a Hilbert space H , a sequence (d_k) of unit vectors in H and a linear operator $i : H \rightarrow X$ satisfy $i(d_k) = e_k$ and $i^*(f_k) = d_k$ for all k . Then for each $\varepsilon > 0$ and $x \in X$, $\|x - K_n^\varepsilon(x)\| \leq \varepsilon$ for n sufficiently large and*

$$\#\{n : K_n^\varepsilon(x) \neq K_{n-1}^\varepsilon(x)\} \leq \inf\{2\varepsilon^{-2}\|v\|_H^2 : v \in H, \|i(v) - x\| \leq \varepsilon/4\}.$$

Proof. For $x \in X$ let $v \in H$ be such that $\|i(v) - x\| \leq \varepsilon/4$ and let $v_n = \sum_{k=1}^n g_k^\varepsilon(x) d_k$. For each n we have $i(v_n) = x_n^\varepsilon$, $|g_n^\varepsilon(x)| = \|v_n - v_{n-1}\|_H$ and $v_n \neq v_{n-1}$ implies $f_n(v_n) = f_n(x)$. Since $i^*(f_n) = d_n$, for each $w \in H$ we have $f_n(i(w)) = \langle w, d_n \rangle_H$. Therefore we check easily that if $v_n \neq v_{n-1}$ then

$$\begin{aligned} \|v - v_{n-1}\|_H^2 &= \|v - v_n\|_H^2 + \|v_n - v_{n-1}\|_H^2 + 2\operatorname{Re}\langle v - v_n, v_n - v_{n-1} \rangle_H \\ &= \|v - v_n\|_H^2 + \|v_n - v_{n-1}\|_H^2 + 2\operatorname{Re}\langle v - v_n, g_n^\varepsilon(x) d_n \rangle_H \\ &= \|v - v_n\|_H^2 + \|v_n - v_{n-1}\|_H^2 + 2\operatorname{Re}(\overline{g_n^\varepsilon(x)} f_n(i(v - v_n))) \\ &\geq \|v - v_n\|_H^2 + \|v_n - v_{n-1}\|_H^2 - 2\|v_n - v_{n-1}\|_H \|i(v) - x\| \\ &\geq \|v - v_n\|_H^2 + \|v_n - v_{n-1}\|_H^2 - \frac{\varepsilon}{2} \|v_n - v_{n-1}\|_H \end{aligned}$$

(the inequalities hold since $\|f_n\| \leq 1$ and $\|i(v) - x\| \leq \varepsilon/4$). But $v_n \neq v_{n-1}$ implies that $\|v_n - v_{n-1}\|_H \geq \varepsilon$ and hence $\|v_n - v_{n-1}\|_H^2 - (\varepsilon/2)\|v_n - v_{n-1}\|_H \geq \varepsilon^2/2$. Thus

$$\|v - v_{n-1}\|_H^2 \geq \|v - v_n\|_H^2 + \varepsilon^2/2 \quad \text{whenever } v_n \neq v_{n-1}.$$

Summing these inequalities over all n with $v_n \neq v_{n-1}$ we conclude that

$$\|v\|_H^2 \geq \frac{\varepsilon^2}{2} \#\{n : v_n \neq v_{n-1}\} \geq \frac{\varepsilon^2}{2} \#\{n : x_n^\varepsilon \neq x_{n-1}^\varepsilon\}.$$

This proves the second statement of Theorem 3. And it follows that there exists N such that $x_n^\varepsilon = x_N^\varepsilon$ for all $n > N$. Therefore $|f_n(x - x_N^\varepsilon)| < \varepsilon$ for $n > N$. Because (f_k) is norming this proves that $\|x - x_N^\varepsilon\| \leq \varepsilon$. This completes the proof.

REMARK 3. If a sequence (f_k) is norming and a sequence (e_k) is linearly dense in a Banach space X then for the sequence (e_k, f_k) there exist $i : H \rightarrow X$ and (d_k) which satisfy the assumptions of Theorem 3 if and only if the matrix $(a_{n,k}) = (f_n(e_k))$ is positive definite and the entries on the diagonal are equal to 1.

An immediate consequence of Theorem 3 is the following

COROLLARY 5. *If h is as in Example 2, (t_k) is a dense sequence in \mathbb{T} and ε is a fixed positive number then K_n^ε corresponding to (h_{t_k}, δ_{t_k}) has the property that for each $x \in C(\mathbb{T})$ the inequality $\|K_n^\varepsilon(x) - x\| \leq \varepsilon$ holds for all n large enough.*

Proof. To apply Theorem 3 it is enough to take for $i : H \rightarrow X$ the operator defined in the proof of Corollary 4 and to put $(d_k) = (b_{t_k})$.

Acknowledgments. The first author's research for the present article was carried out at the Department of Mathematics of Texas A&M University. He wishes to acknowledge the hospitality and to thank Professors A. G. Poltoratskii and J. Zinn from that Department for numerous illuminating discussions which led to some of the results in the article.

References

- [1] Y. Censor, *Row-action methods for huge and sparse systems and their applications*, SIAM Rev. 23 (1981), 444–466.
- [2] S. Kaczmarz, *Angenäherte Auflösung von Systemen linearer Gleichungen*, Bull. Acad. Polon. Sci. Lett. A 35 (1937), 355–357.
- [3] J. Mycielski, *A learning theory for linear operators*, Proc. Amer. Math. Soc. 103 (1988), 547–550.
- [4] J. Mycielski and S. Świerczkowski, *Uniform approximation with linear combinations of reproducing kernels*, Studia Math. 121 (1996), 105–114.
- [5] D. Sarason, *Sub-Hardy Hilbert Spaces in the Unit Disk*, Wiley, New York, 1994.

Institute of Mathematics
 Warsaw University
 Banacha 2
 02-097 Warszawa, Poland
 E-mail: kwapstan@mimuw.edu.pl

Department of Mathematics
 University of Colorado
 Boulder, CO 80309-0395, U.S.A.
 E-mail: jmyciel@euclid.colorado.edu

Received March 13, 2001
Revised version June 21, 2001

(4705)