RESEARCH

Real Analysis Exchange
Vol. 36(2), 2010/2011, pp. 361–372

Jan Mycielski, Department of Mathematics, University of Colorado Boulder, CO 80309–0395, U.S.A. email: Jan.Mycielski@Colorado.edu

LEARNING THEOREMS

Abstract

We will prove learning theorems that could explain, if only a little, how some organisms generalize information that they get from their senses.

1 Theorems and open problems

Given a metric space M, an infinite sequence $x_0, x_1, \dots$ of points of M and an unknown real valued function $f : M \to \mathbb{R}$, suppose that we have learned a sequence of n data points $(x_0, f(x_0)), \dots, (x_{n-1}, f(x_{n-1}))$. How to predict the value $f(x_n)$? Assuming some regularity of f the simplest way that appears reasonable is the following. Define a function $f_n : M \to \mathbb{R}$ $(n > 0)$ such that for every $x \in M$ we pick the last term x_k of the sequence $x_0, \dots, x_{n-1}$ among those which are the nearest to x and let $f_n(x) = f(x_k)$.

Theorem 0. *If M is compact and f is continuous, then*

$$\lim_{n\to\infty} |f_n(x_n) - f(x_n)| = 0.$$

This easily follows from the facts that f is uniformly continuous and the sets $\{x_0, \dots, x_{n-1}\}$ approximate the set $\{x_0, x_1, \dots\}$ in Hausdorff's distance.

The main purpose of this paper is to prove other theorems of that kind related to the Laws of Large Numbers. We add also a little improvement of a convergence theorem of the Kaczmarz-Agmon Projection Algorithm.

Let λ be a Radon probability measure in the Euclidean space $\mathbb{R}^d$, and $x_0, x_1, \dots$ be independent random variables taking values in $\mathbb{R}^d$ with distribution λ. Let P be the product measure λ^ω in $(\mathbb{R}^d)^\omega$. With $M = \mathbb{R}^d$ and the same definition of f_n as above (thus f_n depends on $(x_0, \dots, x_{n-1})$) we have the following theorem.

Mathematical Reviews subject classification: Primary: 26, 28; Secondary: 41
Key words: laws of large numbers, approximation theory
Received by the editors September 10, 2010
Communicated by: R. Daniel Mauldin

Theorem 1. *If $f : \mathbb{R}^d \to \mathbb{R}$ is λ-measurable then, for every $\varepsilon > 0$,*

$$\lim_{n\to\infty} P(|f_n(x_n) - f(x_n)| < \varepsilon) = 1.$$

Remark. By the theorem of Fubini, *if*

$$P(|f_n(x_n) - f(x_n)| < \varepsilon) \geq 1 - \delta,$$

then

$$P(\lambda\{x : |f_n(x) - f(x)| < \varepsilon\} \geq 1 - \sqrt{\delta}) \geq 1 - \sqrt{\delta}.$$

Theorem 1 will be proved in Section 2. It is a simple consequence of the Besicovitch generalization of the Lebesgue Density Theorem to all Radon measures in $\mathbb{R}^d$. (See P. Mattila [4] and, for related results, S. G. Krantz and T. D. Parsons [3]. That generalization is no longer true for all Radon measures in the infinite dimensional Hilbert space l^2, see P. Mattila and R. D. Mauldin [5].)

It would be interesting to estimate the rate of convergence in Theorem 1; it seems that $f_n \to f$ *in measure* λ *for* P*-almost all sequences* $(x_0, x_1, \dots)$. Recently D. H. Fremlin proved the latter in the case $d = 1$ (see [2, Corollary 3B]) and, for all d, in the case when when λ is the Lebesgue measure in the unit cube $[0,1]^d$ and $\lambda(\mathbb{R}^d - [0,1]^d) = 0$ ([2, Theorem 5B]).

We will show in Section 4 that, even in the case when M is the cube $[0,1]^d$ with the Euclidean metric and λ is the Lebesgue measure in $[0,1]^d$, convergence almost everywhere may fail: For every $0 \leq a < 1$ there exist closed subsets E of $[0,1]^d$ with $\lambda(E) = a$, such that if f is the characteristic function of E then $f_n(x) \to f(x)$ fails P-almost surely for almost all $x \in E$.

Let us return to the general case of a Radon measure λ in $\mathbb{R}^d$. If f is bounded, convergence almost everywhere can be secured by a more sophisticated algorithm:

$$\bar{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x).$$

Theorem 2. *If the function $f : \mathbb{R}^d \to \mathbb{R}$ is λ-measurable and bounded, then for P-almost all sequences $x_0, x_1, \dots$ the sequence $\bar{f}_1, \bar{f}_2, \dots$ converges to f λ-almost everywhere.*

However, the evaluation of $\bar{f}_n$ requires more computation than the evaluation of f_n. Hence some learning organisms may be able to apply f_n but not $\bar{f}_n$.

I do not know if *Theorem 2 can be extended to all λ-integrable f*. However, it fails for some λ-measurable f; see Section 4.

Now we will state variants of Theorems 1 and 2 suggested to the author by Roy O. Davies. Let S be a probability measure space with measure μ and $P = \mu^\omega$. Let $P_0, P_1, \ldots$ be a sequence of finite or countable partitions of S into measurable sets such that $P_0 = \{S\}$ and P_{n+1} is finer than P_n for $n = 0, 1, \ldots$.

We assume that the σ-field F of subsets of S generated by $P_0 \cup P_1 \cup \ldots$ is such that μ is the closure of μ restricted to F, i.e., μ is defined over all subsets of S of the form $A \,\Delta\, N$, where $A \in F$ and N is any subset of a null set of F, such that $\mu(A \,\Delta\, N) = \mu(A)$, where Δ is the symmetric difference of sets.

Then, given a μ-measurable function $f : S \to \mathbb{R}$ and a sequence $x_0, x_1, \ldots$ of points of S, we define functions $f_n : S \to \mathbb{R}$ $(n > 0)$: For every $x \in S$ we choose the least set $Q \in P_0 \cup P_1 \cup \ldots$ such that $x \in Q$ and $\{x_0, \ldots, x_{n-1}\} \cap Q$ is not empty. Then we choose the largest $k < n$ such that $x_k \in Q$, and we put $f_n(x) = f(x_k)$.

Theorem 3. *Theorem 1 is true in this setting.*

Recently D. W. Stroock (to appear in a new edition of [9]) extended Theorems 1 and 3 showing that $E[|f_n(x_n) - f(x_n)|^p] \to 0$, for all $p \geq 1$, under additional assumptions but in a more general setting turning them into *tracking theorems*.

Again we do not know if Theorem 3 can be strengthened to claim that $f_n \to f$ *in measure almost surely*. Again convergence almost everywhere can fail but, if f is bounded, it holds for the first means; in other words:

Theorem 4. *Theorem 2 is true in this setting.*

Again we do not know if Theorem 4 generalizes to all integrable f, but we will show examples of Lebesgue-measurable unbounded functions f for which convergence $\bar{f}_n(x) \to f(x)$ fails almost surely for all x in a set of measure as close to 1 as we wish.

There are essentially only two other learning theorems in the style of the above. The first expresses the convergence of the Kaczmarz Projection Algorithm. For completeness let me state them here with a little refinement in the second one.

Let H be a real Hilbert space, $p \in H$ and $f_0, f_1, \ldots$ a sequence in H such that, for all n, f_{n+1} is an orthogonal projection of f_n into any hyperplane L_n separating f_n from p or containing p.

Theorem 5. *The following inequality holds:*

$$\sum_{n=0}^{\infty} ||f_{n+1} - f_n||^2 \leq ||p - f_0||^2.$$

The proof will be given in Section 5. The next theorem which pertains to the Kaczmarz–Agmon Algorithm is more important in various applications.

Let $f_0 \in H - B$, where B is a non-empty open ball in H. For all n let f_{n+1} be the orthogonal projection of f_n into any hyperplane L_n that separates f_n from B. If, for some n, f_n is on the boundary of B, we let $f_m = f_n$ for all $m > n$.

Theorem 6. *The following inequality holds*

$$\sum_{n=0}^{\infty} ||f_{n+1} - f_n|| < \frac{R^2}{2r} - \frac{r}{2},$$

where R is the distance of f_0 from the center of B and r is the radius of B. Moreover, the right side is the least upper bound of all possible sums on the left.

The proof will be given in Section 5. This improves a similar theorem in [4], where the term $-r/2$ was missing. For related results and applications see [1,6,7,8]. Now we will show the relation of Theorems 5 and 6 to Theorem 0 and illustrate their applications.

Suppose that $f : [0,1] \to \mathbb{R}$ is a continuous function and we seek polynomials f_n of degrees $< m$ with real coefficients uniformly approximating f, such that each f_n is constructed in terms of $(x_0, f(x_0)), \ldots, (x_{n-1}, f(x_{n-1}))$.

To define appropriate algorithms producing such f_n we need the following concepts. Let H_m be the m-dimensional vector space over $\mathbb{R}$ of all polynomials over $\mathbb{R}$ of degree $< m$. We introduce in H_m some Hilbert norm $||\cdot||$. Various choices of $||\cdot||$ are possible, e.g. the Euclidean norm for the vectors of coefficients or the L^2 norm $(\int h^2)^{1/2}$.

For each point $(x, y) \in [0,1] \times \mathbb{R}$, the set $\{h \in H_m : h(x) = y\}$ is a hyperplane in H_m. Hence there exists a unique $h_{x,y} \in H_m$ of minimum norm such that $h_{x,y}(x) = y$.

Of course $h_{x,y} = y h_{x,1}$. Hence

$$|y| \leq c||h_{x,y}||, \tag{1}$$

where $c = 1/\min_{x\in[0,1]} ||h_{x,1}||$.

Let p_f be *the polynomial of best uniform approximation* of f in H_m and $\delta = \max_{x\in[0,1]} |p_f(x) - f(x)|$. Thus δ is the distance in $C[0,1]$ between H_m and f.

Finally for every $\gamma \geq 0$ we define an algorithm A_γ producing the sequence $f_0, f_1, \ldots$: First, $f_0 \in H_m$ is arbitrary (the default choice is $f_0 = 0$). Then, given $f_n \in H_m$ we put

$$f_{n+1} = f_n - h_{x_n, v_n},$$

where

$$v_n = \begin{cases} e_n - \gamma & \text{if that value is positive,} \\ e_n + \gamma & \text{if that value is negative,} \\ 0 & \text{otherwise,} \end{cases}$$

where e_n is *the error* made by f_n at x_n, that is,

$$e_n = f_n(x_n) - f(x_n).$$

Then, if $|e_n| \leq \gamma$, $f_{n+1} = f_n$ and, if $|e_n| > \gamma$, $f_{n+1}(x_n) = f(x_n) \pm \gamma$. Moreover, if $|e_n| > \gamma \geq \delta$, then one of the two hyperplanes $\{h : h(x_n) = f(x_n) \pm \gamma\}$ in H_m separates f_n from p_f. Let L_n be that hyperplane. Then f_{n+1} is the orthogonal projection of f_n into L_n. If $\gamma = \delta$, then $p_f \in L_n$ and Theorem 5 applies; if $\gamma > \delta$, there exists a ball B of positive radius and center p_f in H_m such that L_n separates f_n from B, and Theorem 6 applies.

To explain their significance let $[x]^+ = \max\{0, x\}$. Then by (1), A_γ implies the inequalities

$$[|e_n| - \gamma]^+ = |v_n| \leq c||h_{x_n, v_n}|| = c||f_{n+1} - f_n||.$$

Hence Theorem 5 yields

$$\sum([|e_n| - \gamma]^+)^2 < \infty$$

and Theorem 6 yields

$$\sum[|e_n| - \gamma]^+ < \infty,$$

respectively.

Thus in both cases $\gamma = \delta$ and $\gamma > \delta$, whenever some error $|e_n|$ is larger than γ the algorithm A_γ is learning.

2 Proofs of Theorems 1 and 3

PROOF OF THEOREM 1. For all $\varepsilon > 0$ and all integers k let

$$S_{k\varepsilon} = f^{-1}[k\varepsilon, (k+1)\varepsilon).$$

For all $x \in \mathbb{R}^d$ let $S_\varepsilon(x) = S_{k\varepsilon}$ such that $x \in S_{k\varepsilon}$. By the Besicovitch-Lebesgue Density Theorem (see e.g. [4, Corollary 2.14]), λ-almost every $x \in \mathbb{R}^d$ is a point of λ-density 1 of its set $S_\varepsilon(x)$, that is,

$$\lim_{r \to 0} \frac{\lambda(B(x, r) \cap S_\varepsilon(x))}{\lambda(B(x, r))} = 1,$$

where $B(x,r) = \{y \in \mathbb{R}^d : |y-x| < r\}$ for all $r > 0$.

Let $a_n(x)$ $(n > 0)$ be the last term of $x_0, \dots, x_{n-1}$ among those which are the nearest to x. Thus $f_n(x) = f(a_n(x))$, and for λ-almost all x and λ^ω-almost all $(x_0, x_1, \dots)$ we have $a_n(x) \to x$. Therefore, if x is a point of λ-density 1 in $S_\varepsilon(x)$,

$$\lim_{n\to\infty} \lambda^n\{(x_0, \dots, x_{n-1}) : a_n(x) \in S_\varepsilon(x)\} = 1.$$

Hence, for λ-almost all x,

$$\lim_{n\to\infty} \lambda^n\{(x_0, \dots, x_{n-1}) : |f_n(x) - f(x)| < \varepsilon\} = 1.$$

Choose any $\delta > 0$. Then for all large enough n,

$$\lambda\{x : \lambda^n\{(x_0, \dots, x_{n-1}) : |f_n(x) - f(x)| < \varepsilon\} > 1-\delta\} > 1-\delta.$$

And, since the random variables x_n and $(x_0, \dots, x_{n-1})$ are independent, for all large enough n, we have

$$\lambda^{n+1}\{(x_0, \dots, x_n) : |f_n(x_n) - f(x_n)| < \varepsilon\} > (1-\delta)^2.$$

Of course this inequality implies Theorem 1. □

PROOF OF THEOREM 3. We can assume without loss of generality that

1° $\mu(A) > 0$ for all $A \in P_0 \cup P_1 \cup \dots$

2° If $\mu(\bigcap_{n=0}^\infty A_n) > 0$, then $\bigcap_{n=0}^\infty A_n$ is of power continuum.

It follows that the system $(S, \mu, P_0, P_1, \dots)$ can be identified with a system of the form $([0,1), \lambda, P_0', P_1', \dots)$ where λ is the Lebesgue measure and all $A' \in P_0' \cup P_1' \cup \dots$ are intervals of the form $[a,b)$ with $0 \le a < b \le 1$, and $b - a = \mu(A)$ if A' corresponds to A.

Let now, for all $x \in S$, $A_n(x)$ be the unique $A \in P_n$ such that $x \in A$. Since the corresponding A' are intervals, the original Lebesgue Density Theorem easily implies

(LDT') *For every μ-measurable set $X \subseteq S$ almost all $x \in X$ have the property*

$$\lim_{n\to\infty} \frac{\mu(X \cap A_n(x))}{\mu(A_n(x))} = 1.$$

Using (LDT') instead of the original theorem of Lebesgue, the proof of Theorem 3 is quite similar to the proof of Theorem 1. So we omit further details. □

Remark. For a generalization of (LDT') due to J. Marcinkiewicz see [9, Theorem 5.2.12].

3 Proofs of Theorems 2 and 4

PROOF OF THEOREM 2. By the theorem of Fubini it suffices to show that, for λ-almost all $x \in \mathbb{R}^d$, for λ^ω-almost all $(x_0, x_1, \dots) \in (\mathbb{R}^d)^\omega$, $\bar{f}_n(x) \to f(x)$, as $n \to \infty$.

We use the notation of Section 2. For every $\varepsilon > 0$, λ-almost every $x \in \mathbb{R}^d$ is a point of λ-density 1 of $S_\varepsilon(x)$. Hence not only do we have that for λ-almost all x, $a_n(x) \to x$, for λ^ω-almost all sequences $x_0, x_1, \dots$, but also the frequency of those terms $a_i(x)$ in the sequence $a_1(x), a_2(x), \dots$ that belong to $S_\varepsilon(x)$ equals 1. By the assumption of Theorem 2, $|f| < A$ for some constant A. Hence there exists a k such that

$$k\varepsilon \le \liminf_{n\to\infty} \bar{f}_n(x) \le \limsup_{n\to\infty} \bar{f}_n(x) \le (k+1)\varepsilon.$$

Therefore $\lim \bar{f}_n(x)$ exists and equals $f(x)$ for almost all x and $(x_0, x_1, \dots)$. □

PROOF OF THEOREM 4. We use (LDT') and the proof is almost the same as the above, so I omit the details. □

4 Counter-examples to convergence almost everywhere

We will show that in Theorems 1 and 3 convergence $f_n \to f$ almost everywhere may fail almost surely (however we do not know if convergence in measure must hold almost surely), and that in Theorems 2 and 4, the assumption that f is bounded cannot be omitted (however we do not know if it cannot be replaced by the assumption that f is integrable).

In this section λ denotes the Lebesgue measure restricted to the unit cube I^d ($I = [0,1]$) and $P = \lambda^\omega$. For any set S, $|S|$ denotes the cardinality of S.

We begin with the construction of a counter-example related to Theorem 1 with divergence over a set of measure a, for any desired $a < 1$.

We will need the following propositions.

Proposition 1. *For every closed set $A \subset I^d$ with $\lambda(A) < 1$ and every $\varepsilon, \rho > 0$ with $\varepsilon < 1 - \lambda(A)$, there exists an open set $V \subset I^d - A$ with $\lambda(V) = \varepsilon$ and $\lambda(\partial V) = 0$ (∂V denotes the boundary of V) such that for all $x \in I^d - A$ and all $r > \rho$*

$$\frac{\lambda(B(x,r) \cap V)}{\lambda(B(x,r))} > \frac{\varepsilon}{2}.$$

PROOF. Let n be a natural number and C the set of cubes disjoint with A of a partition of I^d into n^d congruent cubes. Let n be large enough such that the union of the cubes of C is of measure $> \varepsilon$. Let V be the union of $|C|$ open

sets with boundaries of measure 0, each of measure $\varepsilon/|C|$ located in distinct cubes of C. It is clear that if n is large enough, V satisfies Proposition 1. □

Let $a_1(x), a_2(x), \dots$ be defined as in Section 2 and $V(A, \varepsilon, \rho)$ be the open set given by Proposition 1.

Proposition 2. *For every $x \in I^d - A$, $\varepsilon > 0$ and any natural number N,*

$$\lim_{\rho \to 0} P(|\{a_1(x), a_2(x), \dots\} \cap V(A, \varepsilon, \rho)| > N) = 1.$$

PROOF. Recall three facts:

1° $\lambda(V(A, \varepsilon, \rho)) = \varepsilon$ and hence it does not depend on ρ.

2° $V(A, \varepsilon, \rho)$ is more and more evenly spread in $I^d - A$ as $\rho \to 0$.

3° The points $x_0, x_1, \dots$ are chosen uniformly and independently in I^d.

It is evident that Proposition 2 follows from these facts. □

Let $\alpha = 1 - a$. Now we define recursively three sequences $A_0, A_1, \dots; \rho_0 > \rho_1 > \dots$; and $n(0), n(1), \dots$. Let $A_0 = \emptyset$, $\rho_0 = 1$ and $n(0) = 0$. Assume that A_k, ρ_k and $n(k)$ are given. Then, by Proposition 2, there exists a $\rho > 0$ and an integer $n > n(k)$ such that for all $x \in I^d - A_k$

$$P\left(\{a_{n(k)}(x), a_{n(k)+1}(x), \dots, a_{n-1}(x)\} \cap V\left(A_k, \frac{\alpha}{2^{k+1}}, \rho\right) \neq \emptyset\right) \geq 1 - \frac{1}{k^2}. \quad (2)$$

Let ρ_{k+1} be any such $\rho < \rho_k$ and $n(k+1)$ the least corresponding n. Finally we define

$$A_{k+1} = A_k \cup \overline{V}\left(A_k, \frac{\alpha}{2^{k+1}}, \rho_{k+1}\right),$$

where $\overline{V}$ denotes the closure of V, and

$$A^* = \bigcup_{k=0}^{\infty} A_k.$$

Let $\mathrm{Int}(A^*)$ denote the interior of A^*. Then, by Proposition 1 , $\lambda(\mathrm{Int}(A^*)) = \lambda(A^*)$.

Proposition 3. *$\lambda(A^*) = \alpha$ and for every $x \in I^d - A^*$ almost surely $x_0, x_1, \dots$ is such that for all large enough k*

$$\{a_{n(k)}(x), \dots, a_{n(k+1)-1}(x)\} \cap V\left(A_k, \frac{\alpha}{2^{k+1}}, \rho_{k+1}\right) \neq \emptyset.$$

PROOF.

$$\lambda(A^*) = \sum_{k=0}^{\infty} \lambda\left(V\left(A_k, \frac{\alpha}{2^{k+1}}, \rho_{k+1}\right)\right) = \sum_{k=0}^{\infty} \frac{\alpha}{2^{k+1}} = \alpha.$$

Since $\lim_{n\to\infty} \prod_{k>n} \left(1 - \frac{1}{k^2}\right) = 1$, by (2) we get the second part of Proposition 3. □

Let now f be the characteristic function of the interior of A^* (denoted $\mathrm{Int}(A^*)$). Thus for every x that is of density 1 in $I^d - A^*$, almost surely the sequence $f(a_1(x))$, $f(a_2(x)), \ldots$ diverges (since it contains almost surely infinitely many 0's and, by Proposition 3, infinitely many 1's). Since for every n, $f_n(x) = f(a_n(x))$, the sequence $f_1(x), f_2(x), \ldots$ also diverges. Since $\lambda(I^d - \mathrm{Int}(A^*)) = 1 - \alpha = a$ this concludes our construction.

Now we will construct an example related to Theorem 2, namely a measurable (but unbounded) $f : I^d \to \mathbb{R}$ such that $\bar{f}_n \to f$ almost everywhere fails almost surely.

Let $f(x) = 0$ if $x \in I^d - \mathrm{Int}(A^*)$ and $f(x) = c_k$ if $x \in V\left(A_k, \frac{\alpha}{2^{k+1}}, \rho_{k+1}\right)$. It is clear that if the constants c_k grow sufficiently fast, then by Proposition 3, for all $x \in I^d - \mathrm{Int}(A^*)$, almost surely the means $\bar{f}_n(x)$ will not converge to 0. For example if $c_k \geq n(k)$ then almost surely, for all large enough k,

$$\bar{f}_{n(k)}(x) = \frac{1}{n(k)} \sum_{i=1}^{n(k)} f(a_i(x)) \geq 1.$$

The examples concerning Theorems 3 and 4 are quite similar so we omit the details.

5 Proofs of Theorems 5 and 6 (In Outline)

PROOF OF THEOREM 5. By the definition of f_{n+1} there is a hyperplane L_n which contains p or separates f_n from p such that f_{n+1} is the orthogonal projection of f_n into L_n. Then there exists also a hyperplane L'_n such that $p \in L'_n$ and, if f'_{n+1} denotes the orthogonal projection of f_n into L'_n, then

$$||f'_{n+1} - p|| = ||f_{n+1} - p|| \tag{3}$$

and

$$||f'_{n+1} - f_n|| \geq ||f_{n+1} - f_n||. \tag{4}$$

In order to find such an L'_n it suffices to consider a hyperplane P parallel to L_n and containing p, and then to rotate P around p toward f_n such that the projection f'_{n+1} of f_n into that rotated P satisfies (3). Then the inequality (4) is automatically satisfied.

By (3) and (4) there exists a sequence $f_0 = f'_0, f'_1, f'_2, \dots$ and a sequence of hyperplanes $L'_1, L'_2, \dots$ all containing p, such that f'_{n+1} is the projection of f'_n into L'_n and

$$||f_{n+1} - f_n|| \leq ||f'_{n+1} - f'_n||$$

for all n.

Hence, to prove Theorem 5, we can assume without loss of generality that $p \in L_n$ for all n.

With that condition Theorem 5 reduces immediately to the case of a 2-dimensional H and it follows easily from the Pythagorean Theorem. □

PROOF OF THEOREM 6. By an argument similar to the above one we can assume without loss of generality that all hyperplanes L_n (separating f_n from B) are tangent to B. And again the problem reduces to the case of a 2-dimensional H.

Then by elementary geometric considerations the least upper bound of the sums equals the length of a certain spiral S on the plane $\mathbb{R}^2$ which is defined as follows. We take a straight segment of length $A = \sqrt{R^2 - r^2}$ and place it on $\mathbb{R}^2$ such that it is tangent at one of its ends to a circle C of radius r. Then we think of this segment as a flexible thread and wind it completely around C such that it is always tangent to C. The free end of the thread traces the required spiral S. Thus the far end of S is at the distance R from the center of C, the near end of S at the distance r from this center, and it is easy to see that the length of S is

$$\int_0^{A/r} (A - r\alpha) d\alpha = \frac{R^2}{2r} - \frac{r}{2}.$$

This yields Theorem 6. □

Acknowledgements. I am indebted to D. H. Fremlin for suggesting the generalization of Theorems 1 and 2 to Radon measures and to D. W. Stroock and the referee for insisting that I clarify a former version of the proof of Theorem 1.

References

[1] V. Faber and J. Mycielski, *Applications of learning theorems*, Fund. Inform., **15** (1991), 145–167.

[2] D. H. Fremlin, available at:
http://www.essex.ac.uk/maths/people/fremlin/probGO.pdf.

[3] S. G. Krantz and T. D. Parsons, *Antisocial subcovers of self-centered coverings*, Amer. Math. Monthly, **96(1)** 1986, 45–48.

[4] P. Mattila, *Geometry of Sets and Measures in Euclidean Spaces*, Cambridge Univ. Press, 1995.

[5] P. Mattila and R. D. Mauldin *Measure and dimension functions: Measurability and densities*, Math. Proc. Cambridge Phil. Soc., **121**(1997), 81–100.

[6] J. Mycielski, *Can mathematics explain natural intelligence*, Physica, **22D** (1986), 366–375.

[7] J. Mycielski, *A learning theorem for linear operators*, Proc. Amer. Math. Soc., **103** (1988), 547–550.

[8] J. Mycielski and S. Świerczkowski, *A theory of the neocortex*, Adv. in Appl. Math., **9** (1988), 465–480.

[9] D. W. Stroock, *Probability Theory, an Analytic View*, Cambridge Univ. Press (new edition, to appear).