JOURNAL OF APPROXIMATION THEORY 1, 189–198 (1968)

Entropic Determination of a Metric Nucleus

JAY E. ISRAEL

The University of Western Australia, Nedlands, Western Australia 6009, Australia

A subset of a metric space can have a nucleus, a pre-eminent element at which the mass of the set is concentrated. This notion, introduced by G. Strang in [6], arises from the examination of the structure of a set using ϵ-entropy. The latter, a discrete quantification of the size of a set, first arose in the study of tabulation and information storage (see [5] and [7]).

The nucleus concept applies it to a different problem, that of locating a distinguished element of a set.

Seeking a nucleus can be viewed as examining the structure of a set (say, the set of functions satisfying certain constraints, or the set of admissible functions in some problem) to find a most characteristic element, as contrasted with choosing that member of a special subset minimizing a residual. The nucleus has interesting analytic, probabilistic, and variational properties, some of which are presented in this paper. In addition, specific examples of nuclei in function spaces are established here.

PRELIMINARIES

If A is a subset of a metric space, $N_\epsilon(A)$ is defined to be the smallest number of elements in any covering of A by sets of diameter $\leqslant 2\epsilon$. This quantity is studied in detail in [5], and is known in connection with information theory [7]. As $\epsilon \to 0$, it is taken as a comparative estimate of the size of A. When $N_\epsilon(A)$ is finite for all $\epsilon > 0$, A is called *totally bounded.* Customarily, $\log_2 N_\epsilon(A)$ is called the *ϵ-entropy* of A.

DEFINITION. Let A be a non-void, totally bounded subset of a metric space. Then x is the *nucleus* of A if, for every neighbourhood, V, of x,

$$N_\epsilon(V \cap A) \sim N_\epsilon(A) \qquad (\epsilon \to 0).$$

In heuristic terms, x is the nucleus of A if every neighbourhood of x in A is asymptotically as large as all of A.

It is not difficult to prove ([4], pp. 13–14) that no set has more than one nucleus. What is interesting is that any set should have so unusual a structure.

The simplest example is that of convergent sequences: if A is a sequence $\{a_i\}_{i=1}^{\infty}$ of distinct elements, and if $\lim_{i\to\infty} a_i = x$, then x is the nucleus of A. This is true since for any fixed neighbourhood, V, of x,

$$N_\epsilon(A \cap V) \leqslant N_\epsilon(A) \leqslant N_\epsilon(A - V) + N_\epsilon(A \cap V),$$

and $N_\epsilon(A - V)$ remains bounded while $N_\epsilon(A \cap V)$ does not. The converse of this theorem does not hold, however—there exist non-convergent sequences which have nuclei ([4], pp. 15–17).

Clearly, it is enough to consider neighbourhoods V of the form

$$S(x,r) = \{z \mid d(x,z) < r\},$$

(where d denotes the metric). Furthermore, since N_ϵ assumes only integer values and is a right-continuous function of ϵ, any countable set of values of ϵ can be discarded in proving $N_\epsilon(A \cap V) \sim N_\epsilon(A)$. These facts will be used in the proofs to follow.

Prescribed Ordinates

Fix a finite real interval $[a,b]$, and abscissae $\{t_i\}_{i=0}^{n}$ with $a = t_0 < t_1 < \ldots < t_n = b$. Let $\{x_i\}_{i=0}^{n}$ be real numbers with $|x_{i+1} - x_i| \leqslant L|t_{i+1} - t_i|$, where $L > 0$ is fixed. Denote by A the set of real-valued functions, f, on $[a,b]$, satisfying a Lipschitz condition with coefficient L, and for which $f(t_i) = x_i$ $(i = 0,1,\ldots,n)$. That is,

$$A = \{f\colon [a,b] \to R \mid f(t_i) = x_i \ (i = 0,1,\ldots,n) \text{ and } \forall\, x,y,\ |f(x) - f(y)| \leqslant L|x-y|\}.$$

The metric is $d(f,g) = \sup_t |f(t) - g(t)|$. A has as a nucleus the broken line function connecting the points (t_i,x_i) $(i = 0,1,\ldots,n)$. The proof of this fact is a straightforward extension of that for the case $n = 1$, which will be presented here. Now since translation is an isometry, it is enough to consider the case $t_0 = x_0 = 0$. Likewise, the change of independent variable $\tau = Lt$ maps the set A isometrically onto a similar set in which the Lipschitz coefficient is 1. Consequently, only the case $L = 1$ need be considered.

Theorem 1. *Suppose $t_1 > 0$ and $|x_1| < t_1$. Let A be the set of real-valued functions on $[0,t_1]$ which satisfy a Lipschitz condition with coefficient 1 and connect the origin to the point (t_1,x_1). Then A has as a nucleus the function $y(t) = (x_1/t_1)t$.*

Proof. The proof consists of two parts, a construction and an analysis.

The construction resembles that of [5] and produces an optimal covering of A by $\{\bar{S}(\phi,\epsilon)\}_{\phi\in\Phi_\epsilon}$. (Overscore denotes topological closure.) Φ_ϵ is the finite set of functions to be constructed. Start with the set Φ characterized by the conditions that $\phi \in \Phi$ if

(a) $\phi(t) = 0$ for $0 \leqslant t \leqslant \epsilon$, and
(b) on each of the intervals $(i-1)\epsilon \leqslant t \leqslant i\epsilon$ $(i = 1,2,\ldots,[t_1/\epsilon])$ and $[t_1/\epsilon]\epsilon \leqslant t \leqslant t_1$, ϕ is linear with slope either $+1$ or -1. Discard any value of ϵ dividing t_1 exactly.

The members of Φ trace out a diamond pattern in the part of the $t-x$ plane under scrutiny. Each diamond can be divided into four subdiamonds by joining the midpoints of its opposite sides. The construction of Φ_ϵ depends on which subdiamond contains the point $P = (t_1,x_1)$. Start with the case when (t_1,x_1) falls in the leftmost subdiamond of some diamond. Suppose the leftmost vertex of that diamond is $(p_\epsilon \epsilon, k_\epsilon \epsilon) = Q$. Then in order for ϕ to be in Φ_ϵ,

(a) on $0 \leqslant t \leqslant p_\epsilon \epsilon$, ϕ must agree with a member of Φ, and
(b) on $p_\epsilon \epsilon \leqslant t \leqslant t_1$, ϕ must follow the straight line from Q to P.

If P falls in a different subdiamond, the above construction is applied after a translation of the entire grid moves P into the leftmost subdiamond of the diamond containing it. In the process the mesh point at $(\epsilon,0)$ is moved to either $(0,0)$, $(\epsilon/2,\epsilon/2)$, or $(\epsilon/2,-\epsilon/2)$, and remains connected to the origin by a straight line. Values of ϵ for which P falls on a boundary between subdiamonds are discarded. An extension of Kolmogorov's proof ([5], pp. 285–288) can be applied in all four of these cases to show that $A \subset \bigcup_{\phi \in \Phi_\epsilon} \bar{S}(\phi,\epsilon)$. Thus, $N_\epsilon(A) \leqslant \operatorname{card}(\Phi_\epsilon)$.

For a lower bound on $N_\epsilon(A)$, choose, for each $\epsilon > 0$ still in consideration, an $\epsilon' > \epsilon$ so small that the point P remains in the same subdiamond as initially. Clearly, $\Phi_{\epsilon'}$ is 2ϵ-separated; i.e., $d(\phi,\psi) > 2\epsilon$ for all ϕ, $\psi \in \Phi_{\epsilon'}$ with $\phi \neq \psi$. Furthermore, since the configuration in the $t-x$ plane is the same for $\Phi_{\epsilon'}$ as for Φ_ϵ, $\operatorname{card}(\Phi_{\epsilon'}) = \operatorname{card}(\Phi_\epsilon)$. Now if B is a 2ϵ-separated subset of A, then $N_\epsilon(A) \geqslant \operatorname{card}(B)$, for otherwise some set of a covering by sets of diameter $\leqslant 2\epsilon$ contains more than one member of B. Consequently, $N_\epsilon(A) \geqslant \operatorname{card}(\Phi_{\epsilon'}) = \operatorname{card}(\Phi_\epsilon)$ and, in fact, $N_\epsilon(A) = \operatorname{card}(\Phi_\epsilon)$.

For the analysis part of the proof, let $r > 0$ be arbitrary and take $V = S(y,r)$. We must show that $N_\epsilon(A \cap V) \sim N_\epsilon(A)$ $(\epsilon \to 0)$. Construct Φ_ϵ as before, this time excluding, in addition, every value of ϵ for which any of the mesh points $(i\epsilon, j\epsilon)$ $(i, j$ integers) fall on either of the boundary lines $x = (x_1/t_1)t \pm r$. If $\epsilon' > \epsilon$ is now chosen smaller than before and so small that none of the mesh points between the boundary lines reach these lines, $\Phi_{\epsilon'} \cap V$ is a 2ϵ-separated subset of $A \cap V$, so that the reasoning above yields

$$N_\epsilon(A \cap V) \geqslant \operatorname{card}(\Phi_{\epsilon'} \cap V) = \operatorname{card}(\Phi_\epsilon \cap V).$$

We now have

$$1 \geqslant \frac{N_\epsilon(A \cap V)}{N_\epsilon(A)} \geqslant \frac{\operatorname{card}(\Phi_\epsilon \cap V)}{\operatorname{card}(\Phi_\epsilon)}.$$

But, assigning equal probabilities to each $\phi \in \Phi_\epsilon$, this last ratio is just $\operatorname{Prob}\{\phi \in V\} = \operatorname{Prob}\{d(\phi, y) < r\}$. So it is enough to show that

$$\lim_{\epsilon \to 0} \operatorname{Prob}\{\phi \in V\} = 1.$$

For this, define, for each $t \in [0, t_1]$ and each ϵ still under consideration, a random variable $W(t, \epsilon)$ by evaluating at t a randomly chosen $\phi \in \Phi_\epsilon$. This process can be thought of as a constrained random walk with time and space increments ϵ. The relationships governing sums of random variables can be applied as in [3] (pp. 218–219) to obtain

$$\operatorname{mean}\{W(t,\epsilon)\} = (x_1/t_1)\,t + 0(\epsilon) \quad \text{and}$$

$$\text{variance} \quad \{W(t,\epsilon)\} \leqslant 12 t_1 \epsilon \tag{1}$$

(assuming $\epsilon \leqslant \frac{1}{2}t_1$ which is clearly sufficient).

Choosing ϵ so small that the $0(\epsilon)$ term in (1) is less than $\frac{1}{2}r$, and applying Chebyshev's inequality ([3], pp. 218–219), we have

$$\begin{aligned}\operatorname{Prob}\{|W(t,\epsilon) - (x_1/t_1)\,t| < r\} &\geqslant \operatorname{Prob}\{|W(t,\epsilon) - \operatorname{mean}\{W(t,\epsilon)\}| < \tfrac{1}{2}r\} \\ &\geqslant 1 - 36 r^{-1} t_1 \epsilon.\end{aligned}$$

Since the right side does not depend on t,

$$\operatorname{Prob}\{\max_{0 \leqslant t \leqslant t_1} |W(t,\epsilon) - (x_1/t_1)\,t| < r\} \geqslant 1 - 36 r^{-1} t_1 \epsilon$$

or

$$1 \geqslant \operatorname{Prob}\{d(\phi, y) < r\} \geqslant 1 - 36 r^{-1} t_1 \epsilon \to 1 \qquad (\epsilon \to 0)$$

as required. Q.E.D.

The problem in which more than two ordinates are prescribed ($n > 1$) is handled by repeating the constructions of the theorem for each of the intervals $[t_{i-1}, t_i]$ $(i = 1, 2, \ldots, n)$.

Sum Constraints

In the same setting as before, we seek the nucleus of a set of functions constrained by a condition of the form $\sum_{i=1}^{n} c_i f(t_i) = 1$, where the c_i are fixed constants and the t_i are prescribed abscissae. In order to convey the basic approach as clearly as possible, we first present the case $n = 2$.

Suppose c_1, c_2 are constants with $c_2 \neq 0$, and suppose $0 < t_1 < t_2 = T$.

Using the uniform metric as before, define A to be the set of functions $f:[0,T]\to R$ such that

(a) f satisfies a Lipschitz condition with coefficient 1,
(b) $f(0)=0$, and
(c) $c_1 f(t_1)+c_2 f(t_2)=1$.

The choice of parameters is assumed to be such that A is non-void. The ordinate, x_1, at t_1 is to be thought of as an independent variable, with the ordinate, x_2, at t_2 depending on it according to the relation

$$x_2=x_2(x_1)=c_2^{-1}(1-c_1x_1).$$

THEOREM 2. *The nucleus of A is the function*

$$y(t)=\begin{cases}(\theta/t_1)\,t & \text{for } 0\leqslant t\leqslant t_1\\ \theta+(c_2^{-1}-c\theta)(t_2-t_1)^{-1}(t-t_1) & \text{for } t_1\leqslant t\leqslant t_2\end{cases}$$

where $c=1+c_1c_2^{-1}$ and where θ is the unique solution of

$$\frac{t_1+\theta}{t_1-\theta}\left(\frac{t_2-t_1-x_2(\theta)+\theta}{t_2-t_1+x_2(\theta)-\theta}\right)^c=1 \tag{2}$$

in the interval $I=\{f(t_1)\mid f\in A\}$.

Writing ln *for the natural logarithm, define*

$$\begin{aligned}B(x_1)&=(t_1+x_1)\ln(t_1+x_1)+(t_1-x_1)\ln(t_1-x_1)\\&\quad+(t_2-t_1+x_2-x_1)\ln(t_2-t_1+x_2-x_1)\\&\quad+(t_2-t_1-x_2+x_1)\ln(t_2-t_1-x_2+x_1).\end{aligned}$$

The proof of the theorem follows from a number of propositions, which we present first.

PROPOSITION 1. *The absolute minimum of $B(x_1)$ for $x_1\in I$ occurs when and only when $x_1=\theta$, as characterized in* (2). *Furthermore, θ is interior to I.*

Proof. Since B is continuous on I (it is extended to the end points of I by L'Hospital's rule) it attains its minimum there. Differentiating,

$$B'(x_1)=\ln\left\{\frac{t_1+x_1}{t_1-x_1}\left(\frac{t_2-t_1-x_2+x_1}{t_2-t_1+x_2-x_1}\right)^c\right\}$$

and

$$B''(x_1)=\frac{2t_1}{t_1^2-x_1^2}+\frac{2c^2(t_2-t_1)}{(t_2-t_1)^2-(x_2-x_1)^2}.$$

By the Lipschitz condition, $B''>0$ and $B'(x_1)$ becomes infinite as x_1 approaches an end point of I. Therefore, the minimum of B must occur in the interior of

I. Since $B'' > 0$, there can be only one relative minimum (call it θ). The characterization (2) is obtained by setting $B'(\theta) = 0$. Q.E.D.

PROPOSITION 2. *Suppose that $F(x_1)$ is defined and integrable for $x_1 \in I$, that it is continuous in some neighbourhood of θ, and that $F(\theta) \neq 0$. Then*

$$\int_I F(x_1)\, e^{-B(x_1)/2\epsilon}\, dx_1 \sim F(\theta)\, e^{-B(\theta)/2\epsilon} \left\{\frac{4\pi\epsilon}{B''(\theta)}\right\}^{1/2} \qquad (\epsilon \to 0).$$

Proof. In view of Proposition 1, this follows from the standard result on the method of Laplace for integrals (see [2], p. 63). Q.E.D.

For values of ϵ dividing neither t_1 nor $t_2 - t_1$ evenly, define

$$R(x_1) = (\{t_1{}^2 - x_1{}^2\}\{(t_2 - t_1)^2 - (x_2(x_1) - x_1)^2\})^{-1/2}.$$

Also define

$$g_i(j) = (2j\epsilon)^i\, R(2j\epsilon)\, e^{-B(2j\epsilon)/2\epsilon} \qquad (i = 0, 1, 2)$$

for all integers j such that $2j\epsilon$ is interior to I. The set of all these integers j will be denoted by J, and its smallest and largest elements by α and β, respectively.

PROPOSITION 3. *If $c_1 \neq -c_2$, then*

$$\sum_{j \in J} g_i(j) \sim \int_\alpha^\beta g_i(u)\, du \qquad (\epsilon \to 0).$$

Proof. From the Euler–Maclaurin sum formula,

$$\sum_{j \in J} g_i(j) = \int_\beta^\alpha g_i(u)\, du + \tfrac{1}{2}\{g_i(\alpha) + g_i(\beta)\} + \int_\alpha^\beta (x - [x] - \tfrac{1}{2})\, g_i'(u)\, du.$$

All the terms on the right except the first can, by application of Proposition 2, be shown to be $o(\int_\alpha^\beta g_i(u)du)$. The assumption $c_1 \neq -c_2$ is necessary to insure that the F of Proposition 2 does not vanish at θ. Q.E.D.

Define

$$K(\epsilon) = (2\epsilon)^{t_2/\epsilon}\, (2\pi\epsilon)^{-1}\, t_1^{t_1/\epsilon + 1/2}\, (t_2 - t_1)^{(t_2 - t_1)/\epsilon + 1/2}$$

and $f_i(j, \epsilon) = K(\epsilon)\, g_i(j)$. Then for $c_1 \neq -c_2$ we have just shown

$$\sum_{j \in J} f_i(j, \epsilon) \sim \theta^i\, K(\epsilon) \sqrt{\pi\epsilon}\, R(\theta)\, B''(\theta)^{-1/2}\, e^{-B(\theta)/2\epsilon} \qquad (\epsilon \to 0). \tag{3}$$

Now for $x_1 = 2j\epsilon$ ($j \in J$), join the origin to the point (t_1, x_1) as in the construction of Theorem 1, and similarly join (t_1, x_1) to $(t_2, x_2(x_1))$ to form $\Phi_\epsilon \subset A$. As before, $\{\bar{S}(\phi, \epsilon)\}_{\phi \in \Phi_\epsilon}$ covers A (cf. [5], p. 289, [7], pp. 65–68).

Let $\nu_0\ (j,\epsilon)$ be the number of members of Φ_ϵ passing through $(t_1, 2j\epsilon)$, and, for $i=1,\ 2$, let $\nu_i(j,\epsilon)=(2j\epsilon)^i\nu_0(j,\epsilon)$. A straightforward calculation yields

$$\nu_i(j,\epsilon)=(2j\epsilon)^i\binom{2n_1}{n_1+j}\binom{2(n_2-n_1)}{n_2-n_1+j-\left[\dfrac{1-2c_1\epsilon j}{2c_2\epsilon}\right]}\qquad (i=0,1,2)\quad (4)$$

where $n_k=[t_k/2\epsilon]$ for $k=1,2$.

PROPOSITION 4. *If $c_1\neq -c_2$ then, for $i=0,1,2$,*

$$\sum_{j\in J}\nu_i(j,\epsilon)\sim\sum_{j\in J}f_i(j,\epsilon)\qquad (\epsilon\to 0).$$

Proof. Stirling's formula gives $\nu_i(j,\epsilon)\sim f_i(j,\epsilon)$. For $i=0$ and $i=2$, a procedure similar to that found in [*8*], pp. 142, 265, is employed to extend this asymptotic result to the sum. For $i=1$, the extension follows from the case $i=0$, and from the result $\sum f_1(j,\epsilon)\sim\theta\sum f_0(j,\epsilon)$, a consequence of (3). Q.E.D.

Proof of Theorem 2.

Case 1: $c_1\neq -c_2$.

Using the same reasoning as in Theorem 1, we have

$$1\geqslant\frac{N_\epsilon(A\cap V)}{N_\epsilon(A)}\geqslant\frac{\operatorname{card}(\Phi_\epsilon\cap V)}{\operatorname{card}(\Phi_\epsilon)}\qquad (5)$$

where $V=S(y,r)$ with $r>0$ fixed.

Fix $\delta>0$. Assigning equal probabilities to the elements of Φ_ϵ, let $W(t,\epsilon)$ be the random variable resulting from evaluation of a random member of Φ_ϵ at t. From (3) and Proposition 4,

$$\text{mean } W(t_1,\epsilon)=\frac{\sum_{j\in J}\nu_1(j,\epsilon)}{\sum_{j\in J}\nu_0(j,\epsilon)}\sim\theta\qquad (\epsilon\to 0).$$

Since the right side is constant, we have, in fact, $\lim_{\epsilon\to 0}\text{mean } W(t_1,\epsilon)=\theta$. Next,

$$\text{var } W(t_1,\epsilon)=\frac{\sum(2j\epsilon-\text{mean } W(t_1,\epsilon))^2\nu_0(j,\epsilon)}{\sum\nu_0(j,\epsilon)}\to\frac{\sum\nu_2(j,\epsilon)}{\sum\nu_0(j,\epsilon)}-\theta^2.$$

But since

$$\lim_{\epsilon\to 0}\frac{\sum\nu_2(j,\epsilon)}{\sum\nu_0(j,\epsilon)}=\theta^2$$

by (3) and Proposition 4, we have

$$\lim_{\epsilon\to 0}\text{var } W(t_1,\epsilon)=0.$$

Now let $\Psi_\epsilon = \{\psi \in \Phi_\epsilon | \theta - \frac{1}{2}r < \psi(t_1) < \theta + \frac{1}{2}r\}$ and $\Psi_\epsilon^j = \{\psi \in \Psi_\epsilon | \psi(t_1) = 2j\epsilon\}$. Using Chebyshev's inequality and the results of the last paragraph, there is an $\epsilon_1 > 0$ such that $0 < \epsilon < \epsilon_1$ implies card Ψ_ϵ/card $\Phi_\epsilon \geqslant \sqrt{1-\delta}$.

Note that the estimate of the variance in (1) does not depend on x_1. This means that there is an ϵ_2 independent of j such that $\epsilon < \epsilon_2$ implies $\text{card}(\Psi_\epsilon^j \cap V_j)/\text{card}\,\Psi_\epsilon^j \geqslant \sqrt{1-\delta}$ where $V_j = S(\text{mean}\,\Psi_\epsilon^j, \frac{1}{2}r) \subset S(y,r) = V$. Consequently, since $\Psi_\epsilon = \bigcup_j \Psi_\epsilon^j$, we have $\text{card}(\Psi_\epsilon \cap V) \geqslant \sqrt{1-\delta}\,\text{card}\,\Psi_\epsilon$. Summarizing,

$$\frac{\text{card}(\Phi_\epsilon \cap V)}{\text{card}\,\Phi_\epsilon} \geqslant \frac{\text{card}(\Psi_\epsilon \cap V)}{\text{card}\,\Phi_\epsilon} \geqslant \sqrt{1-\delta}\,\frac{\text{card}(\Psi_\epsilon \cap V)}{\text{card}\,\Psi_\epsilon} \geqslant 1 - \delta$$

whenever $0 < \epsilon < \min(\epsilon_1, \epsilon_2)$. Combining this with (5), we have $N_\epsilon(A \cap V) \sim N_\epsilon(A)$, as required.

Case 2: $c_1 = -c_2$. In this case $x_2 - x_1$ is independent of x_1. The second binomial coefficient in (4) can then be factored out of all the summations that form means and variances at t_1. The remaining sum resembles that treated in [*6*], and the mean and variance go to zero uniformly in t for $0 \leqslant t \leqslant t_1$. Over $t_1 \leqslant t \leqslant t_2$, the average, of course, converges uniformly to the straight line from $(t_1, 0)$ to (t_2, c_2^{-1}). Since $c = 0$ in this case, the result just obtained conforms with the statement of the theorem. Q.E.D.

It is an easy matter to generalize Theorem 2 to the case of a Lipschitz constant not equal to 1 and to the case $t_2 < T$. With a few remarks, furthermore, the method of proof can be extended to the more general case $\sum_{i=1}^n c_i f(t_i) = 1$ (where $0 < t_1 < \ldots < t_n \leqslant T$). In this general case, the variables $x_1, \ldots, x_{n-1}$ are thought of as independent, and x_n as depending on them according to

$$x_n = x_n(x_1, \ldots, x_{n-1}) = c_n^{-1}\left(1 - \sum_{i=1}^{n-1} c_i x_i\right).$$

For completeness, define $t_0 = x_0 = 0$. I is now a polyhedron in $(n-1)$-dimensional space, and is determined by $|x_i - x_{i-1}| \leqslant L(t_i - t_{i-1})$ $(i = 1, 2, \ldots, n)$. It is easy to see that I is convex.

$$B(x_1, \ldots, x_{n-1}) = \sum_{i=1}^n \{(L(t_i - t_{i-1}) + x_i - x_{i-1}) \ln (L(t_i - t_{i-1}) + x_i - x_{i-1})$$
$$+ (L(t_i - t_{i-1}) - x_i + x_{i-1}) \ln (L(t_i - t_{i-1}) - x_i + x_{i-1})\}.$$

If $\mathbf{M}(x_1, \ldots, x_{n-1})$ is the $(n-1) \times (n-1)$ matrix defined by $M_{ij} = d^2 B/dx_i dx_j$, then $\mathbf{M}$ is positive definite throughout I. To see this, calculate the derivatives and verify that $\mathbf{M} = v_n c_n^{-2} cc^T + \mathbf{D}$, where

(a) $v_i = \dfrac{2L(t_i - t_{i-1})}{L(t_i - t_{i-1})^2 - (x_i - x_{i-1})^2}$ $\quad (i = 1, \ldots, n)$
(note: $v_i > 0$ throughout I),

(b) $c^T = (c_1, c_2, \ldots, c_{n-1})$, and

(c) $\mathbf{D}$ is the symmetric, tri-diagonal matrix with diagonal entries $D_{ii} = v_i + v_{i+1}$ and off-diagonal entries $D_{i,i+1} = -v_{i+1}$.

Now $v_n c_n^{-2} cc^T$ is trivially positive semi-definite, and $\mathbf{D}$ is positive definite because its principal submatrices have determinants which can be shown inductively to be positive. Therefore, $\mathbf{M}$ is positive definite.

The positive definiteness of $\mathbf{M}$ guarantees that there is a unique point at which B attains its minimum, so that the multi-dimensional analogue to the method of Laplace ([*2*], pp. 71–72), can be applied to proving the generalization of Theorem 2. We state this generalization as

THEOREM 3. *Let* $L > 0$, *let* $0 < t_1 < t_2 < \ldots < t_n \leqslant T$, *and let* $c_1, c_2, \ldots, c_n$ *be real constants with* $c_n \neq 0$. *Using the uniform metric, define A to be the set of functions,* f: $[0,T] \to R$ *satisfying a Lipschitz condition with coefficient L such that* $f(0) = 0$ *and* $\sum_{i=1}^{n} c_i f(t_i) = 1$. *Assuming the choice of parameters is such that A is non-void, A has as a nucleus the broken line joining the points* $(0,0)$, $(t_1, \theta_1), \ldots, (t_n, \theta_n), (T, \theta_n)$, *the* θ_i *being uniquely characterized by*

$$\frac{L+\gamma_i}{L-\gamma_i} \cdot \frac{L-\gamma_{i+1}}{L+\gamma_{i+1}} \left(\frac{L-\gamma_n}{L+\gamma_n}\right)^{c_i/c_n} = 1 \qquad (i = 1, 2, \ldots, n-1)$$

where

$$\gamma_j = \frac{\theta_j - \theta_{j-1}}{t_j - t_{j-1}} \quad \textit{and} \quad \theta_n = x_n(\theta_1, \ldots, \theta_{n-1}) = c_n^{-1}\left(1 - \sum_{i=1}^{n-1} c_i \theta_i\right).$$

Concluding Remarks

One can verify that if A is any of the sets whose nucleus was determined above, this nucleus maximizes the functional

$$H(f) = -\int \left\{\frac{L+f'(t)}{2L} \ln \frac{L+f'(t)}{2L} + \frac{L-f'(t)}{2L} \ln \frac{L-f'(t)}{2L}\right\} dt \quad \text{over } A.$$

This quantity appears to be an aggregate, or average over t of a quantity resembling the communication entropy of information theory (see [*1*], pp. 5–24). Exactly why a quantity of this particular connotation should arise in the study of nuclei is not completely understood. Perhaps entropy, or uncertainty, is maximized at a nucleus because any neighbourhood of such a point encompasses a great many points. At any rate, the maximization of H forms an interesting and useful part of the theory of nuclei. In particular, it enables one to prove uniform convergence of the nucleus of the previous section for a sequence of sum constraints which are Riemann sums for an integral constraint. But that is a subject for another paper.

A valuable feature of the approach used to prove the theorems in this paper is that it avoids the necessity to compute $N_\epsilon(A \cap V)$. Instead, the probabilistic formulation and the use of Chebyshev's inequality permits direct estimation of $N_\epsilon(A \cap V)/N_\epsilon(A)$ without considering the numerator alone. This alleviates the problem of counting the number of ϕ-functions in a fixed neighbourhood of a suspected nucleus, but there remain many problems in which no way of calculating $N_\epsilon(A)$ to sufficient accuracy is known. In particular, a nucleus for the set of Lipschitz-continuous functions satisfying $f(0) = 0$ and $\int_0^T c(t)f(t)\,dt = 1$ has not been rigorously established.

I am indebted to Professor G. Strang for his help and advice during this research, and to NASA for its very real support.

REFERENCES

1. R. ASH, "Information Theory." Interscience, New York, 1965.
2. N. G. DE BRUIJN, "Asymptotic Methods in Analysis," 2nd ed. Interscience, New York, 1961.
3. W. FELLER, "An Introduction to Probability Theory and Its Applications," vol. I, 2nd ed. Wiley, New York, 1957.
4. J. E. ISRAEL, "An entropic approach to approximation" (M.I.T. doctoral dissertation, 1968).
5. A. N. KOLMOGOROV AND V. M. TIHOMIROV, "ϵ-entropy and ϵ-capacity of sets in functional spaces." *Am. Math. Soc. Transl., Ser. 2* **17** (1961).
6. G. STRANG, "The nucleus of a set." *Canadian Math. Bulletin* vol. **11** (1968).
7. A. G. VITUSHKIN, "Theory of Transmission and Processing of Information." Pergamon Press, New York, 1961.
8. H. S. WILF, "Mathematics for the Physical Sciences." Wiley, New York, 1962.