METRIC ENTROPY CONDITIONS
FOR AN OPERATOR TO BE OF TRACE CLASS

JOSE M. GONZALEZ-BARRIOS AND RICHARD M. DUDLEY

Abstract. Let \( A \) be an operator from one Hilbert space \( H \) into another. It was known that \( A \) is of trace class if and only if the metric entropy of \( A(B) \) is integrable where \( B \) is the unit ball in \( H \). We give a new, general sufficient condition for an integral operator to be of trace class, and examples showing it is sharp but not necessary.

1. Introduction

Kolmogorov's concept of \( \varepsilon \)-entropy (e.g., [KT]), here called metric entropy [Lo], is a measure of the size of a totally bounded metric space \( (S, d) \). Given \( \varepsilon > 0 \), let \( N_M(\varepsilon, S) \) be the smallest number of closed balls \( B(x_i, \varepsilon) := \{ y : d(x_i, y) \leq \varepsilon \}, i = 1, \ldots, n, \) in a covering of \( S \), in other words, the smallest \( n \) such that there exists an \( \varepsilon \)-net \( \{x_1, \ldots, x_n\} \) for \( (S, d) \). The diameter of a set \( A \subset S \) is

\[
\text{diam } A := \sup \{ d(x, y) : x, y \in A \}.
\]

Let \( N_K(\varepsilon, S) \) be the smallest number of sets \( A_i \) with diam \( A_i \leq 2\varepsilon \) that cover \( S \). Let \( D(\varepsilon, S) \) be the largest number of points \( x_j \in S \) such that \( d(x_i, x_j) > \varepsilon \) for all \( i \neq j \). Then it is known and easily checked that

\[
D(2\varepsilon, S) \leq N_K(\varepsilon, S) \leq N_M(\varepsilon, S) \leq D(\varepsilon, S).
\]

Thus as \( \varepsilon \downarrow 0 \), these quantities are of the same order of magnitude, up to a factor of 2 in \( \varepsilon \). Let \( H_K := \log N_K \), \( H_M := \log N_M \), and \( C := \log D \). The first result we will state is

Theorem A. If \( H \) is a Hilbert space, \( B \) its unit ball, and \( A \) a bounded operator from \( H \) into another Hilbert space, then

(i) \( A \) is of trace class if and only if

\[
\int_0^1 H_M(\varepsilon, A(B)) \, d\varepsilon < \infty;
\]
A is a Hilbert-Schmidt operator if and only if
\[ \int_0^1 H_M(\varepsilon^{1/2}, A(B)) \, d\varepsilon < \infty. \]
Here \( H_M(\cdot, \cdot) \) can be replaced by \( H_K(\cdot, \cdot) \) or \( C(\cdot, \cdot) \).

Theorem A reduces easily to the case of selfadjoint compact operators with a basis of eigenvectors and thus to diagonal operators. Then, part (ii) was stated by Sudakov [Su] and is proved in Marcus [Ma]; part (i) was given by Oloff [Ol], see also Carl [C].

An operator \( A \) is Hilbert-Schmidt if and only if \( A^* A \) is of trace class, where \( A^* \) is the adjoint of \( A \). Hilbert-Schmidt operators are interesting to probabilists as the so-called radonifying operators between Hilbert spaces: Sazonov [Sa], Minlos [Min], Kolmogorov [K], and Schwartz [Schw]. On related questions for the isonormal and other Gaussian processes, the metric entropy condition
\[ \int_0^1 H_K(\varepsilon, E)^{1/2} \, d\varepsilon < \infty \]
has been studied by Dudley [Du1, Du2] and Fernique [F]. Then (1.4) implies (1.3) for an ellipsoid \( E = A(B) \) but not conversely; for example, if \( H_K(\varepsilon, E) \sim \varepsilon^{-2} |\log \varepsilon|^\alpha \) as \( \varepsilon \to 0 \), then (1.4) holds just for \( \alpha < -2 \) and (1.3) holds for \( \alpha < -1 \).

The characterization of trace class operators gives a sufficient condition for an integral operator to be of trace class as follows. Let \((X, \mu)\) and \((Y, \nu)\) be two finite measure spaces. Let \( K \in L^2(X \times Y, \mu \times \nu) \). Then, as is well known, an integral operator \( A_K \) from \( L^2(\nu) \) into \( L^2(\mu) \) is defined by
\[ A_K(f)(x) = \int K(x, y)f(y) \, d\nu(y) \]
and is a Hilbert-Schmidt operator. Conversely, any Hilbert-Schmidt operator from \( L^2(\nu) \) into \( L^2(\mu) \) is of the form \( A_K \) for some \( K \in L^2(X \times Y, \mu \times \nu) \). Conditions for \( A_K \) to be of trace class are not as simple. Here is the main new result of this note, a sufficient condition based on metric entropy.

**Theorem 1.** Let \( K_Y := \{K(\cdot, y) : y \in Y\} \). Suppose \( K \) is such that for some \( M < \infty \) and \( r < 2 \), for \( 0 < \varepsilon < 1 \) we have \( N_M(\varepsilon, K_Y) \leq M/\varepsilon^r \). Then \( A_K \) is of trace class.

The next section will give quite short, partly new proofs of Theorem A and some related facts to be given in Theorem B. Then §3 proves Theorem 1 and shows that "\( r < 2 \)" is sharp but that no metric entropy condition on \( K_Y \) characterizes trace class integral operators.

### 2. Statements and proofs for Theorem A

We can assume that the bounded operator \( A \) takes the Hilbert space \( H \) into itself. Let \( B \) be the unit ball of \( H \). Let \( |A| := (A^* A)^{1/2} \). If any of the conditions in Theorem A is to hold, \( A \) must be a compact operator, hence so is \(|A|\). Then there is an orthonormal basis \( \{e_n\} \) of \( H \) with \(|A|e_n = a_ne_n \) for all \( n \) and \( a_n \to 0 \), \( n \to \infty \). For a given orthonormal set \( \{f_n\} \) and bounded sequence of numbers \( c_n > 0 \), define the ellipsoid \( E(\{c_n\}) := E(\{c_n\}, \{f_n\}) := \)
\{\sum_n x_n f_n : \sum_n (x_n/c_n)^2 \leq 1\}$. Then $|A|(B) = E(\{b_n\}, \{f_n\})$, where $f_n$ is the subsequence of those $e_n$ such that $a_n \neq 0$ (so $a_n > 0$) and $b_n$ are those $a_n$. Also, via a partial isometry (e.g., [Scha, p. 4]) $A(B) = E(\{b_n\}, \{g_n\})$ where \{g_n\} is an orthonormal set. So $A(B)$ and $|A|(B)$ are isometric ellipsoids, with the same sequence \{b_n\}. We can assume that $b_n \downarrow 0$ as $n \to \infty$.

For a sequence $b_n \downarrow 0$ and $t > 0$, let $m(t) := \sup\{n : b_n \geq 1/t\}$, or 0 if $b_0 < 1/t$. For $s > 0$ let $I(s) := \int_0^s m(t)/t \, dt$. Theorem A will follow easily from Theorem B. A key step in the proof follows from Mityagin [Mit]. Theorem 2 of Marcus [Ma] includes the case $r = 2$; and Oloff [Ol] includes the general case.

**Theorem B.** Let $E := E(\{b_n\}, \{f_n\})$ for an orthonormal set \{f_n\} and some $b_n \downarrow 0$. Let $0 < r < \infty$. Then the following are equivalent:

(a) $\sum_n b_n^r < \infty$.

(b) $\int_1^\infty m(t)t^{-r-1} \, dt < \infty$.

(c) For some $c > 0$, $\int_0^1 I(ce^{-1/r}) \, de < \infty$.

(d) For all $c > 0$, $\int_0^1 I(ce^{-1/r}) \, de < \infty$.

(e) $\int_0^1 H_M(ce^{-1/r}, E) \, de < \infty$.

In (e), $H_M(\cdot, \cdot)$ can be replaced equivalently by $H_K(\cdot, \cdot)$ or $C(\cdot, \cdot)$.

**Proof.** The series in (a) equals the Riemann-Stieltjes integral $\int_0^\infty t^{-r} \, dm(t)$. So (a) is equivalent to (b) by integration by parts (e.g., [Le, p. 10]).

Next,

$$\int_0^1 I(ce^{-1/r}) \, de = \int_0^1 \int_0^{ce^{-1/r}} \frac{m(t)}{t} \, dt \, de = \int_0^\infty \frac{m(t)}{t} \min\left(1, \frac{c^r}{t^r}ight) \, dt$$

$$= \int_c^\infty \frac{m(t)}{t} \, dt + c^r \int_c^\infty \frac{m(t)}{t^r+1} \, dt,$$

and since $m(t) = 0$ for $0 < t < 1/b_0$, (b), (c), and (d) are all equivalent.

By results of Mityagin [Mit, p. 74], (d) for $c = 8$ implies (e), and (e) implies (c) for $c = 1/2$. So (a) through (e) are equivalent.

In (e), $H_M$ can be replaced by $H_K$ or $C$ by (1), and since all these functions are nonincreasing, integrability is only an issue near 0. So Theorem B is proved.

Carl and Stephani [CS, pp. 118–119] give other relations between semiaxes and metric entropy of ellipsoids.

**Proof of Theorem A.** For (a), $A$ is of trace class by definition iff $\sum_n b_n < \infty$, and Hilbert-Schmidt iff $\sum_n b_n^2 < \infty$, so we can apply Theorem B for $r = 1, 2$. □

### 3. Integral Operators

First we give a

**Proof of Theorem 1.** Let $C := \nu(Y)^{1/2}$. Let $A$ be the union $A := \{CK(\cdot, y) : y \in Y\} \cup \{-CK(\cdot, y) : y \in Y\} \subset L^2(X)$. Then for $0 < \varepsilon < 1$, $N_M(\varepsilon, A) \leq 2N_M(\varepsilon/C, K_Y) \leq D/\varepsilon^r$ where $D = 2MC^r$ if $C \geq 1$ or $\varepsilon/C \leq 1$, and so in
any case for \( \varepsilon \) small enough and thus for \( 0 < \varepsilon < 1 \), possibly with a larger constant \( D < \infty \).

Let \( B \) be the unit ball in \( L^2(Y, \nu) \). It will be shown that all functions in \( A_K(B) \) are in the closed convex hull of \( A \). For \( f \in B \) let \( f = f^+ - f^- \) where \( f^+ := \max(f, 0) \), \( f^- := -\min(f, 0) \). Then

\[
(*) \quad \int K(x, y) f(y) \, d\nu(y) = \int f^+(y) K(x, y) + f^-(y) (-K(x, y)) \, d\nu(y).
\]

Here \( f^+ + f^- \equiv |f| \geq 0 \) and \( \int |f| \, d\nu \leq C(\int |f|^2 \, d\nu)^{1/2} \leq C \) by the Cauchy-Bunyakovsky (-Schwarz) inequality. So multiplying and dividing by \( C \), we will show that \( A_K(f) \) is of the form

\[
A_K(f)(x) = \int g(x) \, dP(g) \quad \text{where} \quad P \text{ is a probability measure on } A.
\]

There are nonnegative measures \( P_1 \) and \( P_2 \) on \( Y \) with \( dP_1 = f^+ \, d\nu/C \) and \( dP_2 = f^- \, d\nu/C \) so that \( (P_1 + P_2)(Y) \leq 1 \). If \( \alpha := 1 - (P_1 + P_2)(Y) > 0 \), take a fixed \( y = z \) and replace \( P_i \) by \( P_i + \alpha \delta_z/2 \), \( i = 1, 2 \). Then \( P_1 + P_2 \) is a probability measure on \( Y \), and

\[
A_K(f)(x) \equiv \int CK(x, y) \, dP_1(y) + \int -CK(x, y) \, dP_2(y),
\]

as desired. By assumption, \( K_Y \) is totally bounded and so separable. The map \( y \mapsto K(\cdot, y) \) is measurable, so \( A_K(f) \) is in the closed convex hull of \( A \) (e.g., [DiU, pp. 42, 48]). Also, \( A \) is bounded in \( L^2(\mu) \), say \( \|x\| \leq T < \infty \) for all \( x \in A \), so \( A_K(B)/T \) is included in the closed convex hull of \( A/T \). It then follows from [Du3, Theorem 5.1] that for any \( t > 2r/(2+r) \), there are constants \( C_1, C_2 < \infty \) such that for \( 0 < \varepsilon < 1 \),

\[
N_M(\varepsilon, A_K(B)/T) \leq C_1(\exp(C_2\varepsilon^{-t})).
\]

Thus \( N_M(\varepsilon, A_K(B)) \leq C_1 \exp(C_3\varepsilon^{-t}), \ 0 < \varepsilon \leq 1 \), where \( C_3 = C_2 T' \). Now \( r < 2 \) implies \( 2r/(2+r) < 1 \), so we can choose \( t < 1 \) and apply Theorem A to conclude that \( A_K \) is of trace class. \( \square \)

Specializing Theorem 1, let \( X = Y = [a, b] \), \( \mu = \nu = \text{Lebesgue measure} \). Say that \( K(\cdot, \cdot) \in \text{Lip}_\alpha \) in the variable \( x \) iff

\[
|K(x + h, y) - K(x, y)| \leq |h|^{\alpha} G(y)
\]

whenever \( x, x + h, y \in [a, b] \), where \( G \in L^2[a, b] \). The condition \( \text{Lip}_\alpha \) in \( y \) is defined symmetrically. Hille and Tamarkin [HT, Theorem 9.1] implies that \( A_K \) is of trace class if \( K(\cdot, \cdot) \in \text{Lip}_\alpha \) in either of its variables and \( \alpha > \frac{1}{2} \).

This follows directly from Theorem 1: for simplicity suppose \( [a, b] = [0, 1] \). Since the adjoint of a trace class operator is of trace class, and since the adjoint of \( A_K \) is \( A_L \) where \( L(x, y) \equiv K(y, x) \), we can assume \( K \) is \( \text{Lip}_\alpha \) in \( x \).

Let \( 0 < \varepsilon \leq 1 \) and \( \gamma := \max(1, \|G\|_2) \). Then for the usual metric on \( [0, 1], N_M((\varepsilon/\gamma)^{1/\alpha}, [0, 1]) \leq (\gamma/\varepsilon)^{1/\alpha} \) and

\[
|K(x + (\varepsilon/\gamma)^{1/\alpha}, y) - K(x, y)| \leq \varepsilon G(y)/\gamma
\]

whenever all the arguments are in \( [0, 1] \), so

\[
\|K(x + (\varepsilon/\gamma)^{1/\alpha}, \cdot) - K(x, \cdot)\|_2 \leq \varepsilon.
\]

Since \( \alpha > \frac{1}{2} \), it follows that \( \frac{1}{\alpha} < 2 \) and Theorem 1 applies.
Smithies [Sm] and Stinespring [St] extended Hille and Tamarkin's result in a different direction. Stinespring showed that $A_K$ is of trace class if $K(x, y)$ is periodic of period 1 in $x$ and
\[
\int_0^1 \int_0^1 \int_0^1 |K(x+h, y) - K(x, y)|^2 \, dy \, dx \, dh < \infty
\]
for some $\beta > 2$.

**Example.** The condition $r < 2$ in Theorem 1 is sharp: Let $K(x, y) := 1_T(x, y)$ where $T := \{(x, y) : 0 \leq y \leq x \leq 1\}$, on $[0, 1] \times [0, 1]$ with Lebesgue measure. Then $A_K(f)(x) = \int_0^x f(y) \, dy$: $A_K$ is the indefinite integral operator. The functions $f_n(y) := e^{2\pi i ny}$, $n = \pm 1, \pm 2, \ldots$, are eigenvalues of $A_K$ and its adjoint and so of $|A_K|$, with eigenvalues $1/(2\pi|n|)$ for the latter, so $A_K$ is not of trace class, as is well known. For this $K$, we have $N_M(\varepsilon, K_Y) \leq 1/\varepsilon^2$ for $0 < \varepsilon \leq 1$. So Theorem 1 fails for $r = 2$.

On the other hand, the condition $r < 2$ is far from necessary, as the following shows:

**Example.** Let $\mu = \nu = \text{Lebesgue measure on } [0, 1]$. Let $r_n$ be independent Rademacher functions, specifically, $r_n(x) = 1$ if the $n$th binary digit is 1 and $r_n(x) = -1$ otherwise. Then for each $n$, $\mu(r_n = 1) = \mu(r_n = -1) = \frac{1}{2}$ and the $r_n$ are orthonormal in $L^2[0, 1]$. Let $\delta > 0$ and $K(x, y) := \sum_{n \geq 1} n^{-1-\delta} r_n(x) r_n(y)$. Then clearly $A_K$ is of trace class. Now $K_Y$ consists of those functions where each $r_n(y)$ can either be $+1$ or $-1$, independently of the others. Thus, given $\varepsilon > 0$, if $2n^{-1-\delta} > \varepsilon$ then $D(\varepsilon, K_Y) \geq 2^n$ since we can choose $r_j = \pm 1$ for $j = 1, \ldots, n$ and get $2^n$ functions at distances more than $\varepsilon$ apart. Thus
\[
D(\varepsilon, K_Y) \geq \exp((\log 2)(2/\varepsilon)^{1/(1+\delta)} - 1).
\]
So we have, for any $r < 1$, examples of trace class operators with $\log D(\varepsilon, K_Y) \geq \alpha \varepsilon^{-r}$ for some $\alpha > 0$ and for $0 < \varepsilon \leq 1$. In this sense $K_Y$ can be about as large as $A_K(B)$ itself can be for $A_K$ of trace class. Also, since $K$ is symmetric, $A_K$ is selfadjoint and the corresponding class of functions $K_X := \{K(x, \cdot) : x \in X\}$ is the same as $K_Y$.

This and the previous example show that no condition on the metric entropy of $K_Y$ can characterize trace class integral operators.

Stinespring's hypothesis (3.1), although the condition $\beta > 2$ is also sharp, fails for the rank 1 operator $A_L$ with $L(x, y) := r_1(x)r_1(y)$. The hypothesis of Theorem 1 also fails for a rank 1 operator $A_K$ with $K(x, y) = f(x)g(y)$ whenever $g$ is not essentially bounded. So there is still apparently much to be done in finding useful conditions for the trace class property of integral operators.

Some of the results of this paper appeared in the first author's thesis [G].

**Acknowledgment**

We thank Michael Marcus for telling us of his paper [Ma] and a referee for remarks that helped shorten the proof of Theorem B.
REFERENCES


CIMAT, GUANAJUATO 36000, MEXICO

Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307

E-mail address: rmd@math.mit.edu