ISRAEL JOURNAL OF MATHEMATICS, Vol. 55, No. 2, 1986

ON HILBERTIAN SUBSETS OF FINITE METRIC SPACES[†]

BY

J. BOURGAIN,[a] T. FIGIEL[b] AND V. MILMAN[c]

[a] *University of Illinois and IHES*; [b] *Polish Academy of Sciences*; *and* [c] *Tel Aviv University*

ABSTRACT

The following result is proved: For every $\varepsilon > 0$ there is a $C(\varepsilon) > 0$ such that every finite metric space (X, d) contains a subset Y such that $|Y| \geqq C(\varepsilon) \log |X|$ and (Y, d_Y) embeds $(1+\varepsilon)$-isomorphically into the Hilbert space l_2.

0. Introduction

Define the Lipschitz distance between two finite metric spaces (X, ρ) and (Y, d) such that $|X| = |Y|$ as $d(X, Y) = \inf \| \psi \|_{\mathrm{Lip}} \| \psi^{-1} \|_{\mathrm{Lip}}$, where the infimum is taken over all one to one and onto maps $\psi : X \to Y$. Remember that

$$\| \psi \|_{\mathrm{Lip}} = \sup_{x \neq y} \frac{d(\psi(x), \psi(y))}{\rho(x, y)}.$$

If $d(X, Y) \leqq c$, we call such spaces c-isomorphic.

The following result can be regarded as an analogue of the famous theorem of Dvoretzky [1] in the linear theory.

THEOREM. *For every $\varepsilon > 0$ there is a $C(\varepsilon) > 0$ such that every finite metric space (X, d) contains a subset Y such that $(Y, d|_Y)$ embeds $(1+\varepsilon)$-isomorphically into the Hilbert space l_2 and*

$$|Y| \geqq C(\varepsilon) \log |X|.$$

Moreover, one can take $C(\varepsilon) = c_1 \varepsilon / \log(c_2/\varepsilon)$ (if $0 < \varepsilon < 1$) where c_1, c_2 are positive constants.

REMARK. We show also that this result cannot be improved in the general

[†] The authors are grateful to Haim Wolfson for some discussions related to the content of this paper.

Received October 27, 1985

case, i.e. the logarithmic order of $|Y|$ can be optimal even if X is 2-isomorphic to a subset of l_2.

1. Proof of the Theorem

Let $\frac{1}{2} > \delta > 0$ be a fixed number and let $(1+\delta)^{m-1} > 4$. (It will turn out that $1+\varepsilon = (1+\delta)^2$.) Let (X, d) be a finite metric space. We shall define inductively decreasing sets $X_0 = X \supset X_1 \supset \cdots \supset X_k$ so that

$$|X_{i+1}| \geqq \frac{1}{m} |X_i| \qquad \text{for } i = 0, 1, \dots .$$

Suppose that X_i has been defined and

$$|X_i| \geqq m^{-i} |X| > m.$$

We pick $x_i \in X_i$ arbitrarily and let

$$d_i = \max\{d(x_i, x) : x \in X_i\}.$$

We also pick $y_i \in X_i$ such that $d(x_i, y_i) = d_i$ and let

$$A_i = \{x \in X_i : d(x_i, x) \leqq \tfrac{1}{4} d_i\}.$$

We let $g(i) = 0$ if $|A_i| > (1/m)|X_i|$ and $g(i) = 1$ otherwise.

If $g(i) = 0$ we put $X_{i+1} = A_i$.

If $g(i) = 1$, we can find an η, $\frac{1}{4} \leqq \eta < 1$ such that the set

$$B_\eta = \{x \in X_i : \eta d_i < d(x_i, x) \leqq (1+\delta)\eta d_i\}$$

satisfies $|B_\eta| \geqq (1/m)|X_i|$ (because $(1+\delta)^{m-1} \geqq 4$).

In this case we let $X_{i+1} = B_\eta$.

In both cases we have

$$|X_{i+1}| \geqq \frac{1}{m} |X_i| \geqq m^{-i-1} |X|$$

and unless $|X| \leqq m^{i+2}$ we can continue this inductive procedure.

Hence, if $m^k < |X| \leqq m^{k+1}$, then we can define x_i, y_i, $g(i)$ for $i = 0, 1, \dots, k-1$. Now we shall consider the sets

$$\tilde{X} = \{x_i : 0 \leqq i < k, g(i) = 1\} \quad \text{and} \quad \tilde{Y} = \{y_i : 0 \leqq i < k, g(i) = 0\}.$$

Clearly, $|\tilde{X}| + |\tilde{Y}| = k$, hence one of them must have at least $\frac{1}{2}k$ elements.

We shall show first that $\tilde{X}$ has a big subset which is $(1+\delta)^2$-isomorphic to a subset of l_2.

Write $\tilde{X} = \{x_{i_1}, x_{i_2}, \dots, x_{i_s}\}$, where $i_1 < i_2 < \cdots < i_s$ and put $z_j = x_{i_j}$ for $j = 1, \dots, s$. Observe that the formula

$$d'(z_i, z_j) = \max\{d(z_i, z_\tau) : i < \tau \leqq s\}$$

for $1 \leqq i < j \leqq s$ defines a metric d' on $\tilde{X}$ which satisfies

$$d(z_i, z_j) \leqq d'(z_i, z_j) \leqq (1+\delta) d(z_i, z_j)$$

and $d'(z_i, z_j) = \delta_i = d'(z_i, z_s)$ for $i < j \leqq s$. Since $(\tilde{X}, d')$ is $(1+\delta)$-isomorphic to $(\tilde{X}, d\,|_{\tilde{X}})$, it suffices to find a subset of $(\tilde{X}, d')$ which is $(1+\delta)$-isomorphic to a subset of l_2. Let

$$\tilde{X}_n = \{z_i \in \tilde{X} : (1+\delta)^{-n} \leqq \delta_i < (1+\delta)^{1-n}\}$$

for each integer n.

Observe that $z_i \in \tilde{X}_n$, $z_j < \tilde{X}_{n+m}$ implies that $i < j$.

Indeed, suppose that $j < i$, then we would have

$$\delta_i = d(x_i, x_\tau) \leqq d(x_i, x_j) + d(x_j, x_\tau) \leqq 2\delta_j$$

and hence

$$(1+\delta)^{-n} \leqq \delta_i \leqq 2\delta_j < 2(1+\delta)^{1-m-n} < \tfrac{1}{2}(1+\delta)^{-n}.$$

Now consider the sets

$$\cup\{\tilde{X}_n : n \equiv \alpha \,(\mathrm{mod}\, m)\}$$

for $\alpha = 0, 1, \dots, m-1$. One of them, call it X', has at least $(1/m)|\tilde{X}|$ elements.

Clearly, if $z_i, z_j \in X'$ and $i < j$, then either $\{z_i, z_j\} \subset \tilde{X}_n$ for some n, or $z_i \in \tilde{X}_n$, $z_j \in \tilde{X}_{n+sm}$, $s \geqq 1$. In the first case $(1+\delta)^{-1} < \delta_i / \delta_j < 1 + \delta$, in the other case

$$\delta_j \leqq (1+\delta)^{1-n-sm} \leqq (1+\delta)^{1-m} \delta_i \leqq \tfrac{1}{3}\delta_i.$$

Now put for $z_i, z_j \in \tilde{X}$, $i < j$

$$d''(z_i, z_j) = \bar{\delta}_i = \min\{\delta_\tau : 1 \leqq \tau \leqq i, z_\tau \in \tilde{X}\}.$$

The function d'' defines a new metric on $\tilde{X}$

$$(1+\delta)^{-1} d'(z_i, z_j) \leqq d''(z_i, z_j) \leqq d'(z_i, z_j).$$

(It is a metric because the numbers $\bar{\delta}_i$ decrease.)

It remains to apply the following fact.

Lemma. *If ρ is a metric on a set $S = \{s_1, s_2, \dots, s_n\}$ such that $\rho(s_i, s_j) = \rho(s_i, s_n)$ if $1 \leqq i < j \leqq n$ and*

$$\rho(s_1, s_n) \geqq \rho(s_2, s_n) \geqq \cdots \geqq \rho(s_n, s_n) = 0$$

then (S, ρ) *is isometric to a subspace of* l_2.

The proof of the lemma is left as an exercise.

We turn our attention to the second structure $\tilde{Y}$.

One may show that $(\tilde{Y}, d_{\tilde{Y}})$ is $\frac{16}{3}$-isomorphic to a subset of the real line (with the standard metric).

Enumerate again $\tilde{Y} = \{y_{j_1}, \ldots, y_{j_t}\}$, $t = k - s$, where $j_1 < j_2 < \cdots < j_t$ and put $y_{j_i} = w_i$. Define $T(w_i) = d(w_i, w_t)$.

Note that if $w_i = y_j$, $j < t$, then

$$\{w_{i+1}, \ldots, w_t\} \subset A_j.$$

Since diam $A_j \leqq 2 \cdot \frac{1}{4} d(x_j, y_j) = \frac{1}{2} d_j$ and $d(w_t, y_j) \geqq d_j - \frac{1}{4} d_j = \frac{3}{4} d_j$ we see that

$$T(w_{i+1}) \leqq \tfrac{1}{2} d_j \leqq \tfrac{2}{3} T(w_i).$$

Hence if $1 \leqq i < j \leqq t$ then $T(w_j) \leqq T(w_i)$ and

$$\tfrac{1}{4} d(w_i, w_j) \leqq \tfrac{1}{3} d(w_i, w_t) \leqq |T(w_i) - T(w_j)| \leqq d(w_i, w_t) \leqq \tfrac{4}{3} d(w_i, w_j).$$

This estimate can be easily improved by passing to a thinner subsequence. Let $v_i = w_{1+(i-1)m}$ for $i = 1, 2, \ldots, [(t-1)/m] = t'$. Then

$$T(v_{i+1}) \leqq (\tfrac{2}{3})^m T(v_i)$$

and, letting $(\frac{2}{3})^m = \alpha$, we can estimate for $1 \leqq i < j \leqq t'$

$$(1 - \alpha) d(v_i, w_t) \leqq T(v_i) - T(v_j) \leqq T(v_i) = d(v_i, w_t).$$

Also

$$d(v_i, v_j) \leqq d(v_i, w_t) + d(v_j, w_t) \leqq (1 + \alpha) d(v_i, w_t)$$

and

$$d(v_i, w_t) \leqq d(v_i, v_j) + d(v_j, w_t),$$

whence

$$d(v_i, v_j) \geqq (1 - \alpha) d(v_i, w_t).$$

It follows that T defines a $(1 + \alpha)/(1 - \alpha)^2$-isomorphic embedding of $V = \{v_1, \ldots, v_{t'}\}$ into l_2. If $0 < \varepsilon < 1$ and $\alpha = (\frac{2}{3})^m < \varepsilon/6$, then we get a $(1 + \varepsilon)$-isomorphic embedding.

Consequently, one of the sets X' and V must have at least $(1/m)(\frac{1}{2}k)$ elements and both admit a $(1+\varepsilon)$-isomorphic embedding into l_2 if $(1+\delta)^2 \leqq 1+\varepsilon$. Since

$$k+1 \geqq (\log m)^{-1} \log |X|$$

and m depends only on ε. The required estimate follows now from a simple computation

2. Construction of an Example

DEFINITION. Denote $\gamma(X,d)$ the Lipschitz distance from (X,d) to a subset of Hilbert space.

PROPOSITION. *There is ε_0 with the following property. If $n \geqq s \geqq 3+2\log_2 n$, then there is a metric d on $Z_n = \{1,\dots,n\}$ s.t.*

$$\gamma(X, d|_X) \geqq 1+\varepsilon_0$$

for every subset $X \subset Z_n$ with $|X| = s$. Moreover, d takes only values 0, 1, 2, *hence* $\gamma_2(Z_n, d) \leqq 2$.

NOTATION. A metric d on X is called a D-metric if $d(x,y) \in D$ for $x, y \in X$. We shall use this convention only for $D = \{0,1,2\}$.

FACT 1. If d is a D-metric on X then either $\gamma(X,d) = 1$ or $\gamma(X,d) \geqq 1+\varepsilon_0$ where $\varepsilon_0 > 0$ is an absolute constant; this distance is attained for some (X,d) where $|X| = 4$.

FACT 2. Let $\mathcal{D}_s$ be the set of all those D-metrics d on the set $\{1,\dots,s\}$ such that $\gamma(X_s, d) = 1$. Then

$$|\mathcal{D}_s| \leqq s!2^s.$$

PROOF OF THE RESULT. Note first that there are exactly 2^N, $N = \binom{n}{2}$, D-metrics on the set Z_n.

We shall estimate $|\xi|$, where

$$\xi = \{d \in \mathcal{D}_n : \exists X \subset Z_n, |X| = s; \gamma(X, d|_X) = 1\}.$$

Using Fact 2 we get

$$|\xi| \leqq \binom{n}{s} s!2^s 2^{n-\binom{s}{2}}.$$

Since $\binom{n}{s} < n^s/s!$, in order that $|\xi| < |\mathcal{D}_n|$, it will suffice that

$$(2n)^s 2^{-\binom{s}{2}} \leqq 1$$

or

$$\log_2 n \geqq \frac{s-3}{2}$$

PROOF OF FACT 1. Let d be a D-metric on a set X.

We may assume that $|X| \geqq 4$, since otherwise X is isometric to a subset of the Euclidean plane.

We shall consider two cases.

Case I. The relation $d(x, y) \leqq 1$ for $x, y \in X$ is transitive, and hence it is an equivalence relation on X.

In this case one shows easily that $\gamma(X, d) = 1$.

Case II. The relation $d(x, y) \leqq 1$ is not transitive. This implies that there exist x, y, $z \in X$ s.t. $d(x, y) = 1$, $d(y, z) = 1$, $d(x, z) = 2$. Pick another point $w \in X$. Let $X_0 = \{x, y, z, w\}$. It is clear that the set (X_0, d) cannot be embedded isometrically into l_2. (One could not have an isometric embedding T with $T(x) = e_1$, $T(y) = 0$, $T(z) = -e_1$ and $\| e_1 - Tw \|$, $\| 0 - Tw \|$, $\| -e_1 - Tw \|$ being either 1 or 2.) Since $\gamma(X, d) \geqq \gamma(X_0, d_{X_0})$ and there are only finitely many types of D-metric spaces on $\{x, y, z, w\}$, the existence of some ε_0 is obvious. This proves Fact 1.

PROOF OF FACT 2. To obtain Fact 2, it suffices from previous discussions to estimate from above the number of equivalence relations, on the set Z_s. Since after a suitable permutation of Z_s the classes of equivalent elements become intervals which can be coded by marking their initial elements, the estimate by $s!2^s$ becomes obvious.

REFERENCE

1. A. Dvoretzky, *Some results on convex bodies and Banach spaces*, Proc. Int. Symp. on Linear Spaces, Jerusalem, 1961, pp. 123–160.