ALMOST ORTHOGONAL SUBMATRICES OF AN ORTHOGONAL MATRIX

BY

M. RUDELSON*

Department of Mathematics, Texas A&M University
College Station, TX 77845, USA
e-mail: Mark.Rudelson@math.tamu.edu

ABSTRACT

Let $t \geq 1$ and let n, M be natural numbers, $n < M$. Let $A = (a_{i,j})$ be an $n \times M$ matrix whose rows are orthonormal. Suppose that the ℓ_2-norms of the columns of A are uniformly bounded. Namely, for all j

$$\sqrt{\frac{M}{n}} \cdot \left(\sum_{i=1}^{n} a_{i,j}^2 \right)^{1/2} \leq t.$$

Using majorizing measure estimates we prove that for every $\varepsilon > 0$ there exists a set $I \subset \{1, \ldots, M\}$ of cardinality at most

$$C \cdot \frac{t^2}{\varepsilon^2} \cdot n \cdot \log \frac{nt^2}{\varepsilon^2}$$

such that the matrix $\sqrt{M/|I|} \cdot A_I^T$, where $A_I = (a_{i,j})_{j \in I}$, acts as a $(1+\varepsilon)$-isomorphism from ℓ_2^n into $\ell_2^{|I|}$.

1. Introduction

We consider the following problem, posed by B. Kashin and L. Tzafriri [K-T]: *Let $\varepsilon > 0$ and let n, M be natural numbers, $n < M$. Given an $n \times M$ matrix A whose rows are orthonormal, what is the smallest cardinality $L(A, \varepsilon)$ of a subset $I \subset \{1, \ldots, M\}$ so that for all $x \in \ell_2^n$*

$$(1.1) \qquad (1 - \varepsilon) \cdot \|x\| \leq \sqrt{\frac{M}{|I|}} \cdot \left\| R_I A^T x \right\| \leq (1 + \varepsilon) \cdot \|x\| .$$

* Research supported in part by a grant of the US–Israel BSF. Part of this research was performed when the author held a postdoctoral position at MSRI. Research at MSRI was supported in part by NSF grant DMS-9022140.

Received January 29, 1998

Here $R_I\colon \mathbb{R}^M \to \mathbb{R}^M$ is the orthogonal projection onto the space $\mathrm{span}\{e_i | \; i \in I\}$, where $\{e_i\}_{i=1}^M$ is the standard basis of $\mathbb{R}^M$. Throughout this paper we denote by $\|\cdot\|$ the standard ℓ_2-norm and by $|I|$ the cardinality of a set I. The normalizing coefficient $\sqrt{M/|I|}$ arises naturally in the case where all $a_{i,j}$ have the same absolute value. Indeed, in this case the norm of each row of the matrix $R_I A^T$ is $(\sum_{j\in I} a_{i,j}^2)^{1/2} = \sqrt{|I|/M}$.

This problem arises from the question of finding a "good" discretization of a given orthonormal system. Namely, let $\{\phi_i(\omega)\}_{i=1}^n$ be an orthonormal system in the space $L_2(X, \Sigma, \mu)$. Find a finite set of points $x_1, \dots, x_m$ of smallest possible cardinality m such that the system

$$\{(\phi_i(x_1), \phi_i(x_2), \dots, \phi_i(x_m))\}_{i=1}^n$$

of vectors in $\mathbb{R}^m$ will be close to an orthogonal system.

Under an additional assumption that all the entries of A have the same absolute value $1/\sqrt{M}$, Kashin and Tzafriri proved that

$$L(A, \varepsilon) \le \frac{c}{\varepsilon^4} \cdot n^2 \log n. \tag{1.2}$$

Moreover, their proof shows that a random subset I of this cardinality satisfies (1.1) with probability close to 1. Clearly, the estimate (1.2) is not optimal. The example of random selection of columns of a rectangular Walsh matrix, considered by Kashin and Tzafriri, suggests that the possible upper bound could be

$$L(A, \varepsilon) \le C(\varepsilon) \cdot n \log n. \tag{1.3}$$

From the other side, simple examples ([K-T], [R]) show that the estimate (1.3) is the best one can obtain by the random selection method.

As was mentioned in [R], the Kashin and Tzafriri problem is dual to that of finding an approximate John's decomposition. Entropy estimates used in [R] for the last problem enabled one to improve (1.2). More precisely, let $t \ge 1$ and suppose that the matrix A satisfies

$$\sqrt{\frac{M}{n}} \cdot \left(\sum_{i=1}^n a_{i,j}^2 \right)^{1/2} \le t$$

for all $j = 1, \dots, M$. Then

$$L(A, \varepsilon) \le C(\varepsilon) \cdot t^2 \cdot n \log^3 n.$$

In order to improve this estimate one can use majorizing measures instead of entropy estimates. The method of majorizing measures, developed by Talagrand ([L-T], [T1]), is extremely useful in obtaining estimates of stochastic processes, related to random selection. A random process, similar to that arising in the Kashin and Tzafriri problem, was considered by Talagrand [T2] for the problem of embedding of a finite dimensional subspace of L_p into ℓ_p^N. For this kind of process Talagrand introduced a special method of constructing majorizing measures. This method (s-separated trees) can be used to prove an estimate

$$L(A,\varepsilon) \le C(\varepsilon) \cdot t^2 \cdot n \log n \cdot (\log\log n)^2$$

for the Kashin and Tzafriri problem. It is unlikely that the $(\log\log n)^2$ factor can be removed by a modification of the s-separated trees method. However, using a different approach based on the explicit construction of a partition tree, we obtained a sharper estimate. More precisely, we prove the following

THEOREM: *Let $t \ge 1$ and let $A = (a_{i,j})$ be an $n \times M$ matrix, whose rows are orthonormal. Suppose that for all j*

$$\sqrt{\frac{M}{n}} \cdot \left(\sum_{i=1}^{n} a_{i,j}^2 \right)^{1/2} \le t. \tag{1.4}$$

Then for every $\varepsilon > 0$ there exists a set $I \subset \{1, \dots, M\}$ so that

$$|I| \le C \cdot \frac{t^2}{\varepsilon^2} \cdot n \cdot \log \frac{nt^2}{\varepsilon^2}, \tag{1.5}$$

and for all $x \in \mathbb{R}^n$

$$(1-\varepsilon) \cdot \|x\| \le \sqrt{\frac{M}{|I|}} \cdot \left\| R_I A^T x \right\| \le (1+\varepsilon) \cdot \|x\|. \tag{1.6}$$

Throughout this paper C, c etc. denote absolute constants whose value may change from line to line.

The main part of the proof is the proof of Lemma 1 below. Our original proof of this lemma used the direct construction of the majorizing measure. It included an explicit construction of a sequence of partitions and putting weights on the elements of each partition. This scheme is based on the Talagrand and Zinn's proof of the majorizing measure theorem of Fernique (Proposition 2.3 and Theorem 2.5 [T4]). The proof was rather involved, since we had to approximate the natural metric of a random process by a family of metrics depending on the elements of the partition. After we had shown our proof to M. Talagrand, he pointed out that the explicit construction of the partition tree may be substituted by applying his general majorizing measure construction (Theorems 4.2, 4.3 and Proposition 4.4 [T4]). This resulted in a considerable simplification of the proof. We present here the argument suggested by Talagrand.

By the duality between the Kashin and Tzafriri problem and approximate John's decompositions, we have the following

COROLLARY: *Let B be a convex body in $\mathbb{R}^n$ and let $\varepsilon > 0$. There exists a convex body $K \subset \mathbb{R}^n$, so that $d(K, B) \le 1 + \varepsilon$ and the number of contact points of K with its John ellipsoid is less than*

$$m(n, \varepsilon) = C(\varepsilon) \cdot n \cdot \log n.$$

2. The random selection method

Clearly, we may assume that

$$M \ge C \cdot \frac{t^2}{\varepsilon^2} \cdot n \cdot \log n$$

for some absolute constant C.

The proof of the Theorem is based on the following iteration procedure. Let $A = (a_{i,j})$ be an $n \times M$ matrix , satisfying (1.4), We define a sequence $\{\varepsilon_i\}_{i=1}^M$ of independent Bernoulli variables taking values ± 1 with probability $1/2$ and put

$$I_1 = \{i \mid \varepsilon_i = 1\}.$$

Then

$$\frac{M}{2} \cdot \left(1 - \frac{1}{\sqrt{M}}\right) \le |I_1| \le \frac{M}{2} \tag{2.1}$$

with probability at least $1/4$. Define

$$W = A^T \mathbb{R}^n$$

and denote by $w(1), \dots, w(M)$ the coordinates of a vector w. We have to estimate

$$\sup_{x \in B_2^n} \left| 2 \left\| R_I A^T x \right\|^2 - \|x\|^2 \right| = \sup_{w \in W \cap B_2^M} \left| 2 \cdot \sum_{i \in I} w^2(i) - \sum_{i=1}^M w^2(i) \right|$$
$$= \sup_{w \in W \cap B_2^M} \left| \sum_{i=1}^M \varepsilon_i w^2(i) \right|.$$

Denote by $\mathbb{E}X$ the expectation of a random variable X. The key step of the proof is the following

LEMMA 1: *Let W be an n-dimensional subspace of $\mathbb{R}^M$. Let $\varepsilon_1, \dots, \varepsilon_M$ be independent Bernoulli variables taking values ± 1 with probability $1/2$. Then*

$$\mathbb{E} \sup_{w \in W \cap B_2^M} \left| \sum_{i=1}^{M} \varepsilon_i w^2(i) \right| \le C\sqrt{\log M} \cdot \left\| P_W : \ell_1^M \to \ell_2^M \right\|.$$

Here $P_W : \mathbb{R}^M \to \mathbb{R}^M$ is the orthogonal projection onto W.

From (1.4) it follows that

$$\left\| P_W \colon \ell_1^M \to \ell_2^M \right\| \le t \cdot \sqrt{\frac{n}{M}},$$

so by Lemma 1 and Chebychev's inequality we have

$$\sup_{x \in B_2^n} \left| 2 \left\| R_{I_1} A^T x \right\|^2 - \|x\|^2 \right| \le C \cdot t \cdot \sqrt{\frac{n}{M}} \cdot \sqrt{\log M} \tag{2.2}$$

with probability more than $3/4$. Thus, there exists a set $I_1 \subset \{1, \dots, M\}$ satisfying (2.1) and (2.2).

Repeating this procedure, we obtain a sequence of sets $\{1, \dots, M\} = I_0 \supset I_1 \supset \cdots \supset I_s$ so that

$$\frac{|I_k|}{2} \cdot \left(1 - \frac{1}{\sqrt{|I_k|}} \right) \le |I_{k+1}| \le \frac{|I_k|}{2} \tag{2.3}$$

and

$$\sup_{x \in B_2^n} \left| 2^k \left\| R_{I_k} A^T x \right\|^2 - 2^{k-1} \left\| R_{I_{k-1}} A^T x \right\|^2 \right| \le C \cdot t \cdot \sqrt{\frac{n}{M/2^k}} \cdot \sqrt{\log |I_{k-1}|}. \tag{2.4}$$

Indeed, at each step of induction we have

$$\frac{1}{2} \|x\| \le 2^{(k-1)/2} \left\| R_{I_{k-1}} A^T x \right\| \le \frac{3}{2} \|x\|. \tag{2.5}$$

Assume for simplicity that $I_{k-1} = \{1, \dots, m\}$ for some $m < M$. Let $W_k = R_{I_{k-1}} A^T \mathbb{R}^M \subset \mathbb{R}^m$ and let $P_{W_k} \colon \mathbb{R}^m \to \mathbb{R}^m$ be the orthogonal projection onto W_k. Then

$$2^{(k-1)/2} R_{I_{k-1}} A^T B_2^n \subset \frac{3}{2} B_2^m \cap W_k,$$

so for a random set $I_k \subset \{1,\dots,m\}$ we have

$$\mathbb{E}\sup_{x\in B_2^n}\left(2^k\left\|R_{I_k}A^Tx\right\|^2-2^{k-1}\left\|R_{I_{k-1}}A^Tx\right\|^2\right)\leq \mathbb{E}\sup_{w\in\frac{3}{2}B_2^m\cap W_k}\sum_{i=1}^m\varepsilon_iw^2(i)$$
$$\leq\frac{9}{4}\mathbb{E}\sup_{w\in B_2^m\cap W_k}\sum_{i=1}^m\varepsilon_iw^2(i).$$

To apply Lemma 1 we need to compute $\|P_{W_k}:\ell_1^m\to\ell_2^m\|$. By (2.5) we have

$$\|P_{W_k}:\ell_1^m\to\ell_2^m\|\leq 2\cdot\left\|\left(2^{(k-1)/2}R_{I_{k-1}}A^T\right)^T:\ell_1^M\to\ell_2^n\right\|\leq 2^{(k+1/2}\cdot t\cdot\sqrt{\frac{n}{M}}.$$

Now (2.4) follows from Lemma 1 and Chebychev's inequality.

Summing up inequalities (2.4) we get

$$\tag{2.6}\begin{aligned}\sup_{x\in B_2^n}\left|2^s\left\|R_{I_s}A^Tx\right\|^2-\|x\|^2\right|&\leq C\cdot t\cdot\sqrt{\frac{n}{M/2^s}}\cdot\sqrt{\log|I_s|}\\&\leq C\cdot t\cdot\sqrt{\frac{n}{M/2^s}}\cdot\sqrt{\log\frac{M}{2^s}}.\end{aligned}$$

We proceed as long as the last expression is smaller than $\varepsilon/2$. In this case

$$c\cdot\frac{t^2}{\varepsilon^2}\cdot n\cdot\log\frac{nt^2}{\varepsilon^2}\leq\frac{M}{2^s}\leq C\cdot\frac{t^2}{\varepsilon^2}\cdot n\cdot\log\frac{nt^2}{\varepsilon^2}.$$

From (2.3) it follows that

$$\frac{M}{2^s}\cdot\left(1-\frac{4}{\sqrt{|I_s|}}\right)\leq|I_s|\leq\frac{M}{2^s},$$

so we obtain (1.5) and

$$\frac{M}{|I_s|}\cdot\left(1-\left(c\cdot\frac{t^2}{\varepsilon^2}\cdot n\cdot\log n\right)^{-1/2}\right)\leq 2^s\leq\frac{M}{|I_s|}.$$

Then, (2.6) implies that

$$\sup_{x\in B_2^n}\left|\frac{M}{|I_s|}\cdot\left\|R_{I_s}A^Tx\right\|^2-\|x\|^2\right|\leq\varepsilon$$

and this completes the proof of the Theorem. ∎

Remark: The random selection method was first used by Talagrand [T3] to simplify the construction of embedding of a finite dimensional subspace of L_1 into ℓ_1^N. The original construction of Bourgain, Lindenstrauss and Milman used the empirical distribution method instead. The advantage of the random selection is that it enables one to deal with random processes having a subgaussian tail estimate, rather than with general Bernoulli processes.

3. Construction of the majorizing measure

The proof of Lemma 1 uses the majorizing measure theorem of Talagrand [T1], [T4]. This theorem provides a bound to

$$\mathbb{E} \sup_{t \in T} X_t$$

for a subgaussian process X_t indexed by points of a metic space T with a metric d through the geometry of this space. However, it turns out that the space T does not have to be assumed metric. The same proof works in the case when d is a quasimetric, i.e. if there exists a constant L such that for any $t, \bar{t}, s \in T$

$$d(t, \bar{t}) \leq L \cdot \big(d(t,s) + d(s, \bar{t})\big).$$

We use the following version of the

MAJORIZING MEASURE THEOREM: *Let (T, d) be a quasimetric space. Let $(X_t)_{t \in T}$ be a collection of mean 0 random variables with the subgaussian tail estimate*

$$\mathcal{P}\left\{|X_t - X_{\bar{t}}| > a\right\} \leq \exp\left(-c \frac{a^2}{d^2(t, \bar{t})}\right),$$

for all $a > 0$. Let $r > 1$ and let k_0 be a natural number so that T can be covered by one ball of radius r^{-k_0}. Let $\{\varphi_k\}_{k=k_0}^{\infty}$ be a sequence of functions from T to $\mathbb{R}^+$, uniformly bounded by a constant depending only on r. Assume that there exists $\sigma > 0$ so that for any k the functions φ_k satisfy the following condition:

for any $s \in T$ and for any points $t_1, \dots, t_N \in B_{r^{-k}}(s)$ with mutual distances at least r^{-k-1} one has

$$\max_{j=1,\dots,N} \varphi_{k+2}(t_j) \geq \varphi_k(s) + \sigma \cdot r^{-k} \cdot \sqrt{\log N}. \tag{3.1}$$

Then

$$\mathbb{E} \sup_{t \in T} X_t \leq C(r, L) \cdot \sigma^{-1}.$$

This version may be obtained as a combination of the majorizing measure theorem of Fernique [L-T] and the general majorizing measure construction of Talagrand (Theorems 2.1 and 2.2 [T1] or Theorems 4.2, 4.3 and Proposition 4.4 [T4]).

To prove Lemma 1 we need some estimates of covering numbers. Denote by $N(B, d, \varepsilon)$ the ε-entropy of B, i.e. the number of ε-balls in the (quasi-) metric d needed to cover the body B. We use the following

LEMMA 2: *Let W be an n-dimensional subspace of $\mathbb{R}^M$ and let P_W be the orthogonal projection onto W.*

(1) $\varepsilon\sqrt{\log N(B_2^M \cap W, \|\cdot\|_\infty, \varepsilon)} \le C \cdot \|P_W : \ell_1^M \to \ell_2^M\| \cdot \sqrt{\log M}$.

(2) *Let $\|\cdot\|_{\mathcal{E}}$ be a norm defined by*

$$\|x\|_{\mathcal{E}} = \left(\sum_{i=1}^{M} x^2(i) \cdot a_i^2\right)^{1/2}.$$

Then

$$\varepsilon\sqrt{\log N(B_2^M \cap W, \|\cdot\|_{\mathcal{E}}, \varepsilon)} \le C \cdot \|P_W \colon \ell_1^M \to \ell_2^M\| \cdot \left(\sum_{i=1}^{M} a_i^2\right)^{1/2}$$

Proof: Both statements follow from the dual Sudakov minoration proved by Pajor and Tomczak-Jaegermann [P-TJ].

(1) Let g be the standard Gaussian vector in $\mathbb{R}^M$. Then $P_W g$ is the standard Gaussian vector in the space W. So,

$$\varepsilon\sqrt{\log N(B_2^M \cap W, \|\cdot\|_\infty, \varepsilon)} \le C \cdot \mathbb{E}\, \|P_W g\|_\infty = C \cdot \mathbb{E} \max_{j=1,\dots,M} |\langle P_W g, e_j\rangle|$$
$$\le C \cdot \sqrt{\log M} \cdot \max_{j=1,\dots,M} \|P_W e_j\| = C \cdot \sqrt{\log M} \cdot \|P_W \colon \ell_1^M \to \ell_2^M\|. \qquad \blacksquare$$

(2) Again dual Sudakov minoration gives

$$\varepsilon\sqrt{\log N(B_2^M \cap W, \|\cdot\|_{\mathcal{E}}, \varepsilon)} \le C \cdot \mathbb{E}\, \|P_W g\|_{\mathcal{E}} \le C \cdot \left(\mathbb{E}\, \|P_W g\|_{\mathcal{E}}^2\right)^{1/2}$$
$$= C \cdot \left(\mathbb{E} \sum_{i=1}^{M} \langle P_W g, e_i\rangle^2 a_i^2\right)^{1/2} \le C \cdot \max_{i=1,\dots,M} \|P_W e_i\| \cdot \left(\sum_{i=1}^{M} a_i^2\right)^{1/2}. \qquad \blacksquare$$

Proof of Lemma 1: Denote

$$W_1 = B_2^M \cap W.$$

We have to estimate the expectation of the supremum over all $w \in W_1$ of a random process

$$V_w = \sum_{i=1}^{M} \varepsilon_i w^2(i).$$

The process V_w has a subgaussian tail estimate

$$\mathcal{P}\left\{V_w - V_{\bar{w}} > a\right\} \le \exp\left(-c\frac{a^2}{\tilde{d}^2(w, \bar{w})}\right),$$

where

$$\tilde{d}(w,\bar{w}) = \left(\sum_{i=1}^{M}\left(w^2(i)-\bar{w}^2(i)\right)^2\right)^{1/2}.$$

Notice that $\tilde{d}$ can be considered as a metric on the set

$$\{(w^2(1),\dots,w^2(M)) \mid w=(w(1),\dots,w(M)) \in W_1\}.$$

However, it is not a metric on W_1 itself. We shall estimate the function $\tilde{d}$ on W_1 by a quasimetric, which is easier to control:

$$\frac{1}{\sqrt{2}}\tilde{d}(w,\bar{w}) \le d(w,\bar{w}) = \left(\sum_{i=1}^{M}\left(w(i)-\bar{w}(i)\right)^2\cdot\left(w^2(i)+\bar{w}^2(i)\right)\right)^{1/2}.$$

Since

$$\begin{aligned} d(w,\bar{w}) &= \left(\sum_{i=1}^{M}\frac{1}{2}\left(w(i)-\bar{w}(i)\right)^2\cdot\left((w(i)+\bar{w}(i))^2+(w(i)-\bar{w}(i))^2\right)\right)^{1/2}\\ &\le\frac{1}{\sqrt{2}}\cdot\left(\tilde{d}(w,\bar{w})+\|w-\bar{w}\|^2_{\ell_4^M}\right)\le(1+\sqrt{2})\cdot d(w,\bar{w}),\end{aligned}$$

we have a generalized triangle inequality for d. Namely, for all $u,w,\bar{w}\in W$

$$d(w,\bar{w}) \le 4\cdot(d(w,u)+d(u,\bar{w})). \tag{3.2}$$

The balls in the quasimetric d are not convex. However, we have the following

LEMMA 3: *For all $w\in W$ and $\rho>0$*

$$\operatorname{conv} B_\rho(w) \subset B_{4\rho}(w).$$

Here we denote by $B_\rho(w)$ a ρ-ball in the quasimetric d.

Proof: Note that since for all $u\in B_\rho(w)$

$$\left(\sum_{i=1}^{M}\left(u(i)-w(i)\right)^2 w^2(i)\right)^{1/2}\le\rho \text{ and } \left(\sum_{i=1}^{M}\left(u(i)-w(i)\right)^4\right)^{1/4}\le(\sqrt{2}\rho)^{1/2},$$

the same inequalities hold also for all $u\in\operatorname{conv}B_\rho(w)$. Since for all $a,b\in\mathbb{R}$, $a^2+b^2\le 4a^2+2(a-b)^2$, for any $u\in\operatorname{conv}B_\rho(w)$ we have

$$\begin{aligned} d(u,w) &\le\left(\sum_{i=1}^{M}\left(u(i)-w(i)\right)^2\cdot\left(4w^2(i)+2(u(i)-w(i))^2\right)\right)^{1/2}\\ &\le 2\cdot\left(\sum_{i=1}^{M}\left(u(i)-w(i)\right)^2\cdot w^2(i)\right)^{1/2}+\sqrt{2}\cdot\left(\sum_{i=1}^{M}\left(u(i)-w(i)\right)^4\right)^{1/2}\\ &\le 4\rho. \qquad\blacksquare\end{aligned}$$

Denote

$$Q = \left\| P_W : \ell_1^M \to \ell_2^M \right\|.$$

Let now r be a natural number to be chosen later. Let k_0 and k_1 be the largest natural numbers so that

$$r^{-k_0} \geq Q \quad \text{and} \quad r^{-k_1} \geq Q/\sqrt{n}.$$

Then $k_1 - k_0 \leq (2\log r)^{-1} \log n$. Notice that $W_1 \subset r^{-k_0} \cdot (B_\infty^M \cap W) \subset B_{r^{-k_0}}(0)$.

Define functions $\varphi_k \colon W_1 \to \mathbb{R}$ by

$$\varphi_k(w) = \min\{\|u\|^2 \;\big|\; u \in \operatorname{conv} B_{8r^{-k}}(w)\} + \frac{k-k_0}{\log M}, \qquad \text{if } k = k_0, \ldots, k_1,$$

$$\varphi_k(w) = 1 + \frac{1}{2\log r} + \sum_{l=k_1}^{k} r^{-l} \cdot \frac{\sqrt{n \cdot \log(1 + 2\sqrt{2} r^l)}}{Q \cdot \sqrt{\log M}}, \qquad \text{if } k > k_1.$$

For any $w \in W_1$ the sequence $\{\varphi_k(w)\}_{k=k_0}^\infty$ is nonnegative nondecreasing and bounded by an absolute constant depending only on r. Indeed, if $k \leq k_1$ then

$$\varphi_k(w) \leq 1 + \frac{1}{2\log r} \cdot \frac{\log n}{\log M}.$$

For $k > k_1$ we have

$$\begin{aligned}\varphi_k(w) &\leq 1 + \frac{1}{2\log r} + \sum_{l=k_1}^{\infty} r^{-l} \cdot \frac{\sqrt{n \cdot \log(1 + 2\sqrt{2} r^l)}}{Q \cdot \sqrt{\log M}} \\ &\leq 1 + \frac{1}{2\log r} + c(r) \cdot r^{-k_1} \cdot \frac{\sqrt{n}}{Q} \cdot \frac{\sqrt{\log(1 + 2\sqrt{2} r^{k_1})}}{\sqrt{\log M}} \leq C(r).\end{aligned}$$

The last inequality follows from the definition of k_1.

To prove Lemma 1 we have to show that condition (3.1) holds for $\{\varphi_k(w)\}_{k=k_0}^\infty$ with $\sigma = (c \cdot Q \cdot \sqrt{\log M})^{-1}$. Let $x \in W_1$ and suppose that the points $x_1, \ldots, x_N \in B_{r^{-k}}(x)$ satisfy

$$d(x_j, x_l) \geq r^{-k-1} \qquad \text{for all } j \neq l.$$

Assume that $k \geq k_1 - 1$. Since

$$d(u, w) \leq \left(\sum_{i=1}^{M} \left(u^2(i) + w^2(i) \right) \right)^{1/2} \cdot \sup_{i=1,\ldots,M} |u(i) - w(i)| \leq \sqrt{2} \cdot \|u - w\|_\infty,$$

the condition (3.1) follows from a simple volume estimate:

$$\begin{aligned} N \leq & N(W_1, d, r^{-k-1}) \leq N\left(W_1, \|\cdot\|_\infty, \frac{r^{-k-1}}{\sqrt{2}}\right) \leq N\left(W_1, \|\cdot\|, \frac{r^{-k-1}}{\sqrt{2}}\right) \\ \leq & \left(1 + \frac{2\sqrt{2}}{r^{-k-1}}\right)^n . \end{aligned}$$

Suppose now that $k_0 \leq k < k_1 - 1$. For $j = 1, \ldots, N$ denote by z_j the point of $\operatorname{conv} B_{8r^{-k-2}}(x_j)$ for which the minimum of $\|z\|$ is attained and denote by u the similar point of $\operatorname{conv} B_{8r^{-k}}(x)$. By (3.2) and Lemma 3 we have for all $j \neq l$

$$d(x_j, x_l) \leq 16 \cdot \Big(d(x_j, z_j) + d(z_j, z_l) + d(z_l, x_l)\Big) \leq 16 \cdot \Big(16 \cdot r^{-k-2} + d(z_j, z_l)\Big),$$

so $d(z_j, z_l) \geq \frac{1}{2} r^{-k-1}$ if $r \geq 512$. Under the same assumption on r we have

$$d(z_j, x) \leq 4\Big(d(z_j, x_j) + d(x_j, x)\Big) \leq 8r^{-k}.$$

Denote

$$\theta = \max_{j=1,\ldots,N} \|z_j\|^2 - \|u\|^2 .$$

We have to prove that

$$(3.3)\quad r^{-k} \cdot \left(c \cdot Q \cdot \sqrt{\log M}\right)^{-1} \cdot \sqrt{\log N} \leq \max_{j=1,\ldots,N} \varphi_{k+2}(x_j) - \varphi_k(x) = \theta + \frac{2}{\log M}.$$

Since

$$\frac{z_j + u}{2} \in \operatorname{conv} B_{8r^{-k}}(x) \quad \text{and} \quad \|u\| \leq \|z_j\| ,$$

we have

$$\left\|\frac{z_j - u}{2}\right\|^2 = \frac{1}{2}\|z_j\|^2 + \frac{1}{2}\|u\|^2 - \left\|\frac{z_j + u}{2}\right\|^2 \leq \|z_j\|^2 - \left\|\frac{z_j + u}{2}\right\|^2 \leq \|z_j\|^2 - \|u\|^2 ,$$

so

$$(3.4)\qquad \|z_j - u\| \leq 2\sqrt{\theta}.$$

Thus, N is bounded by the $\frac{1}{2}r^{-k-1}$-entropy of the set $K = u + 2\sqrt{\theta}B_2^M \cap W$ in the quasimetric d. To estimate this entropy we partition the set K into S disjoint subsets having diameter less than $\frac{1}{16}r^{-k-1}\theta^{-1/2}$ in the ℓ_∞ metric. By part (1) of Lemma 2 we may assume that

$$(3.5)\qquad \frac{1}{16} r^{-k-1} \cdot \theta^{-1/2} \sqrt{\log S} \leq c \cdot Q \cdot \sqrt{\theta} \sqrt{\log M}.$$

If $S \geq \sqrt{N}$, we are done, because in this case (3.5) implies (3.3). Suppose that $S \leq \sqrt{N}$. Then there exists an element of the partition containing at least $\sqrt{N}$ points z_j. Let $J \subset \{1, \dots, N\}$ be the set of the indices of these points. We have

$$\|z_j - z_l\|_\infty \leq \frac{1}{16} r^{-k-1} \cdot \theta^{-1/2} \tag{3.6}$$

for all $j, l \in J$, $j \neq l$. Since $d(z_j, z_l) \geq \frac{1}{2} r^{-k-1}$, we have

(3.7)

$$\begin{aligned}\left(\frac{1}{2} r^{-k-1}\right)^2 &\leq \sum_{i=1}^M \Big(z_j(i) - z_l(i)\Big)^2 \cdot \Big(z_j^2(i) + z_l^2(i)\Big) \\ &\leq \sum_{i=1}^M \Big(z_j(i) - z_l(i)\Big)^2 \cdot \Big[4u^2(i) + z_j^2(i) \cdot \mathbf{1}_{\{i \,|\, |z_j(i)| \geq 2|u(i)|\}}(i) \\ &\qquad + z_l^2(i) \cdot \mathbf{1}_{\{i \,|\, |z_l(i)| \geq 2|u(i)|\}}(i)\Big].\end{aligned}$$

Then (3.4) implies

$$\sum_{i=1}^M z_j^2(i) \cdot \mathbf{1}_{\{i \,|\, |z_j| \geq 2|u(i)|\}}(i) \leq 4 \sum_{\{i \,|\, |z_j| \geq 2|u(i)|\}} \Big(z_j(i) - u(i)\Big)^2 \leq 16\theta. \tag{3.8}$$

Combining (3.6) and (3.8) we get that (3.7) is bounded by

$$2 \cdot 16\theta \cdot \left(\frac{\theta^{-1/2}}{8} r^{-k-1}\right)^2 + 4 \sum_{i=1}^M \Big(z_j(i) - z_l(i)\Big)^2 \cdot u^2(i).$$

Thus, for all $j, l \in J$, $j \neq l$ we have

$$\left(\sum_{i=1}^M \Big(z_j(i) - z_l(i)\Big)^2 \cdot u^2(i)\right)^{1/2} \geq \frac{1}{8} r^{-k-1}.$$

Then part (2) of Lemma 2 implies

$$\frac{1}{8} r^{-k-1} \sqrt{\log |J|} \leq C\sqrt{\theta} \cdot Q \cdot \left(\sum_{i=1}^M u^2(i)\right)^{1/2} \leq C\sqrt{\theta} \cdot Q.$$

Since for all $\theta > 0$

$$2\sqrt{\theta} \leq \sqrt{\log M} \cdot \theta + \frac{1}{\sqrt{\log M}},$$

we get

$$\frac{1}{16} r^{-k-1} \sqrt{\log N} \leq \frac{1}{8} r^{-k-1} \sqrt{\log |J|} \leq C \cdot Q \cdot \sqrt{\log M} \cdot \left(\theta + \frac{1}{\log M}\right). \qquad \blacksquare$$

ACKNOWLEDGEMENT: I would like to thank Joram Lindenstrauss for helpful discussions and Michel Talagrand for permission to present his approach to the proof of Lemma 1.

References

[K-T] B. Kashin and L. Tzafriri, *Some remarks on the restrictions of operators to coordinate subspaces*, preprint.

[L-T] M. Ledoux and M. Talagrand, *Probability in Banach spaces*, Ergebnisse der Mathematik und ihrer Grenzgebiet, 3 Folge, Vol. 23, Springer, Berlin, 1991.

[P-TJ] A. Pajor and N. Tomczak-Jaegermann, *Subspaces of small codimension of finite dimensional Banach spaces*, Proceedings of the American Mathematical Society **97** (1986), 637–642.

[R] M. Rudelson, *Contact points of convex bodies*, Israel Journal of Mathematics **101** (1997), 93–124.

[T1] M. Talagrand, *Construction of majorizing measures, Bernoulli processes and cotype*, Geometric and Functional Analysis **4** (1994), 660–717.

[T2] M. Talagrand, *Embedding subspaces of L_p in ℓ_p^N*, Operator Theory: Advances and Applications **77** (1995), 311–326.

[T3] M. Talagrand, *Embedding subspaces of L_1 in ℓ_1^N*, Proceedings of the American Mathematical Society **108** (1990), 363–369.

[T4] M. Talagrand, *Majorizing measures: the generic chaining*, Annals of Probability **24** (1996), 1049–1103.