The Johnson-Lindenstrauss lemma almost characterizes Hilbert space, but not quite

William B. Johnson *
Texas A& M University
johnson@math.tamu.edu

Assaf Naor†
Courant Institute
naor@cims.nyu.edu

Abstract

Let X be a normed space that satisfies the Johnson-Lindenstrauss lemma (J-L lemma, in short) in the sense that for any integer n and any $x_1, \ldots, x_n \in X$ there exists a linear mapping $L : X \to F$, where $F \subseteq X$ is a linear subspace of dimension $O(\log n)$, such that $\|x_i - x_j\| \leq \|L(x_i) - L(x_j)\| \leq O(1) \cdot \|x_i - x_j\|$ for all $i, j \in \{1, \ldots, n\}$. We show that this implies that X is almost Euclidean in the following sense: Every n-dimensional subspace of X embeds into Hilbert space with distortion $2^{2^{O(\log^* n)}}$. On the other hand, we show that there exists a normed space Y which satisfies the J-L lemma, but for every n there exists an n-dimensional subspace $E_n \subseteq Y$ whose Euclidean distortion is at least $2^{\Omega(\alpha(n))}$, where α is the inverse Ackermann function.

1 Introduction

The J-L lemma [24] asserts that if H is a Hilbert space, $\varepsilon > 0$, $n \in \mathbb{N}$, and $x_1, \ldots, x_n \in H$ then there exists a linear mapping (even a multiple of an orthogonal projection) $L : H \to F$, where $F \subseteq H$ is a linear subspace of dimension $O(c(\varepsilon) \log n)$, such that for all $i, j \in \{1, \ldots, n\}$ we have

$$\|x_i - x_j\| \leq \|L(x_i) - L(x_j)\| \leq (1 + \varepsilon)\|x_i - x_j\|. \tag{1}$$

This fact has found many applications in mathematics and computer science, in addition to the original application in [24] to a Lipschitz extension problem. The widespread applicability of the J-L lemma in computer science can be (somewhat simplistically) attributed to the fact that it can be viewed as a compression scheme which helps to reduce significantly the space required for storing multidimensional data. We shall not attempt to list here all the applications of the J-L lemma to areas ranging from nearest neighbor search to machine learning—we refer the interested reader to [27, 20, 28, 18, 43, 19, 1] and the references therein for a partial list of such applications.

The applications of (1) involve various requirements from the mapping L. While some applications just need the distance preservation condition (1) and not the linearity of L, most applications require L to be linear. Also, many applications are based on additional information that comes from the proof of the J-L lemma, such as the fact that L arises with high probability from certain distributions over linear mappings. The linearity of L is useful, for example, for fast evaluation of the images $L(x_i)$, and also because these images behave well when additive noise is applied to the initial vectors $x_1, \ldots, x_n$.

*Research supported in part by NSF grants DMS-0503688 and DMS-0528358.
†Research supported in part by NSF grants CCF-0635078 and DMS-0528387.

Due to the usefulness of the J-L lemma there has been considerable effort by researchers to prove such a dimensionality reduction theorem in other normed spaces. All of these efforts have thus far resulted in negative results which show that the J-L lemma fails to hold true in certain non-Hilbertian settings. In [13] Charikar and Sahai proved that there is no dimension reduction via linear mappings in L_1. This negative result was extended to any L_p, $p \in [1,\infty] \setminus \{2\}$, by Lee, Mendel and Naor in [30]. Negative results for dimension reduction without the requirement that the embedding L is linear are known only for the spaces L_1 [9, 31, 30] and L_∞ [7, 25, 3, 33]. Here we show that the negative results for linear dimension reduction in L_p spaces are a particular case of a much more general phenomenon: A normed space that satisfies the J-L lemma is very close to being Euclidean in the sense that all of its n-dimensional subspaces are isomorphic to Hilbert space with distortion $2^{2^{O(\log^*(n))}}$. Here, and in what follows, if $x \geq 1$ then $\log^*(x)$ is the unique integer k such that if we define $a_1 = 1$ and $a_{i+1} = e^{a_i}$ (i.e. a_i is an exponential tower of height i), then $a_k < x \leq a_{k+1}$.

In order to state our results we recall the following notation: The Euclidean distortion of a finite dimensional normed space X, denoted $c_2(X)$, is the infimum over all $D > 0$ such that there exists a linear mapping $S : X \to \ell_2$ which satisfies $\|x\| \leq \|S(x)\| \leq D\|x\|$ for all $x \in X$. Note that in the computer science literature the notation $c_2(X)$ deals with bi-Lipschitz embeddings, but in the context of normed spaces it can be shown that the optimal bi-Lipschitz embedding may be chosen to be linear (this is explained for example in [6, Chapter 7]). The parameter $c_2(X)$ is also known as the Banach-Mazur distance between X and Hilbert space.

Theorem 1.1. *For every $D, K > 0$ there exists a constant $c = c(K, D) > 0$ with the following property. Let X be a Banach space such that for every $n \in \mathbb{N}$ and every $x_1, \ldots, x_n \in X$ there exists a linear subspace $F \subseteq X$, of dimension at most $K \log n$, and a linear mapping $S : X \to F$ such that $\|x_i - x_j\| \leq \|S(x_i) - S(x_j)\| \leq D\|x_i - x_j\|$ for all $i, j \in \{1, \ldots, n\}$. Then for every $k \in \mathbb{N}$ and every k-dimensional subspace $E \subseteq X$, we have*

$$c_2(E) \leq 2^{2^{c \log^*(k)}}. \tag{2}$$

The proof of Theorem 1.1 builds on ideas from [13, 30], while using several fundamental results from the local theory of Banach spaces. Namely, in [30] the L_1 point-set from [13] was analyzed via an analytic argument which extends to any L_p space, $p \neq 2$, rather than the linear programming argument in [13]. In Section 2 we construct a variant of this point-set in any Banach space, and use it in conjunction with some classical results in Banach space theory to prove Theorem 1.1.

The fact that the bound on $c_2(E)$ in (2) is not $O(1)$ is not just an artifact of our iterative proof technique: There do exist non-Hilbertian Banach spaces which satisfy the J-L lemma!

Theorem 1.2. *There exist two universal constants $D, K > 0$ and a Banach space X such that for every $n \in \mathbb{N}$ and every $x_1, \ldots, x_n \in X$ there exists a linear subspace $F \subseteq X$, of dimension at most $K \log n$, and a linear mapping $S : X \to F$ such that $\|x_i - x_j\| \leq \|S(x_i) - S(x_j)\| \leq D\|x_i - x_j\|$ for all $i, j \in \{1, \ldots, n\}$. Moreover, for every integer n the space X has an n-dimensional subspace $F_n \subseteq X$ with*

$$c_2(F_n) \geq 2^{c\alpha(n)}, \tag{3}$$

where $c > 0$ is a universal constant and $\alpha(n) \to \infty$ is the inverse Ackermann function.

We refer the readers to Section 3 for the definition of the inverse Ackermann function. The Banach space X in Theorem 1.2 is the 2-convexification of the Tsirelson space [14], denoted $T^{(2)}$, which we shall now define. The definition below, due to Figiel and Johnson [16], actually gives the dual to the space constructed by Tsirelson (see the book [12] for a comprehensive discussion). Let c_{00} denote the space of

all finitely supported sequences of real numbers. The standard unit basis of c_{00} is denoted by $\{e_i\}_{i=1}^{\infty}$. Given $A \subseteq \mathbb{N}$ we denote by P_A the restriction operator to A, i.e. $P_A\left(\sum_{i=1}^{\infty} x_i e_i\right) = \sum_{i\in A} x_i e_i$. Given two finite subsets $A, B \subseteq \mathbb{N}$ we write $A < B$ if $\max A < \min B$. Define inductively a sequence of norms $\{\|\cdot\|_m\}_{m=0}^{\infty}$ by $\|x\|_0 = \|x\|_{c_0} = \max_{j\geq 1} |x_j|$, and

$$\|x\|_{m+1} = \max\left\{\|x\|_m, \frac{1}{2}\sup\left\{\sum_{j=1}^{n}\left\|P_{A_j}(x)\right\|_m : \ n\in\mathbb{N},\ A_1,\ldots,A_n \subseteq \mathbb{N} \text{ finite}, \{n\} < A_1 < A_2 < \cdots < A_n\right\}\right\} \quad (4)$$

Then for each $x \in c_{00}$ the sequence $\{\|x\|_m\}_{m=0}^{\infty}$ is nondecreasing and bounded from above by $\|x\|_{\ell_1} = \sum_{j=1}^{\infty} |x_j|$. It follows that the limit $\|x\|_T := \lim_{m\to\infty} \|x\|_m$ exists. The space $X = T^{(2)}$ from Theorem 1.2 is the completion of c_{00} under the norm:

$$\left\|\sum_{j=1}^{\infty} x_j e_j\right\|_{T^{(2)}} := \left\|\sum_{j=1}^{\infty} |x_j|^2 e_j\right\|_T^{1/2}. \quad (5)$$

The proof of the fact that $T^{(2)}$ satisfies the J-L lemma consists of a concatenation of several classical results, some of which are quite deep. The lower bound (3) follows from the work of Bellenot [5]. The details are presented in Section 3.

2 Proof of Theorem 1.1

Let $(X, \|\cdot\|)$ be a normed space. The Gaussian type 2 constant of X, denoted $T_2(X)$, is the infimum over all $T > 0$ such that for every $n \in \mathbb{N}$ and every $x_1, \ldots, x_n \in X$ we have,

$$\mathbb{E}\left\|\sum_{i=1}^{n} g_i x_i\right\|^2 \leq T^2 \sum_{i=1}^{n} \|x_i\|^2. \quad (6)$$

Here, and in what follows, $g_1, \ldots, g_n$ denote i.i.d. standard Gaussian random variables. The cotype 2 constant of X, denoted $C_2(X)$, is the infimum over all $C > 0$ such that for every $n \in \mathbb{N}$ and every $x_1, \ldots, x_n \in X$ we have,

$$\sum_{i=1}^{n} \|x_i\|^2 \leq C^2 \mathbb{E}\left\|\sum_{i=1}^{n} g_i x_i\right\|^2. \quad (7)$$

A famous theorem of Kwapien [29] (see also the exposition in [36, Theorem 3.3]) states that

$$c_2(X) \leq T_2(X) \cdot C_2(X). \quad (8)$$

An important theorem of Tomczak-Jaegermann [40] states that if the Banach space X is d-dimensional then there exist $x_1, \ldots, x_d, y_1 \ldots, y_d \in X \setminus \{0\}$ for which

$$\mathbb{E}\left\|\sum_{i=1}^{d} g_i x_i\right\|^2 \geq \frac{T_2(X)^2}{2\pi} \sum_{i=1}^{d} \|x_i\|^2 \quad \text{and} \quad \sum_{i=1}^{d} \|y_i\|^2 \geq \frac{C_2(X)^2}{2\pi} \mathbb{E}\left\|\sum_{i=1}^{d} g_i y_i\right\|^2. \quad (9)$$

In other words, for d-dimensional spaces it suffices to consider $n = d$ in (6) and (7) in order to compute $T_2(X)$ and $C_2(X)$ up to a universal factor. For our purposes it suffices to use the following simpler fact due to Figiel, Lindenstrauss and Milman [17, Lemma 6.1]: If $\dim(X) = d$ then there exist $x_1, \ldots, x_{d(d+1)/2}, y_1, \ldots, y_{d(d+1)/2} \in X \setminus \{0\}$ for which

$$\mathbb{E}\left\|\sum_{i=1}^{d(d+1)/2} g_i x_i\right\|^2 = T_2(X)^2 \sum_{i=1}^{d(d+1)/2} \|x_i\|^2 \quad \text{and} \quad \sum_{i=1}^{d(d+1)/2} \|y_i\|^2 = C_2(X)^2 \mathbb{E}\left\|\sum_{i=1}^{d(d+1)/2} g_i y_i\right\|^2. \tag{10}$$

We note, however, that it is possible to improve the constant terms in Theorem 1.1 if we use (9) instead of (10) in the proof below. We shall now sketch the proof of (10), taken from [17, Lemma 6.1], since this type of finiteness result is used crucially in our proof of Theorem 1.1.

We claim that if $m > d(d+1)/2$ and $u_1, \ldots, u_m \in X$ then there are $v_1, \ldots, v_{m-1}, w_1, \ldots, w_{m-1} \in X$ such that

$$\mathbb{E}\left\|\sum_{i=1}^{m-1} g_i v_i\right\|^2 + \mathbb{E}\left\|\sum_{i=1}^{m-1} g_i w_i\right\|^2 = \mathbb{E}\left\|\sum_{i=1}^{m} g_i u_i\right\|^2 \quad \text{and} \quad \sum_{i=1}^{m-1} \|v_i\|^2 + \sum_{i=1}^{m-1} \|w_i\|^2 = \sum_{i=1}^{m} \|u_i\|^2. \tag{11}$$

Note that (11) clearly implies (10) since it shows that in the definitions (6) and (7) we can take $n = d(d+1)/2$ (in which case the infima in these definitions are attained by a simple compactness argument).

To prove (11) we can think of X as $\mathbb{R}^d$, equipped with a norm $\|\cdot\|$. The random vector $\sum_{i=1}^m g_i u_i = \left(\sum_{i=1}^d g_i u_{ij}\right)_{j=1}^d$ has a Gaussian distribution with covariance matrix $A = (\sum_{i=1}^m u_{ij}u_{ik})_{j,k=1}^d = \sum_{i=1}^m u_i \otimes u_i$. Thus the symmetric matrix A is in the cone generated by the symmetric matrices $\{u_i \otimes u_i\}_{i=1}^m$. By Caratheodory's theorem for cones (see e.g. [15]) we may reorder the vectors u_i so as to find scalars $c_1 \geq c_2 \geq \cdots \geq c_m \geq 0$ with $c_i = 0$ for $i > d(d+1)/2$, such that $A = \sum_{i=1}^m c_i u_i \otimes u_i$. This sum contains at most $d(d+1)/2$ nonzero summands. Define $v_i := \sqrt{c_i/c_1} \cdot u_i$ (so that there are at most $d(d+1)/2 \leq m-1$ nonzero v_i) and $w_i = \sqrt{1 - c_i/c_1} \cdot u_i$ (so that $w_1 = 0$). The second identity in (11) is trivial with these definitions. Now, the random vector $\sum_{i=1}^m g_i v_i$ has covariance matrix $\frac{1}{c_1}\sum_{i=1}^m c_i u_i \otimes u_i = \frac{1}{c_1}A$ and the random vector $\sum_{i=1}^m g_i w_i$ has covariance matrix $\sum_{i=1}^m (1 - c_i/c_1) u_i \otimes u_i = (1 - 1/c_1)A$. Thus $\mathbb{E}\left\|\sum_{i=1}^{m-1} g_i v_i\right\|^2 = \frac{1}{c_1}\mathbb{E}\left\|\sum_{i=1}^m g_i u_i\right\|^2$ and $\mathbb{E}\left\|\sum_{i=1}^{m-1} g_i w_i\right\|^2 = (1 - 1/c_1)\mathbb{E}\left\|\sum_{i=1}^m g_i u_i\right\|^2$. This completes the proof of (11).

We are now in position to prove Theorem 1.1. Define

$$\Delta(n) := \Delta_X(n) := \sup\{c_2(F) : F \subseteq X \text{ linear subspace, } \dim(F) \leq n\}. \tag{12}$$

Note that by John's theorem [21] (see also the beautiful exposition in [4]) $\Delta(n) \leq \sqrt{n}$. Our goal is to obtain a much better bound on $\Delta(n)$. To this end let $F \subseteq X$ be a linear subspace of dimension $k \leq n$. Let m be the integer satisfying $2^{m-1} < k(k+1)/2 \leq 2^m$. We shall use the vectors from (10). By adding some zero vectors so as to have 2^m vectors, and labeling them (for convenience) by the subsets of $\{1, \ldots, m\}$, we obtain $\{x_A\}_{A \subseteq \{1,\ldots,m\}}, \{y_A\}_{A \subseteq \{1,\ldots,m\}} \subseteq X$ such that

$$\mathbb{E}\left\|\sum_{A \subseteq \{1,\ldots,m\}} g_A x_A\right\|^2 = T_2(F)^2 \sum_{A \subseteq \{1,\ldots,m\}} \|x_A\|^2 > 0 \tag{13}$$

$$\sum_{A \subseteq \{1,\ldots,m\}} \|y_A\|^2 = C_2(F)^2 \mathbb{E}\left\|\sum_{A \subseteq \{1,\ldots,m\}} g_A y_A\right\|^2 > 0. \tag{14}$$

For every $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_m) \in \{-1, 1\}$ and $A \subseteq \{1, \ldots, m\}$ consider the Walsh function $W_A(\varepsilon) = \prod_{i \in A} \varepsilon_i$. For every $g = \{g_A\}_{A \subseteq \{1,\ldots,m\}}$ define $\Phi_g, \Psi_g : \{-1, 1\}^m \to F$ by

$$\Phi_g(\varepsilon) \coloneqq \sum_{A \subseteq \{1,\ldots,m\}} g_A W_A(\varepsilon) x_A \quad \text{and} \quad \Psi_g(\varepsilon) \coloneqq \sum_{A \subseteq \{1,\ldots,m\}} g_A W_A(\varepsilon) y_A. \tag{15}$$

Thus Φ_g, Ψ_g are random F-valued functions given by the random Fourier expansions in (15)—the randomness is with respect to the i.i.d. Gaussians $g = \{g_A\}_{A \subseteq \{1,\ldots,m\}}$. These random functions induce the following two random subsets of F:

$$U_g \coloneqq \{\Phi_g(\varepsilon)\}_{\varepsilon \in \{-1,1\}^m} \cup \{x_A\}_{A \subseteq \{1,\ldots,m\}} \cup \{0\} \quad \text{and} \quad V_g \coloneqq \{\Psi_g(\varepsilon)\}_{\varepsilon \in \{-1,1\}^m} \cup \{y_A\}_{A \subseteq \{1,\ldots,m\}} \cup \{0\}.$$

Then $|U_g|, |V_g| \leq 2^{m+1} + 1 \leq 2k(k+1) + 1 \leq 2(n+1)^2$. By the assumptions of Theorem 1.1 it follows that there exist two subspaces $E_g, E'_g \subseteq X$ with $\dim(E_g), \dim(E'_g) \leq K \log\left(2(n+1)^2\right) \leq 4K \log(n+1)$ and two linear mappings $L_g : X \to E_g$, $L'_g : X \to E'_g$, which satisfy

$$x, y \in U_g \implies \|x - y\| \leq \|L_g(x) - L_g(y)\| \leq D\|x - y\|. \tag{16}$$

and

$$x, y \in V_g \implies \|x - y\| \leq \|L'_g(x) - L'_g(y)\| \leq D\|x - y\|. \tag{17}$$

Moreover, by the definition of $\Delta(\cdot)$ there are two linear mappings $S_g : E_g \to \ell_2$ and $S'_g : E'_g \to \ell_2$ which satisfy

$$x \in E_g \implies \|x\| \leq \|S_g(x)\|_2 \leq 2\Delta\left(4K \log(n+1)\right)\|x\|, \tag{18}$$

and

$$x \in E'_g \implies \|x\| \leq \|S'_g(x)\|_2 \leq 2\Delta\left(4K \log(n+1)\right)\|x\|. \tag{19}$$

By the orthogonality of the Walsh functions we see that

$$\mathbb{E}_\varepsilon \left\| S_g\left(L_g\left(\Phi_g(\varepsilon)\right)\right)\right\|_2^2 = \mathbb{E}_\varepsilon \left\| \sum_{A \subseteq \{1,\ldots,m\}} g_A W_A(\varepsilon) S_g(L_g(x_A)) \right\|_2^2 = \sum_{A \subseteq \{1,\ldots,m\}} g_A^2 \left\| S_g(L_g(x_A))\right\|_2^2, \tag{20}$$

and

$$\mathbb{E}_\varepsilon \left\| S'_g\left(L_g\left(\Psi_g(\varepsilon)\right)\right)\right\|_2^2 = \mathbb{E}_\varepsilon \left\| \sum_{A \subseteq \{1,\ldots,m\}} g_A W_A(\varepsilon) S_g(L_g(y_A)) \right\|_2^2 = \sum_{A \subseteq \{1,\ldots,m\}} g_A^2 \left\| S_g(L_g(y_A))\right\|_2^2. \tag{21}$$

A combination of the bounds in (16) and (18) shows that for all $A \subseteq \{1, \ldots, m\}$ we have $\left\|S_g(L_g(x_A))\right\|_2 \leq 2D\Delta\left(4K \log(n+1)\right)\|x_A\|$ and for all $\varepsilon \in \{-1, 1\}^m$ we have $\left\|S_g\left(L_g\left(\Phi_g(\varepsilon)\right)\right)\right\|_2 \geq \left\|\Phi_g(\varepsilon)\right\|$. Thus (20) implies that

$$\mathbb{E}_\varepsilon \left\|\Phi_g(\varepsilon)\right\|^2 \leq 4D^2\Delta\left(4K \log(n+1)\right)^2 \sum_{A \subseteq \{1,\ldots,m\}} g_A^2 \|x_A\|^2. \tag{22}$$

Arguing similarly, while using (17), (19) and (21), we see that

$$\mathbb{E}_\varepsilon \left\|\Psi_g(\varepsilon)\right\|^2 \geq \frac{1}{4D^2\Delta\left(4K\log(n+1)\right)^2} \sum_{A\subseteq\{1,\ldots,m\}} g_A^2 \|y_A\|^2. \tag{23}$$

Taking expectation with respect to the Gaussians $\{g_A\}_{A\subseteq\{1,\ldots,m\}}$ in (22) we see that

$$4D^2\Delta\left(4K\log(n+1)\right)^2 \sum_{A\subseteq\{1,\ldots,m\}} \|x_A\|^2 \geq \mathbb{E}_g\mathbb{E}_\varepsilon \left\|\sum_{A\subseteq\{1,\ldots,m\}} g_A W_A(\varepsilon) x_A\right\|^2 = \mathbb{E}_g \left\|\sum_{A\subseteq\{1,\ldots,m\}} g_A x_A\right\|^2, \tag{24}$$

where we used the fact that for each fixed $\varepsilon \in \{-1,1\}^m$ the random variables $\{W_A(\varepsilon)g_A\}_{A\subseteq\{1,\ldots,m\}}$ have the same joint distribution as the random variables $\{g_A\}_{A\subseteq\{1,\ldots,m\}}$. Similarly, taking expectation in (23) yields

$$\sum_{A\subseteq\{1,\ldots,m\}} \|y_A\|^2 \leq 4D^2\Delta\left(4K\log(n+1)\right)^2 \mathbb{E}_g \left\|\sum_{A\subseteq\{1,\ldots,m\}} g_A y_A\right\|^2. \tag{25}$$

Combining (24) with (13) and (25) with (14) we get the bounds:

$$T_2(F), C_2(F) \leq 2D\Delta\left(4K\log(n+1)\right).$$

In combination with Kwapien's theorem (8) we deduce that

$$c_2(F) \leq T_2(F)C_2(F) \leq 4D^2\Delta\left(4K\log(n+1)\right)^2.$$

Since F was an arbitrary subspace of X of dimension at most n, it follows that

$$\Delta(n) \leq 4D^2\Delta\left(4K\log(n+1)\right)^2. \tag{26}$$

Iterating (26) $\log^*(n)$ times implies that

$$\Delta(n) \leq 2^{2^{c(K,D)\log^*(n)}},$$

as required. □

3 Proof of Theorem 1.2

We shall now explain why the 2-convexification of the Tsirelson space $T^{(2)}$, as defined in the introduction, satisfies the J-L lemma. First we give a definition. Given an increasing sequence $h(n)$ with $0 \leq h(n) \leq n$, say that a Banach space is h-Hilbertian provided that for every finite dimensional subspace E of X there are subspaces F and G of E such that $E = F \oplus G$, $\dim(F) = O(h(\dim E))$ and $c_2(G) = O(1)$[1]. If the Banach space X is log-Hilbertian, then X satisfies the J-L lemma. Indeed, take $x_1, \ldots, x_n \in X$ and let E be their span. Write $E = F \oplus G$ as above and decompose each of the x_i accordingly, i.e. $x_i = y_i \oplus z_i$ where $y_i \in F$ and $z_i \in G$. Since $c_2(G) = O(1)$, by the J-L lemma we can find a linear operator $L : G \to G'$, where $G' \subseteq G$ is a subspace of dimension $O(\log n)$, such that $\|z_i - z_j\| = \Theta(\|L(z_i) - L(z_j)\|)$ for all $i, j \in \{1, \ldots, n\}$. The linear operator

[1]Here the direct sum notation means as usual that for every $y \oplus z \in F \oplus G$ we have $\|y \oplus z\|_X = \Theta\left(\|y\|_X + \|z\|_X\right)$, where the implied constants are independent of E

$L' : E \to F \oplus G'$ given by $L'(y \oplus z) = y \oplus L(z)$ has rank $O(\log n)$ and satisfies $\|x_i - x_j\| = \Theta(\|L'(x_i) - L'(x_j)\|)$, as required.

We now explain why $T^{(2)}$ satisfies the J-L lemma. In [22] Johnson defined the following modification of the Tsirelson space. As in the case of the Tsirelson space, the construction consists of an inductive definition of a sequence of norms on c_{00}. Once again we set $|||x|||_0 = \|x\|_{c_0}$ and

$$|||x|||_{m+1} = \max\left\{|||x|||_m, \frac{1}{2}\sup\left\{\sum_{j=1}^{(n+1)^n} |||P_{A_j}(x)|||_m : \ n \in \mathbb{N},\ A_1, \ldots, A_{(n+1)^n} \subseteq [n, \infty) \text{ finite \& disjoint}\right\}\right\} \quad (27)$$

We then define $\|x\|_{\mathscr{T}} := \lim_{m\to\infty} |||x|||_m$, and the modified space $\mathscr{T}^{(2)}$ as the completion of c_{00} under the norm:

$$\left\|\sum_{j=1}^{\infty} x_j e_j\right\|_{\mathscr{T}^{(2)}} := \left\|\sum_{j=1}^{\infty} |x_j|^2 e_j\right\|_{\mathscr{T}}^{1/2}. \quad (28)$$

In [23] Johnson proved that a certain subspace Y of $\mathscr{T}^{(2)}$ (spanned by a subsequence of the unit vector basis) is log-Hilbertian. In [10] Casazza, Johnson and Tzafriri showed that it is not necessary to pass to the subspace Y, and in fact $\mathscr{T}^{(2)}$ itself has the desired decomposition property. Finally, a deep result of Casazza and Odell [11] shows that $T^{(2)}$ is just $\mathscr{T}^{(2)}$ with an equivalent norm. This concludes the proof of the fact that $T^{(2)}$ satisfies the J-L lemma.

It remains to establish the lower bound (3). Note that the fact that $c_2\left(T^{(2)}\right) = \infty$ already follows from the original paper of Figiel and Johnson [16]—our goal here is to give a quantitative estimate. This will be a simple consequence of a paper of Bellenot [5]. Define inductively a sequence of functions $\{g_k : \mathbb{N} \to \mathbb{N}\}_{k=0}^{\infty}$ as follows: $g_0(n) = n + 1$ and $g_{i+1}(n) = g_i^{(n)}(n)$, where $g_i^{(n)}$ denotes the n-fold iterate of g_i, i.e. $g_i^{(n+1)}(j) = g_i(g_i^{(n)}(j))$. The inverse Ackermann function is the inverse of the function $n \mapsto g_n(n)$, i.e. its value on $n \in \mathbb{N}$ is the unique integer k such that $g_k(k) < n \leq g_{k+1}(k+1)$. Note that in the literature there are several variants of the inverse Ackermann function, but it is possible to show that they are all the same up to bounded additive terms—see, for example, [2, Appendix B] for a discussion of such issues. In particular, we define $\alpha(n)$ to be the inverse of the function $h(n) = g_n(2)$, but its asymptotic behavior is the same as the inverse Ackermann function (since $g_n(2) > n$, and therefore $g_n(n) < g_n(g_n(2)) = g_{n+1}(2)$). Now, by [5, Proposition 5] for every $k \geq 1$ there exist scalars $\{x_j\}_{j=1}^{g_k(2)} \subseteq \mathbb{R}$ which are not all equal to 0 such that

$$\left\|\sum_{j=1}^{g_k(2)} x_j e_j\right\|_T \leq \frac{k}{2^k} \sum_{j=3}^{g_k(2)} |x_j|. \quad (29)$$

Hence, by the definition (5) we have for all $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_{g_k(2)}) \in \{-1, 1\}^{g_k(2)}$,

$$\left\|\sum_{j=1}^{g_k(2)} \varepsilon_j x_j e_j\right\|_{T^{(2)}}^2 \leq \frac{k^2}{2^{2k}} \sum_{j=1}^{g_k(2)} x_j^2. \quad (30)$$

Let $F \subseteq T^{(2)}$ denote the span of $\{e_1, \ldots, e_{g_k(2)}\}$. Averaging (30) over ε and using the definition of the cotype 2 constant of F, we see that $C_2(F) \geq 2^k/k$, and therefore the Euclidean distortion of F is at least $2^k/k$. Since the dimension of F is $g_2(k)$, this concludes the proof of (3), and hence also the proof of Theorem 1.2. □

4 Remarks and open problems

We end this note with some concluding comments and questions that arise naturally from our work.

1. The space $T^{(2)}$ was the first example of what Pisier [37, Chapter 12] calls weak Hilbert spaces. One of the many equivalents for a Banach space X to be a weak Hilbert is that every finite dimensional subspace E of X can be written as $E = F \oplus G$ with $\dim G \geq \delta \dim E$ for some universal constant $\delta > 0$ and $c_2(G) = O(1)$. It is not known whether every weak Hilbert space is log-Hilbertian or even h-Hilbertian for some $h(n) = o(n)$. However, Nielsen and Tomczak-Jaegermann [34], using the same kind of reasoning that works for $T^{(2)}$ (see [10]), proved that a weak Hilbert space with an unconditional basis is even $2^{O(\alpha(\cdot))}$-Hilbertian.

2. A Banach space X is called asymptotically Hilbertian provied that for each n there is a finite codimensional subspace Y of X so that $\Delta_Y(n) = O(1)$ ($\Delta_Y(n)$ is defined in (12)). Every weak Hilbert space is asymptotically Hilbertian [37, Chapter 14]. The results in [23] and the argument at the beginning of section 3 show that every asymptotically Hilbertian space has a subspace which satisfies the J-L lemma.

3. Does there exist a function $f(n) \uparrow \infty$ so that if X is a Banach space for which $\Delta(n) = O(f(n))$, where $\Delta(n)$ is as in (12), i.e. $c_2(E) = O(f(\dim E))$ for all finite dimensional subspaces E of X, then X satisfies the J-L lemma? An affirmative answer would show that there are natural Banach spaces other than Hilbert spaces, even some Orlicz sequence spaces, which satisfy the J-L lemma.

4. A question which obviously arises from our results is to determine the true rate of "closeness" (in the sense of (2)) between spaces satisfying the J-L lemma and Hilbert space. Which of the bounds $\Delta(n) = 2^{2^{O(\log^*(n))}}$ and $\Delta(n) = 2^{\Omega(\alpha(n))}$ is closer to the truth?

5. Our argument also works when the dimension is only assumed to be reduced to a power of $\log n$, and we get nontrivial bounds even when this dimension is, say, $2^{(\log n)^\beta}$ for some $\beta < 1$. However, except for spaces that are of type 2 or of cotype 2, our proof does not yield any meaningful result when the dimension is lowered to n^γ for some $\gamma \in (0, 1)$. The problem is that in the recursive inequality (26) the term $\Delta\left(4K \log(n+1)\right)$ is squared. This happens since in Kwapien's theorem (8) the Euclidean distortion is bounded by the product of the type 2 and cotype 2 constants rather than by their maximum. While it is tempting to believe that the true bound in Kwapien's theorem should be $c_2(X) = O\left(\max\{T_2(X), C_2(X)\}\right)$, it was shown by Tomczak-Jaegermann [41, Proposition 2] that up to universal constants Kwapien's bound $c_2(X) \leq T_2(X)C_2(X)$ cannot be improved.

6. In [35] Pisier proved that if a Banach space X satisfies $\Delta(n) = o(\log n)$, then X is superreflexive; i.e., X admits an equivalent norm which is uniformly convex. Hence any space satisfying the assumptions of Theorem 1.1 is superreflexive.

7. It is of interest to study dimension reduction into arbitrary low dimensional normed spaces, since this can serve just as well for the purpose of data compression (see [32]). Assume that X is a Banach space such that for every n and $x_1, \ldots, x_n \in X$ there exists a $d(n)$-dimensional Banach space Y and a linear mapping $L : X \to Y$ such that $\|x_i - x_j\| \leq \|L(x_i) - L(x_j)\| \leq D\|x_i - x_j\|$ for all $i, j \in \{1, \ldots, n\}$. Since by John's theorem [21] we have $c_2(Y) \leq \sqrt{d(n)}$ we can argue similarly to the proof in Section 2 (in this simpler case the proof is close to the argument in [30]), while using the result of Tomczak-Jaegermann (9), to deduce that $T_2(F), C_2(F) \leq 2\pi\sqrt{d(n)}$. By Kwapien's theorem we deduce that

$c_2(F) \leq 4\pi^2 d(n)$. If $d(n) \leq n^\gamma$ for some $\gamma \in (0,1)$ which is independent of n and F, the fact that $T_2(F), C_2(F) \leq 2\pi n^{\gamma/2}$ for every n-dimensional subspace $F \subseteq X$ implies (see [42]) that X has type $\frac{2}{1+\gamma} - \varepsilon$ and cotype $\frac{2}{1-\gamma} + \varepsilon$ for every $\varepsilon > 0$. In particular, if $d(n) = n^{o(1)}$ then X has type $2 - \varepsilon$ and cotype $2 + \varepsilon$ for every $\varepsilon > 0$.

8. We do not know of any non-trivial linear dimension reduction result in L_p for $p \in [1, \infty) \setminus \{2\}$. For example, is it possible to embed with $O(1)$ distortion via a linear mapping any n-point subset of L_1 into a subspace of L_1 of dimension, say, $n/4$, or even into $\ell_1^{n/4}$? Remarkably even such modest goals seem to be beyond the reach of current techniques. Clearly n-point subsets of L_1 are in their n-dimensional span, but we do not know if they embed with constant distortion into ℓ_1^d when $d = O(n)$. Schechtman proved in [38] that we can take $d = O(n \log n)$. We refer to [38, 8, 39, 26] for more information on the harder problem of embedding n-dimensional subspaces of L_1 into low dimensional ℓ_1^d. We also refer to these references for similar results in L_p spaces.

References

[1] N. Ailon and B. Chazelle. Approximate nearest neighbors and the fast Johnson-Lindenstrauss transform. In *STOC 2006: ACM Symposium on Theory of Computing*, pages 557–563. 2006.

[2] N. Alon, H. Kaplan, G. Nivasch, M. Sharir, and S. Smorodinsky. Weak ϵ-nets and interval chains. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2008, San Francisco, California, USA, January 20-22, 2008*, pages 1194–1203. SIAM, 2008.

[3] J. Arias-de Reyna and L. Rodríguez-Piazza. Finite metric spaces needing high dimension for Lipschitz embeddings in Banach spaces. *Israel J. Math.*, 79(1):103–111, 1992.

[4] K. Ball. An elementary introduction to modern convex geometry. In *Flavors of geometry*, volume 31 of *Math. Sci. Res. Inst. Publ.*, pages 1–58. Cambridge Univ. Press, Cambridge, 1997.

[5] S. F. Bellenot. The Banach space T and the fast growing hierarchy from logic. *Israel J. Math.*, 47(4):305–313, 1984.

[6] Y. Benyamini and J. Lindenstrauss. *Geometric nonlinear functional analysis. Vol. 1*, volume 48 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2000.

[7] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.

[8] J. Bourgain, J. Lindenstrauss, and V. Milman. Approximation of zonoids by zonotopes. *Acta Math.*, 162(1-2):73–141, 1989.

[9] B. Brinkman and M. Charikar. On the impossibility of dimension reduction in l_1. *J. ACM*, 52(5):766–788 (electronic), 2005.

[10] P. G. Casazza, W. B. Johnson, and L. Tzafriri. On Tsirelson's space. *Israel J. Math.*, 47(2-3):81–98, 1984.

[11] P. G. Casazza and E. Odell. Tsirelson's space and minimal subspaces. In *Texas functional analysis seminar 1982–1983 (Austin, Tex.)*, Longhorn Notes, pages 61–72. Univ. Texas Press, Austin, TX, 1983.

[12] P. G. Casazza and T. J. Shura. *Tsirel′son's space*, volume 1363 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1989. With an appendix by J. Baker, O. Slotterbeck and R. Aron.

[13] M. Charikar and A. Sahai. Dimension reduction in the ℓ_1 norm. In *43rd Annual IEEE Conference on Foundations of Computer Science*, pages 251–260. IEEE Computer Society, 2002.

[14] B. S. Cirel′son. It is impossible to imbed 1_p of c_0 into an arbitrary Banach space. *Funkcional. Anal. i Priložen.*, 8(2):57–60, 1974.

[15] L. Danzer, B. Grünbaum, and V. Klee. Helly's theorem and its relatives. In *Proc. Sympos. Pure Math., Vol. VII*, pages 101–180. Amer. Math. Soc., Providence, R.I., 1963.

[16] T. Figiel and W. B. Johnson. A uniformly convex Banach space which contains no l_p. *Compositio Math.*, 29:179–190, 1974.

[17] T. Figiel, J. Lindenstrauss, and V. D. Milman. The dimension of almost spherical sections of convex bodies. *Acta Math.*, 139(1-2):53–94, 1977.

[18] P. Indyk. Algorithmic applications of low-distortion geometric embeddings. In *42nd IEEE Symposium on Foundations of Computer Science (Las Vegas, NV, 2001)*, pages 10–33. IEEE Computer Soc., Los Alamitos, CA, 2001.

[19] P. Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *J. ACM*, 53(3):307–323 (electronic), 2006.

[20] P. Indyk and R. Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In *STOC '98 (Dallas, TX)*, pages 604–613. ACM, New York, 1999.

[21] F. John. Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*, pages 187–204. Interscience Publishers, Inc., New York, N. Y., 1948.

[22] W. B. Johnson. A reflexive Banach space which is not sufficiently Euclidean. *Studia Math.*, 55(2):201–205, 1976.

[23] W. B. Johnson. Banach spaces all of whose subspaces have the approximation property. In *Special topics of applied mathematics (Proc. Sem., Ges. Math. Datenverarb., Bonn, 1979)*, pages 15–26. North-Holland, Amsterdam, 1980.

[24] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, volume 26 of *Contemp. Math.*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984.

[25] W. B. Johnson, J. Lindenstrauss, and G. Schechtman. On Lipschitz embedding of finite metric spaces In low-dimensional normed spaces. In *Geometrical aspects of functional analysis (1985/86)*, pages 177–184. Lecture Notes in Math., 1267, Springer, Berlin, 1987.

[26] W. B. Johnson and G. Schechtman. Finite dimensional subspaces of L_p. In *Handbook of the geometry of Banach spaces, Vol. I*, pages 837–870. North-Holland, Amsterdam, 2001.

[27] J. M. Kleinberg. Two algorithms for nearest-neighbor search in high dimensions. In *STOC '97 (El Paso, TX)*, pages 599–608 (electronic). ACM, New York, 1999.

[28] E. Kushilevitz, R. Ostrovsky, and Y. Rabani. Efficient search for approximate nearest neighbor in high dimensional spaces. *SIAM J. Comput.*, 30(2):457–474 (electronic), 2000.

[29] S. Kwapień. Isomorphic characterizations of inner product spaces by orthogonal series with vector valued coefficients. *Studia Math.*, 44:583–595, 1972. Collection of articles honoring the completion by Antoni Zygmund of 50 years of scientific activity, VI.

[30] J. R. Lee, M. Mendel, and A. Naor. Metric structures in L_1: dimension, snowflakes, and average distortion. *European J. Combin.*, 26(8):1180–1190, 2005.

[31] J. R. Lee and A. Naor. Embedding the diamond graph in L_p and dimension reduction in L_1. *Geom. Funct. Anal.*, 14(4):745–747, 2004.

[32] J. Matoušek. Bi-Lipschitz embeddings into low-dimensional Euclidean spaces. *Comment. Math. Univ. Carolin.*, 31(3):589–600, 1990.

[33] J. Matoušek. On the distortion required for embedding finite metric spaces into normed spaces. *Israel J. Math.*, 93:333–344, 1996.

[34] N. J. Nielsen and N. Tomczak-Jaegermann. Banach lattices with property (H) and weak Hilbert spaces. *Illinois J. Math.*, 36(3):345–371, 1992.

[35] G. Pisier. Martingales with values in uniformly convex spaces. *Israel J. Math.*, 20(3-4):326–350, 1975.

[36] G. Pisier. *Factorization of linear operators and geometry of Banach spaces*, volume 60 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1986.

[37] G. Pisier. *The volume of convex bodies and Banach space geometry*, volume 94 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1989.

[38] G. Schechtman. More on embedding subspaces of L_p in l_r^n. *Compositio Math.*, 61(2):159–169, 1987.

[39] M. Talagrand. Embedding subspaces of L_1 into l_1^N. *Proc. Amer. Math. Soc.*, 108(2):363–369, 1990.

[40] N. Tomczak-Jaegermann. Computing 2-summing norm with few vectors. *Ark. Mat.*, 17(2):273–277, 1979.

[41] N. Tomczak-Jaegermann. The Banach-Mazur distance between symmetric spaces. *Israel J. Math.*, 46(1-2):40–66, 1983.

[42] N. Tomczak-Jaegermann. *Banach-Mazur distances and finite-dimensional operator ideals*, volume 38 of *Pitman Monographs and Surveys in Pure and Applied Mathematics*. Longman Scientific & Technical, Harlow, 1989.

[43] S. S. Vempala. *The random projection method*. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 65. American Mathematical Society, Providence, RI, 2004. With a foreword by Christos H. Papadimitriou.