Séminaire BOURBAKI Janvier 2011
63ème année, 2010-2011, n° 1033

SPARSE QUADRATIC FORMS AND THEIR GEOMETRIC APPLICATIONS
[after Batson, Spielman and Srivastava]

by **Assaf NAOR**

1. INTRODUCTION

In what follows all matrices are assumed to have real entries, and square matrices are always assumed to be symmetric unless stated otherwise. The support of a $k \times n$ matrix $A = (a_{ij})$ will be denoted below by

$$\mathrm{supp}(A) = \big\{(i,j) \in \{1,\dots,k\} \times \{1,\dots,n\} :\ a_{ij} \neq 0\big\}.$$

If A is an $n \times n$ matrix, we denote the decreasing rearrangement of its eigenvalues by

$$\lambda_1(A) \geqslant \lambda_2(A) \geqslant \cdots \geqslant \lambda_n(A).$$

$\mathbb{R}^n$ will always be assumed to be equipped with the standard scalar product $\langle \cdot,\cdot \rangle$. Given a vector $v \in \mathbb{R}^n$ and $i \in \{1,\dots,n\}$, we denote by v_i the ith coordinate of v. Thus for $u, v \in \mathbb{R}^n$ we have $\langle u, v\rangle = \sum_{i=1}^n u_i v_i$.

Our goal here is to describe the following theorem of Batson, Spielman and Srivastava [BSS], and to explain some of its recently discovered geometric applications. We expect that there exist many more applications of this fundamental fact in matrix theory.

Theorem 1.1. — *For every $\varepsilon \in (0,1)$ there exists $c(\varepsilon) = O(1/\varepsilon^2)$ with the following properties. Let $G = (g_{ij})$ be an $n \times n$ matrix with nonnegative entries. Then there exists an $n \times n$ matrix $H = (h_{ij})$ with nonnegative entries that satisfies the following conditions:*

1. $\mathrm{supp}(H) \subseteq \mathrm{supp}(G)$.
2. *The cardinality of the support of H satisfies $|\mathrm{supp}(H)| \leqslant c(\varepsilon)n$.*
3. *For every $x \in \mathbb{R}^n$ we have*

$$\sum_{i=1}^n \sum_{j=1}^n g_{ij}(x_i - x_j)^2 \leqslant \sum_{i=1}^n \sum_{j=1}^n h_{ij}(x_i - x_j)^2 \leqslant (1+\varepsilon) \sum_{i=1}^n \sum_{j=1}^n g_{ij}(x_i - x_j)^2. \quad (1)$$

Supported in part by NSF grant CCF-0832795, BSF grant 2006009, and the Packard Foundation.

The second assertion of Theorem 1.1 is that the matrix H is *sparse*, yet due to the third assertion of Theorem 1.1 the quadratic form $\sum_{i=1}^n \sum_{j=1}^n h_{ij}(x_i - x_j)^2$ is nevertheless a good approximation of the quadratic form $\sum_{i=1}^n \sum_{j=1}^n g_{ij}(x_i - x_j)^2$. For this reason Theorem 1.1 is called in the literature a *sparsification theorem*.

The bound on $|\mathrm{supp}(H)|$ obtained in [BSS] is

$$|\mathrm{supp}(H)| \leqslant 2 \left\lceil \frac{(\sqrt{1+\varepsilon}+1)^4}{\varepsilon^2} n \right\rceil. \tag{2}$$

Thus $c(\varepsilon) \leqslant 32/\varepsilon^2 + O(1/\varepsilon)$. There is no reason to expect that (2) is best possible, but a simple argument [BSS, Section 4] shows that necessarily $c(\varepsilon) \geqslant 8/\varepsilon^2$.

1.1. Historical discussion

The sparsification problem that is solved (up to constant factors) by Theorem 1.1 has been studied for some time in the theoretical computer science literature. The motivations for these investigations were algorithmic, and therefore there was emphasis on constructing the matrix H quickly. We will focus here on geometric applications of Theorem 1.1 for which the existential statement suffices, but we do wish to state that [BSS] shows that H can be constructed in time $O(n^3|\mathrm{supp}(G)|/\varepsilon^2) = O(n^5/\varepsilon^2)$. For certain algorithmic applications this running time is too slow, and the literature contains works that yield weaker asymptotic bounds on $|\mathrm{supp}(H)|$ but have a faster construction time. While such tradeoffs are important variants of Theorem 1.1, they are not directly relevant to our discussion and we will not explain them here. For the applications described below, even a weaker bound of, say, $|\mathrm{supp}(H)| \leqslant c(\varepsilon) n \log n$ is insufficient.

Benczúr and Karger [BK] were the first to study the sparsification problem. They proved the existence of a matrix H with $|\mathrm{supp}(H)| \leqslant c(\varepsilon) n \log n$, that satisfies the conclusion (1) only for *Boolean* vectors $x \in \{0,1\}^n$. In their series of works on fast solvers for certain linear systems [ST1, ST2, ST3, ST4], Spielman and Teng studied the sparsification problem as stated in Theorem 1.1, i.e., with the conclusion (1) holding for *every* $x \in \mathbb{R}^n$. Specifically, in [ST4], Spielman and Teng proved Theorem 1.1 with the weaker estimate $|\mathrm{supp}(H)| = O\left(n(\log n)^7/\varepsilon^2\right)$. Spielman and Srivastava [SS1] improved this estimate on the size of the support of H to $|\mathrm{supp}(H)| = O(n(\log n)/\varepsilon^2)$. As we stated above, Theorem 1.1, which answers positively a conjecture of Spielman-Srivastava [SS1], is due to Batson-Spielman-Srivastava [BSS], who proved this sharp result via a new deterministic iterative technique (unlike the previous probabilistic arguments) that we will describe below. This beautiful new approach does not only yield an asymptotically sharp bound on $|\mathrm{supp}(H)|$: it gives for the first time a deterministic algorithm for constructing H (unlike the previous randomized algorithms), and it also gives additional results that will be described later. We refer to Srivastava's dissertation [Sr2] for a very nice and more complete exposition of these ideas. See also the work of Kolla-Makarychev-Saberi-Teng [KMST] for additional results along these lines.

1.2. Combinatorial interpretation

Suppose that G is the adjacency matrix of the complete graph, i.e., the diagonal entries of G vanish and $g_{ij} = 1$ if $i \neq j$. Assume also that the matrix H of Theorem 1.1 happens to be a multiple of the adjacency matrix of a d-regular graph $\Gamma = (\{1, \dots, n\}, E)$, i.e., for some $\gamma > 0$ and all $i, j \in \{1, \dots, n\}$ we have $h_{ij} = \gamma$ if $\{i, j\} \in E$ and $h_{ij} = 0$ otherwise. Thus $|\mathrm{supp}(H)| = dn$. By expanding the squares in (1) and some straightforward linear algebra, we see that (1) is equivalent to the bound $(\lambda_1(H) - \lambda_n(H))/(\lambda_1(H) - \lambda_2(H)) \leqslant 1 + \varepsilon$. Thus if ε is small then the graph Γ is a good expander (see [HLW] for background on this topic). The Alon-Boppana bound [Ni] implies that H satisfies $(\lambda_1(H) - \lambda_n(H))/(\lambda_1(H) - \lambda_2(H)) \geqslant 1 + 4(1 - o(1))\sqrt{d}$ as $n, d \to \infty$. This lower bound can be asymptotically attained since if Γ is a Ramanujan graph of Lubotzky-Phillips-Sarnak [LPS] then $\lambda_1(H)/\gamma, \lambda_n(H)/\gamma \in \left[-2\sqrt{d-1}, 2\sqrt{d-1}\right]$. Writing $1 + \varepsilon = \left(d + 2\sqrt{d-1}\right) / \left(d - 2\sqrt{d-1}\right) = 1 + 4(1 + o(1))/\sqrt{d}$, we see that the existence of Ramanujan graphs means that (in this special case of the complete graph) there exists a matrix H satisfying (1) with $|\mathrm{supp}(H)| = dn = 16n(1 + o(1))/\varepsilon^2$. The bound on $|\mathrm{supp}(H)|$ in (2) shows that Thereom 1.1 achieves the optimal Ramanujan bound up to a factor of 2. For this reason Batson-Spielman-Srivastava call the matrices produced by Theorem 1.1 "twice-Ramanujan sparsifiers". Of course, this analogy is incomplete since while the matrix H is sparse, it need not be a multiple of the adjacency matrix of a graph, but rather an adjacency matrix of a weighted graph. Moreover, this graph has bounded average degree, rather than being a regular graph of bounded degree. Such weighted sparse (though non-regular) graphs still have useful pseudorandom properties (see [BSS, Lemma 4.1]). Theorem 1.1 can be therefore viewed as a new deterministic construction of "expander-like" weighted graphs, with very good spectral gap. Moreover, it extends the notion of expander graphs since one can start with an arbitrary matrix G before applying the sparsification procedure, with the quality of the resulting expander (measured in terms of absolute spectral gap) being essentially the same as the quality of G as an expander.

1.3. Structure of this paper.

In Section 2 we state a stronger theorem (Theorem 2.1) of Batson-Spielman-Srivastava [BSS], and prove that it implies Theorem 1.1. Section 3 contains the Batson-Spielman-Srivastava proof of this theorem, which is based on a highly original iterative argument. Section 4 contains an application of Theorem 2.1, due to Srivastava [Sr1], to approximate John decompositions. In section 5 we describe two applications of Theorem 2.1, due to Newman-Rabinovich [NR] and Schechtman [Sche3], to dimensionality reduction problems. Section 6 describes the work of Spielman-Srivastava [SS2] that shows how their proof technique for Theorem 2.1 can be used to prove a sharper version of the Bourgain-Tzafriri restricted invertibility principle. Section 7 contains concluding comments and some open problems.

2. A STRONGER THEOREM

Batson-Spielman-Srivastava actually proved a stronger theorem that implies Theorem 1.1. The statement below is not identical to the statement in [BSS], though it easily follows from it. This formulation is stated explicitly as Theorem 1.6 in Srivastava's dissertation [Sr2].

Theorem 2.1. — *Fix $\varepsilon \in (0,1)$ and $m,n \in \mathbb{N}$. For every $x_1,\ldots,x_m \in \mathbb{R}^n$ there exist $s_1,\ldots,s_m \in [0,\infty)$ such that*

$$\left|\left\{i \in \{1,\ldots,m\} : s_i \neq 0\right\}\right| \leqslant \left\lceil \frac{n}{\varepsilon^2} \right\rceil, \tag{3}$$

and for all $y \in \mathbb{R}^n$ we have

$$(1-\varepsilon)^2 \sum_{i=1}^m \langle x_i, y\rangle^2 \leqslant \sum_{i=1}^m s_i \langle x_i, y\rangle^2 \leqslant (1+\varepsilon)^2 \sum_{i=1}^m \langle x_i, y\rangle^2. \tag{4}$$

2.1. Deduction of Theorem 1.1 from Theorem 2.1

Let $G = (g_{ij})$ be an $n \times n$ matrix with nonnegative entries. Note that the diagonal entries of G play no role in the conclusion of Theorem 1.1, so we may assume in what follows that $g_{ii} = 0$ for all $i \in \{1,\ldots,n\}$.

The degree matrix associated to G is defined as usual by

$$D_G = \begin{pmatrix} \sum_{j=1}^n g_{1j} & 0 & \ldots & \ldots & 0 \\ 0 & \sum_{j=1}^n g_{2j} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \sum_{j=1}^n g_{3j} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & \ldots & 0 & \sum_{j=1}^n g_{nj} \end{pmatrix}, \tag{5}$$

and the Laplacian associated to G is defined by

$$\Delta_G = D_G - G = \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n g_{ij}(e_i - e_j) \otimes (e_i - e_j), \tag{6}$$

where $e_1,\ldots,e_n \in \mathbb{R}^n$ is the standard basis of $\mathbb{R}^n$. In the last equation in (6), and in what follows, we use standard tensor notation: for $x,y \in \mathbb{R}^n$ the linear operator $x \otimes y : \mathbb{R}^n \to \mathbb{R}^n$ is given by $(x \otimes y)(z) = \langle x, z\rangle y$.

Theorem 2.1, applied to the vectors $\{\sqrt{g_{ij}}\,(e_i - e_j) :\ i,j \in \{1,\ldots,n\} \wedge i<j\} \subseteq \mathbb{R}^n$, implies that there exist $\{s_{ij} :\ i,j \in \{1,\ldots,n\}\ \wedge\ i<j\} \subseteq [0,\infty)$, at most $\lceil n/\varepsilon^2 \rceil$ of which are nonzero, such that for every $y \in \mathbb{R}^n$ we have

$$\langle \Delta_G y, y\rangle \leqslant \sum_{i=1}^{n-1} \sum_{j=i+1}^n s_{ij} g_{ij} \langle e_i - e_j, y\rangle^2 \leqslant \left(\frac{1+\varepsilon}{1-\varepsilon}\right)^2 \langle \Delta_G y, y\rangle. \tag{7}$$

Extend $(s_{ij})_{i<j}$ to a symmetric matrix by setting $s_{ii} = 0$ and $s_{ji} = s_{ij}$ if $i > j$, and define $H = (h_{ij})$ by $h_{ij} = s_{ij} g_{ij}$. Then $\mathrm{supp}(H) \subseteq \mathrm{supp}(G)$ and $|\mathrm{supp}(H)| \leqslant 2\lceil n/\varepsilon^2 \rceil$.

A straightforward computation shows that $\langle \Delta_G y, y\rangle = \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n g_{ij}(y_i-y_j)^2$ and $\sum_{i=1}^{n-1}\sum_{j=i+1}^n s_{ij}g_{ij}\langle e_i-e_j,y\rangle^2 = \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n h_{ij}(y_i-y_j)^2$. Thus, due to (7) Theorem 1.1 follows, with the bound on $|\mathrm{supp}(H)|$ as in (2). □

3. PROOF OF THEOREM 2.1

Write $A=\sum_{i=1}^m x_i\otimes x_i$. Note that it suffices to prove Theorem 2.1 when A is the $n\times n$ identity matrix I. Indeed, by applying an arbitrarily small perturbation we may assume that A is invertible. If we then set $y_i = A^{-1/2}x_i$ then $\sum_{i=1}^m y_i\otimes y_i = I$, and the conclusion of Theorem 2.1 for the vectors $\{y_1,\ldots,y_m\}$ implies the corresponding conclusion for the original vectors $\{x_1,\ldots,x_m\}$.

The situation is therefore as follows. We are given $x_1,\ldots,x_n\in\mathbb{R}^n$ satisfying

$$\sum_{i=1}^m x_i\otimes x_i = I. \tag{8}$$

Our goal is to find $\{s_i\}_{i=1}^m\subseteq[0,\infty)$ such that at most $\lceil n/\varepsilon^2\rceil$ of them are nonzero, and

$$\frac{\lambda_1\left(\sum_{i=1}^n s_ix_i\otimes x_i\right)}{\lambda_n\left(\sum_{i=1}^n s_ix_i\otimes x_i\right)} \leqslant \left(\frac{1+\varepsilon}{1-\varepsilon}\right)^2. \tag{9}$$

For the ensuing argument it will be convenient to introduce the following notation:

$$\theta = \frac{1+\varepsilon}{1-\varepsilon}. \tag{10}$$

The proof constructs by induction $\{t_k\}_{k=1}^\infty\subseteq[0,\infty)$ and $\{y_k\}_{k=1}^\infty\subseteq\{x_1,\ldots,x_m\}$ with the following properties. Setting $A_0=0$ and $A_i=\sum_{j=1}^i t_jy_j\otimes y_j$ for $i\in\mathbb{N}$, the following inequalities hold true:

$$-\frac{n}{\varepsilon}+i<\lambda_n(A_i)\leqslant\lambda_1(A_i)<\theta\left(\frac{n}{\varepsilon}+i\right), \tag{11}$$

and for every $i\in\mathbb{N}$ we have

$$\sum_{j=1}^n\frac{1}{\theta\left(\frac{n}{\varepsilon}+i\right)-\lambda_j(A_i)} = \sum_{j=1}^n\frac{1}{\theta\left(\frac{n}{\varepsilon}+i-1\right)-\lambda_j(A_{i-1})}, \tag{12}$$

and

$$\sum_{j=1}^n\frac{1}{\lambda_j(A_i)-\left(-\frac{n}{\varepsilon}+i\right)} \leqslant \sum_{j=1}^n\frac{1}{\lambda_j(A_{i-1})-\left(-\frac{n}{\varepsilon}+i-1\right)}. \tag{13}$$

(The sums in (12) and (13) represent the traces of certain matrices constructed from the A_i, and we will soon see that this is the source of their relevance.)

If we continue this construction for $k=\lceil n/\varepsilon^2\rceil$ steps, then by virtue of (11) we would have

$$\frac{\lambda_1(A_k)}{\lambda_n(A_k)} \leqslant \frac{\theta\left(\frac{n}{\varepsilon}+\frac{n}{\varepsilon^2}\right)}{\frac{n}{\varepsilon^2}-\frac{n}{\varepsilon}} = \left(\frac{1+\varepsilon}{1-\varepsilon}\right)^2.$$

By construction $A_k = \sum_{i=1}^m s_i x_i \otimes x_i$ with $s_1, \ldots, s_m \in [0, \infty)$ and at most k of them nonzero. Thus, this process would prove the desired inequality (9).

Note that while for our purposes we just need the spectral bounds in (11), we will need the additional conditions on the resolvent appearing in (12) and (13) in order for us to be able to perform the induction step. Note also that due to (11) all the summands in (12) and (13) are positive.

Suppose that $i \geqslant 1$ and we have already constructed the scalars $t_1, \ldots, t_{i-1} \in [0, \infty)$ and vectors $y_1, \ldots, y_{i-1} \in \{x_1, \ldots, x_m\}$, and let A_{i-1} be the corresponding positive semidefinite matrix. The proof of Theorem 2.1 will be complete once we show that we can find $t_i \geqslant 0$ and $y_i \in \{x_1, \ldots, x_m\}$ so that the matrix $A_i = A_{i-1} + t_i y_i \otimes y_i$ satisfies the conditions (11), (12), (13).

It follows from the inductive hypotheses (11) and (13) that

$$\begin{aligned} 0 < \frac{1}{\lambda_n(A_{i-1}) - \left(-\frac{n}{\varepsilon} + i - 1\right)} &\leqslant \sum_{j=1}^n \frac{1}{\lambda_j(A_{i-1}) - \left(-\frac{n}{\varepsilon} + i - 1\right)} \\ &\leqslant \sum_{j=1}^n \frac{1}{\lambda_j(A_0) - \left(-\frac{n}{\varepsilon}\right)} = \varepsilon < 1. \end{aligned} \tag{14}$$

Hence, since $A_i - A_{i-1}$ is positive semidefinite, $\lambda_n(A_i) \geqslant \lambda_n(A_{i-1}) > -\frac{n}{\varepsilon} + i$, implying the leftmost inequality in (11).

It will be convenient to introduce the following notation:

$$a = \sum_{j=1}^n \frac{1}{\theta\left(\frac{n}{\varepsilon} + i - 1\right) - \lambda_j(A_{i-1})} - \sum_{j=1}^n \frac{1}{\theta\left(\frac{n}{\varepsilon} + i\right) - \lambda_j(A_{i-1})} > 0, \tag{15}$$

and

$$b = \sum_{j=1}^n \frac{1}{\lambda_j(A_{i-1}) - \left(-\frac{n}{\varepsilon} + i\right)} - \sum_{j=1}^n \frac{1}{\lambda_j(A_{i-1}) - \left(-\frac{n}{\varepsilon} + i - 1\right)} > 0. \tag{16}$$

Note that (16) makes sense since, as we have just seen, (14) implies that we have $\lambda_n(A_{i-1}) > -\frac{n}{\varepsilon} + i$. This, combined with (11), shows that the matrices $\theta\left(\frac{n}{\varepsilon} + i\right) I - A_{i-1}$ and $A_{i-1} - \left(-\frac{n}{\varepsilon} + i\right) I$ are positive definite, and hence also invertible. Therefore, for every $j \in \{1, \ldots, m\}$ we can consider the following quantities:

$$\alpha_j = \left\langle \left(\theta\left(\frac{n}{\varepsilon} + i\right) I - A_{i-1}\right)^{-1} x_j, x_j \right\rangle + \frac{1}{a} \left\langle \left(\theta\left(\frac{n}{\varepsilon} + i\right) I - A_{i-1}\right)^{-2} x_j, x_j \right\rangle, \tag{17}$$

and

$$\beta_j = \frac{1}{b} \left\langle \left(A_{i-1} - \left(-\frac{n}{\varepsilon} + i\right) I\right)^{-2} x_j, x_j \right\rangle - \left\langle \left(A_{i-1} - \left(-\frac{n}{\varepsilon} + i\right) I\right)^{-1} x_j, x_j \right\rangle. \tag{18}$$

The following lemma contains a crucial inequality between these quantities.

Lemma 3.1. — *We have* $\sum_{j=1}^m \beta_j \geqslant \sum_{j=1}^m \alpha_j$.

Assuming Lemma 3.1 for the moment, we will show now how to complete the inductive construction. By Lemma 3.1 there exists $j \in \{1, \ldots, m\}$ for which $\beta_j \geqslant \alpha_j$. We will fix this j from now on. Denote

$$t_i = \frac{1}{\alpha_j} \quad \text{and} \quad y_i = x_j. \tag{19}$$

The following formula is straightforward to verify—it is known as the Sherman-Morrison formula (see [GV, Section 2.1.3]): for every invertible $n \times n$ matrix A and every $z \in \mathbb{R}^n$ we have

$$(A + z \otimes z)^{-1} = A^{-1} - \frac{1}{1 + \langle A^{-1} z, z \rangle} A^{-1}(z \otimes z) A^{-1}. \tag{20}$$

Note that $\operatorname{tr}\left(A^{-1}(z \otimes z)A^{-1}\right) = \langle A^{-2}z, z\rangle$. Hence, by taking the trace of the identity (20) we have

$$\operatorname{tr}\left((A + z \otimes z)^{-1}\right) = \operatorname{tr}\left(A^{-1}\right) - \frac{\langle A^{-2} z, z \rangle}{1 + \langle A^{-1} z, z \rangle}. \tag{21}$$

Now, for every $t \in (0, 1/\alpha_j]$ we have

$$\begin{aligned}
\sum_{j=1}^{n} \frac{1}{\theta\left(\frac{n}{\varepsilon} + i\right) - \lambda_j(A_{i-1} + t x_j \otimes x_j)} &= \operatorname{tr}\left(\left(\theta\left(\frac{n}{\varepsilon} + i\right) I - A_{i-1} - t x_j \otimes x_j\right)^{-1}\right) \\
&\overset{(21)}{=} \sum_{j=1}^{n} \frac{1}{\theta\left(\frac{n}{\varepsilon} + i\right) - \lambda_j(A_{i-1})} + \frac{\left\langle \left(\theta\left(\frac{n}{\varepsilon} + i\right) I - A_{i-1}\right)^{-2} x_j, x_j \right\rangle}{\frac{1}{t} - \left\langle \left(\theta\left(\frac{n}{\varepsilon} + i\right) I - A_{i-1}\right)^{-1} x_j, x_j \right\rangle} \\
&\leqslant \sum_{j=1}^{n} \frac{1}{\theta\left(\frac{n}{\varepsilon} + i\right) - \lambda_j(A_{i-1})} + \frac{\left\langle \left(\theta\left(\frac{n}{\varepsilon} + i\right) I - A_{i-1}\right)^{-2} x_j, x_j \right\rangle}{\alpha_j - \left\langle \left(\theta\left(\frac{n}{\varepsilon} + i\right) I - A_{i-1}\right)^{-1} x_j, x_j \right\rangle} \qquad (22) \\
&\overset{(17)}{=} \sum_{j=1}^{n} \frac{1}{\theta\left(\frac{n}{\varepsilon} + i\right) - \lambda_j(A_{i-1})} + a \\
&\overset{(15)}{=} \sum_{j=1}^{n} \frac{1}{\theta\left(\frac{n}{\varepsilon} + i - 1\right) - \lambda_j(A_{i-1})}. \qquad (23)
\end{aligned}$$

In (22) we used the fact that $t \leqslant 1/\alpha_j$ and $\alpha_j > \left\langle \left(\theta\left(\frac{n}{\varepsilon} + i\right) I - A_{i-1}\right)^{-1} x_j, x_j \right\rangle$. In particular, there is equality in (22) if $t = 1/\alpha_j$. As $A_i = A_{i-1} + \frac{1}{\alpha_j} x_j \otimes x_j$, this proves (12). Inequality (23) also implies the rightmost inequality in (11). Indeed, assume for contradiction that $\lambda_1\left(A_{i-1} + \frac{1}{\alpha_j} x_j \otimes x_j\right) \geqslant \theta\left(\frac{n}{\varepsilon} + i\right)$. Since by the inductive hypothesis $\lambda_1(A_{i-1}) < \theta\left(\frac{n}{\varepsilon} + i - 1\right) < \theta\left(\frac{n}{\varepsilon} + i\right)$, it follows by continuity that there exists $t \in (0, 1/\alpha_j]$ for which $\lambda_1\left(A_{i-1} + t x_j \otimes x_j\right) = \theta\left(\frac{n}{\varepsilon} + i\right)$. This value of t would make $\sum_{j=1}^{n} 1/\left(\theta\left(\frac{n}{\varepsilon} + i\right) - \lambda_j(A_{i-1} + t x_j \otimes x_j)\right)$ be infinite, contradicting (23) since by the inductive hypothesis all the summands in the right-hand side of (23) are positive and finite.

It remains to prove (13)—this is the only place where the condition $\beta_j \geqslant \alpha_j$ will be used. We proceed as follows.

$$\begin{aligned}
\sum_{j=1}^{n} \frac{1}{\lambda_j(A_i) - \left(-\frac{n}{\varepsilon}+i\right)} &\overset{(19)}{=} \operatorname{tr}\left(\left(A_{i-1} - \left(-\frac{n}{\varepsilon}+i\right) I + \frac{1}{\alpha_j} x_j \otimes x_j\right)^{-1}\right) \\
&\overset{(21)}{=} \sum_{j=1}^{n} \frac{1}{\lambda_j(A_{i-1}) - \left(-\frac{n}{\varepsilon}+i\right)} - \frac{\left\langle \left(A_{i-1} - \left(-\frac{n}{\varepsilon}+i\right) I\right)^{-2} x_j, x_j \right\rangle}{\alpha_j + \left\langle \left(A_{i-1} - \left(-\frac{n}{\varepsilon}+i\right) I\right)^{-1} x_j, x_j \right\rangle} \\
&\overset{(\beta_j \geqslant \alpha_j)}{\leqslant} \sum_{j=1}^{n} \frac{1}{\lambda_j(A_{i-1}) - \left(-\frac{n}{\varepsilon}+i\right)} - \frac{\left\langle \left(A_{i-1} - \left(-\frac{n}{\varepsilon}+i\right) I\right)^{-2} x_j, x_j \right\rangle}{\beta_j + \left\langle \left(A_{i-1} - \left(-\frac{n}{\varepsilon}+i\right) I\right)^{-1} x_j, x_j \right\rangle} \\
&\overset{(18)}{=} \sum_{j=1}^{n} \frac{1}{\lambda_j(A_{i-1}) - \left(-\frac{n}{\varepsilon}+i\right)} - b \\
&\overset{(16)}{=} \sum_{j=1}^{n} \frac{1}{\lambda_j(A_{i-1}) - \left(-\frac{n}{\varepsilon}+i-1\right)}.
\end{aligned}$$

This concludes the inductive construction, and hence also the proof of Theorem 2.1, provided of course that we prove the crucial inequality contained in Lemma 3.1.

Proof of Lemma 3.1. — It is straightforward to check that the identity (8) implies that for every $n \times n$ matrix A we have

$$\sum_{j=1}^{m} \langle Ax_j, x_j \rangle = \operatorname{tr}(A). \tag{24}$$

Hence,

$$\sum_{j=1}^{m} \alpha_j \overset{(17)\wedge(24)}{=} \operatorname{tr}\left(\left(\theta\left(\frac{n}{\varepsilon}+i\right) I - A_{i-1}\right)^{-1}\right) + \frac{\operatorname{tr}\left(\left(\theta\left(\frac{n}{\varepsilon}+i\right) I - A_{i-1}\right)^{-2}\right)}{a}, \tag{25}$$

and,

$$\sum_{j=1}^{m} \beta_j \overset{(18)\wedge(24)}{=} \frac{\operatorname{tr}\left(\left(A_{i-1} - \left(-\frac{n}{\varepsilon}+i\right) I\right)^{-2}\right)}{b} - \operatorname{tr}\left(\left(A_{i-1} - \left(-\frac{n}{\varepsilon}+i\right) I\right)^{-1}\right). \tag{26}$$

Now,

$$\begin{aligned}
\operatorname{tr}\left(\left(\theta\left(\frac{n}{\varepsilon}+i\right) I - A_{i-1}\right)^{-1}\right) &= \sum_{j=1}^{n} \frac{1}{\theta\left(\frac{n}{\varepsilon}+i\right) - \lambda_j(A_{i-1})} \\
&\leqslant \sum_{j=1}^{n} \frac{1}{\theta\left(\frac{n}{\varepsilon}+i-1\right) - \lambda_j(A_{i-1})} \overset{(12)}{=} \sum_{j=1}^{n} \frac{1}{\frac{\theta n}{\varepsilon} - \lambda_j(A_0)} = \frac{\varepsilon}{\theta},
\end{aligned} \tag{27}$$

and

$$\frac{1}{a}\cdot\operatorname{tr}\left(\left(\theta\left(\frac{n}{\varepsilon}+i\right)I-A_{i-1}\right)^{-2}\right)$$
$$\stackrel{(15)}{=}\frac{\sum_{j=1}^{n}\left(\theta\left(\frac{n}{\varepsilon}+i\right)-\lambda_j(A_{i-1})\right)^{-2}}{\theta\sum_{j=1}^{n}\left(\theta\left(\frac{n}{\varepsilon}+i\right)-\lambda_j(A_{i-1})\right)^{-1}\left(\theta\left(\frac{n}{\varepsilon}+i-1\right)-\lambda_j(A_{i-1})\right)^{-1}}\leqslant\frac{1}{\theta}. \quad (28)$$

Hence,

$$\sum_{j=1}^{n}\alpha_j \stackrel{(25)\wedge(27)\wedge(28)}{\leqslant} \frac{1+\varepsilon}{\theta}\stackrel{(10)}{=}1-\varepsilon. \quad (29)$$

In order to use (26), we first bound b as follows.

$$\begin{aligned}
b &\stackrel{(16)}{=} \sum_{j=1}^{n}\frac{1}{\left(\lambda_{j-1}(A_{i-1})-\left(-\frac{n}{\varepsilon}+i\right)\right)\left(\lambda_{j-1}(A_{i-1})-\left(-\frac{n}{\varepsilon}+i-1\right)\right)}\\
&\leqslant \left(\sum_{j=1}^{n}\frac{1}{\lambda_j(A_{i-1})-\left(-\frac{n}{\varepsilon}+i-1\right)}\right)^{1/2}\\
&\qquad\cdot\left(\sum_{j=1}^{n}\frac{1}{\left(\lambda_{j-1}(A_{i-1})-\left(-\frac{n}{\varepsilon}+i\right)\right)^2\left(\lambda_{j-1}(A_{i-1})-\left(-\frac{n}{\varepsilon}+i-1\right)\right)}\right)^{1/2}\\
&\stackrel{(14)}{\leqslant} \sqrt{\varepsilon}\left(\sum_{j=1}^{n}\frac{1}{\left(\lambda_{j-1}(A_{i-1})-\left(-\frac{n}{\varepsilon}+i\right)\right)^2}-b\right)^{1/2}\\
&\leqslant \left(\sum_{j=1}^{n}\frac{1}{\left(\lambda_{j-1}(A_{i-1})-\left(-\frac{n}{\varepsilon}+i\right)\right)^2}-b\right)^{1/2},
\end{aligned}$$

which simplifies to give the bound

$$\frac{1}{b}\sum_{j=1}^{n}\frac{1}{\left(\lambda_{j-1}(A_{i-1})-\left(-\frac{n}{\varepsilon}+i\right)\right)^2}=\frac{\operatorname{tr}\left(\left(A_{i-1}-\left(-\frac{n}{\varepsilon}+i\right)I\right)^{-2}\right)}{b}\geqslant b+1. \quad (30)$$

Hence,

$$\begin{aligned}
\sum_{j=1}^{m}\beta_j &\stackrel{(26)\wedge(30)}{\geqslant} b+1-\sum_{j=1}^{n}\frac{1}{\lambda_j(A_{i-1})-\left(-\frac{n}{\varepsilon}+i\right)}\\
&\stackrel{(16)}{=} 1-\sum_{j=1}^{n}\frac{1}{\lambda_j(A_{i-1})-\left(-\frac{n}{\varepsilon}+i-1\right)}\stackrel{(14)}{\geqslant}1-\varepsilon. \quad (31)
\end{aligned}$$

Lemma 3.1 now follows from (29) and (31). □

Remark 3.2. — In the inductive construction, instead of ensuring equality in (12), we could have ensured equality in (13) and replaced the equality sign in (12) with the inequality sign $\leqslant$. This would be achieved by choosing $t_i = 1/\beta_j$ in (19). Alternatively we could have chosen t_i to be any value in the interval $[1/\beta_j, 1/\alpha_j]$, in which case both inductive conditions (12) and (13) would be with the inequality sign $\leqslant$.

4. APPROXIMATE JOHN DECOMPOSITIONS

Let $B_2^n \subseteq \mathbb{R}^n$ be the unit ball with respect to the standard Euclidean metric. Recall that an ellipsoid $\mathcal{E} = TB_2^n \subseteq \mathbb{R}^n$ is an image of B_2^n under an invertible linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$. Let $K \subseteq \mathbb{R}^n$ be a centrally symmetric (i.e., $K = -K$) convex body. John's theorem [Jo] states that among the ellipsoids that contain K, there exists a unique ellipsoid of minimal volume. This ellipsoid is called the John ellipsoid of K. If the John ellipsoid of K happens to be B_2^n, the body K is said to be in John position. For any K there is a linear invertible transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ such that TK is in John position. The Banach-Mazur distance between two centrally symmetric convex bodies $K, L \subseteq \mathbb{R}^n$, denoted $d_{\mathrm{BM}}(K, L)$, is the infimum over those $s > 0$ for which there exists a linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$ satisfying $K \subseteq TL \subseteq sK$.

John [Jo] proved that if K is in John position then there exist contact points $x_1, \ldots, x_m \in (\partial K) \cap (\partial B_2^n)$ and positive weights $c_1, \ldots, c_m > 0$ such that

$$\sum_{i=1}^{m} c_i x_i = 0, \tag{32}$$

and

$$\sum_{i=1}^{m} c_i x_i \otimes x_i = I. \tag{33}$$

When conditions (32) and (33) are satisfied we say that $\{x_i, c_i\}_{i=1}^m$ form a John decomposition of the identity. It is hard to overstate the importance of John decompositions in analysis and geometry, and we will not attempt to discuss their applications here. Interested readers are referred to [Bal2] for a taste of this rich field.

John proved that one can always take $m \leqslant n(n + 1)/2$. This bound cannot be improved in general (see [PT] for an even stronger result of this type). However, if one allows an arbitrarily small perturbation of the body K, it is possible to reduce the number of contact points with the John ellipsoid to grow linearly in n. This sharp result is a consequence of the Batson-Spielman-Srivastava sparsification theorem 2.1, and it was proved by Srivastava in [Sr1]. The precise formulation of Srivastava's theorem is as follows.

THEOREM 4.1. — *If $K \subseteq \mathbb{R}^n$ is a centrally symmetric convex body and $\varepsilon \in (0, 1)$ then there exists a convex body $L \subseteq \mathbb{R}^n$ with $d_{\mathrm{BM}}(K, L) \leqslant 1 + \varepsilon$ such that L has at most $m = O(n/\varepsilon^2)$ contact points with its John ellipsoid.*

The problem of perturbing a convex body so as to reduce the size of its John decomposition was studied by Rudelson in [Ru1], where the bound $m \leqslant C(\varepsilon)n(\log n)^3$ was obtained via a randomized construction. In [Ru2] Rudelson announced an improved bound of $m \leqslant C(\varepsilon) n \log n (\log\log n)^2$ using a different probabilistic argument based on majorizing measures, and in [Ru3] Rudelson obtained the bound $m = O(\varepsilon^{-2} n \log n)$, which was the best known bound prior to Srivastava's work.

The key step in all of these proofs is to extract from (33) an approximate John decomposition. This amounts to finding weights $s_1, \ldots, s_m \in [0, \infty)$, such that not many of them are nonzero, and such that we have the operator norm bound $\|I - \sum_{i=1}^m s_i x_i \otimes x_i\| \leqslant \varepsilon$. This is exactly what Theorem 2.1 achieves, with $|\{i \in \{1, \ldots, m\} : \ s_i \neq 0\}| \leqslant c(\varepsilon) n$. Prior to the deterministic construction of Batson-Spielman-Strivastava [BSS], such approximate John decompositions were constructed by Rudelson via a random selection argument, and a corresponding operator-valued concentration inequality. In particular, Rudelson's bound [Ru3] $m = O(\varepsilon^{-2} n \log n)$ uses an influential argument of Pisier. Such methods are important to a variety of applications (see [RV, Tr2]), and in particular this is how Spielman-Srivastava [SS1] proved their earlier $O(\varepsilon^{-2} n \log n)$ sparsification theorem. While yielding suboptimal results, this method is important since it has almost linear (randomized) running time. We refer to the recent work of Adamczak, Litvak, Pajor and Tomczak-Jaegermann for deeper investigations of randomized approximations of certain decompositions of the identity (under additional assumptions).

Proof of Theorem 4.1. — Suppose that K is in John position, and let $\{x_i, c_i\}_{i=1}^n$ be the corresponding John-decomposition. Since $\sum_{i=1}^m (\sqrt{c_i} x_i) \otimes (\sqrt{c_i} x_i) = I$, we may use Theorem 2.1 to find $s_1, \ldots, s_m \geqslant 0$, with at most $O(n/\varepsilon^2)$ of them nonzero, such that if we set $A = \sum_{i=1}^m s_i c_i x_i \otimes x_i$, then the matrices $A - I$ and $(1 + \varepsilon/4)I - A$ are positive semidefinite. Thus $\|A - I\| \leqslant \varepsilon/4$.

The rest of the proof follows the argument in [Ru1, Ru2]. Write $\mathcal{E} = A^{1/2} B_2^n$. Then since $\|A - I\| \leqslant \varepsilon/4$ we have

$$\left(1 - \frac{\varepsilon}{4}\right) \mathcal{E} \subseteq B_2^n \subseteq \left(1 + \frac{\varepsilon}{4}\right) \mathcal{E}.$$

Denote $y_i = x_i / \|A^{-1/2} x_i\|_2 \in \partial \mathcal{E}$ and define

$$H = \operatorname{conv}\left(\{\pm y_i\}_{i \in J} \bigcup \left(\frac{1}{1+\varepsilon} K\right)\right),$$

where $J = \{i \in \{1, \ldots, m\} : s_i \neq 0\}$. Then H is a centrally symmetric convex body, and by a straightforward argument one checks (see [Ru1, Ru2]) that $\frac{1}{1+\varepsilon} K \subseteq H \subseteq (1+2\varepsilon) K$.

Set $L = A^{-1/2} H$. Since $K \subseteq B_2^n$ we have $(\partial H) \cap (\partial \mathcal{E}) = \{\pm y_i\}_{i \in J}$, and therefore $(\partial L) \cap (\partial B_2^n) = \{\pm z_i\}_{i \in J}$, where $z_i = A^{-1/2} y_i$. Writing $a_i = \frac{c_i s_i}{2} \|A^{1/2} x_i\|_2$, we have

$$\begin{aligned} \sum_{i \in J} a_i z_i \otimes z_i + \sum_{i \in J} a_i (-z_i) \otimes (-z_i) &= \sum_{i=1}^m s_i c_i (A^{-1/2} x_i) \otimes (A^{-1/2} x_i) \\ &= A^{-1/2} \left(\sum_{i=1}^m s_i c_i x_i \otimes x_i \right) A^{-1/2} = A^{-1/2} A A^{-1/2} = I. \end{aligned}$$

Hence $\{\pm z_i, a_i\}_{i\in J}$ form a John decomposition of the identity consisting of contact points of L and $B_2^n \supseteq L$. By John's uniqueness theorem [Jo] it follows that B_2^n is the John ellipsoid of L. $\square$

Remark 4.2. — Rudelson [Ru2, Ru3] also studied approximate John decompositions for non-centrally symmetric convex bodies. He proved that Theorem 4.1 holds if K is not necessarily centrally symmetric, with $m = O(\varepsilon^{-2} n \log n)$. Note that in the non-symmetric setting one needs to define the Banach-Mazur appropriately: $d_{\mathrm{BM}}(K, L)$ is the infimum over those $s > 0$ for which there exists $v \in \mathbb{R}^n$ and a linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$ satisfying $K + v \subseteq TL \subseteq s(K + v)$. Srivastava [Sr1], based on a refinement of the proof technique of Theorem 2.1, proved that if $K \subseteq \mathbb{R}^n$ is a convex body and $\varepsilon \in (0, 1)$, then there exists a convex body $L \subseteq \mathbb{R}^n$ with $d_{\mathrm{BM}}(K, L) \leqslant \sqrt{5} + \varepsilon$ such that L has at most $m = O(n/\varepsilon^3)$ contact points with its John ellipsoid. Thus, it is possible to get bounded perturbations with linearly many contact points with the John ellipsoid, but it remains open whether this is possible with $1 + \varepsilon$ perturbations. The problem is how to ensure condition (32) for an approximate John decomposition using the Batson-Spielman-Srivastava technique—for symmetric bodies this is not a problem since we can take the reflections of the points in the approximate John decomposition.

5. DIMENSIONALITY REDUCTION IN L_p SPACES

Fix $p \geqslant 1$. In what follows L_p denotes the space of p-integrable functions on $[0, 1]$ (equipped with Lebesgue measure), and ℓ_p^n denotes the space $\mathbb{R}^n$, equipped with the ℓ_p norm $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$. Since any n-dimensional subspace of L_2 is isometric to ℓ_2^n, for any $x_1, \ldots, x_n \in L_2$ there exist $y_1, \ldots, y_n \in \ell_2^n$ satisfying $\|x_i - x_j\|_2 = \|y_i - y_j\|_2$ for all $i, j \in \{1, \ldots, n\}$. But, more is true if we allow errors: the Johnson-Lindenstrauss lemma [JL] says that for every $x_1, \ldots, x_n \in L_2$, $\varepsilon \in (0, 1)$ there exists $k = O(\varepsilon^{-2} \log n)$ and $y_1, \ldots, y_n \in \ell_2^k$ such that $\|x_i - x_j\|_2 \leqslant \|y_i - y_j\|_2 \leqslant (1 + \varepsilon)\|x_i - x_j\|_2$ for all $i, j \in \{1, \ldots, n\}$. This bound on k is known to be sharp up to a $O(\log(1/\varepsilon))$ factor [Al].

In L_p for $p \neq 2$ the situation is much more mysterious. Any n-points in L_p embed isometrically into ℓ_p^k for $k = n(n - 1)/2$, and this bound on k is almost optimal [Bal1]. If one is interested, as in the Johnson-Lindenstrauss lemma, in embeddings of n-point subsets of L_p into ℓ_p^k with a $1 + \varepsilon$ multiplicative error in the pairwise distances, then the best known bound on k, due to Schechtman [Sche2], was

$$k \leqslant \begin{cases} C(\varepsilon) n \log n & p \in [1, 2), \\ C(p, \varepsilon) n^{p/2} \log n & p \in (2, \infty). \end{cases} \tag{34}$$

We will see now how Theorem 2.1 implies improvements to the bounds in (34) when $p = 1$ and when p is an even integer. The bounds in (34) for $p \notin \{1\} \cup 2\mathbb{N}$ remain the best currently known. We will start with the improvement when $p = 1$, which is due to Newman and Rabinovich [NR]. In the case $p \in 2\mathbb{N}$, which is due to Schechtman [Sche3], more is true: the claimed bound on k holds for embeddings of any n-dimensional linear subspace of L_p into ℓ_p^k, and when stated this way (rather than for n-point subsets of L_p) it is sharp [BDGJN].

5.1. Finite subsets of L_1

It is known that a Johnson-Lindenstrauss type result cannot hold in L_1: Brinkman and Charikar [BC] proved that for any $D > 1$ there exists arbitrarily large n-point subsets $\{x_1, \ldots, x_n\} \subseteq L_1$ with the property that if they embed with distortion D into ℓ_1^k then necessarily $k \geqslant n^{c/D^2}$, where $c > 0$ is a universal constant. Here, and in what follows, a metric space (X, d) is said to embed with distortion D into a normed space Y if there exists $f : X \to Y$ satisfying $d(x, y) \leqslant \|f(x) - f(y)\| \leqslant Dd(x, y)$ for all $x, y \in X$. No nontrivial restrictions on bi-Lipschitz dimensionality reduction are known for finite subsets of L_p, $p \in (1, \infty) \smallsetminus \{2\}$. On the positive side, as stated in (34), Schechtman proved [Sche2] that any n-point subset of L_1 embeds with distortion $1 + \varepsilon$ into ℓ_1^k, for some $k \leqslant C(\varepsilon) n \log n$. The following theorem of Newman and Rabinovich [NR] gets the first asymptotic improvement over Schechtman's 1987 bound, and is based on the Batson-Spielman-Srivastava theorem.

Theorem 5.1. — *For any $\varepsilon \in (0, 1)$, any n-point subset of L_1 embeds with distortion $1 + \varepsilon$ into ℓ_1^k for some $k = O(n/\varepsilon^2)$.*

Proof. — Let $f_1, \ldots, f_n \in L_1$ be distinct. By the cut-cone representation of L_1 metrics, there exists nonnegative weights $\{w_E\}_{E \subseteq \{1, \ldots, n\}}$ such that for all $i, j \in \{1, \ldots, n\}$ we have

$$\|f_i - f_j\|_1 = \sum_{E \subseteq \{1, \ldots, n\}} w_E |\mathbf{1}_E(i) - \mathbf{1}_E(j)|. \tag{35}$$

See [DL] for a proof of (35) (see also [Na, Section 3] for a quick proof).

For every $E \subseteq \{1, \ldots, n\}$ define $x_E = \sqrt{w_E} \sum_{i \in E} e_i \in \mathbb{R}^n$ ($e_1, \ldots, e_n$ is the standard basis of $\mathbb{R}^n$). By Theorem 2.1 there exists a subset $\sigma \subseteq 2^{\{1, \ldots, n\}}$ with $|\sigma| = O(n/\varepsilon^2)$, and nonnegative weights $\{s_E\}_{E \in \sigma}$, such that for every $y \in \mathbb{R}^n$ we have

$$\sum_{E \subseteq \{1, \ldots, n\}} w_E \left(\sum_{i \in E} y_i \right)^2 \leqslant \sum_{E \in \sigma} s_E w_E \left(\sum_{i \in E} y_i \right)^2 \leqslant (1 + \varepsilon) \sum_{E \subseteq \{1, \ldots, n\}} w_E \left(\sum_{i \in E} y_i \right)^2. \tag{36}$$

Define $z_1, \ldots, z_n \in \mathbb{R}^\sigma$ by $z_i = (s_E w_E \mathbf{1}_E(i))_{E \in \sigma}$. For $i, j \in \{1, \ldots, n\}$ apply (36) to the vector $y = e_i - e_j$, noting that for all $E \subseteq \{1, \ldots, n\}$, for this vector y we have $\left(\sum_{i \in E} y_i \right)^2 = |\mathbf{1}_E(i) - \mathbf{1}_E(j)|$.

$$\begin{aligned} \|f_i - f_j\|_1 &\overset{(35)}{=} \sum_{E \subseteq \{1, \ldots, n\}} w_E |\mathbf{1}_E(i) - \mathbf{1}_E(j)| \overset{(36)}{\leqslant} \sum_{E \in \sigma} s_E w_E |\mathbf{1}_E(i) - \mathbf{1}_E(j)| \\ &= \|z_i - z_j\|_1 \overset{(36)}{\leqslant} (1 + \varepsilon) \sum_{E \subseteq \{1, \ldots, n\}} w_E |\mathbf{1}_E(i) - \mathbf{1}_E(j)| \overset{(35)}{=} (1 + \varepsilon) \|f_i - f_j\|_1. \quad \square \end{aligned}$$

Remark 5.2. — Talagrand [Ta1] proved that any n-dimensional linear subspace of L_1 embeds with distortion $1+\varepsilon$ into ℓ_1^k, with $k \leqslant C(\varepsilon) n \log n$. This strengthens Schechtman's bound in (34) for n-point subsets of L_1, since it achieves a low dimensional embedding of their span. It would be very interesting to remove the $\log n$ term in Talagrand's theorem, as this would clearly be best possible. Note that n-point subsets of L_1 can conceivably be embedded into ℓ_1^k, with $k \ll n$. Embedding into at least n dimensions (with any finite distortion) is a barrier whenever the embedding proceeds by actually embedding the span of the given n points. The Newman-Rabinovich argument based on sparsification proceeds differently, and one might hope that it could be used to break the n dimensions barrier for n-point subsets of L_1. This turns out to be possible: the forthcoming paper [ANN] shows that for any $D > 1$, any n-point subset of L_1 embeds with distortion D into ℓ_1^k, with $k = O(n/D)$.

5.2. Finite dimensional subspaces of L_p for even p

Given an $n \in \mathbb{N}$ and $\varepsilon \in (0,1)$, what is the smallest $k \in \mathbb{N}$ such that any n-dimensional subspace of L_p linearly embeds with distortion $1+\varepsilon$ into ℓ_p^k? This problem has been studied extensively [Sche1, Sche2, BLM, Ta1, Ta2, JS1, SZ, Zv, JS2], the best known bound on k being as follows.

$$k \leqslant \begin{cases} C(p,\varepsilon) n \log n (\log\log n)^2 & p \in (0,1) \quad \text{[Zv]}, \\ C(\varepsilon) n \log n & p = 1 \quad \text{[Ta1]}, \\ C(\varepsilon) n \log n (\log\log n)^2 & p \in (1,2) \quad \text{[Ta2]}, \\ C(p,\varepsilon) n^{p/2} \log n & p \in (2,\infty) \quad \text{[BLM]}. \end{cases}$$

In particular, Bourgain, Lindenstrauss and Milman [BLM] proved that if $p \in (2,\infty)$ then one can take $k \leqslant C(p,\varepsilon) n^{p/2} \log n$. It was long known [BDGJN], by considering subspaces of L_p that are almost isometric to ℓ_2^n, that necessarily $k \geqslant c(p,\varepsilon) n^{p/2}$. We will now show an elegant argument of Schechtman, based on Theorem 2.1, that removes the $\log n$ factor when p is an even integer, thus obtaining the first known sharp results for some values of $p \neq 2$.

THEOREM 5.3. — *Assume that $p > 2$ is an even integer, $n \in \mathbb{N}$ and $\varepsilon \in (0,1)$. Then any n-dimensional subspace X of L_p embeds with distortion $1+\varepsilon$ into ℓ_p^k for some $k \leqslant (cn/p)^{p/2}/\varepsilon^2$, where c is a universal constant.*

Proof. — By a standard argument (approximating a net in the sphere of X by simple functions), we may assume that $X \subseteq \ell_p^m$ for some finite (huge) $m \in \mathbb{N}$. In what follows, when we use multiplicative notation for vectors in $\mathbb{R}^m$, we mean coordinate-wise products, i.e., for $x, y \in \mathbb{R}^m$, write $xy = (x_1 y_1, \ldots, x_m y_m)$ and for $r \in \mathbb{N}$ write $x^r = (x_1^r, \ldots, x_m^r)$.

Let $u_1, \ldots, u_n$ be a basis of X. Consider the following subspace of $\mathbb{R}^m$:

$$Y = \operatorname{span}\left(\left\{ u_{j_1}^{p_1} u_{j_2}^{p_2} \cdots u_{j_\ell}^{p_\ell} :\ \ell \in \mathbb{N},\ j_1, \ldots, j_\ell \in \{1, \ldots, n\},\ p_1 + \cdots + p_\ell = \frac{p}{2} \right\}\right).$$

Then

$$d = \dim(Y) \leqslant \binom{n+p/2-1}{p/2} \leqslant \left(\frac{10n}{p}\right)^{p/2}.$$

Thinking of Y as a d-dimensional subspace of ℓ_2^m, let $v_1, \ldots, v_d$ be an orthonormal basis of Y. Define $x_1, \ldots, x_m \in Y$ by $x_i = \sum_{j=1}^d \langle v_j, e_i\rangle v_j$, where as usual $e_1, \ldots, e_m$ is the standard coordinate basis of $\mathbb{R}^m$. Note that by definition (since $v_1, \ldots, v_d$ is an orthonormal basis of Y), for every $y \in Y$ and for every $i \in \{1, \ldots, m\}$ we have $\langle x_i, y\rangle = \langle y, e_i\rangle = y_i$. By Theorem 2.1, there exists a subset $\sigma \subseteq \{1, \ldots, m\}$ with $|\sigma| = O(d/(p\varepsilon)^2) \leqslant (cn/p)^{p/2}/\varepsilon^2$, and $\{s_i\}_{i\in\sigma} \subseteq (0, \infty)$, such that for all $y \in Y$ we have

$$\sum_{i=1}^m y_i^2 \leqslant \sum_{i\in\sigma} s_i y_i^2 \leqslant \left(1+\frac{\varepsilon p}{4}\right)\sum_{i=1}^m y_i^2. \tag{37}$$

In particular, since by the definition of Y for every $x \in X$ we have $x^{p/2} \in Y$,

$$\|x\|_p = \left(\sum_{i=1}^m x_i^p\right)^{1/p} \overset{(37)}{\leqslant} \left(\sum_{i\in\sigma} s_i x_i^p\right)^{1/p} \overset{(37)}{\leqslant} \left(1+\frac{\varepsilon p}{4}\right)^{1/p} \left(\sum_{i=1}^m x_i^p\right)^{1/p} \leqslant (1+\varepsilon)\|x\|_p.$$

Thus $x \mapsto (s_i^{1/p} x_i)_{i\in\sigma}$ maps X into $\ell_p^\sigma \subseteq \ell_p^{(cn/p)^{p/2}/\varepsilon^2}$ and has distortion $1+\varepsilon$. □

Remark 5.4. — The bound on k in Theorem 5.3 is sharp also in terms of the dependence on p. See [Sche3] for more information on this topic.

6. THE RESTRICTED INVERTIBILITY PRINCIPLE

In this section square matrices are no longer assumed to be symmetric. The ensuing discussion does not deal with a direct application of the statement of Theorem 2.1, but rather with an application of the method that was introduced by Batson-Spielman-Srivastava to prove Theorem 2.1.

Bourgain and Tzafriri studied in [BT1, BT2, BT3] conditions on matrices which ensure that they have large "well invertible" sub-matrices, where well invertibility refers to control of the operator norm of the inverse. Other than addressing a fundamental question, such phenomena are very important to a variety of interesting applications that we will not survey here.

To state the main results of Bourgain-Tzafriri, we need the following notation. For $\sigma \subseteq \{1, \ldots, n\}$ let $R_\sigma : \mathbb{R}^n \to \mathbb{R}^\sigma$ be given by restricting the coordinates to σ, i.e., $R_\sigma(\sum_{i=1}^n a_i e_i) = \sum_{i\in\sigma} a_i e_i$ (as usual, $\{e_i\}_{i=1}^n$ is the standard coordinate basis of $\mathbb{R}^n$). In matrix notation, given an operator $T : \mathbb{R}^n \to \mathbb{R}^n$, the operator $R_\sigma T R_\sigma^* : \mathbb{R}^\sigma \to \mathbb{R}^\sigma$ corresponds to the $\sigma \times \sigma$ sub-matrix $(\langle Te_i, e_j\rangle)_{i,j\in\sigma}$. The operator norm of T (as an operator from ℓ_2^n to ℓ_2^n) will be denoted below by $\|T\|$, and the Hilbert-Schmidt norm of T will be denoted $\|T\|_{\mathrm{HS}} = \sqrt{\sum_{i=1}^n \sum_{j=1}^n \langle Te_i, e_j\rangle^2}$.

The following theorem from [BT1, BT3] is known as the Bourgain-Tzafriri restricted invertibility principle.

THEOREM 6.1. — *There exist universal constants $c, K > 0$ such that for every $n \in \mathbb{N}$ and every linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$ the following assertions hold true:*

1. *If $\|Te_i\|_2 = 1$ for all $i \in \{1,\dots,n\}$ then there exists a subset $\sigma \subseteq \{1,\dots,n\}$ satisfying*

$$|\sigma| \geqslant \frac{cn}{\|T\|^2}, \tag{38}$$

such that $R_\sigma T^* T R_\sigma^*$ is invertible and

$$\left\|(R_\sigma T^* T R_\sigma^*)^{-1}\right\| \leqslant K. \tag{39}$$

2. *If $\langle Te_i, e_i\rangle = 1$ for all $i \in \{1,\dots,n\}$ then for all $\varepsilon \in (0,1)$ there exists a subset $\sigma \subseteq \{1,\dots,n\}$ satisfying*

$$|\sigma| \geqslant \frac{c\varepsilon^2 n}{\|T\|^2}, \tag{40}$$

such that $R_\sigma T^* T R_\sigma^*$ is invertible and

$$\left\|(R_\sigma T^* T R_\sigma^*)^{-1}\right\| \leqslant 1+\varepsilon. \tag{41}$$

The quadratic dependence on ε in (40) cannot be improved [BHKW]. Observe that (39) is equivalent to the following assertion:

$$\left\|\sum_{i\in\sigma} a_i Te_i\right\|_2^2 \geqslant \frac{1}{K}\sum_{i\in\sigma} a_i^2 \qquad \forall \{a_i\}_{i\in\sigma} \subseteq \mathbb{R}. \tag{42}$$

We note that if T satisfies the assumption of the first assertion of Theorem 6.1 then T^*T satisfies the assumption of the second assertion of Theorem 6.1. Hence, the second assertion of Theorem 6.1 implies the first assertion of Theorem 6.1 with (39) replaced by $\left\|(R_\sigma T^* T R_\sigma^*)^{-1}\right\| \leqslant (1+\varepsilon)\|T\|^2$ and (38) replaced by the condition $|\sigma| \geqslant c\varepsilon^2 n/\|T\|^4$.

In [SS2] Spielman and Srivastava proved the following theorem:

THEOREM 6.2. — *Suppose that $x_1,\dots,x_m \in \mathbb{R}^n \smallsetminus \{0\}$ satisfy*

$$\sum_{i=1}^m x_i \otimes x_i = I. \tag{43}$$

Then for every linear $T : \mathbb{R}^n \to \mathbb{R}^n$ and $\varepsilon \in (0,1)$ there exists $\sigma \subseteq \{1,\dots,m\}$ with

$$|\sigma| \geqslant \left\lfloor \frac{\varepsilon^2\|T\|_{\mathrm{HS}}^2}{\|T\|^2} \right\rfloor, \tag{44}$$

and such that for all $\{a_i\}_{i\in\sigma} \subseteq \mathbb{R}$ we have

$$\left\|\sum_{i\in\sigma} a_i T x_i\right\|_2^2 \geqslant \frac{(1-\varepsilon)^2\|T\|_{\mathrm{HS}}^2}{m}\sum_{i\in\sigma} a_i^2. \tag{45}$$

Theorem 6.2 implies the Bourgain-Tzafriri restricted invertibility principle. Indeed, take $x_i = e_i$ and note that if either $\|Te_i\|_2 = 1$ for all $i \in \{1, \ldots, n\}$ or $\langle Te_i, e_i\rangle = 1$ for all $i \in \{1, \ldots, n\}$ then $\|T\|_{\mathrm{HS}}^2 \geqslant n$. The idea to improve the Bourgain-Tzafriri theorem in terms of Hilbert-Schmidt estimates is due to Vershynin, who proved in [Ve] a statement similar to Theorem 6.2 (with asymptotically worse dependence on ε). Among the tools used in Vershynin's argument is the Bourgain-Tzafriri restricted invertibility theorem itself, but we will see how the iterative approach of Section 3 yields a self-contained and quite simple proof of Theorem 6.2. This new approach of Spielman-Srivastava has other advantages. Over the years, there was interest [BT1, BT3, Tr1, CT] in improving the quantitative estimates in Theorem 6.1 (i.e., the bounds on c, K, and the dependence $|\sigma|$ on ε and $\|T\|$), and Theorem 6.2 yields the best known bounds. Moreover, it is not obvious that the subset σ of Theorem 6.1 can be found in polynomial time. A randomized algorithm achieving this was recently found by Tropp [Tr1], and the work of Spielman-Srivastava yields a determinstic algorithm which finds in polynomial time a subset σ satisfying the assertions of Theorem 6.2.

Before proceeding to an exposition of the proof of Theorem 6.2 in [SS2], we wish to note that another important result of Bourgain-Tzafriri [BT1, BT2] is the following theorem, which is easily seen to imply the second assertion of Theorem 6.1 with the conclusion (41) replaced by $\left\|(R_\sigma T R_\sigma^*)^{-1}\right\| \leqslant 1+\varepsilon$. This theorem is important for certain applications, and it would be interesting if it could be proved using the Spielman-Srivastava method as well.

Theorem 6.3. — *There is a universal constant $c > 0$ such that for every $\varepsilon > 0$ and $n \in \mathbb{N}$ if an operator $T : \mathbb{R}^n \to \mathbb{R}^n$ satisfies $\langle Te_i, e_i\rangle = 0$ for all $i \in \{1, \ldots, n\}$ then there exists a subset $\sigma \subseteq \{1, \ldots, n\}$ satisfying $|\sigma| \geqslant c\varepsilon^2 n$ and $\|R_\sigma T R_\sigma^*\| \leqslant \varepsilon\|T\|$.*

6.1. Proof of Theorem 6.2

The conclusion (45) of Theorem 6.2 is equivalent to the requirement that the matrix

$$A = \sum_{i\in\sigma} (Tx_i) \otimes (Tx_i) \tag{46}$$

has $|\sigma|$ eigenvalues at least $(1-\varepsilon)^2\|T\|_{\mathrm{HS}}^2/m$. Indeed, if B is the $|\sigma| \times n$ matrix whose rows are $\{Tx_i\}_{i\in\sigma}$, then $A = B^*B$. The eigenvalues of A are therefore the same as the eigenvalues of the $|\sigma| \times |\sigma|$ Gram matrix $BB^* = (\langle Tx_i, Tx_j\rangle)_{i,j\in\sigma}$. The assertion that all the eigenvalues of BB^* are at least $(1-\varepsilon)^2\|T\|_{\mathrm{HS}}^2/m$ is identical to (45).

Define

$$k = \left\lfloor \frac{\varepsilon^2\|T\|_{\mathrm{HS}}^2}{\|T\|^2} \right\rfloor. \tag{47}$$

We will construct inductively $y_0, y_1, \ldots, y_k \in \mathbb{R}^n$ with the following properties. We set $y_0 = 0$ and require that $y_1, \ldots, y_k \in \{x_1, \ldots, x_m\}$. Moreover, if for $i \in \{0, \ldots, k\}$ we write

$$b_i = \frac{(1-\varepsilon)}{m}\left(\|T\|_{\mathrm{HS}}^2 - \frac{i}{\varepsilon}\|T\|^2\right), \tag{48}$$

then the matrix

$$A_i = \sum_{j=0}^{i} (Ty_j) \otimes (Ty_j) \qquad (49)$$

has k eigenvalues bigger than b_i and all its other eigenvalues equal 0 (this holds vacuously for $i = 0$). Note that this requirement implies in particular that $y_1, \dots, y_k$ are distinct. Finally, we require that for every $i \in \{1, \dots, k\}$ we have

$$\sum_{j=1}^{m} \left\langle (A_i - b_i I)^{-1} Tx_j, Tx_j \right\rangle < \sum_{j=1}^{m} \left\langle (A_{i-1} - b_{i-1} I)^{-1} Tx_j, Tx_j \right\rangle . \qquad (50)$$

The matrix A_k will then have the form (46) with $|\sigma| = k$, and have k eigenvalues greater than $(1-\varepsilon)^2 \|T\|_{\mathrm{HS}}^2 / m$, as required. It remains therefore to show that for $i \in \{1, \dots, k\}$ there exists a vector y_i satisfying the desired properties, assuming that $y_0, y_1, \dots, y_{i-1}$ have already been selected.

Lemma 6.4. — *Denote*

$$\mu = \sum_{j=1}^{m} \left\langle (A_{i-1} - b_{i-1} I)^{-1} Tx_j, Tx_j \right\rangle - \sum_{j=1}^{m} \left\langle (A_{i-1} - b_i I)^{-1} Tx_j, Tx_j \right\rangle . \qquad (51)$$

(Since $b_i \in (0, b_{i-1})$, the matrix $(A_{i-1} - b_i I)^{-1}$ makes sense in (51).) Then

$$\sum_{j=1}^{m} \left\langle (A_{i-1} - b_i I)^{-1} TT^* (A_{i-1} - b_i I)^{-1} Tx_j, Tx_j \right\rangle$$
$$< -\mu \sum_{j=1}^{m} \left(1 + \left\langle (A_{i-1} - b_i I)^{-1} Tx_j, Tx_j \right\rangle \right) . \qquad (52)$$

Assuming the validity of Lemma 6.4 for the moment, we will show how to complete the inductive construction. By (52) there exists $j \in \{1, \dots, m\}$ satisfying

$$\left\langle (A_{i-1} - b_i I)^{-1} TT^* (A_{i-1} - b_i I)^{-1} Tx_j, Tx_j \right\rangle$$
$$< -\mu \left(1 + \left\langle (A_{i-1} - b_i I)^{-1} Tx_j, Tx_j \right\rangle \right) . \qquad (53)$$

Our inductive choice will be $y_i = x_j$.

The matrix $(A_{i-1} - b_{i-1} I)^{-1} - (A_{i-1} - b_i I)^{-1}$ is positive definite, since by the inductive hypothesis its eigenvalues are all of the form $(\lambda - b_{i-1})^{-1} - (\lambda - b_i)^{-1}$ for some $\lambda \in \mathbb{R}$ that satisfies $\lambda > b_{i-1} > b_i$ or $\lambda = 0$ (and since for $i \leqslant k$ we have $b_i > 0$). Hence $\mu \geqslant 0$. Since the left hand side of (53) is nonnegative, it follows that

$$1 + \left\langle (A_{i-1} - b_i I)^{-1} Tx_j, Tx_j \right\rangle < 0. \qquad (54)$$

Since $A_i = A_{i-1} + (Tx_j) \otimes (Tx_j)$, it follows from (21) that

$$\operatorname{tr}\left((A_i - b_i I)^{-1} \right) - \operatorname{tr}\left((A_{i-1} - b_i I)^{-1} \right) = -\frac{\left\langle (A_{i-1} - b_i I)^{-2} Tx_j, Tx_j \right\rangle}{1 + \left\langle (A_{i-1} - b_i I)^{-1} Tx_j, Tx_j \right\rangle} \overset{(54)}{>} 0, \qquad (55)$$

where in the last inequality of (55) we used the fact that $(A_{i-1} - b_i I)^{-2}$ is positive definite. At the same time, by the inductive hypothesis $\lambda_1(A_{i-1}), \ldots, \lambda_{i-1}(A_{i-1}) > b_{i-1}$, and $\lambda_i(A_{i-1}) = \cdots = \lambda_n(A_{i-1}) = 0$. Since $A_i - A_{i-1}$ is a rank one positive semidefinite matrix, the eigenvalues of A_i and A_{i-1} interlace (see [Bah, Section III.2]; this result goes back to [Wey]), and therefore

$$\lambda_1(A_i) \geqslant \lambda_1(A_{i-1}) \geqslant \lambda_2(A_i) \geqslant \lambda_2(A_{i-1}) \geqslant \cdots \geqslant \lambda_{i-1}(A_{i-1}) \geqslant \lambda_i(A_i), \tag{56}$$

and

$$\lambda_i(A_{i-1}) = \cdots = \lambda_n(A_{i-1}) = \lambda_{i+1}(A_i) = \cdots = \lambda_n(A_i) = 0. \tag{57}$$

Hence,

$$\begin{aligned} 0 &\overset{(55)}{<} \operatorname{tr}\left((A_i - b_i I)^{-1}\right) - \operatorname{tr}\left((A_{i-1} - b_i I)^{-1}\right) \\ &\overset{(57)}{=} \frac{1}{\lambda_i(A_i) - b_i} + \frac{1}{b_i} + \sum_{j=1}^{i-1} \left(\frac{1}{\lambda_j(A_i) - b_i} - \frac{1}{\lambda_j(A_{i-1}) - b_i} \right) \overset{(56)}{\leqslant} \frac{\lambda_i(A_i)}{b_i(\lambda_i(A_i) - b_i)}, \end{aligned}$$

implying that $\lambda_i(A_i) > b_i$.

Therefore, in order to establish the inductive step, it remains to prove (50). To this end, note that due to (43) and (24) for every $n \times n$ matrix A we have

$$\sum_{j=1}^{m} \langle A T x_j, T x_j \rangle = \operatorname{tr}\left(T^* A T\right). \tag{58}$$

Hence (50) is equivalent to the inequality

$$\operatorname{tr}\left(T^* \left(A_{i-1} - b_{i-1} I\right)^{-1} T\right) > \operatorname{tr}\left(T^* \left(A_i - b_i I\right)^{-1} T\right). \tag{59}$$

Now,

$$\begin{aligned} &\operatorname{tr}\left(T^* \left(A_i - b_i I\right)^{-1} T\right) - \operatorname{tr}\left(T^* \left(A_{i-1} - b_i I\right)^{-1} T\right) \\ &\overset{(20)}{=} \quad -\frac{\operatorname{tr}\left(T^* \left(A_{i-1} - b_i I\right)^{-1} \left((T x_j) \otimes (T x_j)\right) \left(A_{i-1} - b_i I\right)^{-1} T\right)}{1 + \left\langle \left(A_{i-1} - b_i I\right)^{-1} T x_j, T x_j \right\rangle} \\ &= \quad -\frac{\left\langle \left(A_{i-1} - b_i I\right)^{-1} T T^* \left(A_{i-1} - b_i I\right)^{-1} T x_j, T x_j \right\rangle}{1 + \left\langle \left(A_{i-1} - b_i I\right)^{-1} T x_j, T x_j \right\rangle} \\ &\overset{(54)\wedge(53)}{<} \quad \mu \overset{(51)\wedge(58)}{=} \operatorname{tr}\left(T^* \left(A_{i-1} - b_{i-1} I\right)^{-1} T\right) - \operatorname{tr}\left(T^* \left(A_{i-1} - b_i I\right)^{-1} T\right). \end{aligned}$$

This proves (59), so all that remains in order to prove Theorem 6.2 is to prove Lemma 6.4.

Proof of Lemma 6.4. — Using (58) we see that our goal (52) is equivalent to the following inequality

$$\begin{aligned} &\operatorname{tr}\left(T^* \left(A_{i-1} - b_i I\right)^{-1} T T^* \left(A_{i-1} - b_i I\right)^{-1} T\right) \\ &\qquad\qquad < -\mu \left(m + \operatorname{tr}\left(T^* \left(A_{i-1} - b_i I\right)^{-1} T\right)\right). \end{aligned} \tag{60}$$

Note that

$$\begin{aligned}&\operatorname{tr}\left(T^*\left(A_{i-1}-b_iI\right)^{-1}TT^*\left(A_{i-1}-b_iI\right)^{-1}T\right)\\&\quad\leqslant\|T\|^2\operatorname{tr}\left(\left(A_{i-1}-b_iI\right)^{-1}TT^*\left(A_{i-1}-b_iI\right)^{-1}\right)=\|T\|^2\operatorname{tr}\left(T^*\left(A_{i-1}-b_iI\right)^{-2}T\right).\quad(61)\end{aligned}$$

The inductive hypothesis (50), or its equivalent form (59), implies that

$$\begin{aligned}&\operatorname{tr}\left(T^*\left(A_{i-1}-b_{i-1}I\right)^{-1}T\right)<\operatorname{tr}\left(T^*\left(A_0-b_0I\right)^{-1}T\right)\\&\qquad=-\frac{1}{b_0}\operatorname{tr}(T^*T)=\frac{\|T\|_{\mathrm{HS}}^2}{b_0}\overset{(48)}{=}-\frac{m}{1-\varepsilon}.\quad(62)\end{aligned}$$

Hence,

$$\operatorname{tr}\left(T^*\left(A_{i-1}-b_iI\right)^{-1}T\right)\overset{(51)\wedge(58)\wedge(62)}{<}-\frac{m}{1-\varepsilon}-\mu.\quad(63)$$

From (61) and (63) we see that in order to prove (60) it suffices to establish the following inequality:

$$\|T\|^2\operatorname{tr}\left(T^*\left(A_{i-1}-b_iI\right)^{-2}T\right)\leqslant\frac{\varepsilon m}{1-\varepsilon}\mu+\mu^2.\quad(64)$$

To prove (64) we first make some preparatory remarks. For $r\in\{0,\dots,i-1\}$ let P_r be the orthogonal projection on the image of A_r and let $Q_r=I-P_r$ be the orthogonal projection on the kernel of A_r. Since $A_0=0$ we have $Q_0=I$. Moreover, because $A_r=A_{r-1}+(Ty_r)\otimes(Ty_r)$ and $A_{r-1},(Ty_r)\otimes(Ty_r)$ are both positive semidefinite, it follows that $\operatorname{Ker}(A_r)=\operatorname{Ker}(A_{r-1})\cap(Tx_r)^\perp$. Therefore

$$\operatorname{tr}(Q_{r-1}-Q_r)=\dim(\operatorname{Ker}(A_{r-1}))-\dim(\operatorname{Ker}(A_r))\leqslant 1.\quad(65)$$

Hence,

$$\begin{aligned}\|Q_rT\|_{\mathrm{HS}}^2&=\operatorname{tr}\left(T^*Q_rT\right)=\|Q_{r-1}T\|_{\mathrm{HS}}^2-\operatorname{tr}\left(T^*(Q_{r-1}-Q_r)T\right)\\&\geqslant\|Q_{r-1}T\|_{\mathrm{HS}}^2-\|T\|^2\operatorname{tr}(Q_{r-1}-Q_r)\overset{(65)}{\geqslant}\|Q_{r-1}T\|_{\mathrm{HS}}^2-\|T\|^2.\quad(66)\end{aligned}$$

Since $Q_0=I$, (66) yields by induction the following useful bound:

$$\|Q_{i-1}T\|_{\mathrm{HS}}^2\geqslant\|T\|_{\mathrm{HS}}^2-(i-1)\|T\|^2.\quad(67)$$

Next, since the nonzero eigenvalues of A_{i-1} are greater than b_{i-1}, the matrix $T^*P_{i-1}\left(\left(A_{i-1}-b_{i-1}I\right)\left(A_{i-1}-b_iI\right)^{-2}\right)P_{i-1}T$ is positive semidefinite. In particular, its trace is nonnegative, yielding the following estimate:

$$\begin{aligned}0&\leqslant\operatorname{tr}\left(T^*P_{i-1}\left(\left(A_{i-1}-b_{i-1}I\right)\left(A_{i-1}-b_iI\right)^{-2}\right)P_{i-1}T\right)\\&=\operatorname{tr}\left(T^*P_{i-1}\left(\frac{\left(A_{i-1}-b_{i-1}I\right)^{-1}-\left(A_i-b_iI\right)^{-1}}{(b_{i-1}-b_i)^2}-\frac{\left(A_{i-1}-b_iI\right)^{-2}}{b_{i-1}-b_i}\right)P_{i-1}T\right),\end{aligned}$$

which rearranges to the following inequality:

$$\begin{aligned}&(b_{i-1}-b_i)\operatorname{tr}\left(T^*P_{i-1}\left(A_{i-1}-b_iI\right)^{-2}P_{i-1}T\right)\\&\quad\leqslant\operatorname{tr}\left(T^*P_{i-1}\left(A_{i-1}-b_{i-1}I\right)^{-1}P_{i-1}T\right)-\operatorname{tr}\left(T^*P_{i-1}\left(A_{i-1}-b_iI\right)^{-1}P_{i-1}T\right).\quad(68)\end{aligned}$$

Since $Q_{i-1}(A_{i-1}-b_{i-1}I)^{-1}Q_{i-1}=-\frac{1}{b_{i-1}}Q_{i-1}$ and $Q_{i-1}(A_{i-1}-b_iI)^{-1}Q_{i-1}=-\frac{1}{b_i}Q_{i-1}$,

$$\begin{aligned}
\mu &= \mathrm{tr}\left(T^*(P_{i-1}+Q_{i-1})\left(A_{i-1}-b_{i-1}I\right)^{-1}(P_{i-1}+Q_{i-1})T\right)\\
&\quad -\mathrm{tr}\left(T^*(P_{i-1}+Q_{i-1})\left(A_{i-1}-b_iI\right)^{-1}(P_{i-1}+Q_{i-1})T\right)\\
&= \mathrm{tr}\left(T^*P_{i-1}\left(A_{i-1}-b_{i-1}I\right)^{-1}P_{i-1}T\right)-\mathrm{tr}\left(T^*P_{i-1}\left(A_{i-1}-b_iI\right)^{-1}P_{i-1}T\right)\\
&\quad +\left(\frac{1}{b_i}-\frac{1}{b_{i-1}}\right)\mathrm{tr}\left(T^*Q_{i-1}T\right)\\
&\overset{(68)}{\geqslant} (b_{i-1}-b_i)\mathrm{tr}\left(T^*P_{i-1}\left(A_{i-1}-b_iI\right)^{-2}P_{i-1}T\right)+\frac{b_{i-1}-b_i}{b_{i-1}b_i}\|Q_{i-1}T\|_{\mathrm{HS}}^2.
\end{aligned}\tag{69}$$

Also $Q_{i-1}(A_{i-1}-b_iI)^{-2}Q_{i-1}=\frac{1}{b_i^2}Q_{i-1}$, and therefore

$$\begin{aligned}
\mathrm{tr}\left(T^*\left(A_{i-1}-b_iI\right)^{-2}T\right) &= \mathrm{tr}\left(T^*P_{i-1}\left(A_{i-1}-b_iI\right)^{-2}P_{i-1}T\right)+\frac{\mathrm{tr}(T^*Q_{i-1}T)}{b_i^2}\\
&= \mathrm{tr}\left(T^*P_{i-1}\left(A_{i-1}-b_iI\right)^{-2}P_{i-1}T\right)+\frac{1}{b_i^2}\|Q_{i-1}T\|_{\mathrm{HS}}^2\\
&\overset{(69)}{\leqslant} \frac{\mu}{b_{i-1}-b_i}+\frac{\|Q_{i-1}T\|_{\mathrm{HS}}^2}{b_i}\left(\frac{1}{b_i}-\frac{1}{b_{i-1}}\right)\\
&\overset{(48)}{=} \frac{\varepsilon m\mu}{(1-\varepsilon)\|T\|^2}+\frac{\|Q_{i-1}T\|_{\mathrm{HS}}^2}{b_i}\left(\frac{1}{b_i}-\frac{1}{b_{i-1}}\right).
\end{aligned}$$

It follows that in order to prove the desired inequality (64), it suffices to show that the following inequality holds true:

$$\|T\|^2\frac{\|Q_{i-1}T\|_{\mathrm{HS}}^2}{b_i}\left(\frac{1}{b_i}-\frac{1}{b_{i-1}}\right)\leqslant\mu^2.\tag{70}$$

Since $b_{i-1}>b_i$ and $T^*P_{i-1}\left(A_{i-1}-b_iI\right)^{-2}P_{i-1}T$ is positive semidefinite, a consequence of (69) is that $\mu\geqslant\|Q_{i-1}T\|_{\mathrm{HS}}^2\left(\frac{1}{b_i}-\frac{1}{b_{i-1}}\right)$. Hence, in order to prove (70) it suffices to show that

$$\|T\|^2\frac{\|Q_{i-1}T\|_{\mathrm{HS}}^2}{b_i}\left(\frac{1}{b_i}-\frac{1}{b_{i-1}}\right)\leqslant\|Q_{i-1}T\|_{\mathrm{HS}}^4\left(\frac{1}{b_i}-\frac{1}{b_{i-1}}\right)^2,$$

or equivalently,

$$\|Q_{i-1}T\|_{\mathrm{HS}}^2\geqslant\|T\|^2\frac{b_{i-1}}{b_{i-1}-b_i}\overset{(48)}{=}\varepsilon\|T\|_{\mathrm{HS}}^2-(i-1)\|T\|^2,$$

which is a consequence of inequality (67), that we proved earlier. $\square$

7. NONLINEAR NOTIONS OF SPARSIFICATION

Quadratic forms such as $\sum_{i=1}^n\sum_{j=1}^n g_{ij}(x_i-x_j)^2$ are expressed in terms of the mutual distances between the points $\{x_1,\ldots,x_n\}\subseteq\mathbb{R}$. This feature makes them very useful for a variety of applications in metric geometry, where the Euclidean distance is replaced by other geometries. We refer to [MN, NS] for a (partial) discussion of such issues. It would be useful to study the sparsification problem of Theorem 1.1 in the non-Euclidean setting as well, although the spectral arguments used by Batson-Spielman-Srivastava seem inadequate for addressing such nonlinear questions.

In greatest generality one might consider an abstract set X, and a symmetric function (kernel) $K : X \times X \to [0,\infty)$. Given an $n \times n$ matrix $G = (g_{ij})$, the goal would be to find a sparse $n \times n$ matrix $H = (h_{ij})$ satisfying

$$\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}K(x_i,x_j) \leqslant \sum_{i=1}^{n}\sum_{j=1}^{n} h_{ij}K(x_i,x_j) \leqslant C\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}K(x_i,x_j), \tag{71}$$

for some constant $C > 0$ and all $x_1,\ldots,x_n \in X$.

Cases of geometric interest in (71) are when $K(x,y) = d(x,y)^p$, where $d(\cdot,\cdot)$ is a metric on X and $p > 0$. When $p \neq 2$ even the case of the real line with the standard metric is unclear. Say that an $n \times n$ matrix $H = (h_{ij})$ is a p-sparsifier with quality C of an $n \times n$ matrix $G = (g_{ij})$ if $\mathrm{supp}(H) \subseteq \mathrm{supp}(G)$ and there exists a scaling factor $\lambda > 0$ such that for every $x_1,\ldots,x_n \in \mathbb{R}$ we have

$$\lambda\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}|x_i-x_j|^p \leqslant \sum_{i=1}^{n}\sum_{j=1}^{n} h_{ij}|x_i-x_j|^p \leqslant C\lambda\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}|x_i-x_j|^p. \tag{72}$$

By integrating (72) we see that it is equivalent to the requirement that for every $f_1,\ldots,f_n \in L_p$ we have

$$\lambda\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}\|f_i-f_j\|_p^p \leqslant \sum_{i=1}^{n}\sum_{j=1}^{n} h_{ij}\|f_i-f_j\|_p^p \leqslant C\lambda\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}\|f_i-f_j\|_p^p. \tag{73}$$

By a classical theorem of Schoenberg [Scho] (see also [WW]), if $q \leqslant p$ then the metric space $(\mathbb{R},|x-y|^{q/p})$ admits an isometric embedding into L_2, which in turn is isometric to a subspace of L_p. It therefore follows from (73) that if H is a p-sparsifier of G with quality C then it is also a q-sparsifier of G with quality C for every $q \leqslant p$. In particular, when $p \in (0,2)$, Theorem 1.1 implies that for every G a p-sparsifier H of quality $1+\varepsilon$ always exists with $|\mathrm{supp}(H)| = O(n/\varepsilon^2)$.

When $p > 2$ it is open whether every matrix G has a good p-sparsifier H. By “good” we mean that the quality of the sparsifier H is small, and that $|\mathrm{supp}(H)|$ is small. In particular, we ask whether every matrix G admits a p-sparisfiers H with quality $O_p(1)$ (maybe even $1+\varepsilon$) and $|\mathrm{supp}(H)|$ growing linearly with n.

It was shown to us by Bo’az Klartag that if $G = (g_ig_j)$ is a product matrix with non-negative entries then Matoušek’s extrapolation argument for Poincaré inequalities [Ma] (see also [NS, Lemma 4.4]) can be used to show that if $q > p$ and H is a p-sparsifier of G with quality C, then H is also a q-sparsifier of G with quality $C'(C,p,q)$. However, we shall now present a simple example showing that a p-sparsifier of G need not be a q-sparsifier of G with quality independent of n for any $q > p$, for some matrix G (which is, of course, not a product matrix). This raises the question whether or not the method of Batson-Spielman-Srivastava, i.e., Theorem 1.1, produces a matrix H which is a $O(1)$-quality p-sparsifier of G for some $p > 2$.

Fix $q > p$, $\varepsilon > 0$ and $n \in \mathbb{N}$. Let $G = (g_{ij})$ be the $n \times n$ adjacency matrix of the weighted n-cycle, where one edge has weight 1, and all remaining edges have weight $(n-1)^{p-1}/\varepsilon$, i.e., $g_{1n} = g_{n1} = 1$,

$$g_{12} = g_{21} = g_{23} = g_{32} = \cdots = g_{n-1,n} = g_{n,n-1} = \frac{(n-1)^{p-1}}{\varepsilon},$$

and all the other entries of G vanish. Let $H = (h_{ij})$ be the adjacency matrix of the same weighted graph, with the edge $\{1, n\}$ deleted, i.e., $h_{1n} = h_{n1} = 0$ and all the other entries of H coincide with the entries of G. It is immediate from the definition that $\sum_{i=1}^n \sum_{j=1}^n g_{ij}|x_i - x_j|^p \geqslant \sum_{i=1}^n \sum_{j=1}^n h_{ij}|x_i - x_j|^p$ for all $x_1, \ldots, x_n \in \mathbb{R}$. The reverse inequality is proved as follows:

$$\begin{aligned}
\sum_{i=1}^n \sum_{j=1}^n g_{ij}|x_i - x_j|^p &= 2|x_1 - x_n|^p + \frac{2(n-1)^{p-1}}{\varepsilon} \sum_{i=1}^{n-1} |x_i - x_{i+1}|^p \\
&\leqslant 2\left(\sum_{i=1}^{n-1} |x_i - x_{i+1}|\right)^p + \frac{2(n-1)^{p-1}}{\varepsilon} \sum_{i=1}^{n-1} |x_i - x_{i+1}|^p \\
&\leqslant (1+\varepsilon)\frac{2(n-1)^{p-1}}{\varepsilon} \sum_{i=1}^{n-1} |x_i - x_{i+1}|^p \\
&= (1+\varepsilon) \sum_{i=1}^n \sum_{j=1}^n h_{ij}|x_i - x_j|^p.
\end{aligned}$$

Hence H is a p-sparsifier of G with quality $1 + \varepsilon$.

For the points $x_i = i$ we have $\sum_{i=1}^n \sum_{j=1}^n g_{ij}|x_i - x_j|^q = 2(n-1)^q + 2(n-1)^p/\varepsilon$, and $\sum_{i=1}^n \sum_{j=1}^n h_{ij}|x_i - x_j|^q = 2(n-1)^p/\varepsilon$. At the same time, if $y_2 = 1$ and $y_i = 0$ for all $i \in \{1, \ldots, n\} \smallsetminus \{2\}$, we have $\sum_{i=1}^n \sum_{j=1}^n g_{ij}|y_i - y_j|^q = \sum_{i=1}^n \sum_{j=1}^n h_{ij}|y_i - y_j|^q > 0$. Thus, the quality of H as a q-sparsifier of G is at least $\varepsilon(n-1)^{q-p}$, which tends to ∞ with n, since $q > p$.

Acknowledgments. This paper is a survey of recent results of various authors, most notably Batson-Spielman-Srivastava [BSS], Spielman-Srivastava [SS1, SS2], Srivastava [Sr1, Sr2], Newman-Rabinovich [NR] and Schechtman [Sche3]. Any differences between the presentation here and the results being surveyed are only cosmetic. I am very grateful to Alexandr Andoni, Tim Austin, Keith Ball, Bo'az Klartag, Ofer Neiman and especially Vincent Lafforgue, Gilles Pisier, Gideon Schechtman and Nikhil Srivastava, for helpful discussions and suggestions.

REFERENCES

[ALPT] R. Adamczak, A. E. Litvak, A. Pajor, N. Tomczak-Jaegermann – *Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles*, J. Amer. Math. Soc. 23 (2010), no. 2, 535-561.

[Al] N. Alon – *Problems and results in extremal combinatorics*, Discrete Math. 273 (2003), no. 1–3, 31-53.

[ANN] A. Andoni, A. Naor, O. Neiman – *On isomorphic dimension reduction in ℓ_1*, preprint (2011).

[Bah] R. Bahtia – *Matrix analysis*, Graduate Texts in Mathematics, 169. Springer-Verlag, New York, 1997.

[Bal1] K. M. Ball – *Isometric embedding in l_p-spaces*, European J. Combin. 11 (1990), no. 4, 305-311.

[Bal2] K. M. Ball – *An elementary introduction to modern convex geometry*, In Flavors of Geometry, 1–58, Math. Sci. Res. Inst. Publ., 31, Cambridge Univ. Press, Cambridge, 1997.

[BSS] J. D. Batson, D. A. Spielman and N. Srivastava – *Twice-Ramanujan sparsifiers*, In Proceedings of the 41st Annual ACM Symposium on Theory of Computing (2009), 255–262. Available at `http://arxiv.org/abs/0808.0163`.

[BK] A. A. Benczúr and D. R. Karger – *Approximatind $s-t$ minimum cuts in $\tilde{O}(n^2)$ time*, In Proceedings of the 28th Annual ACM Symposium on the Theory of Computing (1996), 47–55.

[BDGJN] G. Bennet, L. E. Dor, V. Goodman, W. B. Johnson, C. M. Newman – *On uncomplemented subspaces of L_p, $1 < p < 2$*, Israel J. Math. 26 (1977), no. 2, 178-187.

[BHKW] K. Berman, H. Halpern, V. Kaftal, G. Weiss – *Matrix norm inequalities and the relative Dixmier property*, Integral Equations Operator Theory 11 (1988), no. 1, 28-48.

[BLM] J. Bourgain, J. Lindenstrauss, V. Milman – *Approximation of zonoids by zonotopes*, Acta Math. 162 (1989), no. 1-2, 73-141.

[BT1] J. Bourgain and L. Tzafriri – *Invertibility of large submatrices with applications to the geometry of Banach spaces and harmonic analysis*, Israel J. Math. 57 (1987), 137–224.

[BT2] J. Bourgain and L. Tzafriri – *Restricted invertibility of matrices and applications*, Analysis at Urbana, Vol. II (Urbana, IL, 1986-1987), 61-107, London Math. Soc. Lecture Note Ser., 138, Cambridge Univ. Press, Cambridge, 1989.

[BT3] J. Bourgain and L. Tzafriri – *On a problem of Kadison and Singer*, J. Reine Angew. Math. 420 (1993), 1–43.

[BC] B. Brinkman and M. Charikar – *On the impossibility of dimension reduction in l_1*, J. ACM 52 (2005), no. 5, 766-788.

[CT] P. Casazza and J. Tremain – *Revisiting the Bourgain-Tzafriri restricted invertibility theo- rem*, Operators and Matrices 420 (2009), 97–110.

[DL] M. M. Deza and M. Laurent – *Geometry of cuts and metrics*, Algorithms and Combinatorics, 15, Springer-Verlag, Berlin, 1997.

[GV] G. H. Golub and C. F. Van Loan – *Matrix computations. Third edition*, Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, 1996.

[HLW] S. Hoory, N. Linial, A. Wigderson – *Expander graphs and their applications*, Bull. Amer. Math. Soc. (N.S.) 43 (2006), no. 4, 439-561.

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

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

[JS1] W. B. Johnson and G. Schechtman – *Remarks on Talagrand's deviation inequality for Rademacher functions*, Functional analysis (Austin, TX, 1987/1989), 72-77, Lecture Notes in Math., 1470, Springer, Berlin, 1991.

[JS2] W. B. Johnson and G. Schechtman – *Finite dimensional subspaces of L_p*, Handbook of the geometry of Banach spaces, Vol. I, 837-870, North-Holland, Amsterdam, 2001.

[KMST] A. Kolla, Y. Makarychev, A. Saberi and S-H. Teng – *Subgraph sparsification and nearly optimal ultrasparsifiers*, In Proceedings of the 42nd ACM Symposium on Theory of Computing (2010), 57-66. Available at `http://arxiv.org/abs/0912.1623`.

[LPS] A. Lubotzky, R. Phillips, P. Sarnak – *Ramanujan graphs*, Combinatorica 8 (1988), no. 3, 261-277.

[Ma] J. Matoušek – *On embedding expanders into l_p spaces*, Israel J. Math. 102 (1997), 189-197.

[MN] M. Mendel and A. Naor – *Towards a calculus for non-linear spectral gaps*, In Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms (2010), 236–255. Available at `http://arxiv.org/abs/0910.2041`.

[Na] A. Naor – *L_1 embeddings of the Heisenberg group and fast estimation of graph isoperimetry*, in Proceedings of the International Congress of Mathematicians, Hyderabad India, 2010. Available at `http://www.cims.nyu.edu/~naor/homepagefiles/ICM.pdf`.

[NS] A. Naor and L. Silberman – *Poincaré inequalities, embeddings, and wild groups*, preprint (2010), to appear in Compositio Math. Available at `http://www.cims.nyu.edu/~naor/homepagefiles/wilder.pdf`.

[NR] I. Newman and Y. Rabinovich – *Finite volume spaces and sparsification*, preprint (2010). Available at `http://arxiv.org/abs/1002.3541`.

[Ni] A. Nilli – *On the second eigenvalue of a graph*, Discrete Math. 91 (1991), no. 2, 207-210.

[PT] A. Pełczyński and N. Tomczak-Jaegermann – *On the length of faithful nuclear representations of finite rank operators*, Mathematika 35 (1988), no. 1, 126-143.

[Ru1] M. Rudelson – *Approximate John's decompositions*, Geometric aspects of functional analysis (Israel, 1992-1994), 245-249, Oper. Theory Adv. Appl., 77, Birkhuser, Basel, 1995.

[Ru2] M. Rudelson – *Contact points of convex bodies*, Israel J. Math. 101(1) (1997), 92–124.

[Ru3] M. Rudelson – *Random vectors in the isotropic position*, J. Funct. Anal. 163(1) (1999), 60–72.

[RV] M. Rudelson and R. Vershynin – *Sampling from large matrices: an approach through geometric functional analysis*, J. ACM 54 (2007), no. 4, Art. 21, 19 pp. (electronic).

[Sche1] G. Schechtman – *Fine embeddings of finite-dimensional subspaces of L_p, $1 \leqslant p < 2$, into l_1^m*, Proc. Amer. Math. Soc. 94 (1985), no. 4, 617-623.

[Sche2] G. Schechtman – *More on embedding subspaces of L_p in l_r^n*, Compositio Math. 61 (1987), no. 2, 159-169.

[Sche3] G. Schechtman – *Tight embedding of subspaces of L_p in ℓ_p^n for even p*, Preprint (2010). Available at `http://arxiv.org/abs/1009.1061`.

[SZ] G. Schechtman and A. Zvavitch – *Embedding subspaces of L_p into l_p^N, $0 < p < 1$*, Math. Nachr. 227 (2001), 133142.

[Scho] I. J. Schoenberg – *On certain metric spaces arising from Euclidean spaces by a change of metric and their imbedding in Hilbert space*, Ann. of Math. (2) 38 (1937), no. 4, 787-793.

[ST1] D. A. Spielman and S-H. Teng – *Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems*, In Proceedings of the 36th Annual ACM Symposium on Theory of Computing (2004), 81–90.

[ST2] D. A. Spielman and S-H. Teng – *Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems*, Preprint (2006). Available at `http://arxiv.org/abs/cs/0607105`.

[ST3] D. A. Spielman and S-H. Teng – *A local clustering algorithm for massive graphs and its application to nearly-linear time graph partitioning*, Preprint (2008). Available at `http://arxiv.org/abs/0809.3232`.

[ST4] D. A. Spielman and S-H. Teng – *Spectral sparsification of graphs*, Preprint (2008). Available at `http://arxiv.org/abs/0808.4134`.

[SS1] D. A. Spielman and N. Srivastava – *Graph sparsification by effective resistances*, In Proceedings of the 40th Annual ACM Symposium on Theory of Computing (2008), 563–568. Available at `http://arxiv.org/abs/0803.0929`.

[SS2] D. A. Spielman and N. Srivastava – *An elementary proof of the restricted invertibility theorem.* Preprint (2009), to appear in Israel J. Math. Available at `http://arxiv.org/abs/0911.1114`.

[Sr1] N. Srivastava – *On Contact points of convex bodies*, Preprint (2009). Available at `http://www.cs.yale.edu/homes/srivastava/papers/contact.pdf`.

[Sr2] N. Srivastava – *Spectral sparsification and restricted invertibility*, Ph.D. thesis, Yale University (2010). Available at `http://www.cs.yale.edu/homes/srivastava/dissertation.pdf`.

[Ta1] M. Talagrand – *Embedding subspaces of L_1 into l_1^N*, Proc. Amer. Math. Soc. 108 (1990), no. 2, 363-369.

[Ta2] M. Talagrand – *Embedding subspaces of L_p in l_p^N*, Geometric aspects of functional analysis (Israel, 1992-1994), 311-325, Oper. Theory Adv. Appl., 77, Birkhuser, Basel, 1995.

[Tr1] J. A. Tropp – *Column subset selection, matrix factorization, and eigenvalue optimization*, In Proceedings of the Nineteenth Annual ACM -SIAM Symposium on Discrete Algorithms (2009), 978–986. Available at `http://arxiv.org/abs/0806.4404`.

[Tr2] J. A. Tropp – *User-friendly tail bounds for sums of random matrices*, 2010. Preprint available at `http://arxiv.org/abs/1004.4389`.

[Ve] R. Vershynin – *John's decompositions: selecting a large part*, Israel J. Math. 122 (2001), 253–277. Available at `http://www-personal.umich.edu/~romanv/papers/john.pdf`.

[WW] J. H. Wells and L. R. Williams – *Embeddings and extensions in analysis*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 84. Springer-Verlag, New York-Heidelberg, 1975.

[Wey] H. Weyl – *Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung)*, Math. Ann. 71 (1912), no. 4, 441-479.

[Zv] A. Zvavitch – *More on embedding subspaces of L_p into l_p^N*, $0 < p < 1$, Geometric aspects of functional analysis, 269-280, Lecture Notes in Math., 1745, Springer, Berlin, 2000.

Assaf NAOR

New York University
Courant Institute
251 Mercer street
New York, NY 10012 – USA
E-mail : `naor@cims.nyu.edu`