arXiv:1601.00948v1 [math.FA] 5 Jan 2016

RESTRICTED INVERTIBILITY REVISITED

ASSAF NAOR AND PIERRE YOUSSEF

Dedicated to Jirka Matoušek

Abstract. Suppose that $m,n\in\mathbb{N}$ and that $A:\mathbb{R}^m\to\mathbb{R}^n$ is a linear operator. It is shown here that if $k,r\in\mathbb{N}$ satisfy $k<r\leqslant \mathbf{rank}(A)$ then there exists a subset $\sigma\subseteq\{1,\ldots,m\}$ with $|\sigma|=k$ such that the restriction of A to $\mathbb{R}^\sigma\subseteq\mathbb{R}^m$ is invertible, and moreover the operator norm of the inverse $A^{-1}:A(\mathbb{R}^\sigma)\to\mathbb{R}^m$ is at most a constant multiple of the quantity $\sqrt{mr/((r-k)\sum_{i=r}^m \mathsf{s}_i(A)^2)}$, where $\mathsf{s}_1(A)\geqslant\ldots\geqslant\mathsf{s}_m(A)$ are the singular values of A. This improves over a series of works, starting from the seminal Bourgain–Tzafriri Restricted Invertibility Principle, through the works of Vershynin, Spielman–Srivastava and Marcus–Spielman–Srivastava. In particular, this directly implies an improved restricted invertibility principle in terms of Schatten–von Neumann norms.

1. Introduction

Given $m,n\in\mathbb{N}$, the rank of a linear operator $A:\mathbb{R}^m\to\mathbb{R}^n$ equals the largest possible dimension of a linear subspace $V\subseteq\mathbb{R}^m$ on which A is injective, i.e., the inverse $A^{-1}:A(V)\to V$ exists. The *restricted invertibility problem* asks for conditions on A that ensure a strengthening of this basic fact from linear algebra in two ways, corresponding to additional *structural information* on the subspace $V\subseteq\mathbb{R}^m$ on which A is injective, as well as *quantitative information* on the behavior of the inverse $A^{-1}:A(V)\to V$. Firstly, the goal is to find a large dimensional *coordinate subspace* on which A is invertible, i.e., we wish to find a large subset $\sigma\subseteq\{1,\ldots,m\}$ such that A is injective on $\mathbb{R}^\sigma\subseteq\mathbb{R}^m$. Secondly, rather than being satisfied with mere invertibility we ask for A to be *quantitatively invertible* on $\mathbb{R}^\sigma$ in the sense that the operator norm of the inverse $A^{-1}:A(\mathbb{R}^\sigma)\to\mathbb{R}^\sigma$ is not too large. Obviously, additional assumptions on A are required for such conclusions to hold true.

The following theorem, which is known as the Bourgain–Tzafriri Restricted Invertibility Principle [BT87, BT89, BT91], is a seminal result that addressed the above question and had major influence on subsequent research, with a variety of interesting applications to several areas. Throughout what follows, for $m\in\mathbb{N}$ the standard coordinate basis of $\mathbb{R}^m$ will be denoted by $e_1,\ldots,e_m\in\mathbb{R}^m$.

Theorem 1 (Bourgain–Tzafriri). *There exist two universal constant $c,C\in(0,\infty)$ with the following property. Suppose that $m\in\mathbb{N}$ and that $A:\mathbb{R}^m\to\mathbb{R}^m$ is a linear operator such that the Euclidean norm of the vector $Ae_j\in\mathbb{R}^m$ equals 1 for every $j\in\{1,\ldots,m\}$. Letting $\|A\|$ denote the operator norm of A, there exists a subset $\sigma\subseteq\{1,\ldots,m\}$ with $|\sigma|\geqslant cm/\|A\|^2$ such that A is injective on $\mathbb{R}^\sigma$ and the operator norm of the inverse $A^{-1}:A(\mathbb{R}^\sigma)\to\mathbb{R}^\sigma$ is at most C.*

In what follows, for $p\in[1,\infty]$ and $m\in\mathbb{N}$ the ℓ_p norm of a vector $x\in\mathbb{R}^m$ will be denoted as usual by $\|x\|_p$. Thus $\|x\|_2$ is the Euclidean norm of x. We shall also denote (as usual) by ℓ_p^m the normed space $\mathbb{R}^m$ equipped with the ℓ_p norm. The standard scalar product on $\mathbb{R}^m$ will be denoted $\langle\cdot,\cdot\rangle$. For $k,m,n\in\mathbb{N}$ and a k-dimensional subspace $V\subseteq\mathbb{R}^m$, the Schatten–von Neumann p norm of a linear operator $A:V\to\mathbb{R}^n$ will be denoted below by $\|A\|_{\mathsf{S}_p}$. Thus

$$\|A\|_{\mathsf{S}_p}\stackrel{\mathrm{def}}{=}\Big(\mathbf{Tr}(A^*A)^{\frac{p}{2}}\Big)^{\frac{1}{p}}=\Big(\sum_{j=1}^k \mathsf{s}_j(A)^p\Big)^{\frac{1}{p}},$$

A. N. was supported in part by the NSF, the BSF, the Packard Foundation and the Simons Foundation.

where $\mathsf{s}_1(A) \geqslant \mathsf{s}_2(A) \geqslant \ldots \geqslant \mathsf{s}_k(A)$ denote the singular values of A, i.e., they are the (decreasing rearrangement of the) eigenvalues of the positive semidefinite operator $\sqrt{A^*A} : V \to V^*$. Thus $\|A\|_{\mathsf{S}_\infty} = \mathsf{s}_1(A)$ is the operator norm of A. Also, $\|A\|_{\mathsf{S}_2}$ is the Hilbert–Schmidt norm of A, i.e., for every orthonormal basis $u_1, \ldots, u_k$ of V we have $\|A\|_{\mathsf{S}_2}^2 = \sum_{i=1}^k \sum_{j=1}^n \langle Au_i, e_j\rangle^2 = \sum_{i=1}^k \|Ae_i\|_2^2$. Below it will sometimes be convenient to denote the smallest singular value of A by $\mathsf{s}_{\min}(A) = \mathsf{s}_k(A)$. Thus A is injective if and only if $\mathsf{s}_{\min}(A) > 0$, in which case $\|A^{-1}\|_{\mathsf{S}_\infty} = 1/\mathsf{s}_{\min}(A)$.

Given $m \in \mathbb{N}$ and $\sigma \subseteq \{1, \ldots, m\}$ it will be convenient to denote the formal identity from $\mathbb{R}^\sigma$ to $\mathbb{R}^m$ by $J_\sigma : \mathbb{R}^\sigma \to \mathbb{R}^m$, i.e., $J_\sigma((a_j)_{j\in\sigma}) = \sum_{j\in\sigma} a_j e_j$ for every $(a_j)_{j\in\sigma} \in \mathbb{R}^\sigma$. With this notation, given an operator $A : \mathbb{R}^m \to \mathbb{R}^n$ that is injective on $\mathbb{R}^\sigma$ we can consider the operator $(AJ_\sigma)^{-1} : A(\mathbb{R}^\sigma) \to \mathbb{R}^\sigma$. We shall sometimes drop the need to mention explicitly that A is injective on $\mathbb{R}^\sigma$ by adhering to the convention that if A is not injective on $\mathbb{R}^\sigma$ then $\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty} = \infty$.

Using the above notation, Theorem 1 asserts that if $A : \mathbb{R}^m \to \mathbb{R}^m$ is a linear operator that satisfies $\|Ae_j\|_2 = 1$ for all $j \in \{1, \ldots, m\}$ then there exists $\sigma \subseteq \{1, \ldots, m\}$ with $|\sigma| \gtrsim m/\|A\|_{\mathsf{S}_\infty}$ such that $\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty} \lesssim 1$, or equivalently $\mathsf{s}_{\min}(AJ_\sigma) \gtrsim 1$. Here, and in what follows, we use the following standard asymptotic notation. Given two quantities $K, L \in \mathbb{R}$ the notation $K \lesssim L$ (respectively $K \gtrsim L$) means that there exists a universal constant $c \in (0, \infty)$ such that $K \leqslant cL$ (respectively $K \geqslant cL$). The notation $K \asymp L$ means that both $K \lesssim L$ and $K \gtrsim L$ hold true.

The following theorem is a useful strengthening of the Bourgain–Tzafriri Restricted Invertibility Principle that was discovered by Vershynin in [Ver01].

Theorem 2 (Vershynin)**.** *There exists a universal constant $c \in (0, \infty)$ with the following property. Fix $k, m, n \in \mathbb{N}$. Let $A : \mathbb{R}^m \to \mathbb{R}^n$ be a linear operator with $\|Ae_j\|_2 = 1$ for all $j \in \{1, \ldots, m\}$. Also, let $\Delta : \mathbb{R}^n \to \mathbb{R}^n$ be a positive definite diagonal operator, i.e., there exist $d_1, \ldots, d_n \in (0, \infty)$ such that $\Delta x = (d_1 x_1, \ldots, d_n x_n)$ for every $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Suppose that $k < \|A\Delta\|_{\mathsf{S}_2}^2/\|A\Delta\|_{\mathsf{S}_\infty}^2$ and write $k = (1-\varepsilon)\|A\Delta\|_{\mathsf{S}_2}^2/\|A\Delta\|_{\mathsf{S}_\infty}^2$ where $\varepsilon \in (0,1)$ (thus $\varepsilon = 1 - k\|A\Delta\|_{\mathsf{S}_\infty}^2/\|A\Delta\|_{\mathsf{S}_2}^2$). Then there exists a subset $\sigma \subseteq \{1, \ldots, m\}$ with $|\sigma| = k$ such that $\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty} \leqslant \varepsilon^{-c\log(1/\varepsilon)}$.*

For a linear operator $T : \mathbb{R}^m \to \mathbb{R}^n$, the quantity $\|T\|_{\mathsf{S}_2}^2/\|T\|_{\mathsf{S}_\infty}^2$ is often called the *stable rank* of T, though this terminology sometimes also refers to the quantity $\|T\|_{\mathsf{S}_1}/\|T\|_{\mathsf{S}_\infty}$. In both cases, the use of the term 'stable' in this context expresses the fact that the quantity in question is a robust replacement for the rank of T in the sense that the rank of T could be large due to the fact that T has many positive but nevertheless very small singular values, while if the stable rank of T is large then its singular values are large on average. Below we shall use the terminology 'stable rank' exclusively for the quantity $\|T\|_{\mathsf{S}_2}^2/\|T\|_{\mathsf{S}_\infty}^2$, which we denote by $\mathbf{srank}(T) = \|T\|_{\mathsf{S}_2}^2/\|T\|_{\mathsf{S}_\infty}^2$.

Theorem 1 coincides with the special case $\varepsilon = \frac{1}{2}$ and $\Delta = I_n$ of Theorem 2, where I_n is the identity operator on $\mathbb{R}^n$. However, Theorem 2 improves over Theorem 1 in three ways that are important for geometric applications. Firstly, Theorem 2 treats rectangular matrices while Theorem 1 treats only the case $m = n$. Secondly, even in the special case $\Delta = I_n$ of Theorem 2 the size of the subset $\sigma \subseteq \{1, \ldots, m\}$ is allowed to be arbitrarily close to $\mathbf{srank}(A)$, while in Theorem 1 it can only be taken to be a constant multiple of $\mathbf{srank}(A)$. Lastly, Theorem 2 actually allows for the size of the subset $\sigma \subseteq \{1, \ldots, m\}$ to be arbitrarily close to the supremum of $\mathbf{srank}(A\Delta)$ over all positive definite diagonal operators $\Delta : \mathbb{R}^m \to \mathbb{R}^m$, a quantity that could be much larger than $\mathbf{srank}(A)$.

Remark 3. Theorem 2 is often stated in the literature as a subset selection principle for John decompositions of the identity. Namely, suppose that $k, m, n \in \mathbb{N}$ and $x_1, \ldots, x_m \in \mathbb{R}^n \smallsetminus \{0\}$ satisfy $\sum_{j=1}^m \langle x_j, y\rangle^2 = \|y\|_2^2$ for every $y \in \mathbb{R}^n$. Equivalently, we have $\sum_{j=1}^m x_j \otimes x_j = I_n$, where for $x, y \in \mathbb{R}^n$ the rank-one operator $x \otimes y : \mathbb{R}^n \to \mathbb{R}^n$ is defined as usual by setting $(x \otimes y)(z) = \langle x, z\rangle y$ for every $z \in \mathbb{R}^n$. Suppose that $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear operator satisfying $Tx_1, \ldots, Tx_m \neq 0$, and that $k = (1-\varepsilon)\mathbf{srank}(T)$ for some $\varepsilon \in (0,1)$. Then there exists $\sigma \subseteq \{1,\ldots,m\}$ with $|\sigma| = k$ and

$$\forall \{a_j\}_{j\in\sigma} \subseteq \mathbb{R}, \qquad \Big\| \sum_{j\in\sigma} \frac{a_j}{\|Tx_j\|_2} Tx_j \Big\|_2 \geqslant \varepsilon^{c\log(1/\varepsilon)} \Big(\sum_{j\in\sigma} a_j^2 \Big)^{\frac{1}{2}}.$$

The above formulation is equivalent to Theorem 2 as stated in terms of rectangular matrices by considering the operator $A : \mathbb{R}^m \to \mathbb{R}^n$ that is given by $Ae_j = Tx_j/\|Tx_j\|_2$ for every $j \in \{1,\ldots,m\}$.

A recent breakthrough of Spielman–Srivastava [SS12], that relies nontrivially on a remarkable method for sparsifying quadratic forms that was developed by Batson–Spielman–Srivastava [BSS12] (see also the survey [Nao12]), yielded the following improved restricted invertibility principle, via techniques that are entirely different from those used by Bourgain–Tzafriri and Vershynin.

Theorem 4 (Spielman–Srivastava)**.** *Suppose that $k,m,n \in \mathbb{N}$ and let $A : \mathbb{R}^m \to \mathbb{R}^n$ be a linear operator such that $k < \mathbf{srank}(A)$. Write $k = (1-\varepsilon)\mathbf{srank}(A)$ where $\varepsilon \in (0,1)$. Then there exists a subset $\sigma \subseteq \{1,\ldots,m\}$ with $|\sigma| = k$ such that*

$$\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty} \leqslant \frac{1}{1-\sqrt{1-\varepsilon}} \cdot \frac{\sqrt{m}}{\|A\|_{\mathsf{S}_2}} \leqslant \frac{2\sqrt{m}}{\varepsilon \|A\|_{\mathsf{S}_2}}.$$

In the setting of Theorem 4, since $\|A\|_{\mathsf{S}_2} = \sqrt{m}$ when the columns of A have unit Euclidean norm, Theorem 1 is a special case of Theorem 4. As in the case $\Delta = I_n$ of Theorem 2, the statement of Theorem 4 has the additional feature that the subset $\sigma \subseteq \{1,\ldots,m\}$ can have size arbitrarily close to $\mathbf{srank}(A)$. Moreover, in Theorem 4 the columns of A need not have unit Euclidean norm, and the upper bound on $\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty}$ in terms of ε is much better in Theorem 4 than the corresponding bound in the case $\Delta = I_n$ of Theorem 2; in fact this bound is asymptotically sharp [BHKW88] as $\varepsilon \to 0$. An additional feature of Theorem 4 is that its proof in [SS12] yields a deterministic polynomial time algorithm for finding the subset σ, while previous to [SS12] only a randomized polynomial time algorithm was available [Tro09]. Theorem 2 does have a feature that Theorem 4 does not, namely the size of the subset $\sigma \subseteq \{1,\ldots,m\}$ can be taken to be arbitrarily close to the supremum of $\mathbf{srank}(A\Delta)$ over all positive definite diagonal operators $\Delta : \mathbb{R}^m \to \mathbb{R}^m$, albeit with worse dependence on ε. However, in [You14] it was shown how to combine the features of Theorem 2 and Theorem 4 so as to yield this stronger guarantee with the better dependence on ε that is asserted in Theorem 4. This improvement is important for certain geometric applications [You14]. The new results that are presented below have this stronger "weighted" feature, but for the sake of simplicity of the initial discussion in the Introduction we shall first present all the ensuing statements in their "unweighted" form that corresponds to the way Theorem 4 is stated above.

A different proof of Theorem 4 in the special case $AA^* = I_n$ was found by Marcus, Spielman and Srivastava in [MSS14], using their powerful method of interlacing polynomials [MSS15a, MSS15b]. In fact, their forthcoming work [MSS16] obtains Theorem 5 below, which yields for the first time a restricted invertibility principle for subsets that can be asymptotically larger than the stable rank, with their size depending on the ratio of the Hilbert–Schmidt norm and the Schatten–von Neumann 4 norm. This result was announced by Srivastava in his talk at the conference *Banach Spaces: Geometry and Analysis* (Hebrew University, May 2013), and it is actually a precursor to the outstanding subsequent work [MSS15b]. Its proof will appear for the first time in the forthcoming preprint [MSS16], but we confirmed with the authors that they obtain Theorem 5 as stated below.

Theorem 5 (Marcus–Spielman–Srivastava)**.** *Suppose that $k,m,n \in \mathbb{N}$ and let $A : \mathbb{R}^m \to \mathbb{R}^n$ be a linear operator such that $k < \frac{1}{4}(\|A\|_{\mathsf{S}_2}/\|A\|_{\mathsf{S}_4})^4$. Define $\varepsilon \in (3/4,1)$ by $k = (1-\varepsilon)\|A\|_{\mathsf{S}_2}^4/\|A\|_{\mathsf{S}_4}^4$. Then there exists a subset $\sigma \subseteq \{1,\ldots,m\}$ with $|\sigma| = k$ such that*

$$\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty} \leqslant \frac{1}{\sqrt{1-2\sqrt{1-\varepsilon}}} \cdot \frac{\sqrt{m}}{\|A\|_{\mathsf{S}_2}}. \tag{1}$$

Theorem 5 can be much better than the previously known restricted invertibility principles at detecting large well-invertible sub-matrices. To state a concrete example, suppose that the singular values of A are $\mathsf{s}_1(A) \asymp \sqrt[4]{m}$ and $\mathsf{s}_2(A) \asymp \mathsf{s}_3(A) \asymp \ldots \asymp \mathsf{s}_m(A) = 1$. Then Theorem 4 yields a subset $\sigma \subseteq \{1,\ldots,m\}$ of size of order $\sqrt{m}$ for which the operator norm of the inverse of AJ_σ is $O(1)$, while Theorem 5 yields such a subset whose size is at least a constant multiple of m.

The restriction $k < \frac{1}{4}(\|A\|_{\mathsf{S}_2}/\|A\|_{\mathsf{S}_4})^4$ in Theorem 5 ensures that $\varepsilon > 3/4$, so that the quantity appearing under the square root in (1) is positive. Thus, in the statement of Theorem 5 k cannot be arbitrarily close to the "modified stable rank" $\|A\|_{\mathsf{S}_2}^4/\|A\|_{\mathsf{S}_4}^4$, but this will be remedied below.

It is important to note that the quantity $\|A\|_{\mathsf{S}_2}^4/\|A\|_{\mathsf{S}_4}^4$ is always at least $\mathbf{srank}(A)$. More generally, given $p \in (2,\infty]$, if we define the p-stable rank of A to be the quantity

$$\mathbf{srank}_p(A) \stackrel{\mathrm{def}}{=} \left(\frac{\|A\|_{\mathsf{S}_2}}{\|A\|_{\mathsf{S}_p}}\right)^{\frac{2p}{p-2}}, \tag{2}$$

then in particular $\mathbf{srank}_4(A) = \|A\|_{\mathsf{S}_2}^4/\|A\|_{\mathsf{S}_4}^4$ and $\mathbf{srank}_\infty(A) = \mathbf{srank}(A)$. We claim that

$$p \geqslant q > 2 \implies \mathbf{srank}_p(A) \leqslant \mathbf{srank}_q(A), \tag{3}$$

Indeed, by direct application of Hölder's inequality we have

$$\|A\|_{\mathsf{S}_q} \leqslant \|A\|_{\mathsf{S}_2}^{\frac{2(p-q)}{q(p-2)}} \cdot \|A\|_{\mathsf{S}_p}^{\frac{p(q-2)}{q(p-2)}},$$

which simplifies to give (3). The limit as $p \to 2^+$ of $\mathbf{srank}_p(A)$ can be computed explicitly, yielding the quantity below, denoted $\mathbf{Entrank}(A)$, which we naturally call the entropic stable rank of A.

$$\begin{aligned}\mathbf{Entrank}(A) &\stackrel{\mathrm{def}}{=} \lim_{p\to 2^+} \mathbf{srank}_p(A) = \exp\left(\log\sum_{j=1}^m \mathsf{s}_j(A)^2 - \frac{2\sum_{j=1}^m \mathsf{s}_j(A)^2\log \mathsf{s}_j(A)}{\sum_{j=1}^m \mathsf{s}_j(A)^2}\right)\\ &= \exp\left(\frac{\mathbf{Tr}(A^*A)\log\mathbf{Tr}(A^*A) - \mathbf{Tr}(A^*A\log(A^*A))}{\mathbf{Tr}(A^*A)}\right) = \|A\|_{\mathsf{S}_2}^2 \prod_{j=1}^m \mathsf{s}_j(A)^{-\frac{2\mathsf{s}_j(A)^2}{\|A\|_{\mathsf{S}_2}^2}}.\end{aligned}$$

As we shall explain in the next section, here we obtain an improved restricted invertibility theorem that in particular yields a strengthening of Theorem 5 that allows one to make use of the p-stable rank of A for every $p > 2$, thus producing well-invertible sub-matrices of A of size that can be any integer that is less than the entropic stable rank of A.

1.1. **Restricted invertibility in terms of rank.** Our main new result is the following theorem.

Theorem 6. *Suppose that $k,m,n \in \mathbb{N}$. Let $A : \mathbb{R}^m \to \mathbb{R}^n$ be a linear operator with $\mathbf{rank}(A) > k$. Then there exists a subset $\sigma \subseteq \{1,\ldots,m\}$ with $|\sigma| = k$ such that*

$$\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty} \lesssim \min_{r\in\{k+1,\ldots,\mathbf{rank}(A)\}} \sqrt{\frac{mr}{(r-k)\sum_{i=r}^m \mathsf{s}_i(A)^2}}. \tag{4}$$

Example 7. To illustrate the relation between Theorem 4, Theorem 5 and Theorem 6, consider a linear operator $A : \mathbb{R}^m \to \mathbb{R}^n$ with $\mathsf{s}_j(A) \asymp 1/\sqrt{j}$ for every $j \in \{1,\ldots,m\}$. Thus $\mathbf{rank}(A) = m$, $\mathbf{srank}(A) \asymp \log m$ and $\mathbf{srank}_4(m) \asymp (\log m)^2$. Since $\sqrt{m}/\|A\|_{\mathsf{S}_2} \asymp \sqrt{m/\log m}$, Theorem 4 yields $\sigma \subseteq \{1,\ldots,m\}$ with $|\sigma| \asymp \log m$ and $\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty} \lesssim \sqrt{m/\log m}$, Theorem 5 yields such a subset with $|\sigma| \asymp (\log m)^2$, and Theorem 6 yields such a subset with $|\sigma| \gtrsim \sqrt{m}$. In fact, for every $\varepsilon \in (0,1)$, Theorem 6 yields $\sigma \subseteq \{1,\ldots,m\}$ with $|\sigma| \gtrsim m^{1-\varepsilon}$ such that $\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty} \lesssim \frac{1}{\sqrt{\varepsilon}}\sqrt{m/\log m}$.

Theorem 6 has the feature that it asserts the existence of a coordinate subspace of dimension arbitrarily close to the rank of the given operator on which it is invertible, with quantitative control on the operator norm of the inverse. The rank is not a stable quantity, but it is simple to deduce stable consequences of Theorem 6 that are stronger than Theorem 5. Indeed, continuing with the notation of Theorem 6, for every $p \in (2,\infty)$ we can apply Hölder's inequality to deduce that

$$\begin{aligned}\|A\|_{\mathsf{S}_2}^2 &= \sum_{i=1}^{r-1} \mathsf{s}_i(A)^2 + \sum_{i=r}^{m} \mathsf{s}_i(A)^2 \\ &\leqslant (r-1)^{1-\frac{2}{p}}\bigg(\sum_{i=1}^{r-1}\mathsf{s}_i(A)^p\bigg)^{\frac{2}{p}} + \sum_{i=r}^{m}\mathsf{s}_i(A)^2 \leqslant (r-1)^{1-\frac{2}{p}}\|A\|_{\mathsf{S}_p}^2 + \sum_{i=r}^{m}\mathsf{s}_i(A)^2.\end{aligned}$$

Hence,

$$\sum_{i=r}^{m}\mathsf{s}_i(A)^2 \geqslant \|A\|_{\mathsf{S}_2}^2 - (r-1)^{1-\frac{2}{p}}\|A\|_{\mathsf{S}_p}^2 \overset{(2)}{=} \|A\|_{\mathsf{S}_2}^2\left(1-\left(\frac{r-1}{\mathbf{srank}_p(A)}\right)^{1-\frac{2}{p}}\right). \tag{5}$$

A substitution of (5) into (4) yields the following estimate.

$$\mathsf{s}_{\min}(AJ_\sigma)^2 \gtrsim \max_{r\in\{k+1,\ldots,\mathbf{srank}_p(A)\}}\left(1-\frac{k}{r}\right)\left(1-\left(\frac{r-1}{\mathbf{srank}_p(A)}\right)^{1-\frac{2}{p}}\right)\cdot\frac{\|A\|_{\mathsf{S}_2}^2}{m}. \tag{6}$$

The estimate (6) is nontrivial only when $k < \mathbf{srank}_p(A)$, so write $k = (1-\varepsilon)\mathbf{srank}_p(A)$ for some $\varepsilon \in (0,1)$. One checks that the following choice of $r \in \{k+1,\ldots,\mathbf{srank}_p(A)\}$ attains the maximum in the right hand side of (6), up to universal constant factors. If ε is bounded away from 1, say $\varepsilon \in (0,1/2]$, choose $r \asymp (1-\varepsilon/2)\mathbf{srank}_p(A)$. If $1/2 < \varepsilon \leqslant 1-e^{-p/(p-2)}$ then choose $r \asymp \log(1/(1-\varepsilon))\cdot\mathbf{srank}_p(A)$. If $1-e^{-p/(p-2)} < \varepsilon < 1$ then choose $r \asymp e^{-p/(p-2)}\mathbf{srank}_p(A)$. Thus,

$$\begin{aligned}0<\varepsilon\leqslant\frac{1}{2} &\implies \|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty} \lesssim \sqrt{\frac{p}{p-2}}\cdot\frac{\sqrt{m}}{\varepsilon\|A\|_{\mathsf{S}_2}},\\ \frac{1}{2}<\varepsilon\leqslant 1-e^{-\frac{p}{p-2}} &\implies \|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty} \lesssim \sqrt{\frac{p}{p-2}}\cdot\frac{\sqrt{m}}{\log\left(1/(1-\varepsilon)\right)\|A\|_{\mathsf{S}_2}},\\ 1-e^{-\frac{p}{p-2}}<\varepsilon<1 &\implies \|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty} \lesssim \frac{\sqrt{m}}{\|A\|_{\mathsf{S}_2}}.\end{aligned}$$

A more concise way to write these estimates is as follows.

$$\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty} \lesssim \left(1+\frac{p}{(p-2)\big|\log\left(1-\varepsilon^2\right)\big|}\right)^{\frac{1}{2}}\frac{\sqrt{m}}{\|A\|_{\mathsf{S}_2}}.$$

For ease of future reference, we record the above corollary of Theorem 6 as Theorem 8 below.

Theorem 8 (Restricted invertibility in terms of Schatten–von Neumann norms)**.** *Suppose that $k,m,n \in \mathbb{N}$, $\varepsilon \in (0,1)$ and $p \in (2,\infty)$. Let $A:\mathbb{R}^m \to \mathbb{R}^n$ be a linear operator that satisfies $k \leqslant (1-\varepsilon)\mathbf{srank}_p(A)$. Then there exists a subset $\sigma \subseteq \{1,\ldots,m\}$ with $|\sigma| = k$ such that*

$$\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty} \lesssim \left(1+\frac{p}{(p-2)\big|\log\left(1-\varepsilon^2\right)\big|}\right)^{\frac{1}{2}}\frac{\sqrt{m}}{\|A\|_{\mathsf{S}_2}}.$$

Equivalently, if $k < \mathbf{Entrank}(A)$ then there exists $\sigma \subseteq \{1,\ldots,m\}$ with $|\sigma| = k$ such that

$$\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty} \lesssim \inf_{p>2}\psi_p\left(1-\frac{k}{\mathbf{srank}_p(A)}\right)\frac{\sqrt{m}}{\|A\|_{\mathsf{S}_2}},$$

where $\psi_p : \mathbb{R} \to [0,\infty]$ *is defined by* $\psi_p(\varepsilon) = \infty$ *if* $\varepsilon \leqslant 0$, $\psi_p(x) = (\sqrt{p/(p-2)})/\varepsilon$ *if* $0 < \varepsilon < 1/2$, $\psi_p(\varepsilon) = (\sqrt{p/(p-2)})/\log(1/(1-\varepsilon))$ *if* $1/2 < \varepsilon \leqslant 1 - e^{-p/(p-2)}$ *and* $\psi_p(\varepsilon) = 1$ *if* $\varepsilon > 1 - e^{-p/(p-2)}$.

The case $p = 4$ of Theorem 8 implies (up to constant factors) the conclusion of Theorem 5, though now treating any $\varepsilon \in (0,1)$, i.e., k arbitrarily close to $\mathbf{srank}_4(A)$, while Theorem 5 applies only when $\varepsilon > 3/4$. Theorem 8 can detect the well-invertibility of A on coordinate subspaces that are much larger than those detected by Theorem 5. For example suppose that the singular values of A are $\mathsf{s}_1(A) \asymp \sqrt[3]{m}$ and $\mathsf{s}_2(A) \asymp \mathsf{s}_3(A) \asymp \ldots \asymp \mathsf{s}_m(A) \asymp 1$. Then Theorem 5 yields a subset $\sigma \subseteq \{1,\ldots,m\}$ of size of order $m^{2/3}$ for which the operator norm of the inverse of AJ_σ is $O(1)$, while (the case $p = 3$ of) Theorem 8 yields such a subset whose size is proportional to m.

We shall prove Theorem 6 through an application of Theorem 9 below, which is a restricted invertibility statement of independent interest, in combination with a volumetric argument that leads to Lemma 10 below. Throughout what follows, given $n \in \mathbb{N}$ and a linear subspace $F \subseteq \mathbb{R}^n$, we shall denote the orthogonal projection from $\mathbb{R}^n$ onto F by $\mathsf{Proj}_F : \mathbb{R}^n \to F$.

Theorem 9. *Fix* $k,m,n \in \mathbb{N}$ *and a linear operator* $A : \mathbb{R}^m \to \mathbb{R}^n$ *satisfying* $\mathbf{rank}(A) > k$. *Let* $\omega \subseteq \{1,\ldots,m\}$ *be any subset with* $|\omega| = \mathbf{rank}(A)$ *such that the vectors* $\{Ae_i\}_{i\in\omega} \subseteq \mathbb{R}^n$ *are linearly independent. For every* $j \in \omega$ *let* $F_j \subseteq \mathbb{R}^n$ *be the orthogonal complement of the span of* $\{Ae_i\}_{i\in\omega\smallsetminus\{j\}} \subseteq \mathbb{R}^n$, *i.e.,*

$$F_j \stackrel{\mathrm{def}}{=} \Big(\mathbf{span}\,\{Ae_i\}_{i\in\omega\smallsetminus\{j\}}\Big)^{\perp}. \tag{7}$$

Then there exists a subset $\sigma \subseteq \omega$ *with* $|\sigma| = k$ *such that*

$$\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty} \lesssim \frac{\sqrt{\mathbf{rank}(A)}}{\sqrt{\mathbf{rank}(A)-k}} \cdot \max_{j\in\omega} \frac{1}{\|\mathsf{Proj}_{F_j} Ae_j\|_2}. \tag{8}$$

The link between Theorem 9 and Theorem 6 is furnished through the following lemma.

Lemma 10. *Fix* $r,m,n \in \mathbb{N}$. *Let* $A : \mathbb{R}^m \to \mathbb{R}^n$ *be a linear operator with* $\mathbf{rank}(A) \geqslant r$. *For every* $\tau \subseteq \{1,\ldots,m\}$ *let* $E_\tau \subseteq \mathbb{R}^n$ *be the orthogonal complement of the span of* $\{Ae_j\}_{j\in\tau} \subseteq \mathbb{R}^n$, *i.e.,*[1]

$$E_\tau \stackrel{\mathrm{def}}{=} \Big(\mathbf{span}\,\{Ae_j\}_{j\in\tau}\Big)^{\perp}. \tag{9}$$

Then there exists a subset $\tau \subseteq \{1,\ldots,m\}$ *with* $|\tau| = r$ *such that*

$$\forall\, j \in \tau, \qquad \big\|\mathsf{Proj}_{E_{\tau\smallsetminus\{j\}}} Ae_j\big\|_2 \geqslant \frac{1}{\sqrt{m}}\Big(\sum_{i=r}^m \mathsf{s}_i(A)^2\Big)^{\frac12}. \tag{10}$$

The deduction of Theorem 6 from Theorem 9 and Lemma 10 is simple. Indeed, in the setting of Theorem 6, take $r \in \{k+1,\ldots,\mathbf{rank}(A)\}$ and apply Lemma 10 to obtain a subset $\tau \subseteq \{1,\ldots,m\}$ with $|\tau| = r$ that satisfies (10). This implies in particular that $\{Ae_j\}_{j\in\tau}$ are linearly independent, hence the operator $AJ_\tau : \mathbb{R}^\tau \to \mathbb{R}^n$ has rank r. By Theorem 9 applied with A replaced by AJ_τ, $m = r = \mathbf{rank}(A)$ and $\omega = \tau$, we obtain a further subset $\sigma \subseteq \tau$ with $|\sigma| = k$ such that

$$\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty} \overset{(8)\wedge(10)}{\lesssim} \sqrt{\frac{mr}{(r-k)\sum_{i=r}^m \mathsf{s}_i(A)^2}}.$$

This is precisely the assertion of Theorem 6.

In Section 5 we shall prove the following variant of Theorem 9.

[1] Comparing (7) and (9) we see that $F_j = E_{\omega\smallsetminus\{j\}}$ for every $j \in \omega$.

Theorem 11. *Fix $k,m,n\in\mathbb{N}$ and a linear operator $A:\mathbb{R}^m\to\mathbb{R}^n$ satisfying $\mathbf{rank}(A)>k$. Then there exists a subset $\sigma\subseteq\{1,\ldots,m\}$ with $|\sigma|=k$ such that*

$$\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty}\leqslant \frac{\sqrt{m}}{\sqrt{\mathbf{rank}(A)}-\sqrt{k}}\left(\frac{1}{\mathbf{rank}(A)}\sum_{i=1}^{\mathbf{rank}(A)}\frac{1}{\mathsf{s}_i(A)^2}\right)^{\frac12}. \tag{11}$$

To explain how Theorem 11 relates to Theorem 6, note that in the setting of Theorem 6 we have

$$\sum_{j\in\omega}\frac{1}{\|\mathsf{Proj}_{F_j}Ae_j\|_2^2}=\sum_{i=1}^{\mathbf{rank}(A)}\frac{1}{\mathsf{s}_i(AJ_\omega)^2}. \tag{12}$$

The simple linear-algebraic justification of (12) appears in Section 2.1 below. For simplicity suppose that $\omega=\{1,\ldots,m\}$, so $\mathbf{rank}(A)=m$, and write $k=(1-\varepsilon)m$ for some $\varepsilon\in(0,1)$. Then Theorem 6 yields a subset $\sigma\subseteq\{1,\ldots,m\}$ with $|\sigma|=k$ such that

$$\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty}\lesssim\frac{1}{\sqrt{\varepsilon}}\cdot\max_{j\in\{1,\ldots,m\}}\frac{1}{\|\mathsf{Proj}_{F_j}Ae_j\|_2}, \tag{13}$$

while, due to (12), Theorem 11 yields a subset $\sigma\subseteq\{1,\ldots,m\}$ with $|\sigma|=k$ such that

$$\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty}\leqslant\frac{1}{1-\sqrt{1-\varepsilon}}\left(\frac1m\sum_{i=1}^m\frac{1}{\|\mathsf{Proj}_{F_j}Ae_j\|_2^2}\right)^{\frac12}\asymp\frac1\varepsilon\left(\frac1m\sum_{i=1}^m\frac{1}{\|\mathsf{Proj}_{F_j}Ae_j\|_2^2}\right)^{\frac12}. \tag{14}$$

The estimates (13) and (14) are incomparable since (13) yields a dependence on ε that is better than that of (14) as $\varepsilon\to0$, while the bound in (14) is in terms of the average of the quantities $\{1/\|\mathsf{Proj}_{F_j}Ae_j\|_2^2\}_{j=1}^m$ rather than their maximum. It remains an interesting open question whether one could obtain a restricted invertibility theorem that combines the best terms in (13) and (14).

Remark 12. Theorem 9 is best possible, up to constant factors. Indeed, fix $k,m\in\mathbb{N}$ with $k<m$ and let B be the m by m matrix all of whose diagonal entries equal m and all of whose off-diagonal entries equal -1. Then B is positive definite (diagonal-dominant) and we choose $A=\sqrt{B}$. We are thus in the setting of Theorem 9 with $m=n=\mathbf{rank}(A)$ and $\omega=\{1,\ldots,m\}$. The quantity $1/\|\mathsf{Proj}_{F_j}Ae_j\|_2^2$ is equal to the j'th diagonal entry of $(A^*A)^{-1}=B^{-1}$; see equation (16) in Section 2.1 below for a simple justification of this fact. The matrix B is an invertible circulant matrix, and as such B^{-1} is also a circulant matrix whose diagonal entries equal $2/(m+1)$; see [Dav79, KS12] for more on the explicit evaluation of basic quantities related to circulant matrices, including their inverses and eigenvalues, which we use here. Therefore $1/\|\mathsf{Proj}_{F_j}Ae_j\|_2=\sqrt{2/(m+1)}$ for every $j\in\{1,\ldots,m\}$, so that the right hand side of (8) equals $\sqrt{2m/((m+1)(m-k))}\asymp1/\sqrt{m-k}$. At the same time, take any $\sigma\subseteq\{1,\ldots,m\}$ with $|\sigma|=k$. Then $(AJ_\sigma)^*(AJ_\sigma)=J_\sigma^*BJ_\sigma$ corresponds to a k by k matrix whose diagonal entries equal m and whose off-diagonal entries equal -1. This is again a circulant matrix whose eigenvalues equal to $m+1$ with multiplicity $k-1$ and $m+1-k$ with multiplicity 1. Thus $\mathsf{s}_1(AJ_\sigma)=\ldots=\mathsf{s}_{k-1}(AJ_\sigma)=\sqrt{m+1}$ and $\mathsf{s}_k(AJ_\sigma)=\mathsf{s}_{\min}(AJ_\sigma)=1/\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty}=\sqrt{m+1-k}$. This shows that $\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty}\asymp1/\sqrt{m-k}$, so that (8) is sharp up to constant factors.

1.2. **Remarks on the proofs.** The original proof of Bourgain and Tzafriri of Theorem 1 consists of a beautiful combination of probabilistic, combinatorial and analytic arguments. It proceeds roughly along three steps. Firstly, using random selectors one finds a large collection of columns of A that is "well separated." In the second step one uses the Sauer–Shelah lemma [Sau72, She72] to find a further subset of the columns such that the inverse of the restriction of A to this subset, when viewed as an operator from ℓ_2 to ℓ_1, has small norm; the Sauer–Shelah lemma is discussed in Section 2.4 below, since it plays an important role here as well. The third step of the Bourgain–Tzafriri proof uses tools from functional analysis, specifically the Little Grothendieck's Inequality [Gro53] and the Pietsch Domination Theorem [Pie67], to control the desired Hilbertian operator norm; these analytic tools are used here as well, and are explained in detail in Section 2.2 and Section 2.3 below.

Vershynin's proof of Theorem 2 uses the Bourgain–Tzafriri restricted invertibility theorem as a "black box," alongside with (unpublished) work of Kashin and Tzafriri (see Theorem 2.5 in [Ver01]). A key contribution of Verhynin was the idea to work with the Hilbert–Schmidt norm so as to allow for an iterative argument. As we stated earlier, the proof of Spielman and Srivastava of Theorem 4 is entirely different from the previously used methods in this context, relying on the 'sparsification method' of Batson–Spielman–Srivastava [BSS12]. This refreshing approach led to many important developments, and it was subsequently augmented by the powerful 'method of interlacing polynomials' of Marcus–Spielman–Srivastava, which they used to prove Theorem 5, showing that one could use higher Schatten–von Neumann norms to address the restricted invertibility problem.

Our starting point here was the realization that one could use ideas and techniques that predate the works of Vershynin, Spielman–Srivastava and Marcus–Spielman–Srivastava to obtain asymptotically sharp results such as Theorem 4, and even to strengthen the statement in terms of higher Schatten–von Neumann norms that is contained in Theorem 5. These later results were based on the discovery of powerful new techniques, leading to many additional applications (crowned by the solution of the Kadison–Singer problem [MSS15b]) that are not covered here, but the present work shows how to apply classical methods to improve over the best known bounds on the restricted invertibility problem. Specifically, we rely on the beautiful work of Giannopoulos [Gia96], which treats a seemingly unrelated geometric question (see also [Gia95]), though it is partially inspired by the work of Bourgain–Tzafriri [BT87] itself, as well as the works of Bourgain–Szarek [BS88] and Szarek–Talagrand [ST89] (see also [Sza91]). The key step is to use Giannopoulos' clever iterative application of the Sauer–Shelah lemma (Bourgain–Tzafriri used the Sauer–Shelah lemma only once in their original argument) in the proof of Theorem 9. In fact, one could use a geometric statement of Giannopoulos [Gia96] as a "black box" so as to obtain a shorter proof of Theorem 9; this is carried out in Section 4.1 below, but only after we present a self-contained argument in Section 4.

Theorem 11 is of a different nature, since its proof uses the Marcus–Spielman–Srivastava method of interlacing polynomials. We do not see how to prove it using the classical analytic techniques that are utilized elsewhere in this article, and in fact we do not need it for the applications that are obtained here (as we explained earlier, Theorem 11 is incomparable to Theorem 9, being weaker in terms of the dependence on certain parameters and stronger in other respects). Nevertheless, Theorem 11 certainly belongs to the family of restricted invertibility results that we study here.

Among the interesting questions that arise naturally from the present work, we ask whether Theorem 6, Theorem 8, Theorem 9 and Theorem 11 can be made to be algorithmic. Our current proofs do not yield a polynomial time algorithm that finds the desired coordinate subspace, due to various reasons, including (but not limited to) the use of the Sauer–Shelah lemma (in Theorem 6, Theorem 8 and Theorem 9) and the use of the method of interlacing polynomials (in Theorem 11).

1.3. **Roadmap.** While this article is primarily devoted to new results, it also has an expository component due to the fact that we are using tools and ideas from diverse fields, with which some readers may not be familiar. Being very much inspired by Matoušek's exceptionally clear style of mathematical exposition, we also made an effort for the ensuing arguments to be self-contained by including quick explanations of classical results that are being used. It seems impossible to fully achieve a Matoušek-style exposition, but hopefully his influence helped us to make an important area of mathematics and a collection of powerful and versatile tools accessible to a wider audience.

Section 2 describes auxiliary statements that will be used in the subsequent proofs. These include classical results of major importance to several fields, and we include brief deductions of what we need so as to make this article self-contained. Section 3 contains the proof of Lemma 10. A self-contained proof of Theorem 6, using a clever iterative procedure of Giannopoulus [Gia96], appears in Section 4. This is followed by Section 4.1, where it is shown that Theorem 6 is equivalent to a geometric theorem of Giannopulos [Gia96], thus yielding a shorter (but not self-contained) proof of Theorem 6. Section 5 contains the proof of Theorem 11.

2. Preliminaries

In this section we shall describe several tools that will be used in the ensuing arguments, and derive certain corollaries of them in forms that will be easy to quote as the need arises later.

2.1. **A bit of linear algebra.** We shall start with elementary linear algebraic reasoning that clarifies the meaning of some of the quantities that were discussed in the Introduction. In particular, we shall see why the identity (12) holds true.

We work here in the setting of Theorem 9, namely we are given $k,m,n\in\mathbb{N}$ and a linear operator $A:\mathbb{R}^m\to\mathbb{R}^n$ satisfying $\mathbf{rank}(A)>k$. We are also fixing any subset $\omega\subseteq\{1,\ldots,m\}$ with $|\omega|=\mathbf{rank}(A)$ such that the vectors $\{Ae_i\}_{i\in\omega}\subseteq\mathbb{R}^n$ are linearly independent. For $j\in\omega$ we consider the linear subspace $F_j\subseteq\mathbb{R}^n$ that is defined in (7), namely F_j is the orthogonal complement of the span of $\{Ae_i\}_{i\in\omega\smallsetminus\{j\}}\subseteq\mathbb{R}^n$. For every $j\in\omega$ define a vector $\mathsf{v}_j\in\mathbb{R}^n$ as follows.

$$\mathsf{v}_j\stackrel{\mathrm{def}}{=}\frac{\mathsf{Proj}_{F_j}Ae_j}{\|\mathsf{Proj}_{F_j}Ae_j\|_2^2}\in\mathbb{R}^n. \tag{15}$$

For every $j\in\omega$, since $I_n-\mathsf{Proj}_{F_j}$ is the orthogonal projection onto $\mathbf{span}(\{Ae_i\}_{i\in\omega\smallsetminus\{j\}})\subseteq\mathbb{R}^n$, we know that $I_n-\mathsf{Proj}_{F_j}Ae_j\in\mathbf{span}(\{Ae_i\}_{i\in\omega\smallsetminus\{j\}})$. So, $\{\mathsf{Proj}_{F_j}Ae_j\}_{j\in\omega}\subseteq\mathbf{span}(\{Ae_i\}_{i\in\omega})$, and therefore $\{\mathsf{v}_j\}_{j\in\omega}\subseteq\mathbf{span}(\{Ae_i\}_{i\in\omega})$. For $j\in\omega$ we have $\langle\mathsf{Proj}_{F_j}Ae_j,Ae_j\rangle=\|\mathsf{Proj}_{F_j}Ae_j\|_2^2$, so $\langle\mathsf{v}_j,Ae_j\rangle=1$. Also, because $\mathsf{Proj}_{F_j}Ae_j$ is orthogonal to $\{Ae_i\}_{i\in\omega\smallsetminus\{j\}}$, we have $\langle\mathsf{v}_j,Ae_i\rangle=0$ for every $i\in\omega\smallsetminus\{j\}$. Since $\{Ae_i\}_{i\in\omega}$ is a basis of $\mathbf{span}(\{Ae_i\}_{i\in\omega})$ and $\{\mathsf{v}_j\}_{j\in\omega}\subseteq\mathbf{span}(\{Ae_i\}_{i\in\omega})$, this means that $\{\mathsf{v}_j\}_{j\in\omega}$ is the *unique* dual basis of $\{Ae_i\}_{i\in\omega}$ in $\mathbf{span}(\{Ae_i\}_{i\in\omega})$.

The operator $(AJ_\omega)^*(AJ_\omega):\mathbb{R}^\omega\to\mathbb{R}^\omega$ has rank $|\omega|=\mathbf{rank}(A)$, hence it is invertible. For every $j\in\omega$ we may therefore consider the vector

$$\mathsf{w}_j\stackrel{\mathrm{def}}{=}(AJ_\omega)\big((AJ_\omega)^*(AJ_\omega)\big)^{-1}e_j\in\mathbf{span}(\{Ae_i\}_{i\in\omega}).$$

Observe that for every $i,j\in\omega$ we have

$$\begin{aligned}\langle\mathsf{w}_j,Ae_i\rangle&=\Big\langle(AJ_\omega)\big((AJ_\omega)^*(AJ_\omega)\big)^{-1}e_j,(AJ_\omega)e_i\Big\rangle\\&=\Big\langle(AJ_\omega)^*(AJ_\omega)\big((AJ_\omega)^*(AJ_\omega)\big)^{-1}e_j,e_i\Big\rangle=\langle e_j,e_i\rangle.\end{aligned}$$

By the uniqueness of the dual basis of $\{Ae_i\}_{i\in\omega}$ in $\mathbf{span}(\{Ae_i\}_{i\in\omega})$, we conclude that $\mathsf{v}_j=\mathsf{w}_j$ for every $j\in\omega$. This implies in particular that for every $j\in\omega$ we have

$$\begin{aligned}\frac{1}{\|\mathsf{Proj}_{F_j}Ae_j\|_2^2}&=\|\mathsf{v}_j\|_2^2=\langle\mathsf{w}_j,\mathsf{w}_j\rangle=\Big\langle(AJ_\omega)\big((AJ_\omega)^*(AJ_\omega)\big)^{-1}e_j,(AJ_\omega)\big((AJ_\omega)^*(AJ_\omega)\big)^{-1}e_j\Big\rangle\\&=\Big\langle\big((AJ_\omega)^*(AJ_\omega)\big)^{-1}e_j,(AJ_\omega)^*(AJ_\omega)\big((AJ_\omega)^*(AJ_\omega)\big)^{-1}e_j\Big\rangle=\Big\langle\big((AJ_\omega)^*(AJ_\omega)\big)^{-1}e_j,e_j\Big\rangle.\end{aligned}\tag{16}$$

Consequently,

$$\sum_{j\in\omega}\frac{1}{\|\mathsf{Proj}_{F_j}Ae_j\|_2^2}=\sum_{j\in\omega}\Big\langle\big((AJ_\omega)^*(AJ_\omega)\big)^{-1}e_j,e_j\Big\rangle=\mathbf{Tr}\Big(\big((AJ_\omega)^*(AJ_\omega)\big)^{-1}\Big)=\sum_{i=1}^{\mathbf{rank}(A)}\frac{1}{\mathsf{s}_i(AJ_\omega)^2}.$$

This is precisely the identity (12). The above discussion, and in particular the auxiliary vectors (15) and their properties that were derived above, will play a role in later arguments as well.

2.2. **Grothendieck.** We shall use later the following important theorem of Grothendieck [Gro53].

Theorem 13 (Little Grothendieck Inequality)**.** *Fix $k,m,n\in\mathbb{N}$. Suppose that $T:\mathbb{R}^m\to\mathbb{R}^n$ is a linear operator. Then for every $x_1,\dots,x_k\in\mathbb{R}^m$ there exists $i\in\{1,\dots,m\}$ such that*

$$\sum_{r=1}^k \|Tx_r\|_2^2 \leqslant \frac{\pi}{2}\|T\|^2_{\ell_\infty^m\to\ell_2^n}\sum_{r=1}^k x_{ri}^2. \tag{17}$$

Here $\|T\|_{\ell_\infty^m\to\ell_2^n}\stackrel{\mathrm{def}}{=}\max_{x\in[-1,1]^m}\|Tx\|_2$ is the operator norm of T when it is viewed as an operator from ℓ_∞^m to ℓ_2^n, and $x_{ri}=\langle x_r,e_i\rangle$ is the i'th coordinate of $x_r\in\mathbb{R}^m$.

The literature contains clear expositions of Theorem 13 and its various useful generalizations and equivalent formulations; see e.g. [Pis86, DJT95]. Nevertheless, for the sake of completeness we shall now quickly explain why Theorem 13 holds true, following (a specialization of) the standard proofs of this fact [Pis86, DJT95]. We note that the factor $\pi/2$ in (17) is sharp; see e.g. the remark immediately following the proof of Theorem 5.4 in [Pis86].

To prove Theorem 13, by rescaling both T and $(x_1,\dots,x_k)$ we may assume without loss of generality that $\|T\|_{\ell_\infty^m\to\ell_2^n}=1$ and $\sum_{r=1}^k\|Tx_r\|_2^2=1$. With this normalization, we claim that

$$\sum_{j=1}^m\left(\sum_{r=1}^k (T^*Tx_r)_j^2\right)^{\frac12}\leqslant\sqrt{\frac{\pi}{2}}. \tag{18}$$

Once proven, (18) implies the desired estimate (17) via the following application of Cauchy–Schwarz.

$$\begin{aligned}1=\sum_{r=1}^k\|Tx_r\|_2^2=\sum_{r=1}^k\langle x_r,T^*Tx_r\rangle=\sum_{j=1}^m\sum_{r=1}^k x_{rj}(T^*Tx_r)_j\leqslant\sum_{j=1}^m\left(\sum_{r=1}^k x_{rj}^2\right)^{\frac12}\left(\sum_{r=1}^k (T^*Tx_r)_j^2\right)^{\frac12}\\ \leqslant\max_{i\in\{1,\dots,m\}}\left(\sum_{r=1}^k x_{ri}^2\right)^{\frac12}\sum_{j=1}^m\left(\sum_{r=1}^k(T^*Tx_r)_j^2\right)^{\frac12}\stackrel{(18)}{\leqslant}\sqrt{\frac{\pi}{2}}\cdot\max_{i\in\{1,\dots,m\}}\left(\sum_{r=1}^k x_{ri}^2\right)^{\frac12}.\end{aligned}$$

To prove (18), let $\{g_r\}_{r=1}^k$ be i.i.d. standard Gaussian random variables. For every $j\in\{1,\dots,m\}$ the random variable $\sum_{r=1}^k g_r(T^*Tx_r)_j$ is Gaussian with mean 0 and variance $\sum_{r=1}^k(T^*Tx_r)_j^2$. So,

$$\begin{aligned}\mathbb{E}\left[\sum_{j=1}^m\left|\left(T^*\sum_{r=1}^k g_rTx_r\right)_j\right|\right]=\mathbb{E}\left[\sum_{j=1}^m\left|\sum_{r=1}^k g_r(T^*Tx_r)_j\right|\right]=\sum_{j=1}^m\mathbb{E}\left[\left|\sum_{r=1}^k g_r(T^*Tx_r)_j\right|\right]\\ =\mathbb{E}\big[|g_1|\big]\sum_{j=1}^m\left(\sum_{r=1}^k(T^*Tx_r)_j^2\right)^{\frac12}=\sqrt{\frac{2}{\pi}}\sum_{j=1}^m\left(\sum_{r=1}^k(T^*Tx_r)_j^2\right)^{\frac12}.\end{aligned} \tag{19}$$

Let $z\in\{-1,1\}^m$ be the random vector given by $z_j\stackrel{\mathrm{def}}{=}\mathbf{sign}\Big(\big(T^*\sum_{r=1}^k g_rTx_r\big)_j\Big)$. Then

$$\begin{aligned}\sum_{j=1}^m\left|\left(T^*\sum_{r=1}^k g_rTx_r\right)_j\right|=\left\langle z,T^*\sum_{r=1}^k g_rTx_r\right\rangle=\left\langle Tz,\sum_{r=1}^k g_rTx_r\right\rangle\\ \leqslant\|Tz\|_2\cdot\left\|\sum_{r=1}^k g_rTx_r\right\|_2\leqslant\|T\|_{\ell_\infty^m\to\ell_2^n}\cdot\|z\|_\infty\cdot\left\|\sum_{r=1}^k g_rTx_r\right\|_2=\left\|\sum_{r=1}^k g_rTx_r\right\|_2.\end{aligned} \tag{20}$$

By taking expectations in (20) we see that

$$\sqrt{\frac{2}{\pi}}\sum_{j=1}^m\bigg(\sum_{r=1}^k (T^*Tx_r)_j^2\bigg)^{\frac12} \overset{(19)}{=} \mathbb{E}\bigg[\sum_{j=1}^m\Big|\Big(T^*\sum_{r=1}^k g_rTx_r\Big)_j\Big|\bigg]$$
$$\overset{(20)}{\leqslant} \mathbb{E}\bigg[\Big\|\sum_{r=1}^k g_rTx_r\Big\|_2\bigg] \leqslant \bigg(\mathbb{E}\bigg[\Big\|\sum_{r=1}^k g_rTx_r\Big\|_2^2\bigg]\bigg)^{\frac12} = \sum_{r=1}^k \|Tx_r\|_2^2 = 1,$$

This is precisely the desired estimate (18), thus completing the proof of Theorem 13. □

2.3. **Pietsch.** Another classical tool that will be used later (together with the Little Grothendieck Inequality) is the Pietsch Domination Theorem [Pie67].

Theorem 14 (Pietsch Domination)**.** *Fix $m,n\in\mathbb{N}$ and $M\in(0,\infty)$. Suppose that $T:\mathbb{R}^m\to\mathbb{R}^n$ is a linear operator such that for every $k\in\mathbb{N}$ and $x_1,\ldots,x_k\in\mathbb{R}^m$ there exists $i\in\{1,\ldots,m\}$ with $\sum_{r=1}^k\|Tx_r\|_2^2\leqslant M^2\sum_{r=1}^k x_{ri}^2$. Then there exist $\mu_1,\ldots,\mu_m\in[0,1]$ with $\sum_{i=1}^m\mu_i=1$ such that*

$$\forall\, x\in\mathbb{R}^m,\qquad \|Tx\|_2^2\leqslant M^2\sum_{i=1}^m \mu_i x_i^2.$$

In Banach space theoretic terminology, the assumption on the operator T in Theorem 14 says that T has 2-*summing norm at most* M when it is viewed as an operator from ℓ_∞^m to ℓ_2^n. We refer to the book [DJT95] for much more on this topic, as well as proofs of (more general versions of) the Pietsch Domination Theorem. As before, for the sake of completeness we shall now explain why Theorem 14 holds true, following (a specialization of) the standard proofs [DJT95] of this fact, which amount to an application of the separation theorem (equivalently, Hahn–Banach or duality of linear programming) to appropriately chosen convex sets.

Let $\mathsf{K}\subseteq\mathbb{R}^m$ be the set of all those vectors $y\in\mathbb{R}^n$ for which there exists $k\in\mathbb{N}$ and $x_1,\ldots,x_k\in\mathbb{R}^m$ such that $y_i=\sum_{r=1}^k\|Tx_r\|_2^2-M^2\sum_{r=1}^m x_{ri}^2$ for every $i\in\{1,\ldots,m\}$. It is immediate to check that K is convex, and the assumption on T can be restated as saying that $\mathsf{K}\cap(0,\infty)^m=\varnothing$. By the separation theorem there exist $\mu=(\mu_1,\ldots,\mu_m)\in\mathbb{R}^m$ such that $\sum_{i=1}^m\mu_iy_i<\sum_{i=1}^m\mu_iz_i$ for every $y\in\mathsf{K}$ and $z\in(0,\infty)^m$. In particular, $\mu\neq0$ and $\inf_{z\in(0,\infty)^m}\langle z,\mu\rangle>-\infty$, so necessarily $\mu_i\geqslant0$ for all $i\in\{1,\ldots,m\}$. We may rescale so that $\sum_{i=1}^m\mu_i=1$. If $x\in\mathbb{R}^m$ then $(\|Tx\|_2^2-M^2x_i^2)_{i=1}^m\in\mathsf{K}$, so $\|Tx\|_2^2-M^2\sum_{i=1}^m\mu_ix_i^2=\sum_{i=1}^m\mu_i(\|Tx\|_2^2-M^2x_i^2)\leqslant\inf_{z\in(0,\infty)^m}\sum_{i=1}^m\mu_iz_i=0$. □

The following lemma is a combination of the Little Grothendieck Inequality and the Pietsch Domination Theorem; this is how Theorem 13 and Theorem 14 will be used in what follows.

Lemma 15. *Fix $m,n\in\mathbb{N}$ and $\varepsilon\in(0,1)$. Let $T:\mathbb{R}^n\to\mathbb{R}^m$ be a linear operator. Then there exists a subset $\sigma\subseteq\{1,\ldots,m\}$ with $|\sigma|\geqslant(1-\varepsilon)m$ such that*

$$\|\mathsf{Proj}_{\mathbb{R}^\sigma}T\|_{\mathsf{S}_\infty}\leqslant\sqrt{\frac{\pi}{2\varepsilon m}}\cdot\|T\|_{\ell_2^n\to\ell_1^m}. \tag{21}$$

Proof. Since we have $\|T^*\|_{\ell_\infty^m\to\ell_2^n}=\|T\|_{\ell_2^n\to\ell_1^m}$, an application of Theorem 13 to $T^*:\mathbb{R}^m\to\mathbb{R}^n$ shows that the assumption of Theorem 14 holds true with T replaced by T^* and $M=\sqrt{\pi/2}\cdot\|T\|_{\ell_2^n\to\ell_1^m}$. Hence, Theorem 14 shows that there exists $\mu\in[0,1]^m$ with $\sum_{i=1}^m\mu_i=1$ such that

$$\forall\, y\in\mathbb{R}^m,\qquad \|T^*y\|_2^2\leqslant\frac{\pi}{2}\|T\|_{\ell_2^n\to\ell_1^m}^2\sum_{i=1}^m\mu_iy_i^2. \tag{22}$$

Define

$$\sigma\overset{\mathrm{def}}{=}\bigg\{i\in\{1,\ldots,m\}:\ \mu_i\leqslant\frac{1}{m\varepsilon}\bigg\}. \tag{23}$$

Since μ is a probability measure on $\{1,\ldots,m\}$, by Markov's inequality we have $|\sigma| \geqslant (1-\varepsilon)m$.

Take $x \in \mathbb{R}^n$ and choose $y \in \mathbb{R}^m$ such that $\|y\|_2 = 1$ and $\|\mathsf{Proj}_{\mathbb{R}^\sigma} Tx\|_2 = \langle y, \mathsf{Proj}_{\mathbb{R}^\sigma} Tx\rangle$. Then,

$$\begin{aligned}\|\mathsf{Proj}_{\mathbb{R}^\sigma} Tx\|_2^2 &= \langle y, \mathsf{Proj}_{\mathbb{R}^\sigma} Tx\rangle^2 = \langle T^*\mathsf{Proj}_{\mathbb{R}^\sigma} y, x\rangle^2 \leqslant \|T^*\mathsf{Proj}_{\mathbb{R}^\sigma} y\|_2^2 \cdot \|x\|_2^2 \\ &\overset{(22)}{\leqslant} \frac{\pi}{2}\|T\|^2_{\ell_2^n\to\ell_1^m}\cdot\|x\|_2^2\sum_{i\in\sigma}\mu_i y_i^2 \overset{(23)}{\leqslant} \frac{\pi}{2m\varepsilon}\|T\|^2_{\ell_2^n\to\ell_1^m}\cdot\|x\|_2^2\cdot\|y\|_2^2 = \frac{\pi}{2m\varepsilon}\|T\|^2_{\ell_2^n\to\ell_1^m}\cdot\|x\|_2^2. \end{aligned} \tag{24}$$

Since (24) holds true for every $x \in \mathbb{R}^n$, this completes the proof of the desired estimate (21). □

2.4. **Sauer–Shelah.** The Sauer–Shelah lemma [Sau72, She72] is a fundamental combinatorial principle of wide applicability that will be used crucially later.

Lemma 16 (Sauer–Shelah)**.** *Fix $m,n \in \mathbb{N}$. Suppose that $\Omega \subseteq \{-1,1\}^n$ satisfies $|\Omega| \geqslant \sum_{k=0}^{m-1}\binom{n}{k}$. Then there exists a subset $\sigma \subseteq \{1,\ldots,n\}$ with $|\sigma| \geqslant m$ such that $\mathsf{Proj}_{\mathbb{R}^\sigma}\Omega = \{-1,1\}^\sigma$, i.e., for every $\varepsilon \in \{-1,1\}^\sigma$ there exists $\delta \in \Omega$ such that $\delta_j = \varepsilon_j$ for every $j \in \sigma$. In particular, if $|\Omega| \geqslant 2^{n-1}$ then such a subset $\sigma \subseteq \{1,\ldots,n\}$ exists with $|\sigma| \geqslant \lceil (n+1)/2\rceil \geqslant n/2$.*

It is simple to prove Lemma 16 by induction on n if one strengthens the inductive hypothesis so as to assert that if we denote $\mathbf{sh}(\Omega) = \{\sigma \subseteq \{1,\ldots,n\} :\ \mathsf{Proj}_{\mathbb{R}^\sigma}\Omega = \{-1,1\}^\sigma\}$ then $|\mathbf{sh}(\Omega)| \geqslant |\Omega|$; clearly this would imply Lemma 16 since the number of subsets of $\{1,\ldots,n\}$ of size at most $m-1$ equals $\sum_{k=0}^{m-1}\binom{n}{k}$. This stronger statement is due to Pajor [Paj85], and the resulting very short inductive proof which we shall now sketch for completeness appears as Theorem 1.1 in [ARS02].

The case $n = 1$ holds trivially (here we use the convention that $\{-1,1\}^\varnothing = \varnothing$ and $\mathsf{Proj}_{\mathbb{R}^\varnothing}\Omega = \varnothing$). Assuming the validity of the above statement for n, take $\Omega \subseteq \{-1,1\}^{n+1} = \{-1,1\}^n \times \{-1,1\}$ and denote $\Omega_1 = \{x \in \{-1,1\}^n :\ (x,1) \in \Omega\}$ and $\Omega_{-1} = \{x \in \{-1,1\}^n :\ (x,-1) \in \Omega\}$. Then $|\Omega_1| + |\Omega_{-1}| = |\Omega|$ and by the inductive hypothesis we have $|\mathbf{sh}(\Omega_1)| \geqslant |\Omega_1|$ and $|\mathbf{sh}(\Omega_{-1})| \geqslant |\Omega_{-1}|$. By our definitions we have $\mathbf{sh}(\Omega) \supseteq (\mathbf{sh}(\Omega_1)\cup\mathbf{sh}(\Omega_{-1})) \cup \{\sigma\cup\{n+1\} :\ \sigma \in \mathbf{sh}(\Omega_1)\cap\mathbf{sh}(\Omega_{-1})\}$, so $|\mathbf{sh}(\Omega)| \geqslant |\mathbf{sh}(\Omega_1)\cup\mathbf{sh}(\Omega_{-1})| + |\mathbf{sh}(\Omega_1)\cap\mathbf{sh}(\Omega_{-1})| = |\mathbf{sh}(\Omega_1)| + |\mathbf{sh}(\Omega_{-1})| \geqslant |\Omega_1| + |\Omega_{-1}| = |\Omega|$. □

2.5. **Fan and Hilbert–Schmidt.** We record for ease of future use the following lemma that controls the influence of multiplication by an orthogonal projection on the Hilbert–Schmidt norm of a linear operator. Its proof is a simple consequence of the classical *Fan Maximum Principle* [Fan49], but we couldn't locate a reference where it is stated explicitly in the form that we will use later.

Lemma 17. *Fix $m,n \in \mathbb{N}$ and $r \in \{1,\ldots,n\}$. Let $A : \mathbb{R}^m \to \mathbb{R}^n$ be a linear operator and let $\mathsf{P} : \mathbb{R}^n \to \mathbb{R}^n$ be an orthogonal projection of rank r. Then*

$$\|\mathsf{P}A\|_{\mathsf{S}_2} \geqslant \bigg(\sum_{i=n-r+1}^{m}\mathsf{s}_i(A)^2\bigg)^{\frac12}.$$

Proof. Since $I_n - \mathsf{P}$ is an orthogonal projection of rank $n-r$, by a classical result of Fan [Fan49],

$$\mathbf{Tr}(AA^*(I_n-\mathsf{P})) \leqslant \sum_{i=1}^{n-r}\mathsf{s}_i(AA^*) = \sum_{i=1}^{n-r}\mathsf{s}_i(A)^2 \tag{25}$$

The proof of (25) is simple; see e.g. [Stø13, Lemma 8.1.8] for a short proof and [Bha97, Chapter III] for more general variational principles along these lines. Now, since P is an orthogonal projection,

$$\begin{aligned}\|\mathsf{P}A\|_{\mathsf{S}_2}^2 &= \mathbf{Tr}((\mathsf{P}A)^*(\mathsf{P}A)) = \mathbf{Tr}(A^*\mathsf{P}A) = \mathbf{Tr}(AA^*\mathsf{P}) = \mathbf{Tr}(AA^*) - \mathbf{Tr}(AA^*(I_n-\mathsf{P})) \\ &= \sum_{i=1}^m \mathsf{s}_i(A)^2 - \mathbf{Tr}(AA^*(I_n-\mathsf{P})) \overset{(25)}{\geqslant} \sum_{i=1}^m\mathsf{s}_i(A)^2 - \sum_{i=1}^{n-r}\mathsf{s}_i(A)^2 = \sum_{i=n-r+1}^m \mathsf{s}_i(A)^2.\end{aligned}$$ □

3. Proof of Lemma 10

In this section we shall prove Lemma 10 in a more general weighted form that corresponds to the renormalization step in Vershynin's Theorem, i.e., Theorem 2. Using this weighted version of Lemma 10, one can directly deduce weighted versions of Theorem 6 and Theorem 8 as well, by combining Lemma 18 below with Theorem 9, exactly as we did in the Introduction.

Lemma 18 (weighted version of Lemma 10)**.** *Fix $r,m,n\in\mathbb{N}$. Let $A:\mathbb{R}^m\to\mathbb{R}^n$ be a linear operator with $\mathbf{rank}(A)\geqslant r$. For every $\tau\subseteq\{1,\ldots,m\}$ let $E_\tau\subseteq\mathbb{R}^n$ be defined as in (9), i.e., it is the orthogonal complement of the span of $\{Ae_j\}_{j\in\tau}\subseteq\mathbb{R}^n$. Then for every $d_1,\ldots,d_m\in(0,\infty)$ there exists a subset $\tau\subseteq\{1,\ldots,m\}$ with $|\tau|=r$ such that*

$$\forall\, j\in\tau,\qquad \left\|\mathsf{Proj}_{E_{\tau\smallsetminus\{j\}}}Ae_j\right\|_2\geqslant \frac{d_j}{\sqrt{\sum_{i=1}^m d_i^2}}\left(\sum_{i=r}^m \mathsf{s}_i(A)^2\right)^{\frac12}. \tag{26}$$

Proof. For every $\tau\subseteq\{1,\ldots,m\}$ let $K_\tau\subseteq\mathbb{R}^n$ be the convex hull of the vectors $\{\pm Ae_j/d_j\}_{j\in\tau}$, i.e.,

$$K_\tau\stackrel{\mathrm{def}}{=}\mathbf{conv}\left(\left\{\frac{1}{d_j}Ae_j:\ j\in\tau\right\}\cup\left\{-\frac{1}{d_j}Ae_j:\ j\in\tau\right\}\right). \tag{27}$$

The desired subset $\tau\subseteq\{1,\ldots,m\}$ will be chosen so as to maximize the r-dimensional volume of the convex hull of K_σ over all those subsets σ of $\{1,\ldots,m\}$ of size r. Namely, we shall fix from now on a subset $\tau\subseteq\{1,\ldots,m\}$ with $|\tau|=r$ such that

$$\mathbf{vol}_r(K_\tau)=\max_{\substack{\sigma\subseteq\{1,\ldots,m\}\\|\sigma|=r}}\mathbf{vol}_k(K_\sigma). \tag{28}$$

Take any $\beta\subseteq\{1,\ldots,m\}$ with $|\beta|=r-1$ and fix $i\in\{1,\ldots,m\}\smallsetminus\beta$. Then by the definition (27) we have $K_{\beta\cup\{i\}}=\mathbf{conv}(\{\pm Ae_i/d_i\}\cup K_\beta)$, i.e., $K_{\beta\cup\{i\}}$ is the union of the two cones with base K_β and apexes at $\pm Ae_i/d_i$. Recalling (9), note that $K_\beta\subseteq\mathbf{span}(K_\beta)=E_\beta^\perp$. Hence, the height of these two cones equals the Euclidean length of the orthogonal projection of Ae_i/d_i onto E_β. Therefore,

$$\mathbf{vol}_r\big(K_{\beta\cup\{i\}}\big)=\frac{2\left\|\mathsf{Proj}_{E_\beta}Ae_i\right\|_2\mathbf{vol}_{r-1}(K_\beta)}{rd_i}. \tag{29}$$

Returning to the subset τ that was chosen in (28), we see that if $j\in\tau$ and $i\in\{1,\ldots,m\}$ then

$$\begin{aligned}\frac{2\left\|\mathsf{Proj}_{E_{\tau\smallsetminus\{j\}}}Ae_j\right\|_2\mathbf{vol}_{r-1}\big(K_{\tau\smallsetminus\{j\}}\big)}{rd_j}&\stackrel{(29)}{=}\mathbf{vol}_r(K_\tau)\\ &\stackrel{(28)}{\geqslant}\mathbf{vol}_r\big(K_{(\tau\smallsetminus\{j\})\cup\{i\}}\big)\stackrel{(29)}{=}\frac{2\left\|\mathsf{Proj}_{E_{\tau\smallsetminus\{j\}}}Ae_i\right\|_2\mathbf{vol}_{r-1}\big(K_{\tau\smallsetminus\{j\}}\big)}{rd_i}.\end{aligned} \tag{30}$$

Since we are assuming that $r\leqslant\mathbf{rank}(A)$, we know that $\mathbf{vol}_r(K_\tau)>0$. It therefore follows from (30) that also $\mathbf{vol}_{r-1}\big(K_{\tau\smallsetminus\{j\}}\big)>0$, so me may cancel the quantity $2\mathbf{vol}_{r-1}\big(K_{\tau\smallsetminus\{j\}}\big)/r$ from both sides of (30). Since the resulting estimate holds true for every $i\in\{1,\ldots,m\}$, we conclude that

$$\forall\, j\in\tau,\qquad \frac{\left\|\mathsf{Proj}_{E_{\tau\smallsetminus\{j\}}}Ae_j\right\|_2}{d_j}=\max_{i\in\{1,\ldots,m\}}\frac{\left\|\mathsf{Proj}_{E_{\tau\smallsetminus\{j\}}}Ae_i\right\|_2}{d_i}. \tag{31}$$

Consequently, for every $j\in\tau$ we have

$$\frac{\left\|\mathsf{Proj}_{E_{\tau\smallsetminus\{j\}}}Ae_j\right\|_2^2}{d_j^2}\left(\sum_{i=1}^m d_i^2\right)\stackrel{(31)}{\geqslant}\sum_{i=1}^m\left\|\mathsf{Proj}_{E_{\tau\smallsetminus\{j\}}}Ae_i\right\|_2^2=\left\|\mathsf{Proj}_{E_{\tau\smallsetminus\{j\}}}A\right\|_{\mathsf{S}_2}^2.$$

Equivalently,

$$\forall\, j \in \tau, \qquad \left\|\mathsf{Proj}_{E_{\tau\smallsetminus\{j\}}} Ae_j\right\|_2 \geqslant \frac{d_j}{\sqrt{\sum_{i=1}^m d_i^2}} \left\|\mathsf{Proj}_{E_{\tau\smallsetminus\{j\}}} A\right\|_{\mathsf{S}_2}. \tag{32}$$

Recalling (9), since $|\tau| = r$ we know that $\dim(E_{\tau\smallsetminus\{j\}}) = n-(r-1)$ for every $j \in \tau$. Consequently, $\mathsf{Proj}_{E_{\tau\smallsetminus\{j\}}} : \mathbb{R}^n \to \mathbb{R}^n$ is an orthogonal projection of rank $n-(r-1)$, so that the desired inequality (26) follows from (32) and Lemma 17. □

4. Giannopoulos

In this section we shall prove Theorem 9, following the lines of a clever iterative procedure that was devised by Giannopoulos in [Gia96]. Throughout the ensuing discussion, we may assume in the setting of Theorem 9 that $\omega = \{1,\ldots,m\}$, in which case $\mathbf{rank}(A) = m$. Indeed, there is no loss of generality by doing so because for general $\omega \subseteq \{1,\ldots,m\}$ we could then consider the restricted operator $AJ_\omega : \mathbb{R}^\omega \to \mathbb{R}^n$ in order to obtain Theorem 9 as stated in the Introduction.

Lemma 19. *Fix $n \in \mathbb{N}$ and $m \in \{1,\ldots,n\}$. Let $A : \mathbb{R}^m \to \mathbb{R}^n$ be a linear operator such that the vectors $\{Ae_j\}_{j=1}^m \subseteq \mathbb{R}^n$ are linearly independent. Suppose that $k \in \mathbb{N}\cup\{0\}$ and $\sigma \subseteq \{1,\ldots,m\}$. For $j \in \{1,\ldots,m\}$ recall the definition of the subspace $F_j \subseteq \mathbb{R}^n$ in (7) (with $\omega = \{1,\ldots,m\}$), i.e,*

$$F_j = \Big(\mathbf{span}\,\{Ae_i\}_{i\in\{1,\ldots,m\}\smallsetminus\{j\}}\Big)^{\perp}.$$

Then there exists $\tau \subseteq \sigma$ with $|\tau| \geqslant (1-2^{-k})|\sigma|$ such that for every $\vartheta \subseteq \{1,\ldots,m\}$ that satisfies $\vartheta \supseteq \tau$ and every $a = (a_1,\ldots,a_m) \in \mathbb{R}^m$ there exists an index $j \in \{1,\ldots,m\}$ for which

$$\sum_{i\in\tau}|a_i| \leqslant \frac{\sqrt{|\sigma|}\sum_{r=1}^k 2^{\frac{r}{2}}}{\|\mathsf{Proj}_{F_j}Ae_j\|_2}\Big\|\sum_{i\in\vartheta}a_iAx_i\Big\|_2 + (2^k-1)\sum_{i\in\vartheta\cap(\sigma\smallsetminus\tau)}|a_i|. \tag{33}$$

Proof. It will be convenient to introduce the following notation.

$$M \stackrel{\mathrm{def}}{=} \max_{j\in\{1,\ldots,m\}} \frac{1}{\|\mathsf{Proj}_{F_j}Ae_j\|_2} \qquad \text{and} \qquad \alpha_k \stackrel{\mathrm{def}}{=} \sum_{r=1}^k 2^{\frac{r}{2}}. \tag{34}$$

Throughout we adhere to the convention that an empty sum vanishes, thus in particular $\alpha_0 = 0$.

Under the notation (34), our goal becomes to show that there exists $\tau \subseteq \sigma$ with $|\tau| \geqslant (1-2^{-k})|\sigma|$ such that for every $\vartheta \subseteq \{1,\ldots,m\}$ that satisfies $\vartheta \supseteq \tau$ and every $a \in \mathbb{R}^m$ we have

$$\sum_{i\in\tau}|a_i| \leqslant \alpha_k M\sqrt{|\sigma|}\Big\|\sum_{i\in\vartheta}a_iAx_i\Big\|_2 + (2^k-1)\sum_{i\in\vartheta\cap(\sigma\smallsetminus\tau)}|a_i|. \tag{35}$$

We shall prove this statement by induction on k. The case $k=0$ holds vacuously by taking $\tau = \varnothing$. Assuming the validity of this statement for k, we shall proceed to deduce its validity for $k+1$.

We are given $\tau \subseteq \sigma$ with $|\tau| \geqslant (1-2^{-k})|\sigma|$ such that for every $\vartheta \subseteq \{1,\ldots,m\}$ that satisfies $\vartheta \supseteq \tau$ we know that (35) holds true for every $a \in \mathbb{R}^m$. Observe that if $\tau = \sigma$ then τ itself would satisfy the required statement for $k+1$, so we may assume from now on that $\sigma \smallsetminus \tau \neq \varnothing$.

For every $j \in \{1,\ldots,m\}$ let v_j be given as in (15), i.e.,

$$\mathsf{v}_j \stackrel{\mathrm{def}}{=} \frac{\mathsf{Proj}_{F_j}Ae_j}{\|\mathsf{Proj}_{F_j}Ae_j\|_2^2} \in \mathbb{R}^n. \tag{36}$$

Observe that the denominator in (36) (and also in (33) and (34)) does not vanish since we are assuming in Lemma 19 that $\{Ae_j\}_{j=1}^m$ are linearly independent. Define $\Omega \subseteq \{-1,1\}^{\sigma\setminus\tau}$ as follows.

$$\Omega \stackrel{\mathrm{def}}{=} \left\{\varepsilon \in \{-1,1\}^{\sigma\setminus\tau} : \ \Big\|\sum_{i\in\sigma\setminus\tau} \varepsilon_i \mathsf{v}_i\Big\|_2 \leqslant M\sqrt{2|\sigma\setminus\tau|}\right\}. \tag{37}$$

By the parallelogram identity we have

$$\begin{aligned} M^2|\sigma\setminus\tau| &\stackrel{(34)}{\geqslant} \sum_{i\in\sigma\setminus\tau} \frac{1}{\|\mathsf{Proj}_{F_i} Ae_i\|_2^2} \stackrel{(36)}{=} \sum_{i\in\sigma\setminus\tau} \|\mathsf{v}_i\|_2^2 = \frac{1}{2^{|\sigma\setminus\tau|}} \sum_{\varepsilon\in\{-1,1\}^{\sigma\setminus\tau}} \Big\|\sum_{i\in\sigma\setminus\tau} \varepsilon_i \mathsf{v}_i\Big\|_2^2 \\ &\stackrel{(37)}{\geqslant} \frac{1}{2^{|\sigma\setminus\tau|}} \sum_{\substack{\varepsilon\in\{-1,1\}^{\sigma\setminus\tau}\\ \varepsilon\notin\Omega}} 2M^2|\sigma\setminus\tau| = 2M^2|\sigma\setminus\tau|\left(1-\frac{|\Omega|}{2^{|\sigma\setminus\tau|}}\right). \end{aligned} \tag{38}$$

Since $|\sigma\setminus\tau|>0$, it follows from (38) that $|\Omega|\geqslant 2^{|\sigma\setminus\tau|-1}$.

We can now apply the Sauer–Shelah Lemma, i.e., Lemma 16, thus deducing that there exists a subset $\beta\subseteq\sigma\setminus\tau$ with $|\beta|\geqslant|\sigma\setminus\tau|/2$ such that $\mathsf{Proj}_{\mathbb{R}^\beta}\Omega=\{-1,1\}^\beta$. Defining $\tau^*=\tau\cup\beta$ we shall now proceed to show that τ^* satisfies the inductive hypothesis with k replaced by $k+1$.

Since $\beta\cap\tau=\varnothing$, $\tau\subseteq\sigma$ and $|\beta|\geqslant|\sigma\setminus\tau|/2$ we have

$$|\tau^*| = |\tau|+|\beta| \geqslant |\tau| + \frac{|\sigma|-|\tau|}{2} = \frac{|\tau|+|\sigma|}{2} \geqslant \frac{(1-2^{-k})|\sigma|+|\sigma|}{2} = (1-2^{-k-1})|\sigma|. \tag{39}$$

Next, suppose that $\vartheta\subseteq\{1,\ldots,m\}$ satisfies $\vartheta\supseteq\tau^*$. If $a\in\mathbb{R}^m$ then because $\mathsf{Proj}_{\mathbb{R}^\beta}\Omega=\{-1,1\}^\beta$ there exists $\varepsilon\in\Omega$ such that for every $j\in\beta$ we have $\varepsilon_j=\mathbf{sign}(a_j)$. The fact that $\varepsilon\in\Omega$ means that

$$\Big\|\sum_{i\in\sigma\setminus\tau}\varepsilon_i\mathsf{v}_i\Big\|_2 \leqslant M\sqrt{2|\sigma\setminus\tau|} \leqslant \frac{M\sqrt{2|\sigma|}}{2^{k/2}}, \tag{40}$$

where in the last step of (40) we used the fact that $|\tau|\geqslant(1-2^{-k})|\sigma|$.

The definition (36) of $\{\mathsf{v}_j\}_{j=1}^m$ implies that $\langle \mathsf{v}_i, Ae_j\rangle=\delta_{ij}$ for every $i,j\in\{1,\ldots,m\}$. Hence,

$$\begin{aligned} \sum_{i\in\beta}|a_i| &= \Big\langle \sum_{i\in\beta} a_iAe_i, \sum_{i\in\sigma\setminus\tau}\varepsilon_i\mathsf{v}_i\Big\rangle = \Big\langle \sum_{i\in\vartheta} a_iAe_i, \sum_{i\in\sigma\setminus\tau}\varepsilon_i\mathsf{v}_i\Big\rangle - \sum_{i\in(\vartheta\setminus\beta)\cap(\sigma\setminus\tau)}\varepsilon_i a_i \\ &\leqslant \Big\|\sum_{i\in\vartheta}a_iAe_i\Big\|_2\Big\|\sum_{i\in\sigma\setminus\tau}\varepsilon_i\mathsf{v}_i\Big\|_2 + \sum_{i\in\vartheta\cap(\sigma\setminus\tau^*)}|a_i| \stackrel{(40)}{\leqslant} \frac{M\sqrt{2|\sigma|}}{2^{k/2}}\Big\|\sum_{i\in\vartheta}a_iAe_i\Big\|_2 + \sum_{i\in\vartheta\cap(\sigma\setminus\tau^*)}|a_i|. \end{aligned} \tag{41}$$

The penultimate step of (41) uses the Cauchy–Schwarz inequality and the fact that, by the definition of τ^*, we have $(\vartheta\setminus\beta)\cap(\sigma\setminus\tau)=\vartheta\cap(\sigma\setminus\tau^*)$. Now,

$$\begin{aligned} \sum_{i\in\tau^*}|a_i| = \sum_{i\in\tau}|a_i| + \sum_{i\in\beta}|a_i| &\stackrel{(35)}{\leqslant} \alpha_k M\sqrt{|\sigma|}\Big\|\sum_{i\in\vartheta}a_iAe_i\Big\|_2 + (2^k-1)\sum_{i\in\vartheta\cap(\sigma\setminus\tau)}|a_i| + \sum_{i\in\beta}|a_i| \\ &= \alpha_k M\sqrt{|\sigma|}\Big\|\sum_{i\in\vartheta}a_iAe_i\Big\|_2 + (2^k-1)\sum_{i\in\vartheta\cap(\sigma\setminus\tau^*)}|a_i| + 2^k\sum_{i\in\beta}|a_i|, \end{aligned} \tag{42}$$

where for the last step of (42) recall that $\vartheta\cap(\sigma\setminus\tau)=(\vartheta\cap(\sigma\setminus\tau^*))\cup\beta$. It remains to combine (41) and (42) to deduce that

$$\sum_{i\in\tau^*}|a_i| \leqslant \left(\alpha_k + 2^{\frac{k+1}{2}}\right)M\sqrt{|\sigma|}\Big\|\sum_{i\in\vartheta}a_iAe_i\Big\|_2 + (2^{k+1}-1)\sum_{i\in\vartheta\cap(\sigma\setminus\tau^*)}|a_i|. \tag{43}$$

Recalling the definition of α_k in (34), we have $\alpha_{k+1} = \alpha_k + 2^{(k+1)/2}$, so the validity of (39) and (43) completes the proof that τ^* satisfies the inductive hypothesis with k replaced by $k+1$. $\square$

Lemma 20. *Fix $m, n, t \in \mathbb{N}$ and $\beta \subseteq \{1, \ldots, m\}$. Let $A : \mathbb{R}^m \to \mathbb{R}^n$ be a linear operator such that the vectors $\{Ae_j\}_{j=1}^m \subseteq \mathbb{R}^n$ are linearly independent. Then there exist two subsets $\sigma, \tau \subseteq \beta$ satisfying $\sigma \subseteq \tau$, $|\tau| \geqslant (1 - 2^{-t})|\beta|$ and $|\tau \smallsetminus \sigma| \leqslant |\beta|/4$ such that if we denote $\vartheta = \tau \cup (\{1, \ldots, m\} \smallsetminus \beta)$ then*

$$\left\| \mathsf{Proj}_{\mathbb{R}^\sigma} (AJ_\vartheta)^{-1} \right\|_{\mathsf{S}_\infty} \lesssim \max_{j \in \{1,\ldots,m\}} \frac{2^{\frac{t}{2}}}{\|\mathsf{Proj}_{F_j} Ae_j\|_2},$$

where we recall that the definition of the subspace $F_j \subseteq \mathbb{R}^n$ is given in (7).

Proof. An application of Lemma 19 with $\sigma = \beta$ and $k = t$ produces $\tau \subseteq \beta$ with $|\tau| \geqslant (1 - 2^{-t})|\beta|$ such that if we choose $\vartheta = \tau \cup (\{1, \ldots, m\} \smallsetminus \beta)$ in (33) and continue with the notation in (34) then

$$\forall\, a \in \mathbb{R}^m, \qquad \sum_{i \in \tau} |a_i| \lesssim 2^{\frac{t}{2}} M \sqrt{|\beta|} \left\| \sum_{i \in \vartheta} a_i A e_i \right\|_2. \tag{44}$$

Note that the above choice of ϑ makes the second term in the right hand side of (33) vanish, and this is the only way by which (33) will be used here. However, the more complicated form of (33) was needed in Lemma 19 to allow for the inductive construction to go through.

A different way to state (44) is the following operator norm bound.

$$\left\| \mathsf{Proj}_{\mathbb{R}^\tau} (AJ_\vartheta)^{-1} \right\|_{\ell_2^\vartheta \to \ell_1^\tau} \lesssim 2^{\frac{t}{2}} M \sqrt{|\beta|}.$$

Since $|\tau| \geqslant (1 - 2^{-t})|\beta| \geqslant |\beta|/2$, if we set $\varepsilon \stackrel{\mathrm{def}}{=} |\beta|/(4|\tau|)$ then $\varepsilon \in (0, 1/2)$. We are therefore in position to use Lemma 15, thus producing a subset $\sigma \subseteq \tau$ with $|\tau \smallsetminus \sigma| \leqslant \varepsilon|\tau| = |\beta|/4$ such that

$$\left\| \mathsf{Proj}_{\mathbb{R}^\sigma} (AJ_\vartheta)^{-1} \right\|_{\mathsf{S}_\infty} = \left\| \mathsf{Proj}_{\mathbb{R}^\sigma} \mathsf{Proj}_{\mathbb{R}^\tau} (AJ_\vartheta)^{-1} \right\|_{\mathsf{S}_\infty} \lesssim \frac{2^{\frac{t}{2}} M \sqrt{|\beta|}}{\sqrt{\varepsilon |\tau|}} \asymp 2^{\frac{t}{2}} M. \qquad \square$$

Proof of Theorem 9. Recall that, in the setting of Theorem 9, we are currently assuming without loss of generality that $\omega = \{1, \ldots, m\}$. Choose $r \in \mathbb{N} \cup \{0\}$ such that

$$\frac{1}{2^{2r+1}} \leqslant 1 - \frac{k}{m} \leqslant \frac{1}{2^{2r}}. \tag{45}$$

Denote $\tau_0 \stackrel{\mathrm{def}}{=} \{1, \ldots, m\}$ and $\sigma_0 \stackrel{\mathrm{def}}{=} \varnothing$. We shall construct by induction on $u \in \{0, \ldots, r+1\}$ two subsets $\sigma_u, \tau_u \subseteq \{1, \ldots, m\}$ such that if we denote

$$\beta_u \stackrel{\mathrm{def}}{=} \tau_u \smallsetminus \sigma_u \qquad \text{and} \qquad \forall\, u \in \{1, \ldots, r+1\}, \qquad \vartheta_u \stackrel{\mathrm{def}}{=} \tau_u \cup (\{1, \ldots, m\} \smallsetminus \beta_{u-1}), \tag{46}$$

then the following properties hold true for every $u \in \{1, \ldots, r+1\}$.

(a) $\sigma_u \subseteq \tau_u \subseteq \beta_{u-1}$.
(b) $|\tau_u| \geqslant (1 - 2^{-2r+u-4})|\beta_{u-1}|$ and $|\beta_u| \leqslant \frac{1}{4}|\beta_{u-1}|$.
(c) $\left\| \mathsf{Proj}_{\mathbb{R}^{\sigma_u}} (AJ_{\vartheta_u})^{-1} \right\|_{\mathsf{S}_\infty} \lesssim 2^{r - \frac{u}{2}} M$, where M is defined in (34).

Indeed, assuming inductively that σ_{u-1}, τ_{u-1} have been constructed, the existence of sets σ_u, τ_u with the desired properties follows from an application of Lemma 20 with $\beta = \beta_{u-1}$ and $t = 2r - u + 4$.

Recalling (46), by (a) we have $\beta_{u-1} = \beta_u \cupdot \sigma_u \cupdot (\beta_{u-1} \smallsetminus \tau_u)$ for every $u \in \{1, \ldots, r+1\}$. Hence,

$$|\sigma_u| = |\beta_{u-1}| - |\beta_u| - |\beta_{u-1} \smallsetminus \tau_u| \geqslant |\beta_{u-1}| - |\beta_u| - \frac{|\beta_{u-1}|}{2^{2r-u+4}} \geqslant |\beta_{u-1}| - |\beta_u| - \frac{m}{2^{2r+u+2}}, \tag{47}$$

where the penultimate inequality in (47) uses the first assertion in (b) and the final inequality in (47) uses the fact that, by induction, the second assertion in (b) implies that $|\beta_{u-1}| \leqslant m/4^{u-1}$, since $\beta_0 = \{1, \ldots, m\}$. Observe that the sets $\{\sigma_u\}_{u=1}^{r+1}$ are pairwise disjoint, so if we denote

$$\sigma \stackrel{\mathrm{def}}{=} \bigcupdot_{u=1}^{r+1} \sigma_u, \tag{48}$$

then

$$|\sigma| = \sum_{u=1}^{r+1} |\sigma_u| \stackrel{(47)}{\geqslant} |\beta_0| - |\beta_{r+1}| - \frac{m}{2^{2r+2}} \sum_{u=1}^{\infty} \frac{1}{2^u} \geqslant m - \frac{m}{4^{r+1}} - \frac{m}{2^{2r+2}} = m - \frac{m}{2^{2r+1}} \stackrel{(45)}{\geqslant} k. \tag{49}$$

Next, recalling the definition of ϑ_u in (46), observe that

$$\sigma \subseteq \bigcap_{u=1}^{r+1} \vartheta_u. \tag{50}$$

Indeed, in order to verify the validity of (50) note that due to (a) we have $\sigma_u, \sigma_{u+1}, \ldots, \sigma_{r+1} \subseteq \tau_u$ and $\sigma_1, \ldots, \sigma_{u-1} \subseteq \{1, \ldots, m\} \smallsetminus \beta_{u-1}$ for every $u \in \{1, \ldots, r+1\}$. It follows from (50) that if $a \in \mathbb{R}^\sigma$ then for every $u \in \{1, \ldots, r+1\}$ we have $J_\sigma a \in J_{\vartheta_u} \mathbb{R}^{\vartheta_u} \subseteq \mathbb{R}^m$. Consequently,

$$\mathsf{Proj}_{\mathbb{R}^{\sigma_u}} (AJ_{\vartheta_u})^{-1} (AJ_\sigma) a = \mathsf{Proj}_{\mathbb{R}^{\sigma_u}} J_\sigma a. \tag{51}$$

We therefore have the following estimate.

$$\begin{aligned} \|J_\sigma a\|_2^2 \stackrel{(48)}{=} \Big\| \sum_{u=1}^{r+1} \mathsf{Proj}_{\mathbb{R}^{\sigma_u}} J_\sigma a \Big\|_2^2 = \sum_{u=1}^{r+1} \|\mathsf{Proj}_{\mathbb{R}^{\sigma_u}} J_\sigma a\|_2^2 \stackrel{(51)}{=} \sum_{u=1}^{r+1} \big\|\mathsf{Proj}_{\mathbb{R}^{\sigma_u}} (AJ_{\vartheta_u})^{-1} (AJ_\sigma) a\big\|_2^2 \\ \stackrel{(c)}{\lesssim} \sum_{u=1}^{r+1} 2^{2r-u} M^2 \|(AJ_\sigma)a\|_2^2 \stackrel{(45)}{\asymp} \frac{mM^2}{m-k} \|(AJ_\sigma)a\|_2^2. \end{aligned} \tag{52}$$

Recalling the definition of M in (34), since (52) holds true for every $a \in \mathbb{R}^\sigma$ we conclude that

$$\big\|(AJ_\sigma)^{-1}\big\|_{\mathsf{S}_\infty} \lesssim \frac{\sqrt{m}}{\sqrt{m-k}} \cdot \max_{j \in \{1,\ldots,m\}} \frac{1}{\|\mathsf{Proj}_{F_j} Ae_j\|_2}.$$

This is the desired estimate (8), which, together with (49), concludes the proof of Theorem 9. □

4.1. **Geometric interpretation of Theorem 9.** Theorem 21 below is due to Giannopoulos [Gia96]. It can be viewed as a geometric version of the Sauer–Shelah lemma for ellipsoids. A different geometric version of the Sauer–Shelah lemma was proved by Szarek–Talagrand [ST89].

Theorem 21 (Giannopoulos)**.** *There exists a universal constant $c \in (0, \infty)$ with the following property. Suppose that $m, n \in \mathbb{N}$ and $\varepsilon \in (0,1)$. Let $y_1, \ldots, y_m \in \mathbb{R}^n$ be vectors that satisfy $\|y_i\|_2 \leqslant 1$ for every $i \in \{1, \ldots, m\}$. Denote*

$$\mathcal{E} \stackrel{\mathrm{def}}{=} \Big\{ a = (a_1, \ldots, a_m) \in \mathbb{R}^m;\ \Big\| \sum_{j=1}^m a_j y_j \Big\|_2 \leqslant 1 \Big\}. \tag{53}$$

Then there exists a subset $\sigma \subseteq \{1, \ldots, m\}$ with $|\sigma| \geqslant (1-\varepsilon)m$ such that $\mathsf{Proj}_{\mathbb{R}^\sigma}(\mathcal{E}) \supseteq c\sqrt{\varepsilon} B_2^\sigma$, where $B_2^\sigma = \{x \in \mathbb{R}^\sigma :\ \|x\|_2 \leqslant 1\}$ denotes the unit Euclidean ball in $\mathbb{R}^\sigma$.

In this section we shall show that Theorem 21 is equivalent to Theorem 9, thus in particular describing a shorter proof of Theorem 9 that relies on Theorem 21.

Let us first prove that the validity of Theorem 9 implies the validity of Theorem 21. Suppose that we are in the setting that is described in the statement of Theorem 21. It was observed in [Gia96] that the validity of Theorem 21 under the additional assumption that $y_1,\ldots,y_m$ are linearly independent formally implies the validity of Theorem 21 in the above stated generality. Indeed, this follows by applying (the linear independent case of) Theorem 21 to the linearly independent vectors $y_1+e_{n+1}, y_2+e_{n+2},\ldots,y_m+e_{n+m}\in\mathbb{R}^{n+m}$. So, suppose that $y_1,\ldots,y_m\in\mathbb{R}^n$ are linearly independent and let $x_1,\ldots,x_m\in\mathbf{span}\{y_1,\ldots,y_m\}$ be the corresponding dual basis, i.e.,

$$\forall\, i,j\in\{1,\ldots,m\},\qquad \langle x_i,y_j\rangle=\delta_{ij}. \tag{54}$$

Define a linear operator $A:\mathbb{R}^m\to\mathbb{R}^n$ by setting $Ae_i=x_i$ for every $i\in\{1,\ldots,m\}$. Continuing with the notation for the subspace $F_j\subseteq\mathbb{R}^n$ that is given in (7) (with $\omega=\{1,\ldots,m\}$), we know by (54) that $y_j\in F_j$, so $\langle \mathsf{Proj}_{F_j}x_j,y_j\rangle=\langle x_j,y_j\rangle=1$. Since we are assuming in the setting of Theorem 21 that $\|y_j\|_2\leqslant 1$, this implies that $1=\langle \mathsf{Proj}_{F_j}x_j,y_j\rangle\leqslant\|y_j\|_2\cdot\|\mathsf{Proj}_{F_j}x_j\|_2\leqslant\|\mathsf{Proj}_{F_j}x_j\|_2$.

An application of Theorem 9 now shows that there exists $\sigma\subseteq\{1,\ldots,m\}$ with $|\sigma|\geqslant\lfloor(1-\varepsilon)m\rfloor$ and a universal constant $c\in(0,\infty)$ such that

$$\forall\, b\in\mathbb{R}^\sigma,\qquad \Big\|\sum_{j\in\sigma}b_jx_j\Big\|_2\geqslant c\sqrt{\varepsilon}\Big(\sum_{j\in\sigma}b_j^2\Big)^{\frac12}. \tag{55}$$

We claim that (55) implies that $\mathsf{Proj}_{\mathbb{R}^\sigma}(\mathcal{E})\supseteq c\sqrt{\varepsilon}B_2^\sigma$, where $\mathcal{E}$ is given in (53). Indeed, suppose that $a=\sum_{j\in\sigma}a_je_j\in\mathbb{R}^\sigma$ satisfies

$$a\in c\sqrt{\varepsilon}B_2^\sigma\iff\Big(\sum_{j\in\sigma}a_j^2\Big)^{\frac12}\leqslant c\sqrt{\varepsilon}. \tag{56}$$

Since the vectors $\{x_j\}_{j\in\sigma}\cup\{y_j\}_{j\in\{1,\ldots,m\}\smallsetminus\sigma}$ form a basis of $\mathbf{span}\{y_1,\ldots,y_m\}$, there exists a vector $b=(b_1,\ldots,b_m)\in\mathbb{R}^m$ such that

$$\sum_{j\in\sigma}a_jy_j=\sum_{j\in\sigma}b_jx_j+\sum_{j\in\{1,\ldots,m\}\smallsetminus\sigma}b_jy_j. \tag{57}$$

Denote

$$a^*=(a_1^*,\ldots,a_m^*)\stackrel{\mathrm{def}}{=}\sum_{j\in\sigma}a_je_j-\sum_{j\in\{1,\ldots,m\}\smallsetminus\sigma}b_je_j\in\mathbb{R}^m. \tag{58}$$

Then $\mathsf{Proj}_{\mathbb{R}^\sigma}a^*=a$ and

$$\begin{aligned}\Big\|\sum_{j=1}^m a_j^*y_j\Big\|_2^2&=\Big\langle\sum_{j=1}^m a_j^*y_j,\sum_{j=1}^m a_j^*y_j\Big\rangle\stackrel{(57)\wedge(58)}{=}\Big\langle\sum_{j=1}^m a_j^*y_j,\sum_{j\in\sigma}b_jx_j\Big\rangle\stackrel{(54)\wedge(58)}{=}\sum_{j\in\sigma}a_jb_j\\&\leqslant\Big(\sum_{j\in\sigma}a_j^2\Big)^{\frac12}\Big(\sum_{j\in\sigma}b_j^2\Big)^{\frac12}\stackrel{(56)}{\leqslant}c\sqrt{\varepsilon}\Big(\sum_{j\in\sigma}b_j^2\Big)^{\frac12}\stackrel{(55)}{\leqslant}\Big\|\sum_{j\in\sigma}b_jx_j\Big\|_2\stackrel{(57)\wedge(58)}{=}\Big\|\sum_{j=1}^m a_j^*y_j\Big\|_2.\end{aligned} \tag{59}$$

By cancelling $\big\|\sum_{j=1}^m a_j^*y_j\big\|_2$ from both sides of (59) and recalling (53), we conclude that $a^*\in\mathcal{E}$. Thus $a=\mathsf{Proj}_{\mathbb{R}^\sigma}a^*\in\mathsf{Proj}_{\mathbb{R}^\sigma}(\mathcal{E})$, as required.

Next, we shall prove the converse implication, i.e., that the validity of Theorem 21 implies the validity of Theorem 9. Suppose that we are in the setting of Theorem 9. As we explained in the beginning of Section 4, we may assume without loss of generality that $\omega=\{1,\ldots,m\}$, hence $\mathbf{rank}(A)=m$. Let $M\in(0,\infty)$ be defined as in (34), i.e., $M=\max_{j\in\{1,\ldots,m\}}\|\mathsf{Proj}_{F_j}Ae_j\|_2^{-1}$. Set

$$\forall\, i\in\{1,\ldots,m\},\qquad y_i\stackrel{\mathrm{def}}{=}\frac{\mathsf{Proj}_{F_i}Ae_i}{\|\mathsf{Proj}_{F_i}Ae_i\|_2}\in\mathbb{R}^n.$$

Then by definition $\|y_i\|_2=1$ for every $j\in\{1,\ldots,m\}$, and, by the same reasoning as in the beginning of Section 2.1, we know that $\langle y_j,Ae_j\rangle\geqslant 1/M$ and $\langle y_i,Ae_j\rangle=0$ for every distinct

$i,j \in \{1,\ldots,m\}$. By Theorem 21 applied with $\varepsilon = 1-k/m$ there exists $\sigma \subseteq \{1,\ldots,m\}$ of size $|\sigma| \geqslant (1-\varepsilon)m = k$ such that $\mathsf{Proj}_{\mathbb{R}^\sigma}(\mathcal{E}) \supseteq c\sqrt{\varepsilon}B_2^\sigma$, where $\mathcal{E}$ is defined in (53). Suppose that $a \in \mathbb{R}^\sigma \setminus \{0\}$. Then $c\sqrt{\varepsilon}a/\|a\|_2 \in \mathsf{Proj}_{\mathbb{R}^\sigma}(\mathcal{E})$, which means that there exists $b \in \mathbb{R}^m$ such that $b_j = c\sqrt{\varepsilon}a_j/\|a\|_2$ for every $j \in \sigma$ and (by the definition of $\mathcal{E}$) we have $\|\sum_{i=1}^m b_i y_i\|_2 \leqslant 1$. So,

$$\Big\|\sum_{j\in\sigma} a_j Ae_j\Big\|_2 \geqslant \Big\|\sum_{j\in\sigma} a_j Ae_j\Big\|_2 \cdot \Big\|\sum_{j=1}^m b_j y_j\Big\|_2 \geqslant \Big\langle \sum_{j\in\sigma} a_j Ae_j, \sum_{j=1}^m b_j y_j\Big\rangle$$
$$= \sum_{j\in\sigma} a_j b_j \langle Ae_j, y_j\rangle = \sum_{j\in\sigma} \frac{c\sqrt{\varepsilon}a_j^2}{\|a\|_2}\langle Ae_j, y_j\rangle \geqslant \frac{c\sqrt{\varepsilon}}{M\|a\|_2}\sum_{j\in\sigma} a_j^2 = \frac{c\sqrt{m-k}}{M\sqrt{m}}\|a\|_2.$$

This is precisely the desired conclusion in Theorem 9. □

5. Marcus–Spielman–Srivastava

Our goal here is to prove Theorem 11. This section differs from the previous sections in that we shall use the method of interlacing polynomials of Marcus–Spielman–Srivastava without sketching the proofs of the tools that we quote. The reason for this is that the ideas of Marcus–Spielman–Srivastava are remarkable and deep, but nevertheless elementary and accessible, and their presentation in [MSS15a, MSS15b] and especially in the beautiful survey [MSS14] (which is the main reference in the present section) is already a perfect exposition for a wide mathematical audience.

Suppose that $A:\mathbb{R}^m \to \mathbb{R}^n$ is a linear operator. Let $\mathsf{j}_1,\ldots,\mathsf{j}_k$ be i.i.d. random variables that are distributed uniformly over $\{1,\ldots,m\}$. For every $t \in \{1,\ldots,k\}$ consider the random vector

$$\mathsf{w}_t \stackrel{\mathrm{def}}{=} \sqrt{m}Ae_{\mathsf{j}_t}. \tag{60}$$

Then,

$$\mathbb{E}\big[\mathsf{w}_t \otimes \mathsf{w}_t\big] = \sum_{i=1}^m (Ae_i)\otimes(Ae_i) = AA^*. \tag{61}$$

Denote

$$\gamma \stackrel{\mathrm{def}}{=} \frac{\mathbf{rank}(A)\Big(\sqrt{\mathbf{rank}(A)}-\sqrt{k}\Big)^2}{\sum_{i=1}^{\mathbf{rank}(A)}\frac{1}{\mathsf{s}_i(A)^2}}. \tag{62}$$

With this notation, we shall prove below that

$$\Pr\left[\mathsf{s}_k\Big(\sum_{t=1}^k \mathsf{w}_t\otimes\mathsf{w}_t\Big) \geqslant \gamma\right] > 0. \tag{63}$$

Recalling (60), we see that (62) implies that there exist $j_1,\ldots,j_k \in \{1,\ldots,m\}$ such that

$$\mathsf{s}_k\Big(\sum_{t=1}^k (Ae_{j_t})\otimes(Ae_{j_t})\Big) \geqslant \frac{\gamma}{m} = \frac{\mathbf{rank}(A)\Big(\sqrt{\mathbf{rank}(A)}-\sqrt{k}\Big)^2}{m\sum_{i=1}^{\mathbf{rank}(A)}\frac{1}{\mathsf{s}_i(A)^2}}. \tag{64}$$

The rank of the operator $B \stackrel{\mathrm{def}}{=} \sum_{t=1}^k (Ae_{j_t})\otimes(Ae_{j_t})$ is at most the cardinality of $\sigma \stackrel{\mathrm{def}}{=} \{j_1,\ldots,j_k\}$. At the same time, by (64) we know that $\mathsf{s}_k(B) > 0$, because we are assuming that $k < \mathbf{rank}(A)$. Thus B has rank at least k, implying that the indices $j_1,\ldots,j_k$ are necessarily distinct, or equivalently that $|\sigma| = k$. Consequently $B = (AJ_\sigma)(AJ_\sigma)^*$ and $\mathsf{s}_k(B) = \mathsf{s}_{\min}(B) = \mathsf{s}_{\min}(AJ_\sigma)^2 = 1/\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty}^2$. Therefore (64) is the same as the desired restricted invertibility statement (11) of Theorem 11.

It remains to establish the validity of (63). Denote $Q \stackrel{\mathrm{def}}{=} AA^* : \mathbb{R}^n \to \mathbb{R}^n$ and let $\mathsf{q} : \mathbb{R} \to \mathbb{R}$ be the polynomial that is defined as follows.

$$\forall\, x \in \mathbb{R}, \qquad \mathsf{q}(x) \stackrel{\mathrm{def}}{=} (I - \partial_y)^k \mathbf{det}(xI_n + yQ)\Big|_{y=0},$$

where I denotes the identity operator on the space of polynomials and ∂_y is the differentiation operator with respect to the variable y (and, as before, I_n is the n by n identity matrix). By Theorem 4.5 in [MSS14], all the roots of the degree n polynomial q are real, and we denote their decreasing rearrangement by $\rho_1 \geqslant \rho_2 \geqslant \ldots \geqslant \rho_n$. Thus, ρ_k is the k'th largest root of q. A combination of Theorem 1.7 in [MSS14] and Theorem 4.1 in [MSS14] shows that

$$\Pr\left[\mathsf{s}_k\bigg(\sum_{t=1}^{k} \mathsf{w}_t \otimes \mathsf{w}_t\bigg) \geqslant \rho_k\right] > 0. \tag{65}$$

Consequently, in order to prove (63) it suffices to prove that $\rho_k \geqslant \gamma$, where γ is defined in (62).

Write $Q = U\Delta U^{-1}$, where $U : \mathbb{R}^n \to \mathbb{R}^n$ is an orthogonal matrix and $\Delta : \mathbb{R}^n \to \mathbb{R}^n$ is a diagonal matrix whose diagonal equals $(\mathsf{s}_1(A)^2, \ldots, \mathsf{s}_n(A)^2) \in \mathbb{R}^n$. Then for every $x, y \in \mathbb{R}$ we have

$$\mathbf{det}(xI_n + yQ) = \mathbf{det}\left(U(xI_n + y\Delta)U^{-1}\right) = \prod_{i=1}^{n}\left(x + y\mathsf{s}_i(A)^2\right) = x^{n-\mathbf{rank}(A)} \prod_{i=1}^{\mathbf{rank}(A)} \left(x + y\mathsf{s}_i(A)^2\right),$$

where we used the fact that $\mathsf{s}_i(A) = 0$ when $i > \mathbf{rank}(A)$. Consequently,

$$\mathsf{q}(x) = x^{n-\mathbf{rank}(A)}(I - \partial_y)^k \prod_{i=1}^{\mathbf{rank}(A)} \left(x + y\mathsf{s}_i(A)^2\right)\Big|_{y=0}. \tag{66}$$

We claim that if we denote by D the differentiation operator on the space of polynomials then

$$\mathsf{q}(x) = x^{n-k} \prod_{i=1}^{\mathbf{rank}(A)} \left(I - \mathsf{s}_i(A)^2\mathsf{D}\right) x^k. \tag{67}$$

The identity (67) is proven in the special case $\mathsf{s}_1(A) = \ldots = \mathsf{s}_{\mathbf{rank}(A)}(A) = 1$ in [MSS14]. The validity of (67) in full generality follows from checking that the coefficients of the polynomials that appear in the right hand sides of (66) and (67) are equal to each other. Indeed, starting with (66),

$$\begin{aligned} &x^{n-\mathbf{rank}(A)}(I - \partial_y)^k \prod_{i=1}^{\mathbf{rank}(A)} \left(x + y\mathsf{s}_i(A)^2\right)\Big|_{y=0} \\ &= x^{n-\mathbf{rank}(A)} \sum_{u=0}^{k} \binom{k}{u}(-1)^u \partial_y^u \sum_{\Omega \subseteq \{1,\ldots,\mathbf{rank}(A)\}} x^{\mathbf{rank}(A)-|\Omega|} y^{|\Omega|} \prod_{i\in\Omega} \mathsf{s}_i(A)^2\Big|_{y=0} \\ &= \sum_{\substack{\Omega \subseteq \{1,\ldots,\mathbf{rank}(A)\}\\ |\Omega| \leqslant k}} \frac{(-1)^{|\Omega|} x^{n-|\Omega|} k!}{(k-|\Omega|)!} \prod_{i\in\Omega} \mathsf{s}_i(A)^2, \end{aligned} \tag{68}$$

since $\partial_y^u y^{|\Omega|}|_{y=0} = |\Omega|! \cdot \mathbf{1}_{\{|\Omega|=u\}}$ for every $(u,\Omega) \in \{0,\ldots,k\} \times \{1,\ldots,\mathbf{rank}(A)\}$. At the same time,

$$x^{n-k} \prod_{i=1}^{\mathbf{rank}(A)} \left(I - \mathsf{s}_i(A)^2\mathsf{D}\right) x^k = x^{n-k} \sum_{\Omega \subseteq \{1,\ldots,\mathbf{rank}(A)\}} (-1)^{|\Omega|} \bigg(\prod_{i\in\Omega} \mathsf{s}_i(A)^2\bigg) \mathsf{D}^{|\Omega|} x^k. \tag{69}$$

Since for every for every $(u,\Omega) \in \{0,\ldots,k\} \times \{1,\ldots,\mathbf{rank}(A)\}$ we have $\mathsf{D}^{|\Omega|}x^k = 0$ if $|\Omega| > k$ and $\mathsf{D}^{|\Omega|}x^k = x^{k-|\Omega|}k!/(k-|\Omega|)!$ if $|\Omega| \leqslant k$, the validity of (67) follows by comparing (68) and (69).

Having established the identity (67), we shall proceed to prove the desired estimate $\rho_k \geqslant \gamma$ by applying the barrier method of [BSS12], reasoning along the lines of the argument that is presented in [MSS14]. Following [BSS12, SV13], given a polynomial $f : \mathbb{R} \to \mathbb{R}$ and $\phi \in (0, \infty)$ we consider the corresponding "soft spectral edge" $\mathbf{smin}_\phi(f) \in \mathbb{R}$, which is defined as follows

$$\mathbf{smin}_\phi(f) \stackrel{\mathrm{def}}{=} \inf\left\{b \in \mathbb{R} : \; f'(b) = -\phi f(b)\right\}. \tag{70}$$

As explained in [MSS14, Section 3.2], it is simple to check that for every $\phi \in (0, \infty)$ the smallest real root of f is at least the quantity $\mathbf{smin}_\phi(f)$. Hence, if we define

$$g(x) \stackrel{\mathrm{def}}{=} \prod_{i=1}^{\mathbf{rank}(A)} \left(I - \mathsf{s}_i(A)^2 \mathsf{D}\right) x^k, \tag{71}$$

then it follows from the above discussion and the identity (67) that it suffices to prove that

$$\sup_{\phi \in (0,\infty)} \mathbf{smin}_\phi(g) \geqslant \gamma. \tag{72}$$

Indeed, by (67) the n real roots of q consist of 0 with multiplicity $n - k$ and also the k roots of g (which are therefore necessarily real). Since g has degree k, the validity of (72) would imply that the smallest root of g is at least $\gamma > 0$, so the k'th largest root of q would be at least γ as well.

To prove (72), recall that Lemma 3.8 of [MSS14] asserts that for every polynomial $f : \mathbb{R} \to \mathbb{R}$ all of whose roots are real, and for every $\phi \in (0, \infty)$, we have

$$\mathbf{smin}_\phi\big((I - \mathsf{D})f\big) \geqslant \mathbf{smin}_\phi(f) + \frac{1}{1+\phi}. \tag{73}$$

For $\mathsf{s} \in (0, \infty)$ define $f_\mathsf{s} : \mathbb{R} \to \mathbb{R}$ by setting $f_\mathsf{s}(x) \stackrel{\mathrm{def}}{=} f(\mathsf{s}x)$ for every $x \in \mathbb{R}$. Observe that

$$\forall\, \mathsf{s} \in (0, \infty), \qquad (I - \mathsf{s}\mathsf{D})f = ((I - \mathsf{D})f_\mathsf{s})_{1/\mathsf{s}} \qquad \text{and} \qquad \mathbf{smin}_\phi(f_\mathsf{s}) \stackrel{(70)}{=} \frac{\mathbf{smin}_{\phi/\mathsf{s}}(f)}{\mathsf{s}}. \tag{74}$$

Consequently, for every real-rooted polynomial f and every $\mathsf{s}, \phi \in (0, \infty)$ we have

$$\begin{aligned} \mathbf{smin}_\phi\big((I - \mathsf{s}\mathsf{D})f\big) &\stackrel{(74)}{=} \mathbf{smin}_\phi\big(((I - \mathsf{D})f_\mathsf{s})_{1/\mathsf{s}}\big) \stackrel{(74)}{=} \mathsf{s} \cdot \mathbf{smin}_{\mathsf{s}\phi}\big((I - \mathsf{D})f_\mathsf{s}\big) \\ &\stackrel{(73)}{\geqslant} \mathsf{s}\left(\mathbf{smin}_{\mathsf{s}\phi}(f_\mathsf{s}) + \frac{1}{1 + \mathsf{s}\phi}\right) \stackrel{(74)}{=} \mathbf{smin}_\phi(f) + \frac{1}{\frac{1}{\mathsf{s}} + \phi}. \end{aligned} \tag{75}$$

By iterating (75) we see that

$$\begin{aligned} \mathbf{smin}_\phi(g) &\geqslant \mathbf{smin}_\phi(x^k) + \sum_{i=1}^{\mathbf{rank}(A)} \frac{1}{\frac{1}{\mathsf{s}_i(A)^2} + \phi} \\ &\stackrel{(70)}{=} -\frac{k}{\phi} + \sum_{i=1}^{\mathbf{rank}(A)} \frac{1}{\frac{1}{\mathsf{s}_i(A)^2} + \phi} \geqslant -\frac{k}{\phi} + \frac{\mathbf{rank}(A)}{\phi + \frac{1}{\mathbf{rank}(A)} \sum_{i=1}^{\mathbf{rank}(A)} \frac{1}{\mathsf{s}_i(A)^2}}, \end{aligned} \tag{76}$$

where the last step of (76) holds true due to the convexity of the function $x \mapsto 1/(\phi + x)$ on $(0, \infty)$. One checks that the value of ϕ that maximizes the right hand side of (76) is

$$\phi_{\max} \stackrel{\mathrm{def}}{=} \frac{\sqrt{k}}{\sqrt{\mathbf{rank}(A)} - \sqrt{k}} \left(\frac{1}{\mathbf{rank}(A)} \sum_{i=1}^{\mathbf{rank}(A)} \frac{1}{\mathsf{s}_i(A)^2}\right).$$

The right hand side of (76) equals γ when $\phi = \phi_{\max}$, so $\rho_k \geqslant \mathbf{smin}_{\phi_{\max}}(g) \geqslant \gamma$, as required. $\square$

Remark 22. The above argument actually yields a subset $\sigma \subseteq \{1,\ldots,m\}$ with $|\sigma| = k$ such that

$$\mathsf{s}_{\min}(AJ_\sigma)^2 = \mathsf{s}_k(AJ_\sigma)^2 \geqslant \frac{1}{m}\sup\left\{-\frac{k}{\phi} + \sum_{i=1}^{\mathbf{rank}(A)} \frac{\mathsf{s}_i(A)^2}{\phi + \mathsf{s}_i(A)^2} :\ \phi \in (0,\infty)\right\}. \qquad (77)$$

Indeed, continuing with the above notation, we explained why $\rho_k \geqslant \sup_{\phi\in(0,\infty)} \mathbf{smin}_\phi(g)$, so (77) follows from (65) and the penultimate step in (76).

The estimate (77) is more complicated than the assertion of Theorem 11, but it is sometimes significantly stronger. One such instance is the matrix A of Example 7. In that case, a somewhat tedious but straightforward computation allows one to obtain sharp estimates on the right hand side of (77), yielding bounds that coincide (up to constant factors) with those that are stated in Example 7 as a consequence of Theorem 6, while Theorem 11 yields much weaker bounds. There are also situations in which (77) yields worse bounds than those that follow from Theorem 9, e.g. when $\mathsf{s}_1(A) \asymp \ldots \asymp \mathsf{s}_m(A) \asymp 1$ and $k = (1-\varepsilon)m$ the bound on $\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty}$ that follows from (77) is $O(1/\varepsilon)$ while in the same situation Theorem 9 yields the bound $\|(AJ_\sigma)^{-1}\|_{\mathsf{S}_\infty} \lesssim 1/\sqrt{\varepsilon}$.

Acknowledgements. We thank Bill Johnson for helpful discussions. This work was initiated while we were participating in the workshop *Beyond Kadison–Singer: paving and consequences* at the American Institute of Mathematics. We thank the organizers for the excellent working conditions.

References

[ARS02] R. P. Anstee, L. Rónyai, and A. Sali. Shattering news. *Graphs Combin.*, 18(1):59–73, 2002.

[Bha97] R. Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.

[BHKW88] K. Berman, H. Halpern, V. Kaftal, and G. Weiss. Matrix norm inequalities and the relative Dixmier property. *Integral Equations Operator Theory*, 11(1):28–48, 1988.

[BS88] J. Bourgain and S. J. Szarek. The Banach-Mazur distance to the cube and the Dvoretzky-Rogers factorization. *Israel J. Math.*, 62(2):169–180, 1988.

[BSS12] J. Batson, D. A. Spielman, and N. Srivastava. Twice-Ramanujan sparsifiers. *SIAM J. Comput.*, 41(6):1704–1721, 2012.

[BT87] J. Bourgain and L. Tzafriri. Invertibility of "large" submatrices with applications to the geometry of Banach spaces and harmonic analysis. *Israel J. Math.*, 57(2):137–224, 1987.

[BT89] J. Bourgain and L. Tzafriri. Restricted invertibility of matrices and applications. In *Analysis at Urbana, Vol. II (Urbana, IL, 1986–1987)*, volume 138 of *London Math. Soc. Lecture Note Ser.*, pages 61–107. Cambridge Univ. Press, Cambridge, 1989.

[BT91] J. Bourgain and L. Tzafriri. On a problem of Kadison and Singer. *J. Reine Angew. Math.*, 420:1–43, 1991.

[Dav79] P. J. Davis. *Circulant matrices*. John Wiley & Sons, New York-Chichester-Brisbane, 1979. A Wiley-Interscience Publication, Pure and Applied Mathematics.

[DJT95] J. Diestel, H. Jarchow, and A. Tonge. *Absolutely summing operators*, volume 43 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995.

[Fan49] K. Fan. On a theorem of Weyl concerning eigenvalues of linear transformations. I. *Proc. Nat. Acad. Sci. U. S. A.*, 35:652–655, 1949.

[Gia95] A. A. Giannopoulos. A note on the Banach-Mazur distance to the cube. In *Geometric aspects of functional analysis (Israel, 1992–1994)*, volume 77 of *Oper. Theory Adv. Appl.*, pages 67–73. Birkhäuser, Basel, 1995.

[Gia96] A. A. Giannopoulos. A proportional Dvoretzky-Rogers factorization result. *Proc. Amer. Math. Soc.*, 124(1):233–241, 1996.

[Gro53] A. Grothendieck. Résumé de la théorie métrique des produits tensoriels topologiques. *Bol. Soc. Mat. São Paulo*, 8:1–79, 1953.

[KS12] I. Kra and S. R. Simanca. On circulant matrices. *Notices Amer. Math. Soc.*, 59(3):368–377, 2012.

[MSS14] A. W. Marcus, D. A. Spielman, and N. Srivastava. Ramanujan graphs and the solution of the Kadison–Singer problem. In *Proceedings of the 2014 International Congress of Mathematicians. Volume III*, pages 363–386. 2014. Available at `http://www.icm2014.org/en/vod/proceedings`.

[MSS15a] A. W. Marcus, D. A. Spielman, and N. Srivastava. Interlacing families I: Bipartite Ramanujan graphs of all degrees. *Ann. of Math. (2)*, 182(1):307–325, 2015.

[MSS15b] A. W. Marcus, D. A. Spielman, and N. Srivastava. Interlacing families II: Mixed characteristic polynomials and the Kadison-Singer problem. *Ann. of Math. (2)*, 182(1):327–350, 2015.

[MSS16] A. W. Marcus, D. A. Spielman, and N. Srivastava. Interlacing families III: Improved restricted invertibility estimates, 2016. In preparation.

[Nao12] A. Naor. Sparse quadratic forms and their geometric applications [following Batson, Spielman and Srivastava]. *Astérisque*, (348):Exp. No. 1033, viii, 189–217, 2012. Séminaire Bourbaki: Vol. 2010/2011. Exposés 1027–1042.

[Paj85] A. Pajor. *Sous-espaces l_1^n des espaces de Banach*, volume 16 of *Travaux en Cours [Works in Progress]*. Hermann, Paris, 1985. With an introduction by Gilles Pisier.

[Pie67] A. Pietsch. Absolut p-summierende Abbildungen in normierten Räumen. *Studia Math.*, 28:333–353, 1966/1967.

[Pis86] 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; by the American Mathematical Society, Providence, RI, 1986.

[Sau72] N. Sauer. On the density of families of sets. *J. Combinatorial Theory Ser. A*, 13:145–147, 1972.

[She72] S. Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific J. Math.*, 41:247–261, 1972.

[SS12] D. A. Spielman and N. Srivastava. An elementary proof of the restricted invertibility theorem. *Israel J. Math.*, 190:83–91, 2012.

[ST89] S. J. Szarek and M. Talagrand. An "isomorphic" version of the Sauer-Shelah lemma and the Banach-Mazur distance to the cube. In *Geometric aspects of functional analysis (1987–88)*, volume 1376 of *Lecture Notes in Math.*, pages 105–112. Springer, Berlin, 1989.

[Stø13] E. Størmer. *Positive linear maps of operator algebras*. Springer Monographs in Mathematics. Springer, Heidelberg, 2013.

[SV13] N. Srivastava and R. Vershynin. Covariance estimation for distributions with $2+\varepsilon$ moments. *Ann. Probab.*, 41(5):3081–3111, 2013.

[Sza91] S. J. Szarek. On the geometry of the Banach-Mazur compactum. In *Functional analysis (Austin, TX, 1987/1989)*, volume 1470 of *Lecture Notes in Math.*, pages 48–59. Springer, Berlin, 1991.

[Tro09] J. A. Tropp. Column subset selection, matrix factorization, and eigenvalue optimization. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 978–986. SIAM, Philadelphia, PA, 2009.

[Ver01] R. Vershynin. John's decompositions: selecting a large part. *Israel J. Math.*, 122:253–277, 2001.

[You14] P. Youssef. Restricted invertibility and the Banach-Mazur distance to the cube. *Mathematika*, 60(1):201–218, 2014.

Mathematics Department, Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544-1000, USA

E-mail address: naor@math.princeton.edu

Laboratoire de Probabilités et de Modèles Aléatoires, Université Paris-Diderot, 5 rue Thomas Mann 75205, Paris CEDEX 13, France

E-mail address: youssef@math.univ-paris-diderot.fr