The Restricted Isometry Property of Subsampled Fourier Matrices

Ishay Haviv and Oded Regev

Abstract A matrix $A \in \mathbb{C}^{q \times N}$ satisfies the *restricted isometry property* of order k with constant ϵ if it preserves the ℓ_2 norm of all k-sparse vectors up to a factor of $1 \pm \epsilon$. We prove that a matrix A obtained by randomly sampling $q = O(k \cdot \log^2 k \cdot \log N)$ rows from an $N \times N$ Fourier matrix satisfies the restricted isometry property of order k with a fixed ϵ with high probability. This improves on Rudelson and Vershynin (Comm Pure Appl Math, 2008), its subsequent improvements, and Bourgain (GAFA Seminar Notes, 2014).

1 Introduction

A matrix $A \in \mathbb{C}^{q \times N}$ satisfies the *restricted isometry property* of order k with constant $\epsilon > 0$ if for every k-sparse vector $x \in \mathbb{C}^N$ (i.e., a vector with at most k nonzero entries), it holds that

$$(1 - \epsilon) \cdot \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \epsilon) \cdot \|x\|_2^2 \,. \tag{1}$$

Intuitively, this means that every k columns of A are nearly orthogonal. This notion, due to Candès and Tao [9], was intensively studied during the last decade and found various applications and connections to several areas of theoretical computer science, including sparse recovery [8, 20, 27], coding theory [14], norm embeddings [6, 22], and computational complexity [4, 25, 31].

The original motivation for the restricted isometry property comes from the area of compressed sensing. There, one wishes to compress a high-dimensional sparse vector $x \in \mathbb{C}^N$ to a vector Ax, where $A \in \mathbb{C}^{q \times N}$ is a measurement matrix that enables

A preliminary version appeared in Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 2016, pages 288–297.

I. Haviv (✉)
School of Computer Science, The Academic College of Tel Aviv-Yaffo, Tel Aviv 61083, Israel

O. Regev
Courant Institute of Mathematical Sciences, New York University, New York, NY, USA

B. Klartag, E. Milman (eds.), *Geometric Aspects of Functional Analysis*,
Lecture Notes in Mathematics 2169, DOI 10.1007/978-3-319-45282-1_11

reconstruction of x from Ax. Typical goals in this context include minimizing the number of measurements q and the running time of the reconstruction algorithm. It is known that the restricted isometry property of A, for $\epsilon < \sqrt{2} - 1$, is a sufficient condition for reconstruction. In fact, it was shown in [8, 9, 11, 12] that under this condition, reconstruction is equivalent to finding the vector of least ℓ_1 norm among all vectors that agree with the given measurements, a task that can be formulated as a linear program [13, 16], and thus can be solved efficiently.

The above application leads to the challenge of finding matrices $A \in \mathbb{C}^{q \times N}$ that satisfy the restricted isometry property and have a small number of rows q as a function of N and k. (For simplicity, we ignore for now the dependence on ϵ.) A general lower bound of $q = \Omega(k \cdot \log(N/k))$ is known to follow from [18] (see also [17]). Fortunately, there are matrices that match this lower bound, e.g., random matrices whose entries are chosen independently according to the normal distribution [10]. However, in many applications the measurement matrix cannot be chosen arbitrarily but is instead given by a random sample of rows from a unitary matrix, typically the discrete Fourier transform. This includes, for instance, various tests and experiments in medicine and biology (e.g., MRI [28] and ultrasound imaging [21]) and applications in astronomy (e.g., radio telescopes [32]). An advantage of subsampled Fourier matrices is that they support fast matrix-vector multiplication, and as such, are useful for efficient compression as well as for efficient reconstruction based on iterative methods (see, e.g., [26]).

In recent years, with motivation from both theory and practice, an intensive line of research has aimed to study the restricted isometry property of random submatrices of unitary matrices. Letting $A \in \mathbb{C}^{q \times N}$ be a (normalized) matrix whose rows are chosen uniformly and independently from the rows of a unitary matrix $M \in \mathbb{C}^{N \times N}$, the goal is to prove an upper bound on q for which A is guaranteed to satisfy the restricted isometry property with high probability. Note that the fact that the entries of every row of A are not independent makes this question much more difficult than in the case of random matrices with independent entries.

The first upper bound on the number of rows of a subsampled Fourier matrix that satisfies the restricted isometry property was $O(k \cdot \log^6 N)$, which was proved by Candès and Tao [10]. This was then improved by Rudelson and Vershynin [30] to $O(k \cdot \log^2 k \cdot \log(k \log N) \cdot \log N)$ (see also [15, 29] for a simplified analysis with better success probability). A modification of their analysis led to an improved bound of $O(k \cdot \log^3 k \cdot \log N)$ by Cheraghchi, Guruswami, and Velingker [14], who related the problem to a question on the list-decoding rate of random linear codes over finite fields. Interestingly, replacing the $\log(k \log N)$ term in the bound of [30] by $\log k$ was crucial for their application.[1] Recently, Bourgain [7] proved a bound of $O(k \cdot \log k \cdot \log^2 N)$, which is incomparable to those of [14, 30] (and has a worse dependence on ϵ; see below). We finally mention that the best known lower bound on the number of rows is $\Omega(k \cdot \log N)$ [5].

[1] Note that the list-decoding result of [14] was later improved by Wootters [33] using different techniques.

1.1 Our Contribution

In this work, we improve the previous bounds and prove the following.

Theorem 1.1 (Simplified) *Let $M \in \mathbb{C}^{N\times N}$ be a unitary matrix with entries of absolute value $O(1/\sqrt{N})$, and let $\epsilon > 0$ be a fixed constant. For some $q = O(k\cdot\log^2 k\cdot\log N)$, let $A \in \mathbb{C}^{q\times N}$ be a matrix whose q rows are chosen uniformly and independently from the rows of M, multiplied by $\sqrt{N/q}$. Then, with high probability, the matrix A satisfies the restricted isometry property of order k with constant ϵ.*

The main idea in our proof is described in Sect. 1.3. We arrived at the proof from our recent work on list-decoding [19], where a baby version of the idea was used to bound the sample complexity of learning the class of Fourier-sparse Boolean functions.[2] Like all previous work on this question, our proof can be seen as a careful union bound applied to a sequence of progressively finer nets, a technique sometimes known as chaining. However, unlike the work of Rudelson and Vershynin [30] and its improvements [14, 15], we avoid the use of Gaussian processes, the "symmetrization process," and Dudley's inequality. Instead, we follow and refine Bourgain's proof [7], and apply the chaining argument directly to the problem at hand using only elementary arguments. It would be interesting to see if our proof can be cast in the Gaussian framework of Rudelson and Vershynin.

We remark that the bounds obtained in the previous works [14, 30] have a multiplicative $O(\epsilon^{-2})$ term, whereas a much worse term of $O(\epsilon^{-6})$ was obtained in [7]. In our proof of Theorem 1.1 we nearly obtain the best known dependence on ϵ. For simplicity of presentation we first prove in Sect. 3 our bound with a weaker multiplicative term of $O(\epsilon^{-4})$, and then, in Sect. 4, we modify the analysis and decrease the dependence on ϵ to $O(\epsilon^{-2})$ up to logarithmic terms.

1.2 Related Literature

As mentioned before, one important advantage of using subsampled Fourier matrices in compressed sensing is that they support fast, in fact nearly linear time, matrix-vector multiplication. In certain scenarios, however, one is not restricted to using subsampled Fourier matrices as the measurement matrix. The question then is whether one can decrease the number of rows using another measurement matrix, while still keeping the near-linear multiplication time. For $k < N^{1/2-\gamma}$ where $\gamma > 0$ is an arbitrary constant, the answer is yes: a construction with the *optimal* number

[2] The result in [19] is weaker in two main respects. First, it is restricted to the case that Ax is in $\{0,1\}^q$. This significantly simplifies the analysis and leads to a better bound on the number of rows of A. Second, the order of quantifiers is switched, namely it shows that for any sparse x, a random subsampled A works with high probability, whereas for the restricted isometry property we need to show that a random A works for all sparse x.

$O(k \cdot \log N)$ of rows follows from works by Ailon and Chazelle [1] and Ailon and Liberty [2] (see [6]). For general k, Nelson, Price, and Wootters [27] suggested taking subsampled Fourier matrices and "tweaking" them by bunching together rows with random signs. Using the Gaussian-process-based analysis of [14, 30] and introducing further techniques from [23], they showed that with this construction one can reduce the number of rows by a logarithmic factor to $O(k \cdot \log^2(k \log N) \cdot \log N)$ while still keeping the nearly linear multiplication time. Our result shows that the same number of rows (in fact, a slightly smaller number) can be achieved already with the original subsampled Fourier matrices without having to use the "tweak." A natural open question is whether the "tweak" from [27] and their techniques can be combined with ours to further reduce the number of rows. An improvement in the regime of parameters of $k = \omega(\sqrt{N})$ would lead to more efficient low-dimensional embeddings based on Johnson–Lindenstrauss matrices (see, e.g., [1–3, 22, 27]).

1.3 Proof Overview

Recall from Theorem 1.1 and from (1) that our goal is to prove that a matrix A given by a random sample Q of q rows of M satisfies with high probability that for all k-sparse x, $\|Ax\|_2^2 \approx \|x\|_2^2$. Since M is unitary, the latter is equivalent to saying that $\|Ax\|_2^2 \approx \|Mx\|_2^2$. Yet another way of expressing this condition is as

$$\mathop{\mathbb{E}}_{j \in Q} \left[(|Mx|^2)_j \right] \approx \mathop{\mathbb{E}}_{j \in [N]} \left[(|Mx|^2)_j \right] ,$$

i.e., that a sample $Q \subseteq [N]$ of q coordinates of the vector $|Mx|^2$ gives a good approximation to the average of all its coordinates. Here, $|Mx|^2$ refers to the vector obtained by taking the squared absolute value of Mx coordinate-wise. For reasons that will become clear soon, it will be convenient to assume without loss of generality that $\|x\|_1 = 1$. With this scaling, the sparsity assumption implies that $\|Mx\|_2^2$ is not too small (namely at least $1/k$), and this will determine the amount of additive error we can afford in the approximation above. This is the only way we use the sparsity assumption.

At a high level, the proof proceeds by defining a finite set of vectors $\mathcal{H}$ that forms a *net*, i.e., a set satisfying that any vector $|Mx|^2$ is close to one of the vectors in $\mathcal{H}$. We then argue using the Chernoff-Hoeffding bound that for any fixed vector $h \in \mathcal{H}$, a sample of q coordinates gives a good approximation to the average of h. Finally, we complete the proof by a union bound over all $h \in \mathcal{H}$.

In order to define the set $\mathcal{H}$ we notice that since $\|x\|_1 = 1$, Mx can be seen as a weighted average of the columns of M (possibly with signs). In other words, we can think of Mx as the *expectation* of a vector-valued random variable given by a certain probability distribution over the columns of M. Using the Chernoff-Hoeffding bound again, this implies that we can approximate Mx well by taking the average over a small number of samples from this distribution. We then let $\mathcal{H}$ be the

set of all possible such averages, and a bound on the cardinality of $\mathcal{H}$ follows easily (basically N raised to the number of samples). This technique is sometimes referred to as Maurey's empirical method.

The argument above is actually oversimplified, and carrying it out leads to rather bad bounds on q. As a result, our proof in Sect. 3 is slightly more delicate. Namely, instead of just one set $\mathcal{H}$, we have a sequence of sets, $\mathcal{H}_1, \mathcal{H}_2, \ldots$, each being responsible for approximating a different scale of $|Mx|^2$. The first set $\mathcal{H}_1$ approximates $|Mx|^2$ on coordinates on which its value is highest; since the value is high, we need less samples in order to approximate it well, as a result of which the set $\mathcal{H}_1$ is small. The next set $\mathcal{H}_2$ approximates $|Mx|^2$ on coordinates on which its value is somewhat smaller, and is therefore a bigger set, and so on and so forth. The end result is that any vector $|Mx|^2$ can be approximately decomposed into a sum $\sum_i h^{(i)}$, with $h^{(i)} \in \mathcal{H}_i$. To complete the proof, we argue that a random choice of q coordinates approximates all the vectors in all the $\mathcal{H}_i$ well. The reason working with several $\mathcal{H}_i$ leads to the better bound stated in Theorem 1.1 is this: even though as i increases the number of vectors in $\mathcal{H}_i$ grows, the quality of approximation that we need the q coordinates to provide decreases, since the value of $|Mx|^2$ there is small and so errors are less significant. It turns out that these two requirements on q balance each other perfectly, leading to the desired bound on q.

2 Preliminaries

Notation The notation $x \approx_{\epsilon,\alpha} y$ means that $x \in [(1-\epsilon)y - \alpha, (1+\epsilon)y + \alpha]$. For a matrix M, we denote by $M^{(\ell)}$ the ℓth column of M and define $\|M\|_\infty = \max_{i,j} |M_{i,j}|$.

The Restricted Isometry Property The restricted isometry property is defined as follows.

Definition 2.1 We say that a matrix $A \in \mathbb{C}^{q \times N}$ satisfies the *restricted isometry property* of order k with constant ϵ if for every k-sparse vector $x \in \mathbb{C}^N$ it holds that

$$(1-\epsilon) \cdot \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\epsilon) \cdot \|x\|_2^2.$$

Chernoff-Hoeffding Bounds We now state the Chernoff-Hoeffding bound (see, e.g., [24]) and derive several simple corollaries that will be used extensively later.

Theorem 2.2 *Let $X_1, \ldots, X_N$ be N identically distributed independent random variables in $[0, a]$ satisfying $\mathbb{E}[X_i] = \mu$ for all i, and denote $\overline{X} = \frac{1}{N} \cdot \sum_{i=1}^N X_i$. Then there exists a universal constant C such that for every $0 < \epsilon \leq 1/2$, the probability that $\overline{X} \approx_{\epsilon,0} \mu$ is at least $1 - 2e^{-C \cdot N \mu \epsilon^2 / a}$.*

Corollary 2.3 *Let $X_1, \ldots, X_N$ be N identically distributed independent random variables in $[0, a]$ satisfying $\mathbb{E}[X_i] = \mu$ for all i, and denote $\overline{X} = \frac{1}{N} \cdot \sum_{i=1}^N X_i$.*

Then there exists a universal constant C such that for every $0 < \epsilon \leq 1/2$ and $\alpha > 0$, the probability that $\overline{X} \approx_{\epsilon,\alpha} \mu$ is at least $1 - 2e^{-C \cdot N\alpha\epsilon/a}$.

Proof If $\mu \geq \frac{\alpha}{\epsilon}$ then by Theorem 2.2 the probability that $\overline{X} \approx_{\epsilon,0} \mu$ is at least $1 - 2e^{-C \cdot N\mu\epsilon^2/a}$, which is at least $1 - 2e^{-C \cdot N\alpha\epsilon/a}$. Otherwise, Theorem 2.2 for $\tilde{\varepsilon} = \frac{\alpha}{\mu} > \epsilon$ implies that the probability that $\overline{X} \approx_{\tilde{\varepsilon},0} \mu$, hence $\overline{X} \approx_{0,\alpha} \mu$, is at least $1 - 2e^{-C \cdot N\mu\tilde{\varepsilon}^2/a}$, and the latter is at least $1 - 2e^{-C \cdot N\alpha\epsilon/a}$. ■

Corollary 2.4 *Let $X_1, \ldots, X_N$ be N identically distributed independent random variables in $[-a, +a]$ satisfying $\mathbb{E}[X_i] = \mu$ and $\mathbb{E}[|X_i|] = \tilde{\mu}$ for all i, and denote $\overline{X} = \frac{1}{N} \cdot \sum_{i=1}^{N} X_i$. Then there exists a universal constant C such that for every $0 < \varepsilon' \leq 1/2$ and $\alpha > 0$, the probability that $\overline{X} \approx_{0,\epsilon' \cdot \tilde{\mu} + \alpha} \mu$ is at least $1 - 4e^{-C \cdot N\alpha\epsilon'/a}$.*

Proof The corollary follows by applying Corollary 2.3 to $\max(X_i, 0)$ and to $-\min(X_i, 0)$. ■

We end with the additive form of the bound, followed by an easy extension to the complex case.

Corollary 2.5 *Let $X_1, \ldots, X_N$ be N identically distributed independent random variables in $[-a, +a]$ satisfying $\mathbb{E}[X_i] = \mu$ for all i, and denote $\overline{X} = \frac{1}{N} \cdot \sum_{i=1}^{N} X_i$. Then there exists a universal constant C such that for every $b > 0$, the probability that $\overline{X} \approx_{0,b} \mu$ is at least $1 - 4e^{-C \cdot Nb^2/a^2}$.*

Proof We can assume that $b \leq 2a$. The corollary follows by applying Corollary 2.4 to, say, $\alpha = 3b/4$ and $\epsilon' = b/(4a)$. ■

Corollary 2.6 *Let $X_1, \ldots, X_N$ be N identically distributed independent complex-valued random variables satisfying $|X_i| \leq a$ and $\mathbb{E}[X_i] = \mu$ for all i, and denote $\overline{X} = \frac{1}{N} \cdot \sum_{i=1}^{N} X_i$. Then there exists a universal constant C such that for every $b > 0$, the probability that $|\overline{X}| \approx_{0,b} |\mu|$ is at least $1 - 8e^{-C \cdot Nb^2/a^2}$.*

Proof By Corollary 2.5 applied to the real and imaginary parts of the random variables $X_1, \ldots, X_N$ it follows that for a universal constant C, the probability that $\mathsf{Re}(\overline{X}) \approx_{0,b/\sqrt{2}} \mathsf{Re}(\mu)$ and $\mathsf{Im}(\overline{X}) \approx_{0,b/\sqrt{2}} \mathsf{Im}(\mu)$ is at least $1 - 8e^{-C \cdot Nb^2/a^2}$. By triangle inequality, it follows that with such probability we have $|\overline{X}| \approx_{0,b} |\mu|$, as required. ■

3 The Simpler Analysis

In this section we prove our result with a multiplicative term of $O(\epsilon^{-4})$ in the bound. This will be obtained in Theorem 3.7 as an easy corollary of the following theorem.

Theorem 3.1 *For a sufficiently large N, a matrix $M \in \mathbb{C}^{N \times N}$, and sufficiently small $\epsilon, \eta > 0$, the following holds. For some $q = O(\epsilon^{-3}\eta^{-1} \log N \cdot \log^2(1/\eta))$, let Q*

be a multiset of q uniform and independent random elements of $[N]$*. Then, with probability* $1 - 2^{-\Omega(\epsilon^{-2}\cdot\log N\cdot\log(1/\eta))}$*, it holds that for every* $x \in \mathbb{C}^N$*,*

$$\mathop{\mathbb{E}}_{j\in Q}\left[|(Mx)_j|^2\right] \approx_{\epsilon,\eta\cdot\|x\|_1^2\cdot\|M\|_\infty^2} \mathop{\mathbb{E}}_{j\in[N]}\left[|(Mx)_j|^2\right].$$

Throughout the proof we assume without loss of generality that the matrix $M \in \mathbb{C}^{N\times N}$ satisfies $\|M\|_\infty = 1$. For $\epsilon, \eta > 0$, we denote $t = \log_2(1/\eta)$, $r = \log_2(1/\epsilon^2)$, and $\gamma = \eta/(2t)$.

We now define the approximating vector sets $\mathcal{H}_i$, $i = 1, \ldots, t$, each responsible for coordinates of $|Mx|^2$ of a different scale (the larger the i the smaller the scale). We start by defining the "raw approximations" $\mathcal{G}_i$, which are essentially vectors obtained by averaging a certain number of columns of M. We then define the vectors in $\mathcal{H}_i$ by restricting the vectors in $\mathcal{G}_i$ (actually $\mathcal{G}_{i+r}$) to the set of coordinates B_i where there is a clear "signal" and not just noise. This is necessary in order to make sure that the small coordinates of $|Mx|^2$ are not flooded by noise from the coarse approximations. Details follow.

The Vector Sets $\mathcal{G}_i$ For every $1 \le i \le t + r$, let $\mathcal{G}_i$ denote the set of all vectors $g^{(i)} \in \mathbb{C}^N$ that can be represented as

$$g^{(i)} = \frac{\sqrt{2}}{|F|}\cdot\sum_{(\ell,s)\in F}(-1)^{s/2}\cdot M^{(\ell)} \tag{2}$$

for a multiset F of $O(2^i\cdot\log(1/\gamma))$ pairs in $[N]\times\{0,1,2,3\}$. A trivial counting argument gives the following.

Claim 3.2 *For every* $1 \le i \le t + r$, $|\mathcal{G}_i| \le N^{O(2^i\cdot\log(1/\gamma))}$.

The Vector Sets $\mathcal{H}_i$ For a t-tuple of vectors $(g^{(1+r)}, \ldots, g^{(t+r)}) \in \mathcal{G}_{1+r}\times\cdots\times\mathcal{G}_{t+r}$ and for $1 \le i \le t$, let B_i be the set of all $j \in [N]$ for which i is the smallest index satisfying $|g_j^{(i+r)}| \ge 2\cdot 2^{-i/2}$. For such i, define the vector $h^{(i)}$ by

$$h_j^{(i)} = \min(|g_j^{(i+r)}|^2\cdot\mathbf{1}_{j\in B_i}, 9\cdot 2^{-i}). \tag{3}$$

Let $\mathcal{H}_i$ be the set of all vectors $h^{(i)}$ that can be obtained in this way.

Claim 3.3 *For every* $1 \le i \le t$, $|\mathcal{H}_i| \le N^{O(\epsilon^{-2}\cdot 2^i\cdot\log(1/\gamma))}$.

Proof Observe that every $h^{(i)} \in \mathcal{H}_i$ is fully defined by some $(g^{(1+r)}, \ldots, g^{(i+r)}) \in \mathcal{G}_{1+r}\times\cdots\times\mathcal{G}_{i+r}$. Hence

$$|\mathcal{H}_i| \le |\mathcal{G}_{1+r}|\cdots|\mathcal{G}_{i+r}| \le N^{O(\log(1/\gamma))\cdot(2^{1+r}+2^{2+r}+\cdots+2^{i+r})} \le N^{O(\log(1/\gamma))\cdot 2^{i+r+1}}.$$

Using the definition of r, the claim follows. ∎

Lemma 3.4 *For every $\tilde{\eta} > 0$ and some $q = O(\epsilon^{-3}\tilde{\eta}^{-1}\log N \cdot \log(1/\gamma))$, let Q be a multiset of q uniform and independent random elements of $[N]$. Then, with probability $1 - 2^{-\Omega(\epsilon^{-2}\cdot\log N\cdot\log(1/\gamma))}$, it holds that for all $1 \le i \le t$ and $h^{(i)} \in \mathcal{H}_i$,*

$$\mathop{\mathbb{E}}_{j\in Q}\left[h_j^{(i)}\right] \approx_{\epsilon,\tilde{\eta}} \mathop{\mathbb{E}}_{j\in[N]}\left[h_j^{(i)}\right].$$

Proof Fix an $1 \le i \le t$ and a vector $h^{(i)} \in \mathcal{H}_i$, and denote $\mu = \mathbb{E}_{j\in[N]}[h_j^{(i)}]$. By Corollary 2.3, applied with $\alpha = \tilde{\eta}$ and $a = 9 \cdot 2^{-i}$ (recall that $h_j^{(i)} \le a$ for every j), with probability $1 - 2^{-\Omega(2^i\cdot q\epsilon\tilde{\eta})}$, it holds that $\mathbb{E}_{j\in Q}[h_j^{(i)}] \approx_{\epsilon,\tilde{\eta}} \mu$. Using Claim 3.3, the union bound over all the vectors in $\mathcal{H}_i$ implies that the probability that some $h^{(i)} \in \mathcal{H}_i$ does not satisfy $\mathbb{E}_{j\in Q}[h_j^{(i)}] \approx_{\epsilon,\tilde{\eta}} \mu$ is at most

$$N^{O(\epsilon^{-2}\cdot 2^i\cdot\log(1/\gamma))} \cdot 2^{-\Omega(2^i\cdot q\epsilon\tilde{\eta})} \le 2^{-\Omega(\epsilon^{-2}\cdot 2^i\cdot\log N\cdot\log(1/\gamma))}.$$

We complete the proof by a union bound over i. ∎

Approximating the Vectors *Mx*

Lemma 3.5 *For every vector $x \in \mathbb{C}^N$ with $\|x\|_1 = 1$, every multiset $Q \subseteq [N]$, and every $1 \le i \le t + r$, there exists a vector $g \in \mathcal{G}_i$ that satisfies $|(Mx)_j| \approx_{0,2^{-i/2}} |g_j|$ for all but at most γ fraction of $j \in [N]$ and for all but at most γ fraction of $j \in Q$.*

Proof Observe that for every $\ell \in [N]$ there exist $p_{\ell,0}, p_{\ell,1}, p_{\ell,2}, p_{\ell,3} \ge 0$ that satisfy

$$\sum_{s=0}^{3} p_{\ell,s} = |x_\ell| \quad \text{and} \quad \sqrt{2}\cdot\sum_{s=0}^{3} p_{\ell,s}\cdot(-1)^{s/2} = x_\ell.$$

Notice that the assumption $\|x\|_1 = 1$ implies that the numbers $p_{\ell,s}$ form a probability distribution. Thus, the vector Mx can be represented as

$$Mx = \sum_{\ell=1}^{N} x_\ell \cdot M^{(\ell)} = \sqrt{2}\cdot\sum_{\ell=1}^{N}\sum_{s=0}^{3} p_{\ell,s}\cdot(-1)^{s/2}\cdot M^{(\ell)} = \mathop{\mathbb{E}}_{(\ell,s)\sim D}[\sqrt{2}\cdot(-1)^{s/2}\cdot M^{(\ell)}],$$

where D is the distribution that assigns probability $p_{\ell,s}$ to the pair (ℓ, s).

Let F be a multiset of $O(2^i \cdot \log(1/\gamma))$ independent random samples from D, and let $g \in \mathcal{G}_i$ be the vector corresponding to F as in (2). By Corollary 2.6, applied with $a = \sqrt{2}$ (recall that $\|M\|_\infty = 1$) and $b = 2^{-i/2}$, for every $j \in [N]$ the probability that

$$|(Mx)_j| \approx_{0,2^{-i/2}} |g_j| \tag{4}$$

is at least $1 - \gamma/4$. It follows that the expected number of $j \in [N]$ that do not satisfy (4) is at most $\gamma N/4$, so by Markov's inequality the probability that the

number of $j \in [N]$ that do not satisfy (4) is at most γN is at least $3/4$. Similarly, the expected number of $j \in Q$ that do not satisfy (4) is at most $\gamma|Q|/4$, so by Markov's inequality, with probability at least $3/4$ it holds that the number of $j \in Q$ that do not satisfy (4) is at most $\gamma|Q|$. It follows that there exists a vector $g \in \mathcal{G}_i$ for which (4) holds for all but at most γ fraction of $j \in [N]$ and for all but at most γ fraction of $j \in Q$, as required. ■

Lemma 3.6 *For every multiset $Q \subseteq [N]$ and every vector $x \in \mathbb{C}^N$ with $\|x\|_1 = 1$ there exists a t-tuple of vectors $(h^{(1)}, \ldots, h^{(t)}) \in \mathcal{H}_1 \times \cdots \times \mathcal{H}_t$ for which*

1. $\mathbb{E}_{j\in Q}\left[|(Mx)_j|^2\right] \approx_{O(\epsilon),O(\eta)} \mathbb{E}_{j\in Q}\left[\sum_{i=1}^t h_j^{(i)}\right]$ *and*
2. $\mathbb{E}_{j\in [N]}\left[|(Mx)_j|^2\right] \approx_{O(\epsilon),O(\eta)} \mathbb{E}_{j\in [N]}\left[\sum_{i=1}^t h_j^{(i)}\right]$.

Proof By Lemma 3.5, for every $1 \le i \le t$ there exists a vector $g^{(i+r)} \in \mathcal{G}_{i+r}$ that satisfies

$$|(Mx)_j| \approx_{0,2^{-(i+r)/2}} |g_j^{(i+r)}| \tag{5}$$

for all but at most γ fraction of $j \in [N]$ and for all but at most γ fraction of $j \in Q$. We say that $j \in [N]$ is *good* if (5) holds for every $1 \le i \le t$, and otherwise that it is *bad*. Notice that all but at most $t\gamma$ fraction of $j \in [N]$ are good and that all but at most $t\gamma$ fraction of $j \in Q$ are good. Let $(h^{(1)}, \ldots, h^{(t)})$ and $(B_1, \ldots, B_t)$ be the vectors and sets associated with $(g^{(1+r)}, \ldots, g^{(t+r)})$ as defined in (3). We claim that $h^{(1)}, \ldots, h^{(t)}$ satisfy the requirements of the lemma.

We first show that for every good j it holds that $|(Mx)_j|^2 \approx_{3\epsilon,9\eta} \sum_{i=1}^t h_j^{(i)}$. To obtain it, we observe that if $j \in B_i$ for some i, then

$$2 \cdot 2^{-i/2} \le |g_j^{(i+r)}| \le 3 \cdot 2^{-i/2}. \tag{6}$$

The lower bound follows simply from the definition of B_i. For the upper bound, which trivially holds for $i = 1$, assume that $i \ge 2$, and notice that the definition of B_i implies that $|g_j^{(i+r-1)}| < 2 \cdot 2^{-(i-1)/2}$. Using (5), and assuming that ϵ is sufficiently small, we obtain that

$$\begin{aligned} |g_j^{(i+r)}| &\le |(Mx)_j| + 2^{-(i+r)/2} \le |g_j^{(i+r-1)}| + 2^{-(i+r-1)/2} + 2^{-(i+r)/2} \\ &\le 2^{-i/2}(2^{3/2} + 2^{1/2} \cdot \epsilon + \epsilon) \le 3 \cdot 2^{-i/2}. \end{aligned}$$

Hence, by the upper bound in (6), for a good $j \in B_i$ we have $h_j^{(i)} = |g_j^{(i+r)}|^2$ and $h_j^{(i')} = 0$ for $i' \ne i$. Observe that by the lower bound in (6),

$$|(Mx)_j| \in [|g_j^{(i+r)}| - 2^{-(i+r)/2}, |g_j^{(i+r)}| + 2^{-(i+r)/2}] \subseteq [(1-\epsilon)\cdot|g_j^{(i+r)}|, (1+\epsilon)\cdot|g_j^{(i+r)}|],$$

and that this implies that $|(Mx)_j|^2 \approx_{3\epsilon,0} \sum_{i=1}^{t} h_j^{(i)}$. On the other hand, in case that j is good but does not belong to any B_i, recalling that $t = \log_2(1/\eta)$, it follows that

$$|(Mx)_j| \leq |g_j^{(t+r)}| + 2^{-(t+r)/2} \leq 2 \cdot 2^{-t/2} + 2^{-(t+r)/2} \leq 3 \cdot 2^{-t/2} \leq 3\sqrt{\eta},$$

and thus $|(Mx)_j|^2 \approx_{0,9\eta} 0 = \sum_{i=1}^{t} h_j^{(i)}$.

Finally, for every bad j we have

$$\left| |(Mx)_j|^2 - \sum_{i=1}^{t} h_j^{(i)} \right| \leq \max\left(|(Mx)_j|^2, \sum_{i=1}^{t} h_j^{(i)} \right) \leq 2.$$

Since at most $t\gamma$ fraction of the elements in $[N]$ and in Q are bad, their effect on the difference between the expectations in the lemma can be bounded by $2t\gamma$. By our choice of γ, this is η, completing the proof of the lemma. ■

Finally, we are ready to prove Theorem 3.1.

Proof of Theorem 3.1 By Lemma 3.4, applied with $\tilde{\eta} = \eta/(2t)$, a random multiset Q of size

$$q = O\left(\epsilon^{-3}\eta^{-1} \cdot t \cdot \log N \cdot \log(1/\gamma)\right) = O\left(\epsilon^{-3}\eta^{-1} \log N \cdot \log^2(1/\eta)\right)$$

satisfies with probability $1 - 2^{-\Omega(\epsilon^{-2} \cdot \log N \cdot \log(1/\eta))}$ that for all $1 \leq i \leq t$ and $h^{(i)} \in \mathcal{H}_i$,

$$\mathop{\mathbb{E}}_{j \in Q}\left[h_j^{(i)}\right] \approx_{\epsilon,\eta/t} \mathop{\mathbb{E}}_{j \in [N]}\left[h_j^{(i)}\right],$$

in which case we also have

$$\mathop{\mathbb{E}}_{j \in Q}\left[\sum_{i=1}^{t} h_j^{(i)}\right] \approx_{\epsilon,\eta} \mathop{\mathbb{E}}_{j \in [N]}\left[\sum_{i=1}^{t} h_j^{(i)}\right].$$

We show that a Q with the above property satisfies the requirement of the theorem. Let $x \in \mathbb{C}^N$ be a vector, and assume without loss of generality that $\|x\|_1 = 1$. By Lemma 3.6, there exists a t-tuple of vectors $(h^{(1)}, \ldots, h^{(t)}) \in \mathcal{H}_1 \times \cdots \times \mathcal{H}_t$ satisfying Items 1 and 2 there. As a result,

$$\mathop{\mathbb{E}}_{j \in Q}\left[|(Mx)_j|^2\right] \approx_{O(\epsilon),O(\eta)} \mathop{\mathbb{E}}_{j \in [N]}\left[|(Mx)_j|^2\right],$$

and we are done. ■

3.1 The Restricted Isometry Property

Equipped with Theorem 3.1, it is easy to derive our result on the restricted isometry property (see Definition 2.1) of random sub-matrices of unitary matrices.

Theorem 3.7 *For sufficiently large N and k, a unitary matrix $M \in \mathbb{C}^{N\times N}$ satisfying $\|M\|_\infty \le O(1/\sqrt{N})$, and a sufficiently small $\epsilon > 0$, the following holds. For some $q = O(\epsilon^{-4}\cdot k\cdot \log^2(k/\epsilon)\cdot \log N)$, let $A \in \mathbb{C}^{q\times N}$ be a matrix whose q rows are chosen uniformly and independently from the rows of M, multiplied by $\sqrt{N/q}$. Then, with probability $1 - 2^{-\Omega(\epsilon^{-2}\cdot \log N\cdot \log(k/\epsilon))}$, the matrix A satisfies the restricted isometry property of order k with constant ϵ.*

Proof Let Q be a multiset of q uniform and independent random elements of $[N]$, defining a matrix A as above. Notice that by the Cauchy-Schwarz inequality, any k-sparse vector $x \in \mathbb{C}^N$ with $\|x\|_2 = 1$ satisfies $\|x\|_1 \le \sqrt{k}$. Applying Theorem 3.1 with $\epsilon/2$ and some $\eta = \Omega(\epsilon/k)$, we get that with probability $1-2^{-\Omega(\epsilon^{-2}\cdot \log N\cdot \log(k/\epsilon))}$, it holds that for every $x \in \mathbb{C}^N$ with $\|x\|_2 = 1$,

$$\|Ax\|_2^2 = N\cdot \mathop{\mathbb{E}}_{j\in Q}\left[|(Mx)_j|^2\right] \approx_{\epsilon/2,\epsilon/2} N\cdot \mathop{\mathbb{E}}_{j\in[N]}\left[|(Mx)_j|^2\right] = \|Mx\|_2^2 = 1\ .$$

It follows that every vector $x \in \mathbb{C}^N$ satisfies $\|Ax\|_2^2 \approx_{\epsilon,0} \|x\|_2^2$, hence A satisfies the restricted isometry property of order k with constant ϵ. ■

4 The Improved Analysis

In this section we prove the following theorem, which improves the bound of Theorem 3.1 in terms of the dependence on ϵ.

Theorem 4.1 *For a sufficiently large N, a matrix $M \in \mathbb{C}^{N\times N}$, and sufficiently small $\epsilon, \eta > 0$, the following holds. For some $q = O(\log^2(1/\epsilon)\cdot\epsilon^{-1}\eta^{-1}\log N\cdot\log^2(1/\eta))$, let Q be a multiset of q uniform and independent random elements of $[N]$. Then, with probability $1 - 2^{-\Omega(\log N\cdot\log(1/\eta))}$, it holds that for every $x \in \mathbb{C}^N$,*

$$\mathop{\mathbb{E}}_{j\in Q}\left[|(Mx)_j|^2\right] \approx_{\epsilon,\eta\cdot\|x\|_1^2\cdot\|M\|_\infty^2} \mathop{\mathbb{E}}_{j\in[N]}\left[|(Mx)_j|^2\right]. \tag{7}$$

We can assume that $\epsilon \ge \eta$, as otherwise, one can apply the theorem with parameters $\eta/2, \eta/2$ and derive (7) for ϵ, η as well (because the right-hand size is bounded from above by $\|x\|_1^2\cdot\|M\|_\infty^2$). As before, we assume without loss of generality that $\|M\|_\infty = 1$. For $\epsilon \ge \eta > 0$, we define $t = \log_2(1/\eta)$ and $r = \log_2(1/\epsilon^2)$. For the analysis given in this section, we define $\gamma = \eta/(60(t+r))$. Throughout the proof, we use the vector sets $\mathcal{G}_i$ from Sect. 3 and Lemma 3.5 for this value of γ.

The Vector Sets $\mathcal{D}_{i,m}$ For a $(t+r)$-tuple of vectors $(g^{(1)},\ldots,g^{(t+r)}) \in \mathcal{G}_1 \times \cdots \times \mathcal{G}_{t+r}$ and for $1 \le i \le t$, let C_i be the set of all $j \in [N]$ for which i is the smallest index satisfying $|g_j^{(i)}| \ge 2 \cdot 2^{-i/2}$. For $m = i,\ldots,i+r$ define the vector $h^{(i,m)}$ by

$$h_j^{(i,m)} = |g_j^{(m)}|^2 \cdot \mathbf{1}_{j \in C_i}, \tag{8}$$

and for other values of m define $h^{(i,m)} = 0$. Now, for every m, let $\Delta^{(i,m)}$ be the vector defined by

$$\Delta_j^{(i,m)} = \begin{cases} h_j^{(i,m)} - h_j^{(i,m-1)}, & \text{if } |h_j^{(i,m)} - h_j^{(i,m-1)}| \le 30 \cdot 2^{-(i+m)/2}; \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

Note that the support of $\Delta^{(i,m)}$ is contained in C_i. Let $\mathcal{D}_{i,m}$ be the set of all vectors $\Delta^{(i,m)}$ that can be obtained in this way.

Claim 4.2 *For every $1 \le i \le t$ and $i \le m \le i+r$, $|\mathcal{D}_{i,m}| \le N^{O(2^m \cdot \log(1/\gamma))}$.*

Proof Observe that every vector in $\mathcal{D}_{i,m}$ is fully defined by some $(g^{(1)},\ldots,g^{(m)}) \in \mathcal{G}_1 \times \cdots \times \mathcal{G}_m$. Hence

$$|\mathcal{D}_{i,m}| \le |\mathcal{G}_1| \cdots |\mathcal{G}_m| \le N^{O(\log(1/\gamma)) \cdot (2^1 + 2^2 + \cdots + 2^m)} \le N^{O(\log(1/\gamma)) \cdot 2^{m+1}},$$

and the claim follows. ∎

Lemma 4.3 *For every $\tilde{\varepsilon}, \tilde{\eta} > 0$ and some $q = O(\tilde{\varepsilon}^{-1}\tilde{\eta}^{-1} \log N \cdot \log(1/\gamma))$, let Q be a multiset of q uniform and independent random elements of $[N]$. Then, with probability $1 - 2^{-\Omega(\log N \cdot \log(1/\gamma))}$, it holds that for every $1 \le i \le t$, m, and a vector $\Delta^{(i,m)} \in \mathcal{D}_{i,m}$ associated with a set C_i,*

$$\mathop{\mathbb{E}}_{j \in Q}\left[\Delta_j^{(i,m)}\right] \approx_{0,b} \mathop{\mathbb{E}}_{j \in [N]}\left[\Delta_j^{(i,m)}\right] \quad \text{for} \quad b = O\left(\tilde{\varepsilon} \cdot 2^{-i} \cdot \frac{|C_i|}{N} + \tilde{\eta}\right). \tag{10}$$

Proof Fix i, m, and a vector $\Delta^{(i,m)} \in \mathcal{D}_{i,m}$ associated with a set C_i as in (9). Notice that

$$\mathop{\mathbb{E}}_{j \in [N]}[|\Delta_j^{(i,m)}|] \le 30 \cdot 2^{-(i+m)/2} \cdot \frac{|C_i|}{N}.$$

By Corollary 2.4, applied with

$$\epsilon' = \tilde{\varepsilon} \cdot 2^{(m-i)/2}, \quad \alpha = \tilde{\eta}, \quad \text{and} \quad a = 30 \cdot 2^{-(i+m)/2},$$

we have that (10) holds with probability $1 - 2^{-\Omega(2^m \cdot q\tilde{\varepsilon}\tilde{\eta})}$. Using Claim 4.2, the union bound over all the vectors in $\mathcal{D}_{i,m}$ implies that the probability that some $\Delta^{(i,m)} \in \mathcal{D}_{i,m}$

does not satisfy (10) is at most

$$N^{O(2^m \cdot \log(1/\gamma))} \cdot 2^{-\Omega(2^m \cdot q\widetilde{\varepsilon}\widetilde{\eta})} \leq 2^{-\Omega(2^m \cdot \log N \cdot \log(1/\gamma))} .$$

The result follows by a union bound over i and m. ■

Approximating the Vectors Mx

Lemma 4.4 *For every multiset $Q \subseteq [N]$ and every vector $x \in \mathbb{C}^N$ with $\|x\|_1 = 1$ there exist vector collections $(\Delta^{(i,m)} \in \mathcal{D}_{i,m})_{m=i,\ldots,i+r}$ associated with sets C_i $(1 \leq i \leq t)$, for which*

1. $\mathbb{E}_{j\in[N]}\left[|(Mx)_j|^2\right] \geq \sum_{i=1}^{t} 2^{-i} \cdot \frac{|C_i|}{N} - \eta$,
2. $\mathbb{E}_{j\in Q}\left[|(Mx)_j|^2\right] \approx_{O(\epsilon),O(\eta)} \mathbb{E}_{j\in Q}\left[\sum_{i=1}^{t}\sum_{m=i}^{i+r} \Delta_j^{(i,m)}\right]$, *and*
3. $\mathbb{E}_{j\in[N]}\left[|(Mx)_j|^2\right] \approx_{O(\epsilon),O(\eta)} \mathbb{E}_{j\in[N]}\left[\sum_{i=1}^{t}\sum_{m=i}^{i+r} \Delta_j^{(i,m)}\right]$.

Proof By Lemma 3.5, for every $1 \leq i \leq t + r$ there exists a vector $g^{(i)} \in \mathcal{G}_i$ that satisfies

$$|(Mx)_j| \approx_{0,2^{-i/2}} |g_j^{(i)}| \tag{11}$$

for all but at most γ fraction of $j \in [N]$ and for all but at most γ fraction of $j \in Q$. We say that $j \in [N]$ is *good* if (11) holds for every i, and otherwise that it is *bad*. Notice that all but at most $(t + r)\gamma$ fraction of $j \in [N]$ are good and that all but at most $(t + r)\gamma$ fraction of $j \in Q$ are good. Consider the sets C_i and vectors $h^{(i,m)}, \Delta^{(i,m)}$ associated with $(g^{(1)}, \ldots, g^{(t+r)})$ as defined in (8). We claim that $\Delta^{(i,m)}$ satisfy the requirements of the lemma.

Fix some $1 \leq i \leq t$. For every good $j \in C_i$, the definition of C_i implies that $|g_j^{(i)}| \geq 2 \cdot 2^{-i/2}$, so using (11) it follows that

$$|(Mx)_j| \geq |g_j^{(i)}| - 2^{-i/2} \geq 2^{-i/2}. \tag{12}$$

We also claim that $|(Mx)_j| \leq 3 \cdot 2^{-(i-1)/2}$. This trivially holds for $i = 1$, so assume that $i \geq 2$, and notice that the definition of C_i implies that $|g_j^{(i-1)}| < 2 \cdot 2^{-(i-1)/2}$, so using (11), it follows that

$$|(Mx)_j| \leq |g_j^{(i-1)}| + 2^{-(i-1)/2} \leq 3 \cdot 2^{-(i-1)/2}. \tag{13}$$

Since at most $(t + r)\gamma$ fraction of $j \in [N]$ are bad, (12) yields that

$$\mathop{\mathbb{E}}_{j\in[N]}\left[|(Mx)_j|^2\right] \geq \sum_{i=1}^{t} 2^{-i} \cdot \frac{|C_i|}{N} - (t + r)\gamma/2 \geq \sum_{i=1}^{t} 2^{-i} \cdot \frac{|C_i|}{N} - \eta,$$

as required for Item 1.

Next, we claim that every good j satisfies

$$|(Mx)_j|^2 \approx_{O(\epsilon),O(\eta)} \sum_{i=1}^{t} h_j^{(i,i+r)} \,. \tag{14}$$

For a good $j \in C_i$ and $m \geq i$,

$$\left| |(Mx)_j|^2 - h_j^{(i,m)} \right| \leq 2 \cdot |(Mx)_j| \cdot 2^{-m/2} + 2^{-m} \leq 10 \cdot 2^{-(i+m)/2}, \tag{15}$$

where the first inequality follows from (11) and the second from (13). In particular, for $m = i + r$ (recall that $r = \log_2(1/\epsilon^2)$), we have

$$\left| |(Mx)_j|^2 - h_j^{(i,i+r)} \right| \leq 10 \cdot \epsilon \cdot 2^{-i} \leq 10 \cdot \epsilon \cdot |(Mx)_j|^2 \,,$$

and thus $|(Mx)_j|^2 \approx_{O(\epsilon),0} h_j^{(i,i+r)}$. Since every good j belongs to at most one of the sets C_i, for every good $j \in \bigcup C_i$ we have $|(Mx)_j|^2 \approx_{O(\epsilon),0} \sum_{i=1}^{t} h_j^{(i,i+r)}$. On the other hand, if j is good but does not belong to any C_i, by our choice of t, it satisfies

$$|(Mx)_j| \leq |g_j^{(t)}| + 2^{-t/2} \leq 3 \cdot 2^{-t/2} = 3\sqrt{\eta} \,,$$

and thus $|(Mx)_j|^2 \approx_{0,9\eta} 0 = \sum_{i=1}^{t} h_j^{(i,i+r)}$. This establishes that (14) holds for every good j.

Next, we claim that for every good j,

$$|(Mx)_j|^2 \approx_{O(\epsilon),O(\eta)} \sum_{i=1}^{t} \sum_{m=i}^{i+r} \Delta_j^{(i,m)} \,. \tag{16}$$

This follows since for every $1 \leq i \leq t$, the vector $h^{(i,i+r)}$ can be written as the telescopic sum

$$h^{(i,i+r)} = \sum_{m=i}^{i+r} (h^{(i,m)} - h^{(i,m-1)}) \,,$$

where we used that $h^{(i,i-1)} = 0$. We claim that for every good j, these differences satisfy

$$|h_j^{(i,m)} - h_j^{(i,m-1)}| \leq 30 \cdot 2^{-(i+m)/2},$$

thus establishing that (16) holds for every good j. Indeed, for $m \geq i+1$, (15) implies that

$$|h_j^{(i,m)} - h_j^{(i,m-1)}| \leq 10 \cdot (2^{-(i+m)/2} + 2^{-(i+m-1)/2}) \leq 30 \cdot 2^{-(i+m)/2}, \tag{17}$$

and for $m = i$ it follows from (11) combined with (13).

Finally, for every bad j we have

$$\left||(Mx)_j|^2 - \sum_{i=1}^{t}\sum_{m=i}^{i+r}\Delta_j^{(i,m)}\right| \le 1 + 30\cdot\max_{1\le i\le t}\left(\sum_{m=i}^{i+r}2^{-(i+m)/2}\right) \le 60\,.$$

Since at most $(t+r)\gamma$ fraction of the elements in $[N]$ and in Q are bad, their effect on the difference between the expectations in Items 2 and 3 can be bounded by $60(t+r)\gamma$. By our choice of γ this is η, as required. ■

Finally, we are ready to prove Theorem 4.1.

Proof of Theorem 4.1 Recall that it can be assumed that $\epsilon \ge \eta$. By Lemma 4.3, applied with $\tilde{\varepsilon} = \epsilon/r$ and $\tilde{\eta} = \eta/(rt)$, a random multiset Q of size

$$\begin{aligned} q &= O\Big(\epsilon^{-1}\eta^{-1}\cdot r^2\cdot t\cdot\log N\cdot\log(1/\gamma)\Big)\\ &= O\Big(\log^2(1/\epsilon)\cdot\epsilon^{-1}\eta^{-1}\log N\cdot\log^2(1/\eta)\Big)\end{aligned}$$

satisfies with probability $1 - 2^{-\Omega(\log N\cdot\log(1/\eta))}$, that for every $1 \le i \le t$, m, and $\Delta^{(i,m)} \in \mathcal{D}_{i,m}$ associated with a set C_i,

$$\mathop{\mathbb{E}}_{j\in Q}\left[\Delta_j^{(i,m)}\right] \approx_{0,b_i} \mathop{\mathbb{E}}_{j\in[N]}\left[\Delta_j^{(i,m)}\right] \quad \text{for} \quad b_i = O\Big(\frac{\varepsilon}{r}\cdot 2^{-i}\cdot\frac{|C_i|}{N} + \frac{\eta}{rt}\Big),$$

in which case we also have

$$\mathop{\mathbb{E}}_{j\in Q}\left[\sum_{i=1}^{t}\sum_{m=i}^{i+r}\Delta_j^{(i,m)}\right] \approx_{0,b} \mathop{\mathbb{E}}_{j\in[N]}\left[\sum_{i=1}^{t}\sum_{m=i}^{i+r}\Delta_j^{(i,m)}\right] \quad \text{for} \quad b = O\Big(\epsilon\cdot\sum_{i=1}^{t}2^{-i}\cdot\frac{|C_i|}{N} + \eta\Big)\,. \tag{18}$$

We show that a Q with the above property satisfies the requirement of the theorem. Let $x \in \mathbb{C}^N$ be a vector, and assume without loss of generality that $\|x\|_1 = 1$. By Lemma 4.4, there exist vector collections $(\Delta^{(i,m)} \in \mathcal{D}_{i,m})_{m=i,\ldots,i+r}$ associated with sets C_i $(1 \le i \le t)$, satisfying Items 1, 2, and 3 there. Combined with (18), this gives

$$\mathop{\mathbb{E}}_{j\in Q}\left[|(Mx)_j|^2\right] \approx_{O(\epsilon),O(\eta)} \mathop{\mathbb{E}}_{j\in[N]}\left[|(Mx)_j|^2\right]\,,$$

and we are done. ■

4.1 The Restricted Isometry Property

It is easy to derive now the following theorem. The proof is essentially identical to that of Theorem 3.7, using Theorem 4.1 instead of Theorem 3.1.

Theorem 4.5 *For sufficiently large N and k, a unitary matrix $M \in \mathbb{C}^{N\times N}$ satisfying $\|M\|_\infty \leq O(1/\sqrt{N})$, and a sufficiently small $\epsilon > 0$, the following holds. For some $q = O(\log^2(1/\epsilon)\epsilon^{-2}\cdot k\cdot\log^2(k/\epsilon)\cdot\log N)$, let $A \in \mathbb{C}^{q\times N}$ be a matrix whose q rows are chosen uniformly and independently from the rows of M, multiplied by $\sqrt{N/q}$. Then, with probability $1 - 2^{-\Omega(\log N\cdot\log(k/\epsilon))}$, the matrix A satisfies the restricted isometry property of order k with constant ϵ.*

Acknowledgements We thank Afonso S. Bandeira, Mahdi Cheraghchi, Michael Kapralov, Jelani Nelson, and Eric Price for useful discussions, and anonymous reviewers for useful comments.

Oded Regev was supported by the Simons Collaboration on Algorithms and Geometry and by the National Science Foundation (NSF) under Grant No. CCF-1320188. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

1. N. Ailon, B. Chazelle, The fast Johnson–Lindenstrauss transform and approximate nearest neighbors. SIAM J. Comput. **39**(1), 302–322 (2009). Preliminary version in STOC'06
2. N. Ailon, E. Liberty, Fast dimension reduction using Rademacher series on dual BCH codes. Discrete Comput. Geom. **42**(4), 615–630 (2009). Preliminary version in SODA'08
3. N. Ailon, E. Liberty, An almost optimal unrestricted fast Johnson–Lindenstrauss transform. ACM Trans. Algorithms **9**(3), 21 (2013). Preliminary version in SODA'11
4. A.S. Bandeira, E. Dobriban, D.G. Mixon, W.F. Sawin, Certifying the restricted isometry property is hard. IEEE Trans. Inform. Theory **59**(6), 3448–3450 (2013)
5. A.S. Bandeira, M.E. Lewis, D.G. Mixon, Discrete uncertainty principles and sparse signal processing. CoRR abs/1504.01014 (2015)
6. R. Baraniuk, M. Davenport, R. DeVore, M. Wakin, A simple proof of the restricted isometry property for random matrices. Constr. Approx. **28**(3), 253–263 (2008)
7. J. Bourgain, An improved estimate in the restricted isometry problem, in *Geometric Aspects of Functional Analysis*. Lecture Notes in Mathematics, vol. 2116, pp. 65–70 (Springer, Berlin, 2014)
8. E.J. Candès, The restricted isometry property and its implications for compressed sensing. C. R. Math. **346**(9–10), 589–592 (2008)
9. E.J. Candès, T. Tao, Decoding by linear programming. IEEE Trans. Inform. Theory **51**(12), 4203–4215 (2005)
10. E.J. Candès, T. Tao, Near-optimal signal recovery from random projections: universal encoding strategies? IEEE Trans. on Inform. Theory **52**(12), 5406–5425 (2006)
11. E.J. Candès, M. Rudelson, T. Tao, R. Vershynin, Error correction via linear programming, in *46th Annual IEEE Symposium on Foundations of Computer Science, FOCS*, pp. 295–308 (2005)
12. E.J. Candès, J.K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements. Commun. Pure Appl. Math. **59**(8), 1207–1223 (2006)
13. S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit. SIAM J. Comput. **20**(1), 33–61 (1998)
14. M. Cheraghchi, V. Guruswami, A. Velingker, Restricted isometry of Fourier matrices and list decodability of random linear codes. SIAM J. Comput. **42**(5), 1888–1914 (2013). Preliminary version in SODA'13
15. S. Dirksen, Tail bounds via generic chaining. Electron. J. Prob. **20**(53), 1–29 (2015)

16. D.L. Donoho, M. Elad, V.N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Trans. Inform. Theory **52**(1), 6–18 (2006)
17. S. Foucart, A. Pajor, H. Rauhut, T. Ullrich, The Gelfand widths of ℓ_p-balls for $0 < p \leq 1$. J. Complex. **26**(6), 629–640 (2010)
18. A.Y. Garnaev, E.D. Gluskin, On the widths of Euclidean balls. Sov. Math. Dokl. **30**, 200–203 (1984)
19. I. Haviv, O. Regev, The list-decoding size of Fourier-sparse boolean functions, in *Proceedings of the 30th Conference on Computational Complexity, CCC*, pp. 58–71 (2015)
20. P. Indyk, I. Razenshteyn, On model-based RIP-1 matrices, in *Automata, Languages, and Programming - 40th International Colloquium, ICALP*, pp. 564–575 (2013)
21. A.C. Kak, M. Slaney, *Principles of Computerized Tomographic Imaging* (Society of Industrial and Applied Mathematics, Philadelphia, 2001)
22. F. Krahmer, R. Ward, New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property. SIAM J. Math. Anal. **43**(3), 1269–1281 (2011)
23. F. Krahmer, S. Mendelson, H. Rauhut, Suprema of chaos processes and the restricted isometry property. CoRR abs/1207.0235 (2012)
24. C. McDiarmid, Concentration, in *Probabilistic Methods for Algorithmic Discrete Mathematics*. Algorithms Combination, vol. 16 (Springer, Berlin, 1998), pp. 195–248
25. A. Natarajan, Y. Wu, Computational complexity of certifying restricted isometry property, in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX*, pp. 371–380 (2014)
26. D. Needell, J.A. Tropp, CoSaMP: iterative signal recovery from incomplete and inaccurate samples. Commun. ACM **53**(12), 93–100 (2010)
27. J. Nelson, E. Price, M. Wootters, New constructions of RIP matrices with fast multiplication and fewer rows, in *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, pp. 1515–1528 (2014)
28. D.G. Nishimura, *Principles of Magnetic Resonance Imaging* (Stanford University, Stanford, CA, 2010)
29. H. Rauhut, Compressive sensing and structured random matrices, in *Theoretical Foundations and Numerical Methods for Sparse Recovery*, vol. 9, ed. by M. Fornasier (De Gruyter, Berlin, 2010), pp. 1–92
30. M. Rudelson, R. Vershynin, On sparse reconstruction from Fourier and Gaussian measurements. Commun. Pure Appl. Math. **61**(8), 1025–1045 (2008). Preliminary version in CISS'06
31. A.M. Tillmann, M.E. Pfetsch, The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing. IEEE Trans. Inform. Theory **60**(2), 1248–1259 (2014)
32. S. Wenger, S. Darabi, P. Sen, K. Glassmeier, M.A. Magnor, Compressed sensing for aperture synthesis imaging, in *Proceedings of the International Conference on Image Processing, ICIP*, pp. 1381–1384 (2010)
33. M. Wootters, On the list decodability of random linear codes with large error rates, in *Proceedings of the 45th Annual ACM Symposium on Theory of Computing, STOC*, pp. 853–860 (2013)