UNCERTAINTY PRINCIPLES AND VECTOR QUANTIZATION

YURII LYUBARSKII AND ROMAN VERSHYNIN

Abstract. An abstract form of the Uncertainty Principle set forth by Candes and Tao has found remarkable applications in the sparse approximation theory. This paper demonstates a new connection between the Uncertainty Principle and the vector quantization theory. We show that for frames in $\mathbb{C}^n$ that satisfy the Uncertainty Principle, one can quickly convert every frame representation into a more regular Kashin's representation, whose coefficients all have the same magnitude $O(1/\sqrt{n})$. Information tends to spread evenly among these coefficients. As a consequence, Kashin's representations have a great power for reduction of errors in their coefficients. In particular, scalar quantization of Kashin's representations yields robust vector quantizers in $\mathbb{C}^n$.

1. Introduction

Quantization is a representation of continuous structures with discrete structures. In the practice of computing, quantization allows us to work with real numbers in terms of finite arithmetics. Digital signal processing, which has revolutionized the modern treatment of still images, video and audio, employs quantization as a conversion step from the analog to the digital world. A survey of the state-of-the-art of quantization prior to 1998 as well as outline of its numerous aplications can be found the the paper [24] by Gray and Neuhoff. For more recent developments, we refer the reader to [15] and the references therein.

In this paper, we are interested in robust *vector quantization* procedures. As a typical example, consider the problem of analog-to-digital conversion of signals. We look at signals as vectors in a Hilbert space H, such as $L_2(\mathbb{R})$. To convert a signal $f \in H$ into a digital form, one first represents f by a finite-dimensional vector $x = (a_1, \dots, a_n) \in \mathbb{C}^n$ using, for example, some n coefficients of f with respect to an orthogonal basis (see [41] for a survey of approximation methods). Then one performs scalar quantization of each coefficient a_i, say with accuracy ε. The quantized vector $\hat{x} = (\hat{a}_1, \dots, \hat{a}_n)$, whose coefficients satisfy $|a_i - \hat{a}_i| \sim \varepsilon$, serves as a digital representation of the signal f. The accuracy of such analog-to-digital conversion of the signal f is then determined by the two errors – the error of the finite dimensional approximation of f and the quantization error. By Parseval's identity, the latter is bounded by

$$\|x - \hat{x}\|_2 = \Big(\sum_{i=1}^n |a_i - \hat{a}_i|^2 \Big)^{1/2} \sim \varepsilon \sqrt{n}. \tag{1.1}$$

The bound (1.1) on the quantization error grows with the dimension n, so the quantization scheme suffers from a tradeoff between approximation and quantization. Improving the quality of a finite-dimensional approximation of f forces one to increase the dimension n, but this in turn increases the quantization error.

R.V. was partially supported by NSF grant DMS 0401032 and Alfred P. Sloan Foundation, Yu.L. was partially supported by the Norwegian Research Council project 160192/v30.

In this paper, we propose a quantization procedure for vectors $a \in \mathbb{C}^n$, which reduces the quantization error down to $O(\varepsilon)$ uniformly in all dimensions n. The quantization error will no longer accumulate with the dimension, thus eliminating the tradeoff between approximation and quantization.

Our vector quantization procedure is based on representations in $\mathbb{C}^n$ with respect to redundant systems of vectors (frames). Frame representations are commonly used in signal processing for reduction of errors caused by the noise in the coefficients (in particular quantization errors), see e.g. [11, 23, 10] and references therein. Since the information about f gets spread over several frame coefficients, the errors tend to cancel when the representation is summed up. As a result, increasing the redundancy of the frame gradually reduces the representation error, see [11, 23, 10].

We shall show that already a constant redundancy (actually, any redundancy factor bigger than 1) can reduce the error to the minimal possible order $O(\varepsilon)$. We do this by connecting the vector quantization problem to the Uncertainty Principle in harmonic analysis. An abstract form of the Uncertainty Principle for matrices was recently introduced in [7]. It has been already successfully used in the growing area of compressed sensing and found many applications, see [9].

In this paper, we show that for frames in $\mathbb{C}^n$ that satisfy the Uncertainty Principle, one can quickly convert every frame representation into a new *Kashin's representation* whose coefficients are of the same order $O(1/\sqrt{n})$. The information tends to spread nearly evenly among the coefficients of Kashin's representations.

As a consequence, Kashin's representations withstand errors in their coefficients in a very strong way: the representation error gets bounded by the *average*, rather than the sum, of the errors in the coefficients. These errors may be of arbitrary nature, including scalar quantization, losses or flipping the bits. If one uses Kashin's representations of vectors $a \in \mathbb{C}^n$ instead of the orthonormal basis or frame representations, then the quantization error in (1.1) gets reduced to the minimal possible order $O(\varepsilon)$.

The article is organized as follows. Section 2 introduces Kashin's representations, discusses their relation to convex geometry (Euclidean projections of the cube) and explains how to use Kashin's representations for vector quantization. In Section 3, we discuss the Uncertainty Principle for matrices and frames. Theorems 3.5 and 3.9 state that for frames that satisfy Uncertainty Principle, every frame representation can be replaced with a Kashin's representation. A robust algorithm is given to quickly convert frame into Kashin's representations. In Section 4, we discuss families of matrices and frames that satisfy the Uncertainty Principle. These include: random orthogonal matrices, random partial Fourier matrices, and a large family of matrices with independent entries (subgaussian matrices), in particular random Gaussian and Bernoulli matrices.

Acknowledgements. This project started in 2003 when the second author visited University of California, Davis. The second author gave a talk in March 2006 at the Time-Frequency Brown Bag Seminar at Princeton. He sincerely thanks the participants for their valuable comments, and especially Ingrid Daubechies for her encouragement and interest in this work.

2. Kashin's representations

2.1. **Frame representations.** A sequence $(u_i)_{i=1}^N \subset \mathbb{C}^n$ is called a tight frame if

$$\|x\|_2^2 = \sum_{i=1}^N |\langle x, u_i \rangle|^2 \qquad \text{for all } x \in \mathbb{C}^n. \tag{2.1}$$

This definition differs by a constant normalization factor from the one often used in the literature, but (2.1) will be more convenient for us to work with.

A frame $(u_i)_{i=1}^N \subset \mathbb{C}^n$ can be identified with its *frame matrix* matrix U, the $n \times N$ matrix whose columns are u_i. The following are easily seen to be equivalent:

(1) $(u_i)_{i=1}^N$ is a tight frame in $\mathbb{C}^n$;
(2) every vector $x \in \mathbb{C}^n$ has the frame representation

$$x = \sum_{i=1}^N b_i u_i, \quad \text{where} \quad b_i = \langle x, u_i \rangle; \tag{2.2}$$

(3) the rows of the frame matrix U are orthonormal.

When $N > n$, the tight frames are linearly dependent systems, so various frame coefficients b_i can carry the same information about the vector x. This makes frames withstand noise in coefficients better than orthonormal bases, see [11, 23, 10]. However, using the frame representation (2.2) may not always be the best way to use the redundancy of the frame. Some coefficients b_i may be much bigger than the others, and thus carry more information about x. In order to help information spread in the most uniform way, one should try to make all coefficients of the same magnitude. Such representations will be called Kashin's representations.

2.2. **Kashin's representations.** Consider a sequence $(u_i)_{i=1}^N \subset \mathbb{C}^n$. We shall say that the expansion

$$x = \frac{1}{\sqrt{N}} \sum_{i=1}^N a_i u_i \tag{2.3}$$

is a *Kashin's representation with level* K of a vector $x \in \mathbb{C}^n$ if

$$|a_i| \le K\|x\|_2, \quad i = 1, 2, \dots, N.$$

To explain the factor $1/\sqrt{N}$ in (2.3) note that if all the frame coefficients b_i of a unit vector x are of the same magnitude, then this magnitude can be easily seen to be equal $1/\sqrt{N}$.

One can not always replace tight frame representations with Kashin's representations (with constant level K). However, there are natural classes of frames of constant redundancy $\lambda = N/n$ for which this is possible:

Theorem 2.1 (Existence of Kashin's representations)**.** *There exist tight frames with arbitrarily small redundancy $\lambda = N/n > 1$, and such that every vector in $\mathbb{C}^n$ has a Kashin's representation with level K that depends on λ only (not on n or N).*

This is a reformulation of the classical result of geometric functional analysis due to Kashin [30] (with an optimal dependence on λ given later by Garnaev and Gluskin [17]). To see this, we shall look at Kashin's representations from a geometric viewpoint. Let $Q^N = \{x : \|x\|_\infty \le 1\}$ and $B^n = \{x : \|x\|_2 \le 1\}$ stand for the unit cube and unit Euclidean ball in the spaces $\mathbb{C}^N$ and $\mathbb{C}^n$ respectively. The following observation is straightforward:

Observation 2.2 (Kashin's representations realize Eucliedan projections of the cube)**.** *Consider a sequence $(u_i)_{i=1}^N$ in $\mathbb{C}^n$, and a number $K > 0$. The following are equivalent:*

(i) Every vector $x \in \mathbb{C}^n$ has a Kashin's representation of level K with respect to the system (u_i);

(ii) The $n \times N$ matrix U whose columns are u_i satisfies

$$B^n \subseteq \frac{K}{\sqrt{N}} U(Q^N). \tag{2.4}$$

■

Kashin's theorem [30] states that there exists an orthogonal projection of the unit cube in $\mathbb{C}^N$ onto a subspace of dimension n, which is equivalent to a Euclidean ball (see [36] Section 6 for more general results). In other words, there exists an $n \times N$ matrix U whose rows are orthonormal and such that

$$B^n \subseteq \frac{K}{\sqrt{N}} U(Q^N) \subseteq K B^n, \tag{2.5}$$

where K depends on the redundancy $\lambda = N/n$ only, not on particular values n and N. (Note that the second inclusion in (2.5) trivially holds.) In fact, a random matrix U with orthonormal rows, picked with respect to a rotationally invariant distribution, satisfies (2.5) with high probability (exponential in n).

The first inclusion in (2.5), by Observation 2.2 means that the columns u_i of the matrix U form a system for which every vector has a Kashin's representation. Since the rows of U are orthonormal, (u_i) is a tight frame. This proves Theorem 2.1. ■

In geometric functional analysis, many classes of matrices U are known to realize Euclidean projections of the cube as in (2.5). We discuss these in Section 4.1.

2.3. **Error reduction, vector quantization.** Kashin's representations have a maximal power to reduce errors in the coefficients. Indeed, consider a tight frame $(u_i)_{i=1}^N$ in $\mathbb{C}^n$, but instead of using frame representations we shall use Kashin's representations with some constant level K. So we represent a vector $x \in \mathbb{C}^n$, $\|x\|_2 \le 1$, with its Kashin's coefficients $(a_1, \dots, a_N) \in C^N$, $|a_i| \le K$. Assume these coefficients are damaged (due to quantization, losses, flips of bits, etc.) and we only know noisy coefficients $(\hat{a}_1, \dots, \hat{a}_N) \in C^N$, $|a_i| \le K$. When we try to reconstruct the vector x from the noisy coefficients as $\hat{x} = \frac{1}{\sqrt{N}} \sum_{i=1}^N \hat{a}_i u_i$, the accuracy of this reconstruction is

$$\|x - \hat{x}\|_2 = \frac{1}{\sqrt{N}} \Big\| \sum_{i=1}^N (a_i - \hat{a}_i) u_i \Big\|_2 \le \Big(\frac{1}{N} \sum_{i=1}^N |a_i - \hat{a}_i|^2 \Big)^{1/2}. \tag{2.6}$$

Hence *the error of a Kashin's representation is bounded by the average error of its coefficients.*

In particular, Kashin's representations effectively reduce *quantization errors*. Suppose we need to quantize a vector $x \in \mathbb{C}^n$. We perform a scalar quantization of the coefficients a_i with step ε, i.e. $|a_i - \hat{a}_i| \le \varepsilon$. By (2.6), the quantization error is

$$\|x - \hat{x}\|_2 \le \varepsilon.$$

This is much better than in (1.1): the quantization error no longer grows with the dimension n.

Kashin's decompositions also withstand *arbitrary errors* made to a small fraction of the coefficients a_i. These may include losses of coefficients and arbitrary flips of bits. Suppose at most δN coefficients $(a_1, \dots, a_N)$ are damaged in an arbitrary way, which results in coefficients $(\hat{a}_1, \dots, \hat{a}_N)$. Since all $|a_i| \le K$, we can assume (by truncation) that all $|\hat{a_i}| \le K$. When we try to reconstruct the vector x from the damaged coefficients (as before), the accuracy of the reconstruction can be estimated using (2.6) as

$$\|x - \hat{x}\|_2 \le 2K\delta^{1/2},$$

which is small if the portion δ of the damaged coefficients is small (recall that K is a constant).

By Theorem 2.1, the maximal error reduction effect is achieved using frames with only a *constant redundancy*, say with $N = 2n$. This shows rather unexpectedly that the redundancy can have an abrupt rather than a gradual effect: any redundancy factor $\lambda = N/n > 1$ has the error reduction power of maximal possible order.

3. Computing Kashin's representations

Computing the coefficients a_i of a Kashin's representation of a given vector x can be described as a linear feasibility problem, and can be solved in (weakly) polynomial time using linear programming methods.

In this paper, we shall take a different approach to computing Kashin's representations, by establishing their connection with the Uncertainty Principle. This will have several advantages over the linear programming approach:

(1) Whenever a frame (u_i) satisfies the Uncertainty Principle, one can effectively transform every frame representations into a Kashin's representation. It will take $O(\log N)$ multiplications of the matrix U by a vector.
(2) The Uncertainty Principle will thus be a guarantee that a given frame (u_i) yields a Kashin's representation for every vector. This can help to identify frames that yield Kashin's representations.
(3) The algorithm to transform frame representations into Kashin's representations is simple, natural, and robust. It can thus be implemented on analog devices. Followed by some robust scalar quantization of coefficients (such as one-bit β-quantization [13, 14]), this algorithm yields a *robust one-bit vector quantization scheme for analog-to-digital conversion.* The vector quantization error in this scheme is bounded by the average of the quantization errors of the coefficients, as we discussed in Section 2.3.

3.1. **The Uncertainty Principle.**

The classical uncertainty principle of quantum mechanics has an equivalent harmonic analysis form: no function can be localized both in time and in frequency [27]. A variant of the uncertainty principle due to Donoho and Stark [16] states that if $f \in L^2(\mathbb{C})$ almost vanishes outside a measurable set T and its Fourier transform $\hat{f}$ almost vanishes outside a measurable set Ω, then the product of measures $|T||\Omega|$ is at least almost 1. Donoho and Stark proposed applications of this principle for signal recovery [16].

For signals on discrete domains Ω, no satisfying version of the uncertainty principle was known until three years ago. The uncertainty principle for the discrete Fourier transform in $\mathbb{C}^N$ states that $|\text{supp}(x)||\text{supp}(\hat{x})| \ge N$ for all $x \in \mathbb{C}^N$ (see [16]). This inequality is unfortunately sharp – both terms in this product can be of order $\sqrt{N}$.

In a sequence of recent papers by Candes, Romberg and Tao [4, 7, 3] and by Rudelson and Vershynin [37, 38], it was shown that a much stronger discrete uncertainty principle holds for *random* sets of size proportional to N. Moreover, one of these sets can be arbitrary (non-random), and the other (random) can be almost the whole domain! The following result is a consequence of [37, 38]:

Theorem 3.1 (Uncertainty Principle)**.** *Let $N = (1+\mu)n$ for some integer n and $\mu \in (0,1)$. Then a random subset Ω of $\{0,\dots,N-1\}$ of average cardinality n satisfies the following with high probability. For every $z \in \mathbb{C}^N$,*

$$\mathrm{supp}(z) \subseteq \Omega \quad \textit{implies} \quad |\mathrm{supp}(\hat{z})| > \delta N,$$

where $\delta = c\mu^2/\log^2 N$ and $c > 0$ is an absolute constant.

Moreover, for every $x \in \mathbb{C}^N$, $|\mathrm{supp}(x)| \le \delta N$, one has

$$\|\hat{x} \cdot \mathbf{1}_\Omega\|_2 \le (1 - c\mu)\|x\|_2, \tag{3.1}$$

where $\mathbf{1}_\Omega$ denotes the indicator function of Ω.

The first, qualitative, part of the theorem easily follows from the second, quantitative part with $z = \hat{x}$. If $\mathrm{supp}(\hat{x}) \subseteq \Omega$ and $|\mathrm{supp}(x)| \le \delta N$ then, by the second part, $\|\hat{x}\|_2 = \|\hat{x} \cdot \mathbf{1}_\Omega\|_2 < \|x\|_2$, which would contradict Plancherel's inequality.

We can regard inequality (3.1) as a property of the partial Fourier matrix U, which consists of the rows of the DFT matrix Φ indexed in the random set Ω. Then (3.1) says that $\|Ux\|_2 \le (1 - c\mu)\|x\|_2$ for all vectors $x \in \mathbb{C}^N$ such that $|\mathrm{supp}(x)| \le \delta N$. Now we can abstract from the harmonic analysis in question and introduce a general uncertainty principle as a property of matrices.

Definition 3.2 (Uncertainty Principle for matrices)**.** An $n \times N$ matrix U satisfies the Uncertainty Principle with parameters $\eta, \delta \in (0,1)$ if

$$\|Ux\|_2 \le \eta\|x\|_2 \quad \text{for all } x \in \mathbb{C}^N \text{ such that } |\mathrm{supp}(x)| \le \delta N. \tag{3.2}$$

We will only use the Uncertainty Principle for matrices U with orthonormal or almost orthonormal rows, in which case it is always a nontrivial property.

A related Uniform Uncertainty Principle (UUP) was recently introduced Candes and Tao in the context of the sparse recovery problems [8]. The UUP with parameters $\varepsilon, \delta \in (0,1)$ states that there exists $\lambda > 0$ such that

$$\lambda(1-\varepsilon)\|x\|_2 \le \|Ux\|_2 \le \lambda(1+\varepsilon)\|x\|_2 \quad \text{for all } x \in \mathbb{C}^N \text{ such that } |\mathrm{supp}(x)| \le \delta N.$$

See also [5, 6] for more refined versions. Known also as the Restricted Isometry Condition, UUP was shown in [8] to be a guarantee that one can efficiently solve underdetermined systems of linear equations $Ux = b$ under the assumption that the solution is sparse, $|\mathrm{supp}(x)| \le \delta N$. This is a part of the fast developing area of Compressed Sensing [9].

The Uncertainty Principle is a weaker assumption (thus easier to verify) than the UUP:

Observation 3.3. *For matrices with orthonormal rows, the UUP with parameters (ε, δ) implies the Uncertainty Principle with parameters $\eta = \frac{1+\varepsilon}{1-\varepsilon}\sqrt{\frac{n}{N}}, \delta$.*

The proof is straightforward, so we omit it.

The Uncertainty Principle can be equivalently stated as a property of systems of vectors $(u_i)_{i=1}^N$, which form the columns of the matrix U. We will use it for tight or (almost tight) frames, in which case it is a nontrivial property:

Definition 3.4 (Uncertainty Principle for frames)**.** A system of vectors $(u_i)_{i=1}^N$ in $\mathbb{C}^n$ satisfies the Uncertainty Principle with parameters $\eta, \delta \in (0,1)$ if

$$\Big\| \sum_{i \in \Omega} a_i u_i \Big\|_2 \le \eta \Big(\sum_{i \in \Omega} |a_i|^2 \Big)^{1/2} \tag{3.3}$$

for every subset $\Omega \subset \{1, 2, \dots, N\}$, $|\Omega| \le \delta N$.

3.2. **Converting frame representations into Kashin's representations.** For every tight frame that satisfies the Uncertainty Principle, one can quickly transform frame representations into Kashin's representations.

This transform is natural and fast. We truncate the coefficients of the frame representation (2.2) of x at level $M = \|x\|_2/\sqrt{\delta N}$ in hope to achieve a Kashin's representation with level $1/\sqrt{\delta}$. The truncated representation may sum up to a vector $x^{(1)}$ different from x. So we consider the residual $x - x^{(1)}$, compute its frame representation and again truncate its coefficients, now at a lower level ηM. We continue this process of expansion – truncation – reconstruction, each time reducing the truncation level by the factor of η. The Uncertainty Principle will imply that the norm of the residual reduces by the factor of η at each iteration. So we can compute Kashin's representations of level $K = K(\eta, \delta)$ with accuracy ε in $O(\log(1/\varepsilon))$ iterations.

The analysis of this algorithm yields:

Theorem 3.5 (From frame to Kashin representations)**.** *Let $(u_i)_{i=1}^N$ be a tight frame in $\mathbb{C}^n$ which satisfies the Uncertainty Principle with parameters η, δ. Then each vector $x \in \mathbb{C}^n$ admits a Kashin representation of level $K = (1-\eta)^{-1}\delta^{-1/2}$.*

In order to prove Theorem 3.5, we first introduce the truncation operator for frame representations. Given a number $M > 0$ the one-dimensional truncation at level M is defined as

$$t_M(z) = M\frac{z}{|z|}\min\{|z|, M\}, \quad z \in \mathbb{C} \tag{3.4}$$

Let $(u_i)_{i=1}^N \subset \mathbb{C}^n$ be a tight frame which satisfies the Uncertainty Principle with parameters η, δ. Fix some $C \geq \delta^{-1/2}$.

We start with the frame representation (2.2) and define the truncation operator on $\mathbb{C}^n$ as

$$Tx = \sum_{i=1}^N \hat{b}_i u_i, \quad \text{where} \quad \hat{b}_i = t_M(b_i), \ M = \frac{\|x\|_2}{\sqrt{\delta N}}. \tag{3.5}$$

Due to the Uncertainty Principle, the residual of the truncation is small:

Lemma 3.6 (Truncation error)**.** *In the above notations, for every $x \in \mathbb{C}^n$,*

$$\|x - Tx\|_2 \leq \eta\|x\|_2. \tag{3.6}$$

Proof. Fix $x \in \mathbb{C}^n$. Consider the subset Ω of $\{1, \dots, N\}$ defined as

$$\Omega = \{i : \ b_i \neq \hat{b}_i\} = \{i : \ |b_i| > M\}.$$

Then, by the definition of the tight frame,

$$\|x\|_2^2 = \sum_{i=1}^N |b_i|^2 > |\Omega| M^2,$$

thus

$$|\Omega| \leq \frac{\|x\|_2^2}{M^2} = \delta N.$$

Now we can estimate

$$x - Tx = \sum_{i\in\Omega}(b_i - \hat{b}_i)u_i$$

using the Uncertainty Principle as

$$\|x - Tx\|_2 \le \eta \Big(\sum_{i \in \Omega} |b_i - \hat{b}_i|^2 \Big)^{1/2} \le \eta \Big(\sum_{i \in \Omega} |b_i|^2 \Big)^{1/2} \le \eta \Big(\sum_{i=1}^{N} |b_i|^2 \Big)^{1/2} = \eta \|x\|_2.$$

This completes the proof. ∎

Proof of Theorem 3.5. Given $x \in \mathbb{C}^n$, we put

$$x^{(0)} := x, \quad x^{(k)} := x^{(k-1)} - Tx^{(k-1)}, \quad k = 1, 2, \dots$$

Then for each $r \in \mathbb{N}$,

$$x = \sum_{k=0}^{r} Tx^{(k)} + x^{(r+1)}.$$

It follows by induction from Lemma 3.6 that $\|x^{(k)}\|_2 \le \eta^k \|x\|_2$, thus

$$x = \sum_{k=0}^{\infty} Tx^{(k)}.$$

Furthermore, by the definition of the truncation operator T, each vector $Tx^{(k)}$ has an expansion in the system $(u_i)_{i=1}^N$ with coefficients bounded by $\|x^{(k)}\|_2/\sqrt{\delta N} \le \eta^k \|x\|_2/\sqrt{\delta N}$. Summing up these expansions for $k = 0, 1, 2, \dots$, we obtain an expansion of x in the system $(u_i)_{i=1}^N$ with coefficients bounded by $(1-\eta)^{-1}\|x\|_2/\sqrt{\delta N}$. In other words, x has a Kashin's representation with level $K = (1-\eta)^{-1}\delta^{-1/2}$. This completes the proof. ∎

The proof yields an algorithm to compute Kashin's representations:

Algorithm to compute Kashin's representations

Input:
- A tight frame $(u_i)_{i=1}^N$ in $\mathbb{C}^n$ which satisfies the Uncertainty Principle with parameters $\eta, \delta \in (0,1)$.
- A vector $x \in \mathbb{C}^n$ and a number of iterations r.

Output: Kashin's decomposition of x with level $K = (1-\eta)^{-1}\delta^{-1/2}$ and with accuracy $\eta^r \|x\|_2$. Namely, the algorithm finds coefficients $a_1, \dots, a_N$ such that

$$\Big\| x - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} a_i u_i \Big\|_2 \le \eta^r \|x\|_2 \quad \text{and all } |a_i| \le K\|x\|_2.$$

- Initialize the coefficients and the truncation level:
$$a_i \leftarrow 0, \; i = 1, \dots, N; \quad M \leftarrow \frac{\|x\|_2}{\sqrt{\delta N}}.$$
- Repeat the following r times:
  – Compute the frame representation of x and truncate at level M:
$$b_i \leftarrow \langle x, u_i \rangle; \quad \hat{b}_i \leftarrow t_M(b_i), \quad i = 1, \dots, N.$$
  – Reconstruct and compute the error:
$$Tx \leftarrow \sum_{i=1}^{N} \hat{b}_i u_i; \quad x \leftarrow x - Tx.$$
  – Update Kashin's coefficients and the truncation level:
$$a_i \leftarrow a_i + \sqrt{N}\,\hat{b}_i, \; i = 1, \dots, N; \quad M \leftarrow \eta M.$$

Remark. *(Redistributing information).* One can view this algorithm as a method of redistributing information among the coefficients. At each iteration, it "shaves off" excessive information from the few biggest coefficients (using truncation) and redistributes this excess more evenly. This process is continued until all coefficients have a fair share of the information.

Remark. *(Computing exact Kashin's representations).* With a minor modification, this algorithm can compute an *exact* Kashin's representation after $r = O(\log N)$ iterations. We just do not need to truncate the coefficients b_i during the last iteration.

Indeed, for such r, the error factor satisfies $\eta^r \le \frac{K}{\sqrt{N}}$. Thus, during r-th iteration the frame coefficients b_i are all bounded by $\frac{K}{\sqrt{N}}\|x\|_2$, where x is the initial input vector. So b_i are already sufficiently small, and we will not apply the truncation at the last iteration. This yields an exact Kashin's representation of x with $K' = 2K$.

Remark. *(Robustness).* The algorithm above is robust in the sense of [12].

Specifically, the truncation operation (3.4) may be impossible to perform on a physical signal exactly, because it is expensive to build an analog scheme that has an exact phase transition at the truncation level $|z| = M$. A robust algorithm can not rely on any assumptions on exact phase transitions of the operations it uses. Scalar quantizers that are robust in this sense were first constructed by Daubechies and DeVore in [12] and further developed in [26, 13, 14].

Our algorithm is also robust in this sense: the exact truncation t_M can be replaced with any approximate truncation. Such an approximate truncation at level 1 can be any function $t(z) : \mathbb{C} \to \mathbb{C}$ which satisfies for some $\nu, \tau \in (0, 1)$:

$$|z - t(z)| \le \begin{cases} \nu|z| & \text{if } |z| \le \tau, \\ |z| & \text{for all } z \end{cases} \qquad \text{and } |t(z)| \le 1 \text{ for all } z. \tag{3.7}$$

The approximate truncation at level M is defined as $t_M(u) := M\, t(\frac{u}{M})$. The analysis similar to that above yields:

Theorem 3.7. *The algorithm above remains valid for the approximate truncations as above, for level M replaced with $M' = \tau^{-1}M$, parameter η replaced with $\eta' = \sqrt{\eta^2 + \nu^2}$, and level K replaced with $K' = \tau^{-1}(1 - \eta')^{-1}\delta^{-1/2}$, provided that $\eta' < 1$.*

Moreover, the approximate truncation can be different every time it is called by the algorithm, provided that it satisfy (3.7). This facilitates an implementation of the algorithm on analog devices. In particular, one can use this algorithm to build robust vector quantizers for analog-to-digital conversion, as we described in Section 2.3.

Remark. *(Almost tight frames).* Similar results also hold for frames that are almost, but not exactly, tight. This is important for natural classes of frames, such as random gaussian and subgaussian frames (see Theorem 4.6).

Definition 3.8. For $\varepsilon \in (0, 1)$, a sequence $(u_i)_{i=1}^N \subset \mathbb{C}^n$ is called an ε-tight frame if

$$(1 - \varepsilon)\|x\|_2 \le \Big(\sum_{i=1}^N |\langle x, u_i\rangle|^2\Big)^{1/2} \le (1 + \varepsilon)\|x\|_2 \qquad \text{for all } x \in \mathbb{C}^n. \tag{3.8}$$

An analysis similar to that above yields:

Theorem 3.9. *Let $(u_i)_{i=1}^N \subset \mathbb{C}^n$ be an ε-tight frame, which satisfies the uncertainty principle with parameters η and δ. Then Theorem 3.5 and the algorithm above are valid for M replaced with $M' = \sqrt{1+\varepsilon}\, M$ and η replaced with $\eta' = \sqrt{1+\varepsilon}\, \eta + \varepsilon$, provided that $\eta' < 1$.*

Remark. *(History).* The idea behind Theorem 3.5 is certainly not new. Gluskin [19] suggested to use properties that involved only $\|\cdot\|_2$ norms (like our Uncertainty Principle) to deduce results on Euclidean sections of ℓ_1^n (which by duality is equivalent to Euclidean projections (2.4) of a cube). A similar idea was essentially used by Talagrand in his work on the Λ_1 problem [40].

The algorithm to compute Kashin's representations resembles the Chaining Algorithm of [18], which also detects a few biggest coefficients and iterates on the residual, but serves to *find* all big coeffients rather than to spread them out.

4. Matrices and frames that satisfy the Uncertainty Principle

In this section, we give examples of matrices (equivaletly, frames) that satisfy the Uncertainty Principle. By Observation 2.2, such $n \times N$ matrices U realize Euclidean projection of the cube (2.4). Equivalently, these frames $(u_i)_{i=1}^N$ (the columns of U) yield quickly computable Kashin's representations for every vector $x \in \mathbb{C}^n$.

4.1. **Matrices known to realize Euclidean projections of the cube.** Much attention has been paid to such Euclidean projections of the cube (2.4) in geometric functional analysis. Results in the literature are usually stated in the dual form, about n-dimensional Euclidean subspaces of ℓ_1^N.

Kashin proved (2.4) for a random orthogonal $n \times N$ matrix U (formed by the first n rows of a random matrix in $\mathcal{O}(N)$), with $N = \lambda n$ for arbitrary $\lambda > 1$, and with exponential probability ([30], see also [36] Section 6.) The level K (2.4) depends only on λ; an optimal dependence was given later by Garnaev and Gluskin [17]).

A similar result holds for $U = \frac{1}{\sqrt{N}}\Phi$, where Φ is a random Bernoulli matrix, which means that the entries of Φ are ± 1 symmetric independent random variables. Schechtman [39] first proved this result with $N = O(n)$, and [33] generalized this for $N = \lambda n$ fr arbitrary $\lambda > 1$. The dependence of K on λ was improved recently in [1]. In fact, these results hold for a quite general class of subgaussian matrices (which includes Bernoulli and Gaussian random matrices).

In is unknown whether Kashin's theorem holds for a partial Fourier matrix; this conjecture is known as the Λ_1 problem. Consider the Discrete Fourier Transform in $\mathbb{C}^N$, where $N = O(n)$, given by the orthogonal $N \times N$ matrix Φ. It is unknown whether there exists a submatrix U which consists of some n rows of Φ and such that it realizes an Euclidean proection of the cube in the sense of (2.4).

In the positive direction, a partial result due to Bourgain, later reproved by Talagrand with a general method [40], states that a random partial Fourier matrix U satisfies (2.5) with high probability for $N = O(n)$ and $K = O(\sqrt{\log(N) \log\log(N)})$. It was recently proved in [25] that Bourgain's result holds arbitrarily small redundancy, that is for $N = \lambda n$ with arbitrary $\lambda > 1$, however at the cost of a slightly worse logarithmic factor in K. A similar result can also be deduced from Theorem 4.3 below (along with Theorem 3.5 and Observation 2.2), which is a consequence of the uncertainty principle in [37, 38].

No explicit constructions of matrices U are known. However, there exists small space constructions that use a small number of random bits [2, 28, 29].

4.2. **Random orthogonal matrices.** We consider random $n \times N$ matrices whose rows are orthonormal. Such matrices can be obtained by selecting the first n rows of orthogonal $N \times N$ matrices. Indeed, denote by $\mathcal{O}(N)$ the space of all orthogonal $N \times N$ matrices with the normalized Haar measure. Then

$$\mathcal{O}(n \times N) = \{P_n V; V \in \mathcal{O}(N)\}, \tag{4.1}$$

where $P_n : \mathbb{C}^N \to \mathbb{C}^n$ is the orthogonal projection on the first n coordinates. The probability measure on $\mathcal{O}(n \times N)$ is induced by the Haar measure on $\mathcal{O}(N)$.

Theorem 4.1 (UP for random orthogonal matrices)**.** *Let $\mu > 0$ and $N = (1 + \mu)n$. Then, with probability at least $1 - 2\exp(-c\mu^2 n)$, a random orthogonal $n \times N$ matrix U satisfies the Uncertainty Principle with the parameters*

$$\eta = 1 - \frac{\mu}{4}, \quad \delta = \frac{c\mu^2}{\log(1/\mu)}, \tag{4.2}$$

where $c > 0$ is an absolute constant.

Remark. Assumption $\mu > 0$ is not essential; just expressions for η and δ will look differently. We are most interested in small values of μ when redundancy is small.

The proof of Theorem 4.1 uses a standard scheme in geometric functional analysis – the concentration inequality on the sphere followed by an ε-net argument. Denote by S^{N-1} and σ_{N-1} the unit Euclidean sphere in $\mathbb{C}^N$ and the normalized Lebesgue measure on S^{N-1}.

Lemma 4.2. *For arbitrary $t > 0$, $x \in S^{N-1}$, we have*

$$\mathbb{P}\left\{\|Ux\|_2 > (1+t)\sqrt{\frac{n}{N}}\right\} \le 2\exp(-c_1 t^2 n),$$

where $c_1 > 0$ is an absolute constant.

Proof. We use the representation (4.1) and also the fact that $z = Vx$ is a random vector uniformly distributed on S^{N-1}. Thus Ux is distributed identically with $P_n z$. We also have

$$E := \int_{S^{N-1}} \|P_n z\|_2 \, d\sigma_{N-1}(z) \le \Big(\int_{S^{N-1}} \|P_n z\|_2^2 \, d\sigma_{N-1}(z)\Big)^{1/2} = \sqrt{\frac{n}{N}}.$$

The map $z \mapsto \|P_n z\|$ is a 1-Lipschitz function on S^{N-1}. The the concentration inequality (see e.g. [31] Section 1.3) then implies that this function is well concentrated about its average value E:

$$\mathbb{P}\{\|Ux\|_2 > E + u\} \le \sigma_{N-1}(z \in S^{N-1} : \ |\|P_n z\|_2 - E| > u\}) \le 2\exp(-cu^2 N).$$

Choosing $u = t\sqrt{n/N}$ completes the proof. ■

Proof of Theorem 4.1. Assume η and δ satisfy the assumptions (4.2). We have to prove that (3.2) holds with probability at least $1 - \exp(-c\mu^2 n)$.

Consider the set

$$S := \{x \in S^{N-1}, \ |\text{supp}(x)| \le \delta N\} = \bigcup_{|I| \le \delta N} S_I$$

where the union is over all subsets I of $\{1, \dots, N\}$ of cardinality at most δN, and $S_I = S^{N-1} \cap \mathbb{C}^I$ is the set of all unit vectors whose supports lie in I. Let $\varepsilon > 0$. For each I, we can find an ε-net of S_I in the Euclidean norm, and of cardinality at most $(3/\varepsilon)^{\delta N}$ (see e.g. [36] Lemma 4.16). Taking the union over all sets I with $|I| = \lceil \delta N \rceil$, we conclude by the Stirling's bound on the binomial coefficients that there exist an ε-net $\mathcal{N}$ of S of cardinality

$$|\mathcal{N}| \le \binom{N}{\lceil \delta N \rceil} \Big(\frac{3}{\varepsilon}\Big)^{\delta N} \le \Big(\frac{3e}{\varepsilon\delta}\Big)^{\delta N}.$$

Then using Lemma 4.2, we obtain

$$\mathbb{P}\{\exists y \in \mathcal{N} :\ \|Uy\|_2 > (1+t)\sqrt{\frac{n}{N}}\} \le |\mathcal{N}| \cdot 2\exp(-c_1 t^2 n).$$

Every $x \in S$ can be approximated by some $y \in \mathcal{N}$ within ε in the Euclidean norm, and since U has norm one, we have

$$\|Ux\|_2 \le \|Uy\|_2 + \|U(x-y)\|_2 \le \|Uy\|_2 + \varepsilon.$$

Therefore

$$\mathbb{P}\{\exists x \in S :\ \|Ux\|_2 > (1+t)\sqrt{\frac{n}{N}} + \varepsilon\} \le |\mathcal{N}| \cdot 2\exp(-c_1 t^2 n). \tag{4.3}$$

It now remains to choose parameters appropriately. Let $t = \mu/5$ and $\varepsilon = \mu/8$. Then since $N/n = 1 + \mu$ and by the assumption on η in (4.2), we have

$$(1+t)\sqrt{\frac{n}{N}} + \varepsilon \le \eta.$$

Also, we can estimate the probability in (4.3) as

$$|\mathcal{N}| \cdot 2\exp(-ct^2 n) \le \Big(\frac{24e}{\delta\mu}\Big)^{\delta N} \cdot 2\exp(-c_2 t^2 n) \le 2\exp\big[(2\delta\log(24e/\delta\mu) - c_2\mu^2)n\big], \tag{4.4}$$

where $c_2 = c_1/25$. By our choice of δ, the right hand side of (4.4) is bounded by $2\exp(-c\mu^2 n)$, where $c > 0$ is an absolute constant. We conclude that

$$\mathbb{P}\{\exists x \in S :\ \|Ux\|_2 > \eta\} \le 2\exp(-c\mu^2 n).$$

This completes the proof. ■

4.3. **Random partial Fourier matrices.** An important class of matrices that satisfy the Uncertainty Principle can be obtained by selecting n random rows of an arbitrary orthogonal $N \times N$ matrix Φ whose entries are $O(N^{-1/2})$. Here n can be an arbitrarily big fraction of N, so the Uncertainty Principle will hold for almost square random submatrices. This class includes a random partial Fourier matrix, multiplication by which corresponds to sampling n random frequencies of a signal.

More precisely, we select rows of Φ using random selectors $\delta_1, \dots, \delta_N$ – independent Bernoulli random variables, which take value 1 each with probability n/N. The selected rows will be indexed by a random subset $\Omega = \{i : \delta_i = 1\}$ of $\{1, \dots, N\}$, whose average cardinality is n.

Theorem 4.3 (UP for random partial Fourier matrices). *Let Φ be an orthogonal $N \times N$ matrix with uniformly bounded entries: $|\Phi_{ij}| \le \alpha N^{-1/2}$ for some constant α and all i, j. Let n be an integer such that $N = (1+\mu)n$ for some $\mu \in (0,1]$. Then for each $p \in (0,1)$ there exists a constant $c = c(p,\alpha) > 0$ such that the following holds.*

Let U be a submatrix of Φ formed by selecting a subset of the rows of average cardinality n. Then, with probability at least $1-p$, the matrix U satisfies the Uncertainty Principle with parameters

$$\eta = 1 - \frac{\mu}{4}, \quad \delta = \frac{c\mu^2}{\log^4 N}.$$

Theorem 4.3 is a direct consequence of a slightly stronger result estabished in [7] and improved in [37, 38]. For an operator U on a Euclidean space, $\|\cdot\|$ will denote its operator norm.

Theorem 4.4 (UUP for partial Fourier matrices [37, 38])**.** *Assume the hypotheses of Theorem 4.3 are satisfied. Then there exists a constant $C = C(\alpha) > 0$ such that the following holds. Let $r > 0$ and $\varepsilon \in (0,1)$ be such that*

$$n \geq C\Big(\frac{r\log N}{\varepsilon^2}\Big)\log\Big(\frac{r\log N}{\varepsilon^2}\Big)\log^2 r.$$

Then the random submatrix U satisfies:

$$\mathbb{E}\sup_{|T|\leq r}\|id_T - \frac{N}{n}U_T^*U_T\| \leq \varepsilon. \tag{4.5}$$

Here the supremum is taken over all subsets T of $\{1,\dots,N\}$ with at most r elements, U_T denotes the submatrix of U that consists of the columns of U indexed in T, and id_T denotes the identity on $\mathbb{C}^T$.

Proof of Theorem 4.3. Observe that for a linear operator A on $\mathbb{C}^N$ one has

$$\|id - A^*A\| = \sup_{x\in\mathbb{C}^N,\|x\|_2=1}\big|\langle(id - A^*A)x, x\rangle\big| = \sup_{x\in\mathbb{C}^N,\|x\|_2=1}\big|\|Ax\|_2^2 - \|x\|_2^2\big|. \tag{4.6}$$

We use this observation for $A = \sqrt{\frac{N}{n}}U_T$. Since $U_Tx = Ux$ whenever $\mathrm{supp}(x) \subseteq T$, we obtain

$$\mathbb{E}\sup_{\substack{x\in\mathbb{C}^N,\ \|x\|_2=1,\\ |\mathrm{supp}(x)|\leq r}}\Big|\frac{N}{n}\|Ux\|_2^2 - 1\Big| \leq \varepsilon.$$

By Markov's inequality, with probability at least $1-p$ the random matrix U satisfies:

$$\Big|\frac{N}{n}\|Ux\|_2^2 - 1\Big| < \varepsilon/p \quad \text{for all } x \in \mathbb{C}^N,\ |\mathrm{supp}(x)| \leq r,\ \|x\|_2 = 1.$$

In particular, for such U, one has:

$$\|Ux\|_2 \leq \sqrt{1+\varepsilon/p}\sqrt{\frac{n}{N}}\,\|x\|_2 = \sqrt{\frac{1+\varepsilon/p}{1+\mu}}\,\|x\|_2 \quad \text{for all } x \in \mathbb{C}^N,\ |\mathrm{supp}(x)| \leq r.$$

Then, if we set $\varepsilon = cp\mu$ for an appropriate absolute constant $c > 0$, we can bound the factor

$$\sqrt{\frac{1+\varepsilon/p}{1+\mu}} \leq 1 - \frac{\mu}{4} = \eta.$$

This proves the Uncertainty Principle (3.2) with $\delta = r/N$. To estimate δ note that the condition on n in Theorem 4.4 is satisfied if

$$r \leq \frac{c_1\varepsilon^2 N}{\log^4 N} \tag{4.7}$$

where $c_1 = c_1(\alpha) > 0$. Since we have set $\varepsilon = cp\mu$, condition (4.7) is equivalent to

$$\delta \geq \frac{c\mu^2}{\log^4 N}$$

where $c = c(\alpha) > 0$. This completes the proof of Theorem 4.3. ■

Remarks.

1. *(Computing in almost linear time).* Fourier matrices can be used to compute Kashin's representations in $\mathbb{C}^n$ time almost linear in n. Indeed, let for example $N = 2n$. The columns of the $N \times N$ partial Fourier matrix form a tight frame in $\mathbb{C}^n$. By Theorem 4.3 and Section 3.2, we can convert a frame representation of every vector $x \in \mathbb{C}^n$ into a Kashin's representation with level $K = O(\log^2 n)$, and in time $O(n \log^2 n)$. (Recall that the algorithm makes $O(\log n)$ multiplications by a partial Fourier matrix, and each multiplication can be done using the fast Fourier transform in time $O(n \log n)$).

2. The constant $c = c(\alpha, p)$ depends polynomially on α and polylogarithmically on p. The polynomial dependence on α is straightforward form the proof of Theorem 4.4 in [37, 38]. The proof above gives a polynomial dependence on the probability p. To improve it to a polylogarithmic dependence, one can use an exponential tail estimate, proved in [38] Theorem 3.9, instead of the expectation estimate (4.5).

3. We stated Theorem 4.3 in the range $\mu \in (0, 1]$ which is most interesting for us (where the redundancy factor is small). A similar result holds for arbitrary $\mu > 0$.

4.4. **Subgaussian random matrices.** A large family of matrices with independent random entries satisfies the Uncertainty Principle.

Definition 4.5. A random variable ϕ is called *subgaussian with parameter* β if

$$\mathbb{P}\{|\phi| > u\} \leq \exp(1 - u^2/\beta^2) \quad \text{for all } u > 0.$$

Examples of subgaussian ramdom variables include Gaussian $N(0, 1)$ random variables and bounded random variables.

Theorem 4.6 (UP for random subgaussian matrices)**.** *Let Φ be a $n \times N$ matrix whose entries are independent mean zero subgaussian random variables with parameter β. Assume that $N = \lambda n$ for some $\lambda \geq 2$. Then, with probability at least $1 - \lambda^{-n}$, the random matrix $U = \frac{1}{\sqrt{N}}\Phi$ satisfies the Uncertainty Principle with parameters*

$$\eta = C\beta\sqrt{\frac{\log \lambda}{\lambda}}, \quad \delta = \frac{c}{\lambda}, \tag{4.8}$$

where $C, c > 0$ are absolute constants.

Remark. Theorem 4.6 and Lemma 4.8 below can be deduced from the recent works [34, 35]. However, we feel that it would be helpful to include short and rather standard proofs of these results here.

Theorem 4.6 follows easily from an estimate on the operator norm of subgaussian matrix.

Lemma 4.7. *(*[32] *Fact 2.4) Let $n \geq k$ and Φ be a $n \times k$ matrix whose entries are independent mean zero subgaussian random variables with parameter β. Then*

$$\mathbb{P}\{\|\Phi\| > t\sqrt{n}\} \leq \exp(-c_1 n t^2/\beta^2) \quad \text{for all } t \geq C_1\beta, \tag{4.9}$$

here $C_1, c_1 > 0$ are absolute constants.

Proof of Theorem 4.6. The Uncertainty Principle for the matrix U with parameters η, δ is equivalent to the following norm estimate:

$$\sup_{|I|=\lceil \delta N \rceil} \|\Phi_I\| \le \eta\sqrt{N},$$

where the supremum is over all subsets $I \subset \{1, \dots, N\}$ of cardinality $\lceil \delta N \rceil$, and where Φ_I denotes the submatrix of Φ obtained by selecting the columns in I.

Without loss of generality, $c < 1$. Since Φ_I is a $n \times \lceil \delta N \rceil$ matrix and $c_2 n \le \lceil \delta N \rceil \le n$, Lemma 4.7 applies for Φ_I. Taking the union bound over all I, we conclude that for every $t > C_1\beta$,

$$\begin{aligned} \mathbb{P}\{\exists I :\ \|\Phi_I\| > t\sqrt{n}\} &\le \binom{N}{\lceil \delta N \rceil} \exp(-c_1 n t^2/\beta^2) \\ &\le \exp\big[(\log(e/\delta) - c_1 t^2/\beta^2)n\big] \le \exp(-c_3 n t^2/\beta^2) \end{aligned}$$

if we choose $t = C\beta\sqrt{\log\lambda}$ and use our choice of $\delta = c/\lambda$. (Here $c_3 = c_1/2$ and C are absolute constants). With this choice of t, we can write the estimate above as

$$\mathbb{P}\{\exists I :\ \|\Phi_I\| > C\beta\sqrt{\frac{\log\lambda}{\lambda}}\sqrt{N}\} \le \exp(-c_3 C^2 n \log\lambda) \le \lambda^{-n}$$

provided we choose the absolute constant C sufficiently big. This means that the Uncertainty Principle with parameters (4.8) fails with probability at most λ^{-n}. ■

Unlike random orthogonal or partial Fourier matrices considered in Sections 4.2 and 4.3, subgaussian matrices do not in general have orthonormal rows. Nevertheless, their rows of subgaussian matrices are almost orthogonal, and their columns form almost tight frames as we describe below. So, one can use Theorem 3.9 instead of Theorem 3.5 to sompute Kashin's representations for such almost tight frames.

The almost orthogonality of subgaussian matrices can be expressed as follows:

Lemma 4.8. *Let Φ be a $n \times N$ matrix whose entries are independent mean zero subgaussian random variables with parameter β and with variance* 1. *There exist constants $C = C(\beta)$, $c = c(\beta) > 0$ such that the following holds. Assume that*

$$N > \frac{C}{\varepsilon^2} \log\Big(\frac{2}{\varepsilon}\Big) \cdot n$$

for some $\varepsilon \in (0, 1)$. Then

$$\mathbb{P}\{\|id - \frac{1}{N}\Phi\Phi^*\| > \varepsilon\} \le 2\exp(-cN\varepsilon^2).$$

Remark. The dependence in $C(\beta)$, $c(\beta)$ is polynomial. Explicit bounds can be deduced from [35].

As a straightforward consequence, we obtain:

Corollary 4.9 (Subgaussian frames are almost tight)**.** *Let Φ be a subgaussian matrix as in Lemma 4.8. Then the columns of the matrix $\frac{1}{\sqrt{N}}\Phi$ form an ε-tight frame $(u_i)_{i=1}^N$ in $\mathbb{C}^n$.*

Proof of Lemma 4.8. In this proof, $C_1, C_2, c_1, c_2, \dots$ will denote positive absolute constants. By a duality argument as in (4.6),

$$\|id - \frac{1}{N}\Phi\Phi^*\| = \sup_{x \in S^{n-1}} |\frac{1}{N}\|\Phi^* x\|_2^2 - 1|.$$

Denote the columns of Φ by ϕ_i. Fix a vector $x \in \mathbb{C}^n$, $\|x\|_2 = 1$. Since the entries of the vector ϕ_i are subgaussian with parameter β, the random variable $\langle \phi_i, x \rangle$ is also subgaussian with parameter $C_1\beta$, where C_1 is an absolute constant (see Fact 2.1 in [32]). Moreover, this random variable has mean zero and variance 1. We can use Bernstein's inequality (see [42]) to control the average of the independent mean zero random variables $|\langle \phi_i, x \rangle|^2 - 1$ as

$$\mathbb{P}\{|\frac{1}{N}\|\Phi^* x\|_2^2 - 1| > u\} = \mathbb{P}\Big\{\Big|\frac{1}{N}\sum_{i=1}^N |\langle \phi_i, x \rangle|^2 - 1\Big| > u\Big\} \le 2\exp(-c_1 N u^2/\beta^4)$$

for all $u \le c\beta$, where $c_1 > 0$ is an absolute constant.

Denote $U = \frac{1}{\sqrt{N}}\Phi$. There exists a u-net $\mathcal{N}$ of the sphere S^{n-1} in the Euclidean norm, and with cardinality $|\mathcal{N}| \le (3/u)^n$ (see e.g. [36] Lemma 4.16). Using the probability estimate above, we can take the union bound to estimate the probability of the event

$$A := \{\forall y \in \mathcal{N} : \ |\|U^* y\|_2^2 - 1| \le u\}$$

as

$$\mathbb{P}(A^c) \le (3/u)^n \cdot 2\exp(-c_1 N u^2/\beta^4).$$

Applying Lemma 4.7 with $t = C_1\beta$, we see that the event

$$B := \{\|U^*\| \le C_1\beta\} \quad \text{satisfies} \quad \mathbb{P}(B^c) \le \exp(-c_2 N).$$

Consider a realization of the random variables for which the event $A \cap B$ holds. For every $x \in S^{n-1}$, we can find an element of the net $y \in \mathcal{N}$ such that $\|x - y\|_2 \le u$, which implies by the triangle inequality that

$$\begin{aligned} |\|U^* x\|_2 - 1| &\le |\|U^* y\|_2 - 1| + |\|U^* x\|_2 - \|U^* y\|_2| \\ &\le |\|U^* y\|_2^2 - 1| + \|U^*(x - y)\|_2 \le u + 2C_1\beta u \le C_2\beta u, \end{aligned}$$

where $C_2 = 1 + 2C_1$. Now let $u = \varepsilon/3C_2\beta$. Thus $C_2\beta u = \varepsilon/3 \in (0, 1)$, and the estimate above yields $|\|U^* x\|_2^2 - 1| < \varepsilon$ for all $x \in S^{n-1}$ once the event $A \cap B$ holds. Thus

$$\begin{aligned} \mathbb{P}\{\|id - \frac{1}{N}\Phi\Phi^*\| > \varepsilon\} &\le \mathbb{P}\{\exists x \in S^{n-1} : \ |\|U^* x\|_2^2 - 1| > \varepsilon\} \le \mathbb{P}((A \cap B)^c) \\ &\le (3/u)^n \cdot 2\exp(-c_1 N u^2/\beta^4) + \exp(-c_2 N) \le 2\exp(-cN\varepsilon^2) \end{aligned}$$

by our choice of u and by the assumption on N. ∎

References

[1] S. Artstein-Avidan, O. Friedland, V. Milman, S. Sodin, *Polynomial bounds for large Bernoulli sections of ℓ_1^N*, submitted

[2] S. Artstein-Avidan, V. D. Milman, *Logarithmic reduction of the level of randomness in some probabilistic geometric constructions*, Journal of Functional Analysis, Volume 235 Issue 1 (2006), 297-329

[3] E. J. Candés, J. Romberg, *Quantitative Robust Uncertainty Principles and Optimally Sparse Decompositions*, preprint

[4] E. J. Candés, J. Romberg, T. Tao, *Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information*, IEEE Trans. Inform. Th., to appear

[5] E. J. Candés, J. Romberg, T. Tao, *Stable Signal Recovery from Incomplete and Inaccurate Measurements*, Comm. Pure Appl. Math., to appear

[6] E. J. Candés, M. Rudelson, T. Tao, R. Vershynin, *Error Correction via Linear Programming*, Proc. 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005), IEEE, 2005. pp. 295-308

[7] E. J. Candés, T. Tao, *Near Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?*, IEEE Trans. Inform. Th., to appear

[8] E. J. Candés, T. Tao, *Decoding by Linear Programming*, IEEE Trans. Inform. Th., to appear
[9] Compressed Sensing webpage, `http://www.dsp.ece.rice.edu/CS/`
[10] Z. Cvetković, *Resilience properties of redundant expansions under additive noise and quantization*, IEEE Trans. Inform. Th. 49 (2003), 644–656
[11] I. Daubechies, *Ten lectures on wavelets*, SIAM, Philadelphia, PA, 1992
[12] I. Daubechies, R.DeVore, *Approximating a bandlimited function using very coarsely quantized data:a family of stable sigma-delta modulators of arbitrary order*, Ann. of Math. (2) 158 (2003), 679–710
[13] I. Daubechies, R. DeVore, C. S. Güntürk, V. A. Vaishampayan, *Beta Expansions: A New Approach to Digitally Correct A/D Conversion*, Proceedings of the IEEE International Symposium on Circuits and Systems ISCAS 2002, Phoenix, May 2002
[14] I. Daubechies, R. DeVore, C. S. Güntürk, V. A. Vaishampayan, *A/D Conversion with Imperfect Quantizers*, EEE Trans Inf. Theory 53 (2006), 874-885
[15] I. Daubechies, O. Ylmaz, *Robust and practical analog-to-digital conversion with exponential precision*, IEEE Trans. Inform. Theory 52 (2006), 3533–3545.
[16] D. Donoho, P. B. Stark, *Uncertainty principles and signal recovery*, SIAM J. Appl. Math. 49 (1989), 906–931
[17] A. Yu. Garnaev, E. D. Gluskin, *The widths of a Euclidean ball* (Russian), Dokl. Akad. Nauk SSSR 277 (1984), 1048–1052. English translation: Soviet Math. Dokl. 30 (1984), 200–204
[18] A. Gilbert, M. Strauss, J. Tropp, R. Vershynin, *Algorithmic linear dimension reduction in the ell-1 norm for sparse vectors*, submitted. Conference version: Allerton 2006 (44th Annual Allerton Conference on Communication, Control, and Computing)
[19] E. D. Gluskin, Unpublished communications
[20] V.K.Goyal, *Theoretical Foundations of Transform Coding*, IEEE Signal Processing Magazine, vol. 18, no. 5, Sept. 2001, pp. 9–21
[21] V.K.Goyal, *Multiple Description Coding: Compression Meets the Network*, IEEE Signal Processing Mag., vol. 18, no. 5, Sept. 2001, pp. 74–93
[22] V. K. Goyal, J. Kovacevic, J. A. Kelner, *Quantized Frame Expansions with Erasures*, Applied and Computational Harmonic Analysis 10 (2001), 203–233
[23] V. K. Goyal, M. Vetterli, N. T. Thao, *Quantized overcomplete expansions in* $\mathbb{R}^n$*: analysis, synthesis, and algorithms*, IEEE Trans. Inform. Th. 44 (1998), 16–31
[24] R.M.Gray, D.L.Neuhoff, *Quantization*, IEEE Trans. Inform. Theory 44 (1998), 2325–2383
[25] O. Guedon, S. Mendelson, A. Pajor, N. Tomczak-Jaegermann, *Subspaces and orthogonal decompositions generated by bounded orthogonal systems*, Positivity, to appear
[26] S. Güntúrk, *One-Bit Sigma-Delta Quantization with Exponential Accuracy*, Communications on Pure and Applied Mathematics 56 (2003), 1608-1630
[27] V. Havin, B. Jöricke, *The uncertainty principle in harmonic analysis*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], 28. Springer-Verlag, Berlin, 1994
[28] P. Indyk, *Stable Distributions, Pseudorandom Generators, Embeddings and Data Stream Computation*, 41st Symposium on Foundations of Computer Science, 2000
[29] P. Indyk, *Uncertainty Principles, Extractors, and Explicit Embeddings of L2 into L1*, Electronic Colloquium on Computational Complexity, Report No. 126 (2006)
[30] B. Kashin, *Sections of some finite dimensional sets and classes of smooth functions*, Izv. Acad. Nauk SSSR 41 (1977), 334-351 (Russian)
[31] M. Ledoux, *The Concentration of Measure Phenomenon*, AMS Math. Surveys and Monographs, Vol. 89, 2001
[32] A. E. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, *Smallest singular value of random matrices and geometry of random polytopes*, Adv. Math. 195 (2005), no. 2, 491–523
[33] A. E. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, R. Vershynin, *Euclidean embeddings in spaces of finite volume ratio via random matrices*, J. Reine Angew. Math. 589 (2005), 1–19
[34] S. Mendelson, A. Pajor, N. Tomczak-Jaegermann, *Reconstruction and subgaussian operators*, Geometric and Functional Analysis, to appear
[35] S. Mendelson, A. Pajor, N. Tomczak-Jaegermann, *Uniform uncertainty principle for Bernoulli and subgaussian ensembles*, submitted
[36] G. Pisier, *The volume of convex bodies and Banach space geometry*, Cambridge Tracts in Mathematics 94, Cambridge University Press, 1989

[37] M. Rudelson, R. Vershynin, *Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements*, CISS 2006 (40th Annual Conference on Information Sciences and Systems).

[38] M. Rudelson, R. Vershynin, *On sparse reconstruction from Fourier and Gaussian measurements*, submitted

[39] G. Schechtman, *Random embeddings of Euclidean spaces in sequence spaces*, Israel J. Math. 40 (1981), 187–192

[40] M. Talagrand, *Selecting a proportion of characters*, Israel J. Math. 108 (1998), 173–191

[41] V. N. Temlyakov, *Nonlinear methods of approximation*, Found. Comput. Math. 3 (2003), 33–107.

[42] A.W. van der Vaart, J. A. Wellner, *Weak convergence and empirical processes*, Springer Series in Statistics. Springer-Verlag, New York, 1996

Department of Mathematics, Norwegian University of Science and Technology 7491, Trondheim, Norway

E-mail address: `yura@math.ntnu.no`

Department of Mathematics, University of California, Davis, California, 95616, U.S.A.

E-mail address: `vershynin@math.ucdavis.edu`