

# Uncertainty principles and vector quantization

Yurii Lyubarskii and Roman Vershynin

**Abstract**—Given a frame in  $\mathbb{C}^n$  which satisfies a form of the uncertainty principle (as introduced by Candes and Tao), it is shown how to quickly convert the frame representation of every vector into a more robust Kashin’s representation whose coefficients all have the smallest possible dynamic range  $O(1/\sqrt{n})$ . The 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, including coefficient losses and distortions.

**Index Terms**—Frame representations, Kashin’s representations, restricted isometries, uncertainty principles

## I. INTRODUCTION

Quantization is a representation of continuous structures with discrete structures. 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 digital world. A survey of the state-of-the-art of quantization prior to 1998 as well as outline of its numerous applications can be found the paper [22] by Gray and Neuhoff. For more recent developments, we refer the reader to [15] and references therein.

In this paper, we are interested in robust vector encoding and vector quantization. Orthogonal expansions gives a classical way to encode vectors in finite dimensions. One first chooses a convenient orthonormal basis  $(u_i)_{i=1}^n$  of  $\mathbb{C}^n$ . Then one encodes a vector  $x \in \mathbb{C}^n$  by the coefficients  $(a_i)_{i=1}^n$  of its orthogonal expansion

$$x = \sum_{i=1}^n a_i u_i, \quad \text{where } a_i = \langle x, u_i \rangle.$$

An example of this situation is the discrete Fourier transform. At the next step, one quantizes the coefficients  $a_i$  using a convenient

scalar quantizer (for example, a uniform quantizer with fixed number of levels).

A drawback of orthogonal expansions is that the information contained in the vector  $x$  may get distributed unevenly among the coefficients  $a_i$ , which makes this encoding vulnerable to distortions and losses of the coefficients. For example, if  $x$  is collinear with the first basis vector  $u_1$  then all the coefficients except  $a_1$  are zero. If the first coefficient  $a_1$  is lost (for example due to transmission failure) then we can not reconstruct the vector  $x$  even approximately.

A popular way to improve the stability of vector encoding is to use redundant systems of vectors  $(u_i)_{i=1}^N$  in  $\mathbb{C}^n$  called tight frames. These are generalizations of orthonormal bases in the sense that every vector  $x \in \mathbb{C}^n$  can still be represented as

$$x = \sum_{i=1}^N a_i u_i, \quad \text{where } a_i = \langle x, u_i \rangle, \quad (\text{I.1})$$

but for  $N > n$  frames are clearly linearly dependent systems of vectors. These dependencies cause the information contained in  $x$  to spread among several frame coefficients  $a_i$ , which improves the stability of such representations with respect to errors (for example losses and quantization errors), see e.g. [11], [21], [10] and references therein.

The idea of spreading the information evenly among the coefficients is developed in the present paper, and in a sense it is pushed to its limit. As in the previous approaches, we shall start with a frame  $(u_i)_{i=1}^N$ . But instead of the standard frame expansions (I.1) we will be looking at expansions  $x = \sum_{i=1}^N a_i u_i$  with coefficients having the smallest possible dynamic range  $|a_i| = O(1/\sqrt{N})$ . This ensures that the information contained in  $x$  is spread among the coefficients  $a_i$  nearly uniformly. We call such representations of vector  $x$  *Kashin representations*. In this paper we demonstrate the following:

(a) there exist frames  $(u_i)_{i=1}^N$  in  $\mathbb{C}^n$  with redundancy factor  $N/n$  as close as one likes to 1, and such that every vector  $x \in \mathbb{C}^n$  has a Kashin representation;

Yu. Lyubarskii is with Norwegian University of Science and Technology. He was partially supported by the Norwegian Research Council projects 160192/v30, 10323200.

R. Vershynin is with University of Michigan. He was partially supported by NSF grants DMS 0401032, 0652617, 0918623 and Alfred P. Sloan Foundation.

(b) such frames are those that satisfy a form of the uncertainty principle. More precisely, their matrices satisfy a weak version of the restricted isometry property introduced by Candes and Tao [8]. In particular, many natural random frames have this property;

(c) there is a fast algorithm which converts frame representation (I.1) into a Kashin representation of  $x$ .

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 distortion (e.g. due to scalar quantization) and losses (e.g. due to transmission failures).

The article is organized as follows. Section II introduces Kashin's representations, discusses their relation to convex geometry (Euclidean projections of the cube) and explains how one can use Kashin's representations for vector quantization. In Section III, we discuss the uncertainty principle for matrices and frames. Theorems 3.5 and 3.9 state that for frames that satisfy the uncertainty principle, every frame representation can be replaced by Kashin's representation. A robust algorithm is given to quickly convert frame into Kashin's representations. In Section IV, 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.

## II. KASHIN'S REPRESENTATIONS

### A. Frame representations.

A sequence  $(u_i)_{i=1}^N \subset \mathbb{C}^n$  is called a tight frame if it satisfies Parseval's identity

$$\|x\|_2^2 = \sum_{i=1}^N |\langle x, u_i \rangle|^2 \text{ for all } x \in \mathbb{C}^n. \quad (\text{II.1})$$

This definition differs by a constant normalization factor from one which is often used in the literature, but (II.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 the  $n \times N$  frame matrix  $U$  whose columns are  $u_i$ . The following properties are easily seen to be equivalent:

- 1)  $(u_i)_{i=1}^N$  is a tight frame for  $\mathbb{C}^n$ ;
- 2) every vector  $x \in \mathbb{C}^n$  admits frame representation (I.1);

- 3) the rows of the frame matrix  $U$  are orthonormal;
- 4)  $u_i = Ph_i$  for some orthonormal basis  $(h_i)_{i=1}^N$  of  $\mathbb{C}^N$ , where  $P$  is the orthogonal projection in  $\mathbb{C}^N$  onto  $\mathbb{C}^n$ .

When  $N > n$ , the tight frames are linearly dependent systems, so various coefficients  $a_i$  of the frame representation may carry common information about vector  $x \in \mathbb{C}^n$ . This makes frames withstand noise in coefficients better than orthonormal bases, see [11], [21], [10]. However, using frame representation (I.1) may not always be the best way to use the frame redundancy. Some coefficients  $a_i$  may be much bigger than 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.

### B. Kashin's representations

Consider a sequence  $(u_i)_{i=1}^N \subset \mathbb{C}^n$ . We say that the expansion

$$x = \sum_{i=1}^N a_i u_i, \quad \max_i |a_i| \leq \frac{K}{\sqrt{N}} \|x\|_2 \quad (\text{II.2})$$

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

Kashin's representation produce the smallest possible dynamic range of the coefficients, which is  $\sqrt{n}$  smaller than the dynamic range of the frame representations. This is the content of the following simple observation:

*Observation 2.1 (Optimality):* Let  $(u_i)_{i=1}^N$  be a tight frame in  $\mathbb{C}^n$ . Then:

(a) There exists a vector  $x \in \mathbb{C}^n$  for which the coefficients  $a_i = \langle x, u_i \rangle$  of the frame representation (I.1) satisfy

$$\max_i |a_i| \geq \sqrt{\frac{n}{N}} \|x\|_2.$$

(b) For every vector  $x \in \mathbb{C}^n$ , every representation of the form  $x = \sum_{i=1}^N a_i u_i$  satisfies

$$\max_i |a_i| \geq \frac{1}{\sqrt{N}} \|x\|_2.$$

**Proof.** (a) Since the tight frame satisfies  $\sum_{i=1}^N \|u_i\|_2^2 = n$ , one has  $\max_i \|u_i\|_2 \geq \sqrt{n/N}$ . From this part (a) follows.

(b) The correspondence between tight frames and orthonormal bases (property 4) above) yields Bessel's inequality  $\|x\|_2 \leq (\sum_{i=1}^N |a_i|^2)^{1/2}$ , from which part (b) follows. ■

Not every tight frame admits Kashin's representations with constant level  $K$ ; this is clear if one considers an orthonormal basis in  $\mathbb{C}^n$  repeated  $\sim N/n$  times and properly normalized. Nevertheless, some natural classes of frames do have this property.

We start with the following existence result.

*Theorem 2.2 (Existence):* There exist tight frames in  $\mathbb{C}^n$  with arbitrarily small redundancy  $\lambda = N/n > 1$ , and such that every vector  $x \in \mathbb{C}^n$  has Kashin's representation with level  $K$  that depends on  $\lambda$  only (not on  $n$  or  $N$ ).

**Proof.** This statement is in essence a reformulation of the classical result from geometric functional analysis due to Kashin [28] (with an optimal dependence  $K = O(\sqrt{\frac{1}{\lambda-1} \log \frac{1}{\lambda-1}})$  given later by Garnaev and Gluskin [17]). To see this, we shall look at Kashin's representations from the geometric viewpoint. Let  $Q^N = \{x : \|x\|_\infty \leq 1\}$  and  $B^n = \{x : \|x\|_2 \leq 1\}$  stand for the unit cube and unit Euclidean ball in  $\mathbb{C}^N$  and  $\mathbb{C}^n$  respectively. The observation below follows directly from the definition of Kashin's representations.

*Observation 2.3: (Kashin's representations and projections of the cube):* Consider a tight frame  $(u_i)_{i=1}^N$  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  $(u_i)_{i=1}^N$ ;
- (ii) The  $n \times N$  matrix  $U$  whose columns are  $u_i$  satisfies

$$B^n \subseteq \frac{K}{\sqrt{N}} U(Q^N). \quad (\text{II.3})$$

Inclusion (II.3) yields an equivalence

$$B^n \subseteq \frac{K}{\sqrt{N}} U(Q^N) \subseteq K B^n, \quad (\text{II.4})$$

the second inclusion holds trivially. Since the rows of the frame matrix  $U$  are orthonormal, the operator  $U : \mathbb{C}^N \rightarrow \mathbb{C}^n$  is unitarily equivalent to an orthogonal projection. We thus may say that  $U$  realizes *Euclidean projection of the cube*. We refer the reader to [34] Section 6 for more thorough discussion of this topic.

Kashin's theorem [28] 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 and the coefficient  $K$  depends on the redundancy  $\lambda = N/n$  only. In other words, there exists an  $n \times N$  matrix  $U$  whose rows are orthonormal and which satisfies (II.4).

The first inclusion in (II.4) 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 2.2. ■

In geometric functional analysis, many classes of matrices  $U$  are known to realize Euclidean projections of the cube as in (II.4). We discuss them in more details in Section IV-A. In fact we will see that random matrix  $U$  with orthonormal rows picked with respect to a rotationally invariant distribution satisfies (II.4) with high probability.

**Remark** Since the level  $K$  of Kashin's representation depends on redundancy only, this representation become especially efficient in high dimensions when when the factor  $\sqrt{n}$  in the expression for the dynamic range of the frame expansion overpowers the value of  $K$  (which ideally is  $O(\sqrt{\frac{1}{\lambda-1} \log \frac{1}{\lambda-1}})$ ). Therefore we are interested mainly in low redundant frames just in order to avoid getting too large volumes of information to be transmitted.

### C. Stability, vector quantization

Kashin's representations have 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  $K = O(1)$ . So we represent a vector  $x \in \mathbb{C}^n$ ,  $\|x\|_2 \leq 1$ , with its Kashin's coefficients  $(a_1, \dots, a_N) \in \mathbb{C}^N$ ,  $|a_i| \leq K/\sqrt{N}$ . Assume these coefficients are damaged (due to quantization, losses, flips of bits, etc.) and we only know noisy coefficients  $(\hat{a}_1, \dots, \hat{a}_N) \in \mathbb{C}^N$ . When we try to reconstruct  $x$  from these coefficients as  $\hat{x} = \sum_{i=1}^N \hat{a}_i u_i$ , the accuracy of this reconstruction is

$$\begin{aligned} \|x - \hat{x}\|_2 &= \left\| \sum_{i=1}^N (a_i - \hat{a}_i) u_i \right\|_2 \\ &\leq \left( \sum_{i=1}^N |a_i - \hat{a}_i|^2 \right)^{1/2}. \end{aligned} \quad (\text{II.5})$$

Combined with the fact that the coefficients  $a_i$  have the dynamic range  $O(1/\sqrt{N})$ , this yields greater robustness of Kashin's representations with respect to noise, and in particular to *quantization errors*. Suppose we need to quantize a vector  $x \in \mathbb{C}^n$ . We may do this by quantizing each coefficient  $a_i$  separately by performing a uniform scalar quantization of

the dynamic range  $[-K/\sqrt{N}, K/\sqrt{N}]$  with, say,  $L$  levels. The quantization error for each coefficient is thus  $|a_i - \hat{a}_i| \leq K/L\sqrt{N}$ . By (II.5), this produces the overall quantization error

$$\|x - \hat{x}\|_2 \leq K/L = O(1/L).$$

Similar quantization of frame representations (I.1) would only give the bound

$$\|x - \hat{x}\|_2 \leq \sqrt{n}/L$$

because its dynamic range is  $\sqrt{n}$  larger than that of Kashin's representations (by Observation 2.1).

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  $\delta N$  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  $|a_i| \leq K/\sqrt{N}$ , we can assume (by truncation) that all  $|\hat{a}_i| \leq K/\sqrt{N}$ . When we reconstruct  $x$  from these damaged coefficients (as before), the accuracy of this reconstruction can be estimated using (II.5) as

$$\|x - \hat{x}\|_2 \leq 2K\sqrt{\delta} = O(\sqrt{\delta}).$$

Thus the reconstruction error is small whenever the (related) number of damaged coefficients  $\delta$  is small.

By Theorem 2.2, the maximal error reduction effect is achieved using frames with only a *constant redundancy*, in fact any redundancy factor  $\lambda = N/n > 1$  has the error reduction power of maximal possible order. This is in contrast with traditional methods, in which increasing redundancy of the frame gradually reduces the representation error.

### III. COMPUTING KASHIN'S REPRESENTATIONS

Computing the coefficients  $a_i$  of Kashin's representation (II.2) of a given vector  $x$  can be described as a linear feasibility problem, which can be solved in (weakly) polynomial time using linear programming methods.

In this paper, we 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 Kashin's representation. This 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 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 has a potential to be implemented on analog devices. Followed by some robust scalar quantization of coefficients (such as one-bit  $\beta$ -quantization [13], [14]), this algorithm may be used for robust one-bit vector quantization schemes for analog-to-digital conversion.

#### A. The uncertainty principle

The classical uncertainty principle says that a function and its Fourier transform cannot be simultaneously well-localized. We refer the reader to fundamental monograph [25] for history survey and also for numerous realization of this heuristic rule. In particular a variant of the uncertainty principle due to Donoho and Stark [16] states that if  $f \in L_2(\mathbb{R})$  is "almost concentrated" on a measurable set  $T$  while its Fourier transform  $\hat{f}$  is "almost concentrated" on a measurable set  $\Omega$ , then then the product of measures  $|T||\Omega|$  admits a natural low bound. Donoho and Stark proposed applications of this principle for signal recovery [16].

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

In papers by Candes, Romberg and Tao [4], [7], [3] and by Rudelson and Vershynin [35], [36], a much stronger discrete uncertainty principle was established for *random* sets of size proportional to  $N$ . Moreover, one of these sets (say support of the signal in frequency domain) can be arbitrary (non-random), and the other (random support in time domain) can be almost the whole domain. The following result is a consequence of [35], [36]:

*Theorem 3.1 (uncertainty principle):* Let  $N = (1 + \mu)n$  for some integer  $n$  and  $\mu \in (0, 1)$ . Consider a random subset  $\Omega$  of  $\{0, \dots, N - 1\}$  of average cardinality  $n$ , which is obtained from independent random

$\{0, 1\}$ -valued variables  $\delta_0, \dots, \delta_{N-1}$  with  $\mathbb{E}\delta_i = n/N$  as  $\Omega := \{i : \delta_i = 1\}$ . Then  $\Omega$  satisfies the following with high probability. For every  $z \in \mathbb{C}^N$ ,

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

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

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

$$\|\hat{x} \cdot \mathbf{1}_\Omega\|_2 \leq (1 - c\mu)\|x\|_2, \quad (\text{III.1})$$

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

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

We can regard inequality (III.1) as a property of the partial Fourier matrix  $U$ , which consists of the rows of the DFT (discrete Fourier transform) matrix  $\Phi$  indexed by the random set  $\Omega$ . Then (III.1) says that  $\|Ux\|_2 \leq (1 - c\mu)\|x\|_2$  for all vectors  $x \in \mathbb{C}^N$  such that  $|\text{supp}(x)| \leq \delta N$ .

Now we can abstract from the harmonic analysis in question and introduce a general uncertainty principle (UP) as a property of matrices.

*Definition 3.2 (UP for matrices):* An  $n \times N$  matrix  $U$  satisfies the uncertainty principle with parameters  $\eta, \delta \in (0, 1)$  if, for  $x \in \mathbb{C}^N$ ,

$$|\text{supp}(x)| \leq \delta N \quad \text{implies} \quad \|Ux\|_2 \leq \eta\|x\|_2. \quad (\text{III.2})$$

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 introduced by 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, for  $x \in \mathbb{C}^N$ , the condition  $|\text{supp}(x)| \leq \delta N$  implies

$$\lambda(1 - \varepsilon)\|x\|_2 \leq \|Ux\|_2 \leq \lambda(1 + \varepsilon)\|x\|_2.$$

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,  $|\text{supp}(x)| \leq \delta N$ . 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  $\varepsilon, \delta$  implies the uncertainty principle with parameters  $\eta = \frac{1+\varepsilon}{1-\varepsilon}\sqrt{\frac{n}{N}}$ ,  $\delta$ .

**Proof.** Since the columns  $u_i$  of the matrix  $U$  satisfy  $\sum_{i=1}^N \|u_i\|_2^2 = n$ , there exists a column with norm  $\|u_i\|_2 \leq \sqrt{n/N}$ . This column is a preimage of some 1-sparse unit vector  $e_i$ , i.e.  $u_i = Ue_i$  where  $e_i = (0, \dots, 0, 1, \dots, 0)$  with 1 on the  $i$ -th place. Using the UUP for  $x = e_i$  we obtain

$$\lambda(1 - \varepsilon) \leq \|u_i\|_2 \leq \sqrt{\frac{n}{N}}.$$

Hence  $\lambda \leq \frac{1}{1-\varepsilon}\sqrt{\frac{n}{N}}$ . In view of this estimate, the upper bound in the UUP reads as follows:  $|\text{supp}(x)| \leq \delta N$  implies

$$\|Ux\|_2 \leq \frac{1 + \varepsilon}{1 - \varepsilon} \sqrt{\frac{n}{N}} \|x\|_2.$$

This is what we wanted to prove.  $\blacksquare$

The uncertainty principle can be reformulated 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 (UP 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

$$\left\| \sum_{i \in \Omega} a_i u_i \right\|_2 \leq \eta \left( \sum_{i \in \Omega} |a_i|^2 \right)^{1/2} \quad (\text{III.3})$$

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

## B. Converting frame representations into Kashin's representations

For every tight frame that satisfies the uncertainty principle, one can convert frame representations into Kashin's representations.

The conversion procedure is natural and fast. We truncate the coefficients of the frame representation (I.1) of  $x$  at level  $M = \|x\|_2/\sqrt{\delta N}$  in hope to achieve a Kashin's representation with level  $K = 1/\sqrt{\delta}$ . However, 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  $\eta M$ . We continue this process of expansion, truncation and reconstruction, each time reducing the truncation level by the factor of  $\eta$ .

Using the uncertainty principle, we will be able to show that the norm of the residual reduces by the factor of  $\eta$  at each iteration. So we can compute Kashin's representations of level  $K = K(\eta, \delta)$  with accuracy  $\varepsilon$  in  $O(\log(1/\varepsilon))$  iterations. Our analysis of this algorithm will yield:

**Theorem 3.5 (frame to Kashin conversion):** Let  $(u_i)_{i=1}^N$  be a tight frame in  $\mathbb{C}^n$  which satisfies the Uncertainty Principle with parameters  $\eta, \delta$ . 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 this result, we introduce and study the truncation operator for frame representations. Given a number  $M > 0$ , the one-dimensional truncation at level  $M$  is defined for  $z \in \mathbb{C} \setminus \{0\}$  as

$$t_M(z) = \frac{z}{|z|} \min\{|z|, M\}, \quad (\text{III.4})$$

and  $t_M(0) = 0$ .

Consider a frame  $(u_i)_{i=1}^N$  satisfying the assumptions of the theorem. For every  $x \in \mathbb{C}^n$ , we consider the frame representation

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

and define the truncation operator on  $\mathbb{C}^n$  as

$$Tx = \sum_{i=1}^N \hat{b}_i u_i \text{ where } \hat{b}_i = t_M(b_i) \text{ and } M = \|x\|_2 / \sqrt{\delta N}. \quad (\text{III.5})$$

The uncertainty principle helps us to bound the residual of the truncation:

**Lemma 3.6 (Truncation):** In the above notations, for every vector  $x \in \mathbb{C}^n$  we have

$$\|x - Tx\|_2 \leq \eta \|x\|_2. \quad (\text{III.6})$$

**Proof.** Let  $x \in \mathbb{C}^n$ . Consider the subset  $\Omega \subseteq \{1, \dots, N\}$  defined as

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

By the definition of tight frame, we have

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

thus

$$|\Omega| \leq \|x\|_2^2 / M^2 = \delta N.$$

Using the uncertainty principle, we can estimate the norm of the residual

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

as

$$\begin{aligned} \|x - T_M x\|_2 &\leq \eta \left( \sum_{i \in \Omega} |b_i - \hat{b}_i|^2 \right)^{1/2} \leq \\ &\eta \left( \sum_{i \in \Omega} |b_i|^2 \right)^{1/2} \leq \eta \left( \sum_{i=1}^N |b_i|^2 \right)^{1/2} = \eta \|x\|_2. \end{aligned}$$

This completes the proof.  $\blacksquare$

**Proof of Theorem 3.5** Given  $x \in \mathbb{C}^n$ , for  $k = 1, 2, \dots$  we define the vectors

$$x^{(0)} := x, \quad x^{(k)} := x^{(k-1)} - T x^{(k-1)}.$$

Then, for each  $r = 0, 1, 2, \dots$  we have

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

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

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

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

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

$$\begin{aligned} \left\| x - \sum_{i=1}^N a_i u_i \right\|_2 &\leq \eta^r \|x\|_2, \\ \max_i |a_i| &\leq \frac{K}{\sqrt{N}} \|x\|_2. \end{aligned} \quad (\text{III.7})$$

Initialize the coefficients and the truncation level:

$$a_i \leftarrow 0, \quad i = 1, \dots, N; \quad M \leftarrow \frac{\|x\|_2}{\sqrt{\delta N}}.$$

Repeat the following  $r$  times:

- Compute the frame representation of  $x$  and truncate at level  $M$ :

$$b_i \leftarrow \langle x, u_i \rangle, \hat{b}_i \leftarrow t_M(b_i), i = 1, \dots, N.$$

- 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:

$$a_i \leftarrow a_i + \sqrt{N} \hat{b}_i, \quad i = 1, \dots, N;$$

$$M \leftarrow \eta M.$$

**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  $\eta^r \leq \frac{K}{\sqrt{N}}$ . 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 (III.4) may be impossible to realize 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 should 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 [24], [13], [14].

Our algorithm is also robust in the following sense: the exact truncation  $t_M$  can be replaced by any approximate truncation. Such an approximate truncation at level 1 can be

any function  $t(z) : \mathbb{C} \rightarrow \mathbb{C}$  which satisfies for some  $\nu, \tau \in (0, 1)$ :

$$|z - t(z)| \leq \begin{cases} \nu|z| & \text{if } |z| \leq \tau, \\ |z| & \text{for all } z, \end{cases} \quad (\text{III.8})$$

and  $|t(z)| \leq 1$  for all  $z$ .

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

*Theorem 3.7: (Approximate truncation)* The above algorithm remains valid if one replaces exact truncation by any approximate one and also adjust the parameters: level  $M$  should be replaced with  $M' = \tau^{-1}M$ , parameter  $\eta$  should be replaced with  $\eta' = \sqrt{\eta^2 + \nu^2}$ , finally, level  $K$  is replaced with  $K' = \tau^{-1}(1 - \eta')^{-1}\delta^{-1/2}$ , provided that  $\eta' < 1$ .

Moreover, the approximate truncation can be different each time it is called by the algorithm, provided that it satisfies (III.8). This facilitates the algorithm implementation on analog devices. In particular, one can use this algorithm to build robust vector quantizers for analog-to-digital conversion.

**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  $\varepsilon$ -tight frame if

$$(1 - \varepsilon) \|x\|_2 \leq \left( \sum_{i=1}^N |\langle x, u_i \rangle|^2 \right)^{1/2}$$

$$\leq (1 + \varepsilon) \|x\|_2 \quad \text{for all } x \in \mathbb{C}^n. \quad (\text{III.9})$$

An analysis similar to that above yields:

*Theorem 3.9:* Let  $(u_i)_{i=1}^N \subset \mathbb{C}^n$  be an  $\varepsilon$ -tight frame, which satisfies the uncertainty principle with parameters  $\eta$  and  $\delta$ . Then Theorem 3.5 and the algorithm above are valid for  $M$  replaced with  $M' = \sqrt{1 + \varepsilon} M$  and  $\eta$  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  $\ell_1^n$  (which by duality is equivalent to Euclidean projections (II.3) of a cube). A similar idea was essentially used by Talagrand in his work on the  $\Lambda_1$  problem [38].

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 it serves to *find* all big coefficients rather than to spread them out.

#### IV. MATRICES AND FRAMES THAT SATISFY THE UNCERTAINTY PRINCIPLE

In this section, we give examples of matrices (equivalently, frames) that satisfy the uncertainty principle. By Observation 2.3, such  $n \times N$  matrices  $U$  realize Euclidean projection of the cube (II.3). Equivalently, these frames  $(u_i)_{i=1}^N$  (the columns of  $U$ ) yield quickly computable Kashin's representations for every vector  $x \in \mathbb{C}^n$ .

##### A. Matrices known to realize Euclidean projections of the cube

Much attention has been paid to Euclidean projections of the cube (II.3) in geometric functional analysis. Results in the literature are usually stated in the dual form, about  $n$ -dimensional Euclidean subspaces of  $\ell_1^N$ .

Kashin proved (II.3) for 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 ([28], see also [34] Section 6.) The level  $K$  (II.3) depends only on  $\lambda$ ; an optimal dependence was given later by Garnaeu and Gluskin [17]).

A similar result holds for  $U = \frac{1}{\sqrt{N}}\Phi$ , where  $\Phi$  is a random Bernoulli matrix, which means that the entries of  $\Phi$  are  $\pm 1$  symmetric independent random variables. Schechtman [37] first proved this with  $N = O(n)$ , and in [31] this result is generalized for  $N = \lambda n$  with arbitrary  $\lambda > 1$ . The dependence  $K$  on  $\lambda$  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).

It is unknown whether Kashin's theorem holds for partial Fourier matrix; this conjecture is known as the  $\Lambda_1$  problem. Consider the Discrete Fourier Transform in  $\mathbb{C}^N$ , where  $N = O(n)$ , given by the orthogonal  $N \times N$  matrix  $\Phi$ . It is unknown whether there exists a submatrix  $U$  which consists of some  $n$  rows of  $\Phi$  and such that it realizes an Euclidean projection of the cube in the sense of (II.3).

In the positive direction, a partial result due to Bourgain, later reproved by Talagrand with a general method [38], states that a random partial Fourier matrix  $U$  satisfies (II.4) with high probability for  $N = O(n)$  and

$K = O(\sqrt{\log(N) \log \log(N)})$ . It was recently proved in [23] that Bourgain's result holds for 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 2.3), which is a consequence of the uncertainty principle in [35], [36].

No explicit constructions of matrices  $U$  are known. However, there exists small space constructions that use a small number of random bits [2], [26], [27].

##### B. 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

$$\mathcal{O}(n \times N) = \{P_n V; V \in \mathcal{O}(N)\}, \quad (\text{IV.1})$$

where  $P_n : \mathbb{C}^N \rightarrow \mathbb{C}^n$  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

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

where  $c > 0$  is an absolute constant.

**Remark.** Assumption  $\mu > 0$  is not essential; just expressions for  $\eta$  and  $\delta$  will look differently. We are most interested in small values of  $\mu$  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  $\varepsilon$ -net argument. Denote by  $S^{N-1}$  and  $\sigma_{N-1}$  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

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

where  $c_1 > 0$  is an absolute constant.

**Proof.** We use representation (IV.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

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

The map  $z \mapsto \|P_n z\|$  is a 1-Lipschitz function on  $S^{N-1}$ . The concentration inequality (see e.g. [29] Section 1.3) then implies that this function is well concentrated about its average value  $E$ :

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

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

**Proof of Theorem 4.1.** Assume that  $\eta$  and  $\delta$  satisfy the assumptions (IV.2). We have to prove that (III.2) holds with probability at least  $1 - \exp(-c\mu^2 n)$ .

Consider the set

$$S := \{x \in S^{N-1}, |\text{supp}(x)| \leq \delta N\}.$$

We have

$$S = \bigcup_{|I| \leq \delta N} S_I,$$

here the union is taken over all subsets  $I$  of  $\{1, \dots, N\}$  of cardinality at most  $\delta N$ , 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  $\varepsilon$ -net of  $S_I$  in the Euclidean norm, and of cardinality at most  $(3/\varepsilon)^{\delta N}$  (see e.g. [34] 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 exists an  $\varepsilon$ -net  $\mathcal{N}$  of  $S$  of cardinality

$$|\mathcal{N}| \leq \binom{N}{\lceil \delta N \rceil} \left(\frac{3}{\varepsilon}\right)^{\delta N} \leq \left(\frac{3e}{\varepsilon \delta}\right)^{\delta N}.$$

Then using Lemma 4.2, we obtain

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

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

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

Therefore

$$\mathbb{P}\{\exists x \in S : \|Ux\|_2 > (1+t)\sqrt{\frac{n}{N}} + \varepsilon\} \leq |\mathcal{N}| \cdot 2 \exp(-c_1 t^2 n). \quad (\text{IV.3})$$

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  $\eta$  in (IV.2), we have

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

Also, we can estimate the probability in (IV.3) as

$$|\mathcal{N}| \cdot 2 \exp(-ct^2 n) \leq \left(\frac{24e}{\delta\mu}\right)^{\delta N} \cdot 2 \exp(-c_2 t^2 n) \leq 2 \exp[(2\delta \log(24e/\delta\mu) - c_2 \mu^2)n], \quad (\text{IV.4})$$

where  $c_2 = c_1/25$ . By our choice of  $\delta$ , the right hand side of (IV.4) is bounded by  $2 \exp(-c\mu^2 n)$ , where  $c > 0$  is an absolute constant. We conclude that

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

This completes the proof.  $\blacksquare$

### C. 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  $\Phi$  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 random partial Fourier matrices, multiplication by such matrix corresponds to sampling  $n$  random frequencies of a signal.

More precisely, we select rows of  $\Phi$  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  $\Phi$  be an orthogonal  $N \times N$  matrix with uniformly bounded entries:  $|\Phi_{ij}| \leq \alpha N^{-1/2}$  for some constant  $\alpha$  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  $\Phi$  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 established in [7] and improved in [35], [36]. For an operator  $U$  on a Euclidean space,  $\|\cdot\|$  will denote its operator norm.

*Theorem 4.4: (UUP for partial Fourier matrices [35], [36])*

Assume the hypothesis of Theorem 4.3 is 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

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

Then the random submatrix  $U$  satisfies:

$$\mathbb{E} \sup_{|T| \leq r} \|id_T - \frac{N}{n} U_T^* U_T\| \leq \varepsilon. \quad (\text{IV.5})$$

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

$$\begin{aligned} \|id - A^* A\| &= \sup_{x \in \mathbb{C}^N, \|x\|_2=1} |\langle (id - A^* A)x, x \rangle| \\ &= \sup_{x \in \mathbb{C}^N, \|x\|_2=1} \left| \|Ax\|_2^2 - \|x\|_2^2 \right|. \end{aligned} \quad (\text{IV.6})$$

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

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

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

$$\left| \frac{N}{n} \|Ux\|_2^2 - 1 \right| < \varepsilon/p$$

for all  $x \in \mathbb{C}^N$ ,  $|\text{supp}(x)| \leq r$ ,  $\|x\|_2 = 1$ .

In particular, for such  $U$ , one has:  $\|Ux\|_2 \leq$

$$\sqrt{1 + \varepsilon/p} \sqrt{\frac{n}{N}} \|x\|_2 = \sqrt{\frac{1 + \varepsilon/p}{1 + \mu}} \|x\|_2$$

for all  $x \in \mathbb{C}^N$ ,  $|\text{supp}(x)| \leq r$ .

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 (III.2) with  $\delta = r/N$ . To estimate  $\delta$  we note that

the condition on  $n$  in Theorem 4.4 is satisfied if

$$r \leq \frac{c_1 \varepsilon^2 N}{\log^4 N} \quad (\text{IV.7})$$

where  $c_1 = c_1(\alpha) > 0$ . Since we have set  $\varepsilon = cp\mu$ , condition (IV.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.  $\blacksquare$

#### Remarks.

1. (*Computing in almost linear time*). The Fourier matrices can be used to compute Kashin's representations in  $\mathbb{C}^n$  in 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 III-B, we can convert a frame representation of every vector  $x \in \mathbb{C}^n$  into a Kashin's representation with level  $K = O(\log^2 n)$  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  $\alpha$  and polylogarithmically on  $p$ . The polynomial dependence on  $\alpha$  is straightforward from the proof of Theorem 4.4 in [35], [36]. 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 [36] Theorem 3.9, instead of the expectation estimate (IV.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  $\mu > 0$ .

#### D. Subgaussian random matrices.

A large family of matrices with independent random entries satisfies the uncertainty principle.

*Definition 4.5:* A random variable  $\phi$  is called *subgaussian with parameter  $\beta$*  if

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

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

*Theorem 4.6: (UP for random subgaussian matrices)*

Let  $\Phi$  be a  $n \times N$  matrix whose entries are independent mean zero subgaussian random

variables with parameter  $\beta$ . 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

$$\eta = C\beta\sqrt{\frac{\log \lambda}{\lambda}}, \quad \delta = \frac{c}{\lambda}, \quad (\text{IV.8})$$

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

**Remark.** Theorem 4.6 and Lemma 4.8 below can be deduced from the recent works [32], [33]. 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:* ([30] Fact 2.4) Let  $n \geq k$  and  $\Phi$  be a  $n \times k$  matrix whose entries are independent mean zero subgaussian random variables with parameter  $\beta$ . Then

$$\mathbb{P}\{\|\Phi\| > t\sqrt{n}\} \leq \exp(-c_1 n t^2 / \beta^2) \quad (\text{IV.9})$$

for all  $t \geq C_1\beta$ , here  $C_1, c_1 > 0$  are absolute constants.

**Proof of Theorem 4.6.** The uncertainty principle for the matrix  $U$  with parameters  $\eta, \delta$  is equivalent to the following norm estimate:

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

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

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

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

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

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

provided we choose the absolute constant  $C$  sufficiently big. This means that the uncertainty principle with parameters (IV.8) fails with probability at most  $\lambda^{-n}$ . ■

Unlike random orthogonal or partial Fourier matrices considered in Sections IV-B and IV-C, subgaussian matrices do not in general have orthonormal rows. Nevertheless, the 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 compute Kashin's representations for such almost tight frames.

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

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

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

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

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

**Remark.** The dependence in  $C(\beta)$ ,  $c(\beta)$  is polynomial. Explicit bounds can be deduced from [33].

As a straightforward consequence, we obtain:

*Corollary 4.9:* (Subgaussian frames are almost tight) Let  $\Phi$  be a subgaussian matrix as in Lemma 4.8. Then the columns of the matrix  $\frac{1}{\sqrt{N}}\Phi$  form an  $\varepsilon$ -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 (IV.6),

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

Denote the columns of  $\Phi$  by  $\phi_i$ . Fix a vector  $x \in \mathbb{C}^n$ ,  $\|x\|_2 = 1$ . Since the entries of the vector  $\phi_i$  are subgaussian with parameter  $\beta$ , 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 [30]). Moreover, this random variable has mean zero and variance 1. We can use Bernstein's inequality (see [39]) to control the average of the independent mean zero random variables  $|\langle \phi_i, x \rangle|^2 - 1$  as  $\mathbb{P}\left\{\left|\frac{1}{N} \|\Phi^*x\|_2^2 - 1\right| > u\right\} =$

$$\mathbb{P}\left\{\left|\frac{1}{N} \sum_{i=1}^N |\langle \phi_i, x \rangle|^2 - 1\right| > u\right\} \leq$$

$$2\exp(-c_1 N u^2 / \beta^4)$$

for all  $u \leq c\beta$ , 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  $|\mathcal{N}| \leq (3/u)^n$  (see e.g. [34] Lemma 4.16). Using the probability estimate above, we can take the union bound to estimate the probability of the event

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

as

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

Applying Lemma 4.7 with  $t = C_1\beta$ , we see that the event  $B := \{\|U^*\| \leq C_1\beta\}$  satisfies  $\mathbb{P}(B^c) \leq \exp(-c_2 N)$ . 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  $\|x - y\|_2 \leq u$ , which implies by the triangle inequality that

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

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  $\left| \|U^*x\|_2^2 - 1 \right| < \varepsilon$  for all  $x \in S^{n-1}$  once the event  $A \cap B$  holds. Thus

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

by our choice of  $u$  and by the assumption on  $N$ . ■

#### ACKNOWLEDGMENT

This project started in 2003 when the first author visited University of California, Davis. The second author thanks the participants of Time-Frequency Brown Bag Seminar at Princeton for their valuable comments, and especially Ingrid Daubechies for her encouragement and interest in this work. Both authors thank the referees for their many suggestions that helped to greatly improve the presentation.

#### REFERENCES

- [1] S. Artstein-Avidan, O. Friedland, V. Milman, S. Sodin, *Polynomial bounds for large Bernoulli sections of the cross polytope*, Israel J. Math 156 (2006), 141-151
- [2] S. Artstein-Avidan, V. D. Milman, *Logarithmic reduction of the level of randomness in some probabilistic geometric constructions*, Journal of Functional Analysis 235 (2006), 297-329
- [3] E. J. Candés, J. Romberg, *Quantitative robust uncertainty principles and optimally sparse decompositions*, Found. Comput. Math. 6 (2006), 227-254
- [4] E. J. Candés, J. Romberg, T. Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory 52 (2006), 489-509
- [5] E. J. Candés, J. Romberg, T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math. 59 (2006), 1207-1223
- [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. Theory 52 (2006), 5406-5425
- [8] E. J. Candés, T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory 51 (2005), 4203-4215
- [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*. CBMS-NSF Regional Conference Series in Applied Mathematics, 61. Society for Industrial and Applied Mathematics (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*, IEEE Trans. Inform. Theory 53 (2006), 874-885
- [15] I. Daubechies, O. Yılmaz, *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, *One sketch for all: fast algorithms for compressed sensing*, STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing, 237-246, ACM, New York, 2007
- [19] E. D. Gluskin, Unpublished communications
- [20] V. K. Goyal, J. Kovacevic, J. A. Kelner, *Quantized frame expansions with erasures*, Applied and Computational Harmonic Analysis 10 (2001), 203-233
- [21] 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
- [22] R.M.Gray, D.L.Neuhoff, *Quantization*, IEEE Trans. Inform. Theory 44 (1998), 2325-2383
- [23] O. Guedon, S. Mendelson, A. Pajor, N. Tomczak-Jaegermann, *Subspaces and orthogonal decompositions generated by bounded orthogonal systems*, Positivity 11 (2007), 269-283

- [24] S. Güntürk, *One-bit sigma-delta quantization with exponential accuracy*, Communications on Pure and Applied Mathematics 56 (2003), 1608-1630
- [25] 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
- [26] P. Indyk, *Stable distributions, pseudorandom generators, embeddings and data stream computation*, 41st Symposium on Foundations of Computer Science, 2000
- [27] P. Indyk, *Uncertainty principles, extractors, and explicit embeddings of  $L_2$  into  $L_1$* , Electronic Colloquium on Computational Complexity, Report No. 126 (2006)
- [28] B. Kashin, *Sections of some finite dimensional sets and classes of smooth functions*, Izv. Acad. Nauk SSSR 41 (1977), 334-351 (Russian)
- [29] M. Ledoux, *The concentration of measure phenomenon*, AMS Math. Surveys and Monographs, Vol. 89, 2001
- [30] 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
- [31] 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
- [32] S. Mendelson, A. Pajor, N. Tomczak-Jaegermann, *Reconstruction and subgaussian operators in asymptotic geometric analysis*, Geom. Funct. Anal. 17 (2007), 1248–1282
- [33] S. Mendelson, A. Pajor, N. Tomczak-Jaegermann, *Uniform uncertainty principle for Bernoulli and subgaussian ensembles*, Constr. Approx. 28 (2008), 277–289
- [34] G. Pisier, *The volume of convex bodies and Banach space geometry*, Cambridge Tracts in Mathematics 94, Cambridge University Press, 1989
- [35] M. Rudelson, R. Vershynin, *Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements*, CISS 2006 (40th Annual Conference on Information Sciences and Systems).
- [36] M. Rudelson, R. Vershynin, *On sparse reconstruction from Fourier and Gaussian measurements*, submitted
- [37] G. Schechtman, *Random embeddings of Euclidean spaces in sequence spaces*, Israel J. Math. 40 (1981), 187–192
- [38] M. Talagrand, *Selecting a proportion of characters*, Israel J. Math. 108 (1998), 173–191
- [39] A.W. van der Vaart, J. A. Wellner, *Weak convergence and empirical processes*, Springer Series in Statistics. Springer-Verlag, New York, 1996

Yurii Lyubarskii earned the Ph.D. in mathematics from the Institute of Low Temperature Physics and Engineering in Kharkov, Ukraine, in 1974, and D.Sci. in Mathematics from the Leningrad Branch of Steklov Mathematical Institute in 1990.

During 1971-2001 he was a researcher and a leading researcher at Institute of Low Temperature Physics and Engineering. Since 1996 he is a Professor of Mathematics at Norwegian University of Science and Technology in Trondheim, Norway. His research interests include complex analysis, harmonic analysis, and applications to signal processing.

Roman Vershynin earned the Ph.D. degree in mathematics from the University of Missouri-Columbia in 2000.

He was a postdoctoral researcher at Weizmann Institute of Science, Israel, in 2000-2001, and a PIMS postdoctoral researcher at University of Alberta, Canada, in 2001-2003. He was an Assistant and later Associate Professor of Mathematics at University of California, Davis in 2003-2008. Since 2008 he is a Professor of Mathematics at University of Michigan. His research interests include geometric functional analysis, convex geometry, probability theory, and numerical algorithms.