Uncertainty Principles, Frames, and Quantization

Roman Vershynin   Yurii Lyubarskii

University of California, Davis

Norwegian Institute of Science and Technology

Trends in Harmonic Analysis
Strobl, Austria
June 2007

The Uncertainty Principle

- This talk is about new applications of the Uncertainty Principle.
- Even the fact that the U.P. has *any* applications may be surprising: U.P. is a negative statement (no function can be localized in time and frequency).
- [Donoho-Stark, 1989] turned U.P. into a positive tool for signal reconstruction.
- Recent progress in Compressed Sensing confirmed the usefulness of U.P. for sparse representations of signals.
- This talk: new application of U.P. with the *opposite* goal: for spread representations of signals.
- Will be explained this in the language of frames:

Frame representations

Definition (Frames)

A (Parseval) frame is a system of vectors (u_i) in a Hilbert space, which satisfies Parseval's identity: for every vector x,

$$\|x\|^2 = \sum_i |\langle u_i, x\rangle|^2.$$

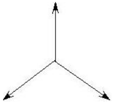
Equivalently, every vector x has the frame representation

$$x = \sum_i \langle u_i, x\rangle u_i.$$

- *Orthonormal bases* satisfy this definition.
  So what makes general frames *different* from orthonormal bases?

Frame representations

- **Answer:** Redundancy.
  Frames need not be linearly independent systems.



- Thus the frame representation

$$x = \sum_i b_i u_i, \quad b_i = \langle u_i, x \rangle$$

  is not unique; one can find other representations

$$x = \sum_i a_i u_i, \quad a_i \neq \langle u_i, x \rangle.$$

- Strangely, we *do not know* any other canonical representation – easy to define and to compute, useful.
- **Why should we care?**

Why non-frame representations?

- **Answer (methodological).**
  Frames are effective precisely because of their redundancy.
  The information about the source x gets *spread* across the frame coefficients. Redundancy $\Rightarrow$ several frame coefficients share common information about x.
- This explains the *effectiveness* of frames (e.g. their stability – resistance to losses/corruption of frame coefficients).
- We should know how to use redundancy directly, by finding useful (non-frame) representations.
- **Answer (practical).**
  Several communities have already been looking for non-frame representations:

Why non-frame representations?

One can pursue two opposite goals in a representation $x = \sum a_i u_i$:

- Sparse representations.
  Most of the coefficients $a_i = 0$.
  (*Sparse Approximation Theory*, *Compressed Sensing*.)
- Democratic representations.
  All coefficients a_i carry an even share of information about x.
  (*Source Coding*).
  Loss of 1% of coefficients $\Rightarrow$ reconstruction error should be 1%.
  Pushes the idea of redundancy to its limit.
- Both types of representations are *non-linear*, in contrast to frame representations.
- These two goals (sparse vs. democratic) are *opposite*.
  However, the main message of this talk: both are based on *the same principle* – the Uncertainty Principle.
- For *sparse representations*, this connection was developed in Compressed Sensing by [Donoho-Stark], [Candes-Romberg-Tao].
  For *democratic representations*, this talk.

Toward democratic representations

- **Goal.** Given a frame $(u_i)_{i=1}^N$ in $\mathbb{R}^n$, *define and compute a democratic representation* of every vector $x \in \mathbb{R}^n$:

$$x = \sum_{i=1}^{N} a_i u_i.$$

- **Democracy:** all coefficients a_i should contain the same (small) amount of information about x. [Calderbank-Daubechies].
- Ideally, $a_i = \pm C$ for some common C, i.e. *one bit* per coefficient. In this case, the representation has to be *approximate*. **Problem**: find such approximations with best accuracy.
- In any case, in democratic representations all a_i should have the same (small) order of magnitude.
- A simple argument shows that this magnitude has to be $\frac{1}{\sqrt{N}}$:

Toward democratic representations

- For any representation

$$x = \sum_{i=1}^{N} a_i u_i,$$

reverse Bessel's inequality holds:

$$\sum_{i=1}^{N} |a_i|^2 \geq \|x\|^2.$$

- So, if all a_i are of the same order, then

$$|a_i| \sim \frac{\|x\|}{\sqrt{N}}.$$

- A representation with coefficients of this (smallest possible) order we shall call Kashin's representation.

Kashin's representations

Definition (Kashin's representations)

Let $(u_i)_{i=1}^N$ in be a frame in $\mathbb{R}^n$, and $x \in \mathbb{R}^n$ be a vector.
A Kashin's representation of x with level K is

$$x = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} a_i u_i, \quad \text{where all } |a_i| \le K \|x\|.$$

- In other words, a Kashin's representations is a representation with coefficients of the *smallest possible order*, i.e. $O(\frac{1}{\sqrt{N}})$.
- **Do Kashin's representations exist?**
  (with constant level K and constant redundancy N/n)
- **Answer**. Yes:

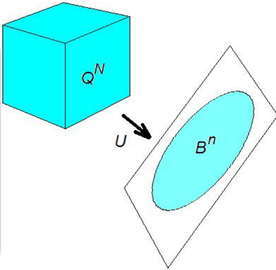
Kashin's representations

Theorem (Existence of Kashin's representations)

For every dimension n and arbitrary redundancy $\lambda > 1$,
there exists a frame $(u_i)_{i=1}^N$ in $\mathbb{R}^n$ with $N/n \leq \lambda$ and such that
every vector in $\mathbb{R}^n$ has a Kashin's representation with level $K = K(\lambda)$.

- Random frames satisfy this theorem (e.g. Gaussian).
- This theorem is an interpretation of a classical result of [Kashin, 1977] in Geometric Functional Analysis:
  "a random projection of a cube looks like a round ball":

$Q^N := \{x \in \mathbb{R}^N : \|x\|_\infty \le 1\} =$ unit cube;
$B^n := \{x :\in \mathbb{R}^n : \|x\|_2 \le 1\} =$ unit Euclidean ball.



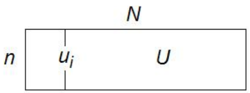
Theorem (Kashin)

There exists a (random) orthogonal projection $U : \mathbb{R}^N \to \mathbb{R}^n$ such that

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

where $K = K(\lambda)$ depends only on the redundancy $\lambda = N/n$.

Let $(u_i)_{i=1}^N$ be the columns of U.
They form a frame (U is an orthogonal projection).



Rewrite the first inclusion for $x \in B^n$ as

$$x = \frac{K}{\sqrt{N}} \sum_{i=1}^N b_i u_i \qquad \text{for some } |b_i| \le 1.$$

This is precisely a Kashin's representation of x. □

Kashin's representations

- **Conclusion.** There exists frames $(u_i)_{i=1}^N$ in every dimension n and with arbitrary small redundancy, such that every vector has a Kashin's representation

$$x = \frac{1}{\sqrt{N}} \sum_1^N b_i u_i, \qquad b_i = O(1).$$

- One can even achieve $b_i = \pm 1$, thus a *one-bit representation*. (Take binary expansion of each b_i; yields approx. representation)
- Thus, *democratic representations do exist*.
- Surprisingly, arbitrary redundancy $\lambda = N/n > 1$ suffices. **Qualitative phenomenon**: Kashin's representations emerge abruptly when we go from bases ($\lambda = 1$) to frames ($\lambda > 1$).
- Redundancy needs not be logarithmic as in sparse representation problems (Compressed Sensing).
- **Question.** Where Kashin's representations can be better than frame representations? **Example.** Source coding:

Example: source coding

- **Source coding.** We need to encode a signal (source) $x \in \mathbb{R}^n$. E.g. internet coding, Analog-to-Digital conversion (audio $\to$ CD).
- **Coding using frames**: use frame coefficients (Bell Labs). We fix a frame $(u_i)_{i=1}^N$ in $\mathbb{R}^n$ and write out the frame representation

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

Coding scheme: $x \mapsto (b_1, \ldots, b_N)$. Thus $\mathbb{R}^n \to \mathbb{R}^N$.

- **Typical observation**: coding with *frames* is more robust than with *bases*, more resistant to errors (losses, bursts, quantization).
- **Reason**: redundancy ($N > n$). Different frame coefficients b_i contain *overlapping information* about x. Thus the information is *protected* when some b_i get corrupted. Let us see how:

Example: source coding

- **Source coding.** We have a source $x \in \mathbb{R}^n$, normalized for convenience: $\|x\| = 1$. We write out its frame representation $x = \sum b_i u_i$ with the frame coefficients $b_i = \langle u_i, x \rangle$. Encode $x \to (b_1, \ldots, b_N)$.
- **Magnitude of the coefficients.** $|b_i| \leq 1$; some may be ~ 1. This determines the complexity (const. bits per coefficient).
- **Errors** corrupt the frame coefficients: $b_i \to \hat{b}_i$. We reconstruct using corrupted coefficients: $x = \sum \hat{b}_i u_i$. Our hope is $\hat{x} \approx x$.
- **Uniform errors** (e.g. quantization): $|b_i - \hat{b}_i| \sim \varepsilon \ll 1$. Accuracy of reconstruction:

$$\|x - \hat{x}\| = \Big\| \sum_{1}^{N} (b_i - \hat{b}_i) u_i \Big\| \leq \Big(\sum_{1}^{N} |b_i - \hat{b}_i|^2 \Big)^{1/2} \sim \varepsilon \sqrt{N}.$$

  Grows with the dimension.
- **Bursty errors**: 1% of the b_i's gets completely corrupted. Reconstruction is impossible if some large coefficient is destroyed (e.g. $|b_i| \sim 1$).

Example: source coding

- Now **replace frame representations by Kahin's** and see how the coding quality changes.
- For a unit source $x \in \mathbb{R}^n$, we write out its Kashin's representation $x = \frac{1}{\sqrt{N}} \sum_1^N a_i u_i$ with $|a_i| \leq K$. Encode $x \to (a_1, \ldots, a_N)$.
- **Errors** corrupt the frame coefficients: $a_i \to \hat{a}_i$. We reconstruct using corrupted coefficients: $x = \frac{1}{\sqrt{N}} \sum_1^N \hat{a}_i u_i$. Our hope is $\hat{x} \approx x$.
- **Accuracy of reconstruction**:

$$\|x - \hat{x}\| = \Big\| \frac{1}{\sqrt{N}} \sum_1^N (a_i - \hat{a}_i) u_i \Big\| \leq \Big(\frac{1}{N} \sum_1^N |a_i - \hat{a}_i|^2 \Big)^{1/2}.$$

- **Uniform errors** (e.g. quantization): $|a_i - \hat{a}_i| \sim \varepsilon \ll 1$.
  Then $\|x - \hat{x}\| \leq \varepsilon$. *Does not grow with the dimension!*
  (Recall: for frame representations, we had $\varepsilon\sqrt{N}$.)
- **Bursty errors**: 1% of the a_i's gets completely corrupted.
  We can a-priori assume that $|\hat{a}_i| \leq K$. Then $\|x - \hat{x}\| \lesssim \sqrt{0.01} \cdot 2K$.
  (Recall: for frame representations, reconstruction was impossible).

Conclusion

- Frames are more robust than bases, because of redundancy.
- Kashin's representations are more robust than frame represen's, because they push the idea of redundancy to its limit:
- Kashin's representations are "maximally democratic". Information is spread uniformly over the coefficients.
- Kashin's representations exist. (There exist frames with arbitrarily small redundancy that yield Kashin's representations).
- **Question.** *How to compute* Kashin's representations?

How to compute Kashin's representations?

- Computing the coefficients of a Kashin's representation

$$x = \frac{1}{\sqrt{N}} \sum_{1}^{N} a_i u_i, \qquad |a_i| \le K,$$

  is a linear feasibility problem
  (linear equation subject to linear constraints $-K \le a_i \le K$).
  Thus can be computed by Linear Programming.
- Will find a much faster way to compute Kashin's representations – roughly as fast as computing *frame* representations.
- Rather than computing from scratch, we will try to convert a frame representation into a Kashin's representation.
- **Problem.** For what frames is a conversion possible?
- **Answer.** Whenever the frame satisfies an Uncertainty Principle (i.e. its matrix satisfies the Restricted Isometry Property in compressed sensing).

Uncertainty Principles

- **Abstract form.** No function is localized in time *and* frequency.
- **In $L^2(\mathbb{R})$.** If f is ε-concentrated on a set T (i.e. $\|f|_{T^c}\| \le \varepsilon\|f\|$) and $\hat{f}$ is δ-concentrated on a set W, then

$$|T||W| > (1-\varepsilon-\delta)^2.$$

  [see Donoho-Stark, 1989]. [Nazarov] – above 1.
- **In discrete spaces.** For $f \in \mathbb{C}^N$,

$$|\operatorname{supp}(f)||\operatorname{supp}(\hat{f})| \ge N.$$

  The equality case is possible. Both x and $\hat{x}$ can be supported by an arithmetic progression of length $\sqrt{N}$) ("spike trains").
- Thus, there is **no discrete U.P.** (for arbitrary sets of size $\sim N$.)
- The situation is much better if one of the sets is random [Candes-Romberg-Tao], [Rudelson-Vershynin]:

Uncertainty Principles

Theorem (Uncertainty Principle [Rudelson-V.)

Choose a subset $\Omega \subset \{1, \ldots, N\}$ *of size* $|\Omega| = (1-\mu)N$ *uniformly at random. Then there is no function* $x \in \mathbb{C}^N$ *for which*

$$\mathrm{supp}(\hat{x}) \subset \Omega \quad \textit{and} \quad |\,\mathrm{supp}(x)| \le \delta N,$$

where $\delta = c\mu^2 \log^{-2} N$.

This theorem follows easily from its quantitative version:

Theorem (Quantitative Uncertainty Principle)

For the random set Ω *as above, and for every sparse function* $x \in \mathbb{C}^N$, $|\,\mathrm{supp}(x)| \le \delta N$, *we have*

$$\|\hat{x}|_\Omega\| \le (1 - c\mu)\|x\|. \tag{UP}$$

Indeed, if *both* $\mathrm{supp}(\hat{x}) \subset \Omega$ and $|\,\mathrm{supp}(x)| \le \delta N$, then (UP) reads as $\|\hat{x}\| \le (1 - c\mu)\|x\| < \|x\|$. But this contradicts Plancherel's identity.

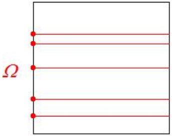
Uncertainty Principles

So, the Quantitative **Uncertainty Principle** says:

$$\|\hat{x}|_\Omega\| \le (1 - c\mu)\|x\| \qquad \text{for all sparse } x,\ |\operatorname{supp}(x)| \le \delta N. \qquad \text{(UP)}$$

(UP) is a property of the *partial Fourier matrix* – the submatrix of DFT with rows in Ω:

$$\|Ux\| \le (1 - c\mu)\|x\|.$$



We can now forget about Fourier Analysis, and think of this as an *abstract property* of a matrix U:

Definition (Uncertainty Principle for matrices)

An $n \times N$ matrix U with orthogonal rows satisfies the U.P. with constants (δ, η) if it *shrinks all sparse vectors*:

$$\|Ux\| \le (1 - \eta)\|x\| \qquad \text{for all sparse } x,\ |\operatorname{supp}(x)| \le \delta N.$$

This U.P. is slightly weaker than the **Restricted Isometry Condition** in Compressed Sensing (where the inequality goes both ways).

Uncertainty Principles

- **U.P. for matrices**:

$$\|Ux\| \le (1-\eta)\|x\| \qquad \text{for all sparse } x,\ |\operatorname{supp}(x)| \le \delta N.$$

U.P. means: U shrinks all sparse vectors.

- Equivalently, we can view U.P. as a property of frames – the columns $(u_i)_{i=1}^N$ of U.
- **U.P. for frames**:

$$\Big\| \sum_{i\in\Omega} b_i u_i \Big\| \le (1-\eta)\Big(\sum_{i\in\Omega} |b_i|^2 \Big)^{1/2} \quad \text{for all small } \Omega,\ |\Omega| \le \delta N.$$

U.P. means: strong reverse Bessel's inequality for sparse coeff's.

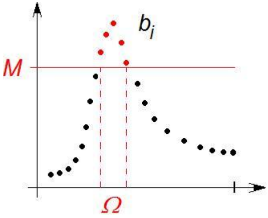
- **Known examples**: random Gaussian, Bernoulli, Fourier. Explicit constructions? (with constant redundancy, δ and η).
- **Main Observation**. If a frame satisfies U.P., then *every frame representation can be converted into a Kashin's representation*. *Spreading algorithm:*

Spreading Algorithm: Frame representations → Kashin's reprs.

Given a frame $(u_i)_{i=1}^N$ that satisfies U.P., and a unit vector x, converts the frame representation of x into Kashin's representation.

1. Expand. Write out the frame representation

$$x = \sum b_i u_i, \qquad b_i = \langle u_i, x \rangle.$$



2. Truncate. Our goal: bring coefficients down to $O(\frac{1}{\sqrt{N}})$. So we *truncate them at level* $M = \frac{1}{\sqrt{\delta N}}$.

This can affect at most δN coefficients. ($\sum |b_i|^2 = 1$, thus at most δN can exceed M.)

3. Compute the residual. By the U.P.,

$$\Big\| \sum_{i \in \Omega} b_i u_i \Big\| \le (1-\eta) \Big(\sum_{i \in \Omega} |b_i|^2 \Big)^{1/2} \le 1 - \eta.$$

We have made progress: the residual is $(1-\eta)$ smaller than x.

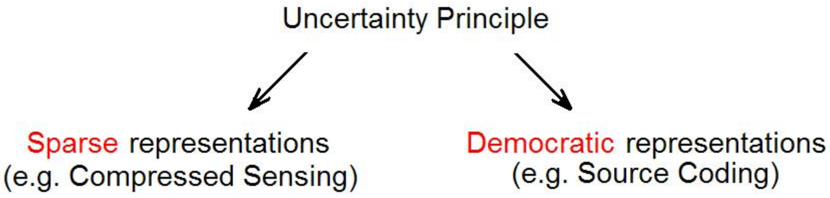
4. Iterate. Repeat the above steps for the residual (with truncation level $(1-\eta)$ times smaller). □

Summary.

- Frames are effective because of their redundancy: same information gets spread across different coefficients.
- However, in frame representations, the information is *not* spread uniformly. (Frame coefficients may have very different magnitude).
- Kashin's representations are those where the information is spread in the most uniform way (democracy). All coefficients are of the smallest possible order $\frac{1}{\sqrt{N}}$.
- Kashin's representations exist.
- Whenever a frame satisfies the Uncertainty Principle, one can quickly convert a frame representation into a Kashin's representation. Spreading algorithm.
- **Examples**: Gaussian, Bernoulli, Fourier (the latter up to $\log^2 N$), all with arbitrary redundancy > 1.
- No explicit examples are known.

Conclusion:

Conclusion:



Paper, slides available at `www.math.ucdavis.edu/~vershynin`