Uncertainty principles and compressed sensing

Jordan Bell

April 21, 2010

Let $Z_N = \mathbb{Z}/N\mathbb{Z}$. Let $\ell^p(Z_N)$ be the set of functions from Z_N to $\mathbb{C}$, and for $T \subseteq Z_N$ let $\ell^p(T)$ be the set of functions $Z_N \to \mathbb{C}$ whose support is a subset of T.

$\|f\|_{\ell^p(T)} = \left(\sum_{x\in T} |f(x)|\right)^{1/p}$. We also define $\|f\|_{\ell^0(Z_N)} = \sum_{x\in \mathrm{supp}(f)} 1$.

$\ell^p(T)$ is a $|T|$-dimensional vector space, for each p. In particular, $\ell^p(Z_N)$ is an N-dimensional vector space and thus $\ell^p(Z_N) \cong \mathbb{C}^N$. We speak interchangeably about a signal of length N, a function from Z_N to $\mathbb{C}$, and a vector of length N; sometimes one of these is more convenient.

For $f \in \ell^1(Z_N)$, the Fourier transform $\hat{f}$ of f is defined by

$$\hat{f}(\xi) = \frac{1}{\sqrt{N}} \sum_{x\in Z_N} f(x) e^{-2\pi i x\xi/N}.$$

$\hat{f} \in \ell^1(Z_N)$.

The Fourier inversion formula is

$$f(x) = \frac{1}{\sqrt{N}} \sum_{\xi\in Z_N} \hat{f}(\xi) e^{2\pi i x\xi/N}.$$

Define $F : \ell^1(Z_N) \to \ell^1(Z_N)$ by $F(f) = \hat{f}$. By the Fourier inversion formula, F is an isomorphism of vector spaces.

In this note we shall state and explain the main points of the proofs of several uncertainty principles for Z_N that are given in Terence Tao's paper *An uncertainty principle for cyclic groups of prime order*, Math. Res. Lett. **12** (2005), no. 1, 121–127 and the April 15, 2007 entry, "Ostrowski lecture: The uniform uncertainty principle and compressed sensing", in Terence Tao's blog (terrytao.wordpress.com). Also see the PDF on Tao's website from the talk by Candès and Tao, "The uniform uncertainty principle and compressed sensing", Harmonic Analysis and Related Topics, Seville, December 5, 2008.

We shall also state and prove a theorem from *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory **52** (2006), no. 2, 489–509 of Candès, Romberg and Tao which gives unique reconstruction. We shall explain why unique reconstruction (even in the special case of the theorem) fails when there is noise. Finally, we'll give some motivation for the Uniform Uncertainty Principle.

For $f \in \ell^1(Z_N)$, given $\Lambda \subseteq Z_N$ and $\{\hat{f}(\xi)\}_{\xi\in\Lambda}$, can we uniquely reconstruct f? We can do this if and only if $\Lambda = Z_N$; this is because by the Fourier inversion formula we can define a function by defining its Fourier transform.

Now say that $f \in \ell^1(Z_N)$ is S-sparse. Given $\Lambda \subset Z_N$ and $\{\hat{f}(\xi)\}_{\xi\in\Lambda}$, can we uniquely reconstruct f? i.e., is there a unique S-sparse function (namely f itself) which agrees with the measurements $\{\hat{f}(\xi)\}_{\xi\in\Lambda}$ we have?

Certainly $|\Lambda| \geq S$ is necessary. Actually it is necessary that $|\Lambda| \geq 2S$ (if $2S \leq N$). Suppose $|\Lambda| < 2S$. The subspace $F(\ell^1(\{1,\ldots,2S\}))$ of $\ell^1(Z_N)$ has dimension $2S$, and the subspace $\ell^1(Z_N - \Lambda)$ has dimension $N - |\Lambda|$. But $2S + N - |\Lambda| > N$, and thus there is a nonzero element in their intersection, say there is a nonzero $f \in \ell^1(\{1,\ldots,2S\})$ such that $\hat{f}|_\Lambda = 0$. Let $f = f_1 - f_2$ where f_1 is supported on $\{1,\ldots,S\}$ and f_2 is supported on $\{S+1,\ldots,2S\}$. $f \neq 0$ so f_1 and f_2 are distinct S-sparse functions such that $\hat{f_1}|_\Lambda = \hat{f_2}|_\Lambda$.

Unique reconstructions fails if and only if there is a nonzero $2S$-sparse function whose Fourier transform vanishes on all of Λ. In other words, unique reconstruction of S-sparse functions measured on Λ holds if and only if for all $2S$-sparse $g \in \ell^1(Z_N)$ there is some $\xi \in \Lambda$ such that $\hat{g}(\xi) \neq 0$.

The uncertainty principle suggests that the Fourier transform of a sparse function will have large support, and if the Fourier transform has large support then for most sets Λ the Fourier transform of the function will not be identically zero on Λ. Generally, the support of a function and the support of its Fourier transform cannot both be small: f and $\hat{f}$ cannot both be compactly supported (Paley-Wiener), and even f and $\hat{f}$ cannot both be supported on sets of finite measure (Benedicks).

Discrete uncertainty principle: If $f \in \ell^1(Z_N)$ is not identically zero, then $|\operatorname{supp}(f)| \cdot |\operatorname{supp}(\hat{f})| \geq N$.

The Plancherel identity tells us that

$$\sum_{x\in Z_N} |f(x)|^2 = \sum_{\xi\in Z_N} |\hat{f}(\xi)|^2.$$

By the Fourier inversion formula,

$$\sup_{x\in\mathbb{Z}/N} |f(x)| \leq \frac{1}{\sqrt{N}} \sum_{\xi\in Z_N} |\hat{f}(\xi)|^2.$$

By the Cauchy-Schwarz inequality,

$$\sum_{\xi\in Z_N} |\hat{f}(\xi)|^2 \leq |\operatorname{supp}(\hat{f})|^{1/2} \cdot \left(\sum_{\xi\in Z_N} |\hat{f}(\xi)|^2 \right)^{1/2}.$$

Therefore

$$N \left(\sup_{x\in Z_N} |f(x)| \right)^2 \leq |\operatorname{supp}(\hat{f})| \cdot \sum_{\xi\in Z_N} |\hat{f}(\xi)|^2.$$

By the Plancherel identity we have

$$N\Big(\sup_{x\in Z_N}|f(x)|\Big)^2 \le |\operatorname{supp}(\hat{f})|\cdot\sum_{x\in Z_N}|f(x)|^2.$$

Also,

$$\sum_{x\in Z_N}|f(x)|^2 \le |\operatorname{supp}(f)|\cdot\big(\sup_{x\in Z_N}|f(x)|\big)^2.$$

Thus

$$N\Big(\sup_{x\in Z_N}|f(x)|\Big)^2 \le |\operatorname{supp}(\hat{f})|\cdot|\operatorname{supp}(f)|\cdot\big(\sup_{x\in Z_N}|f(x)|\big)^2.$$

Since f is not identically 0 this gives

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

Now, we have unique recoverability if and only if for every $2S$-sparse function g has a frequency in Λ. This will happen in particular if $|\Lambda| > N - |\operatorname{supp}(\hat{g})|$. If $|\Lambda| > N - \frac{N}{2S}$ then $|\Lambda| > N - |\operatorname{supp}(\hat{g})|$. Therefore, if $|\Lambda| > N - \frac{N}{2S}$ then by the Discrete Uncertainty Principle we will have unique reconstruction.

The bound $|\Lambda| > N - \frac{N}{2S}$ cannot not be improved in general: When N is a square, let f be the indicator function for $T = \{0, \sqrt{N}, 2\sqrt{N}, \ldots, N-\sqrt{N}\}$ and let $\Lambda = Z_N - T$. Then f is $\sqrt{N}$-sparse so $N - \frac{N}{2S} = N - \frac{\sqrt{N}}{2}$, and $|\Lambda| = N - \sqrt{N}$. We know that $\hat{f} = f$. The zero function and f are both $\sqrt{N}$-sparse functions whose Fourier transforms agree on Λ, so unique reconstruction does not occur. (Here we showed that one can't do better in general than $|\Lambda| > N - \frac{N}{S}$.)

If $N = ab$ and f is the indicator function for the multiples of b, then $|\operatorname{supp}(f)| = a$, and one can show that $|\operatorname{supp}(\hat{f})| = b$. See p. 226 of Audrey Terras, *Fourier analysis on finite groups and applications*, 1999. We can take $|\Lambda| = ab - b$ measurements and still not be able to distinguish f from the zero function.

When N is a prime then these types of counterexamples do not exist.

Uncertainty principle for cyclic groups of prime order: If N is prime and $f \in \ell^1(Z_N)$ is not identically zero, then $|\operatorname{supp}(f)| + |\operatorname{supp}(\hat{f})| \ge N + 1$. This demands far fewer measurements than the Discrete Uncertainty Principle.

We have unique reconstruction if for every $2S$-sparse $g \in \ell^1(Z_N)$, $\hat{g}$ is not identically zero on Λ. It would suffice that for every $2S$-sparse g, $\operatorname{supp}(\hat{g}) > N - |\Lambda|$. By this uncertainty principle we know that $\operatorname{supp}(\hat{g}) > N - \operatorname{supp}(g) \ge N - 2S$, so if $|\Lambda| \ge 2S$ then we have unique reconstruction.

If we take a bump function on $\mathbb{R}$ that is 1 on $\{-\lfloor\sqrt{N}\rfloor, -\lfloor\sqrt{N}\rfloor+1, \ldots, \lfloor\sqrt{N}\rfloor\}$ then the Fourier transform will be concentrated on the same set. This yields a function f on Z_N defined by the values of the bump function on the integers $0, 1, \ldots, N-1$. If there is roundoff error, then f is $\sqrt{N}$-sparse, since it is very small outside of $\{0, 1, \ldots, \lfloor\sqrt{N}\rfloor\}$. If we take $\Lambda = \{\lfloor N/3\rfloor, \lfloor N/3\rfloor + 1, \ldots, \lfloor 2N/3\rfloor\}$ then $\hat{f}$ will be very small on Λ and so by roundoff error in fact zero on Λ. This examples applies whether or not N is prime, so if there is error

then we don't have unique reconstruction for all sets satisfying $|\Lambda| \geq 2S$ when N is prime.

These obstructions to unique reconstruction all involved sets of observed frequencies Λ that were arithmetic progressions. More generally, the problem Λ were structured. It turns out that if we select a set of frequencies uniformly at random among all $\binom{N}{|\Lambda|}$ sets of size $|\Lambda$ we will usually not have these obstructions. ("Most sets" don't have the properties that make unique reconstruction fail.)

Uniform Uncertainty Principle: If Λ is a random set with $|\Lambda| \gg S \log^4 N$, then with probability $1 - O(N^{-A})$ for any fixed A,

$$\sum_{\xi \in \Lambda} |\hat{f}(\xi)|^2 \approx \frac{|\Lambda|}{N} \sum_{x \in Z_N} |f(x)|^2$$

for all $4S$-sparse functions $f \in \ell^1(Z_N)$. $X \approx Y$ means that X and Y differ by at most 10%.

For a single function f we cans show that Λ gets its "fair share" of the ℓ^2 norm using the Chernoff inequality, but it is much harder to prove for all functions since there are so many functions... We need some way of keeping a handle on the set of all S-sparse functions.

Chernoff inequality: Let $X_1, \ldots, X_n$ be independent random variables with $|X_i| \leq K$, mean μ_i and variance σ_i^2. Let $S_n = \sum_{i=1}^n X_i, \mu = \sum_{i=1}^n \mu_i, \sigma^2 = \sum_{i=1}^n \sigma_i^2$. Then for any $\lambda > 0$ one has

$$P(|S_n - \mu| \geq \lambda\sigma) \leq$$
$$C \max(\exp(-c\lambda^2), \exp(-c\lambda\sigma/K))$$

for some absolute constants $C, c > 0$.

Let $n = |\Lambda|$ and take $X_i = |\hat{f}(\xi)|^2$ with probability $\frac{1}{N}$. Then $\mu_i = \frac{1}{N}\ell^2(\hat{f})$ and $\mu = \frac{|\Lambda|}{N}\ell^2(\hat{f})$. When $|\Lambda|$ is small relative to N, then $\sum_{i=1}^n X_i$ and $\sum_{\xi \in \Lambda} |\hat{f}(\xi)|^2$ have nearly the same distribution, because most multisets of size $|\Lambda|$ will not have repeated entries.

To use the Chernoff inequality we want $P(|S_n - \mu| > \delta\mu) \leq P(|S_n - \mu| > \lambda\sigma)$, where $\delta = 0.1$. So let $\lambda = \frac{\delta\mu}{\sigma}$.

The paper *Near-optimal signal recovery from random projections: Universal encoding strategies?*, IEEE Trans. Inform. Theory **52** (2006), no. 12, 5406–5425 by Candès and Tao deals with signals that are not sparse but that obey a "power law", where there are S big components in the signal and then for the rest of the components, the nth component decays $O(n^{-1/p})$ for some $p > 0$. This paper is a tour de force. The "main result of the paper" which states that if a measurement process obeys UUP and ERP then we can reconstruct something which has a very close ℓ^2 norm to the original signal. But to actually show that ERP and UUP hold for the measurement processes (=ensembles=random matrices) they deal with in the paper is a whole other barrel of monkeys. (In this presentation we talked about the Fourier ensemble, which is the hardest ensemble to prove ERP and UUP for.)

Proof of the uncertainty principle for cyclic groups of prime order. First assume what we'll call Chebotarev's lemma: Let p be prime and $1 \leq n \leq p$. Let $x_1, \ldots, x_n$ be distinct elements of Z_p and let $\xi_1, \ldots, x_n$ be distinct elements of Z_p. Then the matrix $(e^{2\pi i x_i \xi_k})_{1 \leq j,k \leq n}$ has nonzero determinant.

Now, let $A, \tilde{A}$ be nonempty subsets of Z_p with $|A| = |\tilde{A}|$. Define $T : \ell^2(A) \to \ell^2(\tilde{A})$ by $Tf = \hat{f}|_{\tilde{A}}$. Thinking of the left-hand side (physical space) as having a basis of Dirac deltas and the right-hand side (frequency space) as having a basis of exponentials, the matrix representation of T is $(e^{2\pi i x_i \xi_k})_{1 \leq j,k \leq n}$ for some choice of x_i and ξ_i. Therefore T is an isomorphism.

Suppose by contradiction that there is a nonzero $f \in \ell^2(Z_p)$ such that $|\operatorname{supp}(f)| + |\operatorname{supp}(\hat{f})| \leq p$. Let $A = \operatorname{supp}(f)$. Then there some $\tilde{A} \subset \ell^2(Z_p)$ that is disjoint from $\operatorname{supp}(\hat{f})$ and such that $|A| = |\tilde{A}|$. $T : \ell^2(A) \to \ell^2(\tilde{A})$ defined by $Tf = \hat{f}|_{\tilde{A}}$ is an isomorphism, but f is a nonzero vector that it sends to zero, a contradiction. Therefore for all nonzero $f \in \ell^2(Z_p)$, $|\operatorname{supp}(f)| + |\operatorname{supp}(\hat{f})| \geq p + 1$. We've seen too that this bound is sharp. Thus the uncertainty principle for cyclic groups of prime order is proved.

Now we'll turn to proving Chebotarev's lemma.

The proof of Chebotarev's lemma itself requires another rather easy lemma (from the "Galois theory of the cyclotomic integers"): Let p is prime, n a positive integer and $P(z_1, \ldots, z_n)$ a polynomial with integer coefficients. If $\omega_1, \ldots, \omega_n$ are pth roots of unity such that $P(\omega_1, \ldots, \omega_n) = 0$, then $P(1, \ldots, 1)$ is a multiple of p.

To prove this, let $\omega = e^{2\pi i/p}$ and let $\omega_j = \omega^{k_j}$ for some integers $0 \leq k_j < p$. Now define

$$Q(z) = P(z^{k_1}, \ldots, z^{k_n}) \qquad \mod z^p - 1.$$

Here we are *not* talking about quotient rings, just the result of polynomial long division. So $Q(\omega) = 0$ and $Q(1) = P(1, \ldots, 1)$ (since the quotient from the long division will be equal to 0 when evaluated at 1). Since p is prime the minimal polynomial of ω is the cyclotomic polynomial $\Phi_p(z) = 1 + z + \cdots + z^{p-1}$; $Q(z)$ must be the product of this with a polynomial over the integers, but since $Q(z)$ has degree at most $p - 1$, it is an integer multiple of the minimal polynomial. Therefore $Q(1) = P(1, \ldots, 1)$ is an integer multiple of $\Phi_p(1) = p$.

Now we have everything we need for the proof of Chebotarev's lemma. Let $\omega_j = e^{2\pi i x_j/p}$. We have to show that

$$\det(\omega_j^{\xi_k})_{1 \leq j,k \leq n}$$

is not a multiple of p. To do this, define

$$D(z_1, \ldots, z_n) = \det(z_j^{\xi_k})_{1 \leq j,k \leq n}$$

and put

$$D(z_1, \ldots, z_n) = P(z_1, \ldots, z_n) \prod_{1 \leq j < j' \leq n} (z_j - z_{j'}).$$

Consider

$$(z_2\frac{d}{dz_2})^1(z_3\frac{d}{dz_3})^2\cdots(z_n\frac{d}{dz_n})^{n-1}D(z_1,\ldots,z_n) \tag{1}$$

evaluated at $z_1 = \ldots = z_n = 1$.

This is equal to

$$(n-1)!(n-2)!\cdots 1P(1,\ldots,1).$$

None of the factorials are multiples of p since $n \leq p$. So to show that $P(1,\ldots,1)$ is not a multiple of p it suffices to show that (1) is not a multiple of p. But (1) is equal to

$$\det(\xi_k^{j-1})_{1\leq j,k\leq n}$$

which is a Vandermonde determinant and is equal to

$$\prod_{1\leq k<k'\leq n}(\xi_k - \xi_{k'}).$$

These factors are all distinct modulo p, hence the product is not a multiple of p.