The uniform uncertainty principle and compressed sensing

Harmonic analysis and related topics, Seville

December 5, 2008

Emmanuel Candés (Caltech), Terence Tao (UCLA)

Uncertainty principles

A basic principle in harmonic analysis is:

> **Uncertainty principle:** (informal) If a function $f : G \to \mathbb{C}$ on an abelian group G is concentrated in a small set, then its Fourier transform $\hat{f} : \hat{G} \to \mathbb{C}$ must be "spread out" over a large set.

There are many results that rigorously capture this sort of principle.

For instance, for the real line $G = \mathbb{R}$, with the standard Fourier transform $\hat{f}(\xi) = \int_{\mathbb{R}} f(x) e^{-2\pi i x\xi}\ dx$, we have

Heisenberg uncertainty principle: If $\|f\|_{L^2(\mathbb{R})} = \|\hat{f}\|_{L^2(\mathbb{R})} = 1$, and $x_0, \xi_0 \in \mathbb{R}$, then $\|(x - x_0)f\|_{L^2(\mathbb{R})} \|(\xi - \xi_0)\hat{f}\|_{L^2(\mathbb{R})} \geq \frac{1}{4\pi}$. (More succinctly: $(\Delta x)(\Delta \xi) \geq \frac{1}{4\pi}$.)

Proof: Normalise $x_0 = \xi_0 = 0$, use the obvious inequality $\int_{\mathbb{R}} |axf(x) + ibf'(x)|^2\ dx \geq 0$, integrate by parts, and optimise in a, b. $\square$

Equality is attained for centred Gaussians

$$f(x) = ce^{-\pi Ax^2}; \quad \hat{f}(\xi) = \frac{c}{\sqrt{A}} e^{-\pi \xi^2/A}$$

when $x_0 = \xi_0 = 0$; this example can be translated and modulated to produce similar examples exist for other x_0, ξ_0.

What about for finite abelian groups G, e.g. cyclic groups $G = \mathbb{Z}/N\mathbb{Z}$?

The Pontryagin dual group $\hat{G}$ of characters $\xi : G \to \mathbb{R}/\mathbb{Z}$ has the same cardinality as G. For $f : G \to \mathbb{C}$, we define the Fourier transform $\hat{f} : \hat{G} \to \mathbb{C}$ as

$$\hat{f}(\xi) := \int_G f(x) e(\xi \cdot x)\, dx$$

where $e(x) := e^{2\pi i x}$ and $dx = \frac{1}{|G|} d\#$ is normalised counting measure.

We have the inversion formula

$$f(x) := \sum_{\xi \in \hat{G}} \hat{f}(\xi) e(-\xi \cdot x)$$

and the Plancherel formula

$$\|f\|_{L^2(G)} = \|\hat{f}\|_{l^2(\hat{G})}.$$

The analogue of Gaussians for finite abelian groups are the indicator functions of subgroups.

If $H \leq G$ is a subgroup of G, define the *orthogonal complement* $H^\perp \leq \hat{G}$ as

$$H^\perp := \{\xi \in \hat{G} : \xi \cdot x = 0 \text{ for all } x \in H\}.$$

We have the Poisson summation formula

$$\widehat{1_H} = \frac{|H|}{|G|} 1_{H^\perp}$$

(in particular, the Fourier transform of 1 is a Dirac mass, and vice versa.) From this and Plancherel we have the basic identity

$$|H| \times |H^\perp| = |G|.$$

More generally, for finite abelian G we have

Donoho-Stark uncertainty principle(1989) For any non-trivial $f : G \to \mathbb{C}$, we have $|\text{supp}(f)||\text{supp}(\hat{f})| \geq |G|$.

Proof: Combine Plancherel's theorem with the Hölder inequality estimates

$$\|f\|_{L^1(G)} \leq |\text{supp}(f)|^{1/2}|G|^{-1/2}\|f\|_{L^2(G)};$$

$$\|\hat{f}\|_{l^2(\hat{G})} \leq |\text{supp}(\hat{f})|^{1/2}\|\hat{f}\|_{l^\infty(\hat{G})}$$

and the Riemann-Lebesgue inequality

$$\|\hat{f}\|_{l^\infty(\hat{G})} \leq \|f\|_{L^1(G)}.$$

□

One can show that equality is attained precisely for the indicators 1_H of subgroups, up to translation, modulation, and multiplication by constants.

One also has a slightly more quantitative variant:

Entropy uncertainty principle If $\|f\|_{L^2(G)} = \|\hat{f}\|_{l^2(\hat{G})} = 1$, then

$$-\int_G |f(x)|^2 \log \frac{1}{|f(x)|}\, d\xi - \sum_{\xi \in G} |\hat{f}(\xi)|^2 \log \frac{1}{|f(\xi)|} \geq 0.$$

Proof Differentiate (!) the Hausdorff-Young inequality $\|\hat{f}\|_{L^{p'}(\hat{G})} \leq \|f\|_{L^p(G)}$ at $p = 2$. $\square$

From Jensen's inequality we have

$$-\int_G |f(x)|^2 \log \frac{1}{|f(x)|}\, d\xi \geq \log \frac{|\mathrm{supp}(f)|}{|G|}$$

$$-\sum_{\xi \in G} |\hat{f}(\xi)|^2 \log \frac{1}{|f(\xi)|} \geq \log |\mathrm{supp}(\hat{f})|$$

and so the entropy uncertainty principle implies the Donoho-Stark uncertainty principle.

Again, equality is attained for indicators of subgroups (up to translation, modulation, and scalar multiplication).

For arbitrary groups G and arbitrary functions f, one cannot hope to do much better than the above uncertainty principles, due to examples such as the subgroup counterexamples $f = 1_H$.

On the other hand, for generic groups and functions, one expects to do a lot better. (For instance, for generic f one has $\mathrm{supp}(f) = G$ and $\mathrm{supp}(\hat{f}) = \hat{G}$.)

So one expects to obtain improved estimates by imposing additional hypotheses on G or f.

For instance, for cyclic groups $G = \mathbb{Z}/p\mathbb{Z}$ of prime order, which have no non-trivial subgroups, we have

Uncertainty principle for $\mathbb{Z}/p\mathbb{Z}$ (T., 2005)
If $f : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ is non-trivial, then $|\mathrm{supp}(f)| + |\mathrm{supp}(\hat{f})| \geq p + 1$.

This is equivalent to an old result of Chebotarev that all minors of the Fourier matrix $(e(x\xi/p))_{1 \leq x, \xi \leq p}$ are non-singular, which is proven by algebraic methods. The result is completely sharp: if A, B are sets with $|A| + |B| \geq p + 1$, then there exists a function f with $\mathrm{supp}(f) = A$ and $\mathrm{supp}(\hat{f}) = B$.

Partial extensions to other groups (Meshulam, 2006).

This uncertainty principle has some amusing applications to arithmetic combinatorics, for instance it implies the Cauchy-Davenport inequality (1813)

$$|A+B| \geq \min(|A|+|B|-1, p)$$

for subsets A, B of $\mathbb{Z}/p\mathbb{Z}$. (Proof: apply the uncertainty principle to functions of the form $f * g$, where f is supported in A, g is supported in B, and $\mathrm{supp}(\hat{f}), \mathrm{supp}(\hat{g})$ are chosen to have as small an intersection as possible.)

Further applications of this type (Sun-Guo, 2008), (Guo, 2008).

The uncertainty principle for $\mathbb{Z}/p\mathbb{Z}$ has the following equivalent interpretation:

> **Uncertainty principle for $\mathbb{Z}/p\mathbb{Z}$** If $f : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ is S-sparse (i.e. $|\text{supp}(f)| \leq S$) and non-trivial, and $\Omega \subset \mathbb{Z}/p\mathbb{Z}$ has cardinality at least S, then $\hat{f}$ does not vanish identically on Ω.

From a signal processing perspective, this means that any S Fourier coefficients of f are sufficient to detect the non-triviality of an S-sparse signal. This is of course best possible.

It is crucial that p is prime. For instance, if N is a square, then $\mathbb{Z}/N\mathbb{Z}$ contains a subgroup of size $\sqrt{N}$, and the indicator function of that subgroup (the Dirac comb) vanishes on $N - \sqrt{N}$ Fourier coefficients despite being only $\sqrt{N}$-sparse.

As a corollary of the uncertainty principle, if $f : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ is an unknown signal which is known to be S-sparse, and we measure $2S$ Fourier coefficients $(\hat{f}(\xi))_{\xi \in \Omega}$ of f, then this uniquely determines f; for if two S-sparse signals f, g had the same Fourier coefficients on Ω, then the $2S$-sparse difference $f - g$ would have trivial Fourier transform on Ω, a contradiction.

This is a prototype of a compressed sensing result. Compressed sensing refers to the ability to reconstruct sparse (or compressed) signals using very few measurements, without knowing in advance the support of the signal. (Note that one normally needs all p Fourier coefficients in order to recover a general signal; the point is that sparse signals have a much lower information entropy and thus are easier to recover than general signals.)

However, this result is unsatisfactory for several reasons:

- It is ineffective. It says that recovery of the S-sparse signal f from $2S$ Fourier coefficients is possible (since f is uniquely determined), but gives no efficient algorithm to actually locate this f.
- It is not robust. For instance, the result fails if p is changed from a prime to a composite number. One can also show that the result is not stable with respect to small perturbations of f, even if one keeps p to be prime.

It turns out that both of these problems can be solved if the frequency set Ω does more than merely detect the presence of a non-trivial sparse signal, but gives an accurate measurement as to how large that signal is. This motivates:

Restricted Isometry Principle (RIP). A set of frequencies $\Omega \subset \mathbb{Z}/N\mathbb{Z}$ is said to obey the RIP with sparsity S and error tolerance δ if one has

$$(1-\delta)\frac{|\Omega|}{N}\|f\|_{L^2(\mathbb{Z}/N\mathbb{Z})}^2 \leq \|\hat{f}\|_{l^2(\Omega)}^2 \leq (1+\delta)\frac{|\Omega|}{N}\|f\|_{L^2(\mathbb{Z}/N\mathbb{Z})}^2$$

for all S-sparse functions f.

Note that the factor $\frac{|\Omega|}{N}$ is natural in view of Plancherel's theorem; it asserts that Ω always captures its "fair share" of the energy of a sparse function. It implies that Ω detects the presence of non-trivial S-sparse functions, but is much stronger than this.

This principle is very useful in compressed sensing, e.g.

Theorem (Candés-Romberg-T., 2005) Suppose $\Omega \subset \mathbb{Z}/N\mathbb{Z}$ obeys the RIP with sparsity $4S$ and error tolerance $1/4$. Then any S-sparse signal f is the unique solution $g : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ to the problem $\hat{g}|_\Omega = \hat{f}|_\Omega$ with minimal $L^1(G)$ norm. In particular, f can be reconstructed from the Fourier measurements $\hat{f}|_\Omega$ by a convex optimisation problem.

Sketch of proof: If a function $g = f + h$ is distinct from f but has the same Fourier coefficients as f on Ω, use the RIP to show that h has a substantial presence outside of the support of f compared to its presence inside this support, and use this to show that $f + h$ must have a strictly larger $L^1(G)$ norm than f. $\square$

Similar arguments show that signal recovery using frequency sets that obey the RIP are robust with respect to noise or lack of perfect sparsity (e.g. if f is merely S-compressible rather than S-sparse, i.e. small outside of a set of size S). There is now a vast literature on how to efficiently perform compressed sensing for various measurement models, many of which obey (or are assumed to obey) the RIP.

On the other hand, the RIP fails for many frequency sets. Consider for instance an S-sparse function $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ that is a bump function adapted to the interval $\{-S/2, \ldots, S/2\}$ in $\mathbb{Z}/N\mathbb{Z}$. Then $\hat{f}$ is concentrated in an interval of length about N/S centred at the frequency origin. If Ω avoids this interval (or intersects it with too high or too low of a density), then the RIP fails.

Variants of this example show that a frequency set must be equidistributed in various senses if it is to obey the RIP.

Uniform uncertainty principle (Candés-T., 2006) A randomly chosen subset Ω of $\mathbb{Z}/N\mathbb{Z}$ of size $CS\log^6 N$ will obey the RIP with high probability $(1 - O(N^{-C}))$.

Informally, a randomly chosen set of size $O(S\log^6 N)$ will always capture its fair share of the energy of any S-sparse function, thus we have a sort of "local Plancherel theorem" for sparse functions that only requires a random subset of the frequencies.

This implies that robust compressed sensing is possible with an oversampling factor of $O(\log^6 N)$. This was improved to $O(\log^5 N)$ (Rudelson-Vershynin, 2008). In practice, numerics show that an oversampling of 4 or 5 is sufficient. A separate argument (Candés-Romberg-Tao, 2006) shows that (non-robust) compressed sensing is possible w.h.p. with an oversampling factor of just $O(\log N)$.

The method of proof is related to Bourgain's solution of the Λ_p problem, which eventually reduced to understanding the behaviour of maximal exponential sums such as

$$\Lambda_p(\Omega) := \sup\{\| \sum_{\xi \in \Omega} c_\xi e(x\xi/N)\|_{L^p(\mathbb{Z}/N\mathbb{Z})} : \|c\|_{l^2(\Omega)} = 1\}$$

for randomly chosen sets Ω. In particular it relies on the use of a chaining argument used by Bourgain (and also simultaneously by Talagrand).

The chaining argument

For each individual S-sparse function f, and a random Ω, it is not hard to show that the desired inequality

$$(1-\delta)\frac{|\Omega|}{N}\|f\|^2_{L^2(\mathbb{Z}/N\mathbb{Z})} \leq \|\hat{f}\|^2_{l^2(\Omega)} \leq (1+\delta)\frac{|\Omega|}{N}\|f\|^2_{L^2(\mathbb{Z}/N\mathbb{Z})}$$

holds with high probability; this is basically the law of large numbers (and is the reason why Monte Carlo integration works), and one can get very good estimates using the Chernoff inequality. The problem is that there are a lot of S-sparse functions in the world, and the total probability of error quickly adds up.

One can partially resolve this problem by discretisation: pick an ε, and cover the space Σ of all S-sparse functions in some suitable metric (e.g. L^2 metric) by an ε-net of functions. But it turns out that there are still too many functions in the net to control, and even after controlling these functions, the S-sparse functions in Σ that are near to the net, but not actually on the net, are still not easy to handle.

The solution is to chain several nets together, or more precisely to chain together 2^{-n}-nets $\mathcal{N}_n$ of Σ for each $n = 1, 2, 3, \ldots$. Instead of controlling the functions f_n in each net $\mathcal{N}_n$ separately, one instead controls the extent to which f_n deviates from its "parent" f_{n-1} in the next coarser net $\mathcal{N}_{n-1}$, defined as the nearest element in $\mathcal{N}_{n-1}$ to f_n. This deviation is much smaller than f itself, in practice, and is easier to control. After getting good control on all of these deviations, one can then control arbitrary functions f by expressing f as a telescoping series of differences $f_n - f_{n-1}$.