Random Sampling and Probability

Richard Bass
University of Connecticut
r.bass@uconn.edu
www.math.uconn.edu/$\sim$bass

This is joint work with Karlheinz (Charlie) Gröchenig, University of Vienna

The problem

The problem: given x_j, $y_j = f(x_j)$, find f.

This is

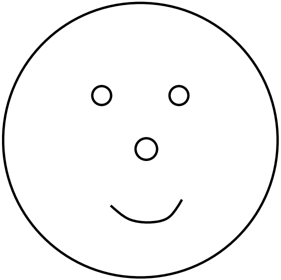
- Classical
- Useful
- Topical: (Candés, Romberg, and Tao; Cucker and Smale; and Smale and Zhou)
- Easy

Refining the problem

We need to restrict the class of f's somehow. We assume our samples lie in $[0,1]^d$ and we look at band-limited f's:

$$f(x) = \sum_{|k| \le M, k \in \mathbb{Z}^d} a_k e^{2\pi i k \cdot x}.$$

If the x_n's are on a lattice, things are easier, but one wants to allow other cases. For example, recording a picture for a passport, one want more detail in certain places, such as the face.



The case of one dimension is well understood, but $d > 1$ is poorly understood. The reason is that a lot is known about the zeros of an analytic function, but very little about the zeros of a holomorphic function on $\mathbb{C}^n$. Yet there are algorithms that work, although no one understands why they do.

The method

The method to find f is this. We have a collection of linear equations:

$$y_n = f(x_n) = \sum_{|k| \le M, k \in \mathbb{Z}^d} a_k e^{2\pi i k \cdot x_n}.$$

If we let $U_{nk} = e^{2\pi i k \cdot x_n}$, and define the matrices a and y in the obvious way, we need to solve

$$Ua = y$$

for a. It is common to look at

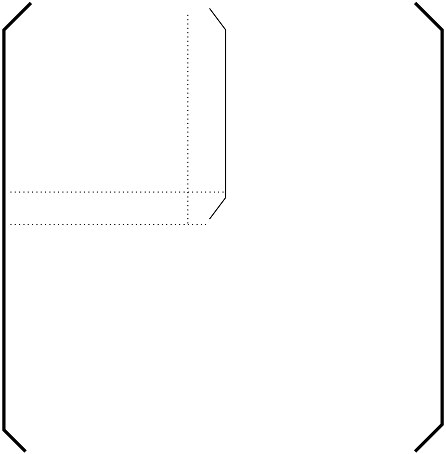
$$U^* Ua = U^* y,$$

and then $T = U^* U$ is a Toeplitz matrix.

First result

Question 1: For what x_n's can we solve this equation?

Theorem 1. *If the x_j's are chosen independently and the distribution of each x_j has a density, then T is invertible with probability 1.*



The idea of the proof is as follows. T is a $(2M+1)^d \times (2M+1)^d$ matrix. We look at the $N \times N$ upper left hand corner and show by induction that each of these is invertible. Suppose we have the $N \times N$ matrix invertible. If we let b_i be the vector consisting of the first N columns in the i^{th} row, then b_{N+1} is a unique linear combination of $b_1, \ldots, b_N$. So the only way we do not have invertibility of the $(N+1) \times (N+1)$ upper left hand corner is if $a_{N+1,N+1}$ is the same linear combination of the $a_{i,N+1}$. Sorting this out, this means that x_{N+1} is a zero for a particular fixed trigonometric polynomial. The set of zeros of a trigonometric polynomial have measure 0, so with probability 1, x_{N+1} is not a zero.

The sampling inequality

Question 2: If one can solve the system of equations, can one do it practically? What is the condition number of T?

The condition number is the ratio of the largest to smallest eigenvalue. The larger the condition number, the harder it is to solve linear equations numerically.

What we prove is known as a sampling inequality:

$$A\|f\|_2^2 \le \sum_{j=1}^{r} |f(x_j)|^2 \le B\|f\|_2^2.$$

If we have a sampling inequality, we know three things.

1. $\kappa(T) \leq B/A$, where κ is the condition number.

(Recall the equation $y = Ua$. The middle expression in the sampling inequality is $\|y\|_2^2$ and the left and right hand sides are $\|a\|_2^2$.)

2. Uniqueness - if $f_1(x_j) = f_2(x_j)$ for all j, the left hand inequality applied to $f_1 - f_2$ implies uniqueness.

3. Stability - if we vary f a little, the samples will only vary a little. This comes from the right hand inequality.

Second result

Theorem 2. *If the x_j are i.i.d. with a uniform distribution, then*

$$(1-\varepsilon) r \|f\|_2^2 \leq \sum_{j=1}^{r} |f(x_j)|^2 \leq (1+\varepsilon) r \|f\|_2^2$$

holds with probability at least

$$1 - c_1 e^{-c_2 r \frac{\varepsilon^2}{1+\varepsilon}}.$$

A corollary is that

$$\limsup_{r\to\infty} \frac{\sup_f |\sum |f(x_j)|^2 - \|f\|_2^2|}{\sqrt{r \log\log r}\|f\|_2^2} = c, \qquad a.s.$$

So

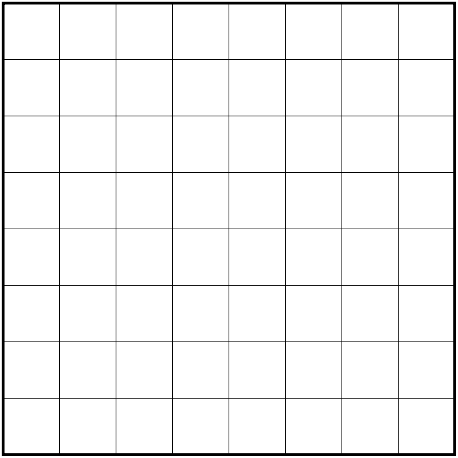
$$\kappa(T) \approx 1 + c\Big(\frac{\log\log r}{r}\Big)^{1/2}.$$

For the lattice case, we have the same without the double logs, so we don't lose much.

I'll sketch two proofs. One gives explicit bounds, the other generalizes. Let

$$\delta = \inf\{s : [0,1]^d \subset \cup_{j=1}^r B(x_j, s)\}.$$

We show that δ will be small when r is large enough, and then apply a theorem of Beurling.



Divide the unit cube into subcubes of side $1/N$. If none are empty, then $\delta(r) \leq \sqrt{d}/N$. The probability that a sample misses a given cube is $1 - N^{-d}$. So the probability that all the samples miss a given cube is $(1 - N^{-d})^r$. There are N^d cubes, so the probability that $\delta > \sqrt{d}/N$ is bounded by

$$N^d(1 - N^{-d})^r.$$

It turns out this sort of analysis is sharp.

The other proof uses three ideas.

First, since $f = \sum a_k e^{2\pi i x \cdot k}$, the L^2 norm of f is the same as the ℓ^2 norm of a. By Cauchy-Schwarz, the L^∞ norm is comparable (with a constant $(2M+1)^{d/2}$ which can be bad, but is fixed). By interpolation all the L^p norms of f are comparable.

Second, Bernstein's inequality - the one for sums of independent random variables, not the one for the derivative of a Fourier series.

$$\mathbb{P}(S_n > \lambda) \le \exp\Big(- \frac{\lambda^2}{2\sigma^2 + \lambda M/3} \Big).$$

Third is the technique of metric entropy.

To get a bound for the sup of Y_f over a class of f's, one gets a bound for the sup over a small finite class C_1 of f's. Then one gets a bound for the sup over a slightly larger finite class C_2 by comparing every f in C_2 to the nearest element in C_1. Continuing, one gets a bound on the sup for a countable dense subset of f's.

Here is a little more detail. For each i, let C_i be a finite set of points, suppose the C_i increase, and the union of the C_i is dense in the space.

Write

$$\mathbb{P}(\sup_{C_{i+1}} |Y_f| > \lambda_i + \sup_{C_i} |Y_f|) \le \mathbb{P}(\sup_{f \in C_{i+1}, g \in C_i} |Y_f - Y_g| > \lambda_i).$$

One can be a little more efficient in realizing that the last sup can be restricted to f's and g's that are close together.

By balancing the selection of the C_i against the values of λ_i, one can get very good estimates.

Generalizations

We can let our class of functions be $\{\sum a_k e_k\}$ for some other basis $\{e_k\}$. We can look at almost periodic functions:

$$\sum a_k e^{2\pi i \lambda_k \cdot x}.$$

where the λ_k are not necessarily elements of $\mathbb{Z}^d$. We can look at algebraic polynomials, we can look at shift invariant functions (e.g., wavelets), and we can look at spherical harmonics.

The infinite-dimensional problem

Now what if we want to let $x_n \in \mathbb{R}^d$, $y_n = f(x_n)$? A suitable class of functions to consider is the band-limited functions, those whose Fourier transform has support in $[-\frac{1}{2}, \frac{1}{2}]^d$. We are now in an infinite dimensional situation.

Negative results

(a) Suppose in each cube we pick r points uniformly. In this case the sampling inequality fails. Look at $d = 1$. Find f such that the zeros of f are near the even integers. By Borel-Cantelli there will be a long string of intervals where all the samples are near the even integers. A shift of f gives a function with a poor constant for the sampling inequality.

One can construct a sequence of functions f_k such that

$$\sum |f_k(x_k)|^2 \leq \frac{\|f_k\|_2^2}{k}.$$

(b) Another idea for random sampling is the spatial Poisson process. Let λ be Lebesgue measure, and we want the number of samples in a set A to be Poisson with parameter $\lambda(A)$, where the numbers of points in disjoint sets are independent. By Borel-Cantelli there will be a large hole with no samples, and by results of Landau a sampling inequality cannot hold.

We could let $d\lambda/dx = o(1+\log^+(|x|))$, and there is the same difficulty, but if

$$d\lambda/dx = c(1+\log^+(|x|))$$

for large enough c, then things are OK.

Positive results

We look at functions most of whose energy is not too far away from the origin. We let

$$\mathcal{B}_{R,\delta} = \Big\{ f \text{ band-limited} : \int_{[-R/2,R/2]^d} |f(x)|^2\,dx \geq (1-\delta)\|f\|_2^2, \|f\|_2 \leq 1 \Big\}.$$

Then the sampling inequality holds.

The idea is to use metric entropy. The spaces $\mathcal{B}_{R,\delta}$ are compact. When we discussed metric entropy, we said we balanced the size of the C_i's against the λ_i's. (We looked at pairs f, g with $f \in C_{i+1}$, $g \in C_i$, and f and g close.)

Given some compact metric space, the covering numbers $N(\varepsilon)$ are defined by

$$N(\varepsilon) = \log M(\varepsilon),$$

where $M(\varepsilon)$ is the fewest number of balls of radius ε that cover the space.

The size of the C_i's is related to the covering numbers.

To get the covering numbers, we estimate eigenvalues.

Let $A_R = QP_RQ$, where

$$\widehat{Qf} = 1_{[-1/2,1/2]^d}\widehat{f}, \qquad P_R f = 1_{[-R/2,R/2]^d} f.$$

Let φ_n be an orthonormal basis for L^2 with respect to A_R. These are products of prolate spheroidal functions.

An argument counting the number of eigenvalues less than ε for A_R leads to a covering number for $\mathcal{B}_{R,\delta} \cap B$, where B is the unit ball in L^2.

The argument goes something like this.

Let λ_n be the eigenfunctions and let

$$S_\delta = \Big\{ c \in \ell^2 : \|c\|_2 \le 1, \sum_n \lambda_n |c_n|^2 \ge 1 - \delta \Big\}.$$

Since

$$\sum_n \lambda_n |c_n|^2 = \sum_{\lambda_n \ge \varepsilon/2} + \sum_{\lambda_n < \varepsilon/2},$$

it suffices to get a covering number for

$$S_\delta^\varepsilon = \Big\{ c \in \ell^2 : \|c\|_2 \le 1, \sum_{\lambda_n \ge \varepsilon/2} \lambda_n |c_n|^2 \ge 1 - \delta - \varepsilon/2 \Big\}.$$

From analysis, there are good estimates for the number of eigenvalues larger than $\varepsilon/2$. So S_δ^ε is a subset of the unit ball in $\mathbb{C}^N$ for the appropriate N. And the covering number for the unit ball in $\mathbb{C}^N$ is easy to compute.

Once one has the covering numbers, one csn apply the metric entropy argument. The result one gets is that the sampling inequality holds, except for an event whose probability goes to 0 exponentially fast in the number of samples.

Some open problems

1. A practical algorithm is not known for reconstructing f in the infinite dimensional case.

2. In the finite dimensional case, our results are asymptotic in the number of samples r. What if r is just barely large enough to insure invertibility of the Toeplitz matrix? What impact does increasing r by 1, or by a factor of 2, have?