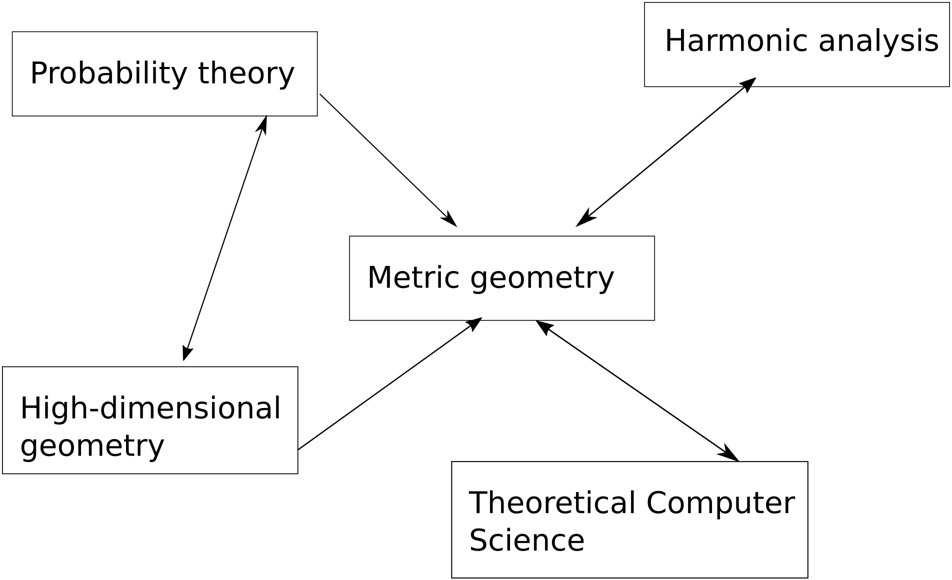
The work of Assaf Naor

Terence Tao

University of California, Los Angeles

Feb 13, 2020


Probability theory
Harmonic analysis
Metric geometry
High-dimensional geometry
Theoretical Computer Science

- One of the themes of Naor's research is the Ribe program that seeks to develop the close analogy between the geometry of Banach spaces, and the geometry of metric spaces.
- Of course, every Banach space $(X, \|\|_X)$ is also a metric space (X, d_X) if one uses the metric $d_X(x, y) := \|x - y\|_X$. But the metric space (X, d_X) loses much of the "linear" structure of the original Banach space $(X, \|\|_X)$.
- For instance, the Banach space has an addition operation $+ : X \times X \to X$, while the metric space is not obviously equipped with one.

Remarkably, the "nonlinear" metric space (X, d_X) still "remembers" pretty much all of the "linear" Banach space structure!

- For instance, it is not difficult to show (cf. the Mazur-Ulam theorem, 1932) that if $(X, \|\|_X), (Y, \|\|_Y)$ are two Banach spaces whose metric spaces $(X, d_X), (Y, d_Y)$ are isometric, then the isometry is affine, and $(X, \|\|_X), (Y, \|\|_Y)$ are isomorphic as Banach spaces.
- A deeper fact is Ribe's theorem from 1976: if $(X, \|\|_X), (Y, \|\|_Y)$ are two Banach spaces whose metric spaces $(X, d_X), (X, d_Y)$ are uniformly homeomorphic (i.e., there is a homomorphism $f : X \to Y$ such that f, f^{-1} are uniformly continuous), then $(X, \|\|_X), (Y, \|\|_Y)$ are locally representable in each other (every finite-dimensional subspace of X is equivalent to one of Y and vice versa, with constants independent of the dimension).

- Because of Ribe's theorem, we know that any property of a Banach space $(X, \|\|_X)$ which is "essentially finite-dimensional" in the sense that it is preserved by local bi-representability, can in principle be expressed purely in terms of the structure of the underlying metric space (X, d_X) (up to uniform homeomorphisms). However, the known proofs of Ribe's theorem do not give a satisfactory way to make such expressions explicit.
- In 1985, Bourgain proposed the Ribe program to firstly locate "good" metric descriptions of as many "essentially finite-dimensional" properties of Banach spaces as possible; and then to use these descriptions to generalise the theory of these properties to wider classes of metric spaces than those arising from Banach spaces. Thus, this program would extend the "linear" theory of Banach spaces to the "nonlinear" setting of metric spaces.

- For instance, an important invariant of a Banach space $(X, \|\|_X)$ is its Radamacher type. We say that X has Radamacher type p if there is a randomised triangle inequality

$$\mathbf{E}\|\sum_{i=1}^n \epsilon_i x_i\|_X \leq C_p(\sum_{i=1}^n \|x_i\|_X^p)^{1/p}$$

for all elements $x_1, \ldots, x_n \in X$ and some constant C_p, where $\epsilon_1, \ldots, \epsilon_n \in \{-1, +1\}$ are independent Bernoulli random variables of mean zero.
- As a key example, $L^p(\mathbf{R})$ is of Radamacher type $\min(p, 2)$ and no better (which implies for instance that the spaces $L^p(\mathbf{R})$, $p \leq 2$ are all inequivalent).
- This sort of type information is very useful for many applications in high dimensional geometry, analysis, and probability.

- It is easy to check that two Banach spaces that are locally representable in each other have the same Radamacher type. Thus, by Ribe's theorem, it should be possible to express the Radamacher type of a Banach space $(X, \|\|_X)$ purely in terms of the metric space (X, d_X) in a manner which is preserved by uniform homeomorphisms. But how?

- One can show that the property of a Banach space $(X, \|\|_X)$ having Radamacher type p is equivalent to an estimate of the form

$$\mathbf{E} d_X(f(\epsilon), f(-\epsilon))^p \leq C_p \sum_{i=1}^n \mathbf{E} d_X(f(\epsilon), f(\pi_i \epsilon))^p \quad (1)$$

  for some C_p and any linear map $f : \{-1, +1\}^n \to X$, where $\epsilon = (\epsilon_1, \ldots, \epsilon_n)$ is a uniform random variable on $\{-1, +1\}^n$ and $\pi_i : \{-1, +1\}^n \to \{-1, +1\}^n$ is the reflection across the $\epsilon_i = 0$ hyperplane (i.e., it replaces ϵ_i with $-\epsilon_i$).
- In 1969, Enflo made the bold proposal to extend this notion to arbitrary metric spaces by dropping the requirement of linearity. Thus, let us say that a metric space (X, d) has Enflo type p if (1) holds for some C_p and *all* maps $f : \{-1, +1\}^n \to X$. Slight variants of Enflo type were also later proposed by Gromov in 1983 and by Bourgain-Milman-Wolfson in 1986.

- It is obvious that if a Banach space has Enflo type *p*, it also has Radamacher type *p*.
- Enflo asked the converse question of whether every Banach space of Radamacher type *p* is also of Enflo type *p*, which would be a satisfactory completion of the Ribe program for the concept of Radamacher type. This remains open, although thanks to the works of Bourgain-Milman-Wolfson and Pisier in 1986 we know that such spaces at least have Enflo type $p - \varepsilon$ for any $\varepsilon > 0$.

Fully resolving Enflo's problem has been quite difficult, but Naor has obtained several partial results towards this problem:

- In 2002, Naor and Schechtman solved Enflo's problem under the additional hypothesis that X is a UMD (unconditional martingale difference) Banach space.
- In 2007 Mendel and Naor were able to solve a variant of Enflo's problem for "scaled Enflo type" - a concept which is also purely metric in nature, though less useful than Enflo type for applications.
- In 2008, Mendel and Naor solved the analogue of Enflo's problem for the complementary notion of Radamacher cotype (which is well adapted to study L^p type spaces for $p \geq 2$ rather than $p \leq 2$).
- In 2012 Hytonen and Naor also solved Enflo's problem for Banach lattices.

These arguments use non-trivial estimates from harmonic analysis, such as Riesz transform inequalities.

- We have focused the first part of Ribe's program - finding descriptions of Banach space concepts, such as Radamacher type, that are valid for arbitrary metric spaces. Now we turn to the second part of Ribe's program - finding analogues of Banach space theorems that hold for arbitrary metric spaces.
- Naor has several results of this nature regarding metric notions of type and cotype, but we will now look at some of his other work in this area, regarding nonlinear versions of Dvoretzky's theorem.

In 1961, Dvoretzky solved a conjecture of Grothendieck by establishing

Dvoretzky's theorem

Let k be a natural number and $D > 1$. If the dimension n of a normed vector space $(V, \|\|_V)$ is sufficiently large depending on k and D, then V has a k-dimensional subspace W which embeds into a Hilbert space H with distortion D (thus there is a linear map $T : W \to H$ with $A\|x\|_V \leq \|Tx\|_H \leq DA\|x\|_V$ for all $x \in W$ and some $A > 0$).

Roughly speaking, any large Banach space contains a large subspace that is approximately Euclidean in nature. The modern proof of this theorem is probabilistic in nature; after suitable normalisation, a random subspace of the space V will work.

Inspired by the Ribe program, in 1986 Bourgain-Figiel-Milman established an analogous theorem for (finite) metric spaces:

Nonlinear Dvoretzky's theorem

Let k be a natural number and $D > 1$. If the cardinality n of a metric space (X, d) is sufficiently large depending on k and D, then X has a subset Y of cardinality k which embeds into a Hilbert space H with distortion D (thus there is a map $T : Y \to H$ with $Ad(x, y) \leq \|Tx - Ty\|_H \leq DAd(x, y)$ for all $x, y \in Y$ and some $A > 0$).

It is then natural to ask for the optimal dependence of parameters for n, k, D. That is to say: given a size n and a distortion D, what is the largest size k for which one can guarantee a subset Y that embeds into Hilbert space with the given distortion?

After work of Mendel-Naor (2007), Naor-T. (2010), Mendel-Naor (2013), we now have a satisfactory answer to this question:

Quantitative nonlinear Dvoretzky's theorem

Let $0 < \varepsilon < 1$, and let (X, d) be a metric space of cardinality n. Then there is a subset Y of X of cardinality $n^{1-\varepsilon}$ that embeds into Hilbert space with distortion $O(1/\varepsilon)$.

The dependence of constants here is basically optimal. There is also a continuous variant of this result, where one works with infinite metric spaces and uses Hausdorff dimension in place of cardinality. The result is proven by randomly “fragmenting” the space X to extract a Cantor-like subset Y.

- In fact, the proof method gives something stronger; the subset Y not only embeds into a Hilbert space, but the metric on this space is comparable to a very simple type of metric d_Y known as an ultrametric, in which the triangle inequality $d_Y(x,z) \leq d_Y(x,y) + d_Y(y,z)$ is upgraded to $d_Y(x,z) \leq \max(d_Y(x,y), d_Y(y,z))$.
- Indeed, nowadays the nonlinear Dvoretzky theorem can be viewed as a corollary of a more powerful statement of Mendel and Naor known as the ultrametric skeleton theorem.

The ultrametric skeleton theorem has a number of further applications in analysis, probability, and theoretical computer science. For instance, it gave a new proof of the difficult majorising measures theorem of Talagrand that computes (up to constants) the expected value of the supremum of a Gaussian process.

Here is a theoretical computer science application, due to Mendel and Naor:

Approximate distance oracles

If $0 < \varepsilon < 1$, then any metric space (X, d) of cardinality n can be preprocessed in time $O(n^2)$ to yield a data structure of size $O(n^{1+\varepsilon})$, such that the distance between any two points in X can be computed in $O(1)$ time using this data structure up to a multiplicative error of $O(1/\varepsilon)$.

- Another application of metric embedding results to theoretical computer science comes from the work of Naor and his coauthors on the embeddability properties of the Heisenberg group

$$H = \begin{pmatrix} 1 & \mathbf{R} & \mathbf{R} \\ 0 & 1 & \mathbf{R} \\ 0 & 0 & 1 \end{pmatrix},$$

  endowed with the Carnot-Carathéodory metric.
- The Heisenberg group arose first in quantum mechanics, and since played an important role in harmonic analysis, several complex variables, and ergodic theory.
- It is the simplest non-trivial example of a nilpotent Lie group - a continuous group that just barely fails to be Euclidean.

- In contrast to the Euclidean metric, the Carnot-Carathéodory metric is anisotropic.
- For instance, the distance between $\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}$ and the identity matrix is comparable to $|x| + |y| + |z|^{1/2}$, in contrast with the Euclidean (Frobenius) distance of $\sqrt{|x|^2 + |y|^2 + |z|^2}$.

- In 1996, Semmes observed (as a corollary of a 1989 differentiation theorem of Pansu) that it was impossible to embed the Heisenberg group H into a Hilbert space with bounded distortion.
- However, finite subsets of H can of course still be embedded. For example, one can try to embed the lattice points B_R of a ball of radius R in H into a Hilbert space, or in a Euclidean space.
- By work of Lee-Naor (2006), Naor-Neiman (2012), Austin-Naor-Tessera (2013), and Naor-Lafforgue (2014) (see also T. (2019)), a satisfactory understanding of the embedding problem was obtained.

- For instance, the optimal distortion for embedding B_R into a Hilbert space is now known to comparable to $\sqrt{\log R}$ for large R.
- The lower bound is established from a certain variant of an isopemetric inequality on the Heisenberg group; the upper bound follows from some general metric embedding theorems of Assuoad (1983).

In 2014, Naor and Young used these Heisenberg isoperimetric inequalities to give a new result on the sparsest cut problem in theoretical computer science:

Sparsest cut problem

Suppose one is given a graph $G = (V, E)$. What is the minimal value of the quantity

$$\frac{\#\{(v_1, v_2) \in V_1 \times V_2 : \{v_1, v_2\} \in E\}}{(\# V_1)(\# V_2)}$$

where $V = V_1 \uplus V_2$ ranges over non-trivial partitions of the vertex set V?

Informally, one wants to break a given graph into two non-empty pieces while cutting as few edges as possible (in a relative sense). This problem is known to be NP-hard (even to approximate), and is related to deep open problems in theoretical computer science such as the Unique Games Conjecture.

While the sparsest cut problem is difficult to solve, Goemans and Linial proposed a semidefinite relaxation of the problem which can be computed (to $o(1)$ precision) in polynomial time. To motivate this relaxation, let us first state a weighted generalisation of the sparsest cut problem:

Weighted sparsest cut problem

Suppose one is given non-negative weights c_{ij}, d_{ij} for $i, j = 1, \ldots, n$. What is the minimal value of the quantity

$$\frac{\sum_{i=1}^n \sum_{j=1}^n c_{ij}|v_i - v_j|^2}{\sum_{i=1}^n \sum_{j=1}^n d_{ij}|v_i - v_j|^2}$$

where $v_1, \ldots, v_n$ takes values in $\{0, 1\}$ and the denominator is non-zero?

The original sparsest cut problem corresponds to the special case when $d_{ij} = 1$ and c_{ij} are the coefficients of the adjacency matrix of the graph.

One can relax this problem by letting $v_1, \ldots, v_n$ take values in a Hilbert space rather than in $\{0, 1\}$, but imposing a negative type condition:

Relaxed sparsest cut problem

Suppose one is given non-negative weights c_{ij}, d_{ij} for $i, j = 1, \ldots, n$. What is the minimal value of the quantity

$$\frac{\sum_{i=1}^n \sum_{j=1}^n c_{ij} \|v_i - v_j\|^2}{\sum_{i=1}^n \sum_{j=1}^n d_{ij} \|v_i - v_j\|^2}$$

where $v_1, \ldots, v_n$ takes values in a Hilbert space and obeys the negative type condition $\|v_i - v_j\|^2 \leq \|v_i - v_k\|^2 + \|v_k - v_j\|^2$ for all i, j, k?

- The relaxed sparsest cut problem can be solved (with $o(1)$ precision) in polynomial time using the technique of semidefinite programming (the problem is equivalent to a convex optimization problem involving semidefinite matrices).
- Unfortunately the solution to the relaxed sparsest cut problem can be lower than the solution to the weighted sparsest cut. The ratio between the two is known as the integrality gap for this relaxation.
- Goemans and Linial conjectured that this gap was bounded, but this was unfortunately disproven by Khot and Vishnoi in 2015.

- In 2008, Arora-Lee-Naor showed that the integrablity gap is at most $\log^{1/2+o(1)} n$, making this algorithm the most accurate polynomial time algorithm currently known for the sparsest cut problem.
- In 2014, Naor-Young used isoperimetric inequalities for the (five-dimensional) Heisenberg group to show that the integrality gap is also at least $c \log^{1/2} n$ for some $c > 0$. (The connection, roughly speaking, comes from the fact that the Carnot-Carathéodory metric on the Heisenberg group is of negative type.)

Now we turn to one last result of Naor, in which Hilbert space methods were used to solve a classical problem in probability.

- Given a random variable X on the real line with density function f, its entropy is defined by the formula

$$\mathrm{Ent}(X) := -\int_{\mathbf{R}} f(x) \log f(x)\, dx.$$

- In the 1940s, Shannon essentially established the inequality

$$\mathrm{Ent}\left(\frac{X_1 + X_2}{\sqrt{2}}\right) \geq \mathrm{Ent}(X)$$

whenever X_1, X_2 were independent copies of X (assuming of course that the entropy is well-defined). (A detailed proof was first given by Stam in 1959.)

- Iterating Shannon's inequality, one can conclude that the entropies

$$\mathrm{Ent}(\frac{X_1 + \cdots + X_n}{\sqrt{n}})$$

were monotone increasing in n, *as long as n was restricted to be a power of* 2.
- In 1978, Lieb conjectured that this monotonicity held for all n. (Note that this is consistent with the central limit theorem, since the normal distribution maximises the entropy amongst all distributions of a given variance.)
- This could not be established by Shannon's methods, and the first solution was given in an influential paper of Artstein, Ball, Barthe, and Naor in 2004.

The argument of Artstein-Ball-Barthe-Naor uses several ingredients that are now standard in the literature.

- The first step is to replace the "logarithmic" expression $\mathrm{Ent}(X) = \int_{\mathbf{R}} f \log f$ with a more tractable "quadratic" expression, namely the Fisher information $J(X) := \int_{\mathbf{R}} \frac{(f')^2}{f}$. The relationship between the two was observed by de Bruijn and by Bakry-Emery in the 1980s; basically, the Fisher information is the rate of increase in the entropy when one applies the Ornstein-Uhlenbeck process.
- From this relationship and the fundamental theorem of calculus, one can deduce the monotonicity of Shannon entropy from an analogous monotonicity for Fisher information.

One can express the Fisher information of random variables such as $X_1 + \cdots + X_{i-1} + \cdots + X_{i+1} + \cdots + X_n$ in terms of various "quadratic" integrals on $\mathbf{R}^n$, involving a function g_i that does not depend on the x_i coordinate, but has mean zero when integrated against a product measure $f^n\ dx$ (the distribution function of $(X_1, \ldots, X_n)$. The claim then boils down to the following basic application of Hilbert space geometry:

Improved triangle inequality

Let $g_1, \ldots, g_n$ be functions in $L^2(f^n\ dx)$, with each g_i of mean zero and independent of the x_i variable. Then

$$\|g_1 + \cdots + g_n\|^2_{L^2(f_n\ dx)} \leq (n-1)(\|g_1\|^2_{L^2(f^n\ dx)} + \cdots + \|g_n\|^2_{L^2(f^n\ dx)}).$$

Note that the triangle inequality and Cauchy-Schwarz would give this inequality but with the $n-1$ factor worsened to n. The crucial improvement to $n-1$ comes from the different invariance properties of each of the $g_1, \ldots, g_n$ which give a tiny bit of "orthogonality" between the $g_1, \ldots, g_n$.

Thanks for listening!