Convexity, Complexity, and High Dimensions

Stanisław Szarek

Université Pierre et Marie Curie, Paris, France
Case Western Reserve University, Cleveland, USA

Madrid, August 25, 2006

http://www.case.edu/artsci/math/szarek/
http://www.institut.math.jussieu.fr/~szarek/

Abstract : We discuss metric, algorithmic, and geometric issues related to broadly understood complexity of high dimensional convex sets. The specific topics include metric entropy and its duality, derandomization of constructions of normed spaces or of convex bodies, and different fundamental questions related to the geometric diversity of such bodies, as measured by various isomorphic (as opposed to isometric) invariants.

Plan of the talk

- introductory remarks about the field (for non-specialists)
- *metric complexity*: the metric entropy of a convex set and related duality issues
- *geometric complexity*: diversity of convex sets
- *algorithmic complexity* of convex sets, *de-randomization*

"Case studies" centered around a concept and related fundamental questions or conjectures.

Introductory remarks, some notation

Perspective : asymptotic geometric analysis (AGA)

(or geometry of Banach spaces, or local theory of normed/Banach spaces, or geometric functional analysis (GFA), or high-dimensional convex geometry)

Setting : fixed finite but usually high dimension

Objects of study : convex bodies, normed spaces, linear operators, matrices,. . .

Typical objective :

$$cf \le \text{invariant} \le Cf$$

- f : an explicit function of the parameters
- $C, c > 0$: constants independent of the particular instance of the problem

This is elsewhere denoted as "invariant $= \Theta(f)$"

In GFA : "invariant $\simeq f$," or worse "$\sim f$"

(can be (mis)understood as $\lim \frac{\text{invariant}}{f} = 1$)

Dictionary: *geometric vs. functional-analytic objects*

- normed space $X \leftrightarrow$ its unit ball B_X
- convex body $K \subset \mathbb{R}^n \leftrightarrow$ its gauge $\|\cdot\|_K$

  (i.e., $\|x\|_K := \inf\{t > 0 : x \in tK\}$)

and so if K is 0-symmetric (usually assumed)

- $K \leftrightarrow$ the normed space $(\mathbb{R}^n, \|\cdot\|_K)$
- pairs of convex sets $\leftrightarrow$ operators

$(K, B) \quad \leftrightarrow \quad \mathrm{Id} : (\mathbb{R}^n, \|\cdot\|_K) \to (\mathbb{R}^n, \|\cdot\|_B)$

- duals of normed spaces $\leftrightarrow$ polars of sets

  the polar of S : $S^\circ := \{x : \forall y \in S \quad \langle x, y \rangle \leq 1\}$

Fundamental concept : Banach-Mazur distance

$$d(K,B) := \mathsf{inf}\{\lambda > 0 : \exists u \in GL(n) \ \ K \subset u(B) \subset \lambda K\}$$

or, in terms of normed spaces,

$$d(X,Y) := \mathsf{inf}\{\|u\| \cdot \|u^{-1}\| \ : \ u \in \mathcal{L}(X,Y)\}$$

(maximum over all isomorphisms)

Groups other than $GL(n)$ are also considered

Metric Complexity

Covering numbers, metric entropy functional, packing numbers

Definition *If X is a metric space, $K \subset X$ - a compact set and $\varepsilon > 0$, the covering number $N(K, \varepsilon)$ is the minimal number of balls of radius ε whose union covers K. We call $\varepsilon \to \log N(K, \varepsilon)$ the metric entropy functional of K.*

If X is a *normed space*, this is about covering K by *translates* of $B = \varepsilon B_X$.

In a *vector space*, $N(K, B)$ is the covering number of a subset K by translates of a convex body B.

A closely related concept: *packing numbers* $M(K, B)$, the *maximal* number of *disjoint* translates of B by elements of K. We have $N(K, 2B) \leq M(K, B) \leq N(K, B)$ if B is symmetric.

Motivation, applications

- the immediate geometric framework
- $\log_2 N(K, \varepsilon)$ = the complexity of K, in bits, at the level of resolution ε
- in coding theory, a self-correcting code is essentially a packing
- in probability theory, metric entropy functionals are very closely related to various invariants of stochastic processes
- metric entropy estimates are crucial in numerous constructions in classical and functional analysis and in operator theory

The duality conjecture (Pietsch 1972)

Conjecture *There exist numerical constants* $a, b \geq 1$ *such that for any dimension* n *and for any two symmetric convex bodies* K, B *in* $\mathbb{R}^n$ *one has*

$$b^{-1} \log N(B^\circ, aK^\circ) \leq \log N(K, B),$$

where $S^\circ := \{x : \forall y \in S \quad \langle x, y \rangle \leq 1\}$.

Remarks

• The bipolar theorem ($(S^\circ)^\circ = S$) allows to exchange the role of the sets K, B and their polars.

• Rescaling allows to introduce an extra parameter ε and to consider $N(K, \varepsilon B)$ etc.

So we are asking whether the two metric entropy functionals, $\log N(K, \varepsilon B)$ and $\log N(B^\circ, \varepsilon K^\circ)$, are – in the appropriate sense – equivalent, uniformly over the dimension of the problem, K, B and $\varepsilon > 0$.

The original functional-analytic formulation

Set, for a compact linear operator $u : X \to Y$

$$N(u, \varepsilon) := N(uB_X, \varepsilon) = N(uB_X, \varepsilon B_Y)$$

The duality conjecture asserts that

$$b^{-1} \log N(u^*, a\varepsilon) \le \log N(u, \varepsilon)$$

uniformly over X, Y, u and ε.

While the general conjecture remains open, there is a substantial progress to report.

Progress report

Theorem 1 *The duality conjecture holds if one of the spaces* X, Y *is a Hilbert space or, equivalently, if one of the sets* K, B *is an ellipsoid.*

[Artstein, Milman, S. (2004)]

Theorem 2 *The duality conjecture holds if one of* X, Y *is a* K*-convex space.*

[A., M., S., Tomczak-Jaegermann, (2004)]

K-convexity

K-convexity $\Leftrightarrow$ absence of large subspaces resembling finite-dimensional ℓ_1-spaces

Other interesting descriptions:

- nontrivial type $p > 1$
- boundedness of the Rademacher (or Gaussian) projection on $L^2(X)$.

K-convex spaces $\supset$ L_p-spaces for $1 < p < \infty$ (classical or non-commutative)

K-convexity can be quantified $\rightarrow$ "K-convexity constant," well behaving with respect to standard functors of functional analysis

Comments on various degrees of generality

- *both* X and Y are Hilbert spaces:

- metric entropy of a Hilbert space operator depends only on its singular values, and the singular values of an operator and its adjoint coincide

- in the "geometric" formulation, K, B (and hence K°, B°) are ellipsoids, with the pair (K, B) affinely equivalent to the pair (B°, K°) (in this order!)

- so duality results hold with $a = b = 1$

- *just one* of the spaces is a Hilbert space:

- the duality theorem expresses what seems to be a rather fundamental property of *all* convex subsets of the Hilbert space.

- the general setting

- a statement about *arbitrary* convex subsets of a *general* normed space.

Related concepts, further problems

A new concept: *convexified packing*

Definition *A sequence* $x_1, \ldots, x_m$ *is called a* convexified B-packing *if*

$$(x_j + B) \ \cap \ \text{conv} \bigcup_{i<j} (x_i + B) \ = \ \emptyset$$

for $j = 2, \ldots, m$. *Next, the* convexified packing number $\widehat{M}(K, B)$ *is the maximal length of a sequence in* K *that is a convexified* B*-packing.*

Unlike for the usual packing or covering, the order is important here.

For this modified notion, the duality holds

Proposition 1 *If $K, B \subset \mathbb{R}^n$ are convex symmetric bodies, then $\widehat{M}(K,B) \le \widehat{M}(B^\circ, K^\circ/2)^2$.*

This is (essentially) a consequence of a Hahn-Banach type separation theorem. Another ingredient of the proof is

Proposition 2 *In a Hilbert space (or a K-convex space), the packing numbers $M(K,\cdot)$ and $\widehat{M}(K,\cdot)$ are equivalent (in the appropriate sense) if the diameter of K and the resolution are comparable.*

Conclusion : the duality of $M(\cdot,\cdot)$, or $N(\cdot,\cdot)$, holds whenever the diameter of K and the resolution are comparable. To obtain the theorems, we use now

Proposition 3 *For a given space X, if the duality conjecture holds for the resolution $\varepsilon = 1$ and all $K \subset X$ verifying $K \subset 4B_X$, then it holds (perhaps with different a, b) for all K and all $\varepsilon > 0$.*

In geometric terms : it is enough to prove the original conjecture for K, B with $K \subset 4B$

Some of the ideas behind Propositions 1 and 2 were already present in the much earlier paper [Bourgain, Pajor, S., Tomczak-Jaegermann (1989)].

Several natural problems remain open:

Are the functionals $M(K, \cdot)$ *and* $\widehat{M}(K, \cdot)$ *always equivalent? equivalent for bounded sets (i.e., if the diameter and the resolution are comparable)? in specific spaces such as* ℓ_1*?*

Positive results in this spirit imply new results on duality.

Geometric Diversity

The theme considered in this part is

To what extent do convex bodies of the same dimension n exhibit common geometric features? In particular, to what extent do they resemble the arguably most regular body, the Euclidean ball B_2^n?

[resp., normed spaces, the Euclidean space ℓ_2^n]

Answer #1 :

The Minkowski compactum and its size

The n*th Minkowski compactum* : the set of (classes of affinely equivalent) symmetric convex bodies in $\mathbb{R}^n$ (or, equivalently, of classes of isometric n-dimensional normed spaces), endowed with the Banach-Mazur distance.

$$\max d(K, B_2^n) = \sqrt{n}, \ \max d(K, B) = \Theta(n)$$

[John (1948), Gluskin (1981)]; the maxima are taken over all 0-symmetric n-dimensional convex bodies K, B.

Answer #2 :

Quotient of a subspace theorem

Every n*-dimensional normed space admits a* subspace of a quotient *whose dimension is* $n/2$ *and whose Banach-Mazur distance to the Euclidean space does not exceed some universal bound* C *(resp.,* $\geq \theta n$ *for* $\theta \in (0,1)$*,* $C = C(\theta) \to 1$ *as* $\theta \to 0^+$*).*

[Milman (1985)] In other words, every n-dimensional symmetric convex body admits a central section and an affine image of that section that is of dimension $\geq n/2$ and that is C-equivalent to a Euclidean ball.

Conclusion : The answer dramatically depends on the invariant that is relevant in a particular context or application.

In what follows we will be mostly interested in the circle of ideas related to the quotient of a subspace theorem and its aftermath.

A digression : The Dvoretzky theorem

Every n-dimensional symmetric convex body K admits a nearly spherical central section whose dimension is of order $\log n$.

[Dvoretzky (1961), Milman (1971)] In general, the logarithmic order is optimal.

The original “almost isometric” formulation:

Given $\varepsilon > 0$,t there exists a $(1+\varepsilon)$-spherical section of K of dimension $k \geq c(\varepsilon) \log n$.

Optimal dependence between k, ε and n : unknown

Further digression : The Knaster problem

Given continuous function on the sphere in $\mathbb{R}^n$ and a configuration of n points on that sphere, is there a rotation of the configuration on which the function is constant?

[Knaster, *The New Scottish Book* (1946)]

An affirmative answer would easily imply the Dvoretzky theorem with very good dependence on the parameters.

Theorem 3 *The answer to the Knaster problem is "no" if the dimension n is large.*

[Kashin, S., CRAS (2003)]

Large: answer is unknown only for n between 4 and $\approx$60 [Hinrichs, Richter (2005)].

However, the following may still be true, and its consequences would be just as good for precise forms of the Dvoretzky theorem.

Is the answer to the Knaster problem affirmative if the configuration $\mathcal{C} \subset S^{n-1}$ consists of (say) $\leq \sqrt{n}$ points? or if $\sum_{x\in\mathcal{C}} x \otimes x = \lambda P$?

Back to the quotient of a subspace theorem

ICM 1986 - Milman's program to further develop the "proportional theory." 1st problem:

Does every n-dimensional normed space admit a quotient of dimension $\geq n/2$ whose cotype 2 constant is bounded by a universal numerical constant?

There was strong "experimental evidence" that such a result may be true, and the answer was affirmative if one replaced the cotype 2 property by a closely related "bounded volume ratio property."

Cotype 2 and Cotype 2 constants

Cotype 2 constants of a space X is the smallest C (if it exists) such that, for every finite sequence (x_j) in X one has

$$\mathrm{Ave}_{\pm}\|\sum_j \pm x_j\|^2 \geq C^{-2}\sum_j \|x_j\|^2$$

(relaxed parallelogram inequality). If such a constant exists, the space is said to have cotype 2.

Examples: classical and non-commutative L_p-spaces for $p \in [1,2]$.

Negative answers: the saturation phenomenon

Theorem 4 $\exists c > 0$ *such that whenever* k, n *are positive integers with*

$$k \leq c\sqrt{n}$$

and W *is a* k*-dimensional normed space, then there exists an* n*-dimensional normed space* X *such that every quotient* Y *of* X *with* $\dim Y \geq n/2$ *contains a contractively complemented subspace isometric to* W*.*

[S., Tomczak-Jaegermann (2004), (2005)]

Relation to Milman's problem

Apply the Theorem with $W = \ell_\infty^k$ for largest possible k, that is with k of order $\sqrt{n}$.

Then every quotient Y of X with $\dim Y \geq n/2$ contains ℓ_∞^k, and so its cotype 2 constant is at least $\sqrt{k} \simeq n^{1/4}$

Another interesting point : the $\sqrt{n}$ threshold, which appears in many places in the theory

Why "saturation"?

Theorem 4 says that the space X is so "saturated" with subspaces isometric to W (copies of W), that such subspaces persist in every "sufficiently large" quotient of X.

Complementability $\Rightarrow$ "every quotient Y of X" can be replaced by "every subspace Y of X."

Thus, in general, passing to large subspaces *or* large quotients can not erase k-dimensional features of a space if k is below certain threshold value.

Any future for the "proportional theory"?

Some positive results in the same direction are still possible (Theorem 4 clarifies which). Here is a sample problem :

Given n*-dimensional normed space* X*, is there a subspace* $Y \subset X$ *with* $m := \dim Y \geq n/2$ *and a basis* $y_1, \ldots, y_m$ *of* Y *such that, denoting by* $y_1^*, \ldots, y_m^*$ *the dual basis of* Y^* *we have*

$$\mathrm{Ave}_{\pm} \| \sum_i \pm y_i \| \cdot \mathrm{Ave}_{\pm} \| \sum_j \pm y_j^* \| \leq Cm \ ?$$

[May be true without passing to a subspace.]

Algorithmic Complexity, Derandomization

The theme of this part is : *How difficult is it to describe a given convex body K?*

Prime example : algorithmic complexity of the *membership oracle*: How difficult is it to decide whether a point belongs to K?

Another class of issues : if the existence of K with certain properties is shown by a non-constructive or probabilistic proof; is it possible to give an *explicit example*, or an efficient *derandomized algorithm*?

The probabilistic method in functional analysis : a glorious history

- Dvoretzky theorem
- Kashin decomposition
- Gluskin random normed spaces
- random factorizations of operators
- poorly reducible matrices
- matrix models in free probability
- saturation phenomenon
- compressed sensing
- . . .

Methodology : produce a random variable whose values were convex sets – or normed spaces, or operators – and to show that the required property is satisfied with nonzero probability (and typically very close to 1)

Challenge : exhibit explicit objects with similar properties

Explicit : feasible in the algorithmic sense

Typical goal : replace random matrices by *pseudorandom* matrices

Will concentrate on : *poorly reducible matrices*, *Kashin decomposition of* ℓ_1^m and *compressed sensing*

Poorly reducible matrices

Definition *A matrix M is said to be reducible if, in some orthonormal basis, it can be written as a block matrix*

$$M = \begin{bmatrix} M_1 & 0 \\ 0 & M_2 \end{bmatrix},$$

where M_1 and M_2 are square matrices.

This is equivalent to M commuting with a nontrivial orthogonal projection.

Experimental mathematics led to (≈1980)

Conjecture *As n increases to ∞, the reducible matrices become more and more dense (in norm) in the space of all $n \times n$ matrices.*

Theorem 5 *There is a computable constant $\delta > 0$ such that for every $n \geq 2$ there is an $n \times n$ (real or complex) matrix of norm one which cannot be approximated within δ by a reducible matrix.*

Probabilistic argument : [Herrero & S. (1986)]

Explicit construction, based on property τ : [Benveniste & S. (≈2002)]

The Kashin decomposition

Given $m = 2n \in 2\mathbb{N}$*, there exist two orthogonal* m*-dimensional subspaces* E_1*,* $E_2 \subset \mathbb{R}^m$ *such that*

$$\frac{1}{8}\|x\|_2 \leq \frac{1}{\sqrt{m}}\|x\|_1 \leq \|x\|_2$$

for all $x \in E_i,\ i = 1, 2$*.*

[Kashin (1977)] In other words, the space ℓ_1^{2n} is an orthogonal (in the ℓ_2^{2n} sense) sum of two *nearly Euclidean* subspaces.

This is surprising since the discrepancy between the ℓ_1 and ℓ_2 norms on $\mathbb{R}^m$ is $\sqrt{m}$.

One way to prove it : show that most of n-dimensional sections of the unit ball of ℓ_1^m are *nearly spherical* if $n = m/2$.

Challenge : find *explicit* sections with that property.

Reality check : explicit nearly Euclidean subspaces of ℓ_1^m are to date known only for dimensions $n = O(\sqrt{m})$.

Progress report on *Kashin decompositions and large nearly Euclidean subspaces of* ℓ_1^m.

• several attempts at partial derandomization of Kashin decomposition, notably [Artstein & Milman (2006)] :
- random walks on expander graphs
- requires $O(m(\log m)^\alpha)$ random bits, vs. $\simeq m^2$ entries in the random matrix approach

• explicit high-dimensional nearly spherical projections of sections of simplices [Ben-Tal & Nemirovski (2001)] via mathematical programming

Why mathematical programming?

Passing to a projection or a section are natural operations in mathematical programming; the following are two sides of the same story:

• every conic quadratic problem can be reduced to linear programming without increasing the size of the problem too much

• a k-dimensional ball is approximable within ε by a projection of a section of a simplex of dimension $O(k \log 1/\varepsilon)$

More details and backround: the talk by Arkadi Nemirovski on Monday

Compressed sensing

Problem from signal processing : efficiently reconstruct sparse vector $x \in \mathbb{R}^N$ by performing $m \ll N$ linear measurements

Sparse: supported on at most s coordinates

Surprise: possible with $m = O(s \log(N/s))$

For best rates: random measurements, Gaussian or Bernoulli $m \times N$ matrices. Many needed techniques were known in GFA starting around 1980.

What about explicit sensing matrices?

<u>*Specific questions*</u> : Find (for all m) explicit $m \times 10m$ matrices $A = A_m$ such that, for any $x \in \mathbb{R}^{10m}$ supported on at most $m/10$ coordinates the following "restricted isometry condition" holds

$$0.9\|x\|_2 \leq \|Ax\|_2 \leq 1.1\|x\|_2$$

How difficult is it to verify the restricted isometry condition? (Appears to be NP-hard.)

More on this: the talk by Emmanuel Candès (section 18) just an hour ago.

If you did not go, bad news : efficient sensing in the past hasn't been worked out yet.