Concentration, results and applications

Gideon Schechtman [1]

Department of Theoretical Mathematics,
Weizmann Institute of Science,
Rehovot, Israel

Concentration inequalities are estimates for the degree of approximation of functions on metric probability spaces around their mean. It turns out that in many natural situations one can give very good such estimates, and that these are extremely useful. We survey here some of the main methods for proving such inequalities and give a few examples to the way these estimates are used.

Contents

[1] the author was partially supported by The Israel Science Foundation founded by The Academy of Science and Humanities

Preprint submitted to Elsevier Preprint *19 April 1999*

1 Introduction: approximate isoperimetric inequalities and concentration

Let $(\Omega, \mathcal{F}, \mu)$ be a probability space where $\mathcal{F}$ is the Borel σ-field with respect to a metric d on Ω . The isoperimetric problem for the probability metric space $(\Omega, \mathcal{F}, \mu, d)$ is: Given $0 < a < 1$ and $\epsilon > 0$, what is

$$\inf\{\mu(A_\epsilon);\ \ A \in \mathcal{F}, \mu(A) = a\}?$$

and for what A is it attained. Here A_ϵ, *the ϵ neighborhood of A*, is defined as $A_\epsilon = \{\omega \in \Omega\ ;\ \ d(\omega, A) < \epsilon\}$.

There are relatively few interesting cases, some of which will be described below, in which the answer to this question is known. However, it turns out that for many applications a solution to a somewhat weaker question is sufficient: Instead of finding the actual infimum of the quantity above it is enough to find a good lower bound to $\mu(A_\epsilon)$, subject to $\mu(A) = a$. We shall refer to such a lower bound as a solution to the *approximate isoperimetric inequality* (for the given space and parameters) provided the solution is optimal except for absolute constants in the "right places".

Let us illustrate the above by the example most relevant for us. The space under question will be $(S^{n-1}, \mathcal{F}, \mu, d)$. Here S^{n-1} is the unit sphere in $\mathbb{R}^n$, d the geodesic distance, $\mathcal{F}$ the Borel σ-field and μ the normalized Haar measure (the unique probability measure on S^{n-1} which is invariant under the orthogonal group). P. Lévy [35] stated and sketched a proof of the isoperimetric inequality for this space. For every a and ϵ the minimal set is an (arbitrary) cap (i.e., a d-ball) of measure a. For a cap B of measure $\frac{1}{2}$, B_ϵ is a cap of radius $\frac{\pi}{2} + \epsilon$. A standard computation then implies that, for $a = \frac{1}{2}$, say, and any ϵ $\mu(A_\epsilon) \geq \mu(B_\epsilon) \geq 1 - \sqrt{\pi/8}\,\mathrm{e}^{-\epsilon^2 n/2}$ for any Bore set $A \subset S^{n-1}$ of measure $\frac{1}{2}$. Any inequality, $\mu(A_\epsilon) \geq 1 - \mathrm{e}^{-c\epsilon^2 n}$, holding for all A with $\mu(A) = \frac{1}{2}$, with c an absolute constant, will be referred to as an approximate isoperimetric inequality (for sets of measure $\frac{1}{2}$) in this case. As we shall see below these inequalities are extremely powerful, the value of the constant c is of little importance for the applications we have in mind, and it is much easier to prove the approximate inequality than the isoperimetric one. Moreover, several proofs of the approximate isoperimetric inequality in this case (and there are many of them) can be generalized to other situations in which no isoperimetric inequality is known.

The importance of the approximate isoperimetric inequalities stems from the fact that they imply the following *concentration phenomenon.*

In the setup above, if $\mu(A_\epsilon) \geq 1-\eta/2$ for all A with $\mu(A) \geq \frac{1}{2}$ and if $f : \Omega \to \mathbb{R}$ is a function with Lipschitz constant 1, i.e., $|f(x) - f(y)| \leq d(x,y)$ for all $x, y \in \Omega$, then $\mu(\{x\ ;\ |f(x) - M| \geq \epsilon\}) \leq \eta$.

Here M denotes the median of the function f, i.e., is defined by $\mu(\{f \geq M\})$, $\mu(\{f \leq M\}) \geq \frac{1}{2}$. This is easily seen (and first noticed by Lévy in the setting of S^{n-1}) by applying the inequality $\mu(A_\epsilon) \geq 1 - \eta/2$ once for the set $\{f \leq M\}$ and once for $\{f \geq M\}$. If η is small this is interpreted as "any such f is almost a constant on almost all of Ω". For example, in the example above we get that any Lipschitz function of constant one, $f : S^{n-1} \to \mathbb{R}$, satisfies $\mu(\{x \in S^{n-1}\ ;\ |f(x) - M| \geq \epsilon\}) \leq 2\,\mathrm{e}^{-\epsilon^2 n/2}$, which is quite counterintuitive.

The median M can be replaced by the expectation of f, $\mathbb{E}f = \int_{S^{n-1}} f\, d\mu$ provided we change the constants $2, \frac{1}{2}$ to other absolute constants. Furthermore, each of these two concentration inequalities is also equivalent (with a change of constants) to $\mu \times \mu(\{(x,y) \in S^{n-1} \times S^{n-1}\ ;\ |f(x) - f(y)| \geq \epsilon\}) \leq C\,\mathrm{e}^{-c\epsilon^2 n}$. This holds not only in this particular example but in great generality (see for example [40] V.4).

The opposite statement to the one in the second to last paragraph also holds.

Concentration implies approximate isoperimetric inequality:
If $\mu(\{x\ ;\ |f(x) - M| \geq \epsilon\}) \leq \eta$ for all Lipschitz function with constant one then $\mu(A_\epsilon) > 1 - \eta$ for all sets A of measure at least $\frac{1}{2}$.

This follows easily by considering the function $f(x) = d(x, A)$.

Vitali Milman realized the relevance of Lévy's concentration inequality to problems in Geometry and Functional Analysis. Using it he found in [39] a new proof of Dvoretzky's theorem [11] on Euclidean section of convex bodies which was much more accessible than the complicated original proof. Much more importantly, his proof is subject to vast variations and generalizations. See Section 3.1 for this proof. Except for using the idea of concentration in many instances himself, Milman also promoted the search for new concentration inequalities and new applications of them.

In this article we survey many (but not all) of the methods of proof of concentration and approximate isoperimetric inequalities. We tried to concentrate mostly on methods which are quite general or that we feel were not explored enough and should become more general. There are many different such methods with some overlap as to the inequalities they prove. Section 2 contains this survey.

In Section 3 we give a sample of applications of concentration inequalities. There are many more such applications. At some points our presentation is very sketchy since on one hand many of the applications need the introduction of quite a lot of tools not directly connected to the main theme here and on the other hand some of the subjects dealt with in this application section are also dealt with, with more details in other articles in this handbook. We hope we give enough to wet the reader's appetite to search for more in the original sources or the other articles of this handbook.

We would like to emphasize that this is far from being a comprehensive survey of the topic of concentration. This author has a soft point for new ideas in proofs and in many instances below preferred to give a glimpse into these ideas by treating a special case or a version of the relevant result which is not necessarily the last word on it rather than to give all the details on the subject.

2 Methods of proof

2.1 Isoperimetric inequalities, Brunn–Minkowski inequality

We start by stating two forms of the classical Brunn–Minkowski inequality. Here $|\cdot|$ denotes Lebesgue measure in $\mathbb{R}^n$ and $A+B$ denotes the Minkowski's addition of sets in $\mathbb{R}^n$; $A+B=\{a+b; a\in A, b\in B\}$.

Theorem 1 *(i) For every n and every two nonempty measurable subsets of $\mathbb{R}^n$ A and B,*

$$|A+B|^{1/n} \geq |A|^{1/n} + |B|^{1/n}. \tag{1}$$

(ii) For every n, every two nonempty measurable subsets of $\mathbb{R}^n$ A and B and every $0<\lambda<1$,

$$|\lambda A + (1-\lambda)B| \geq |A|^{\lambda}|B|^{1-\lambda}. \tag{2}$$

Equality in either inequality holds if and only if A and B are homothetic.

Theorem 1 has many different proofs. We refer to [51] for two of them and for an extensive discussion concerning this theorem. A variation of this theorem was proved by Prékopa and Leindler [45], [33]. One possible proof of their theorem is by induction on the dimension (see e.g. [44]). Theorem 1 is a simple consequence of this theorem.

Theorem 2 *Let f, g, h be integrable non-negative valued functions on $\mathbb{R}^n$ and let $0 < \lambda < 1$. Assume*

$$h(\lambda x + (1-\lambda)y) \geq f(x)^{\lambda} g(y)^{1-\lambda}, \text{ for all } x, y \in \mathbb{R}^n \tag{3}$$

then

$$\int_{\mathbb{R}^n} h \geq \left(\int_{\mathbb{R}^n} f \right)^{\lambda} \left(\int_{\mathbb{R}^n} g \right)^{1-\lambda}. \tag{4}$$

Theorem 1 provides a simple proof of the classical isoperimetric inequality in $\mathbb{R}^n$. To avoid restricting ourselves to bodies for whose surface area is definable we prefer to state it as: *for every $0 < a < \infty$ and every $\epsilon > 0$, among all bodies of volume a in $\mathbb{R}^n$ the ones for which the volume of A_ϵ is minimal are exactly balls of volume A.*

Maurey [38] noticed that Theorem 2 can be used to give a simple proof of the approximate isoperimetric inequality on the sphere (or equivalently for the canonical Gaussian measure on $\mathbb{R}^n$). Recently, Arias-de-Renya, Ball and Villa [4] discovered an even more direct proof of the approximate isoperimetric inequality on the sphere, using Theorem 1. Their proof actually establishes a far reaching generalization originally due to Gromov and Milman [19]. We refer to [23] for a discussion of the notion of uniform convexity. We only recall the following (equivalent) definition for the modulus of convexity δ of a normed space $(X, \|\cdot\|)$:

$$\delta(\epsilon) = \inf \left\{ 1 - \left\| \frac{x+y}{2} \right\| ;\ \|x\|, \|y\| \leq 1, \|x-y\| \geq \epsilon \right\}. \tag{5}$$

Given a norm $\|\cdot\|$ on $\mathbb{R}^n$ we consider, in the following theorem, the set $S = \{x \in \mathbb{R}^n;\ \|x\| = 1\}$ with the metric $d(x, y) = \|x - y\|$ and the Borel probability measure $\mu(A) = \frac{|\{tA;\ 0 \leq t \leq 1\}|}{|\{x;\ \|x\| \leq 1\}|}$.

Theorem 3 *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ and let δ be the modulus of convexity of $(\mathbb{R}^n, \|\cdot\|)$. Then for any Borel set $A \subset S$ and any $\epsilon > 0$,*

$$\mu(A_\epsilon) > 1 - 2\mu(A)^{-1} e^{-2n\delta(\epsilon/2)}. \tag{6}$$

Proof. Let $K = \{x;\ \|x\| \le 1\}$ and ν the normalized Lebesgue measure on K. By considering the set $\{tA;\ \frac{1}{2} \le t \le 1\}$ it is clearly enough to prove that, for $B \subset K$, $\nu(B_\epsilon) > 1 - \nu(B)^{-1} e^{-2n\delta(\epsilon)}$.

Put $C = \{x \in K;\ d(x,B) \ge \epsilon\}$ then, for all $x \in B, y \in C$, $\left\| \frac{x+y}{2} \right\| \le 1 - \delta(\epsilon)$, i.e.,

$$\frac{B+C}{2} \subset (1-\delta(\epsilon))K$$

therefore, by the Brunn–Minkowski inequality,

$$\nu(B)\nu(C) \le (1-\delta(\epsilon))^{2n} \le e^{-2n\delta(\epsilon)}.$$

■

Since for the Euclidean norm on $\mathbb{R}^n$, $\delta(\epsilon) \ge \epsilon^2/8$, we get a simple proof of the approximate isoperimetric inequality for the sphere S^{n-1} (with the Euclidean or geodesic distance and Haar measure) discussed in the introduction.

Corollary 4 *If $A \subset S^{n-1}$ and $\epsilon > 0$ then*

$$\mu(A_\epsilon) > 1 - 2\mu(A)^{-1} e^{-n\epsilon^2/16}.$$

Consequently, if $f : S^{n-1} \to \mathbb{R}$ is a function with Lipschitz constant 1 then

$$\mu(\{x\ ;\ |f(x) - M| \ge \epsilon\}) \le 8e^{-n\epsilon^2/16}.$$

There are several ways to prove the isoperimetric inequality (as opposed to approximate isoperimetric inequalities) on the sphere. Some of them generalize to give isoperimetric inequalities in other situations. We refer to Appendix I in [40] in which Gromov presents a generalization based on Levy's original proof and proves an isoperimetric inequality for Riemannian manifolds in term of their Ricci curvature. A particularly useful instance of this generalization is the case of $O(n)$ equipped with its Haar measure and Euclidean metric (i.e. the Hilbert-Schmidt norm). [13] contains a relatively easy and self contained proof of the isoperimetric inequalities on the sphere by symmetrization. It seems however to be very special to S^{n-1}. We now sketch very briefly a proof by another method of symmetrization which is not very well known and which we think deserves to be better known. It seems to have the potential to generalize to other situations. It is due to Baernstein II and Taylor [6] and is written in detail with indications towards generalizations in [7].

Sketch of proof of Levy's isoperimetric inequality. Given a Hyperplane H through zero in $\mathbb{R}^n$ we denote $S_0 = S^{n-1} \cap H$ and by S_+ and S_- the two open half spheres in the complement of H. Let also $\sigma = \sigma_H$ be the reflection with respect to H. Of course σ is an isometry with respect to the (Euclidean or geodesic) metric on S^{n-1}, it satisfies $\sigma^2 = \mathit{identity}$ and preserves the Haar measure. It also satisfies that if $x, y \in S_+$ then $d(x, y) \leq d(x, \sigma(y))$.

Given a set $A \subseteq S^{n-1}$ we define its *two point symmetrization* A^* with respect to the above decomposition as

$$A^* = [A \cap (S_+ \cup S_0)] \cup [A \cap S_- \cap \sigma(A \cap S_+)] \cup [\sigma(A \cap S_- \setminus \sigma(A \cap S_+))]$$

i.e., we "push up" elements of $A \cap S_-$ into S_+ using σ whenever there is space available. The term symmetrization seems a bit misleading since we desymmetrize as far as symmetry with respect to H is concerned. The point of course is that A^* is closer to cap than A is and in that sense is more symmetric.

Note that if A is Borel, $\mu(A^*) = \mu(A)$. It is also easy to prove that for every $\epsilon > 0$ and for every $A \subseteq S^{n-1}$

$$(A^*)_\epsilon \subseteq (A_\epsilon)^*.$$

In particular,

$$\mu((A^*)_\epsilon) \leq \mu((A_\epsilon)^*) \leq \mu(A_\epsilon).$$

The definition of the symmetrization procedure and the last property hold for any metric probability space (K, μ) admitting an isometric and measure preserving involution σ and any partition of the complement of $K_0 = \{x;\ x = \sigma(x)\}$ into K_-, K_+ provided this involution and partition satisfy the following properties: $K_+ = \sigma(K_-)$ and $d(x, y) \leq d(x, \sigma(y))$ for all $x, y \in K_+$.

To prove the isoperimetric inequality we would like to apply the operation $A \to A^*$ with respect to many hyperplanes, reach a set so that no farther application of this operation improves $\mu(A_\epsilon)$ and prove that such a set must be a cap. We'll sketch in a minute how to do that for S^{n-1} but we would like to emphasis again that this seems plausible in other situations as well and we think it deserves further investigation.

Consider the metric space $\mathcal{C}$ of all closed subsets of S^{n-1} with the Housdorff metric. Fix $A \in \mathcal{C}$ and consider the set $\mathcal{B} \subseteq C$ of all sets $B \in \mathcal{C}$ satisfying:

– For all $\epsilon > 0$ $\ \mu(B_\epsilon) \leq \mu(A_\epsilon)$ and

– $\mu(B) = \mu(A)$.

One checks that the set $\mathcal{B}$ is closed in $\mathcal{C}$.

Fix a point $x_0 \in S^{n-1}$ and let C be the closed cap centered at x_0 with measure $\mu(A)$. It is enough to prove that $C \in \mathcal{B}$. For any hyperplane H with $x_0 \notin H$ we denote by S_+ the open half sphere containing x_0. One now proves that $B \to \mu(B \cap C)$ is upper semi continuous on $\mathcal{C}$. Consequently, $\mu(B \cap C)$ attains its maximum on $\mathcal{B}$, say at B. We shall show that $B \supseteq C$ which will prove the claim. If this is not the case then $\mu(B \setminus C) = \mu(C \setminus B) > 0$. Let $x \in B \setminus C$ and $y \in C \setminus B$ be points of density of the respective sets and let H be the hyperplane perpendicular to the segment $[x, y]$ and crossing it at the midpoint $(x + y)/2$. Let $B(x, r) \subset S_-$, $B(y, r) \subset S_+$ be small balls such that $\mu(B(x, r) \cap (B \setminus C)) > 0.99\mu(B(x, r))$ and $\mu(B(y, r) \cap (C \setminus B)) > 0.99\mu(B(y, r))$. Applying the symmetrization $B \to B^*$ with respect to this hyperplane, most of $B(x, r)$ will be transferred into $B(y, r)$ while no point of $C \cap B$ is transferred to a point which is not in C. Thus, $\mu(B^* \cap C) > \mu(B \cap C)$. Since B^* also belongs to $\mathcal{B}$ we get a contradiction. ■

With a bit more effort the proof above can be adjusted to show that caps are the only solutions to the isoperimetric problem in S^{n-1}.

2.2 Martingales

Recall that for $f \in L_1(\Omega, \mathcal{F}, P)$ and for $\mathcal{G}$, a sub σ-algebra of $\mathcal{F}$, the *conditional expectation*, $\mathbb{E}(f|\mathcal{G})$, of f given $\mathcal{G}$ is the unique $h \in L_1(\Omega, \mathcal{G}, P_{|\mathcal{G}})$ satisfying

$$\int_A h dP = \int_A f dP \quad \text{for all } A \in \mathcal{G}. \tag{7}$$

(h is the Radon-Nikodým derivative of the measure $\nu(A) = \int_A f dP$ on $\mathcal{G}$ with respect to $P_{|\mathcal{G}}$.)

The correspondence $f \to \mathbb{E}(f|\mathcal{G})$ is a linear positive operator of norm one on all the spaces $L_p(\Omega, \mathcal{F}, P)$, $1 \le p \le \infty$. Some additional properties of this operator are:

– If $\mathcal{G}' \subset \mathcal{G}$ is a sub σ-algebra then $\mathbb{E}(\mathbb{E}(f|\mathcal{G})|\mathcal{G}') = \mathbb{E}(f|\mathcal{G}')$.
– If $g \in L_\infty(\Omega, \mathcal{G}, P)$ then $\mathbb{E}(fg|\mathcal{G}) = g\mathbb{E}(f|\mathcal{G})$.
– For the trivial σ-algebra $\mathcal{G} = \{\emptyset, \Omega\}$, $\mathbb{E}(f|\mathcal{G}) = \mathbb{E}f$, the expectation of f.

Given a finite or infinite sequence of σ-algebras, $\mathcal{F}_0, \mathcal{F}_1, \ldots$, a sequence of elements of $L_1(\Omega, \mathcal{F}, P)$, $f_0, f_1, \ldots$, is said to be a *martingale* with respect to

$\mathcal{F}_0, \mathcal{F}_1, \ldots$ if $f_i = \mathbb{E}(f_j|\mathcal{F}_i)$ for all $i \le j$. We shall always assume here that $\mathcal{F}_0$ is the trivial σ-algebra $\{\emptyset, \Omega\}$ and that the sequence is finite with the last terms being $f_n = f$ and $\mathcal{F}_n = \mathcal{F}$. Then, $f_i = \mathbb{E}(f|\mathcal{F}_i)$, $i = 0, 1, \ldots, n$. We also denote $d_i = f_i - f_{i-1}$, $i = 1, 2, \ldots, n$, and call the sequence $\{d_i\}_{i=1}^n$ the martingale difference sequence. One set of examples of a martingale is the following: Let X_i be a sequence of mean zero independent random variables and put $f_i = \sum_{j=0}^i X_j$, then $\{f_i\}$ is a martingale with respect to $\{\mathcal{F}_i\}$ where $\mathcal{F}_i$ is the smallest σ-algebra with respect to which $X_0, \ldots, X_i$ are measurable. In a lot of senses a general martingale resembles this particular set of examples. There are many inequalities estimating the probability of the deviation of $f = f_n$ from $f_0 = \mathbb{E}f$ in terms of the behavior of the sequence $\{d_i\}$. In the next proposition we gather some of them. (i) is due to K. Azuma [5] or[52] p. 238. (ii) and (iii) are due to Pisier [42], (ii) was first used in [24]. (iv) is a generalization to the martingale case of Prokhorov's inequality. In a somewhat weaker form it first appears in [27]. The form here is from [21]

Proposition 5 *(i) For all $t > 0$,*

$$P(\{\omega;\ |f(\omega) - \mathbb{E}f| \ge t\}) \le 2\exp(-t^2/2\sum_{i=1}^n \|d_i\|_\infty^2). \tag{8}$$

(ii) For all $1 < p < 2$ and $t > 0$,

$$P(\{\omega;\ |f(\omega) - \mathbb{E}f| \ge t\}) \le K\exp(-\delta(t/\|\{\|d_i\|_\infty\}\|_{p,\infty})^q) \tag{9}$$

where $q^{-1} + p^{-1} = 1$, K and δ depend only on p and $\|\{a_i\}_{i=1}^n\|_{p,\infty} = \max_{1\le j\le n} j^{1/p} a_j^*$ with $\{a_j^*\}$ denoting the decreasing rearrangement of the sequence $\{|a_j|\}$.

(iii) For all $t > 0$,

$$P(\{\omega;\ |f(\omega) - \mathbb{E}f| \ge t\}) \le K\exp(-\exp(\delta t/\|\{\|d_i\|_\infty\}\|_{1,\infty})) \tag{10}$$

where K and δ are absolute constants.

(iv) Put $M = \max_{1\le i\le n}\|d_i\|_\infty$ and $S^2 = \|\sum_{i=1}^n \mathbb{E}(d_i^2|\mathcal{F}_{i-1})\|_\infty$. Then, for all $t > 0$,

$$P(\{\omega;\ |f(\omega) - \mathbb{E}f| \ge t\}) \le 2\exp\left(-\frac{t}{2M}\cdot \operatorname{arc\,sinh}\left(\frac{Mt}{2S^2}\right)\right). \tag{11}$$

The proofs of these and similar inequalities are usually quite simple. Let us sketch **the proof of (i).** If $\mathcal{F}_i$ is "rich" enough, extreme points in the set $\{d \in L_\infty(\Omega, \mathcal{F}_i, P);\ \mathbb{E}(d|\mathcal{F}_{i-1}) = 0,\ |d| \le a\}$ have constant absolute value equal to a. Consequently for all $\lambda \in \mathbb{R}$,

$$\mathbb{E}(e^{\lambda d_i}|\mathcal{F}_{i-1}) \le \cosh \lambda\|d_i\|_\infty \le e^{\lambda^2\|d_i\|_\infty^2/2}. \tag{12}$$

Extending $\mathcal{F}_i$ (to become rich enough) if necessary, this inequality holds always. It follows that

$$\mathbb{E}e^{\lambda\sum_{i=1}^n d_i} = \mathbb{E}\left(\mathbb{E}(e^{\lambda\sum_{i=1}^{n-1} d_i}|\mathcal{F}_{n-1})\right)e^{\lambda^2\|d_n\|_\infty^2/2}. \tag{13}$$

Iterating this (by applying $\mathbb{E}(\cdot|\mathcal{F}_{n-2})$, then $\mathbb{E}(\cdot|\mathcal{F}_{n-3})$...) we get

$$\mathbb{E}e^{\lambda(f-\mathbb{E}f)} \le e^{\lambda^2\sum_{i=1}^n\|d_i\|_\infty^2/2}. \tag{14}$$

Applying Chebyshev's inequality we get, for positive λ,

$$\begin{aligned} P(\{\omega;\ f(\omega)-\mathbb{E}f \ge t\}) &\le P(\{\omega;\ e^{\lambda(f(\omega)-\mathbb{E}f)-\lambda t} \ge 1\}) \\ &\le e^{-\lambda t}\mathbb{E}e^{\lambda(f-\mathbb{E}f)} \le e^{-\lambda t+\lambda^2\sum_{i=1}^n\|d_i\|_\infty^2/2}. \end{aligned} \tag{15}$$

Minimizing over positive λ and repeating this with negative λ we get the result. ■

V. V. Yurinski [59] was probably the first to use martingale inequalities in the context of Banach space valued random variables. The point is that if X_i are independent Banach space valued random variables and we form the martingale $f_i = \mathbb{E}(\|\sum_{j=1}^n X_j\|\,|\mathcal{F}_i)$ then the martingale differences satisfy $|d_i| \le \|X_i\|$. This can be used to estimate the tail behavior of $\|\sum_{j=1}^n X_j\|$. Maurey [37] noticed that martingale deviation inequalities can be used to prove approximate isoperimetric inequality for the interesting case of the permutation group. We present a somewhat simplified version of his proof with some abstractization ([46] [40]).

The *length* of a finite metric space (Ω, d) is defined as the infimum of $\ell = (\sum_{i=1}^n a_i^2)^{1/2}$ over all sequences $a_1, \ldots, a_n$ of positive numbers satisfying: There exists a sequence $\{\Omega_k\}_{k=0}^n$ of partitions of Ω with

- $\Omega_0 = \{\Omega\}$ and $\Omega_n = \{\{\omega\}\}_{\omega\in\Omega}$.
- Ω_k refines Ω_{k-1}, $k = 1, \ldots, n$.
- If $k = 1, \ldots, n$, $A \in \Omega_{k-1}$, $B, C \subset A$ and $B, C \in \Omega_k$ then there is a one to one map h from B onto C such that $d(\omega, h(\omega)) \le a_k$ for all $\omega \in B$.

The two basic examples we shall deal with are the Hamming cube, H_n, and the permutation group, Π_n. The Hamming cube is the set $\{0,1\}^n$ with the metric $d((\epsilon_i)_{i=1}^n, (\delta_i)_{i=1}^n) = \#\{i;\ \epsilon_i \ne \delta_i\}$. Π_n is the set of permutations of $\{1, 2, \ldots, n\}$ with the metric $d(\pi, \varphi) = \#\{i;\ \pi(i) \ne \varphi(i)\}$. The length is smaller or equal $\sqrt{n}$ in the first case and $2\sqrt{n-1}$ in the second. Let us illustrate this in the second example. Fix $1 \le k \le n-1$ and $i_1, i_2, \ldots, i_k$ distinct elements of

$\{1, 2, \ldots, n\}$. Put

$$A_{i_1, i_2, \ldots, i_k} = \{\pi \in \Pi_n;\ \pi(1) = i_1, \ldots, \pi(k) = i_k\} \tag{16}$$

and let Ω_k be the partition whose atoms are all the sets $A_{i_1, i_2, \ldots, i_k}$ where $(i_1, i_2, \ldots, i_k)$ ranges over all $n!/(n-k)!$ possibilities. It is clear that the first two requirements from $\{\Omega_k\}_{k=0}^{n-1}$ are satisfied (with $n-1$ replacing n). To show that the third one is satisfied with $a_i = 2$ for $i = 1, \ldots, n-1$, let $A = A_{i_1, i_2, \ldots, i_{k-1}} \in \Omega_{k-1}$ and $B = A_{i_1, i_2, \ldots, i_{k-1}, r}, C = A_{i_1, i_2, \ldots, i_{k-1}, s} \in \Omega_k$ and define $h : B \to C$ by $h(\pi) = (r, s) \circ \pi$ (where (r, s) is the transposition of r and s).

We are now ready to state the main theorem of this section.

Theorem 6 *Let (Ω, d) be a finite metric space of length at most ℓ. Let P be the normalized counting measure on Ω. Then,*

(i) Let $f : \Omega \to \mathbb{R}$ satisfy $|f(x) - f(y)| \le d(x, y)$ for all $x, y \in \Omega$. Then for all $t > 0$,

$$P(\{\omega;\ |f(\omega) - \mathbb{E}f| \ge t\}) \le 2\exp(-t^2/2\ell^2). \tag{17}$$

(ii) Let $A \subset \Omega$ with $P(A) \ge 1/2$ then for all $t > 0$

$$P(A_t) \ge 1 - 2\exp(-t^2/8\ell^2). \tag{18}$$

Sketch of proof. Let $\ell = (\sum_{i=1}^n a_i^2)^{1/2}$ with a_i and Ω_i, $i = 0, \ldots, n$, as in the definition of length. Let $\mathcal{F}_i$ be the field generated by Ω_i and form the martingale $f_i = \mathbb{E}(f|\mathcal{F}_i)$, $i = 0, \ldots, n$. Note that f_i is constant on each atom B of Ω_i and that this constant is $f_{i|B} = \mathrm{Ave}_{x \in B} f(x)$. If B, C are two atoms of Ω_i contained in an atom A of Ω_{i-1} then by the third property of the sequence of partitions,

$$|f_{i|B} - f_{i|C}| = |B|^{-1} \left| \sum_{x \in B} f(x) - f(h(x)) \right| \le a_i. \tag{19}$$

Since $f_{i-1|A}$ is the average of $f_{i|B}$ over all atoms B of $\mathcal{F}_i$ which are subsets of A, we get from (19) that $|f_{i-1|A} - f_{i|C}| \le a_i$ and since this holds for all such A and C, $\|d_i\|_\infty \le a_i$. Now apply 5 (i). This proves (i). (ii) follows from (i) as explained in the introduction. ∎

Corollary 7 *Let (Ω, d) be either H_n or Π_n.*

(i) Let $f : \Omega \to \mathbb{R}$ satisfy $|f(x) - f(y)| \le d(x, y)$ for all $x, y \in \Omega$. Then for

all $t > 0$,

$$P(\{\omega;\ |f(\omega) - \mathbb{E}f| \geq t\}) \leq 2\exp(-t^2/8n). \tag{20}$$

(ii) Let $A \subset \Omega$ *with* $P(A) \geq 1/2$ *then for all* $t > 0$

$$P(A_t) \geq 1 - 2\exp(-t^2/32n). \tag{21}$$

By considering a ball in the Hamming metric it is easy to see that, except for the choice of the absolute constants involved, the result for H_n is best possible. In this case, the exact solution to the isoperimetric problem is known as well (and, for sets of measure $2^k/2^n$, is a ball) [20],[14]. For sets of measure of the form $2^k/2^n$ this can also be deduced from the method of two point symmetrization introduced in the previous section. For Π_n the solution to the isoperimetric problem is not known. However, again except for the absolute constants involved, the corollary gives the right result:

Example 8 *Let* n *be odd and define* $A \subset \Pi_{2n}$ *by*

$$A = \{\pi;\ \pi(i) \leq n \textit{ for more than } n/2 \textit{ indices } i \textit{ with } 1 \leq i \leq n\}. \tag{22}$$

Then, $\mu(A) = \frac{1}{2}$ *and for all* $k < n/2$,

$$\begin{aligned} P(A_k^c) &= \tfrac{1}{(2n)!} \textstyle\sum_{l=0}^{[\frac{n}{2}-k]+1} \binom{n}{l} \frac{n!}{(n-l)!} \frac{n!}{l!} n! \\ &= \tfrac{1}{\binom{2n}{n}} \textstyle\sum_{l=0}^{[\frac{n}{2}-k]+1} \binom{n}{l}^2. \end{aligned} \tag{23}$$

For k *with* k/n *bounded away from* 0 *and* 1, *a short computation shows that this is larger than* $e^{-\delta k^2/n}$.

It is also not hard to see that, at least for some a and t, balls are not the solution to the isoperimetric problem $\inf\{P(A_t);\ P(A) = a\}$ on Π_n. We wonder whether there is an equivalent, with constants independent of n, (and hopefully natural) metric on Π_n for which one can solve the isoperimetric problem.

The advantage of the method described above is in its generality; in principle, whenever we have a metric probability space we can estimate its length by trying different sequences of partitions and get some approximate isoperimetric inequality. In reality it turns out that in most specific problems, and in particular when the space is naturally a product space, one gets better results by other methods.

In [53] Talagrand introduced a relatively simple but quite powerful method to prove concentration inequalities which works in many situations in which the probability space is a product space with many components. The proofs, as naive as they may look, are by induction on the number of components. The monograph [57] contains many more instances in which variants of this method work. Another feature in Talagrand's work is the deviation from the traditional way of measuring distances; the "distance" of a point from a set is not always measured by a metric. We start with a small variation on the original theorem of Talagrand taken from [25].

Theorem 9 *Let $\Omega_i \subset X_i$, $i = 1, \ldots, n$, be compact subsets of normed spaces with* $\mathrm{diam}(\Omega_i) \le 1$. *Consider $\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$ as a subset of the ℓ_2 sum $(\sum_{i=1}^n \oplus X_i)_2$. Let μ_i be a probability measure on Ω_i, $i = 1, \ldots, n$, and put $P = \mu_1 \times \mu_2 \times \ldots \mu_n$. For a compact $A \subset \Omega$ denote the convex hall of A by* $\mathrm{conv(A)}$ *and for $x \in \Omega$ put $\varphi(x, A) =$* $\mathrm{dist(x, conv(A))}$ *(with respect to the metric in $(\sum_{i=1}^n \oplus X_i)_2$). Then*

(i)
$$\int e^{\varphi^2(x,A)/4} \le \frac{1}{P(A)}. \tag{24}$$

In particular, for all $t > 0$,

$$P(\{x;\ \varphi(x, A) > t\}) \le \frac{1}{P(A)} e^{-t^2/4}. \tag{25}$$

(ii) If $f : \Omega \to \mathbb{R}$ is convex and Lipschitz (with respect to the metric of $(\sum_{i=1}^n \oplus X_i)_2$) with constant 1 then

$$P(\{x;\ |f(x) - \int f| > t\}) \le 4e^{-ct^2} \tag{26}$$

for all $t > 0$ and some universal $c > 0$.

Sketch of proof. The proof of the first assertion of (i) is by induction. The second assertion of (i) and also (ii) (with a bit more effort) follow as in (15). The other theorems in this section are proved similarly. We shall illustrate the induction step. Assume that $\int e^{\varphi^2(x,A)/4} dP(x) \le \frac{1}{P(A)}$ for all compact $A \subset \Omega = \Omega_1 \times \ldots \times \Omega_n$ and let $A \subset \Omega \times \Omega_{n+1}$. For $\omega \in \Omega_{n+1}$ put $A(\omega) = \{x \in \Omega;\ (x, \omega) \in A\}$ (where, for $x = (x_1, \ldots, x_n) \in \Omega$, $(x, \omega) = (x_1, \ldots, x_n, \omega)$). Put also $B = \cup_{\omega \in \Omega_{n+1}} A(\omega)$. Fix a $y = (x, \omega) \in \Omega \times \Omega_{n+1}$ and notice that $\varphi(y, A) \le \varphi(x, A(\omega))$ provided $A(\omega) \ne \emptyset$. Also, $\varphi(y, A) \le \varphi(x, B) + 1$. From these two inequalities it is easy to deduce that, for all $0 \le \lambda \le 1$, $\varphi^2(y, A) \le \lambda\varphi^2(x, A(\omega)) + (1 - \lambda)\varphi^2(x, B) + (1 - \lambda)^2$. Using Hölder's inequality and the induction hypothesis, one gets, for all $\omega \in \Omega_{n+1}$,

$$\int\limits_{\Omega} e^{\varphi^2((x,\omega),A)/4} \leq \frac{e^{(1-\lambda)^2/4}}{P(B)} \left(\frac{P(A(\omega))}{P(B)}\right)^{-\lambda}. \tag{27}$$

We now use a numerical inequality (which can serve as a good Calculus exercise). For all $0 \leq p \leq 1$,

$$\inf_{0 \leq p \leq 1} p^{-\lambda} e^{(1-\lambda)^2/4} \leq 2 - p.$$

Using this inequality with $p = \frac{P(A(\omega))}{P(B)}$ and integrating (27) over ω, we get

$$\int\limits_{\Omega_{n+1}} \int\limits_{\Omega} e^{\varphi^2((x,\omega),A)/4} \leq \frac{1}{P(B)} \left(2 - \frac{P \times \mu_{n+1}(A)}{P(B)}\right) \leq \frac{1}{P \times \mu_{n+1}(A)}. \tag{28}$$

■

Note that if $X_i = \{-1, 1\}$ with the uniform measure for each i then by Corollary 7 the same conclusion as in Theorem 9*(ii)* holds for any (i.e., not necessarily convex) function satisfying $|f(x) - f(y)| \leq n^{-1/2} \sum |x_i - y_i|$. However, for a convex function, Theorem 9 gives a much better result since $n^{-1/2} \sum |x_i - y_i| \leq (\sum |x_i - y_i|)^{1/2}$.

The theorem above has the disadvantage that, because of the convexity assumption, it applies only to Ω_i's which lie in a linear space. This is taken care of in the next theorem from [57] which surprisingly is extremely applicable.

Given probability spaces $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i = 1, \ldots, n$, form the product space (Ω, P) with $\Omega = \prod_{i=1}^n \Omega_i$ and $P = \prod \mu_i$. For $x, y \in \Omega$ let $U(x,y)$ be the sequence in $\{0,1\}^n$ which realizes the Hamming distance between x and y, i.e., has 0 exactly in the coordinates i where $x_i = y_i$. For a subset A of Ω and for $x \in \Omega$ we set $U(x,A)$ to be the subset of $\{0,1\}^n$ consisting of all sequences $U(x,y)$ for some $y \in A$, i.e.,

$$U(x,A) = \{\{\epsilon_i\}_{i=1}^n \in \{0,1\}^n;\ \text{for some } y \in A, y_i = x_i \text{ iff } \epsilon_i = 0\}.$$

For $x \in \Omega$ and $A \subset \Omega$ let $\varphi(x,A) = \mathrm{d}(0, \mathrm{conv}(\mathrm{U}(\mathrm{x}, \mathrm{A})))$. It should be noted that, in general, $\varphi(x,A)$ is not induced by a metric. i.e., there is no metric d on Ω such that $\varphi(x,A) = \inf\{d(x,y);\ y \in A\}$. This is easily seen to be the case for $\Omega = \{0,1\}^n$ for example.

Theorem 10 *Let $A \subset \Omega$ then*

$$\int e^{\varphi^2(x,A)/4} \leq \frac{1}{P(A)}. \tag{29}$$

In particular, for all $t > 0$,

$$P(\{x;\ \varphi(x,A) > t\}) \leq \frac{1}{P(A)} e^{-t^2/4}. \tag{30}$$

Using the Hahn Banach theorem one can show that

$$\varphi(x,A) = \sup_{\sum \alpha_i^2 = 1} \inf\{\sum_{\{i;y_i \neq x_i\}} \alpha_i;\ y \in A\}. \tag{31}$$

Notice that, if h denotes the Hamming distance on Ω, i.e., $h(x,y) = \#\{i; y_i \neq x_i\}$, then formula 31 implies that $\varphi(x,A) \leq h^{1/2}(x,A)$. Using this inequality and the martingale method of section 2.2 one gets only $P(\varphi(x,A) > t^{1/2}) \leq Ce^{-ct^2/n}$ while Theorem 10 gives $P(\varphi(x,A) > t^{1/2}) < 4e^{-t/4} \leq 4e^{-t^2/4n}$ for t in the relevant range, $0 < t < n$. This illustrates the possible advantage of this inequality over Corollary 7 for H_n. Theorem 10 has many applications. We refer to [57] for some on them. A variant of Theorem 9 and particularly of (26) was recently proved by M. Ledoux ([30] or [31]). The difference is that the convexity assumption on f is weakened to convexity of each variable separately but the conclusion is only a one sided deviation inequality:

$$P(\{x;\ f(x) - \int f > t\}) \leq 4e^{-ct^2}. \tag{32}$$

It is unknown whether a similar lower deviation inequality also holds.

The next result was first proved by Talagrand in [54]. The original proof was very complicated but in [57] Talagrand presented a much simpler inductive proof which we shall sketch here. Consider a product probability space ($\Omega = \prod_{i=1}^n \Omega_i, P = \prod_{i=1}^n \mu_i$). Given a $q \in \mathbb{N}$ and $q+1$ elements of Ω, $x, y^1, \ldots, y^q$, we define the "Hamming distance" of x from the q-tuple $y^1, \ldots, y^q$ by

$$h(x; y^1, \ldots, y^q) = \#\{i;\ x_i \notin \{y_i^1, \ldots, y_i^q\}\}. \tag{33}$$

Given q subsets $A_1, \ldots, A_q$ of Ω, we define

$$h(x; A_1, \ldots, A_q) = \inf\{h(x; y^1, \ldots, y^q);\ y^1 \in A_1, \ldots, y^q \in A_q\}. \tag{34}$$

Theorem 11

$$\int q^{h(x;A_1,\ldots,A_q)} \le \frac{1}{\prod_{j=1}^q P(A_j)}. \tag{35}$$

In particular,

$$P(\{x;\ h(x;A_1,\ldots,A_q) \ge k\}) \le \frac{1}{\prod_{j=1}^q P(A_j)} q^{-k} \tag{36}$$

for all $k \in \mathbb{N}$.

Sketch of proof of the induction step: For $A_1,\ldots,A_q \subset \Omega \times \Omega_{n+1}$ and $\omega \in \Omega_{n+1}$ put

$$A_j(\omega) = \{y \in \Omega;\ (y,\omega) \in A_j\}, \quad j = 1,\ldots,q \tag{37}$$

and

$$B_j = \cup_{u \in \Omega_{n+1}} A_j(u), \quad j = 1,\ldots,q. \tag{38}$$

Fix $\omega \in \Omega$ and $k \in \{1,\ldots,q\}$ and put also

$$C_j = \begin{cases} B_j & \text{if } j \ne k \\ A_k(\omega) & \text{if } j = k. \end{cases} \tag{39}$$

One then shows that

$$h((x,\omega);A_1,\ldots,A_q) \le \min\{1 + h(x;B_1,\ldots,B_q), h(x;C_1,\ldots,C_q)\}. \tag{40}$$

It then follows from the induction hypothesis that

$$\int q^{h((x,\omega);A_1,\ldots,A_q)} \le \frac{1}{\prod_{j=1}^q P(B_j)} \min\left\{q, \min_{1\le k\le q} \frac{P(B_k)}{P(A_k(\omega))}\right\}. \tag{41}$$

If $0 \le h_i \le 1$, $i = 1,\ldots,q$, are functions on a probability space then

$$\int \min\{q, \min_{1\le i\le q} h_i^{-1}\} \le \prod_{i=1}^q \left(\int h_i\right)^{-1}. \tag{42}$$

This follows easily from the inequality $\int h^{-1}(\int h)^q \le 1$ which holds for every function h satisfying $q^{-1} \le h \le 1$. Using (42) and integrating (41) over Ω_{n+1}, we get the assertion for $n+1$. ∎

We shall see in a minute the big advantage of this theorem over the concentration inequality for the Hamming metric. Although it looks like there is not much difference between $h(\cdot; A, A)$, say, and the Hamming distance of a point from a set ($d(\cdot, A)$ of Section 5), it turns out that the last theorem gives much better concentration when it applies. Theorem 11 is still looking for good applications. As far as we know Theorem 11 has basically one application dealing with the tail behavior of norms of sums of independent Banach space valued random variables. This is the original application which led Talagrand to prove this result (see [54] and [57], section 13). This particular application also has a different proof [29].

To illustrate the advantage of Theorem 11 over the basic inequality for the Hamming metric we define a class of functions and state a corollary which amounts to a deviation inequality for this class of functions. For $I \subset \{1, \dots, n\}$ denote

$$\Omega_I = \prod_{i \in I} \Omega_i \quad \text{and} \quad \Omega^* = \cup_{I \subset \{1,\dots,n\}} \Omega_I$$

and let $f : \Omega^* \to \mathbb{R}^+$. We say that f is *monotone* if

$$I \subset J \subseteq \{1,, \dots, n\} \ \text{ implies } \ f((x_i)_{i \in I}) \le f((x_j)_{j \in J}) \tag{43}$$

for all $(x_j)_{j \in J} \in \Omega_J$. We say that f is *subadditive* if for all I, J disjoint subsets of $\{1,, \dots, n\}$ and all $(x_i)_{i \in I \cup J} \in \Omega_{I \cup J}$,

$$f((x_i)_{i \in I \cup J}) \le f((x_i)_{i \in I}) + f((x_j)_{j \in J}). \tag{44}$$

Here is an example of such a function: Let Ω_i be subsets of a normed space $(X, \|\cdot\|)$ and put $f((x_i)_{i \in I}) = \mathrm{Ave}_{\epsilon_i = \pm 1} \|\sum_{i \in I} \epsilon_i x_i\|$.

For $x \in \Omega_I, y \in \Omega_J$ we shall denote by $h(x, y)$ the number of coordinates in which $x_i \ne y_i$ including coordinates in which one or both of x_i, y_i are not defined.

Corollary 12 *Let $f : \Omega^* \to \mathbb{R}^+$ be monotone, subadditive and satisfy $|f(x) - f(y)| \le h(x, y)$ for all $x, y \in \Omega^*$. Then, for all $a > 0$, $1 \le k \le n$ and $q \in \mathbb{N}$,*

$$P(\{x \in \Omega;\ f(x) \ge (q+1)a + k\}) \le P(f \le a)^{-q} q^{-k}. \tag{45}$$

For a being the median of f and $q = 2$, say, one gets $P(f \geq 3a + k) \leq 42^{-k}$. If $a << k << n$ this is much better than what one gets for a general Lipschitz function from, e.g., the martingale method. There one gets $P(f \geq a + k) \leq 2e^{-k^2/4n}$.

Note the resemblance with the situation concerning Theorem 9: In both cases we evaluate the probability of deviation of f from its expectation (or median), a quantity which depends only on the behavior of f on Ω (since the probability measure is supported there). However, by extending f to a larger set (in Theorem 9 the convex hull of Ω, here Ω^*), if possible, using its Lipschitz constant on the larger set and some additional properties of the extended function (there convexity, here monotonicity and subadditivity) we get, in some cases a stronger concentration result than the basic one.

Proof of Corollary 12. For $1 \leq i \leq q$ put $A_i = A = \{x \in \Omega;\ f(x) \leq a\}$. Then

$$\{f(x) \geq (q+1)a + k\} \subseteq \{h(x; A_1, \ldots, A_q) \geq k\}.$$

Indeed, if $h(x; A_1, \ldots, A_q) < k$, let $y^1, \ldots, y^q \in A$ be such that, putting $I = \{i;\ x_i \notin \{y_i^1, \ldots, y_i^q\}\}$, $\#I < k$. The complement of I can be written as $\cup_{j=1}^k J_j$ with $J_j \subseteq \{1, \ldots, n\}$ satisfying $x_i = y_i^j$ for $i \in J_j$. Then, assuming I is not empty,

$$\begin{aligned} f(x) &\leq f(x_{|I}) + \textstyle\sum_{j=1}^q f(x_{|J_j}) \\ &\leq f(x_{|I} + y^1_{|J_1}) + \textstyle\sum_{j=1}^q f(y^j_{|J_j}) \\ &\leq \#I + f(y^1_{|J_1}) + \textstyle\sum_{j=1}^q f(y^j_{|J_j}) \\ &\leq \#I + f(y^1) + \textstyle\sum_{j=1}^q f(y^j) \\ &< k + (q+1)a. \end{aligned} \tag{46}$$

The corollary follows now immediately from Theorem 11. ∎

The paper [57] also contains a generalization of the concentration inequality for the permutation group, Corollary 7. The (inductive) proof of this result is a bit harder than the other proofs surveyed in this section and we shall not reproduce it. This result also awaits good applications.

Equip the symmetric group S_n with its natural probability measure, μ. For $\sigma \in S_n$ and $A \subseteq S_n$ let

$$f(\sigma, A) = \inf \left\{ \sum_{i=1}^{n} s_i^2; (s_1, \ldots, s_n) \in V_A(\sigma) \right\} \tag{47}$$

where $V_A(\sigma)$ is the convex hall of the set

$$\{(s_1, \ldots, s_n) \in \{0,1\}^n; \exists \tau \in A \ s.t. \ \forall i \leq n, s_i = 0 \Rightarrow \tau(i) = \sigma(i)\}.$$

Theorem 13 *for every $A \subset S_n$, $t > 0$*

$$\mu(\{\sigma;\ f(\sigma, A) > t\}) \leq \frac{1}{\mu(A)} e^{-t/16}. \tag{48}$$

The manuscript [57] contains many refinements of Theorems 10, 11 and 13 which we do not reproduce here.

2.4 Spectral methods

Let $(\Omega, \mathcal{F}, \mu)$ be a probability space, $\mathcal{A}$ some set of measurable functions and $\mathcal{E} : \mathcal{A} \to \mathbb{R}^+$ some function (which we shall refer to as *energy function*). For $f \in L_2(\Omega)$ denote by $\sigma^2(f)$ the variance of f,

$$\sigma^2(f) = \int (f - \mathbb{E}f)^2 d\mu = \int f^2 d\mu - (\int f d\mu)^2 \tag{49}$$

and, for $f \in L^2 \log L$ (i.e. $\int f^2 \log^+ f d\mu < \infty$), denote by $\epsilon(f)$ the entropy of f^2,

$$\epsilon(f) = \int f^2 \log f^2 d\mu - \int f^2 d\mu \log(\int f^2 d\mu) \tag{50}$$

(which is necessarily finite). We say that $(\mathcal{A}, \mathcal{E})$ satisfy a *Poincaré inequality* with constant C if

$$\sigma^2(f) \leq C\mathcal{E}(f) \quad \text{for all } f \in \mathcal{A}. \tag{51}$$

We say that $(\mathcal{A}, \mathcal{E})$ satisfy a *logarithmic Sobolev inequality* with constant C if

$$\epsilon(f) \leq C\mathcal{E}(f) \quad \text{for all } f \in \mathcal{A}. \tag{52}$$

The main example of an energy function $\mathcal{E}$ is related to the gradient or generalization of it. If d is a metric on Ω (and $\mathcal{F}$ the Borel σ-field), define the *norm of the gradient* at $x \in \Omega$ by

$$|\nabla f(x)| = \limsup_{y \to x} \frac{|f(x) - f(y)|}{d(x,y)}. \tag{53}$$

Note that $\nabla f(x)$ by itself is not defined. The reason for this notation is of course that if (Ω, d) is a Riemannian manifold (in particular if it is $\mathbb{R}^n$ with the Euclidean distance) and if f is differentiable at x then $|\nabla f(x)|$ is the Euclidean norm of the gradient of f at x. Define now

$$\mathcal{E}(f) = \int |\nabla f(x)|^2 d\mu(x). \tag{54}$$

The classical Poincaré (or Rayleigh–Ritz) inequality says that, in the case of a compact Riemannian manifold, (51) is satisfied with $C = \lambda_1^{-1}$, λ_1 being the first positive eigenvalue of the Laplacian on $L_2(\Omega, \mu)$.

We shall only deal here with the energy function (54). [31] contains many other examples and a comprehensive treatment of the subject of this section.

If $\mathcal{A}$ is the set of bounded Lipschitz functions on (Ω, d), the norm of the gradient satisfies the chain rule: If $\phi \in C^1(\mathbb{R})$ and $f \in \Omega$ then $\phi \circ f \in \Omega$ and

$$|\nabla(\phi \circ f)(x)| \le |\nabla f(x)||\phi'(f(x))| \tag{55}$$

and consequently

$$\mathcal{E}(\phi \circ f) \le \|f\|_{\mathrm{Lip}}^2 \int |\phi'(f(x))|^2 d\mu(x) \tag{56}$$

where $\|f\|_{\mathrm{Lip}}$ denotes the Lipschitz constant of f. The next theorem, basically due to Gromov and Milman, shows that Poincaré inequality implies concentration.

Theorem 14 *Let $(\Omega, \mathcal{F}, \mu, d)$ be a probability metric space. Let $\mathcal{A}$ be the set of bounded Lipschitz functions on (Ω, d) and let $\mathcal{E}$ be defined by (54). Assume that $(\mathcal{A}, \mathcal{E})$ satisfies the Poincaré inequality (51). Then for all $|\lambda| < 2/\sqrt{C}$ and every bounded f with Lipschitz constant* 1

$$\mathbb{E}e^{\lambda(f - \mathbb{E}f)} \le \frac{240}{4 - C\lambda^2}. \tag{57}$$

In particular

$$P(|f - \mathbb{E}f| > t) \leq 240 e^{-\sqrt{\frac{2}{C}}t} \quad \textit{for all } t > 0. \tag{58}$$

Proof. By (51) and (56)

$$\mathbb{E}e^g - (\mathbb{E}e^{g/2})^2 \leq C\mathcal{E}(e^{g/2}) \leq \frac{C}{4}\|g\|_{Lip}\mathbb{E}e^g$$

for any $g \in \mathcal{A}$. In particular, for any λ,

$$\mathbb{E}e^{\lambda f} - (\mathbb{E}e^{\frac{\lambda}{2}f})^2 \leq \frac{C\lambda^2}{4}\mathbb{E}e^{\lambda f}$$

or

$$\mathbb{E}e^{\lambda f} \leq \frac{1}{1 - \frac{C\lambda^2}{4}}(\mathbb{E}e^{\frac{\lambda}{2}f})^2.$$

Iterating we get for every n,

$$\mathbb{E}e^{\lambda f} \leq \prod_{k=o}^{n-1}\left(\frac{1}{1 - \frac{C\lambda^2}{4^{k+1}}}\right)^{2^k}\left(\mathbb{E}e^{\frac{\lambda}{2^n}f}\right)^{2^n}$$

which tends to

$$\prod_{k=o}^{\infty}\left(\frac{1}{1 - \frac{C\lambda^2}{4^{k+1}}}\right)^{2^k} e^{\lambda \mathbb{E}f}.$$

■

Remark 15 *1. A simple limiting argument shows now that the assumption that f is bounded is superfluous.*
2. The simple example of the exponential distribution on $\mathbb{R}$ shows that (except for the absolute constants involved) one can't improve the concentration function e^{-ct}. As we shall see below, what looks like a slight change, logarithmic Sobolev inequality instead of Poincaré inequality, changes the behavior of the concentration function from e^{-ct} to e^{-ct^2}.

The next theorem is apparently due to Herbst.

Theorem 16 *Let $(\Omega, \mathcal{F}, \mu, d)$ be a probability metric space. Let $\mathcal{A}$ be the set of bounded Lipschitz functions on (Ω, d) and let $\mathcal{E}$ be defined by (54). Assume* *that $(\mathcal{A}, \mathcal{E})$ satisfies the logarithmic Sobolev inequality (52) then for all $\lambda \in \mathbb{R}$ and every bounded f with Lipschitz constant* 1

$$\mathbb{E}e^{\lambda(f-\mathbb{E}f)} \leq e^{C\lambda^2/4}. \tag{59}$$

In particular

$$P(|f - \mathbb{E}f| > t) \leq 2e^{-t^2/C} \quad \textit{for all } t > 0. \tag{60}$$

Proof. Put $h(\lambda) = \mathbb{E}e^{\lambda f}$, then

$$\epsilon(e^{\lambda f/2}) = \mathbb{E}\lambda f e^{\lambda f} - \mathbb{E}e^{\lambda f} \log(\mathbb{E}e^{\lambda f}) = \lambda h'(\lambda) - h(\lambda)\log(h(\lambda)). \tag{61}$$

Also, from (56), we get,

$$\mathcal{E}(e^{\lambda f/2}) \leq \frac{\lambda^2}{4}\mathbb{E}e^{\lambda f} = \frac{\lambda^2}{4}h(\lambda). \tag{62}$$

Combining (61), (62) and (52) we get

$$\lambda h'(\lambda) - h(\lambda)\log(h(\lambda)) \leq \frac{\lambda^2 C}{4}h(\lambda)$$

or, putting $k(\lambda) = \lambda^{-1}\log h(\lambda)$ (and, by continuity, $k(0) = \mathbb{E}f$),

$$k'(\lambda) = \frac{1}{\lambda}\frac{h'(\lambda)}{h(\lambda)} - \frac{1}{\lambda^2}\log h(\lambda) \leq \frac{C}{4}, \quad \text{for all } \lambda \in \mathbb{R}.$$

It follows that $k(\lambda) - k(0) \leq \frac{C\lambda}{4}$ and thus

$$\mathbb{E}e^{\lambda(f-\mathbb{E}f)} = e^{\lambda(k(\lambda)-k(0))} \leq e^{C\lambda^2/4}.$$

■

Remark 17 *A simple limiting argument shows that here too the assumption that f is bounded is superfluous.*

Both Poincaré inequality and logarithmic Sobolev inequality carry over nicely to product spaces in the following sense: For $i = 1, 2, \ldots, n$, let $(\Omega_i, \mathcal{F}_i, \mu_i)$ be a probability space, $\mathcal{A}_i$ some set of measurable functions on Ω_i and $\mathcal{E}_i : \mathcal{A}_i \to \mathbb{R}^+$ some energy function. Put $(\Omega, P) = \prod_{i=1}^n (\Omega_i, \mu_i)$. Given a function f on Ω we denote by f_i the same function considered as a function of the i-th variable only, keeping all other variables fixed. Define $\mathcal{E}(f) = \mathbb{E}_P \sum_{i=1}^n \mathcal{E}_i(f_i)$. Let $\mathcal{A}$ denote the set of all functions f such that (for all $x_1, \dots, x_n$ and) for all i, f_i is in $\mathcal{A}_i$.

One can prove that

$$\sigma^2(f) \le \mathbb{E}_P \sum_{i=1}^n \sigma^2(f_i) \quad \text{and} \quad \epsilon(f) \le \mathbb{E}_P \sum_{i=1}^n \epsilon(f_i), \tag{63}$$

from which the following proposition easily follows.

Proposition 18 *Assume $(\mathcal{A}_i, \mathcal{E}_i)$, $i = 1, \dots, n$, all satisfy Poincaré inequality (resp. logarithmic Sobolev inequality) with a common constant, C. Then $(\mathcal{A}, \mathcal{E})$ satisfies Poincaré inequality (resp. logarithmic Sobolev inequality) with the same constant, C.*

Example 19 *The symmetric exponential measure on $\mathbb{R}$, i.e. the measure with density $\frac{1}{2}e^{-|t|}$, satisfies Poincaré inequality with constant 4. Consequently, the same is true for the measure on $\mathbb{R}^n$ which is the n fold product of this measure.*

The canonical Gaussian measure on $\mathbb{R}$ and thus on $\mathbb{R}^n$ satisfies logarithmic Sobolev inequality with constant 2.

The proof of both statements can be found in [31]. The second one is due to Gross and, in view of Theorem 16, implies the concentration inequality for γ_n, the Gaussian measure on $\mathbb{R}^n$: If $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz with constant one with respect to the Euclidean metric then

$$\gamma_n \left(|f - \int f \, d\gamma_n| > t \right) \le Ce^{-ct^2}.$$

From this it is not hard to get the concentration inequality for S^{n-1}. One uses Lemma 22 below.

We would also like to state a theorem first proved by Talagrand [56] which "interpolates" between the last two theorems. See [8] and [31] for a relatively simple proof along the lines of the proofs of the last two theorems. We state it only for a specific probability measure P on $\mathbb{R}^n$, the product of the measures with density $\frac{1}{2}e^{-|t|}$ on $\mathbb{R}$. See [31] for generalizations.

Theorem 20 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function satisfying*

$$|f(x) - f(y)| \le \alpha \|x - y\|_2 \quad \textit{and} \quad |f(x) - f(y)| \le \beta \|x - y\|_1 \,. \tag{64}$$

Then, with the probability introduced above,

$$P\left(|f(x) - \mathbb{E}f| > r\right) \leq C\exp(-c\min(r/\beta, r^2/\alpha^2)) \tag{65}$$

for some absolute positive constants C, c and all $r > 0$.

Remark 21 *Although it deals with a different probability measure, Theorem 20 also implies the concentration inequality for the Gaussian measure on $\mathbb{R}^n$ (and thus, via Lemma 22 below, also for the Haar measure on S^{n-1}). This follows from a simple transference of the Gaussian measure to the product of the symmetrized exponential measure discussed above. Thus, Theorem 20 can be considered as a strengthening of these inequalities. We refer to [56] and [31] for that and further discussion.*

Although the methods in this and the previous section are specialized to product measures, there is a way to transfer such results to some other situations. In particular to the case of unit balls of ℓ_p^n spaces equipped with the normalized Lebesgue measure. The basic tool is the following simple result: Consider the measure $\mu(A) = \frac{|\{tA;\ 0\leq t\leq 1\}|}{|\{x;\ \|x\|_p\leq 1\}|}$ on the surface of the ℓ_p^n ball, $0 < p < \infty$. Consider also n independent random variables $X_1, X_2, \ldots, X_n$ each with density function $c_p e^{-|t|^p}$, $t \in \mathbb{R}$. (Note that necessarily $c_p = p/2\Gamma(1/p)$.)

Lemma 22 *Put $S = (\sum_{i=1}^n |X_i|^p)^{1/p}$. Then $\left(\frac{X_1}{S}, \frac{X_2}{S}, \ldots, \frac{X_n}{S}\right)$ induces the measure μ on ∂B_p^n. Moreover, $\left(\frac{X_1}{S}, \frac{X_2}{S}, \ldots, \frac{X_n}{S}\right)$ is independent of S.*

See [49] for a proof. This lemma is used there to compute the tail behavior of the ℓ_q norm on the ℓ_p^n ball. Recently ([50]) this result was strengthen, in the case $p = 1$, $q = 2$, to give a concentration inequality for general Lipschitz functions, with respect to the Euclidean metric, on the ℓ_1^n ball B_1^n. The proof combines most of the results of this section and we shall not give it here.

Theorem 23 *There exist positive constants C, c such that if $f : \partial B_1^n \to \mathbb{R}$ satisfies $|f(x) - f(y)| \leq \|x - y\|_2$ for all $x, y, \in \partial B_1^n$ then, for all $t > 0$,*

$$\mu(\{x;\ |f(x) - \mathbb{E}f| > t\}) \leq C\exp(-ctn). \tag{66}$$

2.5 Bounds on Gaussian processes

As we shall see below, in the application sections, concentration inequalities are used mostly to find a point ω, in the metric probability space under consideration, in which a big collection of functions $\{G_t(\omega)\}_{t\in T}$ are each close to its mean. There may be other ways to reach such a conclusion. Assuming the means of all the functions under consideration are zero, it would be enough, for example, to prove that $\mathbb{E}\sup_{t\in T}|G_t|$ is small (then, for a set of ω's of measure at least 1/2, $\sup_{t\in T}|G_t(\omega)|$ is at most 2×small).

When T is a metric space and G_t a Gaussian process (meaning that any finite linear combination of the G_t's has a Gaussian distribution) the evaluation of $\mathbb{E}\sup_{t\in T}|G_t|$ is an extensively studied subject in Probability (having to do with the existence of a continuous version of the process). See for example [32]. There are well studied connections between the quantity $\mathbb{E}\sup_{t\in T}|G_t|$ and the entropy (or covering) function of the metric space T as well as with other properties of T. A recent achievement in this area is Talagrand's majorizing measure theorem which relates the boundedness of $\mathbb{E}\sup_{t\in T}|G_t|$ to the existence of a certain measure (called majorizing measure) on T and gives new ways to estimate this quantity. This subject is reviewed in [28] and we'll not get into it any further here. We only remark that the proofs in this area are very much connected with concentration properties of Gaussian variables.

2.6 Other tools

We dealt above mostly with geometric and probabilistic tools to prove concentration and approximate isoperimetric inequalities. There are many other methods and results that are not discussed here for lack of space. In particular we didn't discuss at all combinatorial methods. For example the (exact) isoperimetric inequality for the Hamming cube (from which Corollary 7 for that case follows) was first proved by Harper [20] (see also [14] for a simpler proof) by combinatorial methods.

There are also geometrical and probabilistic methods we didn't discuss. [43] contains a yet another nice probabilistic proof of Corollary 4 due to Maurey and Pisier. It uses properties very special to Gaussian variables and thus does not seem to generalize much.

[48] contains a generalization of Corollary 4 to harmonic measures on S^{n-1}. The proof is by reduction to the Haar measure.

A new probabilistic method which emerged recently is that of *transportation cost*, see [36], [58] and [10]. This seems very much related to Kantorovich's solution of Monge's "mass transport" problem although, as far as I know, no concrete relation has been found yet.

This short list is far from exhausting all the sources on this vast subject.

3 Applications

3.1 Dvoretzky–like theorems

The introduction of the method(s) of concentration of measure into Banach Space Theory was initiated by Milman in his proof [39] of Dvoretzky's theorem concerning spherical sections of convex bodies [11]. Although this topic is extensively reviewed in the article [15] in this handbook, I would like to begin the applications section with a statement of the theorem and a brief description of its proof.

Theorem 24 *For all $\epsilon > 0$ there exists a constant $c = c(\epsilon) > 0$ such that for any n-dimensional normed space X there exists a subspace Y of dimension $k \geq c \log n$ such that the Banach–Mazur distance $d(Y, \ell_2^k) \leq 1 + \epsilon$.*

See [23] for the definition of the Banach–Mazur distance. The one to one correspondence between n-dimensional normed spaces and n-dimensional symmetric convex bodies (and the fact that every $2n$-dimensional ellipsoid has an n-dimensional section which is a multiple of the canonical Euclidean ball) easily shows that the theorem above is equivalent to the following geometrical statement. By a convex body in $\mathbb{R}^n$ we mean a compact convex set with non-empty interior.

Theorem 25 *For all $\epsilon > 0$ there exists a constant $c = c(\epsilon) > 0$ such that every centrally symmetric convex body K admits a $k \geq c \log n$ central section K_0 and a positive number r satisfying $rB \subset K_0 \subset (1 + \epsilon)rB$, where B is the canonical Euclidean ball in the subspace spanned by K_0.*

Sketch of proof. Since the statement of each of the two theorems is invariant under invertible linear transformations, we may assume that the unit ball K of $X = (\mathbb{R}^n, \|\cdot\|)$ satisfies $B_2^n \subset K$ and the canonical Euclidean ball B_2^n in $\mathbb{R}^n$ is (the) ellipsoid of maximal volume among all ellipsoids inscribed in K. (It is a theorem of F. John that the maximal volume ellipsoid is uniquely determined but we do not need this fact here.) A relatively easy theorem of Dvoretzky and Rogers [12] (see also [40] p.10) implies now that $\mathbb{E} = \mathbb{E}\|\cdot\| = \int_{S^{n-1}} \|x\| d\mu(x) > c\sqrt{\frac{\log n}{n}}$ for some absolute constant c.

Denoting by ν the normalized Haar measure on the orthogonal group $O(n)$ and applying Corollary 4 to the function $x \to \|x\|$, which is Lipschitz with constant one, we get that, for every fixed $x \in S^{n-1}$,

$$\nu(\{u;\ |\|ux\| - \mathbb{E}| > \epsilon\mathbb{E}\}) = \mu(\{x \in S^{n-1};\ |\|x\| - \mathbb{E}| > \epsilon\mathbb{E}\})$$

$$< e^{-c\epsilon^2 \mathbb{E}^2 n} < e^{-c\epsilon^2 \log n}.$$

Fix a k-dimensional subspace $V_0 \subset \mathbb{R}^n$ and an ϵ net $\mathcal{N}$ in $V_0 \cap S^{n-1}$ of cardinality smaller than $(3/\epsilon)^k$. The existence of such a net follows from an easy volume argument (see [40] p.7). It then follows that if $(3/\epsilon)^k e^{-c\epsilon^2 \log n} < 1$, i.e., if k is no larger than a constant depending on ϵ times $\log n$, then

$$\nu(\{u;\ |\|ux\| - \mathbb{E}| > \epsilon\mathbb{E},\ \text{for some } x \in \mathcal{N}\} < 1$$

which implies that there is a $u \in O(n)$ such that

$$(1-\epsilon)\mathbb{E} \le \|ux\| \le (1+\epsilon)\mathbb{E}, \quad \text{for all} \ \ x \in \mathcal{N}.$$

It now follows from a successive approximation argument that similar inequalities hold for all $x \in S^{n-1}$ which implies the conclusion of the theorem for the subspace uV_0. ■

We next state another application of the concentration inequality on the Euclidean sphere. This Lemma of Johnson and Lindenstrauss is much simpler but has a lot of applications including "real life" ones like efficient algorithms for detecting clusters.

Theorem 26 *Let $x_1, x_2, \ldots, x_n$ be points in some Hilbert space. If $k \ge \frac{c}{\epsilon^2}\log n$ (with $c > 0$ an absolute constant), then there are $y_1, y_2, \ldots, y_n \in \ell_2^k$ satisfying*

$$\|x_i - x_j\| \le \|y_i - y_j\| \le (1+\epsilon)\|x_i - x_j\| \tag{67}$$

for all $1 \le i \ne j \le n$.

Sketch of proof. We may assume that the points x_i are in ℓ_2^n. Fix a $k < n$ and a rank k orthogonal projection P_0 on ℓ_2^n. When u ranges over $O(n)$, $P = uP_0u^{-1}$ ranges over all rank k orthogonal projections. It is not hard to check that, for all $x \in S^{n-1}$, $\mathbb{E} = \int_{O(n)} \|uP_0u^{-1}x\| d\nu(u)$ is of the order $\sqrt{k/n}$ and thus, for every $x \in S^{n-1}$,

$$\nu(\{u;\ |\|uP_0u^{-1}x\| - \mathbb{E}| > \epsilon\mathbb{E}\}) = \mu(\{x;\ |\|P_0x\| - \mathbb{E}| > \epsilon\mathbb{E}\}) < e^{-c\epsilon^2 k}.$$

It follows that, if $k \le \frac{C}{\epsilon^2}\log n$, there is a $u \in O(n)$ for which

$$(1-\epsilon)\mathbb{E} \le \left\| uP_0u^{-1}\left(\frac{x_i - x_j}{\|x_i - x_j\|}\right)\right\| \le (1+\epsilon)\mathbb{E}$$

for all $i \neq j$. The range of uP_0u^{-1} is k-dimensional. Take $y_i = \frac{uP_0u^{-1}x_i}{(1-\epsilon)\mathbb{E}}$. ■

3.2 Fine embeddings of subspaces of L_p in l_p^n

When specializing the proof of Theorem 24 to the case of $X = \ell_r^n$, one sees quite easily that *if $1 \leq r < 2$ then for all $\epsilon > 0$ there exists a constant $c = c(r, \epsilon) > 0$ such that for all n there exists a subspace Y of ℓ_r^n of dimension $k \geq cn$ whose Banach–Mazur distance to Euclidean space, $d(Y, \ell_2^k) \leq 1 + \epsilon$. (For $2 < r < \infty$ the same holds with $k \geq cn^{2/r}$.)* This subject is extensively reviewed in [15].

Since it is known (and follows from the existence of p-stable random variables, see below) that for $r < p \leq 2$ ℓ_p embeds isometrically into $L_r[0,1]$, it is natural to ask whether a similar statement holds with 2 replaced by p, i.e., whether, for $r < p < 2$, ℓ_p^k $(1+\epsilon)$-embeds into ℓ_r^n for k proportional to n. Noticing that Gaussian variables are very different from p-stable ones for $p < 2$ (the first decay exponentially while the latter only polynomially), and that the concentration inequality behind the proof of Theorem 24 has very much to do with the exponential decay of Gaussian variables, one's first guess would be that the answer to the question above is negative (and probably that k can only be some logarithmic function of n).

It turns out, however, that the answer to the question above is positive. It was proved in [24] that *for $1 \leq p < 2$ and for every n and ϵ, ℓ_1^n contains a subspace Y with $d(Y, \ell_p^k) < 1 + \epsilon$ where $k \geq c(p, \epsilon)n$.* This was the first result concerning "tight embeddings" that didn't deal with Euclidean spaces. It was proved using certain approximation of p-stable random variables and concentration inequalities for martingales as discussed in Section 2.2. This result lead to a series of generalizations and results of similar nature. We refer to [26] for a survey of this topic. Here we only deal with two such examples of generalizations. We would first like to mention a result of Pisier [42], generalizing the result above from the side of the containing space, ℓ_1^n.

Recall that a random variable h whose characteristic function is given by $\mathbb{E}e^{ith} = e^{-c|t|^p}$, for some positive constant c, is called (symmetric) p-stable. Lévy proved the existence of such variables for $0 < p \leq 2$. (There are no such variables for $p > 2$.) A p-stable variable has r-th moment for all $r < p$ but doesn't have p-th moment. For $1 < p < 2$ we'll denote from now on by h the p-stable variable whose first moment is equal to 1. This defines its distribution uniquely. If $h, h_1, \ldots, h_n$ are independent and identically distributed then it is easy to see (compute the characteristic function) that $\sum_{i=1}^n \alpha_i h_i$ also has the same distribution as h as long as $\sum_{i=1}^n |\alpha_i|^p = 1$. In particular the span of $h_1, \ldots, h_n$ in $L_1[0,1]$ is isometric to ℓ_p^n.

For $1 < p < 2$, the *stable type p constant* of a Banach space X, denoted $ST_p(X)$, is the smallest constant C such that,

$$\mathbb{E}\|\sum h_i x_i\| \leq Cn^{1/p} \sup_{1\leq i\leq n} \|x_i\| \tag{68}$$

for all finite sequences $\{x_1, \ldots, x_n\}$ of elements of X. (This is an equivalent definition to the more common one where $n^{1/p}\sup_{1\leq i\leq n}\|x_i\|$ is replaced with $(\sum_{i=1}^n \|x_i\|^p)^{1/p}$.) Pisier's result is:

Theorem 27 *For each $1 < p < 2$ and $\epsilon > o$ there is a positive constant $c = c(p, \epsilon)$ such that any Banach space X contains a subspace Y satisfying $d(Y, \ell_p^k)$ as long as*

$$k < cST_p(X)^{p/(p-1)}. \tag{69}$$

Since it is easy to see that $ST_p(\ell_1^n) \geq n^{(p-1)/p}$, this implies the result of [24] referred to above.

A brief sketch of the proof. Pick a finite sequence, $x_1, x_2, \ldots, x_n$, of elements of X for which $\max\|x_i\| = 1$ and $\mathbb{E}\|\sum h_i x_i\| \geq \frac{1}{2}n^{1/p}ST_p(X)$. Let $u_1, u_2, \ldots$ be a sequence of independent random variables each uniformly distributed over the set of $2n$ elements $\{\pm x_1, \pm x_2, \ldots, \pm x_n\}$. Put also $\Gamma_j = \sum_{i=1}^j A_i$, $j = 1, 2, \ldots$, where the A_j's are independent (and independent of the sequence $\{u_i\}$) canonical exponential variables, i.e., $P(A_i > t) = e^{-t}$, $t > 0$. We shall use a representation theorem for p-stable variables, due to Lepage, Woodroofe and Zinn [34] which says in particular that, for some constant c_p depending only on p,

$$S = \sum_{j=1}^{\infty} \Gamma_j^{-1/p} u_j \text{ has the same distribution as } c_p n^{-1/p} \sum_{i=1}^{n} h_i x_i \tag{70}$$

and in particular, $\mathbb{E}\|S\| \geq \frac{c_p}{2}ST_p(X)$. Note that for any functional x^*, $x^*(S)$ is a p-stable variable. If $S_1, \ldots, S_k$ are independent and all have the same distribution as S then it is easily seen that if $\sum_{i=1}^k |\alpha_i|^p = 1$ then $\sum_{i=1}^k \alpha_i S_i$ has the same distribution as S and in particular $\mathbb{E}\|\sum_{i=1}^k \alpha_i S_i\| = \mathbb{E}\|S\|$.

The next step is to replace the random coefficients $\{\Gamma_j^{-1/p}\}$ with the deterministic sequence $\{j^{-1/p}\}$. Put $R = \sum_{j=1}^{\infty} j^{-1/p}u_j$ and let $R_1, \ldots, R_k$ be independent and all have the same distribution as R. A computation using the explicit distribution of Γ_j shows that

$$C = \mathbb{E}\|S_i - R_i\| < \infty$$

and it follows that, if $\sum_{i=1}^k |\alpha_i|^p = 1$,

$$\begin{aligned}\Big|\mathbb{E}\|\textstyle\sum_{i=1}^k \alpha_i S_i\| - \mathbb{E}\|\sum_{i=1}^k \alpha_i R_i\|\Big| &\le C\textstyle\sum_{i=1}^k |\alpha_i| \\ &\le Ck^{(p-1)/p}(\textstyle\sum_{i=1}^k |\alpha_i|^p)^{1/p} \\ &< Cc^{(p-1)/p} ST_p(X)\end{aligned} \tag{71}$$

by the choice of k.

Note that $\sum_{i=1}^k |\alpha_i|^p = 1$ implies that $\|\{\alpha_i j^{-1/p}\}\|_{p,\infty} = 1$ and thus Proposition 5(ii) implies that for all such $\{\alpha_i\}$ and for all $t > 0$,

$$P\left(\Big|\|\sum \alpha_i R_i\| - \mathbb{E}\|\sum \alpha_i R_i\|\Big| > t\right) \le K\exp\left(-\delta t^{p/(p-1)}\right). \tag{72}$$

This last equation is of course the place where the method of concentration enters, which was the main thing we wanted to illustrate here. The rest of the proof goes along similar lines to the end of the proof of Theorem 24: Note that it follows from (71), that, for c small enough, $\mathbb{E}\|\sum \alpha_i R_i\|$ is of order $ST_p(X)$. Choose an ϵ net in the sphere of ℓ_p^k of cardinality smaller than $(3/\epsilon)^k$. Then, with high probability, $\|\sum \alpha_i R_i\|$ is of order $ST_p(X)$ for all sequences $\{\alpha_i\}$ in the net. By a successive approximation the same holds now for all sequences $\{\alpha_i\}$ in the sphere of ℓ_p^n which completes the proof. ■

Another way to generalize the result of Schechtman and Johnson [24] (that ℓ_p^{cn} nicely embeds in ℓ_1^n) is from the side of the embedded space, ℓ_p^n. After some initial work by Schechtman (mostly [47]) on embedding finite dimensional subspaces of $L_p[0,1]$ in low dimensional ℓ_r^n spaces in which a new class of "random embeddings" (which were not related to p stable variables) were introduced, Bourgain, Lindenstrauss and Milman [9] proved that, for $0 < r < 2$, every k dimensional subspace of $L_r[0,1]$ $(1+\epsilon)$-embeds in ℓ_r^n provided n/k is at least a certain power of $\log n$ times a constant (depending only on r and ϵ). See also [25] for a different proof. All the proofs involved use concentration in one way or another. The result of [9] mentioned above was improved and simplified by Talagrand [55]. Since his proof has to do with bounds on Gaussian processes and is related to Section 2.5, we would like to briefly review it. As we have already advertized, the article [26] has more on that subject. Here we shall deal only with the case $r = 1$.

Theorem 28 *For every ϵ there is a constant $C(\epsilon)$ such that for all n, every n dimensional subspace Y of $L_1[0,1]$ is $(1+\epsilon)$-isomorphic to a subspace of $\ell_1^{Cn\log n}$.*

We remark in passing that one of the main open problems in this area is whether the factor $\log n$ is needed. Besides concentration inequalities the proof uses some other heavy tools and is discussed in [26]. We shall only touch the idea involving bounds on Gaussian procceses.

The idea of the proof. By crude approximation we may assume that Y is a subspace of ℓ_1^m for some finite (but huge) m. We would like to show that a restriction to a "random" subset of cardinality $Cn \log n$ of the coordinates is a good isomorphism when restricted to Y. Of course this is wrong in general (for instance if Y has an element which is supported on only one coordinate, this element would most probably be sent to zero by such a restriction). The idea is to first "change the density" and send Y to an isometric subspace whose elements are "spread out" over the m coordinates. The idea that this may work was the point of [47]. It will be dealt with in [26] and will not be discussed here any further. We'll concentrate in describing how to evaluate the norm of the random restriction on Y and the norm of its inverse assuming Y is already in good position.

We do it inductively, restricting first to a random set of about half the coordinates where each coordinate is choosen with probability $1/2$ and the different choices are independent. Equivalently, let $\{\epsilon_i\}_{i=1}^m$ be independent variables each taking the values -1 and 1 with probability $1/2$ each. We would like to evaluate the restriction to the set $A = \{i;\ \epsilon_i = 1\}$. If we could show that

$$\sup_{x\in Y,\|x\|\le 1} \left| 2\sum_{i\in A} |x_i| - \sum_{i=1}^{m} |x_i| \right| < \epsilon(n,m) \tag{73}$$

with $\epsilon(n,m)$ "very small" when n/m is small, then this would mean that (2 times) the restriction to A is very close to an isometry. Iterating this would lead, depending on the behavior of $\epsilon(n,m)$, to the desired random restriction onto a small set of coordinates. Note that the quantity in (73) is equal to $\sup_{x\in Y,\|x\|\le 1} |\sum_{i=1}^m \epsilon_i |x_i||$ and in particular is the same for A and its complement. Since we are interested in only one set A, of cardinality at most $m/2$ satisfying (73), it is enough to establish

$$\mathbb{E} \sup_{x\in Y,\|x\|\le 1} \left| \sum_{i=1}^{m} \epsilon_i |x_i| \right| \le \epsilon(n,m). \tag{74}$$

This quantity is dominated by a similar one with independent standard Gaussian variables g_i's replacing the ϵ_i's. So the problem reduces to estimating

$$\mathbb{E} \sup_{x\in Y,\|x\|\le 1} \left| \sum_{i=1}^{m} g_i |x_i| \right|$$

i.e. the expectation of the supremum of a specific Gaussian process. This makes the connection with Section 2.5. We shall not go into more details here.

Theorem 28 has a nice geometrical interpretation which is obtained by looking at the polar body to the unit ball of Y.

Corollary 29 *Let K be the (Minkowski) sum of segments in $\mathbb{R}^n$ (or limit of such bodies, these are called zonoids). Then, for every ϵ, there is a body L in $\mathbb{R}^n$ which is the sum of at most $C(\epsilon)n\log n$ segments and which ϵ approximates K in the sense that*

$$L \subset K \subset (1+\epsilon)L.$$

3.3 Selecting good substructures

Given a sequence of independent vectors $\{x_1, x_2, \ldots, x_n\}$ in a normed space X and an $\epsilon > 0$, what is the largest cardinality k such that there are k disjoint blocks $y_1, y_2, \ldots, y_k$ which are $(1+\epsilon)$-unconditional or $(1+\epsilon)$-symmetric?

Recall that by disjoint blocks we mean vectors of the form $y_i = \sum_{j\in\sigma_i} a_j x_j$, $i = 1, \ldots, k$, with $\sigma_1, \sigma_2, \ldots, \sigma_k$ disjoint subsets of $1, 2, \ldots, n$. $y_1, y_2, \ldots, y_k$ is said to be $(1+\epsilon)$-unconditional (resp. $(1+\epsilon)$-symmetric) if

$$\|\sum_{i=1}^{k} \epsilon_i b_i y_i\| \le (1+\epsilon)\|\sum_{i=1}^{k} b_i y_i\|$$

for all signs $\{\epsilon_i\}$ and all coefficients $\{b_i\}$. (resp. if

$$\|\sum_{i=1}^{k} \epsilon_i b_i y_{\pi(i)}\| \le (1+\epsilon)\|\sum_{i=1}^{k} b_i y_i\|$$

for all signs $\{\epsilon_i\}$, all permutations π of $1, 2, \ldots, k$ and all coefficients $\{b_i\}$.)

These problems and various variations thereof were treated quite successfully by concentration of measure methods. The point is that, fixing a partition $\sigma_1, \sigma_2, \ldots, \sigma_k$ of $\{1, 2, \ldots, n\}$ and coefficients $\{\{a_j\}_{j\in\sigma_i}\}_{i=1}^{k}$, the norms

$$\|\sum_{i=1}^{k} b_i e_i\|_u = \mathrm{Ave}_\epsilon \|\sum_{i=1}^{k} b_i \sum_{j\in\sigma_i} \epsilon_j a_j x_j\|$$

and

$$\|\sum_{i=1}^{k} b_i e_i\|_s = \mathrm{Ave}_{\epsilon,\pi}\|\sum_{i=1}^{k} b_i \sum_{j\in\sigma_i} \epsilon_j a_{\pi(j)} x_j\|$$

on $\mathbb{R}^n$ are 1-unconditional and 1-symmetric respectively. If we can find signs $\{\{\epsilon_j\}_{j\in\sigma_i}\}_{i=1}^k$ such that, for all $\{b_i\}$, $\|\sum_{i=1}^k b_i \sum_{j\in\sigma_i} \epsilon_j a_j x_j\|/\|\sum_{i=1}^k b_i e_i\|_u$ is appropriately close to one, then we found disjoint blocks of length k which are $(1+\epsilon)$-unconditional. A similar statement holds for the symmetric case.

For lack of space we shall not review all that is known about this subject. The unconditional case was first treated by Amir and Milman in [2],[3]. Gowers improved some of their quantitative estimates ([16],[17]) and in some instances got, except for possible log factors, the best possible estimates. The symmetric case was treated by Maurey [37] and was the motivation for proving Corollary 7 (for Π_n).

We were dealing here only with applications to functional analysis and convexity. There are many applications to other areas which we shall not expend on. There are applications to graph theory (see e.g. the construction of expander graphs in [1]), to other combinatorial questions, computer science, mathematical physics and probability (in particular to estimating the tail behavior of random variables of the form $\|\sum \epsilon_i X_i\|$ for independent vector valued random variables $\{X_i\}$). [57] contains many applications of the material of Section 2.3.

References

[1] N. Alon and V. D. Milman, λ_1, isoperimetric inequalities for graphs, and superconcentrators, *J. Combin. Theory Ser. B* **38** (1985) 73–88.

[2] D. Amir, V. D. Milman, Unconditional and symmetric sets in n-dimensional normed spaces, *Israel J. Math.* **37** (1980) 3–20.

[3] D. Amir, V. D. Milman, A quantitative finite-dimensional Krivine theorem, *Israel J. Math.* **50** (1985) 1–12.

[4] J. Arias-de-Reyna, K. Ball and R. Villa, Concentration of the distance in finite dimensional normed spaces.

[5] K. Azuma, Weighted sums of certain dependent random variables, *Tôhoku Math. J.* **19** (1967) 357–367.

[6] A. Baernstein II, B. A. Taylor, Spherical rearrangements, subharmonic functions, and $*$-functions in n-space, *Duke Math. J.* **43** (1976) 245–268.

[7] Y. Benyamini, Two point symmetrization, the isoperimetric inequality on the sphere and some applications, *Longhorn Notes, Univ. of Texas, Texas Funct. Anal. Seminar,* (1983-1984) 53–76.

[8] S. Bobkov and M. Ledoux, Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution, *Probab. Theory Related Fields* **107** (1997) 383–400.

[9] J. Bourgain, J. Lindenstrauss and V. D. Milman, Approximation of zonoids by zonotopes *Acta Math.* **162** (1989) 73–141.

[10] A. Dembo, Information inequalities and concentration of measure, *Ann. Probab.* **25** (1997) 927–939.

[11] A. Dvoretzky, Some results on convex bodies and Banach spaces, *Proc. Symp. on Linear Spaces,* (Jerusalem, 1961) 123–160.

[12] A. Dvoretzky and C. A. Rogers, Absolute and unconditional convergence in normed linear spaces, *Proc. Nat. Acad. Sci. USA* **36** (1950) 192–197.

[13] T. Figiel, J. Lindenstrauss and V. D. Milman, The dimension of almost spherical sections of convex bodies, *Acta Math.* **139** (1977) 53–94.

[14] P. Frankl and Z. Füredi, A short proof for a theorem of Harper about Hamming-spheres, *Discrete Math.* **34** (1981) 311–313.

[15] A. A. Giannopoulos and V. M. Milman, this handbook.

[16] W. T. Gowers, Symmetric block bases in finite-dimensional normed spaces, *Israel J. Math.* **68** (1989) 193–219.

[17] W. T. Gowers, Symmetric block bases of sequences with large average growth, *Israel J. Math.* **69** (1990) 129–151.

[18] M. Gromov and V. D. Milman, A topological application of the isoperimetric inequality, *Amer. J. Math.* **105** (1983) 843–854.

[19] M. Gromov and V. D. Milman, Generalization of the spherical isoperimetric inequality to uniformly convex Banach spaces, *Compositio Math.* **62** (1987) 263–282.

[20] L. H. Harper, Optimal numberings and isoperimetric problems on graphs, *J. Combinatorial Theory* **1** (1966) 385–393.

[21] P. Hitczenko, Best constants in martingale version of Rosenthal's inequality, *Ann. Probab.* **18** (1990) 1656–1668.

[22] W. B. Johnson and J. Lindenstrauss, Extensions of Lipschitz mappings into a Hilbert space, *Contemp. Math.* **26** (1984) 189–206

[23] W. B. Johnson and J. Lindenstrauss, this handbook.

[24] W. B. Johnson and G. Schechtman, l_p^m into l_1^n, *Acta Math.* **149** (1982) 71–85.

[25] W. B. Johnson and G. Schechtman, Remarks on Talagrand's deviation inequality for Rademacher functions, in: *Functional analysis (Austin, TX, 1987/1989)* (Lecture Notes in Math., 1470, Springer, Berlin, 1991) 72–77.

[26] W. B. Johnson and G. Schechtman, this handbook.

[27] W. B. Johnson, G. Schechtman and J. Zinn, Best constants in moment inequalities for linear combinations of independent and exchangeable random variables, *Ann. Probab.* **13** (1985) 234–253.

[28] S. Kwapień, this handbook.

[29] S. Kwapień, J. Szulga, Hypercontraction methods in moment inequalities for series of independent random variables in normed spaces, *Ann. Probab.* **19** (1991) 369–379.

[30] M. Ledoux, On Talagrand's deviation inequalities for product measures, *ESAIM Probab. Statist.* **1** (1995/97) 63–87.

[31] M. Ledoux, Concentration of measure and logarithmic Sobolev inequalities, Preprint (1997).

[32] M. Ledoux and M. Talagrand, *Probability in Banach spaces, Isoperimetry and processes* (Springer-Verlag, Berlin, 1991).

[33] L. Leindler, On a certain converse of Hölder's inequality II, *Acta Sci. Math. (Szeged)* **33** (1972) 217–223.

[34] R. Lepage, M. Woodroofe and J. Zinn, Convergence to a stable distribution via order statistics, *Ann. Probab.* **9** (1981) 624–632.

[35] P. Lévy, *Problmes concrets d'analyse fonctionnelle* (French) 2d ed. (Gauthier-Villars, Paris, 1951).

[36] K. Marton, A measure concentration inequality for contracting Markov chains, *Geom. Funct. Anal.* **6** (1996) 556–571.

[37] B. Maurey, Construction de suites symétriques. (French) *C. R. Acad. Sci. Paris Sér. A-B* **288** (1979) 679–681.

[38] B. Maurey, Some deviation inequalities, *Geom. Funct. Anal.* **1** (1991) 188–197.

[39] V. D. Milman, A new proof of the theorem of A. Dvoretzky on sections of convex bodies, *Funct. Anal. Appl.* **5** (1971) 28–37.

[40] V. D. Milman and G. Schechtman, *Asymptotic theory of finite–dimensional normed spaces* (Lecture Notes in Mathematics, 1200. Springer-Verlag, Berlin-New York, 1986).

[41] V. D. Milman and G. Schechtman, An "isomorphic" version of Dvoretzky's theorem, *C. R. Acad. Sci. Paris Sér. I Math.* **321** (1995) 541–544.

[42] G. Pisier, On the dimension of the l_p^n-subspaces of Banach spaces, for $1 \leq p < 2$, *Trans. Amer. Math. Soc.* **276** (1983) 201–211.

[43] G. Pisier, Probabilistic methods in the geometry of Banach spaces, *CIME, Varenna, 1985* (Lecture Notes in Mathematics, 1206. Springer-Verlag, 1986).

[44] G. Pisier, *The volume of convex bodies and Banach space geometry* Cambridge University Press (1989).

[45] A. Prékopa, On logarithmic concave measures and functions, *Acta Sci. Math. (Szeged)* **34** (1973) 335–343.

[46] G. Schechtman, Lévy type inequality for a class of finite metric spaces, in: *Martingale theory in harmonic analysis and Banach spaces (Cleveland, Ohio, 1981)* (Lecture Notes in Math., 939, Springer, Berlin-New York, 1982) 211–215.

[47] G. Schechtman, More on embedding subspaces of L_p in l_r^n, *Compositio Math.* **61** (1987) 159–169.

[48] G. Schechtman and M. Schmuckenschläger, A concentration inequality for harmonic measures on the sphere, in: Geometric aspects of functional analysis (Israel, 1992–1994), *Oper. Theory Adv. Appl.* **77** (1995) 255–273.

[49] G. Schechtman and J. Zinn, On the volume of the intersection of two L_p^n balls *Proc. Amer. Math. Soc.* **110** (1990) 217–224.

[50] G. Schechtman and J. Zinn, Concentration on the ℓ_1^n ball, in preparation.

[51] R. Schneider, *Convex bodies: the Brunn-Minkowski theory* (Encyclopedia of Mathematics and its Applications, 44. Cambridge University Press, Cambridge, 1993).

[52] W. F. Stout, *Almost sure convergence* (Probability and Mathematical Statistics, Vol. 24. Academic Press 1974).

[53] M. Talagrand, An isoperimetric theorem on the cube and the Kintchine-Kahane inequalities *Proc. Amer. Math. Soc.* **104** (1988) 905–909.

[54] M. Talagrand, Isoperimetry and integrability of the sum of independent Banach-space valued random variables, *Ann. Probab.* **17** (1989) 1546–1570.

[55] M. Talagrand, Embedding subspaces of L_1 into l_1^N, *Proc. Amer. Math. Soc.* **108** (1990) 363–369.

[56] M. Talagrand, A new isoperimetric inequality and the concentration of measure phenomenon, *Geometric aspects of functional analysis (Israel, 1989–90)*, (Lecture Notes in Math., Springer) **1469** (1991) 94–124.

[57] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, *Inst. Hautes Études Sci. Publ. Math.* **81** (1995) 73–205.

[58] M. Talagrand, Transportation cost for Gaussian and other product measures, *Geom. Funct. Anal.* **6** (1996) 587–600.

[59] V. V. Yurinsky, Exponential bounds for large deviations, *Theor. Probab. Appl.* **19** (1974) 154–155.