On the role of chaining methods in concentration inequalities for polynomials

Radosław Adamczak

University of Warsaw

Chaining Methods
and their Applications to Computer Science
Harvard University 2016

Definition

For a random variable X, define the Orlicz norm

$$\|X\|_{\psi_2} = \inf\{a > 0 \colon \mathbb{E}\exp(|X|^2/a^2) \leq 2\}.$$

[1] Throughout the talk C will stand for an absolute numerical constant, the value of which may change between occurrences. Similarly C_α denotes constants depending only on α.

Definition

For a random variable X, define the Orlicz norm

$$\|X\|_{\psi_2} = \inf\{a > 0 \colon \mathbb{E}\exp(|X|^2/a^2) \le 2\}.$$

- By Chebyshev's inequality we have

$$\mathbb{P}(|X| \ge t) \le 2\exp(-t^2/\|X\|_{\psi_2}^2).$$

- Conversely, if for all $t > 0$, $\mathbb{P}(|X| \ge t) \le 2\exp(-t^2/K^2)$, then $\|X\|_{\psi_2} \le CK$(integration by parts)[1].
- Also,

$$\frac{1}{C}\|X\|_{\psi_2} \le \sup_{p\ge 2} \frac{\|X\|_p}{\sqrt{p}} \le C\|X\|_{\psi_2},$$

where $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$.
- If $\|X\|_{\psi_2} < \infty$ we say that X is subgaussian.

---

[1] Throughout the talk C will stand for an absolute numerical constant, the value of which may change between occurrences. Similarly C_α denotes constants depending only on α.

For sums of i.i.d. random variables, we have the classical bound based on the Laplace transform

Theorem

If $X_1, \ldots, X_n$ are independent random variables, $S_n = X_1 + \ldots + X_n$, then

$$\|S_n - \mathbb{E}S_n\|_{\psi_2} \leq C\sqrt{\sum_{i=1}^{n} \|X_i\|_{\psi_2}^2}.$$

In particular,

$$\mathbb{P}(|S_n - \mathbb{E}S_n| \geq t) \leq 2\exp\Big(-\frac{t^2}{C\sum_{i=1}^{n}\|X_i\|_{\psi_2}^2}\Big).$$

For sums of i.i.d. random variables, we have the classical bound based on the Laplace transform

Theorem

If $X_1, \ldots, X_n$ are independent random variables, $S_n = X_1 + \ldots + X_n$, then

$$\|S_n - \mathbb{E}S_n\|_{\psi_2} \leq C\sqrt{\sum_{i=1}^{n} \|X_i\|_{\psi_2}^2}.$$

In particular,

$$\mathbb{P}(|S_n - \mathbb{E}S_n| \geq t) \leq 2\exp\Big(-\frac{t^2}{C\sum_{i=1}^n \|X_i\|_{\psi_2}^2}\Big).$$

Question:

What concentration do we get for polynomials in independent subgaussian random variables?

Theorem (Hanson-Wright, Borell, Ledoux, Arcones-Giné, Latała)

Let $X_1, \ldots, X_n$ be independent mean zero variables with $\|X_i\|_{\psi_2} \le L$. Let $A = (a_{ij})_{i,j=1}^n$ and

$$Z = \sum_{i,j=1}^{n} a_{ij} X_i X_j.$$

Then for $t \ge 0$,

$$\mathbb{P}(|Z - \mathbb{E}Z| \ge t) \le 2\exp\Big(-\frac{1}{C}\min\Big(\frac{t^2}{L^4\|A\|_{HS}^2}, \frac{t}{L^2\|A\|}\Big)\Big).$$

Above $\|A\|_{HS} = \sqrt{\sum_{i,j=1}^n a_{ij}^2}$, $\|A\| = \sup_{x\in S^{n-1}} |Ax|$ are the Hilbert-Schmidt and operator norms of the matrix.

How to generalize this to polynomials of higher degree? Need to introduce counterparts of $\|A\|_{HS}$ and $\|A\|$. We do it via appropriate injective tensor product norms.

We have the following ugly-looking definition:

How to generalize this to polynomials of higher degree? Need to introduce counterparts of $\|A\|_{HS}$ and $\|A\|$. We do it via appropriate injective tensor product norms.

We have the following ugly-looking definition:
Let $A = (a_{i_1\ldots i_d})_{i_1,\ldots,i_d=1}^n$ be a d-indexed array of real numbers.
For a partition $\mathcal{J} = \{J_1, \ldots, J_k\}$ of $\{1, \ldots, d\}$ we define

$$\|A\|_{\mathcal{J}} = \sup\Big\{ \sum_{i_1,\ldots,i_d} a_{i_1\ldots i_d} \prod_{j=1}^{k} x^{(j)}_{(i_r)_{r\in J_j}} \colon \sum_{(i_r)_{r\in J_j}} |x^{(j)}_{(i_r)_{r\in J_j}}|^2 \le 1, j = 1, \ldots, k \Big\}$$

How to generalize this to polynomials of higher degree? Need to introduce counterparts of $\|A\|_{HS}$ and $\|A\|$. We do it via appropriate injective tensor product norms.

We have the following ugly-looking definition:
Let $A = (a_{i_1 \dots i_d})_{i_1,\dots,i_d=1}^n$ be a d-indexed array of real numbers.
For a partition $\mathcal{J} = \{J_1, \dots, J_k\}$ of $\{1, \dots, d\}$ we define

$$\|A\|_{\mathcal{J}} = \sup\Big\{ \sum_{i_1,\dots,i_d} a_{i_1 \dots i_d} \prod_{j=1}^{k} x^{(j)}_{(i_r)_{r\in J_j}} \colon \sum_{(i_r)_{r\in J_j}} |x^{(j)}_{(i_r)_{r\in J_j}}|^2 \le 1, j = 1, \dots, k \Big\}$$

Examples:

$\|(a_{ij})\|_{\{1,2\}} = \sup\{\sum_{ij} a_{ij} x_{ij} \colon \sum_{ij} x_{ij}^2 \le 1\} = \|A\|_{HS}$

$\|(a_{ij})\|_{\{1\}\{2\}} = \sup\{\sum_{ij} a_{ij} x_i y_j \colon \sum_i x_i^2 \le 1, \sum_j y_j^2 \le 1\} = \|A\|$

$\|(a_{ijk})\|_{\{1,2\}\{3\}} = \sup\{\sum_{ijk} a_{ijk} x_{ij} y_k \colon \sum_{ij} x_{ij}^2, \sum_k y_k^2 \le 1\}$

$\|(a_{ijk})\|_{\{1\}\{2\}\{3\}} = \sup\{\sum_{ijk} a_{ijk} x_i y_j z_k \colon \sum_i x_i^2, \sum_j y_j^2, \sum_k z_k^2 \le 1\}$

$\|(a_{ijk})\|_{\{1,2,3\}} = \sup\{\sum_{ijk} a_{ijk} x_{ijk} \colon \sum_{ijk} x_{ijk}^2 \le 1\} = \sqrt{\sum_{ijk} a_{ijk}^2}$

How to generalize this to polynomials of higher degree? Need to introduce counterparts of $\|A\|_{HS}$ and $\|A\|$. We do it via appropriate injective tensor product norms.

We have the following ugly-looking definition:
Let $A = (a_{i_1\dots i_d})_{i_1,\dots,i_d=1}^n$ be a d-indexed array of real numbers.
For a partition $\mathcal{J} = \{J_1, \dots, J_k\}$ of $\{1, \dots, d\}$ we define

$$\|A\|_{\mathcal{J}} = \sup\Big\{ \sum_{i_1,\dots,i_d} a_{i_1\dots i_d} \prod_{j=1}^{k} x^{(j)}_{(i_r)_{r\in J_j}} \colon \sum_{(i_r)_{r\in J_j}} |x^{(j)}_{(i_r)_{r\in J_j}}|^2 \le 1, j = 1, \dots, k \Big\}$$

Examples:
$\|(a_{ij})\|_{\{1,2\}} = \sup\{\sum_{ij} a_{ij}x_{ij} \colon \sum_{ij} x_{ij}^2 \le 1\} = \|A\|_{HS}$
$\|(a_{ij})\|_{\{1\}\{2\}} = \sup\{\sum_{ij} a_{ij}x_i y_j \colon \sum_i x_i^2 \le 1, \sum_j y_j^2 \le 1\} = \|A\|$
$\|(a_{ijk})\|_{\{1,2\}\{3\}} = \sup\{\sum_{ijk} a_{ijk}x_{ij}y_k \colon \sum_{ij} x_{ij}^2, \sum_k y_k^2 \le 1\}$
$\|(a_{ijk})\|_{\{1\}\{2\}\{3\}} = \sup\{\sum_{ijk} a_{ijk}x_i y_j z_k \colon \sum_i x_i^2, \sum_j y_j^2, \sum_k z_k^2 \le 1\}$
$\|(a_{ijk})\|_{\{1,2,3\}} = \sup\{\sum_{ijk} a_{ijk}x_{ijk} \colon \sum_{ijk} x_{ijk}^2 \le 1\} = \sqrt{\sum_{ijk} a_{ijk}^2}$
We let P_d be the set of all partitions of $\{1, \dots, d\}$.

Define $P(x_1,\ldots,x_n) = \sum_{ij} a_{ij}x_ix_j$, $X = (X_1,\ldots,X_n)$. Then the Hanson-Wright inequality can be written as

$$\mathbb{P}(|P(X) - \mathbb{E}P(X)| \geq t) \leq 2\exp\Big(-\frac{1}{C}\frac{t^2}{L^4\|D^2P\|^2_{\{1,2\}}} \wedge \frac{t}{L^2\|D^2P\|_{\{1\}\{2\}}}\Big),$$

where $L = \max_i \|X_i\|_{\psi_2}$.

One can generalize this to arbitrary polynomials (Latała '06 for tetrahedral polynomials in Gaussian variables, Wolff-RA '13 for arbitrary polynomials in subgaussian variables):

Theorem (Wolff-RA '13)

Let $X = (X_1,\ldots,X_n)$, where X_i are independent, with $\|X_i\|_{\psi_2} \leq L$. Let $P\colon \mathbb{R}^n \to \mathbb{R}$ be a polynomial of degree d. Then for any $t > 0$,

$$\mathbb{P}(|P(X) - \mathbb{E}P(X)| \geq t)$$
$$\leq 2\exp\Big(-\frac{1}{C_d}\min_{1\leq i\leq d}\min_{\mathcal{J}\in P_i}\Big(\frac{t}{L^i\|\mathbb{E}D^iP(X)\|_{\mathcal{J}}}\Big)^{\frac{2}{\#\mathcal{J}}}\Big)$$

Theorem

$X = (X_1, \ldots, X_n)$, X_i independent, $\|X_i\|_{\psi_2} \le L$, deg $P = d$*. Then*

$$\mathbb{P}(|P(X) - \mathbb{E}P(X)| \ge t) \le 2\exp\Big(-\frac{1}{C_d}\min_{1\le i\le d}\min_{\mathcal{J}\in P_i}\Big(\frac{t}{L^i\|\mathbb{E}D^i f(X)\|_{\mathcal{J}}}\Big)^{\frac{2}{\#\mathcal{J}}}\Big)$$

Remarks:

- If X_i are standard Gaussian, the ineq. can be reversed.
- The proof of the above version is a rather tedious reduction to the case of multilinear forms in Gaussian variables due to Latała (using general decoupling inequalities, contraction and some combinatorics/linear algebra; **no chaining**).
- Chaining enters the game for Gaussian variables, this is what we will focus on now.

We will consider the following framework:

- $g_i^{(j)}$, $i = 1, \ldots, n, j = 1, \ldots, d$ are i.i.d. standard Gaussian variables
- $A = (a_{i_1,\ldots,i_d})_{i_1,\ldots,i_d=1}^n$ is a d-indexed array of real numbers
- $Z = \sum_{i_1\ldots,i_d=1}^n a_{i_1\ldots i_d} g_{i_1}^{(1)} \cdots g_{i_d}^{(d)}$.

We want to prove the result by Latała (2006):

$$\mathbb{P}(|Z| \geq t) \leq 2\exp\Big(-\frac{1}{C_d}\min_{\mathcal{J}\in P_d}\frac{t^{2/\#\mathcal{J}}}{\|A\|_{\mathcal{J}}^{2/\#\mathcal{J}}}\Big).$$

To get tail estimates it is enough to bound moments, use Chebyshev in L_p and optimize in p. Therefore, our tail inequality is equivalent to

$$\|Z\|_p \leq C_d \sum_{\mathcal{J}\in P_d} p^{\#\mathcal{J}/2}\|A\|_{\mathcal{J}}.$$

$$\|Z\|_p \leq C_p \sum_{\mathcal{J} \in P_d} p^{\#\mathcal{J}/2} \|A\|_{\mathcal{J}}.$$

- For $d = 1$ this is trivial $\| \sum_i a_i g_i \|_p \leq C p^{1/2} \sqrt{\sum_i a_i^2}$.
- For higher d – induction. For $d = 2$ it is still easy:

$$\mathbb{E}|Z|^p = \mathbb{E}|\sum_{ij} a_{ij} g_i^{(1)} g_j^{(2)}|^p \leq C^p p^{p/2} \mathbb{E}|\sum_j (\sum_i a_{ij} g_i)^2|^{p/2}$$

$$= C^p p^{p/2} \mathbb{E} \sup_{|x| \leq 1} |\sum_i (\sum_j a_{ij} x_j) g_i|^p.$$

From Gaussian concentration we always have

$$\| \sup_{t \in T} |\sum_i t_i g_i| \|_p \leq \mathbb{E} \sup_{t \in T} |\sum_i t_i g_i| + C\sqrt{p} \sup_{t \in T} |t|.$$

In our case $\sup_{t \in T} |t| = \|A\|_{\{1\}\{2\}}$ and

$$\mathbb{E} \sup_{t \in T} |\sum_i t_i g_i| \leq \| \sup_{t \in T} |\sum_i t_i g_i| \|_2 = \|A\|_{\{1,2\}}, \text{ so indeed}$$

$$\|Z\|_p \leq C(\sqrt{p} \|A\|_{\{1,2\}} + p \|A\|_{\{1\}\{2\}}).$$

Let's try to repeat this argument with $d = 3$.

$$\mathbb{E}|\sum_{ijk} a_{ijk} g_i^{(1)} g_j^{(2)} g_k^{(3)}|^p$$
$$\leq C^p p^{p/2} \mathbb{E}\|(\sum_k a_{ijk} g_k)_{ij}\|_{\{1,2\}}^p + C^p p^p \mathbb{E}\|(\sum_k a_{ijk} g_k)_{ij}\|_{\{1\}\{2\}}^p$$
$$=: C^p (p^{p/2} \mathbb{E} X^p + p^p \mathbb{E} Y^p)$$

The first term is easy

$$X = \sup_{\sum_{ij} x_{ij}^2 \leq 1} \sum_k \sum_{ij} a_{ijk} x_{ij} g_k,$$

so

$$\|X\|_p \leq \|X\|_2 + \sqrt{p} \sup_{\sum_{ij} x_{ij}^2 \leq 1} \sqrt{\sum_k (\sum_{ij} a_{ijk} x_{ij})^2} = \|X\|_2 + \sqrt{p} \|A\|_{\{1,2\}\{3\}}.$$

Moreover $\|X\|_2^2 = \mathbb{E} \sum_{ij} (\sum_k a_{ijk} g_k)^2 = \sum_{ijk} a_{ijk}^2 = \|A\|_{\{1,2,3\}}^2$.

What about

$$Y = \sup_{\sum_i x_i^2, \sum_j y_j^2 \le 1} \sum_k \sum_{ij} a_{ijk} x_i y_j g_k ?$$

The concentration part works:

$$\|Y\|_p \le \mathbb{E}Y + C\sqrt{p} \sup_{\sum_i x_i^2, \sum_j y_j^2 \le 1} \sqrt{\sum_k \Big(\sum_{ij} a_{ijk} x_i y_j\Big)^2}$$

$$= \mathbb{E}Y + \sqrt{p}\|A\|_{\{1\}\{2\}\{3\}}.$$

The main difficulty (and the only thing that remains) is to estimate $\mathbb{E}Y$.

What about

$$Y = \sup_{\sum_i x_i^2, \sum_j y_j^2 \le 1} \sum_k \sum_{ij} a_{ijk} x_i y_j g_k?$$

The concentration part works:

$$\|Y\|_p \le \mathbb{E}Y + C\sqrt{p} \sup_{\sum_i x_i^2, \sum_j y_j^2 \le 1} \sqrt{\sum_k (\sum_{ij} a_{ijk} x_i y_j)^2}$$

$$= \mathbb{E}Y + \sqrt{p}\|A\|_{\{1\}\{2\}\{3\}}.$$

The main difficulty (and the only thing that remains) is to estimate $\mathbb{E}Y$.
For higher d similar difficulties, gets more technical but the ideas are the same, so we stick to $d = 3$.

To match the powers of p, we need an estimate of the form

$$\mathbb{E} Y = \mathbb{E} \sup_{|x|,|y|\le 1} \sum_k \sum_{i,j} a_{ijk} x_i y_j g_k$$
$$\le C\Big(\frac{1}{\sqrt{p}}\|A\|_{\{1,2,3\}} + \|A\|_{\{1,3\}\{2\}} + \|A\|_{\{1\}\{2,3\}} + \sqrt{p}\|A\|_{\{1\}\{2\}\{3\}}\Big).$$

It will be convenient to work with a more general quantity. For $T \subseteq B_2^n \times B_2^n$ consider

$$F(T) = \mathbb{E} \sup_{(x,y)\in T} \sum_k \sum_{i,j} a_{ijk} x_i y_j g_k$$

$$F(T) = \mathbb{E} \sup_{(x,y)\in T} \sum_k \sum_{i,j} a_{ijk} x_i y_j g_k$$

Assume $T = \bigcup_{m=1}^{N} ((b^m, c^m) + T_m)$, $(b^m, c^m) \in T$, $T_m \subseteq B_2^n \times B_2^n$

$$F(T) = \mathbb{E} \sup_{(x,y)\in T} \sum_k \sum_{i,j} a_{ijk} x_i y_j g_k$$

Assume $T = \bigcup_{m=1}^N ((b^m, c^m) + T_m)$, $(b^m, c^m) \in T$, $T_m \subseteq B_2^n \times B_2^n$
We can write for any (x', y'),

$$F(T) = \mathbb{E} \max_{m\le N} \sup_{(x,y)\in(b^m,c^m)+T_m} \sum_k \sum_{i,j} a_{ijk} x_i y_j g_k$$

$$= \mathbb{E} \max_{m\le N} \sup_{(x,y)\in(b^m,c^m)+T_m} \sum_k \sum_{i,j} a_{ijk} (x_i y_j - x_i' y_j') g_k.$$

By concentration

$$F(t) \le \max_{m\le N} \mathbb{E} \sup_{(x,y)\in(b^m,c^m)+T_m} \sum_k \sum_{i,j} a_{ijk} (x_i y_j - x_i' y_j') g_k$$

$$+ C\sqrt{\log N} \Delta_A(T)$$

$$= \max_{m\le N} F((b^m, c^m) + T_m) + C\sqrt{\log N} \Delta_A(T),$$

where $\Delta_A(T) = \sup_{(x,y),(x',y')\in T} \sqrt{\sum_k (\sum_{ij} a_{ijk} (x_i y_j - x_i' y_j'))^2}$.

$$F(T) = \mathbb{E} \sup_{(x,y)\in T} \sum_k \sum_{i,j} a_{ijk} x_i y_j g_k$$

Assume $T = \bigcup_{m=1}^N ((b^m, c^m) + T_m)$, $(b^m, c^m) \in T$, $T_m \subseteq B_2^n \times B_2^n$
We can write for any (x', y'),

$$F(T) = \mathbb{E} \max_{m\le N} \sup_{(x,y)\in(b^m,c^m)+T_m} \sum_k \sum_{i,j} a_{ijk} x_i y_j g_k$$
$$= \mathbb{E} \max_{m\le N} \sup_{(x,y)\in(b^m,c^m)+T_m} \sum_k \sum_{i,j} a_{ijk} (x_i y_j - x'_i y'_j) g_k.$$

By concentration

$$F(t) \le \max_{m\le N} \mathbb{E} \sup_{(x,y)\in(b^m,c^m)+T_m} \sum_k \sum_{i,j} a_{ijk} (x_i y_j - x'_i y'_j) g_k$$
$$+ C\sqrt{\log N}\Delta_A(T)$$
$$= \max_{m\le N} F((b^m, c^m) + T_m) + C\sqrt{\log N}\Delta_A(T),$$

where $\Delta_A(T) = \sup_{(x,y),(x',y')\in T} \sqrt{\sum_k (\sum_{ij} a_{ijk}(x_i y_j - x'_i y'_j))^2}$.

$$F(T) \leq \max_{m \leq N} F((b^m, c^m) + T_m) + C\sqrt{\log N}\Delta_A(T),$$

where $\Delta_A(T) = \sup_{(x,y),(x',y')\in T} \sqrt{\sum_k (\sum_{ij} a_{ijk}(x_i y_j - x'_i y'_j))^2}$.
Moreover, as $T_m \subseteq B_2^n \times B_2^n$, $(b^m, c^m) \in T$

$$\begin{aligned}
&F((b^m, c^m) + T_m) - F(T_m) \\
&\leq \mathbb{E} \sup_{(x,y)\in T_m} \Big(\sum_k \sum_{ij} a_{ijk} x_i c_j^m g_k + \sum_k \sum_{ij} a_{ijk} b_i^m y_j g_k \Big) \\
&\leq \mathbb{E} \sqrt{\sum_i (\sum_{jk} a_{ijk} c_j^m g_k)^2} + \mathbb{E} \sqrt{\sum_j (\sum_{ik} a_{ijk} b_i^m g_k)^2} \\
&\leq \sup_{(b,c)\in T} \Big(\sqrt{\sum_{ik} (\sum_j a_{ijk} c_j)^2} + \sqrt{\sum_{jk} (\sum_i a_{ijk} b_i)^2} \Big) =: s_A(T)
\end{aligned}$$

So if $T = \bigcup_{m=1}^{N}((b^m, c^m) + T_m)$, we end up with

$$F(T) \leq \max_{m \leq N} F(T_m) + s_A(T) + C\sqrt{\log N}\Delta_A(T),$$

where

$$s_A(T) = \sup_{(b,c)\in T} \Big(\sqrt{\sum_{ik} (\sum_j a_{ijk} c_j)^2} + \sqrt{\sum_{jk} (\sum_i a_{ijk} b_i)^2} \Big)$$

and

$$\Delta_A(T) = \sup_{(x,y),(x',y')\in T} \sqrt{\sum_k (\sum_{ij} a_{ijk}(x_i y_j - x'_i y'_j))^2}.$$

The goal is to find a good partition of T into 'better' sets, replace T with T_m and then iterate.
There are three items we need to control: $N, \Delta_A(T_m), s_A(T_m)$.

The main technical Lemma:

Lemma (Latała '06)

Let $p \geq 1$ and let $T \subset B_2^n \times B_2^n$ satisfy $2 \leq \#T < \infty$, $T - T \subseteq B_2^n \times B_2^n$. Then for any $l \geq 0$ there exist finite sets T_m and points $(b^m, c^m) \in T$ such that

(i) $2 \leq N \leq \exp(C2^{2l}p)$

(ii) $T = \bigcup_{m=1}^{N}((b^m, c^m) + T_m)$, $T_m - T_m \subseteq T - T$, $\#T_m \leq \#T - 1$,

(iii) $\Delta_A(T_m) \leq 2^{-2l}p^{-1}\|A\|_{\{1,2,3\}}$,

(iv) $s_A(T_m) \leq 2^{-l}p^{-1/2}\|A\|_{\{1,2,3\}}$,

Note that by separability and homogeneity,

$$\mathbb{E}Y = F(B_2^n \times B_2^n) = 4 \sup\{F(T) \colon \#T < \infty, T \subseteq 2^{-1}(B_2^n \times B_2^n)\}.$$

By using the lemma we can replace F successively by smaller sets. Note that if $\#T = 1$, then $F(T) = 0$.

We have $s_A(T) \le s_A(B_2^n \times B_2^n) \le \|A\|_{\{1\}\{2,3\}} + \|A\|_{\{2\}\{1,3\}}$, $\Delta_A(T) \le \Delta_A(B_2^n \times B_2^n) \le 2\|A\|_{\{1\}\{2\}\{3\}}$. Therefore

$$
\begin{aligned}
F(T) &\overset{l=1}{\lesssim} \max_{m_1} F(T_{m_1}) + \|A\|_{\{1\}\{2,3\}} + \|A\|_{\{2\}\{1,3\}} + \sqrt{p}\|A\|_{\{1\}\{2\}\{3\}} \\
&\overset{l=2}{\lesssim} \max_{m_1,m_2} F(T_{m_1,m_2}) + \max_{m_1} s_A(T_{m_1}) + \sqrt{2^4 p} \max_{m_1} \Delta_A(T_{m_1}) \\
&+ \|A\|_{\{1\}\{2,3\}} + \|A\|_{\{2\}\{1,3\}} + \sqrt{p}\|A\|_{\{1\}\{2\}\{3\}} \\
&\lesssim \max_{m_1,m_2} F(T_{m_1,m_2}) + 2^{-2}p^{-1/2}\|A\|_{\{1,2,3\}} + \sqrt{2^4 p 2^{-4} p^{-1}}\|A\|_{\{1,2,3\}} \\
&+ \|A\|_{\{1\}\{2,3\}} + \|A\|_{\{2\}\{1,3\}} + \sqrt{p}\|A\|_{\{1\}\{2\}\{3\}} \\
&\lesssim \cdots \lesssim \\
&\lesssim \|A\|_{\{1,2,3\}} \sum_{l=1}^{\infty} 2^{-l} p^{-1/2} + \|A\|_{\{1\}\{2,3\}} + \|A\|_{\{2\}\{1,3\}} \\
&+ \sqrt{p}\|A\|_{\{1\}\{2\}\{3\}} \\
&\lesssim p^{-1/2}\|A\|_{\{1,2,3\}} + \|A\|_{\{1\}\{2,3\}} + \|A\|_{\{2\}\{1,3\}} + \sqrt{p}\|A\|_{\{1\}\{2\}\{3\}}.
\end{aligned}
$$

How to prove the technical lemma?

Lemma (Latała '06)

Let $p \geq 1$ and let $T \subset B_2^n \times B_2^n$ satisfy $\# T \geq 2$, $T - T \subseteq B_2^n \times B_2^n$. Then for any $l \geq 0$ there exist finite sets T_m and points $(b^m, c^m) \in T$ such that

(i) $2 \leq N \leq \exp(C2^{2l}p)$

(ii) $T = \bigcup_{m=1}^{N}((b^m, c^m) + T_m)$, $T_m - T_m \subseteq T - T$, $\# T_m \leq \# T - 1$,

(iii) $\Delta_A(T_m) \leq 2^{-2l}p^{-1}\|A\|_{\{1,2,3\}}$,

(iv) $s_A(T_m) \leq 2^{-l}p^{-1/2}\|A\|_{\{1,2,3\}}$,

Let us focus on (iv) first. Recall

$$
\begin{aligned}
s_A(S) &:= \sup_{(x,y)\in S}\Big(\sqrt{\sum_{ik}(\sum_j a_{ijk}y_j)^2} + \sqrt{\sum_{jk}(\sum_i a_{ijk}x_i)^2}\Big) \\
&= \sup_{(x,y)\in S} |||(x,y)|||
\end{aligned}
$$

$$|||(x,y)||| := \Big(\sqrt{\sum_{ik}(\sum_{j} a_{ijk}y_j)^2} + \sqrt{\sum_{jk}(\sum_{i} a_{ijk}x_i)^2}\Big)$$

We want to decompose $T \subseteq B_2^n \times B_2^n \subseteq \sqrt{2}B_2^{2n}$ into $N \leq \exp(C2^{2l}p)$ sets with $\mathrm{diam}(S_i, |||\cdot|||) \leq 2^{-l}p^{-1/2}\|A\|_{\{1,2,3\}}$.

Theorem (Pajor, Tomczak-Jaegermann)

If $T \subseteq B_2^n$ and $\|\cdot\|$ is a norm on $\mathbb{R}^n$, then for $\varepsilon \in (0,1)$, T can be partitioned into $N \leq \exp(C\varepsilon^{-2})$ sets $S_1, \ldots, S_N$ such that for all i, $\mathrm{diam}(S_i, \|\cdot\|) \leq \varepsilon\mathbb{E}\|G_n\|$, where G_n is the standard Gaussian vector in $\mathbb{R}^n$.

We have $2n$ instead of n, $\varepsilon = 2^{-l}p^{-1/2}$ and

$$\mathbb{E}|||G_{2n}||| = \mathbb{E}\Big(\sqrt{\sum_{ik}(\sum_{j} a_{ijk}g_j)^2} + \sqrt{\sum_{jk}(\sum_{i} a_{ijk}g_i)^2}\Big)$$

$$\leq 2\sqrt{\sum_{ijk} a_{ijk}^2} = 2\|A\|_{\{1,2,3\}}.$$

Thus we partition T into sets $S_1, \ldots, S_{N_1}$, $2 \leq N_1 \leq \exp(C2^{2l}p)$, assume wlog that they are pairwise disjoint and nonempty (so $\#S_i < \#T$), shift each of them and write

$$S_m = (b^m, c^m) + \tilde{T}_m, \text{ with } (b^m, c^m) \in S_m.$$

These are not the sets T_m yet. We also need (iv), i.e.

$$\Delta_A(T_m) := \sup_{(x,y),(x',y') \in T_m} \sqrt{\sum_k (\sum_{ij} a_{ijk}(x_i y_j - x'_i y'_j))^2} \leq 2^{-2l} p^{-1} \|A\|_{\{1,2,3\}}.$$

To this end we further partition the sets $\tilde{T}_m$ using ideas similar as in the proof of the dual Sudakov minoration (to bound from above the covering numbers, we bound from below the Gaussian measure of balls corresponding to our metric).

Lemma (Latała '06)

Let $\|\cdot\|$ be a norm on $\mathbb{R}^{n^2}$ and let ρ be a metric on $B_2^n \times B_2^n$ given by

$$\rho((x,y)-(x',y')) = \|x\otimes y - x'\otimes y'\|, \textit{ where } x\otimes y = (x_i y_j)_{i,j=1}^n.$$

Then for every set $T \subseteq B_2^n \times B_2^n$ and any $\varepsilon \in (0,1)$ there exists a partition of T into $N \leq \exp(C\varepsilon^{-2})$ sets S_i such that

$$\mathrm{diam}(S_i,\rho) \leq \varepsilon \sup_{(x,y)\in T} \mathbb{E}\|x\otimes G_n\| + \varepsilon \sup_{(x,y)\in T} \mathbb{E}\|G_n \otimes y\| + \varepsilon^2 \mathbb{E}\|G_n \otimes G_n'\|,$$

where G_n, G_n' are i.i.d. standard Gaussian vectors in $\mathbb{R}^n$.

In our case we want to partition the sets $\tilde{T}_m$, the metric comes from $\|\cdot\| = \sqrt{\sum_k (\sum_{ij} a_{ijk} x_{ij})^2}$ and $\varepsilon = 2^{-l} p^{-1/2}$.

We want to partition the sets $\tilde{T}_m$ into $N_2 \leq \exp(C2^{2l}p)$ sets $T_{m,r}$, $\|\cdot\| = \sqrt{\sum_k(\sum_{ij} a_{ijk}x_{ij})^2}$, $\varepsilon = 2^{-l}p^{-1/2}$.

$$\Delta_A(T_{m,r}) \leq 2^{-l}p^{-1/2} \sup_{(x,y)\in\tilde{T}_m} \mathbb{E}\|x \otimes G_n\| + 2^{-l}p^{-1/2} \sup_{(x,y)\in\tilde{T}_m} \mathbb{E}\|G_n \otimes y\|$$
$$+ 2^{-2l}p^{-1}\mathbb{E}\|G_n \otimes G'_n\|.$$

We have

$$\sup_{(x,y)\in\tilde{T}_m} \mathbb{E}\|x \otimes G_n\| = \sup_{(x,y)\in\tilde{T}_m} \mathbb{E}\sqrt{\sum_k(\sum_{ij} a_{ijk}x_i g_j)^2}$$
$$\leq \sup_{(x,y)\in\tilde{T}_m} \sqrt{\sum_{jk}(\sum_i a_{ijk}x_i)^2} \leq s_A(\tilde{T}_m) \leq 2^{-l}p^{-1/2}\|A\|_{\{1,2,3\}}.$$

The second term is analogous. For the third term,

$$\mathbb{E}\|G_n \otimes G'_n\| = \mathbb{E}\sqrt{\sum_k(\sum_{ij} a_{ijk}g_i g'_j)^2} \leq \|A\|_{\{1,2,3\}}.$$

Altogether we get $\Delta_A(T_{m,r}) \leq 3 \cdot 2^{-2l}p^{-1}\|A\|_{\{1,2,3\}}$. Now we relabel the family $T_{m,r}$ to T_m ($N \leq N_1 N_2 \leq \exp(C2^{2l}p)$). $\square$

Further developments:

- Two-sided moment estimates for multilinear forms in nonnegative indep. random variables with log-concave tail (i.e. $t \mapsto \log \mathbb{P}(|X| \geq t)$ is concave) – Latała-Łochowski
- Recently generalized by Meller to nonnegative variables s.t. $\|X\|_{2p} \leq C\|X\|_p$.
- The case of symmetric random variables with log-concave tails ($t \mapsto \log \mathbb{P}(|X| \geq t)$ is concave) is known only for $d = 1$ (Gluskin-Kwapień), $d = 2$ (Latała) and $d = 3$ (Latała-RA), except for multilinear forms in exponential variables, known for all d (Latała-RA).
- Multilinear forms in variables with log-convex tails – Hitczenko,Montgomery-Smith,Oleszkiewicz (for $d = 1$), Kolesko-Latała (general d).
- Counterparts for U-statistics – RA, Latała-RA
- Upper bounds for functions in dependent random variables with bounded derivatives of higher order, under log-Sobolev type assumptions – Wolff-RA

Different approaches, with other quantities controlling the tail

- Kim-Vu (better dependence of constants on the degree, less precise for fixed degree)
- Schudy-Sviridenko

Some open problems

- two sided bounds for moments of polynomials in $\{0,1\}$-valued variables (applications to random graphs)
- symmetric variables with log-concave tails, $d > 3$ (in particular Rademachers)
- polynomials with coefficients in a Banach space (some applications in statistics); the case of Hilbert spaces and Gaussian variables can be deduced from Latała's result
    - a special case: good inequalities for quadratic forms with matrix-coefficients

Thank you