Nemirovski's Inequalities Revisited

Lutz Dümbgen, Sara A. van de Geer, and Jon A. Wellner
University of Bern, ETH Zürich, and University of Washington

August 21, 2008

Abstract

An important tool for statistical research are moment inequalities for sums of independent random vectors. Nemirovski and coworkers (1983, 2000) derived one particular type of such inequalities: For certain Banach spaces $(\mathbb{B}, \|\cdot\|)$ there exists a constant $K = K(\mathbb{B}, \|\cdot\|)$ such that for arbitrary independent and centered random vectors $X_1, X_2, \ldots, X_n \in \mathbb{B}$, their sum S_n satisfies the inequality $\mathbb{E}\,\|S_n\|^2 \le K \sum_{i=1}^n \mathbb{E}\,\|X_i\|^2$. We present and compare three different approaches to obtain such inequalities: Nemirovski's results are based on deterministic inequalities for norms. Another possible vehicle are type and cotype inequalities, a tool from probability theory on Banach spaces. Finally, we use a truncation argument plus Bernstein's inequality to obtain another version of the moment inequality above. Interestingly, all three approaches have their own merits.

1 Introduction

A major theme in current statistical research concerns problems in which the "sample size" (or number of independent units) n is small or moderate, say on the order of 10^2 or 10^4, while the number d of items measured for each independent unit is large, say on the order of 10^6 or 10^7. Studies of the properties of statistical methods for such problems often rely on "maximal inequalities" for sums of independent random variables. Such inequalities and related tools are the subject of empirical process theory as developed in Dudley (1999), Pollard (1990), van de Geer (2000), and van der Vaart and Wellner (1996). The maximal inequalities developed in empirical process theory typically involve large (uncountable) classes of functions, are usually formulated in terms of (uniform) covering numbers or bracketing entropy numbers, and are built up from basic inequalities for finite classes of functions via chaining arguments; see e.g. section 2.2 of van der Vaart and Wellner (1996), section 3 of Pollard (1990), section 5.1 of de la Peña and Giné (1999), or section 3.2 of van de Geer (2000). Similar inequalities for finite classes of functions (or sums of independent random vectors) have been derived by way of probabilistic methods for Banach spaces, and via deterministic inequalities for norms. One interesting inequality of the latter type due to Nemirovski and Yudin (1983) and Nemirovski (2000) was used by Greenshtein and Ritov

(2004) in their study of the "persistence properties" of lasso methods for regression in small n, large d problems.

Our goals in this paper are to compare the inequalities resulting from the three different approaches (deterministic inequalities for norms, probabilistic methods for Banach spaces, empirical process theory) in the case of finite classes of functions or random vectors and to refine or improve the constants involved in each case. The improved constants in these inequalities for finite classes of functions may be of interest for development of sharpened versions of the empirical process inequalities for large classes of functions $\mathcal{F}$ with explicit constants.

Generally we are aiming at inequalities of the following type: Let $X_1, X_2, \ldots, X_n$ be stochastically independent random vectors with values in a (real) Banach space $(\mathbb{B}, \|\cdot\|)$ such that $\mathbb{E}\, X_i = 0$ and $\mathbb{E}\, \|X_i\|^2 < \infty$. With $S_n := \sum_{i=1}^n X_i$ we want to show that

$$\mathbb{E}\, \|S_n\|^2 \;\le\; K \sum_{i=1}^n \mathbb{E}\, \|X_i\|^2 \tag{1.1}$$

for some constant K depending only on $(\mathbb{B}, \|\cdot\|)$.

An important special case are Hilbert spaces $(\mathbb{B}, \langle\cdot,\cdot\rangle, \|\cdot\|)$. Here inequality (1.1) turns out to be an equality with constant $K = 1$, because

$$\mathbb{E}\, \|S_n\|^2 \;=\; \sum_{i,j=1}^n \mathbb{E}\langle X_i, X_j\rangle \;=\; \sum_{i=1}^n \mathbb{E}\, \|X_i\|^2$$

by independence and mean zero of the random vectors X_i.

For statistical applications, the case $\mathbb{B} = \mathbb{R}^d$ and $\|\cdot\| = \|\cdot\|_r$ for some $r \in [1, \infty]$ is of particular interest. Here the r-norm of a vector $x \in \mathbb{R}^d$ is defined as

$$\|x\|_r \;:=\; \begin{cases} \Big(\sum_{j=1}^d |x_j|^r\Big)^{1/r} & \text{if } 1 \le r < \infty, \\ \max_{1\le j\le d} |x_j| & \text{if } r = \infty. \end{cases} \tag{1.2}$$

Nemirovski's inequality in the form stated in Nemirovski (2000) and used by Greenshtein and Ritov (2004) says that (1.1) holds with $K = C\min(r, \log(d))$ if $d \ge 2$ for some universal, but unspecified constant C.

The remainder of this paper is organized as follows: In Section 2 we review several deterministic inequalities for norms and, in particular, key arguments of Nemirovski (2000). Our exposition includes explicit and improved constants. While finishing the present paper we became aware of yet unpublished work of Nemirovski (2004) and Juditsky and Nemirovski (2008) who also improved some inequalities of Nemirovski (2000). Rio (2008) uses similar methods in a different context. In Section 3 we present inequalities of type (1.1) which follow from type and co-type inequalities developed in probability theory on Banach spaces. Section 4 presents an alternative approach for the special case of $\mathbb{B} = \mathbb{R}^d$ and $\|\cdot\| = \|\cdot\|_\infty$ which is based on truncation and Bernstein's inequality. Finally, in Section 5 we compare the inequalities resulting from these three approaches. In that section we relax the assumption that $\mathbb{E}\, X_i = 0$ for a more thorough understanding of the differences between the three approaches. Proofs are deferred to Section 6.

2 Nemirovski's approach: Deterministic inequalities for norms

In the this section we review and refine inequalities of type (1.1) based on deterministic inequalities for norms. The considerations for $\mathbb{B} = \mathbb{R}^d$ and $\|\cdot\| = \|\cdot\|_r$ follow closely the arguments of Nemirovski (2000).

2.1 Some inequalities for $\mathbb{R}^d$ and the norms $\|\cdot\|_r$

Throughout this subsection let $\mathbb{B} = \mathbb{R}^d$, equipped with one of the norms $\|\cdot\|_r$ defined in (1.2).

A first solution. Recall that for any $x \in \mathbb{R}^d$,

$$\|x\|_r \ \le\ \|x\|_q \ \le\ d^{1/q-1/r}\|x\|_r \quad \text{for } 1 \le q < r \le \infty. \tag{2.1}$$

Moreover, as mentioned before,

$$\mathbb{E}\,\|S_n\|_2^2 \ =\ \sum_{i=1}^n \mathbb{E}\,\|X_i\|_2^2.$$

Thus for $1 \le q < 2$,

$$\mathbb{E}\,\|S_n\|_q^2 \ \le\ (d^{1/q-1/2})^2\,\mathbb{E}\,\|S_n\|_2^2 \ =\ d^{2/q-1}\sum_{i=1}^n \mathbb{E}\,\|X_i\|_2^2 \ \le\ d^{2/q-1}\sum_{i=1}^n \mathbb{E}\,\|X_i\|_q^2,$$

whereas for $2 < r \le \infty$,

$$\mathbb{E}\,\|S_n\|_r^2 \ \le\ \mathbb{E}\,\|S_n\|_2^2 \ =\ \sum_{i=1}^n \mathbb{E}\,\|X_i\|_2^2 \ \le\ d^{1-2/r}\sum_{i=1}^n \mathbb{E}\,\|X_i\|_r^2.$$

Thus we may conclude that (1.1) holds with

$$K = K_N(d,r) \ :=\ \begin{cases} d^{2/r-1} & \text{if } 1 \le r \le 2, \\ d^{1-2/r} & \text{if } 2 \le r \le \infty. \end{cases}$$

Example 2.1 In case of $1 \le r \le 2$, the preceding result is sharp, as can be seen from the following example: Let $b_1, b_2, \ldots, b_d$ be the standard basis of $\mathbb{R}^d$, and for $i = 1, 2, \ldots, d$ let $X_i := \epsilon_i b_i$ with independent Rademacher variables $\epsilon_1, \epsilon_2, \ldots, \epsilon_d \sim \mathrm{Unif}\{-1, 1\}$. Then $\mathbb{E}\, X_i = 0$, $\|X_i\|_r = 1$ and $\|S_d\|_r = d^{1/r}$, so that $\mathbb{E}\,\|S_d\|_r^2 = d^{2/r-1}\sum_{i=1}^d \mathbb{E}\,\|X_i\|_r^2$.

A refinement for $r > 2$. In what follows we shall obtain a substantially smaller constant $K_N(d,r)$ for large r. The main ingredient is the following result:

Lemma 2.2 *For arbitrary fixed $r \in [2,\infty)$ and $x \in \mathbb{R}^d \setminus \{0\}$ let*

$$h(x) \ := \ 2\|x\|_r^{2-r} \bigl(|x_i|^{r-2} x_i\bigr)_{i=1}^d$$

while $h(0) := 0$. Then for arbitrary $x, y \in \mathbb{R}^d$,

$$\|x\|_r^2 + h(x)^\top y \ \le \ \|x+y\|_r^2 \ \le \ \|x\|_r^2 + h(x)^\top y + (r-1)\|y\|_r^2.$$

Nemirovski and Yudin (1983) and Nemirovski (2000) stated Lemma 2.1 with the factor $r-1$ on the right side replaced with Cr for some (absolute) constant $C > 1$. Lemma 2.2, which is a special case of the more general Lemma 2.6 in the next subsection, may be applied to the partial sums $S_0 := 0$ and $S_k := \sum_{i=1}^k X_i$, $1 \le k \le n$, to show that for $2 \le r < \infty$,

$$\begin{aligned} \mathbb{E}\,\|S_k\|_r^2 \ &\le \ \mathbb{E}\bigl(\|S_{k-1}\|_r^2 + h(S_{k-1})^\top X_k + (r-1)\|X_k\|_r^2\bigr) \\ &= \ \mathbb{E}\,\|S_{k-1}\|_r^2 + \mathbb{E}\,h(S_{k-1})^\top\,\mathbb{E}\,X_k + (r-1)\,\mathbb{E}\,\|X_k\|_r^2 \\ &= \ \mathbb{E}\,\|S_{k-1}\|_r^2 + (r-1)\,\mathbb{E}\,\|X_k\|_r^2, \end{aligned}$$

and inductively we obtain a second candidate for $K(d,r)$:

$$\mathbb{E}\,\|S_n\|_r^2 \ \le \ (r-1)\sum_{i=1}^n \mathbb{E}\,\|X_i\|_r^2 \quad \text{for } 2 \le r < \infty.$$

Finally, we apply (2.1) again: For $2 \le q \le r \le \infty$ with $q < \infty$,

$$\mathbb{E}\,\|S_n\|_r^2 \ \le \ \mathbb{E}\,\|S_n\|_q^2 \ \le \ (q-1)\sum_{i=1}^n \mathbb{E}\,\|X_i\|_q^2 \ \le \ (q-1)d^{2/q-2/r}\sum_{i=1}^n \mathbb{E}\,\|X_i\|_r^2.$$

This inequality entails our first ($q = 2$) and second ($q = r < \infty$) preliminary result, and we arrive at the following refinement:

Theorem 2.3 *For arbitrary $r \in [2,\infty]$,*

$$\mathbb{E}\,\|S_n\|_r^2 \ \le \ K_N(d,r)\sum_{i=1}^n \mathbb{E}\,\|X_i\|_r^2$$

with

$$K_N(d,r) \ := \ \inf_{q\in[2,r]\cap\mathbb{R}} (q-1)d^{2/q-2/r}.$$

This constant $K_N(d,r)$ satisfies the (in)equalities

$$K_N(d,r) \begin{cases} = & d^{1-2/r} & \text{if } d \le 7 \\ \le & r-1 & \\ \le & 2e\log d - e & \text{if } d \ge 3, \end{cases}$$

and

$$K_N(d,\infty) \ \ge\ 2e\log d - 3e.$$

Corollary 2.4 *In case of $\mathbb{B} = \mathbb{R}^d$, $d \ge 3$, and $\|\cdot\| = \|\cdot\|_\infty$, inequality (1.1) holds with constant $K = 2e\log d - e$. If the X_i's are also identically distributed, then*

$$\mathbb{E}\,\|n^{-1/2}S_n\|_\infty^2 \ \le\ (2e\log d - e)\,\mathbb{E}\,\|X_1\|_\infty^2.$$

Remark 2.5 At least for $r = \infty$ and large d, the constant $K = 2e\log d - e$ cannot be improved substantially. For let $X_1,\ldots,X_n$ be independent, identically distributed on $\{-1,1\}^d$. Then $\mathbb{E}\,X_i = 0$ and $\|X_i\|_\infty = 1$, while $n^{-1/2}S_n$ converges in distribution to a standard Gaussian random vector $Z \in \mathbb{R}^d$ as $n \to \infty$. But it is well-known that $\|Z\|_\infty = \sqrt{2\log d} + o_p(1)$ as $d \to \infty$. Hence any surrogate $K_*(d,\infty)$ for $K_N(d,\infty)$ will satisfy

$$\liminf_{d\to\infty} \frac{K_*(d,\infty)}{2\log d} \ \ge\ 1.$$

2.2 Arbitrary L_r-spaces

Lemma 2.2 is a special case of a more general inequality: Let (T,Σ,μ) be a σ-finite measure space, and for $1 \le r < \infty$ let $L_r(\mu)$ be the set of all measurable functions $f : T \to \mathbb{R}$ with finite (semi-) norm

$$\|f\|_r \ :=\ \Big(\int |f|^r\,d\mu\Big)^{1/r},$$

where two such functions are viewed as equivalent if they coincide almost everywhere with respect to μ. In what follows we investigate the functional

$$f \ \mapsto\ V(f) := \|f\|_r^2$$

on $L_r(\mu)$. Note that $(\mathbb{R}^d, \|\cdot\|_r)$ corresonds to $(L_r(\mu), \|\cdot\|_r)$ if we take $T = \{1,2,\ldots,d\}$ equipped with counting measure μ.

Note again that $V(\cdot)$ is convex, so for arbitrary $f,g \in L_r(\mu)$, the directional derivative

$$DV(f,g) \ :=\ \lim_{t\downarrow 0} t^{-1}\big(V(f+tg) - V(f)\big)$$

exists and is a sublinear function of g. Moreover it is well known from convex analysis that

$$V(f) + DV(f,g) \ \le \ V(f+g).$$

The next theorem provides an explicit expression for $DV(f,g)$ and an upper bound for $V(f+g)$ which improves an inequality of Nemirovski and Yudin (1983).

Lemma 2.6 *Let $r \ge 2$. Then for arbitrary $f, g \in L_r(\mu)$,*

$$DV(f,g) \ = \ \int h(f) g \, d\mu \quad \textit{with} \quad h(f) \ := \ 2|f|^{r-2} f \ \in \ L_q(\mu),$$

where $q := r/(r-1)$. Moreover,

$$V(f) + DV(f,g) \ \le \ V(f+g) \ \le \ V(f) + DV(f,g) + (r-1)V(g).$$

Remark 2.7 The upper bound for $V(f+g)$ is sharp in the following sense: Suppose that $\mu(T) < \infty$, and let $f, g_o : T \to \mathbb{R}$ be measurable such that $|f| \equiv |g_o| \equiv 1$ and $\int f g_o \, d\mu = 0$. Then our proof of Lemma 2.6 reveals that

$$\frac{V(f + tg_o) - V(f) - DV(f, tg_o)}{V(tg_o)} \ \to \ r - 1 \quad \text{as } t \to 0.$$

Remark 2.8 In case of $r = 2$, Lemma 2.6 is well known and easily verified. Here the upper bound for $V(f+g)$ is even an equality, i.e.

$$V(f+g) \ = \ V(f) + DV(f,g) + V(g).$$

Lemma 2.6 leads directly to the following result:

Corollary 2.9 *In case of $\mathbb{B} = L_r(\mu)$, inequality (1.1) is satisfied with $K = r - 1$.*

2.3 A connection to geometrical functional analysis

For any Banach space $(\mathbb{B}, \|\cdot\|)$ and Hilbert space $(\mathbb{H}, \langle\cdot,\cdot\rangle, \|\cdot\|)$, their Banach-Mazur distance $D(\mathbb{B}, \mathbb{H})$ is defined to be the infimum of

$$\|T\| \cdot \|T^{-1}\|$$

over all linear isomorphisms $T : \mathbb{B} \to \mathbb{H}$, where $\|T\|$ and $\|T^{-1}\|$ denote the usual operator norms

$$\begin{aligned} \|T\| \ &:= \ \sup\bigl\{\|Tx\| : x \in \mathbb{B}, \|x\| \le 1\bigr\}, \\ \|T^{-1}\| \ &:= \ \sup\bigl\{\|T^{-1}y\| : y \in \mathbb{H}, \|y\| \le 1\bigr\}. \end{aligned}$$

(If no such bijection exists, one defines $D(\mathbb{B},\mathbb{H}) := \infty$.) Given such a bijection T,

$$\begin{aligned} \mathbb{E}\,\|S_n\|^2 &\le \|T^{-1}\|^2\,\mathbb{E}\,\|TS_n\|^2 \\ &= \|T^{-1}\|^2 \sum_{i=1}^n \mathbb{E}\,\|TX_i\|^2 \\ &\le \|T^{-1}\|^2\|T\|^2 \sum_{i=1}^n \mathbb{E}\,\|X_i\|^2. \end{aligned}$$

This leads to the following observation:

Corollary 2.10 *For any Banach space $(\mathbb{B}, \|\cdot\|)$ and any Hilbert space $(\mathbb{H}, \langle,\cdot,\cdot,\rangle, \|\cdot\|)$ with finite Banach-Mazur distance $D(\mathbb{B},\mathbb{H})$, inequality (1.1) is satisfied with $K = D(\mathbb{B},\mathbb{H})^2$.*

A famous result from geometrical functional analysis is John's theorem (cf. Tomczak-Jaegermann 1989, Johnson and Lindenstrauss 2001) for finite-dimensional normed spaces. It entails that $D(\mathbb{B}, \mathbb{R}^{\dim(\mathbb{B})}) \le \sqrt{\dim(\mathbb{B})}$, where $\mathbb{R}^d$ is equipped with the standard inner product. This entails the following fact:

Corollary 2.11 *For any normed space $(\mathbb{B}, \|\cdot\|)$ with finite dimension, inequality (1.1) is satisfied with $K = \dim(\mathbb{B})$.*

3 The probabilistic approach: Type and co-type inequalities

3.1 Rademacher type and cotype inequalities

Let $\{\epsilon_i\}$ denote a sequence of independent Rademacher random variables. Let $1 \le p < \infty$. A Banach space $\mathbb{B}$ with norm $\|\cdot\|$ is said to be of *(Rademacher) type* p if there is a constant T_p such that for all finite sequences $\{x_i\}$ in $\mathbb{B}$,

$$\mathbb{E}\Big\|\sum_{i=1}^n \epsilon_i x_i\Big\|^p \le T_p^p \sum_{i=1}^n \|x_i\|^p.$$

Similarly, for $1 \le q < \infty$, $\mathbb{B}$ is of *(Rademacher) cotype* q if there is a constant C_q such that for all finite sequences $\{x_i\}$ in $\mathbb{B}$,

$$\mathbb{E}\Big\|\sum_{i=1}^n \epsilon_i x_i\Big\|^q \ge C_q^{-q} \left(\sum_{i=1}^n \|x_i\|^q\right)^{1/q}.$$

Ledoux and Talagrand (1991), page 247, note that type and cotype properties appear as dual notions: If a Banach space $\mathbb{B}$ is of type p, its dual space $\mathbb{B}'$ is of cotype $q = p/(p-1)$.

One of the basic results concerning Banach spaces with type p and cotype q is the following proposition (due to Hoffmann-Jørgensen):

Proposition 3.1 *(Ledoux and Talagrand, 1991, Proposition 9.11, page 248).*
If $\mathbb{B}$ is of type $p \geq 1$ with constant T_p, then

$$\mathbb{E}\,\|S_n\|^p \;\leq\; (2T_p)^p \sum_{i=1}^{n} \mathbb{E}\,\|X_i\|^p.$$

If $\mathbb{B}$ is of cotype $q \geq 1$ with constant C_q, then

$$\mathbb{E}\,\|S_n\|^q \;\geq\; (2C_q)^{-q} \sum_{i=1}^{n} \mathbb{E}\,\|X_i\|^q.$$

As shown in Ledoux and Talagrand (1991), page 27, the Banach space $L_r(\mu)$ with $1 \leq r < \infty$ (cf. section 2.2) is of type r when $r \leq 2$ and of type 2 for $r \geq 2$. Similarly, $L_r(\mu)$ is co-type r for $r \geq 2$ and co-type 2 for $r \leq 2$.

In case of $r \geq 2 = p$, explicit values for the constant T_p in Proposition 3.1 can be obtained from the optimal constants in Khintchine's inequalities due to Haagerup (1982).

Lemma 3.2 *For $2 \leq r < \infty$, the space $L_r(\mu)$ is of type 2 with constant $T_2 = B_r$, where*

$$B_r \;:=\; 2^{1/2} \left(\frac{\Gamma((r+1)/2)}{\sqrt{\pi}} \right)^{1/r}.$$

Corollary 3.3 *For $\mathbb{B} = L_r(\mu)$, $2 \leq r < \infty$, inequality (1.1) is satisfied with $K = 4B_r^2$.*

Remark 3.4 Note that $B_2 = 1$ and

$$\frac{B_r}{\sqrt{r}} \;\to\; \frac{1}{\sqrt{e}} \quad \text{as } r \to \infty.$$

Thus for large values r, the conclusion of Corollary 3.3 is weaker than the one of Corollary 2.9.

3.2 Gaussian type 2 inequalities

Another connecting link is via Gaussian type constants. Here one replaces the Rademacher sequence $\{\epsilon_i\}$ with a sequence $\{Z_i\}$ of independent standard Gaussian random variables. A Banach space $\mathbb{B}$ is called *Gaussian type* p, if there exists a constant $T_p > 0$ such that

$$\mathbb{E}\Big\|\sum_i Z_i x_i\Big\|^p \;\leq\; T_p^p \sum_i \|x_i\|^p$$

for arbitrary fixed finite sequences $\{x_i\}$ in $\mathbb{B}$. Let $T_2^R(\mathbb{B})$ and $T_2^G(\mathbb{B})$ be the smallest possible Rademacher and Gaussian type 2 constants, respectively. Then

$$T_2^G(\mathbb{B}) \;\leq\; T_2^R(\mathbb{B}) \;\leq\; \sqrt{\pi/2}\, T_2^G(\mathbb{B}); \tag{3.1}$$

see e.g. (Pisier, 1986, Proposition 3.2, page 187) and our proof in Section 6.

For the special space $\ell_\infty^d := (\mathbb{R}^d, \|\cdot\|_\infty)$, it follows from Sidák's theorem (Šidák, 1968) that

$$T_2^G(\ell_\infty^d) \;=\; c_d \;:=\; \sqrt{\mathbb{E}\max_{1\le j\le d} Z_j^2}, \tag{3.2}$$

and

$$c_d \;\le\; \sqrt{2\log d} \quad \text{if } d \ge 3. \tag{3.3}$$

Combining these facts shows that

$$T_2^R(\ell_\infty^d) \;\le\; \sqrt{\pi/2}\, c_d \;\le\; \sqrt{\pi \log d} \quad \text{if } d \ge 3.$$

Using this result together with the Hoffmann-Jørgensen inequality (Proposition 3.1) yields another Nemirovski type inequality:

Corollary 3.5 *If $(\mathbb{B}, \|\cdot\|) = \ell_\infty^d$ for some $d \ge 3$, then inequality (1.1) holds with*

$$K \;=\; K_{T2}(d,\infty) := 2\pi c_d^2 \;\le\; 4\pi\log d.$$

4 The empirical process approach: Truncation and Bernstein's inequality

The random vectors $X_i \in \mathbb{R}^d$ are split into two random vectors via truncation. Namely, let $X_i = X_i^{(a)} + X_i^{(b)}$ with

$$X_i^{(a)} \;:=\; 1_{[\|X_i\|_\infty \le \kappa_o]} X_i \quad \text{and} \quad X_i^{(b)} \;:=\; 1_{[\|X_i\|_\infty > \kappa_o]} X_i$$

for some constant $\kappa_o > 0$ to be specified later. Then we write $S_n = A_n + B_n$ with the centered random sums

$$A_n \;:=\; \sum_{i=1}^n (X_i^{(a)} - \mathbb{E}\, X_i^{(a)}) \quad \text{and} \quad B_n \;:=\; \sum_{i=1}^n (X_i^{(b)} - \mathbb{E}\, X_i^{(b)}).$$

The sum A_n involves centered random vectors in $[-2\kappa_o, 2\kappa_o]^d$ and will be treated by means of Bernstein type inequalities. To this end we introduce the *linexp* function

$$\mathrm{e}(L) \;:=\; \exp(1/L) - 1 - 1/L, \quad L > 0.$$

Lemma 4.1 *Let $Z \in [-\kappa, \kappa]$ have mean zero and variance σ^2. Then for any $L > 0$,*

$$\log \mathbb{E}\exp\Big(\frac{Z}{\kappa L}\Big) \;\le\; \log\Big(1 + \frac{\sigma^2 \mathrm{e}(L)}{\kappa^2}\Big) \;\le\; \frac{\sigma^2\mathrm{e}(L)}{\kappa^2}.$$

Lemma 4.2 *Suppose that $X_i = (X_{i,j})_{j=1}^d$ satisfies $\|X_i\|_\infty \le \kappa$ and $\max_{1\le j\le d} \mathrm{Var}(X_{i,j}) \le \sigma_i^2$ for all i. Let $\Gamma := \sum_{i=1}^n \sigma_i^2$. Then for $d \ge 3$ and any $L > 0$,*

$$\sqrt{\mathbb{E}\,\|S_n\|_\infty^2} \ \le\ \kappa L \log(2d) + \frac{\Gamma L\, \mathrm{e}(L)}{\kappa}.$$

Theorem 4.3 *In case of $(\mathbb{B}, \|\cdot\|) = \ell_\infty^d$ for some $d \ge 3$, inequality (1.1) holds with*

$$K \ = \ K_{TB}(d,\infty) := \bigl(1 + 3.46\sqrt{\log(2d)}\bigr)^2.$$

If the random vectors X_i are symmetrically distributed around 0, one may even set

$$K \ = \ \bigl(1 + 2.9\sqrt{\log(2d)}\bigr)^2.$$

5 Comparisons

In this section we compare the three approaches just described in the special case of $(\mathbb{B}, \|\cdot\|) = \ell_\infty^d$ with $d \ge 3$. As to the random vectors X_i, we broaden our point of view and consider three different cases:

General case: The random vectors X_i are independent with $\mathbb{E}\,\|X_i\|_\infty^2 < \infty$ for all i.

Centered case: In addition, $\mathbb{E}\,X_i = 0$ for all i.

Symmetric case: In addition, $\mathcal{L}(X_i) = \mathcal{L}(-X_i)$ for all i.

In view of the general case, we reformulate inequality (1.1) as follows:

$$\mathbb{E}\,\|S_n - \mathbb{E}\,S_n\|_\infty^2 \ \le\ K \sum_{i=1}^n \mathbb{E}\,\|X_i\|_\infty^2. \tag{5.1}$$

One reason for this extension is that in some applications, particularly in connection with empirical processes, it is easier and more natural to work with uncentered summands X_i. Let us discuss briefly the consequences of this extension in the three frameworks:

Nemirovski's approach: Between the centered and symmetric case there is no difference. If (1.1) holds in the centered case for some K, then in the general case

$$\mathbb{E}\,\|S_n - \mathbb{E}\,S_n\|_\infty^2 \ \le\ K \sum_{i=1}^n \mathbb{E}\,\|X_i - \mathbb{E}\,X_i\|_\infty^2 \ \le\ 4K \sum_{i=1}^n \mathbb{E}\,\|X_i\|_\infty^2.$$

The latter inequality follows from the general fact that

$$\mathbb{E}\,\|Y - \mathbb{E}\,Y\|^2 \ \le\ \mathbb{E}\bigl((\|Y\| + \|\,\mathbb{E}\,Y\|)^2\bigr) \ \le\ 2\,\mathbb{E}\,\|Y\|^2 + 2\|\,\mathbb{E}\,Y\|^2 \ \le\ 4\,\mathbb{E}\,\|Y\|^2.$$

This looks rather crude at first glance, but in case of the maximum norm and high dimension d, the factor 4 cannot be reduced. For let $Y \in \mathbb{R}^d$ have independent components $Y_1, \ldots, Y_d \in \{-1, 1\}$ with $\mathbb{P}(Y_j = 1) = 1 - \mathbb{P}(Y_j = -1) = p \in [1/2, 1)$. Then $\|Y\|_\infty \equiv 1$, while $\mathbb{E}\, Y = (2p-1)_{j=1}^d$ and

$$\|Y - \mathbb{E}\, Y\|_\infty \ = \ \begin{cases} 2(1-p) & \text{if } Y_1 = \cdots = Y_d = 1, \\ 2p & \text{else.} \end{cases}$$

Hence

$$\frac{\mathbb{E}\, \|Y - \mathbb{E}\, Y\|_\infty^2}{\mathbb{E}\, \|Y\|_\infty^2} \ = \ 4\bigl((1-p)^2 p^d + p^2(1-p^d)\bigr).$$

If we set $p = 1 - d^{-1/2}$ for $d \geq 4$, then the latter ratio converges to 4 as $d \to \infty$.

The approach via Rademacher and Gaussian type 2 inequalities: The first part of Proposition 3.1, involving the Rademacher type constant T_p, remains valid if we drop the assumption that $\mathbb{E}\, X_i = 0$ and replace S_n with $S_n - \mathbb{E}\, S_n$. Thus there is no difference between the general and the centered case. In the symmetric case, however, the factor 2^p in Proposition 3.1 becomes superfluous. Thus, if (1.1) holds with a certain constant K in the general and centered case, we may replace K with $K/4$ in the symmetric case.

The approach via truncation and Bernstein's inequality: Our proof for the centered case does not utilize that $\mathbb{E}\, X_i = 0$, so again there is no difference between the centered and general case. However, in the symmetric case, the truncated random vectors $1\{\|X_i\|_\infty \leq \kappa\} X_i$ and $1\{\|X_i\|_\infty > \kappa\} X_i$ are centered, too, which leads to the substantially smaller constant K in Theorem 4.3.

Summaries and comparisons. Table 1 summarizes the constants $K = K(d, \infty)$ we have found so far by the three different methods and for the three different cases. Table 2 contains the corresponding limits

$$K^* \ := \ \lim_{d \to \infty} \frac{K(d, \infty)}{\log d}.$$

Interestingly, there is no global winner among the three methods. But for the centered case, Nemirovski's approach yields asymptotically the smallest constants. In particular,

$$\begin{aligned}
\lim_{d \to \infty} \frac{K_{TB}(\infty, d)}{K_N(\infty, d)} \ &= \ \frac{3.46^2}{2e} \doteq 2.20205, \\
\lim_{d \to \infty} \frac{K_{T2}(\infty, d)}{K_N(\infty, d)} \ &= \ \frac{2\pi}{e} \doteq 2.31145, \\
\lim_{d \to \infty} \frac{K_{TB}(\infty, d)}{K_{T2}(\infty, d)} \ &= \ \frac{3.46^2}{4\pi} \doteq 0.95267.
\end{aligned}$$

Truncation and Bernstein's inequality yields a better constant than the type 2 inequalities for extremely large d. It is easily checked that the former wins for dimensions larger than approximately 9.40433×10^{71}! Figure 1 shows the constants $K(d,\infty)$ for the centered case over a certain range of dimensions d.

	General case	Centered case	Symmetric case
Nemirovski	$8e\log d - 4e$	$2e\log d - e$	$2e\log d - e$
Type 2 inequalities	$2\pi c_d^2 \le 4\pi\log d$	$2\pi c_d^2 \le 4\pi\log d$	$\pi\log d$
Truncation/Bernstein	$\left(1+3.46\sqrt{\log(2d)}\right)^2$	$\left(1+3.46\sqrt{\log(2d)}\right)^2$	$\left(1+2.9\sqrt{\log(2d)}\right)^2$

Table 1: The different constants $K(d,\infty)$.

	General case	Centered case	Symmetric case
Nemirovski	$8e \doteq 21.7463$	$2e \doteq \mathbf{5.4366}$	$2e \doteq 5.4366$
Type 2 inequalities	$4\pi \doteq 12.5664$	$4\pi \doteq 12.5664$	$\pi \doteq \mathbf{3.1416}$
Truncation/Bernstein	$3.46^2 = \mathbf{11.9716}$	$3.46^2 = 11.9716$	$2.9^2 = 8.41$

Table 2: The different limits K^*.

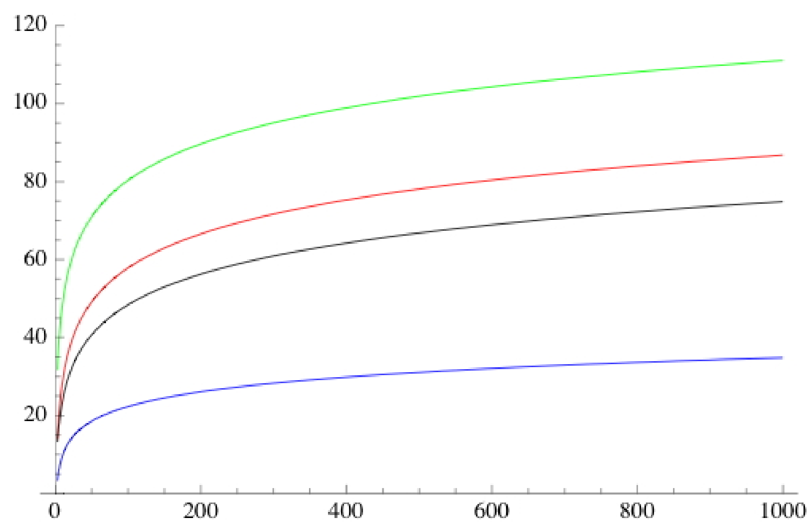

Figure 1: Comparison of $K(d,\infty)$ obtained via the three proof methods: Blue (bottom) = Nemirovski; Red and Black (middle) = type 2 inequalities; Green (top) = truncation and Bernstein inequality

6 Proofs

6.1 Proofs for Section 2

Proof of (2.1). In case of $r=\infty$, the asserted inequalities read

$$\|x\|_\infty \ \le\ \|x\|_q \ \le\ d^{1/q}\|x\|_\infty \quad \text{for } 1 \le q < \infty$$

and are rather obvious. For $1 \le q < r < \infty$, note first that

$$\begin{aligned} \|x\|_q^q &= \|x\|_r^q \sum_{i=1}^d \bigl(|x_i|/\|x\|_r\bigr)^q \\ &\ge \|x\|_r^q \sum_{i=1}^d \bigl(|x_i|/\|x\|_r\bigr)^r \\ &= \|x\|_r^q, \end{aligned}$$

because $|x_i|/\|x_i\|_r \le 1$. Moreover, it follows from Jensen's inequality that

$$\begin{aligned} \|x\|_q^r &= d^{r/q}\Bigl(\sum_{i=1}^d d^{-1}|x_i|^q\Bigr)^{r/q} \\ &\le d^{r/q}\sum_{i=1}^d d^{-1}\bigl(|x_i|^q\bigr)^{r/q} \\ &\le d^{r/q-1}\sum_{i=1}^d |x_i|^r \\ &= \bigl(d^{1/q-1/r}\|x\|_r\bigr)^r. \qquad \Box \end{aligned}$$

Proof of Lemma 2.6. In case of $r = 2$, $V(f+g)$ is equal to $V(f) + DV(f,g) + V(g)$. In case of $r \ge 2$ and $\|f\|_r = 0$, both $DV(f,g)$ and $\int h(f) g\, d\mu$ are equal to zero, and the asserted inequalities reduce to the trivial statement that $V(g) \le (r-1)V(g)$. Thus let us restrict our attention to the case $r > 2$ and $\|f\|_r > 0$.

Note first that the mapping

$$\mathbb{R} \ni t \ \mapsto\ h_t := |f + tg|^r$$

is pointwise twice continuously differentiable with derivatives

$$\begin{aligned} \dot{h}_t &= r|f+tg|^{r-1}\mathrm{sign}(f+tg)g \ = \ r|f+tg|^{r-2}(f+tg)g, \\ \ddot{h}_t &= r(r-1)|f+tg|^{r-2}g^2. \end{aligned}$$

By means of the inequality $|x+y|^b \le 2^{b-1}\bigl(|x|^b + |y|^b\bigr)$ for real numbers x, y and $b \ge 1$, a consequence of Jensen's inequality, we can conclude that for any bound $t_o > 0$,

$$\begin{aligned} \max_{|t|\le t_o} |\dot{h}_t| &\le r2^{r-2}\bigl(|f|^{r-1}|g| + t_o^{r-1}|g|^r\bigr), \\ \max_{|t|\le t_o} |\ddot{h}_t| &\le r(r-1)2^{r-3}\bigl(|f|^{r-2}|g|^2 + t_o^{r-2}|g|^r\bigr). \end{aligned}$$

The latter two envelope functions belong to $L_1(\mu)$. This follows from Hölder's inequality which we rephrase for our purposes in the form

$$\int |f|^{\lambda r}|g|^{(1-\lambda)r}\,d\mu \;\le\; \|f\|_r^{\lambda r}\|g\|_r^{(1-\lambda)r} \quad \text{for } 0 \le \lambda \le 1.$$

Hence we may conclude via dominated convergence that

$$t \;\mapsto\; \tilde{v}(t) := \|f + tg\|_r^r$$

is twice continuously differentiable with derivatives

$$\begin{aligned} \tilde{v}'(t) &= r\int |f+tg|^{r-2}(f+tg)g\,d\mu, \\ \tilde{v}''(t) &= r(r-1)\int |f+tg|^{r-2}g^2\,d\mu. \end{aligned}$$

This entails that

$$t \;\mapsto\; v(t) := V(f+tg) \;=\; \tilde{v}(t)^{2/r}$$

is continuously differentiable with derivative

$$v'(t) \;=\; (2/r)\tilde{v}(t)^{2/r-1}\tilde{v}'(t) \;=\; \int h(f+tg)g\,d\mu.$$

For $t = 0$ this entails the asserted expression for $DV(f,g)$. Moreover, $v(t)$ is twice continuously differentiable on the set $\{t \in \mathbb{R} : \|f+tg\|_r > 0\}$ which equals either $\mathbb{R}$ or $\mathbb{R} \setminus \{t_o\}$ for some $t_o \neq 0$. On this set the second derivative equals

$$\begin{aligned} v''(t) &= (2/r)\tilde{v}(t)^{2/r-1}\tilde{v}''(t) + (2/r)(2/r-1)\tilde{v}(t)^{2/r-2}\tilde{v}'(t)^2 \\ &= 2(r-1)\int \frac{|f+tg|^{r-2}}{\|f+tg\|_r^{r-2}}g^2\,d\mu - 2(r-2)\Big(\int \frac{|f+tg|^{r-2}(f+tg)}{\|f+tg\|_r^{r-1}}g\,d\mu\Big)^2 \\ &\le 2(r-1)\|g\|_r^2 = 2(r-1)V(g) \end{aligned}$$

by virtue of Hölder's inequality. Consequently,

$$\begin{aligned} V(f+g) - V(f) - DV(f,g) &= v(1) - v(0) - v'(0) \\ &= \int_0^1 (v'(t) - v'(0))\,dt \\ &\le 2(r-1)V(g)\int_0^1 t\,dt \\ &= (r-1)V(g). \qquad \Box \end{aligned}$$

Proof of Theorem 2.3. The first part is an immediate consequence of the considerations preceding the theorem. It remains to prove the (in)equalities and expansion for $K_N(d,r)$. Note that $K_N(d,r)$ is the infimum of $h(q)d^{-2/r}$ over all real $q \in [2,r]$, where $h(q) := (q-1)d^{2/q}$ satisfies the equation

$$h'(q) \;=\; \frac{d^{2/q}}{q^2}\big((q-\log d)^2 - (\log d - 2)\log d\big).$$

Since $7 < e^2 < 8$, this shows that h is strictly increasing on $[2,\infty)$ if $d \le 7$. Hence

$$K_N(d,r) \;=\; h(2)d^{-2/r} \;=\; d^{1-2/r} \quad \text{if } d \le 7.$$

For $d \ge 8$, one can easily show that $\log d - \sqrt{(\log d - 2)\log d} < 2$, so that h is strictly decreasing on $[2, r_d]$ and strictly increasing on $[r_d, \infty)$, where

$$r_d \;:=\; \log d + \sqrt{(\log d - 2)\log d} \;\begin{cases} <\; 2\log d, \\ >\; 2\log d - 2. \end{cases}$$

Thus for $d \ge 8$,

$$K_N(d,r) \;=\; \begin{cases} h(r)d^{-2/r} \;=\; r-1 \;<\; 2\log d - 1 & \text{if } r \le r_d, \\ h(r_d)d^{-2/r} \;\le\; h(2\log d) \;=\; 2e\log d - e & \text{if } r \ge r_d. \end{cases}$$

Moreover, one can verify numerically that $K_N(d,r) \le d \le 2e\log d - e$ for $3 \le d \le 7$.

Finally, for $d \ge 8$, the inequalities $r'_d := 2\log d - 2 < r_d < r''_d := 2\log d$ yield

$$K_N(d,\infty) = h(r_d) \;\ge\; (r'_d - 1)d^{2/r''_d} \;=\; 2e\log d - 3e,$$

and for $1 \le d \le 7$, the inequality $d = K_N(d,\infty) \ge 2e\log(d) - 3e$ is easily verified. □

6.2 Proofs for Section 3

Proof of Lemma 3.2. Let $x_1, \ldots, x_n$ be fixed functions in $L_r(\mu)$. Then by Haagerup (1981), for any $t \in T$,

$$\Big\{\mathbb{E}\Big|\sum_{i=1}^n \epsilon_i x_i(t)\Big|^r\Big\}^{1/r} \;\le\; B_r\Big(\sum_{i=1}^n |x_i(t)|^2\Big)^{1/2}. \tag{6.1}$$

To use inequality (6.1) for finding an upper bound for the type constant for L_r, rewrite it as

$$\mathbb{E}\Big|\sum_{i=1}^n \epsilon_i x_i(t)\Big|^r \;\le\; B_r^r\Big(\sum_{i=1}^n |x_i(t)|^2\Big)^{r/2}.$$

It follows from Fubini's theorem and the previous inequality that

$$\begin{aligned}
\mathbb{E}\Big\|\sum_{i=1}^n \epsilon_i x_i\Big\|_r^r &= \mathbb{E}\int\Big|\sum_{i=1}^n \epsilon_i x_i(t)\Big|^r d\mu(t) \\
&= \int \mathbb{E}\Big|\sum_{i=1}^n \epsilon_i x_i(t)|^r d\mu(t) \\
&\le B_r^r \int\Big(\sum_{i=1}^n |x_i(t)|^2\Big)^{r/2} d\mu(t).
\end{aligned}$$

Using the triangle inequality (or Minkowski's inequality), we obtain

$$\begin{aligned}
\Big\{\mathbb{E}\Big\|\sum_{i=1}^n \epsilon_i x_i\Big\|_r^r\Big\}^{2/r} &\le B_r^2\Big\{\int\Big(\sum_{i=1}^n |x_i(t)|^2\Big)^{r/2} d\mu(t)\Big\}^{2/r} \\
&\le B_r^2 \sum_{i=1}^n \Big(\int |x_i(t)|^r\, d\mu(t)\Big)^{2/r} \\
&= B_r^2 \sum_{i=1}^n \|x_i\|_r^2.
\end{aligned}$$

Furthermore, since $g(v) = v^{2/r}$ is a concave function of $v \ge 0$, the last display implies that

$$\mathbb{E}\Big\|\sum_{i=1}^n \epsilon_i x_i\Big\|_r^2 \;\le\; \Big\{\mathbb{E}\Big\|\sum_{i=1}^n \epsilon_i x_i\Big\|_r^r\Big\}^{2/r} \;\le\; B_r^2 \sum_{i=1}^n \|x_i\|_r^2. \qquad \Box$$

Proof of (3.2). Let $x_i \in \ell_\infty^d$ for $i = 1, \ldots, n$. Then $V := \sum_{i=1}^n Z_i x_i$ has distribution $N_d(0, \Sigma)$ with covariance matrix

$$\Sigma \;=\; \sum_{i=1}^n x_i x_i'.$$

Thus we want to show that

$$\mathbb{E}\max_{1\le j\le d} V_j^2 \;\le\; c_d^2 \sum_{i=1}^n \|x_i\|_\infty^2$$

with $c_d^2 = \mathbb{E}\max_{1\le j\le d} Z_j^2$. We may assume without loss of generality that $\sum_{i=1}^n \|x_i\|_\infty^2 = 1$. For otherwise we could replace x_i with $x_i/\sqrt{\sum_{i=1}^n \|x_i\|_\infty^2}$. Then it remains to be shown that

$$\mathbb{E}\max_{1\le j\le d} V_j^2 \;=\; \mathbb{E}\Big\|\sum_{i=1}^n Z_i x_i\Big\|_\infty^2 \;\le\; c_d^2. \tag{6.2}$$

Now we have

$$\mathrm{Var}(V_j) \;=\; \sum_{i=1}^n x_{ij}^2 \;\le\; \sum_{i=1}^n \|x_i\|_\infty^2 \;=\; 1. \tag{6.3}$$

By (Šidák, 1968, Corollary 3, page 1428) and (6.3) it follows that

$$\begin{aligned} \mathbb{P}\Big(\max_{1\le j\le d}|V_j|<t\Big) &= \mathbb{P}\big(|V_1|\le t,\ldots,|V_d|<t\big) \\ \ge \prod_{j=1}^{d} P(|V_j|<t) &\ge \prod_{j=1}^{d} P(|Z_j|<t) = \mathbb{P}\Big(\max_{1\le j\le d}|Z_j|<t\Big). \end{aligned}$$

This implies that (6.2) holds and hence the conclusion, since

$$\begin{aligned} \mathbb{E}\max_{1\le j\le d}|V_j|^2 &= 2\int_0^\infty tP(\max_{1\le j\le d}|V_j|\ge t)\,dt \\ &\le 2\int_0^\infty tP(\max_{1\le j\le d}|Z_j|\ge t)\,dt = \mathbb{E}\max_{1\le j\le d}|Z_j|^2. \end{aligned}$$

To prove the inequality in (3.2) we will use the upper bound of Exercise 2.3.5 of van der Vaart and Wellner (1996), which holds, in fact, for every $t_o>0$. Thus for every $t_o>0$

$$\begin{aligned} c_d^2\equiv\mathbb{E}\max_{1\le j\le d}|Z_j|^2 &\le t_o^2+d\int_{t_o}^\infty P(Z>t)dt^2 \\ &= t_o^2+2d\int_{t_o}^\infty t(1-\Phi(t))dt \\ &\le t_o^2+2d\int_{t_o}^\infty \phi(t)dt \qquad \text{(by Mills' ratio)} \\ &= t_o^2+2d(1-\Phi(t_o)). \end{aligned}$$

Evaluating this bound at $t_o=\sqrt{2\log(d/\sqrt{2\pi})}$ yields

$$\begin{aligned} c_d^2 &\le 2\log(d/\sqrt{2\pi})+2d(1-\Phi(\sqrt{2\log(d/\sqrt{2\pi})})) \\ &\le 2\log d-2\frac{1}{2}\log(2\pi)+2d\frac{\phi(\sqrt{2\log(d/\sqrt{2\pi})})}{\sqrt{2\log(d/\sqrt{2\pi})}} \\ &= 2\log d-\log(2\pi)+\frac{\sqrt{2}}{\sqrt{\log(d/\sqrt{2\pi})}} \qquad (6.4) \\ &\le 2\log d \end{aligned}$$

where the last inequality holds if

$$\frac{\sqrt{2}}{\sqrt{\log(d/\sqrt{2\pi})}}\le\log(2\pi),$$

or equivalently if

$$\log d \ \ge\ \frac{2}{(\log(2\pi))^2}+\frac{\log(2\pi)}{2} \ =\ 1.51104...,$$

and hence if $d\ge 5>e^{1.51104...}\doteq 4.53$. The claimed inequality is easily verified numerically for $d=3,4$. (It fails for $d=2$.) As can be seen from (6.4), $2\log d-\log(2\pi)$ gives a good approximation to $\mathbb{E}\max_{1\le j\le d}Z_j^2$ for large d.

Proof of (3.1). Let $x_i \in \mathbb{B}$ and let $\{\epsilon_i\}$ and $\{Z_i\}$ be sequences of independent Rademacher and $N(0,1)$ random variables which are themselves independent. Then, since $\mathbb{E}\,|Z_i| = \sqrt{2/\pi}$ for all i, Jensen's inequality yields

$$\begin{aligned}
\mathbb{E}\Big\|\sum_{i=1}^n \epsilon_i x_i\Big\|^2 &= \mathbb{E}_\epsilon\Big\|\sum_{i=1}^n \epsilon_i \frac{\mathbb{E}\,|Z_i|}{\sqrt{2/\pi}} x_i\Big\|^2 \\
&= \frac{\pi}{2}\,\mathbb{E}_\epsilon\,\mathbb{E}_Z\Big\|\sum_{i=1}^n \epsilon_i Z_i x_i\Big\|^2 \\
&= \frac{\pi}{2}\,\mathbb{E}\Big\|\sum_{i=1}^n Z_i x_i\Big\|^2 \qquad \big(\text{since } \epsilon_i Z_i \overset{d}{=} Z_i\big) \\
&\leq \frac{\pi}{2} T_2^G(\mathbb{B})^2 \sum_{i=1}^n \|x_i\|^2.
\end{aligned}$$

This implies that $T_2^R(\mathbb{B}) \leq \sqrt{\pi/2}\, T_2^G(\mathbb{B})$. To prove the left inequality in (3.1), note that

$$\begin{aligned}
\mathbb{E}\Big\|\sum_{i=1}^n Z_i x_i\Big\|^2 &= \mathbb{E}\Big\|\sum_{i=1}^n \epsilon_i Z_i x_i\Big\|^2 = \mathbb{E}_Z\,\mathbb{E}_\epsilon\Big\|\sum_{i=1}^n \epsilon_i (Z_i x_i)\Big\|^2 \\
&\leq \mathbb{E}_Z\Big\{T_2^R(\mathbb{B})^2 \sum_{i=1}^n Z_i^2 \|x_i\|^2\Big\} = T_2^R(\mathbb{B})^2 \sum_{i=1}^n \|x_i\|^2,
\end{aligned}$$

whence $T_2^G(B) \leq T_2^G(B)$. □

6.3 Proofs for Section 4

Proof of Lemma 4.1. It follows from $\mathbb{E}\,Z = 0$, the Taylor expansion of the exponential function and the inequality $\mathbb{E}\,|Z|^m \leq \sigma^2\kappa^{m-2}$ for $m \geq 2$ that

$$\begin{aligned}
\mathbb{E}\exp\Big(\frac{Z}{\kappa L}\Big) &= 1 + \mathbb{E}\Big\{\exp\Big(\frac{Z}{\kappa L}\Big) - 1 - \frac{Z}{\kappa L}\Big\} \\
&\leq 1 + \sum_{m=2}^\infty \frac{1}{m!}\frac{\mathbb{E}\,|Z|^m}{(\kappa L)^m} \leq 1 + \frac{\sigma^2}{\kappa^2}\sum_{m=2}^\infty \frac{1}{m!}\frac{1}{L^m} = 1 + \frac{\sigma^2 \mathrm{e}(L)}{\kappa^2}.
\end{aligned}$$

□

Proof of Lemma 4.2. Applying Lemma 4.1 to the j-th components $X_{i,j}$ of X_i and $S_{n,j}$ of S_n yields for all $L > 0$,

$$\begin{aligned}
\log\,\mathbb{E}\exp\Big(\frac{\pm S_{n,j}}{\kappa L}\Big) &= \sum_{i=1}^n \log\mathbb{E}\exp\Big(\frac{\pm X_{i,j}}{\kappa L}\Big) \\
&\leq \sum_{i=1}^n \frac{\mathrm{Var}(X_{i,j})\mathrm{e}(L)}{\kappa^2} \leq \sum_{i=1}^n \frac{\sigma_i^2 \mathrm{e}(L)}{\kappa^2} = \frac{\Gamma \mathrm{e}(L)}{\kappa^2}.
\end{aligned}$$

Equivalently

$$\mathbb{E}\exp\Bigl(\frac{\pm S_{n,j}}{\kappa L}\Bigr) \quad\le\quad \exp\Bigl(\frac{\Gamma \mathrm{e}(L)}{\kappa^2}\Bigr).$$

Hence

$$\begin{aligned}
\mathbb{E}\exp\Bigl(\frac{\|S_n\|_\infty}{\kappa L}\Bigr)-1 &= \mathbb{E}\max_{1\le j\le d}\Bigl(\exp\Bigl(\frac{|S_{n,j}|}{\kappa L}\Bigr)-1\Bigr)\\
&\le \sum_{j=1}^d \mathbb{E}\Bigl(\exp\Bigl(\frac{|S_{n,j}|}{\kappa L}\Bigr)-1\Bigr)\\
&\le \sum_{j=1}^d \mathbb{E}\Bigl(\exp\Bigl(\frac{S_{n,j}}{\kappa L}\Bigr)+\exp\Bigl(\frac{-S_{n,j}}{\kappa L}\Bigr)-1\Bigr)\\
&\le 2d\exp\Bigl(\frac{\Gamma \mathrm{e}(L)}{\kappa^2}\Bigr)-d.
\end{aligned}$$

Since $z \mapsto h(z) := \log^2(z)$ is increasing on $[1,\infty)$ and concave on $[e,\infty)$, it follows from Jensen's inequality that $Y := \|S_n\|_\infty/(\kappa L)$ satisfies

$$\begin{aligned}
\mathbb{E}\,Y^2 = \mathbb{E}\,h(\exp(Y)) &\le \mathbb{E}\,h\bigl(e+\exp(Y)-1\bigr)\\
&\le h\bigl(e+\mathbb{E}\exp(Y)-1\bigr)\\
&\le \log^2\Bigl\{e+2d\exp\Bigl(\frac{\Gamma \mathrm{e}(L)}{\kappa^2}\Bigr)-d\Bigr\}\\
&\le \Bigl\{\log(2d)+\frac{\Gamma \mathrm{e}(L)}{\kappa^2}\Bigr\}^2 \qquad \text{(because } d\ge e).
\end{aligned}$$

This entails that

$$\mathbb{E}\,\|S_n\|_\infty^2 \;\le\; \Bigl\{\kappa L\log(2d)+\frac{\Gamma L\,\mathrm{e}(L)}{\kappa}\Bigr\}^2,$$

which is equivalent to the inequality stated in the lemma. □

Proof of Theorem 4.3. For fixed $\kappa_o > 0$ we split S_n into $A_n + B_n$ as described before. Let us bound the sum B_n first: For this term we have

$$\begin{aligned}
\|B_n\|_\infty &\le \sum_{i=1}^n \bigl\{1_{[\|X_i\|_\infty>\kappa_o]}\|X_i\|_\infty + \mathbb{E}(1_{[\|X_i\|_\infty>\kappa_o]}\|X_i\|_\infty)\bigr\}\\
&= \sum_{i=1}^n \bigl\{1_{[\|X_i\|_\infty>\kappa_o]}\|X_i\|_\infty - \mathbb{E}(1_{[\|X_i\|_\infty>\kappa_o]}\|X_i\|_\infty)\bigr\}\\
&\qquad + 2\sum_{i=1}^n \mathbb{E}(1_{[\|X_i\|_\infty>\kappa_o]}\|X_i\|_\infty)\\
&=: B_{n1}+B_{n2}.
\end{aligned}$$

Therefore, since $\mathbb{E}\, B_{n1} = 0$,

$$\begin{aligned}
\mathbb{E}\,\|B_n\|_\infty^2 &\le \mathbb{E}(B_{n1} + B_{n2})^2 \;=\; \mathbb{E}\, B_{n1}^2 + B_{n2}^2 \\
&= \sum_{i=1}^n \mathrm{Var}\big(1_{[\|X_i\|_\infty > \kappa_o]}\|X_i\|_\infty\big) + 4\Big(\sum_{i=1}^n \mathbb{E}(\|X_i\|_\infty 1_{[\|X_i\|_\infty > \kappa_o]})\Big)^2 \\
&\le \sum_{i=1}^n \mathbb{E}\,\|X_i\|_\infty^2 + 4\Big(\sum_{i=1}^n \frac{\mathbb{E}\,\|X_i\|_\infty^2}{\kappa_o}\Big)^2 \\
&= \Gamma + 4\frac{\Gamma^2}{\kappa_o^2},
\end{aligned}$$

where we define $\Gamma := \sum_{i=1}^n \sigma_i^2$ with $\sigma_i^2 := \mathbb{E}\,\|X_i\|_\infty^2$.

The first sum, A_n, may be bounded by means of Lemma 4.2 with $\kappa = 2\kappa_o$, utilizing the bound

$$\mathrm{Var}(X_{i,j}^{(a)}) \;=\; \mathrm{Var}\big(1_{[\|X_i\|_\infty \le \kappa_o]} X_{i,j}\big) \;\le\; \mathbb{E}\big(1_{[\|X_i\|_\infty \le \kappa_o]} X_{i,j}^2\big) \;\le\; \sigma_i^2.$$

Thus

$$\mathbb{E}\,\|A_n\|_\infty^2 \;\le\; \Big\{2\kappa_o L\log(2d) + \frac{\Gamma L\, \mathrm{e}(L)}{2\kappa_o}\Big\}^2.$$

Combining the bounds we find that

$$\begin{aligned}
\sqrt{\mathbb{E}\,\|S_n\|_\infty^2} &\le \sqrt{\mathbb{E}\,\|A_n\|_\infty^2} + \sqrt{\mathbb{E}\,\|B_n\|_\infty^2} \\
&\le 2\kappa_o L\log(2d) + \frac{\Gamma L \mathrm{e}(L)}{2\kappa_o} + \sqrt{\Gamma} + 2\frac{\Gamma}{\kappa_o} \\
&= \alpha\kappa_o + \frac{\beta}{\kappa_o} + \sqrt{\Gamma},
\end{aligned}$$

where $\alpha := 2L\log(2d)$ and $\beta := \Gamma(L\,\mathrm{e}(L) + 4)/2$. This bound is minimized if $\kappa_o = \sqrt{\beta/\alpha}$ with minimum value

$$2\sqrt{\alpha\beta} + \sqrt{\Gamma} \;=\; \big(1 + 2\sqrt{L^2\mathrm{e}(L) + 4L}\sqrt{\log(2d)}\big)\sqrt{\Gamma},$$

and for $L = 0.407$ the latter bound is not greater than

$$\big(1 + 3.46\sqrt{\log(2d)}\big)\sqrt{\Gamma}.$$

In the special case of symmetrically distributed random vectors X_i, our treatment of the sum B_n does not change, but in the bound for $\mathbb{E}\,\|A_n\|_\infty^2$ one may replace $2\kappa_o$ with κ_o. Thus

$$\begin{aligned}
\sqrt{\mathbb{E}\,\|S_n\|_\infty^2} &\le \kappa_o L\log(2d) + \frac{\Gamma L \mathrm{e}(L)}{\kappa_o} + \sqrt{\Gamma} + 2\frac{\Gamma}{\kappa_o} \\
&= \alpha'\kappa_o + \frac{\beta'}{\kappa_o} + \sqrt{\Gamma} \qquad \big(\text{with } \alpha' := L\log(2d), \beta' := \Gamma(L\,\mathrm{e}(L) + 2)\big) \\
&= \big(1 + 2\sqrt{L^2\mathrm{e}(L) + 2L}\sqrt{\log(2d)}\big)\sqrt{\Gamma} \qquad \big(\text{if } \kappa_o = \sqrt{\beta'/\alpha'}\big).
\end{aligned}$$

For $L = 0.5$ the latter bound is not greater than

$$\big(1 + 2.9\sqrt{\log(2d)}\big)\sqrt{\Gamma}. \qquad \square$$

6.4 Proofs for Section 5

Notes on the crossings claimed at the end of section 5. First we show that

$$K_{TB}(\infty, d) = (1 + 3.46\sqrt{\log(2d)})^2 \ \leq \ 4\pi \log d =: K'_{T2}(\infty, d)$$

for $d \geq 9.40433 * 10^{71}$. Letting $y \equiv \sqrt{\log(2d)}$, $a = 3.46$, $b^2 = 4\pi$, $c = 4\pi \log(2)$, the inequality above becomes

$$(1 + ay)^2 \leq b^2 y^2 - c.$$

Thus we seek a solution of

$$(1 + ay)^2 = b^2 y^2 - c;$$

or, equivalently of

$$(b^2 - a^2) y^2 - 2ay - (c + 1) = 0.$$

This has positive solution

$$\begin{aligned} y_0 &= \frac{2a + \sqrt{4a^2 + 4(b^2 - a^2)(c+1)}}{2(b^2 - a^2)} \\ &= \frac{2(3.46) + \sqrt{4(3.46)^2 + 4(4\pi - (3.46)^2)(4\pi \log(2) + 1)}}{2(4\pi - (3.46)^2)} \\ &= 12.9003 \ldots. \end{aligned}$$

Thus the claimed inequality holds if $d \geq d_0 \equiv \exp(y_0^2)/2 \approx 9.40433 * 10^{71}$.

Now we show that

$$K_{TB}(\infty, d) = (1 + 3.46\sqrt{\log(2d)})^2 \ \leq \ 2\pi c_d^2 = K_{T2}(\infty, d) \tag{6.5}$$

for $d \geq 5.82884 \cdot 10^{86}$. From a lower bound similar to the upper bound in (6.4) we have

$$c_d^2 = \mathbb{E} \max_{1 \leq j \leq d} |Z_j|^2 \geq 2 \log d - \log(2\pi).$$

Thus to prove (6.5) it suffices to show that

$$K_{TB}(\infty, d) = (1 + 3.46\sqrt{\log(2d)})^2 \ \leq \ 2\pi(2 \log d - \log(2\pi)) \leq 2\pi c_d^2 =: K''_{TB}(\infty, d).$$

Letting $y \equiv \sqrt{\log(2d)}$, $a = 3.46$, $b^2 = 4\pi$, $c = 2\pi \log(8\pi)$, the inequality in the last display becomes

$$(1 + ay)^2 \leq b^2 y^2 - c,$$

which is exactly the same form as in the previous crossing argument with only a different value of the constant c. Thus we seek a solution of

$$(1 + ay)^2 = b^2 y^2 - c;$$

or, equivalently of

$$(b^2 - a^2)y^2 - 2ay - (c + 1) = 0.$$

This has positive solution

$$\begin{aligned} y_0 &= \frac{2a + \sqrt{4a^2 + 4(b^2 - a^2)(c + 1)}}{2(b^2 - a^2)} \\ &= \frac{2(3.46) + \sqrt{4(3.46)^2 + 4(4\pi - (3.46)^2)(2\pi \log(8\pi) + 1)}}{2(4\pi - (3.46)^2)} \\ &= 14.159 \ldots . \end{aligned}$$

Thus the claimed inequality holds if $d \geq d_0 \equiv \exp(y_0^2)/2 \approx 5.82884 \times 10^{86}$. $\square$

Acknowledgements. The authors are grateful to Ilya Molchanov for drawing their attention to Banach-Mazur distances, and to Stanislaw Kwapien and Vladimir Koltchinskii for pointers concerning type and co-type proofs and constants. This research was initiated during the opening week of the program on "Statistical Theory and Methods for Complex, High-Dimensional Data" held at the *Isaac Newton Institute for Mathematical Sciences* from 7 January to 27 June, 2008, and was made possible in part by the support of the Isaac Newton Institute for visits of various periods by all three authors. The research of Wellner was also supported in part by NSF grant DMS-0503822.

References

DE LA PEÑA, V. H. and GINÉ, E. (1999). *Decoupling*. Probability and its Applications (New York), Springer-Verlag, New York. From dependence to independence, Randomly stopped processes, U-statistics and processes, Martingales and beyond.

DUDLEY, R. M. (1999). *Uniform Central Limit Theorems*, vol. 63 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

GREENSHTEIN, E. and RITOV, Y. (2004). Persistence in high-dimensional linear predictor selection and the virtue of overparametrization. *Bernoulli* **10** 971–988.

HAAGERUP, U. (1981). The best constants in the Khintchine inequality. *Studia Math.* **70** 231–283.

JOHNSON, W. B. and LINDENSTRAUSS, J. (2001). Basic concepts in the geometry of Banach spaces. In *Handbook of the geometry of Banach spaces, Vol. I*. North-Holland, Amsterdam, 1–84.

JUDITSKY, A. and NEMIROVSKI, A. S. (2008). Large deviations of vector-valued martingales in 2-smooth normed spaces. Tech. rep., Georgia Institute of Technology.

LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces*, vol. 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin. Isoperimetry and processes.

NEMIROVSKI, A. S. (2000). Topics in non-parametric statistics. In *Lectures on Probability Theory and Statistics (Saint-Flour, 1998)*, vol. 1738 of *Lecture Notes in Mathematics*. Springer, Berlin, 85–277.

NEMIROVSKI, A. S. (2004). Regular Banach spaces and large deviations of random sums. Working paper.

NEMIROVSKI, A. S. and YUDIN, D. B. (1983). *Problem Complexity and Method Efficiency in Optimization*. John Wiley and Sons, Chichester.

PISIER, G. (1986). Probabilistic methods in the geometry of Banach spaces. In *Probability and analysis (Varenna, 1985)*, vol. 1206 of *Lecture Notes in Math.* Springer, Berlin, 167–241.

POLLARD, D. (1990). *Empirical processes: theory and applications*. NSF-CBMS Regional Conference Series in Probability and Statistics, 2, Institute of Mathematical Statistics, Hayward, CA.

RIO, E. (2008). Moment inequalities for sums of dependent random variables under projective conditions. *J. Theor. Probab.* **21** to appear.

ŠIDÁK, Z. (1968). On multivariate normal probabilities of rectangles: Their dependence on correlations. *Ann. Math. Statist.* **39** 1425–1434.

TOMCZAK-JAEGERMANN, N. (1989). *Banach-Mazur distances and finite-dimensional operator ideals*, vol. 38 of *Pitman Monographs and Surveys in Pure and Applied Mathematics*. Longman Scientific & Technical, Harlow.

VAN DE GEER, S. A. (2000). *Applications of empirical process theory*, vol. 6 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics, Springer-Verlag, New York. With applications to statistics.