Nemirovski's Inequalities Revisited

Lutz Dümbgen, Sara A. van de Geer, Mark C. Veraar, and Jon A. Wellner

1. INTRODUCTION. Our starting point is the following well-known theorem from probability: Let $X_1, \ldots, X_n$ be independent random variables with finite second moments, and let $S_n = \sum_{i=1}^n X_i$. Then

$$\mathrm{Var}(S_n) = \sum_{i=1}^n \mathrm{Var}(X_i). \tag{1}$$

If we suppose that each X_i has mean zero, $\mathbb{E}X_i = 0$, then (1) becomes

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

This equality generalizes easily to vectors in a Hilbert space $\mathbb{H}$ with inner product $\langle \cdot, \cdot \rangle$: If the X_i's are independent with values in $\mathbb{H}$ such that $\mathbb{E}X_i = 0$ and $\mathbb{E}\|X_i\|^2 < \infty$, then $\|S_n\|^2 = \langle S_n, S_n \rangle = \sum_{i,j=1}^n \langle X_i, X_j \rangle$, and since $\mathbb{E}\langle X_i, X_j \rangle = 0$ for $i \neq j$ by independence,

$$\mathbb{E}\|S_n\|^2 = \sum_{i,j=1}^n \mathbb{E}\langle X_i, X_j \rangle = \sum_{i=1}^n \mathbb{E}\|X_i\|^2. \tag{3}$$

What happens if the X_i's take values in a (real) Banach space $(\mathbb{B}, \|\cdot\|)$? In such cases, in particular when the square of the norm $\|\cdot\|$ is not given by an inner product, we are aiming at inequalities of the following type: Let $X_1, X_2, \ldots, X_n$ be independent random vectors with values in $(\mathbb{B}, \|\cdot\|)$ with $\mathbb{E}X_i = 0$ and $\mathbb{E}\|X_i\|^2 < \infty$. With $S_n := \sum_{i=1}^n X_i$ we want to show that

$$\mathbb{E}\|S_n\|^2 \le K \sum_{i=1}^n \mathbb{E}\|X_i\|^2 \tag{4}$$

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

For statistical applications, the case $(\mathbb{B}, \|\cdot\|) = \ell_r^d := (\mathbb{R}^d, \|\cdot\|_r)$ 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

$$\|x\|_r := \begin{cases} \left(\sum_{j=1}^d |x_j|^r \right)^{1/r} & \text{if } 1 \le r < \infty, \\ \max_{1 \le j \le d} |x_j| & \text{if } r = \infty. \end{cases} \tag{5}$$

An obvious question is how the exponent r and the dimension d enter an inequality of type (4). The influence of the dimension d is crucial, since current statistical research

doi:10.4169/000298910X476059

often involves small or moderate "sample size" n (the number of independent units), 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. The following two examples for the random vectors X_i provide lower bounds for the constant K in (4):

Example 1.1 (A lower bound in ℓ_r^d). Let $b_1, b_2, \ldots, b_d$ denote the standard basis of $\mathbb{R}^d$, and let $\epsilon_1, \epsilon_2, \ldots, \epsilon_d$ be independent *Rademacher variables*, i.e., random variables taking the values $+1$ and -1 each with probability $1/2$. Define $X_i := \epsilon_i b_i$ for $1 \le i \le n := d$. Then $\mathbb{E}X_i = 0$, $\|X_i\|_r^2 = 1$, and $\|S_n\|_r^2 = d^{2/r} = d^{2/r-1}\sum_{i=1}^n \|X_i\|_r^2$. Thus any candidate for K in (4) has to satisfy $K \ge d^{2/r-1}$.

Example 1.2 (A lower bound in ℓ_∞^d). Let $X_1, X_2, X_3, \ldots$ be independent random vectors, each uniformly distributed on $\{-1, 1\}^d$. Then $\mathbb{E}X_i = 0$ and $\|X_i\|_\infty = 1$. On the other hand, according to the Central Limit Theorem, $n^{-1/2}S_n$ converges in distribution as $n \to \infty$ to a random vector $Z = (Z_j)_{j=1}^d$ with independent, standard Gaussian components, $Z_j \sim N(0, 1)$. Hence

$$\sup_{n\ge1} \frac{\mathbb{E}\|S_n\|_\infty^2}{\sum_{i=1}^n \mathbb{E}\|X_i\|_\infty^2} = \sup_{n\ge1} \mathbb{E}\|n^{-1/2}S_n\|_\infty^2 \ge \mathbb{E}\|Z\|_\infty^2 = \mathbb{E} \max_{1\le j\le d} Z_j^2.$$

But it is well known that $\max_{1\le j\le d}|Z_j| - \sqrt{2\log d} \to_p 0$ as $d \to \infty$. Thus candidates $K(d)$ for the constant in (4) have to satisfy

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

At least three different methods have been developed to prove inequalities of the form given by (4). The three approaches known to us are:

(a) deterministic inequalities for norms;
(b) probabilistic methods for Banach spaces;
(c) empirical process methods.

Approach (a) was used by Nemirovski [**14**] to show that in the space ℓ_r^d with $d \ge 2$, inequality (4) holds with $K = C\min(r, \log(d))$ for some universal (but unspecified) constant C. In view of Example 1.2, this constant has the correct order of magnitude if $r = \infty$. For statistical applications see Greenshtein and Ritov [**7**]. Approach (b) uses special moment inequalities from probability theory on Banach spaces which involve nonrandom vectors in $\mathbb{B}$ and Rademacher variables as introduced in Example 1.1. Empirical process theory (approach (c)) in general deals with sums of independent random elements in infinite-dimensional Banach spaces. By means of chaining arguments, metric entropies, and approximation arguments, "maximal inequalities" for such random sums are built from basic inequalities for sums of independent random variables or finite-dimensional random vectors, in particular, "exponential inequalities"; see, e.g., Dudley [**4**], van der Vaart and Wellner [**26**], Pollard [**21**], de la Pena and Giné [**3**], or van de Geer [**25**].

Our main goal in this paper is to compare the inequalities resulting from these different approaches and to refine or improve the constants K obtainable by each method. 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 [**14**]. Our exposition includes explicit and improved constants. While finishing the present paper we became aware of yet unpublished work of Nemirovski [**15**] and Juditsky and

Nemirovski [**12**] who also improved some inequalities of [**14**]. Rio [**22**] uses similar methods in a different context. In Section 3 we present inequalities of type (4) which follow from type and cotype inequalities developed in probability theory on Banach spaces. In addition, we provide and utilize a new type inequality for the normed space ℓ_∞^d. To do so we utilize, among other tools, exponential inequalities of Hoeffding [**9**] and Pinelis [**17**]. In Section 4 we follow approach (c) and treat ℓ_∞^d by means of a truncation argument and Bernstein's exponential 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. Most proofs are deferred to Section 6.

2. NEMIROVSKI'S APPROACH: DETERMINISTIC INEQUALITIES FOR NORMS. In this section we review and refine inequalities of type (4) based on deterministic inequalities for norms. The considerations for $(\mathbb{B}, \|\cdot\|) = \ell_r^d$ follow closely the arguments of [**14**].

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 (5). For $x \in \mathbb{R}^d$ we think of x as a column vector and write $x^\top$ for the corresponding row vector. Thus $xx^\top$ is a $d \times d$ matrix with entries $x_i x_j$ for $i, j \in \{1, \ldots, d\}$.

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

$$\|x\|_r \le \|x\|_q \le d^{1/q-1/r}\|x\|_r \quad \text{for } 1 \le q < r \le \infty. \tag{6}$$

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 (4) holds with

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

Example 1.1 shows that this constant $\widetilde{K}(d, r)$ is indeed optimal for $1 \le r \le 2$.

A refinement for $r > 2$. In what follows we shall replace $\widetilde{K}(d, r) = d^{1-2/r}$ with substantially smaller constants. The main ingredient is the following result:

Lemma 2.1. *For arbitrary fixed $r \in [2, \infty)$ and $x \in \mathbb{R}^d \setminus \{0\}$ let*

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

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

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

[**16**] and [**14**] stated Lemma 2.1 with the factor $r - 1$ on the right side replaced with Cr for some (absolute) constant $C > 1$. Lemma 2.1, which is a special case of the more general Lemma 2.4 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 $2 \le r < \infty$,

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

and inductively we obtain a second candidate for K in (4):

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

Finally, we apply (6) 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 2.2. *For arbitrary $r \in [2, \infty]$,*

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

with

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

This constant $K_{\mathrm{Nem}}(d, r)$ satisfies the (in)equalities

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

and

$$K_{\mathrm{Nem}}(d, \infty) \ge 2e\log d - 3e.$$

Corollary 2.3. *In the case* $(\mathbb{B}, \|\cdot\|) = \ell_\infty^d$ *with* $d \geq 3$*, inequality* (4) *holds with constant* $K = 2e\log d - e$*. If the* X_i*'s are also identically distributed, then*

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

Note that

$$\lim_{d\to\infty} \frac{K_{\mathrm{Nem}}(d,\infty)}{2\log d} = \lim_{d\to\infty} \frac{2e\log d - e}{2\log d} = e.$$

Thus Example 1.2 entails that for large dimension d, the constants $K_{\mathrm{Nem}}(d,\infty)$ and $2e\log d - e$ are optimal up to a factor close to $e \doteq 2.7183$.

2.2. Arbitrary L_r-spaces. Lemma 2.1 is a special case of a more general inequality: Let (T, Σ, μ) be a σ-finite measure space, and for $1 \leq r < \infty$ let $L_r(\mu)$ be the set of all measurable functions $f : T \to \mathbb{R}$ with finite 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 $(\mathbb{R}^d, \|\cdot\|_r)$ corresponds to $(L_r(\mu), \|\cdot\|_r)$ if we take $T = \{1, 2, \ldots, d\}$ equipped with counting measure μ.

Note that $V(\cdot)$ is convex; thus for fixed $f, g \in L_r(\mu)$, the function

$$v(t) := V(f + tg) = \|f + tg\|_r^2, \qquad t \in \mathbb{R}$$

is convex with derivative

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

By convexity of v, the directional derivative $DV(f, g) := v'(0)$ satisfies

$$DV(f, g) \leq v(1) - v(0) = V(f + g) - V(f).$$

This proves the lower bound in the following lemma. We will prove the upper bound in Section 6 by computation of v'' and application of Hölder's inequality.

Lemma 2.4. *Let* $r \geq 2$*. Then for arbitrary* $f, g \in L_r(\mu)$*,*

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

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

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

Remark 2.5. 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 2.4 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 2.6. If $r = 2$, Lemma 2.4 is well known and easily verified. Here the upper bound for $V(f+g)$ is even an equality, i.e.,

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

Remark 2.7. Lemma 2.4 improves on an inequality of [**16**]. After writing this paper we realized Lemma 2.4 is also proved by Pinelis [**18**]; see his (2.2) and Proposition 2.1, page 1680.

Lemma 2.4 leads directly to the following result:

Corollary 2.8. *In the case $\mathbb{B} = L_r(\mu)$ for $r \geq 2$, inequality* (4) *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

$$\|T\| := \sup\big\{\|Tx\| : x \in \mathbb{B}, \|x\| \leq 1\big\},$$
$$\|T^{-1}\| := \sup\big\{\|T^{-1}y\| : y \in \mathbb{H}, \|y\| \leq 1\big\}.$$

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

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

This leads to the following observation:

Corollary 2.9. *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* (4) *is satisfied with $K = D(\mathbb{B}, \mathbb{H})^2$.*

A famous result from geometrical functional analysis is John's theorem (see [**24**], [**11**]) for finite-dimensional normed spaces. It entails that $D(\mathbb{B}, \ell_2^{\dim(\mathbb{B})}) \leq \sqrt{\dim(\mathbb{B})}$. This entails the following fact:

Corollary 2.10. *For any normed space* $(\mathbb{B}, \|\cdot\|)$ *with finite dimension, inequality* (4) *is satisfied with* $K = \dim(\mathbb{B})$.

Note that Example 1.1 with $r = 1$ provides an example where the constant $K = \dim(\mathbb{B})$ is optimal.

3. THE PROBABILISTIC APPROACH: TYPE AND COTYPE 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 $\mathbb{B}$,

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

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 $\mathbb{B}$,

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

Ledoux and Talagrand [**13**, p. 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:

Proposition 3.1. [**13**, Proposition 9.11, p. 248]. *If* $\mathbb{B}$ *is of type* $p \ge 1$ *with constant* T_p*, then*

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

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

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

As shown in [**13**, p. 27], the Banach space $L_r(\mu)$ with $1 \le r < \infty$ (cf. Section 2.2) is of type $\min(r, 2)$. Similarly, $L_r(\mu)$ is cotype $\max(r, 2)$. If $r \ge 2 = p$, 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 [**8**].

Lemma 3.2. *For* $2 \le r < \infty$*, 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* (4) *is satisfied with* $K = 4B_r^2$.

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 of r, the conclusion of Corollary 3.3 is weaker than that of Corollary 2.8.

3.2. The Space ℓ_∞^d. The preceding results apply only to $r < \infty$, so the special space ℓ_∞^d requires different arguments. First we deduce a new type inequality based on Hoeffding's [**9**] exponential inequality: if $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are independent Rademacher random variables, $a_1, a_2, \ldots, a_n$ are real numbers, and $v^2 := \sum_{i=1}^n a_i^2$, then the tail probabilities of the random variable $\left|\sum_{i=1}^n a_i \epsilon_i\right|$ may be bounded as follows:

$$\mathbb{P}\left(\left|\sum_{i=1}^n a_i \epsilon_i\right| \geq z\right) \leq 2\exp\left(-\frac{z^2}{2v^2}\right), \quad z \geq 0. \tag{7}$$

At the heart of these tail bounds is the following exponential moment bound:

$$\mathbb{E}\exp\left(t\sum_{i=1}^n a_i \epsilon_i\right) \leq \exp(t^2 v^2/2), \quad t \in \mathbb{R}. \tag{8}$$

From the latter bound we shall deduce the following type inequality in Section 6:

Lemma 3.4. *The space* ℓ_∞^d *is of type* 2 *with constant* $\sqrt{2\log(2d)}$.

Using this upper bound together with Proposition 3.1 yields another Nemirovski-type inequality:

Corollary 3.5. *For* $(\mathbb{B}, \|\cdot\|) = \ell_\infty^d$, *inequality* (4) *is satisfied with* $K = K_{\text{Type2}}(d, \infty) = 8\log(2d)$.

Refinements. Let $T_2(\ell_\infty^d)$ be the optimal type-2 constant for the space ℓ_∞^d. So far we know that $T_2(\ell_\infty^d) \leq \sqrt{2\log(2d)}$. Moreover, by a modification of Example 1.2 one can show that

$$T_2(\ell_\infty^d) \geq c_d := \sqrt{\mathbb{E}\max_{1\leq j\leq d} Z_j^2}. \tag{9}$$

The constants c_d can be expressed or bounded in terms of the distribution function Φ of $N(0,1)$, i.e., $\Phi(z) = \int_{-\infty}^z \phi(x)\,dx$ with $\phi(x) = \exp(-x^2/2)/\sqrt{2\pi}$. Namely, with $W := \max_{1\leq j\leq d} |Z_j|$,

$$c_d^2 = \mathbb{E}(W^2) = \mathbb{E}\int_0^\infty 2t 1_{[t\leq W]}\,dt = \int_0^\infty 2t\mathbb{P}(W \geq t)\,dt,$$

and for any $t > 0$,

$$\mathbb{P}(W \geq t) \begin{cases} = 1 - \mathbb{P}(W < t) = 1 - \mathbb{P}(|Z_1| < t)^d = 1 - (2\Phi(t) - 1)^d, \\ \leq d\mathbb{P}(|Z_1| \geq t) = 2d(1 - \Phi(t)). \end{cases}$$

These considerations and various bounds for Φ will allow us to derive explicit bounds for c_d.

On the other hand, Hoeffding's inequality (7) has been refined by Pinelis [**17, 20**] as follows:

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} a_i \epsilon_i\right| \geq z\right) \leq 2K(1 - \Phi(z/v)), \quad z > 0, \tag{10}$$

where K satisfies $3.18 \leq K \leq 3.22$. This will be the main ingredient for refined upper bounds for $T_2(\ell_\infty^d)$. The next lemma summarizes our findings:

Lemma 3.6. *The constants c_d and $T_2(\ell_\infty^d)$ satisfy the following inequalities:*

$$\sqrt{2\log d + h_1(d)} \leq c_d \leq \begin{cases} T_2(\ell_\infty^d) \leq \sqrt{2\log d + h_2(d)}, & d \geq 1 \\ \sqrt{2\log d}, & d \geq 3 \\ \sqrt{2\log d + h_3(d)}, & d \geq 1 \end{cases} \tag{11}$$

where $h_2(d) \leq 3$, $h_2(d)$ becomes negative for $d > 4.13795 \times 10^{10}$, $h_3(d)$ becomes negative for $d \geq 14$, and $h_j(d) \sim -\log\log d$ as $d \to \infty$ for $j = 1, 2, 3$.

In particular, one could replace $K_{\text{Type2}}(d, \infty)$ in Corollary 3.5 with $8\log d + 4h_2(d)$.

4. THE EMPIRICAL PROCESS APPROACH: TRUNCATION AND BERNSTEIN'S INEQUALITY. An alternative to Hoeffding's exponential tail inequality (7) is a classical exponential bound due to Bernstein (see, e.g., [**2**]): Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables with mean zero such that $|Y_i| \leq \kappa$. Then for $v^2 = \sum_{i=1}^n \mathrm{Var}(Y_i)$,

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} Y_i\right| \geq x\right) \leq 2\exp\left(-\frac{x^2}{2(v^2 + \kappa x/3)}\right), \qquad x > 0. \tag{12}$$

We will not use this inequality itself but rather an exponential moment inequality underlying its proof:

Lemma 4.1. *For $L > 0$ define*

$$\mathrm{e}(L) := \exp(1/L) - 1 - 1/L.$$

Let Y be a random variable with mean zero and variance σ^2 such that $|Y| \leq \kappa$. Then for any $L > 0$,

$$\mathbb{E}\exp\Big(\frac{Y}{\kappa L}\Big) \leq 1 + \frac{\sigma^2\mathrm{e}(L)}{\kappa^2} \leq \exp\Big(\frac{\sigma^2\mathrm{e}(L)}{\kappa^2}\Big).$$

With the latter exponential moment bound we can prove a moment inequality for random vectors in $\mathbb{R}^d$ with bounded components:

Lemma 4.2. *Suppose that* $X_i = (X_{i,j})_{j=1}^d$ *satisfies* $\|X_i\|_\infty \le \kappa$, *and let* Γ *be an upper bound for* $\max_{1\le j\le d} \sum_{i=1}^n \mathrm{Var}(X_{i,j})$. *Then for any* $L > 0$,

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

Now we return to our general random vectors $X_i \in \mathbb{R}^d$ with mean zero and $\mathbb{E}\|X_i\|_\infty^2 < \infty$. They are split into two random vectors via truncation: $X_i = X_i^{(a)} + X_i^{(b)}$ 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

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

The sum A_n involves centered random vectors in $[-2\kappa_o, 2\kappa_o]^d$ and will be treated by means of Lemma 4.2, while B_n will be bounded with elementary methods. Choosing the threshold κ and the parameter L carefully yields the following theorem.

Theorem 4.3. *In the case* $(\mathbb{B}, \|\cdot\|) = \ell_\infty^d$ *for some* $d \ge 1$, *inequality* (4) *holds with*

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

If each of the random vectors X_i *is symmetrically distributed around* 0, *one may even set*

$$K = K_{\mathrm{TrBern}}^{(\mathrm{symm})}(d, \infty) = \bigl(1 + 2.9\sqrt{\log(2d)}\bigr)^2.$$

5. COMPARISONS. In this section we compare the three approaches just described for the space ℓ_∞^d. 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, X_i is symmetrically distributed around 0 for all i.

In view of the general case, we reformulate inequality (4) as follows:

$$\mathbb{E}\|S_n - \mathbb{E}S_n\|_\infty^2 \le K \sum_{i=1}^n \mathbb{E}\|X_i\|_\infty^2. \tag{13}$$

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 cases there is no difference. If (4) holds in the centered case for some K, then in the general case

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

The latter inequality follows from the general fact that

$$\mathbb{E}\|Y - \mathbb{E}Y\|^2 \leq \mathbb{E}\big((\|Y\| + \|\mathbb{E}Y\|)^2\big) \leq 2\mathbb{E}\|Y\|^2 + 2\|\mathbb{E}Y\|^2 \leq 4\mathbb{E}\|Y\|^2.$$

This looks rather crude at first glance, but in the 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

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

Hence

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

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

The approach via Rademacher 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 centered cases. In the symmetric case, however, the factor 2^p in Proposition 3.1 becomes superfluous. Thus, if (4) holds with a certain constant K in the general and centered cases, 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 cases. 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

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

	General case	Centered case	Symmetric case
Nemirovski	$8e\log d - 4e$	$2e\log d - e$	$2e\log d - e$
Type-2 inequalities	$8\log(2d)$ $8\log d + 4h_2(d)$	$8\log(2d)$ $8\log d + 4h_2(d)$	$2\log(2d)$ $2\log d + h_2(d)$
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 2. The different limits K^*.

	General case	Centered case	Symmetric case
Nemirovski	$8e \doteq 21.7463$	$2e \doteq \mathbf{5.4366}$	$2e \doteq 5.4366$
Type-2 inequalities	**8.0**	8.0	**2.0**
Truncation/Bernstein	$3.46^2 = 11.9716$	$3.46^2 = 11.9716$	$2.9^2 = 8.41$

case, Nemirovski's approach yields asymptotically the smallest constants. In particular,

$$\lim_{d\to\infty} \frac{K_{\text{TrBern}}(d, \infty)}{K_{\text{Nem}}(d, \infty)} = \frac{3.46^2}{2e} \doteq 2.20205,$$

$$\lim_{d\to\infty} \frac{K_{\text{Type2}}(d, \infty)}{K_{\text{Nem}}(d, \infty)} = \frac{4}{e} \doteq 1.47152,$$

$$\lim_{d\to\infty} \frac{K_{\text{TrBern}}(d, \infty)}{K_{\text{Type2}}(d, \infty)} = \frac{3.46^2}{8} \doteq 1.49645.$$

The conclusion at this point seems to be that Nemirovski's approach and the type 2 inequalities yield better constants than Bernstein's inequality and truncation. Figure 1 shows the constants $K(d, \infty)$ for the centered case over a certain range of dimensions d.

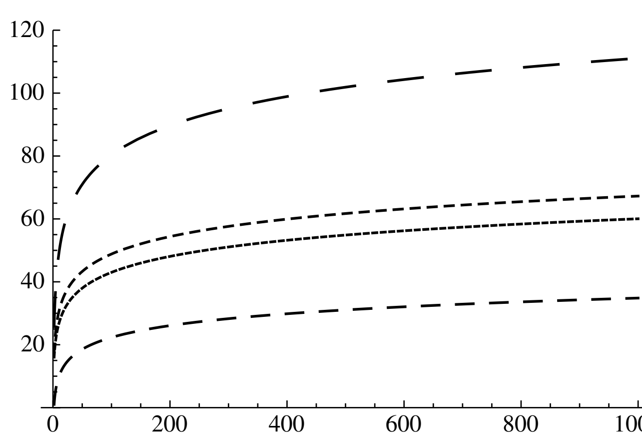

Figure 1. Comparison of $K(d, \infty)$ obtained via the three proof methods: Medium dashing (bottom) = Nemirovski; Small and tiny dashing (middle) = type 2 inequalities; Large dashing (top) = truncation and Bernstein inequality.

6. PROOFS.

6.1. Proofs for Section 2.

Proof of (6). In the case $r = \infty$, 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 $1 \le q < r < \infty$, (6) is an easy consequence of Hölder's inequality. ■

Proof of Lemma 2.4. In the case $r = 2$, $V(f+g)$ is equal to $V(f) + DV(f,g) + V(g)$. If $r \ge 2$ 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

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

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

$$\max_{|t|\le t_o} |\dot{h}_t| \le r2^{r-2}\big(|f|^{r-1}|g| + t_o^{r-1}|g|^r\big),$$
$$\max_{|t|\le t_o} |\ddot{h}_t| \le r(r-1)2^{r-3}\big(|f|^{r-2}|g|^2 + t_o^{r-2}|g|^r\big).$$

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

$$\int |f|^{(1-\lambda)r}|g|^{\lambda r}\,d\mu \le \|f\|_r^{(1-\lambda)r}\|g\|_r^{\lambda r} \quad \text{for } 0 \le \lambda \le 1. \tag{14}$$

Hence we may conclude via dominated convergence that

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

is twice continuously differentiable with derivatives

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

This entails that

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

is continuously differentiable with derivative

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

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

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

by virtue of Hölder's inequality (14) with $\lambda = 2/r$. Consequently, by using

$$v'(t) - v'(0) = \int_0^t v''(s)\,ds \leq 2(r-1)V(g)t,$$

we find that

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

■

Proof of Theorem 2.2. The first part is an immediate consequence of the considerations preceding the theorem. It remains to prove the (in)equalities and expansion for $K_{\text{Nem}}(d, r)$. Note that $K_{\text{Nem}}(d, r)$ 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 $d \leq 7$. Hence

$$K_{\text{Nem}}(d, r) = h(2)d^{-2/r} = d^{1-2/r} \quad \text{if } d \leq 7.$$

For $d \geq 8$, 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 $d \geq 8$,

$$K_{\text{Nem}}(d,r) = \begin{cases} h(r)d^{-2/r} = r - 1 < 2\log d - 1 & \text{if } r \leq r_d, \\ h(r_d)d^{-2/r} \leq h(2\log d) = 2e\log d - e & \text{if } r \geq r_d. \end{cases}$$

Moreover, one can verify numerically that $K_{\text{Nem}}(d,r) \leq d \leq 2e\log d - e$ for $3 \leq d \leq 7$.

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

$$K_{\text{Nem}}(d,\infty) = h(r_d) \geq (r_d' - 1)d^{2/r_d''} = 2e\log d - 3e,$$

and for $1 \leq d \leq 7$, the inequality $d = K_{\text{Nem}}(d,\infty) \geq 2e\log(d) - 3e$ is easily verified. ■

6.2. Proofs for Section 3.

Proof of Lemma 3.2. The following proof is standard; see, e.g., [**1**, p. 160], [**13**, p. 247]. Let $x_1, \ldots, x_n$ be fixed functions in $L_r(\mu)$. Then by [**8**], for any $t \in T$,

$$\left\{\mathbb{E}\left|\sum_{i=1}^{n} \epsilon_i x_i(t)\right|^r\right\}^{1/r} \leq B_r\left(\sum_{i=1}^{n} |x_i(t)|^2\right)^{1/2}. \tag{15}$$

To use inequality (15) for finding an upper bound for the type constant for L_r, rewrite it as

$$\mathbb{E}\left|\sum_{i=1}^{n} \epsilon_i x_i(t)\right|^r \leq B_r^r\left(\sum_{i=1}^{n} |x_i(t)|^2\right)^{r/2}.$$

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

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

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

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

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

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

Proof of Lemma 3.4. For $1 \leq i \leq n$ let $x_i = (x_{im})_{m=1}^d$ be an arbitrary fixed vector in $\mathbb{R}^d$, and set $S := \sum_{i=1}^n \epsilon_i x_i$. Further let S_m be the mth component of S with variance $v_m^2 := \sum_{i=1}^n x_{im}^2$, and define $v^2 := \max_{1\leq m\leq d} v_m^2$, which is not greater than $\sum_{i=1}^n \|x_i\|_\infty^2$. It suffices to show that

$$\mathbb{E}\|S\|_\infty^2 \leq 2\log(2d)v^2.$$

To this end note first that $h : [0,\infty) \to [1,\infty)$ with

$$h(t) := \cosh(t^{1/2}) = \sum_{k=0}^\infty \frac{t^k}{(2k)!}$$

is bijective, increasing, and convex. Hence its inverse function $h^{-1} : [1,\infty) \to [0,\infty)$ is increasing and concave, and one easily verifies that

$$h^{-1}(s) = \big(\log(s + (s^2-1)^{1/2})\big)^2 \leq (\log(2s))^2.$$

Thus it follows from Jensen's inequality that for arbitrary $t > 0$,

$$\begin{aligned}\mathbb{E}\|S\|_\infty^2 &= t^{-2}\mathbb{E}h^{-1}\big(\cosh(\|tS\|_\infty)\big) \leq t^{-2}h^{-1}\big(\mathbb{E}\cosh(\|tS\|_\infty)\big) \\ &\leq t^{-2}\big(\log\big(2\mathbb{E}\cosh(\|tS\|_\infty)\big)\big)^2.\end{aligned}$$

Moreover,

$$\mathbb{E}\cosh(\|tS\|_\infty) = \mathbb{E}\max_{1\leq m\leq d}\cosh(tS_m) \leq \sum_{m=1}^d \mathbb{E}\cosh(tS_m) \leq d\exp(t^2v^2/2),$$

according to (8), whence

$$\mathbb{E}\|S\|_\infty^2 \leq t^{-2}\log\big(2d\exp(t^2v^2/2)\big)^2 = \big(\log(2d)/t + tv^2/2\big)^2.$$

Now the assertion follows if we set $t = \sqrt{2\log(2d)/v^2}$. $\blacksquare$

Proof of (9). We may replace the random sequence $\{X_i\}$ in Example 1.2 with the random sequence $\{\epsilon_i X_i\}$, where $\{\epsilon_i\}$ is a Rademacher sequence independent of $\{X_i\}$. Thereafter we condition on $\{X_i\}$, i.e., we view it as a deterministic sequence such that $n^{-1}\sum_{i=1}^n X_i X_i^\top$ converges to the identity matrix I_d as $n \to \infty$, by the strong law of large numbers. Now Lindeberg's version of the multivariate Central Limit Theorem shows that

$$\sup_{n\geq 1}\frac{\mathbb{E}\big\|\sum_{i=1}^n \epsilon_i X_i\big\|_\infty^2}{\sum_{i=1}^n \|X_i\|_\infty^2} \geq \sup_{n\geq 1}\mathbb{E}\Big\|n^{-1/2}\sum_{i=1}^n \epsilon_i X_i\Big\|_\infty^2 \geq c_d^2. \qquad \blacksquare$$

Inequalities for Φ. The subsequent results will rely on (10) and several inequalities for $1-\Phi(z)$. The first of these is:

$$1-\Phi(z) \le z^{-1}\phi(z), \qquad z>0, \tag{16}$$

which is known as *Mills' ratio*; see [**6**] and [**19**] for related results. The proof of this upper bound is easy: since $\phi'(z) = -z\phi(z)$ it follows that

$$1-\Phi(z) = \int_z^\infty \phi(t)\,dt \le \int_z^\infty \frac{t}{z}\phi(t)\,dt = \frac{-1}{z}\int_z^\infty \phi'(t)\,dt = \frac{\phi(z)}{z}. \tag{17}$$

A very useful pair of upper and lower bounds for $1-\Phi(z)$ is as follows:

$$\frac{2}{z+\sqrt{z^2+4}}\phi(z) \le 1-\Phi(z) \le \frac{4}{3z+\sqrt{z^2+8}}\phi(z), \qquad z>-1; \tag{18}$$

the inequality on the left is due to Komatsu (see, e.g., [**10**, p. 17]), while the inequality on the right is an improvement of an earlier result of Komatsu due to Szarek and Werner [**23**].

Proof of Lemma 3.6. To prove the upper bound for $T_2(\ell_\infty^d)$, let $(\epsilon_i)_{i\ge1}$ be a Rademacher sequence. With S and S_m as in the proof of Lemma 3.4, for any $\delta>0$ we may write

$$\begin{aligned}
\mathbb{E}\|S\|_\infty^2 &= \int_0^\infty 2t\mathbb{P}\left(\sup_{1\le m\le d}|S_m|>t\right)dt \\
&\le \delta^2 + \int_\delta^\infty 2t\mathbb{P}\left(\sup_{1\le m\le d}|S_m|>t\right)dt \\
&\le \delta^2 + \sum_{m=1}^d \int_\delta^\infty 2t\mathbb{P}\big(|S_m|>t\big)\,dt.
\end{aligned}$$

Now by (10) with v^2 and v_m^2 as in the proof of Lemma 3.4, followed by Mills' ratio (16),

$$\begin{aligned}
\int_\delta^\infty 2t\mathbb{P}(|S_m|>t)\,dt &\le \int_\delta^\infty \frac{4Kv_m}{\sqrt{2\pi}t} t e^{-t^2/(2v_m^2)}\,dt \\
&= \frac{4Kv_m}{\sqrt{2\pi}}\int_\delta^\infty e^{-t^2/(2v_m^2)}\,dt = 4Kv_m^2\int_\delta^\infty \frac{e^{-t^2/(2v_m^2)}}{\sqrt{2\pi}v_m}\,dt \\
&= 4Kv_m^2(1-\Phi(\delta/v_m)) \le 4Kv^2(1-\Phi(\delta/v)).
\end{aligned} \tag{19}$$

Now instead of the Mills' ratio bound (16) for the tail of the normal distribution, we use the upper bound part of (18). This yields

$$\int_\delta^\infty 2t\mathbb{P}(|S_m|>t)\,dt \le 4Kv^2(1-\Phi(\delta/v)) \le \frac{4cv^2}{3\delta/v+\sqrt{\delta^2/v^2+8}}e^{-\delta^2/(2v^2)},$$

where we have defined $c := 4K/\sqrt{2\pi} = 12.88/\sqrt{2\pi}$, and hence

$$\mathbb{E}\|S\|^2 \le \delta^2 + \frac{4cdv^2}{3\delta/v+\sqrt{\delta^2/v^2+8}}e^{-\delta^2/(2v^2)}.$$

Taking

$$\delta^2 = v^2 2 \log\left(\frac{cd/2}{\sqrt{2\log(cd/2)}}\right)$$

gives

$$\begin{aligned}
\mathbb{E}\|S\|^2 \ \le\ v^2 & \Bigg\{ 2\log d + 2\log(c/2) - \log(2\log(dc/2)) \\
& \qquad + \frac{8\sqrt{2\log(cd/2)}}{3\sqrt{2\log\left(\frac{cd}{2\sqrt{2\log(cd/2)}}\right)} + \sqrt{2\log\left(\frac{cd}{2\sqrt{2\log(cd/2)}}\right) + 8}} \Bigg\} \\
=: v^2 & \big\{ 2\log d + h_2(d) \big\}
\end{aligned}$$

where it is easily checked that $h_2(d) \le 3$ for all $d \ge 1$. Moreover $h_2(d)$ is negative for $d > 4.13795 \times 10^{10}$. This completes the proof of the upper bound in (11).

To prove the lower bound for c_d in (11), we use the lower bound of [**13**, Lemma 6.9, p. 157] (which is, in this form, due to Giné and Zinn [**5**]). This yields

$$c_d^2 \ge \frac{\lambda}{1+\lambda} t_o^2 + \frac{1}{1+\lambda} d \int_{t_o}^{\infty} 4t(1-\Phi(t))\,dt \tag{20}$$

for any $t_o > 0$, where $\lambda = 2d(1-\Phi(t_o))$. By using Komatsu's lower bound (18), we find that

$$\begin{aligned}
\int_{t_o}^{\infty} t(1-\Phi(t))\,dt &\ge \int_{t_o}^{\infty} \frac{2t}{t+\sqrt{t^2+4}}\phi(t)\,dt \\
&\ge \frac{2t_o}{t_o+\sqrt{t_o^2+4}} \int_{t_o}^{\infty} \phi(t)\,dt \\
&= \frac{2}{1+\sqrt{1+4/t_o^2}}(1-\Phi(t_o)).
\end{aligned}$$

Using this lower bound in (20) yields

$$\begin{aligned}
c_d^2 &\ge \frac{\lambda}{1+\lambda} t_o^2 + \frac{1}{1+\lambda} d \frac{8}{1+\sqrt{1+4/t_o^2}}(1-\Phi(t_o)) \\
&= \frac{2d(1-\Phi(t_o))}{1+2d(1-\Phi(t_o))}\left\{ t_o^2 + \frac{4}{1+\sqrt{1+4/t_o^2}} \right\} \\
&\ge \frac{\frac{4d}{t_o+\sqrt{t_o^2+4}}\phi(t_o)}{1+\frac{4d}{t_o+\sqrt{t_o^2+4}}\phi(t_o)} \left\{ t_o^2 + \frac{4}{1+\sqrt{1+4/t_o^2}} \right\}.
\end{aligned} \tag{21}$$

Now we let $c \equiv \sqrt{2/\pi}$ and $\delta > 0$ and choose

$$t_o^2 = 2\log\left(\frac{cd}{(2\log(cd))^{(1+\delta)/2}}\right).$$

For this choice we see that $t_o \to \infty$ as $d \to \infty$,

$$4d\phi(t_o) = \frac{2d}{\sqrt{2\pi}} \cdot \frac{(2\log(cd))^{(1+\delta)/2}}{cd} = 2(2\log(cd))^{(1+\delta)/2},$$

and

$$\frac{4d\phi(t_o)}{t_o} = \frac{2(2\log(cd))^{(1+\delta)/2}}{\{2\log(cd/(2\log(cd))^{(1+\delta)/2})\}^{1/2}} \to \infty$$

as $d \to \infty$, so the first term on the right-hand side of (21) converges to 1 as $d \to \infty$, and it can be rewritten as

$$\begin{aligned}
&A_d\left\{t_o^2 + \frac{4}{1+\sqrt{1+4/t_o^2}}\right\}\\
&\quad = A_d\left\{2\log\left(\frac{cd}{(2\log(cd))^{(1+\delta)/2}}\right) + \frac{4}{1+\sqrt{1+4/t_o^2}}\right\}\\
&\quad \sim 1\cdot\big\{2\log d + 2\log c - (1+\delta)\log(2\log(cd)) + 2\big\}.
\end{aligned}$$

To prove the upper bounds for c_d, we will use the upper bound of [**13**, Lemma 6.9, p. 157] (which is, in this form, due to Giné and Zinn [**5**]). For every $t_o > 0$

$$\begin{aligned}
c_d^2 \equiv \mathbb{E}\max_{1\le j\le d}|Z_j|^2 &\le t_o^2 + d\int_{t_o}^{\infty} 2t P(|Z_1| > t)\,dt\\
&= t_o^2 + 4d\int_{t_o}^{\infty} t(1-\Phi(t))\,dt\\
&\le t_o^2 + 4d\int_{t_o}^{\infty} \phi(t)\,dt \qquad \text{(by Mills' ratio)}\\
&= t_o^2 + 4d(1-\Phi(t_o)).
\end{aligned}$$

Evaluating this bound at $t_o = \sqrt{2\log(d/\sqrt{2\pi})}$ and then using Mills' ratio again yields

$$\begin{aligned}
c_d^2 &\le 2\log\big(d/\sqrt{2\pi}\big) + 4d\left(1-\Phi\left(\sqrt{2\log\big(d/\sqrt{2\pi}\big)}\right)\right)\\
&\le 2\log d - 2\frac{1}{2}\log(2\pi) + 4d\frac{\phi\left(\sqrt{2\log\big(d/\sqrt{2\pi}\big)}\right)}{\sqrt{2\log\big(d/\sqrt{2\pi}\big)}}\\
&= 2\log d - \log(2\pi) + \frac{2\sqrt{2}}{\sqrt{\log\big(d/\sqrt{2\pi}\big)}} \qquad (22)\\
&\le 2\log d,
\end{aligned}$$

where the last inequality holds if

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

or equivalently if

$$\log d \geq \frac{8}{(\log(2\pi))^2} + \frac{\log(2\pi)}{2} = 3.28735\ldots,$$

and hence if $d \geq 27 > e^{3.28735\ldots} \doteq 26.77$. The claimed inequality is easily verified numerically for $d = 3, \ldots, 26$. (It fails for $d = 2$.) As can be seen from (22), $2\log d - \log(2\pi)$ gives a reasonable approximation to $\mathbb{E}\max_{1\leq j\leq d} Z_j^2$ for large d. Using the upper bound in (18) instead of the second application of Mills' ratio and choosing $t_o^2 = 2\log(cd/\sqrt{2\log(cd)})$ with $c := \sqrt{2/\pi}$ yields the third bound for c_d in (11) with

$$\begin{aligned} h_3(d) = &-\log(\pi) - \log(\log(cd)) \\ &+ \frac{8}{3\sqrt{1 - \frac{\log(2\log(cd))}{2\log(cd)}} + \sqrt{1 - \frac{\log(2\log(cd))}{2\log(cd)} + \frac{4}{\log(cd)}}}. \end{aligned}$$ ■

6.3. Proofs for Section 4.

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

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

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

$$\mathbb{E}\exp\left(\frac{\pm S_{n,j}}{\kappa L}\right) = \prod_{i=1}^{n}\mathbb{E}\exp\left(\frac{\pm X_{i,j}}{\kappa L}\right) \leq \prod_{i=1}^{n}\exp\left(\frac{\mathrm{Var}(X_{i,j})\mathrm{e}(L)}{\kappa^2}\right) \leq \exp\left(\frac{\Gamma \mathrm{e}(L)}{\kappa^2}\right).$$

Hence

$$\mathbb{E}\cosh\left(\frac{\|S_n\|_\infty}{\kappa L}\right) = \mathbb{E}\max_{1\leq j\leq d}\cosh\left(\frac{S_{n,j}}{\kappa L}\right) \leq \sum_{j=1}^{d}\mathbb{E}\cosh\left(\frac{S_{n,j}}{\kappa L}\right) \leq d\exp\left(\frac{\Gamma \mathrm{e}(L)}{\kappa^2}\right).$$

As in the proof of Lemma 3.4 we conclude that

$$\begin{aligned} \mathbb{E}\|S_n\|_\infty^2 &\leq (\kappa L)^2\left(\log\left(2\mathbb{E}\cosh\left(\frac{\|S_n\|_\infty}{\kappa L}\right)\right)\right)^2 \\ &\leq (\kappa L)^2\left(\log(2d) + \frac{\Gamma \mathrm{e}(L)}{\kappa^2}\right)^2 \\ &= \left(\kappa L\log(2d) + \frac{\Gamma L\,\mathrm{e}(L)}{\kappa}\right)^2, \end{aligned}$$

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

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

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

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

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

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

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

Thus

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

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, because $\mathbb{E}X_i^{(a)} = 0$. 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

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

ACKNOWLEDGMENTS. The authors owe thanks to the referees for a number of suggestions which resulted in a considerable improvement in the article. The authors are also grateful to Ilya Molchanov for drawing their attention to Banach-Mazur distances, and to Stanislaw Kwapien and Vladimir Koltchinskii for pointers concerning type and cotype 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 Dümbgen, van de Geer, and Wellner. The research of Wellner was also supported in part by NSF grants DMS-0503822 and DMS-0804587. The research of Dümbgen and van de Geer was supported in part by the Swiss National Science Foundation.

REFERENCES

1. A. Araujo and E. Giné, *The Central Limit Theorem for Real and Banach Valued Random Variables*, Wiley Series in Probability and Mathematical Statistics, John Wiley, New York, 1980.
2. G. Bennett, Probability inequalities for the sum of independent random variables, *J. Amer. Statist. Assoc.* **57** (1962) 33–45. `doi:10.2307/2282438`
3. V. H. de la Peña and E. Giné, *Decoupling: From Dependence to Independence*, Probability and its Applications, Springer-Verlag, New York, 1999.
4. R. M. Dudley, *Uniform Central Limit Theorems*, Cambridge Studies in Advanced Mathematics, vol. 63, Cambridge University Press, Cambridge, 1999.
5. E. Giné and J. Zinn, Central limit theorems and weak laws of large numbers in certain Banach spaces, *Z. Wahrsch. Verw. Gebiete* **62** (1983) 323–354. `doi:10.1007/BF00535258`
6. R. D. Gordon, Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument, *Ann. Math. Statistics* **12** (1941) 364–366. `doi:10.1214/aoms/1177731721`
7. E. Greenshtein and Y. Ritov, Persistence in high-dimensional linear predictor selection and the virtue of overparametrization, *Bernoulli* **10** (2004) 971–988. `doi:10.3150/bj/1106314846`
8. U. Haagerup, The best constants in the Khintchine inequality, *Studia Math.* **70** (1981) 231–283.
9. W. Hoeffding, Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.* **58** (1963) 13–30. `doi:10.2307/2282952`
10. K. Itô and H. P. McKean, Jr., *Diffusion Processes and their Sample Paths*, Classics in Mathematics, Springer-Verlag, Berlin, 1974.
11. W. B. Johnson and J. Lindenstrauss, Basic concepts in the geometry of Banach spaces, in *Handbook of the Geometry of Banach Spaces*, vol. I, North-Holland, Amsterdam, 2001, 1–84.
12. A. Juditsky and A. S. Nemirovski, Large deviations of vector-valued martingales in 2-smooth normed spaces, Tech. report, Georgia Institute of Technology, Atlanta, GA, 2008.
13. M. Ledoux and M. Talagrand, *Probability in Banach Spaces: Isoperimetry and Processes*, Ergebnisse der Mathematik und ihrer Grenzgebiete 3. Folge / A Series of Modern Surveys in Mathematics, vol. 23, Springer-Verlag, Berlin, 1991.
14. A. S. Nemirovski, Topics in non-parametric statistics, in *Lectures on Probability Theory and Statistics (Saint-Flour, 1998)*, Lecture Notes in Mathematics, vol. 1738, Springer, Berlin, 2000, 85–277.

15. ———, Regular Banach spaces and large deviations of random sums (working paper) 2004.
16. A. S. Nemirovski and D. B. Yudin, *Problem Complexity and Method Efficiency in Optimization*, John Wiley, Chichester, UK, 1983.
17. I. Pinelis, Extremal probabilistic problems and Hotelling's T^2 test under a symmetry condition, *Ann. Statist.* **22** (1994) 357–368. `doi:10.1214/aos/1176325373`
18. ———, Optimum bounds for the distributions of martingales in Banach spaces, *Ann. Probab.* **22** (1994) 1679–1706. `doi:10.1214/aop/1176988477`
19. ———, Monotonicity properties of the relative error of a Padé approximation for Mills' ratio, *J. Inequal. Pure Appl. Math.* **3/2** (2002).
20. ———, Toward the best constant factor for the Rademacher-Gaussian tail comparison, *ESAIM Probab. Stat.* **11** (2007) 412–426. `doi:10.1051/ps:2007027`
21. D. Pollard, *Empirical Processes: Theory and Applications*, NSF-CBMS Regional Conference Series in Probability and Statistics, 2, Institute of Mathematical Statistics, Hayward, CA, 1990.
22. E. Rio, Moment inequalities for sums of dependent random variables under projective conditions, *J. Theoret. Probab.* **22** (2009) 146–163. `doi:10.1007/s10959-008-0155-9`
23. S. J. Szarek and E. Werner, A nonsymmetric correlation inequality for Gaussian measure, *J. Multivariate Anal.* **68** (1999) 193–211. `doi:10.1006/jmva.1998.1784`
24. N. Tomczak-Jaegermann, *Banach-Mazur Distances and Finite-Dimensional Operator Ideals*, Pitman Monographs and Surveys in Pure and Applied Mathematics, vol. 38, Longman Scientific & Technical, Harlow, UK, 1989.
25. S. A. van de Geer, *Applications of Empirical Process Theory*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 6, Cambridge University Press, Cambridge, 2000.
26. A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer Series in Statistics, Springer-Verlag, New York, 1996.

LUTZ DÜMBGEN received his Ph.D. from Heidelberg University in 1990. From 1990–1992 he was a Miller research fellow at the University of California at Berkeley. Thereafter he worked at the Universities of Bielefeld, Heidelberg, and Lübeck. Since 2002 he has been professor of statistics at the University of Bern. His research interests are nonparametric, multivariate, and computational statistics.
Institute of Mathematical Statistics and Actuarial Science, University of Bern, Alpeneggstrasse 22, CH-3012 Bern, Switzerland
duembgen@stat.unibe.ch

SARA A. VAN DE GEER obtained her Ph.D. at Leiden University in 1987. She worked at the Center for Mathematics and Computer Science in Amsterdam, at the Universities of Bristol, Utrecht, Leiden, and Toulouse, and at the Eidgenössische Technische Hochschule in Zürich (2005–present). Her research areas are empirical processes, statistical learning, and statistical theory for high-dimensional data.
Seminar for Statistics, ETH Zurich, CH-8092 Zurich, Switzerland
geer@stat.math.ethz.ch

MARK C. VERAAR received his Ph.D. from Delft University of Technology in 2006. In the year 2007 he was a postdoctoral researcher in the European RTN project "Phenomena in High Dimensions" at the IMPAN institute in Warsaw (Poland). In 2008 he spent one year as an Alexander von Humboldt fellow at the University of Karlsruhe (Germany). Since 2009 he has been Assistant Professor at Delft University of Technology (the Netherlands). His main research areas are probability theory, partial differential equations, and functional analysis.
Delft Institute of Applied Mathematics, Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands
m.c.veraar@tudelft.nl, mark@profsonline.nl

JON A. WELLNER received his B.S. from the University of Idaho in 1968 and his Ph.D. from the University of Washington in 1975. He got hooked on research in probability and statistics during graduate school at the UW in the early 1970s, and has enjoyed both teaching and research at the University of Rochester (1975–1983) and the University of Washington (1983–present). If not for probability theory and statistics, he might be a ski bum.
Department of Statistics, Box 354322, University of Washington, Seattle, WA 98195-4322
jaw@stat.washington.edu