Nemirovski’s inequality revisited: some comparisons

Jon A. Wellner

University of Washington
• joint work with Lutz Dümbgen, Sara van de Geer, and Mark Veraar

• Talk at meeting on Advances in Stochastic Inequalities and their Applications BIRS, Banff, Alberta, 7-12 June 2009

• Email: jaw@stat.washington.edu
  http://www.stat.washington.edu/jaw/jaw.research.html
Outline

- Introduction:
  Bounds for sums of independent random elements
Outline

• Introduction:
  Bounds for sums of independent random elements

• Nemirovski’s inequality
Outline

• Introduction:
  Bounds for sums of independent random elements
• Nemirovski’s inequality
• Proofs of Nemirovski’s inequality:
  ◦ Via deterministic inequalities for norms
  ◦ Via probabilistic methods for Banach spaces
  ◦ Via truncation and Bernstein’s inequality (empirical process methods).
Outline

• Introduction:
  Bounds for sums of independent random elements
• Nemirovski’s inequality
• Proofs of Nemirovski’s inequality:
  ◦ Via deterministic inequalities for norms
  ◦ Via probabilistic methods for Banach spaces
  ◦ Via truncation and Bernstein’s inequality (empirical process methods).
• Comparisons in three settings
Outline

• Introduction:
  Bounds for sums of independent random elements

• Nemirovski’s inequality

• Proofs of Nemirovski’s inequality:
  ◦ Via deterministic inequalities for norms
  ◦ Via probabilistic methods for Banach spaces
  ◦ Via truncation and Bernstein’s inequality (empirical process methods).

• Comparisons in three settings

• Problems and further issues
1. Introduction

- Let $X_1, \ldots, X_n$ be independent random variables with $EX_i^2 < \infty$, $S_n = \sum_{i=1}^{n} X_i$. Then

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

- If $E(X_i) = 0$ for $1 \leq i \leq n$, then (1) becomes

$$ES_n^2 = \sum_{i=1}^{n} EX_i^2.$$  \hfill (2)

- If $X_1, \ldots, X_n$ are independent with values in a Hilbert space $\mathcal{H}$ with inner product $\langle \cdot, \cdot \rangle$, and have $EX_i = 0$ and $E\|X_i\|^2 < \infty$, then

$$E\|S_n\|^2 = \sum_{i=1}^{n} E\|X_i\|^2.$$  \hfill (3)
• What if the $X_i$’s are independent with values in a (real) Banach space $(\mathbb{B}, \| \cdot \|)$? Let $X_1, \ldots, X_n$ be independent random vectors with values in $\mathbb{B}$ with $EX_i = 0$ and $E\|X_i\|^2 < \infty$. Let $S_n = \sum_{i=1}^{n} X_i$. We want inequalities of the form

$$E\|S_n\|^2 \leq K \sum_{i=1}^{n} E\|X_i\|^2$$

(4)

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

• Of special interest: $(\mathbb{B}, \| \cdot \|) = \ell^d_r \equiv (\mathbb{R}^d, \| \cdot \|_r)$ for $r \in [1, \infty]$ where

$$\|x\|_r = \begin{cases} \left( \sum_{j=1}^{d} |x_j|^r \right)^{1/r} & \text{if } 1 \leq r < \infty, \\ \max_{1 \leq j \leq d} |x_j| & \text{if } r = \infty. \end{cases}$$
2. Nemirovski’s inequality

**Theorem 1.** (Nemirovski’s inequality)
Let $X_1, \ldots, X_n$ be independent random vectors in $\mathbb{R}^d$, $d \geq 3$, with $EX_i = 0$ and $E\|X_i\|_2^2 < \infty$. Then for every $r \in [2, \infty]$

$$E\left\|\sum_{i=1}^n X_i\right\|_r^2 \leq K_{Nem}(d, r) \sum_{i=1}^n E\|X_i\|_r^2$$

where $\|\cdot\|_r$ is the $\ell_r$ norm, $\|x\|_r \equiv \left\{\sum_1^d |x_j|^r\right\}^{1/r}$, and where

$$K_{Nem}(d, r) = \inf_{q \in [2, r] \cap \mathbb{R}} (q - 1)d^{2/q - 2/r} \begin{cases} = d^{1-2/r}, & d \leq 7 \\ \leq r - 1, & \text{for all } d \\ \leq 2e \log d - e, & d \geq 3 \end{cases} \leq \min\{r - 1, 2e \log d - e\}.$$
Corollary 1. \((r = \infty \text{ version of Nemirovski’s inequality:})\)
Under the assumptions of Theorem 1

\[
E\left\| \sum_{i=1}^{n} X_i \right\|_\infty^2 \leq (2e \log d - e) \sum_{i=1}^{n} E\left\| X_i \right\|_\infty^2
\]

where \(\| \cdot \|_\infty\) is the \(\ell_\infty\) norm, \(\| x \|_\infty \equiv \max\{ |x_j| : 1 \leq j \leq d \}\).
3. Three Proofs of Nemirovski’s inequality

Proof 1: via deterministic inequalities for norms:

For given \( r \in [2, \infty) \) consider the map \( V_r \) from \( \mathbb{R}^d \) to \( \mathbb{R} \) defined by

\[
V_r(x) \equiv \|x\|_r^2.
\]

Then \( V_r \) is continuously differentiable with Lipschitz continuous derivative \( \nabla V_r \). Furthermore

\[
V_r(x + y) \leq V_r(x) + y' \nabla V_r(x) + (r - 1)V_r(y)
\]  

(5)

for an absolute constant \( C \). Thus, writing

\[
\sum_{i=1}^{n} X_i = \sum_{i=1}^{n-1} X_i + X_n,
\]

it follows from (5) that

\[
V_r(\sum_{i=1}^{n} X_i) \leq V_r(\sum_{i=1}^{n-1} X_i) + X_n' \nabla V_r(\sum_{i=1}^{n-1} X_i) + (r - 1)V_r(X_n).
\]
Taking expectations across this inequality and using $X_n$ and \(\sum_{i=1}^{n-1} X_i\) independent and $E(X_n) = 0$ yields

\[
EV_r \left( \sum_{i=1}^{n} X_i \right) \leq E \left\{ V_r \left( \sum_{i=1}^{n-1} X_i \right) + X'_n \nabla V_r \left( \sum_{i=1}^{n-1} X_i \right) \right\} \\
+ (r - 1) EV_r (X_n) \\
= EV_r \left( \sum_{i=1}^{n-1} X_i \right) + (r - 1) E \|X_n\|_r^2.
\]

By recursion and the definition of $V_r(x)$ this yields

\[
E \| \sum_{i=1}^{n} X_i \|_r^2 \leq (r - 1) \sum_{i=1}^{n} E \|X_i\|_r^2,
\]

so the claim holds with $r - 1$ rather than $K_{Nem}(r, d)$. 
To show that we can replace $r - 1$ by $K_{Nem}(d, r)$ we use the following elementary inequalities: for $1 \leq q \leq r$

$$\|x\|_r \leq \|x\|_q \leq d^{(1/q)-(1/r)} \|x\|_r$$

for all $x \in \mathbb{R}^d$ (by Hölder’s inequality). Thus for $2 \leq q \leq r \leq \infty$ with $q < \infty$,

$$E\|S_n\|_r^2 \leq E\|S_n\|_q^2 \leq (q - 1) \sum_{i=1}^n E\|X_i\|_q^2$$

$$\leq (q - 1)d^{2/q-2/r} \sum_{i=1}^n E\|X_i\|_r^2.$$

This implies

$$E\|S_n\|_r^2 \leq K_{Nem}(d, r) \sum_{i=1}^n E\|X_i\|_r^2.$$
where

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

\[
= d^{1 - 2/r} , \quad d \leq 7
\]

\[
\leq r - 1 , \quad \text{for all } d
\]

\[
\leq 2e \log d - e , \quad d \geq 3
\]

since \(q = 2\) achieves the inf taking \(q = r\)

taking \(q = 2 \log d\).
Proof 2: via probabilistic methods for Banach spaces:
Let \( \{\epsilon_i\} \) be a sequence of independent Rademacher random variables, and let \( 1 \leq p < \infty \). A Banach space \( \mathcal{B} \) 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 \( \mathcal{B} \),

\[
E \left\| \sum_{i=1}^{n} \epsilon_i x_i \right\|^p \leq T_p^{p} \sum_{i=1}^{n} \| x_i \|^p.
\]

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

\[
E \left\| \sum_{i=1}^{n} \epsilon_i x_i \right\|^q \geq C_q^{-q} \sum_{i=1}^{n} \| x_i \|^q.
\]

\( \mathcal{B} = L_r(\mu) \) with \( 1 \leq r < \infty \) is type \( \min\{r, 2\} \) and cotype \( \max\{r, 2\} \).
The following proposition is an elementary consequence of a symmetrization inequality.

**Proposition.** If $\mathbb{B}$ is of type $p \geq 1$ with constant $T_p$, then

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

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

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

is the optimal constant in Khintchine’s inequality due to Haagerup (1981). Hence for $X_1, \ldots, X_n$ independent in $L_r(\mu)$ with $EX_i = 0$ and $E\|X_i\|_r^2 < \infty$,

$$E\|S_n\|_r^2 \leq 4B_r \sum_{i=1}^n E\|X_i\|_r^2.$$
The Banach space $\ell^d_r$ can be viewed as $L_r(\mu)$ with $\mu$ counting measure on $\{1, \ldots, d\}$, so the Corollary covers the case $\ell^d_r$ with $r < \infty$.

What about $\ell^d_\infty = (\mathbb{R}^d, \| \cdot \|_\infty)$? This case requires a separate treatment. Here is one basic result:

**Lemma 2.1.** $\ell^d_\infty$ is type 2 with constant $T_2(\ell^d_\infty) \leq \sqrt{2 \log(2d)}$.

This yields the following Nemirovski-type inequality:

**Corollary 2.1.** For $(\mathbb{B}, \| \cdot \|) = \ell^d_\infty$, inequality (4) holds with $K \equiv K_{Type2}(d, \infty) = 8 \log(2d)$.

**Proof.** For $1 \leq i \leq n$ let $x_i = (x_{ij})_{j=1}^d$ be fixed vectors in $\mathbb{R}^d$, and set

$$S \equiv \sum_{i=1}^n \epsilon_i x_i, \quad S_j = j^{th} - \text{component of } S$$

so $\text{Var}(S_j) \equiv v_j^2 = \sum_{i=1}^n x_{ij}^2$. 
Then
\[ v^2 \equiv \max_{1 \leq j \leq d} v_j^2 \leq \sum_{i=1}^{n} \|x_i\|_\infty^2, \]
and it suffices to show that
\[ E\|S\|_\infty^2 \leq 2 \log(2d)v^2. \]

Define \( h : [0, \infty) \to [1, \infty) \) by \( h(t) = \cosh(\sqrt{t}). \) Then \( h \) is one-to-one, increasing, and convex. Thus \( h^{-1} : [1, \infty) \to [0, \infty) \) is increasing, concave, and
\[ h^{-1}(s) = \left( \log(s + (s^2 - 1)^{1/2}) \right)^2 \leq (\log(2s))^2. \]

Thus by Jensen’s inequality, for arbitrary \( t > 0, \)
\[
E\|S\|_\infty^2 = t^{-2} Eh^{-1}(\cosh(\|tS\|_\infty)) \leq t^{-2} h^{-1}(E \cosh(\|tS\|_\infty)) \\
\leq t^{-2} (\log(2E \cosh(\|tS\|_\infty)))^2. \quad (7)
\]
Furthermore,

\[ E \cosh(\|tS\|_{\infty}) = E \max_{1 \leq j \leq d} \cosh(tS_j) \leq \sum_{j=1}^{d} E \cosh(tS_j) \leq d \exp(t^2 v^2 / 2) \]  

(8)

by using the exponential moment bound

\[ E \exp \left( t \sum_{i=1}^{n} x_{ij} \epsilon_i \right) \leq \exp(t^2 v_j^2 / 2) \leq \exp(t^2 v^2 / 2) \]

which is the basis of Hoeffding’s inequality

\[ P \left( \left| \sum_{i=1}^{n} x_{ij} \epsilon_i \right| \geq z \right) \leq 2 \exp \left( -\frac{z^2}{2v_j^2} \right), \quad z > 0. \]
Combining (8) with (7) yields

\[ E\|S\|_\infty^2 \leq t^{-2} \left( \log(2d \exp(t^2v^2/2)) \right)^2 = \left( \frac{\log(2d)}{t} + \frac{tv^2}{2} \right)^2 \]

\[ = 2 \log(2d)v^2 \]

by choosing \( t = \sqrt{2 \log(2d)/v^2} \).

Refinements: Hoeffding’s inequality

\[ P \left( \left| \sum_{i=1}^{n} a_i \epsilon_i \right| \geq z \right) \leq 2 \exp \left( -\frac{z^2}{2} \right), \quad z > 0 \]

for constants \( a_1, \ldots, a_n \) with \( \sum_{i=1}^{n} a_i^2 = 1 \) has been refined by Pinelis (1994, 2007): for a constant \( K \) with \( 3.18 \leq K \leq 3.22 \),

\[ P \left( \left| \sum_{i=1}^{n} a_i \epsilon_i \right| \geq z \right) \leq 2K(1 - \Phi(z)), \quad z > 0. \]
Pinelis’s inequality can be used to obtain refined bounds for $T_2(\ell^d_\infty)$. To state the result, let

$$c_d^2 \equiv E \max_{1 \leq j \leq d} Z_j^2$$

where $Z_1, \ldots, Z_d$ are i.i.d. $N(0, 1)$.

**Proposition:** The constants $c_d$ and $T_2(\ell^d_\infty)$ satisfy the following inequalities:

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

where ...
\[ h_2(d) \equiv 2 \log(c/2) - \log(2 \log(dc/2)) \]

\[ + \frac{8 \sqrt{2 \log(cd/2)}}{3 \sqrt{2 \log \left( \frac{cd}{2 \sqrt{2 \log(cd/2)}} \right)} + \sqrt{2 \log \left( \frac{cd}{2 \sqrt{2 \log(cd/2)}} \right)} + 8 \]

\[ h_3(d) \equiv - \log(\pi) - \log(\log(cd)) \]

\[ + \frac{8}{3 \sqrt{1 - \frac{\log(2 \log(cd))}{2 \log(cd)}} + \sqrt{1 - \frac{\log(2 \log(cd))}{2 \log(cd)}} + \frac{4}{\log(cd)} \]

where \( h_2(d) \leq 3, h_2(d) < 0 \) for \( d > 4.13795 \times 10^{10} \), \( h_3(d) < 0 \) for \( d \geq 14 \), and \( h_j(d) \sim - \log \log d \) as \( d \to \infty \) for \( j = 1, 2, 3 \).
Proof 3: via truncation and Bernstein’s inequality

Let $Y_1, \ldots, Y_n$ be independent random variables with mean zero satisfying $|Y_i| \leq \kappa$. Then the usual form of Bernstein’s inequality is as follows: for $v^2 = \sum_{i=1}^{n} \text{Var}(Y_i)$,

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

We will not use this inequality itself, but rather an exponential moment inequality which is implicit in its proof.

Lemma 3.1 For $L > 0$ define $e(L) \equiv \exp(1/L) - 1 - 1/L$. Let $Y$ be a random variable with mean zero and variance $\sigma^2$ such that $|Y| \leq \kappa$. Then for any $L > 0$,

$$E \exp \left( \frac{Y}{\kappa L} \right) \leq 1 + \frac{\sigma^2 e(L)}{\kappa^2} \leq \exp \left( \frac{\sigma^2 e(L)}{\kappa^2} \right).$$
This exponential moment inequality yield the following second moment bound for sums of random vectors in $\mathbb{R}^d$ with bounded components:

Lemma 3.2 Suppose that $X_i = (X_{i,j})_{j=1}^d$ satisfies $\|X_i\|_\infty \leq \kappa$ and suppose that $\Gamma \geq \max_{1 \leq j \leq d} \sum_{i=1}^n \text{Var}(X_{i,j})$. Then for any $L > 0$

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

Now consider again our general random vectors $X_i \in \mathbb{R}^d$ with mean zero and $E\|X_i\|_\infty^2 < \infty$. We decompose these as $X_i = X_i^{(a)} + X_i^{(b)}$ via truncation with

$$X_i^{(a)} \equiv X_i 1\{\|X_i\|_\infty \leq \kappa_0\}, \quad X_i^{(b)} \equiv X_i 1\{\|X_i\|_\infty > \kappa_0\}$$

where $\kappa_0$ is a constant to be specified later.
Then $S_n = A_n + B_n$ for centered random sums

$$A_n \equiv \sum_{i=1}^{n} (X_i^{(a)} - EX_i^{(a)}) , \quad B_n \equiv \sum_{i=1}^{n} (X_i^{(b)} - EX_i^{(b)}).$$

The sum $A_n$ involves centered random vectors in $[-2\kappa_0, 2\kappa_0]^d$ and will be treated by means of Lemma 3.2. The sum $B_n$ will be treated directly. Choosing the truncation level $\kappa_0$ and the parameter $L$ carefully yields the following theorem.

**Theorem 3.1** In the case $(\mathbb{B}, \| \cdot \|) = \ell_\infty^d$ for some $d \geq 1$, inequality (4) holds with

$$K = K_{TrBern}(d, \infty) \equiv (1 + 3.46\sqrt{\log(2d)})^2.$$
If the random vectors $X_i$ are all symmetric about 0, then (4) holds with

$$K = K_{TrBern}^{(symm)}(d, \infty) \equiv (1 + 2.9 \sqrt{\log(2d)})^2.$$
4. Comparisons in three settings

Three different setting in which to compare the methods:

- General case:
  The $X_i$’s are independent with $E\|X_i\|_\infty^2 < \infty$ for $1 \leq i \leq n$. 
4. Comparisons in three settings

Three different setting in which to compare the methods:

• **General case:**
  The $X_i$’s are independent with $E\|X_i\|_\infty^2 < \infty$ for $1 \leq i \leq n$.

• **Centered case:** In addition, $EX_i = 0$ for all $1 \leq i \leq n$. 
4. Comparisons in three settings

Three different settings in which to compare the methods:

- **General case:** The \( X_i \)'s are independent with \( E\|X_i\|_\infty^2 < \infty \) for \( 1 \leq i \leq n \).
- **Centered case:** In addition, \( E X_i = 0 \) for all \( 1 \leq i \leq n \).
- **Symmetric case:** In addition, \( X_i \) is symmetrically distributed around 0 for \( 1 \leq i \leq n \).
4. Comparisons in three settings

Three different settings in which to compare the methods:

- **General case:**
  The $X_i$’s are independent with $E\|X_i\|_\infty^2 < \infty$ for $1 \leq i \leq n$.

- **Centered case:** In addition, $EX_i = 0$ for all $1 \leq i \leq n$.

- **Symmetric case:** In addition, $X_i$ is symmetrically distributed around $0$ for $1 \leq i \leq n$.

<table>
<thead>
<tr>
<th></th>
<th>General</th>
<th>Centered</th>
<th>Symmetric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nem</td>
<td>$8e \log d - 4e$</td>
<td>$2e \log d - e$</td>
<td>$2e \log d - e$</td>
</tr>
<tr>
<td>Type-2</td>
<td>$8 \log(2d)$</td>
<td>$8 \log(2d)$</td>
<td>$2 \log(2d)$</td>
</tr>
<tr>
<td></td>
<td>$8 \log d + 4h_2(d)$</td>
<td>$8 \log d + 4h_2(d)$</td>
<td>$2 \log d + h_2(d)$</td>
</tr>
<tr>
<td>TrBern</td>
<td>$(1 + 3.46 \sqrt{\log(2d)})^2$</td>
<td>$(1 + 3.46 \sqrt{\log(2d)})^2$</td>
<td>$(1 + 2.9 \sqrt{\log(2d)})^2$</td>
</tr>
</tbody>
</table>

Table 4: The different constants $K(d, \infty)$. 
Define

\[ K^* \equiv \lim_{d \to \infty} \frac{K(d, \infty)}{\log d}. \]

<table>
<thead>
<tr>
<th></th>
<th>General</th>
<th>Centered</th>
<th>Symmetric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nem</td>
<td>(8e \div 21.7463)</td>
<td>(2e \div 5.4366)</td>
<td>(2e \div 5.4366)</td>
</tr>
<tr>
<td>Type-2</td>
<td>(8.0)</td>
<td>(8.0)</td>
<td>(2.0)</td>
</tr>
<tr>
<td>TrBern</td>
<td>(3.46^2 = 11.9716)</td>
<td>(3.46^2 = 11.9716)</td>
<td>(2.9^2 = 8.41)</td>
</tr>
</tbody>
</table>

Table 5: The different limits \(K^*\).
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.
Full paper, to appear in the *American Mathematical Monthly* available at:

- arXiv:math.ST/0807.2245