Sums of Norm Spheres Are Norm Shells and Lower Triangle Inequalities Are Sharp

By Lutz Mattner

Universität Trier
September 1, 2009

SumsOfNormSpheres.FINAL.20090901.tex

The statements in the title are explained and proved, as a little exercise in elementary normed vector space theory at the level of Chapter 5 of Dieudonné’s Foundations of Mathematical Analysis. A connection to recent moment bounds for submartingales is sketched.

1. A stochastic introduction: From moment bounds for submartingales to elementary normed space theory. This note treats from Section 2 onwards an elementary and natural but seemingly not too well-known exercise in analysis, suitable for mathematics students after their first year at university. The present somewhat lengthy but less elementary and less detailed introductory section explains our probabilistic motivation for undertaking the exercise, but it is not logically necessary for understanding the remaining two sections.

Our motivation stems from moment bounds in probability. Referring to the appropriate sections of [1], [2] and [6] for good general introductions to this topic, we restrict our attention here to the somewhat specialized subtopic of finding the optimal lower and upper bounds for the $r$th absolute moment $\mathbb{E}|S_n|^r$, with $r \in [1, \infty[$, of a sum $S_n = \sum_{i=1}^{n} X_i$ of real-valued random variables, given $n \in \mathbb{N}$, given only the individual moments $\mathbb{E}|X_1|^r, \ldots, \mathbb{E}|X_n|^r$ of the same order, and given some structural assumption on the process $(X_1, \ldots, X_n)$. Structural assumptions of interest include the following ones, ordered here according to increasing generality,

- IIDC: $X_1, \ldots, X_n$ are independent, identically distributed, and centred
- IC: $X_1, \ldots, X_n$ are independent and centred
- MG: $(S_1, \ldots, S_n)$ is a martingale

with “centred” meaning $\mathbb{E}X_i = 0$ for each $i$.

AMS 2000 subject classifications: Primary 46B20; secondary 26D99, 60E15.

Keywords and phrases: Normed vector space, moment inequalities.
For example, under the assumption IIDC, bounding $E|S_n|^r$ in terms of $E|X_1|^r = \ldots = E|X_n|^r$ is equivalent to bounding $E\left|\frac{1}{n}\sum_{i=1}^{n} Y_i - EY_1\right|^r$ for not necessarily centred IID random variables $Y_i$ in terms of $E|Y_1 - EY_1|^r$, and this is of interest as one reasonable way of assessing the quality of $\frac{1}{n}\sum_{i=1}^{n} Y_i$ as a statistical estimator for $EY_1$.

The best known and most convenient optimal bounds of the kind considered here refer to the case of $r = 2$, namely

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

valid and then trivially optimal under each of IIDC, IC and MG. Further optimal bounds are rare. Restricting our attention to $r \in [1,2]$ here (for brevity, and since for $r > 2$ the more effective inequalities of Rosenthal [6, pp. 59, 83] involving moments of orders 2 and $r$ are available) and under the structures considered, the only ones known to me are the trivial upper bound

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

valid (by the usual triangle inequality for expectations, without using any structural assumption) and optimal under each of IID, IC, MG (as can be seen by considering independent random variables $X_i$ with $P(X_i = -a_i/p) = P(X_i = a_i/p) = p/2$ and $P(X_i = 0) = 1 - p$, and letting $p \to 0$), the lower bound of Hornich type

$$
E|S_n| \geq c_n E|X_1| \quad \text{under IIDC}
$$

where, writing $b(n, p; k) := \binom{n}{k}p^k(1-p)^{n-k}$ for binomial probabilities, we have put $c_n := n b(n, \lfloor n/2 \rfloor; \lfloor n/2 \rfloor)$, which is $\sim \sqrt{\frac{2n}{\pi}}$ for $n \to \infty$, and finally

$$
E|S_n| \geq \max \left\{ E|X_k| - \sum_{i=1}^{k-1} E|X_i| \right\}^{n} \cup \left\{ \frac{E|X_k|}{2} \right\}^{n} \quad \text{under MG}
$$

where $\{b_k\}_{k=n_1}^{n_2}$ denotes the possibly empty set $\{b_k : k \in \mathbb{N}, n_1 \leq k \leq n_2\}$. See [4] for (3) and [5] for (4), and both for more detailed discussions and appropriate references.

Generally speaking one should expect that proving optimal bounds as considered here, or equivalently solving associated extremal problems subject to constraints, should be the most difficult under IC and easiest under MG, since $E|S_n|^r$ is under MG a linear functional of the law of $(S_1, \ldots, S_n)$ which varies in a convex set, while under IC it is an $n$-linear functional of the
individual laws $P_1, \ldots, P_n$, namely an integral with respect to the product measure $P_1 \otimes \ldots \otimes P_n$, and finally under IIDC it is an integral with respect to a "power measure" $P^{\otimes n}$. It is therefore surprising and certainly an instance of luck that (3) has been proved. The paper [5] resulted from an attempt to find an analogue of (3) under IC. Being unsuccessful in that respect, the authors of [5] of course modified the assumptions and finally came up with (4) and with a similarly strange result for submartingales. Oddly enough, during the many pizza evenings it took them, they never thought of first solving their moment problems under the simplest of all structural assumptions, namely

N: No assumption, i.e., $X_1, \ldots, X_n$ are arbitrary random variables

Returning to general $r \in [1, \infty]$ now and considering $(\mathbb{E}|S_n|^r)^{1/r}$ instead of $\mathbb{E}|S_n|^r$, this amounts to writing down the best upper and lower bounds for $L^r$ norms of sums given the norms of the summands, and the nonsurprising result, see (6) below, turns out to be the same more generally on every normed space of dimension at least two. The proof of the optimality of the lower bound is a not completely trivial and nice exercise in analysis, as we try to show below.

2. An elementary analytic introduction. Let $(E, \| \cdot \|)$ be a normed vector space over the real numbers $\mathbb{R}$. Then, for $x, y \in E$, the triangle inequalities

\begin{equation}
\left| \|x\| - \|y\| \right| \leq \|x + y\| \leq \|x\| + \|y\| \tag{5}
\end{equation}

provide optimal lower and upper bounds for $\|x + y\|$ in terms of $\|x\|$ and $\|y\|$. Namely, ignoring the trivial case where $E$ is zero-dimensional and given any positive numbers $a, b \in [0, \infty]$, we get $x, y \in E$ with $\|x\| = a$, $\|y\| = b$, and $\|x + y\| = |a - b|$ or $\|x + y\| = a + b$, by choosing some $u \in E$ with $\|u\| = 1$ and then putting $x := au$ and $y := -bu$ or $y := bu$, respectively.

How about more than two vectors? For $n \geq 2$ and $x_1, \ldots, x_n \in E$, the inequalities (5) generalize quite obviously to

\begin{equation}
\max_{k=1}^n \left( \|x_k\| - \sum_{i \neq k} \|x_i\| \right) \leq \left\| \sum_{i=1}^n x_i \right\| \leq \sum_{i=1}^n \|x_i\| \tag{6}
\end{equation}

with $c_+ := \max\{c, 0\}$ for $c \in \mathbb{R}$, and with "$i \neq k$" of course meaning "$i \in \{1, \ldots, n\} \setminus \{k\}$". Here the left hand inequality in (6) follows from observing

$$
\left\| \sum_{i=1}^n x_i \right\| = \left\| x_k + \sum_{i \neq k} x_i \right\| \geq \|x_k\| - \left\| \sum_{i \neq k} x_i \right\| \geq \|x_k\| - \sum_{i \neq k} \|x_i\|
$$
for each \(k \in \{1, \ldots, n\}\). But can such a simple-minded approach yield optimal bounds for \(\| \sum_{i=1}^{n} x_i \|\) given \(\|x_1\|, \ldots, \|x_n\|\) also if \(n \geq 3\)? For the upper bound in (5), the answer is obviously “yes”, by choosing the \(x_i\) to be appropriate positive multiples of some nonzero \(u \in E\), as above for the special case of \(n = 2\). For the lower bound, however, the answer is clearly “no” when \(E\) is one-dimensional: Then, for example, if \(n = 3\) and \(\|x_1\| = \|x_2\| = \|x_3\| = 1\), we necessarily have \(\|x_1 + x_2 + x_3\| \in \{1, 3\}\), whereas (6) only yields the trivial lower bound \(\|x_1 + x_2 + x_3\| \geq 0\). The purpose of the remainder of this note is to show that in higher dimensions the inequalities (5) are indeed optimal:

**Proposition.** Let \((E, \| \cdot \|)\) be a normed vector space over \(\mathbb{R}\) with dimension at least 2. Then, for every choice of \(a, b \in [0, \infty[\), we have

\[
\{ x + y : x, y \in E, \|x\| = a, \|y\| = b \}
\]

\[
= \{ z \in E : |a - b| \leq \|z\| \leq a + b \}
\]

and, more generally, for every integer \(n \geq 2\) and every \(a \in [0, \infty[^n\),

\[
\left\{ \sum_{i=1}^{n} x_i : x_i \in E \text{ and } \|x_i\| = a_i \text{ for each } i \right\}
\]

\[
= \left\{ z \in E : \max_{k=1}^{n} (a_k - \sum_{i \neq k} a_i)_+ \leq \|z\| \leq \sum_{i=1}^{n} a_i \right\}
\]

Writing

\[
S_r := \{ x \in E : \|x\| = r \}
\]

for the possibly degenerate norm sphere of radius \(r \in [0, \infty[\) around \(0 \in E\), identity (7) says that the sum of two norm spheres is the closed norm shell around 0 with inner and outer radii obtained from the inequalities (5), and identity (8) is the corresponding statement for the sum of \(n\) spheres, with the radii of the resulting shell obtained from (6).

Proving the proposition turns out to be a little exercise in elementary normed vector space theory rather exactly at the level of [3, section 5.1], and accordingly we have written the present sections 2 and 3 with a potential student of that reference in mind.

### 3. Proofs.

**Lemma 1.** Let \((E, \| \cdot \|)\) be a normed vector space over \(\mathbb{R}\) with dimension at least 2. Then, for each \(r \in [0, \infty[\), the set \(S_r\) from (9) is connected.
Proof. Let \(x, y \in E\) with \(x \neq y\). Then we can choose \(z \in E\) linearly independent of \(y - x\), and observe that then the continuous paths

\[
[0, 1] \ni t \mapsto x + t(y - x) + \min\{t, 1 - t\} \alpha z =: \gamma_\alpha(t)
\]

with \(\alpha \in \mathbb{R}\) have trajectories disjoint except for their endpoints \(x\) and \(y\), that is, \(\gamma_\alpha([0, 1]) \neq \gamma_\beta([0, 1])\) for \(\alpha \neq \beta\). Hence, if also \(x, y \neq 0\), then at most one \(\gamma_\alpha\) passes through 0. This shows that \(E \setminus \{0\}\) is connected. Hence so is its image \(S_r\) under the continuous map \(E \setminus \{0\} \ni x \mapsto rx/\|x\|\). \(\square\)

Remark. The above proof yields in fact that \(E \setminus A\) is connected whenever \(A\) is at most countable. See [3, exercise 5 in section 5.1] for further such results.

Lemma 2. Let \(I \subset \mathbb{R}\) be a nonempty compact interval and let \(f, g : I \to \mathbb{R}\) be continuous functions with \(f \leq g\). Then \(\bigcup_{x \in I}[f(x), g(x)] = [\min f, \max g]\).

Proof. Let \(A\) and \(B\) denote the two sets claimed to be equal. Trivially, \(A \subset B\). Let \(x, y \in I\) with \(f(x) = \min f\) and \(g(y) = \max g\). Then

\[
z(t) := \begin{cases} f(x) + 2t(g(x) - f(x)) & \text{for } t \in [0, \frac{1}{2}] \\ g(x + 2(t - \frac{1}{2})(y - x)) & \text{for } t \in [\frac{1}{2}, 1] \end{cases}
\]

defines a continuous function \(z : [0, 1] \to A\), with \(z(0) = \min f\) and \(z(1) = \max g\). Hence, by the intermediate value theorem, we have \(B \subset A\). \(\square\)

Proof of the Proposition. Let \(A\) and \(B\) denote the two sides of \((7)\). We have \(A \subset B\) by \((5)\). So let \(z_0 \in B\). Using the notation from \((9)\), we consider the continuous function

\[
S_a \ni x \mapsto \|z_0 - x\| =: h(x)
\]

If \(z_0 \neq 0\), we put \(u := z_0/\|z_0\|\), otherwise we choose an arbitrary \(u \in E\) with \(\|u\| = 1\). Then \(x_1 := au\) and \(x_2 := -au\) belong to \(S_a\), and the assumption \(|a - b| \leq \|z_0\| \leq a + b\) yields \(h(x_1) = \|z_0 - au\| = \|z_0\| - a \leq b\) and \(h(x_2) = \|z_0 + au\| = \|z_0\| + a \geq b\). By Lemma 1 and by the continuity of \(h\), the image \(h(S_a)\) is connected, so there exists an \(x \in S_a\) with \(h(x) = b\), that is, \(y := z_0 - x \in S_b\). Thus \(z_0 = x + y \in A\), and \((7)\) is proved.

If \(n = 2\), then \((8)\) is just \((7)\) with \(a_1, a_2 \in \text{in place of } a, b\). So assume that \(n \geq 2\), \(a_1, \ldots, a_{n+1} \in [0, \infty]\), and that \((5)\) holds. To prepare for an application of Lemma 2 we let \(\alpha := \max_{k=1}^n(a_k - \sum_{i \in \{1, \ldots, n\} \setminus \{k\}} a_i)\), \(\beta := \sum_{i=1}^n a_i\), and \(f(r) := |r - a_{n+1}|\) and \(g(r) := r + a_{n+1}\) for \(r \in [\alpha, \beta]\); and observe that

\[
\min f = \begin{cases} 0 & \text{if } \alpha \leq a_{n+1} \leq \beta \\ \alpha - a_{n+1} & \text{if } a_{n+1} \leq \alpha \\ a_{n+1} - \beta & \text{if } a_{n+1} \geq \beta \end{cases} = \max_{k=1}^{n+1} \left( a_k - \sum_{i \in \{1, \ldots, n+1\} \setminus \{k\}} a_i \right)
\]
and \(\max g = \sum_{i=1}^{n+1} a_i\). Using now the notations \(S_a + S_b\) and \(\sum_{i=1}^{n} S_a_i\) for the left hand sides of (7) and (8), we may write the inductive hypothesis (8) as \(\sum_{i=1}^{n} S_a_i = \bigcup_{r \in [\alpha, \beta]} S_r\) and obtain

\[
\sum_{i=1}^{n+1} S_a_i = \bigcup_{r \in [\alpha, \beta]} S_r + S_{a_{n+1}} \\
= \bigcup_{r \in [\alpha, \beta]} \left( S_r + S_{a_{n+1}} \right) \\
= \bigcup_{r \in [\alpha, \beta]} \bigcup_{\varphi \in [f(r), g(r)]} S_{\varphi} \quad \text{[by (7)]} \\
= \bigcup_{\varphi \in [\min f, \max g]} S_{\varphi} \quad \text{[by Lemma 2]}
\]

and the last set equals the right hand side of (8) with \(n + 1\) in place of \(n\).

**References.**

[1] Y.S. Chow and H. Teicher (1997). *Probability Theory. Independence, Interchangeability, Martingales.* Third Edition. Springer, N.Y.

[2] V. H. de la Peña and E. Giné (1999). *Decoupling: From Dependence to Independence.* Springer, N.Y.

[3] J. Dieudonné (1969). *Foundations of Modern Analysis.* 9th, enlarged and corrected printing, Academic Press, New York and London. Available at [www.archive.org/details/FoundationsOfModernAnalysis](www.archive.org/details/FoundationsOfModernAnalysis)

[4] L. Mattner (2003). Mean absolute deviations of sample means and minimally concentrated binomials. *Ann. Probab.* 31, 914-925.

[5] L. Mattner and U. Rösler. Optimal \(L^1\)-bounds for submartingales. Preprint. [arXiv:0809.3522](arXiv:0809.3522)

[6] V.V. Petrov (1995). *Limit Theorems of Probability. Sequences of independent random variables.* Oxford University Press, Oxford.