COMBINATORICA
Akadémiai Kiadó – Springer-Verlag

COMBINATORICA 12 (3) (1992) 303–316

ON THE PRODUCT OF SIGN VECTORS AND UNIT VECTORS*

RON HOLZMAN** and DANIEL J. KLEITMAN

Received September 28, 1989
Revised February 2, 1990

For a fixed unit vector $a=(a_1,\dots,a_n)\in S^{n-1}$, consider the 2^n sign vectors $\epsilon=(\varepsilon_1,\dots,\varepsilon_n)\in\{\pm1\}^n$ and the corresponding scalar products $\epsilon\cdot a=\sum_{i=1}^n\varepsilon_i a_i$. The question that we address is: for how many of the sign vectors must $\epsilon\cdot a$ lie between -1 and 1. Besides the straightforward interpretation in terms of the sums $\sum\pm a_i$, this question has appealing reformulations using the language of probability theory or of geometry.

The natural conjectures are that at least $\frac12$ the sign vectors yield $|\epsilon\cdot a|\le 1$ and at least $\frac38$ of the sign vectors yield $|\epsilon\cdot a|<1$ (the latter excluding the case when $|a_i|=1$ for some i). These conjectured lower bounds are easily seen to be the best possible. Here we prove a lower bound of $\frac38$ for both versions of the problem, thus completely solving the version with strict inequality. The main part of the proof is cast in a more general probabilistic framework: it establishes a sharp lower bound of $\frac38$ for the probability that $|X+Y|<1$, where X and Y are independent random variables, each having a symmetric distribution with variance $\frac12$.

We also consider an asymptotic version of the question, where $n\to\infty$ along a sequence of instances of the problem satisfying $\|a\|_\infty\to 0$. Our result, best expressed in probabilistic terms, is that the distribution of $\epsilon\cdot a$ converges to the standard normal distribution, and in particular the fraction of sign vectors yielding $\epsilon\cdot a$ between -1 and 1 tends to $\sim$68%.

1. Introduction

The following unsolved problem was presented in [2] and attributed to B. Tomaszewski. Consider n real numbers $a_1,\dots,a_n$ satisfying $\sum_{i=1}^n a_i^2=1$. Of the 2^n sums of the form $\sum\pm a_i$, is it possible that there are more with $|\sum\pm a_i|>1$ than there are with $|\sum\pm a_i|\le 1$?

The natural conjecture is that it is impossible. Formally, let $S^{n-1}=\{a=(a_1,\dots,a_n)\in\mathbb{R}^n:\sum_{i=1}^n a_i^2=1\}$ and let $\{\pm1\}^n=\{\epsilon=(\varepsilon_1,\dots,\varepsilon_n):\varepsilon_i=1\text{ or }-1,\ i=1,\dots,n\}$. For $\epsilon\in\{\pm1\}^n$ and $a\in S^{n-1}$, $\epsilon\cdot a=\sum_{i=1}^n\varepsilon_i a_i$ is their scalar product.

Conjecture 1.1. *For all* $a\in S^{n-1}$, $|\{\epsilon\in\{\pm1\}^n:|\epsilon\cdot a|\le 1\}|/2^n\ge\frac12$.

One can verify the conjecture for low values of n. In particular, for $n=2$ and $a_1,a_2>0$, we have $|\epsilon\cdot a|\le 1$ if and only if $\varepsilon_1\ne\varepsilon_2$, so the lower bound of $\frac12$ is attained. For large values of n, it is instructive to think about the special case when $a_1=\dots=$

AMS subject classification code (1991): 60 C 05, 05 A 20

* This research was supported in part by the Institute for Mathematics and its Applications with funds provided by the National Science Foundation.

** Incumbent of the Robert Edward and Roselyn Rich Manson Career Development Chair.

$a_n = \sqrt{\frac{1}{n}}$. Here $|\epsilon \cdot a| \le 1$ if and only if $|\sum_{i=1}^n \varepsilon_i| \le \sqrt{n}$; by the normal approximation to the binomial distribution, the latter holds true for $\sim 68\%$ of the sign vectors ϵ.

We want to suggest three appealing interpretations of the conjecture:

(i) Sum partitions. The conjecture asserts that there are many ways to partition a sum $\sum a_i$ into two roughly equal partial sums. More explicitly, "many" means at least half of all partitions, and "roughly equal" means (under the normalization $\sum a_i^2 = 1$) differing by at most 1.

(ii) Chebyshev-type inequality. When $\{\pm 1\}^n$ is viewed as a probability space, with all ϵ equiprobable, $\epsilon \cdot a$ becomes a random variable. Denoting $X = \epsilon \cdot a$ we have $E(X) = 0$ and $\mathrm{Var}(X) = \sum a_i^2 = 1$. The conjecture states that X lies within one standard deviation of its mean with probability $\ge \frac{1}{2}$. This is not true of a general random variable; Chebyshev's inequality yields a lower bound of 0 for this probability. Yet, according to the conjecture, this is true for X that is a linear combination of mutually independent random variables each assuming the values ± 1 with equal probabilities.

(iii) A ball and a cube. Consider an n-dimensional ball and a smallest n-dimensional cube containing it. The conjecture asserts that for any pair of parallel supporting hyperplanes of the ball, at least half the vertices of the cube lie between (or on) the two hyperplanes.

A variant of the problem addressed by Conjecture 1.1 is obtained by insisting on $|\epsilon \cdot a| < 1$ instead of $|\epsilon \cdot a| \le 1$. To get a meaningful lower bound for this version, one must exclude those $a \in S^{n-1}$ with $|a_i| = 1$ for some i. Next, consider $n = 4$ and $a_1 = \cdots = a_4 = \frac{1}{2}$; in this case just 6 out of 16 sign vectors yield $|\epsilon \cdot a| < 1$. We shall prove that this is the lowest possible:

Theorem 1.2. *For all $a \in S^{n-1}$ with $|a_i| < 1$, $i = 1, \ldots, n$,*

$$|\{\epsilon \in \{\pm 1\}^n : |\epsilon \cdot a| < 1\}|/2^n \ge \frac{3}{8}.$$

Our method of proof does not seem to lead to a better lower bound when $|\epsilon \cdot a| \le 1$ is under consideration. Thus, for the original problem addressed by Conjecture 1.1 we only have the following immediate consequence of Theorem 1.2:

Corollary 1.3. *For all $a \in S^{n-1}$, $|\{\epsilon \in \{\pm 1\}^n : |\epsilon \cdot a| \le 1\}|/2^n \ge \frac{3}{8}$.*

Our strategy for proving Theorem 1.2 grew out of the following idea. Suppose that the set $[n] = \{1, \ldots, n\}$ is partitioned into two subsets I and $\overline{I}$ so that $\sum_{i \in I} a_i^2 = \sum_{i \in \overline{I}} a_i^2 = \frac{1}{2}$. Define two random variables over the probability space $\{\pm 1\}^n$ (with all ϵ equiprobable) by $X = \sum_{i \in I} \varepsilon_i a_i$, $Y = \sum_{i \in \overline{I}} \varepsilon_i a_i$. Then X and Y are independent and each has a symmetric distribution with variance $\frac{1}{2}$. By Chebyshev's inequality $P\{|X| < 1\} \ge \frac{1}{2}$, $P\{|Y| < 1\} \ge \frac{1}{2}$. It follows by independence and symmetry that at least $\frac{1}{8}$ of the time X and Y are < 1 in absolute value and have opposite signs (or one of them is 0). Since $X + Y = \epsilon \cdot a$, we obtain that $P\{|\epsilon \cdot a| < 1\} \ge \frac{1}{8}$.

There are two aspects of this argument that call for an improvement. First, a "balanced" partition of $[n]$, i.e., one with $\sum a_i^2 = \frac{1}{2}$ over each part, may not exist. Second, the lower bound of $\frac{1}{8}$ is considerably lower than we would like, and this is because the argument fails to exploit all combinations of values of X and Y that entail $|\epsilon \cdot a| < 1$.

The structure of our proof is the following. In Section 2 we show that if a balanced partition exists and X and Y are defined as above then $P\{|X+Y|<1\} \geq \frac{3}{8}$. This takes care of the second point made in the last paragraph. With respect to the first point, we define X and Y in the same manner using any partition of $[n]$ into I and $\bar{I}$. For any such partition $P\{|X+Y|<1\} = P\{|\epsilon \cdot a|<1\}$, and we look for partitions that will enable us to establish a good lower bound for this probability. In Section 3 we handle several special cases in which we can show that $P\{|X+Y|< 1\} \geq \frac{3}{8}$ using partitions with $|I| \leq 3$. These special cases are characterized, roughly speaking, by the concentration of a significant part of the sum $\sum a_i^2$ in one, two or three of the summands. In Section 4 we show that if none of the special cases occurs then for some j with $a_j^2 < \frac{1}{12}$ there exists a partition of $[n]\backslash\{j\}$ into I and $\bar{I}$ such that $\sum_{i\in I} a_i^2 \leq \frac{1}{2}$ and $\sum_{i\in \bar{I}} a_i^2 \leq \frac{1}{2}$; this situation is close enough to the balanced case to permit us to obtain the lower bound of $\frac{3}{8}$ by an extension of the same argument.

A related question that we also address in this paper is: can we say more about the problem if n is assumed to be large? The fact that n is large is not helpful by itself, since the distribution of $\epsilon \cdot a$ remains the same if $a = (a_1, \ldots, a_n)$ is replaced by $(a_1, \ldots, a_n, 0, \ldots, 0)$. On the other hand, we have observed above that if n is large and the a_i, $i = 1, \ldots, n$, are equal, then $|\epsilon \cdot a| \leq 1$ holds for $\sim 68\%$ of the sign vectors ϵ. The following theorem, which we prove in Section 4, establishes this phenomenon under a more general assumption on the behavior of the a_i.

Theorem 1.4. *For $n = 1, 2, \ldots$ let $a_n = (a_{n1}, \ldots, a_{nn}) \in S^{n-1}$ and assume that $\max_{i=1,\ldots,n} |a_{ni}| \to 0$ as $n \to \infty$. Over the probability space $\{\pm 1\}^n$ (with all $\epsilon_n \in \{\pm 1\}^n$ equiprobable) define the random variable $X_n = \epsilon_n \cdot a_n$. Then as $n \to \infty$ the distributions of X_n tend to the standard normal distribution function Φ. In particular, $P\{|X_n| \leq 1\}$ as well as $P\{|X_n| < 1\}$ tend to $\Phi(1) - \Phi(-1) \approx 0.6826$.*

2. The Balanced Case

In this section we assume that $a = (a_1, \ldots, a_n) \in S^{n-1}$ and $[n]$ is partitioned into I and $\bar{I}$ so that $\sum_{i\in I} a_i^2 = \sum_{i\in \bar{I}} a_i^2 = \frac{1}{2}$. For the random variables $X = \sum_{i\in I} \varepsilon_i a_i$ and $Y = \sum_{i\in \bar{I}} \varepsilon_i a_i$ defined over the probability space $\{\pm 1\}^n$ (with all $\epsilon = (\varepsilon_1, \ldots, \varepsilon_n) \in \{\pm 1\}^n$ equiprobable), we prove that $P\{|X+Y|<1\} \geq \frac{3}{8}$. The example mentioned in the Introduction, with $a_1 = \cdots = a_4 = \frac{1}{2}$, indicates that this lower bound is sharp.

The first idea in the proof is to forget about the particular way X and Y are defined, and maintain only some of their properties. Thus, we shall establish the following more general fact.

Proposition 2.1. *Let X and Y be independent random variables, each assuming finitely many real values and having a symmetric distribution with variance $\frac{1}{2}$. Then $P\{|X+Y|<1\} \geq \frac{3}{8}$.*

In anticipation of the need for an extension of this result (that will arise in Section 4), we shall actually prove here:

Proposition 2.2. *Let X and Y be independent random variables, each assuming finitely many real values and having a symmetric distribution, with*

$$6-4\sqrt{2} \le \mathrm{Var}(X) \le \frac{1}{2} \text{ and } 6-4\sqrt{2} \le \mathrm{Var}(Y) \le \frac{1}{2}.$$

Let a be a real number with $|a|<1-\sqrt{\frac{1}{2}}$. Then $P\{-1+a<X+Y<1+a\}\ge\frac{3}{8}$.

Proof. Let $x_1,\dots,x_k$ be all the distinct values that $|X|$ assumes, and let $p_i = P\{|X|=x_i\}$, $i=1,\dots,k$. Similarly, let $y_1,\dots,y_\ell$ be all the distinct values that $|Y|$ assumes, and let $q_j=P\{|Y|=y_j\}$, $j=1,\dots,\ell$. Denote $P=P\{-1+a<X+Y<1+a\}$. Then:

$$P=\sum_{i=1}^{k}\sum_{j=1}^{\ell} c_{ij}p_iq_j, \tag{1}$$

where c_{ij} is $\frac{1}{4}$ times the number of valid inequalities among

$$\begin{aligned}
&(\alpha)\ x_i+y_j<1-a,\\
&(\beta)\ x_i+y_j<1+a,\\
&(\gamma)\ |x_i-y_j|<1-a,\\
&(\delta)\ |x_i-y_j|<1+a.
\end{aligned}$$

Given the x_i and the y_j, (1) can be regarded as defining a bilinear function $P\colon D_1\times D_2\to\mathbb{R}$, where D_1 and D_2 are polytopes in $\mathbb{R}^k$ and $\mathbb{R}^\ell$ respectively defined by:

$$p=(p_1,\dots,p_k)\in D_1 \iff \begin{cases} p_i\ge 0, & i=1,\dots,k\\ \sum_{i=1}^k p_i=1 \\ \sum_{i=1}^k x_i^2p_i=\mathrm{Var}(X)\end{cases}$$

$$q=(q_1,\dots,q_\ell)\in D_2 \iff \begin{cases} q_j\ge 0, & j=1,\dots,\ell\\ \sum_{j=1}^\ell q_j=1 \\ \sum_{j=1}^\ell y_j^2q_j=\mathrm{Var}(Y)\end{cases}$$

A minimum of P in $D_1\times D_2$ can be found at a point $(p,q)\in\operatorname{ext}D_1\times\operatorname{ext}D_2$, where $\operatorname{ext}D_i$ is the set of extreme points of D_i (this is a consequence of the bilinearity of P). If $p\in\operatorname{ext}D_1$ then at least $k-2$ of the inequalities $p_i\ge 0$ in the definition of D_1 must be satisfied by p with equality. An analogous statement holds for $q\in\operatorname{ext}D_2$.

The upshot of all this is that we may assume that $k\le 2$ and $\ell\le 2$. In fact, we shall assume from now on that $k=\ell=2$ (if for instance $k=1$, we add an arbitrary x_2 with probability 0). Updating our notation, we have now:

$$\begin{aligned}
&x_1<x_2, \quad p=P\{|X|=x_1\}, \quad 1-p=P\{|X|=x_2\};\\
&y_1<y_2, \quad q=P\{|Y|=y_1\}, \quad 1-q=P\{|Y|=y_2\}.
\end{aligned}$$

By assumption,

$$px_1^2+(1-p)x_2^2\le\frac{1}{2} \quad\text{and}\quad qy_1^2+(1-q)y_2^2\le\frac{1}{2}. \tag{2}$$

It follows that

(3) $$x_1 \leq \sqrt{\frac{1}{2}} \quad \text{and} \quad y_1 \leq \sqrt{\frac{1}{2}}.$$

Similarly, from the lower bound of $6-4\sqrt{2}$ on the variances we deduce that

(4) $$x_2 \geq 2-\sqrt{2} \quad \text{and} \quad y_2 \geq 2-\sqrt{2}.$$

We assume w.l.o.g. that $a \geq 0$. Thus, the following implications among the inequalities that determine c_{ij} in (1) hold: $(\alpha) \Rightarrow (\beta)$ and (γ), (β) or $(\gamma) \Rightarrow (\delta)$. We shall use these implications in our arguments without explicit mention.

From (3) and the assumption that $a < 1-\sqrt{\frac{1}{2}}$ it follows that $|x_1 - y_1| < 1-a$, which accounts for

(5) $$P \geq \frac{1}{2}pq.$$

From here on the argument to show that $P \geq \frac{3}{8}$ splits into several cases.

Case I. At least one of the following holds:

$$\begin{array}{ll} \text{(i)} & x_2 - y_1 < 1-a, \\ \text{(ii)} & x_2 + y_1 < 1+a, \\ \text{(iii)} & y_2 - x_1 < 1-a, \\ \text{(iv)} & y_2 + x_1 < 1+a. \end{array}$$

W.l.o.g. we assume that (i) or (ii) holds. In this case (5) can be improved to get

(6) $$P \geq \frac{1}{2}pq + \frac{1}{2}(1-p)q = \frac{1}{2}q.$$

Now, if $q \geq \frac{3}{4}$ then (6) implies $P \geq \frac{3}{8}$. We assume henceforth that

(7) $$q < \frac{3}{4}.$$

Putting this into (2) yields $y_2 < \sqrt{2}$, which in turn implies, using (4), that

(8) $$y_2 - x_2 < \sqrt{2} - (2-\sqrt{2}) < 1.$$

Our assumption that (i) or (ii) holds implies that

(9) $$x_2 - y_2 < 1-a.$$

Indeed, if (i) holds then $x_2 - y_2 < x_2 - y_1 < 1-a$; if (ii) holds then, using (4), $x_2 - y_2 \leq x_2 + y_1 - y_2 < 1+a-(2-\sqrt{2}) < 1-a$. From (8) and (9) we deduce that $|x_2 - y_2| < 1+a$, and so (6) can be improved to get

(10) $$P \geq \frac{1}{2}q + \frac{1}{4}(1-p)(1-q).$$

Now, if (iii) or (iv) holds then we may add $\frac{1}{2}p(1-q)$ to the right-hand-side of (10); also, from (9) and the fact that $y_2 - x_2 < 1-a$ (which follows from (iii) or (iv) in the same manner as (9) follows from (i) or (ii)) we deduce that $|x_2 - y_2| < 1-a$, which justifies an additional term of $\frac{1}{4}(1-p)(1-q)$; thus $P \geq \frac{1}{2}q + \frac{1}{2}(1-p)(1-q) + \frac{1}{2}p(1-q) = \frac{1}{2}$. So we assume henceforth that (iii) and (iv) fail, i.e.,

(11) $$y_2 \geq 1-a+x_1 \quad \text{and} \quad y_2 \geq 1+a-x_1.$$

This entails in particular that $y_2 \geq 1$, so from (2) we deduce that

(12) $$q \geq \frac{1}{2}.$$

If $x_1 + y_1 < 1+a$ then (10) can be improved to yield $P \geq \frac{1}{2}q + \frac{1}{4}(1-p)(1-q) + \frac{1}{4}pq = \frac{3}{8} + \frac{1}{2}(p+\frac{1}{2})(q-\frac{1}{2}) \geq \frac{3}{8}$ (using (12) for the last inequality). Thus we assume that $x_1 + y_1 \geq 1+a$. Coupled with the first part of (11) this implies that

(13) $$y_2 \geq 2 - y_1.$$

We state and prove now an auxiliary claim which will be used to complete the argument here and will also serve us later on.

Claim 2.3. *Let u satisfy $qy_1^2 + (1-q)y_2^2 \leq 1-u^2$, and suppose $y_2 \geq 2-y_1$. Then $q \geq \frac{1+u}{2}$.*

Proof. Put $t = 1-y_1$. Then $y_2 \geq 1+t$ and so

$$q(1-t)^2 + (1-q)(1+t)^2 \leq 1-u^2.$$

From here we obtain

$$q \geq \frac{(1+t)^2 - 1 + u^2}{4t}.$$

Thus it suffices to show that

$$\frac{(1+t)^2 - 1 + u^2}{4t} \geq \frac{1+u}{2}.$$

The latter is equivalent to $2t + t^2 + u^2 \geq 2t(1+u)$, which in turn is equivalent to $(t-u)^2 \geq 0$. ∎

To complete our argument for Case I, we use the claim with $u = \sqrt{\frac{1}{2}}$ to conclude that $q \geq \frac{1+\sqrt{\frac{1}{2}}}{2}$, which contradicts (7).

From now on we assume that Case I does not occur, i.e.,

(14) $$x_2 \geq 1-a+y_1 \quad \text{and} \quad y_2 \geq 1-a+x_1,$$

(15) $$x_2 \geq 1+a-y_1 \quad \text{and} \quad y_2 \geq 1+a-x_1.$$

This implies in particular that $x_2 \geq 1$ and $y_2 \geq 1$, so from (2) we obtain that

(16) $$p \geq \frac{1}{2} \quad \text{and} \quad q \geq \frac{1}{2}.$$

Case II. Not Case I and $x_1+y_1<1-a$.

In this case (5) can be improved to get

$$P \geq pq. \tag{17}$$

Now, if $|x_2-y_2|<1-a$ then (17) can be improved to get $P \geq pq+\frac{1}{2}(1-p)(1-q)=\frac{1}{3}+\frac{3}{2}(p-\frac{1}{3})(q-\frac{1}{3})$; using (16) this implies $P\geq\frac{3}{8}$. Thus we assume that $|x_2-y_2|\geq 1-a$, and actually, w.l.o.g., $y_2\geq 1-a+x_2$. With the first part of (15) this yields $y_2 \geq 2-y_1$, and so by Claim 2.3 we have $q\geq\frac{1+\sqrt{\frac{1}{2}}}{2}$. Putting this and $p\geq\frac{1}{2}$ into (17) we get $P\geq\frac{1+\sqrt{\frac{1}{2}}}{4}>\frac{3}{8}$, completing the argument for Case II.

Case III. Not Case I and $x_1+y_1\geq 1+a$.

Put $t=\frac{1+a}{2}-x_1$. Then $x_1=\frac{1+a}{2}-t$ and, using the first part of (14) and the assumption of this case, $x_2\geq 1-a+y_1\geq 2-x_1=\frac{3-a}{2}+t$. So (2) yields

$$p\left(\frac{1+a}{2}-t\right)^2+(1-p)\left(\frac{3-a}{2}+t\right)^2\leq\frac{1}{2}.$$

From here we obtain

$$p \geq \frac{\frac{7-6a+a^2}{4}+(3-a)t+t^2}{2-2a+4t}. \tag{18}$$

We also have $y_1\geq 1+a-x_1=\frac{1+a}{2}+t$ and $y_2\geq 1-a+x_1=\frac{3-a}{2}-t$, so (2) yields

$$q\left(\frac{1+a}{2}+t\right)^2+(1-q)\left(\frac{3-a}{2}-t\right)^2\leq\frac{1}{2}.$$

From here we obtain

$$q \geq \frac{\frac{7-6a+a^2}{4}-(3-a)t+t^2}{2-2a-4t}. \tag{19}$$

Given (5), (18) and (19), it suffices to show that

$$\frac{1}{2}\cdot\frac{\frac{7-6a}{4}+(3-a)t}{2-2a+4t}\cdot\frac{\frac{7-6a}{4}-(3-a)t}{2-2a-4t}\geq\frac{3}{8}.$$

This is equivalent to $(\frac{7-6a}{4})^2-(3-a)^2t^2\geq 3[(1-a)^2-4t^2]$, or, after rearranging, $(\frac{7-6a}{4})^2-3(1-a)^2+[12-(3-a)^2]t^2\geq 0$. Thus, it suffices to show that $(\frac{7-6a}{4})^2-3(1-a)^2\geq 0$. But the latter is equivalent to $1+12a(1-a)\geq 0$, and the argument for Case III is complete.

Case IV. Not Case I and $1-a\leq x_1+y_1<1+a$.

In this case (5) can be improved to get

$$P\geq\frac{3}{4}pq. \tag{20}$$

We are going to distinguish three subcases, according to the number of valid inequalities among

$$\text{(v)}\quad x_2 - y_1 < 1 + a,$$
$$\text{(vi)}\quad y_2 - x_1 < 1 + a.$$

Note that if (vi) fails then $y_2 \geq 1+a+x_1 \geq 2-y_1$, so by Claim 2.3 we have $q \geq \frac{1+\sqrt{\frac{1}{2}}}{2}$. Similarly, the failure of (v) implies $p \geq \frac{1+\sqrt{\frac{1}{2}}}{2}$.

Subcase 0. None of (v), (vi) holds.

Then $P \geq \frac{3}{4}pq \geq \frac{3}{4}\left(\frac{1+\sqrt{\frac{1}{2}}}{2}\right)^2 > \frac{3}{8}$.

Subcase 1. Exactly one of (v), (vi) holds.

We assume w.l.o.g. that $x_2 - y_1 < 1 + a \leq y_2 - x_1$. Then (20) can be improved to obtain $P \geq \frac{3}{4}pq + \frac{1}{4}(1-p)q = \frac{1}{2}(p+\frac{1}{2})q$. Since $p \geq \frac{1}{2}$ and $q \geq \frac{1+\sqrt{\frac{1}{2}}}{2}$, we get $P \geq \frac{1+\sqrt{\frac{1}{2}}}{4} > \frac{3}{8}$.

Subcase 2. Both (v) and (vi) hold.

Then (20) can be improved to get

$$P \geq \frac{3}{4}pq + \frac{1}{4}(1-p)q + \frac{1}{4}p(1-q) = \frac{1}{4}p + \frac{1}{4}q + \frac{1}{4}pq. \tag{21}$$

Now, if $|x_2 - y_2| < 1 + a$ then (21) can be improved to obtain $P \geq \frac{1}{4}p + \frac{1}{4}q + \frac{1}{4}pq + \frac{1}{4}(1-p)(1-q) = \frac{1}{4} + \frac{1}{2}pq$; using (16) this implies $P \geq \frac{3}{8}$. Thus we assume that $|x_2 - y_2| \geq 1+a$, and actually, w.l.o.g., $x_2 \geq 1+a+y_2$. With the second part of (14) this implies $x_2 \geq 2+x_1 \geq 2$, which in turn implies, via (2), that $p \geq \frac{7}{8}$. Putting this and $q \geq \frac{1}{2}$ into (21) we get $P \geq \frac{29}{64}$. ∎

Remarks.

1. If we were interested only in the case $a=0$ (as in Proposition 2.1, which suffices for the balanced case), then Case IV would not exist and the proof would consist of the first three cases, with some obvious simplifications.
2. The assumption that the variances of X and Y do not exceed $\frac{1}{2}$ is essential for the result: it is easy to construct an example where the variances of X and Y are $\frac{1}{2}+\varepsilon$ and $\frac{1}{2}-\varepsilon$ respectively, and $P\{|X+Y|<1\} < \frac{3}{8}$. The other quantitative assumptions (the lower bounds on the variances, the upper bound on $|a|$) can probably be relaxed at the cost of an increased number of cases in the argument.
3. Simple examples show that the independence assumption cannot be waived, and the symmetry assumption cannot be replaced by the assumption that the expectations are 0. The assumption of finitely many values, however, can be dispensed with by means of an approximation argument. As the result of Proposition 2.1 seems interesting in its own right, we shall state its extension to arbitrary random variables (omitting the details of the proof).

Theorem 2.4. *Let X and Y be independent real valued random variables, each having a symmetric distribution with variance $\frac{1}{2}$. Then $P\{|X+Y|<1\}\geq\frac{3}{8}$.*

3. Three Special Cases

In the previous section we handled those $a=(a_1,\ldots,a_n)\in S^{n-1}$ that admit a partition of $[n]$ into I and $\bar{I}$ so that $\sum_{i\in I}a_i^2=\sum_{i\in\bar{I}}a_i^2=\frac{1}{2}$. Intuitively, such a partition will exist, at least in an approximate sense, if the summands in $\sum a_i^2$ are all small. At the other extreme, if a_1^2, say, is close to 1 then such a partition is precluded. In this section we shall handle this and other cases with a small number of summands accounting for a significant part of the sum $\sum a_i^2$. As before, we shall define random variables $X=\sum_{i\in I}\varepsilon_i a_i$ and $Y=\sum_{i\in\bar{I}}\varepsilon_i a_i$, but here I will consist of those big summands. The variances of X and Y will no longer be $\frac{1}{2}$, in general; this works to our disadvantage. On the other hand, we shall be able to exploit some additional information derived from the fact that X is a sum of a small number of terms. In each of the three cases considered, we shall prove that $P\{|X+Y|<1\}\geq\frac{3}{8}$.

We assume $a=(a_1,\ldots,a_n)\in S^{n-1}$ and, for our convenience,

$$a_1\geq a_2\geq\cdots\geq a_n>0. \tag{22}$$

Case A. $\frac{1}{2}\leq a_1<1$.

We define two random variables over the probability space $\{\pm1\}^n$ (with all $\epsilon=(\varepsilon_1,\ldots,\varepsilon_n)\in\{\pm1\}^n$ equiprobable): $X=\varepsilon_1 a_1$ and $Y=\sum_{i=2}^n\varepsilon_i a_i$. We denote $P=P\{|X+Y|<1\}$, and want to prove that $P\geq\frac{3}{8}$.

The random variable Y is independent of X, it assumes finitely many values and is symmetrically distributed with variance $1-a_1^2$. As in the proof of Proposition 2.2, we may replace Y by a random variable (that we continue to denote by Y) having these same properties and the additional property that $|Y|$ assumes just two values. Thus, we have:

$$y_1<y_2,\quad q=P\{|Y|=y_1\},\quad 1-q=P\{|Y|=y_2\}.$$

The information on the variance yields

$$qy_1^2+(1-q)y_2^2=1-a_1^2. \tag{23}$$

The random variable X remains unchanged. Thus, it assumes the values $\pm a_1$ with equal probabilities. From (23) and the convention $y_1<y_2$ it follows that $y_1\leq\sqrt{1-a_1^2}<1$. Since, by assumption, also $a_1<1$, we have $|y_1-a_1|<1$, which accounts for

$$P\geq\frac{1}{2}q. \tag{24}$$

Now, if $|y_2-a_1|<1$ then (24) can be improved to get $P\geq\frac{1}{2}q+\frac{1}{2}(1-q)=\frac{1}{2}$, so we assume that $|y_2-a_1|\geq1$. Since $a_1-y_2<a_1<1$, this means that

$$y_2\geq1+a_1. \tag{25}$$

As $a_1 \geq \frac{1}{2}$, putting (25) into (23) yields $q \geq \frac{2}{3}$. Now, if $y_1 + a_1 < 1$ then (24) can be improved to obtain $P \geq \frac{1}{2}q + \frac{1}{2}q = q \geq \frac{2}{3}$, so we assume that $a_1 \geq 1 - y_1$. Together with (25) this implies that $y_2 \geq 2 - y_1$. Using Claim 2.3 with $u = a_1$ we deduce that $q \geq \frac{1+a_1}{2} \geq \frac{3}{4}$ and therefore, by (24), $P \geq \frac{3}{8}$.

Case B. $a_1 < \frac{1}{2}$, $a_1^2 + a_2^2 \geq \frac{2}{9}$ and $a_1 - a_2 \leq \frac{1}{3}$.

We define the random variables $X = \varepsilon_1 a_1 + \varepsilon_2 a_2$ and $Y = \sum_{i=3}^{n} \varepsilon_i a_i$. We want to prove that $P = \mathbf{P}\{|X+Y| < 1\} \geq \frac{3}{8}$. As before, we maintain X and replace Y to obtain the following situation:

$$x_1 = a_1 - a_2, \quad x_2 = a_1 + a_2, \quad \mathbf{P}\{|X| = x_1\} = \mathbf{P}\{|X| = x_2\} = \frac{1}{2};$$
$$y_1 < y_2, \quad q = \mathbf{P}\{|Y| = y_1\}, \quad 1 - q = \mathbf{P}\{|Y| = y_2\}.$$

Moreover, with $u = \sqrt{a_1^2 + a_2^2}$ we have

$$qy_1^2 + (1-q)y_2^2 = 1 - u^2. \tag{26}$$

As $a_2 \leq a_1 < \frac{1}{2}$ we have $u^2 = a_1^2 + a_2^2 < \frac{1}{2}$. Thus

$$\frac{2}{9} \leq u^2 < \frac{1}{2}. \tag{27}$$

From the inequalities $a_1^2 + a_2^2 \leq (a_1 + a_2)^2 \leq 2(a_1^2 + a_2^2)$ and (27) we deduce

$$\frac{\sqrt{2}}{3} \leq x_2 \leq \sqrt{2}u < 1. \tag{28}$$

From (26) and the convention $y_1 < y_2$ it follows that $y_1 < 1$; from (28), also $x_1, x_2 < 1$. Thus, $|y_1 - x_1| < 1$ and $|y_1 - x_2| < 1$, which accounts for

$$P \geq \frac{1}{2}q. \tag{29}$$

Suppose that $y_2 - x_1 < 1$. This implies that also $y_2 - x_2 < 1$. Since $x_1 - y_2 < x_2 - y_2 < x_2 < 1$, we have then $|y_2 - x_1| < 1$ and $|y_2 - x_2| < 1$, and so (29) can be improved to get $P \geq \frac{1}{2}q + \frac{1}{2}(1-q) = \frac{1}{2}$. Thus we may assume

$$y_2 \geq 1 + x_1. \tag{30}$$

Suppose now that $y_1 \geq 1 - x_1$. We shall prove then that $q \geq \frac{3}{4}$, implying by (29) that $P \geq \frac{3}{8}$. Indeed, putting our lower bounds on y_1, y_2 and u^2 into (26) yields

$$q(1 - x_1)^2 + (1-q)(1 + x_1)^2 \leq \frac{7}{9}.$$

This implies that

$$q \geq \frac{\frac{2}{9} + 2x_1 + x_1^2}{4x_1}.$$

Since $x_1 = a_1 - a_2 \le \frac{1}{3}$, we have $x_1(1-x_1) \le \frac{2}{9}$, and so it follows that

$$q \geq \frac{x_1(1-x_1) + 2x_1 + x_1^2}{4x_1} = \frac{3}{4}.$$

Thus we may assume that $y_1 + x_1 < 1$. Hence (29) can be improved to obtain

$$P \geq \frac{1}{2}q + \frac{1}{4}q = \frac{3}{4}q. \tag{31}$$

Now, if $q \ge \frac{1}{2}$ then $P \ge \frac{3}{8}$, so we may assume that $q < \frac{1}{2}$. Using (26) and (27) this implies that $y_2 < \frac{\sqrt{14}}{3}$. This, together with (28), implies that $|y_2 - x_2| < 1$. Hence (31) can be improved to get

$$P \geq \frac{3}{4}q + \frac{1}{4}(1-q) = \frac{1}{4} + \frac{1}{2}q. \tag{32}$$

Suppose that $y_1 \ge 1 - x_2$. Then, by (28), $y_1 \ge 1 - \sqrt{2}u > 0$. Since $y_2 \ge 1$ by (30), we get from (26) that

$$q(1-\sqrt{2}u)^2 + 1 - q \leq 1 - u^2.$$

This implies that

$$q \geq \frac{u^2}{2\sqrt{2}u - 2u^2} = \frac{u}{2(\sqrt{2}-u)} \geq \frac{\frac{\sqrt{2}}{3}}{2(\sqrt{2} - \frac{\sqrt{2}}{3})} = \frac{1}{4}.$$

Hence by (32) $P \ge \frac{3}{8}$. Thus we may assume that $y_1 + x_2 < 1$. Therefore (32) can be improved to obtain

$$P \geq \frac{1}{4} + \frac{1}{2}q + \frac{1}{4}q = \frac{1}{4} + \frac{3}{4}q. \tag{33}$$

Since $y_2 \ge 1$, (26) and (27) imply that $q \ge \frac{2}{9}$. Thus (33) yields $P \ge \frac{5}{12}$.

Case C. $a_1^2 + a_2^2 < \frac{2}{9}$, $a_1^2 + a_2^2 + a_3^2 \ge \frac{1}{4}$ and $a_2 + a_3 - a_1 \ge \sqrt{\frac{6}{5}} - 1$.

We define the random variables $X = \varepsilon_1 a_1 + \varepsilon_2 a_2 + \varepsilon_3 a_3$ and $Y = \sum_{i=4}^{n} \varepsilon_i a_i$. Again, we want to prove that $P = \mathbf{P}\{|X+Y| < 1\} \ge \frac{3}{8}$. As in the proof of Proposition 2.2, an argument based on bilinearity shows that we can replace X and Y to obtain the following situation:

$$\begin{aligned} &x_1 < x_2, \quad p = \mathbf{P}\{|X| = x_1\}, \quad 1-p = \mathbf{P}\{|X| = x_2\}; \\ &y_1 < y_2, \quad q = \mathbf{P}\{|Y| = y_1\}, \quad 1-q = \mathbf{P}\{|Y| = y_2\}. \end{aligned}$$

The values x_1 and x_2 are among those numbers of the form $\pm a_1 \pm a_2 \pm a_3$ which are nonnegative. Since $a_2 + a_3 - a_1 \ge \sqrt{\frac{6}{5}} - 1$, these numbers are

$$a_2 + a_3 - a_1 \leq a_1 + a_3 - a_2 \leq a_1 + a_2 - a_3 < a_1 + a_2 + a_3,$$

and we have

$$x_1 \geq \sqrt{\frac{6}{5}} - 1. \tag{34}$$

Also, since $a_1+a_2+a_3 \le \frac{3}{2}(a_1+a_2) \le \frac{3}{2}\sqrt{2(a_1^2+a_2^2)} < \frac{3}{2}\sqrt{2\cdot\frac{2}{9}} = 1$, we have

$$x_2 < 1. \tag{35}$$

Denoting $u = \sqrt{a_1^2+a_2^2+a_3^2}$ we have

$$qy_1^2 + (1-q)y_2^2 = 1 - u^2, \tag{36}$$

and, by assumption,

$$u \ge \frac{1}{2}. \tag{37}$$

The arguments leading to formulae (29) and (30) in Case B are valid here too. Thus, we have

$$P \ge \frac{1}{2}q, \tag{38}$$

because $|y_1-x_1|<1$ and $|y_1-x_2|<1$, and we may assume

$$y_2 \ge 1 + x_1. \tag{39}$$

Together with (34) the latter implies that $y_2^2 \ge \frac{6}{5}$. Using now (36) and (37) it follows that

$$q \ge \frac{3}{8}. \tag{40}$$

If $x_1 \ge 1-y_1$ then, by (39), $y_2 \ge 2-y_1$; using Claim 2.3, (36), (37) and (38) this implies $P \ge \frac{3}{8}$. Thus we may assume that $x_1+y_1<1$. Hence (38) can be improved to obtain

$$P \ge \frac{1}{2}q + \frac{1}{2}pq. \tag{41}$$

Now, if $x_2+y_1<1$ then (41) can be improved to get $P \ge \frac{1}{2}q+\frac{1}{2}pq+\frac{1}{2}(1-p)q = q \ge \frac{3}{8}$. So we may assume

$$x_2 \ge 1 - y_1. \tag{42}$$

Similarly, if $y_2-x_2<1$ then, since $x_2-y_2<x_2<1$, we have $|y_2-x_2|<1$; therefore (41) can be improved to obtain $P \ge \frac{1}{2}q+\frac{1}{2}pq+\frac{1}{2}(1-p)(1-q) = \frac{1}{2}+p(q-\frac{1}{2}) \ge \frac{1}{2}-\frac{1}{8}p \ge \frac{3}{8}$. Thus we may assume that $y_2 \ge 1+x_2$. Coupled with (42) this implies $y_2 \ge 2-y_1$, which implies (as above) $P \ge \frac{3}{8}$.

4. Proofs of Theorems 1.2 and 1.4

Proof of Theorem 1.2. Let $a = (a_1, \dots, a_n) \in S^{n-1}$. We assume that (22) holds and $a_1 < 1$. Furthermore, we assume that none of Cases A, B or C of Section 3 occurs, since in these cases the lower bound of $\frac{3}{8}$ has been established. As Case A does not occur, we have

$$a_1 < \frac{1}{2}. \tag{43}$$

We claim now that

$$a_3 < \sqrt{\frac{1}{12}}. \tag{44}$$

Suppose not, i.e., $a_3 \geq \sqrt{\frac{1}{12}}$. Then $a_1 - a_2 \leq a_1 - a_3 < \frac{1}{2} - \sqrt{\frac{1}{12}} < \frac{1}{3}$, and so the failure of Case B means that $a_1^2 + a_2^2 < \frac{2}{9}$. Since also $a_1^2 + a_2^2 + a_3^2 \geq 3a_3^2 \geq \frac{1}{4}$ and $a_2 + a_3 - a_1 \geq 2a_3 - a_1 > 2\sqrt{\frac{1}{12}} - \sqrt{\frac{2}{9}} > \sqrt{\frac{6}{5}} - 1$, the conditions of Case C hold, contradicting our assumption.

We choose now a partition of $[n] \setminus \{3\} = \{1, 2, 4, \dots, n\}$ into I and $\overline{I}$ so that

$$M = \max\{\sum_{i \in I} a_i^2, \sum_{i \in \overline{I}} a_i^2\}$$

is the smallest possible among all such partitions. We claim that

$$M \leq \frac{1}{2}. \tag{45}$$

To show this, assume w.l.o.g. that $M = \sum_{i \in I} a_i^2$, and let $j \in I$. By the choice of the partition, $\sum_{i \in \overline{I}} a_i^2 + a_j^2 \geq M$. Therefore, $a_j^2 - a_3^2 \geq M - \sum_{i \in \overline{I}} a_i^2 - a_3^2 = M - (1 - \sum_{i \in I} a_i^2) = 2M - 1$. If $a_j \leq a_3$ for any $j \in I$, then this implies (45). If, on the other hand, $a_j > a_3$ for all $j \in I$, then $I \subseteq \{1, 2\}$, and so $M \leq a_1^2 + a_2^2 < \frac{1}{2}$, by (43).

Using this partition, we define two random variables over the probability space $\{\pm 1\}^n$ (with all $\epsilon = (\varepsilon_1, \dots, \varepsilon_n) \in \{\pm 1\}^n$ equiprobable): $X = \sum_{i \in I} \varepsilon_i a_i$ and $Y = \sum_{i \in \overline{I}} \varepsilon_i a_i$. Then

$$\begin{aligned}
P\{|\epsilon \cdot a| < 1\} &= P\{|X + Y + \varepsilon_3 a_3| < 1\} \\
&= P\{\varepsilon_3 = 1\} P\{|X + Y + \varepsilon_3 a_3| < 1 \mid \varepsilon_3 = 1\} \\
&\quad + P\{\varepsilon_3 = -1\} P\{|X + Y + \varepsilon_3 a_3| < 1 \mid \varepsilon_3 = -1\} \\
&= \frac{1}{2} P\{-1 - a_3 < X + Y < 1 - a_3\} \\
&\quad + \frac{1}{2} P\{-1 + a_3 < X + Y < 1 + a_3\} \\
&= P\{-1 + a_3 < X + Y < 1 + a_3\}
\end{aligned}$$

(the last equality holding because $X+Y$ is symmetric). It is easy to verify that the conditions of Proposition 2.2 are satisfied. In particular, (45) says that the variances are at most $\frac{1}{2}$; by (44) and (45) they are at least $1-\frac{1}{2}-\frac{1}{12}=\frac{5}{12}>6-4\sqrt{2}$; by (44) we have $a_3<\sqrt{\frac{1}{12}}<1-\sqrt{\frac{1}{2}}$. Thus, the above probability must be at least $\frac{3}{8}$. ∎

Proof of Theorem 1.4. The proof is a straightforward application of the central limit theorem for triangular arrays (see [1, Theorem 27.2]), which we quote here.

Theorem. *Suppose that for each n the random variables $X_{n1},\dots,X_{nr_n}$ are independent, with $E(X_{ni})=0$ and $s_n^2=\sum_{i=1}^{r_n}\text{Var}(X_{ni})$. Suppose further that the Lindeberg condition holds:*

$$\lim_{n\to\infty}\sum_{i=1}^{r_n}\frac{1}{s_n^2}\int_{|X_{ni}|\geq\varepsilon s_n}X_{ni}^2dP=0$$

for all $\varepsilon>0$. Then the distributions of $\frac{1}{s_n}\sum_{i=1}^{r_n}X_{ni}$ tend to the standard normal distribution.

In our application $r_n=n$ and $X_{ni}=\varepsilon_{ni}a_{ni}$. As $a_n\in S^{n-1}$ we have $s_n=1$. To verify the Lindeberg condition, suppose $\varepsilon>0$ is given. We know that $\max_{i=1,\dots,n}|a_{ni}|\to 0$, so we can find N so that $|a_{ni}|<\varepsilon$ for all $n\geq N$ and all $i=1,\dots,n$. This implies that the event $|X_{ni}|\geq\varepsilon s_n$ is null for all $n\geq N$ and all $i=1,\dots,n$, and therefore $\sum_{i=1}^{n}\frac{1}{s_n^2}\int_{|X_{ni}|\geq\varepsilon s_n}X_{ni}^2dP=0$ for $n\geq N$. Thus the central limit theorem applies and tells us that the distributions of $\epsilon_n\cdot a_n$ tend to the standard normal distribution. ∎

Acknowledgements. We are very grateful to J.R. Griggs for passing the problem and the information he had about it to us, and to Y. Peres for the geometrical interpretation suggested in the Introduction.

References

[1] P. Billingsley: *Probability and measure*, Wiley, New York, 1979.

[2] R. K. Guy: Any answers anent these analytical enigmas?, *Amer. Math. Monthly* **93** (1986), 279–281.

Ron Holzman

Dept. of Applied Mathematics
and Computer Science
The Weizmann Institute of Science
Rehovot 76100,
Israel
maholz@weizmann

Daniel J. Kleitman

Dept. of Mathematics
Massachusetts Institute of Technology
Cambridge, MA 02139,
U. S. A.
djk@math.mit.edu