On subspaces spanned by random selections of ±1 vectors

A. M. Odlyzko
AT&T Bell Laboratories
Murray Hill, NJ 07974

ABSTRACT

Suppose that vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are chosen at random from the ±1 vectors of length n. The probability that these at least are ±1 vector in the subspace (over the ????) spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_p$ that is different from the $\pm \mathbf{v}_j$ is shown to be

$$4 \binom{P}{3} \left(\frac{3}{4} \right)^n + O \left(\left(\frac{7}{10} \right)^n \right) \quad \text{as} \quad n \to \infty ,$$

uniformly for $p \leq n - 10n/(\log n)$. Moreover, the main term in this estimate is the probability that some ???? of the $\mathbf{v}_j$ contain another ±1 vector in their linear span. This result answers a question that arose in the work of Kanter and Sompolinsky on associative memories.

On subspaces spanned by random selections of ±1 vectors

A. M. Odlyzko
AT&T Bell Laboratories
Murray Hill, NJ 07974

1. Introduction

The work of Kantor and Sompolinsky [ ] on associative memories gives rise to the following question. Let the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p$ be chosen randomly from among $\{\pm 1\}^n$ (the ±1 vectors of length n). What is the probability that the subspace spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_p$??? the reals contains a ±1 vector different from $\pm\mathbf{v}_1, \ldots, \pm\mathbf{v}_p$? (The reals can be replaced by any field of characteristic zero, since the answers are the same.) Some of the results of [ ] seemed to suggest that if $p, n \to \infty$ while $p/n \to \alpha$ for some α, $0 < \alpha < 1$, then this probability might tend to 0 for $\alpha < 1 - 2/\pi$ and might tend to 1 for $\alpha > 1 - 2/\pi$. However, G. Kalai and N. Limial conjectured that this is not the case, and that in fact this probability is dominated by the probability that some 3 of the $\mathbf{v}_j$ have a linear combination that is a ±1 vector different from the $\mathbf{v}_j$. We will prove this conjecture here.

Theorem. If $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are chosen at random from $\{\pm 1\}^n$, then the probability P that the linear subspace spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_p$ over the reals contain a ±1 vector different from the $\pm\mathbf{v}_j$ equals

$$P = P_3 + O\left(\left(\frac{7}{10} \right)^n \right) \quad \text{as} \quad n \to \infty , \tag{1.1}$$

where

$$P_3 = 4 \binom{p}{3} \left(\frac{3}{4} \right)^n + O\left(\left(\frac{7}{10} \right)^n \right) \quad \text{as} \quad n \to \infty \tag{1.2}$$

is the probability that some subset of 3 of the $\mathbf{v}_j$ *has a linear combination in* $\{\pm 1\}^n$ *that differs from the* $\pm\, \mathbf{v}_j$*, and where the above estimates are uniform for*

$$p \leq n - 10\, n\, (\log n)^{-1}\ . \tag{1.3}$$

In light of the above result, the Kanter-Sompolinsky results of [ ] have now been taken to suggest a different result. Let Q be the probability that when $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are chosen at random from $\{\pm 1\}^n$, and V is the linear space spanned by the $\mathbf{v}_j$, then there is a vector $\mathbf{w}\ \varepsilon\ \{\pm 1\}^n$ more of whose neighbors (i.e., vectors $\mathbf{u}\ \varepsilon\ \{\pm 1\}^n$ that differ from $\mathbf{w}$ in one coordinate) is in V, but such that $\mathbf{w}$ is closer to V (in the sense of ordinary Euclidean distance) than any of its neighbors. The current conjecture, based on the results of [ ], is that $Q \to 0$ as $p, n \to \infty$ with $p/n \to \alpha$ for $\alpha < 1 - 2/\pi$, and $Q \to 1$ for $\alpha > 1 - 2/\pi$. Our method do not shed any light on this conjecture.

The error term $O((7/10)^n)$ in our Theorem can be substantially improved with additional effort. On the other hand, the limitation $p \leq n - 10\, n\, (\log n)^{-1}$ seems hard to improve (except for the value of the constant 10). When $p = n$, a result of Komlós [ ] implies that the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent with probability $\to 1$ as $n \to \infty$, so that $V = \{\pm 1\}^n$. It would be interesting to find out just how large p has to be so that $P \to 1$, but this problem appears very hard.

The present work is closely related to that of Komlós. The distribution of determinants of matrices whose entries are drawn from some common distribution is only known in a few special cases [ ], such as when they are all normal. The problem of determinants of random ±1 matrices (or of (0,1)-matrices, since there is a well-known correspondence between the two problems) has been of substantial interest for a long time [ ]. Komlós [ ] was the first one to show that the probability of a random $n \times n$ ±1 matrix being singular $\rightarrow 0$ as $n \rightarrow \infty$. We later [ ] extended this result to the case where the entries are drawn from any non-degenerate distribution. Finally, in [ ], he developed a simplified method that enabled him to show that the probability of a random $n \times n$ ±1 matrix being singular is $O(n^{-1/2})$ as $n \rightarrow \infty$. (It is conjectured that this probability is $O(n^2 2^{-n})$, so that such matrices are singular primarily when two rows or columns are equal to each other or the negations of each other.) Parts of our proof use techniques very similar to those of [ ].

There are many other open problems about 0,1 or ±1 random variables. For example, L. Babai has conjectured that the characteristic polynomials of adjacency matrices of random undirected graphs (i.e., of random symmetric (0,1)-matrices with 0's on the diagonal) are irreducible as the dimension $\rightarrow \infty$. (If true, this would say that testing for graph isomorphism is easy most of the time.) This author has also conjectured that polynomials of degree n with coefficients 0,1 and constant term 1 are irreducible with probability $\rightarrow 1$ as $n \rightarrow \infty$.

Some additional results on convex combinations of vertices of an n-cube and linear subspaces can be found in [ ].

2. Proof of Theorem

Let P_m denote the probability that there is a linear combination of some m out of the p random vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p \in \{\pm 1\}^n$ which is in $\{\pm 1\}^n$, and such that all m coefficients in this combination are nonnegative. We will estimate P_3 and show that $P_2, P_4, P_5, \ldots, P_P$ are negligible.

We start with the bounds for $P_5, P_6, \ldots$. Our basic tool will be the following lemma, which was proved by Erdös [ ], but is usually referred to as the Littlewood-Offord Lemma after the people who first raised the problem and proved a weaker form of the result [ ]. (The most general result of this type is due to Kleitman [ ].)

Lemma 2.1. Suppose that $x_1, \ldots, x_m \in \mathbb{R} \setminus \{0\}$, $y \in \mathbb{R}$. *Then*

$$\left| \left\{ (\varepsilon_1, \ldots, \varepsilon_m) : \varepsilon_i = \pm 1 \text{ for all } i, \ \sum_i \varepsilon_i x_i = y \right\} \right| \le \binom{m}{\lfloor m/2 \rfloor} . \qquad (2.1)$$

We now use Lemma 2.1 to prove the following result.

Proposition 2.2. If

$$5 \le m \le p \le n - 10\, n\, (\log n)^{-1} ,$$

then

$$P_m \le (0.69)^n \quad \text{as} \quad n \to \infty \qquad (2.2)$$

for n sufficiently large.

Proof of Proposition 2.2. We clearly have

$$P_m \leq \binom{P}{m} R_m , \tag{2.3}$$

where R_m is the probability that a random $m \times n$ ± 1 matrix M will have some combination of its m rows with all coefficients $\neq 0$ in $\{\pm 1\}^n$. Denote the rows of M by $\mathbf{w}_1 , \ldots , \mathbf{w}_m$. Suppose that $0 < q < n - m$, and assume that the first $m + q$ columns of M have ????? m. If columns $j_1 < \ldots < j_m \leq m + q$ of M are linearly independent, then for each choice of $\alpha_1 , \ldots , \alpha_m \, \varepsilon \, \{\pm 1\}$, there will be a unique set of coefficients $x_1 , \ldots , x_m$ with the j_g-th coordinate of $x_1 \mathbf{w}_1 + \ldots + x_m \mathbf{w}_m$ equal to α_g. Thus there will be at most 2^m sets $x_1 , \ldots , x_m \, \varepsilon \, \mathbb{R} \setminus \{0\}$ with the property that the first $m + q$ coordinates of $x_1 \mathbf{w}_1 + \ldots + x_m \mathbf{w}_m$ are all ± 1. Consider now a fixed choice of $x_1 , \ldots , x_m$. The probability that the j-th coordinate of $x_1 \mathbf{w}_1 + \ldots + x_m \mathbf{w}_m$ equals 1 for $m + q < j \leq n$, as the j-th column of M varies, is

$$2^{-m} \binom{m}{\lfloor m/2 \rfloor}$$

by the Littlewood-Offord lemma, and similarly for the probability that this coordinate equals -1. Since columns j, $m + q < j \leq n$, are independent of each other, we obtain

$$R_m \leq 2^m \binom{m+q}{q} \left[2^{-m} \binom{m}{\lfloor m/2 \rfloor} \right]^{n-m-q} + Q_{m,q} , \tag{2.4}$$

where $Q_{m,q}$ is the probability that a random $m \times (m + q)$ ± 1 matrix has rank $< m$.

We next bound $Q_{m,q}$. We have

$$Q_{m,q} \leq \sum_{k=1}^{m-1} (m-k) \binom{m}{k} \binom{m+q}{k} Q_{m,q,k} , \tag{2.5}$$

where $Q_{m,q,k}$ is the probability that a random $m \times (m+q)$ ± 1 matrix will have the property that every k of its rows are linearly independent, the upper left $k \times k$ submatrix has rank k, and the $(k+1)$-st row is linearly dependent on the first k rows. Given a matrix satisfying the above properties, the rows $\mathbf{u}_1, \ldots, \mathbf{u}_{k+1}$ of the upper left $(k+1) \times k$ submatrix determine unique nonzero coefficients $x_1, \ldots, x_{k+1}$ such that $\sum_{i=1}^{k+1} x_i \mathbf{u}_i = (0, \ldots, 0)$. Given $x_1, \ldots, x_{k+1}$, the probability that this same relation will also hold in columns $k+1, \ldots, m+q$ is

$$\leq \left[2^{-k-1} \binom{k+1}{\lfloor (k+1)/2 \rfloor} \right]^{m+q-k} , \tag{2.6}$$

and so this is a bound for $Q_{m,q,k}$. Therefore, combining (2.5) and (2.6), we obtain

$$Q_{m,q} \leq m \sum_{k=1}^{m-1} \binom{m}{k} \binom{m+q}{k} \left[2^{-k-1} \binom{k+1}{\lfloor (k+1)/2 \rfloor} \right]^{m+q-k} ,$$

and so, by (2.3) and (2.4),

$$\begin{aligned} P_m \leq\ & 2^m \binom{P}{m} \binom{m+q}{q} \left[2^{-m} \binom{m}{\lfloor m/2 \rfloor} \right]^{n-m-q} \\ & + m \binom{P}{m} \sum_{k=1}^{m-1} \binom{m}{k} \binom{m+q}{k} \left[2^{-k-1} \binom{k+1}{\lfloor (k+1)/2 \rfloor} \right]^{m+q-k} . \end{aligned} \tag{2.7}$$

For $5 \leq m \leq n/1000$, we select

$$q = \left\lfloor \frac{65n}{100} \right\rfloor .$$

We find that

$$2^m \binom{P}{m} \binom{m+q}{q} \le 2^m \binom{n}{m}^2 \le 2^{3n/1000} ,$$

$$\left[2^{-m} \binom{m}{\lfloor m/2 \rfloor} \right]^{n-m-q} \le \left(\frac{5}{16} \right)^{0.348n} ,$$

$$\binom{P}{m} \binom{m}{k} \binom{m+q}{k} \le 2^m \binom{P}{m} \binom{m+q}{m} \le 2^{3n/1000} ,$$

and for $1 \le k \le m-1$,

$$\left[2^{-k-1} \binom{k+1}{\lfloor (k+1)/2 \rfloor} \right]^{m+q-k} \le 2^{-q} ,$$

so that in this range

$$P_m \le O(0.67^n) \quad \text{as} \quad n \to \infty . \tag{2.8}$$

We next consider $n/1000 < m \le p < n$. This time we use the inequality

$$P_m \le \binom{P}{m} R_m' + Q_{p,q} , \tag{2.9}$$

where R_m' is the probability that a random $m \times n$ ± 1 matrix has some nonzero combination of its rows in $\{\pm 1\}^n$, and that its first $p+q$ columns have rank m. Then, by the previous argument,

$$R_{m'} \le 2^{p+q} \binom{p+q}{m} \left[2^{-m} \binom{m}{\lfloor m/2 \rfloor} \right]^{n-p-q}$$

$$\le 2^{2n} \left[2^{-m} \binom{m}{\lfloor m/2 \rfloor} \right]^{n-p-q} \le 2^{2n} \left[8 n^{-1/2} \right]^{n-p-q} \qquad (2.10)$$

for large enough n. On the other hand,

$$Q_{p,q} \le n \sum_{k=1}^{p-1} \binom{P}{k} \binom{p+q}{k} \left[2^{-k-1} \binom{k+1}{\lfloor (k+1)/2 \rfloor} \right]^{p+q-k} .$$

Now for large n,

$$\sum_{k=\lfloor n^{1/2} \rfloor + 1}^{p-1} \binom{P}{k} \binom{p+q}{k} \left[2^{-k-1} \binom{k+1}{\lfloor (k+1)/2 \rfloor} \right]^{p+q-k}$$

$$\le n\, 2^{2n}\, n^{-(p+q-k)4} \le n\, 2^{2n}\, n^{-n/10000} \le 2^{-n} .$$

At the same time, for some positive constant C,

$$\sum_{k=1}^{\lfloor n^{1/2} \rfloor} \binom{P}{k} \binom{p+q}{k} \left[2^{-k-1} \binom{k+1}{\lfloor (k+1)/2 \rfloor} \right]^{p+q-k}$$

$$\leq \sum_{k=1}^{\lfloor n^{1/2} \rfloor} \frac{n^{2k}}{(k!)^2} \left[2^{-k-1} \binom{k+1}{\lfloor (k+1)/2 \rfloor} \right]^{p+q-n^{1/2}}$$

$$\leq e^{C n^{1/2} \log n} \sum_{k=1}^{\lfloor n^{1/2} \rfloor} \left[2^{-k-1} \binom{k+1}{\lfloor (k+1)/2 \rfloor} \right]^{p+q}$$

$$\leq n^{1/2} e^{C n^{1/2} \log n} 2^{-p-q} .$$

Combining all these estimates we obtain

$$P_m \leq \binom{P}{m} 2^{2n} \left[8n^{-1/2} \right]^{n-p-q} + 2^{-n} + e^{n^{2/3}} 2^{-p-q}$$

$$\leq 2^{3n} \left[8n^{-1/2} \right]^{n-p-q} + e^{n^{2/3}} 2^{1-p-q} , \qquad (2.11)$$

valid for sufficiently large n and $n/1000 < m \leq p < n$. We now select q so that

$$n - p - q = \left\lfloor \frac{7n \log 2}{\log n} \right\rfloor ,$$

and obtain the claim of Proposition 2.2. ■

We now proceed to consider P_2, P_3, and P_4. If $\mathbf{v}$, $\mathbf{w}$ ε $\{\pm 1\}^n$, then the only way to have $\alpha \mathbf{v} + \beta \mathbf{w}$ ε $\{\pm 1\}^n$ for $\alpha\beta \neq 0$ is if $\mathbf{v} = \pm \mathbf{w}$, to an even that has probability 2^{1-n}. Therefore

$$P_2 = O\left(n^2 \, 2^{-n} \right) \quad \text{as} \quad n \to \infty . \qquad (2.12)$$

Proposition 2.3. We have, for $3 \le p \le n$,

$$P_3 = 4 \binom{P}{3} \left(\frac{3}{4}\right)^n + O\left(4^P \left(\frac{5}{8}\right)^n\right) \quad \text{as} \quad n \to \infty . \tag{2.13}$$

Proof. By (2.3), $P_3 \le \binom{P}{3} R_3$. Since multiplying any collection of rows or columns of a ± 1 matrix by ± 1's does not change the property that some ± 1 vector is in the span of rows of the matrix, we have

$$R_3 = 2^{2-2n} N , \tag{2.14}$$

where N is the number of $3 \times n$ ± 1 matrices M with rows $\mathbf{v}_1, \ldots, \mathbf{v}_3$ for which $\mathbf{v}_1 = (1,1, \ldots, 1)$, the first column equals $(1,1,1)^T$, and such that for some α, β, γ with $\alpha\beta\gamma \ne 0$, we have $\alpha\mathbf{v}_1 + \beta\mathbf{v}_2 + \gamma\mathbf{v}_3 \in \{\pm 1\}^n$. We will estimate the number of such matrices M.

Let $\mathbf{v}_m = (v_{m1}, \ldots, v_{mn})$, and suppose that $\alpha v_{11} + \beta v_{21} + \gamma v_{31} = u$. If M contains a column of the form $(1, -1, -1)^T$ or $(1, 1, -1)^T$, say in the r-th position, and $\alpha v_{1r} + \beta v_{2r} + \gamma v_{3r} = X$, then $x = -u$, since if $x = u$, subtracting this equation from $\alpha v_{11} + \beta v_{21} + \gamma v_{31} = u$ would give $\beta = 0$ or $\gamma = 0$, which is impossible. Similarly, if M contains the column $(1, -1, -1)^T$, say in the r-th position, and another column, say the g-th one, equals $(1, -1, 1)^T$ or $(1, 1, -1)^T$, then $\alpha v_{1r} + \beta v_{2r} + \gamma v_{3r} = u$. Therefore we cannot have all 4 possible columns appearing in M, since then we would have the 4 equations

$$\alpha v_{11} + \beta v_{21} + \gamma v_{31} = u ,$$

$$\alpha v_{1r} + \beta v_{2r} - \gamma v_{3r} = -u ,$$

$$\alpha v_{1s} - \beta v_{2s} + \gamma v_{3s} = -u ,$$

$$\alpha v_{1t} - \beta v_{2t} - \gamma v_{3t} = u ,$$

and adding them shows that $\alpha = 0$, which is a contradiction. On the other hand, for any selection of 3 out of the 4 possible columns of M (always including the first column $(1, 1, 1)^T$), the matrices consisting of precisely those columns, will have the required property, since we will obtain a nonsingular system of 3 equations in 3 unknowns. For any particular choice of 3 columns to appear in M, we will have 3^{n-1} choices of M. There are 3 possible choices of 3 out of 4 columns (since $(1, 1, 1)^T$ always has to be included). If only 2 different columns appear in M, then some two of $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ an equal, and there are $O(2^n)$ such matrices M. Hence we conclude that

$$N = 3^n + O(2^n) , \tag{2.15}$$

and so

$$R_3 = 4(3/4)^n + O(2^{-n}) . \tag{2.16}$$

By the analysis above,

$$P_3 \le 4 \binom{P}{3} \left(\frac{3}{4} \right)^n + O\left(P^3 \, 2^{-n} \right) . \tag{2.17}$$

To get a lower bound for P_3, consider the probability that 2 sets of 3 vectors each, $\mathbf{v}_{i_1}$, $\mathbf{v}_{i_2}$, $\mathbf{v}_{i_3}$ and $\mathbf{v}_{j_1}$, $\mathbf{v}_{j_2}$, $\mathbf{v}_{j_3}$ simultaneously have linear combinations with nonzero coefficients that are in $\{\pm 1\}^n$. If $I = \{i_1, i_2, i_3\} \cap \{j_1, j_2, j_3\}$ is empty or has exactly one element, this probability is R_3^2. If I contains 2 elements, then this probability is 2^{3-3n} times N_2, the number of $4 \times n$ ± 1 matrices with the first row and column 1 and with the property that the submatrices formed by deleting the third or the fourth row have at most 3 distinct columns. If we let k denote the number of 1's, in the second now, we obtain

$$\leq 2^{\max(k,\, n-k) + 1}$$

choices for each of the third and fourth rows, so

$$N_2 \leq 4 \sum_{k=1}^{n} \binom{n}{k} 4^{\max(k,\, n-k)}$$

$$\leq 8 \sum_{k=0}^{n} \binom{n}{k} 4^{k} = 8 \cdot 5^{n} . \tag{2.18}$$

Therefore the probability of finding 2 sets of 3 vectors, each of which gives the desired combination, is $O(P^4 (5/8)^n)$, and therefore we obtain the estimate (2.13) of Proposition 2.3. ■

Proposition 2.4. If $4 \leq p \leq n$, *then*

$$P_4 = O\left(P^4 \left(\frac{5}{8} \right)^n \right) \quad \text{as} \quad n \to \infty . \tag{2.19}$$

Sketch of proof. The proof of this result uses the same ideas as that of Proposition 2.3, but is considerably easier, since only a weak upper bound is required. We reduce the problem of bounding P_4 to that of counting the number of $4 \times n$ ± 1 matrices M with rows $\mathbf{v}_1, \ldots, \mathbf{v}_4$ such that $\mathbf{v}_1 = (1, 1, \ldots, 1)$, the first column of μ is $(1, 1, 1, 1)^T$, and for some nonzero $\alpha, \beta, \gamma, \delta$, $\alpha \mathbf{v}_1 + \beta \mathbf{v}_2 + \gamma \mathbf{v}_3 + \delta \mathbf{v}_4 \ \varepsilon \ \{\pm 1\}^n$. A short argument then shows that such a matrix cannot have more than 5 distinct columns, which then immediately yields (2.19). (A more careful argument shows that such a matrix cannot have more than 4 distinct columns, which then gives $P_4 = O(P^4 2^{-n})$.) ∎

The Theorem easily follows from all the estimates that have been obtained.

Acknowledgements

The author thanks H. Sompolinsky for informing him of the problem and helpful conversations.

References

1. T. W. Anderson, *Introduction to Multivariate Statistical Analysis*, Wiley, 1958.

2. P. Erdös, On a lemma of Littlewood and Offord, Bull. Am. Math. Soc. *5* (1945), 898-902.

3. G. E. Forsythe and J. W. Tukey, The extent of *n* random unit vectors, (Abstract), *Bull. Am. Math. Soc. 58* (1952), 502.

4. Z. Fü redi, Random polytopes in the *d*-dimensional cube, *Discr. Comp. Geometry*, to appear.

5. I. Kanter and H. Sompolinsky, Associative recall of memory without errors, to appear.

6. D. J. Kleitman, On a lemma of Littlewood and Offord on the distribution of linear combinations of vectors, *Advances Math. 5* (1970), 155-157.

7. J. Komlós, On the determinant of (0,1) matrices, *Studia Sci. Math. Hungar. 2* (1967), 7-21.

8. J. Komlós, On the determinant of random matrices, *Studia Sci. Math. Hungar. 3* (1968), 387-399.

9. J. Komlós, unpublished manuscript (1977).

10. J. E. Littlewood and C. Offord, On the number of real roots of a random algebraic equation. III. *Mat. Sbornik 12* (1943), 277-286.

11. N. Metropolis, Spectra of determinant values in (0,1) matrices, pp. 271-276 in *Computers in Number Theory*, A. O. L. Atkin and B. J. Birch, eds., Academic Press, 1971.

12. H. Nyquist, S. O. Rice, and J. Riordan, The distribution of random determinants, *Quart. Appl. Math. 12* (1954), 97-104.

13. A. M. Odlyzko, On the ranks of some (0,1)-matrices with constant row sums, *J. Australian Math. Soc. A 31* (1981), 193-201.

14. A. Prékopa, On random determinants I, *Studia Sci. Math. Hungar. 2* (1967), 125-132.

15. G. Szekeres and P. Turan, On an extremal problem in the theory of determinants, *Math. Naturwiss. Anz. Ungar. Akad. Wiss. 56* (1937), 796-806 (in Hungarian).

16. P. Turan, On a problem in the theory of determinants, *Acta Math. Sinica 5* (1955), 411-423 (in Chinese).