Selected Mathematical and Algorithmic Applications of Linear Algebra

Jiří Matoušek

Version 10/V/2009

Introduction

Some years ago I started gathering nice applications of linear algebra. Here are some pieces from the collection. The applications belong mostly to the main fields of my mathematical interests—combinatorics, geometry, and computer science. Most of them are mathematical, in proving theorems, and some include clever ways of computing things, i.e., algorithms. The appearance of linear-algebraic methods is often unexpected.

I assume background of basic linear algebra, a bit of familiarity with polynomials, and some graph-theoretical and geometric terminology.

I wanted each section to be essentially self-contained. This is kind of opposite to a typical mathematical textbook, where material is developed gradually, and if one wants to make sense of something on page 123, one usually has to understand the previous 122 pages, or with some luck, suitable 38 pages.

After several sections the reader may spot certain common patterns in the presented proofs, which could be discussed at great length, but I have decided to leave out any general accounts on linear algebraic methods.

Nothing in this text is original, and some of the examples are rather well known and appear in many publications (including, in few cases, other books of mine). Several general reference books are listed below. I've also added references to the original sources where I could find them. However, I've kept the historical notes at a minimum and I've put only a limited effort into tracing the origins of the ideas (many apologies to authors whose work is quoted badly or not at all—I will be glad to hear about such cases).

I would appreciate to learn about mistakes, suggestions of how to improve the exposition, and candidates for additions to the collection.

Further reading. An excellent textbook on linear-algebraic methods in combinatorics is

> L. Babai, P. Frankl: *Linear Algebra Methods in Combinatorics (Preliminary version 2).* Department of Computer Science, The University of Chicago, 1992.

Unfortunately, it has never been published officially and it can be obtained, with some effort, as lecture notes of the University of Chicago. It contains several of the topics discussed here, a lot of other material in a similar spirit, and a very nice exposition of some parts of linear algebra.

Algebraic graph theory is treated, e.g, in the books

N. Biggs: *Algebraic Graph Theory*, 2nd edition, Cambridge Univ. Press, Cambridge, 1993

and

C. Godsil, G. Royle: *Algebraic Graph Theory*, Springer, New York, NY 2001.

Contents

1 Are These Distances Euclidean?

Can we find three points $\mathbf{p}, \mathbf{q}, \mathbf{r}$ in the plane whose mutual Euclidean distances are all 1's? Of course we can—the vertices of an equilateral triangle.

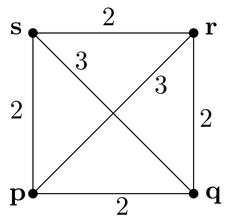
Can we find $\mathbf{p}, \mathbf{q}, \mathbf{r}$ with $\|\mathbf{p}-\mathbf{q}\| = \|\mathbf{q}-\mathbf{r}\| = 1$ and $\|\mathbf{p}-\mathbf{r}\| = 3$? No, since the direct path from $\mathbf{p}$ to $\mathbf{r}$ can't be longer than the path via $\mathbf{q}$; these distances violate the **triangle inequality**, which in the Euclidean case tells us that

$$\|\mathbf{p}-\mathbf{r}\| \leq \|\mathbf{p}-\mathbf{q}\| + \|\mathbf{q}-\mathbf{r}\|$$

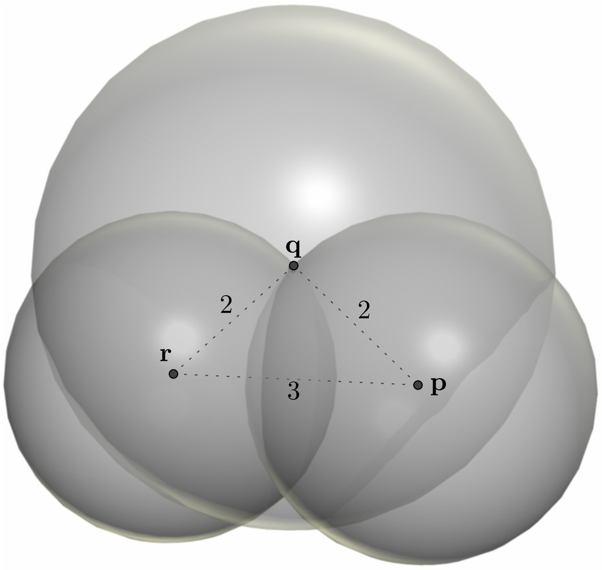
for any three points $\mathbf{p}, \mathbf{q}, \mathbf{r}$ (in any Euclidean space).

It turns out that the triangle inequality is the *only* obstacle for three points: Whenever nonnegative real numbers x, y, z satisfy $x \leq y + z$, $y \leq x + z$, and $z \leq x + y$, then there are $\mathbf{p}, \mathbf{q}, \mathbf{r} \in \mathbb{R}^2$ such that $\|\mathbf{p}-\mathbf{q}\| = x$, $\|\mathbf{q}-\mathbf{r}\| = y$, and $\|\mathbf{p}-\mathbf{r}\| = z$. These are well known conditions for the existence of a triangle with given side lengths.

What about prescribing distances for four points? We need to look for points in $\mathbb{R}^3$; that is, we ask for a tetrahedron with given side lengths. Here the triangle inequality is a necessary, but not sufficient condition. For example, if we require the distances as indicated in the picture,



then there is no violation of a triangle inequality, yet there are no corresponding $\mathbf{p}, \mathbf{q}, \mathbf{r}, \mathbf{s} \in \mathbb{R}^3$. Geometrically, if we construct the triangle $\mathbf{pqr}$, then the spheres around $\mathbf{p}, \mathbf{q}, \mathbf{r}$ that would have to contain $\mathbf{s}$ have no common intersection:



This is a rather ad-hoc argument. Linear algebra provides a very elegant characterization of the systems of numbers that can appear as Euclidean distances, using the notion of a *semidefinite matrix*. The characterization works for any number of points; if there are $n+1$ points, we want them to live in $\mathbb{R}^n$. The formulation becomes more convenient to state if we number the desired points starting from 0:

Theorem. *Let m_{ij}, $i,j = 0,1,\ldots,n$ be nonnegative real numbers with $m_{ij} = m_{ji}$ for all i,j and $m_{ii} = 0$ for all i. Then points $\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_n \in \mathbb{R}^n$ with $\|\mathbf{p}_i - \mathbf{p}_j\| = m_{ij}$ for all i,j exist if and only if the $n \times n$ matrix G with*

$$g_{ij} = \frac{1}{2}\left(m_{0i}^2 + m_{0j}^2 - m_{ij}^2\right)$$

is positive semidefinite.

Let us note that the triangle inequality doesn't appear explicitly in the theorem—it is hidden in the condition of positive semidefiniteness (you may want to check this for the case $n = 2$).

The proof of the theorem relies on the following characterization of positive semidefinite matrices.

Fact. *An real symmetric $n \times n$ matrix A is positive semidefinite iff there exists an $n \times n$ real matrix X such that $A = X^T X$.*

Reminder of a proof. If $A = X^T X$, then for every $\mathbf{x} \in \mathbb{R}^d$ we have $\mathbf{x}^T A\mathbf{x} = (X\mathbf{x})^T(X\mathbf{x}) = \|X\mathbf{x}\|^2 \geq 0$, and so A is positive semidefinite.

Conversely, every real symmetric square matrix A is *diagonalizable*, i.e., it can be written as $A = T^{-1}DT$ for a nonsingular $n \times n$ matrix T and a diagonal matrix D (with the eigenvalues of A on the diagonal). Moreover, an inductive proof of this fact even yields T *orthogonal*, i.e., such that $T^{-1} = T^T$ and thus $A = T^T DT$. Then we can set $X := RT$, where $R = \sqrt{D}$ is the diagonal matrix having the square roots of the eigenvalues of A on the diagonal; here we use the fact that A, being positive semidefinite, has all eigenvalues nonnegative.

It turns out that one can even require X to be upper triangular, and in such case one speaks about a *Cholesky factorization* of A. $\square$

Proof of the theorem. First we check necessity; that is, if $\mathbf{p}_0, \ldots, \mathbf{p}_n \in \mathbb{R}^n$ are given points and $m_{ij} := \|\mathbf{p}_i - \mathbf{p}_j\|$, then G as in the theorem is positive semidefinite.

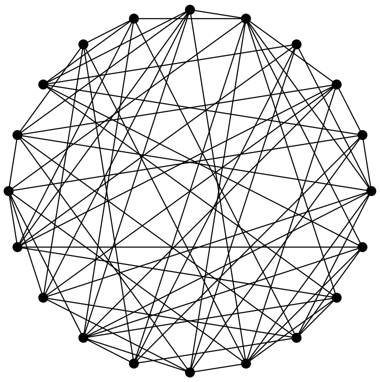
For this, we need the *cosine theorem*, with tells us that $\|\mathbf{x}-\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\langle \mathbf{x} \,|\, \mathbf{y}\rangle$ for any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Thus, if we define $\mathbf{x}_i := \mathbf{p}_i - \mathbf{p}_0$, $i = 1, 2, \ldots, n$, we get that $\langle \mathbf{x}_i \,|\, \mathbf{x}_j\rangle = \frac{1}{2}(\|\mathbf{x}_i\|^2 + \|\mathbf{x}_j\|^2 - \|\mathbf{x}_i - \mathbf{x}_j\|^2) = g_{ij}$. So G is the **Gram matrix** of the vectors $\mathbf{x}_i$, we can write $G = X^T X$, and hence G is positive semidefinite.

Conversely, if G is positive semidefinite, we can decompose it as $G = X^T X$ for some $n \times n$ real matrix X. Then we let $\mathbf{p}_i \in \mathbb{R}^n$ be the ith column of X for $i = 1, 2, \ldots, n$, while $\mathbf{p}_0 := \mathbf{0}$. Reversing the above calculation, we arrive at $\|\mathbf{p}_i - \mathbf{p}_j\| = m_{ij}$, and the proof is finished. $\square$

The theorem solves the question for points living in $\mathbb{R}^n$, the largest dimension one may ever need for $n + 1$ points. One can also ask when the desired points can live in $\mathbb{R}^d$ with some given d, say $d = 2$. An extension of the above argument shows that the answer is positive iff $G = X^T X$ for some matrix X of rank at most d.

2 Where's the Triangle?

Does a given graph contain a triangle, i.e., three vertices u, v, w, every two of them connected by an edge? This question is not entirely easy to answer for graphs with many vertices and edges. For example, where is a triangle in this graph?



An obvious algorithm for finding a triangle inspects every triple of vertices, and thus it needs roughly n^3 operations for an n-vertex graph (there are $\binom{n}{3}$ triples to look at, and $\binom{n}{3}$ is approximately $n^3/6$ for large n). Is there a significantly faster method?

There is, but surprisingly, the only known approach for breaking the n^3 barrier is algebraic, based on fast matrix multiplication.

To explain it, we assume for notational convenience that the vertex set of the given graph G is $\{1, 2, \ldots, n\}$, and we define the **adjacency matrix** of G as the $n \times n$ matrix A with

$$a_{ij} = \begin{cases} 1 & \text{if } i \neq j \text{ and } \{i, j\} \in E(G), \\ 0 & \text{otherwise.} \end{cases}$$

The key insight is to understand the square $B := A^2$. By the definition of matrix multiplication we have $b_{ij} = \sum_{k=1}^{n} a_{ik}a_{kj}$, and

$$a_{ik}a_{kj} = \begin{cases} 1 & \text{if the vertex } k \text{ is adjacent to both } i \text{ and } j, \\ 0 & \text{otherwise.} \end{cases}$$

So b_{ij} counts the number of common neighbors of i and j.

Finding a triangle is equivalent to finding two adjacent vertices i, j with a common neighbor k. So we look for two indices i, j such that both $a_{ij} \neq 0$ and $b_{ij} \neq 0$.

To do this, we need to compute the matrix $B = A^2$. If we do the matrix multiplication according to the definition, we need about n^3 arithmetic operations and thus we save nothing compared to the naive method of inspecting all triples of vertices.

However, ingenious algorithms are known that multiply $n \times n$ matrices asymptotically faster. The first one, due to Strassen, needs roughly $n^{2.807}$ arithmetic operations. For readers who haven't seen it I can recommend looking it up (Wikipedia?); it is based on a simple but very ingenious trick.

The infimum of numbers ω for which there exists a matrix multiplication algorithm using $O(n^\omega)$ operations is often referred to as the *exponent of matrix multiplication*. Its value is unknown (the common belief is that it equals 2); the current best upper bound is roughly 2.376.

Many computational problems are known where fast matrix multiplication brings asymptotic speedup. Finding triangles is among the simplest of them, and several other, more sophisticated algorithms of this kind appear later in this collection.

Remarks. The described method for finding triangles is the fastest known for *dense* graphs, i.e., graphs that have relatively many edges compared to the number of vertices. Another ingenious algorithm, for example, can detect a triangle in time $O(m^{2\omega/(\omega+1)})$, where m is the number of edges (we will not discuss that algorithm here).

One can try to use similar methods for detecting subgraphs other than the triangle; there is an extensive literature concerning this problem. For example, a cycle of length 4 can be detected in time $O(n^2)$, much faster than a triangle!

Sources. A. Itai, M. Rodeh: Finding a minimum circuit in a graph, *SIAM J. Comput.*, 7,4(1978) 413–423.

Among the numerous papers dealing with fast detection of a fixed subgraph in a given graph, we mention

T. Kloks, D. Kratsch, H. Müller: Finding and counting small induced subgraphs efficiently, *Inform. Process. Lett.* 74,3-4(2000) 115--121.

which can be used as a starting point for further explorations of the topic.

The first "fast" matrix multiplication algorithm is due to

V. Strassen: Gaussian elimination is not optimal, *Numer. Math.* 13(1969) 354–356.

The asymptotically fastest known matrix multiplication algorithm is from

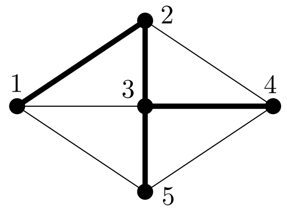
D. Coppersmith, S. Winograd, Matrix multiplication via arithmetic progressions, *J. Symbolic Computation* 9(1990), 251-280.

An interesting new method, which provides similarly fast algorithms in a different way, appeared in

H. Cohn, R. Kleinberg, B. Szegedy, C. Umans: Group-theoretic algorithms for matrix multiplication, in *Proc. 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2005, 379–388.

3 Counting Spanning Trees

A **spanning tree** of a graph G is a connected subgraph of G that has the same vertex set as G and contains no cycles. The next picture shows a 5-vertex graph with one of the possible spanning trees marked thick.



What is the number $\kappa(G)$ of spanning trees of a given graph G? Here is the answer:

Theorem (Matrix-tree theorem). *Let G be a graph on the vertex set $\{1, 2, \dots, n\}$, and let L be the* **Laplace matrix** *of G, i.e., the $n \times n$ matrix whose entry ℓ_{ij} is given by*

$$\ell_{ij} := \begin{cases} \deg(i) & \textit{if } i = j, \\ -1 & \textit{if } \{i, j\} \in E(G), \\ 0 & \textit{otherwise,} \end{cases}$$

where $\deg(i)$ is the number of neighbors (degree) of the vertex i in G. Let L^- be the $(n-1) \times (n-1)$ matrix obtained by deleting the last row and last column of L. Then

$$\kappa(G) = \det(L^-).$$

For example, for the G in the picture we have

$$L = \begin{pmatrix} 3 & -1 & -1 & 0 & -1 \\ -1 & 3 & -1 & -1 & 0 \\ -1 & -1 & 4 & -1 & -1 \\ 0 & -1 & -1 & 3 & -1 \\ -1 & 0 & -1 & -1 & 3 \end{pmatrix}, \quad L^- = \begin{pmatrix} 3 & -1 & -1 & 0 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 4 & -1 \\ 0 & -1 & -1 & 3 \end{pmatrix},$$

and $\det(L^-) = 45$ (can you check the number of spanning trees directly?).

I still remember my amazement when I saw the matrix-tree theorem for the first time. I believe it remains one of the most impressive uses of determinants. It is rather well known, but the forthcoming proof, hopefully, is not among those presented most often, and moreover, it resembles the proof of the *Gessel–Viennot lemma*, which is a powerful general tool in enumeration.

Proof. We begin with the usual expansion of $\det(L^-)$ according to the definition of a determinant as a sum over all permutations of $\{1, 2, \ldots, n-1\}$:

$$\det(L^-) = \sum_{\pi} \operatorname{sgn}(\pi) \prod_{i=1}^{n-1} \ell_{i,\pi(i)}. \tag{1}$$

Here $\operatorname{sgn}(\pi)$ is the sign of the permutation π, which can be defined as $(-1)^t$, where t is an integer such that one can obtain π from the identity permutation by t transpositions.

We now write each diagonal entry ℓ_{ii} of L^- in (1) as a sum of 1's, e.g., instead of 3 we write $(1+1+1)$. Then we multiply out the parentheses, so that each of the products in (1) is further expanded as a sum of products, where the factors in the products are only 1's and -1's. Let us call the resulting sum the **superexpansion** of $\det(L^-)$.

Graphically, each nonzero term in the superexpansion is obtained by selecting one 1 or -1 in each row and in each column of L^-. One of such selections is marked by circling the selected items:

$$\begin{pmatrix} 1+1+1 & -1 & -1 & 0 \\ -1 & 1+1+1 & -1 & -1 \\ -1 & -1 & 1+1+1+1 & -1 \\ 0 & -1 & -1 & 1+1+1 \end{pmatrix}.$$

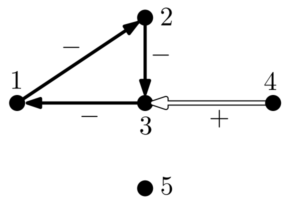
The sign of such a term is $(-1)^m \operatorname{sgn}(\pi)$, where m is the number of -1's factors and π is the corresponding permutation. In the example, $m = 3$

and $\pi = (2,3,1,4)$, with sign $+1$, so the term contributes a -1 to the superexpansion.

Next, we associate a combinatorial object with each term in the superexpansion. The object is a *directed graph* (or *digraph* for short) on the vertex set $\{1, 2, \ldots, n\}$, and moreover, each directed edge is either *positive* or *negative*. The rules for creating this signed digraph are as follows:

- If there is a circled -1 in row i and column j, make a *negative* directed edge from i to j.
- If the kth "1" in the diagonal entry ℓ_{ii} is circled, make a *positive* directed edge from i to the kth smallest neighbor of i in G (the vertices of G are numbered, so we can talk about the kth smallest neighbor).

For the term shown by the circles above, we thus obtain the following signed digraph (negative edges are shown black and positive edges white):



Let $\mathcal{D}$ denote the set of all signed digraphs D obtained in this way from the terms of the superexpansion. It is easy to see that each $D \in \mathcal{D}$ comes from exactly one term of the superexpansion. We can thus talk about $\mathrm{sgn}(D)$, meaning the sign of the corresponding term, and write π_D for the associated permutation.

We divide $\mathcal{D}$ into three parts as follows:

- $\mathcal{T}$, the $D \in \mathcal{D}$ with no directed cycle.
- $\mathcal{D}^+$, the $D \in \mathcal{D}$ with $\mathrm{sgn}(D) = +1$ and at least one directed cycle.
- $\mathcal{D}^-$, the $D \in \mathcal{D}$ with $\mathrm{sgn}(D) = -1$ and at least one directed cycle.

Here is a plan for the rest of the proof. We will show that the "acyclic objects" in $\mathcal{T}$ all have positive signs and they are in one-to-one correspondence with the spanning trees of G—thus they count what we want. Then, by constructing a suitable bijection, we will prove that $|\mathcal{D}^+| = |\mathcal{D}^-|$—so the "cyclic objects" cancel out. We then have $\det(L^-) = \sum_{D\in\mathcal{D}} \operatorname{sgn}(D) = |\mathcal{T}| + |\mathcal{D}^+| - |\mathcal{D}^-| = |\mathcal{T}|$ and the theorem follows.

To realize this plan, we first collect several easy properties of the signed digraphs in D.

(i) *If $i \to j$ is a directed edge, then $\{i, j\}$ is an edge of G.* (Clear.)

(ii) *Every vertex, with the exception of n, has exactly one outgoing edge, while n has no outgoing edge.* (Obvious.)

(iii) *All ingoing edges of n are positive.* (Clear.)

(iv) *No vertex has more than one negative ingoing edge.* This is because two negative ingoing edges $j \to i$ and $k \to i$ would mean two circled entries ℓ_{ji} and ℓ_{ki} in the ith column.

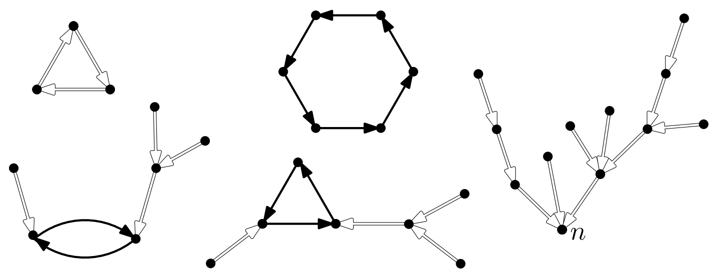
(v) *If a vertex i has a negative ingoing edge, then the outgoing edge is also negative.* Indeed, a negative ingoing edge $j \to i$ means that the off-diagonal entry ℓ_{ji} is circled, and hence none of the 1's in the diagonal entry ℓ_{ii} may be circled (which would be the only way of getting a positive outgoing edge from i).

Claim A. *These properties characterize $\mathcal{D}$. That is, if D is a signed digraph satisfying (i)–(v), then $D \in \mathcal{D}$.*

Proof. Given D, we determine the circled entry in each row i, $1 \le i \le n-1$, of L^-. We look at the single outgoing edge $i \to j$. If it's positive, we circle the appropriate 1 in ℓ_{ii}, and if it's negative, we circle ℓ_{ij}. We can't have two circled entries in a single column, since they would correspond to the situations excluded in (iv) or (v). $\square$

Next, we use (i)–(v) to describe the structure of D.

Claim B. *Each $D \in \mathcal{D}$ has the structure illustrated in the next picture.*



(a) *The vertex set is partitioned into one or more subsets $V_1, V_2, \ldots, V_k$ corresponding to the components of D, with no edges connecting different V_i's. If V_1 is the subset containing the vertex n, then the subgraph on V_0 is a tree with all edges directed towards n. The subgraph on every other V_i contains a single directed cycle of length at least 2, and a tree (possibly empty) attached to each vertex of the cycle, with edges directed towards the cycle.*

(b) *The edges not belonging to the directed cycles are all positive, and in each directed cycle either all edges are positive or all edges are negative.*

(c) *Conversely, each possible D with this structure and satisfying (i) above belongs to $\mathcal{D}$.*

Sketch of proof. Part (a), describing the structure of the digraph, is a straightforward consequence of (ii) (a single outgoing edge for every vertex except for n), and we leave it as an exercise. (If we added a directed loop to n, then every vertex has exactly one outgoing edge, and we get a so-called *functional digraph*, for which the structure as in (a) is well known.)

Concerning (b), if we start at a negative edge and walk on, condition (v) implies that we are going to encounter only negative edges. We thus can't reach n, since its incoming edges are positive, and so at some point we start walking around a negative cycle. Finally, a negative edge can't enter such a negative cycle from outside by (iv).

As for (c), if D has the structure as described in (a) and (b), the conditions (ii)–(iv) are obviously satisfied and Claim A applies. This proves Claim B. $\square$

The first item in our plan of the proof is now very easy to complete.

Corollary. *All $D \in \mathcal{T}$ have a positive sign and they are in one-to-one correspondence with the spanning trees of G.*

Proof. If $D \in \mathcal{D}$ has no directed cycles, then D is a tree with positive edges directed towards the vertex n. Moreover, π_D is the identity permutation since all the circled elements in the term corresponding to D lie on the diagonal of L^-. Thus $\mathrm{sgn}(D) = +1$, and if we forget the orientations of the edges, we arrive at a spanning tree of D. Conversely, given a spanning tree of G, we can orienting its edges towards n, and we obtain a $D \in \mathcal{T}$. $\square$

It remains to deal with the "cyclic objects". For $D \in \mathcal{D}^+ \cup \mathcal{D}^-$, let the *smallest cycle* be the directed cycle that contains the vertex with the smallest number (among all vertices in cycles). Let $\overline{D}$ be obtained from D by changing the signs of all edges in the smallest cycle.

Obviously $\overline{\overline{D}} = D$, and for $D \in \mathcal{D}$ we have $\overline{D} \in \mathcal{D}$ as well, as can be seen using Claim B. The following claim then shows that the mapping sending D to $\overline{D}$ is a bijection between $\mathcal{D}^+$ and $\mathcal{D}^-$, which is all that we need to finish the proof of the theorem.

Claim C. $\operatorname{sgn}(\overline{D}) = -\operatorname{sgn}(D)$.

Proof. We have $\operatorname{sgn}(D) = \operatorname{sgn}(\pi_D)(-1)^m$, where m is the number of negative edges of D and π_D is the associated permutation.

Let $i_1, i_2, \ldots, i_s$ be the vertices of the smallest cycle of D, numbered so that the directed edges of the cycle are $i_1 \to i_2$, $i_2 \to i_3, \ldots$, $i_{s-1} \to i_s$, $i_s \to i_1$.

For the permutation π_D, the i_j are fixed points, i.e., $\pi_D(i_j) = i_j$, $j = 1, 2, \ldots, s$. On the other hand, for $\pi_{\overline{D}}$ we have $\pi_{\overline{D}}(i_1) = i_2, \ldots, \pi_{\overline{D}}(i_{s-1}) = i_s$, $\pi_{\overline{D}}(i_s) = i_1$. Otherwise, π_D and $\pi_{\overline{D}}$ coincide. So $\pi_{\overline{D}}$ has an extra cycle of length s, and thus it can be obtained from π_D by $s-1$ transpositions. Hence $\operatorname{sgn}(\pi_{\overline{D}}) = (-1)^{s-1} \operatorname{sgn}(\pi_D)$, and

$$\operatorname{sgn}(\overline{D}) = \operatorname{sgn}(\pi_{\overline{D}})(-1)^{m+s} = (-1)^{s-1} \operatorname{sgn}(\pi_D)(-1)^{m+s} = -\operatorname{sgn}(D).$$

Claim C, and thus also the theorem, are proved. □

Sources. The theorem is usually attributed to

G. Kirchhoff: Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Verteilung galvanischer Ströme geführt wird, *Ann. Phys. Chem.* 72(1847) 497–508,

while

J. J. Sylvester: On the change of systems of independent variables, *Quart. J. Pure Appl. Math.* 1(1857) 42-56

is regarded as the first complete proof.

The above proof mostly follows

A. T. Benjamin, N. T. Cameron: Counting on determinants, *Amer. Math. Monthly* 112(2005) 481-492;

however, some simple but still not completely obvious steps seem to be missing in that paper's presentation. Benjamin and Cameron attribute the proof to

S. Chaiken: A Combinatorial proof of the all-minors matrix tree theorem, *SIAM J. Alg. Disc. Methods* 3(1982) 319–329,

but it may not be easy to find it there, since the paper deals with a more general setting.

4 The End of the Small Coins

An internet shop was processing m orders, each of them asking for various products. Suddenly, all coins with values below 1 Euro were taken out of circulation, and all prices had to be rounded, up or down, to whole Euros.

How can the shop round the prices so that the total price of each order is not affected by much? This rounding problem and similar questions is studied in *discrepancy theory*. Here we present a nice theorem with a linear-algebraic proof.

Theorem. *If at most t pieces of each product have been ordered in total, and if no order asks for more than one piece of each product, then it is possible to round the prices so that the total price of each order changes by no more than t Euros.*

It is interesting that the bound on the rounding error depends neither on the total number of orders nor on the number of different products.

A mathematical formulation of the problem. Let us call the products $1, 2, \ldots, n$ and let c_j be the price of the jth product. We can assume that each $c_j \in (0, 1)$ (because only the rounding plays a role in the problem). No order contains more than one product of each kind, and so we can represent the ith order as a set $S_i \subseteq \{1, 2, \ldots, n\}$, $i = 1, 2, \ldots, m$. The theorem now asserts that if no j is in more than t sets, then there are numbers $z_1, z_2, \ldots, z_n \in \{0, 1\}$ such that

$$\left| \sum_{j \in S_i} c_j - \sum_{j \in S_i} z_j \right| \leq t, \quad \text{for every } i = 1, 2, \ldots, m.$$

Proof. For every index $j \in \{1, 2, \ldots, n\}$ we introduce a real variable $x_j \in [0, 1]$, with initial value c_j. This variable will change during the proof and at the end, each x_j will have the value 0 or 1, which we will then use for z_j.

In each step, some of the variables x_j are already fixed, while the others are "floating". At the beginning, all the x_j are floating. The fixed x_j have values 0 or 1, and they won't change any more. The values of the floating variables are from the interval $(0, 1)$. In each step, at least one floating variable becomes fixed.

Let us call a set S_i *dangerous* if it contains more than t indices j for which x_j is still floating. The other sets are *safe*. We will keep the following condition satisfied:

$$\sum_{j \in S_i} x_j = \sum_{j \in S_i} c_j \text{ for all dangerous } S_i. \tag{2}$$

Let F be a set of indices of all floating variables and let us consider (2) as a system of linear equations with the floating variables as unknowns (while the values of the fixed variables are regarded as constants). This system surely has a solution—the current values of the floating variables. Since we assume that all floating variables lie in the interval $(0, 1)$, this solution is an interior point of the $|F|$-dimensional cube $[0, 1]^{|F|}$. We need to prove that there is a solution at the boundary of this cube as well, i.e., such that at least one of the variables attains value 0 or 1.

The crucial observation is that there are always fewer dangerous sets than floating variables, since every dangerous set needs more than t floating variables, while each floating variable contributes to at most t dangerous sets. Thus, the considered system of equations has fewer equations than unknowns, and so the solution space has dimension at least 1. Hence there is a straight line (a one-dimensional affine subspace) passing through the considered solution such that all points of this line are solutions, too. This line intersects the boundary of the cube at some point $\mathbf{y}$. We use the coordinates of $\mathbf{y}$ as the values of the floating variables in the next step. But all the floating variables x_j for which the corresponding value of $\mathbf{y}$ is 0 or 1 become fixed.

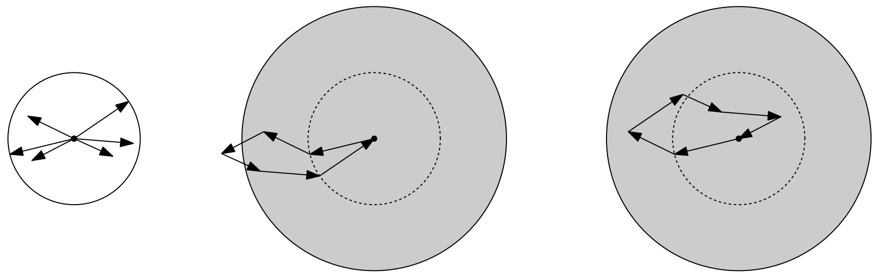
We repeat the described procedure until all variables become fixed. We claim that if we take the final value of x_j for z_j, $j = 1, 2, \ldots, n$, then $\left|\sum_{j \in S_i} c_j - \sum_{j \in S_i} z_j\right| \leq t$ for every $i = 1, 2, \ldots, m$ as we wanted.

To see this, let us consider a set S_i. At the moment when it ceased to be dangerous, we still had $\sum_{j \in S_i} c_j - \sum_{j \in S_i} x_j = 0$ according to (2), and S_i contained the indices of at most t floating variables. The value of each of these floating variables has not changed by more than 1 in the rest of the process (it could have been 0.001 and later be fixed to 1). This finishes the proof. □

Sources. J. Beck, T. Fiala: "Integer making" theorems, *Discr. Appl. Math*, 3(1981), 1–8.

5 Walking in the Yard

A mathematically inclined prison guard forces a prisoner to take a walk under the following strict instructions. The prisoner receives a finite set M of vectors, each of length at most 10m. He must start the walk in the center of a circular prison yard of radius 20m, then move by some vector $\mathbf{v}_1 \in M$, then by some other vector $\mathbf{v}_2 \in M$, etc., using each vector in M exactly once. The vectors in M sum up to $\mathbf{0}$, so the prisoner will again finish in the center. However, he must not cross the boundary of the yard any time during the walk (if he does, the guard will start shooting without warning).



the set M a wrong order a correct order

The following theorem shows that a safe walk is possible for every finite M, and it also works for yards that are d-dimensional balls.

Theorem. *Let M be an arbitrary set of n vectors in $\mathbb{R}^d$ such that $\|\mathbf{v}\| \le 1$ for every $\mathbf{v} \in M$, where the norm $\|\mathbf{v}\|$ of $\mathbf{v}$ is the usual Euclidean length, and $\sum_{\mathbf{v}\in M} \mathbf{v} = \mathbf{0}$. Then it is possible to arrange all vectors of M into a sequence $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ in such a way that $\|\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_k\| \le d$ for every $k = 1, 2, \ldots, n$.*

In the example in the picture, the vectors can even be arranged so that the path lies within a circle of radius 1, but for an arbitrary set of vectors radius 1 may be impossible (find an example). For the plane, the smallest possible radius of the yard is known: $\sqrt{5}/2 \approx 1.118$. For a general dimension d, the best known lower bound is of order $\sqrt{d}$. It is not known whether the theorem can be improved to d-dimensional spherical yards of radius $\mathrm{O}(\sqrt{d}\,)$.

The proof below actually yields a more general statement, for an arbitrary (not necessarily circular) yard: If $B \subset \mathbb{R}^d$ is a bounded convex set containing the origin, and M is a set of n vectors with $\mathbf{v} \in B$ for every $\mathbf{v} \in M$ and $\sum_{\mathbf{v}\in M} \mathbf{v} = 0$, then there is an ordering $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ of the vectors from M such that $\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_k \in dB$ for all $k = 1, 2, \ldots, n$, where $dB = \{d\mathbf{x} : \mathbf{x} \in B\}$. In this more general setting, the constant d *cannot* be improved. To see this for $d = 2$, we take B as an equilateral triangle centered at the origin.

We start the proof with a simple general lemma (which was implicitly used in Section 4).

Lemma. *Let $A\mathbf{x} = \mathbf{b}$ be a system of m linear equations in $n \le m$ unknowns, and let us suppose that it has a solution $\mathbf{x}_0 \in [0,1]^n$. Then there is a solution $\tilde{\mathbf{x}} \in [0,1]^n$ in which at least $n - m$ components are 0's or 1's.*

Proof. We proceed by induction on $m - n$. For $n = m$ there is nothing to prove, so let $n > m$. Then the solution space has dimension at least 1, and so it contains a line passing through $\mathbf{x}_0$. This line intersects the boundary of the cube $[0,1]^n$; let $\mathbf{y}$ be an intersection point. Thus, $y_i \in \{0, 1\}$ for some index i.

Let us set up a new linear system with $n - 1$ unknowns that is obtained from $A\mathbf{x} = \mathbf{b}$ by fixing the value of x_i to y_i. This new system satisfies the assumption of the lemma (a solution lying in $[0,1]^{n-1}$ is obtained from $\mathbf{y}$ by deleting y_i), and so, by the inductive assumption, it has a solution with at least $n - m - 1$ components equal to 0 or 1. Together with y_i this gives a solution of the original system with $n - m$ or more 0's and 1's. $\square$

Proof of the theorem. The rough idea is this: The set M is "very good" because its vectors sum up to $\mathbf{0}$, and thus the sum has norm 0. We introduce a weaker notion of a "good" set of vectors. The definition is chosen so that if K is a good set, then the sum of all of its vectors has norm at most d. Moreover, and this is the heart of the proof, we will show that every good set K of $k > d$ vectors has a good subset of $k - 1$ vectors. This will allow us to find the desired ordering of the vectors of M by induction.

Here is the definition: A set $K = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$ of $k \ge d$ vectors in $\mathbb{R}^d$, each of length at most 1, is called **good** if there exist coefficients $\alpha_1, \ldots, \alpha_k$ satisfying

$$\begin{aligned} \alpha_i &\in [0,1], \quad i = 1, 2, \ldots, k \\ \alpha_1\mathbf{w}_1 + \alpha_2\mathbf{w}_2 + \cdots + \alpha_k\mathbf{w}_k &= \mathbf{0} \qquad (3) \\ \alpha_1 + \alpha_2 + \cdots + \alpha_k &= k - d. \qquad (4) \end{aligned}$$

We note that if the right-hand side of (4) were k instead of $k - d$, then all the α_i's would have to be 1's and thus (3) would simply mean $\sum_{i=1}^{k} \mathbf{w}_i = \mathbf{0}$.

But since $\sum_{i=1}^{n} \alpha_i$ is $k-d$, most of the α_i's must be close to 1, but there is some freedom left.

First let us check that if $K = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$ is good, then $\|\mathbf{w}_1 + \mathbf{w}_2 + \cdots + \mathbf{w}_k\| \leq d$. Indeed, we have

$$\begin{aligned} \left\| \sum_{i=1}^{k} \mathbf{w}_i \right\| &= \left\| \sum_{i=1}^{k} \mathbf{w}_i - \sum_{i=1}^{k} \alpha_i \mathbf{w}_i \right\| \\ &\leq \sum_{i=1}^{k} \|(1-\alpha_i)\mathbf{w}_i\| = \sum_{i=1}^{k} (1-\alpha_i)\|\mathbf{w}_i\| \\ &\leq \sum_{i=1}^{k} (1-\alpha_i) = d. \end{aligned}$$

Next, we have the crucial claim.

Claim. *If $K = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$ is a good set of $k > d$ vectors, then there is some i such that $K \setminus \{\mathbf{w}_i\}$ is a good set of $k-1$ vectors.*

Proof of the claim. We consider the following system of linear equations for unknowns $x_1, \ldots, x_k$:

$$x_1\mathbf{w}_1 + x_2\mathbf{w}_2 + \cdots + x_k\mathbf{w}_k = \mathbf{0} \tag{5}$$

$$x_1 + x_2 + \cdots + x_k = k - d - 1. \tag{6}$$

Here (5) is an equality of two d-dimensional vectors and thus it actually represents d equations. The last equation (6) is like the condition (4), except that the right-hand side is $k-d-1$; this is a preparation for showing that a suitable subset of $k-1$ vectors in K is good.

The above system has $d+1$ equations for k unknowns. If $\alpha_1, \ldots, \alpha_k$ are the coefficients witnessing that K is good, then by setting $x_i := \frac{k-d-1}{k-d}\alpha_i$ we obtain a solution of (5),(6) lying in $[0,1]^k$.

Thus by the lemma there is also a solution $\tilde{\mathbf{x}} \in [0,1]^k$ with at least $k-d-1$ components equal to 0 or 1. We want to see that at least one component of $\tilde{\mathbf{x}}$ has to be 0. Indeed, if all the $k-d-1$ components guaranteed by the lemma happen to be 1, then all of the remaining $d+1$ components must be 0, since all components add up to $k-d-1$ by (6).

Now it is easy to check that for any index i with $\tilde{x}_i = 0$ the set $K \setminus \{\mathbf{w}_i\}$ is good. Indeed, the remaining components of $\tilde{\mathbf{x}}$ can be used in the role of the α_i in the definition of a good set. This proves the claim.

The proof of the theorem is finished easily by induction. We start with the set $M_n := M$, which is obviously good. Using the claim, we find a vector in M_n whose removal produces a good set. We call this vector $\mathbf{v}_n$, and we let $M_{n-1} := M_n \setminus \{\mathbf{v}_n\}$. Similarly, having constructed the good set M_k, we find a vector $\mathbf{v}_k \in M_k$ such that $M_{k-1} := M_k \setminus \{\mathbf{v}_k\}$ is good, and so on, all the way down to M_d.

We are left with the set M_d of d vectors, and we number these arbitrarily $\mathbf{v}_1, \ldots, \mathbf{v}_d$. For $k \leq d$ we obviously have $\|\mathbf{v}_1 + \cdots + \mathbf{v}_k\| \leq k \leq d$, and for $k > d$ the norm of the sum of all vectors in M_k is at most d since M_k is a good set. The theorem is proved. □

Sources. The theorem is sometimes called the Steinitz lemma since Steinitz gave a first complete proof of a weaker version in 1913, following an incomplete proof of Lévy from 1905. The above proof is from

V. S. Grinberg, S. V. Sevastyanov: The value of the Steinitz constant (in Russian), *Funk. Anal. Prilozh.* 14(1980), 56–57.

For background and several results of a similar nature see

I. Bárány: On the power of linear dependencies, in Gy. O. H. Katona, M. Grötschel editors, *Building bridges*, Springer, Berlin 2008, 31–46.

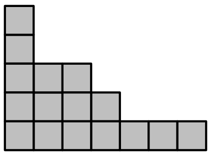
6 More Bricks—More Walls?

One of the classical topics in enumeration are **integer partitions**. For example, there are five partitions of the number 4:

$$
\begin{aligned}
4 &= 1+1+1+1+1 \\
4 &= 2+1+1 \\
4 &= 2+2 \\
4 &= 3+1 \\
4 &= 4.
\end{aligned}
$$

The order of the addends in a partition doesn't matter, and it is customary to write them in a nonincreasing order as we did above.

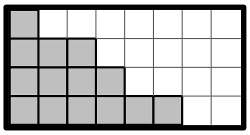
A partition of n is often represented graphically by its **Ferrers diagram**, which one can think of as a nondecreasing wall built of n bricks. For example, the following Ferrers diagram



corresponds to $16 = 5 + 3 + 3 + 2 + 1 + 1 + 1$.

How can we determine or estimate $p(k)$, the number of partitions of the integer k? This is a surprisingly difficult enumeration problem, ultimately solved by a formula of Hardy and Ramanujan. The asymptotics of $p(k)$ is $p(k) \sim \frac{1}{4k\sqrt{3}} e^{\pi\sqrt{2k/3}}$, where $f(k) \sim g(k)$ means $\lim_{k\to\infty} \frac{f(k)}{g(k)} = 1$.

Here we consider another matter, the number $p_{w,h}(k)$ of partitions of k with at most w addends, none of them exceeding h. In other words, $p_{w,h}(k)$ is the number of ways to build a nonincreasing wall out of k bricks inside a box of width w and height h:



Here is the main result of this section:

Theorem. *For every $w \geq 1$ and $h \geq 1$ we have*

$$p_{w,h}(0) \leq p_{w,h}(1) \leq \cdots \leq p_{w,h}\left(\lfloor \tfrac{wh}{2} \rfloor\right)$$

and

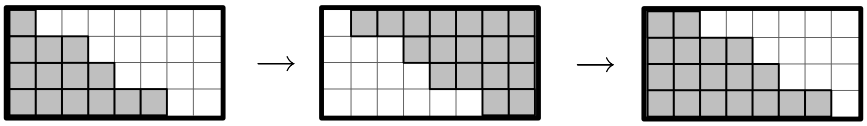
$$p_{w,h}\left(\lceil \tfrac{wh}{2} \rceil\right) \geq p_{w,h}\left(\lceil \tfrac{wh}{2} \rceil + 1\right) \geq \cdots \geq p_{w,h}(wh-1) \geq p_{w,h}(wh).$$

That is, $p_{w,h}(k)$ as a function of k is nondecreasing for $k \leq \frac{wh}{2}$ and nonincreasing for $k \geq \frac{wh}{2}$.

So the first half of the theorem tells us that with more bricks we can build more (or rather, at least as many) walls. This goes on until half of the box is filled with bricks; after that, we already have too little space and the number of possible walls starts decreasing.

Actually, once we know that $p_{w,h}(k)$ is nondecreasing for $k \leq \frac{wh}{2}$, it is easily seen that it must be nonincreasing for $k \geq \frac{wh}{2}$, because $p_{w,h}(k) =$

$p_{w,h}(wh-k)$, as can be seen using the following bijection transforming walls with k bricks into walls with $wh-k$ bricks:



bricks $\leftrightarrow$ nonbricks turn by 180 degrees

The theorem is one of the results that look intuitively obvious but are surprisingly hard to prove. The great Cayley used this as a fact requiring no proof in his 1856 memoir, and only about twenty years later did Sylvester discover the first proof.

One would naturally expect such a combinatorial problem to have a combinatorial solution, perhaps simply an injective map assigning to every wall of k bricks a wall of $k+1$ bricks (for $k+1 \leq \frac{wh}{2}$). But to my knowledge, nobody has managed to discover a proof of this kind, and estimating $p_{w,h}(k)$ or expressing it by a formula doesn't seem to lead to the goal either.

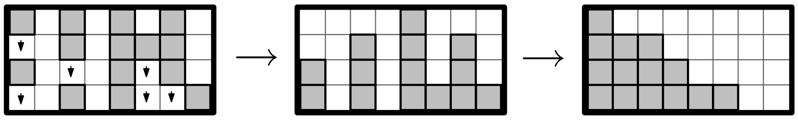
Earlier proofs of the theorem used relatively heavy mathematical tools, essentially representations of Lie algebras. The proof shown here is a result of several simplifications of the original ideas, and it uses "only" matrix-rank arguments.

Functions, or sequences, that are first nondecreasing and then, from some point on, nonincreasing, are called **unimodal** (and so are functions that begin as nonincreasing and continue as nondecreasing). There are many important results and conjectures in various areas of mathematics asserting that certain quantities form an unimodal sequence, and the proof below contains tools of general applicability.

Preliminary considerations. Let us write $n := wh$ for the area of the box, and let us fix a numbering of the n squares in the box by the numbers $1, 2, \ldots, n$.

To prove the theorem, we will show that $p_{w,h}(k) \leq p_{w,h}(\ell)$ for $0 \leq k < \ell \leq \frac{n}{2}$.

The first step is to view a wall in the box as an *equivalence class*. Namely, we start with an arbitrary set of k bricks filling some k squares in the box, and then we tidy them up into a nonincreasing wall:



First we push down the bricks in each column, and then we rearrange the columns into a nonincreasing order.

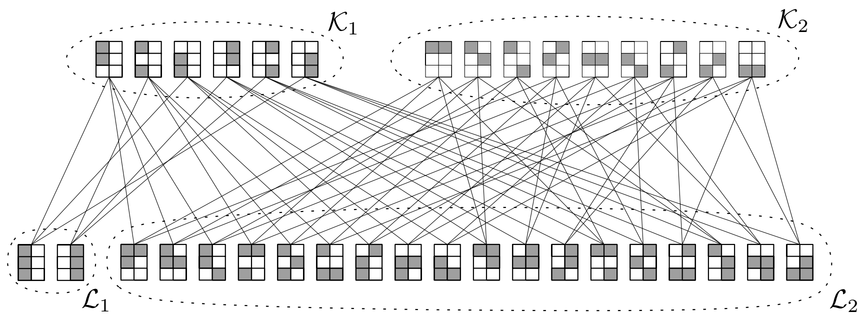
Let us call two k-element subsets $K, K' \subseteq \{1, 2, \ldots, n\}$, understood as sets of k squares in the box, **wall-equivalent** if they lead to the same nonincreasing wall. This indeed defines an equivalence on the set $\mathcal{K}$ of all k-element subsets of $\{1, 2, \ldots, n\}$. Let the equivalence classes be $\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_r$, where $r := p_{w,h}(k)$.

Let's phrase the definition of the wall-equivalence differently, in a way that will be more convenient latter. Let π be a permutation of the n squares in the box; let us say that π **doesn't break columns** if it corresponds to first permuting the squares in each column arbitrarily, and then permuting the columns. It is easily seen that two subsets $K, K' \in \mathcal{K}$ are wall-equivalent exactly if $K' = \pi(K)$ for some permutation that doesn't break columns.[1]

Next, let $\mathcal{L}$ be the set of all ℓ-element subsets of $\{1, 2, \ldots, n\}$, and let it be divided similarly into $s := p_{w,h}(\ell)$ classes $\mathcal{L}_1, \ldots, \mathcal{L}_s$ according to wall-equivalence. The goal is to prove that $r \leq s$.

Let us consider the bipartite graph G with vertex set $\mathcal{K} \cup \mathcal{L}$ and with edges corresponding to inclusion; i.e., a k-element set $K \in \mathcal{K}$ is connected to an ℓ-element set $L \in \mathcal{L}$ by an edge if $K \subseteq L$. Here is a small-scale illustration ($w = 2$, $h = 3$, $k = 2$, $\ell = 3$):

[1] In a more mature mathematical language, the permutations that don't break columns form a permutation group acting on $\mathcal{K}$, and the classes of the wall-equivalence are the orbits of this action. Some things in the sequel could (should?) also be phrased in the language of actions of permutation groups, but I decided to avoid this terminology, with the hope of deterring slightly fewer students.



Claim. *For every i and j, all $L \in \mathcal{L}_j$ have the same number d_{ij} of neighbors in $\mathcal{K}_i$.*

> *Proof.* Let $L, L' \in \mathcal{L}_j$, and let us fix some permutation π that doesn't break columns and such that $L' = \pi(L)$. For $K \in \mathcal{K}_i$, we have $\pi(K) \in \mathcal{K}_i$ as well (by the alternative description of the wall-equivalence), and it is easily seen that $K \mapsto \pi(K)$ defines a bijection between the neighbors of L lying in $\mathcal{K}_i$ and the neighbors of L' lying in $\mathcal{K}_i$. $\square$

Let us now pass to a more general setting for a while: Let U, V be disjoint finite sets, let $(U_1, \ldots, U_r, V_1, \ldots, V_s)$ be a partition of $U \cup V$ with $U = U_1 \cup \cdots \cup U_r$ and $V = V_1 \cup \cdots \cup V_s$, where the U_i and V_j are all nonempty, and let G be a bipartite graph on the vertex set $U \cup V$ (with all edges going between U and V). We call the partition $(U_1, \ldots, U_r, V_1, \ldots, V_s)$ ***V*-degree homogeneous** w.r.t. G if the condition as in the claim holds, i.e., all vertices in V_j have the same number d_{ij} of neighbors in U_i, for all i and j. In such case, we call the matrix $D = (d_{ij})_{i=1}^{r}{}_{j=1}^{s}$ the ***V*-degree matrix** of the partition (with respect to G).

In the setting introduced above, we have a bipartite graph with a V-degree homogeneous partition, and we would like to conclude that r, the number of the U-pieces, can't be smaller than s, the number of V-pieces. The next lemma gives a sufficient condition, which we'll then be able to verify for our particular G. The condition essentially says that V is at least as large as U for a "linear-algebraic reason".

To formulate the lemma, we set up a $|U| \times |V|$ matrix B, with rows indexed by the vertices in U and columns indexed by the vertices in V, whose entries b_{uv} are given by

$$b_{uv} := \begin{cases} 1 & \text{if } \{u, v\} \in E(G) \\ 0 & \text{otherwise.} \end{cases}$$

Lemma. *Let G be a bipartite graph as above, let $(U_1, \ldots, U_r, V_1, \ldots, V_s)$ be a V-degree homogeneous partition of its vertices, and let's suppose that the rows of the matrix B are linearly independent. Then $r \leq s$.*

Proof. This powerful statement is quite easy to prove. We will show that the $r \times s$ V-degree matrix D has linearly independent rows, which means that it can't have fewer columns than rows, and thus $r \leq s$ indeed.

Let $B[U_i, V_j]$ denote the submatrix of B consisting of the entries b_{uv} with $u \in U_i$ and $v \in V_j$; schematically

	V_1	V_2	V_3	V_4
U_1	$B[U_1, V_1]$	$B[U_1, V_2]$		
U_2				
U_3				$B[U_3, V_4]$

The V-degree homogeneity condition translates to the matrix language as follows: The sum of each of the columns of $B[U_i, V_j]$ equals d_{ij}.

For a vector $\mathbf{x} \in \mathbb{R}^r$, let $\tilde{\mathbf{x}} \in \mathbb{R}^{|U|}$ be the vector indexed by the vertices in U obtained by replicating $|U_i|$-times the component x_i; that is, $\tilde{x}_u = x_i$ for all $u \in U_i$, $i = 1, 2, \ldots, r$.

For this $\tilde{\mathbf{x}}$, we consider the product $\tilde{\mathbf{x}}^T B$. Its vth component equals $\sum_{u \in U} \tilde{x}_u b_{uv} = \sum_{i=1}^{r} x_i \sum_{u \in U_i} b_{uv} = \sum_{i=1}^{r} x_i d_{ij} = (\mathbf{x}^T D)_j$. Hence $\mathbf{x}^T D = \mathbf{0}$ implies $\tilde{\mathbf{x}}^T B = \mathbf{0}$.

Let's assume for contradiction that the rows of D are linearly dependent; that is, there is a nonzero $\mathbf{x} \in \mathbb{R}^r$ with $\mathbf{x}^T D = \mathbf{0}$. Then $\tilde{\mathbf{x}} \neq \mathbf{0}$ but, as we've just seen, $\tilde{\mathbf{x}}^T B = \mathbf{0}$. This contradicts the linear independence of the rows of B and proves the lemma. $\square$

Proof of the theorem. We return to the particular bipartite graph G introduced above, with vertex set $\mathcal{K} \cup \mathcal{L}$ and with the $\mathcal{L}$-degree homogeneous partition $(\mathcal{K}_1, \ldots, \mathcal{K}_r, \mathcal{L}_1, \ldots, \mathcal{L}_s)$ according to the wall-equivalence. For applying the lemma, it remains to show that the rows of the corresponding matrix B are linearly independent.

This result, known as **Gottlieb's theorem**,[2] has proved useful in several other applications as well. Explicitly, it tells us that *for* $0 \le k < \ell \le \frac{n}{2}$, *the zero-one matrix* B *with rows indexed by* $\mathcal{K}$ *(all* k*-subsets of* $\{1, 2, \ldots, n\}$*), columns indexed by* $\mathcal{L}$ *(all* ℓ*-subsets), and the nonzero entries corresponding to containment, has linearly independent rows.*

Several proofs are known; here we present one which shows that, under favorable circumstances, the above lemma can be reversed.

Proof of Gottlieb's theorem. For contradiction, we assume that $\mathbf{y}^T B = \mathbf{0}$ for some nonzero vector $\mathbf{y}$. The components of $\mathbf{y}$ are indexed by k-element sets; let us fix some $K_0 \in \mathcal{K}$ with $y_{K_0} \neq 0$.

Next, we partition both $\mathcal{K}$ and $\mathcal{L}$ into $k+1$ classes according to the size of the intersection with K_0 (this partition has nothing to do with the partition of $\mathcal{K}$ and $\mathcal{L}$ considered earlier—we just re-use the same letters):

$$\mathcal{K}_i := \{K \in \mathcal{K} : |K \cap K_0| = i\},\ \ \mathcal{L}_j := \{L \in \mathcal{L} : |L \cap K_0| = j\},\ i, j = 0, 1, \ldots, k.$$

Every $\mathcal{K}_i$ and every $\mathcal{L}_j$ is nonempty—here we use the assumption $k < \ell \le \frac{n}{2}$ (if, for example, we had $k+\ell > n$, we would get $\mathcal{L}_0 = \emptyset$, since there wouldn't be enough room for an ℓ-element L disjoint from K_0).

Here, for a change, we will need that this partition is $\mathcal{K}$-degree homogeneous (with respect to the same bipartite graph as above, with edges representing inclusion). That is, every $K \in \mathcal{K}_i$ has the same number d_{ij} of neighbors in $\mathcal{L}_j$. More explicitly, d_{ij} is the number of ways of extending a k-element set K with $|K \cap K_0| = i$ to an ℓ-element $L \supset K$ with $|L \cap K_0| = j$; this number is clearly independent of the specific choice of K.

By this description, we have $d_{ij} = 0$ for $i > j$, and thus the $\mathcal{K}$-degree matrix D is upper triangular. Moreover, $d_{ii} \neq 0$ for all $i = 0, 1, \ldots, k$, and so D is non-singular.

Using the vector $\mathbf{y}$, we are going to exhibit a nonzero $\mathbf{x} = (x_0, x_1, \ldots, x_k)$ with $\mathbf{x}^T D = \mathbf{0}$, which will be a contradiction. A suitable $\mathbf{x}$ is obtained by summing the components of $\mathbf{y}$ over the classes $\mathcal{K}_i$:

$$x_i := \sum_{K \in \mathcal{K}_i} y_K.$$

We have $\mathbf{x} \neq \mathbf{0}$, since the class $\mathcal{K}_k$ contains only K_0, and so $x_k = y_{K_0} \neq 0$.

For every j we calculate

$$0 \;=\; \sum_{L \in \mathcal{L}_j} (\mathbf{y}^T B)_L = \sum_{L \in \mathcal{L}_j} \sum_{K \in \mathcal{K}} y_K b_{KL} = \sum_{K \in \mathcal{K}} y_K \sum_{L \in \mathcal{L}_j} b_{KL}$$

[2] This is not the only theorem associated with Gottlieb's name, though.

$$= \sum_{i=0}^{k} \sum_{K \in \mathcal{K}_i} y_K d_{ij} = \sum_{i=0}^{k} x_i d_{ij} = (\mathbf{x}^T D)_j.$$

Hence $\mathbf{x}^T D = \mathbf{0}$, and this is the promised contradiction to the non-singularity of D. Gottlieb's theorem, as well as our main theorem, are proved. $\square$

Another example. For readers familiar with the notion of graph isomorphism, the following might be a rewarding exercise in applying the method shown above: Prove that if $g_n(k)$ stands for the number of nonisomorphic graphs with n vertices and k edges, then the sequence $g_n(0), g_n(1), \ldots, g_n(\binom{n}{2})$ is unimodal.

Sources. As was mentioned above, the theorem was implicitly assumed without proof in

A. Cayley: A second memoir on quantics, *Phil. Trans. Roy. Soc.* 146 (1856) 101–126.

The word "quantic" in the title means, in today's terminology, a homogeneous multivariate polynomial, and Cayley was interested in quantics that are invariant under the action of linear transformations. The first proof of the theorem was obtained in

J. J. Sylvester: Proof of the hitherto undemonstrated fundamental theorem of invariants, *Philos. Mag.* 5(1878) 178–188.

A substantially more elementary proof than the previous ones, phrased in terms of group representations, was obtained in

R. P. Stanley: Some aspects of groups acting on finite posets, *J. Combinatorial Theory Ser. A* 32(1982) 132–161.

Our presentation is based on that of Babai and Frankl in their textbook cited in the introduction.

Gottlieb's theorem was first proved in

D. H. Gottlieb: A certain class of incidence matrices, *Proc. Amer. Math. Soc* 17(1066) 1233–1237.

The proof presented above rephrases an argument from

C. D. Godsil: Tools from linear algebra, Chapter 31 of *R. Graham, M. Grötschel, L. Lovász, editors, Handbook of Combinatorics*, North-Holland, Amsterdam, 1995, pages 1705–1748.

For an introduction to integer partitions see

G. Andrews, K. Eriksson: *Integer partitions*, Cambridge University Press, Cambridge 2004, (this is a very accessible source), or Wilf's lecture notes at `http://www.math.upenn.edu/~wilf/PIMS/PIMSLectures.pdf`.

7 Equilateral Sets

An *equilateral set* in $\mathbb{R}^d$ is a set of points $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ such that all pairs $\mathbf{p}_i, \mathbf{p}_j$ of distinct points have the same distance.

Intentionally we haven't said what distance we mean: This will play a key role in this section. If one considers the most usual *Euclidean* distance, then it is not too hard to prove that an equilateral set in $\mathbb{R}^d$ can have $d+1$ points but no more.

As an aside, let us sketch the classical proof that there can't be more than $d+1$ points. Let the points be $\mathbf{p}_1$ through $\mathbf{p}_{n+1}$, translate them so that $\mathbf{p}_{n+1} = \mathbf{0}$, re-scale so that the interpoint distances are 1, and set up the matrix (the Gram matrix) G with $g_{ij} = \langle \mathbf{p}_i \,|\, \mathbf{p}_j \rangle$ (scalar product). Using the equilaterality condition one calculates that $G = \frac{1}{2}(I_n + J_n)$, where J_n is the all-ones matrix, and thus $\mathrm{rank}(G) = n$. On the other hand, we have $G = P^T P$, where P is the $d \times n$ matrix with the vector $\mathbf{p}_i$ as the ith column, and thus $\mathrm{rank}(G) \le d$, which gives $n \le d$.

(And, by the way, how do we prove rigorously that a $(d+1)$-point equilateral set is possible? We can take, e.g., the vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_d$ of the standard basis plus the point $(-t, -t, \ldots, -t)$ for a suitable $t > 0$—even if we are too lazy to calculate the right t, it is easy to prove its existence.)

Approximately equilateral sets. We will now relax the condition that all interpoint distances must be exactly the same: We will require them to be only approximately the same. With a sufficiently strong notion of "approximately the same" we will show that the size of such an approximately equilateral set in $\mathbb{R}^d$ is bounded by a constant multiple of d. The proof relies on a neat linear algebra trick. We will then use the result in the proof of the main theorem of this section.

Rank Lemma. *Let A be a real symmetric $n \times n$ matrix, not equal to the zero matrix. Then*

$$\mathrm{rank}(A) \ge \frac{\left(\sum_{i=1}^n a_{ii}\right)^2}{\sum_{i,j=1}^n a_{ij}^2}.$$

Proof. Linear algebra teaches us that A as in the lemma has n real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. If $\mathrm{rank}(A) = r$, then exactly r of the λ_i are nonzero; we may suppose that $\lambda_i \ne 0$ for $1 \le i \le r$, while $\lambda_i = 0$ for $i > r$.

Let us write down the Cauchy–Schwarz inequality

$$(\sum_{i=1}^{r} x_i y_i)^2 \leq (\sum_{i=1}^{r} x_i^2)(\sum_{i=1}^{r} y_i^2)$$

for $x_i = \lambda_i$, $y_i = 1$; we obtain $(\sum_{i=1}^{r} \lambda_i)^2 \leq r \sum_{i=1}^{r} \lambda_i^2$. Dividing by $\sum_{i=1}^{r} \lambda_i^2$ yields the following inequality for the rank in terms of eigenvalues:

$$\operatorname{rank}(A) \geq \frac{\left(\sum_{i=1}^{n} \lambda_i\right)^2}{\sum_{i=1}^{n} \lambda_i^2}. \tag{7}$$

(We have extended the summation all the way to n, since λ_{r+1} through λ_n are 0's.)

The last inequality can be converted to the inequality in the Rank Lemma in three easy steps: First, the sum of all eigenvalues of A equals the *trace* of A, i.e. $\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} a_{ii}$ (a standard linear algebra fact). This takes care of the numerator in (7). Second, the eigenvalues of A^2 are $\lambda_1^2, \ldots, \lambda_n^2$, as one can recall or immediately check, and thus $\sum_{i=1}^{n} \lambda_i^2 = \operatorname{trace}(A^2)$. Third, $\operatorname{trace}(A^2) = \sum_{i,j=1}^{n} a_{ij}^2$, as one easily calculates. This brings the denominator into the desired form. □

Corollary—A small perturbation of I_n has a large rank. *Let A be a symmetric $n \times n$ matrix with $a_{ii} = 1$, $i = 1, 2, \ldots, n$ and $|a_{ij}| \leq 1/\sqrt{n}$ for all $i \neq j$. Then $\operatorname{rank}(A) \geq \frac{n}{2}$.* □

Proposition (on approximately equilateral sets). *Let $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n \in \mathbb{R}^d$ be points such that for every $i \neq j$ we have*

$$1 - \frac{1}{\sqrt{n}} \leq \|\mathbf{p}_i - \mathbf{p}_j\|^2 \leq 1 + \frac{1}{\sqrt{n}}.$$

Then $n \leq 2(d+2)$. (Note that, for technical convenience, we bound the squared *Euclidean distances.)*

Proof. Let A be the $n \times n$ matrix with $a_{ij} = 1 - \|\mathbf{p}_i - \mathbf{p}_j\|^2$. The assumptions of the proposition immediately give that A meets the assumptions of the above corollary, and thus $\operatorname{rank}(A) \geq \frac{n}{2}$.

It remains to bound $\operatorname{rank}(A)$ from above in terms of d. For $i = 1, 2, \ldots, n$ let $f_i : \mathbb{R}^d \to \mathbb{R}$ be the function defined by $f_i(\mathbf{x}) = 1 - \|\mathbf{x} - \mathbf{p}_i\|^2$; so the ith row of A is $\big(f_i(\mathbf{p}_1), f_i(\mathbf{p}_2), \ldots, f_i(\mathbf{p}_n)\big)$.

We rewrite $f_i(\mathbf{x}) = 1 - \|\mathbf{x}\|^2 - \|\mathbf{p}_i\|^2 + 2(p_{i1}x_1 + p_{i2}x_2 + \cdots + p_{id}x_d)$, where p_{ik} is the kth coordinate of $\mathbf{p}_i$. Then it becomes clear that each

f_i is a linear combination of the following $d+2$ functions: the constant function 1, the function $\mathbf{x} \mapsto \|\mathbf{x}\|^2$, and the "coordinate functions" $\mathbf{x} \mapsto x_k$, $k = 1, 2, \ldots, d$. Hence the vector space generated by the f_i has dimension at most $d+2$, and so has the column space of A. Thus $\operatorname{rank}(A) \leq d+2$, and the proposition is proved. $\square$

Other kinds of distance. Equilateral sets become much more puzzling if one considers other notions of distance in $\mathbb{R}^d$.

First, as a cautionary tale, let us consider the ℓ_∞ ("ell infinity") distance, where the distance of two points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ is defined as $\|\mathbf{x} - \mathbf{y}\|_\infty = \max\{|x_i - y_i| : i = 1, 2, \ldots, d\}$. Then the "cube" $\{0,1\}^d$ is an equilateral set with as many as 2^d points! (Which turns out to be the largest possible example in $\mathbb{R}^d$ with the ℓ_∞ distance, but this is not the story we want to narrate here.)

The distance we really want to focus on here is the ℓ_1 *distance*, given by

$$\|\mathbf{x} - \mathbf{y}\|_1 = |x_1 - y_1| + |x_2 - y_2| + \cdots + |x_d - y_d|.$$

Then the following is an example of an equilateral set with $2d$ points: $\{\mathbf{e}_1, -\mathbf{e}_1, \mathbf{e}_2, -\mathbf{e}_2, \ldots, \mathbf{e}_d, -\mathbf{e}_d\}$. A widely believed conjecture states that this is as many as one can ever get, but until about 2001, no upper bound better than $2^d - 1$ (exponential!) was known.

We will present an ingenious proof of a polynomial upper bound, $O(d^4)$. The proof of the current best bound, $O(d \log d)$, uses a number of additional ideas and it is considerably more technical.

Theorem. *For every $d \geq 1$, no equilateral set in $\mathbb{R}^d$ with the ℓ_1 distance has more than $100d^4$ points.*

The main reason why for the ℓ_1 distance one can't imitate the proof for the Euclidean case sketched above or something similar seems to be this: The functions $\varphi_a\colon \mathbb{R} \to \mathbb{R}$, $\varphi_a(x) = |x - a|$, $a \in \mathbb{R}$, are all linearly independent—unlike the functions $\psi_a(x) = (x-a)^2$ that generate a vector space of dimension only 3.

The forthcoming proof has an interesting twist: In order to establish a bound on *exactly* equilateral sets for the "unpleasant" ℓ_1 distance, we use *approximately* equilateral sets but for the "pleasant" Euclidean distance. Here is a tool for such a passage.

Lemma (on approximate embedding). *For every two natural numbers d, q there exists a mapping $f_{d,q}\colon [0,1]^d \to \mathbb{R}^{dq}$ such that for every $\mathbf{x}, \mathbf{y} \in$*

$[0,1]^d$

$$\|\mathbf{x}-\mathbf{y}\|_1 - \tfrac{2d}{q} \le \tfrac{1}{q}\|f_{d,q}(\mathbf{x}) - f_{d,q}(\mathbf{y})\|^2 \le \|\mathbf{x}-\mathbf{y}\|_1 + \tfrac{2d}{q}.$$

Let us stress that we take *squared* Euclidean distances in the target space. If we wanted instead that the ℓ_1 distance $\|\mathbf{x}-\mathbf{y}\|_1$ be reasonably close to the Euclidean distance of the images for all $\mathbf{x},\mathbf{y}$, the task becomes impossible.

Our proof of the lemma below is somewhat simple-minded. By more sophisticated methods one can reduce the dimension of the target space considerably, and this is also how the d^4 bound in the theorem can be improved.

Proof of the lemma. First we consider the case $d=1$. For $x \in [0,1]$, $f_{1,q}(x)$ is the q-component zero/one vector starting with a segment of $\lfloor qx \rfloor$ ones, followed by $q - \lfloor qx \rfloor$ zeros. Then $\|f_{1,q}(x) - f_{1,q}(y)\|^2$ is the number of position where one of $f_{1,q}(x)$, $f_{1,q}(y)$ has 1 and the other 0, and thus it equals $\Big|\lfloor qx \rfloor - \lfloor qy \rfloor\Big|$. This differs from $q|x-y|$ at most by 2, and we are done with the $d=1$ case.

For larger d, $f_{d,1}(\mathbf{x})$ is defined as the dq-component vector obtained by concatenating $f_{1,q}(x_1)$, $f_{1,q}(x_2)$,..., $f_{1,q}(x_d)$. The error bound is obvious using the 1-dimensional case. $\square$

Proof of the theorem. For contradiction, let us assume that there exists an equilateral set in $\mathbb{R}^d$ with the ℓ_1 distance that has at least $100d^4$ points. After possibly discarding some points we may assume that it has exactly $n := 100d^4$ points.

We re-scale the set so that the interpoint distances become $\frac{1}{2}$, and we translate it so that one of the points is $(\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2})$. Then the set is fully contained in $[0,1]^d$.

We use the lemma on approximate embedding with $q := 40d^3$. Applying the mapping $f_{d,q}$ to our set, we obtain an n-point set in $\mathbb{R}^{qd}$, for which the squared Euclidean distance of every two points is between $\frac{q}{2} - 2d$ and $\frac{q}{2} + 2d$. After re-scaling by $\sqrt{2/q}$, we get an approximately equilateral set with squared Euclidean interpoint distances between $1 - \frac{4d}{q}$ and $1 + \frac{4d}{q}$. We have $\frac{4d}{q} = \frac{1}{10d^2} = \frac{1}{\sqrt{n}}$, and thus the proposition on approximately equilateral sets applies and shows that $n \le 2(dq+2)$. But this is a contradiction, since $n = 100d^4$, while $2(dq+2) = 2(40d^4+2) < 100d^4$. The theorem is proved. $\square$

Sources. N. Alon and P. Pudlák: Equilateral sets in l_p^n, *Geometric and Functional Analysis* 13(2003), 467–482.

Our presentation via an approximate embedding is slightly different.

8 Rotating the Cube

First we state two beautiful geometric theorems. Since we need them only for motivation, we will not discuss the proofs, which involve methods of algebraic topology. Let $S^{n-1} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = 1\}$ stand for the unit sphere in $\mathbb{R}^n$, where $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$ denotes the Euclidean norm. Thus, for example, S^2 is the usual 2-dimensional unit sphere in the 3-dimensional space.

(T1) For every continuous function $f\colon S^2 \to \mathbb{R}$ there exist three mutually orthogonal unit vectors $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ with $f(\mathbf{p}_1) = f(\mathbf{p}_2) = f(\mathbf{p}_3)$.

(T2) Let $\alpha \in (0, 2]$ and let $f\colon S^{n-1} \to \mathbb{R}^{n-1}$ be an arbitrary continuous mapping. Then there are two points $\mathbf{p}, \mathbf{q} \in S^n$ whose Euclidean distance is exactly α and such that $f(\mathbf{p}) = f(\mathbf{q})$. In popular terms, at any given moment there are two places on the Earth's surface that are exactly 1234 km apart and have the same temperature and the same barometric pressure.

Theorem (T2) probably motivated Bronisław Knaster to pose the following question in 1947:

Knaster's question. *Is it true that for every continuous mapping $f\colon S^{n-1} \to \mathbb{R}^m$, where $n-1 \geq m \geq 1$, and every set K of $n-m+1$ points on S^{n-1} there exists a rotation ρ of $\mathbb{R}^n$ around the origin such that all points of the rotated set ρK have the same value of f?*

It is easily seen that a positive answer to Knaster's question for all m, n would contain both (T1) and (T2) as special cases. In particular, the second theorem deals exactly with the case $m = n-1$ of Knaster's question.

Somewhat disappointingly, though, the claim in Knaster's question does *not* hold for all n, m, as was discovered in the 1980s. Actually, it almost *never* holds: By now counterexamples are known for every n and m such that $n-1 > m \geq 2$, and also for $m = 1$ and all n sufficiently large.[3]

[3] This doesn't kill the question, though: It remains to understand for which sets K the claim does hold, and this question is very interesting and very far from solved.

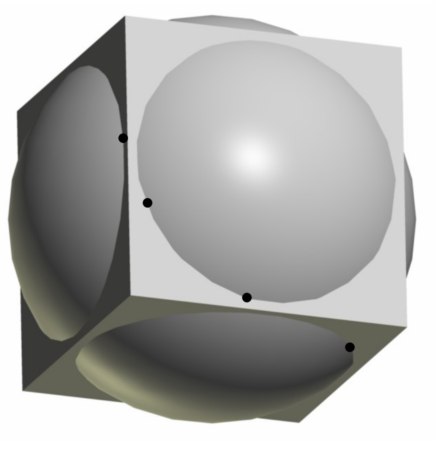
Here we discuss a counterexample for the last of these cases, namely, $m = 1$ (with some suitable large n). It was found only in 2003, after almost all of the other cases had been settled.

Theorem. *There exist an integer n, a continuous function $f\colon S^{n-1} \to \mathbb{R}$, and an n-point set $K \subset S^{n-1}$ such that for every rotation ρ of $\mathbb{R}^n$ around $\mathbf{0}$, the function f attains at least two distinct values on ρK.*

The function f in the proof is very simple, namely, $f(\mathbf{x}) = \|\mathbf{x}\|_\infty := \max\{|x_1|, |x_2|, \ldots, |x_n|\}$. The sophistication is in constructing K and proving the required property.

Some geometric intuition, not really necessary. The maximum value of f on S^{n-1} is obviously 1, attained at the points $\pm\mathbf{e}_1, \ldots, \pm\mathbf{e}_n$. With a little more effort one finds that the minimum of f on S^{n-1} equals $n^{-1/2}$, attained at points of the form $(\pm n^{-1/2}, \pm n^{-1/2}, \ldots, \pm n^{-1/2})$.

Let us now consider the function $f(\mathbf{x}) = \|\mathbf{x}\|_\infty$ on all of $\mathbb{R}^n$. Then the set $\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_\infty = 1\}$ is the surface of the unit cube $[-1, 1]^n$, and more generally, the level set $\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_\infty = t\}$ is the surface of the scaled cube $[-t, t]^n$. Thus, if K is a point set on S^{n-1}, finding a rotation ρ such that f is constant on ρK can be reformulated as follows: Find a scaling factor t and a rotation of the scaled cube $[-t, t]^n$ such that all points of K lie on the surface of the rotated and scaled cube.



In the proof of the theorem, K is chosen as the disjoint union of two sets K_1 and K_2, and these are constructed in such a way that if K_1 should lie on the surface of a rotated and scaled cube, then the scaling factor t has to be *large* (which means, geometrically, that the points of K_1 must be placed far from the corners of the cube), while for K_2 the scaling factor has to be *small* (the points of K_2 must be close to the cube corners). Hence it is impossible for both K_1 and K_2 to lie on the surface of the same scaled and rotated cube.

Preliminaries. In the theorem we deal with a point set K in the $(n-1)$-dimensional unit sphere and its rotated copies ρK. In the proof it will be more convenient to work with a set $\overline{K}$ living in the unit sphere S^{d-1} of a suitable lower dimension. Then, instead of rotations, we consider isometries $\varphi\colon \mathbb{R}^d \to \mathbb{R}^n$, that is, linear maps such that $\|\varphi(\mathbf{x})\| = \|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{R}^d$. If φ_0 is one such isometry, then $K := \varphi_0(\overline{K})$ is a point set in S^{n-1}, and the sets $\varphi(\overline{K})$ for all other isometries $\varphi\colon \mathbb{R}^d \to \mathbb{R}^n$ are exactly all rotated copies of K (and their mirror reflections—but for the purposes of the proof we can ignore the mirror reflections).

We need one more definition. Let $X \subseteq \mathbb{R}^n$ be a set and let $\delta > 0$ be a real number. A set $N \subseteq X$ is called *δ-dense in X* if for every $\mathbf{x} \in X$ there exists $\mathbf{y} \in N$ such that $\|\mathbf{x} - \mathbf{y}\| \leq \delta$.

Lemma K1.

(i) *Let $\varphi\colon \mathbb{R}^d \to \mathbb{R}^n$ be an isometry. Then there exists $\mathbf{x} \in S^{d-1}$ such that $\|\varphi(\mathbf{x})\|_\infty \geq \sqrt{d/n}$.*

(ii) *Let, moreover, $\overline{K}_1 \subset S^{d-1}$ be a $\frac{1}{2}$-dense set in S^{d-1}. Then there exists $\overline{\mathbf{p}} \in \overline{K}_1$ with $\|\varphi(\overline{\mathbf{p}})\|_\infty \geq \frac{1}{2}\sqrt{d/n}$.*

Proof. We begin with part (i). Let A be the matrix of the isometry φ with respect to the standard bases; i.e., the ith column of A is the vector $\varphi(\mathbf{e}_i) \in \mathbb{R}^n$, $i = 1, 2, \ldots, d$. Since φ preserves the Euclidean norm, the columns of A are unit vectors in $\mathbb{R}^n$, and thus

$$\sum_{i=1}^{n} \sum_{j=1}^{d} a_{ij}^2 = d. \tag{8}$$

Let $\mathbf{a}_i \in \mathbb{R}^d$ denote the ith *row* of A. For $\mathbf{x} \in \mathbb{R}^d$, the ith coordinate of $\varphi(\mathbf{x})$ is the scalar product $\langle \mathbf{a}_i \,|\, \mathbf{x} \rangle$, and thus $\|\varphi(\mathbf{x})\|_\infty = \max\{|\langle \mathbf{a}_i \,|\, \mathbf{x} \rangle| : i = 1, 2, \ldots, n\}$.

Now (8) tells us that $\sum_{i=1}^n \|\mathbf{a}_i\|^2 = d$, and thus there is an i_0 with $\|\mathbf{a}_{i_0}\| \geq \sqrt{d/n}$. Setting $\mathbf{x} := \mathbf{a}_{i_0}/\|\mathbf{a}_{i_0}\|$, we have $\|\varphi(\mathbf{x})\|_\infty \geq \langle \mathbf{a}_{i_0} \,|\, \mathbf{x} \rangle = \|\mathbf{a}_{i_0}\| \geq \sqrt{d/n}$, which finishes the proof of part (i).

We proceed with part (ii), which is the result that we will actually use later on. The proof is somewhat more clever than one might perhaps expect at first sight.

In the setting of (ii), we let $M := \sup\{\|\varphi(\mathbf{x})\|_\infty : \mathbf{x} \in S^{d-1}\}$, and let $\mathbf{x}_0 \in S^{d-1}$ be a point where M is attained.[4] By part (i) we have $M \geq \sqrt{d/n}$.

Since $\overline{K}_1$ is $\frac{1}{2}$-dense, we can choose a point $\overline{\mathbf{p}} \in \overline{K}_1$ with $\|\mathbf{x}_0 - \overline{\mathbf{p}}\| \leq \frac{1}{2}$. If, by chance, $\overline{\mathbf{p}} = \mathbf{x}_0$, we are done, and so we may assume $\overline{\mathbf{p}} \neq \mathbf{x}_0$ and let $\mathbf{v} := (\mathbf{x}_0 - \overline{\mathbf{p}})/\|\mathbf{x}_0 - \overline{\mathbf{p}}\| \in S^{d-1}$ be the unit vector in direction $\mathbf{x}_0 - \overline{\mathbf{p}}$. Then $\|\varphi(\mathbf{v})\|_\infty \leq M$ by the choice of M, and thus $\|\varphi(\mathbf{x}_0 - \overline{\mathbf{p}})\|_\infty \leq \frac{1}{2}M$. Then, using the triangle inequality for the $\|.\|_\infty$ norm, we have

$$\|\varphi(\overline{\mathbf{p}})\|_\infty \geq \|\varphi(\mathbf{x}_0)\|_\infty - \|\varphi(\mathbf{x}_0 - \overline{\mathbf{p}})\|_\infty \geq M - \tfrac{1}{2}M = \tfrac{1}{2}M \geq \tfrac{1}{2}\sqrt{n/d}.$$

This proves part (ii). $\square$

Lemma K2. *Let $\overline{K}_2$ be a set of m distinct points of the unit circle $S^1 \subset \mathbb{R}^2$. If t is a number such that there exists an isometry $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^n$ with $\|\varphi(\overline{\mathbf{p}})\|_\infty = t$ for all $\overline{\mathbf{p}} \in \overline{K}_2$, then $t \leq \sqrt{8/m}$.*

Proof. We begin in the same way as in the proof of Lemma K1, this time setting $d = 2$: A is the matrix of φ and $\mathbf{a}_i \in \mathbb{R}^2$ is its ith row. By (8) we have $\sum_{i=1}^n \|\mathbf{a}_i\|^2 = 2$. We are going to bound the left-hand side from below in terms of m and t.

Since the ith coordinate of $\varphi(\overline{\mathbf{p}})$ equals $\langle \mathbf{a}_i \,|\, \overline{\mathbf{p}}\rangle$, the condition $\|\varphi(\overline{\mathbf{p}})\|_\infty = t$ for all $\overline{\mathbf{p}} \in \overline{K}_2$ can be reformulated as follows:

(C1) For every $\overline{\mathbf{p}} \in \overline{K}_2$ there exists an $i = i(\overline{\mathbf{p}})$ with $|\langle \mathbf{a}_i \,|\, \overline{\mathbf{p}}\rangle| = t$.

(C2) For all $\overline{\mathbf{p}} \in \overline{K}_2$ and all i we have $|\langle \mathbf{a}_i \,|\, \overline{\mathbf{p}}\rangle| \leq t$.

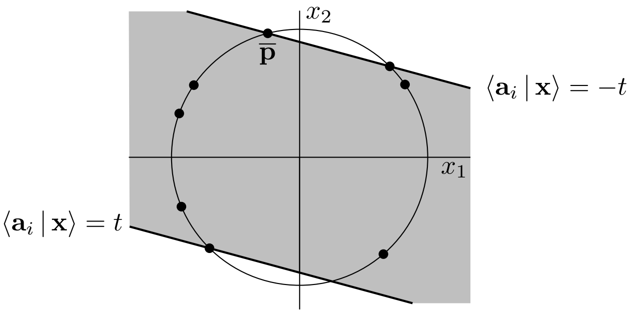
From (C1) we can infer that

$$\text{if } i = i(\overline{p}) \text{ for some } \overline{\mathbf{p}} \in \overline{K}_2, \text{ then } \|\mathbf{a}_i\| \geq t. \tag{9}$$

Indeed, $\overline{\mathbf{p}}$ is a unit vector, so $|\langle \mathbf{y} \,|\, \overline{\mathbf{p}}\rangle| \leq \|\mathbf{y}\|$ for all $\mathbf{y}$, and thus $|\langle \mathbf{a}_i \,|\, \overline{\mathbf{p}}\rangle| = t$ implies $\|\mathbf{a}_i\| \geq t$.

It remains to show that there are *many* distinct i with $i = i(\overline{\mathbf{p}})$ for some $\overline{\mathbf{p}} \in \overline{K}_2$. To this end, we observe that any given i can serve as $i(\overline{\mathbf{p}})$ for at most 4 distinct points $\overline{\mathbf{p}}$. This can be seen from the following geometric picture:

[4] The maximum is attained because S^{d-1} is compact. Readers not familiar enough with compactness may as well consider $\mathbf{x}_0$ such that $\|\varphi(\mathbf{x}_0)\|_\infty \geq 0.99M$, say, which clearly exists. Then the constants in the proof need a minor adjustment.



The condition $i = i(\overline{\mathbf{p}})$ means that the point $\overline{\mathbf{p}}$ lies on one of the lines $\{\mathbf{x} \in \mathbb{R}^2 : \langle \mathbf{a}_i \,|\, \mathbf{x}\rangle = t\}$ and $\{\mathbf{x} \in \mathbb{R}^2 : \langle \mathbf{a}_i \,|\, \mathbf{x}\rangle = -t\}$, and (C2) implies that all points of $\overline{K}_2$ lie within the parallel strip between these two lines. In this situation, the boundary of such a parallel strip can contain at most 4 points of $\overline{K}_2$ (actually, at most 2 points provided that $\overline{K}_2$ is chosen in a suitably general position).

Consequently, there are at least $m/4$ distinct vectors of Euclidean norm at least t among the $\mathbf{a}_i$, and so $\sum_{i=1}^{n} \|\mathbf{a}_i\|^2 \geq tm/4$. Since we already know that the left-hand side equals 2, we arrive at the claim of Lemma K2. $\square$

Two ways of making δ-dense sets. The last missing ingredient for the proof of the theorem is a way of making a $\frac{1}{2}$-dense set $\overline{K}_1$ in S^{d-1}, as in Lemma K1(ii), that is not too large. More precisely, it will be enough to know that for every $d \geq 1$ such a $\overline{K}_1$ exists of size at most $g(d)$, for an arbitrary function g.

This is a well-known geometric result. One somewhat sloppy but quick way of proving it starts by observing that the integer grid $\mathbb{Z}^d$ is $\sqrt{d}$-dense in $\mathbb{R}^d$ (actually $\frac{1}{2}\sqrt{d}$-dense). If we re-scale it by $1/(4\sqrt{d})$ and intersect it with the cube $[-1,1]^d$, we have a $\frac{1}{4}$-dense set N_0 in that cube, of size at most $(8\sqrt{d}+1)^d$. Finally, for every point $\mathbf{x} \in N_0$ that has distance at most $\frac{1}{4}$ to S^{d-1}, we choose a point $\mathbf{y} \in S^{d-1}$ at most $\frac{1}{4}$ apart from $\mathbf{x}$, and we let $N \subset S^{d-1}$ consist of all these $\mathbf{y}$. It is easily checked that N is $\frac{1}{2}$-dense in S^{d-1}. This yields $g(d)$ of order $d^{O(d)}$.

Another proof, the standard "textbook" one, uses a greedy algorithm and a volume argument. We place the first point $\mathbf{p}_1$ to S^{d-1} arbitrarily, and having already chosen $\mathbf{p}_1, \ldots, \mathbf{p}_{i-1}$, we place $\mathbf{p}_i$ to S^{d-1} so that it has distance at least $\frac{1}{2}$ from $\mathbf{p}_1, \ldots, \mathbf{p}_{i-1}$. This process finishes as soon as we can no longer place the next point, i.e., the resulting set is $\frac{1}{2}$-dense. To estimate the number m of points produced in this way, we observe that the balls of radius $\frac{1}{4}$ around the $\mathbf{p}_i$ are all disjoint and contained in the ball of radius $\frac{5}{4}$ around $\mathbf{0}$. Thus, the total volume of the small balls is at most the volume of the large ball, and this gives $m \le 5^d$, a better estimate than for the grid-based argument.

Proof of the theorem. We choose an even $n \ge 2g(100)$, where $g(d)$ is the function as above, we let $\overline{K}_1$ be a $\frac{1}{2}$-dense set in S^{99} of size at most $\frac{n}{2}$, and $\overline{K}_2$ is a set of $\frac{n}{2}$ points in S^1. We let $K := K_1 \cup K_2$, where $K_1, K_2 \subset S^{n-1}$ are isometric images of $\overline{K}_1$ and $\overline{K}_2$, respectively.

Lemma K1(ii) shows that for every rotation ρ there is a point $\mathbf{p} \in \rho K_1$ with $\|\mathbf{p}\|_\infty \ge \frac{1}{2}\sqrt{100/n} > 4n^{-1/2}$. On the other hand, if ρ is a rotation such that $\|\mathbf{p}\|_\infty$ equals the same number t for all $\mathbf{p} \in \rho K_2$, then $t \le \sqrt{16/n} = 4n^{-1/2}$ by Lemma K2. This proves that $K = K_1 \cup K_2$ cannot be rotated so that all of its points have the same $\|.\|_\infty$ norm. $\square$

Sources. B. S. Kashin, S. J. Szarek: The Knaster problem and the geometry of high-dimensional cubes, *C. R. Acad. Sci. Paris, Ser. I* 336(2003), 931–936.

9 Set Pairs and Exterior Products

We prove yet another theorem about intersection properties of sets.

Theorem. *Let $A_1, A_2, \ldots, A_n$ be k-element sets, let $B_1, B_2, \ldots, B_n$ be ℓ-element sets, and let*

(i) *$A_i \cap B_i = \emptyset$ for all $i = 1, 2, \ldots, n$, while*

(ii) *$A_i \cap B_j \ne \emptyset$ for all i, j with $1 \le i < j \le n$.*

Then $n \le \binom{k+\ell}{k}$.

It is easy to understand where $\binom{k+\ell}{k}$ comes from: Let $X := \{1, 2, \ldots, k+\ell\}$, let $A_1, A_2, \ldots, A_n$ be a list of all k-element subsets of X, and let's set $B_i := X \setminus A_i$ for every i. Then the A_i and B_i meet the conditions of the theorem and n equals $\binom{k+\ell}{k}$.

The perhaps surprising thing is that we can't produce more sets satisfying (i) and (ii) even if we use a much larger ground set (note that the theorem doesn't put any restrictions on the number of elements in the union of the A_i and B_i; it only limits their size and intersection pattern).

The so-called *set-pair method* applies the above theorem and similar results in graph and hypergraph theory, combinatorial geometry, and theoretical computer science, and it yields a number of interesting results. We won't discuss these applications here, though. The theorem is included mainly because of the proof method, where we briefly meet a remarkable mathematical object, the exterior algebra of a vector space.

The theorem is known in the literature as the **skew Bollobás theorem**. Bollobás originally proved a weaker (non-skew) version, where condition (ii) is strengthened to

(ii′) $A_i \cap B_j \neq \emptyset$ for all $i, j = 1, 2, \ldots, n$, $i \neq j$.

That version has a short probabilistic (or, if you prefer, double-counting) proof. However, for the skew version only linear-algebraic proofs are known—one of them uses the so-called polynomial method, and a second one, shown here, is a simple instance of another powerful method.

We begin with a simple claim asserting the existence of arbitrarily many vectors "in general position".

Claim. *For every $d \geq 1$ and every $m \geq 1$ there exist vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m \in \mathbb{R}^d$ such that every d or fewer among them are linearly independent.*

Proof. We fix m distinct and nonzero real numbers $t_1, t_2, \ldots, t_m$ arbitrarily and set $\mathbf{v}_i := (t_i, t_i^2, \ldots, t_i^d)$ (these are points on the so-called **moment curve** in $\mathbb{R}^d$).

Since this construction is symmetric, it suffices to check linear independence of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d$ (we assume $m \geq d$, for otherwise, the result is trivial). So let $\sum_{j=1}^d \alpha_j \mathbf{v}_j = \mathbf{0}$. This means $\sum_{j=1}^d \alpha_j t_i^j = 0$ for all i, i.e., $t_1, \ldots, t_d$ are roots of the polynomial $p(x) := \alpha_d x^d + \alpha_{d-1} x^{d-1} + \cdots + \alpha_1 x$. But 0 is another root, so we have $d+1$ distinct roots altogether, and since $p(x)$ has degree at most d, it cannot have $d+1$ distinct roots unless it's the zero polynomial. So $\alpha_1 = \alpha_2 = \cdots = \alpha_d = 0$.

Alternatively, one can prove the linear independence of the $\mathbf{v}_i$ using the Vandermonde determinant (usually computed in introductory courses of linear algebra).

Yet another proof follows easily by induction if one believes that $\mathbb{R}^d$ is not the union of finitely many $(d-1)$-dimensional linear subspaces. (But proving this rigorously is probably as complicated as the proof above.) □

On permutations and signs. We recall that the sign of a permutation $\pi\colon \{1, 2, \ldots, d\} \to \{1, 2, \ldots, d\}$ can be defined as $\operatorname{sgn}(\pi) = (-1)^{\operatorname{inv}(\pi)}$, where $\operatorname{inv}(\pi) = |\{(i,j) : 1 \leq i < j \leq d \text{ and } \pi(i) > \pi(j)\}|$ is the number of *inversions* of π.

Let d be a fixed integer and let $\mathbf{s} = (s_1, s_2, \ldots, s_k)$ be a sequence of integers from $\{1, 2, \ldots, d\}$. We analogously define the sign of $\mathbf{s}$ as

$$\operatorname{sgn}(\mathbf{s}) := \begin{cases} (-1)^{\operatorname{inv}(\mathbf{s})} & \text{if all terms in } \mathbf{s} \text{ are distinct,} \\ 0 & \text{otherwise,} \end{cases}$$

where $\operatorname{inv}(\mathbf{s}) = |\{(i,j) : 1 \le i < j \le k \text{ and } s_i > s_j\}|$.

If we regard a permutation π as the sequence $(\pi(1), \pi(2), \ldots, \pi(d))$, then both definitions of the sign agree, of course.

The exterior algebra of a finite-dimensional vector space. In 1844 Hermann Grassmann, a high-school teacher in Stettin (in Prussia at that time, then in Germany, and nowadays in Poland spelled Szczecin), published a book proposing a new algebraic foundation for geometry. He developed foundations of linear algebra more or less as we know it today, and went on to introduce "exterior product" of vectors, providing a unified and coordinate-free treatment of lengths, areas, and volumes. His revolutionary mathematical discoveries were not appreciated during his lifetime (he became famous as a linguist), but later on, they were completed and partially re-developed by others. They belong among the fundamental concepts of modern mathematics, with many applications e.g. in differential geometry, algebraic geometry, and physics.

Here we will build the **exterior algebra** (also called the **Grassmann algebra**) of a finite-dimensional space in a minimalistic way (which is not the most conceptual one), checking only the properties we need for the proof of the above theorem.

Proposition. *Let V be a d-dimensional vector space.*[5] *Then there is a countable sequence $W_0, W_1, W_2, \ldots$ of vector spaces (among with only $W_0, \ldots, W_d$ really matter) and a binary operation $\wedge$ ("exterior product" or "wedge product") on $W_0 \cup W_1 \cup W_2 \cdots$ with the following properties:*

(EA1) $\dim W_k = \binom{d}{k}$. *In particular, W_1 is isomorphic to V, while $W_k = \{\mathbf{0}\}$ for $k > d$.*

(EA2) *If $\mathbf{u} \in W_k$ and $\mathbf{v} \in W_\ell$, then $\mathbf{u} \wedge \mathbf{v} \in W_{k+\ell}$.*

(EA3) *The exterior product is* **associative**, *i.e.*, $(\mathbf{u} \wedge \mathbf{v}) \wedge \mathbf{w} = \mathbf{u} \wedge (\mathbf{v} \wedge \mathbf{w})$.

(EA4) *The exterior product is* **bilinear**, *i.e.*, $(\alpha\mathbf{u} + \beta\mathbf{v}) \wedge \mathbf{w} = \alpha(\mathbf{u} \wedge \mathbf{w}) + \beta(\mathbf{v} \wedge \mathbf{w})$ *and* $\mathbf{u} \wedge (\alpha\mathbf{v} + \beta\mathbf{w}) = \alpha(\mathbf{u} \wedge \mathbf{v}) + \beta(\mathbf{u} \wedge \mathbf{w})$.

[5] Over any field, but we'll use only the real case.

(EA5) *The exterior product reflects linear dependence in the following way: For any* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d \in W_1$*, we have* $\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \cdots \wedge \mathbf{v}_d = \mathbf{0}$ *if and only if* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d$ *are linearly dependent.*

Proof. Let $\mathcal{F}_k$ denote the set of all k-element subsets of $\{1, 2, \ldots, k\}$. For each $k = 0, 1, \ldots, d$ we fix some $\binom{d}{k}$-dimensional vector space W_k, and let us fix a basis $(\mathbf{b}_K : K \in \mathcal{F}_k)$ of W_k. Here $\mathbf{b}_K$ is just a *name* for a vector in the basis, which will be notationally more convenient than the usual indexing of a basis by integers $1, 2, \ldots$. We set, trivially, $W_{d+1} = W_{d+2} = \cdots = \{\mathbf{0}\}$.

We first define the exterior product on the basis vectors. Let $K, L \subseteq \{1, 2, \ldots, d\}$, where $s_1 < s_2 < \cdots < s_k$ are the elements of K in increasing order and $t_1 < \cdots < t_\ell$ are the elements of L in increasing order. Then we set

$$\mathbf{b}_K \wedge \mathbf{b}_L := \begin{cases} \operatorname{sgn}((s_1, s_2, \ldots, s_k, t_1, t_2, \ldots, t_\ell))\mathbf{b}_{K \cup L} & \text{if } k + \ell \leq d, \\ \mathbf{0} \in W_{k+\ell} & \text{if } k + \ell > d. \end{cases}$$

We note that, in particular, for $K \cap L \neq \emptyset$ we have $\mathbf{b}_K \wedge \mathbf{b}_L = \mathbf{0}$, since then the sequence $(s_1, s_2, \ldots, s_k, t_1, t_2, \ldots, t_\ell)$ has a repeated term and thus its sign is 0. The signs are a bit tricky, but they are crucial for the good behavior of the exterior product with respect to linear independence, i.e., (EA5).

We extend $\wedge$ to all vectors bilinearly: If $\mathbf{u} \in W_k$ and $\mathbf{v} \in W_\ell$, we write them in the appropriate bases as $\mathbf{u} = \sum_{K \in \mathcal{F}_k} \alpha_K \mathbf{b}_K$, $\mathbf{v} = \sum_{L \in \mathcal{F}_\ell} \beta_L \mathbf{b}_L$, and we put

$$\mathbf{u} \wedge \mathbf{w} := \sum_{K \in \mathcal{F}_k, L \in \mathcal{F}_\ell} \alpha_K \beta_L (\mathbf{b}_K \wedge \mathbf{b}_L).$$

Now (EA1), (EA2), and (EA4) (bilinearity) are clear.

As for the associativity (EA3), it suffices to check it for basis vectors, i.e., to verify

$$(\mathbf{b}_K \wedge \mathbf{b}_L) \wedge \mathbf{b}_M = \mathbf{b}_K \wedge (\mathbf{b}_L \wedge \mathbf{b}_M) \tag{10}$$

for all K, L, M. The interesting case is when K, L, M are pairwise disjoint and $|K| + |L| + |M| \leq d$. Then, obviously, both sides of (10) are $\pm \mathbf{b}_{K \cup L \cup M}$, and it suffices to check that the signs match.

To this end, we let $s_1 < \cdots < s_k$ be the elements of K in increasing order, and similarly for $t_1 < \cdots < t_\ell$ and L and for $z_1 < \cdots < z_m$ and M. By counting the inversions of the appropriate sequences, we find that $(\mathbf{b}_K \wedge \mathbf{b}_L) \wedge \mathbf{b}_M = (-1)^N \mathbf{b}_{K \cup L \cup M}$, where $N = \operatorname{inv}((s_1, \ldots, s_k, t_1, \ldots, t_\ell)) +$

$\mathrm{inv}((s_1,\ldots,s_k,z_1,\ldots,z_m))+\mathrm{inv}((t_1,\ldots,t_\ell,z_1,\ldots,z_m))$, and the right-hand side of (10) comes out the same.

Next, if K,L,M are not pairwise disjoint or $k+\ell+m>d$, it is easily checked that both sides of (10) are $\mathbf{0}\in W_{k+\ell+m}$. Finally, having checked (10), it is routine to verify associativity in general—one just writes out the three participating vectors in the respective bases, expands both sides using bilinearity, and uses (10).

It remains to prove (EA5), which is the most interesting part where, finally, the choice of the sign turns from a hassle into a blessing.

Let $\mathbf{v}_1,\ldots,\mathbf{v}_d\in W_1$ be arbitrary, and let's write them in the basis $\mathbf{b}_{\{1\}},\ldots,\mathbf{b}_{\{d\}}$ of W_1 as

$$\mathbf{v}_i=\sum_{j=1}^{d}a_{ij}\mathbf{b}_{\{j\}}.$$

Then, using bilinearity and associativity, we have

$$\mathbf{v}_1\wedge\mathbf{v}_2\wedge\cdots\wedge\mathbf{v}_d=\sum_{j_1,j_2,\ldots,j_d=1}^{n}a_{1j_1}a_{2j_2}\cdots a_{dj_d}\mathbf{b}_{j_1}\wedge\mathbf{b}_{j_2}\wedge\cdots\wedge\mathbf{b}_{j_d}.$$

By the definition of the exterior product of basis vectors, all terms on the right-hand side where some two j_i coincide are $\mathbf{0}$. What remains is a sum over all d-tuples of distinct j_i's, in other words, over all permutations of $\{1,2,\ldots,d\}$:

$$\mathbf{v}_1\wedge\cdots\wedge\mathbf{v}_d=\sum_{\pi}a_{1\pi(1)}a_{2\pi(2)}\cdots a_{d\pi(d)}\mathbf{b}_{\pi(1)}\wedge\mathbf{b}_{\pi(2)}\wedge\cdots\wedge\mathbf{b}_{\pi(d)}.$$

By considerations very similar to those in checking the associativity, we find that $\mathbf{b}_{\pi(1)}\wedge\mathbf{b}_{\pi(2)}\wedge\cdots\wedge\mathbf{b}_{\pi(d)}=\mathrm{sgn}(\pi)\mathbf{b}_{\{1,2,\ldots,d\}}$. Then the last sum transforms into $\det(A)\mathbf{b}_{\{1,2,\ldots,d\}}$, which is $\mathbf{0}$ exactly if the $\mathbf{v}_i$ are linearly dependent. The proposition is proved. □

With just a little more effort, (EA5) can be extended to any number of vectors; i.e., $\mathbf{v}_1,\ldots\mathbf{v}_n\in W_1$ are linearly dependent exactly if their exterior product is $\mathbf{0}$ (we won't need this but not mentioning it seems inappropriate).

Proof of the theorem. Let $d:=k+\ell$ and let us consider the exterior algebra of $\mathbb{R}^d$ as in the proposition, with the vector spaces $W_0,W_1,\ldots$ and the operation $\wedge$. Let us assume, without loss of generality, that $A_1\cup\cdots\cup A_n\cup B_1\cup\cdots\cup B_n=\{1,2,\ldots,m\}$ for some integer m, and let us fix m

vectors $\mathbf{v}_1, \dots \mathbf{v}_m \in W_1 \cong \mathbb{R}^d$ in general position according to the claim above (every d or fewer linearly independent).

For a subset $A = \{i_1, i_2, \dots, i_r\} \subseteq \{1, 2, \dots, m\}$, where $i_1 < i_2 < \dots < i_r$, we set

$$\mathbf{w}_A := \mathbf{v}_{i_1} \wedge \mathbf{v}_{i_2} \wedge \dots \wedge \mathbf{v}_{i_r}.$$

Thus, $\mathbf{w}_A \in W_r$. For $A, B \subseteq \{1, 2, \dots, m\}$ with $|A| + |B| = d$, (EA3) and (EA5) yield

$$\mathbf{w}_A \wedge \mathbf{w}_B = \begin{cases} \pm\mathbf{w}_{A \cup B} \neq \mathbf{0} & \text{for } A \cap B = \emptyset, \\ \mathbf{0} & \text{for } A \cap B \neq \emptyset. \end{cases}$$

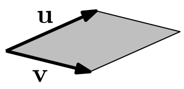
We claim that the n vectors $\mathbf{w}_{A_1}, \mathbf{w}_{A_2}, \dots, \mathbf{w}_{A_n} \in W_k$ are linearly independent. This will prove the theorem, since $\dim(W_k) = \binom{d}{k} = \binom{k+\ell}{k}$.

So let's assume $\sum_{i=1}^n \alpha_i \mathbf{w}_{A_i} = \mathbf{0}$. Assuming that, for some j, we already know that $\alpha_i = 0$ for all $i > j$ (for $j = n$ this is a void assumption), we show that $\alpha_j = 0$ as well. To this end, we consider the exterior product

$$\begin{aligned} \mathbf{0} \wedge \mathbf{w}_{B_j} &= \mathbf{0} = \left(\sum_{i=1}^n \alpha_i \mathbf{w}_{A_i} \right) \wedge \mathbf{w}_{B_j} \\ &= \sum_{i=1}^n \alpha_i (\mathbf{w}_{A_i} \wedge \mathbf{w}_{B_j}) = \alpha_j (\mathbf{w}_{A_j} \wedge \mathbf{w}_{B_j}), \end{aligned}$$

since $\mathbf{w}_{A_i} \wedge \mathbf{w}_{B_j} = 0$ for $i < j$ (using $A_i \cap B_j \neq \emptyset$), $\alpha_i = 0$ for $i > j$ by the inductive assumption, and $\mathbf{w}_{A_i} \wedge \mathbf{w}_{B_i} \neq \mathbf{0}$ since $A_i \cap B_i = \emptyset$. Thus, $\alpha_j = 0$, and the theorem is proved. □

The geometry of the exterior product at a glance. Some low-dimensional instances of the exterior product correspond to familiar concepts. First let $d = 2$ and let's identify W_1 with $\mathbb{R}^d$ so that $(\mathbf{b}_{\{1\}}, \mathbf{b}_{\{2\}})$ corresponds to the standard orthonormal basis $(\mathbf{e}_1, \mathbf{e}_2)$. Then it can be shown that $\mathbf{u} \wedge \mathbf{v} = \pm a \cdot \mathbf{e}_1 \wedge \mathbf{e}_2$, where a is the area of the parallelogram spanned by $\mathbf{u}$ and $\mathbf{v}$.



In $\mathbb{R}^3$, again making a similar identification of W_1 with $\mathbb{R}^3$, it turns out that $\mathbf{u} \wedge \mathbf{v}$ is closely related to the *cross product* of $\mathbf{u}$ and $\mathbf{v}$ (often used in physics), and $\mathbf{u} \wedge \mathbf{v} \wedge \mathbf{w} = \pm a \cdot \mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_3$, where a is the volume of the parallelepiped spanned by $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$. The latter, of course, is an instance of a general rule; in $\mathbb{R}^d$, the volume of the parallelepiped spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_d \in \mathbb{R}^d$ is $|\det(A)|$, where A is the matrix with the $\mathbf{v}_i$ as columns, and we've already verified that $\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_d = \det(A) \cdot \mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_d$.

These are only the first indications that the exterior algebra has a very rich geometric meaning. Generally, one can think of $\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k \in W_k$ as representing, uniquely up to a scalar multiple, the k-dimensional subspace of $\mathbb{R}^d$ spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_k$. However, by far not all vectors in W_k correspond to k-dimensional subspaces in this way; W_k can be thought of as a "closure" that completes the set of all k-dimensional subspaces into a vector space.

Sources. Bollobás' theorem was proved in

B. Bollobás: On generalized graphs, *Acta Math. Acad. Sci. Hung.* 16(1965), 447–452.

The exterior algebra method was introduced in this area by Lovász:

L. Lovász: Flats in matroids and geometric graphs, in *Combinatorial surveys (Proc. Sixth British Combinatorial Conf., Royal Holloway Coll., Egham, 1977)*, Academic Press, London, 1977, 45–86.

This paper contains a version of the Bollobás theorem for vector subspaces, and the proof implies the skew Bollobás theorem easily, but explicitly that theorem seems to appear first in

P. Frankl: An extremal problem for two families of sets, *European J. Combin.* 3,2(1982) 125–127,

where it's proved via *symmetric* tensor products (while the exterior product can be interpreted as an *antisymmetric* tensor product). The method with exterior products was also discovered independently by Kalai and used with great success in the study of convex polytopes and geometrically defined simplicial complexes:

G. Kalai: Intersection patterns of convex sets, *Israel J. Math.* 48(1984) 161–174.

Applications of the set-pair method are surveyed in two papers of Tuza, among which the second one

Zs. Tuza: Applications of the set-pair method in extremal problems, II, in *Combinatorics, Paul Erdős is eighty, Vol. 2*, J. Bolyai Math. Soc., Budapest, 1996, 459–490

has a somewhat wider scope.