The spectral theorem and Hadamard's determinant problem

Chapter 7

A fundamental theorem of linear algebra asserts that every symmetric real matrix A can be diagonalized. That is, for every such matrix A there is a nonsingular real matrix Q such that

$$Q^{-1}AQ = \begin{pmatrix} \lambda_1 & & & & & & \\ & \ddots & & & & & \\ & & \lambda_1 & & & O & \\ & & & \lambda_2 & & & \\ & & & & \ddots & & \\ & & & & & \lambda_t & \\ & & O & & & & \ddots \\ & & & & & & & \lambda_t \end{pmatrix}$$

is in diagonal form. The (real) λ_i's are the eigenvalues of A, and the columns of Q form a basis of eigenvectors. We will make use of this result in several chapters to come.

What's more, the matrix Q can be chosen to be an *orthogonal* matrix, which means that $Q^T = Q^{-1}$, or equivalently that the columns of Q form an orthonormal basis with respect to the usual inner product.

Theorem 1. *For every real symmetric matrix A there is a real orthogonal matrix Q such that Q^TAQ is diagonal.*

Moving Q and Q^T to the right-hand side we may equivalently express the theorem as a representation of A as a linear combination of matrices P_i that correspond to projections onto the eigenspaces C_{λ_i},

$$\begin{aligned} A &= \lambda_1 P_1 + \cdots + \lambda_t P_t, \\ I_n &= P_1 + \cdots + P_t, \end{aligned}$$

with $P_iP_j = \delta_{ij}P_i$ for all i and j. In this form the statement is usually called the *spectral theorem*.

The standard proofs of the theorem proceed by induction on the order of A (with some care in the presence of multiple eigenvalues), construct the basis of eigenvectors step by step, and use the fact that the characteristic polynomial splits into linear factors over the field $\mathbb{C}$ of complex numbers.

© Springer-Verlag GmbH Germany, part of Springer Nature 2018
M. Aigner, G. M. Ziegler, *Proofs from THE BOOK*, https://doi.org/10.1007/978-3-662-57265-8_7

The following proof due to Herb Wilf does it in one stroke and is truly inspired. It is very different from the usual proofs: It does not even refer to the eigenvalues, but instead employs an elegant compactness argument in a surprising way.

■ **Proof.** We start with some preliminary facts. Let $O(n) \subseteq \mathbb{R}^{n\times n}$ be the set of real orthogonal matrices of order n. Since

$$(PQ)^{-1} \;=\; Q^{-1}P^{-1} \;=\; Q^TP^T \;=\; (PQ)^T$$

for $P, Q \in O(n)$, we see that the set $O(n)$ is a group. Regarding any matrix in $\mathbb{R}^{n\times n}$ as a vector in $\mathbb{R}^{n^2}$, we find that $O(n)$ is a compact set. Indeed, as the columns of an orthogonal matrix $Q = (q_{ij})$ are unit vectors, we have $|q_{ij}| \le 1$ for all i and j, thus $O(n)$ is bounded. Furthermore, the set $O(n)$ is defined as a subset of $\mathbb{R}^{n^2}$ by the equations

$$x_{i1}x_{j1} + x_{i2}x_{j2} + \cdots + x_{in}x_{jn} = \delta_{ij} \quad \text{for } 1 \le i, j \le n,$$

hence it is closed, and thus compact.

The Heine–Borel theorem
Every closed and bounded subset of a vector space $\mathbb{R}^N$ is compact.

For any real square matrix A let $\mathrm{Od}(A) = \sum_{i\neq j} a_{ij}^2$ be the sum of the squares of the *off-diagonal* entries. Suppose we can prove the following.

Lemma. *If A is a real symmetric $n \times n$ matrix that is not diagonal, that is,* $\mathrm{Od}(A) > 0$*, then there exists $U \in O(n)$ such that* $\mathrm{Od}(U^TAU) < \mathrm{Od}(A)$.

Given the lemma, the theorem follows in three quick steps. Let A be a real symmetric $n \times n$ matrix.

(A) Consider the map $f_A : O(n) \to \mathbb{R}^{n\times n}$ with $f_A(P) := P^TAP$. The map f_A is continuous on the compact set $O(n)$, and so the image $f_A(O(n))$ is compact.

(B) The function $\mathrm{Od} : f_A(O(n)) \to \mathbb{R}$ is continuous, hence it assumes a minimum, say at $D = Q^TAQ \in f_A(O(n))$.

(C) The value $\mathrm{Od}(D)$ must be zero, and hence D is a *diagonal* matrix as required.

Indeed, if $\mathrm{Od}(D) > 0$, then applying the Lemma we find $U \in O(n)$ with $\mathrm{Od}(U^TDU) < \mathrm{Od}(D)$. But

$$U^TDU = U^TQ^TAQU = (QU)^TA(QU)$$

is in $f_A(O(n))$ (remember $O(n)$ is a group!) with Od-value smaller than that of D — contradiction, and end of proof.

It remains to prove the lemma, and for this we use a very clever method attributed to Carl Gustav Jacob Jacobi. Suppose that $a_{rs} \neq 0$ for some $r \neq s$. Then we claim that the matrix U that agrees with the identity matrix except that $u_{rr} = u_{ss} = \cos\vartheta$, $u_{rs} = \sin\vartheta$, $u_{sr} = -\sin\vartheta$ does the job, for some choice of the (real) angle ϑ:

$$U = \begin{pmatrix} & & & r & & s & \\ 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \cos\vartheta & & \sin\vartheta & r \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \\ & & & -\sin\vartheta & & \cos\vartheta & s \\ & & & & & & & 1 \\ & & & & & & & & \ddots \\ & & & & & & & & & 1 \end{pmatrix}$$

"Diagonalizing by applying a rotation and removing off-diagonal elements"

Clearly, U is orthogonal for any ϑ.
Now let us compute the (k,ℓ)-entry $b_{k\ell}$ of U^TAU. We have

$$b_{k\ell} = \sum_{i,j} u_{ik}a_{ij}u_{j\ell}. \tag{1}$$

For $k,\ell \notin \{r,s\}$ we get $b_{k\ell} = a_{k\ell}$. Furthermore, we have

$$\begin{aligned} b_{kr} &= \sum_{i=1}^{n} u_{ik} \sum_{j=1}^{n} a_{ij}u_{jr} \\ &= \sum_{i=1}^{n} u_{ik}(a_{ir}\cos\vartheta - a_{is}\sin\vartheta) \\ &= a_{kr}\cos\vartheta - a_{ks}\sin\vartheta \quad (\text{for } k \neq r,s). \end{aligned}$$

Similarly, one computes

$$b_{ks} = a_{kr}\sin\vartheta + a_{ks}\cos\vartheta \quad (\text{for } k \neq r,s).$$

It follows that

$$\begin{aligned} b_{kr}^2 + b_{ks}^2 &= a_{kr}^2\cos^2\vartheta - 2a_{kr}a_{ks}\cos\vartheta\sin\vartheta + a_{ks}^2\sin^2\vartheta \\ &\quad + a_{kr}^2\sin^2\vartheta + 2a_{kr}a_{ks}\sin\vartheta\cos\vartheta + a_{ks}^2\cos^2\vartheta \\ &= a_{kr}^2 + a_{ks}^2, \end{aligned}$$

and by symmetry

$$b_{r\ell}^2 + b_{s\ell}^2 = a_{r\ell}^2 + a_{s\ell}^2 \quad (\text{for } \ell \neq r,s).$$

We conclude that the function Od, which sums the squares of the off-diagonal values, agrees for A and U^TAU except for the entries at (r,s) and (s,r), for *any* ϑ. To conclude the proof we now show that ϑ_0 can be chosen suitably as to make $b_{rs} = 0$, which will result in

$$\mathrm{Od}(U^TAU) = \mathrm{Od}(A) - 2a_{rs}^2 < \mathrm{Od}(A)$$

as required.


Jacques Hadamard

Using (1) we find

$$b_{rs} \;=\; (a_{rr} - a_{ss}) \sin\vartheta \cos\vartheta + a_{rs}(\cos^2\vartheta - \sin^2\vartheta).$$

For $\vartheta = 0$ this becomes a_{rs}, while for $\vartheta = \frac{\pi}{2}$ it is $-a_{rs}$. Hence by the intermediate value theorem there is some ϑ_0 between 0 and $\frac{\pi}{2}$ such that $b_{rs} = 0$, and we are through. □

So this was beautiful, and we want to immediately apply the theorem to a famous (and unsolved) problem.

The Hadamard determinant problem
How large can $\det A$ *be on the set of all real* $n \times n$ *matrices* $A = (a_{ij})$ *with* $|a_{ij}| \leq 1$ *for all* i *and* j*?*

Since the determinant is a continuous function in the a_{ij} (considered as variables) and the matrices form a compact set in $\mathbb{R}^{n^2}$, this maximum must exist. Furthermore, the maximum is attained for some matrix all of whose entries are $+1$ or -1, because the function $\det A$ is linear in each single entry a_{ij} (if we keep all other entries fixed). Thus we can start with any matrix A and move one entry after the other to $+1$ or to -1, in every single step not decreasing the determinant, until we arrive at a ± 1-matrix. In the search for the largest determinant we may thus assume that all entries of A are ± 1.

Here is the trick: Instead of A we consider the matrix $B = A^TA = (b_{ij})$. That is, if $c_j = (a_{1j}, a_{2j}, \ldots, a_{nj})^T$ denotes the j-th column vector of A, then $b_{ij} = \langle c_i, c_j\rangle$, the inner product of c_i and c_j. In particular,

$$b_{ii} = \langle c_i, c_i\rangle = n \;\text{ for all }\; i,$$

and

$$\text{trace } B = \sum_{i=1}^{n} b_{ii} = n^2, \tag{2}$$

which will come in handy in a moment.

Now we can go to work. First of all, from $B = A^TA$ we get $|\det A| = \sqrt{\det B}$. Since multiplication of a column of A by -1 turns $\det A$ into $-\det A$, we see that the maximum problem for $\det A$ is the same as for $\det B$. Furthermore, we may assume that A is nonsingular, and hence that B is nonsingular as well.

Since $B = A^TA$ is a symmetric matrix the spectral theorem tells us that for some $Q \in O(n)$,

$$Q^TBQ = Q^TA^TAQ = (AQ)^T(AQ) = \begin{pmatrix} \lambda_1 & & & \\ & \ddots & & O \\ & & \ddots & \\ & O & & \lambda_n \end{pmatrix}, \tag{3}$$

where the λ_i are the eigenvalues of B. Now, if d_j denotes the j-th column vector of AQ (which is nonzero since A is nonsingular), then

$$\lambda_j = \langle d_j, d_j \rangle = \sum_{i=1}^{n} d_{ij}^2 > 0.$$

Thus $\lambda_1, \ldots, \lambda_n$ are positive real numbers and

$$\det B = \lambda_1 \cdots \lambda_n, \ \text{trace}\, B = \sum_{i=1}^{n} \lambda_i.$$

Whenever such a product and sum of positive numbers turn up, it is always a good idea to try the arithmetic-geometric mean inequality (see Chapter 20). In our case this gives with (2)

$$\det B = \lambda_1 \cdots \lambda_n \leq \Big(\frac{\sum_{i=1}^{n} \lambda_i}{n} \Big)^n = \Big(\frac{\text{trace}\, B}{n} \Big)^n = n^n, \qquad (4)$$

and out comes Hadamard's upper bound

$$|\det A| \ \leq \ n^{n/2}. \qquad (5)$$

When do we have equality in (5) or, what is the same, in (4)? Easy enough: if and only if the geometric mean of the λ_i's equals the arithmetic mean, or equivalently, if and only if $\lambda_1 = \cdots = \lambda_n = \lambda$. But then trace $B = n\lambda = n^2$, and so $\lambda_1 = \cdots = \lambda_n = n$. Looking at (3) this means $Q^T B Q = nI_n$, where I_n is the $n \times n$ identity matrix. Now recall $Q^T = Q^{-1}$, multiply by Q on the left, by Q^{-1} on the right, to obtain

$$B = nI_n.$$

Statements (5) and (6) form an instance of *Hadamard's inequality*: The absolute value of the determinant of a matrix is at most the product of the lengths of its columns, with equality if and only if the columns are pairwise orthogonal.

Going back to A this means that

$$|\det A| = n^{n/2} \iff \langle c_i, c_j \rangle = 0 \ \text{ for } \ i \neq j. \qquad (6)$$

Matrices A with ± 1-entries that achieve equality in (5) are aptly called *Hadamard matrices*. So an $n \times n$ matrix A with ± 1-entries is a Hadamard matrix if and only if

$$A^T A \ = \ A A^T \ = \ nI_n.$$

This leads to another unsolved and apparently very difficult problem:

> *For which n does a Hadamard matrix of size $n \times n$ exist?*

A short argument shows that if n is greater than 2, then it must be a multiple of 4. Indeed, suppose that A is an $n \times n$ Hadamard matrix, $n \geq 2$, whose rows are the vectors $r_1, \ldots, r_n$. Clearly, multiplication of any row or column by -1 gives another Hadamard matrix. So we may assume that the first row consists of 1's only. Since $\langle r_1, r_i \rangle = 0$ for $i \neq 1$, every other

row must contain $\frac{n}{2}$ 1's and $\frac{n}{2}$ -1's; in particular, n must be even. Assume now that $n > 2$ and consider rows r_2 and r_3, and denote by a, b, c, d the numbers of columns that have $\begin{smallmatrix}+1\\+1\end{smallmatrix}$, $\begin{smallmatrix}+1\\-1\end{smallmatrix}$, $\begin{smallmatrix}-1\\+1\end{smallmatrix}$, and $\begin{smallmatrix}-1\\-1\end{smallmatrix}$ in rows 2 and 3, respectively. Then from $\langle r_1, r_2\rangle = 0$ and $\langle r_1, r_3\rangle = 0$ we get

$$a + b = c + d = a + c = b + d = \frac{n}{2},$$

which gives $b = c$, $a = d$. But from $\langle r_2, r_3\rangle = 0$ we also have $a+d = b+c$, resulting in $2a = 2b$. We conclude that $a = b = c = d = \frac{n}{4}$. Thus the order of the Hadamard matrix is either $n = 1$ or $n = 2$, or $n = a+b+c+d = 4a$, a multiple of 4.

Does a Hadamard matrix exist for all $n = 4a$? No one knows. The answer is yes for n up to the current record $n = 664$, and for certain infinite series such as the powers of 2 (see the box). But the general answer seems at present out of reach.

Hadamard matrices exist for all $n = 2^m$

Consider an m-set X and index the 2^m subsets $C \subseteq X$ in any way $C_1, \ldots, C_{2^m}$. The matrix $A = (a_{ij})$ is defined as

$$a_{ij} = (-1)^{|C_i \cap C_j|}.$$

We want to verify $\langle r_i, r_j\rangle = 0$ for $i \neq j$. From the definition,

$$\langle r_i, r_j\rangle = \sum_k (-1)^{|C_i\cap C_k| + |C_j \cap C_k|}. \qquad (*)$$

Now, as $C_i \neq C_j$ there exists an element $a \in X$ with $a \in C_i \backslash C_j$ or $a \in C_j \backslash C_i$; suppose $a \in C_i \backslash C_j$. Half the subsets of X contain a, and half do not. Let C run through all subsets that contain a, then the pairs $\{C, C\backslash a\}$ will comprise all subsets of X. But for each such pair $\{C, C\backslash a\}$, $|C_i \cap C| + |C_j \cap C|$ and $|C_i \cap (C\backslash a)| + |C_j \cap (C\backslash a)|$ have different parity, and so the corresponding terms in $(*)$ will sum to 0. But then the whole sum is 0, as required.

For $n = 4$, with the numbering $C_1 = \varnothing$, $C_2 = \{1\}$, $C_3 = \{2\}$, $C_4 = \{1, 2\}$ this yields the matrix

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

For $n = 4a$ we have thus reduced the original problem to the existence of Hadamard matrices. But how large can $\det A$ be when n is *not* a multiple of 4? This is again a hard problem, but maybe we can find a good *lower* bound for the maximum. Here is a method that often proves successful — and it does in our case.

Let us look at *all* 2^{n^2} matrices with ± 1-entries and consider some averages of the determinant. The arithmetic mean $\frac{1}{2^{n^2}} \sum_A \det A$ is 0 (clear?), so this is no big help. But if we consider the *mean square average* instead,

$$D_n \;\coloneqq\; \sqrt{\frac{\sum_A (\det A)^2}{2^{n^2}}},$$

$$\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

Optimal matrices for $n = 2, 3$, and 4, with determinants 2, 4, and 16.

then things brighten up. Clearly,

$$\max_A \det A \;\geq\; D_n,$$

so this will give us a lower bound for the maximum.

The following stunningly simple calculation of D_n^2 probably appeared first in an article by George Szekeres and Paul Turán. We learnt it from a beautiful paper of Herb Wilf who heard it from Mark Kac. In the words of Mark Kac: "*Just write* $(\det A)^2$ *out twice, interchange summation, and everything simplifies.*" So we want to do just that.

From the definition of the determinant we get

$$\begin{aligned} D_n^2 &= \frac{1}{2^{n^2}} \sum_A \Big(\sum_\pi (\operatorname{sign}\pi) a_{1\pi(1)} a_{2\pi(2)} \cdots a_{n\pi(n)} \Big)^2 \\ &= \frac{1}{2^{n^2}} \sum_A \sum_\sigma \sum_\tau (\operatorname{sign}\sigma)(\operatorname{sign}\tau) a_{1\sigma(1)} a_{1\tau(1)} \cdots a_{n\sigma(n)} a_{n\tau(n)}, \end{aligned}$$

where σ and τ run independently through all permutations of $\{1, \ldots, n\}$. Interchange of summation yields

$$D_n^2 = \frac{1}{2^{n^2}} \sum_{\sigma,\tau} (\operatorname{sign}\sigma)(\operatorname{sign}\tau) \Big(\sum_A a_{1\sigma(1)} a_{1\tau(1)} \cdots a_{n\sigma(n)} a_{n\tau(n)} \Big).$$

This doesn't look too promising, but wait. Look at a fixed pair (σ, τ). The inner sum $\sum_A$ is really a summation over n^2 variables, one for each a_{ij}:

$$\sum_{a_{11}=\pm 1} \sum_{a_{12}=\pm 1} \cdots \sum_{a_{nn}=\pm 1} a_{1\sigma(1)} a_{1\tau(1)} \cdots a_{n\sigma(n)} a_{n\tau(n)}. \tag{7}$$

Suppose $\sigma(i) = k \neq \tau(i)$. Then every summand contains a_{ik}, and therefore the whole sum has the *factor* $\sum_{a_{ik}=\pm 1} a_{ik} = 0$, and hence is 0 as well. The only way that the sum fails to be 0 is when $\sigma = \tau$, and everything simplifies indeed: For $\sigma = \tau$, the inner product is 1 as is the term $(\operatorname{sign}\sigma)^2$. The sum in (7) is therefore

$$\sum_{a_{11}=\pm 1} \cdots \sum_{a_{nn}=\pm 1} 1 = 2^{n^2},$$

and wrapping things up we obtain

$$D_n^2 = \frac{1}{2^{n^2}} \sum_\sigma 2^{n^2} = n!,$$

and thus the following result.

Theorem 2. *There exists an $n \times n$ matrix with entries ± 1 whose determinant is greater than $\sqrt{n!}$.*

It is a characteristic feature of averaging that, while we learn that such a matrix exists, we have no clue how to construct it efficiently. But, surprisingly, the bound is quite good. Invoking Stirling's formula from page 13 we have

$$\sqrt{n!} \sim (2\pi n)^{\frac{1}{4}} \left(\frac{n}{e}\right)^{\frac{n}{2}},$$

and this is not too bad in comparison to the upper bound $n^{n/2}$.

Using the biquadratic mean average Szekeres and Turán got the even better lower bound $\frac{1}{4}\sqrt{n!}\sqrt{n}$, but the correct growth for the maximum as n goes to infinity is still not known.

References

[1] J. HADAMARD: *Résolution d'une question relative aux déterminants,* Bulletin des Sciences Mathématiques **17** (1893), 240-246.

[2] G. SZEKERES & P. TURÁN: *An extremal problem in the theory of determinants,* in: "Collected Papers of Paul Turán" (P. Erdős, ed.), Akadémiai Kiadó, Budapest 1990, Vol. 1, pp. 81-87.

[3] H. WILF: *An algorithm-inspired proof of the spectral theorem in E^n,* Amer. Math. Monthly **88** (1981), 49-50.

[4] H. WILF: *Some examples of combinatorial averaging,* Amer. Math. Monthly **92** (1985), 250-261.