RÓBERT FREUD

Linear Algebra: A Problem-Centered Approach Solutions to Selected Exercises

The Reader is advised to consult the solution only after having solved the problem or at least having worked hard on it. It is worth first checking the hint or answer given to many exercises in the other two free online materials to see if the calculation or proof runs in the right direction.

After having arrived at a solution, comparing it to the one given in this booklet can help to make sure that all details are correct and no case was overlooked. You may find that your argument can be simplified; this is okay! One often only really appreciates a neat solution after "suffering" a lot with a more complicated approach.

There can be many correct solutions that differ significantly from the one in this booklet. In this case, however, it is worth checking the validity of the argument very carefully.

CONTENTS

1. Determinants

• **1.1.5 (b)** Answer: $2\lceil (n-4)/5 \rceil + 1$ where $\lceil x \rceil$ is the ceiling of x, i.e., the smallest integer greater than or equal to $\geq x$.

Proof. For a pair $a < b$ in a fixed permutation let $m(a, b)$ denote the number of elements c located between a and b and satisfying $a < c < b$. We shall say that these are the strong elements when swapping a and b. E.g., $m(2, 6) = 2$ in the permutation 3165472 as 5 and 4 are strong when swapping 6 and 2. The swap of a and b amends the number of inversions by $2m(a, b)+1$ since there is a change in the relation of a and b to the strong elements and to each other.

Let M be the maximum of all values $m(a, b)$ in a given permutation. By the previous observation, we have to show that (i) $M \geq \lceil (n-4)/5 \rceil$ for every permutation, and (ii) there exists a permutation where $M = \lceil (n-4)/5 \rceil$.

To prove (i), we fix a permutation. We assume that neither 1, nor n is in the first or last position. The remaining cases can be settled by similar but simpler considerations and we get an even larger M; we leave the details to the Reader

Let k and r be the first and last elements, respectively, and assume that n and 1 are in (say) reverse order, so the permutation is of the form $k \ldots n \ldots 1 \ldots r$.

We show that except k, n, 1, and r, every element is strong in at least one of the five swaps (A) k and n; (B) k and 1; (C) n and 1; (D) n and r; (E) 1 and r.

An element positioned between k and n is strong at (A) or (B) depending on whether it is greater or smaller than k. Every element standing between n and 1 is strong at (C) [some of them may be strong at (B) and/or (D), too]. Finally, an element between 1 and r is strong at (D) or (E) depending on whether it is greater or smaller than r.

So $n-4 \leq m(k, n)+m(1, k)+m(1, n)+m(r, n)+m(1, r) \leq 5M$, implying $M \geq \lceil (n-4)/5 \rceil$ as stated.

Turning to (ii), it clearly suffices to deal with the case $n = 5t + 4$. Analyzing the argument in (i), we get that

$$3t{+}3, \ldots, 4t{+}3 \,\big|\, t{+}1, t, \ldots, 1 \,\big|\, 2t{+}3, \ldots, 3t{+}2 \,\big|\, 5t{+}4, \ldots, 4t{+}4 \,\big|\, t{+}2, \ldots, 2t{+}2$$

is an appropriate permutation.

• **1.1.7 (c)** Answer: If k is odd, then every even $n > k$; if $k \neq 0$ is even, then every $n > k$ except $n = 2k + 1$; and if $k = 0$, then every $n > 0$.

Necessity. Obviously, $n > k$. The total number of permutations is $nk/2$, thus n must be even if k is odd. Finally, the first element is in inversion with k elements, these have to be $1, 2, \ldots, k$, so the first element is $k+1$. Similarly, the last element is $n - k$. Therefore $k + 1 \neq n - k$, i.e., $n \neq 2k + 1$ for $n > 1$.

Sufficiency. We denote by $/s, t\backslash$ if there exists a permutation of $1, 2, \ldots, s$ where every element is in inversion with t other elements. We need two observations:

(A) $/c, k - d\backslash, /d, k - c\backslash \Rightarrow /c + d, k\backslash$;

(B) $/f, k\backslash, /g, k\backslash \Rightarrow /uf + vg, k\backslash$ where u and v are arbitrary positive integers.

To verify (A), let $i_1, \ldots, i_c$ and $j_1, \ldots, j_d$ be permutations guaranteeing $/c, k - d\backslash$ and $/d, k - c\backslash$, respectively. Then $d + i_1, \ldots, d + i_c, j_1, \ldots, j_d$ implies $/c + d, k\backslash$. As for (B), we apply the first permutation for the blocks $1, 2, \ldots, f$; $f + 1, \ldots, 2f$; $\ldots$; $(u - 1)f + 1, \ldots, uf$, and then apply the second permutation for the blocks $uf + 1, \ldots, uf + g$, etc.

We turn to the proof of sufficiency by induction on k. The initial cases $k = 0$ and 1 are obvious. For the induction step, we assume that the statement holds for every $k' < k$ and for every appropriate n. Consider now k and assume first that $k < n \leq 2k$ and kn is even. Write $n = c + d$, $1 \leq c, d \leq k$ with c even (n has several such representations in general). Then $c > k - d$ and $d > k - c$, so both $/c, k - d\backslash$ and $/d, k - c\backslash$ hold, the second one due to $d(k - c) \equiv kn \equiv 0 \pmod 2$. These imply $/c + d, k\backslash = /n, k\backslash$ by (A). (If $d = 2(k - c) + 1$, then $/d, k - c\backslash$ is false, but we can work with another representation $n = c + d$. If there was only one such representation, then $/n, k\backslash$ follows easily directly.) Finally, we can handle the cases $n > 2k$ using $n \leq 2k$ and applying (B).

Note that instead of (A) we could have relied on the property $/n, k\backslash \iff /n, n - 1 - k\backslash$ obtained by reversing the permutation.

• **1.4.13 (a)** If P chooses the zero matrix, then R has to change n elements to eliminate the zero rows and columns, e.g., by creating the identity matrix. We show that this is the worst case. Let r be the maximal number of rows and columns so that the determinant formed of the r^2 elements in their intersections is not zero. Then $n - r$ steps are sufficient. Assume, e.g., that the $r \times r$ determinant D_r in the upper left corner is not 0. R expands the $(r+1)\times(r+1)$ determinant D_{r+1} in the upper left corner along its last (r+1st) row. The coefficient of $\alpha_{r+1,r+1}$ in the expansion is $\pm D_r \neq 0$. Therefore, R can change $\alpha_{r+1,r+1}$ so that the arising determinant $D'_{r+1} \neq 0$. Now R extends D'_{r+1} with the first $r + 2$ elements of the next ($r + 2$nd) row and column

into an $(r+2)\times(r+2)$ determinant and repeats the previous argument for $a_{r+2,r+2}$. Continuing the algorithm, R gets an $n\times n$ matrix A' with a non-zero determinant.

In the solution, we operated with the determinant rank of a matrix, see Section 3.4. To prove the weaker assertion that n steps always suffice, we could have used induction and a simplified version of the above argument.

• **(b)** If P chooses the zero matrix, then n^2-n steps are not enough as P can mark all elements of $n-1$ columns and the skipped zero column yields a 0 determinant. We show by induction that n^2-n+1 steps always suffice. This is obvious for $n=1$. For $n>1$, consider the $n\times n$ matrix A and the n^2-n+1 positions given by P. There must be a row where R can change every element, otherwise there would be only $n(n-1)$ positions marked by P. Also, there exists a column where not every element can be modified. If (say) the first row and column satisfy these conditions, then R modifies the first row by replacing the first element by 1 and the other elements by 0. Deleting the first row and column from A, we obtain an $(n-1)\times(n-1)$ matrix B. As not all of the altogether $2n-1$ elements in the first row and column of A are marked, at least $(n^2-n+1)-(2n-2)=(n-1)^2-(n-1)+1$ positions are marked in B. By the induction hypothesis, R can transform B into a matrix B' with a non-zero determinant. Performing all these changes, A is transformed into an $n\times n$ matrix A' where the first row is the modified version of the first row of A, the last $n-1$ elements of the first column are the same as in A, and the remaining part is B'. Expanding A' along its first row we get $\det A'=\det B'\neq 0$.

• **(ca)** This is the same as Exercise 1.2.7.

• **(cb)** If P chooses the identity matrix, then $(n^2-n)/2$ steps are not enough as P can mark all elements above the main diagonal. We show by induction that $1+(n^2-n)/2$ steps always suffice. This is obvious for $n=1$. For $n>1$, consider the $n\times n$ matrix A and the $1+(n^2-n)/2$ positions given by P.

If every column contains a position marked by P, then R can change these to achieve a zero sum in every column. Adding all rows to the last one, we get a zero row, so the determinant is 0.

If all elements in a column are 0, then no change is needed.

If no position is marked in (say) the first column and, e.g., $\alpha_{11}\neq 0$, then R subtracts a suitable multiple of the first row from the other rows to turn the other elements of the first column into 0. Deleting the first row and column of A, the result is an $(n-1)\times(n-1)$ matrix B_1 with at least

$$1+(n^2-n)/2-(n-1)=1+[(n-1)^2-(n-1)]/2$$

positions marked by P. By the induction hypothesis, R can transform B_1 into a matrix B' with a non-zero determinant. Registering all the above changes in A, we get a matrix A' with $\det A' = \alpha_{11} \det B' = 0$.

• **1.4.14** The requirement is satisfied by matrices with zero determinant but with no zero cofactor A_{ij}. Modifying α_{ij}, the expansion along the ith row (or jth column) yields another value, so the new determinant is not zero. To construct such an $n \times n$ matrix, take an $(n-1) \times (n-1)$ matrix with a non-zero determinant and extend it with an nth row and column so that the sum of elements in each row and column of the resulting $n \times n$ matrix is 0 (cf. Exercise 1.4.11).

• **1.5.6** The jth element in the ith row is the sum $1 + \alpha_i\beta_j + \ldots + \alpha_i^{n-1}\beta_j^{n-1}$ of n terms by the summation formula for geometric sequences. We can write the determinant as the sum of n^n determinants in each of which the jth element in the ith row is $\alpha_i^k\beta_j^k$ with some $0 \le k \le n-1$. Many of these determinants are 0 as they have two identical rows. In the remaining determinants the elements of the jth column are $\alpha_1^{\pi(1)}\beta_j^{\pi(1)}, \alpha_2^{\pi(2)}\beta_j^{\pi(2)}, \ldots, \alpha_n^{\pi(n)}\beta_j^{\pi(n)}$ where $\pi(1), \ldots, \pi(n)$ is a permutation of $0, 1, \ldots, n-1$. Factoring out $\alpha_1^{\pi(1)}, \ldots, \alpha_n^{\pi(n)}$ from the rows and performing $I(\pi)$ row swaps we arrive at $V(\beta_1, \ldots, \beta_n)$. So the original determinant is

$$V(\beta_1, \ldots, \beta_n) \sum (-1)^{I(\pi)} \alpha_1^{\pi(1)} \cdot \ldots \cdot \alpha_n^{\pi(n)} = V(\beta_1, \ldots, \beta_n) V(\alpha_1, \ldots, \alpha_n).$$

Another option is to use the Product Theorem 2.2.4 for determinants (see Exercise 2.2.8).

2. Matrices

• **2.1.18** Exactly the scalar multiples of the identity matrix $A = \lambda I$ commute with every matrix of the same size. They clearly satisfy the requirement. For the converse, assume that $AB = BA$ holds for every B. Let $i \ne j$ and consider the matrix B where the jth element of the ith row is 1 and all other elements are 0. Then the jth column in AB is the same as the ith column in A and all other elements are 0. Similarly, the ith row in BA is the same as the jth row of A and all other elements are 0. The equality $AB = BA$ implies $\alpha_{ii} = \alpha_{jj}$, $k \ne i \Rightarrow \alpha_{ki} = 0$, and $m \ne j \Rightarrow \alpha_{jm} = 0$. This holds for every $i \ne j$, so all elements in the main diagonal of A are equal and all other elements are 0. Thus, $A = \lambda I$.

• **2.2.13** We prove by contradiction. If all elements in a row or column of A are 0, then $\det A = 0$, so A has no inverse. Assume now that there are at least two non-zero elements in some row or column, say $\alpha_{24} > 0$ and $\alpha_{27} > 0$. Multiplying the second row of A with all columns of A^{-1} except its second column, we always get 0 by $AA^{-1} = I$. As all elements are non-negative, the 4th and 7th elements in these columns of A^{-1} must be 0. This means that except perhaps the second elements, all other elements in the 4th and 7th rows of A^{-1} are 0. But then one of these rows is a scalar multiple of the other row, so $\det A^{-1} = 0$, a contradiction.

3. Systems of Linear Equations

• **3.1.17** Double cannot find out the parity of Triple's integers since, e.g., the answer will always be "even" both if Triple's integers were all even or all odd. Triple can determine the parity of Double's integers, e.g., with the 5 questions

$$x_5 + x_1 + x_2 =? = b_1,\ x_1 + x_2 + x_3 =? = b_2,\ x_2 + x_3 + x_4 =? = b_3,$$
$$x_3 + x_4 + x_5 =? = b_4,\ x_4 + x_5 + x_1 =? = b_5$$

where all sums are modulo 2. This system has a unique solution as it can be shown by verifying that its determinant is not zero modulo 2, but also by direct manipulations:

$$b_1 + b_2 + b_3 + b_4 + b_5 = 3(x_1 + x_2 + x_3 + x_4 + x_5) = x_1 + x_2 + x_3 + x_4 + x_5,$$

so $b_2 + b_4 = x_1 + x_2 + 2x_3 + x_4 + x_5 = b_1 + b_2 + b_3 + b_4 + b_5 + x_3$, hence $x_3 = b_1 + b_3 + b_5$, and we get the values of the other unknowns similarly. 4 questions do not suffice since a system with 4 equations and 5 unknowns cannot have a unique solution.

• **3.2.12 (a)** The sum of the n fundamental Lagrange polynomials is the interpolation polynomial f satisfying $f(\gamma_1) = \ldots = f(\gamma_n) = 1$. The constant polynomial $f = 1$ has this property. There is exactly one such polynomial of degree not greater than $n - 1$, so $\sum_{i=1}^{n} L_i = 1$.

• **(b)** The expression (b1) is the value of the polynomial $\sum_{i=1}^{n} L_i = 1$ assumed at ν, i.e., it equals 1. The sum in (b2) is the coefficient of the term of degree $n - 1$ in the same polynomial, so it is 0.

• **3.4.13 (c)** Let $\gamma_1, \ldots, \gamma_n$ be distinct non-zero scalars (assuming that $|F| \geq n + 1$), and $\alpha_{ij} = \gamma_j^{i-1}$ if $i \leq r$, and 0 otherwise. The (row) rank of this $k \times n$

matrix A is not greater than r since it has only r non-zero rows. We show that any r columns are independent, i.e., the (column) rank of any $k \times r$ submatrix B is r. The rank of B does not change if we delete its zero rows and we get an $r \times r$ matrix C. The determinant of C is a Vandermonde determinant generated by distinct elements, so it is not 0. Thus the (determinant) rank of C is r.

• **3.4.14 (c)** Let $\gamma_1, \ldots, \gamma_n$, and $\delta_1, \ldots, \delta_k$ be distinct non-zero scalars and $\alpha_{ij} = 1 + (\delta_i\gamma_j) + \ldots + (\delta_i\gamma_j)^{r-1}$. By Exercise 1.5.6, every $r \times r$ minor in this matrix is the product of two Vandermonde determinants with distinct generators, so none of these minors is 0. We have to show that every $(r+1) \times (r+1)$ minor M is 0. We can write M as the sum of r^{r+1} determinants D where the elements in the ith row of D are the terms $(\delta_i\gamma_j)^s$ $(j = 1, \ldots, n)$ for some fixed $0 \leq s \leq r-1$. By the pigeonhole principle, two rows of D belong to the same s, so these rows are constant multiples of each other. Therefore $D = 0$, hence also $M = 0$.

• **3.4.18 (a)** The sum of such matrices preserves this property, so we cannot obtain every matrix as such a sum.

• **(b)** By part (a) it is sufficient to investigate sums $B + C$ where every row of B and every column of C is an arithmetic sequence, so $\beta_{ij} = x_i + (j-1)y_i$ and $\gamma_{ij} = v_j + (i-1)z_j$, $1 \leq i \leq k, 1 \leq j \leq n$. A representation $A = B + C$ means the solvability of the system of equations $\alpha_{ij} = x_i + (j-1)y_i + v_j + (i-1)z_j$ where x_i, y_i, v_j, and z_j are the unknowns. There are $2k + 2n$ unknowns and kn equations. There are more equations than unknowns for most values of n and k, as $kn > 2k + 2n \iff (k-2)(n-2) > 4$. In this case, the rows of the coefficient matrix are linearly dependent. Therefore, some row of the coefficient matrix is a linear combination of the other rows. If the system is solvable, then the same relation must hold for the corresponding constants α_{ij}. So, the system cannot be solvable for every choice of α_{ij}. Hence, not all matrices A can be represented as a sum of matrices of the given form.

• **(c)** Let $\gamma_1, \ldots, \gamma_n$ be distinct non-zero scalars. We show that every $k \times n$ matrix A is the sum of n matrices M_r, $r = 1, 2, \ldots, n$, where every row of M_r is a geometric sequence with quotient γ_r. Let the elements in the first row of M_r be $x_r, x_r\gamma_r, \ldots, x_r\gamma_r^{n-1}$. Then the first row of $A = \sum_{r=1}^{n} M_r$ means the system of equations $\alpha_{1j} = \sum_{r=1}^{n} x_r\gamma_r^{j-1}$, $j = 1, 2, \ldots, n$, where $x_1, \ldots, x_n$ are the unknowns. The determinant of the system is $V(\gamma_1, \ldots, \gamma_n) \neq 0$, so there is a solution. The same argument works for the other rows, too. If we do not want to allow $x_r = 0$ as then every term of the geometric sequence is 0, we

substitute such a matrix M_r with the sum of two matrices where the zero rows are replaced by any geometric sequence and its negative, and the other rows are the halves of the corresponding rows of M_r.

• **3.4.19 (a)** R can replace every constant by 0, so k steps are sufficient. This is also necessary if every coefficient is 0 and the constants are not 0.

• **(b)** If $k > n$, then changing a single constant on the right-hand side is sufficient. As $k > n$, the rows of the coefficient matrix are linearly dependent. Therefore, there is a $1 \leq i \leq k$ such that the ith row is a linear combination of the other rows. If the system is solvable, then the ith constant β_i can be expressed the same way as the linear combination of the other constants. So, replacing β_i by any value not equal to the given linear combination of the other constants, the system cannot have a solution.

We show that if $k \leq n$, then $n - k + 2$ is the minimal number of steps. We need that many steps if P presented a system of equations where any k columns of the coefficient matrix are linearly independent and the constants on the right-hand side are 0. If there remain k independent columns, then the system is still solvable for any constants on the right-hide side. Therefore R has to modify at least $n - k + 1$ columns plus at least one 0 constant on the right-hand side.

Essentially the same argument yields that $n - k + 2$ steps always suffice. For a simpler technical handling, assume first that there were no row swaps in the elimination process leading to the "unreduced" echelon form of the augmented matrix. As we reach this form by top-down manipulations, we do not add a scalar multiple of the last row to other rows. So, modifying the last row in the echelon form, the corresponding modifications affect only the last row of the original augmented matrix. Thus if we only replace elements in the last row, then we can work in the echelon form. If the ordinary elimination needs row swaps, then we simply ignore them during the elimination, and so the ith leader will not necessarily be in the ith row and the zero rows will not necessarily be at the bottom of the matrix. This simply means a permutation of the rows and the above argument is valid for the row occurring as last in this permutation.

We show that we can achieve a forbidden row modifying at most $n-k+2$ element in the last row of the echelon form (in the above sense). If the constant on the right-hand side was 0, we replace it by a non-zero scalar. This in itself is sufficient if the last row contained no leader. If there is a leader in the last row, it is preceded by at least $k-1$ zeros. So there are at most $n-k+1$ non-zero elements in the last row, and replacing them by 0 we obtain a forbidden row.

• **3.5.8 (c)** Let (say) the last unknown be a free parameter in the general solution of the system of equations $A\mathbf{x} = \mathbf{0}$. Then the last row in the matrices B satisfying $AB = 0$ can be an arbitrary vector $\mathbf{s}$. If also $BA = 0$, then multiplying the last row in B with A we get $\mathbf{s}A = \mathbf{0}$ for every $\mathbf{s}$. This can only hold for $A = 0$.

4. Vector Spaces

• **4.2.12 (e)** To prove by contradiction, we assume $V = \cup_{i=1}^{k} W_i$ where W_i are non-trivial subspaces in V. We may assume that k is the minimal such value. Then $W_1 \not\subseteq \cup_{i=2}^{k} W_i$ since otherwise also $V = \cup_{i=2}^{k} W_i$ would hold. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors satisfying $\mathbf{u} \in W_1$ but $\mathbf{u} \notin W_i$ for $i > 1$ and $\mathbf{v} \notin W_1$. Each of the infinitely many vectors $\mathbf{v} + \lambda\mathbf{u}$ ($\lambda \in F$) is contained in (at least) one of the finitely many subspaces W_j, so two such vectors must be in the same W_i for some $1 \le i \le k$: $\mathbf{v} + \lambda\mathbf{u} \in W_i$ and $\mathbf{v} + \lambda'\mathbf{u} \in W_i$ for some $\lambda \neq \lambda'$. If $i = 1$, then $\mathbf{v} \in W_1$, a contradiction. If $i > 1$, then subtracting the two vectors we get $(\mathbf{v} + \lambda\mathbf{u}) - (\mathbf{v} + \lambda'\mathbf{u}) = (\lambda - \lambda')\mathbf{u} \in W_i$, so $\mathbf{u} \in W_i$, a contradiction.

• **(f)** We adapt the argument of part (e). We have now $|F|$ vectors $\mathbf{v} + \lambda\mathbf{u}$. If $|F| > k - 1$, then we arrive at a contradiction as before. So, $k \ge |F| + 1$.

• **(g)** Since V has a non-trivial subspace, we can find vectors $\mathbf{b} \neq \mathbf{0}$ and $\mathbf{c} \neq \alpha\mathbf{b}$ in V. Let W be a maximal subspace containing no non-zero vectors of the form $\lambda\mathbf{b} + \mu\mathbf{c}, \lambda, \mu \in F$ (i.e., any larger subspace must contain such a non-zero vector). The existence of W is obvious if $|V| < \infty$, and can be proved by the standard set-theoretic methods, e.g., by Zorn's lemma, if $|V| = \infty$. Then V is the union of the $k + 1$ non-trivial subspaces

$$W' = \{\vartheta\mathbf{c} + \mathbf{w} \mid \vartheta \in F, \mathbf{w} \in W\}$$

and

$$W_\gamma = \{\vartheta(\mathbf{b} + \gamma\mathbf{c}) + \mathbf{w} \mid \vartheta \in F, \mathbf{w} \in W\}, \gamma \in F.$$

It is easy to see that these are subspaces. We show that none of them equals V. E.g., $\mathbf{c} \notin W_\gamma$. Otherwise $\mathbf{c} = \vartheta(\mathbf{b} + \gamma\mathbf{c}) + \mathbf{w}$ would imply $\mathbf{w} = -\vartheta\mathbf{b} + (1 - \vartheta\gamma)\mathbf{c}$. By the condition on W, this can happen only if $\mathbf{w} = \mathbf{0}$. Since $\mathbf{b} \neq \mathbf{0}$ and $\mathbf{c} \neq \alpha\mathbf{b}$ we infer $\vartheta = 1 - \vartheta\gamma = 0$, a contradiction. We can verify $\mathbf{b} \notin W'$ similarly.

Finally we show that every $\mathbf{v} \in V$ is in the union of the $k+1$ subspaces. As $W \subseteq W'$, this holds if $\mathbf{v} \in \mathbf{W}$. Otherwise, let $U = \{\vartheta\mathbf{v} + \mathbf{w} \mid \vartheta \in F, \mathbf{w} \in W\}$. Then U is a subspace and contains a vector $\lambda\mathbf{b} + \mu\mathbf{c} \neq \mathbf{0}$ by the maximal

property of W: $\vartheta\mathbf{v}+\mathbf{w}=\lambda\mathbf{b}+\mu\mathbf{c}$. Here $\vartheta\neq 0$ by the condition of W. So

$$\mathbf{v}=\frac{\lambda}{\vartheta}\mathbf{b}+\frac{\mu}{\vartheta}\mathbf{c}-\frac{1}{\vartheta}\mathbf{w}.$$

This implies $\mathbf{v}\in W'$ if $\lambda=0$ and $\mathbf{v}\in W_{\mu/\lambda}$ otherwise.

• *Remark*: It is worth analyzing how this proof implements the idea given in the hint. The proof itself could be accomplished much more comfortably using the dimension and the factor space or direct sum but we did not want to rely on these here.

• **4.4.11** There are many ways to treat each part, e.g., the truth of (a) is obvious from the definition and it is easy to find a counterexample to (c). For a unified discussion, however, it is worth to consider the $k\times m$ matrix where the columns are the m vectors in question. The independence of the vectors means that this matrix has column rank m. Switching to the determinant rank, the determinant of a matrix with integer entries is an integer, so it is irrelevant whether we consider the rank over $\mathbf{R}$ or $\mathbf{Q}$, thus (a) and (b) are true. In the case of F_p we have to investigate the divisibility of the determinant by p. There are non-zero even integers, so (c) is false. An odd integer cannot be zero, so (d) is true. Finally, a non-zero integer has only finitely many prime divisors, so (e) is true. (Cf. Exercises 3.4.8 and 4.6.16.)

• **4.6.7 (a)** Let $\mathbf{b}_1,\dots,\mathbf{b}_{100}$ be a basis in V. Then any 100 from the infinitely many vectors $\mathbf{v}_\gamma=\mathbf{b}_1+\gamma\mathbf{b}_2+\gamma^2\mathbf{b}_3+\ldots+\gamma^{99}\mathbf{b}_{100}, \gamma\in\mathbf{R}$, form a basis. It is sufficient to show that they are independent. This follows as the determinant of the corresponding homogeneous system of linear equations is a Vandermonde determinant generated by distinct elements, so it differs from zero.

• **(b)** Over F_2, a basis and the sum of all basis elements satisfy the condition. We show that this is the maximum. Let $\mathbf{v}_1,\dots,\mathbf{v}_{100}$ be the basis of the first 100 such vectors, and write $\mathbf{v}_{101}$ as their linear combination: $\mathbf{v}_{101}=\sum_{i=1}^{100}\alpha_i\mathbf{v}_i$. If some $\alpha_i=0$, then disregarding $\mathbf{v}_i$, the other 100 vectors $\mathbf{v}_j$ are clearly dependent, a contradiction. So every $\alpha_i=1$, i.e., $\mathbf{v}_{101}$ is uniquely determined, so no larger system can satisfy the requirements.

The 101 vectors consisting of a basis and the sum of its elements satisfy the condition over F_{97}, too. We prove that there are no more suitable vectors in this case either. Similar to the argument above for F_2, no coefficient α_i is zero in the representation $\mathbf{v}_{101}=\sum_{i=1}^{100}\alpha_i\mathbf{v}_i$. Replacing each of the first 100 vectors with its suitable scalar multiple, we can achieve $\alpha_i=1$ for every i, i.e., $\mathbf{v}_{101}=\sum_{i=1}^{100}\mathbf{v}_i$. Assume that there exists some $\mathbf{v}_{102}=\sum_{i=1}^{100}\beta_i\mathbf{v}_i$. By the pigeon hole

principle, $\beta_i = \beta_j$ for some $i \neq j$, say, $\beta_1 = \beta_2$. Then $\mathbf{v}_{102} - \beta_1 \mathbf{v}_{101}$ is a linear combination of $\mathbf{v}_3, \ldots, \mathbf{v}_{100}$, so $\mathbf{v}_3, \ldots, \mathbf{v}_{102}$ are dependent, a contradiction.

Over F_{101} there exist 102 vectors with this property: Let $\mathbf{v}_1, \ldots, \mathbf{v}_{100}$ be any basis, $\mathbf{v}_{101} = \sum_{i=1}^{100} \mathbf{v}_i$, and $\mathbf{v}_{102} = \sum_{i=1}^{100} i\mathbf{v}_i$. It is easy to see that they satisfy the condition. We show that this is the maximum. Similar to the argument for F_{97}, we can assume that $\mathbf{v}_{101} = \sum_{i=1}^{100} \mathbf{v}_i$ and $\mathbf{v}_{102} = \sum_{i=1}^{100} \beta_i \mathbf{v}_i$, where β_i are distinct non-zero elements of F_{101}, hence they form a permutation of 1,2,...,100. Assume the existence of $\mathbf{v}_{103} = \sum_{i=1}^{100} \gamma_i \mathbf{v}_i$. Then also γ_i form a permutation of 1,2,...,100. Further, writing $\gamma_i = \delta_i \beta_i$, also δ_i have to be distinct (and non-zero): If, e.g., $\delta_1 = \delta_2$, then $\mathbf{v}_{103} - \delta_1 \mathbf{v}_{102}$ is a linear combination of $\mathbf{v}_3, \ldots, \mathbf{v}_{100}$, so $\mathbf{v}_3, \ldots, \mathbf{v}_{100}, \mathbf{v}_{102}, \mathbf{v}_{103}$ are dependent, a contradiction. The equalities $\gamma_i = \delta_i \beta_i$ in F_{101} mean modulo 101 congruences. Taking their product for $i = 1, 2, \ldots, 100$, Wilson's theorem yields a contradiction:

$$-1 \equiv 100! \equiv \prod_{i=1}^{100} \gamma_i \equiv \prod_{i=1}^{100} \delta_i \prod_{i=1}^{100} \beta_i \equiv (100!)^2 \equiv (-1)^2 = 1 \pmod{101}.$$

• **4.6.16 (c)** Let the rank of a 0–1 matrix be s over F_2 and t over $\mathbf{Q}$. Taking s linearly independent columns over F_2 they have $2^s - 1$ non-trivial linear combinations. If the matrix has more than $2^s - 1$ columns, there must occur a zero column or two identical columns among them, therefore there cannot be more than $2^s - 1$ independent columns over $\mathbf{Q}$, so $t \leq 2^s - 1$. Conversely, we show that for any $(s \leq) t \leq 2^s - 1$ there exists a $t \times t$ zero-one matrix of ranks s over F_2 and t over $\mathbf{Q}$. This implies that $s = 10$ is the minimal value providing the difference 1000 between the two ranks and then the size of the matrix is $t = s + 1000 = 1010$.

It is enough to deal with the case $t = 2^s - 1$. If $s \leq t' < 2^s - 1$, then we delete $2^s - 1 - t'$ columns from the suitable $(2^s - 1) \times (2^s - 1)$ matrix A so that the remaining part still contains s independent columns over F_2. All columns were independent over $\mathbf{Q}$, hence so are the remaining t' columns, too. Therefore this $(2^s - 1) \times t'$ matrix B has ranks t' over $\mathbf{Q}$ and s over F_2. We keep now s rows from B that are independent over F_2. These are independent over $\mathbf{Q}$, too, therefore we can extend them by $t' - s$ further rows of B so that these t' rows are independent over $\mathbf{Q}$. We obtain a $t' \times t'$ matrix of ranks t' over $\mathbf{Q}$ and s over F_2 as required.

Let now be $t = 2^s - 1$ and we construct a $t \times t$ zero-one matrix of ranks s over F_2 and t over $\mathbf{Q}$. The first s entries of the rows are the binary digits of the numbers $1, 2, \ldots, 2^s - 1$, so the first s elements of the first, second, and third rows are $(0, 0, \ldots, 0, 0, 1)$, $(0, 0, \ldots, 0, 1, 0)$, and $(0, 0, \ldots, 0, 1, 1)$, etc.

The other columns are the (further) non-trivial linear combinations over F_2 of the first s columns (with at least two non-zero coefficients). So we obtain a $(2^s-1)\times(2^s-1)$ zero-one matrix. We show that its columns are independent over $\mathbf{Q}$, thus its rank is 2^s-1 over $\mathbf{Q}$. This and the construction imply that its rank is s over F_2.

For a simpler exposition we insert a zero row to the top of our matrix, this does not influence the independence relations of the columns. By induction on s, we show that in every column of the arising $2^s\times(2^s-1)$ matrix A_s the half of the 2^s elements are 1 and the other half are 0. This is obvious for $s=1$. Now we assume its validity for $s-1$ and let the first $s-1$ columns be $\mathbf{a}_1,\dots,\mathbf{a}_{s-1}$ (each has 2^{s-1} components). For s, the construction of the first s columns (each having 2^s components) is

$$\mathbf{b}_1=\begin{pmatrix}\mathbf{0}\\ \mathbf{1}\end{pmatrix},\ \mathbf{b}_2=\begin{pmatrix}\mathbf{a}_1\\ \mathbf{a}_1\end{pmatrix},\ \dots,\mathbf{b}_s=\begin{pmatrix}\mathbf{a}_{s-1}\\ \mathbf{a}_{s-1}\end{pmatrix}.$$

Since exactly the half of the components of $\mathbf{a}_i$ are 1, this property gets inherited by the vectors $\mathbf{b}_j$, too, so the first s columns in A_s satisfy the condition. The other columns are obtained as non-trivial linear combinations over F_2 of the first s columns, so they are the sums of some vectors $\mathbf{b}_j$. If $\mathbf{b}_1$ does not occur among the terms, then the upper and lower halves of this sum are equal just like in the terms, so we can apply the induction hypothesis. If we add $\mathbf{b}_1$ to such a sum, then the 1s and 0s get swapped in the lower half but as they number the same, we are done now, too.

We shall need that (for $s>1$) any two columns of A_s have exactly 2^{s-2} components both of which are 1. Let $\mathbf{u}$ and $\mathbf{v}$ be two arbitrary columns and let there be x such components. Then each of $\mathbf{u}$ and $\mathbf{v}$ has further $2^{s-1}-x$ components 1. The sum $\mathbf{u}+\mathbf{v}$ occurs among the columns of A_s, so 2^{s-1} of its components are 1. On the other hand, a component of $\mathbf{u}+\mathbf{v}$ is 1 exactly if that component of $\mathbf{u}$ and $\mathbf{v}$ are not equal. So the number of components 1 in $\mathbf{u}+\mathbf{v}$ is $2(2^{s-1}-x)=2^{s-1}$. This gives $x=2^{s-2}$ as stated.

We shall use the inner product (see Sections 7.1, 8.1, and 9.4 for details) to verify the independence of the columns $\mathbf{v}_1,\dots,\mathbf{v}_t$ of the matrix A_s over $\mathbf{Q}$. (The statement is obvious for $s=1$, so we may assume $s\ge 2$, though the proof below formally works for $s=1$, too.) Consider $\sum_{i=1}^t\alpha_i\mathbf{v}_i=\mathbf{0}$ and take the inner product of both sides with an arbitrary $\mathbf{v}_j$. The inner product of two 0–1 vectors is the number of common components 1. This means by the previous two paragraphs that $\alpha_j 2^{s-1}+\sum_{i\ne j}\alpha_i 2^{s-2}=0$, i.e., $2^{s-2}(\alpha_j+\sum_{i=1}^t\alpha_i)=0$. This holds for every j implying $\alpha_i=0$ for every i as stated.

We note that apart from the order of rows and columns, an alternative description of A_s is the following: the indices of the rows and columns are the

subsets and non-empty subsets of $X = \{1, 2, \ldots, s\}$, and $\alpha_{Y,Z} = |Y \cap Z| \bmod 2$ (where $Y, Z \subseteq X$).

5. Linear maps

• **5.5.9** The equality

$$\lambda_1 \mathcal{A}_{i_1 j_1} + \lambda_2 \mathcal{A}_{i_2 j_2} + \ldots + \lambda_m \mathcal{A}_{i_m j_m} = \mathcal{O}$$

means that for arbitrary complex numbers $\alpha_{i_r}, \alpha_{j_r}$ $(1 \leq r \leq m)$ we have

$$\lambda_1 \begin{pmatrix} \alpha_{i_1} \\ \alpha_{j_1} \end{pmatrix} + \lambda_2 \begin{pmatrix} \alpha_{i_2} \\ \alpha_{j_2} \end{pmatrix} + \ldots + \lambda_m \begin{pmatrix} \alpha_{i_m} \\ \alpha_{j_m} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

i.e.

$$\lambda_1 \alpha_{i_1} + \lambda_2 \alpha_{i_2} + \ldots + \lambda_m \alpha_{i_m} = 0, \quad \lambda_1 \alpha_{j_1} + \lambda_2 \alpha_{j_2} + \ldots + \lambda_m \alpha_{j_m} = 0.$$

This shows that if, e.g., the index i_1 differs from all other indices i_r, then necessarily $\lambda_1 = 0$: we obtain it by substituting $\alpha_{i_1} = 1, \alpha_{i_2} = \ldots = \alpha_{i_m} = 0$. The same holds for the indices j_r, too.

• **(a)** We apply the above to the case $m = 3$. Clearly, no two maps $\mathcal{A}_{ij}$ are scalar multiples of each other, so any two of them are linearly independent. This implies that if the linear combination of three maps $\mathcal{A}_{ij}$ is $\mathcal{O}$ and one of the coefficients λ_t is 0, then the other two coefficients must be 0, too. Therefore, to verify independence, it is sufficient to show that for any three (distinct) pairs $(i_1, j_1), (i_2, j_2), (i_3, j_3)$ of indices, necessarily one of the three indices i or one of the three indices j differs from the other two corresponding indices. If the indices i are not all the same, then at least one of them occurs only once. It they are all equal, then the indices j must be all distinct.

• **(b)** e.g., $\mathcal{A}_{11} + \mathcal{A}_{22} - \mathcal{A}_{12} - \mathcal{A}_{21} = \mathcal{O}$.

• **(c)** We show that, e.g., the maps $\mathcal{A}_{ij}$ belonging to $(ij) = (11), (12), (13), (14), (24), (34)$, and (44) are linearly independent in $\mathrm{Hom}\,(V_1\,,\, V_2)$. Only one index i is 2, 3, or 4, so the coefficients of the last three maps must be 0. Similarly, only one index j is 1, 2, or 3, so the first three coefficients are 0, too. So, the remaining middle coefficient must be 0, as well.

Now we prove that any eight maps $\mathcal{A}_{ij}$ are linearly dependent. Since the dimension of the vector space $\mathrm{Hom}\,(V_1\,,\, V_2)$ is $4 \cdot 2 = 8$, it is sufficient to show that the subspace spanned by all maps $\mathcal{A}_{ij}$ does not contain ev-

ery element of $\mathrm{Hom}\,(V_1\,,\,V_2)$. We verify that, e.g., $\mathcal{A}\begin{pmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\\ \alpha_4\end{pmatrix} = \begin{pmatrix}\alpha_1\\ 0\end{pmatrix}$ cannot be written in the form $\mathcal{A} = \sum_{1\le i,j\le 4}\lambda_{ij}\mathcal{A}_{ij}$. We apply both sides to the unit vectors. If $\alpha_1 = 1, \alpha_2 = \alpha_3 = \alpha_4 = 0$, then the two coordinates of the image vectors yield the equations $\lambda_{11}+\lambda_{12}+\lambda_{13}+\lambda_{14} = 1$ and $\lambda_{11}+\lambda_{21}+\lambda_{31}+\lambda_{41} = 0$. In the case of $\alpha_2 = 1, \alpha_1 = \alpha_3 = \alpha_4 = 0$ we obtain $\lambda_{21}+\lambda_{22}+\lambda_{23}+\lambda_{24} = \lambda_{12}+\lambda_{22}+\lambda_{32}+\lambda_{42} = 0$. For the other two unit vectors, the results are $\sum_{j=1}^4\lambda_{3j} = \sum_{i=1}^4\lambda_{i3} = 0$ and $\sum_{j=1}^4\lambda_{4j} = \sum_{i=1}^4\lambda_{i4} = 0$. Adding all equations for the first coordinates, we get $\sum_{1\le i,j\le 4}\lambda_{ij} = 1$, whereas adding the equations for the second coordinates yields $\sum_{1\le i,j\le 4}\lambda_{ij} = 0$, a contradiction.

• The exercise can be handled more conveniently with the matrices of linear maps (see Section 5.7). The map $\mathcal{A}_{ij}$ corresponds to a 2×4 (complex) matrix where the ith element of the first row and the jth element of the second row are 1, all other entries are 0. Then the subspace spanned by the matrices of all maps $\mathcal{A}_{ij}$ is the set of matrices where the sums of the two rows are equal.

Using matrices, we can also handle the generalization of the problem. Let $V_1 = T^n, V_2 = T^k$ and

$$\mathcal{A}_{i_1,i_2,\dots,i_k}\begin{pmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_n\end{pmatrix} = \begin{pmatrix}\alpha_{i_1}\\ \vdots\\ \alpha_{i_k}\end{pmatrix}, \qquad 1\le i_j\le n, \quad j = 1,\dots,k.$$

Then these n^k linear maps have the following properties.

(A) Any three (distinct) maps are linearly independent in $\mathrm{Hom}\,(V_1\,,\,V_2)$ (if $nk\ge 3$).
(B) There exist four linearly independent maps (except if n or k is 1).
(C) The maximal number of linearly independent maps is $(n-1)k+1$.

• **5.6.24** Let F_u denote the set of quaternions $\alpha+\beta u$ where u is a fixed quaternion and $\alpha,\ \beta\in\mathbf{R}$. By the statement in the hint, F_v is a subalgebra of the quaternion algebra isomorphic to the complex numbers. Assume that the quaternions w and z are nth roots of v. Since v is not a real number, the same holds for w and z, too. Thus both subalgebras F_w and F_z have dimension 2, contain $\mathbf{R}$ and $(w^n = z^n =)v$, therefore the intersection of F_w and F_z has dimension 2. This implies $F_w = F_z = F_v$, hence the nth roots of v are in F_v. As F_v is isomorphic to the complex numbers, every non-zero element in it has exactly n (distinct) nth roots in F_v. So v has exactly n (distinct) nth roots among the quaternions.

6. Eigenvalue, Minimal Polynomial

• **6.3.15** Let $r^*(x) = r(x^2)$ for any polynomial r. Then $r(\mathcal{A}^2) = r^*(\mathcal{A})$, so

$$r(\mathcal{A}^2) = \mathcal{O} \iff m_{\mathcal{A}} \mid r^*. \tag{S.6.1}$$

We shall need the following statement:

Lemma: If $\lambda \neq 0$, then the multiplicities of the roots λ^2 in r and λ in r^* are equal.

Proof of the lemma: Let j be the multiplicity of λ^2 in r, i.e., $r = (x-\lambda^2)^j h$ where $h(\lambda^2) \neq 0$. Then $r^* = (x^2 - \lambda^2)^j h^* = (x - \lambda)^j (x + \lambda)^j h^*$, so the multiplicity of λ in r^* is (exactly) j since $-\lambda \neq \lambda$ and $h^*(\lambda) = h(\lambda^2) \neq 0$.

As a further preparation, we separate the largest power of x in $m_{\mathcal{A}}$:

$$m_{\mathcal{A}} = x^k g \quad \text{where} \quad k \geq 0 \text{ and } g(0) \neq 0\,. \tag{S.6.2}$$

To prove *necessity*, we assume that the minimal polynomials of $\mathcal{A}$ and $\mathcal{A}^2$ are equal and verify (i), (ii), and (iii).

(i) If $m_{\mathcal{A}}(\lambda) = 0$, then λ is an eigenvalue of $\mathcal{A}$. The two minimal polynomials are equal, so the eigenvalues are the same, hence λ is an eigenvalue of $\mathcal{A}^2$, too. Therefore, a square root λ_1 of λ is an eigenvalue of $\mathcal{A}$ (by Exercise 6.1.4c), so $m_{\mathcal{A}}(\lambda_1) = 0$. Repeating the argument, we obtain that also $\lambda_2, \lambda_3, \ldots$ are roots of $m_{\mathcal{A}}$ where λ_{i+1} is one of the two values of $\sqrt{\lambda_i}$.

The number of roots of $m_{\mathcal{A}}$ is finite, so $\lambda_i = \lambda_j$ for some $i > j$. Raising this equality to the exponent 2^i, we get $\lambda = \lambda^{2^{i-j}}$, so $\lambda\big(1-\lambda^{2^{i-j}-1}\big) = 0$. This means that λ is 0 or a root of unity of odd order.

(ii) For a proof by contradiction, assume $x^2 \,|\, m_{\mathcal{A}}$, i.e., $k \geq 2$ in (S.6.2). We show that $\mathcal{A}^2$ is a root of the polynomial $t = m_{\mathcal{A}}/x = x^{k-1} g$ contradicting $m_{\mathcal{A}^2} = m_{\mathcal{A}}$.

We shall repeatedly use (S.6.1) and (S.6.2). By the condition, $\mathcal{A}^2$ is a root of $m_{\mathcal{A}}$, further

$$m_{\mathcal{A}}(\mathcal{A}^2) = \mathcal{O} \iff m_{\mathcal{A}} \mid m_{\mathcal{A}}^* \iff x^k g \mid x^{2k} g^*$$

implying $g \mid g^*$ by $(g, x) = 1$. Similarly,

$$t(\mathcal{A}^2) = \mathcal{O} \iff m_{\mathcal{A}} \mid t^* \iff x^k g \mid x^{2k-2} g^*,$$

and the last divisibility holds due to $g \mid g^*$ and $k \leq 2k - 2$. So, $\mathcal{A}^2$ is a root of t.

(iii) The statement is obvious for 0. Let $\lambda \neq 0$ and assume that the multiplicities of λ and λ^2 in $m_{\mathcal{A}}$ are k and j. We prove $j \geq k$ first.

Due to $m_{\mathcal{A}}(\mathcal{A}^2) = \mathcal{O}$, we have $m_{\mathcal{A}} \mid m_{\mathcal{A}}^*$, so the multiplicity of λ in $m_{\mathcal{A}}^*$ is at least k. On the other hand, by the lemma, the multiplicity of λ in $m_{\mathcal{A}}^*$ is exactly j. Combining the two results, we get $j \geq k$.

Consider now the sequence $\lambda, \lambda^2, \lambda^4, \lambda^8, \ldots$. We just proved that no element has a smaller multiplicity in $m_{\mathcal{A}}$ than the preceding one. But this is a purely periodic sequence by (i), so all multiplicities must be equal.

To prove *sufficiency*, we have to check that conditions (i), (ii), and (iii) imply

$$\text{(a) } m_{\mathcal{A}}(\mathcal{A}^2) = \mathcal{O} \qquad \text{and} \qquad \text{(b) } (\mathcal{A}^2) = \mathcal{O} \Longrightarrow m_{\mathcal{A}} \mid s\,.$$

By (S.6.1), statements (a) and (b) are equivalent to

$$\text{(a1) } m_{\mathcal{A}} \mid m_{\mathcal{A}}^* \qquad \text{and} \qquad \text{(b1) } m_{\mathcal{A}} \mid s^* \Longrightarrow m_{\mathcal{A}} \mid s.$$

We verify (a1) first. By (iii), for every root, also its square is a root of $m_{\mathcal{A}}$ with the same multiplicity. We show that the squares of distinct roots are distinct. Equivalently, if $\lambda \neq 0$ is a root, then $-\lambda$ cannot be a root of $m_{\mathcal{A}}$. By (i), λ is a root of unity of odd order. Then the order of $-\lambda$ is the double of the order of λ. But by (i), the order of a root of $m_{\mathcal{A}}$ cannot be even.

By the above, replacing every root by its square in the factorization of $m_{\mathcal{A}}$, we get the same polynomial:

$$m_{\mathcal{A}} = \prod_{i=1}^{r} (x - \lambda_i)^{k_i} = \prod_{i=1}^{r} (x - \lambda_i^2)^{k_i}\,.$$

This implies

$$m_{\mathcal{A}}^* = \prod_{i=1}^{r} (x^2 - \lambda_i^2)^{k_i} = \prod_{i=1}^{r} (x - \lambda_i)^{k_i} \prod_{i=1}^{r} (x + \lambda_i)^{k_i} = m_{\mathcal{A}} \prod_{i=1}^{r} (x + \lambda_i)^{k_i}\,,$$

so $m_{\mathcal{A}} \mid m_{\mathcal{A}}^*$.

Turning to (b1), we have to show that the multiplicity of every root is at least as large in s as in $m_{\mathcal{A}}$.

If 0 is a root of $m_{\mathcal{A}}$, then it is not a multiple root by (ii). So it is sufficient to verify that 0 is a root of s. By $m_{\mathcal{A}}(0) = 0$, 0 is an eigenvalue of $\mathcal{A}$, so it is an eigenvalue of $\mathcal{A}^2$, too. The condition $m_{\mathcal{A}} \mid s^*$ implies $s(\mathcal{A}^2) = \mathcal{O}$ by (S.6.1), so the eigenvalues of $\mathcal{A}^2$ are roots of s, thus $s(0) = 0$.

Let now $\mu \neq 0$ be a root of $m_{\mathcal{A}}$ with multiplicity j. Conditions (i) and (iii) imply that also one of the square roots of μ is a root of $m_{\mathcal{A}}$ with the same multiplicity. Let λ denote this square root of μ, so $\mu = \lambda^2$.

Let the multiplicity of $\mu = \lambda^2$ in s be j'. Then, by the lemma, also the multiplicity of λ in s^* is j'. Since $m_{\mathcal{A}} \mid s^*$, comparing the multiplicities of λ

we obtain $j \le j'$. This means that the multiplicity of μ in s is not smaller than in $m_{\mathcal{A}}$, as claimed.

• **6.4.10 (a)** The image of any linear map is a subspace, so $U = \operatorname{Im} f(\mathcal{A})$ is a subspace. We show that U is $\mathcal{A}$-invariant. Let $\mathbf{u} \in U$, i.e. $\mathbf{u} = f(\mathcal{A})\mathbf{x}$ for some $\mathbf{x} \in V$. Then $\mathcal{A}\mathbf{u} = \mathcal{A}\big(f(\mathcal{A})\mathbf{x}\big) = \big(\mathcal{A}f(\mathcal{A})\big)\mathbf{x} = \big(f(\mathcal{A})\mathcal{A}\big)\mathbf{x} = f(\mathcal{A})(\mathcal{A}\mathbf{x})$, so $\mathcal{A}\mathbf{u} \in U$. (We could have referred to Exercise 6.4.6 with $\mathcal{B} = f(\mathcal{A})$.)

• **(b)** We verify first that $(f\,,\, m_{\mathcal{A}}) = (g\,,\, m_{\mathcal{A}})$ implies $\operatorname{Im} f(\mathcal{A}) = \operatorname{Im} g(\mathcal{A})$. Let $(f\,,\, m_{\mathcal{A}}) = d$; we have to show $\operatorname{Im} f(\mathcal{A}) = \operatorname{Im} d(\mathcal{A})$. Using $f(\mathcal{A})\mathbf{x} = d(\mathcal{A})\big((f/d)(\mathcal{A})\mathbf{x}\big)$ we obtain $\operatorname{Im} f(\mathcal{A}) \subseteq \operatorname{Im} d(\mathcal{A})$. On the other hand, $d = sf + tm_{\mathcal{A}}$ implies

$$d(\mathcal{A})\mathbf{z} = f(\mathcal{A})\big(s(\mathcal{A})\mathbf{z}\big) + t(\mathcal{A})\big(m_{\mathcal{A}}(\mathcal{A})\mathbf{z}\big) = f(\mathcal{A})\big(s(\mathcal{A})\mathbf{z}\big) + \mathbf{0},$$

so $\operatorname{Im} d(\mathcal{A}) \subseteq \operatorname{Im} f(\mathcal{A})$.

To prove the converse, we assume $\operatorname{Im} f(\mathcal{A}) = \operatorname{Im} g(\mathcal{A})$ and will deduce $(f\,,\, m_{\mathcal{A}}) = (g\,,\, m_{\mathcal{A}})$. By the above, it is sufficient to prove that if d_1 and d_2 are divisors of the minimal polynomial $m_{\mathcal{A}}$ satisfying $\operatorname{Im} d_1(\mathcal{A}) = \operatorname{Im} d_2(\mathcal{A})$, then d_1 and d_2 differ only in a constant factor. Let $\operatorname{Im} d_1(\mathcal{A}) = \operatorname{Im} d_2(\mathcal{A})$ and $m_{\mathcal{A}} = h_1 d_1 = h_2 d_2$. Since $\mathcal{O} = m_{\mathcal{A}}(\mathcal{A}) = h_1(\mathcal{A})d_1(\mathcal{A})$, we get $\operatorname{Im} d_1(\mathcal{A}) \subseteq \operatorname{Ker} h_1(\mathcal{A})$. But $\operatorname{Im} d_1(\mathcal{A}) = \operatorname{Im} d_2(\mathcal{A})$, so $\operatorname{Im} d_2(\mathcal{A}) \subseteq \operatorname{Ker} h_1(\mathcal{A})$ implying $h_1(\mathcal{A})d_2(\mathcal{A}) = \mathcal{O}$. Therefore $h_1 d_1 = m_{\mathcal{A}} \,|\, h_1 d_2$, so $d_1 \,|\, d_2$. We get $d_2 \,|\, d_1$ similarly, hence d_1 is a constant multiple of d_2.

• **(c)** Let d_i be pairwise non-associate divisors of the minimal polynomial. By (b), the subspaces $\operatorname{Im} d_i(\mathcal{A})$ are all distinct.

• **(d)** We claim that $\mathcal{A}$ has only trivial invariant subspaces if and only if $m_{\mathcal{A}}$ is irreducible (over F) and $\deg m_{\mathcal{A}} = \dim V$. (As mentioned at the answer to Exercise 6.4.9d, this is equivalent to the irreducibility of $k_{\mathcal{A}}$.)

If $m_{\mathcal{A}}$ is reducible, then $\operatorname{Im} d(\mathcal{A})$ is a non-trivial invariant subspace for a non-trivial $d \,|\, m_{\mathcal{A}}$. If $\deg m_{\mathcal{A}} < \dim V$, then $\langle \mathbf{u}, \mathcal{A}\rangle$ is a non-trivial invariant subspace for any $\mathbf{u} \neq \mathbf{0}$ since $\dim\langle \mathbf{u}, \mathcal{A}\rangle \le \deg m_{\mathcal{A}}$.

To prove the converse, assume that $m_{\mathcal{A}}$ is irreducible, $\deg m_{\mathcal{A}} = \dim V$, and let U be an $\mathcal{A}$-invariant subspace. We have to show that if U contains a vector $\mathbf{u} \neq \mathbf{0}$, then $U = V$. The tightest invariant subspace containing $\mathbf{u}$ is $\langle \mathbf{u}, \mathcal{A}\rangle \subseteq U$, so it is sufficient to verify $\langle \mathbf{u}, \mathcal{A}\rangle = V$. For the sake of convenience, we use the notion and simple properties of the order introduced in Section 6.5 but this just makes wording easier, the proof could be formulated without it, as well. Since $o_{\mathcal{A}}(\mathbf{u})$ divides the minimal polynomial, it can only be the minimal polynomial itself (by $\mathbf{u} \neq \mathbf{0}$). So $\dim\langle \mathbf{u}, \mathcal{A}\rangle = \deg o_{\mathcal{A}}(\mathbf{u}) = \deg m_{\mathcal{A}} = \dim V$. This implies $\langle \mathbf{u}, \mathcal{A}\rangle = V$ (as $\dim V$ is finite).

7. Bilinear Functions

• **7.1.9 (b)** We work with the matrices of the functions. Let $[\mathbf{A}] = (\alpha_{ij})_{1\le i,j\le n}$. We define the bilinear function $\mathbf{A}_{ij}$ by a matrix where the jth element in the ith row is α_{ij} and all other entries are 0. Then $\mathbf{A}_{ij}(\mathbf{u},\mathbf{v}) = \Phi(\mathbf{u})\Psi(\mathbf{v})$ where Φ assumes α_{ij} at $\mathbf{b}_i$ and 0 at the other basis elements, and Ψ assumes 1 at $\mathbf{b}_j$ and 0 at the other basis elements. Since $\mathbf{A} = \sum_{1\le i,j\le n} \mathbf{A}_{ij}$, we get that $\mathbf{A}$ has the desired representation (with $r = n^2$ terms).

We show that the minimal value of r is the rank $\mathrm{rk}([\mathbf{A}])$ of the matrix of $\mathbf{A}$. The rank of a matrix of a bilinear function of the form $\Phi(\mathbf{u})\Psi(\mathbf{v})$ is at most 1 and the rank of the sum of matrices is not greater than the sum of the ranks, so the rank of the matrix of $\mathbf{A}(\mathbf{u},\mathbf{v}) = \sum_{m=1}^{r} \Phi_m(\mathbf{u})\Psi_m(\mathbf{v})$ is at most r, i.e., $\mathrm{rk}([\mathbf{A}]) \le r$.

We have to show that $\mathbf{A}$ can be represented as a sum of $\mathrm{rk}([\mathbf{A}])$ such terms. This means that any matrix A of rank r is the sum of r dyads, i.e., products of $n \times 1$ column vectors and $1 \times n$ row vectors. Let $A = (\alpha_{ij})_{1\le i,j\le n}$ and let C_j and R_i be the jth column and ith row vectors of A. Pick an $\alpha_{ij} \neq 0$ and form $B = \big((1/\alpha_{ij})C_j\big)R_i$. Then B is a dyad with the same ith row and jth column as A. Therefore the ith row and jth column of $A' = A - B$ are zero. We show below that $\mathrm{rk}(A') = \mathrm{rk}(A) - 1$. So, repeating the procedure for A' (or induction on r) yields the statement.

We subtract suitable scalar multiples of the jth column from the other columns of A to get zeros in the ith row except for the jth entry. Then we subtract suitable scalar multiples of the (new) ith row from the other rows to achieve zeros in the jth column except for α_{ij} in the ith row. These transformations have not affected the rank, so $\mathrm{rk}(A_1) = \mathrm{rk}(A)$ for the new matrix A_1. Further, A' and A_1 only differ in the jth entry of the ith row which is 0 in A' and $\alpha_{ij} \neq 0$ in A_1. Consider a maximal linearly independent set of rows in A'. This cannot contain the ith row which is zero. But the same set of rows and the ith row are independent in A_1. This proves $\mathrm{rk}(A) = \mathrm{rk}(A_1) = \mathrm{rk}(A') + 1$ as stated.

Note that the proof made no use of A being a square matrix: we obtained that any matrix A is the sum of dyads and the minimal number of terms in this sum is the rank of A. (This also holds for the zero matrix by the usual convention of defining the empty sum as zero.)

• **7.2.9** The second proof of Theorem 7.2.3 yields that the vectors $\mathbf{A}$-orthogonal to $\mathbf{v}$ form a subspace W of dimension not less than $n-1$. Another option is to consider the homogeneous linear equation $[\mathbf{v}]^T[\mathbf{A}][\mathbf{w}] = 0$ where the variables

are the coordinates of $\mathbf{w}$. There are n unknowns and one equation, so there are at least $n-1$ free parameters. Therefore the subspace of the solutions is of dimension at least $n-1$. If $\mathbf{A}(\mathbf{v},\mathbf{v}) \neq 0$, then $\mathbf{v} \notin W$ implies $\dim W = n-1$. If $\mathbf{v} = \mathbf{0}$ or $\mathbf{A} = \mathbf{0}$, then $W = V$ trivially. This means that both $\dim W = n-1$ and $\dim W = n$ are possible.

We investigate in detail when $W = V$ occurs apart from the trivial cases $\mathbf{v} = \mathbf{0}$ and $\mathbf{A} = \mathbf{0}$. As $\mathbf{v}$ is $\mathbf{A}$-orthogonal to every vector, so $\mathbf{A}(\mathbf{v},\mathbf{v}) = 0$ holds, too.

First we show that if $\mathbf{A}$ is positive or negative semidefinite, then $W = V$. Assume, e.g., $\mathbf{A}(\mathbf{x},\mathbf{x}) \geq 0$ for every $\mathbf{x}$ and consider an $\mathbf{A}$-orthogonal basis $\mathbf{c}_1, \ldots, \mathbf{c}_n$ where $\mathbf{A}(\mathbf{c}_i,\mathbf{c}_i) = 0$ for $1 \leq i \leq t$ and $\mathbf{A}(\mathbf{c}_i,\mathbf{c}_i) > 0$ for $t < i \leq n$. Then $\mathbf{A}(\mathbf{v},\mathbf{v}) = 0 \iff \mathbf{v} \in \langle \mathbf{c}_1, \ldots, \mathbf{c}_t \rangle$. This implies that $\mathbf{v}$ is $\mathbf{A}$-orthogonal to every $\mathbf{c}_j$, $1 \leq j \leq n$, so $\mathbf{v}$ is $\mathbf{A}$-orthogonal to every element of the vector space.

Let now $\mathbf{A}$ be indefinite and $\mathbf{c}_1, \ldots, \mathbf{c}_n$ an $\mathbf{A}$-orthogonal basis where $\mathbf{A}(\mathbf{c}_i,\mathbf{c}_i) = 0$ holds exactly for $1 \leq i \leq t$ (also $t = 0$ is possible). If $\mathbf{v} \in \langle \mathbf{c}_1, \ldots, \mathbf{c}_t \rangle$, then similar to the previous paragraph, $\mathbf{v}$ is $\mathbf{A}$-orthogonal to every $\mathbf{c}_j$, so $W = V$. If $\mathbf{v} \notin \langle \mathbf{c}_1, \ldots, \mathbf{c}_t \rangle$, then $\mathbf{v}$ cannot be $\mathbf{A}$-orthogonal to each of $\mathbf{c}_{t+1}, \ldots, \mathbf{c}_n$, hence $W \neq V$ (and so $\dim W = n-1$).

• **7.2.10** Every symmetric bilinear function has a diagonal matrix where the first r entries in the main diagonal are 1s, the next s entries are -1s, and the last t entries are 0s. Here r, s, t are unique non-negative integers (by the law of inertia) satisfying $r + s + t = n$. We can characterize such a matrix by a sequence of n points and 2 bars: we draw r points for the 1s, then insert a bar, draw s points for the -1s, insert a bar, and draw t points for the 0s. This means that we have to select the positions of the 2 bars from the $n+2$ places, so the number of such sequences is $\binom{n+2}{2}$.

• **7.3.13** If $\mathbf{A}$ is not indefinite, then the argument in the solution to Exercise 7.2.9 yields that every $\mathbf{v} \neq \mathbf{0}$ can be extended to an $\mathbf{A}$-orthogonal basis. If $\mathbf{A}$ is indefinite, then this is not true. Let $\mathbf{c}_1$ and $\mathbf{c}_2$ be $\mathbf{A}$-orthogonal vectors satisfying $\mathbf{A}(\mathbf{c}_1,\mathbf{c}_1) = 1$ and $\mathbf{A}(\mathbf{c}_2,\mathbf{c}_2) = -1$. We show that $\mathbf{v} = \mathbf{c}_1 + \mathbf{c}_2 \neq \mathbf{0}$ cannot be extended to an $\mathbf{A}$-orthogonal basis. To obtain a contradiction, assume that $\mathbf{v} = \mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_n$ is an $\mathbf{A}$-orthogonal basis. Since $\mathbf{A}(\mathbf{v},\mathbf{v}) = 0$, this implies $\mathbf{A}(\mathbf{v},\mathbf{d}_i) = 0$ for every i, so $\mathbf{v}$ is $\mathbf{A}$-orthogonal to every vector in V. However, e.g., $\mathbf{A}(\mathbf{c}_1,\mathbf{v}) = \mathbf{A}(\mathbf{c}_1,\mathbf{c}_1) = 1$, a contradiction.

• **7.3.14 (a)** If $\mathbf{A} = \mathbf{0}$, then $\operatorname{Ker} \tilde{\mathbf{A}} = V$. If $\mathbf{A}$ is definite, then $\operatorname{Ker} \tilde{\mathbf{A}} = \mathbf{0}$. If $\mathbf{A}$ is semidefinite and $\mathbf{c}_1, \ldots, \mathbf{c}_t$ are the elements satisfying $\tilde{\mathbf{A}}(\mathbf{c}_i) = 0$ in

an $\mathbf{A}$-orthogonal basis, then $\operatorname{Ker}\tilde{\mathbf{A}} = \langle \mathbf{c}_1, \dots, \mathbf{c}_t \rangle$ (see, e.g., at the solution of Exercise 7.2.9). Finally we show that $\operatorname{Ker}\tilde{\mathbf{A}}$ is not a subspace if $\mathbf{A}$ is indefinite. Let $\mathbf{c}_1$ and $\mathbf{c}_2$ be $\mathbf{A}$-orthogonal vectors satisfying $\mathbf{A}(\mathbf{c}_1, \mathbf{c}_1) = 1$ and $\mathbf{A}(\mathbf{c}_2, \mathbf{c}_2) = -1$. Then $\mathbf{v} = \mathbf{c}_1 + \mathbf{c}_2$ and $\mathbf{z} = \mathbf{c}_1 - \mathbf{c}_2$ are in the kernel, but $\mathbf{v} + \mathbf{z} = 2\mathbf{c}_1 \notin \operatorname{Ker}\tilde{\mathbf{A}}$.

• **(b)** The kernel of definite and semidefinite forms is a proper subspace, so it cannot contain a basis of V. If $\mathbf{A} = \mathbf{0}$, then the kernel is the entire V, so every basis works. Finally, we show that also the kernel of indefinite forms contains a basis of V. Let the first r entries in the main diagonal of a diagonal matrix of $\mathbf{A}$ be 1, the next s entries -1, and the remaining $t = n - r - s$ entries 0 ($r \geq 1, s \geq 1$ as $\mathbf{A}$ is indefinite). Let $\mathbf{m}_1, \dots, \mathbf{m}_r, \mathbf{n}_1, \dots, \mathbf{n}_s, \mathbf{o}_1, \dots, \mathbf{o}_t$ be a corresponding $\mathbf{A}$-orthogonal basis where $\mathbf{A}(\mathbf{m}_i, \mathbf{m}_i) = 1, \mathbf{A}(\mathbf{n}_j, \mathbf{n}_j) = -1, \mathbf{A}(\mathbf{o}_k, \mathbf{o}_k) = 0$. Then $\mathbf{m}_i \pm \mathbf{n}_j$ and $\mathbf{o}_k$ are in the kernel. These vectors span V, so they contain a basis of V.

• **(c)** If the kernel is a subspace, then the maximum is its dimension. This is 0 for definite forms; the number of zeros in the main diagonal of a diagonal matrix for semidefinite forms; and n for $\mathbf{A} = \mathbf{0}$. Finally, if $\mathbf{A}$ is indefinite, then the kernel contains a basis of V by part (b), so the maximum is n.

• **(d)** If the kernel is a subspace, then its dimension is the maximum. If $\mathbf{A}$ is indefinite, then the maximum is $n - \max(r, s)$ using the notation in (b). Let, e.g., $r \geq s$, then $\mathbf{m}_i + \mathbf{n}_i$, $1 \leq i \leq s$, and $\mathbf{o}_k$, $1 \leq k \leq t$, are independent and the $s + t = n - r$ dimensional subspace spanned by them is in the kernel. On the other hand, if the dimension of a subspace U is greater than $n - r$, then $\dim U + \dim\langle \mathbf{m}_1, \dots, \mathbf{m}_r \rangle > (n - r) + r = n$, so the intersection of the two subspaces contains a vector $\mathbf{z} \neq \mathbf{0}$. But any $\mathbf{z} \neq \mathbf{0} \in \langle \mathbf{m}_1, \dots, \mathbf{m}_r \rangle$ satisfies $\mathbf{A}(\mathbf{z}, \mathbf{z}) > 0$, so $\mathbf{z} \notin \operatorname{Ker}\tilde{\mathbf{A}}$. Therefore the subspace U containing $\mathbf{z}$ cannot be part of the kernel.

8. Euclidean Spaces

• **8.2.17** We may assume that the vectors have unit length.

• **(a)** We show that if the angle is 60° between any two vectors, then they are linearly independent. This implies that there can be at most n such vectors.

By the conditions, $\mathbf{v}_i \cdot \mathbf{v}_i = 1$ and $\mathbf{v}_i \cdot \mathbf{v}_t = 1/2$ if $i \neq t$. Assume $\sum_{i=1}^{k} \lambda_i \mathbf{v}_i = \mathbf{0}$ and take its inner product with each $\mathbf{v}_j$ ($j = 1, \dots, k$). Multiplying the equations by 2, we obtain a system of linear equations $\lambda_j + \sum_{i=1}^{k} \lambda_i = 0, j = 1, \dots, k$. Dividing the sum of the equations by $k + 1$, we obtain $\sum_{i=1}^{k} \lambda_i = 0$.

Plugging back into the equations yields $\lambda_j = 0$. So, the system has only a trivial solution, i.e., the vectors $\mathbf{v}_i$ are independent, as claimed.

Now we prove by induction, that there exist n vectors $\mathbf{v}_i$ in an n-dimensional Euclidean space V with an angle of 60 degrees between any two of them. This is obvious for $n \leq 2$. Let $n > 2$ and $\mathbf{e}_1, \ldots, \mathbf{e}_n$ an orthonormal basis in V. By the induction hypothesis, in the $n-1$-dimensional Euclidean space $\langle \mathbf{e}_1, \ldots, \mathbf{e}_{n-1} \rangle$ there exist suitable vectors $\mathbf{v}_1, \ldots \mathbf{v}_{n-1}$. We show that extending them by $\mathbf{v}_n = \alpha \mathbf{e}_n + \beta(\mathbf{v}_1 + \ldots + \mathbf{v}_{n-1})$ with a suitable α and β, we get a system with the required property. We have to verify

(i) $\mathbf{v}_j \cdot \mathbf{v}_n = 1/2$ for every $j \leq n-1$ and (ii) $\mathbf{v}_n \cdot \mathbf{v}_n = 1$.

We denote $\mathbf{v}_1 + \ldots + \mathbf{v}_{n-1}$ by $\mathbf{s}$. Then $\mathbf{v}_j \cdot \mathbf{e}_n = 0$ implies that (i) is equivalent to $1/2 = \beta(\mathbf{v}_j \cdot \mathbf{s}) = n\beta/2$, so $\beta = 1/n$. Thus, (ii) can be written as $1 = \alpha^2 + \beta^2 ||\mathbf{s}||^2 = \alpha^2 + (n-1)/(2n)$. This gives $\alpha = \pm\sqrt{(n+1)/(2n)}$.

Note, that this method provides a recursive construction for the vectors $\mathbf{v}_i$, we could have even written an explicit formula for them. Analyzing the argument, it turns out that there is essentially a unique choice for $\mathbf{v}_n$, so this yields another proof that there cannot be more than n vectors with this property.

• **(b)** There are three unit vectors in the plane with an angle of 120 degrees between any two of them, so we clearly have three such vectors in every V of dimension at least 2. We show that there do not exist four such vectors. Let $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$ have this property. Then

$$||\mathbf{a} + \mathbf{b} + \mathbf{c}||^2 = ||\mathbf{a}||^2 + ||\mathbf{b}||^2 + ||\mathbf{c}||^2 + 2\mathbf{a}\cdot\mathbf{b} + 2\mathbf{a}\cdot\mathbf{c} + 2\mathbf{b}\cdot\mathbf{c} = 3 - 3 = 0,$$

so $\mathbf{a} + \mathbf{b} + \mathbf{c} = \mathbf{0}$. If there are four such vectors, then the sum of any three of them is $\mathbf{0}$. This easily implies that each vector is $\mathbf{0}$, a contradiction.

• *Remark*: We can prove the generalization similarly:

(a) The maximum is n for any acute angle.

(b) If Φ is an obtuse angle, then the number of vectors is bounded independently of the dimension and the bound depends only on Φ: the achievable maximum is $\lfloor 1 - 1/\cos\Phi \rfloor$.

(c) The remaining angles are obvious: for 0° there are infinitely many vectors; for 90° the maximum is n; and for 180° the maximum is 2.

• **8.4.14 (a)** The transformations $\mathcal{A} = \lambda\mathcal{I}$ clearly work. We show that no other transformation has this property. Note that for any two independent vectors, we can define an inner product where these vectors are orthogonal: just use a basis containing these two vectors. Consider now $\mathcal{A} \neq \lambda\mathcal{I}$. We know

that then $\mathbf{e}$ and $\mathcal{A}\mathbf{e}$ are linearly independent for some vector $\mathbf{e}$. To achieve a contradiction, assume that $\mathcal{A}^*$ does not depend on the selection of the inner product. Then $0 \neq (\mathcal{A}\mathbf{e})\cdot(\mathcal{A}\mathbf{e}) = \mathbf{e}\cdot(\mathcal{A}^*\mathcal{A}\mathbf{e})$ for every inner product. The vectors $\mathbf{e}$ and $\mathcal{A}^*\mathcal{A}\mathbf{e}$ cannot be independent as then they would be orthogonal for some inner product. Therefore $\mathcal{A}^*\mathcal{A}\mathbf{e} = \beta\mathbf{e}$. Consider now an inner product where the independent vectors $(1/\gamma)\mathbf{e}$ and $\mathcal{A}\mathbf{e}$ are elements of an orthonormal basis. Then $1 = (\mathcal{A}\mathbf{e})\cdot(\mathcal{A}\mathbf{e}) = \mathbf{e}\cdot(\mathcal{A}^*\mathcal{A}\mathbf{e}) = \mathbf{e}\cdot(\beta\mathbf{e}) = \beta|\gamma|^2$. But this cannot hold for an arbitrary γ.

• **(b)** We prove sufficiency first. We assume $S_1 = \lambda S_2$ for the inner products (then λ has to be a positive real number) and let $\mathcal{A} \in \operatorname{Hom} V$ be an arbitrary transformation. We denote the adjoints of $\mathcal{A}$ according to S_1 and S_2 by $\mathcal{A}_1$ and $\mathcal{A}_2$. We have to show that they are equal. For any $\mathbf{u}, \mathbf{v} \in V$, we have $S_2(\mathcal{A}\mathbf{u}, \mathbf{v}) = S_2(\mathbf{u}, \mathcal{A}_2\mathbf{v})$ and $S_1(\mathcal{A}\mathbf{u}, \mathbf{v}) = S_1(\mathbf{u}, \mathcal{A}_1\mathbf{v})$. Substituting $S_1 = \lambda S_2$ into the second equality and dividing by λ, we obtain $S_2(\mathcal{A}\mathbf{u}, \mathbf{v}) = S_2(\mathbf{u}, \mathcal{A}_1\mathbf{v})$. Thus $S_2(\mathbf{u}, \mathcal{A}_2\mathbf{v}) = S_2(\mathbf{u}, \mathcal{A}_1\mathbf{v})$. This holds for every $\mathbf{u}$ and $\mathbf{v}$, therefore $\mathcal{A}_1 = \mathcal{A}_2$.

To prove necessity, we assume that for every transformation $\mathcal{A}$, its adjoints according to the inner products S_1 and S_2 are the same, and want to show $S_1 = \lambda S_2$. Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ and $\mathbf{f}_1, \dots, \mathbf{f}_n$ be orthonormal bases according to S_1 and S_2. By the Gram–Schmidt orthogonalization, we may assume $\langle \mathbf{e}_1, \dots, \mathbf{e}_i \rangle = \langle \mathbf{f}_1, \dots, \mathbf{f}_i \rangle$ for every $1 \leq i \leq n$.

First we verify $\mathbf{f}_n = \lambda_n \mathbf{e}_n$ (for some λ_n). If this is not the case, then $\mathbf{f}_n$ and $\mathbf{e}_n$ are linearly independent. Let $\mathcal{A}$ be a linear transformation satisfying $\mathcal{A}\mathbf{e}_n = \mathbf{e}_{n-1}$ and $\mathcal{A}\mathbf{f}_n = \mathbf{e}_n$. Using S_1, we get $0 = \mathbf{e}_{n-1}\cdot\mathbf{e}_n = (\mathcal{A}\mathbf{e}_n)\cdot\mathbf{e}_n = \mathbf{e}_n\cdot(\mathcal{A}^*\mathbf{e}_n)$, so $\mathcal{A}^*\mathbf{e}_n \in \langle \mathbf{e}_1, \dots, \mathbf{e}_{n-1} \rangle = \langle \mathbf{f}_1, \dots, \mathbf{f}_{n-1} \rangle$. Then, by S_2, we have $\mathcal{A}^*\mathbf{e}_n \perp \mathbf{f}_n$, i.e., $0 = \mathbf{f}_n\cdot(\mathcal{A}^*\mathbf{e}_n) = (\mathcal{A}\mathbf{f}_n)\cdot\mathbf{e}_n = \mathbf{e}_n\cdot\mathbf{e}_n$, a contradiction. Thus $\mathbf{f}_n = \lambda_n \mathbf{e}_n$, indeed.

Continuing the process (or by induction), we obtain $\mathbf{f}_i = \lambda_i \mathbf{e}_i$ for every i. We show that all values $|\lambda_i|$ are equal. Let $\mathcal{A}$ be a linear transformation satisfying $\mathcal{A}\mathbf{f}_1 = \mathbf{f}_2$. By S_2, we have $1 = \mathbf{f}_2\cdot\mathbf{f}_2 = (\mathcal{A}\mathbf{f}_1)\cdot\mathbf{f}_2 = \mathbf{f}_1\cdot(\mathcal{A}^*\mathbf{f}_2)$. If $\mathbf{v} = \mathcal{A}^*\mathbf{f}_2 = \sum_{i=1}^n \alpha_i \mathbf{f}_i$, then $\mathbf{f}_1\cdot\mathbf{v} = \alpha_1$, so $\alpha_1 = 1$. Repeating all this for S_1, we get $\mathbf{f}_2\cdot\mathbf{f}_2 = (\lambda_2\mathbf{e}_2)\cdot(\lambda_2\mathbf{e}_2) = |\lambda_2|^2$, whereas $\mathbf{f}_1\cdot\mathbf{v} = (\lambda_1\mathbf{e}_1)\cdot(\sum_{i=1}^n \alpha_i\lambda_i\mathbf{e}_i) = \overline{\lambda_1}\lambda_1\alpha_1 = |\lambda_1|^2$. So $|\lambda_1| = |\lambda_2|$. We obtain similarly that all $|\lambda_i|$ are equal.

Let λ denote the common value of all $|\lambda_i|^2$. We show that $S_1 = \lambda S_2$. Let $\mathbf{c}$ and $\mathbf{d}$ be two arbitrary vectors and express them as the linear combination of the basis vectors $\mathbf{f}_i$ and $\mathbf{e}_i$: $\mathbf{c} = \sum_{i=1}^n \gamma_i \mathbf{f}_i = \sum_{i=1}^n \gamma_i\lambda_i\mathbf{e}_i$, $\mathbf{d} = \sum_{i=1}^n \delta_i \mathbf{f}_i = \sum_{i=1}^n \delta_i\lambda_i\mathbf{e}_i$. The S_2 inner product of $\mathbf{c}$ and $\mathbf{d}$ is $\rho_2 = \sum_{i=1}^n \overline{\gamma_i}\delta_i$. The S_1 inner product is $\rho_1 = \sum_{i=1}^n |\lambda_i|^2\overline{\gamma_i}\delta_i = \lambda\rho_2$. as claimed.

9. Combinatorial Applications

• **9.1.1 (a)** We can simply ask the value of each x_i, so 20 questions are sufficient. But 19 questions are not enough even if Alice declares that she will answer 0 to every question. This means a homogeneous system of 19 linear equations with 20 unknowns, so it has a non-trivial (rational, hence integer) solution besides the trivial one. Therefore the integers x_i cannot be determined uniquely.

• **(b)** Two questions suffice: (i) $x_1 + x_2 + \ldots + x_{20}$, and if the answer is N, then (ii) $x_1 + x_2(N+1) + x_3(N+1)^2 + \ldots + x_{20}(N+1)^{19}$. One question is not enough: Let this be $c_1x_1 + c_2x_2 + \ldots + c_{20}x_{20}$ where, e.g., c_1 and c_2 are (say) positive. Then we get the same answer for $x_1 = 2c_2, x_2 = c_1, x_3 = \ldots = 1$ and $x_1 = c_2, x_2 = 2c_1, x_3 = \ldots = 1$.

• **(c)** One question works. Assume first that every x_i is positive. Then $U = x_1 + (x_1 + x_2)^2 + \ldots + (x_1 + x_2 + \ldots + x_{20})^{20}$ contains all information. Since $(x_1 + x_2 + \ldots + x_{20})^{20} < U < (1 + x_1 + x_2 + \ldots + x_{20})^{20}$, the integer part of the 20th root of U is $x_1 + \ldots + x_{20}$. Replacing $x_1 + \ldots + x_{20}$ by this value, we continue the procedure to obtain the value of $x_1 + \ldots + x_{19}$. Subtracting it from $x_1 + \ldots + x_{20}$, we get x_{20}. We can determine the other values similarly. If x_i can be negative, then replace x_i in U by $(3x_i + 1)^2$.

• **9.1.2** *First proof*: (i) We verify the statement first for rational numbers. Multiplying the numbers by a constant or adding a constant to them preserve the conditions. So it is sufficient to prove the statement for non-negative integers one of which is 0. By checking parity, we find that all numbers are even. Dividing by 2, the quotients are again all even, etc. Therefore each integer must be 0.

(ii) Turning to the general case, the given 13 real numbers span an at most 13-dimensional subspace U in the vector space $\mathbf{R}$ over $\mathbf{Q}$. Express the 13 numbers as linear combinations of a basis in U. Then the conditions are satisfied in every component. Since the coefficients of the basis vectors are rational numbers, they must be equal in every component by (i). So the 13 numbers must be equal.

• *Second proof*: Let $x_1, x_2, \ldots, x_{13}$ be the numbers. Applying a translation, we may assume $x_1 = 0$. The conditions require that certain sums of the numbers x_i are equal. After ordering, this is a homogeneous system of 13 linear equations with coefficients 0 and ± 1 and 12 unknowns as we fixed $x_1 = 0$. We have to show that the system only has a trivial solution.

By (i) in the first proof, this holds in **Q**. Solving the system by Gaussian elimination, we perform the four elementary operations with the rational coefficients, so we remain within **Q**. Therefore we get the same echelon form also if we are looking for real solutions. (Cf. Exercises 3.4.8, 4.4.11, and 4.6.16.)

• *Third proof*: By Exercises 3.4.8 and 4.4.11, if the system of linear equations in the second proof only has a trivial solution in the modulo 2 field F_2, then the same holds in **R**, too. For F_2 the statement asserts that if the corresponding six-term sums have the same parity, then all 13 numbers have the same parity, which is obvious.

• *Fourth proof*: Instead of linear algebra, we use elementary number theory for the transition from **Q** to **R**. We approximate the given real numbers by rationals and apply that the statement is true for these approximating fractions.

Lemma: For any real numbers $x_1, x_2, \ldots, x_m$ and positive integer N, there exist integers $a_1, a_2, \ldots, a_m$ and a positive integer $b \le N^m$ satisfying

$$\left| x_i - \frac{a_i}{b} \right| < \frac{1}{Nb} \qquad i = 1, 2, \ldots, m.$$

Proof of the lemma: We construct the vectors $\mathbf{v}_t = (\{tx_1\}, \ldots, \{tx_m\})$ for every integer $1 \le t \le N^m + 1$, where (*) $\{y\} = y - \lfloor y \rfloor$ is the fractional part of y, i.e., its distance from the closest integer to the left. By the pigeonhole principle, there exist $t \ne s$ such that the vectors $\mathbf{v}_t$ and $\mathbf{v}_s$ differ in every component by less than $1/N$. Substituting (*) yields $|bx_i - a_i| < 1/N$ for every i with $b = |s - t|$ and suitable integers a_i. Dividing by b, we get the statement of the lemma. ■

Returning to the exercise, we apply the lemma for the given real numbers x_i and $N \ge 13$. Replacing x_i by its approximating fraction a_i/b in each equality of the six-term sums, we get that the difference of the two sides is less than $12/(Nb) < 1/b$. Since both sides are fractions with denominators b, the two sides must be equal. This means that the approximating fractions satisfy the conditions of the exercise, so they all must be equal. Performing this for an arbitrarily large N, we get that that the given real numbers x_i must be equal, too.

• **9.1.4 (a)/(i)** Answer: $\lceil \log_2 m \rceil$. We show first that $\lceil \log_2 m \rceil$ terms are sufficient. If m is not a power of two, then let $r_{ij}2^{i-1}$ be the jth coordinate of the ith term where r_{ij} is the ith digit from the right in the binary representation of j ($1 \le i \le \lceil \log_2 m \rceil$, $1 \le j \le m$). The coordinates in each vector assume

only two values since $r_{ij} = 0$ or 1. And the sum of the jth coordinates in the vectors is just the binary representation of j, as required. If m is a power of two, then we perform the above construction for $\begin{pmatrix} 0 \\ 1 \\ \vdots \\ m-1 \end{pmatrix}$ instead of the given vector $\mathbf{z} = \begin{pmatrix} 1 \\ 2 \\ \vdots \\ m \end{pmatrix}$ and add 1 to each coordinate of the (say) first vector in the sum.

We turn to verify that less than $\lceil \log_2 m \rceil$ terms are not sufficient. In the sum of t boring vectors, every coordinate is a t-term sum where each term can assume at most two values. Therefore the coordinates can assume at most 2^t distinct values. All the m coordinates of $\mathbf{z}$ are distinct, so its representation as a sum of t boring vectors has to satisfy $2^t \geq m$, i.e., $t \geq \lceil \log_2 m \rceil$.

• **(a)/(ii)** Answer: $m - 1$. We show first that $m - 1$ terms are sufficient:

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \vdots \\ \beta_m \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_1 \\ \beta_1 \\ \vdots \\ \beta_1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \beta_3 - \beta_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ \beta_4 - \beta_1 \\ \vdots \\ 0 \end{pmatrix} + \ldots + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ \beta_m - \beta_1 \end{pmatrix}.$$

We turn to verify that less than $m - 1$ terms are not sufficient. Adding a constant to every coordinate of a vector, both the original and the new vectors are the sums of the same number of boring vectors. Therefore we can restrict ourselves to vectors with first coordinate 0. Similarly, we can also assume this for the boring vectors in the sum.

Assume that every $\mathbf{v} = \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_m \end{pmatrix}$ is the sum of t boring vectors $\mathbf{u}_j$ with first coordinate 0. The other coordinates of the jth term are 0 or some real number x_j. Considering the coordinates of $\sum_{j=1}^{t} \mathbf{u}_j = \mathbf{v}$ and deleting the trivial equality $0 + 0 + \ldots + 0 = 0$ for the first coordinate, we obtain a system of $m - 1$ linear equations with t unknowns where every coefficient on the left-hand side is 0 or 1 depending on whether that particular coordinate of $\mathbf{u}_j$ is

0 or x_j. According to all possible selections of the coefficients, the number of such systems of linear equations is $2^{t(m-1)}$ (or $\binom{2^{m-1}+t-1}{t}$ if we disregard the order of terms). If every vector $\mathbf{v}$ can be represented, then at least one of the systems has to be solvable for any right-hand side $\beta_2, \beta_3, \ldots, \beta_m$. If $t < m-1$, then there are fewer unknowns than equations, so no system can be solvable for every right-hand side. Therefore, those right-hand sides $\mathbf{v}' = \begin{pmatrix} \beta_2 \\ \beta_3 \\ \vdots \\ \beta_m \end{pmatrix}$ for which a given system is solvable, form a proper subspace in $\mathbf{R}^{m-1}$. A vector space over $\mathbf{R}$ cannot be the union of finitely many proper subspaces (Exercise 4.2.12e). Thus there must exist right-hand sides $\mathbf{v}'$ for which none of the systems has a solution, i.e., there exist vectors $\mathbf{v}$ not representable as the sum of t boring vectors.

• **(b)** Answers: (i): $\lceil \log_k m \rceil$. (ii): $\lceil (m-1)/(k-1) \rceil$ (for $m \geq k$). The proofs run similar to the case $k = 2$ in (a). In (i), we can use a number system of base k. In (ii), every boring term generates $k-1$ unknowns since its coordinates can assume $k-1$ distinct non-zero values.

• **(c)/(i)** Answer: $\lceil \log_k s \rceil$, where $s = \min(m, p)$. This follows similar to the previous arguments noting that the repetitive values in the coordinates do not count separately.

• **(c)/(ii)** The previous argument can be applied if p is sufficiently large compared to m (Exercise 4.2.12f). If, however, p is small, then not only the method fails, but there is also a change in the answer. Since $\lceil \log_k p \rceil$ boring vectors are always sufficient by the construction in (i), the minimum in the case of $p \leq k^{(m-k)/(k-1)}$ will definitely be smaller than $\lceil (m-1)/(k-1) \rceil$ obtained for real numbers.

• **9.1.5 (a)** Answer: 5. To show that 5 rounds suffice, label the students by the binary representations of the integers from 0 to 31. In every round, two teams of 16 members compete. In the ith round, we divide the students into the two teams depending on whether the ith binary digit of a student is 0 or 1.

To prove that 4 rounds are not enough, we may assume that every student participates in one of the teams in every round (we can put the idling students into either team). One of the teams in the first round contains at least 16 students. At least 8 of them are in the same team in the second round. At least 4 of these students are teammates in the third round, so at least two of them are not opponents in the fourth round either.

• **(b)** Answer: 31. A possible schedule: First round: The first team consists of a single student and the other 31 students form the second team. Second round: The first team consist of one of the latter 31 students and the remaining 30 students form the second team, etc.

To show that $r < 31$ rounds are not sufficient, we consider a complete graph of 32 vertices corresponding to the 32 students. We write a real number x_i to be specified later on the ith vertex ($i = 1, 2, 3, \ldots, 32$) and write the product $x_i x_j$ on the edge connecting the ith and jth vertices. The sum S of the products on all edges is

$$S = \sum_{1\le i<j\le 32} x_i x_j = \frac{1}{2}\Big[(\sum_{m=1}^{32} x_m)^2 - \sum_{m=1}^{32} x_m^2\Big]. \tag{S.9.1}$$

We compute the contribution of the two teams to S in the r rounds: let S_k be the sum of products $x_i x_j$ on the edges connecting students opponent in the kth round. Since any two students are opponents in exactly one round, we have

$$S = \sum_{k=1}^{r} S_k. \tag{S.9.2}$$

In the kth round, every student i in the first team T_{1k} is an opponent of every student j in the second team T_{2k}. Therefore S_k is the sum of all products $x_i x_j$ with $i \in T_{1k}$ and $j \in T_{2k}$, so

$$S_k = (\sum_{i\in T_{1k}} x_i)(\sum_{j\in T_{2k}} x_j). \tag{S.9.3}$$

Combining (S.9.1)–(S.9.3), we obtain

$$\frac{1}{2}\Big[(\sum_{m=1}^{32} x_m)^2 - \sum_{m=1}^{32} x_m^2\Big] = \sum_{k=1}^{r}\Big(\sum_{i\in T_{1k}} x_i\Big)\Big(\sum_{j\in T_{2k}} x_j\Big). \tag{S.9.4}$$

Equality (S.9.4) holds for arbitrary real numbers $x_1, x_2, \ldots, x_{32}$. We choose them to satisfy

$$\sum_{m=1}^{32} x_m = 0 \qquad \text{and} \qquad \sum_{i\in T_{1k}} x_i = 0, \quad k = 1, 2, \ldots, r. \tag{S.9.5}$$

(S.9.5) is a homogeneous system of $r+1<32$ linear equations with 32 unknowns. Therefore, it has a non-trivial solution. Substituting it into (S.9.4), we get

$$\sum_{m=1}^{32} x_m^2 = 0.$$

This is a contradiction as the sum of squares of real numbers cannot be 0 unless each number is 0.

• *Remark*: We proved in (b) that partitioning a complete graph of n vertices into bipartite graphs without common edges, we need at least $n-1$ bipartite graphs. If, however, the bipartite graphs can have edges in common, then this number can be reduced to $\lceil \log_2 n \rceil$ by the argument in (a).

• **9.1.6 (a)** The answer is yes to both questions.

(a1) A construction for countably many sets: H_i consists of the products of three distinct positive primes one of which is the ith prime p_i. Then $H_i \cap H_j = \{p_i p_j p_m \mid m \neq i, j\}$, but the only element in $H_i \cap H_j \cap H_m$ is $p_i p_j p_k$.

(a2) To construct continuum many sets, we switch from the positive integers to the cube lattice in $\mathbf{R}^3$ (having the same cardinality). Fix continuum many vectors $\mathbf{v}_\alpha \in \mathbf{R}^3$ with no 0 coordinate such that any three $\mathbf{v}_\alpha$ are linearly independent: e.g., $\mathbf{v}_\alpha = (1, \alpha, \alpha^2)$, $\alpha \in \mathbf{R} \setminus \{0\}$ (see the solution of Exercise 4.6.7a). Let H_α be the lattice points between two remote planes symmetric to the origin with normal vector $\mathbf{v}_\alpha$. The intersection of two sets H_α consists of the infinitely many lattice points in the interior of an infinite prism with four sides. The intersection of three sets H_α comprises the finitely many points of the interior of a parallelepiped.

• **(b)** There exist countably many sets but not more.

(b1) To construct countably many sets, we partition the positive integers first into countably many disjoint infinite subsets, e.g., along the maximal power of two dividing the number: $U_k = \{2^k t \mid (t,2) = 1\}$. Changing the indices from the non-negative integers to their two-element subsets, we relabel these countably many disjoint subsets U_k as $V_{\{i,j\}}$ with $i \neq j$. Then the sets $H_i = \cup_{j \neq i} V_{\{i,j\}}$ meet the requirements: $|H_i \cap H_j| = |V_{\{i,j\}}| = \infty$, but no three H_i share a common element.

(b2) Every positive integer can be an element of at most two sets. So the countably many positive integers altogether can only be elements of countably many sets.

Remark: We can prove similar results for a general $k > 0$ instead of 2 when any k sets share infinitely many common elements, but the intersection of any $k+1$ sets is finite or empty.

• **9.2.16** *First proof*: Following the hint, we have to determine the number of sequences having n elements $+1$ and $n-1$ elements -1 so that the sum of the first k elements is strictly positive for every $1 \le k \le 2n-1$. Clearly, the sequences starting with -1 are bad as they do not satisfy the last condition. We can establish a bijection between the bad sequences starting with $+1$ and the sequences starting with -1 by switching the sign of the numbers occurring till the first partial sum 0. Therefore, the total number of bad sequences is the double of the sequences starting with -1, so it is $2\binom{2n-2}{n-2}$. Subtracting it from the number of all sequences yields the number of good sequences: $\binom{2n-1}{n-1} - 2\binom{2n-2}{n-2} = \binom{2n-2}{n-1}/n$.

• *Second proof*: By the argument in the hint, the power series $A(z) = \sum_{n=1}^{\infty} \alpha_n z^n$ satisfies $A^2(z) = A(z) - z$. Solving this quadratic equation, $A(z) = (1 \pm \sqrt{1-4z})/2$. Expanding $(1-4z)^{1/2}$ into binomial series, we obtain $A(z) = (1 \pm \sum_{n=0}^{\infty} \binom{1/2}{n}(-4z)^n)/2$. The negative sign is needed before the $\sum$ as the constant term is 0 in the power series of $A(z)$ (or because all further coefficients α_n are positive). So, $\alpha_n = \binom{1/2}{n}4^n(-1)^{n+1}/2$. Performing some technical transformations, we obtain $\alpha_n = \binom{2n-2}{n-1}/n$.

• *Third proof*: Let β_n be the number of possible products of n factors where also the order of factors counts, then $\beta_n = n!\alpha_n$. We prove the recursion (*) $\beta_n = (4n-6)\beta_{n-1}$. Consider any product of the numbers $a_1, a_2, \ldots, a_{n-1}$, this means $n-2$ multiplications. We can multiply a_n by any factor of the first multiplication from the left or from the right, by any factor of the second multiplication from the left or from the right, etc., and finally by the complete product from the left or from the right. This means $4(n-2)+2 = 4n-6$ possibilities to insert a_n thus proving (*). Applying (*) repeatedly, we obtain $\beta_n = 2^{n-1}(2n-3)!!$ yielding $\alpha_n = \binom{2n-2}{n-1}/n$.

• **9.3.2** Assuming that there is only a trivial solution, we shall force a contradiction. We show that the congruence

$$F(\mathbf{x}) = \prod_{i=1}^{k}\bigl(1 - f_i^{p-1}(x_1, x_2, \ldots, x_t)\bigr) \equiv \prod_{j=1}^{t}(1 - x_j^{p-1}) = G(\mathbf{x}) \pmod p$$

holds for all values of x_i, $1 \le i \le t$. If every $x_i \equiv 0 \pmod p$, then both sides are congruent to 1. In every other case $f_i \not\equiv 0$ and $x_j \not\equiv 0$ for some i and j.

By Fermat's little theorem, $f_i^{p-1} \equiv 1$ and $x_j^{p-1} \equiv 1 \pmod{p}$. Therefore both sides contain a factor $0 \pmod{p}$, so both sides are congruent to 0. This means that the polynomials F and G considered over the modulo p field $\mathbf{Z}_p$ assume the same values everywhere. Let H^* be the reduced form of a polynomial H over $\mathbf{Z}_p$ obtained by replacing every x_i^p in H with x_i as long as possible. Clearly, the exponents of x_i in the terms of H^* are at most $p-1$, and H and H^* assume the same values everywhere. It can be easily proven by induction on the number of variables that if the polynomials H and K assume the same values everywhere, then the (formal) polynomials H^* and K^* are equal (i.e., they have the same coefficients).

We saw that F and G assume the same values everywhere, therefore the polynomials F^* and G^* are equal. Hence, also $\deg G^* = \deg F^*$. However, by $G = G^*$ and the condition $\sum_{i=1}^{k} \deg f_i < t$, this leads to a contradiction:

$$\deg G^* = \deg G = (p-1)t > (p-1)\left(\sum_{i=1}^{k} \deg f_i\right) = \deg F \geq \deg F^* .$$

• **9.4.10** The maximum is k. There are several examples of k such subsets: the one- and two-element subsets containing a fixed element x_1; here the subset $\{x_1\}$ can be replaced by its complement; some further examples are the (non-degenerate) *finite projective planes* for certain values of k (see Exercise 9.5.11 and the Remark below at the end of the solution). We present two proofs that more than k subsets cannot satisfy the condition.

• *First proof*: Following the hint, we verify that the usual 0–1 vectors $\mathbf{h}_1, \ldots, \mathbf{h}_n$ in $\mathbf{R}^k$ corresponding to the sets $H_1, \ldots, H_n$ are linearly independent. This implies $n \leq \dim \mathbf{R}^k = k$.

Consider a linear combination $\delta_1\mathbf{h}_1 + \ldots + \delta_n\mathbf{h}_n = \mathbf{0}$ with real coefficients δ_j. Taking its inner product with itself, we get

$$(\sum_{j=1}^{n} \delta_j)^2 + \sum_{j=1}^{n} \delta_j^2(|H_j| - 1) = 0.$$

The sum of non-negative numbers can only be 0 if every term is 0, further $|H_j| > 1$ for at least $n-1$ values of j by the condition. Therefore every $\delta_j = 0$ as stated.

• *Second proof*: Let us call points the elements $x_1, x_2, \ldots, x_k$ of the set X. The *degree* $d(x)$ of a point $x \in X$ is the number of subsets H_j containing x: $d(x) = \big|\{j \,|\, x \in H_j\}\big|$.

We verify $x \notin H_j \Rightarrow d(x) \leq |H_j|$. If $x \in H_t \cap H_i$ for $t \neq i$, then $H_t \cap H_j \neq H_i \cap H_j$ since x is the only element in $H_t \cap H_i$. Therefore $H_t \cap H_j$ consists of a different point of H_j for each subset H_t containing x. This proves $d(x) \leq |H_j|$.

For a proof by contradiction, we assume $k < n$. The inequality $d(x) \leq |H_j|$ for $x \notin H_j$ implies

$$\frac{d(x)}{n - d(x)} < \frac{|H_j|}{k - |H_j|}. \tag{S.9.6}$$

Summing these inequalities for every pair $x \notin H_j$, we obtain $\sum_{x \in X} d(x) < \sum_{j=1}^{n} |H_j|$. This is a contradiction since equality must hold by the definition of degree.

• *Remark*: We can read further information from each proof. The first proof can be generalized for the case when any two subsets share (not necessarily one but) the same number of elements (this argument is needed to solve Exercise 9.4.11). The second proof reveals the systems of subsets in the maximal case $n = k$. The denominator of the fraction on the left-hand side of (S.9.6) is 0 if x is an element of all the k subsets. This means the one- and two-element subsets containing x which was the first example listed at the beginning of the solution. Apart from this trivial solution, the fractions make sense and we arrive at a contradiction the same way except if $d(x) = |H_j|$ for every pair $x \notin H_j$. Besides calling the elements $x_i \in X$ points, let us call the subsets H_j lines. Then the conditions are that any two points are contained in exactly one line and any two lines share exactly one common point. These structures are the finite projective planes; the second example at the beginning of the solution was a degenerate projective plane when all but one points are on one line.

• **9.4.12** All subsets of at most m elements satisfy the condition, there are $\sum_{i=0}^{m} \binom{k}{i}$ such subsets. We show that this is the maximum. Consider the 0–1 vectors $\mathbf{h}_1, \ldots, \mathbf{h}_n$ in $\mathbf{R}^k$ corresponding to the sets $H_1, \ldots, H_n$ and let $\beta_1, \ldots, \beta_m$ be all values of $|H_t \cap H_j|$ for $t \neq j$.

By the condition, $\prod_{u=1}^{m}(\mathbf{h}_t \cdot \mathbf{h}_j - \beta_u) = 0$ for every $t \neq j$. We define polynomials $f_1, \ldots, f_n$ in k variables of degree m with real coefficients as $f_j(\mathbf{x}) = \prod_{u=1}^{m}(\mathbf{x} \cdot \mathbf{h}_j - \beta_u)$. Then $f_j(\mathbf{h}_t) = 0$ for $t \neq j$. If we could show that the polynomials f_j are linearly independent, then their number n is at most the dimension of the vector space V of all polynomials in k variables of degree not greater than m. To prove independence, we would need $f_j(\mathbf{h}_j) \neq 0$ for every j (see the refined argument below), but this is not necessarily true. Therefore we have to modify our construction.

The problematic condition $f_j(\mathbf{h}_j) = 0$ means that $|H_j|$ equals some β_u. Therefore we omit such a(n eventual) factor: Let the polynomial g_j be the product of the factors $(\mathbf{x}\cdot\mathbf{h}_j - \beta_u)$ where $\beta_u \neq |H_j|$. Then, assuming $|H_1| \leq |H_2| \leq \ldots \leq |H_n|$, we have

$$g_j(\mathbf{h}_t) \begin{cases} = 0, & \text{if } t < j; \\ \neq 0, & \text{if } t = j. \end{cases}$$

The polynomials g_j are linearly independent: If they satisfy $\sum_{j=1}^{n} \lambda_j g_j = 0$, then substituting $\mathbf{h}_1, \ldots, \mathbf{h}_n$ in this order into the equation, we obtain $\lambda_1 = \ldots = \lambda_n = 0$.

Every g_j is a polynomial in k variables of degree at most m. As they are independent, $n \leq \dim V$. A natural basis in V consists of all monomials $x_1^{r_1} \cdot \ldots \cdot x_k^{r_k}$ where $\sum_{i=1}^{k} r_i \leq m$. To each such monomial, we prepare a sequence of m bullets $\bullet$ and k bars $|$: we start with r_1 bullets followed by a bar, then we draw r_2 bullets followed by a bar, etc., finally the kth bar is followed by $m - \sum_{i=1}^{k} r_i$ bullets. E.g., for $k = 3$ and $m = 4$, the monomial $x_1^2 x_3$ corresponds to the sequence $\bullet\bullet||\bullet|\bullet$. Thus we created a bijection between the elements of the basis and the sequences. There are $\binom{m+k}{m}$ sequences, so $n \leq \dim V = \binom{m+k}{m}$. Unfortunately, this is greater than the upper bound we want to prove.

To obtain the sharp upper bound, observe that $\mathbf{h}_1, \ldots, \mathbf{h}_n$ are the only vectors we substitute into the polynomials and all coordinates of these vectors are 0 or 1. Since $0^2 = 0$ and $1^2 = 1$, we get the same values of substitution if we reduce the higher powers of the variables to the first one. So it is sufficient to consider polynomials that are linear in every variable.

Accordingly, we define the polynomials G_j $(j = 1, 2, \ldots, n)$ so that we write x_s instead of x_s^2 as long as possible for all variables x_s in the polynomial g_j. Then, by the above, $G_j(\mathbf{h}_t) = g_j(\mathbf{h}_t)$ for every t and j. Repeating the previous argument for G_j instead of g_j yields that the polynomials G_j are linearly independent. On the other hand, the polynomials G_j are in the subspace W spanned by $1, x_1, \ldots, x_k, x_1x_2, \ldots, x_{k-1}x_k, x_1x_2x_3, \ldots, x_1x_2 \cdot \ldots \cdot x_m$. These $\sum_{i=0}^{m} \binom{k}{i}$ monomials form a natural basis in W, so $n \leq \dim W = \sum_{i=0}^{m} \binom{k}{i}$ as stated.

• **9.4.18** Let $E_1, \ldots, E_b$ and $O_1, \ldots, O_c$ be the subsets where $|E_i|$ is even and $|O_j|$ is odd $(b + c = n)$ and let the corresponding 0–1 vectors in $(F_2)^k$ be $\mathbf{e}_i$ and $\mathbf{o}_j$. By the condition, the size of the intersection of any two distinct subsets is even. This means that the vectors $\mathbf{e}_1, \ldots, \mathbf{e}_b, \mathbf{o}_1, \ldots, \mathbf{o_c}$ are pairwise orthogonal. Moreover, the vectors $\mathbf{e}_i$ are orthogonal to themselves.

Let U_e and U_o be the subspaces spanned by the vectors $\mathbf{e}_i$ and $\mathbf{o}_j$ and denote their dimensions by e and o. By the proof of the Oddtown theorem, the vectors $\mathbf{o}_j$ are linearly independent, so $o = c$. Clearly, $b \le 2^e$, thus $n = b + c \le o + 2^e$.

We prove $U_e \cap U_o = \mathbf{0}$. Let $\mathbf{x} \in U_e \cap U_o$, then $\mathbf{x} = \sum_{i=1}^{b} \lambda_i \mathbf{e}_i = \sum_{j=1}^{c} \mu_j \mathbf{o}_j$. Taking the inner product of both sides and $\mathbf{o}_m$ yields $\mu_m = 0$ for any $1 \le m \le c$. So $\mathbf{x} = \mathbf{0}$ as stated.

By the condition, $\langle U_e, U_o \rangle \subseteq U_e^{\perp}$. Thus $e + o = \dim\langle U_e, U_o \rangle \le k - e$, hence $e \le \lfloor (k-o)/2 \rfloor$. This implies

$$n \le o + 2^e \le o + 2^{\lfloor (k-o)/2 \rfloor} \le 2^{\lfloor k/2 \rfloor} + \begin{cases} 0, & \text{if } k \text{ is even;} \\ 1, & \text{if } k \text{ is odd.} \end{cases}$$

Finally, to achieve this bound, we use the construction with couples in the proof of the Eventown theorem if k is even, and extend this collection of subsets with the entire X (i.e., with the club of all inhabitants) if k is odd.

• **9.5.5** Following the hint, we show the equivalence of the three conditions cyclically.

I. We prove first that if $f(A) = J$, then every eigenvector of A is an eigenvector of J. This immediately follows from the definition of eigenvectors.

II. Next we assume that every eigenvector of A is an eigenvector of J, and verify that G is a regular and connected graph. The only eigenvectors of J with a non-zero eigenvalue are the scalar multiples of $\mathbf{j}$ (with all coordinates 1). Since the eigenvectors of A span the entire vector space, $\mathbf{j}$ has to be an eigenvector of A. As the coordinates of $A\mathbf{j}$ are the degrees of the vertices, this implies that G is regular.

If G is not connected, denote one of its components by G_1 and let $\mathbf{u}$ be a vector with ith coordinate $u_i = 1$ or 0 depending on whether or not the ith vertex is in G_1. Due to regularity, also $\mathbf{u}$ is an eigenvector of A but it is not an eigenvector of J. This proves that G has to be connected.

III. Finally, we assume that G is regular and connected, and construct a polynomial f satisfying $f(A) = J$. Due to regularity, $\mathbf{j}$ is an eigenvector of A. Let $\mathbf{j}, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be orthogonal eigenbasis of A with eigenvalues $d, \lambda_2, \ldots, \lambda_n$. Then $\mathbf{v}_i \in \langle \mathbf{j} \rangle^{\perp}$, so $J\mathbf{v}_i = \mathbf{0}$ for every $2 \le i \le n$.

We show that $d \ne \lambda_i$ $(2 \le i \le n)$. This means that if $\mathbf{x}$ is an eigenvector with eigenvalue d, then $\mathbf{x}$ is a scalar multiple of $\mathbf{j}$. Consider a vertex belonging to a maximal coordinate of $\mathbf{x}$ and its d neighbors. For a simpler notation, let these be the first $d+1$ vertices. Then, by Exercise 9.5.2, $d \cdot x_1 = x_2 + \ldots + x_{d+1} \le d \cdot x_1$, so $x_1 = x_2 = \ldots = x_{d+1}$. Repeating the argument with $x_2, \ldots, x_{d+1}$

instead of x_1, etc. and using the connectedness of G, we obtain that every x_i is equal. Thus $\mathbf{x}$ is a scalar multiple of $\mathbf{j}$.

Let f be a(n interpolation) polynomial satisfying $f(d) = n$ and $f(\lambda_2) = \ldots = f(\lambda_n) = 0$. Such a polynomial exists as $d \neq \lambda_i$. Then $f(A)\mathbf{j} = f(d)\mathbf{j} = n\mathbf{j}$ and $f(A)\mathbf{v}_i = f(\lambda_i)\mathbf{v}_i = \mathbf{0}$. So $f(A)$ and J map the basis elements $\mathbf{j}, \mathbf{v}_2, \ldots, \mathbf{v}_n$ to the same vectors, therefore $f(A) = J$.

• **9.5.10** Following the hint, we prove (i) first. Let A be a non-negative symmetric matrix, Λ its maximal eigenvalue, $\mathbf{x} \geq \mathbf{0}$, $\mathbf{x} \neq \mathbf{0}$, and $A\mathbf{x} \geq \tau\mathbf{x}$. Let $\mathbf{b}_1, \ldots, \mathbf{b}_n$ be an orthonormal eigenbasis with eigenvalues $\lambda_1, \ldots, \lambda_n$ and $\mathbf{x} = \sum_{i=1}^n \xi_i \mathbf{b}_i$. We can assume that $\mathbf{x}$ is normed, i.e., $\mathbf{x}\cdot\mathbf{x} = \sum_{i=1}^n \xi_i^2 = 1$. Then

$$\tau = \mathbf{x}\cdot(\tau\mathbf{x}) \leq \mathbf{x}\cdot(A\mathbf{x}) = \sum_{i=1}^n \lambda_i \xi_i^2 \leq \Lambda \sum_{i=1}^n \xi_i^2 = \Lambda\,.$$

We apply (i) to verify (ii). Let $\mathbf{z}$ be an eigenvector of A' belonging to the maximal eigenvalue Λ': $A'\mathbf{z} = \Lambda'\mathbf{z}$. We may assume that $\mathbf{z}$ has a positive coordinate. We define a non-negative vector $\mathbf{x}$ by keeping the non-negative coordinates of $\mathbf{z}$ and replacing the (eventual) negative coordinates by 0. Then $A\mathbf{x} \geq A'\mathbf{x} \geq \Lambda'\mathbf{x}$. By (i), this implies $\Lambda \geq \Lambda'$.

We turn to prove the original statement of the exercise by induction on the number of vertices. Let $k = \Lambda + 1$, we have to show that the graph has a k-coloring. If the graph consists of a single vertex, then its only eigenvalue is $\Lambda = 0$. So $k = \Lambda + 1 = 1$ and one color clearly works. Consider now a graph G with n vertices. We define a graph G_1 of $n-1$ vertices by deleting a vertex $v \in G$ of minimal degree and the edges containing v. The degree of v is at most $\Lambda = k - 1$, by Exercise 9.5.9. Finally, Let G' be the union of G_1 and v as an isolated vertex. A simple computation shows that, apart from 0 and its multiplicity, all eigenvalues of G_1 and G' are the same. Denoting the maximal eigenvalue of G' (and G_1) by Λ', (ii) implies $\Lambda' \leq \Lambda$. So, by the induction hypothesis, G_1 can be colored by at most $\Lambda' + 1 \leq \Lambda + 1 = k$ colors. Since the deleted vertex v had at most $k-1$ neighbors in G, at least one of the k colors was not used by the neighbors. Thus the coloring of G_1 can be extended to a coloring of G.

• **9.6.3** Consider a finite field F_2 of p^2 elements and its subfield F_1 of p elements. The multiplicative group of a finite field is cyclic, so F_2 has an element Δ such that every non-zero element in F_2 is a power of Δ. Pick an arbitrary $\Theta \in F_2 \setminus F_1$, and let $\gamma_1, \ldots, \gamma_p$ be the elements of F_1. Write the elements $\Theta + \gamma_i$ as $\Theta + \gamma_i = \Delta^{a_i}$ defining thus p integers a_i between 1 and $p^2 - 1$.

We show that these meet the requirement, i.e., the sums $a_i + a_j$ are pairwise incongruent modulo $p^2 - 1$.

Assume $a_i + a_j \equiv a_k + a_l \pmod{p^2-1}$. By the definition of integers a_i, this means $(\Theta + \gamma_i)(\Theta + \gamma_j) - (\Theta + \gamma_k)(\Theta + \gamma_l) = 0$. The left-hand side is a polynomial of Θ with coefficients from F_1 and of degree at most 1 as Θ^2 gets canceled. It cannot have degree 1 (or 0) since this would imply $\Theta \in F_1$, so it must be the zero polynomial (with all coefficients 0). Then, e.g., by the uniqueness of the root factors, $\{\gamma_i, \gamma_j\} = \{\gamma_k, \gamma_l\}$, and so the same holds for the numbers a_i, too, as claimed.

• **9.6.4** Following the hint, let g be a primitive root modulo p, and let a_i be the modulo $p(p-1)$ solution of the system of congruences $x \equiv i \pmod{p-1}$, $x \equiv g^i \pmod{p}$, $i = 1, 2, \ldots, p-1$. It suffices to verify that for any c, the congruence $c \equiv a_i + a_j \pmod{p(p-1)}$ can hold with at most one (unordered) pair $\{i, j\}$ (allowing also $i = j$). By the definition of a_i, this congruence is equivalent to the system of congruences $c \equiv i + j \pmod{p-1}$, $c \equiv g^i + g^j \pmod{p}$. The first congruence can be written as $g^c \equiv g^i g^j \pmod{p}$. Hence we know both the sum and product of the numbers g^i and g^j modulo p. By the Vieta formulas on roots and coefficients, the residue classes g^i and g^j are the uniquely determined solutions of the quadratic congruence $z^2 - cz + g^c \equiv 0 \pmod{p}$, as p is a prime. Therefore also i and j are unique.

• **9.6.7 (c)** Let $Z = \sum_{i=1}^{s} a_i$ and consider the random variable η which assumes each of the 2^s sums u_j (also including 0 and Z) with probability 2^{-s}. The expectation is $E(\eta) = Z/2$, since pairing up the complementary sums u_j, the sum of every pair is Z. To compute the variance, we introduce the random variables ξ_i, $i = 1, 2, \ldots, s$, where ξ_i assumes each of the values a_i and 0 with probability $1/2$. Then the variables ξ_i are independent and their sum is η, hence we get

$$D^2(\eta) = \sum_{i=1}^{s} D^2(\xi_i) = \frac{1}{4} \sum_{i=1}^{s} a_i^2 < \frac{sn^2}{4}\,.$$

We apply now Chebyshev's inequality: the probability of $|\eta - E(\eta)| > cD(\eta)$ is less than c^{-2}. For $c = 2$, $E(\eta) = Z/2$ and $D(\eta) < n\sqrt{s}/2$ imply that more than 75% of the 2^s distinct sums u_j are in the interval with center $Z/2$ and of length $2n\sqrt{s}$. Therefore

$$\frac{3 \cdot 2^s}{4} \le 2n\sqrt{s}, \quad \text{i.e.} \quad 2^s \le \frac{8n\sqrt{s}}{3}. \tag{S.9.7}$$

Taking the logarithms in (S.9.7), we obtain

$$s < \log_2 n + \frac{\log_2 s}{2} + \log_2\left(\frac{8}{3}\right). \tag{S.9.8}$$

Using $s \le n$, the right-hand side of (S.9.8) is less than $2\log_2 n$ (for $n > 7$), so $s < 2\log_2 n$, i.e., $\log_2 s < 1 + \log_2\log_2 n$. Substituting this bound into the right-hand side of (S.9.8), we get the desired estimate.

• **9.6.9** Following the hint, consider the positive integers up to n that only have digits less than $d/2$ in the number system with base d and the sum of the squares of digits is a fixed q. If three such integers form an arithmetic progression, then the same must hold for every digit since there is no overflow to the next digit due to the restriction on the digits. Therefore every digit of the second integer is the arithmetic mean of the corresponding digits of the other two integers. Using that the sum of the squares of digits in each integer is q, a simple calculation yields that the three integers must be equal. (In other words: Considering the three integers as vectors where the coordinates are the digits, one vector is the half of the sum of the other two and each vector has the same Euclidean norm. This can happen only if the three vectors are equal.)

For a given d, the number of digits is $u \approx (\log n)/(\log d)$ and q can assume at most $ud^2/4$ values. Uniting our sets for all possible values of q, we obtain every integer having all digits less than $d/2$. This gives altogether about $n/2^u$ integers. Therefore there is a q for which the corresponding set contains at least $n/(2^{u-2}ud^2)$ integers. The maximum of this expression occurs when $\log d \approx \sqrt{\log n}$, and we obtain the value claimed in the exercise as a maximum.

• **9.6.10** *First proof*: More generally, let $k > 2$ and $n > 0$ be arbitrary integers (in the exercise, $k = 18$ and $n = 2000$). We consider all the 2^n colorings of $1, 2, \ldots, n$ and give an upper bound for the number of bad colorings containing a k-term monochromatic arithmetic progression (k-MCAP, for short). If this bound is less than 2^n, then there must be some coloring without a k-MCAP.

For a given k-MCAP, its elements can have 2 colors, and the other $n-k$ elements can be colored in 2^{n-k} ways. If the first term of the k-MCAP is j, and the difference is d, then the last term is $j + (k-1)d \le n$, hence $d \le \lfloor (n-j)/(k-1) \rfloor$. Thus at most $H = 2^{n-k+1}\sum_{j=1}^{n-1} \lfloor (n-j)/(k-1) \rfloor$ colorings contain a k-MCAP. Evaluating the inner sum S, we obtain $S \le \sum_{j=1}^{n-1}(n-j)/(k-1) < n^2/2(k-1)$, hence $H \le 2^{n-k+1}n^2/2(k-1)$. Thus, if

$$\frac{2^{n-k+1}n^2}{2(k-1)} < 2^n, \quad \text{i.e.} \quad n < 2^{k/2}\sqrt{k-1},$$

then there exists a coloring without a k-MCAP.

• *Second proof*: More generally, for $k = p+1$ where p is a prime, we can color the integers $1, 2, \ldots, p(2^p - 1)$ red and blue with no k-MCAP. (In the exercise, $p = 17$.)

We consider the field F of 2^p elements: $F = \{0, \Delta, \Delta^2, \ldots, \Delta^{2^p-1} = 1\}$. Let $b_1, \ldots, b_p$ be a basis of F (as a vector space) over (its subfield) $\mathbf{Z}_2$ and let W be the $(p-1)$-dimensional subspace where the coefficient of b_p is 0.

We define a coloring for $1, 2, \ldots, p(2^p - 1)$ without a $(p+1)$-MCAP: an integer t should be red, if $\Delta^t \in W$, and blue otherwise. Assume that we have a $p+1$-term red AP $a, a+d, \ldots, a+pd$. Then $\Delta^a, \Delta^{a+d}, \ldots, \Delta^{a+pd} \in W$. The first p elements are therefore linearly dependent, i.e., for some coefficients $\gamma_i \in \{0, 1\}$ not all 0, we have $\sum_{i=0}^{p-1} \gamma_i \Delta^{a+id} = 0$. Dividing by Δ^a we obtain that $\Theta = \Delta^d$ is the root of some non-zero polynomial over $\mathbf{Z}_2$ of degree at most $p-1$. But $\deg \Theta \mid p$, hence $\deg \Theta = 1$. Therefore, $\Theta \in \mathbf{Z}_2$, thus $\Theta = \Delta^d = 1$. This implies $2^p - 1 \mid d$, hence $a + pd \geq 1 + pd > p(2^p - 1)$, a contradiction.

For blue APs we proceed similarly. Now $\Delta^a, \Delta^{a+d}, \ldots, \Delta^{a+pd} \notin W$, hence the p differences of the consecutive elements must be in W. Therefore these p elements $\Delta^{a+id}(\Delta^d - 1) \in W$, $i = 0, \ldots, p-1$ are linearly dependent. If $\Delta^d = 1$, then we are done as in the red case, otherwise we can divide by $\Delta^d - 1$ and arrive at a contradiction the same way as before.

• *Further proofs*: We can verify by direct calculations that the other three constructions in the hint contain no 18-MCAP.

• *Remark*: Comparing the effectiveness of the various proofs, the first proof yields no construction, just proves the existence of a suitable sequence. Neither is the second proof a practical construction, further it works only for special values of k. However, by using that the primes occur fairly densely among the integers, we can get a somewhat weaker result for every k.

The other three proofs are explicit constructions. For a general k, however, they yield much weaker upper bounds for n.

Generalizing the third proof, we can take two primes around k and $k/2$ instead of 17 and 7, and the coloring is k-MCAP-free up to about $n = k^3/2$.

In the general version of the fourth proof we can work with the primes between $\sqrt{k}$ and $2\sqrt{k}$ (instead of 2, 3, 5, and 7), and the coloring is k-MCAP-free up to about $n = e^{\sqrt{k}}$.

The generalization of the fifth proof guarantees no k-MCAP up to about $n = 2k^3$.

This shows that the first two proofs guarantee the existence of good colorings up to a much larger n than the constructions of the other three proofs. We note that the first proof can be refined to give an even better bound by using the Lovász Local Lemma.

• **9.7.7 (a)** Referring to the hint, we only prove the observation indicated there. So we assume that the angles of a triangle T are linearly independent over $\mathbf{Q}$ and T is partitioned into similar triangles. We want to show that the small triangles must be similar to T and do not cut the angles of T.

Let $\alpha_1, \alpha_2, \alpha_3$ be the angles of the small triangles, then the angles of T are of the form $\sum_{i=1}^3 k_i\alpha_i$ with integers $k_i \geq 0$. Adding the angles of T, we get $\pi = \sum_{i=1}^3 m_i\alpha_i$ with integers $m_i \geq 0$. If, e.g., $m_1 = 0$, then each angle of T is a linear combination of α_2 and α_3. This is impossible since the angles of T are linearly independent. So every $m_i \geq 1$. By $\pi = \sum_{i=1}^3 \alpha_i$, this implies $m_i = 1$. Thus the angles of T are $\alpha_1, \alpha_2, \alpha_3$, i.e., T is similar to the small triangles. We also obtained that the small triangles cannot cut the angles of T.

• **(b)** We verify the two statements indicated in the hint. Let T be a triangle where both the angles and sides are linearly independent over $\mathbf{Q}$. Assume that n is not a square, i.e., $\sqrt{n}$ is irrational, but T can be partitioned into n congruent triangles S.

By (a), the independence of the angles implies that S and T are similar. Let the sides of S and T be a_1, a_2, a_3 and A_1, A_2, A_3. Then $A_i = a_i\sqrt{n}$, further $A_i = \sum_{j=1}^3 k_{ij}a_j$ with integers $k_{ij} \geq 0$. This means that a_1, a_2, a_3 is a non-trivial solution of the homogeneous system of equations

$$\begin{aligned}(k_{11}-\sqrt{n})x_1+ \quad k_{12}x_2 \quad + \quad k_{13}x_3 \quad &= 0\\ k_{21}x_1 \quad +(k_{22}-\sqrt{n})x_2 + \quad k_{23}x_3 \quad &= 0\\ k_{31}x_1 \quad + \quad k_{32}x_2 \quad + (k_{33}-\sqrt{n})x_3 &= 0\end{aligned}$$

So the coefficient matrix

$$A = \begin{pmatrix} k_{11}-\sqrt{n} & k_{12} & k_{13} \\ k_{21} & k_{22}-\sqrt{n} & k_{23} \\ k_{31} & k_{32} & k_{33}-\sqrt{n} \end{pmatrix}$$

has rank $\mathrm{rk}(A) < 3$. Since $\sqrt{n}$ is irrational, the first two rows can only be linearly dependent if $k_{13} = k_{23} = 0$. But then the first and third rows cannot be linearly dependent, so $\mathrm{rk}(A) > 1$, i.e., $\mathrm{rk}(A) = 2$. Thus there is one free parameter in the general solution of the system. Let this be (say) a_3 and choose $a_3 = 1$. Then a_1 and a_2 will be of the form $c+d\sqrt{n}$ with some rational numbers c and d. Therefore every a_i is in the two-dimensional subspace $\langle 1, \sqrt{n}\rangle$. This contradicts the linear independence of the sides a_i.

We have to show that there exists a triangle where both the angles and sides are linearly independent. By the law of sines, the latter is equivalent to the independence of the sines of the angles. We prove that if the sine of an

angle in a triangle is a rational number different from $0, \pm 1/2$, and ± 1, and the sine of another angle is transcendental, then both the angles and their sines are linearly independent over $\mathbf{Q}$.

As a preparation, we verify that if $\sin\gamma$ is algebraic, then also $\sin(s\gamma)$ is algebraic for any $s \in \mathbf{Q}$. Since $\sin\gamma$ is algebraic, so are also $\cos\gamma = \pm\sqrt{1-\sin^2\gamma}$ and $z = \cos\gamma + i\sin\gamma$. Therefore z^s and its imaginary part $\sin(s\gamma)$ are algebraic, too. We obtain similarly that if both $\sin\gamma_1$ and $\sin\gamma_2$ are algebraic, then also $\sin(\gamma_1+\gamma_2)$ is algebraic. We can summarize the above that the angles with an algebraic sine form a subspace in the usual vector space of $\mathbf{R}$ over $\mathbf{Q}$.

Returning to the statement, we assume that the angles of a triangle satisfy $\sin\alpha_1 = r$ and $\sin\alpha_2 = t$, where r is a rational number not equal to $0, \pm 1/2, \pm 1$, and t is transcendental. We verify the independence of the angles first. To obtain a contradiction, assume that a non-trivial combination of α_1, α_2, and $\alpha_3 = \pi - (\alpha_1+\alpha_2)$ with rational coefficients is 0. We can order it as $r_0\pi + r_1\alpha_1 = r_2\alpha_2$, where $r_i \in \mathbf{Q}$ and not all are 0. By the argument above, the sine of the left-hand side is algebraic, but the sine of the right-hand side is transcendental if $r_2 \neq 0$. Thus $r_2 = 0$. This, however, implies $\alpha_1/\pi \in \mathbf{Q}$ contradicting Exercise 9.7.2b. This proves that the angles α_i are independent.

Turning to the sines, assume that $\sin\alpha_1, \sin\alpha_2$, and $\sin\alpha_3 = \sin\alpha_1\cos\alpha_2 + \cos\alpha_1\sin\alpha_2$ are linearly dependent. Since $\sin\alpha_2/\sin\alpha_1$ is irrational (moreover, transcendental), linear dependence implies $\sin\alpha_1\cos\alpha_2 + \cos\alpha_1\sin\alpha_2 = r_1\sin\alpha_1 + r_2\sin\alpha_2$ with $r_i \in \mathbf{Q}$. After ordering, we get $\sin\alpha_1(\cos\alpha_2 - r_1) = \sin\alpha_2(-\cos\alpha_1 + r_2)$. Squaring both sides and replacing $\sin^2\alpha_2$ by $1-\cos^2\alpha_2$, we get a quadratic equation with algebraic coefficients for $\cos\alpha_2$ (the coefficient $\sin^2\alpha_1 + (r_2 - \cos\alpha_1)^2$ of $\cos^2\alpha_2$ is not 0). By the quadratic formula, its root $\cos\alpha_2$ is algebraic. Then also $\sin\alpha_2$ is algebraic, a contradiction.

• **(c)** If $n = k^2$, then dividing the sides of any triangle into k equal parts, we get a suitable partition.

If $n = k^2 + m^2$ with $k, m > 0$, then consider a right triangle T with legs k and m. Drawing the altitude of the hypotenuse, we get two right triangles T_k and T_m similar to T with hypotenuses k and m. Dividing the sides of T_k and T_m into k and m equal parts, we partition them into k^2 and m^2 similar triangles with hypotenuses of unit length, so all small triangles are congruent. Thus we partitioned T into $n = k^2 + m^2$ congruent triangles similar to T. (In the special case $k = m$, we work with an isosceles right triangle.)

For $n = 3k^2$, we start with a regular triangle R. Its medians divide R into six congruent right triangles with angles 30 and 60 degrees. If we consider one of the two halves after cutting R by its median, this right triangle H has the same angles, too. Therefore H can be partitioned into 3 congruent triangles

similar to H. Dividing each of the three triangles into k^2 parts, we get a partition of H into $3k^2$ congruent triangles similar to H.

10. Codes

• **10.2.1** We consider encoding functions first. An encoding function of a (k, n) code is any injective function $\varphi : F^n \to F^k$. Let $F^n = \{\mathbf{v}_1, \ldots, \mathbf{v}_{2^n}\}$. We assign any $\mathbf{c}_1 \in F^k$ to $\mathbf{v}_1$, any $\mathbf{c}_2 \neq \mathbf{c}_1$ to $\mathbf{v}_2$, etc. So the number of encoding functions is $\Gamma = \prod_{i=1}^{2^n}(2^k - i + 1)$.

The encoding functions of linear codes are injective linear maps $\mathcal{A} : F^n \to F^k$. These can be characterized by assigning linearly independent vectors $\mathbf{c}_1, \ldots, \mathbf{c}_n \in F^k$ to a basis $\mathbf{b}_1, \ldots, \mathbf{b}_n \in F^n$. Therefore $\mathbf{c}_1$ can be any non-zero vector in F^k, $\mathbf{c}_2$ any vector outside $\langle \mathbf{c}_1 \rangle$, and in general, $\mathbf{c}_j$ any vector outside $\langle \mathbf{c}_1, \ldots, \mathbf{c}_{j-1} \rangle$. So the number of linear maps is $\Lambda = \prod_{j=0}^{n-1}(2^k - 2^j)$.

As Λ is the product of some factors of Γ, we have $\Lambda \mid \Gamma$, so (b) is true.

Turning to the codes, a (k, n) code is any 2^n-element subset of F^k. Therefore the number of all codes is $\gamma = \binom{2^k}{2^n} = \prod_{i=1}^{2^n} \frac{2^k - i + 1}{i}$.

A (k, n) linear code is any n-dimensional subspace in F^k. Such a subspace is spanned by n linearly independent vectors $\mathbf{c}_1, \ldots, \mathbf{c}_n \in F^k$. We determined above that the number of such vector systems is Λ (also the order of the vectors matters). But the same subspace W is spanned by any system of n linearly independent vectors in W. Applying the previous argument to W instead of F^k, the number of such systems is $\prod_{j=0}^{n-1}(2^n - 2^j)$. So the number of linear codes is $\lambda = \prod_{j=0}^{n-1} \frac{2^k - 2^j}{2^n - 2^j}$.

We obtained that both the numerator and denominator of λ are products of some factors of the numerator and denominator of γ. This, however, does not imply $\lambda \mid \gamma$. E.g., for $n = 3$,

$$\frac{\gamma}{\lambda} = \frac{2^{k-4}(2^k - 3)(2^k - 5)(2^k - 6)(2^k - 7)}{3 \cdot 5}.$$

The factors in the numerator do not cover all possible residues mod 5; none of them is divisible by 5 if $2^k \equiv 4 \pmod 5$. As $2^6 \equiv 4 \pmod 5$, γ/λ is not an integer for $(k,n) = (6,3)$. So (a) is false.

• **10.4.3 (a)** The field F^q has exactly one 2^v-element subfield F^v for every $v \mid q$. The multiplicative group G_v of the subfield F^v is a 2^v-1-element subgroup in the cyclic multiplicative group of F^q. Hence $\Delta^{(2^q-1)/(2^v-1)}$ is a generator of G_v. In other words, G_v consists of the powers Δ^i where $(2^q-1)/(2^v-1) \mid i$. So $\Delta^i \in F^v \iff (2^q-1)/(2^v-1) \mid i$. Therefore, the tightest subfield S containing Δ^i is obtained by taking the minimal v satisfying the two divisibility conditions of the exercise. The dimension of S is this minimal v on the one hand, and is $\deg m_i$ on the other hand. Thus $\deg m_i$ is the minimal v satisfying the conditions.

• **(b)** We saw in the proof of Theorem 10.4.2 that Θ and Θ^2 have the same minimal polynomial. This implies that every $\Delta^{i\cdot 2^j}$ is a root of m_i. How many distinct values are among these powers of Δ? The first repetition occurs when $i \cdot 2^v - i$ is divisible by $o(\Delta) = 2^q - 1$ for the first time. This is the smallest v satisfying $(2^q-1)/(2^v-1) \mid i$. We proved in (a) that the minimal such v is $\deg m_i$. This means that there are $\deg m_i$ distinct values among the powers of Δ in question, these all are roots of m_i, so m_i cannot have other roots.

• **10.4.4 (a)** Similar to the Example after Theorem 10.4.1, we can take $m_1 = x^4 + x + 1$. For $\Theta = \Delta^3$ we get $\Theta^5 = \Delta^{15} = 1$, so Θ is a root of the cyclotomic polynomial $f = x^4 + x^3 + x^2 + x + 1$. This is irreducible over $F = \mathbf{Z}_2$, too, thus $m_3 = f$. Similarly, $\Psi = \Delta^5$ satisfies $\Psi^3 = \Delta^{15} = 1$, so Ψ is a root of the irreducible polynomial $h = x^2 + x + 1$, thus $m_5 = h$. Hence $s = \deg[m_1, m_3, m_5] = 4 + 4 + 2 = 10$.

• **(b)** Both 3 and 5 are coprime to the size $2^q - 1$ of the multiplicative group of the field F^q, so both Δ^3 and Δ^5 generate this group. This implies that their minimal polynomials have degree q.

By Exercise 10.4.2, we have to show that Δ, Δ^3, and Δ^5 have distinct minimal polynomials. If $m_1 = m_3$, then Δ^3 is a root of m_1. By Exercise 10.4.3b, this implies $3 \equiv 2^j \pmod{2^q - 1}$. As the numbers on both sides of the congruence are between 0 and $2^q - 1$, this would mean $3 = 2^j$ which is absurd. Assuming $m_1 = m_5$ leads to a similar contradiction. Finally, assuming $m_3 = m_5$ implies that Δ^5 is a root of m_3. The roots of m_3 are the powers of Δ with exponents

$$3, 6, 12, \ldots, 3 \cdot 2^{q-2}, 3 \cdot 2^{q-1} = 2^q + 2^{q-1} \equiv 2^{q-1} + 1 \pmod{2^q - 1}.$$

Apart from the last one, all these exponents are between 0 and 2^q-1, and none of them is 5. The mod 2^q-1 remainder $2^{q-1}+1$ of the last exponent is not 5 either as $q>3$. Thus also $m_3=m_5$ leads to a contradiction.

• **10.4.6** We show $\deg m_i = q$ for every $i \le 2t-1$ first. Assuming the converse, Exercise 10.4.3a implies that q has a proper divisor v satisfying $(2^q-1)/(2^v-1) \mid i$. Since $v \le q/2$, we get

$$i \ge (2^q-1)/(2^v-1) \ge (2^q-1)/(2^{q/2}-1) = 2^{q/2}+1.$$

By the condition, however, $i \le 2t-1 \le 2^{q/2}-1$, a contradiction.

Next we show $m_i \ne m_r$ for any two distinct odd integers i and r between 1 and $2t-1$. If $m_i = m_r$, then Δ^r is a root of m_i. By Exercise 10.4.3b, this means $r \equiv i \cdot 2^j \pmod{2^q-1}$ for some $0 \le j < q$. If $j \le q/2$, then $i \cdot 2^j \le (2^{q/2}-1)2^{q/2} < 2^q-1$, so the congruence can be replaced by equality which is clearly false. If $j > q/2$, we rewrite the congruence into the form $i2^j - y2^q = r - y$. The left-hand side is divisible by 2^j and is not zero due to $j < q$, so its absolute value is at least $2^j > 2^{q/2}$. But the absolute value of the right-hand side is smaller: $0 < r \le 2t-1 < 2^{q/2}$ and $j < q$ imply $0 \le y < i < 2^{q/2}$. The contradiction proves $m_i \ne m_r$.

So $m_1, m_3, \ldots, m_{2t-1}$ are distinct polynomials of degree q, hence $s = tq$ by Exercise 10.4.2.

• **10.4.9 (a)** We identify $F_{(k)}[x]$ with the factor ring $R_k = F[x]/(x^k-1)$, i.e., we consider the polynomials of degree at most $k-1$ as remainders in the long division by x^k-1. This has the advantage that also multiplication makes sense in $F_{(k)}[x]$ and we obtain a ring (and even an algebra).

In this setting, a cyclic code means that a codeword multiplied by x is a codeword: $x(\gamma_0+\gamma_1 x+\ldots+\gamma_{k-1}x^{k-1}) = \gamma_0 x+\gamma_1 x^2+\ldots+\gamma_{k-2}x^{k-1}+\gamma_{k-1}x^k = \gamma_{k-1} + \gamma_0 x + \gamma_1 x^2 + \ldots + \gamma_{k-2}x^{k-1}$.

In a linear code, the sum of codewords is a codeword. Thus cyclic codes can be characterized by the property that a codeword multiplied by any polynomial is a codeword. Therefore a code C is cyclic if and only if C is an *ideal* in the ring R_k.

Due to long division, $F[x]$ is a Euclidean domain, hence every ideal is a principal ideal. This gets inherited to the factor ring R_k, so a cyclic code C consists of the multiples of a suitable polynomial g. By Exercise 10.2.8c, we can assume $g \mid x^k-1$.

Turning to the converse, assume that the generator polynomial g of a polynomial code C divides x^k-1, and we have to show that C is cyclic. Let the polynomial $c = \gamma_0 + \gamma_1 x + \ldots + \gamma_{k-1}x^{k-1}$ be a codeword, i.e., $g \mid c$. Then

also the cyclic permutation $c' = \gamma_{k-1}+\gamma_0 x+\ldots+\gamma_{k-2}x^{k-1} = xc-\gamma_{k-1}(x^k-1)$ is divisible by g, so c' is a codeword.

• **10.4.10** By Exercise 10.3.9, we have to construct an $m \times k$ quasi-parity-check matrix in which any $d-1$ columns are independent. The first column can be any non-zero vector. Assume that we have already selected the first j columns. Then the $j+1$st column cannot be a linear combination of any at most $d-2$ previous columns. This excludes the zero vector, the j column vectors, their $\binom{j}{2}$ pairwise sums, etc. These are altogether at most $\sum_{i=0}^{d-2}\binom{j}{i}$ bad vectors (the actual number is smaller if certain sums coincide which may occur for $i \geq d/2$). If $\sum_{i=0}^{d-2}\binom{j}{i} < 2^m$, then we have not banned all vectors in F^m, so we can select a $j+1$st column. By the condition, this works even for $j = k-1$. Hence we can construct a suitable $m \times k$ quasi-parity-check matrix.

• **10.4.11 (a)** It is worth to start with the (forbidden) case $q = 1$ when the statement is obvious. Assume its truth for q, so the minimal weight of non-zero codewords is 2^{q-1}. In the transition from q to $q+1$, the new generator matrix will have twice as many rows and by one more column. If $G(1,q) = (\,\mathbf{1} \quad \mathbf{a}_1 \quad \ldots \quad \mathbf{a}_q\,)$, then $G(1,q+1) = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{a}_1 & \ldots & \mathbf{a}_q \\ \mathbf{1} & \mathbf{1} & \mathbf{a}_1 & \ldots & \mathbf{a}_q \end{pmatrix}$. The upper and lower halves of the second column in the new matrix are $\mathbf{0}$ and $\mathbf{1}$, and both halves are the corresponding columns of the old matrix for the other columns. This means that we get the new codewords partly by doubling the old ones and partly by extending an old codeword with its complement. The weight is 2^q in the second case and, by the induction hypothesis, is at least $2 \cdot 2^{q-1} = 2^q$ in the first case.

• **(b)** Let $d(m,q)$ denote the minimal distance of codewords. We prove $d(m,q) = 2^{q-m}$ by induction on $q+m$. The case $m = 1$ was settled in (a). So we can assume $q > m \geq 2$. Reorder the columns of the matrix $G(m,q)$ so that the columns with all 0s in the upper half should stand at the end (these are the original second column and its products with other columns). Combined with the argument in (a), we obtain the matrix $\begin{pmatrix} G(m,q-1) & 0 \\ G(m,q-1) & G(m-1,q-1) \end{pmatrix}$. This generator matrix belongs to a code of type C_5 in Exercise 10.2.7/III. As shown there, the minimal distance is

$$d(m,q) = \min\bigl(2d(m,q-1), d(m-1,q-1)\bigr).$$

By the induction hypothesis, we infer

$$d(m,q) = \min\bigl(2 \cdot 2^{q-1-m}, 2^{(q-1)-(m-1)}\bigr) = 2^{q-m}.$$

A. Basic Algebra

• **A.3.14 (b)** Let T_n denote the sum of all primitive nth roots of unity. Obviously, $T_1 = 1$. We show

I. $T_p = -1$ and $T_{p^k} = 0$ if p is a prime and $k > 1$, and

II. $T_k T_m = T_{km}$ if $(k, m) = 1$.

These clearly imply $T_n = (-1)^r$ if n is the product of r distinct primes, and 0 otherwise. We note that this is the so-called Möbius function $\mu(n)$ playing an important role in number theory. It occurs also in the solution of Exercise A.11.12b.

I. Let S_n denote the sum of all nth roots of unity. By Exercise A.3.12, $S_n = 0$ for $n > 1$. For $n = p$, all pth roots of unity are primitive, except $z = 1$, as $z^p = 1 \iff o(z) \mid p \iff o(z) = p$ or 1. So, $T_p = S_p - 1 = 0 - 1 = -1$.

For $n = p^k$ with $k > 1$,

$$\begin{aligned} z^{p^k} = 1 &\iff o(z) \mid p^k \\ &\iff o(z) = p^k \text{ or } o(z) \mid p^{k-1} \\ &\iff o(z) = p^k \text{ or } z^{p^{k-1}} = 1 \end{aligned}$$

implying $T_{p^k} = S_{p^k} - S_{p^{k-1}} = 0 - 0 = 0$.

II. We prove that if z and w are primitive kth and mth roots of unity, and $(k, m) = 1$, then zw is a primitive kmth root of unity. Let $o(z) = k$ and $o(w) = m$, we claim $o(zw) = km$. By Exercise A.3.13(b), $o(zw) \mid km$. For the converse divisibility, let $o(wz) = t$, so $1 = (wz)^t$. To eliminate w, we raise this equality to the mth power: $1 = z^{tm} w^{tm} = z^{tm}$. This implies $o(z) = k \mid tm$, so $k \mid t$ as $(k, m) = 1$. We obtain $m \mid t$ similarly. Thus $km = [k, m] \mid t$, as stated.

Now we show that, conversely, every primitive kmth root of unity u has a unique decomposition $u = zw$ where z and w are primitive kth and mth roots of unity. As $(k, m) = 1$, there exist integers r and s satisfying $1 = rk + ms$. Clearly $(r, m) = (s, k) = 1$. Then $u = u^{ms+kr} = u^{ms} u^{kr}$. We claim that $z = u^{ms}$ and $w = u^{kr}$ are primitive kth and mth roots of unity. By Exercise A.3.13(a),

$$o(z) = o(u^{ms}) = \frac{o(u)}{(o(u), ms)} = \frac{km}{(km, ms)} = \frac{k}{(k, s)} = k,$$

and similarly, $o(w) = m$. To prove the uniqueness, assume $z_1 w_1 = z_2 w_2$ where $o(z_j) = k$ and $o(w_j) = m$. Then for $v = z_1/z_2 = w_2/w_1$ we have $v^k = v^m = 1$, so $o(v) \mid (k, m) = 1$. Thus $v = 1$ yielding $z_1 = z_2$ and $w_1 = w_2$.

Let $z_1, z_2, \ldots, z_a$, $w_1, w_2, \ldots, w_b$, and $u_1, u_2, \ldots, u_c$ be all primitive kth, mth, and kmth roots of unity, resp. We just showed that the numbers u_h are exactly the products $z_i w_j$. So,

$$T_{km} = \sum_{h=1}^{c} u_h = \sum_{\substack{1\le i\le a\\ 1\le j\le b}} z_i w_j = \left(\sum_{i=1}^{a} z_i\right)\left(\sum_{j=1}^{b} w_j\right) = T_k T_m.$$

• **A.3.16** Let $U = \sum_{k=1}^{n} \cos(kx)$. If $x = 2m\pi$ for some integer m, then $\cos(kx) = 1$, so $U = n$. Hence, we may assume $x \neq 2m\pi$.

First solution: Putting $z = \cos x + i \sin x$, U is the real part of the geometric series $V = \sum_{k=1}^{n} z^k$. As $z \neq 1$, $V = z\dfrac{z^n - 1}{z - 1}$. To express $\operatorname{Re} V$ in a nicer form, let $w = \cos(x/2) + i \sin(x/2)$. Substituting $z = w^2$ into V, factoring out w^n from the numerator and w from the denominator, and using $1/w = \overline{w}$, we obtain

$$V = w^2 \frac{w^{2n} - 1}{w^2 - 1} = w^{2+n-1} \frac{w^n - (1/w)^n}{w - (1/w)} = w^{n+1} \frac{w^n - (\overline{w})^n}{w - \overline{w}} =$$
$$= w^{n+1} \frac{2i \sin(nx/2)}{2i \sin(x/2)} = \Big(\cos\big((n+1)x/2\big) + i \sin\big((n+1)x/2\big)\Big) \frac{\sin(nx/2)}{\sin(x/2)}.$$

So $U = \operatorname{Re} V = \dfrac{\sin(nx/2)\cos((n+1)x/2)}{\sin(x/2)}$.

Second solution: Following the second option in the hint, we consider $W = \sin(x/2)U$. Since $x \neq 2m\pi$, $\sin(x/2) \neq 0$. Applying the trigonometric identity in the hint, we have $\sin(x/2)\cos(kx) = \dfrac{\sin\big((2k+1)x/2\big) - \sin\big((2k-1)x/2\big)}{2}$. Adding these equalities for $k = 1, 2, \ldots, n$, we get a telescoping sum, so

$$W = \frac{\sin\big((2n+1)x/2\big) - \sin(x/2)}{2} = \sin(nx/2)\cos\big((n+1)x/2\big)$$

(in the last step we applied the same trigonometric identity once again).

• **A.8.10 (a)** If $|G| = 1$ or p for some prime p and H is a subgroup, then $|H| = 1$ or $|H| = |G|$ by Lagrange's Theorem. Therefore $H = e$ or $H = G$. Thus G has only trivial subgroups if $|G| = 1$ or p.

Now we show that no other groups have this property. Assume that the only subgroups of G are itself and e. Then necessarily $G = \langle g \rangle$ for any $g \neq e$.

We have to prove that $|G| = o(g)$ can be neither infinite, nor a composite integer. If $o(g) = \infty$, then $\langle g^2 \rangle$ is a non-trivial subgroup, a contradiction. Similarly, if $o(g)$ is composite and d is its non-trivial divisor, then $\langle g^d \rangle$ is a non-trivial subgroup, a contradiction again.

• **(b)** A finite group has only finitely many subsets, so it can have only finitely many subgroups. To prove the converse, we show that any infinite group has infinitely many subgroups. If $o(g) = \infty$ for some $g \in G$, then the subgroups $\langle g^k \rangle$, $k = 1, 2, 3, \ldots$ are all distinct. If every element in G has a finite order (but $|G| = \infty$), then take $g_1 = e$, and if $g_1, \ldots, g_i$ have already been selected, then let g_{i+1} be an arbitrary element outside the union of the cyclic subgroups $\langle g_1 \rangle, \ldots, \langle g_i \rangle$. The orders of the elements are finite, so these cyclic subgroups are finite. As $|G| = \infty$, this procedure cannot get stuck. Clearly, the obtained subgroups $\langle g_i \rangle$, $i = 1, 2, 3, \ldots$ are all distinct.

• **A.10.1** The conditions imply that M is an n-dimensional vector space over L, so any $n+1$ elements are linearly dependent. Hence, also $1, \Theta, \Theta^2, \ldots, \Theta^n$ are linearly dependent, i.e., $\sum_{i=0}^{n} \gamma_i \Theta^i = 0$ for some scalars $\gamma_0, \gamma_1, \ldots \gamma_n \in L$ not all zero. This means that Θ is a root of the non-zero polynomial $f = \sum_{i=0}^{n} \gamma_i x^i \in L[x]$. So Θ is an algebraic element of degree at most n over L.

To prove $\deg \Theta \,|\, n$, consider the chain of extensions $L \subseteq L(\Theta) \subseteq M$. By the Tower Theorem, $n = \deg(M : L) = \deg\big(M : L(\Theta)\big) \cdot \deg\big(L(\Theta) : L\big)$, and the second factor is $\deg \Theta$.

• **A.10.16** $|z| = 1$ implies $\overline{z} = 1/z$, so $\operatorname{Re} z = (z + 1/z)/2 \in \mathbf{Q}(z)$. Therefore $\mathbf{Q}(\operatorname{Re} z) \subseteq \mathbf{Q}(z)$. Obviously, $\mathbf{Q}(\operatorname{Re} z) \subseteq \mathbf{R}$, so $\mathbf{Q}(\operatorname{Re} z) \subseteq \mathbf{Q}(z) \cap \mathbf{R}$.

To prove the opposite containment, consider a real number w in $\mathbf{Q}(z)$. We have to verify $w \in \mathbf{Q}(\operatorname{Re} z)$. To make the idea more clear cut, first assume that z is an algebraic number. Then $w = \sum_{i=0}^{n-1} \alpha_i z^i$ where $\alpha_i \in \mathbf{Q}$ and $n = \deg z$. As $w \in \mathbf{R}$,

$$2w = w + \overline{w} = \sum_{i=0}^{n-1} \alpha_i (z^i + \overline{z}^i) = \sum_{i=0}^{n-1} \alpha_i \Big(z^i + \frac{1}{z^i} \Big) .$$

We show that $z^i + (1/z^i)$ is a polynomial with rational coefficients of $z + (1/z) = 2\operatorname{Re} z$, so the same applies for $2w$ and also for w, yielding $w \in \mathbf{Q}(2\operatorname{Re} z) = \mathbf{Q}(\operatorname{Re} z)$.

We prove this by induction on i. The statement is trivial for $i = 1$, and $z^2 + (1/z^2) = (z + 1/z)^2 - 2$ for $i = 2$. Assuming its validity for $i - 1$ and i and using $z^{i+1} + (1/z^{i+1}) = \big(z^i + (1/z^i)\big)\big(z + (1/z)\big) - \big(z^{i-1} + (1/z^{i-1})\big)$, it follows also for $i + 1$.

We proceed similarly if z is transcendental. Let w be a real number and $w \in \mathbf{Q}(z)$, i.e., $w = g(z)/h(z)$ where $g, h \in \mathbf{Q}[x]$. Since $w = \overline{w}$ and $\overline{z} = 1/z$, we get $g(z)/h(z) = g(1/z)/h(1/z)$, implying $g(z)h(1/z) = g(1/z)h(z)$. Denoting the common value of the two sides by u, we have $2u = g(z)h(1/z)+g(1/z)h(z)$. Performing the multiplications, $2u = \sum_i \gamma_i\big(z^i + (1/z^i)\big)$ for some $\gamma_i \in \mathbf{Q}$. As seen before, $2u$ is a polynomial with rational coefficients of $z + (1/z)$, so $u \in \mathbf{Q}(\operatorname{Re} z)$. Similarly, $v = h(z)h(1/z) \in \mathbf{Q}(\operatorname{Re} z)$. Finally, $w = u/v$, so w is an element of $\mathbf{Q}(\operatorname{Re} z)$.

• **A.10.17** Let $f = \Theta_0 + \Theta_1 x + \ldots + \Theta_n x^n$ where each Θ_i is algebraic and assume $f(\Psi) = 0$. Consider the chain of extensions

$$M_0 = \mathbf{Q}, \quad M_{i+1} = M_i(\Theta_i), \text{ for } i = 0, 1, \ldots, n, \quad \text{and} \quad M_{n+2} = M_{n+1}(\Psi)\,.$$

Every link is an extension of finite degree: $\deg(M_{i+1} : M_i) \le \deg \Theta_i$ for every $0 \le i \le n$ (equality does not hold necessarily since the degree of Θ_i over M_i can be smaller than its degree over $\mathbf{Q}$), and $\deg(M_{n+2} : M_{n+1}) \le n$. By the Tower Theorem, the degree of $M_{n+2} : \mathbf{Q}$ is finite. This implies that every element in M_{n+2}, hence also Ψ is an algebraic number.

• **A.11.8** If f divides $g_k = x^{p^k} - x$, then f has a root in the field M_k of p^k elements by Exercise A.11.7 (moreover f is a product of root factors over M_k). The degree of such a root is $\deg f$, and the degree of every element in M_k divides k.

To prove the converse, assume that $n = \deg f \,|\, k$. Then the factor ring $F_p[x]/(f)$ is a finite field of p^n elements. This field M_n is an extension of F_p with a root Θ of f: $M_n = F_p(\Theta)$. So (e.g., by Exercise A.11.7) Θ is a root of $g_n = x^{p^n} - x$. Since f and g_n have a root in common and f is irreducible, so $f \,|\, g_n$. Further, $n \,|\, k \Rightarrow p^n - 1 \,|\, p^k - 1 \Rightarrow g_n \,|\, g_k$. Hence, also $f \,|\, g_k$ follows.

• **A.11.12 (a)** The generators of the multiplicative group in a field of p^k elements are the roots of primitive polynomials of degree k. Due to irreducibility, two primitive polynomials cannot share a root, and they cannot have multiple roots either. The number of primitive elements is $\varphi(p^k - 1)$, each primitive polynomial has k distinct roots, so the number of (monic) primitive polynomials is $\varphi(p^k - 1)/k$.

• **(b)** Let I_k denote the number of monic irreducible polynomials over F_p of degree k. Following the hint, the irreducible factors of $x^{p^k} - x$ are exactly the polynomials with degrees dividing k, by Exercise A.11.8. Each such polynomial occurs once in the factorization of $x^{p^k} - x$, since $x^{p^k} - x$ has no multiple roots. Comparing the degrees, we get $p^k = \sum_{d \,|\, k} d I_d$.

To express I_k, we use the *Möbius inversion formula*: If $h(n)$ is an arbitrary complex-valued function defined on the positive integers and $H(n) = \sum_{d\,|\,n} h(d)$, then $h(n) = \sum_{d\,|\,n} \mu(d)H(n/d)$. (For the definition of $\mu(n)$ see the answer to this exercise or the solution to Exercise A.3.14(b)). We say that $H(n)$ is the *summation function* (with respect to the divisors) of $h(n)$, and $h(n)$ is the *inversion function* of $H(n)$. The proof of the formula relies on the following basic property of $\mu(n)$: $S_n = \sum_{d\,|\,n} \mu(d) = 0$ if $n > 1$, and $S_1 = 1$.

Returning to the recursion $p^k = \sum_{d\,|\,k} dI_d$, this means that the summation function of $h(n) = nI_n$ is $H(n) = \sum_{d\,|\,n} dI_d = p^n$. So, by the Möbius inversion formula, $nI_n = h(n) = \sum_{d\,|\,n} \mu(d)H(n/d) = \sum_{d\,|\,n} \mu(d)p^{n/d}$, i.e., $I_k = (1/k)\sum_{d\,|\,k} \mu(d)p^{k/d}$.