1 Inner product spaces

1. (Standard) **inner product** on $\mathbb{R}^n$: assigns to a pair of vectors $\mathbf{x}, \mathbf{y}$ the number $\langle \mathbf{x}|\mathbf{y}\rangle = x_1y_1 + x_2y_2 + \cdots + x_ny_n$.

2. (Euclidean) **length** of a vector $\mathbf{x}$ (also called the **norm**):

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}|\mathbf{x}\rangle} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

3. Geometric interpretation: $\langle \mathbf{x}|\mathbf{y}\rangle = \|\mathbf{x}\| \cdot \|\mathbf{y}\| \cdot \cos\varphi$, where φ is the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$.

4. In a "pure" vector space we do not have terms like "length" and "angle". These can be elegantly introduced by means of an inner product defined on the space.

5. An **inner product space** is a vector space V over $\mathbb{R}$ or over $\mathbb{C}$ and a map $V \times V \to \mathbb{R}$ (or $\to \mathbb{C}$), called the **inner product**, denoted by $\langle \mathbf{u}|\mathbf{v}\rangle$ (not uniform in the literature, also $\langle \mathbf{u}, \mathbf{v}\rangle$, $\mathbf{u} \cdot \mathbf{v}$ etc.). Axioms:

(PD)	$\langle \mathbf{v}|\mathbf{v}\rangle \geq 0$, with equality only for $\mathbf{v} = \mathbf{0}$,
(L1)	$\langle a\mathbf{u}|\mathbf{v}\rangle = a\langle \mathbf{u}|\mathbf{v}\rangle$ (for a real or complex number a),
(L2)	$\langle \mathbf{u}+\mathbf{v}|\mathbf{w}\rangle = \langle \mathbf{u}|\mathbf{w}\rangle + \langle \mathbf{v}|\mathbf{w}\rangle$,
(C)	$\langle \mathbf{v}|\mathbf{u}\rangle = \overline{\langle \mathbf{u}|\mathbf{v}\rangle}$ (thus $\langle \mathbf{v}|\mathbf{u}\rangle = \langle \mathbf{u}|\mathbf{v}\rangle$ in the field of real numbers).

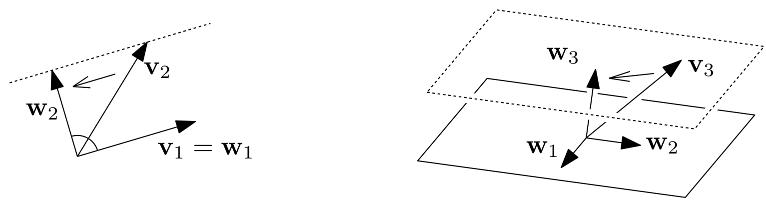
6. The standard inner product on $\mathbb{R}^n$ is the most common, but this is not the only option for an inner product on $\mathbb{R}^n$. For example, in the plane we can also define $\langle \mathbf{x} \,|\, \mathbf{y}\rangle = x_1y_1 + \frac{1}{3}x_1y_2 + \frac{1}{3}x_2y_1 + x_2y_2$ (this is related to positive definite matrices, which we will discuss later).

7. A **norm** on a vector space V (over $\mathbb{R}$ or over $\mathbb{C}$) is a map $V \to \mathbb{R}$, denoted by $\|\mathbf{v}\|$ etc.. Axioms: $\|\mathbf{v}\| \geq 0$, with equality only for $\mathbf{v} = \mathbf{0}$, $\|a\mathbf{v}\| = |a| \cdot \|\mathbf{v}\|$ (a is a real or complex number), triangle inequality $\|\mathbf{u}\| + \|\mathbf{v}\| \geq \|\mathbf{u}+\mathbf{v}\|$. The norm $\|\mathbf{v}\|$ has the meaning of a "length" of a vector $\mathbf{v}$. Any inner product determines the norm $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}|\mathbf{v}\rangle}$, but a norm seldom comes from an inner product. From the standard inner product on $\mathbb{R}^n$ mentioned above we get the Euclidean norm (by the Pythagorean theorem, $\|\mathbf{v}\|$ is just the length of the vector) and the Euclidean distance (the distance between points $\mathbf{u}$ and $\mathbf{v}$ is $\|\mathbf{u} - \mathbf{v}\|$).

8. **Cauchy–Schwarz inequality** $\langle \mathbf{u}|\mathbf{v}\rangle \le \|\mathbf{u}\| \cdot \|\mathbf{v}\|$. Proof: the quadratic polynomial $p(t) = \langle \mathbf{u} + t\mathbf{v}|\mathbf{u} + t\mathbf{v}\rangle$ must have a non-positive discriminant. Geometric meaning, connection with cosine and Pythagorean theorems. Vectors $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** if $\langle \mathbf{u}|\mathbf{v}\rangle = 0$.

9. **Orthogonal system** (nonzero pairwise orthogonal vectors), **orthonormal system** (when additionally unit vectors), their linear independence. Expression of a vector $\mathbf{v}$ in an orthonormal basis $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$: i-th coordinate is $\langle \mathbf{v}|\mathbf{b}_i\rangle$. The coordinates are sometimes called **Fourier coefficients** of the vector $\mathbf{v}$ w.r.t. the basis B.

10. **Gram–Schmidt orthogonalization:** an algorithm that converts a given basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ into an orthogonal basis $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$; the linear hull of the first k vectors is maintained for all k. Geometric illustration:



2nd step (calculate $\mathbf{w}_2$) 3rd step (calculate $\mathbf{w}_3$)

Theorem: Extendability of any orthonormal system to an orthonormal basis (in finite-dimensional spaces!). Note: Gram–Schmidt orthogonalization is numerically unstable, but stable variants are known.

11. **Orthogonal complement** of a set M:

$$M^\perp = \{\mathbf{v} \in V : \langle \mathbf{v}|\mathbf{x}\rangle = 0 \text{ for all } \mathbf{x} \in M\}.$$

12. Another way of viewing a homogeneous system of linear equations $A\mathbf{x} = \mathbf{0}$: solution set = orthogonal complement of the rowspace of the matrix A.

13. Orthogonal complement properties (all in finite dimension):

    (i) $M^\perp$ is a subspace.

    (ii) If $M_1 \subseteq M_2$, then $M_2^\perp \subseteq M_1^\perp$.

    (iii) $M^\perp = (\operatorname{span} M)^\perp$.

(iv) If U is a subspace, then $\left(U^\perp\right)^\perp = U$.

(v) $\dim(U^\perp) = \dim(V) - \dim(U)$.

(i)–(iii) are easy, while (iv),(v) follow from the extendability of an orthogonal basis.

14. An **orthogonal matrix** (silly but traditional terminology) is a square matrix A such that $AA^T = I_n$. Observation: a square matrix has orthonormal columns iff $A^{-1} = A^T$. Therefore, if a square matrix has orthonormal rows, then it also has orthonormal columns.

15. **Orthogonal projection** onto a subspace W; the projection of a point $\mathbf{x}$ is the point from the whole of W that is closest to $\mathbf{x}$. Uniqueness; expression by a formula.

2 Determinant

16. We now assign to each square matrix A a wonderful number, called the **determinant**:

$$\det(A) = \sum_{p \in S_n} \operatorname{sgn}(p) \prod_{i=1}^{n} a_{i,p(i)}$$

(an expression with $n!$ terms).

17. Example: for a 2×2 matrix we have $\det(A) = a_{11}a_{22} - a_{12}a_{21}$.

18. The determinant of a triangular matrix is the product of the elements on the diagonal.

19. $\det(A^T) = \det(A)$ (proof by shuffling of the product and sum in the definition of the determinant).

20. When rearranging the columns according to a permutation q, the determinant is multiplied by $\operatorname{sgn}(q)$ (proof similar to the previous one).

    - Corollary: Swapping two columns (or two rows) changes the sign of the determinant.

- Consequence of the consequence: If the matrix A has two identical columns (or rows), then $\det(A) = 0$.

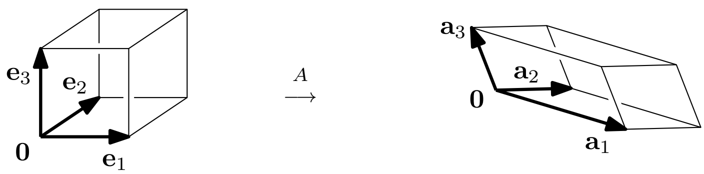
21. The determinant is a linear function of each of its rows.

22. Consequence: Effect of elementary row operations (scaling a row by a number t multiplies the determinant by t, adding the j-th row to the i-th row does not change the determinant). The same for columns.

23. Calculation of $\det(A)$ by Gaussian elimination.

    - Consequence: A square matrix A is nonsingular (i.e. invertible) iff $\det(A) \neq 0$.
    - Consequence: The rank of a matrix does not change by moving to a larger field; e.g. vectors with rational components that are linearly independent over $\mathbb{Q}$ are also linearly independent over $\mathbb{R}$.

24. Geometric meaning of the determinant: The linear map $\mathbb{R}^n \to \mathbb{R}^n$ corresponding to the matrix A transforms the unit cube to a parallelepiped of volume $|\det(A)|$:

    (and changes the area or volume of a general set in the proportion $1 : |\det(A)|$). Informal justification.

25. Note: The sign of the determinant is given by the orientation of the image of the standard basis. Nonsingular $n \times n$ matrices A, B satisfy $\operatorname{sgn}(\det(A)) = \operatorname{sgn}(\det(B))$ iff they can be connected by a "continuous path" of nonsingular matrices.

26. Theorem (determinant products): $\det(AB) = \det(A)\det(B)$. Proof: for singular A easy; nonsingular A can be expressed as a product of elementary matrices using Gaussian elimination (corresponding to row operations), and thus multiplication by A corresponds to a sequence of elementary row operations of the matrix B.

27. Determinant expansion along the i-th row:

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}),$$

where A_{ij} means the matrix formed from A by omitting the i-th row and j-th column. Proof: By linearity of the determinant as a function of a row, it is enough to verify the case when the i-th row is the standard basis vector $\mathbf{e}_j$.

28. Formula for the inverse matrix of a nonsingular matrix A: the (i,j)-entry is $(-1)^{i+j}\det(A_{ji})/\det(A)$ (proves again the existence of an inverse matrix).

29. **Cramer's rule:** If A is a nonsingular square matrix, then the (only) solution of the system $A\mathbf{x} = \mathbf{b}$ has the i-th component equal $\det(A_{i\to\mathbf{b}})/\det(A)$, where the square matrix $A_{i\to\mathbf{b}}$ is obtained from A by replacing the i-th column by the vector $\mathbf{b}$. Completely impractical for calculation, but useful for deriving properties of the solution (and also shows that the determinant arises naturally when solving a system of linear equations).

3 Eigenvalues

30. Eigenvalues are related to many questions in geometry (e.g. how isometries of Euclidean space look), physics (how a bell sounds), graph theory (how good a given graph is as a phone connection diagram), etc.

31. We will find eigenvalues through the investigation of the structure of *endomorphisms*, i.e., linear maps of a vector space V into itself. Note that for a map $X \to X$ several new questions arise that for a general map $X \to Y$ do not make sense, for example about fixed points and iterations. Such questions for linear maps can be solved by eigenvalues.

32. Consider a linear map $f\colon V \to V$ on finite-dimensional V over field $\mathbb{K}$; we want to find a basis $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ so that the matrix f w.r.t. this basis is "simple". Here it is essential that the matrix will depend on only one basis of V! (Recommended to consider: If $f\colon V \to V$ is a linear map of rank r, then two bases of V can be chosen so that the matrix f with respect to them is the matrix I_r padded with zeros at the bottom and right.)

33. Recall: the **change of basis matrix** T from a basis $B = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ to a basis $B' = (\mathbf{v}'_1, \mathbf{v}'_2, \ldots, \mathbf{v}'_n)$ has in the j-th column the coordinates of $\mathbf{v}_j$ w.r.t. B'. Claim: the change of basis matrix from B' to B is T^{-1}. Proof: by a direct calculation, or via isomorphisms with $\mathbb{K}^n$.

34. Therefore, if A is a matrix of a map $f\colon V \to V$ w.r.t. a basis B, then the matrix of f w.r.t. a basis B' is TAT^{-1}, where T is the change of basis matrix from B to B'. Square matrices A and A' are called **similar** if $A' = TAT^{-1}$ for some nonsingular matrix T.

35. Our goal in matrix language: Given a square matrix A, find a similar matrix A' that has a "simple" shape (we will see that often A' is diagonal, although not always). If you don't like linear maps, you can take this as a starting point.

36. For example, a diagonal shape is good for quickly calculating matrix powers (i.e. iterations of the linear map), and it is also possible to see from it how iterations will behave. For if $A = TDT^{-1}$ with D diagonal, then $A^k = TD^kT^{-1}$, and D^k is diagonal and has diagonal entries the k-th powers of the corresponding entries of D.

37. Warning: Elementary row operations *do not preserve* similarity of matrices! Now we have to modify matrices much more carefully!!

38. What does the linear map $\mathbb{R}^n \to \mathbb{R}^n$ with a diagonal matrix represent? Stretches or shortens, and possibly reflects, in the direction of coordinate axes. To diagonalize a matrix of a general linear map we need the directions of the "correct axes" in which the map stretches or shortens while maintaining these directions. This leads to the definition of eigenvalues and eigenvectors.

If $f\colon V \to V$ is a linear map, where V is a vector space over a field $\mathbb{K}$, then a number $\lambda \in \mathbb{K}$ is called an **eigenvalue** of the map f if and only if there exists a *nonzero* vector $\mathbf{v} \in V$ such that $f(\mathbf{v}) = \lambda\mathbf{v}$. An **eigenvector** associated with λ is any non-zero $\mathbf{v}$ satisfying $f(\mathbf{v}) = \lambda\mathbf{v}$.

Comments.

- Hence $\mathbf{v}$ is that "good direction" in which f acts as scaling by λ.

- If $\mathbf{v}$ is an eigenvector and $t \in \mathbb{K}$ is nonzero, then $t\mathbf{v}$ is an eigenvector too.
- Beware: $\mathbf{v}$ must not be $\mathbf{0}$, but λ can be 0!
- Any eigenvector $\mathbf{v}$ generates a 1-dimensional **invariant subspace**. In general, a subspace W of a space V is called an invariant subspace of a map f when $f(W) \subseteq W$.

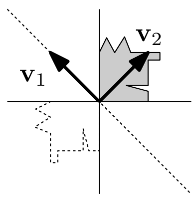
39. For a square matrix A, the eigenvalues and eigenvectors are defined as for the linear map specified by A. Explicitly:

    If A is a square matrix over a field $\mathbb{K}$, then a number $\lambda \in \mathbb{K}$ is called an **eigenvalue** of the matrix A if there exists a vector $\mathbf{v} \neq \mathbf{0}$ satisfying the equation
    $$A\mathbf{v} = \lambda\mathbf{v}.$$

    Again, sworn opponents of linear representations can be content with this matrix definition of eigenvalues.

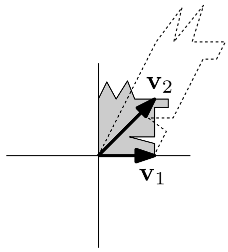
40. Examples of what can happen in the plane:

    - The matrix $\begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$ represents reflection in the line $y = -x$:

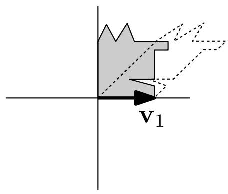
    Eigenvalues 1 (eigenvector $\mathbf{v}_1 = (-1, 1)$) and -1 ($\mathbf{v}_2 = (1, 1)$); $(\mathbf{v}_1, \mathbf{v}_2)$ form a basis, and the map with respect to this basis has diagonal matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

    - The matrix $\begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$ represents skewing and stretching:



Eigenvalues 1 ($\mathbf{v}_1 = (1, 0)$) and 2 ($\mathbf{v}_2 = (1, 1)$); $(\mathbf{v}_1, \mathbf{v}_2)$ again form a basis, and the map has with respect to this basis the diagonal matrix $\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$.

- Rotation around the origin by an angle α has no (real) eigenvalues if α is not a multiple of π, and the matrix representing this transformation is not similar to any diagonal matrix. *But* if we allow complex numbers, this matrix can be diagonalized!
- The matrix $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ represents skewing:



The only eigenvalue is 1 and the only eigenvector $(1, 0)$ (up to a scalar multiple); the matrix cannot be diagonalized – even complex numbers do not help.

41. Two more exotic examples:

- V = the space of all real functions on $[0, 1]$ having continuous derivatives of all orders; the differentiation operator $D\colon V \to V$, $f \mapsto f'$, is a linear map. Every $\lambda \in \mathbb{R}$ is an eigenvalue, with associated eigenvector the function $x \mapsto e^{\lambda x}$. Important for solving linear differential equations with constant coefficients.
- V = space of all infinite real sequences $(y_0, y_1, y_2, \ldots)$ satisfying $y_{n+2} = y_{n+1} + y_n$ for every $n = 0, 1, \ldots$ (like the recurrence

for Fibonacci numbers). $P\colon V \to V$ is the left-shift operator, $(y_0, y_1, y_2, \ldots) \mapsto (y_1, y_2, y_3, \ldots)$. Eigenvectors are obviously scalar multiples of sequences of the form $(\lambda^0, \lambda^1, \lambda^2, \ldots)$. We ask for what λ is such a sequence in V. This will result in two eigenvalues $\lambda_1, \lambda_2 = (1 \pm \sqrt{5})/2$.

42. Observation: Let $f\colon V \to V$ be linear. A basis w.r.t. which f has a diagonal matrix exists iff there is a basis consisting of eigenvectors. The corresponding diagonal matrix has precisely the eigenvalues of f on its diagonal.

43. Proposition: If $\lambda_1, \ldots, \lambda_k$ are *pairwise distinct* eigenvalues of a map f (or a matrix A), and $\mathbf{v}_i$ is some nonzero eigenvector associated with λ_i, then $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly independent. Proof by induction on k.

44. Consequence: If A is an $n \times n$ matrix with n pairwise distinct eigenvalues, then A is diagonalizable. (Opposite implication does not hold!)

45. This is a simple sufficient condition for diagonalizability. Another, which we prove later, says that every *symmetric* square matrix is diagonalizable.

46. We now express the eigenvalues of a matrix as the roots of a polynomial. Note that for fixed λ, $A\mathbf{v} = \lambda\mathbf{v}$ is a homogeneous system of n equations with n unknowns, the components of the vector $\mathbf{v}$. The matrix of this sytem is $A - \lambda I_n$, and hence λ is an eigenvalue only when $A - \lambda I_n$ singular, i.e. if and only if $\det(A - \lambda I_n) = 0$.

    The **characteristic polynomial** of a square matrix A is defined by $p_A(t) = \det(A - tI_n)$, where t is a variable.

    According to the definition of the determinant, the characteristic polynomial is indeed a polynomial, which has degree exactly n and whose roots are precisely the eigenvalues of A.

47. If A and B are similar matrices, then $p_A(t) = p_B(t)$, and therefore A and B have the same eigenvalues. So we can also talk about the characteristic polynomial $p_f(t)$ of a linear map $f\colon V \to V$ on a space of finite dimension.

48. How to find the eigenvalues and characteristic polynomial of a given matrix:

- " Manually": We can compute $\det(A - tI_n)$ by elimination; t of course we have to treat as an unknown, so we work with matrices whose entries are polynomials in the variable t (and not just numbers as usual). Gaussian elimination has to be modified so that it does not use division! In simple cases we can find $p_A(t)$ in this way.
- $p_A(t)$ can also be determined by suitably modifying the matrix A while maintaining similarity. The matrix is converted to a shape in which $p_A(t)$ can be "seen". See, for example, textbooks on numerical mathematics. The roots of $p_A(t)$ are generally obtained by numerical methods.
- In "real" applications, when the eigenvalues of e.g. 1000×1000 matrices need to be found, they are determined by other (mainly iterative) procedures that do not use the characteristic polynomial at all (e.g. by the *QR algorithm*).
- Important note: Determining eigenvalues is a computationally "manageable" problem (there are polynomial-time algorithms that are reasonably effective in practice, albeit complicated), as opposed to difficult problems such as graph coloring and other NP-complete tasks. Eigenvalues are sometimes used in algorithms for the approximate solution of some such difficult tasks.

49. *Important coefficients of the characteristic polynomial.* Let's write

$$p_A(t) = (-1)^n t^n + c_{n-1} t^{n-1} + \cdots + c_1 t + c_0.$$

Then, as can be seen from the definition of the determinant, $c_0 = \det(A)$ and $c_{n-1} = (-1)^{n-1}.\operatorname{trace}(A)$, where the number $\operatorname{trace}(A) = a_{11} + a_{22} + \cdots + a_{nn}$ is called the **trace** of the matrix A. So determinants and traces of similar matrices are equal (which can be easily seen in other ways), and we can talk about the determinant or trace of a linear map $f\colon V \to V$ on a finite-dimensional vector space.

50. *Review of polynomials.* **Fundamental theorem of algebra:** A polynomial of degree at least 1 with real or complex coefficients has at least one complex root (relatively difficult, here without proof). If $p(x)$ has a root α, then $p(x) = (x - \alpha)q(x)$ for some polynomial $q(x)$ (this is true over every field and is easy). Consequence (by induction): A polynomial $p(x)$ of degree n with real or complex coefficients can be written in the form $p(x) = a_n(x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_n)$, where $\alpha_1, \ldots, \alpha_n$ are complex

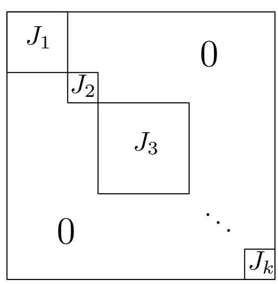
numbers. Alternatively: $p(x) = a_n(x-\beta_1)^{r_1}(x-\beta_2)^{r_2}\cdots(x-\beta_k)^{r_k}$, where $\beta_1,\ldots,\beta_k$ are pairwise distinct complex numbers and $r_1+r_2+..+r_k = n$. Here r_i is the **multiplicity** of the root β_i.

51. Comment: If the number λ is the root of multiplicity r of the characteristic polynomial $p_A(t)$, we say that λ is the eigenvalue of the matrix A of **algebraic multiplicity** r. (In particular, if λ is not an eigenvalue at all, it has algebraic multiplicity 0.) If A is diagonalizable, then the algebraic multiplicity of λ indicates how many times λ appears on the diagonal in the diagonal form.

52. Over the complex numbers we can decompose the characteristic polynomial as a product of linear factors: $p_A(t) = (-1)^n(t-\lambda_1)(t-\lambda_2)\cdots(t-\lambda_n)$. We then have $\det(A) = \lambda_1\lambda_2\cdots\lambda_n$ and $\text{trace}(A) = \lambda_1+\lambda_2+\cdots+\lambda_n$ (each eigenvalue taken with its algebraic multiplicity). For a diagonal (or diagonalizable) matrix this can be seen directly.

53. Matrices that cannot be diagonalized: Their eigenvectors do not form a basis; there must be an eigenvalue λ of algebraic multiplicity $r > 1$ such that the dimension of the solution set of the system $(A-\lambda I_n)\mathbf{x} = 0$ is smaller than r.

54. Theorem (**Jordan normal form**): Let A be a complex $n\times n$ matrix. Then there exists a matrix J similar to A, called the Jordan normal form of A, which has the following shape:



where $J_1, J_2, \ldots, J_k$ are the **Jordan cells**; the matrix J_i is of order $n_i\times n_i$

($n_1 + n_2 + \cdots + n_k = n$) and takes the form

$$J_i = \begin{pmatrix} \lambda_i & 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & \lambda_i & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & \lambda_i & 1 & 0 & \dots & 0 \\ & & & \vdots & & & \\ 0 & 0 & 0 & 0 & \dots & \lambda_i & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & \lambda_i \end{pmatrix}.$$

These λ_i need not to be distinct; in total, each eigenvalue of the matrix A appears on the diagonal of the Jordan normal form J as many times as its algebraic multiplicity. (In particular, for a diagonalizable matrix A each n_i is equal to 1.) Furthermore, J is uniquely determined up to rearrangement of the Jordan cells J_i, so the list $(\lambda_1, n_1), \dots, (\lambda_k, n_k)$ unambiguously represents an equivalence class of similar matrices. It should be emphasized that similarity of matrices is taken here over the field of complex numbers, even if all elements of the original matrix A are real. We will not prove the theorem (elaborate proof).

55. Jordan cells larger than 1×1 are what "prevents diagonalization". In some respects they are "rare", e.g. for a randomly generated matrix A they occur with small probability, but there are many natural examples. For example: V the vector space of polynomials of degree at most 3, $D\colon V \to V$ the differentiation map. The matrix is similar to the 4×4 Jordan cell with eigenvalue 0 on the diagonal.

56. Definition: Let V be a real vector space with an inner product. A linear map $f\colon V \to V$ is called **orthogonal** if it preserves the inner product, i.e. if $\langle f(\mathbf{u}) \,|\, f(\mathbf{v})\rangle = \langle \mathbf{u} \,|\, \mathbf{v}\rangle$ for all $\mathbf{u}, \mathbf{v} \in V$.

57. Proposition (orthogonal map and orthogonal matrix): A linear map $f\colon V \to V$, where V is a finite-dimensional real vector space with inner product, is orthogonal if and only if its matrix A with respect to some orthonormal basis is orthogonal (i.e. $AA^T = I_n$). In the proof is used a *lemma:* If A and B are $n\times n$ matrices such that $\mathbf{x}^T A\mathbf{y} = \mathbf{x}^T B\mathbf{y}$ for any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, then $A = B$.

58. Note: There are analogous terms and results for the field of complex numbers: we are talking about *unitary* maps and matrices.

59. A mechanics note: An orthogonal map evidently also maintains lengths, $\|f(\mathbf{v})\| = \|\mathbf{v}\|$, and for the case of the space $\mathbb{R}^n$ with standard inner product f is thus an **isometry** fixing the origin. It can even be shown, and it is not too hard, that any isometry $f\colon \mathbb{R}^n \to \mathbb{R}^n$ (i.e. a map satisfying $\|f(\mathbf{u}) - f(\mathbf{v})\| = \|\mathbf{u} - \mathbf{v}\|$ for every $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$), for which $f(\mathbf{0}) = \mathbf{0}$, must be linear, and thus an orthogonal map. Hence the motion of rigid bodies, for example, in $\mathbb{R}^3$ is described using orthogonal matrices.

60. Now with the help of orthogonal matrices we give the proof promised earlier that symmetric matrices are diagonalizable – in fact, a little more. Theorem: The eigenvalues of a symmetric real $n \times n$ matrix A are all real, and there exists a (real) orthogonal matrix T such that TAT^{-1} is diagonal.

61. The main steps of the proof:

    - Every eigenvalue is real: Count in two ways $\overline{\mathbf{v}}^T A\mathbf{v}$, where $\mathbf{v}$ is some (possibly complex) eigenvector.
    - The rest is by induction on n: For the induction step take some unit eigenvector $\mathbf{v}$ as the first column and extend to an orthogonal matrix S, and consider how $SAS^{-1} = SAS^T$ looks.

4 Positive definite matrices

62. A *symmetric* real $n \times n$ matrix A is called

    - **positive definite**, if $\mathbf{x}^T A\mathbf{x} > 0$ for all nonzero $\mathbf{x} \in \mathbb{R}^n$, and
    - **positive semidefinite**, if $\mathbf{x}^T A\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$.

    Positive definite matrices are analogous in many ways to positive reals (probably the best analogy that can be defined for matrices).

63. Proposition: The following are equivalent for a square real symmetric matrix A:

    (i) A is positive semidefinite.

    (ii) All the eigenvalues of A are nonnegative.

    (iii) There exists a matrix U such that $U^T U = A$.

Likewise, for positive definiteness the eigenvalues are strictly positive, and the matrix U has the same rank as A.

64. Note: The equivalence of (i) with (iii) intuitively says that a matrix is positive semidefinite just when it has a "square root". The matrix U in (iii) can even be taken to be upper triangular, in which case we have the *Cholesky decomposition* of the matrix A (this term is mostly used for positive definite matrices).

65. Another equivalent condition for positive semidefiniteness of an $n \times n$ matrix A:

    (iv) For $k = 1, 2, \ldots, n$ it holds that $\det(A_k) \geq 0$, where A_k is the matrix formed from A by deleting the last $n - k$ rows and $n - k$ columns.

66. Positive definiteness appears in analysis in the criterion for a local extreme value of a function of several variables.

67. Relation to spaces with an inner product: If A is a positive definite $n \times n$ matrix, then the expression $\langle \mathbf{x} \,|\, \mathbf{y} \rangle = \mathbf{x}^T A \mathbf{y}$ defines an inner product on $\mathbb{R}^n$ (in fact all possible inner products on $\mathbb{R}^n$ take this form).

68. Important method in optimization and other algorithms: **semidefinite programming** = searching for the maxima of a linear function over the set of all positive semidefinite matrices whose elements satisfy given linear equations and inequalities. An effective algorithm is known.

69. A geometric example (rod construction in Euclidean space): Given a symmetric real $(n{+}1)\times(n{+}1)$ matrix $M = (m_{i,j})$, when do there exist points $\mathbf{x}_0$, $\mathbf{x}_1$, …, $\mathbf{x}_n \in \mathbb{R}^d$ so that $\|\mathbf{x}_i - \mathbf{x}_j\| = m_{ij}$, for all i, j? Answer: Define an auxiliary $n{\times}n$ matrix G, $g_{ij} = \frac{1}{2}(m_{0i}^2 + m_{j0}^2 - m_{ij}^2)$. If such $\mathbf{x}_i$ exist and $\mathbf{x}_0 = \mathbf{0}$, then $g_{ij} = \langle \mathbf{x}_i \,|\, \mathbf{x}_j \rangle$. These exist if and only if $G = U^T U$ for some $d{\times}n$ matrix U. In particular, for $d = n$ these $\mathbf{x}_i$ exist iff G is positive semidefinite.

5 Quadratic forms

70. A **quadratic form** on $\mathbb{R}^n$ is a function $f\colon \mathbb{R}^n \to \mathbb{R}$ of the form

$$f(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=i}^{n} a_{ij} x_i x_j$$

(beware, the second sum is from i, not from 1!) for some numbers $a_{ij} \in \mathbb{R}$. It is therefore a quadratic polynomial, where each single term has degree 2 and sometimes called the **analytical expression of a quadratic form**. A quadratic form can also be written in matrix form $f(\mathbf{x}) = \mathbf{x}^T B\mathbf{x}$, where B is the symmetric **matrix of a quadratic form**, given by

$$b_{ij} = \begin{cases} a_{ii} & \text{for } i = j, \\ a_{ij}/2 & \text{for } i < j, \\ a_{ji}/2 & \text{for } i > j. \end{cases}$$

71. Note: The quadratic form f is **positive definite** if $f(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ (like for a matrix). Similarly, positive semidefinite when nonnegative.

72. More generally, for a vector space V over a field $\mathbb{K}$ we define:

    - A **bilinear form** as any map $b\colon V \times V \to \mathbb{K}$ such that $b(a_1\mathbf{u}_1 + a_2\mathbf{u}_2, \mathbf{v}) = a_1 b(\mathbf{u}_1, \mathbf{v}) + a_2 b(\mathbf{u}_2, \mathbf{v})$ (i.e. b is linear in the first component) and $b(\mathbf{u}, a_1\mathbf{v}_1 + a_2\mathbf{v}_2) = a_1 b(\mathbf{u}, \mathbf{v}_1) + a_2 b(\mathbf{u}, \mathbf{v}_2)$ (b is linear in the second component).
    - A **quadratic form** as any map $f\colon V \to \mathbb{K}$ given by the formula $f(\mathbf{v}) = b(\mathbf{v}, \mathbf{v})$ for some bilinear form b.

    Then, for a given basis $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ of a finite-dimensional vector space V, we define the **matrix $B = (b_{ij})$ of the quadratic form** b by the formula $b_{ij} = \frac{1}{2}(f(\mathbf{v}_i + \mathbf{v}_j) - f(\mathbf{v}_i) - f(\mathbf{v}_j))$.

73. What if the basis changes? Let $(\mathbf{v}_1, \dots, \mathbf{v}_n)$ be the old basis, $(\mathbf{v}'_1, \dots, \mathbf{v}'_n)$ the new one, and T the change of basis matrix (i.e. $\mathbf{v}_j = t_{1j}\mathbf{v}'_1 + t_{2j}\mathbf{v}'_2 + \dots + t_{nj}\mathbf{v}'_n$). Then we get $B' = S^T BS$, where $S = T^{-1}$ is the change of basis matrix in the opposite way, B is the matrix of the bilinear/quadratic form with respect to the old basis and B' its matrix with respect to the new basis. (Note, for a linear map $V \to V$ it was $A' = TAT^{-1}$: here it is different!)

74. By changing the basis, we would like to bring the matrix of quadratic form into a “nice” shape, similarly to how we did so for endomorphisms. It will work out a lot more easily:

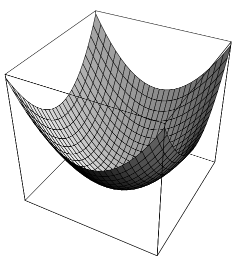
    Theorem **(Sylvester’s law of inertia for quadratic forms)**: For each quadratic form f on a finite-dimensional real vector space there is a basis for which f has a diagonal matrix whose diagonal entries are one, minus one or zero. In addition, the number of ones and the number of minus ones in this diagonal matrix is the same for each such basis (hence “inertia”).

75. More or less the same in matrix language: For each symmetric real matrix B there is a nonsingular matrix S (whose columns in addition are orthogonal to each other) for which the matrix $S^T BS$ is diagonal each of whose entries is $+1$, -1 or 0. The number of entries equal to $+1$ and equal to -1 does not depend on the choice of such S.

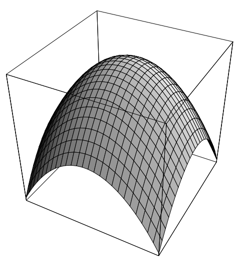
76. The easy part of the proof is the existence of S: From the section on eigenvalues we know that there exists an orthonormal T such that $D = T^T BT$ is diagonal and has the eigenvalues of B on the diagonal (because B is real symmetric). It remains to decompose D as $D = U^T D_0 U$, where U is diagonal with the square roots of the absolute values of the eigenvalues on the diagonal and D_0 is diagonal as in the theorem. Inertia is more elaborate.

77. Note: For positive definite forms, we get only ones on the diagonal (and the form can then be calculated by the standard scalar product); for positive semidefinite forms the diagonal contains only ones and zeros.

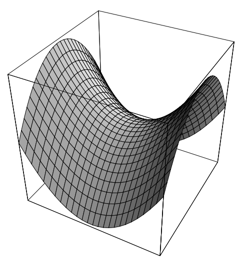
78. For $n = 2$ the theorem says that every quadratic form $f\colon \mathbb{R}^2 \to \mathbb{R}$ can be converted by a linear transformation to just one of the following types (depicted alongside them are their graphs):



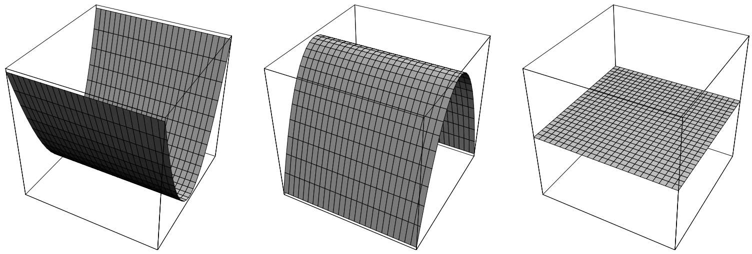
$x_1^2 + x_2^2$



$-x_1^2 - x_2^2$



$x_1^2 - x_2^2$



x_1^2 $-x_1^2$ 0