Chapter 4

Eigenvalues

1 The Spectral Theorem

Hermitian Spaces

Given a $\mathbb{C}$-vector space $\mathcal{V}$, an **Hermitian inner product** in $\mathcal{V}$ is defined as a Hermitian symmetric sesquilinear form such that the corresponding Hermitian quadratic form is positive definite. A space $\mathcal{V}$ equipped with an Hermitian inner product $\langle \cdot, \cdot \rangle$ is called a **Hermitian space.**[1]

The inner square $\langle \mathbf{z}, \mathbf{z} \rangle$ is interpreted as the square of the **length** $|\mathbf{z}|$ of the vector $\mathbf{z}$. Respectively, the **distance** between two points $\mathbf{z}$ and $\mathbf{w}$ in an Hermitian space is defined as $|\mathbf{z}-\mathbf{w}|$. Since the Hermitian inner product is positive, distance is well-defined, symmetric, and positive (unless $\mathbf{z} = \mathbf{w}$). In fact it satisfies the **triangle inequality**[2]:

$$|\mathbf{z} - \mathbf{w}| \leq |\mathbf{z}| + |\mathbf{w}|.$$

This follows from the **Cauchy − Schwarz inequality**:

$$|\langle \mathbf{z}, \mathbf{w} \rangle|^2 \leq \langle \mathbf{z}, \mathbf{z} \rangle \ \langle \mathbf{w}, \mathbf{w} \rangle,$$

where the equality holds if and only if $\mathbf{z}$ and $\mathbf{w}$ are linearly dependent. To derive the triangle inequality, write:

$$\begin{aligned} |\mathbf{z} - \mathbf{w}|^2 &= \langle \mathbf{z} - \mathbf{w}, \mathbf{z} - \mathbf{w} \rangle = \langle \mathbf{z}, \mathbf{z} \rangle - \langle \mathbf{z}, \mathbf{w} \rangle - \langle \mathbf{w}, \mathbf{z} \rangle + \langle \mathbf{w}, \mathbf{w} \rangle \\ &\leq |\mathbf{z}|^2 + 2|\mathbf{z}||\mathbf{w}| + |\mathbf{w}|^2 = (|\mathbf{z}| + |\mathbf{w}|)^2. \end{aligned}$$

[1]Other terms used are **unitary space** and finite dimensional **Hilbert space**.

[2]This makes a Hermitian space a **metric space**.

To prove the Cauchy–Schwarz inequality, note that it suffices to consider the case $|\mathbf{w}| = 1$. Indeed, when $\mathbf{w} = \mathbf{0}$, both sides vanish, and when $\mathbf{w} \neq \mathbf{0}$, both sides scale the same way when $\mathbf{w}$ is normalized to the unit length. So, assuming $|\mathbf{w}| = 1$, we put $\lambda := \langle \mathbf{w}, \mathbf{z} \rangle$ and consider the **projection** $\lambda\mathbf{w}$ of the vector $\mathbf{z}$ to the line spanned by $\mathbf{w}$. The difference $\mathbf{z} - \lambda\mathbf{w}$ is **orthogonal** to $\mathbf{w}$: $\langle \mathbf{w}, \mathbf{z} - \lambda\mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{z} \rangle - \lambda \langle \mathbf{w}, \mathbf{w} \rangle = 0$. From positivity of inner squares, we have:

$$0 \leq \langle \mathbf{z} - \lambda\mathbf{w}, \mathbf{z} - \lambda\mathbf{w} \rangle = \langle \mathbf{z}, \mathbf{z} - \lambda\mathbf{w} \rangle = \langle \mathbf{z}, \mathbf{z} \rangle - \lambda \langle \mathbf{z}, \mathbf{w} \rangle.$$

Since $\langle \mathbf{z}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{z} \rangle} = \bar{\lambda}$, we conclude that $|\mathbf{z}|^2 \geq |\langle \mathbf{z}, \mathbf{w} \rangle|^2$ as required. Notice that the equality holds true only when $\mathbf{z} = \lambda\mathbf{w}$.

All Hermitian spaces of the same dimension are **isometric** (or **Hermitian isomorphic**), i.e. isomorphic through isomorphisms respecting Hermitian inner products. Namely, as it follows from the Inertia Theorem for Hermitian forms, every Hermitian space has an **orthonormal basis**, i.e. a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ such that $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0$ for $i \neq j$ and $= 1$ for $i = j$. In the coordinate system corresponding to an orthonormal basis, the Hermitian inner product takes on the standard form:

$$\langle \mathbf{z}, \mathbf{w} \rangle = \bar{z}_1 w_1 + \cdots + \bar{z}_n w_n.$$

An orthonormal basis is not unique. Moreover, as it follows from the proof of Sylvester's rule, one can start with any basis $\mathbf{f}_1, \dots, \mathbf{f}_n$ in $\mathcal{V}$ and then construct an orthonormal basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ such that $\mathbf{e}_k \in \operatorname{Span}(\mathbf{f}_1, \dots, \mathbf{f}_k)$. This is done inductively; namely, when $\mathbf{e}_1, \dots, \mathbf{e}_{k-1}$ have already been constructed, one subtracts from $\mathbf{f}_k$ its projection to the space $\operatorname{Span}(\mathbf{e}_1, \dots, \mathbf{e}_{k-1})$:

$$\tilde{\mathbf{f}}_k = \mathbf{f}_k - \langle \mathbf{e}_1, \mathbf{f}_k \rangle \mathbf{e}_1 - \cdots - \langle \mathbf{e}_{k-1}, \mathbf{f}_k \rangle \mathbf{e}_{k-1}.$$

The resulting vector $\tilde{\mathbf{f}}_k$ lies in $\operatorname{Span}(\mathbf{f}_1, \dots, \mathbf{f}_{k-1}, \mathbf{f}_k)$ and is orthogonal to $\operatorname{Span}(\mathbf{f}_1, \dots, \mathbf{f}_{k-1}) = \operatorname{Span}(\mathbf{e}_1, \dots, \mathbf{e}_{k-1})$. Indeed,

$$\langle \mathbf{e}_i, \tilde{\mathbf{f}}_k \rangle = \langle \mathbf{e}_i, \mathbf{f}_k \rangle - \sum_{j=1}^{k-1} \langle \mathbf{e}_j, \mathbf{f}_k \rangle \langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0$$

for all $i = 1, \dots, k-1$. To construct $\mathbf{e}_k$, one normalizes $\tilde{\mathbf{f}}_k$ to the unit length:

$$\mathbf{e}_k := \tilde{\mathbf{f}}_k / |\tilde{\mathbf{f}}_k|.$$

The above algorithm of replacing a given basis with an orthonormal one is known as **Gram–Schmidt orthogonalization**.

EXERCISES

316. Prove that if two vectors $\mathbf{u}$ and $\mathbf{v}$ in an Hermitian space are orthogonal, then $|\mathbf{u}|^2 + |\mathbf{v}|^2 = |\mathbf{u} - \mathbf{v}|^2$. Is the converse true? ✓

317. Prove that for any vectors $\mathbf{u}$, $\mathbf{v}$ in an Hermitian space,

$$|\mathbf{u}+\mathbf{v}|^2 + |\mathbf{u}-\mathbf{v}|^2 = 2|\mathbf{u}|^2 + 2|\mathbf{v}|^2.$$

Find a geometric interpretation of this fact. ✓

318. Apply Gram–Schmidt orthogonalization to the basis $\mathbf{f}_1 = \mathbf{e}_1 + 2i\mathbf{e}_2 + 2i\mathbf{e}_3$, $\mathbf{f}_2 = \mathbf{e}_1 + 2i\mathbf{e}_2$, $\mathbf{f}_3 = \mathbf{e}_1$ in the coordinate Hermitian space $\mathbb{C}^3$.

319. Apply Gram–Schmidt orthogonalization to the standard basis $\mathbf{e}_1, \mathbf{e}_2$ of $\mathbb{C}^2$ to construct an orthonormal basis of the Hermitian inner product $\langle \mathbf{z}, \mathbf{w} \rangle = \bar{z}_1 w_1 + 2\bar{z}_1 w_2 + 2\bar{z}_2 w_1 + 5\bar{z}_2 w_2$.

320. Let $\mathbf{f} \in \mathcal{V}$ be a vector in an Hermitian space, $\mathbf{e}_1, \dots, \mathbf{e}_k$ an orthonormal basis in a subspace $\mathcal{W}$. Prove that $\mathbf{u} = \sum \langle \mathbf{e}_i, \mathbf{v} \rangle \mathbf{e}_i$ is the point of $\mathcal{W}$ closest to $\mathbf{v}$, and that $\mathbf{v} - \mathbf{u}$ is orthogonal to $\mathcal{W}$. (The point $\mathbf{u} \in \mathcal{W}$ is called the **orthogonal projection** of $\mathbf{v}$ to $\mathcal{W}$.)

321.* Let $\mathbf{f}_1, \dots, \mathbf{f}_N$ be a finite sequence of vectors in an Hermitian space. The Hermitian $N \times N$-matrix $\langle \mathbf{f}_i, \mathbf{f}_j \rangle$ is called the **Gram matrix** of the sequence. Show that two finite sequences of vectors are isometric, i.e. obtained from each other by a unitary transformation, if and only if their Gram matrices are the same.

Normal Operators

Our next point is that ***an Hermitian inner product on a complex vector space allows one to identify sesquilinear forms on it with linear transformations.***

Let $\mathcal{V}$ be an Hermitian vector space, and $T : \mathcal{V} \mapsto \mathcal{V}$ a $\mathbb{C}$-linear transformation. Then the function $\mathcal{V} \times \mathcal{V} \to \mathbb{C}$

$$T(\mathbf{w}, \mathbf{z}) := \langle \mathbf{w}, T\mathbf{z} \rangle$$

is $\mathbb{C}$-linear in $\mathbf{z}$ and anti-linear in $\mathbf{w}$, i.e. it is sesquilinear.

In coordinates, $\mathbf{w} = \sum_i w_i \mathbf{e}_i$, $\mathbf{z} = \sum_j z_j \mathbf{e}_j$, and

$$\langle \mathbf{w}, T\mathbf{z} \rangle = \sum_{i,j} \overline{w_i}\, \langle \mathbf{e}_i, T\mathbf{e}_j \rangle z_j,$$

i.e., $T(\mathbf{e}_i, \mathbf{e}_j) = \langle \mathbf{e}_i, T\mathbf{e}_j \rangle$ form the coefficient matrix of the sesquilinear form. On the other hand, $T\mathbf{e}_j = \sum_i t_{ij} \mathbf{e}_i$, where $[t_{ij}]$ is the matrix of the linear transformation T with respect to the basis $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$. Note that ***if the basis is orthonormal,*** then $\langle \mathbf{e}_i, T\mathbf{e}_j \rangle = t_{ij}$, i.e. ***the two matrices coincide.***

Since t_{ij} could be arbitrary, it follows that every sesquilinear form on $\mathcal{V}$ is uniquely represented by a linear transformation.

Earlier we have associated with a sesquilinear form its Hermitian adjoint (by changing the order of the arguments and conjugating the value). When the sesquilinear form is obtained from a linear transformation T, the adjoint corresponds to another linear transformation denoted $T^\dagger$ and called **Hermitian adjoint** to T. Thus, by definition,

$$\langle \mathbf{w}, T\mathbf{z} \rangle = \overline{\langle \mathbf{z}, T^\dagger \mathbf{w} \rangle} = \langle T^\dagger \mathbf{w}, \mathbf{z} \rangle \quad \text{for all } \mathbf{z}, \mathbf{w} \in \mathcal{V}.$$

Of course, we also have $\langle T\mathbf{w}, \mathbf{z} \rangle = \langle \mathbf{w}, T^\dagger \mathbf{z} \rangle$ (check this!), and either identity completely characterizes $T^\dagger$ in terms of T.

The matrix of $T^\dagger$ in an orthonormal basis is obtained from that of T by complex conjugation and transposition:

$$t^\dagger_{ij} := \langle \mathbf{e}_i, T^\dagger \mathbf{e}_j \rangle = \langle T\mathbf{e}_i, \mathbf{e}_j \rangle = \overline{\langle \mathbf{e}_j, T\mathbf{e}_i \rangle} =: \overline{t_{ji}}.$$

`Definition.` A linear transformation on an Hermitian vector is called **normal** (or a **normal operator**) if it commutes with its adjoint: $T^\dagger T = TT^\dagger$.

`Example 1.` A scalar operator is normal. Indeed, $(\lambda I)^\dagger = \overline{\lambda}\, I$, which is also scalar, and scalars commute.

`Example 2.` A linear transformation on an Hermitian space is called **Hermitian** if it coincides with its Hermitian adjoint: $S^\dagger = S$. A Hermitian operator[3] is normal.

`Example 3.` A linear transformation is called anti-Hermitian if it is opposite to its adjoint: $A^\dagger = -A$. Multiplying an Hermitian operator by $\sqrt{-1}$ yields an anti-Hermitian one, and *vice versa* (because $(\sqrt{-1}I)^\dagger = -\sqrt{-1}I$). Anti-Hermitian operators are normal.

`Example 4.` Every linear transformation $T : \mathcal{V} \to \mathcal{V}$ can be uniquely written as the sum of Hermitian and anti-Hermitian operators: $T = S + Q$, where $S = (T + T^\dagger)/2 = S^\dagger$, and $Q = (T - T^\dagger)/2 = -Q^\dagger$. We claim that *an operator is normal whenever its Hermitian and anti-Hermitian parts commute.* Indeed, $T^\dagger = S - Q$, and

$$TT^\dagger - T^\dagger T = (S + Q)(S - Q) - (S - Q)(S + Q) = 2(QS - SQ).$$

[3]The term **operator** in Hermitian geometry is synonimous to *linear map.*

`Example 5.` An invertible linear transformation $U : \mathcal{V} \to \mathcal{V}$ is called **unitary** if it preserves inner products:

$$\langle U\mathbf{w}, U\mathbf{z}\rangle = \langle \mathbf{w}, \mathbf{z}\rangle \text{ for all } \mathbf{w}, \mathbf{z} \in \mathcal{V}.$$

Equivalently, $\langle \mathbf{w}, (U^\dagger U - I)\mathbf{z}\rangle = 0$ for all $\mathbf{w}, \mathbf{z} \in \mathcal{V}$. Taking $\mathbf{w} = (U^\dagger U - I)\mathbf{z}$, we conclude that $(U^\dagger U - I)\mathbf{z} = \mathbf{0}$ for all $\mathbf{z} \in \mathcal{V}$, and hence $U^\dagger U = I$. Thus, for a unitary map U, $U^{-1} = U^\dagger$. The converse statement is also true (and easy to check by starting from $U^{-1} = U^\dagger$ and reversing our computation). Since every invertible transformation commutes with its own inverse, we conclude that *unitary transformations are normal.*

EXERCISES

322. Generalize the construction of Hermiatian adjoint operators to the case of operators $A : \mathcal{V} \to \mathcal{W}$ between two different Hermitian spaces. Namely, show that $A^\dagger : \mathcal{W} \to \mathcal{V}$ is uniquely determined by the identity $\langle A^\dagger \mathbf{w}, \mathbf{v}\rangle_{\mathcal{V}} = \langle \mathbf{w}, A\mathbf{v}\rangle_{\mathcal{W}}$ for all $\mathbf{v} \in \mathcal{V}$ and $\mathbf{w} \in \mathcal{W}$.

323. Show that the matrices of $A : \mathcal{V} \to \mathcal{W}$ and $A^\dagger : \mathcal{W} \to \mathcal{V}$ in orthonormal bases of $\mathcal{V}$ and $\mathcal{W}$ are obtained from each other by transposition and complex conjugation.

324. The **trace** of a square matrix A is defined as the sum of its diagonal entries, and is denoted $\operatorname{tr} A$. Prove that $\langle A, B\rangle := \operatorname{tr}(A^\dagger B)$ defines an Hermitian inner product on the space $\operatorname{Hom}(\mathbb{C}^n, \mathbb{C}^m)$ of $m \times n$-matrices.

325. Let $A_1, \dots, A_k : \mathcal{V} \to \mathcal{W}$ be linear maps between Hermitian spaces. Prove that if $\sum A_i^\dagger A_i = 0$, then $A_1 = \cdots = A_k = 0$.

326. Let $A : \mathcal{V} \to \mathcal{W}$ be a linear map between Hermitian spaces. Show that $B := A^\dagger A$ and $C = AA^\dagger$ are Hermitian, and that the corresponding Hermitian forms $B(\mathbf{x}, \mathbf{x}) := \langle \mathbf{x}, B\mathbf{x}\rangle$ in $\mathcal{V}$ and $C(\mathbf{y}, \mathbf{y}) := \langle \mathbf{y}, C\mathbf{y}\rangle$ in $\mathcal{W}$ are non-negative. Under what hypothesis about A is the 1st of them positive? the 2nd one? both?

327. Let $\mathcal{W} \subset \mathcal{V}$ be a subspace in an Hermitian space, and let $P : \mathcal{V} \to \mathcal{V}$ be the map that to each vector $\mathbf{v} \in \mathcal{V}$ assigns its orthogonal projection to $\mathcal{W}$. Prove that P is an Hermitian operator, that $P^2 = P$, and that $\operatorname{Ker} P = \mathcal{W}^\perp$.(It is called the **orthogonal projector** to $\mathcal{W}$.)

328. Prove that an $n \times n$-matrix is unitary if and only if its rows (or columns) form an orthonormal basis in the coordinate Hermitian space $\mathbb{C}^n$.

329. Prove that the determinant of a unitary matrix is a complex number of absolute value 1.

330. Prove that the **Cayley transform**: $C \mapsto (I-C)/(I+C)$, well-defined for linear transformations C such that $I+C$ is invertible, transforms unitary operators into anti-Hermitian and *vice versa.* Compute the square of the Cayley transform. ✓

331. Prove that the commutator $AB - BA$ of anti-Hermitian operators A and B is anti-Hermitian.

332. Give an example of a normal 2×2-matrix which is not Hermitian, anti-Hermitian, unitary, or diagonal.

333. Prove that for any $n \times n$-matrix A and any complex numbers α, β of absolute value 1, the matrix $\alpha A + \beta A^\dagger$ is normal.

334. Prove that $A : \mathcal{V} \to \mathcal{V}$ is normal if and only if $|A\mathbf{x}| = |A^\dagger \mathbf{x}|$ for all $\mathbf{x} \in \mathcal{V}$.

The Spectral Theorem for Normal Operators

Let $A : \mathcal{V} \to \mathcal{V}$ be a linear transformation, $\mathbf{v} \in \mathcal{V}$ a vector, and $\lambda \in \mathbb{C}$ a scalar. The vector $\mathbf{v}$ is called an **eigenvector** of A with the **eigenvalue** λ, if $\mathbf{v} \neq \mathbf{0}$, and $A\mathbf{v} = \lambda\mathbf{v}$. In other words, A preserves the line spanned by the vector $\mathbf{v}$ and acts on this line as the multiplication by λ.

`Theorem.` ***A linear transformation $A : \mathcal{V} \to \mathcal{V}$ on a finite dimensional Hermitian vector space is normal if and only if $\mathcal{V}$ has an orthonormal basis of eigenvectors of A.***

`Proof.` In one direction, the statement is almost obvious: If a basis consists of eigenvectors of A, then the matrix of A in this basis is diagonal. When the basis is orthonormal, the matrix of the Hermitian adjoint operator $A^\dagger$ in this basis is Hermitian adjoint to the matrix of A and is also diagonal. Since all diagonal matrices commute, we conclude that A is normal. Thus, it remains to prove that, conversely, every normal operator has an orthonormal basis of eigenvectors. We will prove this in four steps.

`Step 1.` *Existence of eigenvalues.* We need to show that there exists a scalar $\lambda \in \mathbb{C}$ such that the system of linear equations $A\mathbf{x} = \lambda\mathbf{x}$ has a non-trivial solution. Equivalently, this means that the linear transformation $\lambda I - A$ has a non-trivial kernel. Since $\mathcal{V}$ is finite dimensional, this can be re-stated in terms of the determinant of the matrix of A (in any basis) as

$$\det(\lambda I - A) = 0.$$

This relation, understood as an equation for λ, is called the **characteristic equation** of the operator A. When $A = 0$, it becomes

$\lambda^n = 0$, where $n = \dim \mathcal{V}$. In general, it is a degree-n polynomial equation

$$\lambda^n + p_1\lambda^{n-1} + \cdots + p_{n-1}\lambda + p_n = 0,$$

where the coefficients $p_1, \ldots, p_n$ are certain algebraic expressions of matrix entries of A (and hence are complex numbers). According to the Fundamental Theorem of Algebra, this equation has a complex solution, say λ_0. Then $\det(\lambda_0 I - A) = 0$, and hence the system $(\lambda_0 I - A)\mathbf{x} = \mathbf{0}$ has a non-trivial solution, $\mathbf{v} \neq \mathbf{0}$, which is therefore an eigenvector of A with the eigenvalue λ_0.

`Remark.` Solutions to the system $A\mathbf{x} = \lambda_0\mathbf{x}$ form a linear subspace $\mathcal{W}$ in $\mathcal{V}$, namely the kernel of $\lambda_0 I - A$, and eigenvectors of A with the eigenvalue λ_0 are exactly all non-zero vectors in $\mathcal{W}$. Slightly abusing terminology, $\mathcal{W}$ is called the **eigenspace** of A corresponding to the eigenvalue λ_0. Obviously, $A(\mathcal{W}) \subset \mathcal{W}$. Subspaces with such property are called **A-invariant**. Thus eigenspaces of a linear transformation A are A-invariant.

`Step 2.` *$A^\dagger$-invariance of eigenspaces of A.* Let $\mathcal{W} \neq \{\mathbf{0}\}$ be the eigenspace of a normal operator A, corresponding to an eigenvalue λ. Then for every $\mathbf{w} \in \mathcal{W}$,

$$A(A^\dagger\mathbf{w}) = A^\dagger(A\mathbf{w}) = A^\dagger(\lambda\mathbf{w}) = \lambda(A^\dagger\mathbf{w}).$$

Therefore $A^\dagger\mathbf{w} \in \mathcal{W}$, i.e. the eigenspace $\mathcal{W}$ is $A^\dagger$-invariant.

`Step 3.` *Invariance of orthogonal complements.* Let $\mathcal{W} \subset \mathcal{V}$ be a linear subspace. Denote by $\mathcal{W}^\perp$ the **orthogonal complement** of the subspace $\mathcal{W}$ with respect to the Hermitian inner product:

$$\mathcal{W}^\perp := \{\mathbf{v} \in \mathcal{V} \mid \langle \mathbf{w}, \mathbf{v} \rangle = 0 \ \text{ for all } \mathbf{w} \in \mathcal{W}.\}$$

Note that if $\mathbf{e}_1, \ldots, \mathbf{e}_k$ is a basis in $\mathcal{W}$, then $\mathcal{W}^\perp$ is given by k linear equations $\langle \mathbf{e}_i, \mathbf{v} \rangle = 0$, $i = 1, \ldots, k$, and thus has dimension $\geq n - k$. On the other hand, $\mathcal{W} \cap \mathcal{W}^\perp = \{\mathbf{0}\}$, because no vector $\mathbf{w} \neq \mathbf{0}$ can be orthogonal to itself: $\langle \mathbf{w}, \mathbf{w} \rangle > 0$. It follows from dimension counting formulas that $\dim \mathcal{W}^\perp = n - k$. Moreover, this implies that $\mathcal{V} = \mathcal{W} \oplus \mathcal{W}^\perp$, i.e. the whole space is represented as the direct sum of two orthogonal subspaces.

We claim that *if a subspace is both A- and $A^\dagger$-invariant, then its orthogonal complement is also A- and $A^\dagger$-invariant.* Indeed, suppose that $A^\dagger(\mathcal{W}) \subset \mathcal{W}$, and $\mathbf{v} \in \mathcal{W}^\perp$. Then for any $\mathbf{w} \in \mathcal{W}$, we have: $\langle \mathbf{w}, A\mathbf{v} \rangle = \langle A^\dagger\mathbf{w}, \mathbf{v} \rangle = 0$, since $A^\dagger\mathbf{w} \in \mathcal{W}$. Therefore $A\mathbf{v} \in \mathcal{W}^\perp$, i.e. $\mathcal{W}^\perp$ is A-invariant. By the same token, if $\mathcal{W}$ is A-invariant, then $\mathcal{W}^\perp$ is $A^\dagger$-invariant.

Step 4. *Induction on* $\dim \mathcal{V}$. When $\dim \mathcal{V} = 1$, the theorem is obvious. Assume that the theorem is proved for normal operators in spaces of dimension $< n$, and prove it when $\dim \mathcal{V} = n$.

According to Step 1, a normal operator A has an eigenvalue λ. Let $\mathcal{W} \neq \{\mathbf{0}\}$ be the corresponding eigenspace. If $\mathcal{W} = \mathcal{V}$, then the operator is scalar, $A = \lambda I$, and *any* orthonormal basis in $\mathcal{V}$ will consist of eigenvectors of A. If $\mathcal{W} \neq \mathcal{V}$, then (by Steps 2 and 3) both $\mathcal{W}$ and $\mathcal{W}^{\perp}$ are A- and $A^{\dagger}$-invariant and have dimensions $< n$. The *restrictions* of the operators A and $A^{\dagger}$ to each of these subspaces still satisfy $AA^{\dagger} = A^{\dagger}A$ and $\langle A^{\dagger}\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}, A\mathbf{y}\rangle$ for all $\mathbf{x}, \mathbf{y}$. Therefore these restrictions remain adjoint to each other normal operators on $\mathcal{W}$ and $\mathcal{W}^{\perp}$. Applying the induction hypothesis, we can find orthonormal bases of eigenvectors of A in each $\mathcal{W}$ and $\mathcal{W}^{\perp}$. The union of these bases form an orthonormal basis of eigenvectors of A in $\mathcal{V} = \mathcal{W} \oplus \mathcal{W}^{\perp}$. □

Remark. Note that Step 1 is based on the Fundamental Theorem of Algebra, but does not use normality of A and applies to any $\mathbb{C}$-linear transformation. Thus, ***every linear transformation on a complex vector space has eigenvalues and eigenvectors.*** Furthermore, Step 2 actually applies to any commuting transformations and shows that ***if*** $AB = BA$ ***then eigenspaces of*** A ***are*** B***-invariant.*** The fact that $B = A^{\dagger}$ is used in Step 3.

Corollary 1. ***A normal operator has a diagonal matrix in a suitable orthonormal basis.***

Corollary 2. ***Let*** $A : \mathcal{V} \to \mathcal{V}$ ***be a normal operator,*** λ_i ***distinct roots of its characteristic polynomial,*** m_i ***their multiplicities, and*** $\mathcal{W}_i$ ***corresponding eigenspaces. Then*** $\dim \mathcal{W}_i = m_i$***, and*** $\sum \dim \mathcal{W}_i = \dim \mathcal{V}$***.***

Indeed, this is true for transformations defined by any diagonal matrices. For normal operators, in addition $\mathcal{W}_i \perp \mathcal{W}_j$ when $i \neq j$. In particular we have the following corollary.

Corollary 3. ***Eigenvectors of a normal operator corresponding to different eigenvalues are orthogonal.***

Here is a matrix version of the Spectral Theorem.

Corollary 4. ***A square complex matrix*** A ***commuting with its Hermitian adjoint*** $A^{\dagger}$ ***can be transformed to a diagonal form by transformations*** $A \mapsto U^{\dagger}AU$ ***defined by unitary matrices*** U***.***

Note that for unitary matrices, $U^{\dagger} = U^{-1}$, and therefore the above transformations coincide with similarity transformations $A \mapsto$

$U^{-1}AU$. This is how the matrix A of a linear transformation changes under a change of basis. When both the old and new bases are orthonormal, the transition matrix U must be unitary (because in both old and new coordinates the Hermitian inner product has the same standard form: $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\dagger \mathbf{y}$). The result follows.

EXERCISES

335. Prove that the characteristic polynomial $\det(\lambda I - A)$ of a square matrix A does not change under similarity transformations $A \mapsto C^{-1}AC$ and thus depends only on the linear operator defined by the matrix.

336. Show that if $\lambda^n + p_1\lambda^{n-1} + \cdots + p_n$ is the characteristic polynomial of a matrix A, then $p_n = (-1)^n \det A$, and $p_1 = -\operatorname{tr} A$, and conclude that the trance is invariant under similarity transformations.

337. Prove that $\operatorname{tr} A = -\sum \lambda_i$, where λ_i are the roots of $\det(\lambda I - A) = 0$.↯

338. Prove that if A and B are normal and $AB = 0$, then $BA = 0$. Does this remain true without the normality assumption?

339.* Let operator A be normal. Prove that the set of complex numbers $\{\langle \mathbf{x}, A\mathbf{x} \rangle \mid |\mathbf{x}| = 1\}$ is a convex polygon whose vertices are the eigenvalues of A. ↯

340. Prove that two (or several) commuting normal operators have a common orthonormal basis of eigenvectors. ↯

341. Prove that if A is normal and $AB = BA$, then $AB^\dagger = B^\dagger A$, $A^\dagger B = BA^\dagger$, and $A^\dagger B^\dagger = B^\dagger A^\dagger$.

Unitary Transformations

Note that if λ is an eigenvalue of a unitary operator U then $|\lambda| = 1$. Indeed, if $\mathbf{x} \neq \mathbf{0}$ is a corresponding eigenvector, then $\langle \mathbf{x}, \mathbf{x} \rangle = \langle U\mathbf{x}, U\mathbf{x} \rangle = \lambda\bar{\lambda}\langle \mathbf{x}, \mathbf{x} \rangle$, and since $\langle \mathbf{x}, \mathbf{x} \rangle \neq 0$, it implies $\lambda\bar{\lambda} = 1$.

Corollary 5. ***A transformation is unitary if and only if in some orthonormal basis its matrix is diagonal, and the diagonal entries are complex numbers of absolute value*** 1***.***

On the complex line $\mathbb{C}$, multiplication by λ with $|\lambda| = 1$ and **arg** $\lambda = \theta$ defines the rotation through the angle θ. We will call this transformation on the complex line a **unitary rotation**. We arrive therefore to the following geometric characterization of unitary transformations.

Corollary 6. ***Unitary transformations in an Hermitian space of dimension*** n ***are exactly unitary rotations (through possibly different angles) in*** n ***mutually perpendicular complex directions.***

Orthogonal Diagonalization

Corollary 7. ***A linear operator is Hermitian (respectively anti-Hermitian) if and only if in some orthonormal basis its matrix is diagonal with all real (respectively imaginary) diagonal entries.***

Indeed, if $A\mathbf{x} = \lambda\mathbf{x}$ and $A^\dagger = \pm A$, we have:

$$\lambda\langle \mathbf{x}, \mathbf{x}\rangle = \langle \mathbf{x}, A\mathbf{x}\rangle = \langle A^\dagger \mathbf{x}, \mathbf{x}\rangle = \pm\bar{\lambda}\langle \mathbf{x}, \mathbf{x}\rangle.$$

Therefore $\lambda = \pm\bar{\lambda}$ provided that $\mathbf{x} \neq \mathbf{0}$, i.e. eigenvalues of an Hermitian operator are real and of anti-Hermitian imaginary. *Vice versa*, a real diagonal matrix is obviously Hermitian, and imaginary anti-Hermitian.

Recall that (anti-)Hermitian operators correspond to (anti-) Hermitian forms $A(\mathbf{x}, \mathbf{y}) := \langle \mathbf{x}, A\mathbf{y}\rangle$. Applying the Spectral Theorem and reordering the basis eigenvectors in the monotonic order of the corresponding eigenvalues, we obtain the following classification results for forms.

Corollary 8. ***In a Hermitian space of dimension n, an Hermitian form can be transformed by unitary changes of coordinates to exactly one of the normal forms***

$$\lambda_1|z_1|^2 + \cdots + \lambda_n|z_n|^2, \quad \lambda_1 \geq \cdots \geq \lambda_n.$$

Corollary 9. ***In a Hermitian space of dimension n, an anti-Hermitian form can be transformed by unitary changes of coordinates to exactly one of the normal forms***

$$i\omega_1|z_1|^2 + \cdots + i\omega_n|z_n|^2, \quad \omega_1 \geq \cdots \geq \omega_n.$$

Uniqueness follows from the fact that eigenvalues and dimensions of eigenspaces are determined by the operators in a coordinate-less fashion.

Corollary 10. ***In a complex vector space of dimension n, a pair of Hermitian forms, of which the first one is positive definite, can be transformed by a choice of a coordinate system to exactly one of the normal forms:***

$$|z_1|^2 + \cdots + |z_n|^2, \quad \lambda_1|z_1|^2 + \cdots + \lambda_n|z_n|^2, \quad \lambda_1 \geq \cdots \geq \lambda_n.$$

This is the **Orthogonal Diagonalization Theorem** for Hermitian forms. It is proved in two stages. First, applying the Inertia Theorem to the positive definite form one transforms it to the standard form; the 2nd Hermitian form changes accordingly but remains arbitrary at this stage. Then, applying Corollary 8 of the Spectral Theorem, one transforms the 2nd Hermitian form to its normal form by transformations preserving the 1st one.

Note that one can take the positive definite sesquilinear form corresponding to the 1st Hermitian form for the Hermitian inner product, and describe the 2nd form as $\langle \mathbf{z}, A\mathbf{z} \rangle$, where A is an operator Hermitian with respect to this inner product. The operator, its eigenvalues, and their multiplicities are thus defined by the given pair of forms in a coordinate-less fashion. This guarantees that pairs with different collections $\lambda_1 \geq \cdots \geq \lambda_n$ of eigenvalues are non-equivalent to each other.

EXERCISES

342. Prove that all roots of characteristic polynomials of Hermitian matrices are real.

343. Find eigenspaces and eigenvalues of an orthogonal projector to a subspace $\mathcal{W} \subset \mathcal{V}$ in an Hermitian space.

344. Prove that every Hermitian operator P satisfying $P^2 = P$ is an orthogonal projector. Does this remain true if P is not Hermitian?

345. Prove directly, i.e. not referring to the Spectral Theorem, that every Hermitian operator has an orthonormal basis of eigenvectors. ↯

346. Prove that if $(\lambda - \lambda_1) \cdots (\lambda - \lambda_n)$ is the characteristic polynomial of a normal operator A, then $\sum |\lambda_i|^2 = \mathrm{tr}(A^\dagger A)$.

347. Classify up to linear changes of coordinates pairs (Q, A) of forms, where S is positive definite Hermitian, and A anti-Hermitian.

348. An Hermitian operator S is called **positive** (written: $S \geq 0$) if $\langle \mathbf{x}, S\mathbf{x} \rangle \geq 0$ for all $\mathbf{x}$. Prove that for every positive operator S there is a unique positive **square root** (denoted by $\sqrt{S}$), i.e. a positive operator whose square is S.

349.** Prove the **Singular Value Decomposition Theorem**: For a rank r linear map $A : \mathcal{V} \to \mathcal{W}$ between Hermitian spaces, there exist orthonormal bases $\mathbf{v}_1, \dots, \mathbf{v}_n$ in $\mathcal{V}$ and $\mathbf{w}_1, \dots, \mathbf{w}_m$ in $\mathcal{W}$, and reals $\mu_1 \geq \cdots \geq \mu_r > 0$, such that $A\mathbf{v}_1 = \mu_1 \mathbf{w}_1, \dots, A\mathbf{v}_r = \mu_r \mathbf{w}_r, \quad A\mathbf{v}_{r+1} = \cdots = A\mathbf{v}_n = \mathbf{0}$. ↯

350. Prove that for every complex $m \times n$-matrix A of rank r, there exist unitary $m \times m$- and $n \times n$-matrices U and V, and a diagonal $r \times r$-matrix M with positive diagonal entries, such that $A = U^\dagger \begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix} V$. ↯

351. Using the Singular Value Decomposition Theorem with $m = n$, prove that every linear transformation A of an Hermitian space has a **polar decomposition** $A = SU$, where S is positive, and U is unitary.

352. Prove that the polar decomposition $A = SU$ is unique when A is invertible; namely $S = \sqrt{AA^*}$, and $U = S^{-1}A$. What are polar decompositions of non-zero 1×1-matrices?

2 Euclidean Geometry

Euclidean Spaces

Let $\mathcal{V}$ be a real vector space. A **Euclidean inner product** (or **Euclidean structure**) on $\mathcal{V}$ is defined as a positive definite symmetric bilinear form $\langle \cdot, \cdot \rangle$. A real vector space equipped with a Euclidean inner product is called a **Euclidean space**. A Euclidean inner product allows one to talk about distances between points and angles between directions:

$$|\mathbf{x} - \mathbf{y}| = \sqrt{\langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle}, \quad \cos\theta(\mathbf{x}, \mathbf{y}) := \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{|\mathbf{x}|\ |\mathbf{y}|}.$$

It follows from the Inertia Theorem that ***every finite dimensional Euclidean vector space has an orthonormal basis.*** In coordinates corresponding to an orthonormal basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ the inner product is given by the standard formula:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i,j=1}^{n} x_i y_j \langle \mathbf{e}_i, \mathbf{e}_j \rangle = x_1 y_1 + \cdots + x_n y_n.$$

Thus, every Euclidean space $\mathcal{V}$ of dimension n can be identified with the **coordinate Euclidean space** $\mathbb{R}^n$ by an isomorphism $\mathbb{R}^n \to \mathcal{V}$ respecting inner products. Such an isomorphism is not unique, but can be composed with any invertible linear transformation $U : \mathcal{V} \to \mathcal{V}$ preserving the Euclidean structure:

$$\langle U\mathbf{x}, U\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathcal{V}.$$

Such transformations are called **orthogonal**.

A Euclidean structure on a vector space $\mathcal{V}$ allows one to identify the space with its dual $\mathcal{V}^*$ by the rule that to a vector $\mathbf{v} \in \mathcal{V}$ assigns the linear function on $\mathcal{V}$ whose value at a point $\mathbf{x} \in \mathcal{V}$ is equal to the inner product $\langle \mathbf{v}, \mathbf{x} \rangle$. Respectively, given a linear map $A : \mathcal{V} \to \mathcal{W}$ between Euclidean spaces, the adjoint map $A^t : \mathcal{W}^* \to \mathcal{V}^*$ can be considered as a map between the spaces themselves: $A^t : \mathcal{W} \to \mathcal{V}$. The defining property of the adjoint map reads:

$$\langle A^t \mathbf{w}, \mathbf{v} \rangle = \langle \mathbf{w}, A\mathbf{v} \rangle \quad \text{for all } \mathbf{v} \in \mathcal{V} \text{ and } \mathbf{w} \in \mathcal{W}.$$

Consequently matrices of adjoint maps A and A^t with respect to orthonormal bases of the Euclidean spaces $\mathcal{V}$ and $\mathcal{W}$ are transposed to each other.

As in the case of Hermitian spaces, one easily derives that a linear transformation $U : \mathcal{V} \to \mathcal{V}$ is orthogonal if and only if $U^{-1} = U^t$. In the matrix form, the relation $U^t U = I$ means that columns of U form an orthonormal set in the coordinate Euclidean space.

Our goal here is to develop the spectral theory for **real normal operators**, i.e. linear transformations $A : \mathcal{V} \to \mathcal{V}$ on a Eucledean space commuting with their transposed operators: $A^t A = A A^t$. Symmetric ($A^t = A$), anti-symmetric ($A^t = -A$), and orthogonal transformations are examples of normal operators in Euclidean geometry.

The right way to proceed is to consider Euclidean geometry as Hermitian geometry, equipped with an additional, *real* structure, and apply the Spectral Theorem of Hermitian geometry to real normal operators extended to the complex space.

EXERCISES

353. Prove the Cauchy-Schwartz inequality for Euclidean inner products: $\langle \mathbf{x}, \mathbf{y} \rangle^2 \leq \langle \mathbf{x}, \mathbf{x} \rangle \langle \mathbf{y}, \mathbf{y} \rangle$, strictly, unless $\mathbf{x}$ and $\mathbf{y}$ are proportional, and derive from this that the angle between non-zero vectors is well-defined. ϟ

354. For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, put $\langle \mathbf{x}, \mathbf{x} \rangle = \sum_{i=1}^n 2x_i^2 - 2\sum_{i=1}^{n-1} x_i x_{i+1}$. Show that the corresponding symmetric bilinear form defines on $\mathbb{R}^n$ a Euclidean structure, and find the angles between the standard coordinate axes in $\mathbb{R}^n$. ✓

355. Prove that $\langle \mathbf{x}, \mathbf{x} \rangle := 2\sum_{i \leq j} x_i x_j$ defines in $\mathbb{R}^n$ a Euclidean structure, find pairwise angles between the standard coordinate axes, and show that permutations of coordinates define orthogonal transformations. ϟ

356. In the standard Euclidean space $\mathbb{R}^{n+1}$ with coordinates $x_0, \dots, x_n$, consider the hyperplane H given by the equation $x_0 + \cdots + x_n = 0$. Find explicitly a basis $\{\mathbf{f}_i\}$ in H, in which the Euclidean structure has the same form as in the previous exercise, and then yet another basis $\{\mathbf{h}_i\}$ in which it has the same form as in the exercise preceding it. ✓

357. Prove that if U is orthogonal, then $\det U = \pm 1$. ϟ

358. Provide a geometric description of orthogonal transformations of the Euclidean plane. Which of them have determinant 1, and which -1? ✓

359. Prove that an $n \times n$-matrix U defines an orthogonal transformation in the standard Euclidean space $\mathbb{R}^n$ if and only if the columns of U form an orthonormal basis.

360. Show that rows of an orthogonal matrix form an orthonormal basis.

Complexification

Since $\mathbb{R} \subset \mathbb{C}$, every complex vector space can be considered as a real vector space simply by "forgetting" that one can multiply by non-real scalars. This operation is called **realification**; applied to a $\mathbb{C}$-vector space $\mathcal{V}$, it produces an $\mathbb{R}$-vector space, denoted $\mathcal{V}^{\mathbb{R}}$, of real dimension twice the complex dimension of $\mathcal{V}$.

In the reverse direction, to a real vector space $\mathcal{V}$ one can associate a complex vector space, $\mathcal{V}^{\mathbb{C}}$, called the **complexification** of $\mathcal{V}$. As a real vector space, it is the direct sum of two copies of $\mathcal{V}$:

$$\mathcal{V}^{\mathbb{C}} := \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x}, \mathbf{y} \in \mathcal{V}\}.$$

Thus the addition is performed componentwise, while the multiplication by complex scalars $\alpha + i\beta$ is introduced with the thought in mind that $(\mathbf{x}, \mathbf{y})$ stands for $\mathbf{x} + i\mathbf{y}$:

$$(\alpha + i\beta)(\mathbf{x}, \mathbf{y}) := (\alpha\mathbf{x} - \beta\mathbf{y}, \beta\mathbf{x} + \alpha\mathbf{y}).$$

This results in a $\mathbb{C}$-vector space $\mathcal{V}^{\mathbb{C}}$ whose complex dimension equals the real dimension of $\mathcal{V}$.

`Example.` $(\mathbb{R}^n)^{\mathbb{C}} = \mathbb{C}^n = \{\mathbf{x} + i\mathbf{y} \mid \mathbf{x}, \mathbf{y} \in \mathbb{R}^n\}$.

A productive point of view on complexification is that it is a complex vector space with an *additional structure* that "remembers" that the space was constructed from a real one. This additional structure is the operation of **complex conjugation** $(\mathbf{x}, \mathbf{y}) \mapsto (\mathbf{x}, -\mathbf{y})$.

The operation in itself is a map $\sigma : \mathcal{V}^{\mathbb{C}} \to \mathcal{V}^{\mathbb{C}}$, satisfying $\sigma^2 = \mathrm{id}$, which is **anti-linear** over $\mathbb{C}$. The latter means that $\sigma(\lambda\mathbf{z}) = \bar{\lambda}\sigma(\mathbf{z})$ for all $\lambda \in \mathbb{C}$ and all $\mathbf{z} \in \mathcal{V}^{\mathbb{C}}$. In other words, σ is $\mathbb{R}$-linear, but anti-commutes with multiplication by i: $\sigma(i\mathbf{z}) = -i\sigma(\mathbf{z})$.

Conversely, let $\mathcal{W}$ be a complex vector space equipped with an anti-linear operator whose square is the identity[4]:

$$\sigma : \mathcal{W} \to \mathcal{W}, \quad \sigma^2 = \mathrm{id}, \quad \sigma(\lambda\mathbf{z}) = \bar{\lambda}\sigma(\mathbf{z}) \ \text{ for all } \lambda \in \mathbb{C},\ \mathbf{z} \in \mathcal{W}.$$

Let $\mathcal{V}$ denote the *real* subspace in $\mathcal{W}$ that consists of all σ-invariant vectors. We claim that $\mathcal{W}$ ***is canonically identified with the complexification of*** $\mathcal{V}$***:*** $\mathcal{W} = \mathcal{V}^{\mathbb{C}}$***.*** Indeed, every vector $\mathbf{z} \in \mathcal{W}$ is uniquely written as the sum of σ-invariant and σ-anti-invariant vectors:

$$\mathbf{z} = \frac{1}{2}(\mathbf{z} + \sigma\mathbf{z}) + \frac{1}{2}(\mathbf{z} - \sigma\mathbf{z}).$$

[4]An transformation whose square is the identity is called an **involution**.

Since $\sigma i = -i\sigma$, multiplication by i transforms σ-invariant vectors to σ-anti-invariant ones, and *vice versa*. Thus, $\mathcal{W}$ as a real space is the direct sum $\mathcal{V} \oplus (i\mathcal{V}) = \{\mathbf{x} + i\mathbf{y} \mid \mathbf{x}, \mathbf{y} \in \mathcal{V}\}$, where multiplication by i acts in the required for the complexification fashion: $i(\mathbf{x} + i\mathbf{y}) = -\mathbf{y} + i\mathbf{x}$.

The construction of complexification and its abstract description in terms of the complex conjugation operator σ are the tools that allow one to carry over results about complex vector spaces to real vector spaces. The idea is to consider real objects as complex ones *invariant* under the complex conjugation σ, and apply (or improve) theorems of complex linear algebra in a way that would *respect* σ.

`Example.` A real matrix can be considered as a complex one. This way an $\mathbb{R}$-linear map defines a $\mathbb{C}$-linear map (on the complexified space). More abstractly, given an $\mathbb{R}$-linear map $A : \mathcal{V} \to \mathcal{V}$, one can associate to it a $\mathbb{C}$-linear map $A^{\mathbb{C}} : \mathcal{V}^{\mathbb{C}} \to \mathcal{V}^{\mathbb{C}}$ by $A^{\mathbb{C}}(\mathbf{x}, \mathbf{y}) := (A\mathbf{x}, A\mathbf{y})$. This map is *real* in the sense that it commutes with the complex conjugation: $A^{\mathbb{C}}\sigma = \sigma A^{\mathbb{C}}$.

Vice versa, let $B : \mathcal{V}^{\mathbb{C}} \to \mathcal{V}^{\mathbb{C}}$ be a $\mathbb{C}$-linear map that commutes with σ: $\sigma(B\mathbf{z}) = B\sigma(\mathbf{z})$ for all $\mathbf{z} \in \mathcal{V}^{\mathbb{C}}$. When $\sigma(\mathbf{z}) = \pm\mathbf{z}$, we find $\sigma(B\mathbf{z}) = \pm B\mathbf{z}$, i.e. the subspaces $\mathcal{V}$ and $i\mathcal{V}$ of real and imaginary vectors are B-invariant. Moreover, since B is $\mathbb{C}$-linear, we find that for $\mathbf{x}, \mathbf{y} \in \mathcal{V}$, $B(\mathbf{x} + i\mathbf{y}) = B\mathbf{x} + iB\mathbf{y}$. Thus $B = A^{\mathbb{C}}$ where the linear operator $A : \mathcal{V} \to \mathcal{V}$ is obtained by restricting B to $\mathcal{V}$.

Our nearest goal is to obtain real analogues of the Spectral Theorem and its corollaries. One way to do it is to combine corresponding complex results with complexification. Let $\mathcal{V}$ be a Euclidean space. We extend the inner product to the complexification $\mathcal{V}^{\mathbb{C}}$ in such a way that it becomes an Hermitian inner product. Namely, for all $\mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}' \in \mathcal{V}$, put

$$\langle \mathbf{x} + i\mathbf{y}, \mathbf{x}' + i\mathbf{y}' \rangle = \langle \mathbf{x}, \mathbf{x}' \rangle + \langle \mathbf{y}, \mathbf{y}' \rangle + i\langle \mathbf{x}, \mathbf{y}' \rangle - i\langle \mathbf{y}, \mathbf{x}' \rangle.$$

It is straightforward to check that this form on $\mathcal{V}^{\mathbb{C}}$ is sesquilinear and Hermitian symmetric. It is also positive definite since $\langle \mathbf{x} + i\mathbf{y}, \mathbf{x} + i\mathbf{y} \rangle = |\mathbf{x}|^2 + |\mathbf{y}|^2$. Note that changing the signs of $\mathbf{y}$ and $\mathbf{y}'$ preserves the real part and reverses the imaginary part of the form. In other words, for all $\mathbf{z}, \mathbf{w} \in \mathcal{V}^{\mathbb{C}}$, we have:

$$\langle \sigma(\mathbf{z}), \sigma(\mathbf{w}) \rangle = \overline{\langle \mathbf{z}, \mathbf{w} \rangle} \ (= \langle \mathbf{w}, \mathbf{z} \rangle).$$

This identity expresses the fact that the Hermitian structure of $\mathcal{V}^{\mathbb{C}}$ came from a Euclidean structure on $\mathcal{V}$. When $A : \mathcal{V}^{\mathbb{C}} \to \mathcal{V}^{\mathbb{C}}$ is a *real* operator, i.e. $\sigma A\sigma = A$, the Hermitian adjoint operator $A^\dagger$ is also real.[5] Indeed, since $\sigma^2 = \mathrm{id}$, we find that for all $\mathbf{z}, \mathbf{w} \in \mathcal{V}^{\mathbb{C}}$

$$\langle \sigma A^\dagger \sigma \mathbf{z}, \mathbf{w}\rangle = \langle \sigma\mathbf{w}, A^\dagger\sigma\mathbf{z}\rangle = \langle A\sigma\mathbf{w}, \sigma\mathbf{z}\rangle = \langle \sigma A\mathbf{w}, \sigma\mathbf{z}\rangle = \langle \mathbf{z}, A\mathbf{w}\rangle,$$

i.e. $\sigma A^\dagger \sigma = A^\dagger$. In particular, complexifications of orthogonal ($U^{-1} = U^t$), **symmetric** ($A^t = A$), **anti-symmetric** ($A^t = -A$), **normal** ($A^tA = AA^t$) operators in a Euclidean space are respectively unitary, Hermitian, anti-Hermitian, normal operators on the complexified space, commuting with the complex conjugation.

EXERCISES

361. Consider $\mathbb{C}^n$ as a real vector space, and describe its complexification.

362. Let σ be the complex conjugation operator on $\mathbb{C}^n$. Consider $\mathbb{C}^n$ as a real vector space. Show that σ is symmetric and orthogonal.

363. On the complex line $\mathbb{C}^1$, find all involutions σ anti-commuting with the multiplication by i: $\sigma i = -i\sigma$.

364. Let σ be an involution on a complex vector space $\mathcal{W}$. Considering $\mathcal{W}$ as a real vector space, find eigenvalues of σ and describe the corresponding eigenspaces. ✓

The Real Spectral Theorem

`Theorem.` ***Let $\mathcal{V}$ be a Euclidean space, and $A : \mathcal{V} \to \mathcal{V}$ a normal operator. Then in the complexification $\mathcal{V}^{\mathbb{C}}$, there exists an orthonormal basis of eigenvectors of $A^{\mathbb{C}}$ which is invariant under complex conjugation and such that the eigenvalues corresponding to conjugated eigenvectors are conjugated.***

`Proof.` Applying the complex Spectral Theorem to the normal operator $B = A^{\mathbb{C}}$, we obtain a decomposition of the complexified space $\mathcal{V}^{\mathbb{C}}$ into a direct orthogonal sum of eigenspaces $\mathcal{W}_1, \dots, \mathcal{W}_r$ of B corresponding to distinct complex eigenvalues $\lambda_1, \dots, \lambda_r$. Note that if $\mathbf{v}$ is an eigenvector of B with an eigenvalue μ, then $B\sigma\mathbf{v} = \sigma B\mathbf{v} = \sigma(\mu\mathbf{v}) = \bar{\mu}\sigma\mathbf{v}$, i.e. $\sigma\mathbf{v}$ is an eigenvector of B with the conjugate eigenvalue $\bar{\mu}$. This shows that if λ_i is a non-real eigenvalue, then its conjugate $\bar{\lambda}_i$ is also one of the eigenvalues of B (say, λ_j), and the corresponding eigenspaces are conjugated: $\sigma(\mathcal{W}_i) = \mathcal{W}_j$. By the

[5] This is obvious in the matrix form: In a real orthonormal basis of $\mathcal{V}$ (which is a complex orthonormal basis of $\mathcal{V}^{\mathbb{C}}$) A has a real matrix, so that $A^\dagger = A^t$. Here we argue the "hard way" in order to illustrate how various aspects of σ-invariance fit together.

same token, if λ_k is real, then $\sigma(\mathcal{W}_k) = \mathcal{W}_k$. This last equality means that $\mathcal{W}_k$ itself is the complexification of a real space, namely of the σ-invariant part of $\mathcal{W}_k$. It coincides with the space $\mathrm{Ker}(\lambda_k I - A) \subset \mathcal{V}$ of real eigenvectors of A with the eigenvalue λ_k. Thus, to construct a required orthonormal basis, we take: for each real eigenspace $\mathcal{W}_k$, a Euclidean orthonormal basis in the corresponding real eigenspace, and for each pair $\mathcal{W}_i, \mathcal{W}_j$ of complex conjugate eigenspaces, an Hermitian orthonormal basis $\{\mathbf{f}_\alpha\}$ in $\mathcal{W}_i$ and the conjugate basis $\{\sigma(\mathbf{f}_\alpha)\}$ in $\mathcal{W}_j = \sigma(\mathcal{W}_i)$. The vectors of all these bases altogether form an orthonormal basis of $\mathcal{V}^{\mathbb{C}}$ satisfying our requirements. $\square$

`Example 1.` Identify $\mathbb{C}$ with the Euclidean plane $\mathbb{R}^2$ in the usual way, and consider the operator $(x+iy) \mapsto (\alpha+i\beta)(x+iy)$ of multiplication by given complex number $\alpha + i\beta$. In the basis $1, i$, it has the matrix

$$A = \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix}.$$

Since A^t represents multiplication by $\alpha - i\beta$, it commutes with A. Therefore A is normal. It is straightforward to check that

$$\mathbf{z} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -i \end{bmatrix} \quad \text{and} \quad \bar{\mathbf{z}} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ i \end{bmatrix}$$

are complex eigenvectors of A with the eigenvalues $\alpha + i\beta$ and $\alpha - i\beta$ respectively, and form an Hermitian orthonormal basis in $(\mathbb{R}^2)^{\mathbb{C}}$.

`Example 2.` If A is a linear transformation in $\mathbb{R}^n$, and λ_0 is a non-real root of its characteristic polynomial $\det(\lambda I - A)$, then the system of linear equations $A\mathbf{z} = \lambda_0 \mathbf{z}$ has non-trivial solutions, which cannot be real though. Let $\mathbf{z} = \mathbf{u} + i\mathbf{v}$ be a complex eigenvector of A with the eigenvalue $\lambda_0 = \alpha + i\beta$. Then $\sigma\mathbf{z} = \mathbf{u} - i\mathbf{v}$ is an eigenvector of A with the eigenvalue $\bar{\lambda}_0 = \alpha - i\beta$. Since $\lambda_0 \neq \bar{\lambda}_0$, the vectors $\mathbf{z}$ and $\sigma\mathbf{z}$ are linearly independent over $\mathbb{C}$, and hence the real vectors $\mathbf{u}$ and $\mathbf{v}$ must be linearly independent over $\mathbb{R}$. Consider the plane $\mathrm{Span}(\mathbf{u}, \mathbf{v}) \subset \mathbb{R}^n$. Since

$$A(\mathbf{u} - i\mathbf{v}) = (\alpha - i\beta)(\mathbf{u} - i\mathbf{v}) = (\alpha\mathbf{u} - \beta\mathbf{v}) - i(\beta\mathbf{u} + \alpha\mathbf{v}),$$

we conclude that A preserves this plane and in the basis $\mathbf{u}, -\mathbf{v}$ in it (note the sign change!) acts by the matrix $\begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix}$. If we assume in addition that A is normal (with respect to the standard Euclidean structure in $\mathbb{R}^n$), then the eigenvectors $\mathbf{z}$ and $\sigma\mathbf{z}$ must be Hermitian orthogonal, i.e.

$$\langle \mathbf{u} - i\mathbf{v}, \mathbf{u} + i\mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{u}\rangle - \langle \mathbf{v}, \mathbf{v}\rangle + 2i\langle \mathbf{u}, \mathbf{v}\rangle = 0.$$

We conclude that $\langle \mathbf{u}, \mathbf{v}\rangle = 0$ and $|\mathbf{u}|^2 - |\mathbf{v}|^2 = 0$, i.e. $\mathbf{u}$ and $\mathbf{v}$ are orthogonal and have the same length. Normalizing the length to 1, we obtain an orthonormal basis of the A-invariant plane, in which the transformation A acts as in Example 1. The geometry of this transformation is known to us from studying geometry of complex numbers: It is the composition of the rotation through the angle $\mathbf{arg}(\lambda_0)$ with the expansion by the factor $|\lambda_0|$. We will call such a transformation of the Euclidean plane a **complex multiplication** or **multiplication by a complex scalar**, λ_0.

`Corollary 1.` ***Given a normal operator on a Euclidean space, the space can be represented as a direct orthogonal sum of invariant lines and planes, on each of which the transformation acts as multiplication by a real or complex scalar respectively.***

`Corollary 2.` ***A transformation in a Euclidean space is orthogonal if and only if the space can be represented as the direct orthogonal sum of invariant lines and planes on each of which the transformation acts as multiplication by*** ± 1 ***and rotation respectively.***

`Corollary 3.` ***In a Euclidean space, every symmetric operator has an orthonormal basis of eigenvectors.***

`Corollary 4.` ***Every quadratic form in a Euclidean space of dimension*** n ***can be transformed by an orthogonal change of coordinates to exactly one of the normal forms:***

$$\lambda_1 x_1^2 + \cdots + \lambda_n x_n^2, \quad \lambda_1 \geq \cdots \geq \lambda_n.$$

`Corollary 5.` ***In a Euclidean space of dimension*** n***, every anti-symmetric bilinear form can be transformed by an orthogonal change of coordinates to exactly one of the normal forms***

$$\langle \mathbf{x}, \mathbf{y}\rangle = \sum_{i=1}^{r} \omega_i (x_{2i-1} y_{2i} - x_{2i} y_{2i-1}), \quad \omega_1 \geq \cdots \geq \omega_r > 0, \quad 2r \leq n.$$

Corollary 6. ***Every real normal matrix*** A ***can be written in the form*** $A = U^t M U$ ***where*** U ***is an orthogonal matrix, and*** M ***is block-diagonal matrix with each block either of size*** 1***, or of size*** 2 ***of the form*** $\begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix}$***, where*** $\alpha^2+\beta^2 \neq 0$***.***

If A ***is symmetric, then only blocks of size*** 1 ***are present (i.e.*** M ***is diagonal).***

If A ***is anti-symmetric, then blocks of size*** 1 ***are zero, and of size*** 2 ***are of the form*** $\begin{bmatrix} 0 & -\omega \\ \omega & 0 \end{bmatrix}$***, where*** $\omega > 0$***.***

If A ***is orthogonal, then all blocks of size*** 1 ***are equal to*** ± 1***, and blocks of size*** 2 ***have the form*** $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$***, where*** $0 < \theta < \pi$***.***

EXERCISES

365. Prove that an operator on a Euclidean vector space is normal if and only if it is the sum of commuting symmetric and anti-symmetric operators.

366. Prove that in the complexification $(\mathbb{R}^2)^{\mathbb{C}}$ of a Euclidean plane, all rotations of $\mathbb{R}^2$ have a common basis of eigenvectors, and find these eigenvectors. ↯

367. Prove that an orthogonal transformation in $\mathbb{R}^3$ is either the rotation through an angle $\theta, 0 \leq \theta \leq \pi$, about some axis, or the composition of such a rotation with the reflection about the plane perpendicular to the axis.

368. Find an orthonormal basis in $\mathbb{C}^n$ in which the transformation defined by the cyclic permutation of coordinates: $(z_1, z_2, \dots, z_n) \mapsto (z_2, \dots, z_n, z_1)$ is diagonal.

369. In the coordinate Euclidean space $\mathbb{R}^n$ with $n \leq 4$, find real and complex normal forms of orthogonal transformations defined by various permutations of coordinates.

370. Transform to normal forms by orthogonal transformations:

(a) $x_1x_2 + x_3x_4$, (b) $2x_1^2 - 4x_1x_2 + x_2^2 - 4x_2x_3$,
(c) $5x_1^2 + 6x_2^2 + 4x_3^2 - 4x_1x_2 - 4x_1x_3$.

371. Show that any anti-symmetric bilinear form on $\mathbb{R}^2$ is proportional to

$$\det[\mathbf{x}, \mathbf{y}] = \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} = x_1y_2 - x_2y_1.$$

Find the operator corresponding to this form, its complex eigenvalues and eigenvectors. ✓

372. In Euclidean spaces, classify all operators which are both orthogonal and anti-symmetric.

373. Recal that a biliniar form on $\mathcal{V}$ is called **non-dedgenerate** if the corresponding linear map $\mathcal{V} \to \mathcal{V}^*$ is an isomorphism, and **degenerate** otherwise. Prove that all non-degenerate anti-symetric bilinear forms on $\mathbb{R}^{2n}$ are equivalent to each other, and that all antisymmetric bilinear forms on $\mathbb{R}^{2n+1}$ are degenerate.

374. Derive Corollaries 1 – 6 from the Real Spectral Theorem.

375. Let $\mathcal{U}$ and $\mathcal{V}$ be two subspaces of dimension 2 in the Euclidean 4-space. Consider the map $T : \mathcal{V} \to \mathcal{V}$ defined as the composition: $\mathcal{V} \subset \mathbb{R}^4 \to \mathcal{U} \subset \mathbb{R}^4 \to \mathcal{V}$, where the arrows are the orthogonal projections to $\mathcal{U}$ and $\mathcal{V}$ respectively. Prove that T is positive, and that its eigenvalues have the form $\cos\phi$, $\cos\psi$ where ϕ, ψ are certain angles, $0 \le \phi, \psi \le \pi/2$.

376. Solve **Gelfand's problem**: In the Euclidean 4-space, classify pairs of planes passing through the origin up to orthogonal transformations of the space. ϟ

Courant–Fischer's Minimax Principle

One of the consequences (equivalent to Corollary 4) of the Spectral Theorem is that a pair (Q, S) of quadratic forms in $\mathbb{R}^n$, of which the first one is positive definite, can be transformed by a linear change of coordinates to the normal form:

$$Q = x_1^2 + \cdots + x_n^2, \quad S = \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2, \quad \lambda_1 \ge \cdots \ge \lambda_n.$$

The eigenvalues $\lambda_1 \ge \ldots \lambda_n$ form the **spectrum** of the pair (Q, S). The following result gives a coordinate-less, geometric description of the spectrum (and thus implies the **Orthogonal Diagonalization Theorem** as it was stated in the Introduction).

`Theorem.` ***The k-th greatest spectral number is given by***

$$\lambda_k = \max_{\mathcal{W}:\ \dim \mathcal{W} = k} \ \min_{\mathbf{x} \in \mathcal{W} - \mathbf{0}} \frac{S(\mathbf{x})}{Q(\mathbf{x})},$$

where the maximum is taken over all k***-dimensional subspaces*** $\mathcal{W} \subset \mathbb{R}^n$***, and minimum over all non-zero vectors in the subspace.***

`Proof.` When $\mathcal{W}$ is given by the equations $x_{k+1} = \cdots = x_n = 0$, the minimal ratio $S(\mathbf{x})/Q(\mathbf{x})$ (achieved on vectors proportional to $\mathbf{e}_k$) is equal to λ_k because

$$\lambda_1 x_1^2 + \cdots + \lambda_k x_k^2 \ge \lambda_k (x_1^2 + \cdots + x_k^2) \ \text{ when } \lambda_1 \ge \cdots \ge \lambda_k.$$

Therefore it suffices to prove for every other k-dimensional subspace $\mathcal{W}$ the minimal ratio cannot be greater than λ_k. For this, denote by $\mathcal{V}$ the subspace of dimension $n-k+1$ given by the equations $x_1 = \cdots = x_{k-1} = 0$. Since $\lambda_k \geq \cdots \geq \lambda_n$, we have:

$$\lambda_k x_k^2 + \cdots + \lambda_n x_n^2 \leq \lambda_k (x_k^2 + \cdots + \mathbf{x}_n^2),$$

i.e. for all non-zero vectors $\mathbf{x}$ in $\mathcal{V}$ the ratio $S(\mathbf{x})/Q(\mathbf{x}) \leq \lambda_k$. Now we invoke the dimension counting argument: $\dim \mathcal{W} + \dim \mathcal{V} = k + (n-k+1) = n+1 > \dim \mathbb{R}^n$, and conclude that $\mathcal{W}$ has a non-trivial intersection with $\mathcal{V}$. Let $\mathbf{x}$ be a non-zero vector in $\mathcal{W} \cap \mathcal{V}$. Then $S(\mathbf{x})/Q(\mathbf{x}) \leq \lambda_k$, and hence the minimum of the ratio S/Q on $\mathcal{W} - \mathbf{0}$ cannot exceed λ_k. □

Applying Theorem to the pair $(Q, -S)$ we obtain yet another characterization of the spectrum:

$$\lambda_k = \min_{\mathcal{W}:\ \dim \mathcal{W} = n-k+1} \ \max_{\mathbf{x} \in \mathcal{W} - \mathbf{0}} \frac{S(\mathbf{x})}{Q(\mathbf{x})}.$$

Formulating some applications, we assume that the space $\mathbb{R}^n$ is Euclidean, and refer to the spectrum of the pair (Q, S) where $Q = |\mathbf{x}|^2$, simply as the spectrum of S.

`Corollary 1.` ***When a quadratic form increases, its spectral numbers do not decrease: If*** $S \leq S'$ ***then*** $\lambda_k \leq \lambda'_k$ ***for all*** $k = 1, \ldots, n$.

`Proof.` Indeed, since $S/Q \leq S'/Q$, the minimum of the ratio S/Q on every k-dimensional subspace $\mathcal{W}$ cannot exceed that of S'/Q, which in particular remains true for that $\mathcal{W}$ on which the maximum of S/Q equal to λ_k is achieved.

The following result is called **Cauchy's interlacing theorem**.

`Corollary 2.` ***Let*** $\lambda_1 \geq \cdots \geq \lambda_n$ ***be the spectrum of a quadratic form*** S, ***and*** $\lambda'_1 \geq \cdots \geq \lambda'_{n-1}$ ***be the spectrum of the quadratic form*** S' ***obtained by restricting*** S ***to a given hyperplane*** $\mathbb{R}^{n-1} \subset \mathbb{R}^n$ ***passing through the origin. Then:***

$$\lambda_1 \geq \lambda'_1 \geq \lambda_2 \geq \lambda'_2 \geq \cdots \geq \lambda_{n-1} \geq \lambda'_{n-1} \geq \lambda_n.$$

`Proof.` The maximum over all k-dimensional subspaces $\mathcal{W}$ cannot be smaller than the maximum (of the same quantities) over subspaces lying inside the hyperplane. This proves that $\lambda_k \geq \lambda'_k$. Applying the same argument to $-S$ and subspaces of dimension $n-k-1$, we conclude that $-\lambda_{k+1} \geq -\lambda'_k$. □

An **ellipsoid** in a Euclidean space is defined as the level-1 set $E = \{\mathbf{x} \mid S(\mathbf{x}) = 1\}$ of a positive definite quadratic form, S. It follows from the Spectral Theorem that every ellipsoid can be transformed by an orthogonal transformation to **principal axes**: a normal form

$$\frac{x_1^2}{\alpha_1^2} + \cdots + \frac{x_n^2}{\alpha_n^2} = 1, \quad 0 < \alpha_1 \le \cdots \le \alpha_n.$$

The vectors $\mathbf{x} = \pm\alpha_k \mathbf{e}_k$ lie on the ellipsoid, and their lengths α_k are called **semiaxes** of E. They are related to the spectral numbers $\lambda_1 \ge \cdots \ge \lambda_k > 0$ of the quadratic form by $\alpha_k^{-1} = \sqrt{\lambda_k}$. From Corollaries 1 and 2 respectively, we obtain:

Given two concentric ellipsoids enclosing one another, the semiaxes of the inner ellipsoid do not exceed corresponding semiaxes of the outer:

If $E' \subset E$*, then* $\alpha'_k \le \alpha_k$ *for all* $k = 1, \dots, n$.

Semiaxes of a given ellipsoid are interlaced by semiaxes of any section of it by a hyperplane passing through the center:

If $E' = E \cap \mathbb{R}^{n-1}$*, then* $\alpha_k \le \alpha'_k \le \alpha_{k+1}$ *for* $k = 1, \dots, n-1$.

EXERCISES

377. Prove that every ellipsoid in $\mathbb{R}^n$ has n pairwise perpendicular hyperplanes of bilateral symmetry.

378. Given an ellipsoid $E \subset \mathbb{R}^3$, find a plane passing through its center and intersecting E in a circle. ↯

379. Formulate and prove counterparts of Courant–Fischer's minimax principle and Cauchy's interlacing theorem for Hermitian forms.

380. Prove that semiaxes $\alpha_1 \le \alpha_2 \le \dots$ of an ellipsoid in $\mathbb{R}^n$ and semiaxes $\alpha'_k \le \alpha'_2 \le \dots$ of its section by a linear subspaces of *co*dimension k are related by the inequalities: $\alpha_i \le \alpha'_i \le \alpha_{i+k}$, $i = 1, \dots, n-k$.

381. From the Real Spectral Theorem, derive the Orthogonal Diagonalization Theorem as it is formulated in the Introduction, i.e. for pairs of quadratic forms on $\mathbb{R}^n$, one of which is positive definite.↯

Small Oscillations

Let us consider the system of n identical masses m positioned at the vertices of a regular n-gon, which are cyclically connected by n identical elastic springs, and can oscillate in the direction perpendicular to the plane of the n-gon.

Assuming that the amplitudes of the oscillation are small, we can describe the motion of the masses as solutions to the following system of n second-order Ordinary Differential Equations (ODE for short) expressing Newton's law of motion (mass × acceleration = force):

$$\begin{array}{lcl} m\ddot{x}_1 & = & -k(x_1 - x_n) - k(x_1 - x_2), \\ m\ddot{x}_2 & = & -k(x_2 - x_1) - k(x_2 - x_3), \\ & \cdots & \\ m\ddot{x}_{n-1} & = & -k(x_{n-1} - x_{n-2}) - k(x_{n-1} - x_n) \\ m\ddot{x}_n & = & -k(x_n - x_{n-1}) - k(x_n - x_1). \end{array}.$$

Here $x_1, \dots, x_n$ are the displacements of the n masses in the direction perpendicular to the plane, and k characterizes the rigidity of the springs. [6]

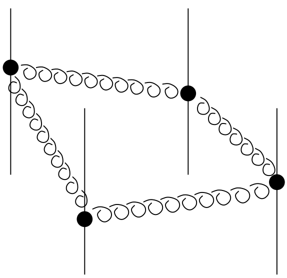

Figure 43

In fact the above ODE system can be read off a pair of quadratic forms: the **kinetic energy**

$$K(\dot{\mathbf{x}}) = \frac{m\dot{x}_1^2}{2} + \frac{m\dot{x}_2^2}{2} + \cdots + \frac{m\dot{x}_n^2}{2},$$

and the **potential energy**

$$P(\mathbf{x}) = k\frac{(x_1 - x_2)^2}{2} + k\frac{(x_2 - x_3)^2}{2} + \cdots + k\frac{(x_n - x_1)^2}{2}.$$

[6]More precisely (see Figure 43, where $n = 4$), we may assume that the springs are stretched, but the masses are confined on the vertical rods and can only slide along them without friction. When a string of length L is horizontal ($\Delta x = 0$), the stretching force T is compensated by the reactions of the rods. When $\Delta x \neq 0$, the horizontal component of the stretching force is still compensated, but the vertical component contributes to the right hand side of Newton's equations. When Δx is small, the contribution equals approximately $-T(\Delta x)/L$ (so that $k = -T/L$).

Namely, for any conservative mechanical system with quadratic kinetic and potential energy functions

$$K(\dot{\mathbf{x}}) = \frac{1}{2}\langle \dot{\mathbf{x}}, M\dot{\mathbf{x}}\rangle,\ \ P(\mathbf{x}) = \frac{1}{2}\langle \mathbf{x}, Q\mathbf{x}\rangle$$

the equations of motion assume the form

$$M\ddot{\mathbf{x}} = -Q\mathbf{x}.$$

A linear change of variables $\mathbf{x} = C\mathbf{y}$ transforms the kinetic and potential energy functions to a new form with the matrices $M' = C^tMC$ and $Q' = C^tQC$. On the other hand, the same change of variables transforms the ODE system $M\ddot{\mathbf{x}} = -Q\mathbf{x}$ to $MC\ddot{\mathbf{y}} = -QC\mathbf{y}$. Multiplying by C^t we get $M'\ddot{\mathbf{y}} = -Q'\mathbf{y}$ and see that the relationship between K, P and the ODE system is preserved. The relationship is therefore *intrinsic*, i.e. independent on the choice of coordinates.

Since the kinetic energy is positive we can apply the Orthogonal Diagonalization Theorem in order to transform K and P simultaneously to

$$\frac{1}{2}(\dot{X}_1^2 + ... + \dot{X}_n^2),\ \text{and}\ \frac{1}{2}(\lambda_1 X_1^2 + ... + \lambda_n X_n^2).$$

The corresponding ODE system splits into unlinked 2-nd order ODEs

$$\ddot{X}_1 = -\lambda_1 X_1,\quad ...,\quad \ddot{X}_n = -\lambda_n X_n.$$

When the potential energy is also positive, we obtain a system of n unlinked **harmonic oscillators** with frequencies $\omega = \sqrt{\lambda_1}, ..., \sqrt{\lambda_n}$.

`Example 1:` *Harmonic oscillators.* The equation $\ddot{X} = -\omega^2 X$ has solutions

$$X(t) = A\,\mathbf{cos}\,\omega t + B\,\mathbf{sin}\,\omega t,$$

where $A = X(0)$ and $B = \dot{X}(0)/\omega$ are arbitrary real constants. It is convenient to plot the solutions on the **phase plane** with coordinates $(X, Y) = (X, \dot{X}/\omega)$. In such coordinates, the equations of motion assume the form

$$\begin{array}{rcr} \dot{X} & = & \omega Y \\ \dot{Y} & = & -\omega X \end{array},$$

and the solutions

$$\left[\begin{array}{c} X(t) \\ Y(t) \end{array}\right] = \left[\begin{array}{rr} \mathbf{cos}\,\omega t & \mathbf{sin}\,\omega t \\ -\mathbf{sin}\,\omega t & \mathbf{cos}\,\omega t \end{array}\right] \left[\begin{array}{c} X(0) \\ Y(0) \end{array}\right].$$

In other words (see Figure 44), the motion on the phase plane is described as ***clockwise rotation with the angular velocity*** ω. Since there is one trajectory through each point of the phase plane, the general theory of Ordinary Differential Equations (namely, the theorem about uniqueness and existence of solutions with given initial conditions) guarantees that these are all the solutions to the ODE $\ddot{X} = -\omega^2 X$.

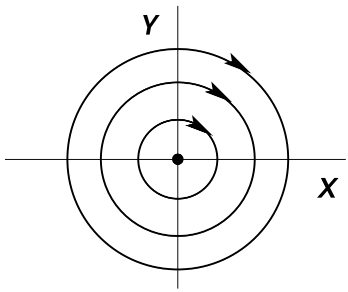

Figure 44

Let us now examine the behavior of our system of n masses cyclically connected by the springs. To find the common orthogonal basis of the pair of quadratic forms K and P, we first note that, since K is proportional to the standard Euclidean structure, it suffices to find an orthogonal basis of eigenvectors of the symmetric matrix Q.

In order to give a concise description of the ODE system $m\ddot{\mathbf{x}} = Q\mathbf{x}$, introduce operator $T : \mathbb{R}^n \to \mathbb{R}^n$ which cyclically shifts the coordinates: $T(x_1, x_2, \dots, x_n)^t = (x_2, \dots, x_n, x_1)$. Then $Q = k(T + T^{-1} - 2I)$. Note that the operator T is obviously orthogonal, and hence unitary in the complexification $\mathbb{C}^n$ of the the space $\mathbb{R}^n$. We will now construct its basis of eigenvectors, which should be called the **Fourier basis**.[7] Namely, let $x_k = \zeta^k$ where $\zeta^n = 1$. Then the sequence $\{x_k\}$ is repeating every n terms, and $x_{k+1} = \zeta x_k$ fo all $k \in \mathbb{Z}$. Thus $T\mathbf{x} = \zeta\mathbf{x}$, where $\mathbf{x} = (\zeta, \zeta^2, \dots, \zeta^n)t$. When $\zeta = e^{2\pi\sqrt{-1}\,l/n}$, $l = 1, 2, \dots, n$ runs various nth roots of unity, we obtain n eigenvectors of the operator T, which corresponds to different eigenvalues, and hence are linearly independent. They are automatically pairwise Hermitian orthogonal (since T is unitary), and happen to have the same Hermitian inner square, equal to n. Thus, when divided

[7] After French mathematician Joseph **Fourier** (1768–1830), and by analogy with the theory of Fourier series.

by $\sqrt{n}$, these vectors form an orthonormal basis in $\mathbb{C}^n$. Besides, this basis is invariant under complex conjugation (because replacing the eigenvalue ζ with $\bar{\zeta}$ also conjugates the corresponding eigenvector).

Now, applying this to $Q = k(T + T^{-1} - 2I)$, we conclude that Q is diagonal in the Fourier basis with the eigenvalues

$$k(\zeta+\zeta^{-1}-2) = 2k(\mathbf{cos}(2\pi l/n)-1) = -4k\,\mathbf{sin}^2\,\pi l/n,\ \ l = 1, 2, \ldots, n.$$

When $\zeta \neq \bar{\zeta}$, this pair of roots of unity yields the same eigenvalue of Q, and the real and imaginary parts of the Fourier eigenvector $\mathbf{x} = (\zeta, \ldots, \zeta^n)^t$ span in $\mathbb{R}^n$ the 2-dimensional eigenplane of the operator Q. When $\zeta = 1$ or -1 (the latter happens only when n is even), the corresponding eigenvalue of Q is 0 and $-4k$ respectively, and the eigenspace is 1-dimensional (spanned the respective Fourier vectors $(1, \ldots, 1)^t$ and $(-1, 1, \ldots, -1, 1)^t$. The whole systems decomposes into superposition of independent "modes of oscillation" (patterns) described by the equations

$$\ddot{X}_l = -\omega_l^2 X_l, \quad \text{where } \omega_l = 2\sqrt{\frac{k}{m}}\left|\mathbf{sin}\,\pi\frac{l}{n}\right|, \quad l = 1, \ldots, n.$$

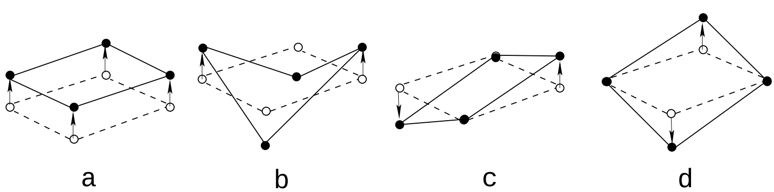

Figure 45

`Example 2:` $n = 4$. Here $z = 1, -1, \pm i$. The value $z = 1$ corresponds to the eigenvector $(1, 1, 1, 1)$ and the eigenvalue 0. This "mode of oscillation" is described by the ODE $\ddot{X} = 0$, and actually corresponds to the steady translation of the chain as a whole with the constant speed (Figure 45a). The value $\zeta = -1$ corresponds to the eigenvector $(-1, 1, -1, 1)$ (Figure 45b) with the frequency of oscillation $2\sqrt{k/m}$. The values $\zeta = \pm i$ correspond to the eigenvectors $(\pm i, -1, \mp i, 1)$. Their real and imaginary parts $(0, -1, 0, 1)$ and $(1, 0, -1, 0)$ (Figures 45cd) span the plane of modes of oscillation with the same frequency $\sqrt{2k/m}$. The general motion of the system is the superposition of these four patterns.

`Remark.` In fact the oscillatory system we've just studied can be considered as a model of sound propagation in a one-dimensional crystal. One can similarly analyze propagation of sound waves in 2-dimensional membranes of rectangular or periodic (toroidal) shape, or in similar 3-dimesnional regions. Physicists often call the resulting picture: the superposition of independent sinusoidal waves, an *ideal gas of phonons*. Here "ideal gas" refers to independence of eigenmodes of oscillators (therefore behaving as noninterracting particles of a sparse gas), and "phonons" emphasises that the "particles" are rather *bells* producing sound waves of various frequencies.

The mathematical aspect of this theory is even more general: the Orthogonal Diagonalization Theorem guarantees that ***small oscillations in any conservative mechanical system near a local minimum of potential energy are described as superpositions of independent harmonic oscillations.***

EXERCISES

382. A mass m is suspended on a weightless rod of length l (as a clock **pendulum**), and is swinging without friction under the action of the force of gravity mg (where g is the **gravitation constant**). Show that the Newton equation of motion of the pendulum has the form $l\ddot{x} = -g \sin x$, where x is the angle the rod makes with the downmward vertical direction, and show that the frequency of small oscillations of the pendulum near the lower equlibrium ($x = 0$) is equal to $\sqrt{g/l}$. ↯

383. In the mass-spring chain (studied in the text) with $n = 3$, find frequencies and describe explicitly the modes of oscillations.

384. The same, for 6 masses positioned at the vertices of the regular hexagon (like the 6 carbon atoms in benzene molecules).

385.* Given n numbers $C_1, \dots, C_n$ (real or complex), we form from them an infinite periodic sequence $\{C_k\}$, $\dots, C_{-1}, C_0, C_1, \dots, C_n, C_{n+1}, \dots$, where $C_{k+n} = C_k$. Let C denote the $n \times n$-matrix whose entries $c_{ij} = C_{j-i}$. Prove that all such matrices (corresponding to different n-periodic sequences) are normal, that they commute, and find their common eigenvectors. ↯

386.* Study small oscillations of a 2-dimansional crystal lattice of toroidal shape consisting of $m \times n$ identical masses (positioned in m circular "rows" and n circular "columns", each interacting only with its four neighbors).

387. Using Courant–Fischer's minimax principle, explain why a cracked bell sounds lower than the intact one.

3 Jordan Canonical Forms

Characteristic Polynomials and Root Spaces

Let $\mathcal{V}$ be a finite dimensional $\mathbb{K}$-vector space. We do not assume that $\mathcal{V}$ is equipped with any structure in addition to the structure of a $\mathbb{K}$-vector space. In this section, we study geometry of linear operators on $\mathcal{V}$. In other words, we study the problem of classification of linear operators $A : \mathcal{V} \to \mathcal{V}$ up to **similarity** transformations $A \mapsto C^{-1}AC$, where C stands for arbitrary invertible linear transformations of $\mathcal{V}$.

Let $n = \dim \mathcal{V}$, and let A be the matrix of a linear operator with respect to some basis of $\mathcal{V}$. Recall that

$$\det(\lambda I - A) = \lambda^n + p_1\lambda^{n-1} + \cdots + p_{n-1}\lambda + p_n$$

is called the **characteristic polynomial** of A. In fact it does not depend on the choice of a basis. Indeed, under a change $\mathbf{x} = C\mathbf{x}'$ of coordinates, the matrix of a linear operator $\mathbf{x} \mapsto A\mathbf{x}$ is transformed into the matrix $C^{-1}AC$ similar to A. We have:

$$\det(\lambda I - C^{-1}AC) = \det[C^{-1}(\lambda I - A)C] = \\ (\det C^{-1})\det(\lambda I - A)(\det C) = \det(\lambda I - A).$$

Therefore, the characteristic polynomial of a linear operator is well-defined (by the geometry of A). In particular, ***coefficients of the characteristic polynomial do not change under similarity transformations.***

Let $\lambda_0 \in \mathbb{K}$ be a root of the characteristic polynomial. Then $\det(\lambda_0 I - A) = 0$, and hence the system of homogeneous linear equations $A\mathbf{x} = \lambda_0\mathbf{x}$ has a non-trivial solution, $\mathbf{x} \neq \mathbf{0}$. As before,we call any such solution an **eigenvector** of A, and call λ_0 the corresponding **eigenvalue**. All solutions to $A\mathbf{x} = \lambda_0\mathbf{x}$ (including $\mathbf{x} = \mathbf{0}$) form a linear subspace in $\mathcal{V}$, called the **eigenspace** of A corresponding to the eigenvalue λ_0.

Let us change slightly our point of view on the eigenspace. It is the null space of the operator $A - \lambda_0 I$. Consider powers of this operator and their null spaces. If $(A - \lambda_0 I)^k\mathbf{x} = 0$ for some $k > 0$, then $(A - \lambda_0 I)^l\mathbf{x} = 0$ for all $l \geq k$. Thus the null spaces are nested:

$$\operatorname{Ker}(A - \lambda_0 I) \subset \operatorname{Ker}(A - \lambda_0 I)^2 \subset \cdots \subset \operatorname{Ker}(A - \lambda_0 I)^k \subset \ldots$$

On the other hand, since $\dim \mathcal{V} < \infty$, nested subspaces must stabilize, i.e. starting from some $m > 0$, we have:

$$\mathcal{W}_{\lambda_0} := \operatorname{Ker}(A - \lambda_0 I)^m = \operatorname{Ker}(A - \lambda_0 I)^{m+1} = \cdots$$

We call the subspace $\mathcal{W}_{\lambda_0}$ a **root space** of the operator A, namely, the root space corresponding to the root λ_0 of the characteristic polynomial.

Note that if $\mathbf{x} \in \mathcal{W}_{\lambda_0}$, then $A\mathbf{x} \in \mathcal{W}_{\lambda_0}$, because $(A - \lambda_0 I)^m A\mathbf{x} = A(A - \lambda_0 I)^m \mathbf{x} = A\mathbf{0} = \mathbf{0}$. Thus a root space is A-invariant. Denote by $\mathcal{U}_{\lambda_0}$ the range of $(A - \lambda_0 I)^m$. It is also A-invariant, since if $\mathbf{x} = (A - \lambda_0 I)^m \mathbf{y}$, then $A\mathbf{x} = A(A - \lambda_0 I)^m \mathbf{y} = (A - \lambda_0 I)^m (A\mathbf{y})$.

`Lemma.` $\mathcal{V} = \mathcal{W}_{\lambda_0} \oplus \mathcal{U}_{\lambda_0}$.

`Proof.` Put $B := (A - \lambda_0 I)^m$, so that $\mathcal{W}_{\lambda_0} = \operatorname{Ker} B$, $\mathcal{U}_{\lambda_0} = B(\mathcal{V})$. Let $\mathbf{x} = B\mathbf{y} \in \operatorname{Ker} B$. Then $B\mathbf{x} = \mathbf{0}$, i.e. $\mathbf{y} \in \operatorname{Ker} B^2$. But $\operatorname{Ker} B^2 = \operatorname{Ker} B$ by the assumption that $\operatorname{Ker} B = \mathcal{W}_{\lambda_0}$ is the root space. Thus $\mathbf{y} \in \operatorname{Ker} B$, and hence $\mathbf{x} = B\mathbf{y} = \mathbf{0}$. This proves that $\operatorname{Ker} B \cap B(\mathcal{V}) = \{\mathbf{0}\}$. Therefore the subspace in $\mathcal{V}$ spanned by $\operatorname{Ker} B$ and $\mathcal{B}(\mathcal{V})$ is their direct sum. On the other hand, for any operator, $\dim \operatorname{Ker} B + \dim B(\mathcal{V}) = \dim \mathcal{V}$. Thus, the subspace spanned by $\operatorname{Ker} B$ and $B(\mathcal{V})$ is the whole space $\mathcal{V}$.

`Corollary 1.` ***For any*** $\lambda \neq \lambda_0$***, the root space*** $\mathcal{W}_\lambda \subset \mathcal{U}_{\lambda_0}$***.***

`Proof.` Indeed, $\mathcal{W}_\lambda$ is invariant with respect to $A - \lambda_0 I$, but contains no eigenvectors of A with eigenvalue λ_0. Therefore $A - \lambda_0 I$ and all powers of it are invertible on $\mathcal{W}_\lambda$. Thus $\mathcal{W}_\lambda$ lies in the range $\mathcal{U}_{\lambda_0}$ of $B = (A - \lambda_0 I)^m$.

`Corollary 2.` ***Suppose that*** $(\lambda - \lambda_1)^{m_1} \dots (\lambda - \lambda_r)^{m_r}$ ***is the characteristic polynomial of*** $A : \mathcal{V} \to \mathcal{V}$***, where*** $\lambda_1, \dots, \lambda_r \in \mathbb{K}$ ***are pairwise distinct roots. Then*** $\mathcal{V}$ ***is the direct sum of root spaces:***

$$\mathcal{V} = \mathcal{W}_{\lambda_1} \oplus \cdots \oplus \mathcal{W}_{\lambda_r}.$$

`Proof.` From Corollary 1, it follows by induction on r, that $\mathcal{V} = \mathcal{W}_{\lambda_1} \oplus \cdots \oplus \mathcal{W}_{\lambda_r} \oplus \mathcal{U}$, where $\mathcal{U} = \mathcal{U}_{\lambda_1} \cap \cdots \cap \mathcal{U}_{\lambda_r}$. In particular, $\mathcal{U}$ is A-invariant as the intersection of A-invariant subspaces, but contains no eigenvectors of A with eigenvalues $\lambda_1, \dots, \lambda_r$. Picking bases in each of the direct summands $\mathcal{W}_{\lambda_i}$ and in $\mathcal{U}$, we obtain a basis in $\mathcal{V}$, in which the matrix of A is block-diagonal. Therefore the characteristic polynomial of A is the product of the characteristic polynomials of A restricted to the summands. So far we haven't used the hypothesis that the characteristic polynomial of A factors into a product of $\lambda - \lambda_i$. Invoking this hypothesis, we see that the factor of the characteristic polynomial, corresponding to $\mathcal{U}$ must have degree 0, and hence $\dim \mathcal{U} = 0$.

`Remarks.` (1) We will see later that dimensions of the root spaces coincide with multiplicities of the roots: $\dim \mathcal{W}_{\lambda_i} = m_i$.

(2) The restriction of A to $\mathcal{W}_{\lambda_i}$ has the property that some power of $A - \lambda_i I$ vanishes. A linear operator some power of which vanishes is called **nilpotent**. Our next task will be to study the geometry of nilpotent operators.

(3) Our assumption that the characteristic polynomial factors completely over $\mathbb{K}$ is automatically satisfied in the case $\mathbb{K} = \mathbb{C}$ due to the Fundamental Theorem of Algebra. Thus, we have proved for every linear operator on a finite dimensional complex vector space, that the space decomposes in a canonical fashion into the direct sum of invariant subspaces on each of which the operator differs from a nilpotent one by scalar summand.

EXERCISES

388. Let $A, B : \mathcal{V} \to \mathcal{V}$ be two commuting linear operators, and p and q two polynomials in one variable. Show that the operators $p(A)$ and $q(B)$ commute.

389. Prove that if A commutes with B, then root spaces of A are B-invariant.

390. Let λ_0 be a root of the characteristic polynomial of an operator A, and m its multiplicity. What are possible values for the dimension of the eigenspace corresponding to λ_0?

391. Let $\mathbf{v} \in \mathcal{V}$ be a non-zero vector, and $\mathbf{a} : \mathcal{V} \to \mathbb{K}$ a non-zero linear function. Find eigenvalues and eigenspaces of the operator $\mathbf{x} \mapsto \mathbf{a}(\mathbf{x})\mathbf{v}$.

Nilpotent Operators

`Example.` Introduce a nilpotent linear operator $N : \mathbb{K}^n \to \mathbb{K}^n$ by describing its action on vectors of the standard basis:

$$N\mathbf{e}_n = \mathbf{e}_{n-1}, \quad N\mathbf{e}_{n-1} = \mathbf{e}_{n-2}, \quad \dots, \quad N\mathbf{e}_2 = \mathbf{e}_1, \quad N\mathbf{e}_1 = \mathbf{0}.$$

Then $N^n = 0$ but $N^{n-1} \neq 0$. We will call N, as well as any operator similar to it, a **regular nilpotent** operator. The matrix of N in the standard basis has the form

$$\left[\begin{array}{ccccc} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & & \dots & & \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{array}\right].$$

It has the range of dimension $n-1$ spanned by $\mathbf{e}_1, \dots, \mathbf{e}_{n-1}$ and the null space of dimension 1 spanned by $\mathbf{e}_1$.

Proposition. ***Let*** $N : \mathcal{V} \to \mathcal{V}$ ***be a nilpotent operator on a*** $\mathbb{K}$***-vector space of finite dimension. Then the space can be decomposed into the direct sum of*** N***-invariant subspaces, on each of which*** N ***is regular.***

Proof. We use induction on $\dim \mathcal{V}$. When $\dim \mathcal{V} = 0$, there is nothing to prove. Now consider the case when $\dim \mathcal{V} > 0$.

The range $N(\mathcal{V})$ is N-invariant, and $\dim N(\mathcal{V}) < \dim V$ (since otherwise N could not be nilpotent). By the induction hypothesis, the space $N(\mathcal{V})$ can be decomposed into the direct sum of N-invariant subspaces, on each of which N is regular. Let l be the number of these subspaces, $n_1, \dots, n_l$ their dimensions, and $\mathbf{e}_1^{(i)}, \dots, \mathbf{e}_{n_i}^{(i)}$ a basis in the ith subspace such that N acts on the basis vectors as in Example:

$$\mathbf{e}_{n_i}^{(i)} \mapsto \cdots \mapsto \mathbf{e}_1^{(i)} \mapsto \mathbf{0}.$$

Since each $\mathbf{e}_{n_i}^{(i)}$ lies in the range of N, we can pick a vector $\mathbf{e}_{n_i+1}^{(i)} \in \mathcal{V}$ such that $N\mathbf{e}_{n_i+1}^{(i)} = \mathbf{e}_{n_i}^{(i)}$. Note that $\mathbf{e}_1^{(1)}, \dots, \mathbf{e}_1^{(l)}$ form a basis in $(\operatorname{Ker} N) \cap N(\mathcal{V})$. We complete it to a basis

$$\mathbf{e}_1^{(1)}, \dots, \mathbf{e}_1^{(l)}, \mathbf{e}_1^{(l+1)}, \dots, \mathbf{e}_1^{(r)}$$

of the whole null space $\operatorname{Ker} N$. We claim that *all the vectors* $\mathbf{e}_j^{(i)}$ *form a basis in* $\mathcal{V}$, and therefore the $l + r$ subspaces

$$\operatorname{Span}(\mathbf{e}_1^{(i)}, \dots, \mathbf{e}_{n_i}^{(i)}, \mathbf{e}_{n_i+1}^{(i)}), \quad i = 1, \dots, l, l+1, \dots, r,$$

(of which the last $r - l$ are 1-dimensional) form a decomposition of $\mathcal{V}$ into the direct sum with required properties.

To justify the claim, notice that the subspace $\mathcal{U} \subset \mathcal{V}$ spanned by $n_1 + \cdots + n_l = \dim N(\mathcal{V})$ vectors $\mathbf{e}_j^{(i)}$ with $j > 1$ is mapped by N onto the space $N(\mathcal{V})$. Therefore: (a) those vectors form a basis of $\mathcal{U}$, (b) $\dim \mathcal{U} = \dim N(\mathcal{V})$, and (c) $\mathcal{U} \cap \operatorname{Ker} N = \{\mathbf{0}\}$. On the other hand, vectors $\mathbf{e}_j^{(i)}$ with $j = 1$ form a basis of $\operatorname{Ker} N$, and since $\dim \operatorname{Ker} N + \dim N(\mathcal{V}) = \dim \mathcal{V}$, together with the above basis of $\mathcal{U}$, they form a basis of $\mathcal{V}$. $\square$

Corollary 1. ***The matrix of a nilpotent operator in a suitable basis is block diagonal with regular diagonal blocks*** (as in Example) ***of certain sizes*** $n_1 \geq \cdots \geq n_r > 0$.

The basis in which the matrix has this form, as well as the decomposition into the direct sum of invariant subspaces as described in Proposition, are not canonical, since choices are involved on each step of induction. However, the dimensions $n_1 \geq \cdots \geq n_r > 0$ of the subspaces turn out to be uniquely determined by the geometry of the operator.

To see why, introduce the following **Young tableaux** (Figure 46). It consists of r rows of identical square cells. The lengths of the rows represent dimensions $n_1 \geq n_2 \geq \cdots \geq n_r > 0$ of invariant subspaces, and in the cells of each row we place the basis vectors of the corresponding subspace, so that the operator N sends each vector to its left neighbor (and those of the leftmost column to $\mathbf{0}$).

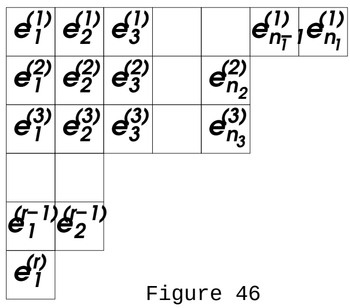

Figure 46

The format of the tableaux is determined by the **partition** of the total number n of cells (equal to $\dim \mathcal{V}$) into the sum $n_1 + \cdots + n_r$ of positive integers. Reading the same format *by columns*, we obtain another partition $n = m_1 + \cdots + m_d$, called **transposed** to the first one, where $m_1 \geq \cdots \geq m_d > 0$ are the heights of the columns. Obviously, two transposed partitions determine each other.

It follows from the way how the cells are filled with vectors $\mathbf{e}_j^{(i)}$, that the vectors in the columns 1 through k form a basis of the space $\operatorname{Ker} N^k$. Therefore

$$m_k = \dim \operatorname{Ker} N^k - \dim \operatorname{Ker} N^{k-1}, \quad k = 1, \dots, d.$$

Corollary 2. ***Consider the flag of subspaces defined by a nilpotent operator*** $N : \mathcal{V} \to \mathcal{V}$***:***

$$\operatorname{Ker} N \subset \operatorname{Ker} N^2 \subset \cdots \subset \operatorname{Ker} N^d = \mathcal{V}$$

and the partition of $n = \dim V$ ***into the summands*** $m_k = \dim \operatorname{Ker} N^k - \dim \operatorname{Ker} N^{k-1}$***. The summands of the transposed***

partition $n = n_1 + \cdots + n_r$ ***are the dimensions of the regular nilpotent blocks of*** N (described in Proposition and Corollary 1).

Corollary 3. ***The number of equivalence classes of nilpotent operators on a vector space of dimension*** n ***is equal to the number of partitions of*** n***.***

EXERCISES

392. Let $\mathcal{V}_n \subset \mathbb{K}[x]$ be the space of all polynomials of degree $< n$. Prove that the differentiation $\frac{d}{dx} : \mathcal{V}_n \to \mathcal{V}_n$ is a regular nilpotent operator.

393. Find all matrices commuting with a regular nilpotent one.

394. Is there an $n \times n$-matrix A such that $A^2 \neq 0$ but $A^3 = 0$: (a) if $n = 2$? (b) if $n = 3$?

395. Classify similarity classes of nilpotent 4×4-matrices.

396. An operator is called **unipotent**, if it differs from the identity by a nilpotent operator. Prove that the number of similarity classes of *unipotent* $n \times n$-matrices is equal to the number of partitions of n, and find this number for $n = 5$.

The Jordan Canonical Form Theorem

We proceed to classification of linear operators $A : \mathbb{C}^n \to \mathbb{C}^n$ up to similarity transformations.

Theorem. ***Every complex matrix is similar to a block-diagonal normal form with each diagonal block of the form:***

$$\begin{bmatrix} \lambda_0 & 1 & 0 & \dots & 0 \\ 0 & \lambda_0 & 1 & \dots & 0 \\ & & \dots & & \\ 0 & 0 & \dots & \lambda_0 & 1 \\ 0 & 0 & \dots & 0 & \lambda_0 \end{bmatrix}, \qquad \lambda_0 \in \mathbb{C},$$

and such a normal form is unique up to permutations of the blocks.

The block-diagonal matrices described in the theorem are called **Jordan canonical forms** (or **Jordan normal forms**). Their diagonal blocks are called **Jordan cells**.

It is instructive to analyze a Jordan canonical form before going into the proof of the theorem. The characteristic polynomial of a Jordan cell is $(\lambda - \lambda_0)^m$ where m is the size of the cell. The characteristic polynomial of a block-diagonal matrix is equal to the product of characteristic polynomials of the diagonal blocks. Therefore the characteristic polynomial of the whole Jordan canonical form is the product of factors $(\lambda - \lambda_i)^{m_i}$, one per Jordan cell. Thus the diagonal entries of Jordan cells are roots of the characteristic polynomial. After subtracting the scalar matrix $\lambda_0 I$, Jordan cells with $\lambda_i = \lambda_0$ (and only these cells) become nilpotent. Therefore the root space $\mathcal{W}_{\lambda_0}$ is exactly the direct sum of those subspaces on which the Jordan cells with $\lambda_i = \lambda_0$ operate.

`Proof of Theorem.` Everything we need has been already established in the previous two subsections.

Thanks to the Fundamental Theorem of Algebra, the characteristic polynomial $\det(\lambda I - A)$ of a complex $n \times n$-matrix A factors into the product of powers of distinct linear factors: $(\lambda - \lambda_1)^{n_1} \cdots (\lambda - \lambda_r)^{n_r}$. According to Corollary 2 of Lemma, the space $\mathbb{C}^n$ is decomposed in a canonical fashion into the direct sum $\mathcal{W}_{\lambda_1} \oplus \cdots \oplus \mathcal{W}_{\lambda_r}$ of A-invariant root subspaces. On each root subspace $\mathcal{W}_{\lambda_i}$, the operator $A - \lambda_i I$ is nilpotent. According to Proposition, the root space $\mathcal{W}_{n_i}$ is represented (in a non-canonical fashion) as the direct sum of invariant subspaces on each of which $A - \lambda_i I$ acts as a regular nilpotent operator. Since the scalar operator $\lambda_i I$ leaves every subspace invariant, this means that $\mathcal{W}_{\lambda_i}$ is decomposed into the direct sum of A-invariant subspaces, on each of which A acts as a Jordan cell with the eigenvalue $\lambda_0 = \lambda_i$. Thus, existence of a basis in which A is described by a Jordan normal form is established.

To prove uniqueness, note that the root spaces $\mathcal{W}_{\lambda_i}$ are intrinsically determined by the operator A, and the partition of $\dim \mathcal{W}_{\lambda_i}$ into the sizes of Jordan cells with the eigenvalue λ_i is uniquely determined, according to Corollary 2 of Proposition, by the geometry of the operator $A - \lambda_i I$ nilpotent on $\mathcal{W}_{\lambda_i}$. Therefore the exact structure of the Jordan normal form of A (i.e. the numbers and sizes of Jordan cells for each of the eigenvalues λ_i) is uniquely determined by A, and only the ordering of the diagonal blocks remains ambiguous. □

`Corollary 1.` ***Dimensions of root spaces*** $\mathcal{W}_{\lambda_i}$ ***coincide with multiplicities of*** λ_i ***as roots of the characteristic polynomial.***

`Corollary 2.` ***If the characteristic polynomial of a complex matrix has only simple roots, then the matrix is*** **diagonalizable*****, i.e. is similar to a diagonal matrix.***

`Corollary 3.` ***Every operator*** $A : \mathbb{C}^n \to \mathbb{C}^n$ ***in a suitable basis is described by the sum*** $D + N$ ***of two commuting ma-***

trices, of which D is diagonal, and N strictly upper triangular.

Corollary 4. ***Every operator on a complex vector space of finite dimension can be represented as the sum $D+N$ of two commuting operators, of which D is diagonalizable and N nilpotent.***

Remark. We used that $\mathbb{K}=\mathbb{C}$ only to factor the characteristic polynomial of the matrix A into linear factors. Therefore the same results hold true over any field $\mathbb{K}$ such that all non-constant polynomials from $\mathbb{K}[\lambda]$ factor into linear factors. Such fields are called **algebraically closed**. In fact (see [8]) every field $\mathbb{K}$ is contained in an algebraically closed field. Thus every linear operator $A:\mathbb{K}^n\to\mathbb{K}^n$ can be brought to a Jordan normal form by transformations $A\mapsto C^{-1}AC$, where however entries of C and scalars λ_0 in Jordan cells may belong to a larger field $\mathbb{F}\supset\mathbb{K}$.[8] We will see how this works when $\mathbb{K}=\mathbb{R}$ and $\mathbb{F}=\mathbb{C}$.

EXERCISES

397. Find Jordan normal forms of the following matrices: ✓

$$(a)\begin{bmatrix}1&2&0\\0&2&0\\-2&-2&1\end{bmatrix},\quad (b)\begin{bmatrix}4&6&0\\-3&-5&0\\-3&-6&1\end{bmatrix},\quad (c)\begin{bmatrix}13&16&16\\-5&-7&-6\\-6&-8&-7\end{bmatrix},$$

$$(d)\begin{bmatrix}3&0&8\\3&-1&-6\\-2&0&-5\end{bmatrix},\quad (e)\begin{bmatrix}-4&2&10\\-4&3&7\\-3&1&7\end{bmatrix},\quad (f)\begin{bmatrix}7&-12&-2\\3&-4&0\\-2&0&2\end{bmatrix},$$

$$(g)\begin{bmatrix}-2&8&6\\-4&10&6\\4&-8&-4\end{bmatrix},\quad (h)\begin{bmatrix}0&3&3\\-1&8&6\\2&-14&-10\end{bmatrix},\quad (i)\begin{bmatrix}1&1&-1\\-3&-3&3\\-2&-2&2\end{bmatrix},$$

$$(j)\begin{bmatrix}1&-1&2\\3&-3&6\\2&-2&4\end{bmatrix},\quad (k)\begin{bmatrix}-1&1&1\\-5&21&17\\6&-26&-21\end{bmatrix},\quad (l)\begin{bmatrix}3&7&-3\\-2&-5&2\\-4&-10&3\end{bmatrix},$$

$$(m)\begin{bmatrix}8&30&-14\\-6&-19&9\\-6&-23&11\end{bmatrix},\quad (n)\begin{bmatrix}9&22&-6\\-1&-4&1\\8&16&-5\end{bmatrix},\quad (o)\begin{bmatrix}4&5&-2\\-2&-2&1\\-1&-1&1\end{bmatrix}.$$

398. Compute powers of Jordan cells. ↯

399. Prove that if some power of a complex matrix is the identity, then the matrix is diagonalizable.

[8] For this, $\mathbb{F}$ does not have to be algebraically closed, but only needs to contain all roots of $\det(\lambda I-A)$.

400. Prove that transposed square matrices are similar.

401. Prove that $\operatorname{tr} A = \sum \lambda_i$ and $\det A = \prod \lambda_i$, where $\lambda_1, \dots, \lambda_n$ are all roots of the characteristic polynomial (repeated according to their multiplicities).

402. Prove that a square matrix satisfies its own characteristic equation; namely, if p denotes the characteristic polynomial of a matrix A, then $p(A) = 0$. (This identity is called the **Hamilton–Cayley equation**.)

The Real Case

Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be an $\mathbb{R}$-linear operator. It acts[9] on the complexification $\mathbb{C}^n$ of the real space, and commutes with the complex conjugation operator $\sigma : \mathbf{x} + i\mathbf{y} \mapsto \mathbf{x} - i\mathbf{y}$.

The characteristic polynomial $\det(\lambda I - A)$ has real coefficients, but its roots λ_i can be either real or come in pairs of complex conjugated roots (of the same multiplicity). Consequently, the complex root spaces $\mathcal{W}_{\lambda_i}$, which are defined as null spaces in $\mathbb{C}^n$ of sufficiently high powers of $A - \lambda_i I$, come in two types. If λ_i is real, then the root space is *real* in the sense that it is σ-invariant, and thus is the complexification of the real root space $\mathcal{W}_{\lambda_i} \cap \mathbb{R}^n$. If λ_i is not real, and $\bar{\lambda}_i$ is its complex conjugate, then $\mathcal{W}_{\lambda_i}$ and $\mathcal{W}_{\bar{\lambda}_i}$ are different root spaces of A, but they are transformed into each other by σ. Indeed, $\sigma A = A\sigma$, and $\sigma\lambda_i = \bar{\lambda}_i\sigma$. Therefore, if $\mathbf{z} \in \mathcal{W}_{\lambda_i}$, i.e. $(A - \lambda_i I)^d \mathbf{z} = \mathbf{0}$ for some d, then $\mathbf{0} = \sigma(A - \lambda_i I)^d \mathbf{z} = (A - \bar{\lambda}_i I)^d \sigma\mathbf{z}$, and hence $\sigma\mathbf{z} \in \mathcal{W}_{\bar{\lambda}_i}$. This allows one to obtain the following improvement for the Jordan Canonical Form Theorem applied to real matrices.

`Theorem.` ***A real linear operator*** $A : \mathbb{R}^n \to \mathbb{R}^n$ ***can be represented by the matrix in a Jordan normal form with respect to a basis of the complexified space*** $\mathbb{C}^n$ ***invariant under complex conjugation.***

`Proof.` In the process of construction bases in $\mathcal{W}_{\lambda_i}$ in which A has a Jordan normal form, we can use the following procedure. When λ_i is real, we take the real root space $\mathcal{W}_{\lambda_i} \cap \mathbb{R}^n$ and take in it a real basis in which the matrix of $A - \lambda_i I$ is block-diagonal with regular nilpotent blocks. This is possible due to Proposition applied to the case $\mathbb{K} = \mathbb{R}$. This real basis serves then as a σ-invariant complex basis in the complex root space $\mathcal{W}_{\lambda_i}$. When $\mathcal{W}_{\lambda_i}$ and $\mathcal{W}_{\bar{\lambda}_i}$ is a pair of complex conjugated root spaces, then we take a required basis in

[9]Strictly speaking, it is the complexification $A^{\mathbb{C}}$ of A that acts on $\mathbb{C}^n = (\mathbb{R}^n)^{\mathbb{C}}$, but we will denote it by the same letter A.

one of them, and then apply σ to obtain such a basis in the other. Taken together, the bases form a σ-invariant set of vectors. $\square$

Of course, for each Jordan cell with a non-real eigenvalue λ_0, there is another Jordan cell of the same size with the eigenvalue $\bar{\lambda}_0$. Moreover, if $\mathbf{e}_1, \dots, \mathbf{e}_m$ is the basis in the A-invariant subspace of the first cell, i.e. $A\mathbf{e}_k = \lambda_0 \mathbf{e}_k + \mathbf{e}_{k-1}$, $k = 2, \dots, m$, and $A\mathbf{e}_1 = \lambda_0 \mathbf{e}_1$, then the A-invariant subspace corresponding to the other cell comes with the complex conjugate basis $\bar{\mathbf{e}}_1, \dots, \bar{\mathbf{e}}_m$, where $\bar{\mathbf{e}}_k = \sigma \mathbf{e}_k$. The direct sum $\mathcal{U} := \text{Span}(\mathbf{e}_1, \dots, \mathbf{e}_m, \bar{\mathbf{e}}_1, \dots, \bar{\mathbf{e}}_m)$ of the two subspaces is both A- and σ-invariant and thus is a complexification of the real A-invariant subspace $\mathcal{U} \cap \mathbb{R}^n$. We use this to describe a real normal form for the action of A on this subspace.

Namely, let $\lambda_0 = \alpha - i\beta$, and write each basis vector $\mathbf{e}_k$ in terms of its real and imaginary part: $\mathbf{e}_k = \mathbf{u}_k - i\mathbf{v}_k$. Then the real vectors $\mathbf{u}_k$ and $\mathbf{v}_k$ form a real basis in the real part of the complex 2-dimensional space spanned by $\mathbf{e}_k$ and $\bar{e}_k = \mathbf{u}_k + i\mathbf{v}_k$. Thus, we obtain a basis $\mathbf{u}_1, \mathbf{v}_1, \mathbf{u}_2, \mathbf{v}_2, \dots, \mathbf{u}_m, \mathbf{v}_m$ in the subspace $\mathcal{U} \cap \mathbb{R}^n$. The action of A on this basis is found from the formulas:

$$A\mathbf{u}_1 + iA\mathbf{v}_1 = (\alpha - i\beta)(\mathbf{u}_1 + i\mathbf{v}_1) = (\alpha\mathbf{u}_1 + \beta\mathbf{v}_1) + i(-\beta\mathbf{u}_1 + \alpha\mathbf{v}_1),$$
$$A\mathbf{u}_k + iA\mathbf{v}_k = (\alpha\mathbf{u}_k + \beta\mathbf{v}_k + \mathbf{u}_{k-1}) + i(-\beta\mathbf{u}_k + \alpha\mathbf{v}_k + \mathbf{v}_{k-1}),\ k > 1.$$

Corollary 1. ***A linear operator*** $A : \mathbb{R}^n \to \mathbb{R}^n$ ***is represented in a suitable basis by a block-diagonal matrix with the diagonal blocks that are either Jordan cells with real eigenvalues, or have the form (***$\beta \neq 0$***):***

$$\begin{bmatrix} \alpha & -\beta & 1 & 0 & 0 & & \dots & & 0 & 0 \\ \beta & \alpha & 0 & 1 & 0 & & \dots & & 0 & 0 \\ 0 & 0 & \alpha & -\beta & 1 & 0 & & \dots & & 0 \\ 0 & 0 & \beta & \alpha & 0 & 1 & & \dots & & 0 \\ & & & & & \dots & & & & \\ 0 & & & \dots & & 0 & \alpha & -\beta & 1 & 0 \\ 0 & & & \dots & & 0 & \beta & \alpha & 0 & 1 \\ 0 & 0 & & \dots & & & 0 & 0 & \alpha & -\beta \\ 0 & 0 & & \dots & & & 0 & 0 & \beta & \alpha \end{bmatrix}.$$

Corollary 2. ***If two real matrices are related by a complex similarity transformation, then they are related by a real similarity transformation.***

Remark. The proof meant here is that if two real matrices are similar over $\mathbb{C}$ then they have the same Jordan normal form, and thus they are similar over $\mathbb{R}$ to the same real matrix, as described in Corollary 1. However, Corollary 2 can be proved directly, without a reference to the Jordan Canonical Form Theorem. Namely, if two real matrices, A and A', are related by a complex similarity transformation: $A' = C^{-1}AC$, we can rewrite this as $CA' = AC$, and taking $C = B + iD$ where B and D are real, obtain: $BA' = AB$ and $DA' = AD$. The problem now is that neither B nor D is guaranteed to be invertible. Yet, there must exist an invertible linear combination $E = \lambda B + \mu D$, for if the polynomial $\det(\lambda B + \mu D)$ of λ and μ vanishes identically, then $\det(B + iD) = 0$ too. For invertible E, we have $EA' = AE$ and hence $A' = E^{-1}AE$.

EXERCISES

403. Classify all linear operators in $\mathbb{R}^2$ up to linear changes of coordinates.

404. Consider real traceless 2×2-matrices $\begin{bmatrix} a & b \\ c & -a \end{bmatrix}$ as points in the 3-dimensional space with coordinates a, b, c. Sketch the partition of this space into similarity classes.

4 Linear Dynamical Systems

We consider here applications of Jordan Canonical Forms to dynamical systems, i.e. models of evolutionary processes. An evolutionary system can be described mathematically as a set, X, of *states* of the system, and a collection of maps $g_{t_1}^{t_2} : X \to X$ which describe the evolution of the states from the time moment t_1 to the time moment t_2, whereas the composition of $g_{t_2}^{t_3}$ with $g_{t_1}^{t_2}$ is required to coincide with $g_{t_1}^{t_3}$.

One says that the dynamical sustem is **time-independent** (or *stationary*), if the maps g_{t_1,t_2} depend only on the time increment $t = t_2 - t_1$, i.e. $g_{t_1}^{t_2} = g_0^{t_2-t_1}$. Omitting the subscript 0, we then have a family of maps $g^t : X \to X$ satisfying $g^t g^u = g^{t+u}$.

We will consider time-independent **linear dynamical systems**, i.e. such systems where the states form a vector space, $\mathcal{V}$, and the evolution maps are linear, and examine both cases: of **discrete time** or **continuous time**. In the latter case, time takes on integer values $t \in \mathbb{Z}$, and the evolution is described by **iterations** of an invertible linear map $G : \mathcal{V} \to \mathcal{V}$. Namely, if the state of the system at $t = 0$ is $\mathbf{x}(0) \in \mathcal{V}$, then the state of the system at the moment $n \in \mathbb{Z}$ is $\mathbf{x}(n) = G^n x(0)$. In the case of continuous time, a time-independent linear dynamical systems are described by a system of constant coefficients **linear** ordinary differential equations (ODE for short): $\dot{\mathbf{x}} = A\mathbf{x}$, and the evolution maps are found by solving the system.

We begin with the case of discrete time.

Iterations of Linear Maps

`Example 1:` *Fibonacci numbers.*. The sequence of integers

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \dots,$$

formed by adding the previous two terms to obtaine the next one, is well-known under the name of **Fibonacci sequence**.[10] It is the solution of the 2nd order **linear recursion relation** $f_{n+1} = f_n + f_{n-1}$, satisfying the intial conditions $f_0 = 0, f_1 = 1$. The sequence can be recast as a trajectory of a linear dynamical system on the plane as follows. Append the pair f_n, f_{n+1} of consecutive terms of the sequence into a 2-column $(x_n, y_n)^t$ and express the next pair as

[10] After Leonardo **Fibonacci** (c. 1170 – c. 1250).

a function of the previous one:

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \end{bmatrix}, \quad \text{where } \begin{bmatrix} x_n \\ y_n \end{bmatrix} := \begin{bmatrix} f_n \\ f_{n+1} \end{bmatrix}.$$

We will now derive the general formula for the numbers f_n (and for all solutions to the recursion relation).

The characteristic polynomial of the matrix $G := \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$ is $\lambda^2 - \lambda - 1$. It has two roots $\lambda_\pm = (1 \pm \sqrt{5})/2$. The corresponding eigenvectors can be taken in the form $\mathbf{v}_\pm = \begin{bmatrix} 1 \\ \lambda_\pm \end{bmatrix}$. Let $\mathbf{v}(0)$ be the vector of initial conditions (it is $(0,1)^t$ for the Fibonacci sequence), and let it be written as a linear combination of the eigenvectors: $\mathbf{v}(0) = C_+\mathbf{v}_+ + C_-\mathbf{v}_-$. Then $\mathbf{v}(n) := G^n\mathbf{v}(0) = C_+\lambda_+^n\mathbf{v}_+ + C_-\lambda_-^n\mathbf{v}_-$. Taking the first component of this vector equality, we obtain the general formula, depending on two arbitrary constants, $C_\pm$, for solutions of the recursion relation $x_{n+1} = x_n + x_{n-1}$:

$$x_n = C_+\left(\frac{1+\sqrt{5}}{2}\right)^n + C_-\left(\frac{1-\sqrt{5}}{2}\right)^n.$$

For the Fibonacci sequence *per se*, from $\lambda_+C_+ + \lambda_-C_- = 1$, $C_+ + C_- = 0$, we find $C_\pm = \pm 1/\sqrt{5}$, and hence

$$f_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right].$$

Note that $|\lambda_-| < 1$, and $|\lambda_+| > 1$. Consequently, as n increases indefinitely, the second'summand will tends to 0, and the first to infinity. Thus asymptotically (for large n) $f_n \approx [(1+\sqrt{5})/2]^n/\sqrt{5}$, while consecutive ratios f_{n+1}/f_n tend to $\lambda_+ = (1+\sqrt{5})/2$, known as **golden ratio**. In fact the fractions $3/2, 5/3, 8/5, 13/8, 21/13, 34/21, \dots$, i.e. f_{n+1}/f_n, are the best possible rational approximations to the golden ratio with the denominators not eexceeding f_n.[11]

[11]The golden ratio often occurs in Nature. For example, on a pine cone, or a pineapple fruit, count the numbers of spirals (formed by the scales near the butt) going in clockwise and in counter-clockwise directions. Most likely you'll find two consquitive Fibonacci numbers: 5 and 8, or 8 and 13. It is not surprising that such observations led to the mystical believes of our ancestors into magical properties of the golden ratio. In fact we shouldn't judge them too harshely: Inspite of some plausible scientific models, the causes for occurences of Fibonacci numbers and the golden ratio in various plants are still not entirely clear.

In general, an (invertible) linear map $G : \mathbb{R}^m \to \mathbb{R}^m$ defined a linear dynamical system: Given the initial state $\mathbf{v}(0)$, the state at the discrete time moment $n \in \mathbb{Z}$ is defined by $\mathbf{v}(n) = G^n\mathbf{v}(0)$. In order to efficiently compute the state $\mathbf{v}(n)$, we need therefore to compute powers of a linear map.

According to the general theory, there exists an invertible complex matrix C such that $G = CJC^{-1}$, where J is one of the Jordam Canonical Forms. Therefore

$$G^n = (CJC^{-1})(CJC^{-1})\dots(CJC^{-1}) = CJ^nC^{-1},$$

and the problem reduces to that for J. Recall that $J = D + N$, the sum of a diagonal matrix D (whose diagonal entries are eigenvalues of G), and of a nilpotent Jordan matrix N (in particular, $N^m = 0$), commuting with D. Thus, by the binomial formula, we have a finite sum:

$$J^n = (D+N)^n = D^n + nD^{n-1}N + \binom{n}{2}D^{n-1}N^2 + \binom{n}{3}D^{n-3}N^3 + \cdots .$$

`Example 2:` *Powers of Jordan cells.* In fact, to find explicitly the powers of J^n it suffcices to do this for each Jordan cell ceparately:

$$\begin{bmatrix} \lambda & 1 & 0 & 0 & \dots \\ 0 & \lambda & 1 & 0 & \dots \\ 0 & 0 & \lambda & 1 & \dots \\ 0 & 0 & 0 & \lambda & \dots \\ & \dots & & & \end{bmatrix}^n = \begin{bmatrix} \lambda^n & n\lambda^{n-1} & \binom{n}{2}\lambda^{n-2} & \binom{n}{3}\lambda^{n-3} & \dots \\ 0 & \lambda^n & n\lambda^{n-1} & \binom{n}{2}\lambda^{n-2} & \dots \\ 0 & 0 & \lambda^n & n\lambda^{n-1} & \dots \\ 0 & 0 & 0 & \lambda^n & \dots \\ & \dots & & & \end{bmatrix}$$

`Example 3.` Let us find the general formula for solutions of the recursive sequence $a_{n+1} = 4a_n - 6a_{n-1} + 4a_{n-2} - a_{n-3}$. Put $\mathbf{v}(n) = (a_{n-3}, a_{n-2}, a_{n-1}, a_n)^t$, Then $\mathbf{v}(n+1) = G\mathbf{v}(n)$, where

$$G = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 4 & -6 & 4 \end{bmatrix}.$$

The reader is asked to check that

$$\det(\lambda I - G) = \lambda^4 - 4\lambda^3 + 6\lambda^2 - 4\lambda + 1 = (\lambda - 1)^4.$$

Therefore G has only one eigenvalue, $\lambda = 1$, or multiplicity 4. Besides, the rank of $I - G$ is at least 3 due to the 3×3-identity matrix which occurs in the right upper corner of G. This shows that the eigenspace has dimension 1, and thus the Jordan canonical form of G has only one Jordan cell, J, of size 4, with the eigenvalue 1. Examining the matrix entries of J^n from Example 2 (with $\lambda = 1$, $m = 4$), we see that they all have the form $\binom{n}{k}$, where $k = 0, 1, 2, 3$. Note that as functions of n these binomial coefficients form a basis in the space of polynomilas of the form $C_0 + C_1 n + C_2 n^2 + C_3 n^3$. We conclude that matrix entries of $G^n = CJ^nC^{-1}$ must be, as functions of n, polynomilas of degree ≤ 3. In particular, a_n must be a polynomial of n of degree ≤ 3. Since solution sequences $\{a_n\}$ must depend on 4 arbitrary parameters a_0, a_1, a_2, a_3, and thus form a 4-dimensional space (the same as the space of polynomials of degree ≤ 3), we finally conlude that *the general solution to the recursion equation has the form* $a_n = C_0 + C_1 n + C_2 n^2 + C_3 n^3$, *where* C_i *are arbitrary constants.* Given specific values of a_0, a_1, a_2, a_3, one can find the corresponding values of C_0, C_1, C_2, C_3 by plugging $n = 0, 1, 2, 3$, and solving the system of 4 linear equations.

As it follows from Example 2 (and is illustrated by Examples 1 and 3), ***for any discrete dymanical system*** $\mathbf{v}(n+1) = G\mathbf{v}(n)$***, components of the vector trajectories*** $\mathbf{v}(n) = G^n\mathbf{v}(0)$ ***as functions of*** n ***are linear combinations of the monomials*** $\lambda_i^n n^{k-1}$***, where*** λ_i ***is a roots of the characteristic polynomial of the operator*** G***, and*** k ***does not exceed the multiplicity of this root.*** In particular, if all roots are simple, the egneral solution has the form $\mathbf{v}(n) = \sum_{i=1}^m C_i \lambda_i^n \mathbf{v}_i$, where $\mathbf{v}_i$ is an eigenvector corresponding to the eigenvalue λ_i, and C_i are arbitrary constants.

EXERCISES

405. Find the general formula for the sequence a_n such that $a_0 = 2$, $a_1 = 1$, and $a_{n+1} = a_n + a_{n-1}$ for all $n \geq 1$.

406. Find all solutions to the recursion equation $a_{n+1} = 5a_n + 6a_{n-1}$.

407. For the recursion equation $a_{n+1} = 6a_n - 9a_{n-1}$, express the general solutions as a function of n, a_0 and a_1. ✓

408. Given two sequences $\{x_n\}, \{y_n\}$ satisfying $x_{n+1} = 4x_n + 3y_n$, $y_{n+1} = 3x_n + 2y_n$, and such that $x_0 = y_0 = 13$. Find the limit of the ratio x_n/y_n as n tends to: (a) $+\infty$; (b) $-\infty$. ✓

409. Prove that for a given $G : \mathcal{V} \to \mathcal{V}$, all trajectories $\{\mathbf{x}(n)\}$ of a given dynamical system $\mathbf{x}(n+1) = G\mathbf{x}(n)$ form a linear subspace of dimension $\dim \mathcal{V}$ in the space of all vector-valued sequences.

410.* For integer $n > 0$, prove that $\left(\frac{3+\sqrt{17}}{2}\right)^n + \left(\frac{3-\sqrt{17}}{2}\right)^n$ is an odd integer.

Linear ODE Systems

Let

$$\begin{aligned} \dot{x}_1 &= a_{11}x_1 + ... + a_{1n}x_n \\ &... \\ \dot{x}_2 &= a_{n1}x_1 + ... + a_{nn}x_n \end{aligned}$$

be a linear homogeneous system of ordinary differential equations with constant (possibly complex) coefficients a_{ij}. It can be written in the matrix form as $\dot{\mathbf{x}} = A\mathbf{x}$.

Consider the infinite matrix series

$$e^{tA} := I + tA + \frac{t^2A^2}{2} + \frac{t^3A^3}{6} + ... + \frac{t^kA^k}{k!} + ...$$

If M is an upper bound for the absolute values of the entries of A, then the matrix entries of $t^kA^k/k!$ are bounded by $n^kt^kM^k$. It is easy to deduce from this that the series converges (at least as fast as the series for e^{ntM}).

`Proposition.` ***The solution to the system*** $\dot{\mathbf{x}} = A\mathbf{x}$ ***with the initial condition*** $\mathbf{x}(0)$ ***is given by the formula*** $\mathbf{x}(t) = e^{tA}\mathbf{x}(0)$***.***

`Proof.` Differentiating the series $\sum_0^\infty t^kA^k/k!$ we find

$$\frac{d}{dt}e^{tA} = \sum_{k=1}^{\infty} \frac{t^{k-1}A^k}{(k-1)!} = \sum_{k=0}^{\infty} \frac{t^kA^{k+1}}{k!} = Ae^{tA}$$

and hence $\frac{d}{dt}e^{tA}\mathbf{x}(0) = A(e^{tA}\mathbf{x}(0))$. Thus $\mathbf{x}(t)$ satisfies the ODE system. At $t = 0$ we have $e^{0A}\mathbf{x}(0) = I\mathbf{x}(0) = \mathbf{x}(0)$ and therefore the initial condition is also satisfied. □

`Remark.` This Proposition defins the exponential fucntion $A \mapsto e^A$ of a matrix (or of a linear operator). The exponential fucntion satisfies $e^Ae^B = e^{A+B}$ ***provided that*** A ***and*** B ***commute***. This can be proved the same way as this is done in the supplement *Complex Numbers* for the exponential function of the complex argument.

The proposition reduces the problem of solving the ODE system $\dot{\mathbf{x}} = A\mathbf{x}$ to computation of the exponential function e^{tA} of a matrix. Notice that if $A = CBC^{-1}$ then

$$A^k = CBC^{-1}CBC^{-1}CBC^{-1}... = CBBB...C^{-1} = CB^kC^{-1}$$

and therefore $\exp(tA) = C^{-1}\exp(tB)C$. This observation reduces computation of e^{tA} to that of e^{tB} where the Jordan normal form of A can be taken on the role of B.

`Example 4.` Let Λ be a diagonal matrix with the diagonal entries $\lambda_1, ..., \lambda_n$. Then Λ^k is a diagonal matrix with the diagonal entries $\lambda_1^k, ..., \lambda_n^k$ and hence

$$e^{t\Lambda} = I + t\Lambda + \frac{t^2}{2}\Lambda^2 + ... = \begin{bmatrix} e^{\lambda_1 t} & 0 & ... \\ & ... & \\ ... & 0 & e^{\lambda_n t} \end{bmatrix}.$$

`Example 5.` Let N be a nilpotent Jordan cell of size m. We have (for $m = 4$):

$$N = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad N^2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad N^3 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and $N^4 = 0$. Generalizing to arbitrary m, we find:

$$e^{tN} = I + tN + \frac{t^2}{2}N^2 + \frac{t^3}{6}N^3 + ... = \begin{bmatrix} 1 & t & \frac{t^2}{2} & ... & \frac{t^{m-1}}{(m-1)!} \\ 0 & 1 & t & \frac{t^2}{2} & ... \\ 0 & 0 & 1 & t & ... \\ & & ... & & \\ & ... & & 0 & 1 \end{bmatrix}.$$

Let $\lambda I + N$ be the Jordan cell of size m with the eigenvalue λ. Then $e^{t(\lambda I+N)} = e^{t\lambda I}e^{tN} = e^{\lambda t}e^{tN}$. Here we use the multiplicative property of the matrix exponential function, valid since I and N commute.

Finally, let $A = \begin{bmatrix} B & \mathbf{0} \\ \mathbf{0} & D \end{bmatrix}$ be a block-diagonal square matrix. Then $A^k = \begin{bmatrix} B^k & \mathbf{0} \\ \mathbf{0} & D^k \end{bmatrix}$ and respectively $e^{tA} = \begin{bmatrix} e^{tB} & \mathbf{0} \\ \mathbf{0} & e^{tD} \end{bmatrix}$. Together with Examples 4 and 5, this shows how to compute the exponential function e^{tJ} for any Jordan normal form J: each Jordan cell has the form $\lambda I + N$ and should be replaced by $e^{\lambda t}e^{tN}$. Since any square matrix A can be reduced to one of the Jordan normal matrices J by similarity transformations, we conclude that $e^{tA} = Ce^{tJ}C^{-1}$ with suitable invertible C.

Applying Example 3 to ODE systems $\dot{\mathbf{x}} = A\mathbf{x}$ we arrive at the following conclusion. Suppose that the characteristic polynomial of A has n distinct roots $\lambda_1, ..., \lambda_n$. Let C be the matrix whose columns are the corresponding n complex eigenvectors $\mathbf{v}_1, ..., \mathbf{v}_n$. Then the solution with the initial condition $\mathbf{x}(0)$ is given by the formula

$$\mathbf{x}(t) = C \begin{bmatrix} e^{\lambda_1 t} & 0 & ... \\ & ... & \\ ... & 0 & e^{\lambda_n t} \end{bmatrix} C^{-1}\mathbf{x}(0).$$

Notice that $C^{-1}\mathbf{x}(0)$ here (as well as $\mathbf{x}(0)$) is a column $\mathbf{c} = (c_1, ..., c_n)^t$ of arbitrary constants, while the columns of $Ce^{t\Lambda}$ are $e^{\lambda_i t}\mathbf{v}_i$. We conclude that the general solution formula reads

$$\mathbf{x}(t) = c_1 e^{\lambda_1 t}\mathbf{v}_1 + ... + c_n e^{\lambda_n t}\mathbf{v}_n.$$

This formula involves eigenvalues λ_i of A, the corresponding eigenvectors $\mathbf{v}_i$, and the arbitrary complex constants c_i, to be found from the system of linear equations $\mathbf{x}(0) = c_1\mathbf{v}_1 + ... + c_n\mathbf{v}_n$.

`Example 6.` The ODE system

$$\begin{aligned} \dot{x}_1 &= 2x_1 + x_2 \\ \dot{x}_2 &= x_1 + 3x_2 - x_3 \\ \dot{x}_3 &= 2x_2 + 3x_3 - x_1 \end{aligned} \quad \text{has the matrix } A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & -1 \\ -1 & 2 & 3 \end{bmatrix}.$$

The characteristic polynomial of A is $\lambda^3 - 8\lambda^2 + 22\lambda - 20$. It has a real root $\lambda_0 = 2$, and factors as $(\lambda - 2)(\lambda^2 - 6\lambda + 10)$. Thus is has two complex conjugate roots $\lambda_\pm = 3 \pm i$. The eigenvectors $\mathbf{v}_0 = (1, 0, 1)^t$ and $\mathbf{v}_\pm = (1, 1 \pm i, 2 \mp i)^t$ are found from the systems of linear equations:

$$\begin{aligned} 2x_1 + x_2 &= 2x_1 \\ x_1 + 3x_2 - x_3 &= 2x_2 \\ -x_1 + 2x_2 + 3x_3 &= 2x_3 \end{aligned} \quad \text{and} \quad \begin{aligned} 2x_1 + x_2 &= (3 \pm i)x_1 \\ x_1 + 3x_2 - x_3 &= (3 \pm i)x_2 \\ -x_1 + 2x_2 + 3x_3 &= (3 \pm i)x_3 \end{aligned}.$$

Thus the general complex solution is a linear combination of the solutions

$$e^{2t}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \; e^{(3+i)t}\begin{bmatrix} 1 \\ 1+i \\ 2-i \end{bmatrix}, \; e^{(3-i)t}\begin{bmatrix} 1 \\ 1-i \\ 2+i \end{bmatrix}$$

with arbitrary complex coefficients. Real solutions can be extracted from the complex ones by taking their real and imaginary parts:

$$e^{3t}\begin{bmatrix} \mathbf{cos}\, t \\ \mathbf{cos}\, t - \mathbf{sin}\, t \\ 2\,\mathbf{cos}\, t + \mathbf{sin}\, t \end{bmatrix},\ e^{3t}\begin{bmatrix} \mathbf{sin}\, t \\ \mathbf{cos}\, t + \mathbf{sin}\, t \\ 2\,\mathbf{sin}\, t - \mathbf{cos}\, t \end{bmatrix}.$$

Thus the general real solution is described by the formulas

$$\begin{array}{rcl} x_1(t) &=& c_1e^t + c_2e^{3t}\,\mathbf{cos}\, t + c_3e^{3t}\,\mathbf{sin}\, t \\ x_2(t) &=& c_2e^{3t}(\mathbf{cos}\, t - \mathbf{sin}\, t) + c_3e^{3t}(2\,\mathbf{cos}\, t + \mathbf{sin}\, t) \\ x_3(t) &=& c_1e^t + c_2e^{3t}(\mathbf{cos}\, t + \mathbf{sin}\, t) + c_3e^{3t}(2\,\mathbf{sin}\, t - \mathbf{cos}\, t) \end{array},$$

where c_1, c_2, c_3 are arbitrary *real* constants. At $t = 0$ we have

$$x_1(0) = c_1 + c_2,\ x_2(0) = c_2 + 2c_3,\ x_3(0) = c_1 + c_2 - c_3.$$

Given the initial values $(x_1(0), x_2(0), x_3(0))$ the corresponding constants c_1, c_2, c_3 can be found from this system of linear algebraic equations.

In general, even if the characteristic polynomial of A has multiple roots, our theory (and Examples 4, 5) show that ***all components of solutions to the ODE system*** $\dot{\mathbf{x}} = A\mathbf{x}$ ***are expressible as linear combinations of functions*** $t^ke^{\lambda t}$, $t^ke^{at}\,\mathbf{cos}\, bt$, $t^ke^{at}\,\mathbf{sin}\, bt$***, where*** λ ***are real eigenvalues,*** $a \pm ib$ ***are complex eigenvalues, and*** $k = 0, 1, 2, ...$ ***is to be smaller than the multiplicity of the corresponding eigenvalue.*** This observation suggests to approach the ODE systems with multiple eigenvalues in the following way avoiding explicit similarity transformation to the Jordan normal form: look for the general solution in the form of linear combinations of these functions with arbitrary coefficients by substituting them into the equations, and find the relations between the arbitrary constants from the resulting system of linear algebraic equations.

`Example 7.` The ODE system

$$\begin{array}{rcl} \dot{x}_1 &=& 2x_1 + x_2 + x_3 \\ \dot{x}_2 &=& -3x_1 - 2x_2 - 3x_3 \\ \dot{x}_3 &=& 2x_1 + 2x_2 + 3x_3 \end{array} \quad \text{has the matrix } A = \begin{bmatrix} 2 & 1 & 1 \\ -3 & -2 & -3 \\ 2 & 2 & 3 \end{bmatrix}$$

with the characteristic polynomial $\lambda^3 - 3\lambda^2 + 3\lambda - 1 = (\lambda - 1)^3$. Thus we can look for solutions in the form

$$x_1 = e^t(a_1 + b_1t + c_1t^2),\ x_2 = e^t(a_2 + b_2t + c_2t^2),\ x_3 = e^t(a_3 + b_3t + c_3t^2).$$

Substituting into the ODE system, and introducing the notation $\sum a_i =: A$, $\sum b_i =: B$, $\sum c_i =: C$, we get (omitting the factor e^t):

$$\begin{aligned}(a_1+b_1)+(b_1+2c_1)t+c_1t^2 &= (a_1+A)+(b_1+B)t+(c_1+C)t^2,\\ (a_2+b_2)+(b_2+2c_2)t+c_2t^2 &= -(3A-a_2)-(3B-b_2)t-(3C-c_2)t^2,\\ (a_3+b_3)+(b_3+2c_3)t+c_3t^2 &= (2A+a_3)+(2B+b_3)t+(2C+c_3)t^2.\end{aligned}$$

Introducing the notation $P := A+Bt+Ct^2$, we rewrite the system of 9 linear equations in 9 unknowns in the form

$$b_1+2c_1t = P,\ b_2+2c_2t = -3P,\ b_3+2c_3t = 2P.$$

This yields $b_1 = A, b_2 = -3A, b_3 = 2A$, and (since $B = A-3A+2A = 0$), we find $c_1 = c_2 = c_3 = 0$. Thus, the general solution, depending on 3 arbitrary constants $a_1, a-2, a_3$, reads:

$$\begin{aligned}x_1 &= e^t(a_1+(a_1+a_2+a_3)t)\\ x_2 &= e^t(a_2-3(a_1+a_2+a_3)t)\\ x_3 &= e^t(a_3+2(a_1+a_2+a_3)t).\end{aligned}$$

In particular, since t^2 does not occur in the formulas, we can conclude that the Jordan form of our matrix has only Jordan cells of size 1 or 2 (and hence one of each, for the total size must be 3):

$$J = \begin{bmatrix} 1 & 1 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}.$$

EXERCISES

411.* Give an example of (non-commuting) 2×2 matrices A and B for which $e^Ae^B \neq e^{A+B}$. ↯

412. Solve the following ODE systems. Find the solution satisfying the initial condition $x_1(0) = 1,\ x_2(0) = 0,\ x_3(0) = 0$:

(a) $\lambda_1 = 1$
$\dot{x}_1 = 3x_1 - x_2 + x_3$
$\dot{x}_2 = x_1 + x_2 + x_3$
$\dot{x}_3 = 4x_1 - x_2 + 4x_3$

(b) $\lambda_1 = 1$
$\dot{x}_1 = -3x_1 + 4x_2 - 2x_3$
$\dot{x}_2 = x_1 + x_3$
$\dot{x}_3 = 6x_1 - 6x_2 + 5x_3$

(c) $\lambda_1 = 1$
$\dot{x}_1 = x_1 - x_2 - x_3$
$\dot{x}_2 = x_1 + x_2$
$\dot{x}_3 = 3x_1 + x_3$

(d) $\lambda_1 = 2$
$\dot{x}_1 = 4x_1 - x_2 - x_3$
$\dot{x}_2 = x_1 + 2x_2 - x_3$
$\dot{x}_3 = x_1 - x_2 + 2x_3$

(e) $\lambda_1 = 1$
$\dot{x}_1 = -x_1 + x_2 - 2x_3$
$\dot{x}_2 = 4x_1 + x_2$
$\dot{x}_3 = 2x_1 + x_2 - x_3$

$$\begin{array}{ll} (f)\ \ \lambda_1 = 2 & (g)\ \ \lambda_1 = 1 \\ \dot{x}_1 = 4x_1 - x_2 & \dot{x}_1 = 2x_1 - x_2 - x_3 \\ \dot{x}_2 = 3x_1 + x_2 - x_3 & \dot{x}_2 = 2x_1 - x_2 - 2x_3 \\ \dot{x}_3 = x_1 + x_3 & \dot{x}_3 = -x_1 + x_2 + 2x_3 \end{array}.$$

413. Find the Jordan canonical forms of the 3×3-matrices of the ODE systems (a–g) from the previous exercise.

414.* Prove the following identity: $\det e^A = e^{\operatorname{tr} A}$. ↯

Higher Order Linear ODE

A linear homogeneous constant coefficient n-th order ODE

$$\frac{d^n}{dt^n}x + a_1\frac{d^{n-1}}{dt^{n-1}}x + ... + a_{n-1}\frac{d}{dt}x + a_n x = 0$$

can be rewritten, by introducing the notations

$$x_1 := x,\ x_2 := \dot{x},\ x_2 := \ddot{x}, ..., x_{n-1} := d^{n-1}x/dt^{n-1},$$

as a system $\dot{\mathbf{x}} = A\mathbf{x}$ of n first order ODE with the matrix

$$A = \left[\begin{array}{ccccc} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & 0 & ... \\ & & ... & & \\ 0 & ... & 0 & 0 & 1 \\ -a_n & -a_{n-1} & ... & -a_2 & -a_1 \end{array}\right],$$

Then our theory of linear ODE system applies. There are however some simplifications which are due to the special form of the matrix A. First, computing the characteristic polynomial of A we find

$$\det(\lambda I - A) = \lambda^n + a_1\lambda^{n-1} + ... + a_{n-1}\lambda + a_n.$$

Thus the polynomial can be easily read off the high order differential equation. Next, let $\lambda_1, ..., \lambda_r$ be the roots of the characteristic polynomial, and $m_1, ..., m_r$ their multiplicities ($m_1 + ... + m_r = n$). Then it is clear that the solutions $x(t) = x_1(t)$ must have the form

$$e^{\lambda_1 t}P_1(t) + ... + e^{\lambda_r t}P_r(t),$$

where $P_i = a_0 + a_1 t + ... + a_{m_i-1}t^{m_i-1}$ is a polynomial of degree $< m_i$. The total number of arbitrary coefficients in these polynomials equals $m_1 + ... + m_r = n$. On the other hand, the general solution to the n-th order ODE must depend on n arbitrary initial values $(x(0), \dot{x}(0), ..., x^{(n-1)}(0))$. This can be justified by the general Existence and Uniqueness Theorem for solutions of ODE systems. We conclude that ***each of the functions***

$$e^{\lambda_1 t},\ e^{\lambda_1 t}t,\ ...,\ e^{\lambda_1 t}t^{m_1-1},\ ...,\ e^{\lambda_r t},\ e^{\lambda_r t}t,\ ...,\ e^{\lambda_r t}t^{m_r-1}$$

must satisfy our differential equation, and any (complex) solution is uniquely written as a linear combination of these functions with suitable (complex) coefficients. (In other words, these functions form a *basis* of the space of complex solutions to the differential equation.)

`Example 8:` $x^{(xii)} - 3x^{(viii)} + 3x^{(iv)} - x = 0$ has the characteristic polynomial $\lambda^{12} - 3\lambda^8 + 3\lambda^4 - 1 = (\lambda-1)^3(\lambda+1)^3(\lambda-i)^3(\lambda+i)^3$. The following 12 functions form therefore a complex basis of the solution space:

$$e^t,\ te^t,\ t^2e^t,\ e^{-t},\ te^{-t},\ t^2e^{-t},\ e^{it},\ te^{it},\ t^2e^{it},\ e^{-it},\ te^{-it},\ t^2e^{-it}.$$

Of course, a basis of the space of real solutions is obtained by taking real and imaginary parts of the complex basis:

$$e^t,\ te^t,\ t^2e^t,\ e^{-t},\ te^{-t},\ t^2e^{-t},$$
$$\cos t,\ \sin t,\ t\cos t,\ t\sin t,\ t^2\cos t,\ t^2\sin t.$$

`Remark.` The fact that the functions $e^{\lambda_i t}t^k,\ k < m_i,\ i = 1, ..., r$, form a basis of the space of solutions to the differential equation with the characteristic polynomial $(\lambda - \lambda_1)^{m_1}...(\lambda - \lambda_r)^{m_r}$ is not hard to check directly, without a reference to linear algebra and Existence and Uniqueness Theorem. However, it is useful to understand how this property of the equation is related to the Jordan structure of the corresponding matrix A. In fact the Jordan normal form of the matrix A consists of exactly r Jordan cells — one cell of size m_i for each eigenvalue λ_i. This simplification can be explained as follows. For every λ, the matrix $\lambda I - A$ has rank $n - 1$ *at least* (due to the presence of the $(n-1) \times (n-1)$ identity submatrix in the right upper corner of A). This guarantees that the eigenspaces of A have dimension 1 *at most*, and hence that A cannot have more than one Jordan cell corresponding to the same root of the characteristic polynomial. Using this property, the reader can check now that the formulation of the Jordan Theorem in terms of differential equations given in the Introduction is indeed equivalent to the matrix formulation given in the previous Section.

EXERCISES

415. Solve the following ODE, and find the solution satisfying the initial condition $x(0) = 1, \dot{x}(0) = 0, ..., x^{(n-1)}(0) = 0$: $\quad (a)\ \ x^{(3)} - 8x = 0,$

$$(b)\ \ x^{(4)} + 4x = 0, \quad (c)\ \ x^{(6)} + 64x = 0, \quad (d)\ \ x^{(5)} - 10x^{(3)} + 9x = 0,$$
$$(e)\ \ x^{(3)} - 3\dot{x} - 2x = 0, \quad (f)\ \ x^{(5)} - 6x^{(4)} + x^{(3)} = 0,$$
$$(g)\ \ x^{(5)} + 8x^{(3)} + 16\dot{x} = 0, \quad (h)\ \ x^{(4)} + 4\ddot{x} + 3x = 0.$$

416. Rewrite the ODE system

$$\begin{array}{c} \ddot{x}_1 + 4\dot{x}_1 - 2x_1 - 2\dot{x}_2 - x_2 = 0 \\ \ddot{x}_1 - 4\dot{x}_1 - \ddot{x}_2 + 2\dot{x}_2 + 2x_2 = 0 \end{array}$$

of two 2-nd order equations in the form of a linear ODE system $\dot{\mathbf{x}} = A\mathbf{x}$ of four 1-st order equations and solve it.

417. Show that every polynomial $\lambda^n + a_1\lambda^{n-1} + \cdots + a_n$ is the charactersitic polynomial of an $n \times n$-matrix.