March 26, 2018

Preface

Exercises in the collectiont of all exercises are included for one of the following reasons:

- to allow to practice course material,
- to give applications of the course material,
- to provide more details on results discussed in the lectures,
- to give details on side remarks in the lectures,
- to provide details on results that are useful for the final assignment.

For each lecture you are supposed to do a few selected exercises from this collection.

Most chapters (titled 'Lecture n - ...' in this collection) start with a brief introduction. This text is included in order to give a context for the exercises, i.e., to settle notation, to provide a brief review of the required theory and a motivation for the exercises. The text is not meant as an easy introduction to the theory; for this, please consult the lectures, the transparencies or text books. Nevertheless, if you understand the theory, the text may provide a convenient summary.

The exercises in Lecture 0.
The set of exercises in this preliminary chapter forms an overview of the Linear Algebra material from bachelor courses, material that will be used during this Numerical Linear Algebra course (in the lectures, exercises, assignments or final report) and that is supposed to be known. This set also fixes the notation. A few items, as Schur decomposition, might not have been discussed (or only briefly) in a standard bachelor course. They are included since this overview seems to be the appropriate place for them. But, they will be introduced properly in the lectures, when needed.

Lecture 0 – Preliminaries

Scalars in $\mathbb{C}$ and $\mathbb{R}$ are denoted by lower Greek letters, as λ.

High dimensional vectors and matrices are denoted by bold face letters, lower case letters are used for vectors and capitals for matrices. If, for instance, n is large (high), then $\mathbf{x}, \mathbf{y}, \ldots$ are vectors in $\mathbb{C}^n$ (or $\mathbb{R}^n$) and $\mathbf{A}, \mathbf{V}, \ldots$ are $n \times k$ matrices. Low dimensional vectors and matrices are denoted by standard letters: $x, y, \ldots$ or $\vec{x}, \vec{y}, \ldots$ are k-vectors for small (low) k, $A, S, \ldots$ are $k \times \ell$ matrices, with ℓ small as well. In many of our applications, $n \in \mathbb{N}$ will be large, and $k \in \mathbb{N}$ will be modest.[1]

Spaces are denoted with calligraphic capitals, as $\mathcal{V}$.

We view an n-vector as a **column vector**, that is, as an $n \times 1$ matrix. Our notation is column vector oriented, that is, we denote **row vectors** ($1 \times n$ matrices) as $\mathbf{x}^*$, with $\mathbf{x}$ a column vector.

(i.e., all matrix entries are in $\mathbb{R}$),

Let $\mathbf{A} = (A_{ij})$ be an $n \times k$ matrix: $\mathbf{A} = (A_{ij})$ indicates that A_{ij} is the (i, j)-entry of $\mathbf{A}$. The matrix $\mathbf{A}$ is said te be **real** if all entries $A_{ij} \in \mathbb{R}$. With $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k]$ or $\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_k]$ we settle the notation for the columns of $\mathbf{A}$: the jth column equals $\mathbf{a}_j$. The absolute value and the complex conjugate are entry-wise operations: $|\mathbf{A}| \equiv (|A_{ij}|)$ and $\bar{\mathbf{A}} \equiv (\bar{A}_{ij})$. The **transpose** $\mathbf{A}^{\mathrm{T}}$ of the matrix $\mathbf{A}$ is the $k \times n$ matrix with (i, j)-entry A_{ji}: $\mathbf{A}^{\mathrm{T}} \equiv (A_{ji})$. $\mathbf{A}^{\mathrm{H}}$ is the **adjoint** or **Hermitian conjugate** of $\mathbf{A}$: $\mathbf{A}^{\mathrm{H}} \equiv \bar{\mathbf{A}}^{\mathrm{T}}$. We will also use the notation $\mathbf{A}^*$ for $\mathbf{A}^{\mathrm{H}}$: $\mathbf{A}^* = \mathbf{A}^{\mathrm{H}}$.[2]

We follow MATLAB's notation to describe matrices that are formed from other matrices: consider an $n \times k$ matrix $\mathbf{A} = (A_{ij})$ and an $m \times l$ matrix $\mathbf{B} = (B_{ij})$. If $m = n$, then $[\mathbf{A}, \mathbf{B}]$ is the $n \times (k + l)$ matrix with (i, j) entry equal to $A_{i,j}$ if $j \leq k$ and $B_{i,j-k}$ if $j > k$: $\mathbf{A}$ is extended with the columns from $\mathbf{B}$. If $k = l$, then $[\mathbf{A}; \mathbf{B}]$ is the $(n + m) \times k$ matrix with (i, j) entry equal to $A_{i,j}$ if $i \leq n$ and $B_{i-n,j}$ if $i > n$: $\mathbf{A}$ is extended with the rows from $\mathbf{B}$. Note that $[\mathbf{A}; \mathbf{B}] = [\mathbf{A}^{\mathrm{T}}\ \mathbf{B}^{\mathrm{T}}]^{\mathrm{T}}$. If $I = (i_1, i_2, \ldots, i_p)$ is a sequence of numbers $i_r \in \{1, 2, \ldots, n\}$ and $J = (j_1, j_2, \ldots, j_q)$ is a sequence of numbers j_s in $\{1, 2, \ldots, k\}$, then $\mathbf{A}(I, J)$ is the $p \times q$ matrix with (r, s) entry equal to A_{i_r, j_s}. Note that entries of $\mathbf{A}$ can be used more than once.

Below, we collect a number of standard results in Linear Algebra that will be frequently used. The statements are left to the reader as an exercise.

A Spaces

Let $\mathcal{V}$ and $\mathcal{W}$ be linear subspace of $\mathbb{C}^n$.
Then $\mathcal{V} + \mathcal{W}$ is the subspace $\mathcal{V} + \mathcal{W} \equiv \{\mathbf{x} + \mathbf{y} \,\big|\, \mathbf{x} \in \mathcal{V}, \mathbf{y} \in \mathcal{W}\}$.
We put $\mathcal{V} \oplus \mathcal{W}$ for the subspace $\mathcal{V} + \mathcal{W}$ if $\mathcal{V} \cap \mathcal{W} = \{\mathbf{0}\}$.

---

[1] We distinguish high and low dimensionality to indicate differences in efficiency. A dimension k is 'low', if the solution of k-dimensional problems of a type that we want to solve numerically can be computed in a split second with a computer and standard software. The dimension is 'high' if more computational time is required or non-standard software has to be used. For **linear system**s, that is,

solve $Ax = b$ for x, where A is a given $k \times k$ matrix and b is a given k-vector,

k small is like $k \leq 1000$. For **eigenvalue problem**s, that is,

find a non-trivial vector x and a scalar λ such that $Ax = \lambda x$, where A a given $k \times k$ matrix,

k small is like $k \leq 100$. From a pure mathematical point of view 'low' and 'high' dimensionality does not have a meaning (in pure mathematics, 'low' would mean finite, while 'high' would be infinitely dimensional. The problems that we will solve are all finite dimensional). In a mathematical statement the difference between low and high dimensionality does not play a role. But in its interpretation for practical use, it does.

[2] Formally, $\mathbf{A}^*$ is defined with respect to inner products: if $(\cdot, \cdot)_X$ and $(\cdot, \cdot)_Y$ are inner product on a linear space $\mathcal{X}$ and on a linear space $\mathcal{Y}$, respectively, and $\mathbf{A}$ linearly maps $\mathcal{X}$ to $\mathcal{Y}$, then $\mathbf{A}^*$ is the linear map from $\mathcal{Y}$ to $\mathcal{X}$ for which $(\mathbf{A}\mathbf{x}, \mathbf{y})_Y = (\mathbf{x}, \mathbf{A}^*\mathbf{y})_X$ for all $\mathbf{x} \in \mathcal{X}$ and $\mathbf{y} \in \mathcal{Y}$. With respect to the standard inner product $(x, y) \equiv y^{\mathrm{H}} x$ on $\mathcal{X} \equiv \mathbb{C}^k$ and on $(\mathbf{x}, \mathbf{y}) \equiv \mathbf{y}^{\mathrm{H}}\mathbf{x}$ on $\mathcal{Y} \equiv \mathbb{C}^n$, we have that $\mathbf{A}^* = \mathbf{A}^{\mathrm{H}}$. With $\mathbf{A}^*$, we will (implicitly) refer to standard inner product, unless explicitly stated otherwise.

Exercise 0.1.

(a) $\mathcal{V} + \mathcal{W}$ is a linear subspace.

(b) Suppose $\mathcal{V} \cap \mathcal{W} = \{\mathbf{0}\}$. Then $\dim(\mathcal{V}) + \dim(\mathcal{W}) = \dim(\mathcal{V} \oplus \mathcal{W})$

(c) Suppose $\mathcal{V} \cap \mathcal{W} = \{\mathbf{0}\}$. Then $\mathcal{V} \oplus \mathcal{W} = \mathbb{C}^n$ if and only if $\dim(\mathcal{V}) + \dim(\mathcal{W}) = n$.

(d) If $\dim(\mathcal{V}) + \dim(\mathcal{W}) > n$, then $\mathcal{V} \cap \mathcal{W} \neq \{\mathbf{0}\}$.

If $\mathbf{x}$ and $\mathbf{y}$ are n-vectors (i.e., in $\mathbb{C}^n$), then we put $\|\mathbf{x}\|_2 \equiv \sqrt{\mathbf{x}^*\mathbf{x}}$ and $\mathbf{y} \perp \mathbf{x}$ if $\mathbf{y}^*\mathbf{x} = 0$.

Exercise 0.2.

(a) The map $(\mathbf{x}, \mathbf{y}) \rightsquigarrow \mathbf{y}^*\mathbf{x}$ from $\mathbb{C}^n \times \mathbb{C}^n$ to $\mathbb{C}$ defines an **inner product** on $\mathbb{C}^n$:
   1) $\mathbf{x}^*\mathbf{x} \geq 0$ and $\mathbf{x}^*\mathbf{x} = 0$ if and only if $\mathbf{x} = \mathbf{0}$ $\quad (\mathbf{x} \in \mathbb{C}^n)$,
   2) $\mathbf{x} \rightsquigarrow \mathbf{y}^*\mathbf{x}$ is a linear map from $\mathbb{C}^n$ to $\mathbb{C}$ for all $\mathbf{y} \in \mathbb{C}^n$,
   3) $(\mathbf{y}^*\mathbf{x})^- = \mathbf{x}^*\mathbf{y}$ $\quad (\mathbf{x}, \mathbf{y} \in \mathbb{C}^n)$.

(b) The map $\mathbf{x} \rightsquigarrow \|\mathbf{x}\|_2$ from $\mathbb{C}^n$ to $\mathbb{C}$ defines an **norm** on $\mathbb{C}^n$:
   1) $\|\mathbf{x}\|_2 \geq 0$ and $\|\mathbf{x}\|_2 = 0$ if and only if $\mathbf{x} = \mathbf{0}$ $\quad (\mathbf{x} \in \mathbb{C}^n)$,
   2) $\|\alpha\mathbf{x}\|_2 = |\alpha| \, \|\mathbf{x}\|_2$ $\quad (\alpha \in \mathbb{C}, \mathbf{x} \in \mathbb{C}^n)$,
   3) $\|\mathbf{x} + \mathbf{y}\|_2 \leq \|\mathbf{x}\|_2 + \|\mathbf{y}\|_2$ $\quad (\mathbf{x}, \mathbf{y} \in \mathbb{C}^n)$.

(c) $|\mathbf{y}^*\mathbf{x}| \leq \|\mathbf{x}\|_2 \, \|\mathbf{y}\|_2$ $\quad (\mathbf{x}, \mathbf{y} \in \mathbb{C}^n)$ (**Cauchy–Schwartz**).

(d) If $\mathbf{x} \perp \mathbf{y}$ then $\|\mathbf{x} + \mathbf{y}\|_2^2 = \|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2$ $\quad (\mathbf{x}, \mathbf{y} \in \mathbb{C}^n)$ (**Pythagoras**).

We put
$\mathbf{v} \perp \mathcal{W}$ if $\mathbf{v} \perp \mathbf{w}$ $(\mathbf{w} \in \mathcal{W})$, $\quad \mathcal{V} \perp \mathcal{W}$ if $\mathbf{v} \perp \mathbf{W}$ $(\mathbf{v} \in \mathcal{V})$, and $\mathcal{V}^\perp \equiv \{\mathbf{y} \in \mathbb{C}^n \,\big|\, \mathbf{y} \perp \mathcal{V}\}$.
Let $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_k]$ be a $n \times k$ matrix with columns $\mathbf{v}_1, \ldots, \mathbf{v}_k$. Then

$$\mathrm{span}(\mathbf{V}) \equiv \mathrm{span}(\mathbf{v}_1, \ldots, \mathbf{v}_k) \equiv \left\{ \sum_{j=1}^{k} \alpha_j \mathbf{v}_j \,\Big|\, \alpha_j \in \mathbb{C} \right\}.$$

We put $\mathbf{x} \perp \mathbf{V}$ if $\mathbf{x} \perp \mathrm{span}(\mathbf{V})$. Moreover, $\mathbf{V}^\perp \equiv \{\mathbf{y} \in \mathbb{C}^n \,\big|\, \mathbf{y} \perp \mathbf{V}\}$.

Exercise 0.3.

(a) $\dim(\mathcal{V}) = n - \dim(\mathcal{V}^\perp)$.

(b) $\mathbf{x} \perp \mathbf{V} \quad \Leftrightarrow \quad \mathbf{x} \perp \mathbf{v}_i$ for all $i = 1, \ldots, k \quad \Leftrightarrow \quad \mathbf{V}^*\mathbf{x} = 0$.

(c) $\dim(\mathrm{span}(\mathbf{V})) \leq k$.

The angle $\angle(\mathbf{x}, \mathbf{y})$ between two non-trivial n-vectors $\mathbf{x}$ and $\mathbf{y}$ is in $[0, \frac{1}{2}\pi]$ such that

$$\cos\angle(\mathbf{x}, \mathbf{y}) = \frac{|\mathbf{y}^*\mathbf{x}|}{\|\mathbf{y}\|_2 \, \|\mathbf{x}\|_2}.$$

B Matrices.

Let $\mathbf{A} = (A_{ij})$ be an $n \times k$ matrix. We will view the matrix $\mathbf{A}$ as map from $\mathbb{C}^k$ to $\mathbb{C}^n$ defined by the matrix-vector multiplication: $x \rightsquigarrow \mathbf{A}x$ $(x \in \mathbb{C}^k)$. Note that $\mathbf{A}$ is real if and only of matrix-vector multiplication maps $\mathbb{R}^k$ to $\mathbb{R}^n$.

The **column (row) rank** of $\mathbf{A}$ is the maximum number of linearly independent columns (rows) of the matrix $\mathbf{A}$.

Theorem 0.1 *The row rank of a matrix is equal to the column rank.*

The above theorem allows us to talk about the **rank** of a matrix.

The **range** $\mathcal{R}(\mathbf{A})$ of $\mathbf{A}$ is $\{\mathbf{A}y \,\big|\, y \in \mathbb{C}^k\}$.
The **null space** $\mathcal{N}(\mathbf{A})$ or **kernel** of $\mathbf{A}$ is $\{x \in \mathbb{C}^k \,\big|\, \mathbf{A}x = \mathbf{0}\}$.

Theorem 0.2 $\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{A}^*)^\perp$.

Exercise 0.4.

(a) $\mathcal{R}(\mathbf{A}) = \text{span}(\mathbf{A})$.

(b) the rank of $\mathbf{A}$ equals $\dim(\mathcal{R}(\mathbf{A}))$.

(c) Prove Theorem 0.2: $\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{A}^*)^\perp$.

(d) $\dim(\mathcal{R}(\mathbf{A})) = n - \dim(\mathcal{N}(\mathbf{A}))$.

Exercise 0.5.

(a) $\mathbb{A} : \mathbb{C}^k \to \mathbb{C}^n$ is a linear map $\quad\Leftrightarrow$

for some $n \times k$ matrix $\mathbf{A}$ we have that $\mathbb{A}(x) = \mathbf{A}x$ for all $x \in \mathbb{C}^k$.

$\mathbf{A}$ is the matrix of $\mathbb{A}$ with respect to the standard basis in $\mathbb{C}^k$ and $\mathbb{C}^n$.

(b) Let $v_1, \ldots, v_k$ be a basis of $\mathbb{C}^k$ and $\mathbf{w}_1, \ldots, \mathbf{w}_n$ a basis of $\mathbb{C}^n$. Let $V \equiv [v_1, \ldots, v_k]$ and $\mathbf{W} \equiv [\mathbf{w}_1, \ldots, \mathbf{w}_n]$. Then V and $\mathbf{W}$ are non-singular and $\mathbf{W}^{-1}\mathbf{A}V$ is the matrix of the map $x \rightsquigarrow \mathbf{A}x$ from $\mathbb{C}^k$ to $\mathbb{C}^n$ with respect to the V and $\mathbf{W}$ basis.

Exercise 0.6. Let $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_k]$ be an $n \times k$ matrix and $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_k]$ and $m \times k$ matrix. Let $D \equiv \text{diag}(\lambda_1, \ldots, \lambda_k)$ be an $k \times k$ diagonal matrix with diagonal entries λ_j.

(a) $\mathbf{A}^{**} = \mathbf{A}$.

(b) $(\mathbf{B}\mathbf{A}^*)^* = \mathbf{A}\mathbf{B}^*$.

(c) If $x = (x_1, \ldots, x_k)^{\mathrm{T}}$, then $\mathbf{A}x = \sum_{j=1}^k \mathbf{a}_j x_j = (\sum_{j=1}^k \mathbf{a}_j e_j^*)x$, whence $\mathbf{A} = \sum_{j=1}^k \mathbf{a}_j e_j^*$. Here, e_j is the jth standard basis vector in $\mathbb{C}^k$ ($e_1 \equiv (1, 0, 0, \ldots, 0)^{\mathrm{T}}$, $e_2 \equiv (0, 1, 0, \ldots, 0)^{\mathrm{T}}$, etc.).

(d) $\mathbf{A}\mathbf{B}^* = \sum_{j=1}^k \mathbf{a}_j \mathbf{b}_j^*$.

(e) $\mathbf{a}_j \mathbf{b}_j^*$ are $n \times m$ rank one matrices.

(f) $\mathbf{A}D\mathbf{B}^* = \sum_{j=1}^k \lambda_j \mathbf{a}_j \mathbf{b}_j^*$.

Exercise 0.7. Let the $n \times n$ matrix $\mathbf{U} = (u_{ij})$ be **upper triangular**, i.e., $u_{ij} = 0$ if $i > j$.

(a) $\mathbf{U}^{-1}$ is upper triangular and $\mathbf{U}^*$ is lower triangular.

(b) If in addition the diagonal of $\mathbf{U}$ is the identity matrix, then the diagonal of $\mathbf{U}^{-1}$ is the identity matrix as well.

(c) The product of upper triangular matrices is upper triangular as well.

If $\mathbf{A}$ is an $n \times n$ matrix, then the **determinant** $\det(\mathbf{A})$ is the volume of the 'block' $\{\mathbf{A}\mathbf{x} \,\big|\, \mathbf{x} = (x_1, \ldots, x_n)^{\mathrm{T}}, x_i \in [0, 1]\}$. The **trace** $\text{trace}(\mathbf{A})$ of $\mathbf{A}$ is the sum of its diagonal entries.

Theorem 0.3 *If* $\mathbf{A}$ *is* $n \times k$ *and* $\mathbf{B}$ *is* $k \times n$, *then* $\text{trace}(\mathbf{A}\mathbf{B}) = \text{trace}(\mathbf{B}\mathbf{A})$.
If $n = k$, *then* $\det(\mathbf{A}\mathbf{B}) = \det(\mathbf{A})\det(\mathbf{B})$.

Exercise 0.8. Let $\mathbf{A}$ be an $n \times n$ matrix.

(a) Prove that the following properties are equivalent:

- $\det(\mathbf{A}) \neq 0$.
- $\mathbf{A}$ had full rank.
- $\mathbf{A}$ has a trivial null space: $\mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}$.
- The range of $\mathbf{A}$ is $\mathbb{C}^n$: $\mathcal{R}(\mathbf{A}) = \mathbb{C}^n$.
- $\mathbf{A} : \mathbb{C}^n \to \mathbb{C}^n$ is invertible.
- There is an $n \times n$ matrix, denoted by $\mathbf{A}^{-1}$, for which $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$.

$\mathbf{A}$ is **non-singular** if $\mathbf{A}$ has one of these properties. $\mathbf{A}^{-1}$ is the **inverse** of $\mathbf{A}$.

(b) $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. If $\mathbf{B}$ is $n \times n$ and $\mathbf{B}\mathbf{A} = \mathbf{I}$ or $\mathbf{A}\mathbf{B} = \mathbf{I}$, then $\mathbf{B} = \mathbf{A}^{-1}$.

(c) With Cramer's rule, the inverse of a matrix can be expressed in terms of determinants of submatrices. However, this approach for finding inverses is extremely inefficient and, except for very low dimensions, it is never used in practice. Cramer's rule for $n = 2$:

$$\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}^{-1} = \frac{1}{\alpha\delta - \beta\gamma} \begin{bmatrix} \delta & -\beta \\ -\gamma & \alpha \end{bmatrix}.$$

Exercise 0.9.

(a) Let $\mathbf{A}$, $\mathbf{L}$ and $\mathbf{U}$ be $n \times n$ matrices such that $\mathbf{A} = \mathbf{L}\mathbf{U}$, $\mathbf{L}$ lower triangular with diagonal $\mathbf{I}$ and $\mathbf{U}$ upper triangular. Let μ_j be the (j, j)-entry of $\mathbf{U}$. $\det(\mathbf{A}) = \det(\mathbf{U}) = \mu_1 \cdot \ldots \cdot \mu_n$.

Exercise 0.10. Let $\mathbf{A}$ be an $n \times n$ non-singular matrix.

(a) Prove that $(\mathbf{A}^{\mathrm{T}})^{-1} = (\mathbf{A}^{-1})^{\mathrm{T}}$ and $(\mathbf{A}^{\mathrm{H}})^{-1} = (\mathbf{A}^{-1})^{\mathrm{H}}$.
We will put $\mathbf{A}^{-\mathrm{T}}$ instead of $(\mathbf{A}^{\mathrm{T}})^{-1}$ and $\mathbf{A}^{-\mathrm{H}}$ instead of $(\mathbf{A}^{\mathrm{H}})^{-1}$.

C Orthonormal matrices.

$\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_k]$ is **orthogonal** if $\mathbf{v}_i \perp \mathbf{v}_j$ for all $i, j = 1, \ldots, k$, $i \neq j$.
If $\mathbf{V}$ is orthogonal and, in addition, $\|\mathbf{v}_j\|_2 = 1$ $(j = 1, \ldots, k)$, then $\mathbf{V}$ is **orthonormal**.
In some textbooks, $\mathbf{V}$ is called orthogonal if multiplication by $\mathbf{V}$ preserves orthogonality (see Exercise 1.10.C).

If $\mathbf{V}$ is square and orthonormal, then is said to be **unitary**.

To be able to 'work' with a linear (sub)space $\mathcal{V}$ a basis is needed. For stability reasons, it is preferred to have an orthonormal basis, $\mathbf{v}_1, \ldots, \mathbf{v}_k$, say. Then, $\mathbf{V} \equiv [\mathbf{v}_1, \ldots, \mathbf{v}_k]$ is an orthonormal matrix. Conversely, the columns of an orthonormal matrix $\mathbf{V}$ form an orthonormal basis (of $\text{span}(\mathbf{V})$). Orthonormal basis of $\mathbb{C}^n$ correspond to unitary matrices.

Exercise 0.11. Let $\mathbf{V}$ be an $n \times k$ matrix.

(a) If $\mathbf{V}$ is orthonormal, then $k = \dim(\text{span}(\mathbf{V}))$. Note that $k \leq n$.

(b) $\mathbf{V}$ is orthonormal $\Leftrightarrow$ $\mathbf{V}^*\mathbf{V} = I_k$ the $k \times k$ identity matrix.

(c) If, with $m \leq k$, W is an $k \times m$ orthonormal matrix, then $\mathbf{V}W$ is an $n \times m$ orthonormal matrix. In particular the product of unitary matrices of the same dimension is unitary.

Let $\mathbf{a}_1, \ldots, \mathbf{a}_k$ be non-trivial n-vectors.
The **Gram–Schmidt process** in ALG. 0.1 (see also Exercise 0.12(a)) constructs orthonormal n-vectors $\mathbf{q}_1, \ldots, \mathbf{q}_\ell$ that span the same space as $\mathbf{a}_1, \ldots, \mathbf{a}_k$. The $\mathbf{q}_j$ form the columns of an $n \times \ell$ orthonormal matrix $\mathbf{Q}$. Note that $\ell \leq k$ and $\ell \leq n$, while $\ell < k$ only if the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are linearly dependent. Let R be the $\ell \times k$ matrix with ij entry r_{ij} as computed in the algorithm and 0 if not computed. Then $\mathbf{A} = \mathbf{Q}R$. The following theorem highlights this result.

Theorem 0.4 *Let* $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_k]$ *be an* $n \times k$ *matrix.*
Let $\mathbf{Q}$ *and* R *be as produced by the Gram–Schmidt process applied to the columns of* $\mathbf{A}$.
Then $\mathbf{Q}$ *is orthonormal,* $\text{span}(\mathbf{A}) = \text{span}(\mathbf{Q})$, R *is upper triangular, and* $\mathbf{A} = \mathbf{Q}R$.

Exercise 0.12. ***Proof of Theorem*** **0.4.** Let $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_k]$ be an $n \times k$ matrix.

(a) Suppose $\mathbf{q}_1, \ldots, \mathbf{q}_\ell$ is an orthonormal system, $\ell < k$. For $\mathbf{a}_j \in \mathbb{C}^n$, consider

$$r_{ij} = \mathbf{q}_i^* \mathbf{a}_j \quad (i = 1, \ldots, \ell), \quad \mathbf{v} = \mathbf{a}_j - \sum_{j=1}^{\ell} \mathbf{q}_i \, r_{ij}, \tag{0.1}$$

```
Gram–Schmidt orthonormalisation
r11 = ||a1||2,  q1 = a1/r11,  ℓ = 1.
for j = 2,...,k
  Orthogonalise:
   v = aj
   for i = 1,...,ℓ
     rij = qi* aj,  v ← v − qi rij
   end for
  Normalise:
   rℓ+1,j = ||v||2
   If  rℓ+1,j ≠ 0
     ℓ ← ℓ+1,  qℓ = v/rℓj
   end if
end for
```

Algorithm 0.1. The Gram–Schmidt process constructs an orthonormal basis $\mathbf{q}_1, \ldots, \mathbf{q}_\ell$ for the space spanned by $\mathbf{a}_1, \ldots, \mathbf{a}_k$. Here $\leftarrow$ indicates that the new quantity replaces the old one. If $\mathbf{a}_j$ is in the span of $\mathbf{a}_1, \ldots, \mathbf{a}_{j-1}$, then, $\mathbf{a}_j$ is in the span of $\mathbf{q}_1, \ldots, \mathbf{q}_{\ell-1}$, $r_{\ell\, j} = 0$ and no new orthonormal vector $\mathbf{q}_\ell$ is formed. If the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are linearly independent then ℓ at the end of each loop equals j.

and, if $\|\mathbf{v}\|_2 \neq 0$,

$$r_{\ell+1,j} = \|\mathbf{v}\|_2, \quad \mathbf{q}_{\ell+1} \equiv \frac{\mathbf{v}}{r_{\ell+1,j}}. \tag{0.2}$$

Then, $\mathbf{q}_{\ell+1} \perp \mathrm{span}(\mathbf{q}_1, \ldots, \mathbf{q}_\ell)$, and

$$\mathbf{a}_j = \sum_{i=1}^{\ell+1} \mathbf{q}_i\, r_{ij} = \mathbf{Q}_{\ell+1}\, r_j,$$

where $\mathbf{Q}_{\ell+1} = [\,\mathbf{q}_1, \ldots, \mathbf{q}_{\ell+1}\,]$ and $r_j \in \mathbb{C}^{\ell+1}$ has ith entry r_{ij} as described above in (0.1) and (0.2). In particular, $\mathbf{Q}_{\ell+1}$ is orthonormal and $\mathbf{a}_j \in \mathrm{span}(\mathbf{Q}_{\ell+1})$.

In (0.1), the vector $\mathbf{a}_j$ is **orthogonalised against** $\mathbf{q}_1, \ldots, \mathbf{q}_\ell$, while in (0.2) the vector $\mathbf{v}$ is **normalised**.

(b) Show that (0.1) can be expressed as

$$\mathbf{v} = \mathbf{a}_j - \mathbf{Q}_\ell(\mathbf{Q}_\ell^* \mathbf{a}_j), \tag{0.3}$$

(c) If $\|\mathbf{v}\|_2 = 0$, then $\mathbf{a} = \mathbf{Q}_\ell r'_j$, where r'_j is the ℓ upper part of r_j.

(d) Prove Theorem 0.4: there is an $n \times \ell$ orthonormal matrix $\mathbf{Q}$, with $\ell \leq \min(k, n)$, and an $\ell \times k$ upper triangular matrix R such that

$$\mathbf{A} = \mathbf{Q}R. \tag{0.4}$$

(e) There is an $n \times n$ unitary matrix $\widetilde{\mathbf{Q}}$ and an $n \times k$ upper triangular matrix $\widetilde{\mathbf{R}}$ such that

$$\mathbf{A} = \widetilde{\mathbf{Q}}\widetilde{\mathbf{R}} \tag{0.5}$$

(f) Relate $\mathbf{Q}$ and $\widetilde{\mathbf{Q}}$ and R and $\widetilde{\mathbf{R}}$.

The relation in (0.5) is the **QR-decomposition** or **QR-factorisation** of $\mathbf{A}$. The relation in (0.4) is the **economical** form of the QR-decomposition.

Plane rotations and basic reflections are simple instances of unitary matrices: a two-dimensional plane rotation is given by

$$\begin{bmatrix} c & -s \\ s & c \end{bmatrix} \quad \text{with} \quad c \in [-1,+1],\ s \in \mathbb{C},\ c^2 + |s|^2 = 1. \tag{0.6}$$

Note that c and s can be viewed as $c = \cos(\phi)$ and $s = \sin(\phi)\,\zeta$ for some $\phi \in \mathbb{R}$ and some **sign** ζ, that is, $\zeta \in \mathbb{C}$, $|\zeta| = 1$. An $n \times n$ plane rotation $\mathbf{G}$, also called **Givens rotations**, rotates in some (i,j)-plane, that is, $\mathbf{G}$ is the $n \times n$ identity except at the four entries (i,i), (i,j), (j,i), (j,j), where it is a two dimensional plane rotation. A basic reflection is a diagonal matrix with all diagonal entries $+1$, except for one that equals some sign.

Since the product of $n \times n$ unitary matrices is unitary, the product of plane rotations and basic reflections is unitary. The converse is true as well: any unitary matrix $\mathbf{Q}$ can be expressed as a product of Givens rotations and one basic reflections. Hence (discarding the one reflection), the columns of a unitary matrix $\mathbf{Q}$ can be viewed as a 'rotated' basis of $\mathbb{C}^n$, i.e., as a rotated version of the standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$.

Theorem 0.5 *Let $\mathcal{V}$ be a k-dimensional linear subspace of $\mathbb{C}^n$. Let $\mathbf{b} \in \mathbb{C}^n$.*
For a $\mathbf{b}_0 \in \mathcal{V}$, the following two properties are equivalent:

(i) $\|\mathbf{b} - \mathbf{b}_0\|_2 \leq \|\mathbf{b} - \mathbf{v}\|_2$ *for all* $\mathbf{v} \in \mathcal{V}$.

(ii) $\mathbf{b} - \mathbf{b}_0 \perp \mathcal{V}$.

There is exactly one $\mathbf{b}_0 \in \mathcal{V}$ with one of these equivalent properties.

Exercise 0.13. Let $\mathcal{V}$ be a k-dimensional linear subspace of $\mathbb{C}^n$. Let $\mathbf{b} \in \mathbb{C}^n$.

(a) There is an $n \times k$ orthonormal matrix $\mathbf{V}$ such that $\mathcal{V} = \text{span}\,(\mathbf{V})$.

(b) We have that $\mathbf{b}_0 \equiv \mathbf{V}(\mathbf{V}^*\mathbf{b}) \in \mathcal{V}$ and $\mathbf{b} - \mathbf{b}_0 \perp \mathcal{V}$.

(c) If $\mathbf{x} = \mathbf{y} + \mathbf{z}$ for some $\mathbf{y} \in \mathcal{V}$ and $\mathbf{z} \perp \mathcal{V}$, then $\mathbf{y} = \mathbf{x}_0 \equiv \mathbf{V}(\mathbf{V}^*\mathbf{x})$.

(d) $\mathbb{C}^n = \mathcal{V} \oplus \mathcal{V}^\perp$.

(e) Prove Theorem 0.5.

Exercise 0.14. Let $\mathbf{A}$ be an $n \times k$ matrix.

(a) $\mathcal{R}(\mathbf{A}) = \{\mathbf{A}x \,\big|\, x \perp \mathcal{N}(\mathbf{A})\}$.

(b) For an $x \in \mathbb{C}^k$, let $x_1 \in \mathbb{C}^k$ be such that $x_1 \perp \mathcal{N}(\mathbf{A})$ and $x - x_1 \in \mathcal{N}(\mathbf{A})$.
There is precisely one $k \times n$ matrix, denoted by $\mathbf{A}^\dagger$, for which

$$\mathbf{A}^\dagger\mathbf{y} = 0 \ \text{ if } \ \mathbf{y} \perp \mathcal{R}(\mathbf{A}) \quad \text{and} \quad \mathbf{A}^\dagger(\mathbf{A}\mathbf{x}) = x_1 \quad (x \in \mathbb{C}^k).$$

$\mathbf{A}^\dagger$ is the inverse of $\mathbf{A}$ as a map from $\mathcal{N}(\mathbf{A})^\perp$ to $\mathcal{R}(\mathbf{A})$ with null-space equal to $\mathcal{R}(\mathbf{A})^\perp$.
$\mathbf{A}^\dagger$ is the **Moore–Penrose pseudo inverse** or **generalised inverse** of $\mathbf{A}$.

(c) The following four properties do not involve the notion of orthogonality. They characterise the Moore–Penrose pseudo inverse.

$$\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}, \quad \mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger, \quad (\mathbf{A}\mathbf{A}^\dagger)^* = \mathbf{A}\mathbf{A}^\dagger, \quad (\mathbf{A}^\dagger\mathbf{A})^* = \mathbf{A}^\dagger\mathbf{A}.$$

D Eigenvalues.

Let $\mathbf{A}$ be an $n \times n$ matrix. Let $\lambda \in \mathbb{C}$.
If $\mathbf{x} \in \mathbb{C}^n$, then $(\lambda, \mathbf{x})$ is an **eigenpair** of the matrix $\mathbf{A}$ if $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ and $\mathbf{x} \neq \mathbf{0}$, λ is an **eigenvalue** and $\mathbf{x}$ is an **eigenvector** associated to the eigenvalue λ.
$\mathcal{V}(\lambda) \equiv \{\mathbf{x} \in \mathbb{C}^n \,\big|\, \mathbf{A}\mathbf{x} = \lambda\mathbf{x}\}$ is the **eigenspace associated to** λ. The dimension of $\mathcal{V}(\lambda)$ is the

geometric multiplicity of the eigenvalue λ.
The **characteristic polynomial** P_A is defined by

$$P_A(\zeta) \equiv \det(\zeta\mathbf{I} - \mathbf{A}) \qquad (\zeta \in \mathbb{C}).$$

Exercise 0.15.

(a) $\lambda \in \mathbb{C}$ is an eigenvalue of $\mathbf{A}$ if and only if λ is a root of P_A, i.e., $P_A(\lambda) = 0$.

(b) If P_A has k mutually different complex roots, then $\mathbf{A}$ has at least k eigenvalues.

(c) If $\mathbf{A}$ is real, and $(\lambda, \mathbf{x})$ is an eigenpair of $\mathbf{A}$, then $(\bar{\lambda}, \bar{\mathbf{x}})$ is an eigenpair of $\mathbf{A}$.

The **algebraic multiplicity** of the eigenvalue λ is the multiplicity of the root λ of P_A. λ is a **simple eigenvalue** of $\mathbf{A}$ if its algebraic multiplicity is one. An eigenvalue λ of $\mathbf{A}$ is **semi-simple** if the algebraic multiplicity equals the geometric multiplicity. The matrix $\mathbf{A}$ is **semi-simple** if all of its eigenvalues are semi-simple. If all eigenvalues are simple, then $\mathbf{A}$ is said to be **simple**.

Exercise 0.16.

(a) Any simple eigenvalue is semi-simple.

(b) **Counted according to** algebraic **multiplicity**,[3] $\mathbf{A}$ has n eigenvalues.

(c) Give an example of a 2×2 matrix with an eigenvalue with algebraic multiplicity 2 and geometric multiplicity 1.

(d) For any $n \times n$ matrix $\mathbf{B}$, the two matrices $\mathbf{AB}$ and $\mathbf{BA}$ have the same eigenvalues with equal multiplicity (algebraic, as well as geometric).
The same statement also holds for the non-zero eigenvalues in case $\mathbf{A}$ is $n \times k$ and $\mathbf{B}$ is $k \times n$.

(e) Actually, whether (λ, x), with $\lambda \in \mathbb{C}$ and $x \in \mathbb{C}^n$, is an eigenpair depends on the *linear map* $\mathbb{A} : \mathbb{C}^n \to \mathbb{C}^n$, $\mathbb{A}\,x = \lambda\,x$, and not on the basis that is used to describe vectors in the space. In particular, the value of an eigenvalue does not change if another basis is selected, nor the eigenvector. However, the representation of the vector does change. More specific, if $\mathbf{T}$ is a non-singular $n \times n$ matrix, then $\mathbf{A}$ and $\widetilde{\mathbf{A}} \equiv \mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ (that is, the matrix $\mathbf{A}$ of $\mathbb{A}$ with respect to the standard basis, and the matrix $\widetilde{\mathbf{A}}$ of $\mathbb{A}$ with respect to the basis of columns of $\mathbf{T}$) have the same eigenvalues with equal multiplicity (algebraic, as well as geometric). If $\mathbf{x}$ represents x with respect to the standard basis, then $\widetilde{\mathbf{x}} \equiv \mathbf{T}^{-1}\mathbf{x}$ represents x with respect to the basis of columns of $\mathbf{T}$:

$$\mathbb{A}\,x = \lambda\,x \quad \Leftrightarrow \quad \mathbf{A}\mathbf{x} = \lambda\,\mathbf{x} \quad \Leftrightarrow \quad \widetilde{\mathbf{A}}\widetilde{\mathbf{x}} = \lambda\,\widetilde{\mathbf{x}}.$$

(f) If $\mathbf{A} = (A_{ij})$ is a triangular matrix, the diagonal elements $A_{11}, A_{22}, \ldots, A_{nn}$ are the eigenvalues of $\mathbf{A}$ counted according to multiplicity.

(g) Any non-trivial linear subspace $\mathcal{V}$ of $\mathbb{C}^n$ that is **invariant under multiplication** by $\mathbf{A}$ (i.e., $\mathbf{A}\mathbf{x} \in \mathcal{V}$ for all $\mathbf{x} \in \mathcal{V}$) contains at least one eigenvector of $\mathbf{A}$.

(h) $\mathcal{V}(\lambda) \subset \mathcal{W}(\lambda) \equiv \{\mathbf{w} \in \mathbb{C}^n \,\Big|\, (\mathbf{A} - \lambda\mathbf{I})^k\mathbf{w} = \mathbf{0} \text{ for some } k \in \mathbb{N}\}$

(i) Both $\mathcal{V}(\lambda)$ and $\mathcal{W}(\lambda)$ are linear subspaces of $\mathbb{C}^n$ invariant under multiplication by $\mathbf{A}$.

(j) The dimension of $\mathcal{W}(\lambda)$ equals the algebraic multiplicity of the eigenvalue λ.

(k) To simplify notation, assume 0 is an eigenvalue of $\mathbf{A}$ (otherwise, replace $\mathbf{A}$ by $\mathbf{A} - \lambda\mathbf{I}$). Let $\mathbf{x}$ be a non-trivial vector in $\mathcal{W}(0)$. Let $k \in \mathbb{N}$ be the smallest number for which $\mathbf{A}^k\mathbf{x} = \mathbf{0}$. Assume $\alpha_m\mathbf{A}^m\mathbf{x} + \ldots + \alpha_1\mathbf{A}\mathbf{x} + \alpha_0\mathbf{x} = \mathbf{0}$ for some $\alpha_j \in \mathbb{C}$. Prove that $\alpha_0 = \ldots = \alpha_{k-1} = 0$. Prove that $\mathbf{x} \in \mathcal{W}(\mu) \quad \Leftrightarrow \quad \mu = 0$. In particular, $\mathcal{W}(\lambda) \cap \mathcal{W}(\mu) = \{\mathbf{0}\}$ if $\lambda \neq \mu$.

(l) $\mathbb{C}^n = \bigoplus \mathcal{W}(\lambda)$, where we sum over all different eigenvalues λ of $\mathbf{A}$.

[3] that is, if an eigenvalue has algebraic multiplicity k then this eigenvalue is counted k times. Below, if list the eigenvalues 'counted according to multiplicity', we mean that an eigenvalue with algebraic multiplicity k is listed k times.

If $\mathbf{Q}$ is $n \times k$ orthonormal with $k \leq n$ and S is $k \times k$ upper triangular such that

$$\mathbf{AQ} = \mathbf{Q}S, \tag{0.7}$$

then (0.7) is a **partial Schur decomposition** (or **partial Schur form**) of $\mathbf{A}$ (of order k). If $k = n$, then (0.7) is a **Schur decomposition** of **Schur form**.

Note that, if (0.7) is a partial Schur decomposition, then the space $\mathcal{Q}$ spanned by the columns $\mathbf{q}_1, \ldots, \mathbf{q}_k$ is invariant under multiplication by $\mathbf{A}$ and, restricted to $\mathcal{Q}$, S is the matrix of $\mathbf{A}$ with respect to the basis of $\mathbf{q}_i$. In particular, if $Sy = \lambda\, y$ and $\mathbf{x} \equiv Q\, y$ than $\mathbf{Ax} = \lambda\, \mathbf{x}$ and the diagonal elements of S are k eigenvalues of $\mathbf{A}$.

Exercise 0.17. Suppose we have a partial Schur decomposition (0.7).

(a) The diagonal entries of S are eigenvalues of S and of $\mathbf{A}$.

(b) If $Sy = \lambda y$, then $(\lambda, \mathbf{Q}y)$ is an eigenpair of $\mathbf{A}$

(c) The computation of y with $Sy = \lambda y$ requires the solution of an upper triangular system.

Theorem 0.6 *Let $\mathbf{A}$ be an $n \times n$ matrix.*
1) $\mathbf{A}$ has a Schur decomposition, say $\mathbf{A} = \mathbf{QSQ}^*$. Put $\lambda_i \equiv S_{ii}$ $(i = 1, \ldots, n)$.
2) $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of $\mathbf{A}$ (and of $\mathbf{S}$) counted according to multiplicity. In particular,

$$\det(\mathbf{A}) = \det(\mathbf{S}) = \prod_{j=1}^{n} \lambda_j \quad \textit{and} \quad \operatorname{trace}(\mathbf{A}) = \operatorname{trace}(\mathbf{S}) = \sum_{j=1}^{n} \lambda_j.$$

3) For any given ordering $\lambda_1, \ldots, \lambda_n$ of the eigenvalues of $\mathbf{A}$, counted according to multiplicity, there is a Schur decomposition $\mathbf{A} = \mathbf{QSQ}^*$ such that $S_{ii} = \lambda_i$ for $i = 1, \ldots, n$.

Proof. We leave the proof of 2) as an exercise. For the proof of 3), we refer to Lecture 13.

We apply induction to prove 1) of the theorem.

There is a normalised eigenvector $\mathbf{q}_1$ of $\mathbf{A}$ (cf., Exercise 0.16.(g)). Note that $\mathbf{Aq}_1 = \mathbf{q}_1\lambda_1$ is a partial Schur decomposition of order 1.

Suppose we have a partial Schur decomposition $\mathbf{AQ}_k = \mathbf{Q}_k S_k$ of order k. Note that $\mathbf{Q}_k^\perp$ is a linear subspace of $\mathbb{C}^n$ that is invariant under multiplication by the **deflated** matrix $\widetilde{\mathbf{A}} \equiv (\mathbf{I} - \mathbf{Q}_k\mathbf{Q}_k^*)\mathbf{A}(\mathbf{I} - \mathbf{Q}_k\mathbf{Q}_k^*)$. Therefore (again by (g) of Exercise 0.16), $\widetilde{\mathbf{A}}$ has a normalised eigenvector in $\mathbf{Q}_k^\perp$, say $\mathbf{q}_{k+1}$ with eigenvalue, say λ_{k+1}. Expanding $\mathbf{Q}_k$ to $\mathbf{Q}_{k+1}$ and S_k to S_{k+1},

$$\mathbf{Q}_{k+1} \equiv [\mathbf{Q}_k, \mathbf{q}_{k+1}] \quad \text{and} \quad S_{k+1} \equiv \begin{bmatrix} S_k & \mathbf{Q}_k^*\mathbf{A}\mathbf{q}_{k+1} \\ \vec{0}^* & \lambda_{k+1} \end{bmatrix},$$

leads to the partial Schur decomposition $\mathbf{AQ}_{k+1} = \mathbf{Q}_{k+1}S_{k+1}$ of order $k+1$. □

As observed in Exercise 0.17, eigenvectors of S, whence of $\mathbf{S}$, lead to eigenvectors of $\mathbf{A}$. In particular if all eigenvalues are semi-simple, then there is an upper triangular matrix $\mathbf{X}$ with diagonal $\mathbf{I}$ such that $\mathbf{SX} = \mathbf{X}\Lambda$ with Λ a diagonal matrix with the λ_i on the diagonal: the columns of $\mathbf{X}$ are eigenvectors of $\mathbf{S}$. Moreover, the columns of $\mathbf{QX}$ are eigenvectors of $\mathbf{A}$.

Let $\mathbb{A} : \mathbb{C}^n \to \mathbb{C}^n$ be the linear map with matrix $\mathbf{A}$ w.r.t. the standard basis. Note that $\mathbf{S}$ can be viewed as the matrix of $\mathbb{A}$ with respect to the 'rotated basis' $\mathbf{q}_1, \ldots, \mathbf{q}_n$ of columns of $\mathbf{Q}$: the existence of a Schur decomposition reveals that, with respect to some appropriate (map dependent) orthonormal basis, any linear map with image space equal to domain space can be represented as an upper-triangular matrix. The following theorem states that there exists a basis that gives an even simpler matrix (Jordan form), and in many cases even a diagonal one (if $\mathbf{A}$ is semi-simple). However, for stability reason, the Schur form is often preferred in practise. Moreover, many theoretical results can be as easily proved using a Schur form as using a Jordan form. In practise, the Jordan form is computed from the Schur form.

Without proof, we mention:

Theorem 0.7 *There is a non-singular* $n \times n$ *matrix* $\mathbf{T}$ *such that* $\mathbf{AT} = \mathbf{TJ}$*, where* $\mathbf{J}$ *is a matrix on* **Jordan normal form***, i.e.,* $\mathbf{J}$ *is a block diagonal matrix with Jordan blocks on the diagonal. A* **Jordan block** *is a square matrix of the form* $J_\lambda = \begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}$.

$\mathbf{A}$ is **diagonalizable** if $\mathbf{J}$ is diagonal (i.e., all Jordan blocks in $\mathbf{J}$ are 1×1).

Theorem 0.8 *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix.*
The following two statements are equivalent for any eigenvalue λ *of* $\mathbf{A}$*:*
i) λ *is not semi-simple.*
ii) $\mathbf{Ay} = \lambda\mathbf{y} + \mathbf{x}$ *for some vector* $\mathbf{y} \neq \mathbf{0}$ *and some eigenvector* $\mathbf{x}$ *with eigenvalue* λ*.*
The following three properties are equivalent for $\mathbf{A}$*:*
1) $\mathbf{A}$ *is semi-simple.*
2) $\mathbf{A}$ *is diagonalizable.*
3) *There is a basis of eigenvector of* $\mathbf{A}$*, i.e., there is a basis* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *of* $\mathbb{C}^n$ *such that* $\mathbf{v}_i$ *is an eigenvector of* $\mathbf{A}$ *for all* i*.*

Exercise 0.18. ***Proof of Theorem*** **0.8.**
(a) If an eigenvalue λ of $\mathbf{A}$ shows up in exactly p Jordan blocks in the Jordan normal form, then p is the geometric multiplicity of λ.
(b) Suppose $\mathbf{J}$ is on Jordan normal form. Describe $\mathcal{V}(\lambda)$ and $\mathcal{W}(\lambda)$ in terms of the standard basis vectors $\mathbf{e}_i$.
(c) $\mathbf{A}$ is semi-simple $\Leftrightarrow$ $\mathbf{A}$ is diagonalizable.
(d) Prove Theorem 0.8.

Theorem 0.9 (Cayley-Hamilton)
Let $P_A(\zeta) = \zeta^n + \alpha_{n-1}\zeta^{n-1} + \ldots + \alpha_0$ $(\zeta \in \mathbb{C})$ *be the characteristic polynomial of* $\mathbf{A}$*. Then*

$$P_A(\mathbf{A}) \equiv \mathbf{A}^n + \alpha_{n-1}\mathbf{A}^{n-1} + \ldots + \alpha_0\mathbf{I} = \mathbf{0}. \tag{0.8}$$

The **minimal polynomial** Q_A *of* $\mathbf{A}$ *is the monic non-trivial polynomial* Q *of minimal degree for which* $Q(\mathbf{A}) = \mathbf{0}$*.* Q *is* **monic** *if* $Q(\zeta) = \zeta^k +$ *terms of degree* $< k$*. The minimal polynomial* **factorises** P_A*, i.e.,* $P_A = Q_A R$ *for some polynomial* R *(*R *might be constant* 1*).*

Exercise 0.19. ***Proof of Theorem*** **0.9.** Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $\mathbf{A}$ counted according to algebraic multiplicity.
(a) If $\mathbf{T}$ is a non-singular $n \times n$ matrix and P is a polynomial, then $P(\mathbf{T}^{-1}\mathbf{AT}) = \mathbf{T}^{-1}P(\mathbf{A})\mathbf{T}$.
(b) Let p be a polynomial. Show that

$$p(J) = \begin{bmatrix} p(\lambda) & p'(\lambda) & \frac{p''(\lambda)}{2!} & \frac{p^{(3)}(\lambda)}{3!} \\ 0 & p(\lambda) & p'(\lambda) & \frac{p''(\lambda)}{2!} \\ 0 & 0 & p(\lambda) & p'(\lambda) \\ 0 & 0 & 0 & p(\lambda) \end{bmatrix} \quad \text{if} \quad J = \begin{bmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{bmatrix}. \tag{0.9}$$

(Hint: first consider polynomials of the form $p(\zeta) = \zeta^k$ $(\zeta \in \mathbb{C})$ for $k = 1, 2, \ldots$.)
Generalise this result to Jordan blocks of higher dimension.
(c) If J_λ is a Jordan block of size $\ell \times \ell$, then $P(J_\lambda) = 0$ for any polynomial P of the form $P(\zeta) = (\lambda - \zeta)^\ell Q(\zeta)$ $(\zeta \in \mathbb{C})$, with Q a polynomial.

(d) Use Theorem 0.7 to prove (0.8).

(e) Show that the minimal polynomial factorises the characteristic polynomial.

(f) Show that the degree of the minimal polynomial is at least equal to the number of different eigenvalues of $\mathbf{A}$, with equality if and only if $\mathbf{A}$ is semi-simple. The degree of the minimal polynomial is also called the *degree of* $\mathbf{A}$.

Exercise 0.20. Consider the situation of Theorem 0.9.

(a) Prove that

$$\alpha_0 = \det(\mathbf{A}) = \prod_{j=1}^{n} \lambda_j, \quad \alpha_{n-1} = \text{trace}(\mathbf{A}) = \sum_{j=1}^{n} \lambda_j.$$

(b) Suppose $\mathbf{A}$ is non-singular. Note that then $\alpha_0 \neq 0$. Consider the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$. Show that

$$\mathbf{x} = q(\mathbf{A})\mathbf{b} \quad \text{for some polynomial } \mathbf{q} \text{ of degree} < n.$$

Actually, one can take $q(\zeta) = -\frac{1}{\alpha_0}(\zeta^{n-1} + \alpha_{n-1}\zeta^{n-2} + \ldots + \alpha_1)$. Give also an expression for q in terms of the minimal polynomial.

Exercise 0.21. Let $\mathbf{B}$ be an $n \times n$ matrix that **commutes with** $\mathbf{A}$, i.e., $\mathbf{BA} = \mathbf{AB}$.

(a) Both space $\mathcal{V}(\lambda)$ and $\mathcal{W}(\lambda)$ (w.r.t. $\mathbf{A}$) are invariant under multiplication by $\mathbf{B}$.

(b) The space $\mathcal{V}(\lambda)$ contains an eigenvector of $\mathbf{B}$.

If $\mathbf{y} \in \mathbb{C}^n, \mathbf{y} \neq \mathbf{0}$ and $\mathbf{y}^*\mathbf{A} = \mu\mathbf{y}^*$, then $\mathbf{y}$ is a **left eigenvector** of $\mathbf{A}$ associated to the (left) eigenvalue μ. If we discuss left eigenvectors, then we refer to non-trivial vectors $\mathbf{x}$ for which $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ as **right eigenvectors**. Left and right eigenvectors with different eigenvalues are mutual orthogonal (for a proof, see Exercise 0.22):

Theorem 0.10 *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix.*
1) $\lambda \in \mathbb{C}$ *is a left eigenvalue of* $\mathbf{A}$ *if and only if* λ *is a right eigenvalue of* $\mathbf{A}$.
2) *If* $\mathbf{x}$ *is a right eigenvector with eigenvalue* λ *and* $\mathbf{y}$ *be a left eigenvector with eigenvalue* $\mu \neq \lambda$, *then* $\mathbf{y} \perp \mathbf{x}$.

Corollary 0.11 *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix.*
Suppose $\mathbf{u}$ *is in the span of right eigenvectors* $\mathbf{x}_i$ *of* $\mathbf{A}$ *with eigenvalue* λ_i: $\mathbf{u} = \sum \alpha_i \mathbf{x}_i$.
If λ_i *is simple and* $\mathbf{y}_i$ *is the left eigenvector of* $\mathbf{A}$ *associated with* λ_i *scaled such that* $\mathbf{y}^*\mathbf{x} = 1$, *then* $\alpha_i = \mathbf{y}_i^*\mathbf{u}$.

Exercise 0.22. Let $\mathbf{y}$ be a left eigenvector with eigenvalue μ.

(a) For $\lambda \in \mathbb{C}$, λ left eigenvalue $\quad\Leftrightarrow\quad P_A(\lambda) = 0 \quad\Leftrightarrow\quad \lambda$ is a right eigenvalue.

(b) If $\mathbf{x}$ is a right eigenvector with eigenvalue λ and $\lambda \neq \mu$, then $\mathbf{y} \perp \mathbf{x}$.

(c) If $\mathbf{x}$ is a right eigenvector with eigenvalue μ and there is an n-vector $\mathbf{z}$ such that $\mathbf{A}\mathbf{z} = \mu\mathbf{z} + \mathbf{x}$ ($\mathbf{x}$ is associated with a non-trivial Jordan block J_μ), then $\mathbf{y} \perp \mathbf{x}$.

(d) The subspace $\mathbf{y}^\perp$ is invariant under multiplication by $\mathbf{A}$.

(e) If μ is simple, then $\mathbf{y}^\perp = \bigoplus \mathcal{W}(\lambda)$, where we sum over all eigenvalues λ of $\mathbf{A}$, $\lambda \neq \mu$.

(f) $\{\mathbf{y} \,\big|\, (\mathbf{A}^* - \bar{\mu}\mathbf{I})^\ell \mathbf{y} = \mathbf{0} \text{ for some } \ell \in \mathbb{N}\} \perp \mathcal{W}(\lambda)$ if $\lambda \neq \mu$.

(g) Give an example of a matrix $\mathbf{A}$ with left and right eigenvector $\mathbf{y}$ and $\mathbf{x}$, respectively, both associated to the same eigenvalue λ such that $\mathbf{y} \perp \mathbf{x}$. (Hint: you can find a 2×2 matrix $\mathbf{A}$ with $\lambda = 0$ with this property.)

The **spectrum** $\Lambda(\mathbf{A})$ of $\mathbf{A}$ is the set of all eigenvalues of $\mathbf{A}$.
The **spectral radius** $\rho(\mathbf{A})$ of $\mathbf{A}$ is the absolute largest eigenvalue of $\mathbf{A}$:

$$\rho(\mathbf{A}) = \{|\lambda| \,\big|\, \lambda \in \Lambda(\mathbf{A})\}.$$

For complex numbers x with $|x| < 1$ we have that $x^k \to 0 \ \ (k \to \infty)$ and (geometric series)

$$(1-x)^{-1} = 1 + x + x^2 + x^3 + \ldots.$$

For matrices $\mathbf{A}$, $\rho(\mathbf{A}) < 1$ implies $\mathbf{A}^k \to \mathbf{0} \ \ (k \to \infty)$ and (**Neumann series**)

$$(\mathbf{I} - \mathbf{A})^{-1} = \mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \mathbf{A}^3 + \ldots. \tag{0.10}$$

Theorem 0.12
1) $\mathbf{A}^k\mathbf{x} \to \mathbf{0} \ \ (k \to \infty)$ *for all* $\mathbf{x} \in \mathbb{C}^n \quad \Leftrightarrow \quad \rho(\mathbf{A}) < 1$.
2) *If* $1 \notin \Lambda(\mathbf{A})$, *then* $\mathbf{I} - \mathbf{A}$ *is non-singular.*
3) *If* $\rho(\mathbf{A}) < 1$, *then* $\mathbf{I} + \mathbf{A} + \ldots + \mathbf{A}^k$ *converges to* $(\mathbf{I} - \mathbf{A})^{-1}$.

Exercise 0.23. ***Proof of Theorem*** 0.12**.**

(a) Prove the first statement of the theorem in case $\mathbf{A}$ is a Jordan block J_λ. (Hint: J_λ^k is upper triangular with entries λ^k, $k\lambda^{k-1}$, $\frac{1}{2!}k(k-1)\lambda^{k-2}$, ..., on the main diagonal, first co-diagonal, second co-diagonal, ..., respectively, see (0.9))

(b) Prove the first statement of the theorem for the general case.

(c) Prove the third statement. (Hint: check that $(\mathbf{I} - \mathbf{A})(\mathbf{I} + \mathbf{A} + \ldots + \mathbf{A}^k) = \mathbf{I} - \mathbf{A}^{k+1}$.)

An eigenvalue λ of $\mathbf{A}$ is **dominant** if it is simple and $|\lambda| > |\lambda_j|$ for all other eigenvalues λ_j of $\mathbf{A}$. An eigenvector associated to a dominant eigenvalue is said to be **dominant**.

Convention 0.13 Below, inequalities, as $\mathbf{V} \leq \mathbf{W}$, for matrices, assume $\mathbf{V} = (V_{ij})$ and $\mathbf{W} = (W_{ij})$ to be of the same dimension and refer to a matrix entry-wise inequality, i.e., $V_{ij} \leq W_{ij}$ for all i, j. The matrix $|\mathbf{V}|$ has $|V_{ij}|$ as (i,j)th entry, that is, the absolute value is also matrix entry-wise. We follow the same convention for vector inequalities and absolute value. Note that $\mathbf{v} > \mathbf{0}$ is not the same as $\mathbf{v} \geq \mathbf{0}$, $\mathbf{v} \neq \mathbf{0}$: we have that $\mathbf{v} = (v_1, \ldots, v_k)^{\mathrm{T}} > \mathbf{0}$ if $v_i > 0$ for all i, whereas, $\mathbf{v} \geq \mathbf{0}$, $\mathbf{v} \neq \mathbf{0}$ only implies that $v_i > 0$ for some i.

If $\mathbf{A} \geq \mathbf{0}$, then $\rho(\mathbf{A}) \in \Lambda(\mathbf{A})$, that is, the eigenvalue of $\mathbf{A}$ with largest absolute value is in $[0, \infty)$ (cf., the first claim of the Perron–Frobenius Theorem below). According to the second and third claim of the Perron-Frobenius theorem below, inspection of the graph[4] of $\mathbf{A}$ with $\mathbf{A} \geq \mathbf{0}$, may reveal that $\rho(\mathbf{A})$ is dominant. Irreducibility and a-periodicity play a role.[5]

Theorem 0.14 (Perron–Frobenius) *Let* $\mathbf{A} \geq \mathbf{0}$.
1) *Then* $\rho(\mathbf{A})$ *is an eigenvalue with an eigenvector* $\mathbf{v} \geq \mathbf{0}$.
2) *If* $\mathbf{A} \geq \mathbf{0}$ *is irreducible, then* $\mathbf{v} > \mathbf{0}$ *and* $\rho(\mathbf{A})$ *is a simple eigenvalue.*
3) *If* $\mathbf{A} \geq \mathbf{0}$ *is irreducible and a-periodic, then* $\rho(\mathbf{A})$ *is a dominant eigenvalue.*

Exercise 0.24. ***Proof of Theorem*** 0.14**.** Let $\mathbf{A} \geq \mathbf{0}$. To ease notation (slightly), we assume that $\rho(\mathbf{A}) = 1$ (why is this not a restriction on the generality?).

(a) Put $\nu_k = 1 - 1/k$. Select an $\mathbf{u}$ with $\mathbf{u} \neq \mathbf{0}$, $\mathbf{u} \geq \mathbf{0}$. Show that

$$\mathbf{v}_k \equiv [\mathbf{I} + (\nu_k\mathbf{A}) + (\nu_k\mathbf{A})^2 + (\nu_k\mathbf{A})^3 + \ldots]\,\mathbf{u}$$

exists and conclude that $\nu_k\mathbf{A}\mathbf{v}_k = \mathbf{v}_k - \mathbf{u}$. Show that $\mathbf{u} \leq \mathbf{v}_k \leq \mathbf{v}_{k+1}$ for all $k = 1, 2, \ldots$.

Put $\phi_k \equiv \|\mathbf{v}_k\|_2$. Distinguish two cases: i) $\phi_k \to \infty$ for $k \to \infty$ and ii) $\sup_k \phi_k < \infty$.

---

[4]The directed graph associated to the matrix $\mathbf{A}$ consists of vertices $1, \ldots, n$ and there is an edge from i to j iff $A_{ij} \neq 0$.

[5]A matrix is **irreducible** if for all i, j there is a path in its graph from vertex i to vertex j. The matrix is a-periodic if the greatest common divisor of the length of circular paths is 1.

(b) Assume $\phi_k \to \infty$. Use an compactness argument to conclude that some subsequence of $(\mathbf{v}_k/\phi_k)$ converges so some vector $\mathbf{v}$, say. Prove that $\|\mathbf{v}\|_2 = 1$, $\mathbf{v} \geq \mathbf{0}$ and $\mathbf{A}\mathbf{v} = \mathbf{v}$.

We will now argue that, for $\mathbf{u} > \mathbf{0}$, case ii) does not occur.

(c) Assume $\sup_k \phi_k < \infty$. Use an compactness argument to conclude that some subsequence of $(\mathbf{v}_n)$ converges so some vector $\mathbf{v}$, say. Show that $\mathbf{A}\mathbf{v} = \mathbf{v} - \mathbf{u}$ and $\mathbf{v} \geq \mathbf{0}$.

Let $\mathbf{w}$ be some left eigenvector of $\mathbf{A}$ with eigenvalue λ, $|\lambda| = 1$: $\mathbf{w}^*\mathbf{A} = \lambda\mathbf{w}^*$. Note that such a vector exists. Assume $\mathbf{y} \geq \mathbf{0}$ and $\mathbf{z}$ are vectors for which $\mathbf{A}\mathbf{y} = \mathbf{y} - \mathbf{z}$. Show that

$$|\mathbf{w}|^* \leq |\mathbf{w}|^*\mathbf{A} \quad \text{and} \quad |\mathbf{w}|^*\mathbf{y} \leq |\mathbf{w}|^*\mathbf{A}\mathbf{y} = |\mathbf{w}|^*\mathbf{y} - |\mathbf{w}|^*\mathbf{z}. \tag{0.11}$$

Take $\mathbf{y} = \mathbf{v}$ and $\mathbf{z} = \mathbf{u} > \mathbf{0}$ and conclude that there is a $\mathbf{v} \neq \mathbf{0}$ such that $\mathbf{A}\mathbf{v} = \mathbf{v}$ and $\mathbf{v} \geq \mathbf{0}$. Use (0.11) also to prove the following two statements.

A) 1 is a semi-simple eigenvalue of $\mathbf{A}$ if $\mathbf{v} > \mathbf{0}$ (if not semi-simple, take $\mathbf{z} = \mathbf{v}$),

B) $|\mathbf{w}|^* = |\mathbf{w}|^*\mathbf{A}$ if $\mathbf{v} > 0$ (if $|\mathbf{w}|^* \neq |\mathbf{w}|^*\mathbf{A}$, take $\mathbf{y} = \mathbf{v}$, $\mathbf{z} = \mathbf{0}$).

Now, assume that $\mathbf{A}$ is also irreducible.

(d) Prove that $\mathbf{v} > \mathbf{0}$ and that 1 is semi-simple.

(e) To prove simplicity of the eigenvalue 1, assume $\mathbf{A}\widetilde{\mathbf{v}} = \widetilde{\mathbf{v}}$ for some vector $\widetilde{\mathbf{v}} \neq \mathbf{0}$. First prove that $\widetilde{\mathbf{v}}$ is a scalar multiple of $\mathbf{v}$ if $\widetilde{\mathbf{v}}$ is real. (Hint, for some $\alpha \in \mathbb{R}$, $\mathbf{v} - \alpha\widetilde{\mathbf{v}} \geq \mathbf{0}$, while at least one coordinate of this vector equals 0.) Now observe that, also $\mathbf{A}\overline{\widetilde{\mathbf{v}}} = \overline{\widetilde{\mathbf{v}}}$ and conclude that $\widetilde{\mathbf{v}}$ is a scalar multiple of $\mathbf{v}$. In particular, we see that the eigenvalue 1 has geometric multiplicity 1, and therefore algebraic multiplicity 1: 1 is a simple eigenvalue.

Now, assume $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ for some λ with $|\lambda| = 1$ and $\mathbf{x} \neq \mathbf{0}$.

(f) Show that $|\mathbf{x}|$ is a multiple of $\mathbf{v}$. (Hint: cf., B).) In particular, $x_i \neq 0$ for all i.

Express $\mathbf{x}$ as a 'minimal' sum of 'disjoint' non-trivial vectors of 'constant sign', that is, $\mathbf{x} = \widehat{\mathbf{x}}_1 + \ldots + \widehat{\mathbf{x}}_m$ with, for each $i = 1, \ldots, n$, $\mathbf{e}_i^*\widehat{\mathbf{x}}_p \neq 0$ for exactly one p (disjoint), for each q, $\zeta_q\,\widehat{\mathbf{x}}_q \geq \mathbf{0}$ for some sign ζ_q, i.e., $\zeta_q \in \mathbb{C}$, $|\zeta_q| = 1$ (constant sign), and $\zeta_q \neq \zeta_p$ if $q \neq p$ (minimal). Here, and below, $p, q \in \{1, \ldots, m\}$.

(g) Show that $\mathbf{e}_i^*\mathbf{A}\widehat{\mathbf{x}}_q = 0$ for all $q \neq p$ if $\mathbf{e}_i^*\mathbf{A}\widehat{\mathbf{x}}_p \neq 0$. (Hint: if $\mathbf{e}_i^*\mathbf{A}\widehat{\mathbf{x}}_p \neq 0$ for two ps, then $|\mathbf{e}_i^*\mathbf{A}\mathbf{x}| < \mathbf{e}_i^*\mathbf{A}|\mathbf{x}|$, whence $|\mathbf{x}| = |\mathbf{A}\mathbf{x}| \neq \mathbf{A}|\mathbf{x}|$). Hence, for any sign ζ, $\zeta\,\widehat{\mathbf{x}}_p \geq 0$ if and only if $\zeta\mathbf{A}\widehat{\mathbf{x}}_p \geq \mathbf{0}$. Prove that, for some permutation π of $1, \ldots, m$, $\mathbf{A}\widehat{\mathbf{x}}_p = \lambda\,\widehat{\mathbf{x}}_{\pi(p)}$ for all p.
Use the irreducibility of $\mathbf{A}$ to prove that the permutation π is circular.[6]

Finally, assume that $\mathbf{A}$ is a-periodic.

(h) Conclude that $m = 1$, $\lambda = 1$, and 1 is a dominant eigenvalue of $\mathbf{A}$.

(i) Discuss properties of the absolute largest eigenvalue of the following two matrices

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 0 & 6 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{bmatrix}.$$

Note that the last matrix maps $\mathbf{e}_1$ to $\frac{1}{2}\mathbf{e}_2$. Then to $\frac{1}{6}\mathbf{e}_3$ and then back to $\mathbf{e}_1$. In particular, the sequence $(\|\mathbf{A}^k\mathbf{u}_0\|_2)$ does not need to be monotonic, even not for $k \geq k_0$, k_0 large.

A characteristic polynomial is monic: the leading coefficient is one. Conversely, any monic polynomial is a characteristic polynomial of some suitable matrix. This statement is obvious if the zeros of the polynomial are available: then, we can take the diagonal matrix with the zeros on the diagonal. However, for a suitable matrix, we do not need the zeros.

[6] that is, all values $\pi(1), \pi^2(1), \ldots, \pi^{m-1}(1)$ are different

Let $p(\zeta) = \zeta^n - (\alpha_{n-1}\zeta^{n-1} + \ldots + \alpha_1\zeta + \alpha_0)$ $(\zeta \in \mathbb{C})$ be a polynomial (with $\alpha_j \in \mathbb{C}$). Then

$$\mathbf{H}\begin{bmatrix} \lambda^{n-1} \\ \lambda^{n-2} \\ \vdots \\ \vdots \\ 1 \end{bmatrix} = \lambda \begin{bmatrix} \lambda^{n-1} \\ \lambda^{n-2} \\ \vdots \\ \vdots \\ 1 \end{bmatrix}, \quad \text{where} \quad \mathbf{H} \equiv \begin{bmatrix} \alpha_{n-1} & \alpha_{n-2} & \ldots & \alpha_1 & \alpha_0 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ddots & & \vdots \\ & \ddots & \ddots & \ddots & \vdots \\ & & & 1 & 0 \end{bmatrix}, \tag{0.12}$$

for all zeros λ of p. In particular, the zeros of p are eigenvalues of $\mathbf{H}$ and p is the characteristic polynomial of $\mathbf{H}$. $\mathbf{H}$ is the **companion matrix** of p. Modern software packages as MATLAB compute zeros of polynomials, by forming the companion matrix and applying modern numerical techniques for computing eigenvalues of matrices (as the QR-algorithm, to be discussed in Lecture 4).

Exercise 0.25. Let p a polynomial with companion matrix $\mathbf{H}$ (cf., (0.12)).
Let $\mathbf{x}(\zeta)$ be the vector with coordinates $\zeta^{n-1}, \zeta^{n-2}, \ldots, \zeta, 1$ $(\zeta \in \mathbb{C})$.

(a) Prove that $\mathbf{Hx}(\lambda) = \lambda\mathbf{x}(\lambda) \quad \Leftrightarrow \quad p(\lambda) = 0$.

(b) Prove that p is the characteristic polynomial of $\mathbf{H}$ in case all zeros of p are mutually different.

(c) Suppose $p(\lambda) = p'(\lambda) = 0$. Show that $\mathbf{Hx}'(\lambda) = \lambda\mathbf{x}'(\lambda) + \mathbf{x}$ and conclude that λ is an eigenvalue of $\mathbf{H}$ of algebraic multiplicity at least 2. and that the associated Jordan block J_λ is at least 2×2.

(d) Prove that p is the characteristic polynomial of $\mathbf{H}$ regardless the multiplicity of the zeros.

E Special matrices.

$\mathbf{A}$ is an $n \times n$ matrix.
$\mathbf{A}$ is **Hermitian** (or **self adjoined**) if $\mathbf{A}^* = \mathbf{A}$. $\mathbf{A}$ is **symmetric** if $\mathbf{A}^{\mathrm{T}} = \mathbf{A}$.
Note hat for a real matrix $\mathbf{A}$, $\mathbf{A}$ is symmetric if and only of $\mathbf{A}$ is Hermitian. Often, if a matrix is said to be symmetric, it is implicitly assumed that the matrix is real. If that is not case, the matrix is referred to as a *complex* symmetric matrix, i.e., the possibility that matrix entries are non-real is explicitly mentioned.

A matrix $\mathbf{A}$ is **anti-Hermitian** if $\mathbf{A}^* = -\mathbf{A}$. Sometimes it is convenient to split a (general square) matrix $\mathbf{A}$ into a Hermitian and an anti-Hermitian part:

$$\mathbf{A} = \mathbf{A}_h + \mathbf{A}_a, \quad \text{with} \quad \mathbf{A}_a \equiv \tfrac{1}{2}(\mathbf{A} + \mathbf{A}^*) \quad \text{and} \quad \mathbf{A}_a \equiv \frac{1}{2}(\mathbf{A} - \mathbf{A}^*) \tag{0.13}$$

(check this), as a complex number α can be split onto a real and an imaginary part: $\alpha = \alpha_r + i\alpha_i$ with $\alpha_r = \mathrm{Re}(\alpha)$ and $\alpha_i = \mathrm{Im}(\alpha)$. Here i is the complex number $\sqrt{-1}$.

Theorem 0.15 $\mathbf{A} = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{x}^*\mathbf{A}\mathbf{x} = 0$ *for all* $\mathbf{x} \in \mathbb{C}^n$.
$\mathbf{A}$ *is Hermitian* $\quad \Leftrightarrow \quad \mathbf{x}^*\mathbf{A}\mathbf{x} \in \mathbb{R}$ *for all* $\mathbf{x} \in \mathbb{C}^n$.

Exercise 0.26.

(a) Prove the first statement of Th. 0.15. (Hint: take $\mathbf{x} = \mathbf{y} + \mathbf{z}$ and $\mathbf{x} = \mathbf{y} + i\mathbf{z}$.)

(b) If $\mathbf{A}$ is real, can we conclude that $\mathbf{A} = \mathbf{0}$ whenever $\mathbf{x}^*\mathbf{A}\mathbf{x} = 0$ for all $\mathbf{x} \in \mathbb{R}^n$?

(c) Let $\mathbf{A}$ be real. Show that $\mathbf{A}$ is anti-Hermitian $\quad \Leftrightarrow \quad \mathbf{x}^*\mathbf{A}\mathbf{x} = 0$ for all $\mathbf{x} \in \mathbb{R}^n$

(d) If $\mathbf{A}$ and $\mathbf{H}$ are Hermitian and $\alpha, \beta \in \mathbb{R}$, then $\alpha\mathbf{A} + \beta\mathbf{H}$ is Hermitian.

(e) If $\mathbf{V}$ is an $n \times k$ matrix and $\mathbf{A}$ is Hermitian, then $\mathbf{V}^*\mathbf{A}\mathbf{V}$ is Hermitian.

(f) If $\mathbf{A}$ is anti-Hermitian, then $i\mathbf{A}$ is Hermitian.

(g) Prove the second statement of Th. 0.15:
if $\mathbf{x}^*\mathbf{A}\mathbf{x} \in \mathbb{R}$ for all $\mathbf{x} \in \mathbb{C}^n \quad \Leftrightarrow \quad \mathbf{x}^*\mathbf{A}_\mathrm{a}\mathbf{x} = 0$ for all $\mathbf{x} \in \mathbb{C}^n \quad \Leftrightarrow \quad \mathbf{A} = \mathbf{A}_\mathrm{h}$ is Hermitian.

(h) If $\mathbf{A} = \mathbf{QSQ}^*$ is the Schur decomposition of an Hermitian matrix $\mathbf{A}$, then $\mathbf{S}$ is a real diagonal. In particular, an Hermitian matrix $\mathbf{A}$ is diagonalizable, all eigenvalues are real and $\mathbf{A}$ has an orthonormal basis of eigenvectors, i.e., there is an orthonormal basis of $\mathbb{C}^n$ such that all basis vectors are eigenvectors of $\mathbf{V}$.

$\mathbf{A}$ is a **normal matrix** if $\mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A}$.

Theorem 0.16 *Hermitian and anti-Hermitean matrices are normal.*
If $\mathbf{A}$ is normal, then a vector is a right eigenvector of $\mathbf{A}$ if and only if it is a left eigenvector.
The following properties are equivalent voor a square matrix $\mathbf{A}$:
1) $\mathbf{A}$ *is normal.*
2) $\mathbf{A}_a\mathbf{A}_h = \mathbf{A}_h\mathbf{A}_a$.
3) *There is an orthonormal basis of eigenvector of* $\mathbf{A}$.
4) $\mathbf{A}^* = p(\mathbf{A})$ *for any polynomial p for which $p(\lambda) = \overline{\lambda}$ for all eigenvalues λ of* $\mathbf{A}$.
5) *There is a polynomial p with real coefficients for which* $\mathbf{A}^* = p(\mathbf{A})$.

Exercise 0.27. ***Proof of Theorem*** **0.16.**

(a) Prove the first claim of the theorem.

(b) Subsequentially prove the following impications (see the theorem)
1) $\Rightarrow$ 2), 2) $\Rightarrow$ 3) (Hint: use (b) of Exercise 0.21),
3) $\Rightarrow$ 4), 4) $\Rightarrow$ 5) (Hint: use Lagrange interpolation), 5) $\Rightarrow$ 1).

(c) Prove that left and right eigenvectors coincide in case $\mathbf{A}$ is normal. Does the converse hold?

Assume in the remaining of this exercise that $\mathbf{A}$ is normal

(d) Prove that there is a polynomial p as in 5) with degree $\leq \#\Lambda(\mathbf{A})$, i.e., the number of *different* eigenvalues of $\mathbf{A}$. In particular, the degree of the polynomial p is $\leq$ the degree of the minimal polynomial of $\mathbf{A}$.

(e) If $\mathbf{A}^* = p(\mathbf{A})$ then $p \circ p(\mathbf{A}) = \mathbf{A}$, in particular the minimal polynomial of $\mathbf{A}$ is a polynomial factor of the polynomial $\lambda - p(p(\lambda))$.

For a $n \times n$ (complex) matrix $\mathbf{A}$ the **field of values** is defined by

$$\mathcal{F}(\mathbf{A}) \equiv \{\mathbf{x}^*\mathbf{A}\mathbf{x} \,\big|\, \mathbf{x} \in \mathbb{C}^n, \|\mathbf{x}\|_2 = 1\}.$$

Note that the vectors $\mathbf{x}$ in the definition have norm 1 and do not form a convex set.[7] Nevertheless, the field of values is a convex subset of $\mathbb{C}$ that contains the eigenvalues of $\mathbf{A}$ (see the theorem below), but can be larger than the convex hull of the eigenvalues.

Theorem 0.17 *Let $\mathbf{A}$ be an $n \times n$ matrix.*
The field of values $\mathcal{F}(\mathbf{A})$ of $\mathbf{A}$ is convex and contains all eigenvalues of $\mathbf{A}$. Further

$$\mathcal{F}(\mathbf{Q}^*(\mathbf{A} - \sigma\mathbf{I})\mathbf{Q}) = \mathcal{F}(\mathbf{A}) - \sigma \quad \textit{for all } n \times n \textit{ unitary matrices } \mathbf{Q} \textit{ and } \sigma \in \mathbb{C}. \tag{0.14}$$

If $\mathbf{A}$ is normal, then, $\mathcal{F}(\mathbf{A})$ equals the convex hull of all eigenvalues λ_j of $\mathbf{A}$, that is,

$$\mathcal{F}(\mathbf{A}) = \left\{ \sum_{j=1}^{n} \beta_j \lambda_j \,\Big|\, \beta_j \in [0,1], \sum_{j=1}^{n} \beta_j = 1 \right\}. \tag{0.15}$$

[7] A subset $\mathcal{G}$ of $\mathbb{C}^n$ is convex if, with $\mathbf{x}, \mathbf{y}$ in $\mathcal{G}$, all vectors $\alpha\mathbf{x} + (1-\alpha)\mathbf{y}$ are in $\mathcal{G}$ for all $\alpha \in [0,1]$.

Exercise 0.28. ***Proof of Theorem*** **0.17.**

(a) First assume that $\mathbf{A}$ is normal and prove (0.15). (Hint use 3) of Th. 0.16.)

(b) Prove (0.14).

(c) To prove convexity of $\mathcal{F}(\mathbf{A})$ for general square matrices $\mathbf{A}$, show that it suffices to show convexity for 2×2 matrices.

(d) Show that the field of value of

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

equals the unit disc $\{\zeta\in\mathbb{C}\,\big|\,|\zeta|\leq 1\}$.

(e) Let $\mathbf{A}$ be a 2×2 matrix. For simplicity assume $\mathbf{A}$ is real. Select σ to be the trace of $\mathbf{A}$ en $\mathbf{Q}$ the orthonormal matrix of eigenvectors of $\mathbf{A}_h$. Show that, for some $\alpha,\beta\in\mathbb{R}$,

$$\mathbf{Q}^*(\mathbf{A}-\sigma\mathbf{I})\mathbf{Q}=\begin{bmatrix} \alpha & \beta \\ -\beta & -\alpha \end{bmatrix}.$$

Show that $\mathcal{F}(\mathbf{A})$ is convex.

$\mathbf{A}$ is (semi-) **positive definite** if $\mathbf{x}^*\mathbf{A}\mathbf{x}>0$ ($\mathbf{x}^*\mathbf{A}\mathbf{x}\geq 0$, respectively) for all $\mathbf{x}\in\mathbb{C}^n, \mathbf{x}\neq\mathbf{0}$.

Theorem 0.18 *The following two properties are equivalent for a square matrix* $\mathbf{A}$*:*
1) $\mathbf{A}$ *is (semi)-positive definite.*
2) $\mathbf{A}$ *is Hermitian and* $\lambda>0$ *(*$\lambda\geq 0$*, respectively) for all eigenvalues* λ *of* $\mathbf{A}$.

Exercise 0.29.

(a) Prove Th. 0.18:
$\mathbf{A}$ is positive definite $\Leftrightarrow$ $\mathbf{A}$ is Hermitian and $\lambda>0$ for all eigenvalues λ of $\mathbf{A}$.
$\mathbf{A}$ is semi positive definite $\Leftrightarrow$ $\mathbf{A}$ is Hermitian and $\lambda\geq 0$ for all eigenvalues λ of $\mathbf{A}$.

(b) $\mathbf{A}$ is positive definite $\Leftrightarrow$ $\mathbf{A}=\mathbf{M}\mathbf{M}^*$ for some non-singular $n\times n$ matrix $\mathbf{M}$.

(c) $\mathbf{A}$ is positive definite $\Leftrightarrow$ $\mathbf{A}=\mathbf{L}\mathbf{L}^*$ for some non-singular $n\times n$ lower triangular matrix $\mathbf{L}$. (Hint: apply (0.5) to $\mathbf{M}^*$).

(d) $\mathbf{A}$ is semi positive definite $\Leftrightarrow$ $\mathbf{A}=\mathbf{M}\mathbf{M}^*$ for some $n\times n$ matrix $\mathbf{M}$.

In the above statements, it is essential that the positive definiteness is with respect to complex data: if $\mathbf{A}$ is real and $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}>0$ for all $\mathbf{x}\in\mathbb{R}^n$, $\mathbf{x}\neq\mathbf{0}$, then, we can *not* conclude that $\mathbf{A}$ is symmetric.

(e) Give an example of a non-symmetric 2×2 real matrix $\mathbf{A}$ for which $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}>0$ for all $\mathbf{x}\in\mathbb{R}^2$, $\mathbf{x}\neq\mathbf{0}$.

F Quiz

Exercise 0.30. Let

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

(a) Wat is the Range of A?

(b) What is the Null space of A?

(c) What is the rank of A?

(d) What are the eigenvalues of A?