LISTA 1 – MS993 / MT404 – 2S2016

O conjunto de exercícios listados a seguir constitui uma síntese de elementos básicos de Álgebra Linear que será usado durante este curso MS933/MT404 de *Álgebra Linear Numérica*. Esse material também será útil para o desenvolvimento dos projetos, demais listas e para a prova escrita. Alguns itens, como a decomposição Schur, pode não ter sido discutido (ou apenas brevemente) em um curso básico de Álgebra Linear. Os tópicos abordados na lista serão revisitados ao longo do curso, quando necessário, como por exemplo, Subespaços fundamentais de Matrizes, Normas de Vetores e Matrizes, Matrizes ortonormais, Valores próprios e Matrizes especiais (e.g., matrizes definidas-positivas e simétricas).

Nesse contexto, esta lista 1 está sendo incluída também com a finalidade de dar um contexto para os exercícios, ou seja, para introduzir/revisitar notação, para fornecer uma breve revisão da teoria necessária e uma motivação para estudos teóricos via prática de exercícios representativos.

Como já mencionado, o curso MS993 / MT404 tem como objetivo primário o de fornecer uma visão teórica e com uma ênfase no desenvolvimento de habilidades práticas computacionais de métodos numéricos aplicados para resolver problemas de álgebra linear numérica em grande escala. De todo modo, os exercícios são incluídos por uma das seguintes razões:

- Permitir uma prática sobre aspectos teóricos de tópicos do curso.
- Ampliar o conhecimento sobre livros de *Álgebra Linear Numérica*.
- Estabelecer e fixar notação tipicamente encontrada nos livros de *Álgebra Linear Numérica*.
- Ampliar assuntos discutidos durante as aulas.
- Ganhar alguma familiaridade com o Matlab (mas qualquer outra linguagem pode ser usada).
- Fornecer mais detalhes sobre resultados discutidos.
- Mais detalhes e observações sobre a conexão entre os aspectos teóricos e computacionais.
- Composição da nota final.

O texto não tem a finalidade de ser uma introdução fácil à teoria e os aspectos computacionais para MS993/MT404. Para este fim teremos aulas para introdução e discussão dos conceitos pertinentes, além de apontamentos para motivação do assunto e para auxuliar na fundamentação da teoria. Os livros textos servirão sempre como uma fonte importante para fixação de toda a teoria matemática subjacente.

Atenção: Nessa lista alguns teoremas são apresentados sem demonstração. Assim, pede-se que, além de fazer todos os exercícios propostos, que seja também incluindo a prova correspondente para cada um desses teoremas que foram apenas enunciados.

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.

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 $\mathbf{A}$. 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 $\mathcal{V} \cap \mathcal{W} = \{\mathbf{0}\}$.

Exercise 0.1.
(a) $\mathcal{V} + \mathcal{W}$ is a linear subspace.

---

[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.

(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: $x \rightsquigarrow \mathbf{A}x$ $(x \in \mathbb{C}^k)$.

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 $\{x \in \mathbb{C}^k \,\big|\, \mathbf{A}x = \mathbf{0}\}$.

Exercise 0.4.

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

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

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

(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 $x \in \mathbb{C}^k$.

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

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

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

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 0.2 *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.

```
GRAM–SCHMIDT ORTHONORMALISATION
r_11 = ||a_1||_2,  q_1 = a_1/r_11,  ℓ = 1.
for j = 2, ..., k
  Orthogonalise:
   v = a_j
   for i = 1, ..., ℓ
     r_ij = q_i^* a_j,  v ← v − q_i r_ij
   end for
  Normalise:
   r_{ℓ+1,j} = ||v||_2
   If  r_{ℓ+1,j} ≠ 0
     ℓ ← ℓ + 1,  q_ℓ = v/r_ℓj
   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.

(a) Let $\mathbf{A}$, $\mathbf{L}$ and $\mathbf{U}$ be $n \times n$ matrices such that $\mathbf{A} = \mathbf{LU}$, $\mathbf{L}$ 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

Exercise 0.11. Let $\mathbf{V}$ be an $n \times k$ matrix.
(a) If $\mathbf{V}$ is orthonormal, then $k = \dim(\mathrm{span}(\mathbf{V}))$.
(b) $\mathbf{V}$ is orthonormal $\Leftrightarrow$ $\mathbf{V}^*\mathbf{V} = I_k$ the $k \times k$ identity matrix

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 0.3 *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, $\mathrm{span}(\mathbf{A}) = \mathrm{span}(\mathbf{Q})$, R is upper triangular, and $\mathbf{A} = \mathbf{Q}R$.

A matrix $\mathbf{Q}$ is **unitary** of $\mathbf{Q}$ is square and orthonormal.

Exercise 0.12. ***Proof of Theorem*** 0.3. 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

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

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 0.3: 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.

Theorem 0.4 *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 $\mathcal{V} = \mathrm{span}(\mathbf{V})$.
(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 0.4.

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 $\mathbf{A} = (a_{ij})$ is **real** (i.e., $a_{ij} \in \mathbb{R}$ for all i, j), 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, $\mathbf{A}$ 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) Eigenvalues do not depend on the basis, i.e., if $\mathbf{T}$ is a non-singular $n \times n$ matrix, then $\mathbf{A}$ and $\mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ have the same eigenvalues with equal multiplicity (algebraic, as well as geometric).

(f) 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 $\mathbf{A}$.

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

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

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

(j) 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 $\lambda\neq\mu$.

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

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

$$\mathbf{A}\mathbf{Q}=\mathbf{Q}S, \tag{0.6}$$

then (0.6) is a **partial Schur decomposition** (or **partial Schur form**) of $\mathbf{A}$ (of order k). If $k=n$, then (0.6) is a **Schur decomposition** of **Schur form**.

Theorem 0.5 $\mathbf{A}$ *has a Schur decomposition.*

Proof. Apply induction to k to prove the theorem:

There is a normalised eigenvector $\mathbf{q}_1$ of $\mathbf{A}$. Note that $\mathbf{A}\mathbf{q}_1=\mathbf{q}_1\lambda_1$ is a partial Schur decomposition of order 1.

Suppose we have a partial Schur decomposition $\mathbf{A}\mathbf{Q}_k=\mathbf{Q}_kS_k$ 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 (see (f) 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 $\mathbf{A}\mathbf{Q}_{k+1}=\mathbf{Q}_{k+1}S_{k+1}$ of order $k+1$. □

Exercise 0.17. Suppose we have a partial Schur decomposition (0.6).

(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.

Without proof, we mention:

Theorem 0.6 *There is a non-singular $n\times n$ matrix $\mathbf{T}$ such that $\mathbf{A}\mathbf{T}=\mathbf{T}\mathbf{J}$, 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 0.7 *The following properties are equivalent voor any $n\times n$ matrix $\mathbf{A}$:*

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*** *0.7.*

(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 $\quad\Leftrightarrow\quad$ $\mathbf{A}$ is diagonalizable.

(d) Prove Theorem 0.7.

Theorem 0.8 (Cayley-Hamilton)
Let $P_A(\zeta) = \zeta^n + \alpha_{n-1}\zeta^{n-1} + \ldots + \alpha_0 \quad (\zeta \in \mathbb{C})$ be the characteristic polynomial of $\mathbf{A}$. Then

$$P_A(\mathbf{A}) \equiv \mathbf{A}^n + \alpha_{n-1}\mathbf{A}^{n-1} + \ldots + \alpha_0\mathbf{I} = \mathbf{0}. \tag{0.7}$$

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*** *0.8.* 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 $P(\mathbf{T}^{-1}\mathbf{A}\mathbf{T}) = \mathbf{T}^{-1}P(\mathbf{A})\mathbf{T}$.

(b) Let p be a polynomial. Show that

$$p(J) = \begin{bmatrix} p(\lambda) & p'(\lambda) & p''(\lambda) \\ 0 & p(\lambda) & p'(\lambda) \\ 0 & 0 & p(\lambda) \end{bmatrix} \quad \text{if} \quad J = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}. \tag{0.8}$$

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 $P(\zeta) = (\lambda - \zeta)^\ell Q(\zeta) \quad (\zeta \in \mathbb{C})$, with Q a polynomial.

(d) Use Theorem 0.6 to prove (0.7).

(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 0.8.

(a) Prove that

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

(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 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 eigenvector**s. Left and right eigenvectors with different eigenvalues are mutual orthogonal (for a proof, see Exercise 0.22):

Theorem 0.9 *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 0.10 *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 $\quad\Leftrightarrow\quad P_A(\lambda) = 0 \quad\Leftrightarrow\quad$ λ 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 $\mathbf{Az} = \mu\mathbf{z}+\mathbf{x}$ ($\mathbf{x}$ 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) $\{\mathbf{y} \,\big|\, (\mathbf{A}^* - \bar{\mu}\mathbf{I})^\ell \mathbf{y} = \mathbf{0}$ 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)

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

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

$$(\mathbf{I} - \mathbf{A})^{-1} = \mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \mathbf{A}^3 + \dots. \tag{0.9}$$

Theorem 0.11
1) $\mathbf{A}^k\mathbf{x} \to \mathbf{0} \;\; (k \to 0)$ *for all* $\mathbf{x} \in \mathbb{C}^n \quad\Leftrightarrow\quad \rho(\mathbf{A}) < 1$.
2) *If* $1 \not\in \Lambda(\mathbf{A})$*, then* $\mathbf{I} - \mathbf{A}$ *is non-singular.*
3) *If* $\rho(\mathbf{A}) < 1$*, then* $I + \mathbf{A} + \ldots + \mathbf{A}^k$ *converges to* $(\mathbf{I} - \mathbf{A})^{-1}$*.*

Exercise 0.23. ***Proof of Theorem*** 0.11*.*

(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^n, $n\lambda^{n-1}$, $n(n-1)\lambda^{n-2}$, $\ldots$, on the main diagonal, first co-diagonal, second co-diagonal, $\ldots$, respectively, see (0.8))

(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**.

Theorem 0.12 (Perron–Frobenius) *Let* $\mathbf{A}$ *be such that* $\mathbf{A} = |\mathbf{A}|$*. Then* $\rho(\mathbf{A}) \in \Lambda(\mathbf{A})$*. If, in addition,* $\mathbf{A}$ *is irreducible and a-periodic,*[3] *then* $\rho(\mathbf{A})$ *is a dominant eigenvalue of* $\mathbf{A}$*.*

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.

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

$$\mathbf{H}\begin{bmatrix} \lambda^{n-1} \\ \lambda^{n-2} \\ \vdots \\ \vdots \\ 1 \end{bmatrix} = \lambda \begin{bmatrix} \lambda^{n-1} \\ \lambda^{n-2} \\ \vdots \\ \vdots \\ 1 \end{bmatrix}, \quad \text{where} \quad \mathbf{H} \equiv \begin{bmatrix} \alpha_{n-1} & \alpha_{n-2} & \ldots & \alpha_1 & \alpha_0 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ddots & & \vdots \\ & \ddots & \ddots & \ddots & \vdots \\ & & & 1 & 0 \end{bmatrix}, \tag{0.10}$$

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.

Exercise 0.24. Let p a polynomial with companion matrix $\mathbf{H}$ (cf., (0.10)).

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 $\mathbf{H}\mathbf{x}(\lambda) = \lambda\mathbf{x}(\lambda) \quad \Leftrightarrow \quad p(\lambda) = 0$.

(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 $\mathbf{H}\mathbf{x}'(\lambda) = \lambda\mathbf{x}'(\lambda) + \mathbf{x}$ 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 $\mathbf{A}$ (i.e., all matrix entries are in $\mathbb{R}$), $\mathbf{A}$ 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:

$$\mathbf{A} = \mathbf{A}_h + \mathbf{A}_a, \quad \text{with} \quad \mathbf{A}_a \equiv \tfrac{1}{2}(\mathbf{A} + \mathbf{A}^*) \quad \text{and} \quad \mathbf{A}_{\mathrm{a}} \equiv \frac{1}{2}(\mathbf{A} - \mathbf{A}^*) \tag{0.11}$$

(see Exercise 0.25), 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 $\sqrt{-1}$.

Exercise 0.25.

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

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

[3] 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 $A_{ij} \neq 0$. 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.

(c) If $\mathbf{A}$ is anti-Hermitian, then $i\mathbf{A}$ is Hermitian. Here $i = \sqrt{-1}$.

(d) Any square matrix $\mathbf{A}$ can be written as in (0.11) with $\mathbf{A}_\mathrm{h}$ Hermitian and $\mathbf{A}_\mathrm{a}$ anti-Hermitian.

(e) $\mathbf{A}$ is Hermitian $\quad\Leftrightarrow\quad \mathbf{x}^*\mathbf{A}\mathbf{y} \in \mathbb{R}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$.

(f) 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 $\mathbf{x} \in \mathbb{C}^n$.

(g) If $\mathbf{x}^*\mathbf{A}\mathbf{x} \in \mathbb{R}$ for all $\mathbf{x} \in \mathbb{C}^n \quad\Leftrightarrow\quad \mathbf{A} = \mathbf{A}_\mathrm{h}$ is Hermitian.

(h) If $\mathbf{A} = \mathbf{Q}\mathbf{S}\mathbf{Q}^*$ 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 0.13 *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 for which* $\mathbf{A}^* = p(\mathbf{A})$.

Exercise 0.26. ***Proof of Theorem*** **0.13.**

(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 $\lambda - p(p(\lambda))$.

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

Exercise 0.27.

(a) $\mathbf{A}$ is positive definite $\quad\Leftrightarrow\quad \mathbf{A}$ is Hermitian and $\lambda > 0$ for all eigenvalues λ of $\mathbf{A}$.
(Here, you can use that $\mathbf{A} = \mathbf{0}$ if $\mathbf{x}^*\mathbf{A}\mathbf{x} = 0$ for all $\mathbf{x} \in \mathbb{C}^n$. For a proof, see Exercise 1.9(a).)

(b) $\mathbf{A}$ is semi positive definite $\quad\Leftrightarrow\quad \mathbf{A}$ is Hermitian and $\lambda \geq 0$ for all eigenvalues λ of $\mathbf{A}$.

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

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

(e) $\mathbf{A}$ is semi positive definite $\quad\Leftrightarrow\quad \mathbf{A} = \mathbf{M}\mathbf{M}^*$ 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.

(f) 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 0.28. 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?