Vectors

Linear combinations

D1.7: A linear combination $\lambda_1 v_1 + \cdots + \lambda_n v_n$ is called

1. *affine* if $\lambda_1 + \cdots + \lambda_n = 1$,
2. *conic* if $\lambda_1, \ldots, \lambda_n \geqslant 0$,
3. *convex* if it is both affine and conic.

Scalar products, lengths, angles

O1.10: Let $u, v, w \in \mathbb{R}^m, \lambda \in \mathbb{R}$. Then

1. $v \cdot w = w \cdot v$;
2. $(\lambda v) \cdot w = \lambda(v \cdot w) = v \cdot (\lambda w)$;
3. $u \cdot (v+w) = u \cdot v + u \cdot w$ and $(u+v) \cdot w = u \cdot v + v \cdot w$;
4. $v \cdot v \geqslant 0$, equality iff $v = \mathbf{0}$.

D1.11 (Euclidean norm): Let $v \in \mathbb{R}^m$. Then $\|v\| := \sqrt{v \cdot v}$.

L1.12 (Cauchy-Schwarz inequality): Let $v, w \in \mathbb{R}^m$. Then $|v \cdot w| \leqslant \|v\|\|w\|$. Equivalently $(uw)^2 \leqslant v^2 w^2$. Equality iff $v = \lambda w$.

D1.14 (Angle): Let $v, w \in \mathbb{R}^m$ nonzero. Then $\cos(\alpha) = \frac{v \cdot w}{\|v\|\|w\|}$.

D1.15: Two vectors $v, w \in \mathbb{R}^m$ are called perpendicular/orthogonal iff $v \cdot w = 0$.

L1.16: Let $v, w \in \mathbb{R}^m$. Then $\|v + w\| \leqslant \|v\| + \|w\|$.

Linear independence

L1.19: Let $v_1, \ldots, v_n \in \mathbb{R}^n$. The following statements are equivalent:

1. At least one of the vectors is a linear combination of the other ones.
2. $\mathbf{0}$ is nontrivial linear combination of the vectors.
3. At least one of the vectors is a linear combination of the prevous ones.

L1.21: Let $v_1, \ldots, v_n \in \mathbb{R}^m$ be linearly independent and $\sum_{j=1}^n \lambda_j v_j = \sum_{j=1}^n \mu_j v_j$. Then $\lambda_j = \mu_j$ for all $j \in [n]$.

D1.22 (Span): Let $v_1, \ldots, v_n \in \mathbb{R}^m$. Then $\mathrm{Span}(v_1, \ldots, v_n) := \{\sum_{j=1}^n \lambda_j v_j : \lambda_j \in \mathbb{R}\}$.

L1.23: Let $v_1, \ldots, v_n \in \mathbb{R}^m$ and $v \in \mathbb{R}^m$ a linear combination of $v_1, \ldots v_n$. Then $\mathrm{Span}(v_1, \ldots, v_n) = \mathrm{Span}(v_1, \ldots, v_n, v)$.

Matrices

Linear combinations

D2.3: Let $A = [a_{ij}]_{i=1,j=1}^{mm}$. If $j < i, j = i, j > i$ then a_{ij} is *below*, *on*, *above* the diagonal.

1. If $a_{ii} = 1$ and $a_{ij} = 0$ for $j \neq i$ then $A = I$ is the *identity* matrix.
2. If $a_{ij} = 0$ for $j \neq i$ then A is *diagonal*.
3. If $a_{ij} = 0$ for $j < i$ then A is *upper triangular*.
4. If $a_{ij} = 0$ for $j > i$ then A is *lower triangular*.
5. If $a_{ij} = a_{ji}$ for all i, j then A is *symmetric*.

D2.9: Let $A \in \mathbb{R}^{m \times n}$ with columns $v_1, \ldots, v_n$. Column v_j is independent iff v_j is not a linear combination of $v_1, \ldots, v_{j-1}$. $\mathrm{rank}(A)$ is the number of independent columns.

D2.11: $A^\top$ is the transpose of A.

O2.12: Let $A \in \mathbb{R}^{m \times n}$. Then $(A^\top)^\top = A$.

D2.13: Let $A \in \mathbb{R}^{m \times n}$. Then $R(A) := C(A^\top)$.

Matrix multiplication

D2.16: Let $A \in \mathbb{R}^{a \times n}$ and $B \in \mathbb{R}^{n \times b}$ with columns b_k. Then $AB \in \mathbb{R}^{a \times b}$ has columns Ab_k.

L2.19: Let $A \in \mathbb{R}^{a \times n}, B \in \mathbb{R}^{n \times b}$. Then $(AB)^\top = B^\top A^\top$.

C2.20: Let $I \in \mathbb{R}^{m \times m}$. Then $IA = A$ for $A \in \mathbb{R}^{m \times n}$ and $AI = A$ for $A \in \mathbb{R}^{n \times m}$.

L2.21: Let $A \in \mathbb{R}^{m \times n}$. The following statements are equivalent:

1. $\mathrm{rank}(A) = 1$.
2. There are nonzero $v \in \mathbb{R}^m, w \in \mathbb{R}^n$ such that $A = vw^\top$.

L2.22: Let A, B, C, D matrices sucht that sums and products are defined. Then

1. $A(B+C) = AB + AC$ and $(B+C)D = BD + CD$
2. $(AB)C = A(BC)$.

CR decomposition

T2.23: Let $A \in \mathbb{R}^{m \times n}, \mathrm{rank}(A) = r$. Let $C \in \mathbb{R}^{m \times r}$ be the submatrix of A containing the independent columns. Then there exists a unique $R \in \mathbb{R}^{r \times n}$ such that $A = CR$.

Linear transformations

D2.25: Let $A \in \mathbb{R}^{m \times n}$. $T_A : \mathbb{R}^n \to \mathbb{R}^m$ is defined by $T_A(x) = Ax$.

O2.26: T_A is a linear transformation.

D2.27 (Linear transformation): Let $T : \mathbb{R}^n \to \mathbb{R}^m$. T is called a *linear transformation* iff for all $x, y \in \mathbb{R}^n, \lambda \in \mathbb{R}$

1. $T(x + y) = T(x) + T(y)$ and
2. $T(\lambda x) = \lambda T(x)$.

T2.29: Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. There exists a unique $A \in \mathbb{R}^{m \times n}$ such that $T = T_A$.

L2.30: Let $T_A : \mathbb{R}^n \to \mathbb{R}^a, T_B : \mathbb{R}^b \to \mathbb{R}^n$ be linear transformations. Then $T_A(T_B(x)) = T_{AB}(x)$.

D2.31 (Kernel and image): Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then

1. $\mathrm{Ker}(T) := \{x \in \mathbb{R}^n : T(x) = \mathbf{0}\} \subseteq \mathbb{R}^n$,
2. $\mathrm{Im}(T) := \{T(x) : x \in \mathbb{R}^n\} \subseteq \mathbb{R}^m$.

O2.32: Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and $A \in \mathbb{R}^{m \times n}$ such that $T = T_A$. Then $\mathrm{Im}(T) = C(A)$.

Solving linear equations

Systems of linear equations

O3.2: Let $A \in \mathbb{R}^{m \times n}$. The columns of A are linearly independent iff $Ax = \mathbf{0}$ has a unique solution, $x = \mathbf{0}$.

Gauss elimination

A: Let $A \in \mathbb{R}^{m \times m}$, $b \in \mathbb{R}^m$. Then

```
procedure Gauss-elimination(A, b)
    for j ← 1, ..., m do
        if A_{j,j} = 0 then
            k ← j + 1
            while k ⩽ m ∧ A_{k,j} = 0 do k ← k + 1
            if k > m then return "gibs auf"
            else exchange rows j and k in A, b
        for i ← j + 1, ..., m do
            c ← A_{i,j}/A_{j,j}
            subtract c · row j from row i in A, b
```

L3.3: Let $Ax = b$ be a system of m linear equations in n variables, $M \in \mathbb{R}^{m \times m}$ be a row operation matrix. Let $A' = MA, b' = Mb$. Then $Ax = b$ and $A'x = b'$ have the same solutions.

T3.5: The following statements are equivalent:

1. Gauss elimination succeeds.
2. The columns of A are linearly independent.

T3.6: Gauss elimination is in $O(m^3)$.

Inverse matrices

D3.7: Let $M \in \mathbb{R}^{m \times m}$. M is called *invertible* iff there is an $M^{-1} \in \mathbb{R}^{m \times m}$ such that $MM^{-1} = M^{-1}M = I$.

L3.8: The inverse of a matrix is unique.

L3.9: Let $A, B \in \mathbb{R}^{m \times m}$ be invertible. Then $(AB)^{-1} = B^{-1}A^{-1}$.

L3.10: Let $A \in \mathbb{R}^{m \times m}$ be invertible. Then $(A^\top)^{-1} = (A^{-1})^\top$.

T3.11: Let $A \in \mathbb{R}^{m \times m}$. The following statements are equivalent.

1. A is invertible.
2. For every $b \in \mathbb{R}^m$, $Ax = b$ has a unique solution.
3. The columns of A are linearly independent.

LU and LUP decomposition

T3.13: Let $A \in \mathbb{R}^{n \times n}$ on which Gauss elimination succeeds without row exchanges, resulting in an upper triangular matrix U. Let c_{ij} be the multiple of row j that we subtract from row $i > j$ when we eliminate in column j. Then $A = LU$ where

$$L = \begin{bmatrix} 1 & & & \\ c_{2,1} & 1 & & \\ \vdots & & \ddots & \\ c_{m,1} & \cdots & c_{m,m-1} & 1 \end{bmatrix}.$$

D3.14: A *permutation* of $[m]$ is a bijective function $\pi \colon [m] \to [m]$.

D3.15: Let $\pi \colon [m] \to [m]$ be a permutation. The *permutation matrix* associated with π is $P \in \mathbb{Z}^{m \times m}$ with

$$p_{ij} = \begin{cases} 1 & \text{if} j = \pi(i) \\ 0 & \text{otherwise} \end{cases}.$$

L3.16: Let P be a permutation matrix. Then $P^{-1} = P^\top$.

L3.17: Let $P, P' \in \mathbb{Z}^{n \times n}$ be permutation matrices with associated permutations π, π'. Then PP' is a permutation matrix as well, associated with the permutation $\pi' \circ \pi$.

T3.18: Let $A \in \mathbb{R}^{n \times n}$, $m \geqslant 1$ have linearly independent columns. There exist $P, L, U \in \mathbb{R}^{n \times n}$ such that $PA = LU$ where P is a permutation matrix, L a lower triangular matrix with 1s on the diagonal and U an upper triangular matrix with nonzero diagonal entries.

R (Solving $Ax = b$ from $PA = LU$): Because $P^{-1} = P^\top$ we have $P^\top LUx = b$. Solve $P^\top z = b$ for z be permutation. Solve $Ly = z$ for y using forward substitution and $Ux = y$ for x using backward substitution.

Gauss-Jordan elimination

D3.19 (REF, RREF): Let $R \in \mathbb{R}^{m \times n}$. R is in row echelon form (REF) if the following holds. There exist $r \leqslant m$ column indices $1 \leqslant j_1 \leqslant \cdots \leqslant j_r \leqslant n$ such that the following statements hold:

1. For $i = 1, \ldots, r$ we have $r_{ij_i} = 1$.
2. For all i, j we have $r_{ij} = 0$ whenever $i > r$ or $j < j_i$ or $j = j_k$ for some $k > i$.

If $r = m$, R is in reduced row echelon form (RREF). We use the notation $\mathrm{REF}(j_1, \ldots, j_r)$ and $\mathrm{RREF}(j_1, \ldots, j_r)$.

O3.20: A matrix R in $\mathrm{REF}(j_1, \ldots, j_r)$ has rank r.

A (Gauss-Jordan elimination): Like Gauss elimination, but:

1. Normalize pivot of each row to 1.
2. Eliminate *above* the pivot to get REF.

T3.21 (Gauss-Jordan elimination): Let $A \in \mathbb{R}^{m \times n}$. There exists an invertible matrix $M \in \mathbb{R}^{m \times m}$ such that $R_0 = MA$ is in REF.

L3.22: Let $A \in \mathbb{R}^{m \times n}$, $M \in \mathbb{R}^{m \times m}$ invertible, and $R_0 = MA$ in $\mathrm{REF}(j_1, \ldots, j_r)$. Then A has independent columns $j_1, \ldots, j_r$.

T3.23: Let $A \in \mathbb{R}^{m \times n}$, $\mathrm{rank}(A) = r$ and $b \in \mathbb{R}^m$.

1. Using Gauss-Jordan elimination, A can be transformed into $R_0 = MA$ in REF as given by [T3.21] in time $O(rmn + mn)$.
2. By simultaneously transforming $I \in \mathbb{R}^{m\times m}$ using the same row operations, $M = MI$ can be computed in additional time $O(rm^2 + m^2)$.
3. Given M, the system $Ax = b$ can be solved in $O(m^2)$.

Computing the CR decomposition

T3.24: Let $A \in \mathbb{R}^{m\times n}$ and let $A = CR$ as in [T2.23]. Let $R_0 = MA$ in $\mathrm{REF}(j_1, \dots, j_r)$ be the result of Gauss-Jordan elimination on A [T3.21]. Then R results from R_0 by removing the zero rows at the end (if there are any); in particular R is in $\mathrm{RREF}(j_1, \dots, j_r)$, and C is the submatrix of A with columns $j_1, \dots, j_r$.

The four fundamental subspaces

Vector spaces

D4.1 (Vector space): A *vector space* is a triple $(V, +, \cdot)$ where V is a set and

1. $+ : V \times V \to V$,
2. $\cdot : \mathbb{R} \times V \to V$,

statisfying the following axioms for $u, v, w \in V; \lambda, \mu \in \mathbb{R}$.

1. $v + w = w + v$
2. $u + (v + w) = (u + v) + w$
3. There is $\mathbf{0} \in V$ such that $v + \mathbf{0} = v$.
4. There is $-v \in V$ such that $v + (-v) = \mathbf{0}$.
5. $1 \cdot v = v$
6. $(\lambda\mu)v = \lambda(\mu v)$
7. $\lambda(v + w) = \lambda v + \lambda w$
8. $(\lambda + \mu)v = \lambda v + \mu v$

O4.2: $(\mathbb{R}^m, +, \cdot)$ is a vector space.

D4.3: A *polynomial* p is a sum of the form $p = \sum_{i=0}^m p_i x^i$ for some $m \in \mathbb{N}$. x is a *variable* and $p_0, \dots, p_m \in \mathbb{R}$ are *coefficients* of p. The largest i such that $p_i \neq 0$ is the *degree* of p. The zero polynomial $\mathbf{0} = 0$ has degree -1.

T4.4: Let $\mathbb{R}[x]$ be the set of polynomials in x. Given $p = \sum_{i=0}^m p_i x^i$ and $q = \sum_{i=0}^n q_i x^i$ and $\lambda \in \mathbb{R}$. We define $p + q = \sum_{i=0}^{\max(m,n)} (p_i + q_i)x^i$ and $\lambda p = \sum_{i=0}^m (\lambda P_i)x^i$. Then $(\mathbb{R}[x], +, \cdot)$ is a vector space.

T4.5: $(\mathbb{R}^{m\times n}, +, \cdot)$ is a vector space.

F4.6: Each vector space contains exactly one zero vector.

F4.7: Each v in a vector space has exactly one $-v$.

D4.8: Let V be a vector space. $U \subseteq V$, $U \neq \emptyset$ is called a subspace of V iff for all $v, w \in \mathbb{U}$ and $\lambda \in \mathbb{R}$.

1. $v + w \in U$;
2. $\lambda v \in U$.

L4.9: Let $U \subseteq V$ be a subspace. Then $\mathbf{0} \in V$.

L4.11: Let $A \in \mathbb{R}^{m\times n}$. Then $C(A) = \{Ax : x \in \mathbb{R}^n\}$ is a subspace of $\mathbb{R}^m$.

L4.12: Let V be a vector space and U a subspace. Then U is a vector space.

Bases and dimension

L4.14: Let V be a vector space $G \subseteq V$. Every linear combination of G is in V.

D4.16 (Basis): Let V be a vector space. $B \subseteq V$ is called a basis of V iff B is linearly independent and $\mathrm{Span}(B) = V$.

O4.18: Every set of m linearly independent vectors is a basis of $\mathbb{R}^m$.

L4.19 (Steinitz exchange lemma): Let V be a vector space, $F \subseteq V$ finite and linearly independent and $G \subseteq V$ finite with $\mathrm{Span}(G) = V$. Then

1. $|F| \leqslant |G|$
2. There exists $E \subseteq G$ of size $|G| - |F|$ such that $\mathrm{Span}(F \cup E) = V$.

T4.20: Let V be a vector space and $B, B' \subseteq V$ two finite bases of V. Then $|B| = |B'|$.

D4.21: A vector space V is called finitely generated iff there exists a finite $G \subseteq V$ with $\mathrm{Span}(G) = V$.

T4.22: Let V be a finitely generated vector space and let $G \subseteq V$ be a finite subset with $\mathrm{Span}(G) = V$. Then V has a basis $B \subseteq G$.

D4.23 (Dimension): Let V be a finitely generated vector space. Then $\dim(V)$ is the size of any basis B of V.

L4.24: Let V be a vector space with $\dim(V) = d$.

1. Let $F \subseteq V$ be a set of linearly independent vectors. Then F is a basis of V.
2. Let $G \subseteq V$ be a set of d vectors with $\mathrm{Span}(G) = V$. Then G is a basis of V.

Computing the fundamental subspaces

Computing a basis of $C(A)$ and $R(A)$

T4.25/T4.28: Let $A \in \mathbb{R}^{m\times n}$ and R_0 in REF$(j_1, \dots, j_r)$ the result of Gauss-Jordan elimination on A [T3.21]. Then A has independent columns $j_1, \dots, j_r$ and these form the basis for $C(A)$. The first r rows of R_0 form a basis of $R(A)$. Hence $\dim(C(A)) = \dim(R(A)) = r = \mathrm{rank}(A)$.

L4.27: Let $A \in R^{m\times n}$ and $M \in \mathbb{R}^{m\times m}$ invertible. Then $R(A) = R(MA)$.

T4.29: Let $A \in \mathbb{R}^{m\times n}$. Then $\mathrm{rank}(A) = \mathrm{rank}(A^\top)$.

C4.30: Let $A = CR$ be the CR decomposition A [T2.23]. The columns of C form a basis of $C(A)$ [T4.25]. The rows of R form a basis of $R(A)$ [T4.28], [T3.24].

D4.31: Let $A \in \mathbb{R}m \times n$. Then $N(A) = \{x \in \mathbb{R}^n : Ax = 0\} \subseteq \mathbb{R}^n$

L4.32: Let $A \in \mathbb{R}^{m\times n}$. Then $N(A)$ is a subspace of $\mathbb{R}^n$.

L4.33: Let $A \in \mathbb{R}^{m\times n}$ and $M \in \mathbb{R}^{m\times m}$ invertible. Then $N(A) = N(MA)$.

Computing a basis of $N(A)$

L4.34: Let $R \in \mathbb{R}^{r\times n}$ be in $\mathrm{RREF}(j_1, \dots, j_r)$ (see [D3.19]). Let $j_{r+1} < \cdots < j_n$ denote the indices of the dependent columns. The $r \times r$ submatrix of R formed by the independent columns is I. We let $Q \in \mathbb{R}^{r\times(n-r)}$ denote the submatrix of R formed by the dependent columns. For $x \in \mathbb{R}^n$, let

$$x(I) = \begin{bmatrix} x_{j_1} \\ \vdots \\ x_{j_r} \end{bmatrix} \in \mathbb{R}^r \text{ and } x(Q) = \begin{bmatrix} x_{j_{r+1}} \\ \vdots \\ x_{j_n} \end{bmatrix} \in \mathbb{R}^{n-r}.$$

denote the subvectors of basic and free entries. Let $v_1, \dots, v_{n-r} \in \mathbb{R}^n$ be the vectors defined via $v_i(Q) = e_i$ and $v_i(I) = -Qv_i(Q)$. Then $\{v_1, \dots, v_{n-r}\}$ is a basis of $N(R)$.

T4.35: Let $A \in \mathbb{R}^{m\times n}$ and let R_0 in $\mathrm{REF}(j_1, \dots, j_r)$ be the result of Gauss-Jordan elimination on A (see [T3.21]). Let R in $\mathrm{RREF}(j_1, \dots, j_r)$ be the submatrix of R_0 consisting of the first r rows. The vectors $v_1, \dots, v_{n-r}$ as constructed in [L4.34] form a basis of $N(A) = N(R_0) = N(R)$ and therefore $\dim(N(A)) = n - r = n - \mathrm{rank}(A)$.

L4.36: Let $A \in \mathbb{R}^{m\times n}$. Then $LN(A) := N(A^\top) \subseteq \mathbb{R}^m$.

L4.32 (L4.37): Let $A \in \mathbb{R}^{m\times n}$. Then $N(A)$ $(LN(A))$ is a subspace of R^n $(\mathbb{R}^m)$.

Computing a basis of $LN(A)$

T4.38: Let $A \in \mathbb{R}^{m\times n}$ and let $R_0 = MA$ in $\mathrm{REF}(j_1, \dots, j_r)$ be the result of Gauss-Jordan elimination on A (see [T3.21]). Then the last $m - r$ rows $w_{r+1}, \dots, w_m$ of $M \in \mathbb{R}^{m\times m}$ form a basis of $LN(A)$, and therefore $\dim(LN(A)) = m - r = m - \mathrm{rank}(A)$.

Solution space of $Ax = b$

D4.39 (Solution space): Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$. Then $\mathrm{Sol}(A, b) := \{x \in \mathbb{R}^n : Ax = b\} \subseteq \mathbb{R}^n$.

T4.40: Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and s a solution to $Ax = b$. Then $\mathrm{Sol}(A, b) = \{s + x : x \in N(A)\}$.

L4.41: Let $A \in \mathbb{R}^{m\times n}, \mathrm{rank}(A) = m$. Then $Ax = b$ has a solution for every $b \in \mathbb{R}^m$.

Orthogonality

Orthogonality

T5.1.6 Let $A \in \mathbb{R}^{m\times n}$. $N(A) = C(A^\top)^\perp = R(A)^\perp$.

T5.1.7 Let V, W be orthogonal subspaces of $\mathbb{R}^n$. The following Statements are equivalent:

1. $W = V^\perp$.
2. $\dim(V) + \dim(W) = n$.
3. Every $u \in \mathbb{R}^n$ can be written as $u = v + w$ with unique $v \in V, w \in W$.

L5.1.8 Let V be a subspace of $\mathbb{R}^n$. Then $V = (V^\perp)^\perp$.

T5.1.1 $\{x \in \mathbb{R}^n \mid Ax = b\} = x_1 + N(A)$ where $x_1 \in R(A)$ such that $Ax_1 = b$.

L5.1.11 Let $A \in \mathbb{R}^{m\times n}$. Then $N(A) = N(A^\top A)$ and $C(A^\top) = C(A^\top A)$.

Projections

L5.2.2: Let $a \in \mathbb{R}^m_*$ and $S = \{\lambda a \mid \lambda \in \mathbb{R}\} = C(a)$. $\mathrm{proj}_S(b) = \frac{aa^\top}{a^\top a} b$.

L5.2.3: Let $b \in \mathbb{R}^m$ and $S = C(A)$. Then $\mathrm{proj}_S(b) = A\hat{x}$ where $\hat{x}$ satisfies $A^\top A\hat{x} = A^\top b$.

L5.2.4: $A^\top A$ is invertible iff A has linearly independent columns.

T5.2.6: Let S be a subspace in $\mathbb{R}^n$ and the columns of A are a basis of S. Then $\mathrm{proj}_S(b) = Pb$ where $P = A(A^\top A)^{-1}A^\top$.

R5.2.7:

1. $P^2 = P$.
2. $(I - P)b = \mathrm{proj}_{S^\perp}(b)$.
3. $(I - P)^2 = I - P$.

Least squares approximation

(2): $\min_{\hat{x}\in\mathbb{R}^n} \|A\hat{x} - b\|^2$.

(3): $A\top A\hat{x} = A^\top b$.

F5.3.1: A minimizer of (2) is also a solution of (3). When A has independent columns the unique solution $\hat{x}$ of (2) is given by $\hat{x} = (A^\top A)^{-1}A^\top b$.

Linear regression

Problem: Consider data points $(t_1, b_1), \dots, (t_m, b_m)$. Find $\alpha_0, \alpha_1 \in \mathbb{R}$ such that $b_k \approx \alpha_0 + \alpha_1 t_k$.

Solution: Let

$$b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}, \; A = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \text{ and } \alpha = \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix}.$$

Then $\alpha = (A^\top A)^{-1}A^\top b$ which can be written as

$$\alpha = \begin{bmatrix} m & \sum_{k=1}^m t_k \\ \sum_{k=1}^m t_k & \sum_{k=1}^m t_k^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{k=1}^m b_k \\ \sum_{k=1}^m t_k b_k \end{bmatrix}.$$

R5.3.3: If the columns of A are pairwise orthogonal, then $A^\top A$ is a diagonal matrix, which is easy to invert. In this case this corresponds to $\sum_{k=1}^m t_k = 0$. Then the formula for α simplifies to

$$\alpha = \begin{bmatrix} \frac{1}{m}\sum_{k=1}^m b_k \\ (\sum_{k=1}^m t_k b_k)/(\sum_{k=1}^m t_k^2) \end{bmatrix}.$$

Orthonormal bases, Gram-Schmidt

D5.4.1: Vectors $q_1, \ldots, q_n \in \mathbb{R}^m$ are *orthonormal* iff they are orthogonal and have norm 1. In other words, for all $i, j \in [n]$ we have $q_i^\top q_j = \delta_{ij}$. In this case for Q with columns q_i we have $Q^\top Q = I$.

D5.4.3: $Q \in \mathbb{R}^{n\times n}$ is orthogonal iff $Q^\top Q = I$. In this case $QQ^\top = I, Q^{-1} = Q^\top$ and the columns form an orthonormal basis for $\mathbb{R}^n$.

P5.4.6: Let $Q \in \mathbb{R}^{m\times m}$ be orthogonal and $x, y \in \mathbb{R}^m$. Then $\|Qx\| = \|x\|$ and $(Qx)^\top(Qy) = x^\top y$.

P5.4.7: Let S be a subspace of $\mathbb{R}^m$ and $q_1, \ldots, q_n$ be an orthonormal basis for S. Let $Q \in \mathbb{R}^{m\times n}$ with columns q_i. Then the projection matrix that projects onto S is $QQ^\top$ and the Least squares solution to $Qx = b$ is $\hat{x} = Q^\top b$.

Gram-Schmidt process

A5.4.9: Let $a_1, \ldots, a_n$ be linearly independent. Then

- $q_1 = \frac{a_1}{\|a_1\|}$.
- For $k = 2, \ldots, n$ set
  $q'_k = a_k - \sum_{i=1}^{k-1}(a_k^\top q_i)q_i$
  $q_k = \frac{q'_k}{\|q'_k\|}$.

T5.4.10: The Gram-Schmidt process returns an orthonormal basis for the span of $a_1, \ldots, a_n$.

QR decomposition

D5.4.11: Let $A \in \mathbb{R}^{m\times n}$ have linearly independent columns. The *QR decomposition* is $A = QR$ where $Q \in \mathbb{R}^{m\times n}$ is orthonormal (the output of [A5.4.9]) and $R = Q^\top A$.

L5.4.12: In [D5.4.11] R is upper triangular und invertible. Moreover $QQ^\top A = A$ and hence $A = QR$ is well defined.

F5.4.13: The QR decomposition simplifies some calculations:

- $C(A) = C(Q)$ leads to $\mathrm{proj}_{C(A)}(b) = QQ^\top b$.
- $A^\top A\hat{x} = A^\top b$ becomes $R\hat{x} = Q^\top b$.

Pseudoinverse

D5.5.1: Let $A \in \mathbb{R}^{m\times n}, \mathrm{rank}(A) = n$. Then the *pseudoinverse* $A^+ \in \mathbb{R}^{n\times m}$ of A is $A^+ = (A^\top A)^{-1}A^\top$.

P5.5.2: Let $A \in \mathbb{R}^{m\times n}, \mathrm{rank}(A) = n$. Then $A^+A = I$.

D5.5.3: Let $A \in \mathbb{R}^{m\times n}, \mathrm{rank}(A) = m$. Then $A^+ = A^\top(AA^\top)^{-1}$.

L5.5.4: Let $A \in \mathbb{R}^{m\times n}, \mathrm{rank}(A) = m$. Then $AA^+ = I$.

(10): $\min_{x\in\mathbb{R}^n} \|x\|^2$.

L5.5.5: Let $A \in \mathbb{R}^{n\times n}$, $b \in C(A)$, the (unique) solution to (10) is given by $\hat{x} \in C(A^\top)$ that satisfies the constraint $A\hat{x} = b$.

P5.5.6: For a full row rank matrix A, the unique solution to (10) is given by $\hat{x} = A^+b$.

D5.5.7: Let $A \in \mathbb{R}^{m\times n}$ with CR decomposition $A = CR$. Then $A^+ = R^+C^+ = R^\top(C^\top AR^\top)^{-1}C^\top$.

L5.5.8: Let $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^n$. The (unique) solution to $\min_{x\in\mathbb{R}^n} \|x\|^2$ s.t. $A^\top Ax = A^\top b$, is given by $\hat{x} = A^+b$.

P5.5.9: Let $A \in \mathbb{R}^{m\times n}$, $\mathrm{rank}(A) = r$ and $S \in \mathbb{R}^{m\times r}$, $T \in \mathbb{R}^{r\times n}$ such that $A = ST$. Then $A^+ = T^+S^+$.

T5.5.11: Let $A \in \mathbb{R}^{m\times n}$.

1. $AA^+A = A$.
2. $A^+AA^+ = A^+$.
3. AA^+ is symmetric and projects on $C(A)$.
4. A^+A is symmetric and projects on $C(A^\top)$.
5. $(A^\top)^+ = (A^+)^\top$

P5.5.12: Let $A \in \mathbb{R}^{m\times n}$. Then $f : C(A^\top) \to C(A)$, $f : x \mapsto Ax$ is a bijection.

Farkas' lemma

D5.6.1: Let $A \in \mathbb{Q}^{m\times n}$, $b \in \mathbb{Q}^m$ and $P = \{x \in \mathbb{R}^n \mid Ax \leqslant b\}$. P is called a *polyhedron*. Let $S = [s]$. The *projection of* P on the subspace $\mathbb{R}^s$ associated with the variables in the subset S is $\mathrm{proj}_S(P) := \{x \in \mathbb{R}^s \mid \exists y \in \mathbb{R}^{n-s}$ such that $(x, y) \in P\}$.

P5.6.2: $P \neq \varnothing \iff l \leqslant u \iff 0 \leqslant u - l \iff 0 \leqslant y^\top b$ for all $y \geqslant 0$ such that $y^\top a = 0$.

A: Let $A \in \mathbb{Q}^{m\times n}$, $b \in \mathbb{Q}^m$ and $P = \{x \in \mathbb{R}^n \mid Ax \leqslant b\}$. Let the entries of A be denoted by a_{ij}. Then row i gives us the inequality $\sum_{j=1}^n a_{ij}x_j \leqslant b_i$. Let $\bar{x} = (x_1, \ldots, x_{n-1})$ and $\bar{A}$ consist of the first $n-1$ columns of A. Consider the following algorithm.

1. Partition the indices $M = [m]$ of the rows into three subsets $M_0 = \{i \in M \mid a_{i,n} = 0\}$, $M_+ = \{i \in M \mid a_{i,n} > 0\}$ and $M_- = \{i \in M \mid a_{i,n} < 0\}$.
2. - For every row with index $i \in M_+$ multiply the corresponding constraint by $\frac{1}{a_{in}}$. This gives a new representation of row i as $x_n \leqslant d_i + f_i^\top\bar{x}$ for $i \in M_+$ where $d_i = \frac{b_i}{a_{in}}$, $f_{ij} = -\frac{a_{ij}}{a_{in}}$.
   - Every row with index $k \in M_0$ can be rewritten as $0 \leqslant d_k + f_k^\top\bar{x}$ for $k \in M_0$ where $d_k = b_k$, $f_{kj} = -a_{kj}$.
   - For every row with index $i \in M_-$ multiply the corresponding constraint by $\frac{1}{a_{in}}$. This gives a new representation of row i as $x_n \geqslant d_i + f_i^\top\bar{x}$ for $i \in M_-$ where $d_i = \frac{b_i}{a_{in}}$, $f_{ij} = -\frac{a_{ij}}{a_{in}}$.
3. Return $Q = \{\bar{x} \in \mathbb{R}^{n-1} \mid 0 \leqslant d_k + f_k^\top\bar{x}$ for all $k \in M_0, d_l + f_k^\top\bar{x} \leqslant d_i + f_i^\top\bar{x}$ for all $l \in M_-, i \in M_+\}$

T5.6.3: The set Q returned in Step 3 is a polyhedron. Moreover $Q = \mathrm{proj}_S(P)$, where $S = [n-1]$.

L5.6.4: Let $A \in \mathbb{Q}^{m\times n}$, $b \in \mathbb{Q}^n$ and $P = \{x \in \mathbb{R}^n \mid Ax \leqslant b\}$. Let $S_1 = [n-1]$ and $S_2 = [n-2]$. Then $\mathrm{proj}_{S_2}(P) = \mathrm{proj}_{S_2}(\mathrm{proj}_{S_1}(P))$.

D5.6.5: Let $A \in \mathbb{Q}^{m\times n}$, $b \in \mathbb{Q}^m$ and $P = \{x \in \mathbb{R}^n \mid Ax \leqslant b\}$. For $k \in [j]$ let $A^{(j)}$ be the submatrix of A with column vectors $A_{\cdot k}$. Let $P^{(0)} = P$ and $C^{(0)} = \mathbb{R}^m_+$. Define for $i \in [n] = C^{(i)} = (1)$ and $P^{(i)} = (2)$

$(1) = \{y \in \mathbb{R}^m_+ \mid y^\top A_{\cdot k} = 0$ for all $k = n-i+1, \ldots, n\}$

$(2) = \{\hat{x} \in \mathbb{R}^{n-1} \mid y^\top A^{(n-i)}\hat{x} \leqslant y^\top b$ for all $y \in C^{(i)}\}$

T5.6.6: $\mathrm{proj}_{S_{n-i}}(P) = P^{(i)}$.

Farkas' lemma

T5.6.7: Let $A \in \mathbb{Q}^{m\times n}$, $b \in \mathbb{Q}^m$. Either there exists an $x \in \mathbb{R}^n$ such that $Ax \leqslant b$ or there exists a $y \in \mathbb{R}^m$ such that $y \geqslant \mathbf{0}$, $y^\top A = \mathbf{0}$ and $y^\top b < 0$.

Determinant

D6.0.4: Let $\sigma : [n] \to [n]$ be a permutation of n elements. The sign $\mathrm{sgn}(\sigma)$ counts the parity of the numer of pairs of elements that are out of order after applying σ (1 if even, -1 if odd).

D6.0.6: Let $A \in \mathbb{R}^{n\times n}$. The *determinant* is defined as $\det(A) := \sum_{\sigma\in\Pi_n} \mathrm{sgn}(\sigma)\prod_{i=1}^n A_{i,\sigma(i)}$, where Π_n is the set of all permutations of n elements.

P6.0.7: Let $P \in \mathbb{R}^{m\times m}$ be the permutation matrix corresponding to σ. Then $\det(P) = \mathrm{sgn}(\sigma)$.

P6.0.8: Let $T \in \mathbb{R}^{m\times m}$ be triangular. Then $\det(T) = \prod_{k=1}^n T_{kk}$. In particular $\det(I) = 1$.

T6.0.9: Let $A \in \mathbb{R}^{m\times m}$. Then $\det(A^\top) = \det(A)$.

P6.0.10: Let $Q \in \mathbb{R}^{m\times m}$ be orthogonal. Then $\det(Q) = \pm 1$.

P6.0.11: A matrix $A \in \mathbb{R}^{m\times m}$ is invertible iff $\det(A) \neq 0$.

P6.0.12: Let $A, B \in \mathbb{R}^{m\times m}$. Then $\det(AB) = \det(A)\det(B)$.

P6.0.13: Let $A, B \in \mathbb{R}^{m\times m}$ with $\det(A) \neq 0$. Then $\det(A^{-1}) = \frac{1}{\det(A)}$.

D5.0.15: Let $A \in \mathbb{R}^{m\times m}$ and let $\mathscr{A}_{ij}$ denote the $(m-1)\times(m-1)$ matrix obtained by removing row i and colum j from A. Then we define the co-factors of A as $C_{ij} = (-1)^{i+j}\det(\mathscr{A}_{ij})$.

P5.0.16: Let $A \in \mathbb{R}^{m\times m}$, $i \in [n]$. Then $\det(A) = \sum_{j=1}^n A_{ij}C_{ij}$.

P5.0.17: Let $A \in \mathbb{R}^{m\times m}$ with $\det(A) \neq 0$ and C the matrix with the cofactors of A. Then $A^{-1} = \frac{1}{\det(A)}C^\top$.

P6.0.19 (Cramer's rule): Let $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$ with $\det(A) \neq 0$. Then the solution to $Ax = b$ is given by $x_j = \frac{\det(\mathscr{B}_j)}{\det(A)}$ where $\mathscr{B}_j$ is obtained by replacing the jth column of A with b.

P6.0.21: Let $A \in \mathbb{R}^{n\times n}$ and P a permutation that swaps two elements. Then $\det(PA) = -\det(A)$.

P6.0.22: The determinant is linear in each row or each column. In other words, for any $a_0, \ldots, a_n \in \mathbb{R}^n$ and $\alpha_0, \alpha_1 \in \mathbb{R}$ we have

$$\begin{vmatrix} — & \alpha_0 a_0^\top + \alpha_1 a_1^\top & — \\ — & a_2^\top & — \\ & \vdots & \\ — & a_n^\top & — \end{vmatrix} = \alpha_0 \begin{vmatrix} — & a_0^\top & — \\ — & a_2^\top & — \\ & \vdots & \\ — & a_n^\top & — \end{vmatrix} + \alpha_1 \begin{vmatrix} — & a_1^\top & — \\ — & a_2^\top & — \\ & \vdots & \\ — & a_n^\top & — \end{vmatrix}$$

and the same but transposed.

Eigenvalues and Eigenvectors

Complex numbers

D (Complex numbers):

- $(a+ib) + (x+iy) = (a+x) + i(b+y)$
- $(a+ib)(x+iy) = (ax-by) + i(ay+bx)$
- $(a+ib)(a-ib) = a^2+b^2$
- $\frac{a+ib}{x+iy} = \frac{(x-iy)(a+ib)}{(x-iy)(x+iy)} = \frac{ax+by}{x^2+y^2} + i\frac{bx-ay}{x^2+y^2}$

D (Notation):

- $\Re(a+ib) := a$
- $\Im(a+ib) := b$
- $|z| := \sqrt{a^2+b^2}$ (*modulus*)
- $\overline{a+ib} := a - ib$ (*complex conjugate*)

F7.0.1: Let $\theta \in \mathbb{R}$. Then $e^{i\theta} = \cos\theta + i\sin\theta$.

F7.0.2: A complex number $z \in \mathbb{C}$ can be written as $z = re^{i\theta}$ where $r \geqslant 0$ is the *modulus* and $\theta \in \mathbb{R}$ is the *argument*.

T7.0.3 (Fundamental theorem of algebra): Any degree n non-constant polynomial $P(z) = \alpha_n z^n + \cdots + \alpha_1 z + \alpha_0$ with $\alpha_n \neq 0$ has a zero: $\lambda \in \mathbb{C}$ such that $P(\lambda) = 0$.

C7.0.4: Any degree n non-constant polynomial $P(z) = \alpha_n z^n + \cdots + \alpha_1 z + \alpha_0$ has n zeros: $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$, perhaps with repetitions, such that $P(z) = \alpha_n(z-\lambda_1)\cdots(z-\lambda_n)$. The number of times $\lambda \in \mathbb{C}$ appears in the expansion is called the *algebraic multiplicity* of the zero.

D: Let $A \in \mathbb{C}^{m\times n}$. Then $A^* := \overline{A}^\top$.

F: Let $v, w \in \mathbb{C}^n$. Then $\|v\|^2 = v^*v = \bar{v}\top v$. Furthermore $\langle v, w\rangle = w^*v$.

Introduction to Eigenvalues and Eigenvectors

Problem: Find the explicit representation of the linear recurrence $F_0 = 0, F_1 = 1$ and $F_n = F_{n-1} + F_{n-2}$.

Solution: Let

$$M = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix},\ g_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } g_n = \begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix}.$$

Then $g_n = Mg_{n-1} = \cdots = M^n g_0$. We now solve $0 = \det(M - \lambda I)$ and get $\lambda_1 = \frac{1+\sqrt{5}}{2}$ and $\lambda_2 = \frac{1-\sqrt{5}}{2}$. We find v_k, the non-zero element of $N(M - \lambda_k I)$ for $k = 1, 2$. We write $g_0 = \alpha_1 v_1 + \alpha_2 v_2$ and get $\alpha_1 = \frac{1}{\sqrt{5}}$ and $\alpha_2 = -\frac{1}{\sqrt{5}}$. Then $g_n = A^n g_0 = A^n(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 \lambda_1^n v_1 + \alpha_2 \lambda_2^n v_2$.

D7.1.1: Let $A \in \mathbb{R}^{n\times n}$. Then $\lambda \in \mathbb{C}$ is an *eigenvalue (EW)* and $v \in \mathbb{C}^n \backslash \{0\}$ is an *eigenvector (EV)* of A iff $Av = \lambda v$.

P7.1.2: Let $A \in \mathbb{R}^{n\times n}$. Then $\lambda \in \mathbb{R}$ is a (real) EW of A iff $\det(A - \lambda I) = 0$ and $v \in \mathbb{R}^n$ is an EV associated with λ iff it is a nonzero element of $N(A - \lambda I)$.

P7.1.3: $\det(A - \lambda I)$ is a polynomial in λ of degree n. The coefficient of the λ^n term is $(-1)^n$.

T7.1.4: Every matrix $A \in \mathbb{R}^{n\times n}$ has an EW (perhaps in $\mathbb{C}$).

P7.1.6: Let λ and v be an EW-EV pair of the matrix A. Then for $k \geqslant 1$, λ^k and v are an EW-EV pair of the matrix A^k.

P7.1.7: Let λ and v be an EW-EV pair of the invertible matrix A. Then $\frac{1}{\lambda}$ and v are an EW-EV pair of the matrix A^{-1}.

P7.1.8: Let $A \in \mathbb{R}^{n\times n}$ and let $v_1, \ldots, v_k \in \mathbb{R}^n$ be EWs corresponding to EVs $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$. If $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ are pairwise distinct, the EVs $v_1, \ldots, v_k$ are linearly independent.

T7.1.9: Let $A \in \mathbb{R}^{n\times n}$ with n pairwise distinct real EVs (see [T7.1.8], [C7.0.4]) then there is a basis of $\mathbb{R}^n$ made up of the EVs of A.

P7.1.10: Let $A \in \mathbb{R}^{n\times n}$. Then A and $A^\top$ have the same EWs.

D7.1.11: Let $A \in \mathbb{R}^{n\times n}$. The *trace* of A is $\mathrm{Tr}(A) := \sum_{i=1}^n A_{ii}$.

(33) (Characteristic polynomial of A): $(-1)^n \det(A - zI) = \det(zI - A) = (z - \lambda_1)\cdots(z - \lambda_n)$

P7.1.12 Let $A \in \mathbb{R}^{n\times n}$ and $\lambda_1, \ldots, \lambda_n$ its n EWs as they show up in (33) (meaning that a value may be repeated). Then $\mathrm{Tr}(A) = \sum_{i=1}^n \lambda_i$ and $\det(A) = \prod_{i=1}^n \lambda_i$.

R7.1.13: [P7.1.12] can be useful to check computations.

L7.1.14: Let $A, B, C \in \mathbb{R}^{n\times n}$. Then

1. $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$,
2. $\mathrm{Tr}(ABC) = \mathrm{Tr}(BCA) = \mathrm{Tr}(CAB)$.

R7.1.15: Important words of caution:

1. The EWs of A and $A^\top$ are the same, the EVs might not!
2. The EWs of $A + B$ are not easily computed from the EWs of A and B!
3. The EWs of $A \cdot B$ are not easily computed from the EWs of A and B!
4. Gauss elimination does not preserve EWs or EVs! The EWs are not the diagonal elements of U in the $PA = LU$ factorization.

P7.1.17: Let $Q \in \mathbb{R}^{n\times n}$ be orthogonal and $\lambda \in \mathbb{C}$ an EW of Q. Then $|\lambda| = 1$.

D7.1.20: If, given $A \in \mathbb{R}^{n\times n}$ we can build a basis of $\mathbb{R}^n$ with EVs of A, we say that A has a complete set of real EVs.

P7.1.21: Let P be a projection matrix. Then P has at most two EWs, 0 and 1, and a complete set of real EVs.

D7.1.22: Let $A \in \mathbb{R}^{n\times n}$ and λ an EW of A. Then we call the dimension of $N(A - \lambda I)$ the *geometric multiplicity* of λ.

Diagonalizing and change of basis

T7.2.1: Let $A \in \mathbb{R}^{n\times n}$ be a matrix with a complete set of real EVs (see [D7.1.20]) and let $v_1, \ldots, v_n \in \mathbb{R}^n$ be a basis formed with EVs of A and let $\lambda_1, \ldots, \lambda_n$ be the associated EWs. Let V be the matrix whose columns are the v_is. Then $A = V\Lambda V^{-1}$, where Λ is a diagonal matrix with $\Lambda_{ii} = \lambda_i$.

D7.2.2: $A \in \mathbb{R}^{n\times n}$ is called *diagonalizable* iff there exists an invertible matrix V such that $V^{-1}AV = \Lambda$, where Λ is a diagonal matrix.

D7.2.3: $A, B \in \mathbb{R}^{n\times n}$ are called *similar* iff there exists an invertible matrix S such that $B = S^{-1}AS$.

P7.2.4: Similar matrices have the same EWs.

R7.2.5: If we have a matrix $A \in \mathbb{R}^{n\times n}$ with a complete set of real EVs then [T7.2.1] tells us that the corresponding linear transformation, when viewed in the basis $v_1, \ldots, v_n$ is simply a diagonal matrix.

Symmetric matrices and Spectral theorem

Spectral theorem

T7.3.1: Let $A \in \mathbb{R}^{n\times n}$ be symmetric. Then A has n real EWs and an orthogonal basis made of EVs of A.

C7.3.2: Let $A \in \mathbb{R}^{n\times n}$ be symmetric. Then there exists an orthogonal matrix $V \in \mathbb{R}^{n\times n}$ (whose columns are EVs of A) such that $A = V\Lambda V^\top$, where $\Lambda \in \mathbb{R}^{n\times n}$ is a diagonal matrix with the eigenvalues of A on its diagonal.

R7.3.3: The decomposition in [C7.3.2] and [T7.2.1] is called *eigendecomposition*.

C7.3.4: The rank of a real symmetric matrix A is the number of non-zero eigenvalues (counting repetitions).

R7.3.5: Let $A \in \mathbb{R}^{n\times n}$. Then $\mathrm{rank}(A) = n - \dim(N(A))$ which is the geometric multiplicity of $\lambda = 0$ (see [D7.1.22]). Since symmetric matrices always have a complete set of EWs and EVs, the geometric multiplicities are always the same as the algebraic multiplicities.

P7.3.6: Let $A \in \mathbb{R}^{n\times n}$ be symmetric and let $v_1, \ldots, v_n$ be an orthonormal basis of EVs of A (the columns of V in [C7.3.2]) and $\lambda_1, \ldots, \lambda_n$ the associated EWs. Then $A = \sum_{k=1}^n \lambda_i v_i v_i^\top$.

P7.3.7: Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix and $\lambda \in \mathbb{C}$ an EW of A. Then $\lambda \in \mathbb{R}$.

C7.3.8: Every symmetric matrix $A \in \mathbb{R}^{n\times n}$ has a real EW λ.

Rayleigh quotient

P7.3.10: Let $A \in \mathbb{R}^{n\times n}$. Then the *Rayleigh quotient* defined for $x \in \mathbb{R}^n \backslash \{0\}$ as $R(x) := \frac{x^\top Ax}{x^\top x}$ attains its maximum at $R(v_{\max}) = \lambda_{\max}$ and its minimum at $R(v_{\min}) = \lambda_{\min}$ where $\lambda_{\max}$ ($\lambda_{\min}$) is the largest (smallest) EW.

D7.3.11: $A \in \mathbb{R}^{n\times n}$ is said to be *positive semidefinite (PSD)* (*positive definite (PD)*) iff all its EWs are nonnegative (strictly positive).

P7.3.12: Let $A \in \mathbb{R}^{n\times n}$ be symmetric. Then A is PSD (PD) iff $x^\top Ax \geqslant 0$ ($x^\top Ax > 0$) for all $x \in \mathbb{R}^n \backslash \{0\}$.

F7.3.13: Let $A, B \in \mathbb{R}^{n\times n}$ be PSD (PD). Then their sum is PSD (PD).

D7.3.14 (Gram matrix): Let $v_1, \ldots, v_n \in \mathbb{R}^m$. We define their *Gram matrix* $G \in \mathbb{R}^{n\times n}$ as $G_{ij} = v_i^\top v_j$. Note that if $V \in \mathbb{R}^{m\times n}$ has the v_is as columns then $G = V^\top V$.

R7.3.15: Let $A \in \mathbb{R}^{m\times n}$. As an abuse of notation we also call $AA^\top$ a Gram matrix of A. If $a_1, \ldots, a_n \in \mathbb{R}^m$ are the columns of A then $AA^\top \in \mathbb{R}^{m\times m}$ and $AA^\top = \sum_{i=1}^n a_i a_i^\top$.

P7.3.16: Let $A \in \mathbb{R}^{m\times n}$. Then the non-zero EWs of $A^\top A \in \mathbb{R}^{n\times n}$ are the same as the ones of $AA^\top \in \mathbb{R}^{m\times m}$. Both matrices are symmetric and PSD.

Cholesky decomposition

P7.3.17: Every symmetric PSD matrix M is a gram matrix of an upper triangular matrix C. $M = C^\top C$ is known as the *Cholesky decomposition*.

Singular value decomposition

Singular value decomposition (SVD)

D8.1.1: Let $A \in \mathbb{R}^{m\times n}$. There exist orthogonal matrices $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ such that $A = U\Sigma V^\top$, where $\Sigma \in \mathbb{R}^{m\times n}$ is a diagonal matrix, where the diagonal elements are non-negative and ordered in descending order. The columns $u_1, \ldots, u_m$ ($v_1, \ldots, v_n$) of U (V) are called the *left* (*right*) *singular vectors* of A and are orthonormal. The diagonal elements of Σ, $\sigma_i = \Sigma_{ii}$ are called the *singular values* of A and are ordered as $\sigma_1 \geqslant \cdots \geqslant \sigma_{\min\{m,n\}}$.

R8.1.2: Let $A \in \mathbb{R}^{m\times n}$, $\mathrm{rank}(A) = r$. We can write the SVD in a more compact form $A = U_r\Sigma_r V_r^\top$ where $U_r \in \mathbb{R}^{m\times r}$ ($V_r \in \mathbb{R}^{n\times r}$) contains the first r left (right) singular vectors, and $\Sigma_r \in \mathbb{R}^{r\times r}$ is a diagonal matrix with the first r singular values. This requires considerably less space for a large matrix with small rank.

R8.1.3: Let $A \in \mathbb{R}^{m\times n}$ and $A = U\Sigma V^\top$ be its SVD [D8.1.1]. Then $AA^\top = U(\Sigma\Sigma^\top)U^\top$. Thus the left singular vectors of A, the columns of U are the EVs of $AA^\top$ and the singular values of A are the square-roots of the EWs of $AA^\top$. Note that $\Sigma\Sigma^\top \in \mathbb{R}^{m\times m}$ is diagonal. If $m > n$, A has n singular values and $AA^\top$ has m EWs (which are larger than n but the "missing" ones are 0). Analogously, $A^\top A = V(\Sigma^\top\Sigma)V^\top$), and so the right singular vectors of A, the columns of V, are the EVs of $A^\top A$ and the singular values of A are the square-roots of the EWs of $A^\top A$. Note that $\Sigma^\top\Sigma \in \mathbb{R}^{n\times n}$ is diagonal. If $n > m$ A has m singular values and $A^\top A$ has n EWs (which are larger than m but the "missing" ones are 0). This observation makes it easier to write the singular values/vectors of A in terms of EWs and EVs of $AA^\top$ and $A^\top A$, which are symmetric. This directly implies the uniqueness of singular values and the fact that the rank of a matrix is the number of non-zero singular values.

P8.1.4: Let $A \in \mathbb{R}^{m\times n}$, $\mathrm{rank}(A) = r$. Let $\sigma_1, \ldots, \sigma_r$ be the non-zero singular values of A, $u_1, \ldots, u_r$ ($v_1, \ldots, v_r$) the corresponding left (right) singular vectors. Then $A = \sum_{k=1}^r \sigma_k u_k v_k^\top$.

T8.1.5: Every $A \in \mathbb{R}^{m\times n}$ has an SVD (see [D8.1.1]).

Vector and matrix norms

D (l_p-norm): For $1 \leqslant p \leqslant \infty$ the l_p-norm is given by $\|x\|_p := \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$.

D8.2.1 (Frobenius and Spectral norm): Let $A \in \mathbb{R}^{m\times n}$. Then:

1. $\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2}$,
2. $\|A\|_{op} = \max_{x\in\mathbb{R}^n \text{ s.t. } \|x\|=1n} \|Ax\|$.

P: Let $A \in \mathbb{R}^{m\times n}$ with singular values $\sigma_1, \geqslant \cdots \geqslant \sigma_{\min\{m,n\}}$. Then:

1. $\|A\|_F^2 = \mathrm{Tr}(A^\top A)$
2. $\|A\|_F^2 = \sum_{i=1}^{\min\{m,n\}} \sigma_i^2$
3. $\|A\|_{op} = \sigma_1$
4. $\|A\|_{op} \leqslant \|A\|_F \leqslant \sqrt{\min\{m,n\}}\|A\|_{op}$.

Appendix

Notation

R: The word *iff* stands for "if and only if".

R: The abbreviation *s.t.* stands for "such that".

D: Let $n \in \mathbb{Z}^+$. Then $[n] := \{1, \ldots, n\}$.

D: A function is *bijective* iff it is invertible.

D: Let V be a vector space. Then $V^\perp$ denotes the orthogonal complement of V.

D (Kronecker delta):

$$\delta_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

D: Let $A \in \mathbb{R}^{m \times n}$ with entries a_{ij}. Then

$$\begin{vmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{vmatrix} = \det(A).$$

Symbols

A	Algorithm
C	Corollary
D	Definition
F	Fact
L	Lemma
O	Observation
P	Proposition
R	Remark
T	Theorem