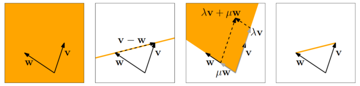
1 Vectors

D 1.4 Linear Combinations: $\sum_{i=1}^{n} \lambda_i \mathbf{v}_i$ are scaled combinations of n vectors $\mathbf{v}_i$.

D 1.7 Combination Types:
(i) **Affine:** $\sum_{i=1}^{n} \lambda_i = 1$
(ii) **Conic:** $\lambda_i \geq 0$ for $i = 1, 2, \ldots, n$
(iii) **Convex:** *Both* Affine and Conic.

D 1.9 Scalar/Dot Product:
$\mathbf{v} \cdot \mathbf{w} := \sum_{i=1}^{m} v_i w_i$
Just multiply component wise, add all components together. Results are in $\mathbb{R}$.



(i) linear (ii) affine (iii) conic (iv) convex

D 1.11 Euclidean norm, squared norm, unit vector: $\|\mathbf{v}\| := \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{\mathbf{v}^\top \mathbf{v}}$, **Squared Norm:** $\|\mathbf{v}\|^2 := \mathbf{v}^\top \mathbf{v}$, **Unit Vector:** $\|\mathbf{u}\| = 1$, computed as $\frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{1}{\|\mathbf{v}\|}\mathbf{v}$ (for any vector $\mathbf{v} \neq 0$)

L 1.12 Cauchy-Schwarz inequality: $|\mathbf{v} \cdot \mathbf{w}| \leq \|\mathbf{v}\|\|\mathbf{w}\|$ for any two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^m$. Equality holds iff $\mathbf{v} = \lambda \mathbf{w}$ or $\mathbf{w} = \lambda \mathbf{v}$ for some scalar $\lambda \in \mathbb{R}$.

D 1.14 Angles: $\cos(\alpha) = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\|\|\mathbf{w}\|} \in [-1, 1]$ If $\mathbf{v} \perp \mathbf{w} \in \mathbb{R}^m$, then $\mathbf{v} \cdot \mathbf{w} = 0$.

D 1.16 Hyperplane through origin: Let $\mathbf{d} \in \mathbb{R}^m, \mathbf{d} \neq 0, H_{\mathbf{d}} = \{\mathbf{v} \in \mathbb{R}^m : \mathbf{v} \cdot \mathbf{d} = 0\}$

L 1.17 Triangle inequality: $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$

D 1.21 Linear (in)dependence: Vectors are linearly **independent** if:
a) No vector is a linear combination of the others.
b) There are no λ_i (except all 0), such that $\sum_{i=1}^{n} \lambda_i \mathbf{v}_i = 0$.

Linearly dependent: At least one vector is a linear combination of the others.
For matrices: columns are linearly independent $\Leftrightarrow (A\mathbf{x} = 0 \Rightarrow \mathbf{x} = 0)$.

L 1.22 Other definitions of linear dependence: Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{R}^m$. Statements are equivalent:
(i) At least one vector is a linear combination of the others.
(ii) There are scalars $\lambda_1, \lambda_2, \ldots, \lambda_n$ besides $0, \ldots, 0$, such that $\sum_{j=1}^{n} \lambda_j \mathbf{v}_j = 0$ (0 is a *non-trivial* linear combination of the vectors)
(iii) At least one of the vectors is a linear combination of the previous ones.

D 1.25 Span: Set of all linear combinations of vectors:
$\text{Span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) := \left\{\sum_{j=1}^{n} \lambda_j \mathbf{v}_j : \lambda_j \in \mathbb{R} \text{ for all } j \in [n]\right\}$

L 1.26: Given a set of vectors: $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^m$ and $\mathbf{v} \in \mathbb{R}^m$ be a linear combination of those vectors, then adding this combination does not change the span:
$\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v})$

Construction of vectors with standard unit vectors: Every vector in a vector space (D 4.1) can be written as: $\mathbf{u} = \sum_{i=1}^{m} u_i \mathbf{e}_i$, where $\mathbf{e}_i$ is a standard unit vector (just one component being 1, all others 0).

L 1.28 Span of m linearly independent vectors is $\mathbb{R}^m$: Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m \in \mathbb{R}^m$ be linearly independent. Then $\text{Span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m) = \mathbb{R}^m$.

2 Matrices

D 2.1 Matrix: $A = [a_{ij}]$ with $i = 1, \ldots, m$ and $j = 1, \ldots, n$ - m rows, n columns (first rows, then columns)

D 2.2 Matrix addition, scalar multiplication: *Addition:* $A + B = [a_{ij} + b_{ij}]$ with $i = 1, \ldots, m$ and $j = 1, \ldots, n$,
Scalar multiplication: $\lambda A = [\lambda a_{ij}]$ with $i = 1, \ldots, m$ and $j = 1, \ldots, n$

D 2.3 Square matrices: *Identity matrix*: square matrix, diagonals 1, $A = AI = IA$;
Diagonal matrix: $a_{ij} = 0$ if $i \neq j$; *Triangle matrix:* **lower** if $a_{ij} = 0$ for $i < j$, **upper** if $a_{ij} = 0$ for $i > j$;
Symmetric matrix: $a_{ij} = a_{ji} \forall i, j, A^\top = A$; *Skew-symmetric matrix:* $a_{ij} = -a_{ji} \forall i, j, A = -A^\top$

D 2.4 Matrix-Vector Product: For $A \in \mathbb{R}^{m \times n}$ and $\mathbf{x} \in \mathbb{R}^n$, the product $A\mathbf{x} \in \mathbb{R}^m$ is computed by taking each row of A and computing its dot product with $\mathbf{x}$. Equivalently, $A\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{a}_j$ where $\mathbf{a}_j$ are the columns of A. Note: $I\mathbf{x} = \mathbf{x}$. **Trace:** Sum of the diagonal entries of a matrix.

O 2.5: Let A be an $m \times n$ matrix. (i) A vector $\mathbf{b} \in \mathbb{R}^m$ is a linear combination of the columns of A iff. there's a vector $\mathbf{x} \in \mathbb{R}^n$ (of suitable scalars), such that $A\mathbf{x} = \mathbf{b}$. (ii) The columns of A are linearly independent iff. $\mathbf{x} = 0$ is the only vector such that $A\mathbf{x} = 0$.

D 2.9 Column space: $C(A) := \{A\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\}$, i.e.: the span (set of all linear combinations) of the column vectors.

D 2.10 Rank: $\text{rank}(A)$:= the number of linearly independent column vectors, also sometimes called column rank.

L 2.11 Independent columns span the column space: Let A be an $m \times n$ matrix with r independent columns, and let C be the $m \times r$ submatrix containing the independent columns. Then $C(A) = C(C)$.

D 2.12 Transpose: Mirror the matrix along its diagonal. $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \leftrightarrow A^\top = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$

O 2.13 Transpose twice & transposing symmetric matrices: $(A^\top)^\top = A$ Moreover, a square matrix A is symmetric iff. $A = A^\top$

D 2.14 Row Space: $R(A) := C(A^\top)$. With the *row* rank of A, being the *column* rank of $A^\top$.

D 2.17 Nullspace: Nullspace contains all input vectors that lead to output vector 0. $N(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = 0\}$. It can easily be obtained from RREF by setting $\mathbf{b} = 0$ in $R\mathbf{x} = \mathbf{b}$. If there are any free variables, choose any real number satisfying the condition.
To find the basis, rewrite and apply this Lemma: $R\mathbf{x} = 0 \Leftrightarrow I \cdot x(I) + Q \cdot x(Q) = 0$ e.g. for $R = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & -2 \end{bmatrix}$ to $I\,[x_1, x_3] + \begin{bmatrix} 2 & 3 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} x_2 \\ x_4 \end{bmatrix} = 0 \Leftrightarrow x(I) = -Q \cdot x(Q)$.
We may then freely choose the *free variables* $x(Q)$. Then find *basis variables* $x(I)$ using the above, typically choose $e_1, \ldots, e_k$ for $x(Q)$ to get the kth vector of basis (k = number of columns of Q). Finally, combine the vectors into one.

D 2.27 Kernel and Image:
- **Kernel:** $N(A) = \mathbf{Ker}(T) := \{\mathbf{x} \in \mathbb{R}^n : T(\mathbf{x}) = 0\} \subseteq \mathbb{R}^n$ (If A is the unique $m \times n$ matrix, such that $T = T_A$)
- **Image:** $C(A) = \mathbf{Im}(T) := \{T(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n\} \subseteq \mathbb{R}^m$ (If A is the unique $m \times n$ matrix, such that $T = T_A$), the set of all outputs that T can produce.

L 2.23: A function $T : \mathbb{R}^n \to \mathbb{R}^m / T : \mathbb{R}^n \to \mathbb{R}$ is a *linear transformation / linear functional* iff. these two linearity axioms hold for all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^n$ and all $\lambda \in \mathbb{R}$:
(i) $T(\mathbf{x} + \mathbf{x}') = T(\mathbf{x}) + T(\mathbf{x}')$
(ii) $T(\lambda \mathbf{x}) = \lambda T(\mathbf{x})$

Visualizing linear transformations: A matrix can be seen as a re-mapping the unit-vectors $\hat{\imath}, \hat{\jmath}, \hat{k}, \ldots$, scaling and re-orienting them. Each column vector can be seen as the new unit vector $\mathbf{e}_i$. For example, $R(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$ would be a rotation matrix, that rotates the plane counterclockwise by θ.
To prove that a transformation T is linear, use Lemma 2.23. $A\mathbf{x} = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{v}_i$, where $\mathbf{v}_i$ is the ith column of A.

O 2.39 Matrix multiplication: $A \times B = C, c_{ij} = \sum_{k=1}^{n} a_{i,k} b_{k,j}$. Dimension restrictions: A is an $m \times n$ matrix, B is $n \times p$, the result C will be $m \times p$. For each entry, multiply the ith row of A with the jth column of B.
This is **NOT** commutative, but associative & distributive. The right-to-left order matters, not the position of any parenthesis.

L 2.40 Matrix multiplication with transposition: $(AB)^\top = B^\top A^\top, (A^\top)^\top = A$

D 2.44 Outer product: $\text{rank}(A) = 1 \Leftrightarrow \exists$ non-zero vectors $\mathbf{v} \in \mathbb{R}^m, \mathbf{w} \in \mathbb{R}^n$ such that A is an outer product, i.e. $A = \mathbf{v}\mathbf{w}^\top$, thus $\text{rank}(\mathbf{v}\mathbf{w}^\top) = 1$.

T 2.46 CR Decomposition: $A = CR$. Get R from (reduced) row echelon form, C is the columns from A where there is a pivot in R. $C \in \mathbb{R}^{m \times r}$, $R \in \mathbb{R}^{r \times n}$ (in RREF), $r = \text{rank}(A)$.
To find **REF** try to create pivots: $R_0 = \begin{bmatrix} \mathbf{1} & 0 & 2 & 3 \\ 0 & \mathbf{1} & 2 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. Use Gauss-Jordan elimination to find it.
RREF is the fully reduced form where all rows with non-zero entries have a leading 1 (pivot), and all other entries in pivot columns are 0.

3 Linear Equations

Solving $A\mathbf{x} = \mathbf{b}$: Overview: Get the system into $A\mathbf{x} = \mathbf{b}$ form. *(Use this if ranks are easy to determine; otherwise proceed with Gaussian Elimination below.)* Three outcomes:
- **No solution:** $\text{rank}(A) < \text{rank}([A|\mathbf{b}])$ (inconsistent)
- **Unique solution:** $\text{rank}(A) = n$ (full column rank, $m \geq n$). Also requires consistency check: $\text{rank}(A) = \text{rank}([A|\mathbf{b}])$.
- **Infinite solutions:** $\text{rank}(A) = \text{rank}([A|\mathbf{b}]) < n$ (underdetermined)

Gaussian Elimination / Gauss-Jordan: **Method:** Transform A into upper-triangle (REF) or fully reduced (RREF) via row operations.
- Augment: $[A|\mathbf{b}]$
- Row reduce using: swap rows, multiply by scalar, add multiple of one row to another
- Back-substitute or read off free variables
- Runtime: $\mathcal{O}(m^3)$ for square matrices

O 2.56 Invertible matrix: Matrix A is invertible, if it is square and there exists B, such that: $AB = I \Leftrightarrow BA = I \Leftrightarrow AB = BA = I$

D 2.57 Inverse matrix: If $AB = I$ for invertible A, then B is its inverse, denoted as A^{-1}.

O 2.58 Inverse of the inverse: $(A^{-1})^{-1} = A$

L 2.59: If A and B are invertible $m \times m$ matrices, AB is also invertible, and $(AB)^{-1} = B^{-1}A^{-1}$

L 2.60: If A is an invertible $m \times m$ matrix, its transpose is also invertible, and $(A^\top)^{-1} = (A^{-1})^\top$

D 3.13 Reduced Row Echelon Form: Let $R = [r_{ij}]$ with $i = 1, \ldots, m$ and $j = 1, \ldots, n$ be an $m \times n$ matrix. R is in RREF, if there is some natural number $r \leq m$ and column indices $1 \leq j_1 < j_2 < \cdots < j_r \leq n$ (*the indices of the "downward step") such that the following two conditions hold:
(i) For every $i \in [r]$, column j_i of R is the standard unit vector $\mathbf{e}_i$.
(ii) All entries r_{ij} "below the staircase" are 0.

L 3.14: A matrix R in RREF $(j_1, j_2, \ldots, j_r)$ has independent **pivot** columns $j_1, j_2, \ldots, j_r$ and therefore rank r.

Gauss-Jordan elimination: Makes Gaussian elimination possible for $m \times n$ matrices and works similarly. Transform the augmented matrix $[A|\mathbf{b}]$ into RREF:
1. Swap rows, so the entry with the largest absolute value is the pivot a_{ij}
2. For each row, use the pivot to clear all entries *below* it using $R_i \leftarrow R_i - \left(\frac{\text{target}}{\text{pivot}}\right) R_{\text{pivot}}$.
3. Normalize all pivots to 1 by dividing the entire row by the pivot value.
4. Clear all entries **above** the pivots using row additions.

After reaching RREF, check the last row(s) $[0 \ldots 0|c]$:
- **No Solution:** If $c \neq 0$ (impossible equation $0 = c$)
- **Unique Solution:** A square identity matrix on the left, the right side are the solved variables $\mathbf{b}_i$.
- **Infinite Solutions:** Row of zeroes $[0 \ldots 0|0]$

A general RREF Structure:
$$\left[\begin{array}{ccc|c} 1 & 0 & \text{free} & b_1 \\ 0 & 1 & \text{free} & b_2 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

4 Four fundamental Subspaces

4.1 Vector Spaces

D 4.1 Vector Space: Vector space is a triple $(V, +, \cdot)$ where V is a set of vectors, satisfying the vector space axioms, *commutativity, associativity, existence of zero and negative vectors and identity element* (1), *compatibility of* $\bigoplus$ with $\cdot$ (in $\mathbb{R}$), *distributivity over* $\bigoplus$ $(\lambda(\mathbf{v} + \mathbf{w}) = \lambda\mathbf{v} + \lambda\mathbf{w})$ and *distributivity over* $+$ (in $\mathbb{R}$) $((\lambda + \mu)\mathbf{v} = \lambda\mathbf{v} + \mu\mathbf{v})$.
To **define a vector space**, we need to define addition and scalar multiplication for the elements in a ***canonical*** way (*according to the accepted standard.*)

D 4.8 Subspace: Let V be a vector space. A nonempty subset $U \subseteq V$ is a subspace of V if these two axioms are true for all $\mathbf{v}, \mathbf{w} \in U$ and all $\lambda \in \mathbb{R}$:

$$\mathbf{v} + \mathbf{w} \in U \quad \text{and} \quad \lambda\mathbf{v} \in U$$

They guarantee that vector addition and scalar multiplication do not take us outside the subspace.

L 4.9 Subspace always has 0: Let $U \subseteq V$ be a subspace of V. Then $0 \in U$.

L 4.11 Column space is a subspace: Let $A \in \mathbb{R}^{m\times n}$, then $C(A) = \{A\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\}$ is a subspace of $\mathbb{R}^m$. Moreover, $R(A) = C(A^\top)$ is a subspace of $\mathbb{R}^n$.

E 4.13 Nullspace is a subspace: Let $A \in \mathbb{R}^{m\times n}$. Then the nullspace $\mathbf{N}(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = 0\}$ is a subspace of $\mathbb{R}^n$.

L 4.14 Subspaces are vector spaces: V is a vector space and U is its subspace. Then U is also a vector space with the same $\bigoplus$ and $\bigodot$ as V.

4.2 Bases and dimension

D 4.18 Basis: Let V be a vector space. A subset $B \subseteq V$ is called a basis of V if B is linearly independent and it spans V : Span$(B) = V$.

L 4.19 Independent columns is a basis: A set of linearly independent vectors that spans a subspace V forms a basis of V. For $\mathbb{R}^m$, the set of standard unit vectors is a basis. For a matrix, all linearly independent columns form a basis of the column space $C(A)$. **Calculating:** If a matrix has full column rank or full row rank, then the basis consists of all the linearly independent column or row vectors.

O 4.20 Non-uniqueness of basis: Every set $B = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\} \subseteq \mathbb{R}^m$ of m linearly independent vectors is a basis of $\mathbb{R}^m$.

D 4.21 Finitely generated vector space: A vector space V is called finitely generated if there exists a finite subset $G \subseteq V$ with Span$(G) = V$. Then V has a basis $B \subseteq G$.

T 4.22 Finitely generated VS has a basis: If V is finitely generated, then V has a basis $B \subseteq V$.

L 4.23 Steinitz exchange lemma: Let $F \subseteq V$ be a finite set of linearly independent vectors, and $G \subseteq V$ a finite set of vectors with Span$(G) = V$. Then $|F| \leq |G|$, and there exists a subset $E \subseteq G$ of size $|G| - |F|$ such that Span$(F \cup E) = V$.

T 4.24 All bases have the same size: All bases of a vector space V have the same size. If B and B' are both bases of V, then $|B| = |B'|$.

D 4.25 Dimension: If V is finitely generated, then $d =$ dim(V) is the size of any basis B of V.

D 4.26 Linear transformation between vector spaces: Let V, W be vector spaces. A function $T : V \to W$ is linear if, for all $\mathbf{x}_1, \mathbf{x}_2 \in V$ and $\lambda_1, \lambda_2 \in \mathbb{R}, T(\lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2) = \lambda_1 T(\mathbf{x}_1) + \lambda_2 T(\mathbf{x}_2)$.

L 4.27 Bijective linear transformations preserve basis: If $T : V \to W$ is a bijective linear map, then $B \subseteq V$ is a basis of $V \Leftrightarrow T(B)$ is a basis of W, and hence dim$(V) =$ dim(W).

D 4.28 Isomorphic vector spaces: $V \cong W \Leftrightarrow \exists T : V \to W$ linear and bijective.

T 4.29 Basis writes vectors as unique linear combinations: Let V be a finitely generated vector space with basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$. Then every $\mathbf{v} \in V$ can be written uniquely as $\mathbf{v} = \sum_{j=1}^m \lambda_j \mathbf{v}_j$, for unique scalars $\lambda_1, \lambda_2, \dots, \lambda_m \in \mathbb{R}$.

L 4.30 Less than dim(V) vectors do not span V: If $|G| <$ dim(V), then Span$(G) \neq V$.

4.3 Computing the three fundamental subspaces

T 4.31 Basis of $C(A)$: Pivot columns of RREF: If R is the RREF of A, then the pivot columns of R form a basis of $C(A)$. Moreover, dim$(C(A)) =$ rank$(A) = r$.

T 4.32 Basis of $\mathbf{R}(A)$: Non-zero rows of RREF of A: Non-zero rows of RREF of A form a basis of $\mathbf{R}(A)$, so dim$(\mathbf{R}(A)) = r$

T 4.33 Row rank equals column rank: rank$(A) =$ rank$(A^\top)$

C 4.34 Rank is at most min of the matrix dimensions: A is an $m \times n$ matrix with rank $r \Rightarrow r \leq$ min(n, m).

L 4.35 Nullspace isomorphism: If $R =$ RREF(A), then $T : N(R) \to \mathbb{R}^{n-r}$ is an isomorphism between $N(R)$ and $\mathbb{R}^{n-r}$. Thus dim$(N(R)) = n - r$.

T 4.36 Basis of $N(A)$: Non-pivot columns of RREF(A): If rank$(A) = r$, then dim$(N(A)) = n - r$.

4.4 All solutions of $A\mathbf{x} = \mathbf{b}$

D 4.37 Solution space: Set of all solutions of $A\mathbf{x} = \mathbf{b}$, thus Sol$(A, \mathbf{b}) := \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{b}\} \subseteq \mathbb{R}^n$.

T 4.38 Solution space from shifting the nullspace: Let $\mathbf{s}$ be some solution of $A\mathbf{x} = \mathbf{b}$, then Sol$(A, \mathbf{b}) := \{\mathbf{s} + \mathbf{x} \in \mathbb{R}^n : \mathbf{x} \in N(A)\}$. We can also compute Sol$(A, \mathbf{b})$, although it is not a subspace. To describe all solutions, we need *some* solutions.

T 4.39 Dimension of a solution space: Let $A \in \mathbb{R}^{m\times n}$ with rank r. If $A\mathbf{x} = \mathbf{b}$ is solvable, then dim$(\mathbf{Sol}(A, \mathbf{b})) = n - r =$ dim$(N(A))$.

T 4.40 Systems of rank m are solvable: Let $A \in \mathbb{R}^{m\times n}$ with rank$(A) = m$. Then $A\mathbf{x} = \mathbf{b}$ is solvable for all $\mathbf{b} \in \mathbb{R}^m$.

T 4.41 Systems of rank less than m are typically unsolvable: Systems of rank $r < m$ are typically unsolvable.

D 4.42 Types of systems: Let $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. The system $A\mathbf{x} = \mathbf{b}$ is classified as:
- $m = n \Rightarrow$ square (A is a square matrix) — **typically solvable if rank $r = m = n$.**
- $m < n \Rightarrow$ underdetermined (A is a wide matrix) — **typically solvable if rank $r = m$.**
- $m > n \Rightarrow$ overdetermined (A is a tall matrix) — **typically unsolvable**.

"Typical" matrices are with $m \leq n$ and have rank $r = m$.

5 Orthogonality and Projections

5.1 Definition

Orthogonality: A geometric and algebraic tool in order to be able to decompose a space into subspaces.

D 5.1.1 Orthogonal subspaces: Two vectors are orthogonal if their scalar product is 0: $\mathbf{v}^\top\mathbf{w} = \sum_{i=1}^n v_i w_i = 0$. Two subspaces are orthogonal if all $\mathbf{v}$ and $\mathbf{w}$ are orthogonal.

L 5.1.2 Orthogonality of bases: Let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ and $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_l$ be bases of subspaces V and W. V and W are orthogonal iff. $\mathbf{v}_i$ and $\mathbf{w}_j$ are orthogonal for all $i \in \{1, \dots, k\}$ and $j \in \{1, \dots, l\}$.

L 5.1.3 Linear independence of bases of orthogonal subspaces: Let V and W be two orthogonal subspaces of $\mathbb{R}^n$ with the bases $\mathbf{v}_1, \dots, \mathbf{v}_k$, $\mathbf{w}_1, \dots, \mathbf{w}_l$. The set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{w}_1, \dots, \mathbf{w}_l\}$ is linearly independent.

C 5.1.4 Combinations of subspaces: Let V and W be orthogonal subspaces. Then $V + W$ is a subspace of $\mathbb{R}^n$, and $V \cap W = \{0\}$ and their direct sum is $V \oplus W = \{\lambda\mathbf{v} + \mu\mathbf{w} : \lambda, \mu \in \mathbb{R}, \mathbf{v} \in V, \mathbf{w} \in W\}$. If dim$(V) = k$ and dim$(W) = l$, then dim$(V \oplus W) = k + l \leq n$, for $V, W \subseteq \mathbb{R}^n$.

D 5.1.5 Orthogonal Complement: Let V be a subspace of $\mathbb{R}^n$, its *orthogonal complement*: $V^\perp = \{\mathbf{w} \in \mathbb{R}^n | \mathbf{w}^\top\mathbf{v} = 0, \forall\mathbf{v} \in V\}$.

T 5.1.6 Relations between subspaces: $N(A) = C(A^\top) = R(A)^\perp$ and $C(A^\top) = N(A)^\perp$

T 5.1.7 Vector decomposition by orthogonal complements: $W = V^\perp \Leftrightarrow$ dim$(V) +$ dim$(W) = n \Leftrightarrow \mathbf{u} = \mathbf{v} + \mathbf{w}, \forall\mathbf{u} \in \mathbb{R}^n$ with unique vectors $\mathbf{v} \in V, \mathbf{w} \in W$.

L 5.1.10 Justification of existing solutions for normal equations: Let $A \in \mathbb{R}^{m\times n}$. Then $N(A) = N(A^\top A)$ and $C(A^\top) = C(A^\top A)$.

5.2 Projections

D 5.2.1 Projections: Projecting a vector onto a subspace is done with proj$_S(\mathbf{b}) =$ argmin$_{\mathbf{p}\in S}\|\mathbf{b} - \mathbf{p}\|$: and yields the closest point in the new subspace S.

L 5.2.2 One-dimensional projection formula: Projection of $\mathbf{b}$ on $S = \{\lambda\mathbf{a} | \lambda \in \mathbb{R}\} = C(\mathbf{a})$: proj$_S(\mathbf{b}) = \frac{\mathbf{a}^\top\mathbf{b}}{\mathbf{a}^\top\mathbf{a}}\mathbf{a}$, where $\mathbf{a} \in \mathbb{R}^m \setminus \{0\}$. We note, that $(\mathbf{b} -$ proj$_S(\mathbf{b})) \perp$ proj$_S(\mathbf{b})$, i.e. the "error-vector" is perpendicular.

L 5.2.3 General Projection Formula: Let S be a subspace in $\mathbb{R}^m$ with basis $\mathbf{a}_1, \dots, \mathbf{a}_n$ that span S. Let A be the matrix with column vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$. The general formula: proj$_S(\mathbf{b}) = A\hat{\mathbf{x}}$, where $\hat{\mathbf{x}}$ satisfies $A^\top A\hat{\mathbf{x}} = A^\top\mathbf{b}$.

L 5.2.4 Properties of $A^\top A$: $A^\top A$ is invertible $\Longleftrightarrow A$ has linearly independent columns. $\Longrightarrow$ $A^\top A$ is a square matrix, symmetric, and invertible.

T 5.2.5 Projection in terms of projection matrix: proj$_S(\mathbf{b}) = P\mathbf{b}$ with projection matrix $P = A(A^\top A)^{-1}A^\top$. A is a matrix given in a task.

6 Applications of Orthogonality and Projections

6.1 Least Squares Approximation

Least Squares: Approximate a solution to a system of equations. Find $\mathbf{x}$ for which $A\mathbf{x}$ is as close as possible to $\mathbf{b}$: min$_{\hat{\mathbf{x}}\in\mathbb{R}^n}\|A\hat{\mathbf{x}} - \mathbf{b}\|^2$. Using the normal equations, we get $A^\top A\hat{\mathbf{x}} = A^\top\mathbf{b}$.

(i) Calculate $M = A^\top A$
(ii) Calculate $\mathbf{b}' = A^\top\mathbf{b}$
(iii) Solve resulting System of Equations $M\hat{\mathbf{x}} = \mathbf{b}'$ as usual.

Linear regression: Application of least squares problem, in which it is to find A and $\mathbf{b}$ such that we can solve the system. We define a matrix $A = \begin{bmatrix} 1 & \mathbf{t}_1 \\ \vdots & \vdots \\ 1 & \mathbf{t}_n \end{bmatrix}$ and a result vector $\mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_n \end{bmatrix}$ where n is the total number of data points and $\mathbf{t}_i$ is the slope of the ith function, where $\mathbf{b}_i$ is its output. The first column is all 1s because the constant element has no scalar. This comes from the following concept: $f(t) = \alpha_0 + \alpha_1 t$, so if the first data point is $(1, 2)$, we get $\alpha + \alpha_1 \cdot 1 = 2$, which will then transform into a SLE with other equations.

L 6.1.2: If A has linearly *dependent* columns, $\mathbf{t}_i = \mathbf{t}_j, \forall i \neq j$.

6.2 The set of all solutions to a system of linear equations

L 6.2.1 Injectivity of A on $C(A^\top)$, uniqueness of solutions: $A \in \mathbb{R}^{m\times n}, \mathbf{x}, \mathbf{y} \in C(A^\top)$: $A\mathbf{x} = A\mathbf{y} \Leftrightarrow \mathbf{x} = \mathbf{y}$, which leads to: $C(A^\top) \cap N(A) = \{0\}$

T 6.2.2 Set of all solutions of linear equations: Suppose the set of all solutions, $\{\mathbf{x} \in \mathbb{R}^n | A\mathbf{x} = \mathbf{b}\} \neq \emptyset$, then $\{\mathbf{x} \in \mathbb{R}^n | \mathbf{x} = \mathbf{b}\} = \mathbf{x}_1 + N(A), \mathbf{x}_1 \in R(A)$ is unique such that $A\mathbf{x}_1 = \mathbf{b}$.

C 6.2.3: Suppose that $\{\mathbf{x} \in \mathbb{R}^n | A\mathbf{x} = \mathbf{b}\} \neq \emptyset$. Then there exists a unique vector $\mathbf{x}_1 \in C(A^\top A)$ such that $A\mathbf{x}_1 = \mathbf{b}$.

T 6.2.4 Linear equations with no solution: For linear equations that have no solutions, these statements are equivalent:

$$\{\mathbf{x} \in \mathbb{R}^n | A\mathbf{x} = \mathbf{b}\} = \emptyset \Longleftrightarrow \{\mathbf{z} \in \mathbb{R}^m | A^\top\mathbf{z} = 0, \mathbf{b}^\top\mathbf{z} = 1\} \neq \emptyset$$

6.3 Orthogonal Bases and Gram Schmidt

D 6.3.1 Orthogonal vectors: $\mathbf{q}_i^\top \mathbf{q}_i = \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$, with δ_{ij} being the *Kronecker delta*.

D 6.3.2 Orthogonal matrix: A square matrix $Q \in R^{n \times m}$ is an *orthogonal matrix* when $Q^\top Q = I$. If it is square, then $QQ^\top = I, Q^{-1} = Q^\top$, and the columns of Q form an orthogonal basis for R^n.

P 6.3.6 Preserving qualities of orthogonal matrices: Orthogonal matrices preserve norm and inner product of vectors: $\|Q\mathbf{x}\| = \|\mathbf{x}\|$ and $(Q\mathbf{x})^\top(Q\mathbf{y}) = \mathbf{x}^\top\mathbf{y}$

P 6.3.7 Least square solution to $Q\mathbf{x} = \mathbf{b}$: The Least Squares solution to $Q\mathbf{x} = \mathbf{b}$, where Q is the matrix whose columns are the vectors forming the orthogonal basis of $S \subseteq \mathbb{R}^m$, is given by $\hat{\mathbf{x}} = Q^\top \mathbf{b}$ and the projection matrix is given by $QQ^\top$.

D 6.3.8 Gram-Schmidt algorithm: This algorithm is used to construct orthogonal bases. We have linearly independent vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$, that span a subspace S, then Gram-Schmidt constructs $\mathbf{q}_1, \dots, \mathbf{q}_n$ by:

1. $\mathbf{q}_1 = \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}$
2. For $k = 2, \dots, n$, $\mathbf{q}_k' = \mathbf{a}_k - \sum_{i=1}^{k-1}(\mathbf{a}_k^\top \mathbf{q}_i)\mathbf{q}_i$
3. Finally, normalize: $\mathbf{q}_k = \frac{\mathbf{q}_k'}{\|\mathbf{q}_k'\|}$

D 6.3.10 QR Decomposition: $A = QR$, where $R = Q^\top A$ and Q is obtained from the Gram-Schmidt process, is made up of the vectors $\mathbf{q}_i$ as columns.

L 6.3.11 Well-Defined QR Decomposition: R is an upper triangular matrix and invertible. Moreover, $QQ^\top A = A$, and hence $A = QR$ is well-defined.

P 6.3.12: This greatly simplifies calculations involving projections and least squares, since $C(A) = C(Q)$, so $\text{proj}_{C(A)}(\mathbf{b}) = QQ^\top \mathbf{b}$ and for least squares, we have $R\hat{\mathbf{x}} = Q^\top \mathbf{b}$. Tis can efficiently be solved using back-substitution because R is triangular.

6.4 Pseudoinverse

D 6.4.1 Left Pseudoinverse (Full column rank): For $A \in \mathbb{R}^{m \times n}$ with full-column $\text{rank}(A) = n$, we get pseudoinverse $A^\dagger \in \mathbb{R}^{n \times m}$ as $A^\dagger = (A^\top A)^{-1} A^\top$. $A^\dagger$ is a left inverse: $A^\dagger A = I$.

D 6.4.3 Right Pseudoinverse (Full row rank): For $A \in \mathbb{R}^{m \times n}$ with full row rank, $\text{rank}(A) = m$ we get $A^\dagger \in \mathbb{R}^{n \times m}$ as $A^\dagger = A^\top (AA^\top)^{-1}$. $A^\dagger$ is a right inverse: $AA^\dagger = I$.

D 6.4.7 CR Decomposition with pseudoinverse: For $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) = r$ and a *CR*-decomposition $A = CR$, we define $A^\dagger = R^\top C^\top$.
In general, $A^\dagger = R^\top (C^\top C)^{-1} = R^\top (C^\top CRR^\top)^{-1} C^\top = R^\top (C^\top AR^\top)^{-1} C^\top$.

L 6.4.8 Unique solution of least squares with pseudoinverse: For any matrix A and vector $\mathbf{x} \in C(A)$, then unique solution of the least squares problem is given by a vector $\hat{\mathbf{x}} \in C(A)$ satisfying $A\hat{\mathbf{x}} = \mathbf{b}$. The solution is $\hat{\mathbf{x}} = A^\dagger \mathbf{b}$, with $A\hat{\mathbf{x}} = \mathbf{b}$, and in the general case $A^\dagger = R^\top C^\top = R^\top (C^\top AR^\top)^{-1} C^\top$.

P 6.4.9 TS Decomposition: For $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) = r$, let $S \in \mathbb{R}^{m \times r}, T \in \mathbb{R}^{r \times n}$ such that $A = ST$. Then, $A^\dagger = T^\dagger S^\dagger$.

T 6.4.10 Properties of Pseudoinverse: Let $A \in \mathbb{R}^{m \times n}$.

- $AA^\dagger A = A$ and $A^\dagger AA^\dagger = A^\dagger$ and $(A^\top)^\dagger = (A^\dagger)^\top$.
- $AA^\dagger$ is symmetric, and the projection matrix for the projection on $C(A)$.
- $A^\dagger A$ is symmetric, and the projection matrix for the projection on $C(A^\top)$.

Moreover, $AA^\dagger = CRR^\top (RR^\top)^{-1} (C^\top C)^{-1} C^\top = C(C^\top C)^{-1} C^\top$, which is the projection onto $C(A)$, and $(AA^\dagger)^\top = AA^\dagger$.

7 The Determinant

The determinant can be understood as a number that corresponds to *how much* the associated linear transformation scales space. For example. a 2D linear transformation with a determinant 2, will scale any area in the space up by 2 *after* the linear transformation has been applied.

7.1 2 times 2

D 7.1.1 2×2 Determinant: For $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $\det(A) = ad - bc$.

L 7.1.2 Multiplication of determinants: $\det(AB) = \det(A) \cdot \det(B)$

L 7.1.3 Invertibility related to the determinant: A matrix $A \in \mathbb{R}^{2 \times 2}$ is invertible iff. $\det(A) \neq 0$.

D 7.2.1 Permutation sign: The sign of a permutation is defined as the number of swaps of rows or columns. det(permutation matrix) $= (-1)^k$det(original matrix), where k is the number of swaps. Even number of swaps $\Rightarrow +1$, odd number $\Rightarrow -1$.

$$\text{sgn}(\sigma) = \begin{cases} 1 & \text{if } |(i,j) \in \{1,\dots,n\} \times \{1,\dots,n\} \text{ such that } i < j \text{ and } \sigma(i) > \sigma(j)| \text{ even} \\ -1 & \text{if } |(i,j) \in \{1,\dots,n\} \times \{1,\dots,n\} \text{ such that } i < j \text{ and } \sigma(i) > \sigma(j)| \text{odd} \end{cases}$$

7.2 General case

D 7.2.3 Determinant big formula: For a square matrix $A \in \mathbb{R}^{m \times m}$, $\det(A) = \sum_{\sigma \in \prod_n} \text{sgn}(\sigma) \prod_{i=1}^n A_{i,\sigma(i)}$, *(Number of permutations: $n!$)*

Determinant Properties:

1. If matrix $T \in \mathbb{R}^{n \times n}$ is triangular, then $\det(T) = \prod_{k=1}^n T_{kk}$, in particular $\det(I) = 1$.
2. Matrix $A \in \mathbb{R}^{n \times n}$, $\det(A) = \det(A^\top)$
3. Matrix $Q \in \mathbb{R}^{n \times n}$ is orthogonal $\iff \det(Q) = 1$ or $\det(Q) = -1$.
4. Matrix $A \in \mathbb{R}^{n \times n}$ is invertible $\iff \det(A) \neq 0$
5. Matrices $A, B \in \mathbb{R}^{n \times n}$, $\det(AB) = \det(A)\det(B)$, in particular $\det(A^n) = \det(A)^n$
6. Matrix $A \in \mathbb{R}^{n \times n}$, $\det(A^{-1}) = \frac{1}{\det(A)}$
7. $\det(\lambda A) = \lambda^n \det(A)$

P 7.2.4 Determinant of orthogonal matrices:

- Given a permutation matrix $P \in \mathbb{R}^{m \times n}$ corresponding to a permutation σ, then $\det(P) = \text{sgn}(\sigma)$. We sometimes also write sgn(P).
- Given a triangular (either upper- or lower) matrix $T \in \mathbb{R}^{n \times n}$ we have $\det(T) = \prod_{k=1}^{n} T_{kk}$, in particular, $\det(I) = 1$.
- If $Q \in \mathbb{R}^{n \times n}$ is an orthogonal matrix then $\det(Q) = 1$ or $\det(Q) = -1$.

$1 = \det(I) = \det(Q^\top Q) = \det(Q^\top)\det(Q) = \det(Q)^2$, so $\det(Q) = \pm 1$. If the determinant is 1, then Q is a rotation matrix. If the determinant is -1 it's a reflection matrix.

P 7.3.2 Cofactor determinant calculation: Let $A \in \mathbb{R}^{n \times n}$, for any $1 \leq i \leq n$, $\det(A) = \sum_{j=1}^{n} A_{ij} C_{ij}$, where the cofactors are $C_{ij} = (-1)^{i+j} \det(A_{ij})$

P 7.3.5 Cramer's Rule: The idea here is that we solve a linear system of type $A\mathbf{x} = \mathbf{b}$, then, due to the determinant being multiplicative, we can solve for each component. The solution $\mathbf{x} \in \mathbb{R}^n$ for $A\mathbf{x} = \mathbf{b}$ is given by $x_j = \frac{\det(A_j)}{\det(A)}$, where A_j is the matrix obtained from A by replacing the j-th column with $\mathbf{b}$.

P 7.3.6 Swapping rows permutation matrix: If $A \in \mathbb{R}^{n \times n}$ and P is a permutation matrix that swaps two elements, meaning that PA corresponds to swapping two rows of A, then $\det(PA) = -\det(A)$.

P 7.3.7 Linearity of the determinant: The determinant is linear in each row (and column). For example:

$$\det \begin{bmatrix} \alpha_0 \mathbf{a}_0^\top + \alpha_1 \mathbf{a}_1^\top \\ \mathbf{a}_2^\top \end{bmatrix} = \alpha_0 \det \begin{bmatrix} \mathbf{a}_0^\top \\ \mathbf{a}_2^\top \end{bmatrix} + \alpha_1 \det \begin{bmatrix} \mathbf{a}_1^\top \\ \mathbf{a}_2^\top \end{bmatrix}$$

8 Eigenvalues and Eigenvectors

8.1 Complex Numbers

Operations: $i^2 = -1$ (**NOT** $i = \sqrt{-1}$, since otherwise $1 = -1$). Complex number $z_j = a_j + b_j i$. *Addition, Subtraction* $(a_1 \pm a_2) + (b_1 \pm b_2)i$. *Multiplication* $(a_1 a_2 - b_1 b_2) + (a_1 b_2 + a_2 b_1)i$. *Division* $\frac{a_1 b_1 + a_2 b_2}{b_1^2 + b_2^2} + \frac{a_2 b_1 - a_1 b_2}{b_1^2 + b_2^2} i$

Parts: $\Re(a + bi) := a$ (Real part), $\Im(a + bi) := b$ (imaginary part), $|z| := \sqrt{a^2 + b^2}$ (modulus), $\overline{a + bi} := a - bi$ (complex conjugate)

R 8.1.1 Euler's formula: For $\theta \in \mathbb{R}$, $e^{i\theta} = \cos\theta + i \sin\theta \Rightarrow e^{i\pi} = -1$

Polar form of a complex number: $z = re^{i\theta}$, $z \in \mathbb{C}$, $r > 0$ is the modulus of z, $\theta \in [0, 2\pi]$.

T 8.1.2 Fundamental Theorem of Algebra: Any degree n non-constant ($n \geq 1$) polynomial $P(z) = \alpha_n z^n + a_{n-1} z^{n-1} + \dots + \alpha_1 z + \alpha_0$, ($\alpha_n \neq 0$) has a zero: there exist $\lambda \in \mathbb{C}$ such that $P(\lambda) = 0$.
A degree n polynomial has at most n distinct zeros (roots).

C 8.1.3 Algebraic multiplicity, number of 0: in polynomial Any degree n non-constant ($n \geq 1$) polynomial has n zeros $\lambda_1, \dots, \lambda_n \in \mathbb{C}$, and $P(z) = \alpha_n (z - \lambda_1)(z - \lambda_2) \dots (z - \lambda_n)$. The number of times $\lambda \in \mathbb{C}$ appears in the expression is called the *algebraic multiplicity* of zero.

Inner product on $\mathbb{C}^n$ and Conjugate Transpose: The inner product on $\mathbb{C}^n$ is given by $\langle v, w \rangle = w^* v$.
$A^* = \bar{A}^\top$

8.2 Introduction to Eigenvectors and Eigenvalues

D 8.2.1 Eigenvector / Eigenvalue pair: Given $A \in \mathbb{R}^{n \times n}$, we say $\lambda \in \mathbb{C}$ is an *eigenvalue* of A and $\mathbf{v} \in \mathbb{C}^n \setminus \{0\}$ is an *eigenvector* of A associated with λ when $A\mathbf{v} = \lambda\mathbf{v}$. $(\lambda, \mathbf{v})$ is an *eigenvalue-eigenvector* pair. If $\lambda \in \mathbb{R}$, then we have a real eigenvalue-eigenvector pair.
Imagine the eigenvectors to be the normalized vectors that ***don't*** *change when applying a linear transformation.*

L 8.2.3 Real Eigenvalues / Eigenvectors: Let $A \in \mathbb{R}^{n \times n}$. Then, $\lambda \in \mathbb{R}$ is a real eigenvalue of A if and only if $\det(A - \lambda I) = 0$. A vector $\mathbf{v} \in \mathbb{R}^n \setminus \{0\}$ is an eigenvector associated with λ if and only if $\mathbf{v} \in N(A - \lambda I)$.

To find an Eigenvalue and Eigenvector of a matrix $M \in \mathbb{R}^{n \times n}$, simply calculate the eigenvalue first, using the zeros of the polynomial obtained from calculating $\det(M - \lambda I)$, which is obtained from L 8.2.3 $\det(M - \lambda I) = 0$. This means, we simply need to calculate the determinant of $M - \lambda I$, which is fairly straightforward. We can then try to find the eigenvectors $\mathbf{v}$ such that $M\mathbf{v} = \lambda\mathbf{v}$, or in other words a non-zero element of $N(M - \lambda I) \setminus \{0\}$, i.e. the null space of $M - \lambda I$. This means we try to find a solution such that $0 = (M - \lambda I)\mathbf{v}$, where $\mathbf{v}$ is not the zero vector.

P 8.2.4 Characteristic polynomial: $(-1)^n \det(A - zI) = \det(zI - A) = (z - \lambda_1)(z - \lambda_2) \dots (z - \lambda_n)$. The coefficient of the λ^n term is $(-1)^n$. Usually determined from $\det(M - \lambda I)$.

T 8.2.5 Existence of eigenvalue: Every matrix $A \in \mathbb{R}^{n \times n}$ has an eigenvalue (perhaps complex-valued).

P 8.2.7 Eigenvalue of orthogonal matrix: If $Q \in \mathbb{R}^{n \times n}$ is an orthogonal matrix. If $\lambda \in \mathbb{C}$ is an eigenvalue of Q, then $|\lambda| = 1$.

L 8.2.8 Complex Eigenvalues exist on conjugate pairs of real A: Let $A \in \mathbb{R}^{n \times n}$. If $(\lambda, \mathbf{v})$ is an eigenvalue-eigenvector pair, then $(\bar{\lambda}, \bar{\mathbf{v}})$ is an eigenvalue-eigenvector pair.

8.3 Properties of Eigenvalues and Eigenvectors

P 8.3.1 Eigenvalue modifications based on the type of matrix:

- If $(\lambda, \mathbf{v})$ is an eigenvalue-eigenvector pair of A, then $(\lambda^k, \mathbf{v})$ is an eigenvalue-eigenvector pair of A^k for $k \geq 1$.
- If $(\lambda, \mathbf{v})$ is an eigenvalue-eigenvector pair of A with $\lambda \neq 0$, then $(\frac{1}{\lambda}, \mathbf{v})$ is an eigenvalue-eigenvector pair of A^{-1}.

L 8.3.2 Linear independence: If $\lambda_1, \ldots, \lambda_n$ are all distinct, the corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent.

T 8.3.3 Existence of basis from Eigenvalues: Let $A \in \mathbb{R}^{n\times n}$ with n distinct real eigenvalues. Then there exists a basis of $\mathbb{R}^n$, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ made of eigenvectors of A.

D 8.3.4 Trace of a matrix: The trace of A is defined by $\text{Tr}(A) = \sum_{i=1}^n A_{ii}$.

L 8.3.5 Transposition equality of Eigenvalues: The eigenvalues of $A \in \mathbb{R}^{n\times n}$ are the same as those of $A^\top$.

L 8.3.6 Determinant and Trace via Eigenvalues: Let $A \in \mathbb{R}^{n\times n}$ and let $\lambda_1, \ldots, \lambda_n$ be its eigenvalues as they appear in the characteristic polynomial. Then, $\det(A) = \prod_{i=1}^n \lambda_i$, $\text{Tr}(A) = \sum_{i=1}^n \lambda_i$.

L 8.3.7 Cyclic invariance of the trace: For $A, B, C \in \mathbb{R}^{n\times n}$: $\text{Tr}(AB) = \text{Tr}(BA)$, then $\text{Tr}(ABC) = \text{Tr}(BCA) = \text{Tr}(CAB)$.

Change of basis: With the linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$ (given by e.g. $\mathbf{x} \in \mathbb{R}^n \to A\mathbf{x} \in \mathbb{R}^m$), for which we want to find a matrix B that maps it from a basis $\mathbf{u}_1, \ldots, \mathbf{u}_n$ to another one $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Now that B helps us map a vector α to a vector β, which has the different basis. We now define U as the matrix whose columns are the first basis and V as the matrix whose columns are the second basis. Now, if $L(\mathbf{x}) = V\beta$ and $\mathbf{x} = U\alpha$, so $\beta = V^{-1}AU\alpha$, now $\beta = V^{-1}AU$.

9 Diagonalization, Singular Value Decomposition

9.1 Diagonalization

T 9.1.1 Diagonalization Theorem, ability changing basis: $A = V\Lambda V^{-1}$, where V's columns are its eigenvectors and Λ is a diagonal matrix with $\Lambda_{ii} = \lambda_i$ and all other entries 0. $A \in \mathbb{R}^{n\times n}$ and has to have a complete set of real eigenvectors (Eigenbasis).
Equivalently, $\Lambda = V^{-1}AV$, since V is invertible.

D 9.1.2 Diagonalizable matrix: A matrix $A \in \mathbb{R}^{n\times n}$ is called *diagonalizable* if there exists an invertible matrix V such that $V^{-1}AV = \Lambda$, where Λ is a diagonal matrix.

D 9.1.3 Complete set of Eigenvectors: If we can find eigenvectors forming a basis of $\mathbb{R}^n$ for A, we say that A has a *complete set of real eigenvectors.*

P 9.1.6 Eigenvalues and Eigenvectors of a projection matrix: Let P be the projection matrix on the subspace $U \subseteq \mathbb{R}^n$. Then P has two eigenvalues, 0 and 1, and a complete set of real eigenvectors.

D 9.1.7 Similar matrices: Matrices $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times n}$ are called *similar* if there exists and invertible matrix S, such that $B = S^{-1}AS$.

P 9.1.8 Similar matrices have the same eigenvalues: Similar matrices $A \in \mathbb{R}^{n\times n}$ and $B = S^{-1}AS \in \mathbb{R}^{n\times n}$ have the same eigenvalues. The matrix A has a complete set of real eigenvectors iff. B does.

D 9.1.10 Geometric multiplicity: Let $A \in \mathbb{R}^{n\times n}$ and let λ be an eigenvalue of A. Then $\dim(N(A - \lambda I))$ is called the *geometric multiplicity* of λ.

L 9.1.11 Complete set of real Eigenvectors: A matrix has a complete set of real eigenvectors iff. all its eigenvalues are real, and the geometric multiplicities equal the algebraic multiplicities for all eigenvalues.

9.2 Symmetric Matrices, Spectral Theorem

T 9.2.1 Spectral Theorem: Any symmetric matrix $A \in \mathbb{R}^{n\times n}$ has n real eigenvalues and an orthogonal basis of $\mathbb{R}^n$ consisting of eigenvectors A.

C 9.2.2 Eigendecomposition: For any symmetric matrix $A \in \mathbb{R}^{n\times n}$, there exists an orthogonal matrix $V \in \mathbb{R}^{n\times n}$ (whose columns are eigenvectors of A) such that $A = V\Lambda V^\top$, where $\Lambda \in \mathbb{R}^{n\times n}$ is diagonal with diagonal entries equal to the eigenvalues of A, and $V^\top V = I$. This decomposition is called the *eigendecomposition.*

C 9.2.4 Rank of real symmetric matrices:
- If A is a real symmetric matrix, then rank(A) is the number of non-zero eigenvalues of A (counting repetitions).
- For a general $n \times n$ matrix, rank$(A) = n - \dim(N(A))$, so the geometric multiplicity of the eigenvalue $\lambda = 0$ equals $\dim(N(A))$.

R 9.2.5: For general $n \times n$ (non-symmetric) matrices, the rank is n minus the dimension of the nullspace, so it is n minus the geometric multiplicity of $\lambda = 0$. Since symmetric matrices always have a complete set of eigenvalues and eigenvectors, the geometric multiplicities are always the same as algebraic multiplicities.

$$\dim(N(A)) + \text{rank}(A) = n$$

P 9.2.6 Rank-One Spectral Decomposition: Let $A \in \mathbb{R}^{n\times n}$ be symmetric, and let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be an orthogonal basis of eigenvectors of A (the columns of V), with associated eigenvalues $\lambda_1, \ldots, \lambda_n$. Then $A = \sum_{k=1}^{n} \lambda_k \mathbf{v}_k \mathbf{v}_k^\top$.
A real symmetric matrix is a weighted sum of orthogonal projections onto its eigenvector directions, with weights given by the eigenvalues.

L 9.2.7 Orthogonality of Eigenvectors: Let $A \in \mathbb{R}^{n\times n}$ be symmetric and let $\lambda_1 \neq \lambda_2 \in \mathbb{R}$ be two distinct eigenvalues of A with corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2$. Then, $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal.

L 9.2.8 Symmetric matrix has real Eigenvalues: A symmetric matrix $A \in \mathbb{R}^{n\times n}$ has only real eigenvalues: $\lambda \in \mathbb{C} \Rightarrow \lambda \in \mathbb{R}$.
If $A\mathbf{v} = \lambda\mathbf{v}$:

$$\bar{\lambda}\|\mathbf{v}\|^2 = \bar{\lambda}\mathbf{v}^*\mathbf{v} = (\lambda\mathbf{v})^*\mathbf{v} = (A\mathbf{v})^*\mathbf{v} = \mathbf{v}^*A^*\mathbf{v} = \mathbf{v}^*A\mathbf{v} = \mathbf{v}^*\lambda\mathbf{v} = \lambda\|\mathbf{v}\|^2$$

$\Longrightarrow$ every symmetric matrix $A \in \mathbb{R}^{n\times n}$ has a real eigenvalue λ **(C 9.2.9)**

P 9.2.10 Rayleigh Quotient: Given a symmetric matrix $A \in \mathbb{R}^{n\times n}$, the Rayleigh Quotient, defined for $\mathbf{x} \in \mathbb{R}^n \setminus \{0\}$, as $R(\mathbf{x}) = \frac{\mathbf{x}^\top A\mathbf{x}}{\mathbf{x}^\top\mathbf{x}}$. The maximum of it is at $R(\mathbf{v}_{\text{max}}) = \lambda_{\text{max}}$ and the minimum correspondingly at the smallest eigenvalue, with λ and $\mathbf{v}$ being the respective minimum and maximum eigenvalue-eigenvector pairs.

D 9.2.11 Positive Semidefinite (PSD) and Positive definite (PD) matrices: A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is said to be PSD if all its eigenvalues are non-negative. If all the eigenvalues of A are strictly positive (no eigenvalue is zero), then we say A is PD.

P 9.2.12 Positivity of the quadratic form: A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is PSD iff. $\mathbf{x}^\top A\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$. Analogously, its PD iff. $\mathbf{x}^\top A\mathbf{x} > 0$ for all $\mathbf{x} \in \mathbb{R}^n$.

D 9.2.13 Gram Matrix: Given vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^n$, their *Gram matrix* is $G \in \mathbb{R}^{n\times n}$ defined by $G_{ij} = \mathbf{v}_i^\top\mathbf{v}_j$ for $i, j \in \{1, \ldots, n\}$. If $V = [\mathbf{v}_1 \ldots \mathbf{v}_n] \in \mathbb{R}^{m\times n}$, then $G = V^\top V$.
If $A = [\mathbf{a}_1 \ldots \mathbf{a}_n] \in \mathbb{R}^{m\times n}$, one also calls $AA^\top$ a Gram matrix (although abuse of notation). Note, that $AA^\top = \sum_{i=1}^n \mathbf{a}_i\mathbf{a}_i^\top$, which will be an $m \times m$ matrix.

P 9.2.15 Same Eigenvalues of transposed matrices: Given a real matrix $A \in \mathbb{R}^{m\times n}$, the non-zero eigenvalues of $A^\top A \in \mathbb{R}^{n\times n}$ and $AA^\top \in \mathbb{R}^{m\times m}$ are the same. Also both are symmetric and PSD.

P 9.2.16 Cholesky Decomposition: Every symmetric PSD matrix M is a gram matrix of an upper triangular matrix C, so that $M = C^\top C$.

9.3 Singular Value Decomposition

D 9.3.1 Singular Value Decomposition: Let $A \in \mathbb{R}^{m\times n}$. There exists 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, in the sense that $\Sigma_{ij} = 0$ when $i \neq j$, and the diagonal elements are non-negative and ordered in descending order. $U^\top U = I$ and $V^\top V = I$.
The columns of U and V are called the left and right singular vectors of A, and the diagonal entries of Σ are the singular values of A, ordered as $\sigma_1 \geq \cdots \geq \sigma_{\text{min}\{m,n\}}$.

R 9.3.2 Compact form of SVD: If A has rank r, then the SVD can be written as $A = U_r\Sigma_r V_r^\top$, where $U_r \in \mathbb{R}^{m\times r}$ and $V_r \in \mathbb{R}^{n\times r}$ have orthonormal columns, and $\Sigma_r \in \mathbb{R}^{r\times r}$ is a diagonal matrix with first r singular values. This representation stores $r(m + n + 1)$ real numbers instead of mn. For small r, this yields substantial savings and motivates low-rank approximations.

T 9.3.3 Every matrix has an SVD: Every matrix $A \in \mathbb{R}^{m\times n}$ has an SVD: $A = U\Sigma V^\top$. Equivalently, every linear transformation is diagonal in orthonormal bases of singular vectors and can be understood in 3 separate steps (the three composing matrices, $V^\top, \Sigma, U$).

P 9.3.4 SVD as a sum of rank-one matrices: Let $A \in \mathbb{R}^{m\times n}$ have rank r, with singular values $\sigma_1, \ldots, \sigma_r$ and corresponding singular vectors $\mathbf{u}_1, \ldots, \mathbf{u}_r$ and $\mathbf{v}_1, \ldots, \mathbf{v}_r$. Then

$$A = \sum_{k=1}^{r} \sigma_k \mathbf{u}_k \mathbf{v}_k^\top$$

We can write any rank-r matrix $A \in \mathbb{R}^{m\times n}$ as a sum of r rank-1 matrices.

10 Strategies

10.1 Systems of Equations

General Solution:
1. Form the augmented matrix $[A|\mathbf{b}]$.
2. Perform **Gauss-Jordan Elimination** to get to RREF.
3. **Consistency Check:** If any row looks like $[0 \ldots 0|\text{non-zero}]$, there is **no solution**.
4. Identify variables: **Pivot variables:** Columns with leading 1s; **Free variables:** Columns without leading 1s.
5. **General Solution:** $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$; $\mathbf{x}_p$: Set free variables to 0, solve for pivots; $\mathbf{x}_h$: Write pivot variables in terms of free variables. Extract free variables as coefficients.

Calculate Inverse Matrix A^{-1}:
1. Form the augmented matrix $[A|I]$.
2. Perform **Gauss-Jordan Elimination**.
3. Once $[I|B]$ is reached, then $B = A^{-1}$. If you get a row of zeros on the left side, A is not invertible (singular).

Linear Independence: To check if vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent:
1. Form matrix $A = [\mathbf{v}_1 \ldots \mathbf{v}_n]$.
2. Perform **Gaussian Elimination** to get REF.
3. If every column has a pivot (no free variables), they are *independent*; otherwise, they are *dependent*.

Calculating the Determinant:
1. **Method A** (2×2) : det $\begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$.
2. **Method B (Triangular):** Product of diagonal entries.
3. **Method C (General** $n \times n$:
   1. Use row operations to convert A to an upper triangular matrix U.
   2. **Track changes:** Row swap: multiply det by -1; Row subtraction $(R_i - kR_j)$: det does **not** change; Scalar multiplication (kR_i): multiply det by k.
4. det(A) = (corresponding factors) $\times \prod \mathbf{u}_{ii}$.

10.2 Fundamental Spaces

Quick Rank Reference: For an $m \times n$ matrix A with rank r:
1. rank(A) = rank$(A^\top)$ (row rank = column rank)
2. rank$(A) \leq \min(m, n)$ (limited by dimensions)
3. rank$(A) = r \Rightarrow \dim(C(A)) = r, \dim(R(A)) = r$
4. rank$(A) = r \Rightarrow \dim(\mathbf{N}(A)) = n - r$ (nullity)
5. Full column rank: rank$(A) = n \Rightarrow$ columns are linearly independent, $\mathbf{N}(A) = \{0\}$
6. Full row rank: rank$(A) = m \Rightarrow A\mathbf{x} = \mathbf{b}$ is solvable for all $\mathbf{b}$
7. Full rank: rank$(A) = m = n \Rightarrow A$ is invertible

Basis for Column Space $C(A)$:
1. Perform **Gaussian Elimination** on A to get R (no need for full RREF).
2. Identify indices of the **pivot columns** in R (e.g. col $1, 3, 4$).
3. **Result:** Select the corresponding columns from the **original** matrix A.

Basis for Row Space $R(A)$: $R(A) = C(A^\top)$
1. Perform **Gaussian Elimination** on A to get R.
2. **Result:** The non-zero rows of R (transposed to be column vectors) from the basis.

Basis for Nullspace $N(A)$:
1. Solve $A\mathbf{x} = 0$ using **Gauss-Jordan** to get RREF.
2. Express pivot variables in terms of free variables.
3. **Result:** The vectors multiplying the free variables form the basis.
4. **Dimension:** $\dim(N(A)) = n - r$ (Columns minus Rank).

10.3 Orthogonality, Projections & Least Squares

Quick Projection of $\mathbf{b}$ onto $\mathbf{a}$: To project a vector $\mathbf{b}$ onto the line spanned by $\mathbf{a}$:
1. **Formula:** $\mathbf{p} = \text{proj}_{\mathbf{a}}(\mathbf{b}) = \frac{\mathbf{a}\cdot\mathbf{b}}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a} = \frac{\mathbf{a}^\top\mathbf{b}}{\|\mathbf{a}\|^2}\mathbf{a}$.
2. **Step 1 (Scalar Part):** Calculate the "overlap": $s = \mathbf{a}^\top\mathbf{b}$.
3. **Step 2 (Normalization):** Calculate squared norm: $n = \mathbf{a}^\top\mathbf{a}$.
4. **Step 3 (Result):** Multiply vector $\mathbf{a}$ by the fraction: $\mathbf{p} = \frac{s}{n}\mathbf{a}$.

Check: The error vector $\mathbf{e} = \mathbf{b} - \mathbf{p}$ must be orthogonal to $\mathbf{a}$ ($\mathbf{a}^\top\mathbf{e} = 0$).

Least Squares Approximation: **Problem:** $A\mathbf{x} = \mathbf{b}$ has no solution $(m > n)$. Find $\hat{\mathbf{x}}$ that minimizes $\|A\mathbf{x} = \mathbf{b}\|^2$.
1. Calculate matrix $M = A^\top A$.
2. Calculate vector $\mathbf{d} = A^\top\mathbf{b}$
3. Solve the system $M\hat{\mathbf{x}} = \mathbf{d}$ (using Gaussian elimination)

Note: If columns of A are independent, $\hat{\mathbf{x}} = (A^\top A)^{-1}A^\top\mathbf{b}$.

Projection of $\mathbf{b}$ onto Subspace S:
1. Find a basis for S and put them as columns in matrix A.
2. Calculate $\hat{\mathbf{x}}$ using *Least Squares*.
3. **Result:** The projection $p = A\hat{\mathbf{x}}$.
4. **Projection Matrix:** $P = A(A^\top A)^{-1}A^\top$.

Gram-Schmidt (Orthonormal Basis):
- **Input:** Independent vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$
- **Output:** Orthonormal vectors $\mathbf{q}_1, \dots, \mathbf{q}_n$.
- Follow **Definition 6.3.8**

10.4 Eigenvalues & Decomposition

Find Eigenvalues and Eigenvectors:
1. **Eigenvalues (λ):** Solve characteristic equation $\det(A - \lambda I) = 0$.
2. **Eigenvectors $(\mathbf{v})$:** For each found λ:
   - Form matrix $(A - \lambda I)$
   - FInd the Nullspace basis of this matrix (solve $(A - \lambda I)\mathbf{v} = 0$).

Diagonalization $(A = V\Lambda V^{-1})$:
1. Find eigenvalues $\lambda_1, \dots, \lambda_n$.
2. Find n independent eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ (If you cannot find n independent vectors, A is not diagonalizable).
3. **Construct Matrices:**
   - Λ : Diagonal matrix with λ_i on diagonal.
   - V : Matrix with eigenvectors $\mathbf{v}_i$ as columns (order must match λ_i).

Spectral Decomposition (Symmetric Matrices): **Condition:** $A = A^\top$; Solved similar to Diagonalization, **but:**
1. Eigenvalues will be *real*
2. Eigenvectors for different λ are automatically orthogonal.
3. **Important:** Normalize the eigenvectors to length 1.
4. **Result:** $A = Q\Lambda Q^\top$ (where Q is orthogonal matrix of normalized eigenvectors).

Singular Value Decomposition (SVD): **Goal:** $A = U\Sigma V^\top$.
1. Compute $M = A^\top A$.
2. Find eigenvalues of $M : \lambda_1, \dots, \lambda_r$ (sorted high to low).
3. **Singular Values:** $\sigma_i = \sqrt{\lambda_i}$, and place these in diagonal Σ.
4. **Find Singular Vectors** (V)**:** Calculate orthonormal vectors of $A^\top A$. These are columns of V.
5. **Left Singular vectors** (U)**:** For non-zero σ_i, calculate $\mathbf{u}_i = \frac{1}{\sigma_i}A\mathbf{v}_i$.
6. (if needed) compute U to be an orthonormal basis using Gram-Schmidt if A is not square / full-rank.

10.5 Quick Sanity Checks

- **Trace:** $\text{Tr}(A) = \sum a_{ii} = \sum \lambda_i$ (Sum of diagonal = sum of eigenvalues).
- **Determinant:** $\det(A) = \prod \lambda_i$ (Product of eigenvalues).
- **Rank:** Rank = Dimension of $C(A)$ = Dimension of $R(A)$ = Number of non-zero singular values.
- **Symmetry:** If A is symmetric, eigenvalues are real, eigenvectors are orthogonal.
- **Orthogonal matrix** Q: $Q^\top Q = I$. Determinant is ± 1. Preserves lengths $(\|Q\mathbf{x}\| = \|\mathbf{x}\|)$.
- A is invertible iff no eigenvalue is zero.
- Eigenvalues of A^k: λ_i^k for each eigenvalue λ_i.
- Eigenvalues of A^{-1}: $\frac{1}{\lambda_i}$ for each eigenvalue $\lambda_i \neq 0$.
- **Skew-symmetric** $(A = -A^\top)$: all eigenvalues are purely imaginary.

10.6 Proof Toolkit (Standard Strategies)

How to prove U is a Subspace (D 4.8): To prove $U \subseteq V$ is a subspace:
1. **Check 1 (Zero):** Show $0 \in U$. (Usually easy, if fails $\to$ not a subspace).
2. **Check 2 (Closure):** Let $\mathbf{u}, \mathbf{v} \in U$ and $\lambda \in \mathbb{R}$. Show $\lambda\mathbf{u} + \mathbf{v} \in U$.

Counter-example: To disprove, find specific vectors where closure fails or show $0 \notin U$.

How to prove Linear Independence (D 1.21/4.17): To prove $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ are L.I.:
1. Set up equation: $\sum_{i=1}^k \lambda_i\mathbf{v}_i = 0$.
2. Show that this implies $\lambda_1 = \dots = \lambda_k = 0$.
3. *Matrix way:* Form $A = [\mathbf{v}_1 \dots \mathbf{v}_k]$. Show $N(A) = \{0\}$ (e.g. rank k).

How to prove Surjectivity / Injectivity: Let $T : V \to W$ be linear (matrix A).
- **Injective (1-to-1):** Show $\text{Ker}(T) = \{0\}$. (Solve $A\mathbf{x} = 0 \implies \mathbf{x} = 0$).
- **Surjective (Onto):** Show $\text{Im}(T) = W$. (Rank = dim(W)).
- **Bijective:** Show both (or if $\dim(V) = \dim(W)$, just one is enough).

Proving Matrix Properties:
- **Symmetric:** Show $A^\top = A$. (Use $(AB)^\top = B^\top A^\top$).
- **Orthogonal:** Show $Q^\top Q = I$. (Cols are orthonormal).
- **Positive Definite:** Show $\mathbf{x}^\top A\mathbf{x} > 0$ for all $\mathbf{x} \neq 0$.

10.7 Advanced Calculation Strategies

Fitting a Polynomial (Least Squares): Task: Fit $p(t) = \alpha_0 + \alpha_1 t + \dots + \alpha_k t^k$ to points $(t_1, y_1), \dots, (t_m, y_m)$.
1. Setup $A\mathbf{x} = \mathbf{b}$ where unknowns $\mathbf{x} = (\alpha_0, \dots, \alpha_k)^\top$.
2. Matrix A (Vandermonde structure):

$$A = \begin{bmatrix} 1 & t_1 & t_1^2 & \dots \\ \vdots & \vdots & \vdots & \\ 1 & t_m & t_m^2 & \dots \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

3. Solve Normal Equations: $A^\top A\hat{\mathbf{x}} = A^\top\mathbf{b}$.

Change of Basis: Let $B_{old} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ and $B_{new} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$. Let T be a transformation with matrix A in standard basis.
- Matrix of T relative to B_{new} is: $D = V^{-1}AV$
- Where $V = [\mathbf{v}_1 \dots \mathbf{v}_n]$ (Cols are basis vectors).
- If B_{new} are eigenvectors, $D = \Lambda$ (Diagonal).

Computing SVD Step-by-Step: Target: $A = U\Sigma V^\top$. (Rank r).
1. **1. Right Singular Vectors** (V)**:** Compute $M = A^\top A$. Find eigenvalues λ_i and orthonormal eigenvectors $\mathbf{v}_i$ of M. Sort $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$. $V = [\mathbf{v}_1 \dots \mathbf{v}_n]$.
2. **2. Singular Values** (Σ)**:** $\sigma_i = \sqrt{\lambda_i}$. Matrix Σ has σ_i on diagonal.
3. **3. Left Singular Vectors** (U)**:** For $i = 1 \dots r$ $(\sigma_i \neq 0)$: $\mathbf{u}_i = \frac{1}{\sigma_i}A\mathbf{v}_i$. For $i > r$: Extend to orthonormal basis of $\mathbb{R}^m$ (Gram-Schmidt on Nullspace of $A^\top$).

10.8 Spectral Theory & Properties

Algebraic vs Geometric Multiplicity: For eigenvalue λ:
- **Alg. Mult.** (n_a)**:** Number of times λ is root of $\det(A - \lambda I)$.
- **Geo. Mult.** (n_g)**:** $\dim(N(A - \lambda I))$ (Num. of independent eigenvectors).
- **Property:** $1 \leq n_g \leq n_a$.
- **Diagonalizable:** iff $\sum n_g = n$ (i.e. $n_g = n_a$ for all λ).

Tricks for 2×2 Eigenvalues: $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Char Poly: $\lambda^2 - \text{Tr}(A)\lambda + \det(A) = 0$.
- $\lambda_{1,2} = \frac{\text{Tr}(A) \pm \sqrt{\text{Tr}(A)^2 - 4\det(A)}}{2}$.
- **Real Eigenvalues:** iff Discriminant $D \geq 0$.
- **One Real Eigenvalue:** iff $D = 0$.
- **Complex Eigenvalues:** iff $D < 0$ (conjugate pair $a \pm bi$).

Positive Definite Matrices (Symmetric A): Check one of these (all equivalent):
1. All eigenvalues $\lambda_i > 0$.
2. $\mathbf{x}^\top A\mathbf{x} > 0$ for all $\mathbf{x} \neq 0$.
3. All pivots (from Gaussian Elim without swap) are > 0.
4. **Sylvester's Criterion:** All upper-left sub-determinants > 0.

Note: For Positive Semidefinite, replace $>$ with $\geq$.

10.9 Matrix Algebra Hacks

Standard Basis Matrices: $E_{ij} = \mathbf{e}_i\mathbf{e}_j^\top$ (Matrix with 1 at i, j, else 0).
- Product: $E_{ij}E_{kl} = (\mathbf{e}_i\mathbf{e}_j^\top)(\mathbf{e}_k\mathbf{e}_l^\top) = \mathbf{e}_i(\mathbf{e}_j^\top\mathbf{e}_k)\mathbf{e}_l^\top$.
- $\mathbf{e}_j^\top\mathbf{e}_k = \delta_{jk}$ (1 if $j = k$, else 0).
- So $E_{ij}E_{kl} = \delta_{jk}E_{il}$. (Zero unless "inner indices" match).

Block Matrices: If $M = \begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$ (Block triangular):
- $\det(M) = \det(A) \cdot \det(D)$.
- Eigenvalues of M are eigenvalues of $A \cup$ eigenvalues of D.

Rank Properties:
- $\text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))$.
- $\text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B)$.
- **Sylvester:** $\text{rank}(A) + \text{rank}(B) - n \leq \text{rank}(AB)$ (where $A : m \times n, B : n \times k$).

11 Typical Exercises

11.1 Projections onto Column Spaces

Exercise: Find projection $\mathbf{p} \in C(Q)$ that minimizes $\|\mathbf{b} - \mathbf{p}\|$: Given: Matrix Q and vector $\mathbf{b}$.
Approach:
1. Recognize this is asking for $\mathbf{p} = \text{proj}_{C(Q)}(\mathbf{b})$ (orthogonal projection onto column space)
2. Use the formula: $\mathbf{p} = Q(Q^\top Q)^{-1}Q^\top\mathbf{b}$
3. If Q has orthonormal columns: $\mathbf{p} = QQ^\top\mathbf{b}$ (much simpler!)
4. Verify: The residual $\mathbf{b} - \mathbf{p}$ should be orthogonal to all columns of Q
5. Check: $\|\mathbf{b} - \mathbf{p}\| = \sqrt{\|\mathbf{b}\|^2 - \|\mathbf{p}\|^2}$

11.2 Proofs with Skew-Symmetric Matrices

Exercise: Prove $\mathbf{x}^\top S\mathbf{x} = 0$ for all $\mathbf{x}$ where $S^\top = -S$: **Key Insight:** Scalar products are always symmetric.
Approach:
1. Start with: $\mathbf{x}^\top S\mathbf{x}$ (this is a scalar, so equals its transpose)
2. Write: $\mathbf{x}^\top S\mathbf{x} = (\mathbf{x}^\top S\mathbf{x})^\top = \mathbf{x}^\top S^\top\mathbf{x}$
3. Substitute the given condition $S^\top = -S$: $\mathbf{x}^\top S\mathbf{x} = \mathbf{x}^\top(-S)\mathbf{x} = -\mathbf{x}^\top S\mathbf{x}$
4. Conclude: Only a scalar equal to its negative is 0, so $\mathbf{x}^\top S\mathbf{x} = 0$

11.3 Least Squares Fitting

Exercise: Minimize $\sum_{k=1}^m (f(x_k) - y_k)^2$ for $f(x) = ax^2 + b$: **Given:** Data points (x_k, y_k) and a model $f(x) = ax^2 + b$.
Approach:
1. Form the design matrix: $A = \begin{bmatrix} x_1^2 & 1 \\ x_2^2 & 1 \\ \vdots & \vdots \\ x_n^2 & 1 \end{bmatrix}$, with $\mathbf{y} = [y_1, \dots, y_n]^\top$
2. Solve the normal equations: $A^\top A\hat{\mathbf{x}} = A^\top\mathbf{y}$
3. Solve for $\hat{\mathbf{x}} = [a, b]^\top$ (can use Gaussian elimination or inversion)
4. Verify with given squared error: Compute $\sum_k (f(x_k) - y_k)^2$

11.4 Matrix Equations from Eigenvector Conditions

Constructing A from orthonormal input–output pairs: Given

$$A\mathbf{v}_1 = \mathbf{w}_1, \quad A\mathbf{v}_2 = \mathbf{w}_2$$

with $\{\mathbf{v}_1, \mathbf{v}_2\}$ orthonormal.
- Form $Q = [\mathbf{v}_1 \mid \mathbf{v}_2]$ (orthogonal $\Rightarrow Q^{-1} = Q^\top$).
- Form $W = [\mathbf{w}_1 \mid \mathbf{w}_2]$.
- Then

$$A = WQ^\top.$$

- A is unique.

11.5 Cauchy-Schwarz and Inequality Proofs

Exercise: Prove $\sum_{i=1}^{n} \frac{a_i^2}{b_i} \geq \frac{(a_1+\cdots+a_n)^2}{b_1+\cdots+b_n}$ **for** $b_i > 0$: **Key Insight:** Recognize this as a weighted Cauchy-Schwarz problem.
Approach:
1. Define vectors: $\mathbf{u} = [\frac{a_1}{\sqrt{b_1}}, \dots, \frac{a_n}{\sqrt{b_n}}]$ and $\mathbf{v} = [\sqrt{b_1}, \dots, \sqrt{b_n}]$
2. Apply Cauchy-Schwarz: $|\mathbf{u} \cdot \mathbf{v}|^2 \leq \|\mathbf{u}\|^2 \|\mathbf{v}\|^2$
3. Compute LHS: $(\mathbf{u} \cdot \mathbf{v})^2 = (\sum_i a_i)^2$
4. Compute RHS: $\|\mathbf{u}\|^2 = \sum_i \frac{a_i^2}{b_i}$ and $\|\mathbf{v}\|^2 = \sum_i b_i$
5. Rearrange to get the desired inequality

11.6 Singular Values and Decompositions

Exercise: Find a non-zero singular value of a given matrix: **Given:** A matrix A, possibly with special structure.
Approach (Method 1 - Slow):
1. Compute $A^\top A$
2. Find eigenvalues of $A^\top A$ (these are σ_i^2)
3. Take square roots to get singular values σ_i

Approach (Method 2 - Faster):
1. If A has a special structure (e.g., orthogonal rows/columns), use that
2. For a rank-1 matrix: $\sigma = \|A\mathbf{v}\|$ for any non-zero $\mathbf{v}$ in the column space
3. Use $\sigma_{\max} = \|A\|$ (spectral norm) $= \sqrt{\lambda_{\max}(A^\top A)}$

12 Requirements Checklist

When can I use this method?:
- **Gaussian Elimination:** Any matrix.
- **Matrix Inversion (A^{-1}):** A must be square ($n \times n$) AND $\det(A) \neq 0$ (Full rank).
- **CR Decomposition:** Any matrix A.
- **QR Decomposition:** A must have linearly independent columns (full column rank) for the standard Gram-Schmidt process.

Diagonalization ($A = V\Lambda V^{-1}$): Requires: $A \in \mathbb{R}^{n\times n}$ must have n linearly independent eigenvectors.
- *Sufficient (but not necessary):* A has n distinct eigenvalues.
- *Necessary and Sufficient:* For every eigenvalue λ, geometric multiplicity = algebraic multiplicity.

Orthogonal Diagonalization ($A = Q\Lambda Q^\top$): Requires: A must be **Symmetric** ($A = A^\top$). *Note:* If A is symmetric, it is *always* diagonalizable with real eigenvalues and orthogonal eigenvectors.

Cholesky Decomposition ($A = LL^\top$): Requires: A must be **Symmetric** AND **Positive Definite** (all $\lambda > 0$ / all pivots > 0).

SVD ($A = U\Sigma V^\top$): Requires: No restrictions! Every matrix $A \in \mathbb{R}^{m\times n}$ has an SVD.

Least Squares ($A^\top A\hat{x} = A^\top b$): Unique Solution Requires: Columns of A must be linearly independent (Full column rank, $N(A) = \{0\}$). If not, infinitely many least-squares solutions exist (use Pseudoinverse).

13 Quick Facts & Properties

Symmetric Matrices ($A = A^\top$):
- Eigenvalues are always **real**.
- Eigenvectors from different eigenspaces are **orthogonal**.
- Always orthogonally diagonalizable: $A = Q\Lambda Q^\top$.
- $\operatorname{rank}(A)$ = number of non-zero eigenvalues (counted with multiplicity).
- $\operatorname{Tr}(A) = \sum_i \lambda_i$, $\quad \det(A) = \prod_i \lambda_i$.
- A is positive definite $\iff$ all eigenvalues > 0.

Orthogonal Matrices ($Q^\top Q = I$):
- Columns form an orthonormal basis for $C(Q)$.
- Preserves norms: $\|Q\mathbf{x}\| = \|\mathbf{x}\|$.
- Preserves dot products: $(Q\mathbf{x}) \cdot (Q\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$.
- $\operatorname{rank}(Q) = n$ (Full column rank).
- $Q^\top$ is the left-inverse ($Q^\top Q = I$).

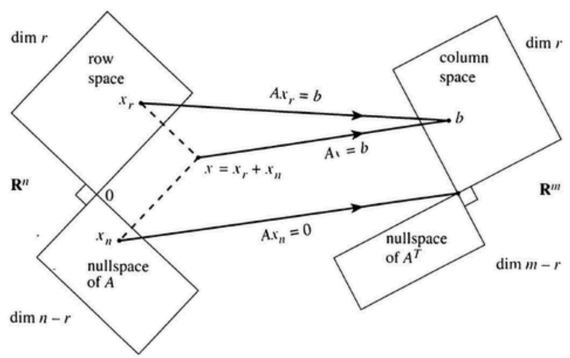
Only if Square (Q is $n \times n$):
- Q is invertible and $Q^{-1} = Q^\top$.
- $QQ^\top = I$ (Rows are also orthonormal).
- $\det(Q) = \pm 1$.
- Eigenvalues satisfy $|\lambda| = 1$.

Skew-Symmetric Matrices ($A^\top = -A$):
- Diagonal entries are all 0.
- If n is odd, then $\det(A) = 0$.
- If n is even, $\det(A) \geq 0$.
- Eigenvalues are purely imaginary or 0.
- $\mathbf{x}^\top A\mathbf{x} = 0$ for all $\mathbf{x} \in \mathbb{R}^n$.
- $\det(A) = \det(-A) = (-1)^n \det(A)$ (useful parity trick).

Projection Matrices ($P^2 = P$):
- Eigenvalues are only 0 or 1.
- Projects onto $C(P)$ along $N(P)$.
- $\operatorname{Tr}(P) = \operatorname{rank}(P)$.
- $I - P$ is also a projection (onto $N(P)$).
- $C(P) \cap N(P) = \{0\}$.

Only if Orthogonal Projection ($P = P^\top$):
- $N(P) = C(P)^\perp$.
- $I - P$ projects onto the orthogonal complement $C(P)^\perp$.
- $\|P\mathbf{x}\| \leq \|\mathbf{x}\|$ for all $\mathbf{x}$ (Non-expansive).

Positive (Semi-)Definite Matrices:
- **Positive definite (PD):** $\mathbf{x}^\top A\mathbf{x} > 0$ for all non-zero $\mathbf{x}$.
- **Positive semidefinite (PSD):** $\mathbf{x}^\top A\mathbf{x} \geq 0$ for all $\mathbf{x}$.
- All eigenvalues ≥ 0 (PD $\iff$ all > 0).
- All pivots ≥ 0 (PD $\iff$ all > 0).
- Diagonal entries satisfy $A_{ii} \geq 0$.
- $\det(A) > 0$ for PD; $\det(A) \geq 0$ for PSD.
- If A is PD, then A^{-1} is also PD.
- $\operatorname{Tr}(A^2) \leq \operatorname{Tr}(A)^2$ for PSD matrices.



13.1 Rapid Operations

Inverse of 2×2: $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \implies A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

Rank Properties:
- $\operatorname{rank}(A) = \operatorname{rank}(A^\top) = \operatorname{rank}(A^\top A) = \operatorname{rank}(AA^\top)$.
- A invertible $\iff$ $\operatorname{rank}(A) = n$ (square case).
- $\dim(\mathcal{N}(A)) = n - \operatorname{rank}(A)$ (Rank–Nullity).
- Full column rank $\implies$ injective.
- Full row rank $\implies$ surjective.
- If $Au = Av$ with $u \neq v$, then A has a non-trivial nullspace.

Determinant Shifts:
- $\det(A^{-1}) = 1/\det(A)$.
- $\det(AB) = \det(A)\det(B)$.
- $\det(kA) = k^n \det(A)$ for $n \times n$ matrices.
- $\det(A^\top) = \det(A)$.
- $\det(A - \lambda I) = 0 \iff \lambda$ is an eigenvalue.

Trace Tricks:
- $\operatorname{Tr}(A + B) = \operatorname{Tr}(A) + \operatorname{Tr}(B)$.
- $\operatorname{Tr}(kA) = k\operatorname{Tr}(A)$.
- $\operatorname{Tr}(ABC) = \operatorname{Tr}(BCA) = \operatorname{Tr}(CAB)$ (cyclic property).
- $\operatorname{Tr}(A^\top A) = \sum_{i,j} a_{ij}^2 \geq 0$.
- $\operatorname{Tr}(A) = \sum_i \lambda_i$ (eigenvalue sum).

Block Matrices: For $M = \begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$ (block triangular):
- $\det(M) = \det(A)\det(D)$.
- Eigenvalues of M = eigenvalues of A $\cup$ eigenvalues of D.
- If A, D invertible: $M^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix}$.

For block diagonal $M = \operatorname{diag}(A, D)$: $M^k = \operatorname{diag}(A^k, D^k)$.

Cross Product (in $\mathbb{R}^3$): $\mathbf{u} \times \mathbf{v} = \det\begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix}$.
- Orthogonal to both $\mathbf{u}$ and $\mathbf{v}$.
- $\|\mathbf{u} \times \mathbf{v}\|$ = area of parallelogram spanned by $\mathbf{u}, \mathbf{v}$.
- $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \det([\mathbf{u}, \mathbf{v}, \mathbf{w}])$ (volume).

Quick Eigenvalue Checks: For 2×2 matrix A:
- $\lambda_1 + \lambda_2 = \operatorname{Tr}(A)$.
- $\lambda_1 \lambda_2 = \det(A)$.

If all row sums equal s: s is an eigenvalue with eigenvector $\mathbf{1}$.

$$A\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} s \\ s \\ \vdots \\ s \end{bmatrix} = s\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

If all column sums equal s: s is an eigenvalue of $A^\top$ (hence also of A).

Nilpotent & Idempotent:
- **Nilpotent:** $A^k = 0$ for some k. All eigenvalues are 0; $\det(A) = 0$; $\operatorname{Tr}(A) = 0$; $\operatorname{rank}(A) < n$.
- Nilpotent matrices are not closed under addition or multiplication.
- **Idempotent:** $A^2 = A$. Eigenvalues are only 0 or 1.
- For idempotent A: $\operatorname{Tr}(A) = \operatorname{rank}(A)$.

Linear Systems & Solutions:
- $m < n \implies$ no linear system $Ax = b$ can have a unique solution.
- If $\operatorname{rank}(A) < n$: Existence of one solution $\implies$ infinitely many solutions.
- Homogeneous system $Ax = 0$ has infinitely many solutions $\iff \operatorname{rank}(A) < n$.