1 Vectors

1.1 Linear Combination (Def. 1.4)

Let $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n \in \mathbb{R}^m$ and $\lambda_1, \lambda_2, ..., \lambda_n \in \mathbb{R}$, then

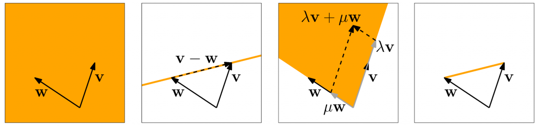
$$\lambda_1 \boldsymbol{v}_1 + \lambda_2 \boldsymbol{v}_2 + ... + \lambda_n \boldsymbol{v}_n$$

is a linear combination of $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$.

1.1.1 Special Linear Combinations (Def. 1.7)

A linear combination $\lambda_1 \boldsymbol{v}_1 + \lambda_2 \boldsymbol{v}_2 + ... + \lambda_n \boldsymbol{v}_n$ is

(ii) *affine* if $\lambda_1 + \lambda_2 + ... + \lambda_n = 1$,
(iii) *conic* if $\lambda_i \geq 0$ for $i = 1, 2, ..., n$,
(iv) *convex* if it is both affine and conic.



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

1.2 Euclidean Norm (Def. 1.11)

$$\|\boldsymbol{v}\| := \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}} = \sqrt{v_1^2 + v_2^2 + ... + v_n^2}.$$

To *normalize* a vector, i.e. make it of *length* one, we can divide by the Euclidean norm: $\boldsymbol{v}' = \boldsymbol{v}/\|\boldsymbol{v}\|$.

1.3 Dot Product and Angles (Def. 1.14)

The angle α between two non-zero vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ is

$$\cos(\alpha) = \frac{\boldsymbol{v} \cdot \boldsymbol{w}}{\|\boldsymbol{v}\|\|\boldsymbol{w}\|},$$

where $-1 \leq \alpha \leq 1$.

1.4 Prove Formula Defines a Scalar Product

Let $\boldsymbol{A} = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}$. Prove that the following formula defines a scalar product in $V = \mathbb{R}^2$:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\boldsymbol{A}} = \langle \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \rangle_{\boldsymbol{A}} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \boldsymbol{A} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

To be a scalar product, three conditions must be satisfied:

(i) **Symmetry:** Since $\boldsymbol{A}^T = \boldsymbol{A}$, the matrix is symmetric. Thus,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\boldsymbol{A}} = \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{y} = (\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{y})^T = \boldsymbol{y}^T \boldsymbol{A}^T \boldsymbol{x} = \boldsymbol{y}^T \boldsymbol{A} \boldsymbol{x}.$$

(ii) **Bilinearity:** Linearity in the first argument holds due to distributivity of matrix multiplication:

$$\begin{aligned}(\alpha \boldsymbol{x} + \beta \boldsymbol{z})^T \boldsymbol{A} \boldsymbol{y} &= (\alpha \boldsymbol{x}^T + \beta \boldsymbol{z}^T) \boldsymbol{A} \boldsymbol{y} \\ &= \alpha \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{y} + \beta \boldsymbol{z}^T \boldsymbol{A} \boldsymbol{y}.\end{aligned}$$

(iii) **Positive Definiteness:** We determine if $\boldsymbol{A}$ is Positive Definite by checking its eigenvalues via $\det(\boldsymbol{A} - \lambda I) = 0$:

$$\det \begin{bmatrix} 3 - \lambda & -1 \\ -1 & 3 - \lambda \end{bmatrix} = (3 - \lambda)^2 - 1 = \lambda^2 - 6\lambda + 8 = (\lambda - 4)(\lambda - 2).$$

The eigenvalues are $\lambda_1 = 4$ and $\lambda_2 = 2$. Since all eigenvalues are strictly positive ($\lambda_i > 0$), $\boldsymbol{A}$ is Positive Definite. By Prop 9.2.12, this implies: $\langle \boldsymbol{x}, \boldsymbol{x} \rangle_{\boldsymbol{A}} = \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} > 0 \quad \forall \boldsymbol{x} \neq 0$.

1.4.1 Cauchy-Schwarz (Lemma 1.12)

For any two vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^m$

$$|\boldsymbol{v}^T \boldsymbol{w}| \leq \|\boldsymbol{v}\| \cdot \|\boldsymbol{w}\|.$$

Equality holds when one is a scalar multiple of the other.

1.4.2 Triangle Inequality (Lemma 1.17)

Let $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^m$. Then

$$\|\boldsymbol{v} + \boldsymbol{w}\| \leq \|\boldsymbol{v}\| + \|\boldsymbol{w}\|.$$

1.4.3 Orthogonal Vectors (Def. 1.15)

Let $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^m$. $\boldsymbol{v}$ and $\boldsymbol{w}$ are orthogonal if and only if

$$\boldsymbol{v} \cdot \boldsymbol{w} = 0,$$

i.e., the angle between them is $\alpha = 90°$.
Note: If $\boldsymbol{v}$ and $\boldsymbol{w}$ are orthogonal, then $\|\boldsymbol{v} + \boldsymbol{w}\|^2 = (\boldsymbol{v} + \boldsymbol{w})^T (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{v}^T \boldsymbol{v} + \boldsymbol{v}^T \boldsymbol{w} + \boldsymbol{w}^T \boldsymbol{v} + \boldsymbol{w}^T \boldsymbol{w} = \|\boldsymbol{v}\|^2 + \|\boldsymbol{w}\|^2$

1.5 Linear Independence (Cor. 1.23)

Let $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n \in \mathbb{R}^m$. They are linearly independent if

(i) none of the vectors is a linear combination of the other ones,
(ii) there are no scalars $\lambda_1, \lambda_2, ..., \lambda_n$ besides $0, 0, ..., 0$ such that $\sum_{j=1}^n \lambda_j \boldsymbol{v}_j = \boldsymbol{0}$,
(iii) none of the vectors is a linear combination of the previous ones.

Following, the columns of a matrix A are linearly independent if there is no $\boldsymbol{x}$ besides $\boldsymbol{0}$ such that $A\boldsymbol{x} = 0$.

1.6 Linear Dependence Example

Let $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{z} \in \mathbb{R}^2$. Prove there exist non-zero scalars c_u, c_v, c_w, c_z such that $c_u + c_v + c_w + c_z = 0$ and $c_u \boldsymbol{u} + c_v \boldsymbol{v} + c_w \boldsymbol{w} + c_z \boldsymbol{z} = \boldsymbol{0}$.

Define augmented vectors in $\mathbb{R}^3$ by appending a 1, e.g., $\boldsymbol{u}' = \begin{bmatrix} u_1 \\ u_2 \\ 1 \end{bmatrix}$. Since there are 4 vectors in $\mathbb{R}^3$, they must be linearly dependent. The relation $c_u \boldsymbol{u}' + c_v \boldsymbol{v}' + c_w \boldsymbol{w}' + c_z \boldsymbol{z}' = \boldsymbol{0}$ implies the vector sum is zero (first two coordinates) and the scalar sum is zero (last coordinate).

1.7 The Span (Def. 1.25)

Let $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n \in \mathbb{R}^m$, their Span is defined as the set of all linear combinations, i.e.,

$$\mathrm{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n) := \left\{ \sum_{j=1}^n \lambda_j \boldsymbol{v}_j : \lambda_j \in \mathbb{R} \forall j \in [n] \right\}.$$

1.8 Span Proof Example

Let $\boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \in \mathbb{R}^2$ be arbitrary and assume $\boldsymbol{v} \neq 0$ and $\boldsymbol{w} \neq \lambda \boldsymbol{v}, \forall \lambda \in \mathbb{R}$. Show that $\mathrm{Span}(\boldsymbol{v}, \boldsymbol{w}) = \mathbb{R}^2$.

To show this, prove that there are λ and μ for each $\boldsymbol{u} \in \mathbb{R}^2$ such that $\boldsymbol{u} = \lambda \boldsymbol{v} + \mu \boldsymbol{w}$, i.e.

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \lambda \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + \mu \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \Rightarrow \begin{matrix} (1) u_1 = \lambda v_1 + \mu w_1 \\ (2) u_2 = \lambda v_2 + \mu w_2 \end{matrix}.$$

Assume $v_1 \neq 0$, thus from (1) follows

$$\lambda = \frac{u_1 - \mu w_1}{v_1}.$$

Using this in (2) (assuming $w_2 - \frac{v_2 w_1}{v_1} \neq 0$), we get

$$u_2 = \left(\frac{u_1 - \mu w_1}{v_1} \right) v_2 + \mu w_2 \Rightarrow \mu = \frac{u_2 - \frac{u_1 v_2}{v_1}}{w_2 - \frac{v_2 w_1}{v_1}}.$$

Thus, choosing λ and μ is possible for all $\boldsymbol{u} \in \mathbb{R}^2$.

2 Matrices (Def. 2.1)

An $m \times n$ matrix is a rectangular array of numbers with m columns and n rows, i.e.,

$$A \in \mathbb{R}^{m \times n} := \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

2.1 The Rank (Def. 2.10)

The *rank* of a matrix $A \in \mathbb{R}^{m \times n}$ is its number of *linearly independent* columns.

2.2 The Transpose (Def. 2.12)

Let $A \in \mathbb{R}^{m \times n}$, the transpose of A, A^T, is equal to A with interchanged rows and columns.

(i) **Reverse Order:** $(AB)^T = B^T A^T$ (Lemma 2.40).
(ii) **Scalars:** $(rA)^T = rA^T$ for any $r \in \mathbb{R}$.
- This implies $(-A)^T = -A^T$.

(iii) **Powers:** $(A^n)^T = (A^T)^n$ for any $n \in \mathbb{N}$.
- Combined with scalars: $((-A)^n)^T = ((-A)^T)^n$.

(iv) **Inverse:** $(A^{-1})^T = (A^T)^{-1}$ (if A is invertible).

2.3 Square Matrix Classes (Def. 2.3)

Let $A \in \mathbb{R}^{m \times m}$, A is

(i) *upper triangular* if all entries below the diag. are 0,
(ii) *lower triangular* if all entries above the diag. are 0,
(iii) *diagonal* if it is both *upper* and *lower* triangular, i.e., the matrix only has non-zero values on its diagonal,
(iv) *symmetric* if the values above and below the diagonal are equal. Let $A \in \mathbb{R}^{m \times m}$, A is symmetric $\iff A = A^T$.

2.4 Matrix-Vector Multiplication (Def 2.4)

The product $A\boldsymbol{x}$ can be understood in three different ways:

2.4.1 Matrix Picture

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m,2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11} \cdot x_1 + ... + a_{1n} \cdot x_n \\ a_{21} \cdot x_1 + ... + a_{2n} \cdot x_n \\ \vdots \\ a_{m1} \cdot x_1 + ... + a_{nm} \cdot x_n \end{bmatrix}.$$

2.4.2 Row Picture

$$\begin{bmatrix} - \boldsymbol{u}_1 - \\ - \boldsymbol{u}_2 - \\ \vdots \\ - \boldsymbol{u}_m - \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_1 \cdot \boldsymbol{x} \\ \boldsymbol{u}_2 \cdot \boldsymbol{x} \\ \vdots \\ \boldsymbol{u}_m \cdot \boldsymbol{x} \end{bmatrix}.$$

2.4.3 Column Picture

$$\begin{bmatrix} | & & | \\ \boldsymbol{w}_1 & \dots & \boldsymbol{w}_n \\ | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \cdot \boldsymbol{w}_1 + x_2 \cdot \boldsymbol{w}_2 + ... + x_n \cdot \boldsymbol{w}_n.$$

2.5 Matrix-Matrix Multiplication (Def. 2.36)

For any two matrices $A \in \mathbb{R}^{m \times k}$ and $B \in \mathbb{R}^{k \times n}$, the product $AB \in \mathbb{R}^{m \times n}$ can also be understood in three ways:

2.5.1 The Matrix Picture

$$\begin{bmatrix} - \boldsymbol{u}_1 - \\ - \boldsymbol{u}_2 - \\ \vdots \\ - \boldsymbol{u}_m - \end{bmatrix} \begin{bmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_1 \cdot \boldsymbol{v}_1 & \boldsymbol{u}_1 \cdot \boldsymbol{v}_2 & \dots & \boldsymbol{u}_1 \cdot \boldsymbol{v}_n \\ \boldsymbol{u}_2 \cdot \boldsymbol{v}_1 & \boldsymbol{u}_2 \cdot \boldsymbol{v}_2 & \dots & \boldsymbol{u}_2 \cdot \boldsymbol{v}_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{u}_m \cdot \boldsymbol{v}_1 & \boldsymbol{u}_m \cdot \boldsymbol{v}_2 & \dots & \boldsymbol{u}_m \cdot \boldsymbol{v}_n \end{bmatrix}.$$

2.5.2 The Row Picture

$$\begin{bmatrix} - \boldsymbol{u}_1 - \\ - \boldsymbol{u}_2 - \\ \vdots \\ - \boldsymbol{u}_m - \end{bmatrix} B = \begin{bmatrix} - \boldsymbol{u}_1 B - \\ - \boldsymbol{u}_2 B - \\ \vdots \\ - \boldsymbol{u}_m B - \end{bmatrix}.$$

2.5.3 The Column Picture

$$A \begin{bmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ A\boldsymbol{v}_1 & A\boldsymbol{v}_2 & \dots & A\boldsymbol{v}_n \\ | & | & & | \end{bmatrix}.$$

2.5.4 Distributivity and Associativity (Lemma 2.42)

Let A, B, C and D be four matrices, then

(i) $A(B + C) = AB + AC$ and $(B + C)D = BD + CD$ (distributivity)
(ii) $(AB)C = A(BC)$ (associativity)

Keep in mind, that **MATRIX MULTIPLICATION IS NOT COMMUTATIVE**.

2.6 CR' Decomposition $A = CR'$ (Theo. 2.46)

Let $A \in \mathbb{R}^{m \times n}$ with rank r. Let $C \in \mathbb{R}^{m \times r}$ be the submatrix of A containing all its independent columns. Then there exists a unique $R' \in \mathbb{R}^{r \times n}$ such that $A = CR'$.

In other words, C describes the linearly independent columns, while R' (R from Gauss-Jordan Elimination, but without the zero rows) shows how to combine them to create A. For instance,

$$\begin{bmatrix} 1 & 2 & 0 & 3 \\ 2 & 4 & 1 & 4 \\ 3 & 6 & 2 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & -2 \end{bmatrix}.$$

2.7 Linear Transformation (Def. 2.21)

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a function. T is called a linear transformation if $\forall \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ and $\forall \lambda, \mu \in \mathbb{R}$,

$$T(\lambda \boldsymbol{v} + \mu \boldsymbol{w}) = \lambda T(\boldsymbol{v}) + \mu T(\boldsymbol{w}).$$

Let $T : \mathbb{R}^n \to \mathbb{R}^m$, is T a linear transformation?
(i) Show $T(\boldsymbol{v} + \boldsymbol{w}) = T(\boldsymbol{v}) + T(\boldsymbol{w})$ for all $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$.
(ii) Show $T(\lambda \boldsymbol{v}) = \lambda T(\boldsymbol{v})$ for all $\boldsymbol{v} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.

2.7.1 $T = T_A$ (Theo. 2.26)

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 $T = T_A$.

2.8 Rotation Matrix

In $\mathbb{R}^{2 \times 2}$, to rotate the xy plane counterclockwise around an angle θ we use

$$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

In $\mathbb{R}^{3 \times 3}$, there are three different rotation matrices, one for each axis:

$$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$$

$$R_y(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$$

$$R_z(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

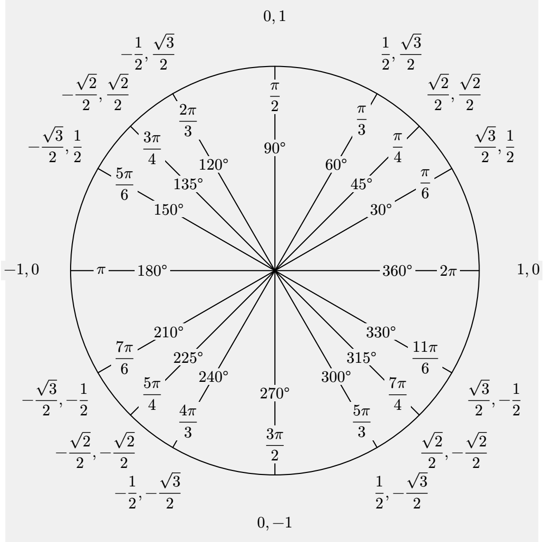
2.8.1 Composition of Rotations

Let $R(\theta_1), R(\theta_2) \in \mathbb{R}^{2\times 2}$ be two rotation matrices, then

$$R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2).$$

Moreover, for every $R(\theta) \in \mathbb{R}^{2\times 2}$, there is some $R(-\theta) \in \mathbb{R}^{2\times 2}$ such that $R(\theta)R(-\theta) = R(0) = I$.

If we need to find an $A \in \mathbb{R}^{2\times 2}$ that satisfies $A^k = I \iff$ k is a multiple of t, then A can be a rotation matrix with $\theta = \frac{2\pi}{t}$. For higher dimensions $A \in \mathbb{R}^{n\times n}$, the same principle applies if we embed the 2D rotation e.g. into the top left corner (like in $R_z(\theta)$ above) and fill the rest of the diagonal with 1s. Below: left $= \cos(\theta)$, right $= \sin(\theta)$.



3 Linear Equations

A system of m linear equations in n variables,

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m, \end{aligned}$$

can be expressed using matrices in the shape $A\boldsymbol{x} = \boldsymbol{b}$:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

3.1 Gauss Elimination

3.1.1 Back Substitution

We can directly solve $A\boldsymbol{x} = \boldsymbol{b}$, when A is an *upper triangular* matrix by solving the equations in reverse order.

$$\left.\begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 0 & 0 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 19 \\ 17 \\ 14 \end{bmatrix}\right\}$$

Solve for x_3 first and then work your way up (*backwards* substitution).

3.1.2 Elimination

Transform $A\boldsymbol{x} = \boldsymbol{b}$ to $U\boldsymbol{x} = \boldsymbol{c}$, where U is upper triangular, by performing

(i) *row exchanges*,
(ii) *scalar multiplication*,
(iii) *row subtraction.*

$$\left[\begin{array}{ccccc|c} u_{11} & \ldots & & & & c_1 \\ 0 & u_{22} & \ldots & & & c_2 \\ 0 & 0 & \ddots & \ldots & & \vdots \\ 0 & 0 & \ldots & u_{rm} & & c_r \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \ldots & 0 & & c_m \end{array}\right]$$

Crucially, $A\boldsymbol{x} = \boldsymbol{b}$ and $U\boldsymbol{x} = \boldsymbol{c}$ have the same solutions (Lemma 3.2).

3.1.3 Elimination and Permutation Matrices

Elmination and row exchanges requires matrices as *linear combinations of the rows* (multiply from the *left*).

row 2 minus $c \cdot$ row 1	exchange row 2 and row 3
$E_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -c & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$	$P_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$

3.2 Inverse Matrices (Def. 2.57)

Let $M \in \mathbb{R}^{m\times m}$, M is invertible, if there exists an $M^{-1} \in \mathbb{R}^{m\times m}$ such that

$$MM^{-1} = M^{-1}M = I.$$

For invertible matrices, the following holds

(i) inverses are unique, i.e. if $AM = MA = I$ and $BM = MB = I$, $A = B$ (Obs. 2.56),
(ii) if A and B are invertible, then $(AB)^{-1} = B^{-1}A^{-1}$ (the product of two invertible matrices is invertible too) (Lemma 2.59),
(iii) if A is invertible, then $(A^T)^{-1} = (A^{-1})^T$ (Lemma 2.60).

3.2.1 Inverse Theo. (Lemma 2.53 / Theo. 3.8)

Let $A \in \mathbb{R}^{m\times m}$, the following is equivalent

(i) A is invertible,
(ii) For every $b \in \mathbb{R}^m$, $A\boldsymbol{x} = \boldsymbol{b}$ has a unique solution $\boldsymbol{x}$,
(iii) The columns of A are linearly independent.

3.2.2 Example

Given $A, B \in \mathbb{R}^{m\times m}$ such that $AB = I$, prove $BA = I$.

(i) Show B has linearly independent columns: Let $\boldsymbol{x} \in \mathbb{R}^m$ such that $B\boldsymbol{x} = 0$, then

$$\boldsymbol{x} = I\boldsymbol{x} = AB\boldsymbol{x} = A\boldsymbol{0} = \boldsymbol{0}.$$

Hence by Obs. 2.5(ii), they are linearly independent.

(ii) Show A also has linearly independent colums: Let $\boldsymbol{y} \in \mathbb{R}^m$ such that $A\boldsymbol{y} = 0$, then by Theorem 3.8 there is some $\boldsymbol{x} \in \mathbb{R}^m$ such that $B\boldsymbol{x} = \boldsymbol{y}$ (because B is invertible), then

$$\boldsymbol{y} = B\boldsymbol{x} = B(I\boldsymbol{x}) = B(AB\boldsymbol{x}) = B(A\boldsymbol{y}) = B\boldsymbol{0} = \boldsymbol{0}.$$

(iii) Show $BA - I = \boldsymbol{0}$:

$$A(BA - I) = ABA - A = IA - A = \boldsymbol{0}.$$

Thus, $BA = I$.

3.2.3 Inverse of a 2×2 matrix

Let $A = \left[\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right]$, then

$$A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} (\text{if } ad - bc \neq 0).$$

3.2.4 Inverse of a $n \times n$ matrix (easy way)

We compute A^{-1} by running Gauss-Jordan elimination on the augmented matrix $[A \mid I]$. If A reduces to I, the right side becomes A^{-1} (Theorem 3.19).

3.2.4.1 Inverse Formula (Prop. 7.3.3)

Let $A \in \mathbb{R}^{n\times n}$ with $\det(A) \neq 0$. Then

$$A^{-1} = \frac{1}{\det(A)} C^T,$$

where C is the matrix of *cofactors*, with entries $C_{ij} = (-1)^{i+j}\det(A_{ij})$ (Def. 7.3.1). Here, A_{ij} is the $(n-1)\times(n-1)$ submatrix obtained by removing row i and column j from A.

3.3 Gauss-Jordan Elimination

3.3.1 Reduced Row Echelon Form (Def. 3.13)

Let $M \in \mathbb{R}^{m\times n}$ with rank r, M is in RREF, if

(i) it contains the unit vectors $e_0, \ldots, e_r$ as its columns in ascending order,
(ii) the columns in between contain values only up to the i-th row, where i refers to the index of the most recent unit vector.

3.3.2 Solving $A\boldsymbol{x} = \boldsymbol{b}$ (Theorem 3.20)

We add the following step to Gauss Elimination:

(i) multiply each row with a pivot p with $1/p \Rightarrow p = 1$
(ii) eliminate any value above the pivot
(iii) (at the end remove any zero-rows, $R_0 \to R$)

A solution to $A\boldsymbol{x} = \boldsymbol{b}$ exists if and only if $b_i = 0$ for all zero-rows $i > r$; then, a specific solution $\boldsymbol{x}$ is found by setting all free variables to 0 and each pivot variable $x_{j_i} = b_i$.

4 The Four Fundamental Subspaces

4.1 Vector Spaces (Def. 4.1)

A vector space is a triple $(V, +, \cdot)$, where

(i) V is a set,
(ii) $+ : V \times V \to V$ a function (vector addition),
(iii) $\cdot : \mathbb{R} \times V \to V$ a function (scalar multiplication).

Such that they satisfy

(i)	$\boldsymbol{v} + \boldsymbol{w} = \boldsymbol{w} + \boldsymbol{v}$	(commutativity)
(ii)	$\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$	(associativity)
(iii)	There is $\boldsymbol{0}$ such that $\boldsymbol{v} + \boldsymbol{0} = \boldsymbol{v}$ for all $\boldsymbol{v}$	(zero vector)
(iv)	There is $-\boldsymbol{v}$ such that $\boldsymbol{v} + (-\boldsymbol{v}) = 0$ for all $\boldsymbol{v}$	(inverse)
(v)	$1 \cdot \boldsymbol{v} = \boldsymbol{v}$ for all $\boldsymbol{v}$	(identity)
(vi)	$(\lambda \cdot \mu)\boldsymbol{v} = \lambda(\mu\boldsymbol{v})$	(compatability in $\mathbb{R}$)
(vii)	$\lambda(\boldsymbol{v} + \boldsymbol{w}) = \lambda\boldsymbol{v} + \lambda\boldsymbol{w}$	(distributivity over $+$)
(viii)	$(\lambda + \mu)\boldsymbol{v} = \lambda\boldsymbol{v} + \mu\boldsymbol{v}$	(distributivity over $+$ in $\mathbb{R}$)

4.2 Subspaces (Def. 4.8) & Proof Strategy

A subset $U \subseteq V$ is a *subspace* of vector space V if it satisfies these three conditions (standard checklist to prove U is a subspace):

(i) **Contains Zero:** $\boldsymbol{0} \in U$ (ensures U is non-empty).
(ii) **Closed under Addition:** $\boldsymbol{u} + \boldsymbol{v} \in U$ for all $\boldsymbol{u}, \boldsymbol{v} \in U$.
(iii) **Closed under Scaling:** $\lambda\boldsymbol{u} \in U$ for all $\boldsymbol{u} \in U$ and $\lambda \in \mathbb{R}$.

4.3 Example

Let V be a vector space and U, W subspaces of V. Show that $U \cup W$ is a subspace of $V \iff U \subseteq W$ or $W \subseteq U$.

"$\Longleftarrow$": Assume $U \subseteq W$, then $U \cup W = W$, which is a subspace of V by assumption. (Analogous for $W \subseteq U$).
"$\Longrightarrow$": Assume $U \cup W$ is a subspace of V and $W \nsubseteq U$, then there is some $\boldsymbol{w} \in W \setminus U$. Let $\boldsymbol{u} \in U$ be arbitrary, then $\boldsymbol{w} + \boldsymbol{u} \in W \cup U$. If $\boldsymbol{w} + \boldsymbol{u} \in U$, then $\boldsymbol{w}$ also $\in U$ (bc of additivity, but this is not possible bc we said $\boldsymbol{w} \in W \setminus U$). Thus $\boldsymbol{w} + \boldsymbol{u} \in W$ must hold, and bc of additivity, $\boldsymbol{u} \in W$. So we started with an arbitrary $\boldsymbol{u} \in U$ and proved it must land in W, as in $\boldsymbol{u} \in W$. Thus, $U \subseteq W$.
Intuition: Union of two subspaces is only a subspace if one is contained in the other (think x-y-axes).

4.4 Bases and Dimensions

4.4.1 Basis (Def. 4.18)

Let V be a vector space. A subset $B \subseteq V$ is called a *basis* of V if B is linearly independent and $\mathrm{Span}(B) = V$.

4.4.2 Dimensions (Def. 4.25)

Let V be a finitely generated vector space. Then $\dim(V) = |B|$, for any basis $B \subseteq V$. Moreover, the size of all bases of a vector space is equal (Theo. 4.24).

4.5 Example

Let $V \subseteq \mathbb{R}^m$, what is the dimension of V?

(i) Find a basis of V,
(ii) show it is a basis (linearly independent & span is V),
(iii) dimension is the cardinality of the base.

4.5.1 Steinitz Exchange Lemma (Lemma 4.23)

Let V be a vector space, $F \subseteq V$ a *finite* set of linearly independent vectors, and $G \subseteq V$ a finite set of vectors with $\mathrm{Span}(G) = V$. Then

(i) $|F| \leq |G|$,
(ii) there is some subset $E \subseteq G$ with $|E| = |G| - |F|$ such that $\mathrm{Span}(F \cup E) = V$.

4.6 The Column Space of A (Def. 2.9)

Let $A \in \mathbb{R}^{m\times n}$. The *column space* of A is the span of its columns.

$$\mathrm{C}(A) := \{A\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^n\} \subseteq \mathbb{R}^m.$$

The set of all vectors you can get by combining the columns of the matrix. For a matrix transformation $A\boldsymbol{x} = \boldsymbol{b}$, the column space is the set of all possible output vectors $\boldsymbol{b}$ for which a solution $\boldsymbol{x}$ exists.

4.6.1 Basis of $\mathrm{C}(A)$ (Theo. 4.31)

The basis of $\mathrm{C}(A)$ is the set of its *linearly independent columns*, and hence $\dim(\mathrm{C}(A)) = \operatorname{rank}(A)$. These can be calculated using Gauss-Jordan (C matrix of $A = CR'$).

4.7 The Row Space of A (Def. 2.14)

Let $A \in \mathbb{R}^{m\times n}$. The *row space* of A is the span of its rows,

$$\mathrm{R}(A) \coloneqq \{A^T\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^m\} = \mathrm{C}(A^T) \subseteq \mathbb{R}^n.$$

It tells about the fundamental, independent equations in a system. If a new equation is a combination of existing rows, it's in the row space. I.e. if a vector satisfies this new equation, it's in the row space.

For example, let $A \Rightarrow R = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \end{bmatrix}$, with row vectors $\boldsymbol{r}_1 = [1, 0, 2]$ and $\boldsymbol{r}_2 = [0, 1, 0]$. A linear combination of these is:

$$c_1[1, 0, 2] + c_2[0, 1, 0] = [c_1, c_2, 2c_1]$$

The row space is the set of vectors $(x, y, z) \in \mathbb{R}^3$ satisfying $z = 2x$ (one less free variable $\Rightarrow$ a plane through the origin).

4.7.1 Basis of $\mathrm{R}(A)$ (Theo. 4.32)

The basis of $\mathrm{R}(A)$ is the set of its *linearly independent rows*, and hence $\dim(\mathrm{R}(A)) = \dim(\mathrm{C}(A)) = \operatorname{rank}(A)$. These correspond to the first r rows of A in REF, where r is the rank of A.
So $R(A) = R(R'^T) = C(R'^T)$ with R from CR decomposition.

4.8 Bases in $A = CR'$ (Theo. 3.18)

Let $A = CR'$, then the columns of C form a basis of $\mathrm{C}(A)$ and the rows of R' form a basis of $\mathrm{R}(A)$.
So $C(A) = C(C)$ and $R(A) = R(R') = C(R'^T)$.

4.9 The Nullspace of A (Def. 2.17)

Let $A \in \mathbb{R}^{m\times n}$. The *nullspace* of A is the set of all solutions to $A\boldsymbol{x} = \boldsymbol{0}$,

$$\mathrm{N}(A) \coloneqq \{\boldsymbol{x} \in \mathbb{R} : A\boldsymbol{x} = 0\} \subseteq \mathbb{R}^n.$$

4.9.1 Basis of $\mathrm{N}(A)$ (Theo. 4.36)

We can calculate a basis of $\mathrm{N}(A)$, by converting A to R' in RREF through Gauss-Jordan and then solving $R'\boldsymbol{x} = \boldsymbol{0}$ with the *special cases* where $\boldsymbol{x}_i = 1$ for every $i = 1, \dots, n$ for every linearly **dependent** column in R'.

For example, $A \in \mathbb{R}^{2\times 4}$,

(i) convert $A \to R'$ through Gauss-Jordan,
(ii) solve $R'\boldsymbol{x} = \boldsymbol{0}$ with special cases (per special case: set one free variable (one variable of the **dependent** columns) e.g. $x_2 = 1$, the other free variables to 0), solve for $\boldsymbol{x}$

$$\underbrace{\begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & -2 \end{bmatrix}}_{A \text{ in RREF } = R'} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \boldsymbol{0} \text{ with } \boldsymbol{x} = \underbrace{\begin{bmatrix} x_1 \\ 1 \\ x_3 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} x_1 \\ 0 \\ x_3 \\ 1 \end{bmatrix}}_{\text{special cases}}.$$

(iii) The solutions $x_1, \dots, x_n$ form a Basis of $\mathrm{N}(R')$,

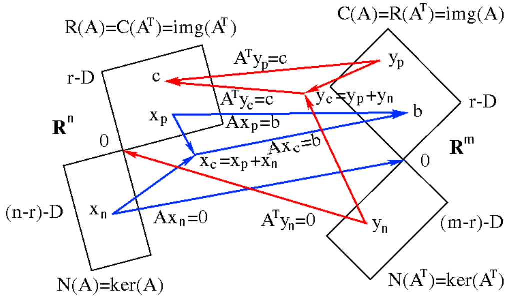
$$\boldsymbol{x}_1 = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{x}_2 = \begin{bmatrix} -3 \\ 0 \\ 2 \\ 1 \end{bmatrix} \Rightarrow \{\boldsymbol{x}_1, \boldsymbol{x}_2\} \text{ is a basis of } \mathrm{N}(R).$$

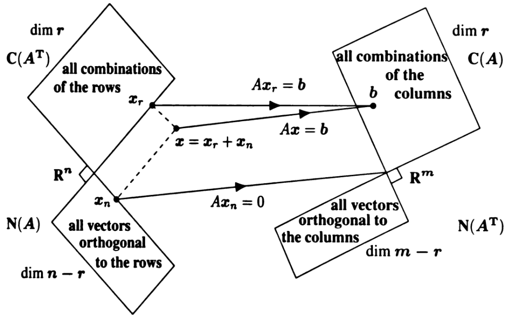
(iv) Any basis of $\mathrm{N}(R')$ is also a basis of $\mathrm{N}(A)$ (Lemma 3.3).

Following, $\dim(\mathrm{N}(A)) = n - \operatorname{rank}(A)$ (Theo. 4.36).

For $A \in \mathbb{R}^{m\times n}$:

subspace	$\mathrm{C}(A)$	$\mathrm{R}(A)$	$\mathrm{N}(A)$	$N(A^T)$
dimensions	r	r	$n - r$	$m - r$
subspace of	$\mathbb{R}^m$	$\mathbb{R}^n$	$\mathbb{R}^n$	$\mathbb{R}^m$





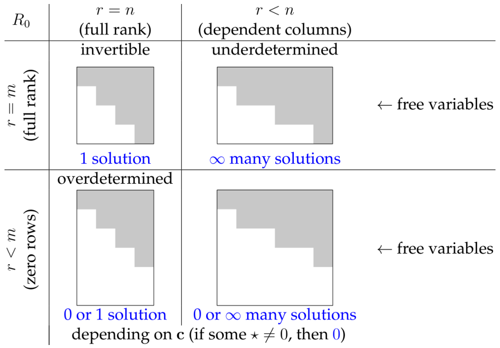
4.10 The Solution Space of $A\boldsymbol{x} = \boldsymbol{b}$ (Def. 4.37)

Let $A \in \mathbb{R}^{m\times n}$, and $\boldsymbol{b} \in \mathbb{R}^m$, then *solution space* of $A\boldsymbol{x} = \boldsymbol{b}$ is the set

$$\operatorname{Sol}(A, \boldsymbol{b}) \coloneqq \{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} = \boldsymbol{b}\} \subseteq \mathbb{R}^n.$$

For any $A\boldsymbol{x} = \boldsymbol{b}$ we have three options, 1. no solutions, 2. one solution and 3. infinite solutions.

(i) If A is not invertible and $\boldsymbol{b} \notin \mathrm{C}(A)$ then no solution exists.
(ii) If A is invertible $\Rightarrow N(A) = \{\boldsymbol{0}\}$ then exactly one solution exists.
(iii) If A is not invertible, but $\boldsymbol{b} \in \mathrm{C}(A)$, then $\exists \boldsymbol{s}$ and we can shift the non-trivial nullspace using $\boldsymbol{s}$ to get inifite solutions: $A\underbrace{(\boldsymbol{s} + \boldsymbol{n})}_{\text{x}} = A\boldsymbol{s} + 0 = \boldsymbol{b}$ with $\boldsymbol{n} \in \mathrm{N}(A)$.



4.10.1 Sol. Space is shifted Nullspace (Theo. 4.38)

Let $A \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. Let $\boldsymbol{s}$ some solution for $\boldsymbol{x}$ to $A\boldsymbol{x} = \boldsymbol{b}$, then

$$\operatorname{Sol}(A, \boldsymbol{b}) = \{\boldsymbol{s} + \boldsymbol{n} : \boldsymbol{n} \in \mathrm{N}(A)\}.$$

5 Orthogonality

5.1 Orthogonal Subspaces (Def. 5.1.1)

Two subspaces V and W are orthogonal if for all $\boldsymbol{v} \in V$ and $\boldsymbol{w} \in W$, $v^T w = 0$.

More specifically, this also holds for the bases, i.e., let $v_1, \dots, v_k$ be a basis of V and $w_1, \dots, w_l$ be a basis of W. V and W are orthogonal if and only if $v_i \cdot w_j = 0\ \forall i \in [k]$ and $j \in [l]$ (Lemma 5.1.2).

5.2 Orthogonal Complement (Def. 5.1.5)

Let V be a subspace of $\mathbb{R}^n$, the orthogonal complement to V is defined as

$$V^\perp \coloneqq \{\boldsymbol{w} \in \mathbb{R}^n : \boldsymbol{w}^T\boldsymbol{v} = 0 \text{ for all } \boldsymbol{v} \in V\}.$$

5.3 Double Complement (Lemma 5.1.8)

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

5.4 Orthogonal Decomp. (Theo. 5.1.7)

Let V, W be subspaces of $\mathbb{R}^n$, then these are equivalent

(i) $W = V^\perp$,
(ii) $\dim(V) + \dim(W) = n$,
(iii) Every $\boldsymbol{u} \in \mathbb{R}^n$ can be written as $\boldsymbol{u} = \boldsymbol{v} + \boldsymbol{w}$ with $\boldsymbol{v} \in V$ and $\boldsymbol{w} \in W$.

5.5 Intersection of Orthogonal Subspaces (Cor. 5.1.4)

Let V and W be orthogonal subspaces. Then their intersection is trivially only at the origin:

$$V \cap W = \{0\}$$

5.6 Orthogonal Matrix Subspaces (Cor. 5.1.9 & Lemma 5.1.10)

Let $A \in \mathbb{R}^{m\times n}$, then

(i) $\mathrm{N}(A) = \mathrm{C}(A^T)^\perp$, also $N(A) = N(A^TA)$,
(ii) $\mathrm{C}(A^T) = \mathrm{N}(A)^\perp$, also $C(A^T) = C(A^TA)$.

5.7 Decomp. of the Sol. Space (Theo. 6.2.2)

Let $A \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. $\operatorname{Sol}(A, \boldsymbol{b}) = \boldsymbol{x}_1 + \mathrm{N}(A)$ with $\boldsymbol{x}_1 \in \mathrm{R}(A)$ such that $A\boldsymbol{x}_1 = b$.

6 Projections

6.1 Projection (Def. 5.2.1)

The projection of a vector $\boldsymbol{b} \in \mathbb{R}^m$ on a subspace S is the point in S closest to $\boldsymbol{b}$:

$$\operatorname{proj}_S(\boldsymbol{b}) = \operatorname{argmin}_{\boldsymbol{p}\in S} \|\boldsymbol{b} - \boldsymbol{p}\|$$

6.2 Normal Equations (Lemma 5.2.3)

The projection of $\boldsymbol{b}$ on $S = C(A)$ is well defined and given by $\boldsymbol{p} = A\hat{\boldsymbol{x}}$, where $\hat{\boldsymbol{x}}$ satisfies the **normal equations**:

$$A^TA\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$$

6.3 Projection Matrix (Theo. 5.2.5)

Let S be a subspace of $\mathbb{R}^m$ and A be a matrix with columns that are a basis of S. The projection of $\boldsymbol{b} \in \mathbb{R}^m$ on S is given by

$$\operatorname{proj}_S(\boldsymbol{b}) = P\boldsymbol{b} \text{ with } P = A(A^TA)^{-1}A^T = QQ^T.$$

Moreover,

(i) A^TA is invertible $\iff$ A has linearly independent columns (Lemma 5.2.4),
(ii) if A has linearly independent columns, then A^TA is square, invertible and symmetric.

6.3.1 Remark 5.2.6

For any projection matrix P and corresponding subspace S,

(i) $$P^2 = \left(A(A^TA)^{-1}A^T\right)^2 = A(A^TA)^{-1}\overbrace{A^TA(A^TA)^{-1}}^{I}A^T = A(A^TA)^{-1}A^T = P$$
(ii) $\operatorname{proj}_{S^\perp}(\boldsymbol{b}) = \boldsymbol{b} - P\boldsymbol{b} = (I - P)\boldsymbol{b}$.

6.4 Matrix for Reflection Through a Plane

To find the matrix B representing a reflection through a plane P passing through the origin:

(i) **Find the normal vector** $\boldsymbol{v}$ from the plane equation ($ax + by + cz = 0 \implies \boldsymbol{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$). $\boldsymbol{v}$ is orthogonal to the plane.
(ii) **Normalize it** to get the unit normal vector $\boldsymbol{n} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}$.
(iii) **Apply the reflection formula:** $B = I - 2\boldsymbol{n}\boldsymbol{n}^T$
Note: This formula works because $\boldsymbol{n}\boldsymbol{n}^T$ is the projection onto the normal, i.e. $I - \boldsymbol{n}\boldsymbol{n}^T$ would be the projection onto the plane. Subtracting it twice reflects the vector across the plane.

7 Linear Regression

7.1 Least Squares Approximation (Section 6.1)

A linear regression through the data points $(t_1, b_1), (t_2, b_2), \dots, (t_m, b_m)$ can be expressed in algebraic terms as minimizing the sum of the squared errors (where α_0 is the intercept, and α_1 the slope),

$$\min_{\alpha_0, \alpha_1} \left\| \boldsymbol{b} - A\underbrace{\begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix}}_{\boldsymbol{\alpha}} \right\|, \text{where } \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \text{ and } A = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}.$$

To minimize error, the error vector must be orthogonal to columns of A, so the minimizer satisfies the **normal equations** (Fact 6.1.1):

$$A^T(\boldsymbol{b} - A\boldsymbol{\alpha}) = 0$$
$$A^TA\boldsymbol{\alpha} = A^T\boldsymbol{b} \text{ ("normal equation", Gauss-J } [A^TA \mid A^T\boldsymbol{b}])$$
$$\boldsymbol{\alpha} = (A^TA)^{-1}A^T\boldsymbol{b}$$
$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} m & \sum_{k=1}^m t_k \\ \sum_{k=1}^m t_k & \sum_{k=1}^m t_k^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{k=1}^m b_k \\ \sum_{k=1}^m t_k b_k \end{bmatrix}.$$

Moreover, the columns of the $m \times 2$ matrix A are linearly dependent $\iff t_i = t_j$ for all $i \neq j$ (Lemma 6.1.2).

7.2 Remark 6.1.3

If the columns of A are pairwise orthogonal (which corresponds to $\sum_{k=1}^m t_k = 0$), then A^TA is a diagonal matrix and we can further simply the expression,

$$\begin{bmatrix}\alpha_0\\ \alpha_1\end{bmatrix} = \begin{bmatrix}\frac{1}{m} & 0\\ 0 & \frac{1}{\sum_{k=1}^m t_k^2}\end{bmatrix}\begin{bmatrix}\sum_{k=1}^m b_k\\ \sum_{k=1}^m t_k b_k\end{bmatrix} = \begin{bmatrix}\frac{1}{m}\sum_{k=1}^m b_k \\ \frac{\sum_{k=1}^m t_k b_k}{\sum_{k=1}^m t_k^2}\end{bmatrix}$$

7.3 Least Squares Example

Find $a, b \in \mathbb{R}$ for the function $f(x) = ax^2 + b$ that minimizes the squared error for the data points $(-1, 2), (0, 1)$, and $(1, 3)$.

We set up the system $A\boldsymbol{x} \approx \boldsymbol{y}$ where unknowns are $\boldsymbol{x} = \left[\begin{smallmatrix}b\\a\end{smallmatrix}\right]$ corresponding to basis functions 1 and x^2:

$$A = \begin{bmatrix}1 & (-1)^2\\ 1 & 0^2\\ 1 & 1^2\end{bmatrix} = \begin{bmatrix}1 & 1\\ 1 & 0\\ 1 & 1\end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix}2\\1\\3\end{bmatrix}.$$

Solve via normal equations $A^T A\boldsymbol{x} = A^T\boldsymbol{y}$:

$$A^T A = \begin{bmatrix}3 & 2\\ 2 & 2\end{bmatrix}, \quad A^T\boldsymbol{y} = \begin{bmatrix}6\\5\end{bmatrix} \Rightarrow \begin{bmatrix}3 & 2\\ 2 & 2\end{bmatrix}\begin{bmatrix}b\\a\end{bmatrix} = \begin{bmatrix}6\\5\end{bmatrix}.$$

Row reduction or substitution yields $b = 1$ and $a = \frac{3}{2}$.

8 Orthonormal Bases

8.1 Orthonormal Vectors (Def. 6.3.1)

Vectors $\boldsymbol{q}_1, ..., \boldsymbol{q}_n \in \mathbb{R}^m$ are orthonormal if they are orthogonal and have norm 1, i.e., if for all $i, j \in [n]$

$$\boldsymbol{q}_i^T\boldsymbol{q}_j = \begin{cases}1 \text{ if } i = j\\ 0 \text{ otherwise}\end{cases}.$$

8.2 Orthogonal Matrix (Def. 6.3.3)

A square matrix $Q \in \mathbb{R}^{n\times n}$ is orthogonal when $Q^T Q = I$. Then $QQ^T = I$ and the columns of Q form an orthonormal basis of $\mathbb{R}^n$.

Additionally, orthogonal matrices *preserve norm and dot product*, i.e.,

$$\|Q\boldsymbol{x}\| = \|\boldsymbol{x}\| \text{ and } (Q\boldsymbol{x})^T(Q\boldsymbol{y}) = \boldsymbol{x}^T\boldsymbol{y},$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ (Prop. 6.3.6).

- If A and B are orthogonal, then so is AB.

8.3 Gram-Schmidt (Alg. 6.3.8)

Given n linearly independent vectors $a_1, ..., a_n$ that span S, $q_1, ..., q_n$ can be constructed as follows:

(i) $q_1 = \frac{a_1}{\|a_1\|}$

(ii) For $k = 2, ..., n$,

$$q_{k'} = a_k - \sum_{i=1}^{k-1} \underbrace{(a_k^T q_i)}_{R_{ik}} q_i \quad \longrightarrow \quad q_k = \frac{q_{k'}}{\underbrace{\|q_{k'}\|}_{\|q_{k'}\| = R_{ii}}}.$$

The vectors $q_1, ..., q_n$ form an orthonormal basis of S (Theorem 6.3.9).

Note: An upper triangular $n \times n$ matrix with non-zero diagonals that does **not** yield the canonical basis after the Gram-Schmidt process is $-I$.

8.4 QR-Decomposition (Def. 6.3.10)

Let $A \in \mathbb{R}^{m\times n}$ with linearly independent columns.

$$A = QR$$

where Q is an $m \times n$ matrix with orthonormal columns and R is an upper triangular matrix given by $R = Q^T A$. Use Gram-Schmidt on the columns of A to get Q.

8.5 Projections with QR (Fact 6.3.12)

(i) Any projection on $C(A)$ can also be done by Q, as $C(A) = C(Q)$, following $\text{proj}_{C(A)}(\boldsymbol{b}) = Q(Q^T Q)^{-1}Q^T\boldsymbol{b} = QQ^T\boldsymbol{b}$.

(ii) The normal equation can be re-written as

$$\begin{aligned} A^T A\hat{x} &= A^T b\\ R^T R\hat{x} &= R^T Q^T\boldsymbol{b} \quad (R^T \text{ is invertible})\\ R\hat{x} &= Q^T\boldsymbol{b}. \text{ (useful for least squares)}\\ A\hat{x} &= QQ^T\boldsymbol{b}\end{aligned}$$

9 Pseudoinverses

9.1 Full Column Rank (Def. 6.4.1)

For any $A \in \mathbb{R}^{m\times n}$ with $\text{rank}(A) = n$,
$A^+_{\text{left}} = (A^T A)^{-1}A^T$ is a **left** inverse, $\Longrightarrow A^+ A = I$.

9.2 Full Row Rank (Def. 6.4.3)

For any $A \in \mathbb{R}^{m\times n}$ with $\text{rank}(A) = m$,
$A^+_{\text{right}} = A^T(AA^T)^{-1}$ is a **right** inverse, $\Longrightarrow AA^+ = I$.

9.3 All Matrices (Def. 6.4.7)

For any $R \in \mathbb{R}^{m\times n}$ with rank r and CR-decomposition $A = CR$,

$$\begin{aligned} A^+ &= (CR)^+ = R^+C^+\\ &= R^T(RR^T)^{-1}(C^T C)^{-1}C^T\\ &= R^T(C^T AR^T)^{-1}C^T.\end{aligned}$$

9.4 Pseudoinverses with rank r (Prop. 6.4.9)

For $A \in \mathbb{R}^{m\times n}$ with rank r, let $S \in \mathbb{R}^{m\times r}$ and $T \in \mathbb{R}^{r\times n}$, such that $A = ST$, then $A^+ = T^+S^+$.

9.5 Optimization Properties (Lemma 6.4.8)

Given $A \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$, the vector $\hat{\boldsymbol{x}} = A^+_{\text{left}}\boldsymbol{b}$ is the unique solution to the problem:

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \|\boldsymbol{x}\|_2 \text{ subject to } A^T A\boldsymbol{x} = A^T\boldsymbol{b}$$

$$\boldsymbol{x} = (A^T A)^{-1}A^T\boldsymbol{b}$$

In words: $A^+_{\text{left}}\boldsymbol{b}$ is the solution to the least squares problem with the smallest norm.

9.6 Properties of A^+ (Theo. 6.4.10)

(i) $AA^+A = A$

(ii) $A^+AA^+ = A^+$. *Proof:* see below.

(iii) AA^+_{left} is symmetric & the projection matrix on $C(A)$.

(iv) $A^+_{\text{right}}A$ is symmetric & the projection matrix on $C(A^T)$.

(v) $(A^T)^+ = (A^+)^T$. *Proof:* see below.

9.7 Proof of Pseudoinverse Properties

(i) Prove that if $\text{rank}(A) = \text{rank}(B) = n$, we have $(AB)^+ = B^+A^+$.

$C(AB) = C(A)$ (as $\text{rank}(B) = n$ implies that $C(B) = \mathbb{R}^n$). Then using Prop. 6.4.9 we get $(AB)^+ = B^+A^+$.

(ii) Prove that $A^+AA^+ = A^+$.

$A^+AA^+ = (CR)^+CR(CR)^+ = R^+(C^+C)(RR^+)C^+ = R^+C^+ = A^+$.

(iii) Prove that $(A^T)^+ = (A^+)^T$.

---

prove for full row & column rank seperately, then use Prop. 6.4.9 to get $(A^T)^+ = (C^T)^+(R^T)^+ = (C^+)^T(R^+)^T = (R^+C^+)^T = (A^+)^T$.

(iv) Prove that A^+A is symmetric and that it is the projection matrix for the subspace $C(A^T)$.

---

projection matrix: $A^+A = (CR)^+CR = R^+C^+CR = R^T(RR^T)^{-1}R. \quad (C(A^T) = C(R^T))$
symmetric: $(A^+A)^T = \left(R^T(RR^T)^{-1}R\right)^T = R^T\left((RR^T)^T\right)^{-1}R = R^T(RR^T)^{-1}R = A^+A$.

10 Certificate of Unsolvability (Theo 6.2.4)

Let $A \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. The system of linear equations has no solution iff there exists non-zero vector $\boldsymbol{z}$ s.t.:

$$\{\boldsymbol{x} \in \mathbb{R}^n \mid A\boldsymbol{x} = \boldsymbol{b}\} = \emptyset$$
$$\Longleftrightarrow \{\boldsymbol{z} \in \mathbb{R}^m \mid A^T\boldsymbol{z} = \boldsymbol{0}, \boldsymbol{b}^T\boldsymbol{z} = 1\} \neq \emptyset$$

10.1 Parametric Solvability Example

For $A = \left[\begin{smallmatrix}2 & 1\\ 1 & 3\\ 0 & 2\end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix}b_1\\ b_2\\ b_3\end{smallmatrix}\right]$. Find b_i s.t. $A\boldsymbol{x} + B\boldsymbol{y} = \boldsymbol{c}$ is solvable for all $\boldsymbol{c}$.

We know $A\boldsymbol{x} + B\boldsymbol{y} = \boldsymbol{c}$ is solvable for all $\boldsymbol{c}$ iff the only solution to $\left[\begin{smallmatrix}A^\top\\ B^\top\end{smallmatrix}\right]\boldsymbol{z} = \boldsymbol{0}$ is with $\boldsymbol{z} = \boldsymbol{0}$.

$$\begin{bmatrix}2 & 1 & 0\\ 1 & 3 & 2\\ b_1 & b_2 & b_3\end{bmatrix}\begin{bmatrix}z_1\\ z_2\\ z_3\end{bmatrix} = \boldsymbol{0} \Longrightarrow \begin{cases}2z_1 + z_2 &= 0\\ z_1 + 3z_2 + 2z_3 &= 0\\ b_1z_1 + b_2z_2 + b_3z_3 &= 0\end{cases}$$

$$\Longrightarrow \begin{cases}2z_1 = -z_2\\ z_1 + 3z_2 = -2z_3\end{cases} \Longrightarrow z_1 = -\frac{1}{2}z_2, \quad z_3 = -\frac{5}{4}z_2$$

Sub. into row 3: $b_1\left(-\frac{1}{2}z_2\right) + b_2z_2 + b_3\left(-\frac{5}{4}z_2\right) = 0$.

$$\left(-\frac{1}{2}b_1 + b_2 - \frac{5}{4}b_3\right)z_2 = 0$$

A non-zero $\boldsymbol{z}$ can only exist if $\left(-\frac{1}{2}b_1 + b_2 - \frac{5}{4}b_3\right) = 0$, which would mean $A\boldsymbol{x} + B\boldsymbol{y} = \boldsymbol{c}$ is **not** solvable. But we want it to be solvable, so we must force $\boldsymbol{z}$ to be 0, which is only possible if $-\frac{1}{2}b_1 + b_2 - \frac{5}{4}b_3 \neq 0 \Longrightarrow -2b_1 + 4b_2 - 5b_3 \neq 0$.

10.2 The Algebraic Certificate

The vector $\boldsymbol{z}$ serves as a "certificate" that $A\boldsymbol{x} = \boldsymbol{b}$ is impossible. If both $\boldsymbol{x}$ and $\boldsymbol{z}$ existed, we would reach the following contradiction:

$$0 = \boldsymbol{0}^T\boldsymbol{x} = (\boldsymbol{z}^T A)\boldsymbol{x} = \boldsymbol{z}^T(A\boldsymbol{x}) = \boldsymbol{z}^T\boldsymbol{b} = 1$$

10.3 Characterizing Solvability

- **Row Independence:** If the rows of A are linearly independent, then $A^T\boldsymbol{z} = \boldsymbol{0}$ only has the trivial solution $\boldsymbol{z} = \boldsymbol{0}$. Since $\boldsymbol{b}^T\boldsymbol{0} = 0 \neq 1$, a certificate $\boldsymbol{z}$ can never exist. Therefore, $A\boldsymbol{x} = \boldsymbol{b}$ always has a solution for every $\boldsymbol{b}$.
- **Linear Independence:** A vector $\boldsymbol{b}$ is linearly independent from the columns of A if and only if $A\boldsymbol{x} = \boldsymbol{b}$ has no solution, which can be verified by finding the certificate $\boldsymbol{z}$.

To find $\boldsymbol{z}$, use Gaussian elimination (RREF) on the augmented matrix:

$$\begin{bmatrix}A^T\\ \boldsymbol{b}^T\end{bmatrix}\boldsymbol{z} = \begin{bmatrix}0\\ 1\end{bmatrix}.$$

10.4 The Set of All Solutions (Theorem 6.2.2)

If the system is solvable, the set of all solutions is a shifted copy of the nullspace:

$$\{\boldsymbol{x} \in \mathbb{R}^n \mid A\boldsymbol{x} = \boldsymbol{b}\} = \boldsymbol{x}_r + N(A)$$

where $\boldsymbol{x}_r$ is the **unique** vector in the row space $R(A) = C(A^T)$ such that $A\boldsymbol{x}_r = \boldsymbol{b}$. This uniqueness is guaranteed because A acts injectively on its row space (Lemma 6.2.1).

Every solution $\boldsymbol{x}$ is a mix of two orthogonal parts:

$$\boldsymbol{x} = \underbrace{\boldsymbol{x}_r}_{\text{Row Space (unique)}} + \underbrace{\boldsymbol{x}_n}_{\text{Nullspace (any)}}$$

(i) **The "Pure" Solution ($\boldsymbol{x}_r$):** There is exactly one solution living entirely in the Row Space (the "Active Zone" of A). This is the solution with the *shortest length* because it contains no "waste" (Lemma 6.4.5).

(ii) **The "Invisible" Noise ($\boldsymbol{x}_n$):** You can add *any* vector $\boldsymbol{x}_n$ from the Nullspace to $\boldsymbol{x}_r$ without changing the result. **Why?** Because the matrix is blind to it:

$$A(\boldsymbol{x}_r + \boldsymbol{x}_n) = A\boldsymbol{x}_r + A\boldsymbol{x}_n = \boldsymbol{b} + 0 = \boldsymbol{b}$$

11 Determinant (Def. 7.2.3)

Let $A \in \mathbb{R}^{n\times n}$, the determinant $\det(A)$ is defined as

$$\det(A) = \sum_{\sigma\in\Pi_n} \text{sgn}(\sigma)\Pi_{i=1}^n A_{i,\sigma(i)},$$

where Π_n is the set of all permutations of n elements. Moreover,

$\det(I) = 1$	(Prop. 7.2.4)
If $A \in \mathbb{R}^{1\times 1}$, $\det(A) = A$	(Def. 7.2.3)
$\det(A^T) = \det(A)$	(Theo. 7.2.5)
If A is triangular, $\det(A) = \Pi_{k=1}^n A_{kk}$	(Prop. 7.2.4)
If A is orthogonal, $\det(A) = 1$ or -1	(Prop. 7.2.4)
A is invertible if and only if $\det(A) \neq 0$	(Theo. 7.2.6)
If $\det(A) \neq 0$, $\det(A^{-1}) = \frac{1}{\det(A)}$	(Theo. 7.2.6)
Given some $B \in \mathbb{R}^{n\times n}$, $\det(AB) = \det(A)\cdot\det(B)$	(Theo. 7.2.6)

- $\det(A + B) \neq \det(A) + \det(B)$
- If any two rows are equal, then $\det(A) = 0$.
- If A has a row of zeros, then $\det(A) = 0$.
- If any Eigenvalue of A is 0, then $\det(A) = 0$.
- If we swap the rows of $A \to B$ once, then $\det(B) = -\det(A)$ (Prop. 7.3.6).
- The determinant is a linear function of each row separately.
  - If a single row of A is multiplied by some scalar t, then $\det(A') = t\cdot\det(A)$.
    - If the whole matrix is multiplied by t (i.e. all n rows are multiplied by t), then $\det(t\cdot A) = t^n\cdot\det(A)$. *Proof:* $\det(tA) = \sum_{\sigma\in\Pi_n}\text{sgn}(\sigma)\prod_{i=1}^n\left(t\cdot a_{i,\sigma(i)}\right) = ... = t^n\sum_{\sigma\in\Pi_n}\text{sgn}(\sigma)\prod_{i=1}^n a_{i,\sigma(i)} = t^n\det(A)$.
  - If a row of A is replaced by the sum of itself and a multiple of another row, the determinant stays unchanged.
- $\det(Q) \in \{1, -1\}$, because $Q^TQ = I$ and $\det(Q^TQ) = \det(Q^T)\det(Q) = \det(I) = \det(Q)^2 = 1$, so $\det(Q) \in \{1, -1\}$.

11.1 Determinant of a 2×2 Matrix

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $\det(A) = ad - bc$.

11.2 Det. through Co-Factors

11.2.1 Co-Factors (Def. 7.3.1)

Let $A \in \mathbb{R}^{n \times n}$, for each $1 \le i, j \le n$, let $\mathcal{A}_{ij}$ denote the matrix obtained by removing the i-th row and j-th column from A. The co-factors or A are $C_{ij} = (-1)^{i+j} \det(\mathcal{A}_{ij})$.

11.2.2 Determinant (Prop. 7.3.2)

We can then rewrite the determinant of A as

$$\det(A) = \sum_{j=1}^{n} A_{ij} C_{ij}, \quad \text{for some } 1 \le i \le n$$

As in, make a $+-+\ldots$ grid, pick a row or column and calculate $\pm A_{ij} \det(\ldots)$ for that whole row or column recursively.

11.3 Cramer's Rule (Prop. 7.3.5)

Let $A \in \mathbb{R}^{n \times n}$, such that $\det(A) \neq 0$ and $\boldsymbol{b} \in \mathbb{R}^n$. Then the solution $\boldsymbol{x} \in \mathbb{R}^n$ for $A\boldsymbol{x} = \boldsymbol{b}$ is given by

$$x_j = \frac{\det(\mathcal{B}_j)}{\det(A)},$$

where $\mathcal{B}_j$ is the matrix obtained by replacing the j-th column of A with $\boldsymbol{b}$.

11.4 Lineary of the Determinant (Prop. 7.3.7)

The determinant is linear in each row or each column, i.e.,

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

11.5 Example

Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathbb{R}^n$ and $M \in \mathbb{R}^{(n-2) \times n}$ be arbitrary and consider the four $n \times n$ matrices

$$A = \begin{bmatrix} —\boldsymbol{v}_1^T— \\ —\boldsymbol{u}_1^T— \\ M \end{bmatrix}, B = \begin{bmatrix} —\boldsymbol{v}_1^T— \\ —\boldsymbol{u}_2^T— \\ M \end{bmatrix},$$

$$C = \begin{bmatrix} —\boldsymbol{v}_2^T— \\ —\boldsymbol{u}_1^T— \\ M \end{bmatrix}, D = \begin{bmatrix} —\boldsymbol{v}_2^T— \\ —\boldsymbol{u}_2^T— \\ M \end{bmatrix}$$

as well as the following $n \times n$ matrix

$$E = \begin{bmatrix} —(\boldsymbol{v}_1 - \boldsymbol{v}_2)^T— \\ —(\boldsymbol{u}_1 - \boldsymbol{u}_2)^T— \\ M \end{bmatrix}.$$

Find a formula for $\det(E)$ in terms of $\det(A)$, $\det(B)$, $\det(C)$ and $\det(D)$.

Use Prop. 7.3.7 to get

$$\det(A) - \det(B) = \det \begin{bmatrix} —\boldsymbol{v}_1^T— \\ —\boldsymbol{u}_1^T— \\ M \end{bmatrix} - \det \begin{bmatrix} —\boldsymbol{v}_1^T— \\ —\boldsymbol{u}_2^T— \\ M \end{bmatrix}$$

$$= \det \begin{bmatrix} —\boldsymbol{v}_1^T— \\ —(\boldsymbol{u}_1 - \boldsymbol{u}_2)^T— \\ M \end{bmatrix}$$

and

$$\det(C) - \det(D) = \det \begin{bmatrix} —\boldsymbol{v}_2^T— \\ —\boldsymbol{u}_1^T— \\ M \end{bmatrix} - \det \begin{bmatrix} —\boldsymbol{v}_2^T— \\ —\boldsymbol{u}_2^T— \\ M \end{bmatrix}$$

$$= \det \begin{bmatrix} —\boldsymbol{v}_2^T— \\ —(\boldsymbol{u}_1 - \boldsymbol{u}_2)^T— \\ M \end{bmatrix}$$

together this gives us

$$\det \begin{bmatrix} —\boldsymbol{v}_1^T— \\ —(\boldsymbol{u}_1 - \boldsymbol{u}_2)^T— \\ M \end{bmatrix} - \det \begin{bmatrix} —\boldsymbol{v}_2^T— \\ —(\boldsymbol{u}_1 - \boldsymbol{u}_2)^T— \\ M \end{bmatrix}$$

$$= \det \begin{bmatrix} —(\boldsymbol{v}_1 - \boldsymbol{v}_2)^T— \\ —(\boldsymbol{u}_1 - \boldsymbol{u}_2)^T— \\ M \end{bmatrix}.$$

Hence, $\det(E) = \det(A) - \det(B) - (\det(C) - \det(D))$.

12 Eigenvalues/vectors (Def. 8.2.1)

Let $A \in \mathbb{R}^{n \times n}$, $\lambda \in \mathbb{C}$ is an eigenvalue of A and $\boldsymbol{v} \in \mathbb{C}^n \setminus \{\boldsymbol{0}\}$ is the associated eigenvector of A, if

$$A\boldsymbol{v} = \lambda \boldsymbol{v}.$$

To calculate eigenvalue/vector pairs, we use Lemma 8.2.3:

(i) $\det(A - \lambda I) = 0 \iff \lambda$ is an eigenvalue of A,
(ii) $\boldsymbol{v}$ is an eigenvector of A (associated with λ) $\iff \boldsymbol{v} \in \mathrm{N}(A - \lambda I)$ and $\boldsymbol{v} \neq \boldsymbol{0}$.

Every matrix $A \in \mathbb{R}^{n \times n}$ has an eigenvalue (Theo. 8.2.5).

12.1 Eigenvector Recurrence in "Closed Form"

Consider the sequence of numbers given by $a_0 = 1$, $a_1 = 1$ and $a_n = -a_{n-1} + 6a_{n-2}$ for $n \ge 2$. Find $\alpha, \beta \in \mathbb{R}$ such that $a_n = \frac{4}{5}\alpha^n + \frac{1}{5}\beta^n$ for all $n \in \mathbb{N}_0$. Prove your answer.

(i) Define sequence algebraically: Let $\boldsymbol{v}_n = \begin{bmatrix} a_{n+1} \\ a_n \end{bmatrix}$. We translate the recurrence $a_n = -a_{n-1} + 6a_{n-2}$ into matrix form.

$$\underbrace{\begin{bmatrix} a_n \\ a_{n-1} \end{bmatrix}}_{\boldsymbol{v}_{n-1}} = \underbrace{\begin{bmatrix} -1 & 6 \\ 1 & 0 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a_{n-1} \\ a_{n-2} \end{bmatrix}}_{\boldsymbol{v}_{n-2}}$$

Note: Since $\boldsymbol{v}_0 = \begin{bmatrix} a_1 \\ a_0 \end{bmatrix}$ has a_0 at the bottom, our target a_n (which we'll need later) will always correspond to the **bottom component** of $\boldsymbol{v}_n$. This is more clear if we shift the above stated matrix equation:

$$\underbrace{\begin{bmatrix} a_{n+1} \\ a_n \end{bmatrix}}_{\boldsymbol{v}_n} = \underbrace{\begin{bmatrix} -1 & 6 \\ 1 & 0 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a_n \\ a_{n-1} \end{bmatrix}}_{\boldsymbol{v}_{n-1}}$$

(iii) Find the corresponding eigenvectors $\boldsymbol{v}_i \in N(A - \lambda_i I)$:

$$(A - \lambda_i I)\boldsymbol{v}_i = 0$$

$$\begin{bmatrix} -3 & 6 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} (\boldsymbol{v}_1)_1 \\ (\boldsymbol{v}_1)_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies \boldsymbol{v}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

$$\begin{bmatrix} 2 & 6 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} (\boldsymbol{v}_2)_1 \\ (\boldsymbol{v}_2)_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies \boldsymbol{v}_2 = \begin{bmatrix} -3 \\ 1 \end{bmatrix},$$

(iv) Express $\boldsymbol{v}_0$ as a linear combination $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2$: We solve the system for the initial vector $\boldsymbol{v}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$:

$$c_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} -3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \implies \begin{cases} 2c_1 - 3c_2 = 1 \\ c_1 + \ c_2 \ = 1 \end{cases}$$

From the second equation, $c_1 = 1 - c_2$. Substituting into the first:

$$2(1 - c_2) - 3c_2 = 1$$

$$2 - 5c_2 = 1 \implies 5c_2 = 1 \implies c_2 = \frac{1}{5}$$

Then, $c_1 = 1 - \frac{1}{5} = \frac{4}{5}$. Thus:

$$\begin{bmatrix} a_1 \\ a_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{4}{5}\boldsymbol{v}_1 + \frac{1}{5}\boldsymbol{v}_2$$

(v) Rewrite $\boldsymbol{v}_n = \begin{bmatrix} a_{n+1} \\ a_n \end{bmatrix} = A^n \boldsymbol{v}_0$ as

$$\boldsymbol{v}_n = A^n \left(\frac{4}{5}\boldsymbol{v}_1 + \frac{1}{5}\boldsymbol{v}_2 \right) = \frac{4}{5}A^n\boldsymbol{v}_1 + \frac{1}{5}A^n\boldsymbol{v}_2$$

$$= \frac{4}{5}\lambda_1^n\boldsymbol{v}_1 + \frac{1}{5}\lambda_2^n\boldsymbol{v}_2 = \frac{4}{5}2^n\begin{bmatrix} 2 \\ 1 \end{bmatrix} + \frac{1}{5}(-3)^n\begin{bmatrix} -3 \\ 1 \end{bmatrix}.$$

(vi) Consider second component: $a_n = \frac{4}{5} \cdot 2^n \cdot 1 + \frac{1}{5} \cdot (-3)^n \cdot 1$.

12.2 Reverse Diagonalization Example

Construct a square matrix B with eigenvalues $0, 1, 2$, such that B is not a diagonal matrix.

Let A be the diagonal matrix with $0, 1, 2$ on its diagonals and let $V = \begin{bmatrix} | & | & | \\ e_1 + e_2 & e_1 - e_2 & e_3 \\ | & | & | \end{bmatrix}$, then $B = VAV^{-1}$. Ensure $\det(V) \neq 0$.

12.3 Quadratic Formula (for finding λ)

To find zeros of $ax^2 + bx + c = 0$ are given by

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

12.4 Special Eigenvalues

(i) If λ and $\boldsymbol{v}$ are an eigenvalue-eigenvector pair of A, then λ^k and $\boldsymbol{v}$ are one for A^k. Induction *Proof:* $A^k\boldsymbol{v} = A(A^{k-1}\boldsymbol{v}) = A(\lambda^{k-1}\boldsymbol{v}) = \lambda^{k-1}(A\boldsymbol{v}) = \lambda^k\boldsymbol{v}$ (Prop. 8.3.1).
(ii) Let A be invertible, if λ and $\boldsymbol{v}$ are an eigenvalue-eigenvector pair of A, then $\frac{1}{\lambda}$ and $\boldsymbol{v}$ are an eigenvalue-eigenvector pair of A^{-1}. *Proof:* $A\boldsymbol{v} = \lambda\boldsymbol{v} \iff \boldsymbol{v} = A^{-1}(\lambda\boldsymbol{v}) \iff \lambda A^{-1}\boldsymbol{v} = \boldsymbol{v} \iff A^{-1}\boldsymbol{v} = \frac{1}{\lambda}\boldsymbol{v}$ (works since $\lambda \neq 0$) (Prop. 8.3.1).
(iii) Let $A \in \mathbb{R}^{n \times n}$, the eigenvalues of A are the same ones as of A^T. *Proof:* $\det(A - zI) = \det((A - zI)^T) = \det(A^T - zI)$ (Lemma 8.3.5).
(iv) Let $Q \in \mathbb{R}^{n \times n}$ be an orthogonal matrix, if λ is an eigenvalue of Q, then $|\lambda| = 1$. *Proof:* $\|\boldsymbol{v}\|^2 = \|Q\boldsymbol{v}\|^2 = \|\lambda\boldsymbol{v}\|^2 = |\lambda| \cdot \|\boldsymbol{v}\|^2$ (Prop. 8.2.7).
(v) Let $A \in \mathbb{R}^{n \times n}$, if $(\lambda, \boldsymbol{v})$ is an eigenvalue-eigenvector pair of A, then $(\overline{\lambda}, \overline{\boldsymbol{v}})$ is an eigenvalue-eigenvector pair of A too. Thus, if $\lambda \in \mathbb{C}$ is an eigenvalue of A, then $\overline{\lambda}$ is also an eigenvalue of A (Lemma 8.2.8). *Proof:* See "Eigenvalues of Sym. Matrices (Cor. 9.2.2)" in the middle of the next page.
(vi) Let $P \in \mathbb{R}^{n \times n}$ be a projection matrix, then P has two distinct eigenvalues, 0 and 1 (every single one of the n eigenvalues is either 0 or 1) and a *complete set* of eigenvectors (Prop. 9.1.6).
(vii) Let $D \in \mathbb{R}^{n \times n}$ be a diagonal matrix, then its eigenvalues are its diagonal entries and the standard basis $(e_1, e_2, \ldots, e_n)$ is a complete set of eigenvectors of D (Example 9.1.4).
(viii) Let $T \in \mathbb{R}^{n \times n}$ be a triangular matrix, then its eigenvalues are its diagonal entries, however, T might not have a complete set of eigenvectors (Example 9.1.5).
(ix) $A + (k \cdot I) \iff$ adding k to all eigenvalues of A. Eigenvectors stay the same. *Proof:* $A\boldsymbol{v} = \lambda\boldsymbol{v} \implies (A + kI)\boldsymbol{v} = (A\boldsymbol{v}) + kI\boldsymbol{v} = \lambda\boldsymbol{v} + k\boldsymbol{v} = (\lambda + k)\boldsymbol{v}$
(x) If matrix A has an eigenvalue λ, then the matrix cA (where c is a scalar number) has the eigenvalue $c\lambda$.
Proof: $(cA)\boldsymbol{v} = c(A\boldsymbol{v}) = c(\lambda\boldsymbol{v}) = (c\lambda)\boldsymbol{v}$.
(xi) For $A \in \mathbb{R}^{2 \times 2}$, $\lambda_1, \lambda_2 = m \pm \sqrt{m^2 - p}$, $m = \frac{a+d}{2}$, $p = \det(A)$

12.5 Important Words Of Caution

(i) Even though the *eigenvalues* of A and A^T are the same, ***the eigenvectors are not***!
(ii) The eigenvalues of $A + B$ ***are not the sum*** of the eigenvalues of A and the eigenvalues of B!
(iii) The eigenvalues of AB ***are not the product*** of the eigenvalues of A and the eigenvalues of B!
(iv) ***Gaussian Elimination does not preserve eigenvalues and eigenvectors***!

12.6 Distinct Eigenvalues (Theo. 8.3.3)

Let $A \in \mathbb{R}^{n \times n}$ with n distinct, real eigenvalues, then there is a basis of $\mathbb{R}^n$ made up of eigenvectors of A. We also say that A has a *complete set* of eigenvectors (Def. 9.1.3).

12.7 Complex Numbers

Complex numbers are of the form $z = (a + ib) \in \mathbb{C}$ with $a, b \in \mathbb{R}$ and $i^2 = -1$. The following operations are defined:

(i) $(a + ib) + (x + iy) = (a + x) + i(b + y)$.
(ii) $(a + ib) \cdot (x + iy) = (ax - by) + i(ay + bx)$.
(iii) $(a + ib) \cdot (a - ib) = a^2 + b^2$.
(iv) $\frac{a+ib}{x+iy} = \frac{(a+ib)(x-iy)}{(x+iy)(x-iy)} = \frac{(ax+by)+i(bx-ay)}{(x^2+y^2)} = \left(\frac{ax+by}{x^2+y^2}\right) + i\left(\frac{bx-ay}{x^2+y^2}\right)$.
(v) $|z| = \sqrt{a^2 + b^2}$ (modulus).
(vi) $\overline{a + ib} = a - ib$ (conjugate).
(vii) $|z|^2 = z\overline{z}$.
(viii) $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$.
(ix) $\frac{1}{z} = \frac{\overline{z}}{|z|^2}$.

A complex number $z = (a + ib) \in \mathbb{C}$ can be written as $z = re^{i\theta}$ where $r = \overline{z}$ and $\theta = \tan^{-1}\left(\frac{b}{a}\right)$.

12.7.1 Complex Matrices/Vectors

Let $A \in \mathbb{C}^{m \times n}$, the *conjugate transpose* $A^* = \overline{A}^T$.

Given $\boldsymbol{v} \in \mathbb{C}^n$, we have

$$\|\boldsymbol{v}\|^2 = \boldsymbol{v}^*\boldsymbol{v} = \overline{\boldsymbol{v}}^T\boldsymbol{v} = \sum_{i=1}^{n} \overline{v}_i v_i = \sum_{i=1}^{n} |v_i|^2.$$

The dot-product in $\mathbb{C}^n$ is given by $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \boldsymbol{w}^*\boldsymbol{v}$.

12.8 Fund. Theorem of Algebra (Cor. 8.1.3)

Any degree $n \ge 1$ polynomial $P(z) = \alpha_n z^n + \ldots + a_1 z + a_0$ with $a_n \neq 0$ has n zeros: $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ such that

$$P(z) = \alpha_n(z - \lambda_1)(z - \lambda_2)\ldots(z - \lambda_n).$$

The number of times $\lambda \in \mathbb{C}$ appears in this expression is called the *algebraic multiplicity* of the zero (i.e. of the $P(\lambda) = 0$). This guarantees that an $n \times n$ matrix always has exactly n eigenvalues (if you count repeats and complex eigenvalues).

12.9 Geometric Multiplicity

Let $A \in \mathbb{R}^{n\times n}$ with eigenval λ_i, we call the dim of $\mathrm{N}(A-\lambda_i I)$ the *geometric multiplicity* of λ_i. Number of lin. indep. eigenvecs for λ_i.

12.10 Characteristic Polynomial

Let $A \in \mathbb{R}^{m\times n}$, the *characteristic polynomial* of A is

$$P(z) = (-1)^n \det(A - zI) = (z-\lambda_1)(z-\lambda_2)...(z-\lambda_n)$$

$$= z^n + \underbrace{\left(-\sum_{i=1}^{n} \lambda_i\right)}_{-\operatorname{Tr}(A)} z^{n-1} + \underbrace{\sum_{k=1}^{n-2} b_k z^k}_{\text{messy middle terms}} + \underbrace{(-1)^n \prod_{i=1}^{n} \lambda_i}_{(-1)^n \det(A)}.$$

12.11 Trace and Determinant (Lemma 8.3.6)

Let $A \in \mathbb{R}^{n\times n}$ and $\lambda_1, ..., \lambda_n$ its n eigenvalues, then

$$\operatorname{Tr}(A) = \sum_{i=1}^{n} \lambda_i \text{ and } \det(A) = \Pi_{i=1}^{n} \lambda_i.$$

Following (Lemma 8.3.7), for matrices A, B and $C \in \mathbb{R}^{n\times n}$,

(i) $\operatorname{Tr}(AB) = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}B_{ji} = \sum_{j=1}^{n}\sum_{i=1}^{n} B_{ji}A_{ij} = \operatorname{Tr}(BA)$.
(ii) $\operatorname{Tr}(A(BC)) = \operatorname{Tr}((BC)A) = \operatorname{Tr}((CA)B)$.

12.12 Diagonalization (Theo. 9.1.1)

Let $A \in \mathbb{R}^{n\times n}$ be a matrix with a complete set of eigenvectors.

Let $V = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix} \in \mathbb{R}^{n\times n}$ be the matrix whose columns are the eigenvectors, and $\Lambda \in \mathbb{R}^{n\times n}$ the matrix whose diagonal entries are the eigenvalues ($\Lambda_{ii} = \lambda_i$ for all $i \in [n]$), then

$$A = V\Lambda V^{-1}$$
$$AV = V\Lambda$$
$$[A\boldsymbol{v}_1, A\boldsymbol{v}_2, ..., A\boldsymbol{v}_n] = [\lambda_1\boldsymbol{v}_1, \lambda_2\boldsymbol{v}_2, ..., \lambda_n\boldsymbol{v}_n]$$

12.13 Diagonalizable Matrix (Def. 9.1.2)

A matrix $A \in \mathbb{R}^{n\times n}$ is said to be *diagonalizable*, if there are n independent eigenvectors (complete set of eigenvectors), and thus there exists an invertible matrix V, such that $V^{-1}AV = \Lambda$, where Λ is a diagonal matrix. *"Can we flatten A into a diagonal matrix?"*

12.14 Complete Set of Eigenvectors (Lemma 9.1.11)

A matrix has a complete set of eigenvectors if all its eigenvalues are real and the geometric multiplicities are the same as the algebraic multiplicities, for all of its eigenvalues.

If given a matrix $A \in \mathbb{R}^{n\times n}$, we can build a basis of $\mathbb{R}^n$ with eigenvectors of A (the eigenvectors are linearly independent), we say that A has a *complete set* of eigenvectors (Def. 9.1.3).

12.15 Similar Matrices (Def. 9.1.7)

Two matrices A and $B \in \mathbb{R}^{n\times n}$ are *similar*, if exists an invertible matrix S, such that $B = S^{-1}AS$.

Similar matrices have the same eigenvalues, Trace and Determinant. *Proof:* $Av = \lambda v \iff \lambda S^{-1}v = S^{-1}Av = S^{-1}A\underbrace{SS^{-1}}_{I}v = B(S^{-1}v)$.

Similar matrices are clones of each other. They represent the exact same linear transformation, just viewed from a different coordinate system (S is a change of basis matrix).

12.16 Characteristic Polynomial Example

Assume that $A, B \in \mathbb{R}^{n\times n}$ are similar, prove that their characteristic polynomials are equal.

As A and B are similar, there exists a matrix S such that $B = S^{-1}AS$. Recall that $\det(S)\det(S^{-1}) = 1$. Thus, $\det(A - zI) = \det(S^{-1})\det(A - zI)\det(S) = \det(S^{-1}(A - zI)S) = \det(S^{-1}AS - zS^{-1}IS) = \det(B - zI)$.

12.17 Spectral Theorem (Theo. 9.2.1)

Any symmetric matrix $A \in \mathbb{R}^{n\times n}$ has n real eigenvalues and an orthonormal basis consisting of its eigenvectors.

12.17.1 Diagonalization for Symmetric Matrices (Cor. 9.2.2)

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 the eigenvectors of A) and a diagonal matrix Λ whose entries are the eigenvalues of A, such that

$$A = V\Lambda V^T \text{ and } V^TV = I.$$

(This is also called the *eigendecomposition*). **Normalize V!**

12.18 Spectral Construction Example

Find a matrix A with orthonormal eigenvectors

$$\boldsymbol{v}_1 = \frac{1}{9}\begin{bmatrix} 1 \\ 8 \\ -4 \end{bmatrix}, \boldsymbol{v}_2 = \frac{1}{9}\begin{bmatrix} -4 \\ 4 \\ 7 \end{bmatrix}, \boldsymbol{v}_3 = \frac{1}{9}\begin{bmatrix} 8 \\ 1 \\ 4 \end{bmatrix}$$

and corresponding eigenvalues $\lambda_1 = 1, \lambda_2 = -1, \lambda_3 = 0$.

Let V be the 3×3 matrix with $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ as its columns and D the diagonal matrix with $\lambda_1, \lambda_2, \lambda_3$ on its diagonal, then $A = VDV^T$.

12.19 Eigenvalues of Sym. Matrices (Cor. 9.2.4)

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

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 the algebraic multiplicities (Remark 9.2.5).

Every symmetric matrix has only real eigenvalues (Lemma 9.2.8). *Proof:* $\overline{\lambda}\|\boldsymbol{v}\|^2 = \overline{\lambda}\boldsymbol{v}^*\boldsymbol{v} = (\lambda\boldsymbol{v})^*\boldsymbol{v} = (A\boldsymbol{v})^*\boldsymbol{v} = \boldsymbol{v}^*A^*\boldsymbol{v} = \boldsymbol{v}^*A\boldsymbol{v} = \boldsymbol{v}^*(\lambda\boldsymbol{v}) = \lambda\|\boldsymbol{v}\|^2 \implies \overline{\lambda} = \lambda$ only holds if $\lambda \in \mathbb{R}$.

12.20 Rayleigh Quotient (Prop. 9.2.10)

Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix, the *Rayleight Quotient*, defined for $\boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$, as

$$R(\boldsymbol{x}) = \frac{\boldsymbol{x}^TA\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}},$$

attains its max at $R(\boldsymbol{v}_{\max}) = \lambda_{\max}$ and its min at $R(\boldsymbol{v}_{\min}) = \lambda_{\min}$, where $\lambda_{\max}$ and $\lambda_{\min}$ are the max and min eigenvalues of A and $\boldsymbol{v}_{\max}$ and $\boldsymbol{v}_{\min}$ their eigenvectors. Useful because it allows to plug in any vector $\boldsymbol{x}$ and get back a number that's like a *weighted average* of the eigenvalue contained in that vector. Measures "how much" the matrix stretches that specific direction $\boldsymbol{x}$.

12.21 Positive (Semi-)Definite (Def. 9.2.11)

A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is said to be *positive semidefinite* (PSD) if all its eigenvalues are ≥ 0 and *positive definite* they are > 0. Moreover, (as per Prop. 9.2.12) A is

(i) PSD $\iff \boldsymbol{x}^TA\boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$,
(ii) PD $\iff \boldsymbol{x}^TA\boldsymbol{x} > 0$ for all $\boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$.

- Given two matrices A and B that are PSD (PD), their sum is also PSD (PD), i.e. they are closed under taking addition.
- A diagonal dominant matrix (diagonal entries are greater than the **absolute** sum of rest of the row's elements) is always PSD!
- Matrix is PSD/PD $\implies$ all diagonal entries are $\geq 0/> 0$ *Proof:* Choose $\boldsymbol{x} = \boldsymbol{e}_i$ in equations (i), (ii) to get the diagonal entry A_{ii} only. (But the other way "$\Longleftarrow$" doesn't hold)!

12.22 Gram Matrix (Def. 9.2.13)

Let $V \in \mathbb{R}^{m\times n}$, the Gram matrix of V is the inner product of the columns of V, i.e., $G = V^TV$.

Sometimes VV^T is also called a Gram matrix of V (the inner product of the rows) (Remark 9.2.14).

12.23 Gram and Eigenvalues (Prop. 9.2.15)

Let $A \in \mathbb{R}^{m\times n}$, the non-zero eigenvalues of A^TA are the same as the ones of AA^T. Both matrices are symmetric and PSD.
Proof: $x^TA^TAx = (Ax)^T(Ax) = \|Ax\|^2 \geq 0$.

12.24 Cholesky Decomposition (Prop. 9.2.16)

Every symmetric, PSD matrix M is a gram matrix of an *upper triangular matrix* C, i.e., $M = C^TC$.

12.24.1 Calculating the Cholesky Decomposition

(i) Let M be symmetric and PSD (PSD because we later do $\sqrt{\lambda}$), the eigendecomposition (Cor. 9.2.2) gives us $M = V\Lambda V^T$
(ii) We build $\Lambda^{\frac{1}{2}}$ by taking the square root of each entry of Λ, following, $M = \left(V\Lambda^{\frac{1}{2}}\right)\left(V\Lambda^{\frac{1}{2}}\right)^T$ ($\sqrt{\text{neg}}$ wouldn't work).
(iii) We then take the QR decomposition $\left(V\Lambda^{\frac{1}{2}}\right)^T = QR$, following, $M = (QR)^T(QR) = R^TQ^TQR = R^TR = C^TC$. *We set* $\left(V\Lambda^{\frac{1}{2}}\right)^T = QR$, *s.t.* $C = R$ *is upper triangular.*

13 Singular Value Decomp. (Def. 9.3.1)

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

$$A = U\Sigma V^T,$$

where $\Sigma \in \mathbb{R}^{m\times n}$ is a diagonal matrix whose entries $\Sigma_{ii} = \sigma_i$ are non-negative and ordered in descending order. Moreover,

(i) $U^TU = V^TV = I$.
(ii) $\sigma_1 \geq ... \geq \sigma_{\min\{m,n\}} \geq 0$.
(iii) The columns of U are called the *left singular vectors* and are orthonormal $\implies U$ has full column rank.
(iv) The columns of V are called the *right singular vectors* and are orthonormal $\implies V$ has full column rank.
(v) The diagonal entries of Σ are called the *singular values*. The singular values of a matrix A are the square roots of the eigenvalues of the symmetric matrix A^TA or AA^T.

13.1 Existence of SVD (Theo. 9.3.3)

Every matrix $A \in \mathbb{R}^{m\times n}$ has a SVD.
As in, **every linear transformation is diagonal when viewed in the bases of the singular vectors**.

13.2 Compact Form if A has Rank r (Remark 9.3.2)

If A has rank r, the SVD can be written in a compact form:

$$A = U_r\Sigma_rV_r^T,$$

where U_r and V_r contain the first r left/right singular vectors respectively and Σ_r contains the first r singular values (which are strictly positive). V_r full col-rank, V_r^T full row-rank, U_r full col-rank.

13.3 Calculating the SVD (Section 9.3)

Let $A \in \mathbb{R}^{m\times n}$ and $A = U\Sigma V^T$ be its SVD, then

$$AA^T = U(\Sigma\Sigma^T)U^T \text{ and } A^TA = V(\Sigma^T\Sigma)V^T.$$

In other words, the SVD of A can be calculated by

(i) Taking the eigendecomposition of AA^T or A^TA.
(ii) U = eigenvector matrix of AA^T,
(iii) V = eigenvector matrix of A^TA,
(iv) singular values = $\sqrt{\text{eigenvalues of } AA^T \text{ or } A^TA}$. Only if A is real & **symmetric**, singular values of A = |eigenvalues| of A.
(v) Σ = descending ordered square roots of eigenvalues of AA^T or A^TA, they are equal. σ_1 is the largest, σ_n the smallest.
(vi) AA^T, A^TA are PSD, thus eigenvalues $\geq 0 \implies$ sing. values ≥ 0.
(vii) If A is invertible, then $A^{-1} = (U\Sigma V^T)^{-1} = V^{T^{-1}}\Sigma^{-1}U^{-1} = V\Sigma^{-1}U^T$ (as $V^TV = I, U^TU = I$)

13.4 Rank r Matrices (Prop. 9.3.4)

Let $A \in \mathbb{R}^{m\times n}$ with rank r and $\sigma_1, ..., \sigma_r$ be the non-zero singular values of A, $\boldsymbol{u}_1, ..., \boldsymbol{u}_r$ the corresponding left singular vectors and $\boldsymbol{v}_1, ..., \boldsymbol{v}_r$ the corresponding right singular vectors. Then $A = \sum_{k=1}^{r} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^T$.

14 Vector and Matrix Norms

14.1 Vector Norms

For $1 \leq p \leq \infty$, the l_p norm is given by

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}},$$

for $p < \infty$, and $\|\boldsymbol{x}\|_\infty = \max_i |x_i|$.

In particular let $\|\boldsymbol{x}\|_p$ for $p = 1$ denote the *Manhattan distance.*

14.1.1 Relations (Prop. 10.0.1)

For all $\boldsymbol{x} \in \mathbb{R}^n$, $\|\boldsymbol{x}\|_2 \leq \|\boldsymbol{x}\|_1 \leq \sqrt{n}\|\boldsymbol{x}\|_2$

14.2 Matrix Norms (Def. 10.0.2)

Let $A \in \mathbb{R}^{m\times n}$ then consider

$$\text{Frobenius norm} \quad \|A\|_F := \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2}$$

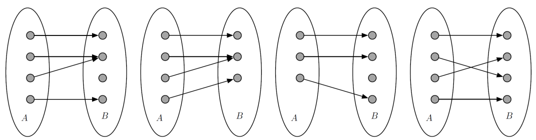
$$\text{operator/spectral norm} \quad \|A\|_{\text{op}} := \max_{\boldsymbol{x}\in\mathbb{R}^n \text{ s.t. } \|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|$$

14.2.1 Properties with Singular Values (Prop. 10.0.3)

Given singular values $\sigma_1 \geq ... \geq \sigma_{\min\{m,n\}}$ of A:

(i) $\|A\|_F^2 = \operatorname{Tr}(A^TA) = \sum_i \sigma_i^2$
(ii) $\|A\|_{\text{op}} = \sigma_1$
(iii) $\|A\|_{\text{op}} \leq \|A\|_F \leq \sqrt{\min\{m,n\}}\|A\|_{\text{op}}$

15 Surjectivity, Injectivity, Bijectivity



weder injektiv noch surjektiv | surjektiv, aber nicht injektiv | injektiv, aber nicht surjektiv | bijektiv