1 Vectors

D 1.4 (Linear Combination):

• let $v, w \in \mathbb{R}^m, \lambda, \mu \in \mathbb{R}$

$\Rightarrow \sum_{i=1}^{n} \lambda_i v_i$ are scaled combinations of n vectors v_i.

D 1.7 (Combination types):

• **Affine Combination:** $\sum_{i=1}^{n} \lambda_i = 1$

Conic Combination: if $\lambda_j \geq 0$ for $j = 1, 2, \ldots, n$

Convex Combination: Affine + Conic

D 1.9 (Scalar/dot product):

• $\mathbf{v} \cdot \mathbf{w} := \sum_{i=1}^{m} v_i w_i$, altetnative notation: $[z_i]_{i=1}^{m} := [v_i + w_i]_{i=1}^{m}$

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: $\|\mathbf{u}\| = 1 = \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{1}{\|\mathbf{v}\|}\mathbf{v}$ (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 $\mathbf{v}, \mathbf{w} \in \mathbb{R}^m$

D 1.14 (Angle between vectors):

• $\cos(\alpha) = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \|\mathbf{w}\|} \in [-1, 1]$

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 1.16 (Triangle inequality):

• $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$

D 1.21 (Linear (in)dependence):

• vectors are linearly dependent if one of them is linear combination of the others: $\mathbf{v}_k = \sum_{\substack{j=1 \\ j \neq k}}^{n} \lambda_j \mathbf{v}_j$

$\Leftrightarrow$ There are scalars $\lambda_1, \lambda_2, \ldots, \lambda_n$ besides $0, 0, \ldots, 0$ such that $\sum_{j=1}^{n} \lambda_j \mathbf{v}_j = \mathbf{0}$. We also say that $\mathbf{0}$ is a nontrivial linear combination of the vectors.

$\Leftrightarrow$ At least one of the vectors is a linear combination of the previous ones.

D 1.25 (Span):

• Span of vectors is a set of all linear combinations of those vectors: $\mathbf{Span}(\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n) := \left\{\sum_{j=1}^{n} \lambda_j \mathbf{v}_j : \lambda_j \in \mathbb{R} \text{ for all } j \in [n]\right\}$

Construction of vectors with standard unit vectors:

• Every target vector can be written as: $\mathbf{u} = \sum_{i=1}^{m} u_i \mathbf{e}_i$, where $\mathbf{e}$ is a standard unit vector.

2 Matrices

D 2.1 (Matrix):

• $A = [a_{ij}]_{i=1,j=1}^{m\ \ n}$ - m rows, n columns (*Zeilen zuerst, Spalten später*)

D 2.2 (Matrix addition, scalat multiplication):

• Addition: $A + B = [a_{ij} + b_{ij}]_{i=1,j=1}^{m\ \ n}$

• Scalar multiplication: $\lambda A = [\lambda a_{ij}]_{i=1,j=1}^{m\ \ n}$

Matrix types:

• **Identity matrix** ($a_{ii} = 1$ for all i): I

• **Diagonal matrix** ($a_{ij} = 0$ for all $i \neq j$): $\text{diag}(d_1, \ldots, d_n)$

• **Upper triangular matrix** ($a_{ij} = 0$ for all $i > j$): U

• **Lower triangular matrix** ($a_{ij} = 0$ for all $i < j$): L

• **Symmetric matrix** ($a_{ij} = a_{ji}$ for all i, j): $A = A^\top$

• **Skew-symmetric matrix** ($a_{ij} = -a_{ji}$ for all i, j): $A = A^\top$

D 2.4 (Matrix-vector product):

• Rows of matrix ($m \times n$) with vector (n elements), i.e. $u_1 = \sum_{i=1}^{m} a_{1,i}\, v_i$, $Ix = x$; **Trace:** Sum of the diagonal entries.

D 2.9 (Column space):

• The column space $\mathbf{C}(A)$ of A is the span (set of all linear combinations) of the columns: $\mathbf{C}(A) := \{A\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\} \subseteq \mathbb{R}^m$

D 2.10 (Rank):

• $\text{rank}(A) :=$ the number of linearly independent column vectors of A.

D 2.11 (Transpose):

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

• $(A^\top)^\top = A$

D 2.13 (Row space):

• $\mathbf{R}(A) := \mathbf{C}(A^\top)$

D 2.17 (Nullspace):

• Nullspace contains all input vectors that lead to output vector $\mathbf{0}$.

$\mathbf{N}(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\} \subseteq \mathbb{R}^n$

D 2.27 (Kernel & Image):

• **Kernel:** $\mathbf{N}(A) = \mathbf{Ker}(T) := \{\mathbf{x} \in \mathbb{R}^n : T(\mathbf{x}) = \mathbf{0}\} \subseteq \mathbb{R}^n$ (If A is the unique $m \times n$ matrix such that $T = T_A$)

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

2.2.2 Working with linear transformations:

• A matrix can be understood as a re-mapping of the unit vectors, scaling and re-orienting them. Each column vector can then be understood as the new unit vector e_i, hence essentially adding another coordinate system to the original one, which is moved and rotated a certain way. The rotation matrix under 2 is such an example. To prove that T is a linear transformation, use $T(x + y) = T(x) + T(y)$ and $T(\lambda x) = \lambda T(x)$. Then insert the linear transformation given by the task and replace x (or whatever variable there is) with $x + y$ or λx. $Ax = \sum_{i=1}^{n} x_i v_i$, where v_i is the i-th column of A.

O 2.39 (Matrix multiplication):

• $A \times B = C$, $c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$. Dimension restrictions: A is $m \times n$, B is $n \times p$, result is $m \times p$. For each entry, multiply the i-th row of A with the j-th column of B.

Not commutative, but associative & distributive.

L 2.40 Matrix multiplication with transposition:

• $(AB)^\top = B^\top A^\top$

D 2.44 Outer product:

• $\text{rank}(A) = 1 \iff \exists$ non-zero vectors $v \in \mathbb{R}^m$, $w \in \mathbb{R}^n$ such that A is an outer product, i.e. $A = vw^\text{T}$, thus $\text{rank}(vw^\text{T}) = 1$.

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), $r = \text{rank}(A)$. **Row Echelon Form:** To find REF, try to create pivots: $R_0 = \begin{bmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. Use Gauss-Jordan elimination to find it (row transformations). **Reduced REF:** RREF is simply REF without any zero rows (i.e. in R_0, R (in RREF) would be R_0 without the last row).

O 2.5.6 (Invertible matrix):

• Matrix A is invertible if it is square and there is B such that: $AB = I \Leftrightarrow BA = I \Leftrightarrow AB = BA = I$

D 2.57 (Inverse matrix and its properties):

• If $AB = I$ for invertible A, then B is its inverse and denoted as A^{-1}.

• $(A^{-1})^{-1} = A$ • $(AB)^{-1} = B^{-1}A^{-1}$ • $(A^\top)^{-1} = (A^{-1})^\top$

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 (the vectors) with two operations $\oplus$ and $\odot$. They are based on algebras called fields and satisfy axioms: *commutativity, associativity, zero vector, negative vector, identity element, compatibility of multiplications of vectors and scalars ($\in \mathbb{R}$), distributivity over $\oplus$ both for vectors and scalars ($\in \mathbb{R}$).*

D 4.8 (Subspace):

• Let V be a vector space. A nonempty subset $U \subseteq V$ is a subspace of V if following axioms are true $\forall \mathbf{v}, \mathbf{w} \in U$ and $\forall \lambda \mathbf{v} \in U$:

• $\mathbf{v} + \mathbf{w} \in U$ • $\lambda \mathbf{v} \in U$.

They guarantee that vector addition and scalar multiplication "doesn't take us outside of a subspace".

L 4.9 (Subspace always has 0):

• Let $U \subseteq V$ be a subspace of a vector space V. Then $\mathbf{0} \in U$ (at least).

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 subspace of $\mathbb{R}^m$.

$\Rightarrow R(A) = C(A^\top)$ is a subspace of $\mathbb{R}^n$.

E 4.13 (The nullspace is a subspace):

• Let $A \in \mathbb{R}^{m \times n}$. Then the nullspace $N(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{v} = 0\}$ 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 $\oplus$ and $\odot$ 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: $\text{Span}(B) = V$.

L 4.19 (Independent columns is a basis):

• Independent columns form basis of column space C(A).

O 4.20 (Non-uniqueness of basis):

• Every set $B = \{\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_n}\} \subseteq \mathbb{R}^m$ of m linearly independent vectors is a basis of $\mathbb{R}^m$.

D 4.21 (Finitely generated vector space):

• There is a finite subset $G \subseteq V$ with $\text{Span}(G) = V$. 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):

• *"exchanging elements between G and F"*

V is finitely generated vector space, $F \subseteq V$ a finite set of lin. independent vectors, and $G \subseteq V$ a finite set of vectors with $\text{Span}(G) = V$, then:

• $|V| \leq |G|$ and • $\exists E \subseteq G$ of size $|G| - |F|$ such that $\text{Span}(F \cup E) = V$.

T 4.24 (All bases have the same size):

• All bases have the same size: $B, B' \in V \Rightarrow |B| = |B'|$.

D 4.25 (Dimension):

• $\dim(V)$ - the dimesions of V. It has a size of arbitrary 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 $x_1, x_2 \in V$ and $\lambda_1, \lambda_2 \in \mathbb{R}$, $T(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 T(x_1) + \lambda_2 T(x_2)$.

L 4.27 (Bijective lin. 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):

• $V \cong W \iff \exists T : V \to W$ linear and bijective.

T 4.29 (Basis writes vectors as a unique lin. combination):

• Let V be a finite-dimensional vector space with basis $B = \{v_1, \ldots, v_m\}$. Then every $v \in V$ can be written uniquely as $v = \sum_{j=1}^m \lambda_j v_j$, for unique scalars $\lambda_1, \ldots, \lambda_m$.

L 4.30 (Less than $\dim(V)$ vectors do not span V):

• If $|G| < \dim V$, then $\text{span}(G) \neq V$.

4.3 Computing the three fundamental subspaces

T 4.31 (Basis of $C(A)$: Pivots columns of RREF):

• R is RREF of A, then all columns at pivots of R form a basis of $C(A)$: $\dim(C(A)) = \text{rank}(A) = r$

T 4.32 (Basis of $R(A)$: Nonzero rows of $\text{RREF}(A)$):

• Nonzero rows of $\text{RREF}(A)$ form a basis of $R(A)$, so, $\dim(R(A)) = r$.

T 4.33 (Row rank equals columns rank):

• $\text{rank}(A) = \text{rank}(A^\top)$

C 4.34 (Rank is at most min of the matrix dimensions):

• A is a $m \times n$ matrix with rank $r \Rightarrow r \leq \min(n, m)$.

L 4.35 (Nullspace isomorphism):

• $R = \text{RREF}(A)$, then $T : N(R) \to \mathbb{R}^{n-r}$ is an isomorphism between $N(R)$ and $\mathbb{R}^{n-r} \Rightarrow \dim(N(R)) = n - r$.

T 4.36 (Basis of $N(A)$: Non-pivot columns of $\text{RREF}(A)$):

• If $\text{rank}(A) = r$, then $\dim(N(A)) = n - r$.

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

D 4.37 (Solution space):

• Solution space of $A\mathbf{x} = \mathbf{b}$:

$\mathbf{Sol}(A, \mathbf{b}) := \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{b}\} \subseteq \mathbb{R}^n$

T 4.38 (Solution space from shifting the nullspace):

• Let $\mathbf{s}$ be some solution of $A\mathbf{x} = \mathbf{b}$, then:

$\mathbf{Sol}(A, \mathbf{b}) := \{\mathbf{s} + \mathbf{x} \in \mathbb{R}^n : \mathbf{x} \in N(A)\}$.

We can also compute $\mathbf{Sol}(A, \mathbf{b})$, although it is not a subspace.

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: $\dim(\mathbf{Sol}(A, \mathbf{b})) = n - r$, and $\dim(\mathbf{Sol}(A, \mathbf{b})) := \dim(N(A))$.

T 4.40 (Systems of rank m are solvable):

• Let $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) = m$, $A\mathbf{x} = \mathbf{b}$ is solvable for all $\mathbf{b} \in \mathbb{R}^m$.

T 4.41 (Systems of rank less than m are typ. unsolvable):

• Systems of rank $r < m$ are typically unsolvable.

D 4.42 (Types of systems):

• Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. The system $A \in \mathbb{R}^{m \times n}$ is called:

• $m = n \Rightarrow$ square (A is a square matrix) $\star$ **typ. solvable**

• $m < n \Rightarrow$ underdetermined (A is a wide matrix) $\star$ **typ. solvable**

• $m > n \Rightarrow$ overdetermined (A is a tall matrix) $\star$ **typ. 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: $v^\top w = \sum_{i=1}^n v_i w_i = 0$. Two subspaces are orthogonal if all v and w are orthogonal.

L 5.1.2 (Orthogonality of bases):

• Let $v_1, ..., v_2$ and $w_1, ..., w_2$ be bases of subspaces W and V. W and V are orthogonal $\Leftrightarrow$ all v_i orthogonal to all w_j

L 5.1.3 (Combinations and interaction of subspaces):

• The set of vectors $\{v_1, ..., v_2, w_1, ..., w_2\}$ are linearly independent.

• The union of bases of two subspaces gives a basis for the new subspace: $V \cup W = V + W = \{\lambda v + \mu w \mid \lambda, \mu \in \mathbb{R},\ v \in V,\ w \in W\}$.

• If V and W are subspaces of $\mathbb{R}^n$, then $V + W$ is a subspace of $\mathbb{R}^n$.

• $V \cap W = \{0\}$ if subspaces are orthogonal.

• $\dim(V) = k$ and $\dim(W) = l$, then $\dim(V + W) = k + l \leq n$.

D 5.1.5 (Orthogonal complement):

• Let V be a subspace of $\mathbb{R}^n$, its **orthogonal complement**: $V^\perp = \{\, w \in \mathbb{R}^n \mid w^T v = 0 \text{ for all } v \in V \,\}$.

T 5.1.6 (Relations between subspaces):

• $N(A) = C(A^\top)^\perp = R(A)^\perp$ and $C(A^\top) = N(A)^\perp$

T 5.1.7 (Vector decomposition by orth. complements):

• $W = V^\perp \Leftrightarrow dim(V) + dim(W) = n \Leftrightarrow$ every $u \in \mathbb{R}^n$ is $u = v + w$, v and w are unique.

L 5.1.10 (Justification of exist. of sol. for normal eq.):

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

5.2 Projections

D 5.2.1 (Projection):

• **Projection** of $b \in \mathbb{R}^m$ on a subspace S (of $\mathbb{R}^m$) is the point in S that is closest to b: $\text{proj}_S(b) = \arg\min_{p \in S} \|b - p\|$.

L 5.2.2 (One-dimensional Projection Formula):

• Projection of b on $S = \{\, \lambda a \mid \lambda \in \mathbb{R} \,\} = C(a)$: $\text{proj}_S(b) = \frac{a a^T}{a^T a} b$.

•"Error vector" ($e = b - p$) is perpendiculat projection: ($e = b - \text{proj}_S(b)) \perp \text{proj}_S(b)$.

L 5.2.3 (General Projection Formula):

• Let S be a subspace in $\mathbb{R}^m$ with a basis $a_1, ..., a_n$ that span S. Let A be the matrix with column vectors $a_1, ..., a_n$.

• The general formula: $\text{proj}_S(b) = A\hat{x}$, where $\hat{x}$ is $A^T A \hat{x} = A^T b$.

L 5.2.4 (Properties of $A^\top A$):

• $A^\top A$ is invertible $\Leftrightarrow$ A has linearly independent columns. $\Rightarrow A^\top A$ is a square matrix, symmetric, invertible.

T 5.2.5 (Projection in terms of projection matrix):

• $\text{proj}_S(b) = Pb$ with projection matrix $P = A(A^{-1}A)A^\top$.

A is matrix given in a task.

6 Applications of Orthogonality and Projections

6.1 Least Squares Approximation

Least Squares:

• Approximate a solution to System of equations: find x for which Ax is as close as possible to b: $\min_{\hat{x} \in \mathbb{R}^n} \|A\hat{x} - b\|^2$

usage:

• find $M = A^\top A$, $b' = A^\top b$, solve $M\hat{x} = b'$

Linear Regression:

• Fitting a parabola

$(t_k, b_k) = \{(0,1),(1,2),(2,5)\},\ b_k \approx \alpha_0 + \alpha_1 t_k + \alpha_2 t_k^2$

$A = \begin{pmatrix}1&0&0\\1&1&1\\1&2&4\end{pmatrix}, b = \begin{pmatrix}1\\2\\5\end{pmatrix}, \hat\alpha = (A^TA)^{-1}A^Tb = \begin{pmatrix}1\\0\\1\end{pmatrix}, \hat b(t) = 1+t^2.$

L 6.1.2:

- A has linearly independent columns $\Leftrightarrow t_i = t_j \forall i \neq j$

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 sol.):

- $A \in \mathbb{R}^{m\times n},\ x, y \in C(A^\top):\ Ax = Ay \Leftrightarrow x = y$

This leads to: $C(A^T) \cap N(A) = \{0\}$

T 6.2.2 (Set of all solution of linear equations):

- Set of all sol. : $\{x \in \mathbb{R}^n | Ax = b\} \neq \emptyset$, then:

$\{x \in \mathbb{R}^n \mid Ax = b\} = x_1 + N(A),\quad x_1 \in R(A)$ is unique s.t. $Ax_1 = b$.

T 6.2.4 (Linear equations with no solution):

- Linear equations has no solution:

$\{x \in \mathbb{R}^n \mid Ax = b\} = \emptyset \iff \{z \in \mathbb{R}^m \mid A^Tz = 0,\ b^Tz = 1\} \neq \emptyset.$

6.3 Orthonormal Bases and Gram Schmidt

D 6.3.1 (Orthonomral vectors):

- $q_i^\top q_j = \delta_{ij} = \begin{cases}0 & i \neq j\\ 1 & i = j\end{cases}$ (orthogonal and have norm 1)

D 6.3.3 (Orthogonal Matrix):

- A square matrix $Q \in \mathbb{R}^{n\times n}$ 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 orthonormal basis for $\mathbb{R}^n$.

- Orthogonal (rotation) matrix example: $R_\theta = \begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$.

P 6.3.6 (Preserving qualities of orthogonal matrices):

- Orthogonal matrices preserve norm and inner product of vectors: $\|Qx\| = \|x\|$ and $(Qx)^\top(Qy) = x^\top y$

P 6.3.7 (Least square solution to $Qx = b$):

- The least square solution to $Qx = b$, where Q is the matrix whose columns are the vectors forming the orthonomal basis of $S \subseteq \mathbb{R}^m$, is given by $\hat x = Q^\top b$ and the projection matrix is given by $QQ^\top$.

D 6.3.8 (Gram-Schmidt algorithm):

- **Gram-Schmidt:** used to construct orthonormal bases.

We have linearly indepenedent vectors $a_1, ..., a_n$ that span a subspace S, then we can construct their orthonormal basis $q_1, ..., q_n$ by:

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

D 6.3.10 (QR-Decomposition):

- $A = QR$, where $R = Q^\top A$, and Q is a matrix with orthonormal columns produced by Gram-Schmidt.

D 6.3.11 (Well-Defined QR Decomposition):

- R - upper-triangluar and invertible matrix $\Rightarrow QQ^\top A = A$, and hence, $A = QR$ is well-defined.

Simplicity of calculation with Q:

- **Projection:** $\text{proj}_{C(A)}(b) = QQ^\top b$, **Least Squares:** $R\hat x = Q^\top b$

This is possible because $C(A) = C(Q)$ and R is triangular - we can use back-substitution with it.

6.4 Pseudoinverses

D 6.4.1 (Left pseudoinverse):

- 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):

- 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 pseudoinverses):

- For $A \in \mathbb{R}^{m\times n}$ with $\text{rank}(A) = r$ and a CR-decomposition $A = CR$, we define $A^\dagger = R^TC^T$. In general, $A^\dagger = R^T(RR^T)^{-1}(C^TC)^{-1}C^T = R^T(C^TCRR^T)^{-1}C^T = R^T(C^TAR^T)^{-1}C^T$.

L 6.4.8 (Unique solution of least sq. with pseudoinverses):

- For any matrix A and vector $b \in \mathcal{C}(A)$, the unique solution of the least squares problem is given by a vector $\hat x \in \mathcal{C}(A^T)$ satisfying $A\hat x = b$. The solution is $\hat x = A^\dagger b$, with $A\hat x = b$, and in the general case $A^\dagger = R^TC^T = R^T(C^TAR^T)^{-1}C^T$.

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 (Pseudoinverses properties):

- Let $A \in \mathbb{R}^{m\times n}$. Then $AA^\dagger A = A$, $A^\dagger AA^\dagger = A^\dagger$, $(A^\dagger)^T = (A^T)^\dagger$.

$AA^\dagger$ is symmetric $\Rightarrow$ projection matrix onto $\mathcal{C}(A)$,

$A^\dagger A$ is symmetric $\Rightarrow$ projection matrix onto $\mathcal{C}(A^T)$.

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

7 The Determinant

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):

- $\det(AB) = \det(A)\det(B)$.

Hence, for an LU-decomposition, $\det(A) = \det(L)\det(U)$.

D 7.2.1 (Permutation sign):

- The sign of a permutation is defined as the number of swaps of rows or columns. $\det(\text{permuted matrix}) = (-1)^k \det(\text{original matrix})$, where k is the number of swaps. Even number of swaps $\Rightarrow +1$, odd number $\Rightarrow -1$. $\text{sgn}(\sigma\circ\gamma) = \text{sgn}(\sigma)\,\text{sgn}(\gamma)$. For all $n \geq 2$, half of the permutations have sign $+1$, half have sign -1.

7.2 General case:

D 7.2.3 (Determinant big formula):

- For a square matrix $A \in \mathbb{R}^{n\times n}$,

$\det(A) = \sum_{\sigma\in\Pi_n}\text{sgn}(\sigma)\prod_{i=1}^n A_{i,\sigma(i)}$. *(Number of permutations: n!)*

- **Determinant Properties:**

1. 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 $A \in \mathbb{R}^{n\times n}$, $\det(A) = \det(A^T)$.
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):

- $1 = \det(I) = \det(Q^TQ) = \det(Q^T)\det(Q) = \det(Q)^2$, so $\det(Q) = \pm 1$. If $\det(Q) = 1$, then Q is a rotation matrix. If $\det(Q) = -1$, then Q is a reflection matrix.

P 7.3.2 (Cofactor determinant calculation):

- **Co-factor method**:

$\det(A) = \sum_{j=1}^n A_{ij}C_{ij}$, where cofactors are $C_{ij} = (-1)^{i+j}\det(A_{ij})$,.

P 7.3.5 (Cramer's Rule):

- **Cramer's Rule**: For $Ax = b$ with $\det(A) \neq 0$, $x_j = \frac{\det(\mathscr{B}_j)}{\det(A)}$, where $\mathscr{B}_j$ is the matrix obtained from A by replacing the j-th column with b.

P 7.3.7 (Linearity of a determinant):

- The determinant is linear in each row (and column). For example, $\det\begin{bmatrix}\alpha_0 a_0^T + \alpha_1 a_1^T\\ a_2^T\end{bmatrix} = \alpha_0 \det\begin{bmatrix}a_0^T\\ a_2^T\end{bmatrix} + \alpha_1 \det\begin{bmatrix}a_1^T\\ a_2^T\end{bmatrix}$.

8 Eigenvalues and Eigenvectors

8.1 Complex Numbers

1. Solve $x^2 + 1 = 0 \Rightarrow x = \sqrt{-1} \Rightarrow \mathbb{C} = \{a + ib : a, b \in \mathbb{R}\}$.
2. $(a + ib) + (x + iy) = (a + x) + i(b + y)$,
3. $(a + ib)(x + iy) = (ax - by) + i(ay + bx)$,
4. $(a + ib)(a - ib) = a^2 + b^2$.
5. $\frac{a+ib}{x+iy} = \frac{(a+ib)(x-iy)}{x^2+y^2} = \frac{ax+by}{x^2+y^2} + i\frac{bx-ay}{x^2+y^2}$.
6. $|z| = \sqrt{a^2 + b^2}$, $z = a + ib$,
7. $\overline{a + ib} = a - ib$.

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:
- $z = re^{i\theta}$, $\quad z \in \mathbb{C}$, $\quad r > 0$ is the modulus of z, $\quad \theta \in [0, 2\pi)$.

T 8.1.2 (Fundamental Theorem of Algebra):
- Any degree n non-constant ($n \geq 1$) polynomial $P(z) = \alpha_n z^n + \alpha_{n-1}z^{n-1} + \cdots + \alpha_1 z + \alpha_0$, $\quad (\alpha_n \neq 0)$ has a zero: there exists $\lambda \in \mathbb{C}$ such that $P(\lambda) = 0$.

$\Rightarrow$ A degree-n polynomial has at most n distinct zeros (roots).

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

Inner product on $\mathbb{C}^n$:
- The inner product on $\mathbb{C}^n$ is given by $\langle v, w \rangle = w^* v$.

Conjugate transpose:
- $A^* = \overline{A}^T$.

8.2 Introduction to Eigenvalues and Eigenvectors

D 8.2.1 (EW/EV pair):
- Given $A \in \mathbb{R}^{n\times n}$, we say $\lambda \in \mathbb{C}$ is an *eigenvalue* of A and $v \in \mathbb{C}^n \setminus \{0\}$ is an *eigenvector* of A associated with λ when $Av = \lambda v$. (λ, v) is an eigenvalue–eigenvector pair. If $\lambda \in \mathbb{R}$, then we have a real eigenvalue–eigenvector pair.

L 8.2.3 (Real EW/EV):
- 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 $v \in \mathbb{R}^n \setminus \{0\}$ is an eigenvector associated with λ if and only if $v \in \mathcal{N}(A - \lambda I)$.

D 8.3.4 (Characteristic Polynomial):
- The characteristic polynomial: $(-1)^n \det(A - zI) = \det(zI - A) = (z-\lambda_1)(z-\lambda_2)\cdots(z-\lambda_n)$. The coefficient of z^n is $(-1)^n$.

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

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

L 8.2.8 (Complex EW exist in conjugate pairs for real A):
- Let $A \in \mathbb{R}^{n\times n}$. If (λ, v) is an eigenvalue–eigenvector pair, then $(\overline{\lambda}, \overline{v})$ is also an eigenvalue–eigenvector pair.

8.3 Properties of Eigenvalues and Eigenvectors

P 8.3.1 (EW modifications based on types of a matrix):
- If (λ, v) is an eigenvalue–eigenvector pair of A, then (λ^k, v) is an eigenvalue–eigenvector pair of A^k for $k \geq 1$.
- If (λ, v) is an eigenvalue–eigenvector pair of A with $\lambda \neq 0$, then $\left(\frac{1}{\lambda}, v\right)$ 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 $v_1, \ldots, v_n$ are linearly independent.

T 8.3.3 (Existence of a basis from EV):
- Let $A \in \mathbb{R}^{n\times n}$ with n distinct real eigenvalues. Then there exists a basis of $\mathbb{R}^n$, $v_1, \ldots, v_n$, made of eigenvectors of A.

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

L 8.3.5 (Transposition equality of EW):
- The eigenvalues of $A \in \mathbb{R}^{n\times n}$ are the same as those of A^T.

L 8.3.6 (Determinant and Trace via EW):
- 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

$\det(A) = \prod_{i=1}^n \lambda_i$, $\mathrm{Tr}(A) = \sum_{i=1}^n \lambda_i$.

L 8.3.7 (Cyclic invariance of the trace):
- For $A, B, C \in \mathbb{R}^{n\times n}$:

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

9 Diagonalizable Matrices, 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 complet set of real eigenvectors (eigenbasis).

Equivalently, $\Lambda = V^{-1}AV$, since V is invertible.

$$\text{Std. coord.} \xrightarrow{V^{-1}} \text{EV. coord.} \xrightarrow{\Lambda} \text{EV. coord.} \xrightarrow{V} \text{Std. coord.}$$

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 EV):
- 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 (Projection and EW/EV):
- Let P be a projection matrix onto a subspace $U \subset \mathbb{R}^n$. 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 an invertible matrix S such that $B = S^{-1}AS$. **P 9.1.8**: Similar matrices have the same eigenvalues.

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

L 9.1.11 (Complete set of real EV):
- A matrix has a complete set of real eigenvectors if and only if 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 orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of 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 $A = V\Lambda V^T$, where $\Lambda \in \mathbb{R}^{n\times n}$ is diagonal with diagonal entries equal to the eigenvalues of A, and $V^TV = I$. This decomposition is called the *eigendecomposition*.

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

P 9.2.6 (Rank-One Spectral Decomposition):
- Let $A \in \mathbb{R}^{n\times n}$ be symmetric, and let $v_1, \ldots, v_n$ be an orthonormal basis of eigenvectors of A (the columns of V), with associated eigenvalues $\lambda_1, \ldots, \lambda_n$. Then $A = \sum_{k=1}^n \lambda_k\, v_k v_k^T$.

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 EV):
- 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 v_1, v_2. Then v_1 and v_2 are orthogonal.

L 9.2.8 (Symmetric matrix has real EW):
- A symmetric matrix $A \in \mathbb{R}^{n\times n}$ has only real eigenvalues: $\lambda \in \mathbb{C} \quad \Rightarrow \lambda \in \mathbb{R}$. Indeed, if $Av = \lambda v$:

$\lambda\|v\|^2 = \overline{\lambda}\, v^*v = (\lambda v)^*v = (Av)^*v = v^*A^*v = v^*Av = v^*\lambda v = \lambda\|v\|^2$.

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

P 9.2.10 (Rayleigh Quotient):
- $A \in \mathbb{R}^{n\times n}$ is symmetric. For $x \in \mathbb{R}^n \setminus \{0\}$, the Rayleigh quotient $R(x) = \frac{x^TAx}{x^Tx}$.

The minimum of $R = R(v_{\min}) = \lambda_{\min}$, and the maximum $R(v_{\max}) = \lambda_{\max}$. Here $\lambda_{\max}/\lambda_{\min}$ are the largest/smallest eigenvalues of A, and $v_{\max}/v_{\min}$ their associated eigenvectors.

D 9.2.11 (PSD and PD matrices):
- $A = A^\top$ • $A \succeq 0$ **(PSD)** $\Leftrightarrow \lambda_i(A) \geq 0$ • $A \succ 0$ **(PD)** $\Leftrightarrow \lambda_i(A) > 0$.

P 9.2.12 (Positivity of the quadratic form):
- Let $A \in \mathbb{R}^{n\times n}$ be symmetric. Then $A \succeq 0 \iff x^\top Ax \geq 0 \quad \forall x \in \mathbb{R}^n$,

and $A \succ 0 \iff x^\top Ax > 0 \quad \forall x \neq 0$.

D 9.2.13 (Gram Matrix):

• Given vectors $v_1, \dots, v_n \in \mathbb{R}^m$, their *Gram matrix* is $G \in \mathbb{R}^{n\times n}$ defined by $G_{ij} = v_i^\top v_j$. If $V = [v_1 \cdots v_n] \in \mathbb{R}^{m\times n}$, then $G = V^\top V$.

• If $A = [a_1 \cdots a_n] \in \mathbb{R}^{m\times n}$, one also calls $AA^\top$ a Gram matrix; note that $AA^\top = \sum_{i=1}^n a_i a_i^\top$. It is $m \times m$ matrix.

P 9.2.15 (Same EV of transposed matrices):

• For 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 upper-triangular matrix C: $M = C^\top C$.

9.3 Singular Value Decomposition

D 9.3.1 (Singular Value Decomposition):

• Let $A \in \mathbb{R}^{m\times n}$. There exist orthogonal matrices $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ and a diagonal matrix $\Sigma \in \mathbb{R}^{m\times n}$ with nonnegative diagonal entries $\sigma_1 \geq \cdots \geq \sigma_{\min(m,n)}$ such that

$$A = U\Sigma V^\top.$$

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.

R 9.3.2 (Compact form of SVD):

• If $\operatorname{rank}(A) = r$, 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 $\Sigma_r = \operatorname{diag}(\sigma_1, \dots, \sigma_r)$. 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 SVD):

• Every matrix $A \in \mathbb{R}^{m\times n}$ has SVD: $A = U\Sigma V^\top$. Equivalently, every linear transformation is diagonal in orthonormal bases of singular vectors.

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 $\sigma_1, \dots, \sigma_r$ and corresponding singular vectors $u_1, \dots, u_r$ and $v_1, \dots, v_r$. Then

$$A = \sum_{k=1}^r \sigma_k \, u_k v_k^\top.$$

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

Algorithms

Gaussian Elimination.

Given $Ax = b$, form the augmented matrix $[A \,|\, b]$ and apply elementary row operations to reach row echelon form (REF): pivot $\to$ swap $\to$ eliminate below $\to$ repeat. If a row $(0 \cdots 0 \,|\, c)$ with $c \neq 0$ appears, the system is inconsistent; otherwise solve by back-substitution.

$$\left[\begin{array}{cc|c} 1 & 1 & 3 \\ 2 & 1 & 4 \end{array}\right] \xrightarrow{R_2 - 2R_1} \left[\begin{array}{cc|c} 1 & 1 & 3 \\ 0 & -1 & -2 \end{array}\right] \Rightarrow (x, y) = (1, 2).$$

Gauss-Jordan Elimination.

Starting from $[A \,|\, b]$, apply Gaussian elimination, then normalize each pivot to 1 and eliminate all other entries in the pivot columns. The resulting reduced row echelon form (RREF) gives the solution directly. Solve:

$$\begin{cases} x + y = 3, \\ 2x + y = 4. \end{cases} \iff \left[\begin{array}{cc|c} 1 & 1 & 3 \\ 2 & 1 & 4 \end{array}\right]$$

Row-reduce to RREF:

$$\left[\begin{array}{cc|c} 1 & 1 & 3 \\ 2 & 1 & 4 \end{array}\right] \xrightarrow{R_2 \leftarrow R_2 - 2R_1} \left[\begin{array}{cc|c} 1 & 1 & 3 \\ 0 & -1 & -2 \end{array}\right]$$

$$\xrightarrow{R_2 \leftarrow -R_2} \left[\begin{array}{cc|c} 1 & 1 & 3 \\ 0 & 1 & 2 \end{array}\right] \xrightarrow{R_1 \leftarrow R_1 - R_2} \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \end{array}\right].$$

$$x = 1, \qquad y = 2.$$

Inverse via Gauss–Jordan.

To compute A^{-1}, form the augmented matrix $[A \,|\, I]$ and apply Gauss–Jordan elimination. If

$$[A \,|\, I] \longrightarrow [I \,|\, B],$$

then $B = A^{-1}$. If I cannot be obtained on the left, A is not invertible.

$$[A \,|\, I]1 = \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array}\right] \xrightarrow{R_2 \leftarrow R_2 - 3R_1} \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & -2 & -3 & 1 \end{array}\right]$$

$$\xrightarrow{R_2 \leftarrow -\frac{1}{2}R_2} \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 1 & \frac{3}{2} & -\frac{1}{2} \end{array}\right] \xrightarrow{R_1 \leftarrow R_1 - 2R_2} \left[\begin{array}{cc|cc} 1 & 0 & -2 & 1 \\ 0 & 1 & \frac{3}{2} & -\frac{1}{2} \end{array}\right].$$

Fitting a line with least squares.

$$\hat{\alpha} = \arg\min_{\alpha \in \mathbb{R}^2} \|A\alpha - b\|^2 = (A^T A)^{-1} A^T b, \quad A = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{pmatrix}$$

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}, b = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \hat{\alpha} = (A^T A)^{-1} A^T b = \begin{pmatrix} \frac{7}{6} \\ \frac{1}{2} \end{pmatrix}, \hat{b}(t) = \frac{7}{6} + \frac{1}{2}t.$$

Forming orthonormal basis via Gram-Schmidt.

Gram-Schmidt used to construct orthonormal bases.
We have linearly indepenedent vectors $a_1, \dots, a_n$ that span a subspace S, then we can construct their orthonormal basis $q_1, \dots, q_n$ by:

• $q_1 = \frac{a_1}{\|a_1\|}$.

• For $k = 2, \dots, n$ do $q_k' = a_k - \sum_{i=1}^{k-1} (a_k^\top q_i)\, q_i$,

• normalise $q_k = \frac{q_k'}{\|q_k'\|}$.

Notes from correction

Notes and ideas

Pythagorean theorem:

• If two vectors are orthogonal, then their squared lengths add:
$\|\mathbf{x}+\mathbf{y}\|^2 = \mathbf{x}^\top\mathbf{x} + 2\underbrace{\mathbf{x}^\top\mathbf{y}}_{=0} + \mathbf{y}^\top\mathbf{y} = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$

Linearity Proof:

• To prove linearity insert arbitrary $\mathbf{v}, \mathbf{w} \in \mathbb{R}$ and $\alpha \in \mathbb{R}$ to $f(\mathbf{v} + \alpha\mathbf{w})$

Non-trivial nullspace and solutions:

• When $m < n$ then there exists a nontrivial null space, which prevents uniqueness of a solution. Let $A \in \mathbb{R}^{m\times n}$, then there exists $\mathbf{x} \neq 0$ but $A\mathbf{x} = 0$:
$\text{rank}(A) \leq m < n \Rightarrow dim(N(A)) = n - \text{rank}(A \geq 1)$

Linear dependence:

• The vectors v_1, v_2, v_3 are *linearly dependent* if there exist scalars a_1, a_2, a_3, not all zero, such that $a_1v_1 + a_2v_2 + a_3v_3 = \mathbf{0}$.

Invertible matrix properties:

• A is invertible
$\Leftrightarrow$ $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$
$\Leftrightarrow$ $N(A) = \{0\}$
$\Leftrightarrow$ $\text{rank}(A) = n$
$\Leftrightarrow$ columns of A are lin. independent

Homogenous system solutions:

• A homogenous system $A\mathbf{x} = 0$ has either one solution $x = 0$ or infinitely many solutions: all scaled $\alpha\mathbf{x}$.

Square a sum:

• $(\sum_{i=1}^n \lambda_i)^2 = (\sum_{i=1}^n \lambda_i)(\sum_{j=1}^n \lambda_j)$
$= \sum_{i=1}^n \sum_{j=1}^n \lambda_i\lambda_j = \sum_{i=1}^n \lambda_i^2 + \sum_{i\neq j} \lambda_i\lambda_j$
That is why $\text{Tr}(A^2) \leq \text{Tr}(A)^2$. When all eigenvalues are zero or at most one is not zero, then this becomes equal.

Dot product concept:

• Dot product tells you how much two vectors point in the same direction.

- $u^T w < 0$: pointing opposite to w
- $u^T w = 0$: perpendicular to w
- $u^T w > 0$: pointing partly in the same direction as w

Prove the dimension of a subspace:

• Write the general matrix, apply the constraints, deconstruct into linear combination of independent variables, prove that they span the space, show that they are linearly independent.

Singular Values inversion:

• Since $\sigma_1 \geq \cdots \geq \sigma_n > 0$, it follows that $\frac{1}{\sigma_1} \leq \cdots \leq \frac{1}{\sigma_n}$, and therefore, when ordered decreasingly, the singular values of A^{-1} are $\frac{1}{\sigma_n}, \ldots, \frac{1}{\sigma_1}$.

Invertible matrix singular values:

• Invertible matrix singular values are strictly positive.

Creative Gauchy-Schwarz Inequality:

• To prove $\sum_{i=1}^n \frac{a_i^2}{b_i} \geq \frac{(\sum_{i=1}^n a_i)^2}{\sum_{i=1}^n b_i}$, apply the Cauchy-Schwarz inequality to the vectors $\left(\frac{a_1}{\sqrt{b_1}}, \ldots, \frac{a_n}{\sqrt{b_n}}\right)$ and $(\sqrt{b_1}, \ldots, \sqrt{b_n})$. Their dot product is $\sum_{i=1}^n a_i$, and their squared norms are $\sum_{i=1}^n \frac{a_i^2}{b_i}$ and $\sum_{i=1}^n b_i$. By Cauchy-Schwarz,
$\left(\sum_{i=1}^n a_i\right)^2 \leq \left(\sum_{i=1}^n \frac{a_i^2}{b_i}\right)\left(\sum_{i=1}^n b_i\right),$
which rearranges by division with a sum of b to the desired inequality.