Review material from 21-241

Department of Mathematical Sciences, Carnegie Mellon University
E-mail address: jkoganem@andrew.cmu.edu.
Date: January 12, 2025.

Part 1. Notation

- $\mathbb{R}^n$: n-dimensional real vector space, consisting of all n-tuples of real numbers. Analogously, $\mathbb{C}^n$ is the $2n$-dimensional complex vector space, consisting of all n-tuples of complex numbers.
- $\mathcal{M}_{m\times n}(\mathbb{F})$: set of $m \times n$ matrices with entries in the field $\mathbb{F}$. We usually only consider the case when $\mathbb{F} = \mathbb{R}$ (the field of real numbers) or $\mathbb{F} = \mathbb{C}$ (the field of complex numbers).
- $\boldsymbol{x}$: we use boldface notation for an element of $\mathbb{R}^n$ for $n \geq 2$. By convention, we think of these as column vectors with respect to the standard basis.
- $\hat{\boldsymbol{x}}$: notation for least squares solution, the vector that minimizes the least squares error $\|A\boldsymbol{x} - \boldsymbol{b}\|^2$.
- $[\boldsymbol{x}]_\alpha$: coordinate vector of $\boldsymbol{x}$ with respect to the ordered basis α. It represents $\boldsymbol{x}$ as a linear combination of the basis vectors in α.
- $\mathrm{Col}(A)$: column space of the matrix A. It is the set of all linear combinations of the columns of A.
- $\mathrm{Null}(A)$: nullspace of the matrix A. It is the set of all vectors $\boldsymbol{x}$ such that $A\boldsymbol{x} = \boldsymbol{0}$.
- dim: dimension of a vector space, which is the number of vectors in a basis for the space.
- $I_{n\times n}$: $n \times n$ identity matrix.
- $[T]_\alpha^\beta$: if $T : V \to W$ is a linear transformation, then $[T]_\alpha^\beta$ denotes the matrix representation of T with respect to the ordered bases α and β. It transforms coordinates from basis α in V to basis β in W.
- proj_W: projection operator onto a given subspace $W \subseteq V$. For example, $\mathrm{proj}_{\mathrm{Col}(A)}\, \boldsymbol{b}$ is the projection of $\boldsymbol{b}$ onto the column space of A.
- $\mathbb{P}_n[x]$: space of polynomials of degree at most n in the variable x.
- $C(\mathbb{R})$: vector space of continuous functions (with codomain $\mathbb{R}$) over the real line. In general, $C(X;Y)$ denotes the set of continuous functions from X to Y (where X and Y are topological spaces so that the notion of continuity makes sense). If $Y = \mathbb{R}$, it is standard to omit Y and simply write $C(X)$.
- $U \oplus W$: (internal) direct sum of vector subspaces $U, W \subseteq V$, which combines two subspaces such that each element in the sum can be uniquely written as a sum of elements from each subspace.
- A^T: transpose of the matrix A.
- $U^\perp$: orthogonal complement of a subspace $U \subseteq \mathbb{R}^n$, the vector subspace of all vectors that are orthogonal to every vector in the U.

Part 2. Concepts to review

Robust summary:

a) Classical solutions to $A\boldsymbol{x} = \boldsymbol{b}$.
   - Existence is guaranteed if A has a pivot in every row/admits full row rank.
   - Uniqueness is guaranteed if A has a pivot in every column/admits full column rank.
   - The number of pivots plus the number of free variables equals the number of columns.
   - The number of pivots equals the dimension of the column space.
   - The number of free variables equals the dimension of the nullspace.
   - Rank-nullity: $\dim \mathrm{Col}(A) + \dim \mathrm{Null}(A) =$ number of columns in A.

b) Least squares solutions to $A\boldsymbol{x} = \boldsymbol{b}$.
   - When $\boldsymbol{b} \notin \mathrm{Col}(A)$, there are no classical solutions to $A\boldsymbol{x} = \boldsymbol{b}$.
   - Least squares solutions satisfy $A\hat{\boldsymbol{x}} = \mathrm{proj}_{\mathrm{Col}(A)}\, \boldsymbol{b}$ or $A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$.
   - Since $\mathrm{Null}(A) = \mathrm{Null}(A^T A)$, when A has independent columns there is a unique least squares solution.
   - Otherwise, there are infinitely many least squares solutions.
   - The least square solution minimizes the least squares error $E : \mathbb{R}^n \to \mathbb{R}$ defined via $E(\boldsymbol{x}) = \|\boldsymbol{b} - A\boldsymbol{x}\|^2$.

c) Spectral theory: finding $\lambda \in \mathbb{R}, \boldsymbol{v} \neq \boldsymbol{0}$ satisfying $A\boldsymbol{v} = \lambda\boldsymbol{v}$.
   - Eigenvectors are non-zero vectors satisfying $A\boldsymbol{v} = \lambda\boldsymbol{v}$.
   - A matrix A is diagonalizable if $A = X\Lambda X^{-1}$ for an invertible matrix X and a diagonal matrix Λ.
   - A matrix A is diagonalizable if and only if the geometric multiplicities of its eigenvalues are all equal their algebraic multiplicities.
   - Eigenvectors corresponding to distinct eigenvalues are linearly independent.
   - Spectral theorem (important!!): if S is a real symmetric matrix, then its eigenvalues are real, its eigenvectors corresponding to district eigenvalues are orthogonal, and one can identify a set of orthonormal eigenvectors forming an orthogonal matrix Q such that $S = Q\Lambda Q^T$. As a result, S can be decomposed as a sum of rank 1 matrices: $S = \lambda_1 \boldsymbol{q}_1\boldsymbol{q}_1^T + \ldots + \lambda_n \boldsymbol{q}_n\boldsymbol{q}_n^T$.

d) Singular value decomposition: finding orthogonal matrices U, V and a rectangular diagonal matrix Σ such that $AV = U\Sigma$.

- A^TA and AA^T are positive semi-definite matrices that share the same positive eigenvalues.
- We can use their eigendecompositions to generalize spectral decomposition for rectangular matrice.
- The singular values of A are square roots of the eigenvalues of A^TA and AA^T.
- The left singular vectors are orthonormal eigenvectors of AA^T.
- The right singular vectors are orthonormal eigenvectors of A^TA.
- The singular vectors corresponding to the positive singular values are related via $A\boldsymbol{v}_i = \sigma_i$ and $A^T\boldsymbol{u}_i = \sigma_i\boldsymbol{v}_i$.
- From an SVD of a matrix A with rank r, we can write A as a sum of rank 1 matrices: $A = \sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^T + \ldots + \sigma_r\boldsymbol{u}_r\boldsymbol{v}_r^T$.

e) Linear transformations and their matrix representations: maps between abstract vector spaces.
- A map $T : V \to W$ is said to be a linear transformation (they're also referred to as linear maps or linear operators) between two vector spaces V, W if $T(x + cy) = T(x) + cT(y)$ for any $x, y \in V, c \in \mathbb{R}$.
- Using the notion of coordinates, we may define the matrix representation of a linear transformation $T : V \to W$ with respect to two ordered basis $\alpha = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ and $\beta = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m\}$ for V, W respectively, as the matrix $[T]_\alpha^\beta \in \mathcal{M}_{m\times n}(\mathbb{R})$ defined via $[T]_\alpha^\beta = \big([T(\boldsymbol{v}_1)]_\beta \bigm| \ldots \bigm| [T(\boldsymbol{v}_n)]_\beta\big)$.
- The coordinates of $T(\boldsymbol{v})$ for any $\boldsymbol{v} \in V$ can be computed via $[T(\boldsymbol{v})]_\beta = [T]_\alpha^\beta[\boldsymbol{v}]_\alpha$.
- Change of basis matrix: if β, β' are two bases for a vector space V, then the change of basis matrix $Q_{\beta\to\beta'}$ is the matrix representation of the identity map $\mathrm{id} : V \to V$ with respect to β, β', i.e. $Q_{\beta\to\beta'} = [I]_\beta^{\beta'}$. By construction, $Q_{\beta\to\beta'}[\boldsymbol{v}]_\beta = [\boldsymbol{v}]_{\beta'}$ for any $\boldsymbol{v} \in V$.

Important matrix decompositions:

a) LU-decomposition: follows from row reduction.
b) QR-decomposition: follows from Gram-Schmidt.
c) Eigen-decomposition: follows from finding eigenvalues and linearly independent eigenvectors.
d) Singular-value decomposition: follows from finding eigenvalues and orthonormal eigenvectors of A^TA and AA^T.

Detailed summary.

a) $A\boldsymbol{x} = \boldsymbol{b}$ and the column picture.
- Any linear system can be viewed as the vector equation
$$x_1\boldsymbol{a}_1 + \ldots + x_n\boldsymbol{a}_n = \boldsymbol{b}, \tag{1.1}$$
or the matrix-vector equation
$$\begin{pmatrix}\boldsymbol{a}_1 & \ldots & \boldsymbol{a}_n\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix} = \boldsymbol{b}. \tag{1.2}$$
The solution set can be found by row reducing the augmented matrix
$$\begin{pmatrix}\boldsymbol{a}_1 & \ldots & \boldsymbol{a}_n & \big| & \boldsymbol{b}\end{pmatrix} \tag{1.3}$$
- A linear system with m equations and n variables can be written as $A\boldsymbol{x} = \boldsymbol{b}$, where $A \in \mathcal{M}_{m\times n}(\mathbb{R}), \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{b} \in \mathbb{R}^m$.
- The augmented matrix associated to the system $A\boldsymbol{x} = \boldsymbol{b}$ is the matrix $\big(A \bigm| \boldsymbol{b}\big) \in \mathcal{M}_{m\times(n+1)}(\mathbb{R})$.
- To simplify the analysis of the linear system, we often row reduce $\big(A \bigm| \boldsymbol{b}\big) \sim \big(\tilde{A} \bigm| \boldsymbol{c}\big)$ where $\tilde{A}$ is the row reduced echelon form of A.
- The number of leading entries in $\tilde{A}$ is the number of pivots in A.
- If the number of leading entries is r, then $n - r$ is the number of free variables.
- Equivalently, the number of pivots plus the number of free variables is equal to the number of columns.
- This implies that the number of pivots in a matrix is less than or equal to n.
- $\tilde{A}$ can have at most one pivot in each row, so the number of pivots is also less than or equal to m.
- Therefore the number of pivots in a matrix is less than or equal to the minimum of m and n.
- A linear system is said to be consistent if it admits at least one solution.
- A linear system is said to be inconsistent if it admits no solutions.

- A linear system is inconsistent if and only if a row of the form $\begin{pmatrix} 0 & \cdots & 0 \mid c \end{pmatrix}$ with $c \neq 0$ appears in the reduced augmented matrix $\left(\tilde{A} \mid \boldsymbol{c}\right)$.
- If a linear system is consistent, it admits a unique solution if and only if it does not admit any free variables.
- Linear combination: if $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$ are p vectors and $c_1, \ldots, c_p$ are p constants, then the vector

$$\boldsymbol{w} = c_1 \boldsymbol{v}_1 + \ldots + c_p \boldsymbol{v}_p \tag{1.4}$$

is called a *linear combination* of the vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$.
- The column picture: $A\boldsymbol{x}$ is a linear combination of the columns of A. $A\boldsymbol{x} = \boldsymbol{b}$ is consistent if and only if $\boldsymbol{b}$ is a linear combination of the columns of A.
- The set of all vectors that are linear combinations of the columns of A is the column space of A.
- If a system is consistent, the general solution is $\boldsymbol{x} = \boldsymbol{x}_p + \boldsymbol{x}_h$, where $\boldsymbol{x}_p$ is any particular solution (non-parametrized) to the inhomogeneous problem and $\boldsymbol{x}_h$ is the general solution (parametrized, unless the only homogeneous solution is the zero solution) to the homogeneous problem.
- The system $A\boldsymbol{x} = \boldsymbol{0}$ is always consistent since $\boldsymbol{0}$ is a trivial solution.
- The number of free parameters in the general solution to the homogeneous problem is the number of free variables.
- The system $A\boldsymbol{x} = \boldsymbol{b}$ is consistent if $\boldsymbol{b} \in \mathrm{Col}(A)$, the column space of A.
- $\mathrm{Col}(A)$ is a vector subspace of $\mathbb{R}^m$.
- The set of solutions to the homogeneous equation $A\boldsymbol{x} = \boldsymbol{0}$ is the nullspace of A.
- If $\boldsymbol{b} \in \mathrm{Col}(A)$, then the system $A\boldsymbol{x} = \boldsymbol{b}$ admits a unique solution if $\mathrm{Null}(A) = \{\boldsymbol{0}\}$, where $\mathrm{Null}(A)$ is the nullspace of A.
- $\mathrm{Null}(A)$ is a vector subspace of $\mathbb{R}^n$.
- $\mathrm{Col}(A) = \mathbb{R}^m$ if and only if A has a pivot in each row. This means that the system $A\boldsymbol{x} = \boldsymbol{b}$ is consistent for every $\boldsymbol{b} \in \mathbb{R}^m$ if and only if there is a pivot in each row of A. In other words, the number of pivots is equal to m. If $m > n$, this cannot happen; when there are more equations than variables, the problem is overspecified so you cannot expect solvability for all $\boldsymbol{b}$.
- $\mathrm{Null}(A) = \{\boldsymbol{0}\}$ if and only if the number of pivots equal to the number of columns. If $m < n$, this cannot happen; when there are less equation than variables, the problem is underspecified and thus free variables are expected to exist.
- $A\boldsymbol{x} = \boldsymbol{b}$ admits a unique solution for all $\boldsymbol{b} \in \mathbb{R}^m$ if and only if $m = n$ and A has exactly m pivots. This is equivalent to requiring A to be invertible.

b) Matrix-vector and matrix-matrix multiplication
- We defined the matrix-product $A\boldsymbol{x}$ so that it is consistent with how we record linear systems.
- This is only well-defined if $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ and $\boldsymbol{x} \in \mathbb{R}^n$. In other words, the number of columns in A must match the number of rows in $\boldsymbol{x}$. The output is a vector in $\mathbb{R}^m$.
- It is defined so that each row of A specifies how to combine the entries in x (row picture).
- It is also defined so that $\boldsymbol{x}$ specifies how to combine the columns of A (column picture).
- The matrix-matrix product AB between two matrices $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ and $B \in \mathcal{M}_{n\times p}(\mathbb{R})$ is defined so that $(AB)\boldsymbol{x} = A(B\boldsymbol{x})$ for any vector $\boldsymbol{x} \in \mathbb{R}^p$.
- It is defined so that each row of A specifies how to combine the rows of B (row picture).
- It is also defined so that each column of B specifies how to combine the columns of A (column picture).
- In general, $AB \neq BA$.
- In general, $AB = AC \;\not\Longrightarrow\; B = C$ and $A\boldsymbol{x} = A\boldsymbol{y} \;\not\Longrightarrow\; \boldsymbol{x} = \boldsymbol{y}$. However, the implication is true if you assume $AB = AC$ for every A or $A\boldsymbol{x} = A\boldsymbol{y}$ for every A.
- In general, $BA = CA \;\not\Longrightarrow\; B = C$ and $A\boldsymbol{x} = B\boldsymbol{x} \;\not\Longrightarrow\; A = B$. However, the implication is true if you assume $BA = CA$ for every A or $A\boldsymbol{x} = B\boldsymbol{x}$ for every $\boldsymbol{x}$.
- Computing AB by focusing on its rows or columns is faster than computing it entry-wise for large matrices.
- Matrix multiplication can be viewed as the left matrix transforming what it is acting on.
- Permutation matrices act on vectors/matrices by swapping rows.
- Diagonal matrices act of vectors/matrices by scaling rows.
- Elementary matrices act on vectors/matrices by adding a scalar copy of one row to another row.

c) Gaussian elimination and LU-decomposition
- Gaussian elimination is an algorithm for simplifying a linear system via elementary row operations.
- Elementary row operations do not change the solution set of a linear system.

- Elementary row operations can be reversed.
- Two systems that can be obtained via row operations are said to be row equivalent.
- Gaussian elimination can be viewed in terms of multiplication on the left by elementary, permutation, and diagonal matrices.
- To perform forward Gaussian elimination, one starts from the left most non-zero column, and with swapping if necessary, uses the upper left most non-zero entry to eliminate the entries below it. One proceed to perform this until one hits the last row.
- Upon completing forward Gaussian elimination, one arrives at an echelon form of the system. Echelon forms are not unique.
- To perform backwards Gaussian elimination, one starts backwards from the last leading entry, scales it to be 1 and then uses it to eliminate the entries above it.
- Upon completing backwards Gaussian elimiantion, one arrives at the row reduced echelon form of the system. The row reduced echelon form is unique.
- If no permutations were performed to reduce A to an echelon form U, then the elimination matrix E that makes $EA = U$ is the product of elementary and diagonal matrices. It is also the product of invertible lower triangular matrices.
- Thus $A = E^{-1}U$ and E^{-1} is a lower trinagular matrix. This is the LU-decomposition of A.
- LU-decompositions can be generalized for rectangular matrices, the only difference is that U is also rectangular and thus the definition of upper triangular needs to be generalized for rectangular matrices.
- LU-decomposition is useful when one wants to solve $A\boldsymbol{x} = \boldsymbol{b}$ for multiple $\boldsymbol{b}$'s. One first solves $L\boldsymbol{y} = \boldsymbol{b}$ and then solves $U\boldsymbol{x} = \boldsymbol{y}$.

d) Inverse of a matrix
- The inverse of a square matrix A is a matrix that undoes the action of A.
- The left inverse of a matrix is automatically the right inverse of a matrix and vice versa.
- Elementary, permutation and diagonal matrices are invertible and their inverses are easy to write down.
- A square matrix $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ is invertible if and only if it is the product of elementary, permutation and diagonal matrices.
- A square matrix $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ is invertible if and only if it is row equivalent to the identity matrix $I_{n\times n}$.
- A square matrix $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ is invertible if and only if its number of pivots is equal to n.
- A square matrix $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ is invertible if and only if $\mathrm{Null}(A) = \{\mathbf{0}\}$.
- A square matrix $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ is invertible if and only if $\mathrm{Col}(A) = \mathbb{R}^n$.
- A square matrix $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ is invertible if and only if $A\boldsymbol{x} = \boldsymbol{b}$ is consistent and admits a unique solution for all $\boldsymbol{b} \in \mathbb{R}^n$. The unique solution in this case is $\boldsymbol{x} = A^{-1}\boldsymbol{b}$.
- A^{-1} can be computed via row reducing the augmented matrix $\big(A \mid I\big)$ (Gauss-Jordan).
- If A, B are two invertible matrices (necessarily square), then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.
- There is a simple formula for the inverse of 2×2 matrices.
- A 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if $ad - bc \neq 0$.

e) Vector spaces and vector subspaces
- Vector spaces is a set V together with operations of vector addition and scalar multiplication that makes it mimic the structure of $\mathbb{R}^n$.
- In most standard examples, the notion of vector addition and scalar multiplication is inherited from the real numbers.
- Canonical examples of vector spaces: $\mathbb{R}^n, \mathcal{M}_{m\times n}(\mathbb{R}), \mathbb{P}_n[x], C(\mathbb{R})$.
- Vectors in general simply refer to elements in a vector space. For example, a vector in $\mathcal{M}_{m\times n}(\mathbb{R})$ is an $m \times n$ real matrix. A vector in $\mathbb{P}_n[x]$ is a polynomial of degree at most n. A vector in $C(\mathbb{R})$ is a continuous function over $\mathbb{R}$.
- A subset W of a vector space V is a vector subspace if W is also a vector space with the operations of vector addition and scalar multiplication inherited from V.
- Criteria for a subset of V to be a subspace: it contains the zero vector 0_V in V, it is closed under vector addition, it is closed under scalar multiplication.
- Canonical examples of vector subspaces: if $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, $\mathrm{Col}(A)$ is a subspace of $\mathbb{R}^m$ and $\mathrm{Null}(A)$ is a subspace of $\mathbb{R}^n$.
- The span of vectors in V is always a vector subspace of V.

f) Linear independence, basis, dimension, rank
- A set of vectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ in a vector space V is said to be linearly independent if
$$c_1\boldsymbol{v}_1 + \ldots + c_k\boldsymbol{v}_k = 0_V \text{ implies } c_1 = \ldots = c_k = 0. \tag{1.5}$$
This is the definition of linear independence in a generic vector space V.
- Any set containing the zero element $\mathbf{0}_V$ is automatically linearly dependent.
- Two vectors are linearly dependent if and only if they are scalar multiples of each other. For more than two vectors, they can be linearly dependent in more complicated ways.
- If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ are vectors in $\mathbb{R}^n$, then they are linearly independent if and only if the matrix
$$A = \left(\boldsymbol{v}_1 \;\middle|\; \ldots \;\middle|\; \boldsymbol{v}_k\right) \tag{1.6}$$
has a trivial nullspace, i.e. $\mathrm{Null}(A) = \{\mathbf{0}\}$. In other words, the only solution to $A\boldsymbol{x} = \mathbf{0}$ is the zero solution $\boldsymbol{x} = \mathbf{0}$. This characterization only works for vectors in $\mathbb{R}^n$.
- If V is a vector space and $\mathcal{B}$ be a collection of vectors in V, then we say $\mathcal{B}$ is a basis for V if $V = \operatorname{span}\mathcal{B}$ and $\mathcal{B}$ is a linearly independent set.
- (Important) The basis theorem: if $\dim V = n$ and $\mathcal{B}$ is a set of n vectors, then $\mathcal{B}$ is a basis for V if $\mathcal{B}$ is a linearly independent set or if $\operatorname{span}\mathcal{B} = V$. The two notions are equivalent (but only when the number of vectors is equal to the dimension of the space).
- Any basis for a vector space V has the same number of elements in them. (For infinite-dimensional vector spaces, this needs to be phrased in terms of the cardinality of a set, but in this class we work exclusively with finite-dimensional vector spaces).
- We use this number (an invariant) to define the dimension of a vector space V. The dimension is defined the be the number of elements in any basis. If this number is finite, we refer to V as a finite-dimensional vector space.
- (Not too important) It's possible for $\dim V = \infty$ (technically, we also need to specify what is the correct notion of a basis, since we need to clarify if we are allowed to take infinite sums or only finite sums when we consider linear combinations), and we refer to these vector spaces as infinite-dimensional vector spaces. The canonical example of an infinite dimensional vector space is $C(\mathbb{R})$, the vector space of continuous functions over the real line. In this class we work exclusively with vector spaces with $\dim V < \infty$.
- The only subspace with dimension 0 is the zero subspace $\{\mathbf{0}_V\}$.
- (Important) If W is a subspace of V, then $\dim W \leq \dim V$.
- If V is a finite dimensional vector space and W is a subspace of V with $\dim W = \dim V$, then $W = V$.
- If V is a finite dimensional vector space and $\dim W < \dim V$, then W must be a proper subspace of V (i.e. $W \subsetneq V$).
- If you have m vectors in a finite-dimensional vector space V with $\dim V = n$, if $m > n$ then the vectors must be linearly dependent, if $m < n$ then they cannot span V (think about what happens in $\mathbb{R}^3$ with 2 or 4 vectors).
- The column rank of A is the number of independent columns in A.
- The row rank of A is the number of independent rows in A.
- The row rank is equal to the column rank, so we simply refer to them as the rank of A.
- The row rank and column rank of a matrix A does not change under elementary row operations.
- This means that if A is row equivalent to B, then $\operatorname{rank} A = \operatorname{rank} B$.
- If $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ and $B \in \mathcal{M}_{n\times p}(\mathbb{R})$, then $\mathrm{Col}(AB)$ is a subspace of $\mathrm{Col}(A)$ inheriting the vector operations from $\mathbb{R}^m$.
- We also have that $\mathrm{Col}(B^TA^T)$ is a subspace of $\mathrm{Col}(B^T)$ as a subspace of $\mathbb{R}^p$ (think about the homework result on row spaces).
- As a result, $\operatorname{rank}(AB) \leq \min\{\operatorname{rank} A, \operatorname{rank} B\}$

g) Column space, nullspace, and the general solution to $A\boldsymbol{x} = \boldsymbol{b}$.
- If W is a subspace of $\mathbb{R}^n$,then a set of vectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ spans W if and only if the system $A\boldsymbol{x} = \boldsymbol{b}$ is consistent for all $\boldsymbol{b} \in W$ where
$$A = \left(\boldsymbol{v}_1 \;\middle|\; \ldots \;\middle|\; \boldsymbol{v}_k\right). \tag{1.7}$$
This is just saying that every vector in W must be a linear combination of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$.
- The column space of a matrix $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ is the set of all linear combinations of the columns of A as vectors in $\mathbb{R}^m$.
- The column space $\mathrm{Col}(A)$ is a subspace of $\mathbb{R}^m$.

- The nullspace of a matrix $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ is the set of vectors $\boldsymbol{x}$ in $\mathbb{R}^n$ for which $A\boldsymbol{x} = \vec{0}$.
- The nullspace $\mathrm{Null}(A)$ is a subspace of $\mathbb{R}^n$.
- In general, A and its row reduced echelon form do not have the same column space.
- If A, B are row equivalent, then they have the same nullspace.
- As a consequence, the nullspace of A is equal to the nullspace of its row reduced echelon form.
- The columns containing the pivots in the row reduced echelon form of A form a basis for the column space of the row reduced echelon form of A.
- The corresponding columns in the original matrix A form a basis for $\mathrm{Col}(A)$.
- Vectors in the nullspace $\mathrm{Null}(A)$ encode the dependencies of the columns of A.
- A basis for the nullspace $\mathrm{Null}(A)$ can be identified by writing every element in the nullspace as a linear combination of vectors, where the weights are specified by the free variables.
- The number of pivots in A is equal to the rank of A.
- The number of free variables in A is equal to the dimension of the nullspace of A.
- (Important) The rank-nullity theorem states that $\dim \mathrm{Col}(A) + \dim \mathrm{Null}(A) = n$ if $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. This is equivalent to saying that the number of pivots plus the number of free variables is equal to the number of columns.
- The rank of an $m \times n$ matrix is equal to the dimension of the column space as a subspace of $\mathbb{R}^m$ (and also the dimension of the row space as a subspace of $\mathcal{M}_{1\times n}(\mathbb{R})$ since row rank equals column rank).
- Since the row rank is equal to the column rank, $\dim \mathrm{Col}(A) = \dim \mathrm{Col}(A^T)$.
- As an application of the rank nullity theorem applied to A and A^T, since $\dim \mathrm{Null}(A), \dim \mathrm{Null}(A^T) \geq 0$, $\dim \mathrm{Col}(A) \leq n$ and $\dim \mathrm{Col}(A^T) \leq m$, but since $\dim \mathrm{Col}(A) = \dim \mathrm{Col}(A^T)$, we see that $\mathrm{rank}\, A \leq \min\{m, n\}$.
- If A is an $m \times n$ matrix with $m > n$, then by the rank-nullity theorem $\dim \mathrm{Col}(A) \leq n < m$, so the column space $\mathrm{Col}(A)$ will always be a strict subspace of $\mathbb{R}^m$ and not the entire space. This implies that $\mathrm{Col}(A) \neq \mathbb{R}^m$ and therefore $A\boldsymbol{x} = \boldsymbol{b}$ cannot admit a solution for all $\boldsymbol{b} \in \mathbb{R}^m$.
- If A is an $m \times n$ matrix with $m < n$, then $\dim \mathrm{Null}(A) = n - \dim \mathrm{Col}(A) \geq n - m > 0$, so the nullspace always contains non-zero vectors as it is not the trivial nullspace. Therefore if $A\boldsymbol{x} = \boldsymbol{b}$ is consistent, the system always admits infinitely many solutions.
- If $m = n$, then $\mathrm{Col}(A) = \mathbb{R}^n$ if and only if $\mathrm{Null}(A) = \{\mathbf{0}\}$. This implies that if $A\boldsymbol{x} = \boldsymbol{b}$ is consistent for every $\boldsymbol{b} \in \mathbb{R}^n$, then the system always admits unique solutions; if $A\boldsymbol{x} = \boldsymbol{c}$ admits a unique solution for some $\boldsymbol{c} \in \mathbb{R}^n$, then $A\boldsymbol{x} = \boldsymbol{b}$ must be consistent for every $\boldsymbol{b} \in \mathbb{R}^n$. On the other hand, if $A\boldsymbol{x} = \boldsymbol{b}$ is not consistent for every $\boldsymbol{b} \in \mathbb{R}^n$, then the system $A\boldsymbol{x} = \boldsymbol{c}$ admits infinitely many solutions if it is consistent; if the system $A\boldsymbol{x} = \boldsymbol{c}$ admits infinitely many solutions for some $\boldsymbol{c} \in \mathbb{R}^n$, then $A\boldsymbol{x} = \boldsymbol{b}$ cannot be consistent for all $\boldsymbol{b} \in \mathbb{R}^n$.
- Criteria for invertibility (you must have $m = n$): if $A \in \mathcal{M}_{n\times n}(\mathbb{R})$, then A is invertible if and only if $\mathrm{Col}(A) = \mathbb{R}^n$ if and only if $\mathrm{Null}(A) = \{\mathbf{0}\}$ if and only if $\mathrm{rank}\, A = \dim \mathrm{Col}(A) = n$ if and only if $\dim \mathrm{Null}(A) = 0$ if and only if all the columns of A are linearly independent if and only if all the rows of A are linearly independent.
- For a given $\boldsymbol{b} \in \mathbb{R}^m$, to identify the general solution to $A\boldsymbol{x} = \boldsymbol{b}$, we first identify a particular solution either via inspection or by row reduction, and then we identify a basis for the nullspace $\mathrm{Null}(A)$. The basis for the nullspace gives an explicit description for the general homogeneous solution, and the general solution to the inhomogeneous solution is $\boldsymbol{x} = \boldsymbol{x}_p + \boldsymbol{x}_h$.

h) Matrix transpose

- The transpose of a matrix $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ is the matrix $A^T \in \mathcal{M}_{n\times m}(\mathbb{R})$ defined via

$$(A^T)_{ij} = A_{ji},\ 1 \leq i \leq n, 1 \leq j \leq m. \tag{1.8}$$

- We always have $(A + cB)^T = A^T + cB^T$ and $(AB)^T = B^T A^T$.
- If A is invertible, then A^T is also invertible (since they are both square matrices with full rank). In this case we also have $(A^T)^{-1} = (A^{-1})^T$, because $A^{-1}A = I$ implies $A^T(A^{-1})^T = I^T = I$.
- If $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, then $A^T A \in \mathcal{M}_{n\times n}(\mathbb{R})$ and $AA^T \in \mathcal{M}_{m\times m}(\mathbb{R})$ are both symmetric matrices.
- An orthogonal matrix Q is a square matrix that satisfies $Q^T Q = I$. It is length preserving and also angle preserving.
- A square matrix is orthogonal if and only if its columns are all mutually orthogonal and their lengths are equal to 1.
- An orthogonal matrix Q also satisfies $QQ^T = I$ (why?).

- In general, if $m \geq n$, then an $m \times n$ matrix Q satisfies $Q^TQ = I$ if and only if its columns are all mutually orthogonal and their lengths are equal to 1. However, if $m > n$ then QQ^T cannot be the identity matrix.

i) The four fundamental subspaces, orthogonal complements
- By rank-nullity, if the $m \times n$ matrix A has rank r, then

$$\dim \mathrm{Col}(A) = r \tag{1.9}$$
$$\dim \mathrm{Null}(A) = n - r \tag{1.10}$$
$$\dim \mathrm{Col}(A^T) = r \tag{1.11}$$
$$\dim \mathrm{Null}(A^T) = m - r. \tag{1.12}$$

- In particular,

$$\dim \mathrm{Col}(A) + \dim \mathrm{Null}(A^T) = m \tag{1.13}$$
$$\dim \mathrm{Col}(A^T) + \dim \mathrm{Null}(A) = n. \tag{1.14}$$

- A vector subspace V is said to be orthogonal to another vector subspace U if every vector in V is perpendicular to every vector in U.
- Only the zero vector $\mathbf{0}$ can lie in the intersection of two orthogonal subspaces.
- The orthogonal complement of a vector subspace U, denoted by $U^\perp$, contains all vectors perpendicular to every vector in U. In this sense the orthogonal complement $U^\perp$ is the largest vector subspace orthogonal to U.
- If U admits a basis $\mathcal{B}$ and a vector $\boldsymbol{v}$ is perpendicular to all the vectors in the basis, then $\boldsymbol{v}$ lies in the orthogonal complement $U^\perp$ (why?).
- If $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, $\mathrm{Col}(A), \mathrm{Null}(A^T)$ are orthogonal complements of each other in $\mathbb{R}^m$, and $\mathrm{Col}(A^T)$ and $\mathrm{Null}(A)$ are orthogonal complements of each other in $\mathbb{R}^n$.
- In general, if V is a subspace of $\mathbb{R}^n$, then $\dim V + \dim V^\perp = n$.
- The orthogonal complement of the orthogonal complement of a subspace is itself: $(V^T)^T = V$. This means is that V is the orthogonal complement of W, then W is also the orthogonal complement of V.

j) Orthogonal projections, orthogonal decomposition
- Let l be a line spanned by a non-zero vector $\boldsymbol{a}$ in $\mathbb{R}^m$ and let $\boldsymbol{b}$ be any other vector in $\mathbb{R}^m$. The projection of $\boldsymbol{b}$ onto l is the vector

$$\boldsymbol{p} = \frac{\boldsymbol{a}\boldsymbol{a}^T}{\boldsymbol{a}^T\boldsymbol{a}}\boldsymbol{b} = \left(\frac{\boldsymbol{b}\cdot\boldsymbol{a}}{\boldsymbol{a}\cdot\boldsymbol{a}}\right)\boldsymbol{a} = \left(\boldsymbol{b}\cdot\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|}\right)\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|}. \tag{1.15}$$

- The derivation is as follows: the projection $\boldsymbol{p}$ lies on l, so $\boldsymbol{p} = \hat{x}\boldsymbol{a}$ for some scalar $\hat{x}$. The error vector $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = \boldsymbol{b} - \hat{x}\boldsymbol{a}$ must be perpendicular to l, so we must have $\boldsymbol{e}$ is perpendicular to $\boldsymbol{a}$. This means that

$$(\boldsymbol{b} - \hat{x}\boldsymbol{a})\cdot\boldsymbol{a} = 0. \tag{1.16}$$

From this it follows that

$$\boldsymbol{b}\cdot\boldsymbol{a} = \hat{x}\boldsymbol{a}\cdot\boldsymbol{a}, \tag{1.17}$$

so

$$\hat{x} = \frac{\boldsymbol{b}\cdot\boldsymbol{a}}{\boldsymbol{a}\cdot\boldsymbol{a}}. \tag{1.18}$$

Then

$$\boldsymbol{p} = \hat{x}\boldsymbol{a} = \left(\frac{\boldsymbol{b}\cdot\boldsymbol{a}}{\boldsymbol{a}\cdot\boldsymbol{a}}\right)\boldsymbol{a}. \tag{1.19}$$

- The projection matrix P associated to the line l maps any vector $\boldsymbol{b}$ to its projection $P\boldsymbol{b}$, and it's given by

$$P = \frac{\boldsymbol{a}\boldsymbol{a}^T}{\boldsymbol{a}^T\boldsymbol{a}}. \tag{1.20}$$

This is a rank 1 square matrix and $\mathrm{Col}(P) = l$, $\mathrm{Null}(P) = l^\perp$. If $\boldsymbol{a} \in \mathbb{R}^n$, then P is an $n \times n$ matrix.
- The projection matrix associated to the orthogonal complement of l is $I - P$.

- In general, we can project onto any vector subspace of $\mathbb{R}^m$. Suppose U is a n-dimensional vector subspace of $\mathbb{R}^m$ and $\mathcal{B}$ is a basis for U. Then consider the matrix A with the basis elements in $\mathcal{B}$ as its columns. Then $U = \mathrm{Col}(A)$ and A is an $m \times n$ matrix with n independent columns. Given a vector $\boldsymbol{b}$ in $\mathbb{R}^m$, the projection of $\boldsymbol{b}$ onto U is given by
$$\boldsymbol{p} = A(A^TA)^{-1}A^T\boldsymbol{b}. \tag{1.21}$$
This is a rank n $m \times m$ matrix with $\mathrm{Col}(P) = U$, $\mathrm{Null}(P) = U^\perp$.
- The projection matrix associated to $U^\perp$ is $I - P$.
- Let's review how we arrived at this. The projection $\boldsymbol{p}$ is in the column space of A, so
$$\boldsymbol{p} = \hat{x}_1\boldsymbol{a}_1 + \ldots + \hat{x}_n\boldsymbol{a}_n \tag{1.22}$$
for some scalars $\hat{x}_1, \ldots, \boldsymbol{x}_n$ and $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ are the columns of A. This is equivalent to writing
$$\boldsymbol{p} = A\hat{\boldsymbol{x}} \tag{1.23}$$
where $(\hat{\boldsymbol{x}})_i = \hat{x}_i$. The error vector $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$ is perpendicular to $\mathrm{Col}(A)$, so it must be in $\mathrm{Null}(A^T)$. This means that
$$A^T(\boldsymbol{b} - \boldsymbol{p}) = A^T(\boldsymbol{b} - A\hat{\boldsymbol{x}}) = 0. \tag{1.24}$$
This means that
$$A^TA\hat{\boldsymbol{x}} = A^T\boldsymbol{b}. \tag{1.25}$$
- In general $\mathrm{Null}(A) = \mathrm{Null}(A^TA)$. This is because if $\boldsymbol{x} \in \mathrm{Null}(A)$, then $A\boldsymbol{x} = \boldsymbol{0}$ and $A^TA\boldsymbol{x} = A^T\boldsymbol{0} = \boldsymbol{0}$. If $\boldsymbol{x} \in \mathrm{Null}(A^TA)$, then $A^TA\boldsymbol{x} = \boldsymbol{0}$, so $A\boldsymbol{x} \in \mathrm{Null}(A^T)$, but since $A\boldsymbol{x}$ is also in the column space of $\mathrm{Col}(A) = (\mathrm{Null}(A^T))^\perp$, we must have $A\boldsymbol{x} = \boldsymbol{0}$. Since A^TA is an $n \times n$ square matrix with the same nullspace as A, A^TA is invertible if we assume the columns of A are independent.
- In other words, if $\mathrm{Null}(A) = \{\boldsymbol{0}\}$, then $\mathrm{Null}(A^TA) = \{\boldsymbol{0}\}$. So if $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ has independent columns, then $A^TA \in \mathcal{M}_{n\times n}(\mathbb{R})$ is invertible. This is equivalent to saying that if A has full column rank, then A^TA has full rank.
- Therefore if we assume that A has independent columns, then A^TA is invertible and
$$\hat{\boldsymbol{x}} = (A^TA)^{-1}A^T\boldsymbol{b}. \tag{1.26}$$
Then
$$\boldsymbol{p} = A\hat{\boldsymbol{x}} = A(A^TA)^{-1}A^T\boldsymbol{b}. \tag{1.27}$$
The projection matrix P is
$$P = A(A^TA)^{-1}A^T. \tag{1.28}$$
- The projection matrix P associated to a subspace V is symmetric: $P^T = P$.
- The projection matrix P associated to a subspace V satisfies $P^2 = P$: the projection of a projection is itself. In fact, $P^k = P$ for all $k \geq 2$.
- The projection $\boldsymbol{p}$ of a vector $\boldsymbol{b}$ onto a subspace V is the "best approxiation" of $\boldsymbol{b}$, in the sense that for all vectors $\boldsymbol{w} \neq \boldsymbol{p}$ in V, we have
$$\|\boldsymbol{b} - \boldsymbol{p}\| < \|\boldsymbol{b} - \boldsymbol{w}\|. \tag{1.29}$$
- As a corollary, the orthogonal projection $\boldsymbol{p}$ is the unique minimizer of the function $E : \mathrm{Col}(A) \to \mathbb{R}$ defined via
$$E(\boldsymbol{w}) = \|\boldsymbol{b} - \boldsymbol{w}\|^2. \tag{1.30}$$
Again, we must assume that A has independent columns.
- Equivalently, $\hat{\boldsymbol{x}}$ is the unique minimizer of the function $E : \mathbb{R}^n \to \mathbb{R}$ defined via
$$E(\boldsymbol{x}) = \|\boldsymbol{b} - A\boldsymbol{x}\|^2. \tag{1.31}$$
Again, we must assume that A has independent columns.
- If U, W are two subspaces of $\mathbb{R}^n$ and $U \cap W = \{\boldsymbol{0}\}$, the direct sum $U \oplus W$ of U, W is the set of vectors z for which z can be written uniquely written as a sum of a vector $u \in U$ and a vector $w \in W$.
- $U \oplus W$ is a subspace of $\mathbb{R}^n$ and $\dim(U \oplus W) = \dim U + \dim W$. In fact, if $\mathcal{B}_U = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ is a basis for U and $\mathcal{B}_W = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_j\}$ is a basis for W, then $\mathcal{B} = \mathcal{B}_U \cup \mathcal{B}_W$ is a basis for $U \oplus W$.

- (Orthogonal decomposition theorem) If V be a subspace of $\mathbb{R}^n$, then $\mathbb{R}^n = V \oplus V^\perp$ and every vector $\boldsymbol{x}$ in $\mathbb{R}^n$ can be uniquely decomposed as

$$\boldsymbol{x} = \boldsymbol{x}_V + \boldsymbol{x}_{V^\perp}, \tag{1.32}$$

where $\boldsymbol{x}_V$ belongs to the subspace V and $\boldsymbol{x}_{V^\perp}$ belongs to the orthogonal complement of V.
- If $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, then $(\mathrm{Null}(A))^\perp = \mathrm{Col}(A^T)$ as subspaces of $\mathbb{R}^n$. Therefore $\mathbb{R}^n = \mathrm{Col}(A^T) \oplus \mathrm{Null}(A)$ and every vector $\boldsymbol{x}$ in $\mathbb{R}^n$ can be written uniquely as

$$\boldsymbol{x} = \boldsymbol{x}_{\mathrm{Col}(A^T)} + \boldsymbol{x}_{\mathrm{Null}(A)}, \tag{1.33}$$

where $\boldsymbol{x}_{\mathrm{Col}(A^T)}$ lives in the column space of A^T and $\boldsymbol{x}_{\mathrm{Null}(A)}$ lives in the nullspace of A.
- Another way of thinking about this is that if P is the $n \times n$ projection associated to $\mathrm{Col}(A^T)$, then

$$\boldsymbol{x} = \underbrace{P\boldsymbol{x}}_{\in \mathrm{Col}(A^T)} + \underbrace{(I-P)\boldsymbol{x}}_{\in \mathrm{Null}(A)}. \tag{1.34}$$

- If $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, then $(\mathrm{Null}(A^T))^\perp = \mathrm{Col}(A)$ as subspaces of $\mathbb{R}^m$. Therefore $\mathbb{R}^m = \mathrm{Col}(A) \oplus \mathrm{Null}(A^T)$ and every vector $\boldsymbol{y}$ in $\mathbb{R}^m$ can be written as

$$\boldsymbol{y} = \boldsymbol{y}_{\mathrm{Col}(A)} + \boldsymbol{y}_{\mathrm{Null}(A^T)}, \tag{1.35}$$

where $\boldsymbol{y}_{\mathrm{Col}(A)}$ lives in the column space of A and $\boldsymbol{y}_{\mathrm{Null}(A^T)}$ lives in the nullspace of A^T.
- Another way of thinking about this is that if P is the $m \times m$ projection associated to $\mathrm{Col}(A)$, then

$$\boldsymbol{y} = \underbrace{P\boldsymbol{y}}_{\in \mathrm{Col}(A)} + \underbrace{(I-P)\boldsymbol{y}}_{\in \mathrm{Null}(A^T)}. \tag{1.36}$$

k) Least squares problems
- The vector $\hat{\boldsymbol{x}}$ is referred to as the *least squares solution* to $A\boldsymbol{x} = \boldsymbol{b}$ if $\boldsymbol{b} \notin \mathrm{Col}(A)$, which satisfies $A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$. The least squares solution is significant because the projection $\boldsymbol{p} = A\hat{\boldsymbol{x}}$ is the best approximation to the vector $\boldsymbol{b}$ in the column space of A, in the sense that

$$\|\boldsymbol{b} - A\hat{\boldsymbol{x}}\| < \|\boldsymbol{b} - A\boldsymbol{x}\|, \tag{1.37}$$

for all vector $\boldsymbol{x} \neq \hat{\boldsymbol{x}}$ in $\mathbb{R}^n$. As a result, the least square solution $\hat{\boldsymbol{x}}$ minimizes the error function

$$E(\boldsymbol{x}) = \|\boldsymbol{b} - A\boldsymbol{x}\|^2, \boldsymbol{x} \in \mathbb{R}^n. \tag{1.38}$$

- To show (1.37) we can use Pythagorean theorem: the vectors $\boldsymbol{b} - A\boldsymbol{x}$, $\boldsymbol{e}$, and $\boldsymbol{p} - A\boldsymbol{x}$ form a right triangle for all vectors $\boldsymbol{x}$ in $\mathbb{R}^n$. Therefore

$$\|b - A\boldsymbol{x}\|^2 = \|\boldsymbol{e}\|^2 + \|\boldsymbol{p} - A\boldsymbol{x}\|^2 \geq \|\boldsymbol{e}\|^2 = \|\boldsymbol{b} - \boldsymbol{p}\|^2, \tag{1.39}$$

where we have equality precisely when $p = A\boldsymbol{x}$, or when $\boldsymbol{x} = \hat{\boldsymbol{x}}$. In other words, the error is minimized precisely when $A\boldsymbol{x}$ is equal to the projection $\boldsymbol{p}$.
- An $m \times n$ matrix Q is said to be semi-orthogonal if $Q^TQ = I_{n\times n}$. When Q is a square matrix, we say that it is orthogonal. Orthogonal matrices satisfy $Q^{-1} = Q^T$ and $Q^TQ = QQ^T = I$.
- If the subspace U has an orthonormal basis, and Q is a matrix with the basis vectors as its columns, then Q is semi-orthogonal and the projection matrix P associated to $U = \mathrm{Col}(Q)$ is

$$P = QQ^T. \tag{1.40}$$

- As a consequence, the projection $\boldsymbol{p}$ of any vector $\boldsymbol{b}$ onto the column space $\mathrm{Col}(Q)$ is

$$\boldsymbol{p} = QQ^T\boldsymbol{b} = (\boldsymbol{q}_1 \cdot \boldsymbol{b})\boldsymbol{q}_1 + \ldots + (\boldsymbol{q}_n \cdot \boldsymbol{b})\boldsymbol{q}_k. \tag{1.41}$$

where $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_k$ are the columns of Q. Another way to write this is

$$\boldsymbol{p} = \mathrm{proj}_U\, \boldsymbol{b} = \mathrm{proj}_{\boldsymbol{q}_1} \boldsymbol{b} + \ldots + \mathrm{proj}_{\boldsymbol{q}_k} \boldsymbol{b}, \tag{1.42}$$

- Parseval's identity gives us the generalization of the Pythagorean theorem:

$$\|\boldsymbol{p}\|^2 = \|\mathrm{proj}_{\boldsymbol{q}_1} \boldsymbol{b}\|^2 + \ldots \|\mathrm{proj}_{\boldsymbol{q}_k} \boldsymbol{b}\|^2 = (\boldsymbol{q}_1 \cdot \boldsymbol{b})^2 + \ldots + (\boldsymbol{q}_k \cdot \boldsymbol{b})^2. \tag{1.43}$$

- Given a set of data points, we can write down an appropriate equation $A^TA\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$ and the least squares solution
$$\hat{\boldsymbol{x}} = \begin{pmatrix} C \\ D \end{pmatrix} \tag{1.44}$$
gives us the coefficients of the trend line that minimizing the sum of the squares of the vertical distances between the data points and the trend line.

l) Gram-Schmidt, QR-decomposition
- Given a basis $\mathcal{B} = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$ for a vector subspace U, we can find an orthogonal basis $\mathcal{B}' = \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n\}$ via the Gram-Schmidt process. The idea is to subtract away the projection of $\boldsymbol{a}_k$ onto the subspace spanned by $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{k-1}$, and the error vector $\boldsymbol{v}_k$ will be perpendicular to $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{k-1}$.
- In matrix form, this can be written as $A = QR$. The matrix R is a square matrix and it records the size of the projections and the process of how to arrive at one basis from the other. The diagonal entries of R records the length of the error vectors $\boldsymbol{v}_k$, so the diagonal entries are positive. This implies that R is invertible.

m) Determinants
- The determinant is a number associated to a square matrix. It can be calculated using a specific formula.
- If a matrix has a row or a column of zeros, then its determinant is 0.
- A matrix B obtained from adding a multiple of a row to another row or a multiple of a column to another column of a matrix A has the same determinant as A.
- A matrix A is invertible if and only if $\det(A) \neq 0$. This is because A has the same determinant as its row echelon form (up to a non-zero constant because of scaling and swapping), and A is invertible if and only if it can be row reduced to the identity matrix. If not, then the row reduced echelon form will contain a row of zeros, and the determinant of that matrix is 0.
- If A, B are both square matrices, then the determinant of AB is equal to the determinant of A times the determinant of B.
- The determinant of A is equal to the determinant of its transpose A^T.
- If A is invertible, then
$$\det A^{-1} = \frac{1}{\det A}. \tag{1.45}$$
This is because $\det A \det A^{-1} = \det I = 1$.

n) Eigenvectors and eigenvalues
- If a non-zero vector $\boldsymbol{v}$ and a scalar λ satisfies
$$A\boldsymbol{v} = \lambda\boldsymbol{v}, \tag{1.46}$$
for a square matrix A, then λ is said to be an eigenvalue of A and $\boldsymbol{v}$ is said to be an eigenvector of A.
- Eigenvectors are not unique: they are vectors in the nullspace of the matrix $A - \lambda I$, so any non-zero scalar multiple of an eigenvector is also an eigenvector.
- The eigenvalues of a matrix are roots to the characteristic polynomial
$$p_A(\lambda) = \det(A - \lambda I), \lambda \in \mathbb{C} \tag{1.47}$$
associated to A. The leading coefficient of this polynomial is $(-1)^n$ and the degree of $p_A(\lambda)$ is n, if A is $n \times n$.
- By the fundamental theorem of algebra, every $n \times n$ square matrix has n eigenvalues (possibly repeating and possibly complex).
- This does not mean that you can always find n linearly independent eigenvectors.
- The trace of a square matrix is the sum of its diagonal elements.
- The trace of an $n \times n$ matrix is also equal to the sum of its n eigenvalues, and the determinant of an $n \times n$ matrix is equal to the product of its n eigenvalues.
- The trace of AB is equal to the trace of BA.
- Matrices sharing the same eigenvalues might not have the same eigenvectors.
- The eigenvalues of a triangular matrix are the entries of its diagonal.
- If the rows of a matrix add up to the same number α, then $\boldsymbol{v} = \mathbf{1}$ (vector containing all 1s) is an eigenvector and the corresponding eigenvalue is $\lambda = \alpha$.
- If $A + \lambda I$ has rows or columns that are multiples of each other, then $-\lambda$ is an eigenvalue.

o) Diagonalization, the spectral theorem

- If $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of a square matrix A and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are a set of eigenvectors, then we can always write
$$AX = X\Lambda, \tag{1.48}$$
where X is the matrix with the eigenvectors as its columns and Λ is the matrix with the corresponding eigenvalues (this is important, the order has to match) on its diagonal.
- If the eigenvectors of A are also linearly independent, then the square matrix X is invertible and one can write
$$A = X\Lambda X^{-1}. \tag{1.49}$$
The process of factoring A this way is referred to as the process of diagonalizing A.
- We say that A is diagonalizable over $\mathbb{R}$ if such an invertible matrix X and diagonal matrix Λ exist and are real matrices.
- We say that A is diagonalizable over $\mathbb{C}$ if X and Λ are complex matrices.
- If A is diagonalizable, then
$$A^k = X\Lambda^k X^{-1}. \tag{1.50}$$
- A is diagonalizable when A has distinct eigenvalues.
- The eigenvalues of a real symmetric matrix are real.
- The eigenvectors of a real symmetric matrix can be chosen to be real.
- (Spectral theorem) The real eigenvectors of a real symmetric matrix are can also be chosen to be orthonormal, so we can write
$$A = Q\Lambda Q^{-1} = Q\Lambda Q^T \tag{1.51}$$
where Q is an orthogonal matrix.
- In this case, the orthogonal matrices Q and Q^T can be thought of as a rotation matrix (maybe times a reflection/permutation matrix depending on the position of its columns). The matrix Λ is stretching the vector $Q^T\boldsymbol{b}$ in the $\boldsymbol{e}_k$ direction by a factor of λ_k.
- Not all matrices are diagonalizable. The classic counterexample is
$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{1.52}$$
The dimension of the nullspace of $A-I$ is 1, so A does not admit two linearly independent eigenvectors.
- A real matrix A is diagonalizable over $\mathbb{R}$ if and only if all of its eigenvalues are real and for every eigenvalue λ of A, the algebraic and geometric multiplicities of λ are equal.

p) The singular-value decomposition
- Any matrix $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ admits a singular-value decomposition of the form $A = U\Sigma V^T$, where U is related to the eigen-decomposition of AA^T and V is related to the eigen-decomposition of A^TA.
- The matrix AA^T is of dimension $m \times m$ and the matrix A^TA is of dimension $n \times n$.
- The matrices AA^T and A^TA share the same positive eigenvalues. The rest of their eigenvalues are zeros.
- If the rank of A is r, then the first r eigenvalues ordered in descending order of AA^T and A^TA are positive and the rest are zero.
- The singular values of A are the square roots of the eigenvalues of AA^T and A^TA.
- The left singular vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ are the eigenvectors of the $m \times m$ square matrix AA^T and the right singular vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are the eigenvectors of the $n \times n$ square matrix A^TA.
- The left singular vectors are orthogonal to each other.
- The right singular vectors are orthogonal to each other.
- The left and right singular vectors are related to each other by $A\boldsymbol{v}_i = \sigma_i\boldsymbol{u}_i$ and $A^T\boldsymbol{u}_i = \sigma_i\boldsymbol{v}_i$ if the singular value σ_i is positive.
- The first r left singular vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ (normalized versions of $A\boldsymbol{v}_1, \ldots, A\boldsymbol{v}_r$) form an orthonormal basis for the column space of A.
- The other $m-r$ left singular vectors $\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_m$ form an orthonormal basis for the nullspace of A^T.
- The first r right singular vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ (normalized versions of $A^T\boldsymbol{u}_1, \ldots, A\boldsymbol{u}_r$) form an orthonormal basis for the column space of A^T.
- The other $n-r$ left singular vectors $\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n$ form an orthonormal basis for the nullspace of A.
- In the SVD of A, the matrix $U \in \mathcal{M}_{m\times m}(\mathbb{R}), \Sigma \in \mathcal{M}_{m\times n}(\mathbb{R}), V \in \mathcal{M}_{n\times n}(\mathbb{R})$.
- The matrices U, V are orthogonal matrices.

- The matrix Σ is a rectangular diagonal matrix.
- From an SVD of a matrix A with rank r, we can write A as a sum of rank 1 matrices: $A = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \ldots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^T$.
- From this we see that we can write a more compact version of SVD using only the first r singular vectors and the positive singular values: $A = U_r \Sigma_r V_r^T$.
- Using the compact SVD, we can define the pseudoinverse of a matrix: $A^+ = V_r \Sigma_r^{-1} U_r$. This is matrix that maps everything from $\mathrm{Col}(A)$ to a vector in $\mathrm{Col}(A^T)$ and everything from $\mathrm{Null}(A^T)$ to the zero vector.
- The projection matrix onto $\mathrm{Col}(A)$ is $A^+ A$ or $U_r U_r^T = \boldsymbol{u}_1 \boldsymbol{u}_1^T + \ldots + \boldsymbol{u}_r \boldsymbol{u}_r^T$.
- The least squares solutions to $A\boldsymbol{x} = \boldsymbol{b}$ with minimum Euclidean norm is $\hat{\boldsymbol{x}} = A^+ \boldsymbol{b} = V_r \Sigma_r^{-1} U_r \boldsymbol{b}$. The minimum of the least squares error is $\left\| \boldsymbol{b} - U_r U_r^T \boldsymbol{b} \right\|^2$.

q) Linear transformations and their matrix representations

- A map $T : V \to W$ is said to be a linear transformation (they're also referred to as linear maps or linear operators) between two vector spaces V, W if $T(x + cy) = T(x) + cT(y)$ for any $x, y \in V, c \in \mathbb{R}$.
- An ordered basis for a vector space is a basis endowed with a specific order.
- If $\alpha = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ is an ordered basis for V, then any vector $\boldsymbol{v} \in V$ can be written uniquely as $\boldsymbol{v} = c_1 \boldsymbol{v}_1 + \ldots + c_n \boldsymbol{v}_n$. We can use this set of unique coefficients to define the coordinate vector of $\boldsymbol{v}$: $[\boldsymbol{v}]_\alpha = \begin{pmatrix} c_1 & \ldots & c_n \end{pmatrix}^T$.
- Using the notion of coordinates, we may define the matrix representation of a linear transformation $T : V \to W$ with respect to two ordered basis $\alpha = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ and $\beta = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m\}$ for V, W respectively, as the matrix $[T]_\alpha^\beta \in \mathcal{M}_{m \times n}(\mathbb{R})$ defined via $[T]_\alpha^\beta = \big([T(\boldsymbol{v}_1)]_\beta \;\big|\; \ldots \;\big|\; [T(\boldsymbol{v}_n)]_\beta \big)$.
- The coordinates of $T(\boldsymbol{v})$ for any $\boldsymbol{v} \in V$ can be computed via $[T(\boldsymbol{v})]_\beta = [T]_\alpha^\beta [\boldsymbol{v}]_\alpha$.
- Change of basis matrix: if β, β' are two bases for a vector space V, then the change of basis matrix $Q_{\beta \to \beta'}$ is the matrix representation of the identity map $\mathrm{id} : V \to V$ with respect to β, β', i.e. $Q_{\beta \to \beta'} = [I]_\beta^{\beta'}$. By construction, $Q_{\beta \to \beta'} [\boldsymbol{v}]_\beta = [\boldsymbol{v}]_{\beta'}$ for any $\boldsymbol{v} \in V$.