Advanced Linear Algebra

Semester 1, 2022

Prof Timothy Moroney

Tarang Janawalkar

Contents

1 Fundamental Concepts of Linear Algebra

1.1 Row Echelon Form

As studied in Linear Algebra, we can solve linear systems by applying the following elementary row operations to any matrix $\mathbf{A}$.

Type I. Exchange any two rows.

Type II. Multiply any row by a constant.

Type III. Add a multiple of one row to another row.

This allows us to reduce $\mathbf{A}$ into **row echelon form** such that the entries below the main diagonal are zero:

$$\mathbf{R}_{\text{ref}} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{mn} \end{bmatrix}$$

1.2 Elementary Matrix

Mathematically, we can represent these row operations as a matrix which is left multiplied to $\mathbf{A}$.

Definition 1.1 (Elementary matrix)**.** An elementary matrix $\mathbf{E}_i$ is constructed by applying a row operation to the elementary matrix $\mathbf{I}_m$. Consider a 3 by 4 matrix $\mathbf{A}$; a common first elementary row operation might be

$$r_2 \leftarrow r_2 - \frac{a_{21}}{a_{11}} r_1$$

which when applied to $\mathbf{I}_3$ yields

$$\mathbf{E}_1 = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{a_{21}}{a_{11}} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

where the 1 subscript simply indicates the first of many elementary row operations. Left multiplying this to an arbitrary $\mathbf{A}$ gives

$$\begin{aligned} \mathbf{E}_1\mathbf{A} &= \begin{bmatrix} 1 & 0 & 0 \\ -\frac{a_{21}}{a_{11}} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \\ &= \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} - \frac{a_{12}a_{21}}{a_{11}} & a_{23} - \frac{a_{13}a_{21}}{a_{11}} & a_{24} - \frac{a_{14}a_{21}}{a_{11}} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \end{aligned}$$

which has the desired result of eliminating the first column of the second row.

1.3 Reduced Row Echelon Form

As there are infinitely many ways to reduce a matrix to row echelon form, we typically reduce $\mathbf{R}_{\text{ref}}$ further into **reduced row echelon form** which is a unique reduction for every $\mathbf{A}$.
This matrix $\mathbf{R}_{\text{rref}}$ (or simply $\mathbf{R}$) generally requires $m \times n$ elementary row operations and is only useful for theoretical analysis. In reduced row echelon form, any entries in the same column as a pivot must be 0, and each pivot is 1.

1.4 Elimination Matrix

The elementary matrices involved in row reduction can be expressed as a single matrix containing every each row operation.

$$\begin{aligned}\mathbf{E}_9\mathbf{E}_8 \dots \mathbf{E}_2\mathbf{E}_1\mathbf{A} &= \mathbf{E}\mathbf{A} \\ &= \mathbf{R}\end{aligned}$$

1.5 Linear Systems

Given the linear system $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ we can augment $\mathbf{A}$ with $\boldsymbol{b}$ to draw conclusions about the solutions. If we left multiply the elimination matrix $\mathbf{E}$ to $\left[\mathbf{A} \mid \boldsymbol{b}\right]$ we can apply the same operations to $\boldsymbol{b}$.

$$\begin{aligned}\mathbf{E}\left[\mathbf{A} \mid \boldsymbol{b}\right] &= \left[\mathbf{E}\mathbf{A} \mid \mathbf{E}\boldsymbol{b}\right] \\ &= \left[\mathbf{R} \mid \boldsymbol{z}\right]\end{aligned}$$

Therefore

$$\mathbf{R}\boldsymbol{x} = \boldsymbol{z}$$

After reducing the matrix $\mathbf{A}$ to $\mathbf{R}$, we can summarise certain characteristics about $\mathbf{A}$.

1.5.1 Basic and Free Variables

Identifying the pivots in $\mathbf{R}$ allows us to determine the dimensions of various subspaces of $\mathbf{A}$.

Definition 1.2 (Basic variables)**.** The columns that a pivot corresponds to are known as basic variables (or leading variables).

Definition 1.3 (Free variables)**.** Any columns not corresponding to any pivots are known as free variables (or parameters). Consequently, any variables that are not basic variables are free variables.

When using backward substitution to solve $\mathbf{R}\boldsymbol{x} = \mathbf{z}$, we assign new variables to any free variables to indicate that they are parameters to the system.

1.5.2 Singular Matrices

An n by n square matrix $\mathbf{A}$ is singular if its associated reduced matrix $\mathbf{R}$ has fewer than n basic variables. It follows that a singular matrix also has a determinant of 0 (as the product of the diagonal is 0) which means it is also noninvertible.

1.6 The Four Fundamental Subspaces

Consider

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 0 & 1 & 4 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{R} = \begin{bmatrix} 1 & 2 & 0 & -2 & -1 \\ 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \mathbf{E} = \begin{bmatrix} 1 & -3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

so that x_1 and x_3 are basic variables, whereas x_2, x_4 and x_5 are free variables.

1.6.1 Row space

The rows containing pivots form the basis vectors for the rowspace of $\mathbf{A}$, denoted $\mathcal{C}\left(\mathbf{A}^\top\right)$.

$$\mathcal{C}\left(\mathbf{A}^\top\right) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \\ -2 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 2 \\ 2 \end{bmatrix} \right\}$$

Note that the row vectors here are represented as column vectors to allows us to conveniently compare these vectors with other spaces.

1.6.2 Null space

The span of vectors that satisfy the homogeneous system $\mathbf{A}\boldsymbol{x} = \mathbf{0}$, form the null space of $\mathbf{A}$, denoted $\mathcal{N}\left(\mathbf{A}\right)$.

$$\mathcal{N}\left(\mathbf{A}\right) = \operatorname{span}\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \\ 0 \\ 1 \end{bmatrix} \right\}$$

1.6.3 Column space

By considering a general vector $\boldsymbol{b}$, we can construct the column space of $\mathbf{A}$, denoted $\mathcal{C}\left(\mathbf{A}\right)$. This is done by augmenting $\mathbf{A}$ with $\boldsymbol{b}$, and appling the same elimination matrix $\mathbf{E}$.

$$\mathbf{E}\left[\mathbf{A} \mid \boldsymbol{b}\right] = \left[\begin{array}{ccccc|c} 1 & 2 & 0 & -2 & -1 & b_1 - 3b_2 \\ 0 & 0 & 1 & 2 & 2 & b_2 \\ 0 & 0 & 0 & 0 & 0 & b_3 \end{array}\right]$$

here we determine any constraints required to make $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ consistent, resulting in

$$\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ 0 \end{bmatrix}$$

This vector will guarantee a consistent solution for any b_1 and b_2, with $b_3 = 0$.
Rewriting $\boldsymbol{b}$ in terms of its two parameters b_1 and b_2, we can construct the basis vectors for the column space of $\mathbf{A}$.

$$\mathcal{C}(\mathbf{A}) = \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

Note that we could have also calculated the column space by finding the rowspace of $\mathbf{A}^\top$.

1.6.4 Left-Null Space

Just as we found the null space for $\mathbf{A}$, we can find a null space for $\mathbf{A}^\top$. This is known as the left-null space, denoted $\mathcal{N}(\mathbf{A}^\top)$.

$$\mathcal{N}(\mathbf{A}^\top) = \text{span}\left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

These four spaces form the fundamental subspaces for any matrix $\mathbf{A}$.

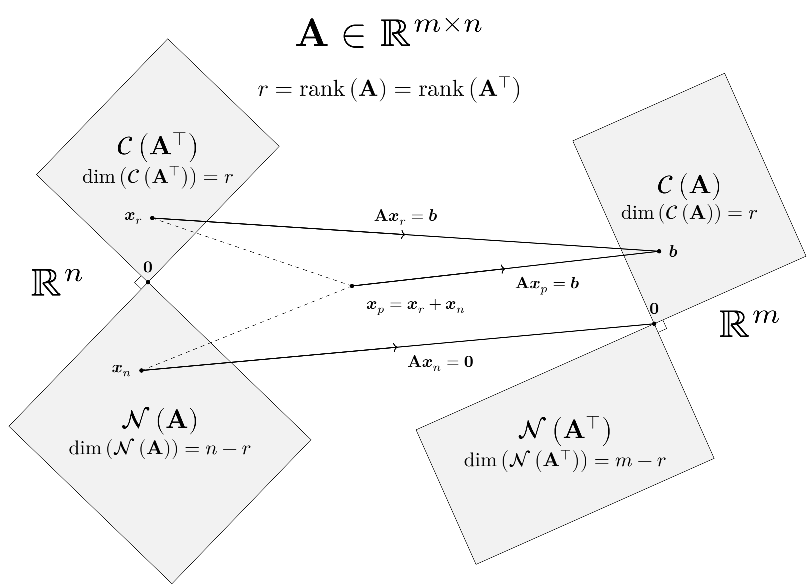

Figure 1: The four fundamental subspaces.

1.6.5 Dimensions of Subspaces

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

Definition 1.4 (Rank)**.** The dimension of the row space is called the **rank** of a matrix.

$$\text{rank}(\mathbf{A}) = \dim(\mathcal{C}(\mathbf{A}^\top)) = r$$

To determine the rank, we can count the number of basic variables in $\mathbf{R}$.
Note that the rank $(\mathbf{A}) = \operatorname{rank}(\mathbf{A}^\top)$.

Definition 1.5 (Nullity)**.** The dimension of the null space is called the **nullity** of the matrix.

$$\operatorname{null}(\mathbf{A}) = \dim(\mathcal{N}(\mathbf{A}))$$

Definition 1.6 (Left nullity)**.** The dimension of the left null space is called the **left nullity** of the matrix.

$$\operatorname{null}(\mathbf{A}^\top) = \dim(\mathcal{N}(\mathbf{A}^\top))$$

Theorem 1.6.1 (Rank-nullity theorem)**.** *The dimension of the domain of* $\mathbf{A}$*,* $\mathbb{R}^n$*, is given by the sum of the dimensions of the row space and null space of* $\mathbf{A}$*.*

$$\begin{aligned}\dim(\mathcal{C}(\mathbf{A}^\top)) + \dim(\mathcal{N}(\mathbf{A})) &= \dim(\mathbb{R}^n)\\ \operatorname{rank}(\mathbf{A}) + \operatorname{null}(\mathbf{A}) &= n\\ r + \operatorname{null}(\mathbf{A}) &= n\end{aligned}$$

Therefore

$$\operatorname{null}(\mathbf{A}) = n - r$$

Corollary 1.6.1.1 (Rank-nullity theorem for the transpose)**.** *The dimension of the codomain of* $\mathbf{A}$*,* $\mathbb{R}^m$*, is given by the sum of the dimensions of the column space and left-null space of* $\mathbf{A}$*.*

$$\begin{aligned}\dim(\mathcal{C}(\mathbf{A})) + \dim(\mathcal{N}(\mathbf{A}^\top)) &= \dim(\mathbb{R}^m)\\ \operatorname{rank}(\mathbf{A}^\top) + \operatorname{null}(\mathbf{A}^\top) &= m\\ r + \operatorname{null}(\mathbf{A}^\top) &= m\end{aligned}$$

Therefore

$$\operatorname{null}(\mathbf{A}^\top) = m - r$$

Theorem 1.6.2 (Orthogonality of subspaces)**.** *The row space and null space are orthogonal complements in* $\mathbb{R}^n$*.*

$$\mathcal{C}(\mathbf{A}^\top)^\perp = \mathcal{N}(\mathbf{A})$$

Similarly, the column space and left-null space are orthogonal complements in $\mathbb{R}^m$*.*

$$\mathcal{C}(\mathbf{A})^\perp = \mathcal{N}(\mathbf{A}^\top)$$

1.7 Consistency of a Linear System

Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, the linear system $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$, with any vector $\boldsymbol{b}$ can be described as follows:

1. Consistent with unique solution:

$$\operatorname{rank}(\mathbf{A}) = \operatorname{rank}([\mathbf{A} \mid \boldsymbol{b}]) = n$$

2. Consistent with infinitely many solutions:

$$\operatorname{rank}(\mathbf{A}) = \operatorname{rank}([\mathbf{A} \mid \boldsymbol{b}])$$

3. Inconsistent with no solutions:

$$\operatorname{rank}(\mathbf{A}) \neq \operatorname{rank}([\mathbf{A} \mid \boldsymbol{b}])$$

However, if $\boldsymbol{b} \in \mathcal{C}(\mathbf{A})$, the system must be consistent.

1.8 General Solution to a Linear System

Given $\boldsymbol{b} \in \mathcal{C}(\mathbf{A})$, the general solution to a system can be expressed as

$$\boldsymbol{x}_g = \boldsymbol{x}_p + \boldsymbol{x}_n$$

where $\boldsymbol{x}_p$ is a particular solution obtained by backward substitution, and $\boldsymbol{x}_n$ represents the linear combination of all null space basis vectors.

1.9 Minimum Norm Solution

In general, the particular solution may contain a linear combination of null space basis vectors requiring $\boldsymbol{x}_p \in \mathbb{R}^n$.
If we consider the solution vector $\boldsymbol{x}_r \in \mathcal{C}(\mathbf{A}^\top)$, then this vector will be the minimum norm solution to $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$.

2 Least Squares

Given an inconsistent system $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$, we can consider an approximate solution such that we minimise the norm of the residual:

$$\|\boldsymbol{b} - \mathbf{A}\boldsymbol{x}\|$$

Therefore we solve the following minimisation problem

$$\boldsymbol{x} = \underset{\boldsymbol{x}^* \in \mathbb{R}^n}{\arg\min} \|\boldsymbol{b} - \mathbf{A}\boldsymbol{x}\|$$

which is known as the **Least Squares** problem.

Theorem 2.0.1 (Minimum norm solution)**.** *The solution to the least squares problem is obtained by the vector $\boldsymbol{x}$ such that $\boldsymbol{b} - \mathbf{A}\boldsymbol{x}$ is orthogonal to the column space of $\mathbf{A}$.*

2.1 Normal Equations

Given that $\boldsymbol{b} - \mathbf{A}\boldsymbol{x}$ is orthogonal to $\mathcal{C}(\mathbf{A})$, it must lie in the left-null space of $\boldsymbol{A}$. By orthogonality, we form the following relationship

$$\mathbf{A}^\top (\boldsymbol{b} - \mathbf{A}\boldsymbol{x}) = 0$$

which is equivalent to solving

$$\mathbf{A}^\top \mathbf{A}\boldsymbol{x} = \mathbf{A}^\top \boldsymbol{b}.$$

This linear equation is known as the **Normal Equations**, which can be thought of as a generalisation of $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$. But whereas $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ can be inconsistent, the normal equations are always consistent.

2.2 Orthogonal Projection

By rearranging the normal equations, we can directly compute the solution vector.

$$\boldsymbol{x} = \left(\mathbf{A}^\top \mathbf{A}\right)^{-1} \mathbf{A}^\top \boldsymbol{b}$$

The projection of $\boldsymbol{b}$ onto $\mathcal{C}\left(\mathbf{A}\right)$ is therefore

$$\begin{aligned}\boldsymbol{b}_P &= \operatorname{proj}_{\mathcal{C}(\mathbf{A})}\left(\boldsymbol{b}\right) \\ &= \mathbf{A}\boldsymbol{x} \\ &= \mathbf{A}\left(\mathbf{A}^\top \mathbf{A}\right)^{-1} \mathbf{A}^\top \boldsymbol{b}\end{aligned}$$

We can identify the matrix

$$\mathbf{P} = \mathbf{A}\left(\mathbf{A}^\top \mathbf{A}\right)^{-1} \mathbf{A}^\top$$

as the orthogonal projector onto the column space of $\mathbf{A}$, so that it operates on $\boldsymbol{b}$:

$$\mathbf{P}\boldsymbol{b} = \operatorname{proj}_{\mathcal{C}(\mathbf{A})}\left(\boldsymbol{b}\right)$$

Definition 2.1 (Idempotent)**.** A matrix $\mathbf{P}$ is idempotent if it satisfies

$$\mathbf{P}^2 = \mathbf{P}$$

Theorem 2.2.1. *Given that $\mathbf{P}\boldsymbol{b}$ produces the projection vector $\boldsymbol{b}_P$, taking the orthogonal projection of $\boldsymbol{b}_P$ again results in the vector $\boldsymbol{b}_P$.*

$$\begin{aligned}\mathbf{P}\boldsymbol{b}_P &= \boldsymbol{b}_P \\ \mathbf{P}\mathbf{P}\boldsymbol{b} &= \mathbf{P}\boldsymbol{b} \\ \mathbf{P}^2\boldsymbol{b} &= \mathbf{P}\boldsymbol{b}\end{aligned}$$

Therefore orthogonal projectors are idempotent matrices as $\mathbf{P}^2 = \mathbf{P}$.

Proof. If we express the orthogonal projector in its full form, we can verify the result from above

$$\begin{aligned}\mathbf{P}^2 &= \mathbf{P}\mathbf{P} \\ &= \mathbf{A}\left(\mathbf{A}^\top \mathbf{A}\right)^{-1} \mathbf{A}^\top \mathbf{A}\left(\mathbf{A}^\top \mathbf{A}\right)^{-1} \mathbf{A}^\top \\ &= \mathbf{A}\cancel{\left(\mathbf{A}^\top \mathbf{A}\right)^{-1} \mathbf{A}^\top \mathbf{A}}\left(\mathbf{A}^\top \mathbf{A}\right)^{-1} \mathbf{A}^\top \\ &= \mathbf{A}\left(\mathbf{A}^\top \mathbf{A}\right)^{-1} \mathbf{A}^\top \\ &= \mathbf{P}\end{aligned}$$

□

2.3 Dependent Columns

If $\mathbf{A}$ has dependent columns, then we will obtain an infinite family of Least Squares solutions. However there will only be one projection vector $\boldsymbol{b}_P$.
This arises from $\mathcal{N}\left(\mathbf{A}\right) = \mathcal{N}\left(\mathbf{A}^\top \mathbf{A}\right)$, so that the solution to the Normal Equations yields

$$\boldsymbol{x} = \boldsymbol{x}_p + \boldsymbol{x}_n$$

where $\boldsymbol{x}_p$ is the particular Least Squares solution, and $\boldsymbol{x}_n$ represents any linear combination of null space vectors.

2.4 Orthogonal Projector onto the Column Space

To obtain a unique solution to the Normal Equations, we can form the matrix $\mathbf{W}$ such that it has full column rank and it spans the column space of $\mathbf{A}$. Then the orthogonal projector is given by

$$\mathbf{P} = \mathbf{W}\left(\mathbf{W}^\top\mathbf{W}\right)^{-1}\mathbf{W}^\top$$

2.5 Orthogonal Projectors onto other Spaces

Given the orthogonal projector $\mathbf{P} = \mathrm{proj}_V$ onto a subspace V, the orthogonal projector $\mathbf{Q} = \mathrm{proj}_{V^\perp}$ onto $V^\perp$ is given by

$$\mathbf{Q} = \mathbf{I} - \mathbf{P}$$

This is because the vector $\boldsymbol{b}$ can be represented as the sum of the projections onto V and $V^\perp$

$$\boldsymbol{b} = \mathbf{P}\boldsymbol{b} + \mathbf{Q}\boldsymbol{b}.$$

The dot product of these two vectors is therefore also zero, given that both projections lie in orthogonal subspaces.

$$\left(\mathbf{P}\boldsymbol{b}\right)^\top\mathbf{Q}\boldsymbol{b} = \mathbf{0}$$

or

$$\mathbf{P}\mathbf{Q} = \mathbf{0}$$

where $\mathbf{0}$ is the zero matrix.

3 Orthogonal Matrices

3.1 Standard Basis Vectors

The standard basis vectors are constructed by placing a 1 in the nth row of the nth basis vector in $\mathbb{R}^n$.

$$\boldsymbol{e}_1 = \begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix},\ \dots,\ \boldsymbol{e}_n = \begin{bmatrix}0\\0\\\vdots\\1\end{bmatrix}$$

3.2 Standard Basis

In $\mathbb{R}^n$, the standard basis S consists of the vectors

$$S = \{\boldsymbol{e}_1,\ \dots,\ \boldsymbol{e}_n\}$$

where the coefficients of a column vector implicitly represent the coefficients of the linear combination of basis vectors in S.
For example, the vector $\boldsymbol{b}$ in the standard basis

$$\boldsymbol{b} = b_1\boldsymbol{e}_1 + \cdots + b_n\boldsymbol{e}_n$$

can be explicitly represented as a column vector

$$(\boldsymbol{b})_S = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

where $b_1, \dots, b_n$ are the entries of $(\boldsymbol{b})_S$ with respect to the standard basis S.

3.3 Change of Basis

To represent the vector $\boldsymbol{b}$ in terms of another basis W, where

$$W = \{\boldsymbol{w}_1, \dots, \boldsymbol{w}_n\}$$

we must determine the components of each basis vector in W that contributes to $\boldsymbol{b}$. This is represented by the system:

$$\begin{aligned} \boldsymbol{b} &= c_1\boldsymbol{w}_1 + \cdots + c_n\boldsymbol{w}_n \\ \boldsymbol{b} &= \begin{bmatrix} | & & | \\ \boldsymbol{w}_1 & \cdots & \boldsymbol{w}_n \\ | & & | \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \\ \boldsymbol{b} &= \mathbf{W}\boldsymbol{c} \end{aligned}$$

The component vector $\boldsymbol{c}$ is precisely the representation of $\boldsymbol{b}$ with respect to the basis W, so that

$$(\boldsymbol{b})_W = \boldsymbol{c}.$$

3.4 Orthonormal Basis

If we consider a basis Q, such that every basis vector is normalised and orthogonal to every other basis vector, then Q is an orthonormal basis.

Definition 3.1 (Kronecker delta)**.** We can summarise such a basis using the Kronecker delta δ_{ij}, which is defined

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

so that in an orthonormal basis Q with basis vectors $\{\boldsymbol{q}_1, \dots, \boldsymbol{q}_n\}$

$$\boldsymbol{q}_i^\top \boldsymbol{q}_j = \delta_{ij}$$

By using an orthonormal basis, we can determine the coefficients of $(\boldsymbol{b})_Q$ without the need to solve a linear system. For example, the ith coefficient can be determined as follows

$$\begin{aligned} \boldsymbol{b} &= c_1\boldsymbol{q}_1 + \cdots + c_n\boldsymbol{q}_n \\ \boldsymbol{q}_i^\top\boldsymbol{b} &= \boldsymbol{q}_i^\top\left(c_1\boldsymbol{q}_1 + \cdots + c_i\boldsymbol{q}_i + \cdots + c_n\boldsymbol{q}_n\right) \\ \boldsymbol{q}_i^\top\boldsymbol{b} &= c_1\cancelto{0}{\boldsymbol{q}_i^\top\boldsymbol{q}_1} + \cdots + c_i\cancelto{1}{\boldsymbol{q}_i^\top\boldsymbol{q}_i} + \cdots + c_n\cancelto{0}{\boldsymbol{q}_i^\top\boldsymbol{q}_n} \\ \boldsymbol{q}_i^\top\boldsymbol{b} &= c_i \\ \mathbf{Q}^\top\boldsymbol{b} &= \boldsymbol{c} \end{aligned}$$

Therefore

$$(\boldsymbol{b})_Q = \mathbf{Q}^\top \boldsymbol{b}$$

3.5 Orthogonal Matrices

To confirm that $Q = \{\boldsymbol{q}_1, \dots, \boldsymbol{q}_n\}$ is an orthonormal basis, we need to confirm that $\boldsymbol{q}_i^\top \boldsymbol{q}_j = \delta_{ij}$ holds for all basis vectors $\boldsymbol{q}_i$. To do so, we can construct the matrix product $\mathbf{Q}^\top \mathbf{Q}$, so that

$$\begin{aligned}
\mathbf{Q}^\top \mathbf{Q} &= \begin{bmatrix} — & \boldsymbol{q}_1 & — \\ & \vdots & \\ — & \boldsymbol{q}_n & — \end{bmatrix} \begin{bmatrix} | & & | \\ \boldsymbol{q}_1 & \cdots & \boldsymbol{q}_n \\ | & & | \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{q}_1^\top \boldsymbol{q}_1 & \cdots & \boldsymbol{q}_1^\top \boldsymbol{q}_n \\ \vdots & \ddots & \vdots \\ \boldsymbol{q}_n^\top \boldsymbol{q}_1 & \cdots & \boldsymbol{q}_n^\top \boldsymbol{q}_n \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} \\
&= \mathbf{I}.
\end{aligned}$$

Definition 3.2 (Orthogonal Matrices)**.** A square matrix $\mathbf{Q}$ is *orthogonal* if it satisfies

$$\mathbf{Q}^\top = \mathbf{Q}^{-1}$$

so that

$$\mathbf{Q}^\top \mathbf{Q} = \mathbf{Q}\mathbf{Q}^\top = \mathbf{I}.$$

3.6 Projection onto a Vector

Consider a subspace which consists of one vector $\boldsymbol{a}$, the projection of $\boldsymbol{b}$ onto $\boldsymbol{a}$ is given by

$$\begin{aligned}
\operatorname{proj}_{\boldsymbol{a}}(\boldsymbol{b}) = \boldsymbol{b}_P &= \boldsymbol{a}\left(\boldsymbol{a}^\top \boldsymbol{a}\right)^{-1} \boldsymbol{a}^\top \boldsymbol{b} \\
&= \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|^2} \boldsymbol{a} \cdot \boldsymbol{b}
\end{aligned}$$

If we instead project $\boldsymbol{b}$ onto a unit vector $\boldsymbol{q}$, then this simplifies to

$$\operatorname{proj}_{\boldsymbol{q}}(\boldsymbol{b}) = \boldsymbol{q}\left(\boldsymbol{q} \cdot \boldsymbol{b}\right)$$

3.7 Gram-Schmidt Process

To convert an arbitrary basis W[1] to an orthonormal basis Q, we must develop a process that is generalisable to any basis.

[1] Here we use W to denote the set of vectors that span $\mathcal{C}(\mathbf{A})$.

We construct Q as follows

$$\begin{aligned} \boldsymbol{v}_1 &= \boldsymbol{w}_1 & \boldsymbol{q}_1 &= \frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|} \\ \boldsymbol{v}_2 &= \boldsymbol{w}_2 - \operatorname{proj}_{\boldsymbol{q}_1}(\boldsymbol{w}_2) & \boldsymbol{q}_2 &= \frac{\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|} \\ \boldsymbol{v}_3 &= \boldsymbol{w}_3 - \operatorname{proj}_{\boldsymbol{q}_1}(\boldsymbol{w}_3) - \operatorname{proj}_{\boldsymbol{q}_2}(\boldsymbol{w}_3) & \boldsymbol{q}_3 &= \frac{\boldsymbol{v}_3}{\|\boldsymbol{v}_3\|} \\ &\vdots & &\vdots \\ \boldsymbol{v}_i &= \boldsymbol{w}_i - \sum_{j=1}^{i-1} \operatorname{proj}_{\boldsymbol{q}_j}(\boldsymbol{w}_i) & \boldsymbol{q}_i &= \frac{\boldsymbol{v}_i}{\|\boldsymbol{v}_i\|}. \end{aligned}$$

Using the projection definition for orthogonal vectors, we can simplify this to

$$\begin{aligned} \boldsymbol{v}_1 &= \boldsymbol{w}_1 & \boldsymbol{q}_1 &= \frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|} \\ \boldsymbol{v}_2 &= \boldsymbol{w}_2 - \boldsymbol{q}_1(\boldsymbol{q}_1 \cdot \boldsymbol{w}_2) & \boldsymbol{q}_2 &= \frac{\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|} \\ \boldsymbol{v}_3 &= \boldsymbol{w}_3 - \boldsymbol{q}_1(\boldsymbol{q}_1 \cdot \boldsymbol{w}_3) - \boldsymbol{q}_2(\boldsymbol{q}_2 \cdot \boldsymbol{w}_3) & \boldsymbol{q}_3 &= \frac{\boldsymbol{v}_3}{\|\boldsymbol{v}_3\|} \\ &\vdots & &\vdots \\ \boldsymbol{v}_i &= \boldsymbol{w}_i - \sum_{j=1}^{i-1} \boldsymbol{q}_j(\boldsymbol{q}_j \cdot \boldsymbol{w}_i) & \boldsymbol{q}_i &= \frac{\boldsymbol{v}_i}{\|\boldsymbol{v}_i\|}. \end{aligned}$$

This produces two sets of mutually orthogonal vectors that span the same space as W. Here $V = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ is an orthogonal basis, and $Q = \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n\}$ is an orthonormal basis.

3.8 QR Decomposition

If we rearrange the steps in the Gram-Schmidt process to solve for $\boldsymbol{w}_i$, we get

$$\begin{aligned} \boldsymbol{w}_1 &= \boldsymbol{q}_1\|\boldsymbol{v}_1\| \\ \boldsymbol{w}_2 &= \boldsymbol{q}_2\|\boldsymbol{v}_2\| + \boldsymbol{q}_1(\boldsymbol{q}_1 \cdot \boldsymbol{w}_2) \\ \boldsymbol{w}_3 &= \boldsymbol{q}_3\|\boldsymbol{v}_3\| + \boldsymbol{q}_1(\boldsymbol{q}_1 \cdot \boldsymbol{w}_3) + \boldsymbol{q}_2(\boldsymbol{q}_2 \cdot \boldsymbol{w}_3) \end{aligned}$$

$$\boldsymbol{w}_i = \boldsymbol{q}_i\|\boldsymbol{v}_i\| + \sum_{j=1}^{i-1} \boldsymbol{q}_j(\boldsymbol{q}_j \cdot \boldsymbol{w}_i)$$

Due to the properties of $\boldsymbol{q}_i$, we can also express $\boldsymbol{w}_i$ as

$$\boldsymbol{w}_i = \sum_{j=1}^{i} \boldsymbol{q}_j(\boldsymbol{q}_j \cdot \boldsymbol{w}_i).$$

And in vector form:

$$\begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_n \end{bmatrix} \begin{bmatrix} \|\boldsymbol{v}_1\| & \boldsymbol{q}_1 \cdot \boldsymbol{w}_2 & \cdots & \boldsymbol{q}_1 \cdot \boldsymbol{w}_n \\ 0 & \|\boldsymbol{v}_2\| & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{q}_{n-1} \cdot \boldsymbol{w}_n \\ 0 & \cdots & 0 & \|\boldsymbol{v}_n\| \end{bmatrix}$$
$$\mathbf{W} = \mathbf{QR}$$

3.9 Least Squares using QR Decomposition

Consider $\boldsymbol{u} \notin \mathcal{C}(\mathbf{A})$, by using the QR decomposition of $\mathbf{A}$, we have

$$\begin{aligned} \mathbf{A}\boldsymbol{c} &= \boldsymbol{u} \\ (\mathbf{QR})\,\boldsymbol{c} &= \boldsymbol{u} \\ \mathbf{Q}^\top\mathbf{QR}\boldsymbol{c} &= \mathbf{Q}^\top\boldsymbol{u} \\ \mathbf{R}\boldsymbol{c} &= \mathbf{Q}^\top\boldsymbol{u}. \end{aligned}$$

Alternatively, consider the Normal Equations,

$$\begin{aligned} \mathbf{A}^\top\mathbf{A}\boldsymbol{c} &= \mathbf{A}^\top\boldsymbol{u} \\ (\mathbf{QR})^\top(\mathbf{QR})\,\boldsymbol{c} &= (\mathbf{QR})^\top\boldsymbol{u} \\ \mathbf{R}^\top\mathbf{Q}^\top\mathbf{QR}\boldsymbol{c} &= \mathbf{R}^\top\mathbf{Q}^\top\boldsymbol{u} \\ \mathbf{R}^\top\mathbf{R}\boldsymbol{c} &= \mathbf{R}^\top\mathbf{Q}^\top\boldsymbol{u} \\ \left(\mathbf{R}^\top\right)^{-1}\mathbf{R}^\top\mathbf{R}\boldsymbol{c} &= \left(\mathbf{R}^\top\right)^{-1}\mathbf{R}^\top\mathbf{Q}^\top\boldsymbol{u} \\ \mathbf{R}\boldsymbol{c} &= \mathbf{Q}^\top\boldsymbol{u} \end{aligned}$$

Therefore by using the QR decomposition of $\mathbf{A}$ we can find the Least Squares solution $\boldsymbol{c}^*$, using backward-substitution.

4 Eigenvalues and Eigenvectors

4.1 Linear Operators

Consider the linear transformation T that maps any vector $\boldsymbol{v} \in V$ to V. T is referred to as a linear operator from V to V:

$$T : V \to V.$$

4.2 The Eigenvalue Problem

Given the linear transformation

$$\boldsymbol{v} \mapsto \mathbf{A}\boldsymbol{v}$$

for $\mathbf{A} \in \mathbb{R}^{n\times n}$, let $\boldsymbol{v}$ be an eigenvector of $\mathbf{A}$ and λ its associated eigenvalue so that the following relationship is satisfied,

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

These eigenvalue and eigenvector pairs form eigenpairs of $\mathbf{A}$.

4.3 Calculating Eigenvalues

By rearranging the eigenvalue problem we have,

$$\begin{aligned}\mathbf{A}\boldsymbol{v} &= \lambda\boldsymbol{v}\\ \mathbf{A}\boldsymbol{v} &= \lambda\mathbf{I}\boldsymbol{v}\\ \lambda\mathbf{I}\boldsymbol{v} - \mathbf{A}\boldsymbol{v} &= \mathbf{0}\\ (\lambda\mathbf{I} - \mathbf{A})\,\boldsymbol{v} &= \mathbf{0}.\end{aligned}$$

For this homogeneous linear system to have non-trivial solutions, the rank of the coefficient matrix $\lambda\mathbf{I} - \mathbf{A}$ must be less than n. This requires the matrix to be singular so that its determinant is equal to 0:

$$\det(\lambda\mathbf{I} - \mathbf{A}) = 0.$$

Definition 4.1 (Characteristic polynomial)**.** Let $P(\lambda)$ denote the resulting polynomial of degree n:

$$P(\lambda) = \det(\lambda\mathbf{I} - \mathbf{A}).$$

We use $\lambda\mathbf{I} - \mathbf{A}$ rather than $\mathbf{A} - \lambda\mathbf{I}$ to ensure that the polynomial is monic, i.e. the leading coefficient is equal to 1.

4.4 Calculating Eigenvectors

To calculate the eigenvector $\boldsymbol{v}_i$ associated with the eigenvalue λ_i, we must find the null space of $\lambda_i\mathbf{I} - \mathbf{A}$. As this yields a basis for the null space, any scalar multiple of an eigenvalue also satisfies the eigenvalue problem.

Definition 4.2 (Eigenspace)**.** The null space of $\lambda_i\mathbf{I} - \mathbf{A}$ is called the eigenspace associated with λ_i.

4.5 Eigen Decomposition

The equations formed by all eigenpairs of the matrix $\mathbf{A}$ can be expressed as

$$\mathbf{AV} = \mathbf{VD}$$

which gives us the eigen decomposition or diagonalisation of $\mathbf{A}$,

$$\mathbf{A} = \mathbf{VDV}^{-1}.$$

Here $\mathbf{V}$ is a matrix comprised of the eigenvectors $\boldsymbol{v}_i$ as its columns, and $\mathbf{D}$ is a diagonal matrix of the eigenvalues λ_i. Mathematically,

$$\mathbf{V} = \begin{bmatrix}\boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n\end{bmatrix} \qquad \mathbf{D} = \operatorname{diag}(\lambda_1,\, \dots,\, \lambda_n)$$

Definition 4.3 (Algebraic multiplicity)**.** The algebraic multiplicity $\mu_{\mathbf{A}}(\lambda_i)$ of an eigenvalue λ_i is the multiplicity of λ_i in the characteristic polynomial (i.e., how many times it is repeated).
For $d \leq n$ distinct eigenvalues,

$$P(\lambda) = 0 = (\lambda - \lambda_1)^{\mu_{\mathbf{A}}(\lambda_1)}\,(\lambda - \lambda_2)^{\mu_{\mathbf{A}}(\lambda_2)} \cdots (\lambda - \lambda_d)^{\mu_{\mathbf{A}}(\lambda_d)}.$$

This requires

$$1 \leq \mu_{\mathbf{A}}(\lambda_i) \leq n$$
$$\mu_{\mathbf{A}} = \sum_{i=1}^{d} \mu_{\mathbf{A}}(\lambda_i) = n$$

Definition 4.4 (Geometric multiplicity)**.** The dimension of the eigenspace associated with λ_i is referred to as the geometric multiplicity of λ_i, denoted $\gamma_{\mathbf{A}}(\lambda_i)$.
As the eigenspace is by definition the null space of $\lambda_i \mathbf{I} - \mathbf{A}$,

$$\gamma_{\mathbf{A}}(\lambda_i) = \operatorname{nullity}(\lambda_i \mathbf{I} - \mathbf{A}).$$

Given $d \leq n$ distinct eigenvalues,

$$1 \leq \gamma_{\mathbf{A}}(\lambda_i) \leq \mu_{\mathbf{A}}(\lambda_i) \leq n$$
$$\gamma_{\mathbf{A}} = \sum_{i=1}^{d} \gamma_{\mathbf{A}}(\lambda_i)$$

so that

$$d \leq \gamma_{\mathbf{A}} \leq n.$$

Theorem 4.5.1 (Eigenvectors of distinct eigenvalues)**.** *Eigenvectors corresponding to distinct eigenvalues are linearly dependent.*

Definition 4.5 (Defective matrix)**.** A defective matrix is a matrix that does not have a complete basis of eigenvectors and is therefore not diagonalisable. In particular, there exists at least one eigenvalue λ_k, where $\gamma_{\mathbf{A}}(\lambda_k) < \mu_{\mathbf{A}}(\lambda_k)$

4.6 Matrix Similarity

Definition 4.6 (Similar matrices)**.** Two $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$ are similar, if there exists an invertible $n \times n$ matrix $\mathbf{P}$ such that

$$\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}.$$

Theorem 4.6.1 (Characteristic polynomial of similar matrices)**.** *Two similar matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *have the same characteristic polynomial* $P(\lambda)$ *so that they share their ranks, determinants, trace, and eigenvalues (including their algebraic and geometric multiplicities).*

Proof. The eigenvalues of $\mathbf{B}$ can be determined using the characteristic polynomial:

$$\begin{aligned}
P_{\mathbf{B}}(\lambda) &= \det(\lambda \mathbf{I} - \mathbf{B}) \\
&= \det(\lambda \mathbf{I} - \mathbf{P}^{-1}\mathbf{A}\mathbf{P}) \\
&= \det(\lambda \mathbf{P}^{-1}\mathbf{I}\mathbf{P} - \mathbf{P}^{-1}\mathbf{A}\mathbf{P}) \\
&= \det\left(\mathbf{P}^{-1}(\lambda \mathbf{I} - \mathbf{A})\mathbf{P}\right) \\
&= \det\left(\mathbf{P}^{-1}\right)\det(\lambda \mathbf{I} - \mathbf{A})\det(\mathbf{P}) \\
&= \det\left(\mathbf{P}^{-1}\mathbf{P}\right)\det(\lambda \mathbf{I} - \mathbf{A}) \\
&= \det(\mathbf{I})\det(\lambda \mathbf{I} - \mathbf{A}) \\
&= \det(\lambda \mathbf{I} - \mathbf{A}) \\
&= P_{\mathbf{A}}(\lambda)
\end{aligned}$$

□

4.7 Constructing a Similar Matrix

Given a diagonalisable matrix $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$, we can construct the matrix $\mathbf{B}$ with arbitrary eigenvectors $\mathbf{W}$, so that

$$\mathbf{B} = \mathbf{W}\mathbf{D}\mathbf{W}^{-1}.$$

Using $\mathbf{D} = \mathbf{V}^{-1}\mathbf{A}\mathbf{V}$,

$$\begin{aligned}\mathbf{B} &= \mathbf{W}\mathbf{D}\mathbf{W}^{-1} \\ &= \mathbf{W}\left(\mathbf{V}^{-1}\mathbf{A}\mathbf{V}\right)\mathbf{W}^{-1} \\ &= \left(\mathbf{V}\mathbf{W}^{-1}\right)^{-1}\mathbf{A}\mathbf{V}\mathbf{W}^{-1}.\end{aligned}$$

Let $\mathbf{P} = \mathbf{V}\mathbf{W}^{-1}$, so that

$$\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}.$$

This result shows us that $\mathbf{A}$ and $\mathbf{B}$ are similar matrices.

4.8 Symmetric Matrices

Consider the symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ where $\mathbf{A}^\top = \mathbf{A}$. Then:

1. $\mathbf{A}$ is always diagonalisable.
2. The eigenvalues and eigenvectors of symmetric matrices are always real.
3. The eigenspaces of a symmetric matrix are orthogonal for $\lambda_i \neq \lambda_j$.

4.9 Orthogonal Matrices

By applying the Gram-Schmidt process to the eigenvectors of a symmetric matrix, we obtain the decomposition

$$\mathbf{A} = \mathbf{Q}\mathbf{D}\mathbf{Q}^\top$$

where $\mathbf{Q}$ is a diagonal matrix that satisfies $\mathbf{Q}^{-1} = \mathbf{Q}^\top$.

4.10 Definite Matrices

Symmetric matrices are also known as definite matrices, and can be classified into four categories.

1. Positive definite matrices: All eigenvalues are positive.
2. Positive semidefinite matrices: All eigenvalues are nonnegative.
3. Negative definite matrices: All eigenvalues are nonpositive.
4. Negative semidefinite matrices: All eigenvalues are negative.

4.11 Eigenbases

When $\mathbf{A}$ has n linearly independent eigenvectors, the columns of $\mathbf{V}$ represent the eigenbasis of $\mathbf{A}$. Let the basis V be the columns of $\mathbf{V}$ so that for the linear system $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$, $[\boldsymbol{x}]_V = \boldsymbol{c}$ and $[\boldsymbol{b}]_V = \boldsymbol{k}$, then

$$\begin{aligned} \mathbf{A}\boldsymbol{x} &= \boldsymbol{b} \\ \mathbf{AV}\boldsymbol{c} &= \mathbf{V}\boldsymbol{k} \\ \mathbf{V}^{-1}\mathbf{AV}\boldsymbol{c} &= \boldsymbol{k} \\ \mathbf{D}\boldsymbol{c} &= \boldsymbol{k} \\ \mathbf{D}\left(\boldsymbol{x}\right)_V &= \left(\boldsymbol{b}\right)_V \end{aligned}$$

so that, with respect to the eigenbasis, a linear map from $\mathbb{R}^n$ to $\mathbb{R}^n$ is represented by the matrix $\mathbf{D}$.

4.12 Matrix Functions

Given a nondefective $\mathbf{A} \in \mathbb{R}^{n\times n}$ with the eigendecomposition $\mathbf{A} = \mathbf{VDV}^{-1}$, we can form the following equalities.

4.12.1 Powers

For $k \in \mathbb{N}$,

$$\mathbf{A}^k = \mathbf{VD}^k\mathbf{V}^{-1} = \mathbf{V}\operatorname{diag}\left(\lambda_1^k,\ \ldots,\ \lambda_n^k\right)\mathbf{V}^{-1}.$$

Proof. We can prove the above theorem using a recursive construction of $\mathbf{A}^k$.

$$\begin{aligned} \mathbf{A}^2 &= \mathbf{AA} = \mathbf{VD}\cancel{\mathbf{V}^{-1}\mathbf{V}}\mathbf{DV}^{-1} = \mathbf{VD}^2\mathbf{V}^{-1} \\ \mathbf{A}^3 &= \mathbf{A}^2\mathbf{A} = \mathbf{VD}^2\cancel{\mathbf{V}^{-1}\mathbf{V}}\mathbf{DV}^{-1} = \mathbf{VD}^3\mathbf{V}^{-1} \end{aligned}$$

so that for a positive integer $k \in \mathbb{N}$,

$$\mathbf{A}^k = \mathbf{A}^{k-1}\mathbf{A} = \mathbf{VD}^{k-1}\cancel{\mathbf{V}^{-1}\mathbf{V}}\mathbf{DV}^{-1} = \mathbf{VD}^k\mathbf{V}^{-1}.$$

□

4.12.2 Polynomials

Let $p\left(t\right) = c_0 + c_1 t + c_2 t^2 + \cdots + c_k t^k$ denote a polynomial of degree k, then the matrix polynomial $p\left(\mathbf{A}\right)$ can be expressed as

$$\begin{aligned} p\left(\mathbf{A}\right) &= c_0\mathbf{I} + c_1\mathbf{A} + c_2\mathbf{A}^2 + \cdots + c_k\mathbf{A}^k \\ &= c_0\mathbf{VIV}^{-1} + c_1\mathbf{VDV}^{-1} + c_2\mathbf{VD}^2\mathbf{V}^{-1} + \cdots + c_k\mathbf{VD}^k\mathbf{V}^{-1} \\ &= \mathbf{V}\left(c_0\mathbf{I} + c_1\mathbf{D} + c_2\mathbf{D}^2 + \cdots + c_k\mathbf{D}^k\right)\mathbf{V}^{-1} \\ &= \mathbf{V}p\left(\mathbf{D}\right)\mathbf{V}^{-1} \\ &= \mathbf{V}\operatorname{diag}\left(p\left(\lambda_1\right),\ \ldots,\ p\left(\lambda_n\right)\right)\mathbf{V}^{-1}. \end{aligned}$$

4.13 Analytic Functions

An analytic function can be represented by its Taylor series expansion, allowing us to operate analytic functions on $\mathbf{A}$.

$$\begin{aligned} f(\mathbf{A}) &= \mathbf{V} f(\mathbf{D}) \mathbf{V}^{-1} \\ &= \mathbf{V} \operatorname{diag}(f(\lambda_1), \ldots, f(\lambda_n)) \mathbf{V}^{-1}. \end{aligned}$$

Theorem 4.13.1 (Cayley-Hamilton theorem)**.** *The characteristic polynomial of a square matrix* $\mathbf{A}$ *(that is not necessarily diagonalisable) is equal to the zero matrix.*

$$P(\mathbf{A}) = \mathbf{0}$$

5 Singular Value Decomposition

The singular value decomposition (SVD) of $\mathbf{A} \in \mathbb{R}^{m \times n}$ is given by

$$\mathbf{A} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^\top$$

where $\mathbf{U} \in \mathbb{R}^{n \times n}$ is an orthogonal matrix, $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times n}$ is a diagonal matrix, and $\mathbf{V} \in \mathbb{R}^{m \times m}$ is an orthogonal matrix. $\mathbf{U}$ is known as the left-singular matrix and $\mathbf{V}$ is the right-singular matrix, corresponding to how these matrices are multiplied to $\boldsymbol{\Sigma}$.
$\boldsymbol{\Sigma}$ consists of the **singular values** σ_i of $\mathbf{A}$, which can be determined using the following process:

$$\begin{aligned} \mathbf{A}^\top \mathbf{A} &= \left(\mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^\top\right)^\top \left(\mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^\top\right) \\ &= \mathbf{V} \boldsymbol{\Sigma}^\top \mathbf{U}^\top \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^\top \\ &= \mathbf{V} \boldsymbol{\Sigma}^\top \boldsymbol{\Sigma} \mathbf{V}^\top \end{aligned}$$

similarly,

$$\begin{aligned} \mathbf{A} \mathbf{A}^\top &= \left(\mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^\top\right) \left(\mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^\top\right)^\top \\ &= \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^\top \mathbf{V} \boldsymbol{\Sigma}^\top \mathbf{U}^\top \\ &= \mathbf{U} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top \mathbf{U}^\top \end{aligned}$$

In both instances, we form an orthogonal eigendecomposition where $\boldsymbol{\Sigma}^\top \boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top$ are the eigenvalues of $\mathbf{A}^\top \mathbf{A}$ and $\mathbf{A} \mathbf{A}^\top$, respectively. But because the eigenvalues of $\mathbf{A}^\top \mathbf{A}$ and $\mathbf{A} \mathbf{A}^\top$ are equal, $\boldsymbol{\Sigma}^\top \boldsymbol{\Sigma} = \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top$.
The singular values are non-negative constants and the entries of $\boldsymbol{\Sigma}$ are always non-increasing: $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ where r is the rank of $\mathbf{A}$.

5.1 Orthonormal Bases for the Fundamental Subspaces

For $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $\operatorname{rank}(\mathbf{A}) = r \leq n$:

$$\mathbf{A} \mathbf{V} = \mathbf{U} \boldsymbol{\Sigma}$$

with

$$\mathbf{A}\mathbf{V} = \begin{bmatrix} | & & | & | & & | \\ \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_r & \boldsymbol{a}_{r+1} & \cdots & \boldsymbol{a}_n \\ | & & | & | & & | \end{bmatrix} \begin{bmatrix} | & & | & | & & | \\ \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_r & \boldsymbol{v}_{r+1} & \cdots & \boldsymbol{v}_n \\ | & & | & | & & | \end{bmatrix}$$

$$\mathbf{U}\boldsymbol{\Sigma} = \begin{bmatrix} | & & | & | & & | \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_r & \boldsymbol{u}_{r+1} & \cdots & \boldsymbol{u}_m \\ | & & | & | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \\ & & & & & \\ & & & & & \end{bmatrix} \begin{array}{l} \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} r \text{ rows} \\ \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} n - r \text{ rows} \\ \left.\vphantom{\begin{matrix} a \\ a \end{matrix}}\right\} m - n \text{ rows} \end{array}$$

where we have r singular values because $\operatorname{rank} \mathbf{A} = \operatorname{rank} \mathbf{A}^\top \mathbf{A}$. If we express each equation separately, then

$$\begin{aligned} \mathbf{A}\boldsymbol{v}_1 &= \sigma_1 \boldsymbol{u}_1 \\ &\vdots \\ \mathbf{A}\boldsymbol{v}_r &= \sigma_2 \boldsymbol{u}_r \\ \mathbf{A}\boldsymbol{v}_{r+1} &= \mathbf{0} \\ &\vdots \\ \mathbf{A}\boldsymbol{v}_n &= \mathbf{0} \end{aligned}$$

which shows us that $\{\boldsymbol{v}_{r+1}, \dots, \boldsymbol{v}_n\}$ forms a basis for the null space of $\mathbf{A}$, requiring the remaining columns of $\mathbf{V}$ to form a basis for the row space of $\mathbf{A}$.
This means that $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_r\}$ also forms a basis for the column space of $\mathbf{A}$, and hence, the remaining columns of $\mathbf{U}$ form a basis for the left-null space of $\mathbf{A}$.
To summarise:

$$\begin{aligned} \mathcal{C}\left(\mathbf{A}^\top\right) &= \operatorname{span}\left(\{\boldsymbol{v}_{i \le r}\}\right) & \mathcal{C}\left(\mathbf{A}\right) &= \operatorname{span}\left(\{\boldsymbol{u}_{i \le r}\}\right) \\ \mathcal{N}\left(\mathbf{A}\right) &= \operatorname{span}\left(\{\boldsymbol{v}_{r < i \le n}\}\right) & \mathcal{N}\left(\mathbf{A}^\top\right) &= \operatorname{span}\left(\{\boldsymbol{u}_{r < i \le m}\}\right) \end{aligned}$$

for $i \in \mathbb{N}$. Additionally, $\mathbf{V}$ forms an orthonormal basis for $\mathbb{R}^n$ while $\mathbf{U}$ forms an orthonormal basis for $\mathbb{R}^m$.

5.2 Singular Bases

Consider the bases V and U from the columns of the orthogonal matrices $\mathbf{V}$ and $\mathbf{U}$, respectively. Let $(\boldsymbol{x})_V = \boldsymbol{c}$ and $(\boldsymbol{b})_U = \boldsymbol{k}$, so that the system $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ becomes

$$\begin{aligned} \mathbf{A}\boldsymbol{x} &= \boldsymbol{b} \\ \mathbf{A}\mathbf{V}\boldsymbol{c} &= \mathbf{U}\boldsymbol{k} \\ \mathbf{U}^\top \mathbf{A}\mathbf{V}\boldsymbol{c} &= \boldsymbol{k} \\ \boldsymbol{\Sigma}\boldsymbol{c} &= \boldsymbol{k} \\ \boldsymbol{\Sigma}\left(\boldsymbol{x}\right)_V &= \left(\boldsymbol{b}\right)_U . \end{aligned}$$

Therefore, with respect to the orthogonal bases V and U, the linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ is represented by the matrix $\mathbf{\Sigma}$.

5.3 Reduced SVD

By ignoring the additional $m-n$ rows in $\mathbf{\Sigma}$, we can form the reduced SVD which removes the additional "0" rows of $\mathbf{\Sigma}$. This results in $\mathbf{U} \in \mathbb{R}^{m\times n}$ and $\mathbf{\Sigma} \in \mathbb{R}^{n\times n}$.

$$\mathbf{U\Sigma} = \begin{bmatrix} | & & | & | & & | \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_r & \boldsymbol{u}_{r+1} & \cdots & \boldsymbol{u}_n \\ | & & | & | & & | \end{bmatrix} \begin{array}{l} \left[\begin{array}{ccccc} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{array}\right] \end{array} \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} r \text{ rows} \\ \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} n-r \text{ rows} \end{array}$$

The reduced SVD also removes the $m-n$ left-null space basis vectors $\{\boldsymbol{u}_{n<i\leq m}\}$ from $\mathbf{U}$.

5.4 Pseudoinverse

Using the orthogonal basis vectors obtained through the SVD of $\mathbf{A} \in \mathbb{R}^{m\times n}$, we can show that $\boldsymbol{v}_i \mapsto \sigma_i \boldsymbol{u}_i$ for all $i \leq r$. Therefore

$$\mathbf{A}\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i.$$

If we consider the inverse mapping $\boldsymbol{u}_i \mapsto \frac{1}{\sigma_i}\boldsymbol{v}_i$, then

$$\mathbf{A}^\dagger \boldsymbol{u}_i = \frac{1}{\sigma_i}\boldsymbol{v}_i$$

where $\mathbf{A}^\dagger$ is the pseudoinverse of $\mathbf{A}$. To determine $\mathbf{A}^\dagger$, the above relationship must hold for all $i \leq r$. If we multiply the RHS by $\boldsymbol{u}_i^\top \boldsymbol{u}_i$

$$\mathbf{A}^\dagger \boldsymbol{u}_i = \frac{1}{\sigma_i}\boldsymbol{v}_i \boldsymbol{u}_i^\top \boldsymbol{u}_i$$

we can show that $\mathbf{A}^\dagger$ takes the form $\frac{1}{\sigma_i}\boldsymbol{v}_i\boldsymbol{u}_i^\top$. Using the Kronecker Delta definition of orthonormal vectors, we know that the product of two orthonormal basis vectors is 1 for $i = j$ and 0 for $i \neq j$. Therefore by taking the sum of all $\frac{1}{\sigma_i}\boldsymbol{v}_i\boldsymbol{u}_i^\top$, we have

$$\begin{aligned} \mathbf{A}^\dagger \boldsymbol{u}_1 &= \left(\frac{1}{\sigma_1}\boldsymbol{v}_1\boldsymbol{u}_1^\top + \cdots + \frac{1}{\sigma_r}\boldsymbol{v}_r\boldsymbol{u}_r^\top\right)\boldsymbol{u}_1 = \frac{1}{\sigma_1}\boldsymbol{v}_1\cancelto{1}{\boldsymbol{u}_1^\top\boldsymbol{u}_1} + \cdots + \frac{1}{\sigma_r}\boldsymbol{v}_r\cancelto{0}{\boldsymbol{u}_r^\top\boldsymbol{u}_1} = \frac{1}{\sigma_1}\boldsymbol{v}_1 \\ &\vdots \\ \mathbf{A}^\dagger \boldsymbol{u}_r &= \left(\frac{1}{\sigma_1}\boldsymbol{v}_1\boldsymbol{u}_1^\top + \cdots + \frac{1}{\sigma_r}\boldsymbol{v}_r\boldsymbol{u}_r^\top\right)\boldsymbol{u}_r = \frac{1}{\sigma_1}\boldsymbol{v}_1\cancelto{0}{\boldsymbol{u}_1^\top\boldsymbol{u}_r} + \cdots + \frac{1}{\sigma_r}\boldsymbol{v}_r\cancelto{1}{\boldsymbol{u}_r^\top\boldsymbol{u}_r} = \frac{1}{\sigma_r}\boldsymbol{v}_r \end{aligned}$$

Therefore the pseudoinverse is given by

$$\mathbf{A}^\dagger = \sum_{i=1}^{r} \frac{1}{\sigma_i}\boldsymbol{v}_i\boldsymbol{u}_i^\top$$

which is equivalent to

$$\mathbf{A}^\dagger = \mathbf{V} \begin{bmatrix} \frac{1}{\sigma_1} & & & & & \\ & \ddots & & & & \\ & & \frac{1}{\sigma_r} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \mathbf{U}^\top$$

If we consider the SVD of $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \mathbf{U}_\Sigma \boldsymbol{\Sigma}_\Sigma \mathbf{V}_\Sigma^\top.$$

we have $\mathbf{U}_\Sigma = \mathbf{I}_m$, $\boldsymbol{\Sigma}_\Sigma = \boldsymbol{\Sigma}$, and $\mathbf{V}_\Sigma = \mathbf{I}_n$. Then the pseudoinverse of this matrix is then

$$\begin{aligned} \boldsymbol{\Sigma}^\dagger &= \sum_{i=1}^{r} \frac{1}{\sigma_i} \boldsymbol{v}_{\Sigma,\, i} \boldsymbol{u}_{\Sigma,\, i}^\top \\ &= \begin{bmatrix} \frac{1}{\sigma_1} & & & & & \\ & \ddots & & & & \\ & & \frac{1}{\sigma_r} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \end{aligned}$$

so that the pseudoinverse of $\mathbf{A}$ can be determined using

$$\mathbf{A}^\dagger = \mathbf{V}\boldsymbol{\Sigma}^\dagger \mathbf{U}^\top.$$

Note that the pseudoinverse can also be obtained using the reduced SVD or by using the first r columns of $\mathbf{U}$ and $\mathbf{V}$ with the $r \times r$ submatrix of $\boldsymbol{\Sigma}$.

5.5 Truncated SVD

By expanding the SVD of $\mathbf{A}$, we can express it as the sum of rank-1 matrices:

$$\mathbf{A} = \sum_{i=1}^{n} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top.$$

However if $r < n$, then

$$\mathbf{A} = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$$

as $\sigma_{r<i\leq n} = 0$. As the singular values are ordered from largest to smallest, if we wished to approximate $\mathbf{A}$ by a matrix of lower rank, we can truncate this sum even further at $i = \nu$ for $\nu < r$, to generate a rank-ν approximation of $\mathbf{A}$:

$$\tilde{\mathbf{A}} = \sum_{i=1}^{\nu} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top.$$

This allows us to represent the matrix $\mathbf{A}$ using only ν singular values and 2ν singular vectors ($\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ for $i = 1 \ldots \nu$). This approximate decomposition is known as the truncated SVD.

5.6 Principal Component Analysis

To summarise, the SVD of $\mathbf{A}$ has three variations in which the dimensions of the decomposition matrices change.

- Full SVD:
$$\underset{m\times n}{\mathbf{A}} = \underset{m\times m}{\mathbf{U}} \underset{m\times n}{\mathbf{\Sigma}} \underset{n\times n}{\mathbf{V}}^\top$$
- Reduced SVD:
$$\underset{m\times n}{\mathbf{A}} = \underset{m\times n}{\mathbf{U}} \underset{n\times n}{\mathbf{\Sigma}} \underset{n\times n}{\mathbf{V}}^\top$$
- Truncated SVD:
$$\underset{m\times n}{\mathbf{A}} = \underset{m\times \nu}{\mathbf{U}} \underset{\nu\times \nu}{\mathbf{\Sigma}} \underset{n\times \nu}{\mathbf{V}}^\top$$

Consider the ith column of $\mathbf{A}$ in the truncated SVD,

$$\tilde{\mathbf{A}}_{:,\,i} = \mathbf{U\Sigma V}^\top_{i,\,:}$$

If we let $\boldsymbol{x}_i = \mathbf{\Sigma V}^\top_{i,\,:}$, then

$$\tilde{\mathbf{A}}_{:,\,i} = \mathbf{U}\boldsymbol{x}_i$$

so that $\boldsymbol{x}_i \in \mathbb{R}^\nu$ is the coordinate vector of $\tilde{\mathbf{A}}_{:,\,i}$ with respect to the basis of left singular vectors $\mathbf{U}_{:,\,i\le\nu}$. In statistics or machine learning, the columns of $\mathbf{U}$ are called "features" as they are the most important singular vectors used to construct $\mathbf{A}$. The vector $\boldsymbol{x}_i$ is then the coordinate vector of $\mathbf{A}$'s projection onto the "feature space".
In data analysis, this process is referred to as Principal Componenet Analysis (PCA) where the matrix $\mathbf{A}$ represents a set of observations in which each column contains a particular explanatory variable that one might be interested in. The model shown above can be used to explain the observations from each variable in that dataset.

6 General Vector Spaces

6.1 Vector Space Axioms

A set V of objects are called "vectors" if the following additive and multiplicative axioms are satisfied $\forall \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ and $\forall k, m \in \mathbb{R}$.

Axiom	Meaning
Closure under vector addition	$\boldsymbol{u}+\boldsymbol{v} \in V$
Commutativity of vector addition	$\boldsymbol{u}+\boldsymbol{v} = \boldsymbol{v}+\boldsymbol{u}$
Associativity of vector addition	$\boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w}) = (\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w}$
Identity element of vector addition	$\exists \mathbf{0} \in V : \boldsymbol{u}+\mathbf{0} = \mathbf{0}+\boldsymbol{u} = \boldsymbol{u}$
Inverse elements of vector addition	$\exists\,(-\boldsymbol{u}) \in V : \boldsymbol{u}+(-\boldsymbol{u}) = \mathbf{0}$

Table 1: Additive axioms.

Axiom	Meaning
Closure under scalar multiplication	$k\boldsymbol{u} \in V$
Distributivity of scalar multiplication with vector addition	$k(\boldsymbol{u}+\boldsymbol{v}) = k\boldsymbol{u}+k\boldsymbol{v}$
Distributivity of scalar multiplication with scalar addition	$(k+m)\boldsymbol{u} = k\boldsymbol{u}+m\boldsymbol{u}$
Associativity of scalar multiplication	$k(m\boldsymbol{u}) = (km)\boldsymbol{u}$
Identity element of scalar multiplication	$1\boldsymbol{u} = \boldsymbol{u}$

Table 2: Multiplicative axioms.

If these axioms are satisfied, the objects called "vectors" and the operators "addition" (+) and "scalar multiplication" (denoted by juxtaposition) form a general vector space V. It follows that the operations of addition and scalar multiplication need not resemble those in $\mathbb{R}^n$ to satisfy the 10 axioms described above. Rather any set with two operations can form a vector space if they satisfy the 10 axioms above.

6.2 Examples of Vector Spaces

Aside from the familiar vector space $\mathbb{R}^n$, we can also consider the following spaces which satisfy the 10 axioms above.

1. The set of $m \times n$ matrices $\mathscr{M}_{mn}$ with matrix addition and scalar multiplication.
2. The set of functions $\mathscr{F}(\Omega) : \Omega \to \mathbb{R}$ with addition and scalar multiplication defined pointwise.