Linear Algebra Cheat Sheet

Range/Span/Column Space

Denoted $\mathcal{R}(\boldsymbol{X})$ is the space or set of all possible linear combinations of the columns of $\boldsymbol{X}$. Thus $\boldsymbol{a} \in \mathcal{R}(\boldsymbol{X})$ if $\boldsymbol{a} = \boldsymbol{X}\boldsymbol{b}$ for some $\boldsymbol{b}$.

Kernel/Null space

Denoted $\mathcal{N}(\boldsymbol{X})$ is the space of all possible linear combinations of vectors orthogonal to the columns of $\boldsymbol{X}$.

Orthogonal complement

Let V be a finite dimensional vector space and W is a subspace of V. Then the orthogonal complement of W, denoted $W^\perp$, is the set of vectors

$$\{\boldsymbol{v} \in V : \boldsymbol{v}^\top \boldsymbol{w} = 0 \text{ for all } \boldsymbol{w} \in W\}.$$

- $W^\perp$ is also a subspace of V
- $(W^\perp)^\perp = W$
- $\dim(W^\perp) = \dim V - \dim W$
- $V = W \oplus W^\perp$

Idempotent matrix

A matrix $\boldsymbol{A}$ such that $\boldsymbol{A} = \boldsymbol{A}^2$ is called **idempotent**. An idempotent matrix $\boldsymbol{A}$ has the following properties:

- $\boldsymbol{A}$ is a square matrix
- the eigenvalues of $\boldsymbol{A}$ are either 0 or 1
- $\operatorname{tr}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A})$
- $\boldsymbol{A}$ is singular unless $\boldsymbol{A} = \boldsymbol{I}$, the identity matrix
- $\boldsymbol{I} - \boldsymbol{A}$ is also an idempotent matrix

Permutation matrix

A permutation matrix $\boldsymbol{P}$ is a square binary matrix that has exactly one entry of 1 in each row and each column and 0s elsewhere. The properties of a permutation matrix include

- Pre-multiplying by $\boldsymbol{P}$ will result in permutation of the rows.
- Post-multiplying by $\boldsymbol{P}$ will result in permutation of the columns.
- $\boldsymbol{P}$ is a orthogonal matrix.
- $\boldsymbol{P}^{-1} = \boldsymbol{P}^\top$.

Rotation matrix

A rotation matrix $\boldsymbol{R}$ is a matrix that is used to perform a rotation in Euclidean space. For example the matrix

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

rotates points in the xy-Cartesian plane (counter)-clockwise through an angle θ about the origin of the Cartesian coordinate system by (pre-) post-multiplying.

Projection matrix

For any $\boldsymbol{v} \in V$, there are unique vectors $\boldsymbol{x} \in W$ and $\boldsymbol{z} \in W^\perp$ such that $\boldsymbol{v} = \boldsymbol{x} + \boldsymbol{z}$. We call $\boldsymbol{x}$ the projection of $\boldsymbol{v}$ onto W and write $\boldsymbol{x} = \boldsymbol{P}_W(\boldsymbol{v})$. In fact, $\boldsymbol{x} = \boldsymbol{P}_W\boldsymbol{v}$ where $\boldsymbol{P}_W$ is the projection matrix that maps any $\boldsymbol{v} \in V$ onto W. Similarly $\boldsymbol{z}$ is the projection of $\boldsymbol{v}$ onto $W^\perp$ and the corresponding projection matrix is referred to as the orthogonal projection matrix, $\boldsymbol{P}_{W^\perp}$. Note $\boldsymbol{z} = \boldsymbol{v} - \boldsymbol{P}_W\boldsymbol{v} = (\boldsymbol{I} - \boldsymbol{P}_W)\boldsymbol{v}$. So $\boldsymbol{P}_{W^\perp} = \boldsymbol{I} - \boldsymbol{P}_W$.
The projection matrix $\boldsymbol{P}_W$ has the following basic properties:

- $\boldsymbol{P}_W$ is idempotent, i.e. $\boldsymbol{P}_W^2 = \boldsymbol{P}_W$,
- $\boldsymbol{P}_W\boldsymbol{x} = \boldsymbol{x}$ for all $\boldsymbol{x} \in W$ (i.e. $\boldsymbol{P}_W$ is the identity operator on W),
- Every vector $\boldsymbol{x} \in V$ may be decomposed uniquely as $\boldsymbol{v} = \boldsymbol{x} + \boldsymbol{z}$ with $\boldsymbol{x} = \boldsymbol{P}_W\boldsymbol{v}$ and $\boldsymbol{z} = \boldsymbol{P}_{W^\perp}\boldsymbol{v} = (\boldsymbol{I} - \boldsymbol{P}_W)\boldsymbol{v}$.
- There exists matrix $\boldsymbol{A}$ with columns corresponding to orthonormal basis of W such that $\boldsymbol{P}_W = \boldsymbol{A}\boldsymbol{A}^\top$.

Projection onto the range of a matrix

Suppose $W = \mathcal{R}(\boldsymbol{X})$. Suppose $\boldsymbol{P}_W$ is a projection matrix onto W and assume $\boldsymbol{X}$ is full-rank, then $\boldsymbol{P}_W = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$. Note there are, of course, other projection matrices onto W. Note that the $\operatorname{tr}(\boldsymbol{P}_W) = \operatorname{rank}(\boldsymbol{X})$.

Diagonal Matrix

- $\boldsymbol{D}\boldsymbol{A}$ (or $\boldsymbol{A}\boldsymbol{D}$) gives a matrix whose *rows* (or *columns*) are those of $\boldsymbol{A}$ multiplied by the respective diagonal elements of $\boldsymbol{D}$

Orthogonality

- $\boldsymbol{v}$ is orthogonal to $\boldsymbol{w}$ if $\boldsymbol{v}^\top\boldsymbol{w} = 0$ (written $\boldsymbol{v} \perp \boldsymbol{w}$).
- If unit vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ satisfy $\boldsymbol{u}^\top\boldsymbol{v} = 0$ then $\boldsymbol{u}$ and $\boldsymbol{v}$ are **orthonormal vectors**.
- If any two of the following conditions are satisfied then the matrix $\boldsymbol{A}$ is an **orthogonal matrix**:
  1. $\boldsymbol{A}$ is square
  2. $\boldsymbol{A}^\top\boldsymbol{A} = \boldsymbol{I}$
  3. $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I}$
  4. $|\boldsymbol{A}| = \pm 1$
- $\boldsymbol{A}^{-1} = \boldsymbol{A}^\top$

Traces

The **trace** of a square $n \times n$ matrix $\boldsymbol{A}$ is the sum of the diagonal elements and denoted $\operatorname{tr}(\boldsymbol{A})$. The trace is not defined for a matrix that is not a square. Let $\lambda_1, \lambda_2, ..., \lambda_n$ be the eigenvalues of $\boldsymbol{A}$.

- $\operatorname{tr}(\boldsymbol{A}) = \operatorname{tr}(\boldsymbol{A}^\top)$
- $\operatorname{tr}(\boldsymbol{A} + \boldsymbol{B}) = \operatorname{tr}(\boldsymbol{A}) + \operatorname{tr}(\boldsymbol{B})$
- $\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{B}\boldsymbol{A}) = \sum_{i=1}^r \sum_{j=1}^c a_{ij}b_{ji}$
- $\operatorname{tr}(\boldsymbol{A}\boldsymbol{A}^\top) = \operatorname{tr}(\boldsymbol{A}^\top\boldsymbol{A}) = \sum_{i=1}^r \sum_{j=1}^c a_{ij}^2$
- $\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}) = \operatorname{tr}(\boldsymbol{B}\boldsymbol{C}\boldsymbol{A}) = \operatorname{tr}(\boldsymbol{C}\boldsymbol{A}\boldsymbol{B})$
- $\boldsymbol{x}^\top\boldsymbol{A}\boldsymbol{y} = \operatorname{tr}(\boldsymbol{x}^\top\boldsymbol{A}\boldsymbol{y}) = \operatorname{tr}(\boldsymbol{A}\boldsymbol{y}\boldsymbol{x}^\top)$
- $\operatorname{tr}(\boldsymbol{A}) = \sum_i^n \lambda_i$
- $\operatorname{tr}\left(\boldsymbol{A}^k\right) = \sum_i^n \lambda_i^k$

Inverses

- The **inverse** of a square matrix $\boldsymbol{A}$ is the unique matrix $\boldsymbol{A}^{-1}$ such that $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$. Note that $\boldsymbol{A}^{-1}$ does not always exist.
- $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.
- $$\begin{bmatrix} \boldsymbol{R} & \boldsymbol{0} \\ \boldsymbol{X} & \boldsymbol{S} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{R}^{-1} & \boldsymbol{0} \\ -\boldsymbol{S}^{-1}\boldsymbol{X}\boldsymbol{R}^{-1} & \boldsymbol{S}^{-1} \end{bmatrix}$$
- $\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ then $\boldsymbol{A}^{-1} = \dfrac{1}{ab - cd}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

 emi.tanaka@sydney.edu.au

- Suppose we have the partitioned matrix $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}$ then $\boldsymbol{A}^{-1}$

$$\begin{bmatrix} (\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1} & -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}(\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1} \\ -\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1} & (\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1} \end{bmatrix}$$

$$\begin{bmatrix} (\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1} & -(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \\ -(\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & (\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1} \end{bmatrix}$$

Woodbury matrix identity

$$(\boldsymbol{A} + \boldsymbol{U}\boldsymbol{B}\boldsymbol{V})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{B}^{-1} + \boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}\boldsymbol{A}^{-1}.$$

Alternatively, [1] deal with

$$(\boldsymbol{A} + \boldsymbol{U}\boldsymbol{B}\boldsymbol{V})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{U}\boldsymbol{B}\boldsymbol{V}(\boldsymbol{I} + \boldsymbol{A}^{-1}\boldsymbol{U}\boldsymbol{B}\boldsymbol{V})^{-1}\boldsymbol{A}^{-1}.$$

Special case:

$$(\boldsymbol{A} + \boldsymbol{u}\boldsymbol{v}^\top)^{-1} = \boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^\top\boldsymbol{A}^{-1}}{1 + \boldsymbol{v}^\top\boldsymbol{A}^{-1}\boldsymbol{u}}$$

Generalised inverse

- For a given matrix $\boldsymbol{A} \in \mathbb{R}^{n\times m}$, if $\boldsymbol{A}^- \in \mathbb{R}^{m\times n}$ is such that
$$\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A} = \boldsymbol{A}$$
then $\boldsymbol{A}^-$ is a **generalised inverse** of $\boldsymbol{A}$. If $\boldsymbol{A}^-$ also satisfies
$$\boldsymbol{A}^-\boldsymbol{A}\boldsymbol{A}^- = \boldsymbol{A}^-$$
then $\boldsymbol{A}^-$ is a **generalised reflexive inverse** of $\boldsymbol{A}$.
If $\boldsymbol{A}^-$ satisfies the above two conditions and also
$$(\boldsymbol{A}\boldsymbol{A}^-)^\top = \boldsymbol{A}\boldsymbol{A}^- \quad \text{and} \quad (\boldsymbol{A}^-\boldsymbol{A})^\top = \boldsymbol{A}^-\boldsymbol{A},$$
then $\boldsymbol{A}^-$ is the **Moore-Penrose pseudoinverse** of $\boldsymbol{A}$.
- $\boldsymbol{A}^-$ is generally not unique (as opposed to $\boldsymbol{A}^{-1}$) although the Moore-Penrose pseudoinverse exists and unique for any matrix.
- For symmetric matrix $\boldsymbol{A}$, $\boldsymbol{G} = \boldsymbol{A}^-$ may not be symmetric although $\boldsymbol{G}^\top$ is still a generalised inverse.

Determinants

- $|\boldsymbol{A}\boldsymbol{B}| = |\boldsymbol{B}\boldsymbol{A}| = |\boldsymbol{A}||\boldsymbol{B}|$.
- If $\boldsymbol{B}$ is obtained from $\boldsymbol{A}$ by swapping two rows then $|\boldsymbol{A}| = -|\boldsymbol{B}|$.
- If $\boldsymbol{B}$ is obtained from $\boldsymbol{A}$ by adding a multiple of one row to another row then $|\boldsymbol{A}| = |\boldsymbol{B}|$.
- If $\boldsymbol{B}$ is obtained from $\boldsymbol{A}$ by taking out a common factor λ from each entry in a row of $\boldsymbol{A}$ then $|\boldsymbol{A}| = \lambda|\boldsymbol{B}|$.
- If $\boldsymbol{A}$ is a triangular matrix then $|\boldsymbol{A}| = a_{11}a_{22}...a_{nn}$.
- $|\boldsymbol{A}| = |\boldsymbol{A}^\top|$.
- $|\boldsymbol{A}^{-1}| = |\boldsymbol{A}|^{-1}$.
- For orthogonal $\boldsymbol{A}$, $|\boldsymbol{A}| = \pm 1$ since $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I}$ implies $|\boldsymbol{A}|^2 = 1$.
- For idempotent $\boldsymbol{A}$, $|\boldsymbol{A}| = 0$ or 1 because $|\boldsymbol{A}|^2 = |\boldsymbol{A}|$.
- $|\boldsymbol{A}| = \prod_{i=1}^n \lambda_i$

Sylvestor's theorem for determinants

$$|\boldsymbol{A} + \boldsymbol{B}\boldsymbol{C}\boldsymbol{D}^\top| = |\boldsymbol{C}^{-1} + \boldsymbol{D}^\top\boldsymbol{A}^{-1}\boldsymbol{B}||\boldsymbol{A}||\boldsymbol{C}|$$

Transpose

- $(\boldsymbol{A} + \boldsymbol{B})^\top = \boldsymbol{A}^\top + \boldsymbol{B}^\top$
- $(\boldsymbol{A}\boldsymbol{B})^\top = \boldsymbol{B}^\top\boldsymbol{A}^\top$
- Transposing a partitioned matrix:
$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} & \boldsymbol{C} \\ \boldsymbol{D} & \boldsymbol{E} & \boldsymbol{F} \end{bmatrix}^\top = \begin{bmatrix} \boldsymbol{A}^\top & \boldsymbol{D}^\top \\ \boldsymbol{B}^\top & \boldsymbol{E}^\top \\ \boldsymbol{C}^\top & \boldsymbol{F}^\top \end{bmatrix}$$
- $\left(\boldsymbol{A}^{-1}\right)^\top = \left(\boldsymbol{A}^\top\right)^{-1}$

Rank/Dimension

The **dimension** of $\mathcal{R}(\boldsymbol{X})$ or the **rank** of the matrix $\boldsymbol{X}$, written as dim($\mathcal{R}(\boldsymbol{X})$) or simply rank($\boldsymbol{X}$), is the number in the minimal (linearly independent) set of columns of $\boldsymbol{X}$ that span $\mathcal{R}(\boldsymbol{X})$. Similar definition applies to any vector space V, where the dimension of V is the number of the vectors in any basis of V.
Suppose that $\boldsymbol{A}$ is an $m \times n$ matrix then

- $\text{rank}(\boldsymbol{A}) \leq \min(m, n)$
- If $\text{rank}(\boldsymbol{A}) = \min(m, n)$ then the matrix $\boldsymbol{A}$ is said to have *full rank*.
- Only a zero matrix has rank zero.
- If $\boldsymbol{A}$ is a square matrix ($m = n$) then $\boldsymbol{A}$ is invertible if and only if $\boldsymbol{A}$ has rank n.
- If $\boldsymbol{B}$ is any $n \times k$ matrix, then $\text{rank}(\boldsymbol{A}\boldsymbol{B}) \leq \min(\text{rank}(\boldsymbol{A}), \text{rank}(\boldsymbol{B}))$.
- If $\boldsymbol{B}$ is any $n \times k$ matrix of rank n, then $\text{rank}(\boldsymbol{A}\boldsymbol{B}) = \text{rank}(\boldsymbol{A})$.
- If $\boldsymbol{C}$ is any $l \times m$ matrix of rank m, then $\text{rank}(\boldsymbol{C}\boldsymbol{A}) = \text{rank}(\boldsymbol{A})$.
- The $\text{rank}(\boldsymbol{A}) = r$ if and only if there exists an invertible $m \times m$ matrix $\boldsymbol{X}$ and an invertible $n \times n$ matrix $\boldsymbol{Y}$ such that $\boldsymbol{X}\boldsymbol{A}\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$.
- $\text{rank}(\boldsymbol{A} + \boldsymbol{B}) \leq \text{rank}\left(\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \end{bmatrix}\right) \leq \text{rank}(\boldsymbol{A}) + \text{rank}(\boldsymbol{B})$
- Sylvestor's rank of nullity: If $\boldsymbol{A}$ is an $m \times n$ matrix and $\boldsymbol{B}$ is any $n \times k$ matrix, $\text{rank}(A) + \text{rank}(B) - n \leq \text{rank}(AB)$.
- The inequality due to Frobenius: $\text{rank}(AB) + \text{rank}(BC) \leq \text{rank}(B) + \text{rank}(ABC)$.
- If $\boldsymbol{A}\boldsymbol{B}\boldsymbol{A} = \boldsymbol{A}$ then $\text{rank}(\boldsymbol{B}\boldsymbol{A}) = \text{rank}(\boldsymbol{A})$
- **Rank-nullity Theorem**: The rank of a matrix plus the nullity of the matrix equals the number of columns of the matrix, i.e.
$$\dim \mathcal{R}(\boldsymbol{A}) + \dim \mathcal{N}(\boldsymbol{A}) = n.$$
- $\text{rank}(\boldsymbol{A}^\top\boldsymbol{A}) = \text{rank}(\boldsymbol{A}\boldsymbol{A}^\top) = \text{rank}(\boldsymbol{A}) = \text{rank}(\boldsymbol{A}^\top)$

Derivatives

Let a be a constant that is not a function of $\boldsymbol{x}$; b a constant that is a function of $\boldsymbol{x}$; vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ functions of $\boldsymbol{x}$, and a matrix $\boldsymbol{A}$ that is not a function of $\boldsymbol{x}$ and a non-singular matrix $\boldsymbol{B}$ that is a function of the scalar s.

- $\frac{\partial a}{\partial \boldsymbol{x}} = \mathbf{0}$
- $\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{I}_m$
- $\frac{\partial \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{A}^\top$
- $\frac{\partial \boldsymbol{x}^\top \boldsymbol{A}}{\partial \boldsymbol{x}} = \boldsymbol{A}$
- $\frac{\partial a\boldsymbol{u}}{\partial \boldsymbol{x}} = a\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{x}}$
- $\frac{\partial b\boldsymbol{u}}{\partial \boldsymbol{x}} = b\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{x}} + \frac{\partial b}{\partial \boldsymbol{x}}\boldsymbol{u}^\top$
- $\frac{\partial \boldsymbol{A}\boldsymbol{u}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{x}}\boldsymbol{A}^\top$
- $\frac{\partial (\boldsymbol{u}+\boldsymbol{v})}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{x}} + \frac{\partial \boldsymbol{v}}{\partial \boldsymbol{x}}$
- $\frac{\partial \boldsymbol{g}(\boldsymbol{u})}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{g}(\boldsymbol{u})}{\partial \boldsymbol{u}}$
- $\frac{\partial \boldsymbol{u}^\top \boldsymbol{v}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{x}}\boldsymbol{v} + \frac{\partial \boldsymbol{v}}{\partial \boldsymbol{x}}\boldsymbol{u}$
- $\frac{\partial \boldsymbol{u}^\top \boldsymbol{A}\boldsymbol{v}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{x}}\boldsymbol{A}\boldsymbol{v} + \frac{\partial \boldsymbol{v}}{\partial \boldsymbol{x}}\boldsymbol{A}^\top\boldsymbol{u}$
- $\frac{\partial \boldsymbol{B}^{-1}}{\partial s} = -\boldsymbol{B}^{-1}\frac{\partial \boldsymbol{B}}{\partial s}\boldsymbol{B}^{-1}$
- $\frac{\partial \log|\boldsymbol{B}|}{\partial s} = \operatorname{tr}\left(\boldsymbol{B}^{-1}\frac{\partial \boldsymbol{B}}{\partial s}\right)$

Non-negative definite matrix

Positive definite matrices

A symmetric $n \times n$ real matrix $\boldsymbol{A}$ is said to be **positive definite** if $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} > 0$ is positive for every non-zero column vector $\boldsymbol{x}$.
A positive definite matrix holds the following properties:

- all its eigenvalues are positive
- every positive definite matrix is invertible and its inverse is also positive definite.
- it has a unique **Cholesky decomposition**: the matrix $\boldsymbol{A}$ is positive definite if and only if there exists a unique lower triangular matrix $\boldsymbol{L}$, with real and strictly positive diagonal elements, such that $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^\top$.
- $\boldsymbol{X}^\top \boldsymbol{A}\boldsymbol{X}$ is positive-semidefinite. If $\boldsymbol{X}$ is invertible, then $\boldsymbol{X}^\top \boldsymbol{A}\boldsymbol{X}$ is positive definite. Note that $\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X}$ does not need to be positive definite.

Positive semi-definite matrices

A symmetric matrix $\boldsymbol{A}$ is **positive semi-definite** if $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \geq 0$ for all $\boldsymbol{x}$ and $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = 0$ for some $\boldsymbol{x} \neq \mathbf{0}$.

- When $\boldsymbol{A}$ is p.(s.)d. so is $\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^\top$ for nonsingular $\boldsymbol{P}$.
- For real $\boldsymbol{X}$, $\boldsymbol{X}^\top \boldsymbol{X}$ is n.n.d. It is p.d. when $\boldsymbol{X}$ has full rank or else it is p.s.d.

Eigen-X

If $\boldsymbol{A}$ is an $n \times n$ matrix, $\boldsymbol{x}$ a non-zero $n \times 1$ column vector and λ is a scalar such that $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$, we call $\boldsymbol{x}$ an **eigenvector** of $\boldsymbol{A}$, and λ the corresponding **eigenvalue** (or λ-eigenvector of A). The set $\{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}\}$ is called the λ-**eigenspace** of $\boldsymbol{A}$ and comprises of the λ-eigenvectors and $\mathbf{0}$.

- A number λ is an eigenvalue of $\boldsymbol{A}$ if and only if $|\boldsymbol{A} - \lambda\boldsymbol{I}_n| = 0$.

Diagonalizable

Diagonalizable Theorem: If $\mathbb{R}^n$ is a basis of $\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n\}$ consisting of eigenvectors of an $n \times n$ matrix $\boldsymbol{A}$ then there exists an invertible matrix $\boldsymbol{P}$ and a diagonal matrix $\boldsymbol{D}$ such that $\boldsymbol{D} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$. And so $\boldsymbol{A}^n = \boldsymbol{P}\boldsymbol{D}^n\boldsymbol{P}^{-1}$.

- Projection matrix are diagonalizable with 0s and 1s on the diagonal.
- Real symmetric matrices are (orthogonally) diagonalizable by orthogonal matrices so $\boldsymbol{D} = \boldsymbol{Q}^\top \boldsymbol{A}\boldsymbol{Q}$ where $\boldsymbol{Q}$ is an orthogonal matrix.
- $\boldsymbol{A}$ is diagonalizable if and only if the sum of the dimensions of its eigenspaces is equal to n.

Illustrations

- **Generation matrix** $\boldsymbol{A}$: relating the frequencies of mating types in one generation to those in another $f^{(i+1)} = \boldsymbol{A}f^{(i)}$.
- **Markov Chain**: $\boldsymbol{x}$ is a *state probability vector* and $\boldsymbol{P}$ is the transition probability matrix which is related by $\boldsymbol{x}_{n+1}^\top = \boldsymbol{x}_n^\top \boldsymbol{P} = \boldsymbol{x}_0^\top \boldsymbol{P}^{n+1}$. Note $\boldsymbol{P}^n\mathbf{1} = \mathbf{1}$.
- **Linear Programming**: below are equivalent.
  1. minimise $f = \boldsymbol{c}^\top \boldsymbol{x}$ subject to $\boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{r}$ and $\boldsymbol{x} \geq \mathbf{0}$
  2. maximise $g = \boldsymbol{r}^\top \boldsymbol{z}$ subject to $\boldsymbol{A}\boldsymbol{z} \leq \boldsymbol{r}$ and $\boldsymbol{z} \geq \mathbf{0}$
- **Graph Theory**: Suppose a set of communication stations $\{S_i\}$. $\boldsymbol{T} = \{t_{ij}\}$ where $t_{ij} = 0$ except $t_{ij} = 1$ if a message can be sent from S_i to S_j. Then $\boldsymbol{T}^r = \{t_{ij}^{(r)}\}$, the element $t_{ij}^{(r)}$ is then the number of ways of getting a message from station i to station j in exactly r steps.

Partitioned Matrix

$\boldsymbol{A}_{r\times c} = \begin{bmatrix} \boldsymbol{\alpha}_1^\top \\ \vdots \\ \boldsymbol{\alpha}_r^\top \end{bmatrix}$ and $\boldsymbol{B}_{c\times s} = \begin{bmatrix} \boldsymbol{b}_1 & \cdots & \boldsymbol{b}_s \end{bmatrix}$:

$$\boldsymbol{A}\boldsymbol{B} = \{\boldsymbol{\alpha}_i^\top \boldsymbol{b}_j\} = \left\{\sum_{k=1}^{c} a_{ik}b_{kj}\right\}$$

Vector/Matrix operations

- A **scalar product** (also called *inner product* or *dot product*) for vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ is written $\boldsymbol{v} \cdot \boldsymbol{w}$ and

$$\boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{v}^\top \boldsymbol{w} = \sum_{i=1}^{n} v_i w_i = v_1 w_1 + .. + v_n w_n.$$

  The (Euclidean) **norm** (sometimes called *length* or *magnitude*) of a vector $\boldsymbol{v}$ is written as $||\boldsymbol{v}||$ and note $\boldsymbol{v}^\top \boldsymbol{v} = \sum_{i=1}^{n} v_i^2 = ||\boldsymbol{v}||^2$.
- A **Hadamard product** is given as $\boldsymbol{A} \cdot \boldsymbol{B} = \{a_{ij}b_{ij}\}$.

Quadratic Form

- The **quadratic form** of a matrix $\boldsymbol{A}$ is given as $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = \sum_i \sum_j x_i x_j a_{ij}$.
- For any particular quadratic form, there is a unique symmetric matrix $\boldsymbol{A}$ for which the quadratic form can be expressed as $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$.
- If we have the quadratic form $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ where $\boldsymbol{A}$ is not symmetric then we can rewrite it as $\boldsymbol{x}^\top \left[\frac{1}{2}(\boldsymbol{A} + \boldsymbol{A}^\top)\right]\boldsymbol{x}$ where $\frac{1}{2}(\boldsymbol{A} + \boldsymbol{A}^\top)$ is a unique symmetric matrix.

Matrices with all element equal

- $\mathbf{1}_n^\top \mathbf{1}_n = n$
- $\mathbf{1}_r \mathbf{1}_s^\top = \boldsymbol{J}_{r\times s}$
- $\boldsymbol{J}_{r\times s}\boldsymbol{J}_{s\times t} = s\boldsymbol{J}_{r\times t}$
- $\mathbf{1}_r^\top \boldsymbol{J}_{r\times s} = r\mathbf{1}_s^\top$
- $\boldsymbol{J}_{r\times s}\mathbf{1}_s = s\mathbf{1}_r$
- $\boldsymbol{J}_n = \mathbf{1}_n\mathbf{1}_n^\top$
- $\boldsymbol{J}_n^2 = n\boldsymbol{J}_n$
- $\bar{\boldsymbol{J}}_n = \frac{1}{n}\boldsymbol{J}_n$
- $\bar{\boldsymbol{J}}_n^2 = \bar{\boldsymbol{J}}_n$
- $\boldsymbol{C}_n = \boldsymbol{I} - \bar{\boldsymbol{J}}_n$
- $\boldsymbol{C} = \boldsymbol{C}^\top = \boldsymbol{C}^2$
- $\boldsymbol{C}\mathbf{1} = \mathbf{0}$
- $\boldsymbol{C}\boldsymbol{J} = \boldsymbol{J}\boldsymbol{C} = \mathbf{0}$
- $\boldsymbol{x}^\top \boldsymbol{C}\boldsymbol{x} = \sum_{i=1}^{n}(x_i - \bar{x})^2$

Vector Space

A **vector space** over $\mathbb{R}$ is a non-empty set V whose elements are called vectors on which two operations are defined, namely *addition of vectors* and *multiplication of a vector by a scalar* satisfying the 10 axioms below.

Axioms of vector space:

For all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ and k, $k1$, $k2 \in \mathbb{R}$,

A1 $\boldsymbol{u} + \boldsymbol{v} \in V$. This property is called *closure under addition*.

A2 $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$. This is called *associative law of addition*.

A3 $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$. That is, addition is *commutative*.

A4 There is a *zero vector* $\mathbf{0} \in V$ with $\boldsymbol{v} + \mathbf{0} = \mathbf{0} + \boldsymbol{v} = \boldsymbol{v}$.

A5 There is a vector which we write as $-\boldsymbol{v}$ and call a negative of $\boldsymbol{v}$, such that $\boldsymbol{v} + -\boldsymbol{v} = \mathbf{0}$.

S1 $k\boldsymbol{v} \in V$. This property is called *closure under multiplication by a scalar*.

S2 $k(\boldsymbol{u} + \boldsymbol{v}) = k\boldsymbol{u} + k\boldsymbol{v}$.

S3 $(k_1 + k_2)\boldsymbol{v} = k_1\boldsymbol{v} + k_2\boldsymbol{v}$.

S4 $(k_1k_2)\boldsymbol{v} = k_1(k_2\boldsymbol{v})$.

S5 $1\boldsymbol{v} = \boldsymbol{v}$.

Special set of vectors

The set of vectors $X = \{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n\}$ in a vector space V is called a **linearly independent** set if the only scalars that satisfy $\sum_{i=1}^n a_i\boldsymbol{v}_i = \mathbf{0}$ are $a_1 = a_2 = ... = a_n = 0$, otherwise it is called a **linearly dependent** set. $\mathcal{R}(X)$ is a subspace of V and is the smallest subspace of V containing the set X.

Basis

. A set of vectors X in a vector space V is called a **basis** of V if $\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n\}$ is a linearly independent set and $\mathcal{R}(X) = V$.

- If an $n \times n$ matrix $\boldsymbol{M}$ has n distinct eigenvalues then the set of n eigenvectors formed by selecting a non-zero vector from each eigenspace is a basis of $\mathbb{R}^n$.
- If $\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n\}$ is a basis of a vector space V then each vector $\boldsymbol{v}$ in V can be expressed as a linear combination of the basis in *exactly one way*.
- If one basis of a vector space V contains n vectors then every basis of V contains n vectors.
- If W is a non-zero subspace of V and V has dimension n then $\dim W \leq n$ with $\dim W = n$ if and only if $W = V$.
- In a vector space V of dimension n, every linearly independent set containing fewer than n vectors can be extended to a basis of V, and every spanning set of V with more than n vectors contains a basis of V where a spanning set is the span of the set of vectors in V. Moreover, every linearly independent set of n vectors is a basis of V and every set of n vectors which spans V is a basis of V.

Vector subspaces

. Let S be a non-empty subset of a vector space V. If S itself satisfies the 10 vector space axioms with the same operations of addition and multiplication by a scalar then S is called a *subspace* of V.
A subset S of a vector space V is a subspace of V if S is not empty and S satisfies A1 and S1. If the zero vector of V is *not* in S then S is not a subspace.
The addition of vector spaces is defined below:

$$V + U = \{\boldsymbol{v} + \boldsymbol{u} : \boldsymbol{v} \in V, \boldsymbol{w} \in U\}.$$

A decomposition of a vector space W to V and U is written as

$$W = V \oplus U$$

where $V \cap U = \{\mathbf{0}\}$. We also say W is the direct sum of V and U.

Symmetric Matrices

- $\boldsymbol{A}$ is symmetric if $\boldsymbol{A} = \boldsymbol{A}^\top$.
- $\boldsymbol{A}\boldsymbol{A}^\top$ and $\boldsymbol{A}^\top\boldsymbol{A}$ are symmetric.
- $\boldsymbol{A}^\top\boldsymbol{A} = \mathbf{0}$ implies $\boldsymbol{A} = \mathbf{0}$.
- $\mathrm{tr}\,(\boldsymbol{A}^\top\boldsymbol{A}) = \mathbf{0}$ implies $\boldsymbol{A} = \mathbf{0}$.

Matrix Factorisation

- Suppose a non-full rank matrix $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{X} & \boldsymbol{Y} \\ \boldsymbol{Z} & \boldsymbol{W} \end{bmatrix}$ where $\boldsymbol{X}$ is full rank. Then

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{Z}\boldsymbol{X}^{-1} \end{bmatrix} \boldsymbol{X} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{X}^{-1}\boldsymbol{Y} \end{bmatrix}$$

- If matrix $\boldsymbol{A}$ is not in above form then there exists permutation matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ such that $\boldsymbol{PAQ}$ is in the above form.

Solving Linear Equations

- The consistent equations $\boldsymbol{Ax} = \boldsymbol{y}$ for $\boldsymbol{y} \neq \mathbf{0}$ have a solution $\boldsymbol{x} = \boldsymbol{Gy}$ if and only if $\boldsymbol{AGA} = \boldsymbol{A}$.
- $\boldsymbol{Ax} = \boldsymbol{y}$ have solutions

$$\tilde{\boldsymbol{x}} = \boldsymbol{Gy} + (\boldsymbol{GA} - \boldsymbol{I})\boldsymbol{z}$$

for $\boldsymbol{G} = \boldsymbol{A}^-$ and any arbitrary vector $\boldsymbol{z}$.

Direct Sum

- $\boldsymbol{A} \oplus \boldsymbol{B} = \begin{bmatrix} \boldsymbol{A} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B} \end{bmatrix}$
- $(\boldsymbol{A} \oplus \boldsymbol{B}) + (\boldsymbol{C} \oplus \boldsymbol{D}) = (\boldsymbol{A} + \boldsymbol{C}) \oplus (\boldsymbol{B} + \boldsymbol{D})$
- $(\boldsymbol{A} \oplus \boldsymbol{B})(\boldsymbol{C} \oplus \boldsymbol{D}) = \boldsymbol{AC} \oplus \boldsymbol{BD}$
- $(\boldsymbol{A} \oplus \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \oplus \boldsymbol{B}^{-1}$

Direct Product

- $(\boldsymbol{A} \otimes \boldsymbol{B})^\top = \boldsymbol{A}^\top \otimes \boldsymbol{B}^\top$
- For vectors $\boldsymbol{x}$ and $\boldsymbol{y}$: $\boldsymbol{x}^\top \otimes \boldsymbol{y} = \boldsymbol{y}\boldsymbol{x}^\top = \boldsymbol{y} \otimes \boldsymbol{x}^\top$
- $\begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 \end{bmatrix} \otimes \boldsymbol{B} = \begin{bmatrix} \boldsymbol{A}_1 \otimes \boldsymbol{B} & \boldsymbol{A}_2 \otimes \boldsymbol{B} \end{bmatrix}$
- $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{X} \otimes \boldsymbol{Y}) = \boldsymbol{AX} \otimes \boldsymbol{BY}$
- $(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$
- $\mathrm{tr}\,(\boldsymbol{A} \otimes \boldsymbol{B}) = \mathrm{tr}\,(\boldsymbol{A})\,\mathrm{tr}\,(\boldsymbol{B})$
- $|\boldsymbol{A}_{p\times p} \otimes \boldsymbol{B}_{m\times m}| = |\boldsymbol{A}|^m|\boldsymbol{B}|^p$
- Eigenvalues of $\boldsymbol{A} \otimes \boldsymbol{B}$ are products of eigenvalues of $\boldsymbol{A}$ with those of $\boldsymbol{B}$.

The matrix $\boldsymbol{X}^\top\boldsymbol{X}$

Suppose $\boldsymbol{G}$ is the generalised inverse of $\boldsymbol{X}^\top\boldsymbol{X}$ then

- $\boldsymbol{G}^\top$ is also a generalised inverse of $\boldsymbol{X}^\top\boldsymbol{X}$.
- $\boldsymbol{XGX}^\top\boldsymbol{X} = \boldsymbol{X}$, i.e. $\boldsymbol{GX}^\top$ is a generalised inverse of $\boldsymbol{X}$.
- $\boldsymbol{XGX}^\top$ is invariant to $\boldsymbol{G}$.
- $\boldsymbol{XGX}^\top$ is symmetric whether $\boldsymbol{G}$ is or not.

Least Squares Equations

The following are invariant to the choice of $(\boldsymbol{X}^\top \boldsymbol{X})^-$:

- the vector of predicted values $\hat{\boldsymbol{y}} = \boldsymbol{X}\left(\boldsymbol{X}^\top \boldsymbol{X}\right)^- \boldsymbol{X}^\top \boldsymbol{y}$;
- The residual sum of squares $(\boldsymbol{y} - \hat{\boldsymbol{y}})^\top (\boldsymbol{y} - \hat{\boldsymbol{y}})$.

References

[1] H V Henderson and Shayle Roy Searle. On Deriving the Inverse of a Sum of Matrices. *Society for Industrial and Applied Mathematics*, 23(1):53–60, 1981.

 emi.tanaka@sydney.edu.au