Definition 1 (Vector space)**.** A vector space is a special collection of vectors that can be:

- added together to produce more vectors;
- scaled by a scalar to produce more vectors.

Each vector space has a corresponding field.

Definition 2 (Field, informal)**.** A field is essentially a set of scalar. For us, normally $\mathbb{R}$ or $\mathbb{C}$. When we don't care which one, we will use the notation $\mathbb{F}$.

Definition 3 (Field)**.** A field is a set $\mathbb{F}$ together with two operations, called addition $+$ and multiplication $\times$ which satisfy the field axioms, which are the following:

1. **Associativity:** $\forall a, b, c \in \mathbb{F}$:
$$a + (b + c) = (a + b) + c, \quad a \times (b \times c) = (a \times b) \times c$$
2. **Commutativity:** $\forall a, b, c \in \mathbb{F}$:
$$a + b = b + a, \quad a \times b = b \times a$$
3. **Additive and multiplicative identity:** $\exists 0 \in \mathbb{F}, 1 \in \mathbb{F}$:
$$a + 0 = a, \quad a \times 1 = a$$
4. **Additive inverses:** $\forall a \in \mathbb{F}, \exists -a \in \mathbb{F}$ such that:
$$a + (-a) = 0$$
5. **Multiplicative inverses:** $\forall a \neq 0 \in \mathbb{F}, \exists \frac{1}{a} \in \mathbb{F}$ such that:
$$a \times \frac{1}{a} = 1$$
6. **Distributivity of multipliation over addition**: $\forall a, b, c \in \mathbb{F}$:
$$a \times (b + c) = (a \times b) + (a \times c)$$

Definition 4 (Vector space)**.** A vector space $\mathcal{V}$ over a field $\mathbb{F}$ is a set of objects (called vectors), together with operations of vector addition $+$ and scalar multiplication $\times$, such that the following for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathcal{V}$ and scalars $a, b \in \mathbb{F}$ hold:

1. **Closure of vector addition:**
$$\mathbf{u} + \mathbf{v} \in \mathcal{V}$$
2. **Commutativity of addition:**
$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$$
3. **Associativity of addition:**
$$\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$$
4. **Identity of addition:**
$$\exists \mathbf{0} \in \mathcal{V} \text{ such that } \mathbf{u} + \mathbf{0} = \mathbf{u} = \mathbf{0} + \mathbf{u}$$
5. **Inverse of addition:**
$$\exists -\mathbf{u} \in \mathcal{V} \text{ such that } \mathbf{u} + (-\mathbf{u}) = \mathbf{0} = (-\mathbf{u}) + \mathbf{u}$$
6. **Closure of scalar multiplication:**
$$a \times \mathbf{u} \in \mathcal{V}$$
7. **Distributive law 1:**
$$a \times (\mathbf{u} + \mathbf{v}) = a \times \mathbf{u} + a \times \mathbf{v}$$
8. **Distributive law 2:**
$$(a + b) \times \mathbf{u} = a \times \mathbf{u} + b \times \mathbf{u}$$
9. **Associative law:**
$$(ab) \times \mathbf{u} = a \times (b \times \mathbf{u})$$
10. **Monoidal law:**
$$1 \times \mathbf{u} = \mathbf{u}$$

Definition 5 (Subspace)**.** A subspace $\mathcal{W}$ of a vector space $\mathcal{V}$ over a field $\mathbb{F}$ is a subset of $\mathcal{W} \subseteq \mathcal{V}$ that is by itself a vector space of $\mathbb{F}$:
$$a\mathbf{u} + b\mathbf{v} \in \mathcal{V}, \quad \forall \mathbf{u}, \mathbf{v} \in \mathcal{W}, \quad \forall a, b \in \mathbb{F}$$

Definition 6 ((non)Trivial subspace)**.** The subsets $\{\mathbf{0}\}$ and $\mathcal{V}$ are always subspaces of $\mathcal{V}$. These are called **trivial subspaces**. Similarly, a subspace $\mathcal{W}$ of $\mathcal{V}$ is said to be **nontrivial** if it is not one of those.

Definition 7 (Proper subspace)**.** A subspace $\mathcal{W}$ of $\mathcal{V}$ is said to be a proper subspace if it is not equal to $\mathcal{V}$, eg. $\mathcal{W} \subset \mathcal{V}$.

Definition 8 (Span)**.** Let $\mathcal{V}$ be a vector space over $\mathbb{F}$ and $\mathcal{S} \subseteq \mathcal{V}$. The span $\text{Span}(\mathcal{S})$ is the intersection of all subspaces that contain $\mathcal{S}$. If $\mathcal{S}$ is non-empty, then $\text{Span}(\mathcal{S})$ is all of the linear combinations of all finitely many vectors in $\mathcal{S}$.
$$\text{Span}(\mathcal{S}) = \begin{cases} \sum_{i=1}^{k} a_i \mathbf{v}_i \mid \mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathcal{S}, a_1, \ldots, a_k \in \mathbb{F}, k \in \mathbb{N} & \text{non-empty} \\ \{\mathbf{0}\} & \text{empty} \end{cases}$$

Definition 9 (Sum of two subspaces)**.** Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be subspaces of a vector space $\mathcal{V}$ over a field $\mathbb{F}$. Then the **sum** of $\mathcal{S}_1$ and $\mathcal{S}_2$ is defined as:
$$\mathcal{S}_1 + \mathcal{S}_2 = \text{Span}(\mathcal{S}_1 \cup \mathcal{S}_2) = \{\mathbf{u} + \mathbf{v} \mid \mathbf{u} \in \mathcal{S}_1, \mathbf{v} \in \mathcal{S}_2\}$$

Definition 10 (Direct sum)**.** If $\mathcal{S}_1 \cap \mathcal{S}_2 = \{\mathbf{0}\}$, then $\mathcal{S}_1 + \mathcal{S}_2$ is referred to as **direct sum**, and is denoted by $\oplus$.

Definition 11 (Linear dependence & independence)**.**

- A finite set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ in a vector space $\mathcal{V}$ over a field $\mathbb{F}$ is **linearly dependent** if and only if there are scalars $a_1, \ldots, a_k \in \mathbb{F}$, **not all zero**, such that $\sum_{i=1}^{k} a_i \mathbf{v}_i = 0$.
- A finite set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is **linearly independent** if they are not linearly dependent, i.e. if $\sum_{i=1}^{k} a_i \mathbf{v}_i = 0$ then we must have $a_1 = \ldots = a_k = 0$.

Definition 12 (Basis)**.** A set of vectors that is linearly independent and spans some vector space forms a **basis** for that vector space. A set $\mathcal{B}$ (which could be countably infinite) is a basis for the vector space $\mathcal{V}$ if and only if:

- $\text{Span}(\mathcal{B}) = \mathcal{V}$;
- $\mathcal{B}$ is linearly independent.

Definition 13 (Finite-dimensional)**.** A vector space is finite-dimensional if it has a finite basis.

Definition 14 (Dimension)**.** The dimension of a vector space $\mathcal{V}$, written as $\dim(\mathcal{V})$, over $\mathbb{F}$ is the number of vectors of any basis of $\mathcal{V}$ over $\mathbb{F}$.

Definition 15 (Orthogonal/orthonormal vectors)**.** A list of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{C}^n$ is orthogonal if:
$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \mathbf{v}_i^* \mathbf{v}_j = \mathbf{v}_j^* \mathbf{v}_i = 0, \quad \forall i, j \in \{1, \ldots, m\}$$
Furthermore, the list is orthonormal if:
$$\|\mathbf{v}_i\|^1 = \langle \mathbf{v}_i, \mathbf{v}_i \rangle = 1, \quad \forall i \in \{1, \ldots, m\}$$

Definition 16 (Linear map)**.** Let $\mathcal{U}$ and $\mathcal{V}$ be vector spaces over the same field $\mathbb{F}$. The mapping $\mathbf{f} : \mathcal{U} \to \mathcal{V}$ is called **linear** if:
$$\mathbf{f}(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha\mathbf{f}(\mathbf{x}) + \beta\mathbf{f}(\mathbf{y}), \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{U}, \quad \forall \alpha, \beta \in \mathbb{F}$$

Definition 17 (Invertible map)**.** Let $\mathcal{U}, \mathcal{V}$ be vector spaces over the same field $\mathbb{F}$. The mapping $\mathbf{f} : \mathcal{U} \to \mathcal{V}$ is called invertible if $\exists! \mathbf{g} : \mathcal{V} \to \mathcal{U}$ such that:

1. $\mathbf{g} \circ \mathbf{f} : \mathcal{U} \to \mathcal{U}, \quad \mathbf{g} \circ \mathbf{f}(\mathbf{u}) = \mathbf{u}, \quad \forall \mathbf{u} \in \mathcal{U}$
2. $\mathbf{f} \circ \mathbf{g} : \mathcal{V} \to \mathcal{V}, \quad \mathbf{f} \circ \mathbf{g}(\mathbf{v}) = \mathbf{v}, \quad \forall \mathbf{v} \in \mathcal{V}$

$\mathbf{f}$ is invertible if it is a bijection.

Definition 18 (Isomorphism for vector spaces)**.** Let $\mathcal{U}, \mathcal{V}$ be vector spaces over the same field $\mathbb{F}$ with the same dimension. The mapping $\mathbf{f} : \mathcal{U} \to \mathcal{V}$ is called an isomoprhism if it is both linear and invertible. In this case, we say that $\mathcal{U}$ and $\mathcal{V}$ are isomorphic.

Definition 19 (Transpose)**.** The transpose of a matrix $\mathbf{A}$, denoted $\mathbf{A}^\top$, is defined as for any $\mathbf{A} \in \mathbb{F}^{m \times n}$:
$$[\mathbf{A}^\top]_{ij} = [\mathbf{A}]_{ji}$$

Definition 20 (Hermitian transpose)**.** The conjugate transpose, adjoint or Hermitian transpose of a matrix $\mathbf{A}$, denoted $\mathbf{A}^*$ (or $\mathbf{A}^H$) is defined as the following: for any $\mathbf{A} \in \mathbb{C}^{m \times n}$:
$$[\mathbf{A}^*]_{ij} = [\bar{\mathbf{A}}]_{ji} \text{ or } \mathbf{A}^* = (\bar{\mathbf{A}})^\top$$

Definition 21 (Symmetric)**.** A square matrix $\mathbf{A} \in \mathbb{F}^{n \times n}$ is symmetric if
$$\mathbf{A}^\top = \mathbf{A}$$

Definition 22 (Skew-symmetric)**.** A square matrix $\mathbf{A} \in \mathbb{F}^{n \times n}$ is skew-symmetric if
$$\mathbf{A}^\top = -\mathbf{A}$$

Definition 23 (Orthogonal)**.** A square matrix $\mathbf{A} \in \mathbb{F}^{n \times n}$ is orthogonal if
$$\mathbf{A}^\top \mathbf{A} = \mathbf{I}$$
where $\mathbf{I}$ is the $n \times n$ identity matrix.

Definition 24 (Hermitian)**.** A square matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ is Hermitian if
$$\mathbf{A}^* = \mathbf{A}$$

Definition 25 (Skew-Hermitian)**.** A square matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is skew-Hermitian if

$$\mathbf{A}^* = -\mathbf{A}$$

Definition 26 (Unitary)**.** A square matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is unitary if

$$\mathbf{A}^*\mathbf{A} = \mathbf{I}$$

where $\mathbf{I}$ is the $n \times n$ identity matrix.

Definition 27 (Normal)**.** A square matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is normal if

$$\mathbf{A}^*\mathbf{A} = \mathbf{A}\mathbf{A}^*$$

Definition 28 (Sum of two matrices)**.** For any $\mathbf{A}, \mathbf{B} \in \mathbb{F}^{m\times n}$, the sum of $\mathbf{A}$ and $\mathbf{B}$ is:

$$[\mathbf{A}+\mathbf{B}]_{ij} = [\mathbf{A}]_{ij} + [\mathbf{B}]_{ij}$$

Definition 29 (Scalar multiplication of matrices)**.** For any $\mathbf{A} \in \mathbb{F}^{m\times n}$, the scalar multiplication of that matrix by λ is defined as:

$$[\lambda\mathbf{A}]_{ij} = \lambda[\mathbf{A}]_{ij}$$

Definition 30 (Matrix inner product)**.** For any $\mathbf{A} \in \mathbb{F}^{m\times n}$ and $\mathbf{B} \in \mathbb{F}^{n\times p}$, we have $\mathbf{AB} \in \mathbb{F}^{m\times p}$, where:

$$[\mathbf{AB}]_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$$

where $[\mathbf{AB}]_{ij}$ is the inner-product of the ith row of $\mathbf{A}$ and the jth column of $\mathbf{B}$.

Definition 31 (Matrix outer product)**.** Let

$$\mathbf{A} = [\mathbf{a}_1 \mid \mathbf{a}_2 \mid \ldots \mid \mathbf{a}_n]$$
$$\mathbf{B} = [\mathbf{b}_1 \mid \mathbf{b}_2 \mid \ldots \mid \mathbf{b}_n]^\top$$

We can combine this to obtain $\mathbf{AB}$:

$$\mathbf{AB} = \sum_{i=1}^{n} \mathbf{a}_i\mathbf{b}_i^\top$$

where $\mathbf{AB}$ is the sum of outer-products of columns of $\mathbf{A}$ and the corresponding rows of $\mathbf{B}$.

Definition 32 (Determinant)**.** The determinant of a matrix $\mathbf{A}$ is a function $\det : \mathbb{F}^{n\times n} \to \mathbb{F}$ defined as (the Leibniz formula):

$$\det \mathbf{A} = \sum_{\pi\in\mathcal{P}} \operatorname{sgn}(\pi) \prod_{i=1}^{n} a_{i\pi_i}$$

Definition 33 (Trace)**.** The trace of a matrix $\mathbf{A}$ is a function $\text{Trace} : \mathbb{F}^{n\times n} \to \mathbb{F}$ that is defined by:

$$\text{Trace}(\mathbf{A}) = \sum_i a_{ii}$$

Definition 34 (Matrix representation)**.** The $m\times n$ matrix $\mathbf{A}$ defined by the scalars a_{ij} is called the **matrix representation** of $\mathbf{f}$ in the ordered bases $\mathcal{B}_\mathcal{U}$ and $\mathcal{B}_\mathcal{V}$.

Definition 35 (Domain of a matrix)**.** The domain of $\mathbf{A}$ is $\text{Domain}(\mathbf{A}) = \mathbb{F}^n$.

Definition 36 (Range of a matrix)**.** The range of $\mathbf{A}$ is

$$\text{Range}(\mathbf{A}) = \{\mathbf{y} \in \mathbb{F}^m \mid \mathbf{y} = \mathbf{Ax} \text{ for some } \mathbf{x} \in \mathbb{F}^n\}$$

Note that $\text{Range}(\mathbf{A})$ is a subspace of $\mathbb{F}^m$ (doesn't have to be m-dimensional, just has to be $\leq m$).

Definition 37 (Rank of a matrix)**.** The rank of a matrix $\mathbf{A}$ is the dimension of the range of that matrix: $\dim(\text{Range}(\mathbf{A})) = \text{Rank}(A)$.

Definition 38 (Full-rank)**.** A full-rank matrix $\mathbf{A} \in \mathbb{F}^{m\times n}$ is a matrix with rank $= \min\{m, n\}$.

Definition 39 (Rank-deficient)**.** A rank-deficient matrix $\mathbf{A} \in \mathbb{F}^{m\times n}$ is a matrix with rank $< \min\{m, n\}$.

Definition 40 (Nullspace)**.** The nullspace of a matrix, denoted $\text{Null}(\mathbf{A})$ or $\text{Kernel}(\mathbf{A})$ is the set of all $\mathbf{x} \in \mathbb{F}^n$ such that $\mathbf{Ax} = \mathbf{0}$:

$$\text{Null}(\mathbf{A}) = \text{Kernel}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{F}^n \mid \mathbf{Ax} = \mathbf{0}\}$$

Note that $\text{Null}(\mathbf{A})$ is a subspace of $\mathbb{F}^n$ (once again, doesn't have to be n-dim.)

Definition 41 (Nullity of a matrix)**.** The nullity of a matrix $\mathbf{A}$ is the dimension of the nullspace of that matrix: $\dim(\text{Null}(\mathbf{A})) = \text{Nullity}(A)$.

Definition 42 (Orthogonal complement)**.** The orthogonal complement of a subspace $\mathcal{S}$, denoted $\mathcal{S}^\perp$, is:

$$\mathcal{S}^\perp = \{\mathbf{v} \mid \langle \mathbf{v}, \mathbf{w}\rangle = 0, \forall \mathbf{w} \in \mathcal{S}\}$$

Essentially all the vectors that are orthogonal to the whole subspace.

Definition 43 (Column space)**.** The column space of a matrix $\mathbf{A}$, denoted $\text{colsp}(\mathbf{A})$, is simply the range of $\mathbf{A}$:

$$\text{colsp}(\mathbf{A}) = \text{Range}(\mathbf{A}) = \{\mathbf{Ax} \mid \mathbf{x} \in \mathbb{C}^m\}$$

Definition 44 (Row space)**.** The row space of a matrix $\mathbf{A}$, denoted $\text{rowsp}(\mathbf{A})$, is simply the range of $\mathbf{A}^\top$:

$$\text{rowsp}(\mathbf{A}) = \text{Range}(\mathbf{A}^\top) = \{\mathbf{A}^\top\mathbf{x} \mid \mathbf{x} \in \mathbb{C}^n\}$$

Definition 45 (Non-singular)**.** $\mathbf{A} \in \mathbb{F}^{n\times n}$ is said to be non-singular if $\mathbf{Ax} = 0 \iff \mathbf{x} = 0$.

Definition 46 (Pseudo-inverse)**.** For any $\mathbf{A} \in \mathbb{F}^{m\times n}$ matrix, $\exists! \mathbf{A}^\dagger \in \mathbb{F}^{n\times m}$ called the pseudo-inverse that satisfies the following four properties;

- $\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}$;
- $\mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger$;
- $(\mathbf{A}\mathbf{A}^\dagger)^* = \mathbf{A}\mathbf{A}^\dagger$;
- $(\mathbf{A}^\dagger\mathbf{A})^* = \mathbf{A}^\dagger\mathbf{A}$.

Definition 47 (Vector norm)**.** Given a vector space $\mathcal{V}$ over $\mathbb{F}$, a norm is a non-negative real-valued function $\|\cdot\| : \mathcal{V} \to [0, \infty)$ with the following properties, namely:

- **Sub-additivity/triangle inequality**: $\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$;
- **Absolute homogeneity**: $\|\alpha\mathbf{u}\| = |\alpha|\|\mathbf{u}\|$;
- **Positive definiteness**: $\|\alpha\mathbf{u}\| = 0 \iff \mathbf{u} = \mathbf{0}$.

Definition 48 (Vector p-norms)**.** The p-norms are the following:

ℓ_1	Manhattan norm	$\|\mathbf{x}\|_1 = \sum_{i=1}^{d} \lvert x_i\rvert$
ℓ_2	Euclidean norm	$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{d} \lvert x_i\rvert^2} = \sqrt{\langle \mathbf{x}, \mathbf{x}\rangle}$
ℓ_∞	max norm	$\|\mathbf{x}\|_\infty = \max_{i=1,\ldots,d} \lvert x_i\rvert$
ℓ_p		$\left(\sum_{i=1}^{d} \lvert x_i\rvert^p\right)^{1/p}$

Definition 49 (Weighted Euclidean norm)**.** Let $\mathbf{W}$ be a diagonal matrix with positive diagonal elements. The **weighted Euclidean norm** is defined as:

$$\|\mathbf{x}\|_\mathbf{W} \triangleq \sqrt{\langle \mathbf{x}, \mathbf{Wx}\rangle}$$

Definition 50 (Frobenius norm)**.** Given any $\mathbf{A} \in \mathbb{C}^{m\times n}$, the ℓ_2 norm of the associated mn-dimensional vector is the Frobenius norm of the matrix:

$$\|\mathbf{A}\|_F \triangleq \left(\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2\right)^{\frac{1}{2}}$$

Definition 51 (Induced matrix norm)**.** Consider an arbitrary matrix $\mathbf{A} \in \mathbb{F}^{m\times s}$. Given any two norms $\|\cdot\|_p$ and $\|\cdot\|_q$ respectively on $\text{Domain}(\mathbf{A})$, $\text{Range}(\mathbf{A})$, the corresponding induced matrix norm is defined as:

$$\|\mathbf{A}\|_{p,q} \triangleq \max_{\substack{\mathbf{x}\in\mathbb{F}^m \\ \mathbf{x}\neq\mathbf{0}}} \frac{\|\mathbf{Ax}\|_q}{\|\mathbf{x}\|_p} = \max_{\substack{\mathbf{x}\in\mathbb{F}^m \\ \|\mathbf{x}\|_p=1}} \|\mathbf{Ax}\|_q$$

A common abbreviation if $p = q$ is to shorten $\|\mathbf{A}\|_{p,p}$ to $\|\mathbf{A}\|_p$.

Definition 52 (Condition of MVP)**.** Let $\mathbf{A} \in \mathbb{C}^{m\times n}$, and consider any vector norm $\|\cdot\|$ with its induced matrix norm. For a given vector $\mathbf{x}$, the condition of MVP for $\mathbf{A}$ is defined as:

$$\kappa(\mathbf{A}; \mathbf{x}) \triangleq \max_{\delta\mathbf{x}} \left(\|\frac{\mathbf{A}\delta\mathbf{x}\|}{\|\delta\mathbf{x}\|} \Big/ \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|} \right) = \frac{\|\mathbf{A}\|\|\mathbf{x}\|}{\|\mathbf{Ax}\|}$$

Definition 53 (Condition Number)**.** In the above, if $m \geq n$ and $\mathbf{A}$ has full column rank, then the condition number of $\mathbf{A}$, relative to $\|\cdot\|$, is defined as:

$$\kappa(\mathbf{A}) = \max_{\mathbf{x}} \kappa(\mathbf{A}; \mathbf{x}) = \|\mathbf{A}\|\|\mathbf{A}^\dagger\|$$

Definition 54 (Well and ill-conditioned)**.** If $\kappa(\mathbf{A})$ is small, $\mathbf{A}$ is said to be well-conditioned. If $\kappa(\mathbf{A})$ is large, $\mathbf{A}$ is ill-conditioned.

Definition 55 (Eigenvalue and eigenvector)**.** Let $\mathbf{A} \in \mathbb{C}^{n\times n}$. If we have:

$$\mathbf{Av} = \lambda\mathbf{v}, \quad \mathbf{v} \in \mathbb{C}^n, \mathbf{v} \neq \mathbf{0}, \lambda \in \mathbb{C}$$

then:

- λ is called an eigenvalue of $\mathbf{A}$;
- $\mathbf{v}$ is called an eigenvector of $\mathbf{A}$ associated with λ;
- the pair $(\lambda, \mathbf{v})$ is an eigenpair for $\mathbf{A}$.

Definition 56 (Spectrum)**.** The spectrum of $\mathbf{A} \in \mathbb{C}^{n\times n}$, denoted by $\text{spec}(\mathbf{A})$, is the set of all eigenvalues of $\mathbf{A}$:

$$\text{spec}(\mathbf{A}) = \{\lambda \in \mathbb{C} \mid \exists \mathbf{v} \neq \mathbf{0}, \mathbf{Av} = \lambda\mathbf{v}\}$$

Definition 57 (Spectral radius)**.** The spectral radius of a matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is the maximum magnitude of an eigenvector in the spectrum of that matrix:

$$\rho(\mathbf{A}) \triangleq \max_{\lambda \in \text{spec}(\mathbf{A})} |\lambda|$$

Definition 58 (Matrix polynomial)**.** Let $\mathbf{A} \in \mathbb{C}^{n\times n}$. Then a matrix polynomial of degree k is defined as:

$$p(\mathbf{A}) = \sum_{i=0}^{k} a_i \mathbf{A}^i$$

for $a_i \in \mathbb{C}, i = 1, 2, \ldots, k$.

Definition 59 (Eigenspace)**.** The eigenspace associated with an eigenvalue λ is the subspace defined as:

$$\begin{aligned}\mathcal{E}_\lambda(\mathbf{A}) &= \text{Null}(\mathbf{A} - \lambda\mathbf{I}) \\ &= \{\mathbf{v} \mid (\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = 0\} \\ &= \{\text{all eigenvectors of } \mathbf{A} \text{ associated with } \lambda\} \cup \{\mathbf{0}\}\end{aligned}$$

Definition 60 (Algebraic multiplicity)**.** The algebraic multiplicity of λ is the multiplicity of λ as a root of the characteristic polynomial.

Definition 61 (Geometric multiplicity)**.** The geometric multiplicty of λ is the dimension of the associated eigenspace:

$$\dim(\mathcal{E}_\lambda(\mathbf{A})) = \dim(\text{Null}(\mathbf{A} - \lambda\mathbf{I}))$$

Definition 62 (Simple eigenvalue)**.** The eigenvalue λ of $\mathbf{A}$ is said to be simple if its algebraic multiplicity is 1.

Definition 63 (Defective matrix)**.** A matrix is defective if it has an eigenvalue λ for which:

$$\dim(\mathcal{E}_\lambda(\mathbf{A})) < m(\lambda)$$

Definition 64 (Similarity transformation)**.** Let $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n\times n}$. We say that $\mathbf{B}$ is similar to $\mathbf{A}$ if there exists a non-singular matrix $\mathbf{S} \in \mathbb{C}^{n\times n}$ such that:

$$\mathbf{B} = \mathbf{S}^{-1}\mathbf{AS}$$

Definition 65 (Diagonalisable matrix)**.** If $\mathbf{A} \in \mathbb{C}^{n\times n}$ is similar to a diagonal matrix (pre and post multiplied), then $\mathbf{A}$ is said to be diagonalisable.

Definition 66 (Jordan block)**.** A **Jordan block** $\mathbf{J}_k(\lambda)$ is a $k \times k$ upper triangular matrix of the form:

$$\mathbf{J}_k(\lambda) = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix}$$

In particular, $\mathbf{J}_1(\lambda) = (\lambda)$ and $\mathbf{J}_2(\lambda) = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$.

Definition 67 (Unitarily diagonalisable matrix)**.** We say that $\mathbf{A}$ is unitarily diagonalisable if it is unitarily similar to a diagonal matrix.

Definition 68 (Compact SVD)**.** Discard zero entries on $\mathbf{\Sigma}$ to get $\mathbf{A} = \mathbf{U}_r\mathbf{\Sigma}_r\mathbf{V}_r^*$, where $\mathbf{U}_r \in \mathbb{C}^{m\times r}, \mathbf{\Sigma}_r \in \mathbb{C}^{r\times r}, \mathbf{V}_r \in \mathbb{C}^{m\times r}$ as:

$$\mathbf{A} = \underbrace{\begin{pmatrix}\mathbf{u}_1 & \ldots & \mathbf{u}_r\end{pmatrix}}_{\mathbf{U}_r} \underbrace{\text{diag}(\sigma_1, \ldots, \sigma_r)}_{\mathbf{\Sigma}_r} \underbrace{\begin{pmatrix}\mathbf{v}_1 & \ldots & \mathbf{v}_r\end{pmatrix}}_{\mathbf{V}_r^*} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^*$$

Definition 69 (Schatten norm)**.** Schatten p-norm of a matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$ is defined by applying vector p-norm to the vector of singular values, i.e.:

$$\|\mathbf{A}\|_p = \left(\sum_{i=1}^{\min m,n} \sigma_i^p\right)^{1/p}$$

Definition 70 (Diagonal matrix)**.** A diagonal matrix $\mathbf{D}$ is of the form:

$$\mathbf{D} = \begin{pmatrix} d_{11} & & & \\ & d_{22} & & \\ & & \ddots & \\ & & & d_{nn} \end{pmatrix}$$

Definition 71 (Block diagonal matrices)**.** A block diagonal matrix $\mathbf{D}$ consists of submatrices like the following:

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_{11} & & & \\ & \mathbf{D}_{22} & & \\ & & \ddots & \\ & & & \mathbf{D}_{kk} \end{pmatrix}$$

Definition 72 (Triangular matrix)**.** A triangular matrix $\mathbf{T}$ is of the following form:

$$\mathbf{T} = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ & t_{22} & \cdots & t_{2n} \\ & & \ddots & \vdots \\ & & & t_{nn} \end{pmatrix}$$

Definition 73 (Block-triangular matrix)**.** A block triangular matrix $\mathbf{T}$ is a matrix of the form:

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} & \cdots & \mathbf{T}_{1k} \\ & \mathbf{T}_{22} & \cdots & \mathbf{T}_{2k} \\ & & \ddots & \vdots \\ & & & \mathbf{T}_{kk} \end{pmatrix}$$

Definition 74 (Permutation matrix)**.** A permutation matrix $\mathbf{P}$ is a matrix where exactly one entry in each row and column is equal to 1, and all other entries are 0. For example:

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ 3 \\ 1 \end{pmatrix}$$

Definition 75 (Hessenberg matrix)**.** A Hessenberg matrix (upper shown here, but lower is easily seen) $\mathbf{A}$ or $\mathbf{H}$ is a matrix of the form:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ & a_{32} & a_{33} & \cdots & a_{3n} \\ & & \ddots & \ddots & \vdots \\ & & & a_{n,n-1} & a_{nn} \end{pmatrix}$$

Definition 76 (Unreduced matrix)**.** A Hessenberg matrix $\mathbf{A}$ is said to be unreduced if all of its super(sub)-diagonal entries are non-zero.

Definition 77 (Projection matrix)**.** A matrix $\mathbf{P} \in \mathbb{C}^{n\times n}$ is a projection, or idempotent, if $\mathbf{P}^2 = \mathbf{P}$.

Definition 78 (Orthogonal projection)**.** A matrix $\mathbf{P} \in \mathbb{C}^{n\times n}$ is an orthogonal projection if $\mathbf{P}^2 = \mathbf{P}$ and $\mathbf{P}^* = \mathbf{P}$.

Definition 79 (Positive (semi-)definite)**.** If $\mathbf{A} \in \mathbb{R}^{n\times n}$ is symmetric, it is positive definite if:

$$\mathbf{A} \succ \mathbf{0} \iff \langle \mathbf{x}, \mathbf{Ax} \rangle > 0, \quad \forall \mathbf{x} \neq \mathbf{0} \in \mathbb{R}^n$$

It is positive semi-definite if:

$$\mathbf{A} \succeq \mathbf{0} \iff \langle \mathbf{x}, \mathbf{Ax} \rangle \geq 0, \quad \forall \mathbf{x} \in \mathbb{R}^n$$

If $\mathbf{A} \in \mathbb{C}^{n\times n}$ is Hermitian (implied), it is positive definite if:

$$\mathbf{A} \succ \mathbf{0} \iff \langle \mathbf{x}, \mathbf{Ax} \rangle > 0, \quad \forall \mathbf{x} \neq \mathbf{0} \in \mathbb{C}^n$$

It is positive semi-definite if:

$$\mathbf{A} \succeq \mathbf{0} \iff \langle \mathbf{x}, \mathbf{Ax} \rangle \geq 0, \quad \forall \mathbf{x} \in \mathbb{C}^n$$

Definition 80 (Loewner Partial-Order).

$$\mathbf{A} \succ \mathbf{B} \iff \mathbf{A} - \mathbf{B} \succ \mathbf{0}$$
$$\mathbf{A} \succeq \mathbf{B} \iff \mathbf{A} - \mathbf{B} \succeq \mathbf{0}$$

Definition 81 (Schur complement). Let $\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{C} \end{pmatrix}$. The Schur complement of $\mathbf{A}$ in $\mathbf{B}$ is $\mathbf{C} - \mathbf{B}^*\mathbf{A}^{-1}\mathbf{B}$.

Definition 82 (Diagonally dominant matrix). A matrix $\mathbf{A}$ is diagonally dominant if it is of the form

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

and the magnitude of the diagonal entry in a row is larger than or equal to the sum of the magnitudes of all the other (non-diagonal) entries in that row:

$$|a_{ii}| \geq \sum_{i \neq j} |a_{ij}|, \quad \forall i.$$

Definition 83 (Banded matrices). A matrix $\mathbf{A}$ is banded if other than inside a band of diagonals, all other elements are nonzero, eg.

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1q} & & & & \\ \vdots & \ddots & \ddots & \ddots & & & \\ a_{p1} & \ddots & \ddots & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots & \ddots & a_{n-q+1,n} \\ & & & \ddots & \ddots & \ddots & \vdots \\ & & & & a_{n,n-p+1} & \cdots & a_{nn} \end{pmatrix}$$

Banded matrices are fantastic for LU decomposition. If $p, q \ll n$ then LU decomposition can be done in merely $\mathcal{O}(n)$ time and storage! But with pivoting, not as fantastic.

$$\mathbf{L} = \begin{pmatrix} 1 & & & & & & \\ \vdots & 1 & & & & & \\ \ell_{p1} & \ddots & 1 & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & \ell_{n,n-p+1} & \cdots & 1 \end{pmatrix}, \mathbf{U} = \begin{pmatrix} u_{11} & \cdots & u_{1q} & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & \ddots & \ddots & u_{n-q+1,n} \\ & & & & & \ddots & \vdots \\ & & & & & & u_{nn} \end{pmatrix}$$

Definition 84 (Non-asymptotic rate of convergence). If $\|\mathbf{T}\| < 1$, from $\|\mathbf{e}_k\| \leq \|\mathbf{T}\|^k \|\mathbf{e}_0\|$, it follows that after $k \geq \log(\varepsilon)/\log(\|\mathbf{T}\|)$, we have $\|\mathbf{e}_k\| \leq \varepsilon \|\mathbf{e}_0\|$ If $\|\mathbf{T}\| < 1$, then the factor $\|\mathbf{T}\|$ is called the non-asymptotic rate of convergence.

Definition 85 (Residual polynomial). A residual polynomial is a polynomial of degree k where $p_k(0) = 1$.

Definition 86 (Chebyshev polynomials of the first kind). Chebyshev polynomials of the first kind are defined recursively as:

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_{k+1}(x) &= 2xT_k(x) - T_{k-1}(x), \quad k \geq 1 \end{aligned}$$

Alternatively, we have explicit expressions:

$$T_k(x) = \begin{cases} \cos(k \arccos(x)) & |x| \leq 1 \\ \frac{1}{2}\left[(x + \sqrt{x^2 - 1})^k + (x + \sqrt{x^2 - 1})^{-k}\right] & |x| \geq 1 \end{cases}$$

Definition 87 (Krylov subspace). The Krylov subspace of order k generated by the matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ and the vector $\mathbf{v} \in \mathbb{C}^n$ is defined as:

$$\mathcal{K}_k(\mathbf{A}, \mathbf{v}) = \mathrm{Span}\{\mathbf{v}, \mathbf{A}\mathbf{v}, \ldots, \mathbf{A}^{k-1}\mathbf{v}\}, \quad k \geq 1$$

and where $\mathcal{K}_0(\mathbf{A}, \mathbf{v}) = \{\mathbf{0}\}$ (since all subspaces have to contain zero).

Definition 88 (Projection method). A projection method consists of a search subspace $\mathcal{K}_k$ with $\dim(\mathcal{K}_k) = k$, a constraint subspace $\mathcal{L}_k$ with $\dim(\mathcal{L}_k) = k$ and the Petrov-Galerkin conditions, which are to find some $\mathbf{x}_k \in \mathbf{x}_0 + \mathcal{K}_k$ such that $\mathbf{r}_k \perp \mathcal{L}_k$. A projection method is orthogonal if we wish to find $\mathcal{L}_k = \mathcal{K}_k$, and oblique if we wish to find $\mathcal{L}_k = \mathbf{A}\mathcal{K}_k$. More formally, let $\mathbf{x}_k = \mathbf{x}_0 + \mathbf{z}_k, \mathbf{z}_k \in \mathcal{K}_k$. Then the Petrov-Galerkin conditions imply $\mathbf{r}_0 - \mathbf{A}\mathbf{z}_k \perp \mathcal{L}_k$. So the projection method is defined as:

$$\text{find } \mathbf{x}_k = \mathbf{x}_0 + \mathbf{z}_k \text{ such that } \begin{cases} \mathbf{z}_k \in \mathcal{K}_k \\ \langle \mathbf{r}_0 - \mathbf{A}\mathbf{z}_k, \mathbf{w} \rangle = 0, \quad \forall \mathbf{w} \in \mathcal{L}_k \end{cases}$$