Fact 1. For a vector space $\mathcal{V}$ over a field $\mathbb{F}$, we have: $a\mathbf{u} + b\mathbf{v} \in \mathcal{V}, \quad \forall \mathbf{u}, \mathbf{v} \in \mathcal{V}, \quad \forall a, b \in \mathbb{F}$

Fact 2 (Various facts about subspaces)**.** The following hold:

- An intersection of subspaces $\mathcal{W} \cap \mathcal{X}$ is always a subspace.
- An union of subspaces $\mathcal{W} \cup \mathcal{X}$ **does not** need to be a subspace.
- A subspace cannot be empty, since a vector space always contains $\mathbf{0}$.

Theorem 1 (Sum of subspaces is smallest subspace)**.** *Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be subspaces of a vector space $\mathcal{V}$ over a field $\mathbb{F}$. Then, $\mathcal{S}_1 + \mathcal{S}_2$ is the **smallest subspace** containing $\mathcal{S}_1$ and $\mathcal{S}_2$.*

Proof. $\mathcal{S}_1 + \mathcal{S}_2$ is trivially a subspace. $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathcal{S}_1 + \mathcal{S}_2$. Conversely, every subspace containing $\mathcal{S}_1, \mathcal{S}_2$ must contain $\mathcal{S}_1 + \mathcal{S}_2$. Hence, $\mathcal{S}_1 + \mathcal{S}_2$ is the smallest subspace that contains $\mathcal{S}_1$ and $\mathcal{S}_2$. ∎

Theorem 2 (Uniquely represented as sum)**.** *Any $\mathbf{w} \in \mathcal{S}_1 \oplus \mathcal{S}_2$ can be **uniquely** represented as: $\mathbf{w} = \mathbf{u} + \mathbf{v}, \quad \mathbf{u} \in \mathcal{S}_1, \quad \mathbf{v} \in \mathcal{S}_2$*

Proof. Proof by contradiction. By the definition of subspace sum, any vector in $\mathcal{S}_1 \oplus \mathcal{S}_2$ can be written as $\mathbf{w} = \mathbf{u}_1 + \mathbf{v}_1, \quad \mathbf{u}_1 \in \mathcal{S}_1, \mathbf{v}_1 \in \mathcal{S}_2$ Suppose we also write $\mathbf{w} = \mathbf{u}_2 + \mathbf{v}_2, \quad \mathbf{u}_2 \in \mathcal{S}_1, \mathbf{v}_2 \in \mathcal{S}_2$ Combining these statements gives: $\mathbf{0} = (\mathbf{u}_1 - \mathbf{u}_2) + (\mathbf{v}_1 - \mathbf{v}_2)$ Clearly, $\mathbf{u}_1 \neq \mathbf{u}_2 \implies \mathbf{v}_1 \neq \mathbf{v}_2$ and vice versa. This implies that: $\mathbf{u}_1 - \mathbf{u}_2 = \mathbf{v}_2 - \mathbf{v}_1 \implies \mathcal{S}_1 \cap \mathcal{S}_2 \neq \{\mathbf{0}\}$ This is a contradiction from the fact that we are doing a direct sum, since the intersection must be zero for a direct sum. Therefore, $\mathbf{u}_1 = \mathbf{u}_2$ and $\mathbf{v}_1 = \mathbf{v}_2$. ∎

Fact 3. Note the following two facts:

- Any $\mathbf{v} \in \mathcal{V}$ can be represented uniquely in terms of elements in $\mathcal{B}$. There is only one and only one way to choose $\mathbf{b}_1, \ldots, \mathbf{b}_k \in \mathcal{B}$ and $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$ such that $\mathbf{v} = \sum_{i=1}^{k} \alpha_i \mathbf{b}_i$.
- Any linearly independent subset of list of $\mathcal{V}$ can be extended, perhaps in may ways, to form a basis of $\mathcal{V}$.

Fact 4. More fun facts about dimensions:

- If $\mathcal{W}$ is a subspace of $\mathcal{V}$, then $\dim(\mathcal{W}) \leq \dim(\mathcal{V})$.
- If $\mathcal{W}$ is a subspace of $\mathcal{V}$ and $\dim(\mathcal{W}) = \dim(\mathcal{V})$, then $\mathcal{W} = \mathcal{V}$.
- If $\dim(\mathcal{V}) = d$, then every system of linearly independent vectors of $\mathcal{V}$ has at most d elements, and any basis of $\mathcal{V}$ has exactly d elements (this is called the Dimension Theorem).
- The only vector space with dimension 0 is $\{0\}$.

Fact 5 (More facts about orthogonal/normal)**.**

- Every orthonormal list of vectors in $\mathbb{C}^n$ is linearly independent;

$$\begin{aligned}\mathbf{0} &= \sum_{i=1}^{n} \alpha_i \mathbf{v}_i \\ \implies 0 &= \left\langle \sum_{i=1}^{n} \alpha_i \mathbf{v}_i, \sum_{i=1}^{n} \alpha_i \mathbf{v}_i \right\rangle \\ &= \sum_{i=1}^{n} |\alpha_i|^2 \\ \implies \alpha_i &= 0, \quad \forall i\end{aligned}$$

- For a list of m orthonormal vectors in $\mathbb{C}^n$, we must have $m \leq n$;
- Any list of m orthonormal vectors in $\mathbb{C}^n$ form a basis for their span as an m-dimensional subspace of $\mathbb{C}^n$.

Theorem 3 (Inverse of a mapping is unique)**.** *An invertible map has a unique inverse.*

Proof. Suppose $\mathbf{f}$ is invertible with inverses $\mathbf{g}_1$ and $\mathbf{g}_2$, so we have: $\mathbf{g}_1 = \mathbf{g}_1 \circ \mathbf{I} = \mathbf{g}_1 \circ (\mathbf{f} \circ \mathbf{g}_2) = (\mathbf{g}_1 \circ \mathbf{f}) \circ \mathbf{g}_2 = \mathbf{I} \circ \mathbf{g}_2 = \mathbf{g}_2$ therefore the inverses $\mathbf{g}_1 = \mathbf{g}_2$ and it is unique. ∎

Theorem 4 (Invertibility of linear operators)**.** *Suppose $\mathcal{V}$ is finite dimensional and $\mathbf{f} : \mathcal{V} \to \mathcal{V}$ is a linear map. Then the following are equivalent:*

- *$\mathbf{f}$ is invertible;*
- *$\mathbf{f}$ is injective;*
- *$\mathbf{f}$ is surjective.*

Proof. Later; see Rank-Nullity Theorem. ∎

Fact 6. Note the following:

- Two finite-dimensional vector spaces over the same field are isomorphic if and only if they have the same dimension.
- Any d-dimensional vector space over $\mathbb{F}$ is isomorphic to $\mathbb{F}^d$.

Fact 7 (Characteristics of matrices)**.** Note the following characteristics of matrices (sum and multiplication):

- $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$;
- $(\mathbf{AB})^* = \mathbf{B}^* \mathbf{A}^*$;
- $(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$;
- $(\mathbf{A} + \mathbf{B})^* = \mathbf{A}^* + \mathbf{B}^*$;

Fact 8 (Facts about determinants:)**.**

- $\det(\mathbf{A}) = \det(\mathbf{A}^\top)$;
- $\det(\mathbf{A}^*) = \det(\bar{\mathbf{A}}) = \overline{\det(\mathbf{A})}$;
- $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$;
- $\det(\mathbf{A}^{-1}) = \frac{1}{\det(\mathbf{A})}$;
- $\det(\alpha \mathbf{A}) = \alpha^n \det(\mathbf{A}), \forall \alpha \in \mathbb{F}$.

Proposition 1. *For any unitary matrix $\mathbf{U} \in \mathbb{C}^{n \times n}$, $|\det(\mathbf{U})| = 1$.*

Proof. Since $\mathbf{U}$ is unitary, we have $\mathbf{U}^*\mathbf{U} = \mathbf{I}$. Therefore:

$$\begin{aligned} 1 &= \det(\mathbf{I}) \\ &= \det(\mathbf{U}^*\mathbf{U}) \\ &= \det(\mathbf{U}^*)\det(\mathbf{U}) \\ &= \overline{\det(\mathbf{U})}\det(\mathbf{U}) \\ &= |\det(\mathbf{U})|^2 \\ \implies |\det(\mathbf{U})| &= \sqrt{1} = 1 \end{aligned}$$

∎

Fact 9 (Properties of trace)**.**

- $\mathrm{Trace}(\mathbf{A}) = \mathrm{Trace}(\mathbf{A}^\top)$;
- $\mathrm{Trace}(\mathbf{A}^*) = \mathrm{Trace}(\bar{\mathbf{A}})$;
- $\mathrm{Trace}(\mathbf{AB}) \neq \mathrm{Trace}(\mathbf{A})\,\mathrm{Trace}(\mathbf{B})$;
- $\mathrm{Trace}(\mathbf{ABC}) = \mathrm{Trace}(\mathbf{CAB}) = \mathrm{Trace}(\mathbf{BCA})$ (this is known as the **cyclic property**).

Theorem 5 (Matrices and linear maps)**.** *Let $\mathbf{f} : \mathbb{F}^n \to \mathbb{F}^m$ be a linear map. Then $\exists! \mathbf{A} \in \mathbb{F}^{m \times n}$ such that: $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x} \quad \forall \mathbf{x} \in \mathbb{F}^n$ Conversely, if $\mathbf{A} \in \mathbb{F}^{m \times n}$ then the function defined above is a **linear map** from $\mathbb{F}^n$ to $\mathbb{F}^m$.*

Theorem 6 (Rank-Nullity Theorem)**.** *Important! If $\mathbf{A} \in \mathbb{F}^{m \times n}$:* $\dim(\mathrm{Range}(\mathbf{A})) + \dim(\mathrm{Null}(\mathbf{A})) = n$

Theorem 7 (Four Fundamental Subspaces)**.** *If $\mathbf{A} \in \mathbb{C}^{m \times n}$ (an m-by-n matrix in complex space), then:* $\mathrm{Null}(\mathbf{A}) = \mathrm{Range}(\mathbf{A}^*)^\perp$ *and* $\mathrm{Null}(\mathbf{A}^*) = \mathrm{Range}(\mathbf{A})^\perp$

Proof. Let $\mathbf{x} \in \mathrm{Null}(\mathbf{A})$. Take any $\mathbf{y} \in \mathrm{Range}(\mathbf{A}^*)$, we have that $\mathbf{y} = \mathbf{A}^*\mathbf{z}$ for some $\mathbf{z}$. So we have: $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{A}^*\mathbf{z} \rangle = \langle \mathbf{A}\mathbf{x}, \mathbf{z} \rangle = 0$ which implies that $\mathbf{x} \in \mathrm{Range}(\mathbf{A}^*)^\perp$, and hence $\mathrm{Null}(\mathbf{A}) \subseteq \mathrm{Range}(\mathbf{A}^*)^\perp$.

Conversely, let $\mathbf{x} \in \mathrm{Range}(\mathbf{A}^*)^\perp$, which means that for any $\mathbf{y} \in \mathrm{Range}(\mathbf{A}^*)$, we have that their inner product $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. In particular, choosing $\mathbf{y} = \mathbf{A}^*\mathbf{A}\mathbf{x}$ implies: $0 = \langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{A}^*\mathbf{A}\mathbf{x} \rangle = \langle \mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{x} \rangle = \|\mathbf{A}\mathbf{x}\|^2$ which gives $\mathbf{A}\mathbf{x} = 0$, which tells us that $\mathrm{Range}(\mathbf{A}^*)^\perp \subseteq \mathrm{Null}(\mathbf{A})$. Since we have shown that they are subsets of each other (a common proof technique), they are equal. The other statement is proved similarly. ∎

Fact 10 (Rank and (col/row)sp)**.** Similar to earlier, we have that: $\dim(\mathrm{colsp}(\mathbf{A})) = \dim(\mathrm{rowsp}(\mathbf{A})) = \mathrm{Rank}(\mathbf{A}) \leq \min\{m, n\}$

Fact 11 (Characteristics of rank)**.**

- $\mathrm{Rank}(\mathbf{A}) = \mathrm{Rank}(\mathbf{A}^\top) = \mathrm{Rank}(\mathbf{A}^*)$;
- $\mathrm{Rank}(\mathbf{A}^*\mathbf{A}) = \mathrm{Rank}(\mathbf{A})$.

Theorem 8 (Full-rank factorisation)**.** *$\mathbf{A}$ has rank r if and only if: $\mathbf{A} = \mathbf{X}\mathbf{Y}^\top$ for some $\mathbf{X} \in \mathbb{C}^{m \times r}, Y \in \mathbb{C}^{n \times r}$ (matrix outer product) each having full rank (independent columns.)*

Theorem 9 (Bounds on rank)**.** *If $\mathbf{A} \in \mathbb{C}^{m \times p}$ and $\mathbf{B} \in \mathbb{C}^{p \times n}$:* $\mathrm{Rank}(\mathbf{A}) + \mathrm{Rank}(\mathbf{B}) - p \leq \mathrm{Rank}(\mathbf{AB}) \leq \min\{\mathrm{Rank}(\mathbf{A}), \mathrm{Rank}(\mathbf{B})\}$

Fact 12 (Short and fat matrices are necessarily singular)**.** $\mathbf{A} \in \mathbb{F}^{m \times n}$ with $m < n$ (short and fat) is necessarily singular

Fact 13 (Equivalent to non-singular)**.**

- $\mathrm{Rank}(\mathbf{A}) = n$;
- $\exists! \mathbf{A}^{-1} \in \mathbb{F}^{n \times n}$ such that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$;
- $\det(\mathbf{A}) \neq 0$
- $\dim(\mathrm{Range}(\mathbf{A})) = n$ and $\dim(\mathrm{Null}(\mathbf{A})) = 0$;
- $\mathrm{Null}(\mathbf{A}) = \{\mathbf{0}\}$;
- $\mathbf{A}$ has linearly independent rows and columns;
- The linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a unique solution for each $\mathbf{b} \in \mathbb{F}^n$.

Fact 14 (Swapping inverse with transpose)**.** You can swap the inverse with the transpose or Hermitian conjugate:

If $\mathbf{A} \in \mathbb{F}^{n \times n}$ is non-singular, then $(\mathbf{A}^{-1})^\top = (\mathbf{A}^\top)^{-1} \triangleq \mathbf{A}^{-\top}$.

If $\mathbf{A} \in \mathbb{C}^{n \times n}$ is non-singular, then $(\mathbf{A}^{-1})^* = (\mathbf{A}^*)^{-1} \triangleq \mathbf{A}^{-*}$.

Fact 15. When $\mathbf{A}$ is full-column rank, we have a left inverse: $\mathbf{A}^\dagger = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*$ and so $\mathbf{A}^\dagger\mathbf{A} = \mathbf{I}$.

Fact 16. When $\mathbf{A}$ is full-row rank, we have a right inverse: $\mathbf{A}^\dagger = \mathbf{A}^*(\mathbf{A}\mathbf{A}^*)^{-1}$ and so $\mathbf{A}\mathbf{A}^\dagger = \mathbf{I}$.

Fact 17 (Pseudoinverse equals inverse)**.** If $\mathbf{A}$ is invertible, its pseudoinverse is its inverse.

Fact 18 (More properties of pseudoinverse)**.**

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

Fact 19. Unlike the inverse, where this is valid: $(\mathbf{AB})^\dagger \neq \mathbf{B}^\dagger\mathbf{A}^\dagger$

Proof. We write:

$$\mathbf{x} = \mathbf{y} + (\mathbf{x} - \mathbf{y}) \implies \|\mathbf{x}\| = \|\mathbf{y} + (\mathbf{x} - \mathbf{y})\| \leq \|\mathbf{y}\| + \|(\mathbf{x}$$
$$\mathbf{y} = \mathbf{x} + (\mathbf{y} - \mathbf{x}) \implies \|\mathbf{y}\| = \|\mathbf{x} + (\mathbf{y} - \mathbf{x})\| \leq \|\mathbf{x}\| + \|(\mathbf{y}$$

Therefore $\|\mathbf{x}\| - \|\mathbf{y}\| \leq \|\mathbf{x} - \mathbf{y}\|$. ∎

Proposition 2 (Equivalence of norms in $\mathbb{C}^d$)**.** *For all $\mathbf{x} \in \mathbb{C}^d$, we have: $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{d}\|\mathbf{x}\|_2 \leq$ $d\|\mathbf{x}\|_\infty$*

Theorem 10 (Unitary invariance of Euclidean norm in $\mathbb{C}^d$)**.** *Given any matrix $\mathbf{U} \in \mathbb{C}^{m \times d}$ with $m \geq d$ and orthonormal columns, we have: $\|\mathbf{U}\mathbf{x}\|_2 = \|\mathbf{x}\|_2$*

Proof. One-liner: $\|\mathbf{U}\mathbf{x}\|_2^2 = \langle \mathbf{U}\mathbf{x}, \mathbf{U}\mathbf{x} \rangle = \langle \mathbf{x}, \mathbf{U}^*\mathbf{U}\mathbf{x} \rangle = \langle \mathbf{x}, \mathbf{x} \rangle = \|\mathbf{x}\|^2$ ∎

Theorem 11 (Unitary invariance of Frobenius norm in $\mathbb{C}^{m \times n}$)**.** *Given any matrix $\mathbf{U} \in \mathbb{C}^{p \times m}$ with $p \geq m$ and orthonormal columns, we have: $\|\mathbf{U}\mathbf{A}\|_F = \|\mathbf{A}\|_F$*

Proof. Another one-liner: $\|\mathbf{U}\mathbf{A}\|_F^2 = \mathrm{Trace}(\mathbf{A}^*\mathbf{U}^*\mathbf{U}\mathbf{A}) = \mathrm{Trace}(\mathbf{A}^*\mathbf{A}) = \|\mathbf{A}\|_F^2$ ∎

Theorem 12 (Sub-multiplicativity of entry-wise matrix norms)**.** *For any two matrices $\mathbf{A} \in \mathbb{C}^{m \times n}, \mathbf{B} \in \mathbb{C}^{n \times p}$:*

$$\|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_F \|\mathbf{B}\|_F$$
$$\|\mathbf{AB}\|_1 \leq \|\mathbf{A}\|_1 \|\mathbf{B}\|_1$$

(note Frobenius norm is the entry-wise ℓ_2 norm)

Theorem 13 (Unitary invariance of induced 2-norm in $\mathbb{F}^{m \times n}$)**.** *Given any matrix $\mathbf{U} \in \mathbb{F}^{p \times m}$ orthonormal columns, we have: $\|\mathbf{U}\mathbf{A}\|_2 = \|\mathbf{A}\|_2$ where the norm here is the induced 2-norm (Euclidean norm)*

Proof. Immediate, by noticing that for any $\mathbf{x} \in \mathbb{F}^m$, we have: $\|\mathbf{U}\mathbf{A}\mathbf{x}\|_2 = \|\mathbf{A}\mathbf{x}\|_2$ ∎

Theorem 14 (All induced matrix norms are submultiplicative)**.** *Let $\|\cdot\|_p, \|\cdot\|_q, \|\cdot\|_r$ be vector norms on, respectively,* $\mathrm{Domain}(\mathbf{B}), \mathrm{Range}(\mathbf{B}), \mathrm{Range}(\mathbf{A})$. *We have: $\|\mathbf{AB}\|_{p,r} \leq \|\mathbf{A}\|_{q,r}\|\mathbf{B}\|_{p,q}$*

Proof. For any $\mathbf{x}$, we have that: $\|\mathbf{AB}\mathbf{x}\|_r \leq \|\mathbf{A}\|_{q,r}\|\mathbf{B}\mathbf{x}\|_q \leq \|\mathbf{A}\|_{q,r}\|\mathbf{B}\|_{p,q}\|\mathbf{x}\|_p$ ∎

Theorem 15 (Equivalence of induced matrix norms)**.** *For any $\mathbf{A} \in \mathbb{C}^{m \times n}$ with $\mathrm{Rank}(\mathbf{A}) = r$, we have:*

$$\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F \leq \sqrt{r}\|\mathbf{A}\|_2$$
$$\|\mathbf{A}\|_\infty \leq \sqrt{n}\|\mathbf{A}\|_2 \leq \sqrt{mn}\|\mathbf{A}\|_\infty$$
$$\|\mathbf{A}\|_1 \leq \sqrt{m}\|\mathbf{A}\|_2 \leq \sqrt{mn}\|\mathbf{A}\|_1$$

Proof. One liner: $1 = \|\mathbf{A}\mathbf{A}^\dagger\| \leq \|\mathbf{A}\|\|\mathbf{A}^\dagger\|$ ∎

Fact 20. $(\lambda, \mathbf{v})$ is an eigenpair $\iff$ $(\lambda\mathbf{I} - \mathbf{A})\mathbf{v} = \mathbf{0}$ and $\mathbf{v} \neq \mathbf{0}$

Fact 21. Eigenvalues are the roots of the characteristic polynomial of $\mathbf{A}$, i.e. $\det(\lambda\mathbf{I}-\mathbf{A}) = 0$. $\det(\lambda\mathbf{I}-\mathbf{A})$ is a polynomial of degree exactly n in λ, i.e.: $\det(\lambda\mathbf{I}-\mathbf{A}) = p_n(\lambda) = \sum_{k=0}^{n} c_k\lambda^i, c_n \neq 0$

Fact 22 (Facts about eigenpairs and conjugates)**.**

- $(\lambda, c\mathbf{v})$ is an eigenpair for $\mathbf{A}$ for any $c \in \mathbb{C}$ e.g. $c = \frac{1}{\|\mathbf{v}\|_2}$;
- If $\mathbf{A} \in \mathbb{C}^{n\times n}$ and $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, then $\bar{\mathbf{A}}\bar{\mathbf{v}} = \bar{\lambda}\bar{\mathbf{v}}$;
- If $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, then $\mathbf{A}\bar{\mathbf{v}} = \bar{\lambda}\bar{\mathbf{v}}$, i.e. for real matrices, if $(\lambda,\mathbf{v})$ is an eigenpair, then so is $(\bar{\lambda},\bar{\mathbf{v}})$;
- If $\mathbf{A} \in \mathbb{C}^{n\times n}$ and λ is an eigenvalue of $\mathbf{A}$, then $\bar{\lambda}$ is an eigenvalue of $\mathbf{A}^*$, but **eigenvectors might not be related**.

Fact 23 (Determinants and trace)**.** Amazingly: $\det(\mathbf{A}) = \prod_{i=1}^{n}\lambda_i$, $\quad \mathrm{Trace}(\mathbf{A}) = \sum_{i=1}^{n}\lambda_i$

Fact 24 (Facts about spectrum and its radius)**.**

- $\mathrm{spec}(\mathbf{A}) = \mathrm{spec}(\mathbf{A}^\top)$;
- $\mathrm{spec}(\bar{\mathbf{A}}) = \mathrm{spec}(\mathbf{A}^*)$;
- $\mathrm{spec}(\mathbf{A}) \neq \mathrm{spec}(\mathbf{A}^*)$ but **always** $\rho(\mathbf{A}) = \rho(\mathbf{A}^*)$;
- $\forall\alpha\in\mathbb{C}$: $\rho(\alpha\mathbf{A}) = |\alpha|\rho(\mathbf{A})$, $\quad \rho(\mathbf{A}^k) = [\rho(\mathbf{A})]^k$

Fact 25. The polynomial is monic if $a_k = 1$. Polynomial factorisation also carries over to matrices: $p(\mathbf{A}) = \prod_{i=1}^{k}(\mathbf{A}-\beta_i\mathbf{I}), \beta_i \in \mathbb{C}, i = 1,\ldots,k$

Theorem 16 (Spectral mapping theorem for matrix polynomials)**.**

- *If $(\lambda,\mathbf{v})$ is an eigenpair $\mathbf{A} \in \mathbb{C}^{n\times n}$, then $(p(\lambda),\mathbf{v})$ is an eigenpair of $p(\mathbf{A})$.*
- *Conversely, if $k \geq 1$ and μ is an eigenvalue of $p(\mathbf{A})$, then there is some eigenvalue of λ of $\mathbf{A}$ such that $\mu = p(\lambda)$.*

Proof. Note that $\mathbf{A}^i\mathbf{v} = \mathbf{A}^{i-1}\mathbf{A}\mathbf{v} = \lambda\mathbf{A}^{i-1}\mathbf{v} = \ldots = \lambda^i\mathbf{v}$. So: $p(\mathbf{A})\mathbf{v} = \sum_{i=0}^{k} a_i\lambda^i\mathbf{v} = p(\lambda)\mathbf{v}$ Let's define $q(t) = p(t) - \mu$. Since $k \geq 1, q(\mathbf{A}) = p(\mathbf{A}) - \mu\mathbf{I}$ has degree k, so it can be factorised as: $p(\mathbf{A}) - \mu\mathbf{I} = q(\mathbf{A}) = \prod_{i=1}^{k}(\mathbf{A}-\beta_i\mathbf{I})$ $p(\mathbf{A})-\mu\mathbf{I}$ is singular so some factor $(\mathbf{A}-\beta_j\mathbf{I})$ must be singular, which means that β_j is an eigenvalue of $\mathbf{A}$. But: $0 = q(\beta_j) = p(\beta_j) - \mu \implies \mu = p(\beta_j)$ ■

Theorem 17. *$\mathbf{A}$ is singular if and only if $0 \in$* $\mathrm{spec}(\mathbf{A})$.

Proof. A lot of if and only ifs:

$$\begin{aligned}\mathbf{A}\text{ is singular} &\iff \exists\mathbf{x}\neq\mathbf{0}\text{ s.t. }\mathbf{A}\mathbf{x} = \mathbf{0}\\ &\iff \exists\mathbf{x}\neq\mathbf{0}\text{ s.t. }\mathbf{A}\mathbf{x} = 0\mathbf{x}\\ &\iff 0 \in \mathrm{spec}(\mathbf{A})\end{aligned}$$

■

Theorem 18. *Let $\mathbf{A} \in \mathbb{C}^{n\times n}$ and $\lambda,\mu\in\mathbb{C}$. Then:* $\lambda \in \mathrm{spec}(\mathbf{A}) \iff \lambda+\mu \in \mathrm{spec}(\mathbf{A}+\mu\mathbf{I})$

Proof. More iffs:

$$\begin{aligned}\lambda\in\mathrm{spec}(\mathbf{A}) &\iff \exists\mathbf{v}\neq\mathbf{0}\text{ s.t. }\mathbf{A}\mathbf{v} = \lambda\mathbf{v}\\ &\iff \mathbf{A}\mathbf{v}+\mu\mathbf{v} = \lambda\mathbf{v}+\mu\mathbf{v}\\ &\iff (\mathbf{A}+\mu\mathbf{I})\mathbf{v} = (\lambda+\mu)\mathbf{v}\\ &\iff \lambda+\mu\in\mathrm{spec}(\mathbf{A}+\mu\mathbf{I})\end{aligned}$$

■

Theorem 19. *If $\mathbf{A} \in \mathbb{C}^{n\times n}$ is Hermitian, then all its eigenvalues are real.*

Proof. Suppose $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, $v \neq \mathbf{0}$. We have: $\lambda\mathbf{v}^*\mathbf{v} = \mathbf{v}^*\mathbf{A}\mathbf{v} = \mathbf{v}^*\mathbf{A}^*\mathbf{v}$ On the other hand: $\lambda\mathbf{v}^*\mathbf{v} = \mathbf{v}^*\mathbf{A}\mathbf{v} \iff (\lambda\mathbf{v}^*\mathbf{v})^* = (\mathbf{v}^*\mathbf{A}\mathbf{v})^* \iff \bar{\lambda}\mathbf{v}^*\mathbf{v} = \mathbf{v}^*\mathbf{A}^*\mathbf{v}$ So $\lambda = \bar{\lambda}$, which means that $\lambda\in\mathbb{R}$. ■

Theorem 20. *If $\mathbf{A} \in \mathbb{C}^{n\times n}$ is Hermitian, then eigenvectors corresponding to distinct eigenvalues are mutually orthogonal.*

Proof. Suppose we have two vectors $\mathbf{v},\mathbf{w}$ such that:

$$\begin{aligned}\mathbf{A}\mathbf{v} = \lambda\mathbf{v}, \mathbf{v}\neq\mathbf{0}\\ \mathbf{A}\mathbf{w} = \mu\mathbf{w}, \mathbf{w}\neq\mathbf{0}\end{aligned}$$

with $\lambda\neq\mu$ (unique eigenpairs). We have:

$$\begin{aligned}\lambda\langle\mathbf{v},\mathbf{w}\rangle &= \langle\lambda\mathbf{v},\mathbf{w}\rangle = \langle\mathbf{A}\mathbf{v},\mathbf{w}\rangle = \langle\mathbf{v},\mathbf{A}^*\mathbf{w}\rangle\\ &= \langle\mathbf{v},\mathbf{A}\mathbf{w}\rangle = \langle\mathbf{v},\mu\mathbf{w}\rangle = \mu\langle\mathbf{v},\mathbf{w}\rangle\end{aligned}$$

So $(\lambda-\mu)\langle\mathbf{v},\mathbf{w}\rangle = 0$, which since $\mu\neq\lambda$, we get $\langle\mathbf{v},\mathbf{w}\rangle = 0$. Recall that an inner product of 0 is equivalent to orthogonality. ■

Theorem 21. *Let $m(\lambda)$ be the algebraic multiplicity of λ. Then there are bounds on the geometric multiplicity: $1 \leq \dim(\mathcal{E}_\lambda(\mathbf{A})) \leq m(\lambda)$*

Fact 26. Two similar matrices share the same spectrum and the same characteristic polynomial.

Theorem 22. *If $\mathbf{A}$ and $\mathbf{B}$ are similar, then they have the same characteristic polynomial.*

Proof.

$$\begin{aligned}p_{\mathbf{B}}(\lambda) &= \det(\mathbf{B}-\lambda\mathbf{I})\\ &= \det(\mathbf{S}^{-1}\mathbf{A}\mathbf{S}-\lambda\mathbf{S}^{-1}\mathbf{S})\\ &= \det(\mathbf{S}^{-1}(\mathbf{A}-\lambda\mathbf{I})\mathbf{S})\\ &= \det(\mathbf{S}^{-1}\det(\mathbf{A}-\lambda\mathbf{I})\det(\mathbf{S}))\\ &= \det(\mathbf{A}-\lambda\mathbf{I}) = p_{\mathbf{A}}(\lambda)\end{aligned}$$

■

Theorem 23. *$(\lambda,\mathbf{v})$ is an eigenpair of $\mathbf{A}$ if and only if $(\lambda,\mathbf{S}^{-1}\mathbf{v})$ is an eigenpair for $\mathbf{B} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$.*

Proof.

$$\begin{aligned}\mathbf{A}\mathbf{v} = \lambda\mathbf{v} &\iff \mathbf{A}\mathbf{S}\mathbf{S}^{-1}\mathbf{v} = \lambda\mathbf{v}\\ &\iff \mathbf{S}^{-1}\mathbf{A}\mathbf{S}\mathbf{S}^{-1}\mathbf{v} = \lambda\mathbf{S}^{-1}\mathbf{v}\\ &\iff \mathbf{B}\mathbf{S}^{-1}\mathbf{v} = \lambda\mathbf{S}^{-1}\mathbf{v}\\ &\iff \mathbf{B}\mathbf{w} = \lambda\mathbf{w}\end{aligned}$$

where we define $\mathbf{w} = \mathbf{S}^{-1}\mathbf{v}$. ■

Theorem 24. *The matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is diagonalisable if and only if it has n linearly independent eigenvectors. In other words, $\mathbf{A}$ is diagonalisable if and only if it is not defective, i.e.:*

$\dim(\mathcal{E}_\lambda(\mathbf{A})) = m(\lambda), \quad \forall\lambda\in\mathrm{spec}(\mathbf{A})$

A simple criterion: if all eigenvalues of $\mathbf{A}$ are simple, then $\mathbf{A}$ is diagonalisable.

Theorem 25 (Eigendecomposition)**.** *Let $\mathbf{A}$ be diagonalisable and define $\mathbf{V} = \begin{pmatrix}\mathbf{v}_1 & \ldots & \mathbf{v}_n\end{pmatrix} \in \mathbb{C}^{n\times n}$ to be the set of linearly independent eigenvectors of $\mathbf{A}$. Then:* $\mathbf{V}^{-1}\mathbf{A}\mathbf{V} = \begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{pmatrix} \triangleq \mathbf{\Lambda}$

Theorem 26 (Schur decomposition/triangularisation)**.** *For any $\mathbf{A} \in \mathbb{C}^{n\times n}$, there exists unitary $\mathbf{U} \in \mathbb{C}^{n\times n}$ such that:* $\mathbf{U}^*\mathbf{A}\mathbf{U} = \begin{pmatrix}\lambda_1 & b_{12} & \ldots & b_{1n}\\ & \lambda_2 & b_{23} & b_{2n}\\ & & \ddots & \vdots\\ & & & \lambda_n\end{pmatrix}$

Theorem 27 (Jordan canonical form)**.** *For any $\mathbf{A} \in \mathbb{C}^{n\times n}$, there is a non-singular $\mathbf{S} \in \mathbb{C}^{n\times n}$, positive integers $k, n_1, n_2, \ldots, n_k$ with $n_1+n_2+\ldots+n_k = n$ and scalars $\lambda_1,\ldots,\lambda_k \in \mathbb{C}$ such that: $\mathbf{A} =$*

$$\mathbf{S}\overbrace{\begin{pmatrix}\mathbf{J}_{n_1}(\lambda_1) & & \\ & \ddots & \\ & & \mathbf{J}_{n_k}(\lambda_k)\end{pmatrix}}^{\mathbf{J}_{\mathbf{A}}}\mathbf{S}^{-1}$$

Fact 27 (Facts about Jordan)**.**

- The Jordan matrix $\mathbf{J}_{\mathbf{A}}$ is uniquely determined up to permutation of Jordan blocks;
- If $\mathbf{A}$ is real and has only real eigenvalues, then $\mathbf{S}$ can be chosen to be real;
- The number of Jordan blocks, k, is the maximum number of linearly independent eigenvectors of $\mathbf{A}$;
- Given an eigenvalue λ, its geometric multiplicity is the number of its corresponding Jordan blocks;
- The sum of the sizes of all Jordan blocks corresponding to an eigenvalue λ is its algebraic multiplicity;
- If an eigenvalue is defective, the size of at least one of its corresponding Joradn blocks is greater than one, so a matrix is diagonalisable if and only if all its Jordan blocks are 1×1.

Fact 28. For diagonalisable matrices, we have: Jordan canonical form $\equiv$ eigendecomposition

Theorem 28. *A matrix is unitarily diagonalisable if and only if it is normal.*

Proof. ($\implies$) Suppose $\mathbf{A}$ is unitarily diagonalisable, that is $\mathbf{U}^*\mathbf{A}\mathbf{U} = \mathbf{\Lambda}$. So we have $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^*$, hence:

$$\begin{aligned}\mathbf{A}\mathbf{A}^* &= \mathbf{U}\mathbf{\Lambda}\mathbf{U}^*\mathbf{U}\mathbf{\Lambda}^*\mathbf{U}^*\\ &= \mathbf{U}\mathbf{\Lambda}\mathbf{\Lambda}^*\mathbf{U}^*\\ &= \mathbf{U}\mathbf{\Lambda}^*\mathbf{\Lambda}\mathbf{U}^*\\ &= \mathbf{U}\mathbf{\Lambda}^*\mathbf{U}^*\mathbf{U}\mathbf{\Lambda}\mathbf{U}^*\\ &= \mathbf{A}^*\mathbf{A}\end{aligned}$$

($\impliedby$) Conversely, suppose $\mathbf{A}$ is normal and consider its Schur decomposition, ie $\mathbf{U}^*\mathbf{A}\mathbf{U} = \mathbf{T}$. We have:

$$\begin{aligned}\mathbf{T}^*\mathbf{T} &= \mathbf{U}^*\mathbf{A}^*\mathbf{U}\mathbf{U}^*\mathbf{A}\mathbf{U}\\ &= \mathbf{U}^*\mathbf{A}^*\mathbf{A}\mathbf{U}\\ &= \mathbf{U}^*\mathbf{A}\mathbf{A}^*\mathbf{U}\\ &= \mathbf{U}^*\mathbf{A}\mathbf{U}\mathbf{U}^*\mathbf{A}^*\mathbf{U}\\ &= \mathbf{T}\mathbf{T}^*\end{aligned}$$

but since $\mathbf{T}$ is upper-triangular, it has to be diagonal. ■

Theorem 29. *For normal matrices: Schur decomposition $\equiv$ eigendecomposition or:* $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^* = \sum_{i=1}^{n}\lambda_i\mathbf{u}_i\mathbf{u}_i^*$

Fact 29 (Some final facts on matrices)**.**

- Among complex matrices, all unitary, Hermitian and skew-Hermitian matrices are normal;
- Among real matrices, all orthogonal, symmetric and skew-symmetric matrices are normal;
- It is **not** the case that all normal matrices are either unitary or (skew-)Hermitian, e.g. $\forall a,b\in\mathbb{C}$, $\begin{bmatrix}a & b\\ -b & a\end{bmatrix}$ is normal and has $\lambda_i = a\pm ib$;
- A normal matrix is Hermitian $\iff$ all its eigenvalues are real

Theorem 30 (Singular value decomposition)**.** *Let $\mathbf{A} \in \mathbb{C}^{m\times n}, q = \min\{m,n\}$ and $\mathrm{Rank}(\mathbf{A}) \triangleq r \leq q$. There exists two unitary matrices $\mathbf{U} \in \mathbb{C}^{m\times m}$ and $\mathbf{V} \in \mathbb{C}^{n\times n}$, and a square diagonal matrix:*

$$\mathbf{\Sigma}_q = \begin{pmatrix}\sigma_1 & & & & & \\ & \sigma_2 & & & & \\ & & \ddots & & & \\ & & & \sigma_r & & \\ & & & & 0 & \\ & & & & & \ddots & \\ & & & & & & 0\end{pmatrix} \in \mathbb{R}^{q\times q}$$

Theorem 31. *For any $\mathbf{A} \in \mathbb{C}^{m\times n}$, we have:*

$\sigma_i = \sqrt{\lambda_i(\mathbf{A}^*\mathbf{A})} = \sqrt{\lambda_i(\mathbf{A}\mathbf{A}^*)}$, $i = 1,2,\ldots,\mathrm{Rank}(\mathbf{A})$

Proof. Assume without loss of generality that $m \geq n$. let $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$. We have:

$$\begin{aligned}\mathbf{A}^*\mathbf{A} &= \mathbf{V}\mathbf{\Sigma}^\top\mathbf{U}^*\mathbf{U}\mathbf{\Sigma}\mathbf{V}^*\\ &= \mathbf{V}\mathbf{\Sigma}^\top\mathbf{\Sigma}\mathbf{V}^*\\ &= \mathbf{V}\mathbf{\Sigma}_n^2\mathbf{V}^*\\ &\implies \ldots\end{aligned}$$

Similarly:

$$\begin{aligned}\mathbf{A}\mathbf{A}^* &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*\mathbf{V}\mathbf{\Sigma}^\top\mathbf{U}^*\\ &= \mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^*\mathbf{U}^*\\ &= \mathbf{U}\begin{pmatrix}\mathbf{\Sigma}_n^2 & \mathbf{0}_{n\times(m-n)}\\ \mathbf{0}_{(m-n)\times n} & \mathbf{0}_{(m-n)\times(m-n)}\end{pmatrix}\mathbf{U}^*\\ &\implies \ldots\end{aligned}$$

■

Theorem 32. *Let $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ be an SVD of $\mathbf{A} \in \mathbb{C}^{m\times n}$ and assume that for some r, we have $\sigma_r \neq 0$ and $\sigma_{r+1} = 0$. (Since singular values are conventionally ordered, this implies all singular values past this point are zero).*

Then we have the following:

- $\mathrm{Rank}(\mathbf{A}) = r$
- $\mathrm{Null}(\mathbf{A}) = \mathrm{Span}\{\mathbf{v}_{r+1},\mathbf{v}_{r+2},\ldots,\mathbf{v}_n\}$
- $\mathrm{Range}(\mathbf{A}) = \mathrm{Span}\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_r\}$.
- $\mathbf{A}^\dagger = \mathbf{V}\mathbf{\Sigma}^\dagger\mathbf{U}^* = \mathbf{V}_r\mathbf{\Sigma}_r^{-1}\mathbf{U}_r^*$

Theorem 33. *Let $\mathbf{A} \in \mathbb{C}^{n\times n}$ be a normal matrix whose (not necessarily distinct) eigenvalues are $\lambda_1,\ldots,\lambda_n$. Show that the singular values of $\mathbf{A}$ are $|\lambda_1|,|\lambda_2|,\ldots,|\lambda_n|$.*

Proof. Since $\mathbf{A}$ is a normal matrix, $\implies$ unitarily diagonalisable as $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^*$, where $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues of $\mathbf{A}$. Now, we have:

$$\begin{aligned}\sigma_i &= \sqrt{\lambda_i(\mathbf{A}^*\mathbf{A})} \\ &= \sqrt{\bar{\lambda}_i(\mathbf{A})\lambda_i(\mathbf{A})} \\ &= \sqrt{|\lambda_i(\mathbf{A})|^2} = |\lambda_i(\mathbf{A})|\end{aligned}$$

■

Theorem 34 (Matrix low-rank approximation: spectral norm). $\|\mathbf{A} - \mathbf{A}_k\|_2 = \min_{\substack{A \in \mathbb{C}^{m\times n} \\ \mathrm{Rank}(\mathbf{B}) \le k}} \|\mathbf{A} - \mathbf{B}\|_2 =$ σ_{k+1} *where* $\|\cdot\|_2$ *is the matrix spectral norm and* $\sigma_{k+1} = 0$ *for* $k = \min\{m, n\}$.

Theorem 35 (Matrix low-rank approximation: Frobenius norm). $\|\mathbf{A} - \mathbf{A}_k\|_{\mathbf{F}} = \min_{\substack{\mathbf{A} \in \mathbb{C}^{m\times n} \\ \mathrm{Rank}(\mathbf{B}) \le k}} \|\mathbf{A} - \mathbf{B}\|_{\mathbf{F}} =$ $\sqrt{\sum_{i=k+1}^{r} \sigma_i}$ *where* $\|\cdot\|_{\mathbf{F}}$ *is the matrix Frobenius norm.*

Theorem 36. $\mathcal{E} = \mathbf{U}\mathcal{E}_0$, *where* $\mathcal{E}_0 = \{\mathbf{y} \in \mathbb{R}^n \mid \sum_{i=1}^{n} \frac{y_i^2}{\sigma_i^2} = 1\}$.

Proof. Suppose $\mathbf{z} \in \mathcal{S}$. We have $\mathbf{A}\mathbf{z} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top\mathbf{z} = \mathbf{U}\mathbf{y}$ where $\mathbf{y} = \mathbf{\Sigma}\mathbf{V}^\top\mathbf{z}$. We just need to show that $\mathbf{y} \in \mathcal{E}_0$. We have $\mathbf{z} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{y}$, which implies that:

$$\begin{aligned}1 &= \|\mathbf{z}\|^2 \\ &= \|\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{y}\|^2 \\ &= \|\mathbf{\Sigma}^{-1}\mathbf{y}\|^2 \\ &= \sum_{i=1}^{n} \frac{y_i^2}{\sigma_i^2}\end{aligned}$$

This implies that $\mathbf{y} \in \mathcal{E}_0$. ■

Fact 30 (Properties of a diagonal matrix).

- $\mathrm{spec}(\mathbf{D}) = \{d_{11}, d_{22}, \dots, d_{nn}\}$;
- $\det(\mathbf{D}) = \prod_{i=1}^{n} d_{ii}$
- $\mathbf{D}$ is non-singular $\iff d_{ii} \neq 0, \forall i$

Fact 31 (Properties of block-diagonal matrices).

- $\mathrm{spec}(\mathbf{D}) = \bigcup_{i=1}^{k} \mathrm{spec}(\mathbf{D}_{ii})$
- $\det(\mathbf{D}) = \prod_{i=1}^{k} \det(\mathbf{D}_{ii})$
- $\mathbf{D}$ is non-singular $\iff \mathbf{D}_{ii}$ is nonsingular

Fact 32 (Properties of a triangular matrix).

- $\mathrm{spec}(\mathbf{T}) = \{t_{11}, t_{22}, \dots, t_{nn}\}$;
- $\det(\mathbf{T}) = \prod_{i=1}^{k} t_{ii}$
- $\mathbf{T}$ is non-singular $\iff$ all $t_{ii} \neq 0$
- $\mathrm{Rank}(\mathbf{T}) \ge$ the number of nonzero t_{ii}. For example, the singular values of the strictly upper triangular matrix: $\begin{pmatrix} 0 & t_{12} & & & \\ & 0 & t_{23} & & \\ & & \ddots & \ddots & \\ & & & & t_{n-1,n} \\ & & & & 0 \end{pmatrix}$ are $0, |t_{12}|, \dots, |t_{n-1,n}|$.
- Sparsity patterns: the inverse of a triangular matrix is triangular.
- The product of two triangular matrices is triangular.

Fact 33 (Properties of block-triangular matrices).

- $\mathrm{spec}(\mathbf{T}) = \bigcap_{i=1}^{k} \mathrm{spec}(\mathbf{T}_{ii})$
- $\det(\mathbf{T}) = \prod_{i=1}^{k} \det(\mathbf{T}_{ii})$
- $\mathbf{T}$ is non-singular $\iff$ all $\mathbf{T}_{ii}$ are non-singular
- $\mathrm{Rank}(\mathbf{T}) \ge \sum_{i=1}^{k} \mathrm{Rank}(\mathbf{T}_{ii})$
- The sparsity pattern is similar to the triangular case, but with respect to blocks.

Fact 34 (Facts about permutation matrices).

- $\mathbf{P}^\top\mathbf{P} = \mathbf{P}\mathbf{P}^\top = \mathbf{I}$, i.e. $\mathbf{P}$ is orthogonal;
- $\det(\mathbf{P}) = \pm 1$, that is, permutation matrices are non-singular
- Left-multiplication of a matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ and $n \times n$ permutation matrix $\mathbf{P}$, i.e. $\mathbf{PA}$, permutes the rows of $\mathbf{A}$;
- Right-multiplication of a matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$, and $n \times n$ permutation matrix $\mathbf{P}$, i.e. $\mathbf{AP}$, permutes the columns of $\mathbf{A}$.
- If $\mathbf{P}$ and $\mathbf{Q}$ are permutation matrices, then so is $\mathbf{PQ}$ and $\mathbf{QP}$ (generally $\mathbf{PQ} \neq \mathbf{QP}$)

Fact 35. The rank of an unreduced matrix is at least $n - 1$ since its first $n - 1$ columns are independent.

Theorem 37 (Various facts about projections which should be proven).

1. $\mathbf{P}\mathbf{v} = \mathbf{v} \iff \mathbf{v} \in \mathrm{Range}(\mathbf{P})$
2. *If* $\mathbf{P}$ *is a projection, then so is* $\mathbf{I} - \mathbf{P}$
3. $\mathrm{Range}(\mathbf{I} - \mathbf{P}) = \mathrm{Null}(\mathbf{P})$
4. $\mathrm{Range}(\mathbf{P}) \cap \mathrm{Range}(\mathbf{I} - \mathbf{P}) = \{\mathbf{0}\}$
5. $\mathrm{Range}(\mathbf{P}) \oplus \mathrm{Range}(\mathbf{I} - \mathbf{P}) = \mathbb{C}^n$
6. $\lambda \in \{0, 1\}$

Proof.

1. ($\implies$) $\mathbf{v} = \mathbf{P}\mathbf{v} \implies \exists \mathbf{w} \in \mathbb{C}^n \text{s.t.} \mathbf{v} = \mathbf{P}\mathbf{w}$, namely $\mathbf{w} = \mathbf{v}$.
   ($\impliedby$) $\mathbf{v} \in \mathrm{Range}(\mathbf{P}) \implies \exists \mathbf{w} \in \mathbb{C}^n$ such that $\mathbf{v} = \mathbf{P}\mathbf{w} \implies \mathbf{P}\mathbf{v} = \mathbf{P}^2\mathbf{w} = \mathbf{P}\mathbf{w} = \mathbf{v}$.
2. $(\mathbf{I} - \mathbf{P})^2 = \mathbf{I} - 2\mathbf{P} + \mathbf{P}^2 = \mathbf{I} - \mathbf{P}$
3. ($\implies$) $\mathbf{v} \in \mathrm{Range}(\mathbf{I} - \mathbf{P}) \implies \exists \mathbf{w} \in \mathbb{C}^n$ such that $\mathbf{v} = (\mathbf{I} - \mathbf{P})\mathbf{w} \implies \mathbf{P}\mathbf{v} = \mathbf{P}(\mathbf{I} - \mathbf{P})\mathbf{w} = (\mathbf{P} - \mathbf{P}^2)\mathbf{w} = \mathbf{0} \implies \mathbf{v} \in \mathrm{Null}(\mathbf{P})$
   ($\impliedby$) $\mathbf{v} \in \mathrm{Null}(\mathbf{P}) \implies \mathbf{P}\mathbf{v} = \mathbf{0} \implies \mathbf{v} = (\mathbf{I} - \mathbf{P})\mathbf{v} \implies \mathbf{v} \in \mathrm{Range}(\mathbf{I} - \mathbf{P})$
4. $\mathbf{v} \in \mathrm{Range}(\mathbf{P}) \cap \mathrm{Range}(\mathbf{I} - \mathbf{P}) \implies \mathbf{v} = \mathbf{P}\mathbf{v} = \mathbf{P}(\mathbf{I} - \mathbf{P})\mathbf{v} = \mathbf{0}$
5. $\mathrm{Range}(\mathbf{P}) \oplus \mathrm{Range}(\mathbf{I} - \mathbf{P}) \subseteq \mathbb{C}^n$, but also $\mathbf{x} = \mathbf{P}\mathbf{x} + (\mathbf{I} - \mathbf{P})\mathbf{x} \implies \mathbb{C}^n \subseteq \mathrm{Range}(\mathbf{P}) \oplus \mathrm{Range}(\mathbf{I} - \mathbf{P})$
6. $\mathbf{P}\mathbf{v} = \lambda\mathbf{v} \implies \mathbf{P}^2\mathbf{v} = \lambda\mathbf{P}\mathbf{v} \implies \mathbf{P}\mathbf{v} = \lambda^2\mathbf{v} \implies \lambda = \lambda^2 \implies \lambda \in \{0, 1\}$.

■

Fact 36 (Facts about orthogonal projections).

- $\mathrm{Range}(\mathbf{P}) \perp \mathrm{Range}(\mathbf{I} - \mathbf{P})$
- $\|\mathbf{v}\|^2 = \|\mathbf{P}\mathbf{v}\|^2 + \|(\mathbf{I} - \mathbf{P})\mathbf{v}\|^2$
- Given any matrix $\mathbf{Q} \in \mathbb{C}^{m\times n}$ with orthonormal columns, $\mathbf{P} = \mathbf{Q}\mathbf{Q}^*$ is an orthogonal projection onto the $\mathrm{Range}(\mathbf{Q})$
- Given any matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$, $\mathbf{P} = \mathbf{A}\mathbf{A}^\dagger$ is an orthogonal projection onto the $\mathrm{Range}(\mathbf{A})$.
- **Rank-one orthogonal projector:** $\mathbf{P} = \mathbf{v}\mathbf{v}^*/\|\mathbf{v}\|^2$ is an orthogonal projection along the direction given by $\mathbf{v} \in \mathbb{C}^n$.

Fact 37 (Facts about positive (semi-)definite matrices).

- $\mathbf{A} \in \mathbb{C}^{m\times n} \implies \mathbf{A}^*\mathbf{A} \succeq \mathbf{0}$
- $\mathbf{A} \in \mathbb{C}^{m\times n} \implies \mathbf{A}\mathbf{A}^* \succeq \mathbf{0}$
- $\mathbf{A} \succ \mathbf{0} \iff \lambda_i(\mathbf{A}) > 0, i = 1, \dots, n$
- $\mathbf{A} \succeq \mathbf{0} \iff \lambda_i(\mathbf{A}) \ge 0, i = 1, \dots, n$
- $\mathbf{A} \prec \mathbf{0} \iff \lambda_i(\mathbf{A}) < 0, i = 1, \dots, n$
- $\mathbf{A} \preceq \mathbf{0} \iff \lambda_i(\mathbf{A}) \le 0, i = 1, \dots, n$
- Every PD matrix is invertible, and its inverse is also PD
- For $\mathbf{A}, \mathbf{B} \succ \mathbf{0}$ and $\alpha > 0$, $\alpha\mathbf{A} \succ \mathbf{0}$ and $\mathbf{A} + \mathbf{B} \succ \mathbf{0}$
- $\mathbf{A} \succeq \mathbf{0} \iff \exists! \mathbf{B} \succeq \mathbf{0}$ such that $\mathbf{B}^2 = \mathbf{A}$ (not to be confused with Cholesky factor)
- For $\mathbf{A} \succ \mathbf{0}$, the Schur decomposition, spectral decomposition and SVD all coincide
- If $\mathbf{A} \succeq \mathbf{0}$, then $\mathbf{B}^*\mathbf{A}\mathbf{B} \succeq \mathbf{0}, \forall \mathbf{B} \in \mathbb{C}^{n\times m}$
- If $\mathbf{A} \succ \mathbf{0}$ and $\mathbf{B}$ has full column rank, then $\mathbf{B}^*\mathbf{A}\mathbf{B} \succ \mathbf{0}$.

Fact 38 (Properties of the Loewner partial-order).

- $\mathbf{A} \succeq \mathbf{B}$, then $\lambda_i(\mathbf{A}) \ge \lambda_i(\mathbf{B}), i = 1, 2, \dots, n$
- $\mathbf{A} \succeq \mathbf{B}$ and $\mathbf{A} \neq \mathbf{B}$, then $\exists i \in \{1, 2, \dots, n\}, \lambda_i(\mathbf{A}) > \lambda_i(\mathbf{B})$
- $\mathbf{A} \succ \mathbf{B}$, then $\lambda_i(\mathbf{A}) > \lambda_i(\mathbf{B}), i = 1, 2, \dots, n$ (after ordering eigenvalues).

Fact 39 (Properties of the Schur complement).

- $\mathbf{M} \succ \mathbf{0} \iff \mathbf{A} \succ \mathbf{0}$ and $\mathbf{C} - \mathbf{B}^*\mathbf{A}^{-1}\mathbf{B} \succ \mathbf{0}$;
- $\mathbf{M} \succeq \mathbf{0} \iff \mathbf{A} \succ \mathbf{0}$ and $\mathbf{C} - \mathbf{B}^*\mathbf{A}^{-1}\mathbf{B} \succeq \mathbf{0}$.

Theorem 38 (Levy-Desplanques Theorem). *A strictly diagonally dominant matrix is non-singular.*

Fact 40 (Implications on positive (semi-)definiteness).

- If $\mathbf{A} \in \mathbb{C}^{n\times n}$ is Hermitian, diagonally dominant with non-negative diagonals $a_{ii} \ge 0\ \forall i$, then $\mathbf{A}$ is positive semi-definite.
- If $\mathbf{A} \in \mathbb{C}^{n\times n}$ is Hermitian, strictly diagonally dominant with positive diagonals $a_{ii} > 0\ \forall i$, then $\mathbf{A}$ is positive definite.

Fact 41. If $\mathbf{A}$ is invertible, then it admits an LU factorisation if and only if all its leading principal minors are nonzero.

Fact 42. We can uniquely write $\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{U}$, where $\mathbf{D}$ is a diagonal matrix and $\mathbf{L}, \mathbf{U}$ are *unit* triangular matrices:

$$\begin{pmatrix} 1 & & & \\ \frac{\ell_{21}}{\ell_{11}} & 1 & & \\ \vdots & \vdots & \ddots & \\ \frac{\ell_{n1}}{\ell_{11}} & \frac{\ell_{n2}}{\ell_{22}} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \ell_{11}u_{11} & & & \\ & \ell_{22}u_{22} & & \\ & & \ddots & \\ & & & \ell_{nn}\end{pmatrix}$$

Theorem 39 (PLU Factorisation). *For each* $\mathbf{A} \in \mathbb{C}^{n\times n}$, *there exists a permutation matrix* $\mathbf{P} \in \mathbb{R}^{n\times n}$, *a unit lower triangular* $\mathbf{L} \in \mathbb{C}^{n\times n}$ *and an upper triangular* $\mathbf{U} \in \mathbb{C}^{n\times n}$ *such that* $\mathbf{A} = \mathbf{P}\mathbf{L}\mathbf{U}$.

Theorem 40 (Cholesky factorisation). *Let* $\mathbf{A} \in \mathbb{C}^{n\times n}$ *be Hermitian. Then the following are true:*

- $\mathbf{A}$ *is positive semidefinite (respectively, positive definite) if and only if there is a lower triangular matrix* $\mathbf{L} \in \mathbb{C}^{n\times n}$ *with nonnegative (respectively, positive) diagonal entries such that* $\mathbf{A} = \mathbf{L}\mathbf{L}^*$;
- *Furthermore, if* $\mathbf{A}$ *is positive definite,* $\mathbf{L}$ *is unique, i.e. there is only one lower triangular matrix* $\mathbf{L}$ *with strictly positive diagonal entries such that* $\mathbf{A} = \mathbf{L}\mathbf{L}^*$;
- $\mathbf{A}$ *is real* $\implies$ $\mathbf{L}$ *is real.*

Proof. Is this even a theorem? ■

Theorem 41 (QR factorisation). *Let* $\mathbf{A} \in \mathbb{C}^{m\times n}$. *Then:*

- *There exists a unitary* $\mathbf{Q} \in \mathbb{C}^{m\times m}$ *and an upper triangular* $\mathbf{R} \in \mathbb{C}^{m\times n}$ *with nonnegative diagonal entries such that* $\mathbf{A} = \mathbf{Q}\mathbf{R}$.
- *If* $m \ge n$, *there exists a* $\mathbf{Q} \in \mathbb{C}^{m\times n}$ *with orthonormal columns and an upper triangular* $\mathbf{R} \in \mathbb{C}^{n\times n}$ *with nonnegative main diagonal entries such that* $\mathbf{A} = \mathbf{Q}\mathbf{R}$. *This is called* ***"Thin QR" or "Reduced QR"***.
- *If* $\mathrm{Rank}(\mathbf{A}) = n$, *then the factors* $\mathbf{Q}$ *and* $\mathbf{R}$ *are uniquely determined and the diagonal entries of* $\mathbf{R}$ *are all positive.*
- *If* $m = n$, *then the factor* $\mathbf{Q}$ *is unitary.*
- *If* $\mathbf{A}$ *is real, then the factors* $\mathbf{Q}$ *and* $\mathbf{R}$ *may be taken to be real.*

Fact 43 (Different forms of QR).

- Let $\mathbf{A} \in \mathbb{C}^{m\times n}$ with $\mathrm{Rank}(\mathbf{A}) = n \le m$. Then: $\mathbf{A} = \mathbf{Q}\mathbf{R} = \begin{pmatrix}\mathbf{Q}_1 & \mathbf{Q}_2\end{pmatrix}\begin{pmatrix}\mathbf{R}_1 \\ \mathbf{0}\end{pmatrix} = \mathbf{Q}_1\mathbf{R}_1$
- Let $\mathbf{A} \in \mathbb{C}^{m\times n}$ with $\mathrm{Rank}(\mathbf{A} = m \le n)$. Then: $\mathbf{A} = \mathbf{Q}\mathbf{R} = \mathbf{Q}\begin{pmatrix}\mathbf{R}_1 & \mathbf{R}_2\end{pmatrix}$

Fact 44 (Permuting QR). Let $\mathbf{A} \in \mathbb{C}^{m\times n}$ with $\mathrm{Rank}(\mathbf{A}) = r < n \le m$. Then we have $r_{ii} = 0$ for some i. One can permute the columns of $\mathbf{A}$ to obtain $\mathbf{A}\mathbf{P} = \mathbf{P}\begin{pmatrix}\mathbf{R}_1 & \mathbf{R}_2 \\ \mathbf{0} & \mathbf{0}\end{pmatrix} = \mathbf{Q}_r\begin{pmatrix}\mathbf{R}_1 & \mathbf{R}_2\end{pmatrix}$, where $\mathbf{P}$ is a permutation matrix, $\mathbf{R}_1 \in \mathbb{C}^{r\times r}$ is non-singular and upper triangular and $\mathbf{Q}_r \in \mathbb{C}^{m\times r}$ has orthonormal columns. Amazingly, this holds in the case where $\mathrm{Rank}(\mathbf{A}) = r < m \le n$.

Fact 45 (Two strokes of luck). We obtain "two strokes of luck" from this result:

1. We can obtain the inverse of $\tilde{\mathbf{L}}^{(i)}$ by simply taking the negative: $\tilde{\mathbf{L}}^{(2)} = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & \frac{-a_{32}^{(1)}}{a_{22}^{(1)}} & 1 & & \\ & \vdots & & \ddots & \\ & \frac{-a_{n2}^{(1)}}{a_{22}^{(1)}} & & & 1 \end{pmatrix} \implies$
   $$[\tilde{\mathbf{L}}^{(2)}]^{-1} = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & \frac{a_{32}^{(1)}}{a_{22}^{(1)}} & 1 & & \\ & \vdots & & \ddots & \\ & \frac{a_{n2}^{(1)}}{a_{22}^{(1)}} & & & 1 \end{pmatrix}$$
2. Let ℓ_k denote a vector with 0s above and at the diagonal, and $\ell_{k+1,k}$ below. It can be seen that a matrix formed with these vectors plus identity gives us the $\mathbf{L}$ matrix:

$$\boldsymbol{\ell}_k = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \ell_{k+1,k} \\ \vdots \\ \ell_{n,k} \end{bmatrix} \implies \left[\tilde{\mathbf{L}}^{(k)}\right]^{-1}\left[\tilde{\mathbf{L}}^{(k+1)}\right]^{-1} =$$

$$=$$

From this we can gather: $[\tilde{\mathbf{L}}^{(1)}]^{-1}[\tilde{\mathbf{L}}^{(2)}]^{-1} = \begin{pmatrix} 1 & & & & \\ \frac{a_{21}}{a_{11}} & 1 & & & \\ \frac{a_{31}}{a_{11}} & \frac{a_{32}^{(1)}}{a_{22}^{(1)}} & 1 & & \\ \vdots & \vdots & & \ddots & \\ \frac{a_{n1}}{a_{11}} & \frac{a_{n2}^{(1)}}{a_{22}^{(1)}} & & & 1 \end{pmatrix}$

Fact 46. For all $\mathbf{b} \in \mathbb{R}^p$ and $\mathbf{B} \in \mathbb{R}^{p\times p}$, we have:

$$f(\mathbf{x}) = \langle \mathbf{b}, \mathbf{x}\rangle \implies \nabla f(\mathbf{x}) = \mathbf{b}$$
$$f(\mathbf{x}) = \langle \mathbf{b}, \mathbf{B}\mathbf{x}\rangle \implies \nabla f(\mathbf{x}) = (\mathbf{B} + \mathbf{B}^\top)\mathbf{x}$$

Proposition 3. *Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be full column rank. We have $\kappa(\mathbf{A}^\top\mathbf{A}) = \kappa^2(\mathbf{A})$.*

Proof. Let $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$ be the economy SVD of $\mathbf{A}$. We have $\mathbf{A}^\top\mathbf{A} = \mathbf{V}\boldsymbol{\Sigma}^2\mathbf{V}^\top$ and hence $\kappa(\mathbf{A}^\top\mathbf{A}) = \frac{\sigma_1^2}{\sigma_n^2} = \kappa^2(\mathbf{A})$. ■

Fact 47 (General stationary iterations). $\mathbf{x}_{k+1} = \mathbf{M}^{-1}(\mathbf{N}\mathbf{x}_k + \mathbf{b}) = \mathbf{x}_k + \mathbf{M}^{-1}(\mathbf{b} - \mathbf{A}\mathbf{x}_k)$

Fact 48 (Types of iteration methods).

- $\mathbf{M} = \mathbf{D}$: Jacobi method (simultaneous relaxation)
- $\mathbf{M} = \mathbf{D} + \mathbf{E}$: Gauss-Seidel method (GS)
- $\mathbf{M} = \omega^{-1}\mathbf{D} + \mathbf{E}$: Successive over-relaxation (SOR). $0 < \omega < 2$ is necessary for convergence (for $\mathbf{A}$ that has nonzero diagonal elements), and sufficient for PD systems. $\omega = 1$ then just Gauss-Seidel - *the best results are usually obtained for* $1 \le \omega < 2$. There is also symmetric SOR (SSOR), and other variants.
- Block version of these splittings.

Proposition 4 (Sufficient condition on convergence). $\|\mathbf{T}\| < 1 \implies \lim_{k\to\infty} \mathbf{e}_k = 0$

Proof.

$$\begin{aligned} \|\mathbf{e}_{k+1}\| &= \left\|\mathbf{T}^{k+1}\mathbf{e}_k\right\| \\ &\le \left\|\mathbf{T}^{k+1}\right\| \|\mathbf{e}_0\| \\ &\le \|\mathbf{T}\|^{k+1} \|\mathbf{e}_0\| \end{aligned}$$

■

Theorem 42 (Necessary and sufficient condition on convergence). $\rho(\mathbf{T}) < 1 \Longleftrightarrow \lim_{k\to\infty} \mathbf{e}_k = 0$

Proof. It follows immediately from the fact that: $\lim_{k\to\infty} \mathbf{A}^k = \mathbf{0} \iff \rho(\mathbf{A}) < 1$ ■

Theorem 43 (Asymptotic rate of convergence). $\limsup_{k\to\infty} \left(\frac{\|\mathbf{e}_k\|}{\|\mathbf{e}_0\|}\right)^{1/k} \le \rho(\mathbf{T}), \quad \forall \mathbf{x}_0$

Theorem 44 (Cayley-Hamilton). *Let $p_n(\lambda) = \sum_{i=0}^n c_i\lambda^i$ be the characteristic polynomial of the matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$. Then we have $p_n(\mathbf{A}) = 0$.*

Theorem 45 (Grade of $\mathbf{v}$ with respect to $\mathbf{A}$). *There exists a positive integer $t \triangleq t(\mathbf{v}, \mathbf{A})$ called the grade of $\mathbf{v}$ with respect to $\mathbf{A}$ such that:* $\dim(\mathcal{K}_k(\mathbf{A}, \mathbf{v})) = \begin{cases} k & k \le t \\ t & k \ge t \end{cases}$ *In words, for all $k \le t$, the vectors forming a Krylov subspace, i.e. $\mathbf{A}^i\mathbf{v}, i = 0, \ldots, k-1$ remian linearly independent, i.e. they form a basis, and hence $\mathcal{K}_{k-1}(\mathbf{A}, \mathbf{v}) \subsetneq \mathcal{K}_k(\mathbf{A}, \mathbf{v})$. After the cutoff, new vectors will be linearly dependent on previous and hence for $k > t$: $\mathcal{K}_{k-1}(\mathbf{A}, \mathbf{v}) = \mathcal{K}_k(\mathbf{A}, \mathbf{v})$*

Proof. Suppose t is the smallest integer such that $\mathbf{A}^t\mathbf{v} + \sum_{i=0}^{t-1}\alpha_i\mathbf{A}^i\mathbf{v} = \mathbf{0}$ for some α_i. In other words, the vectors $\mathbf{v}, \mathbf{A}\mathbf{v}, \mathbf{A}^2\mathbf{v}, \ldots, \mathbf{A}^t\mathbf{v}$ are linearly dependent. So we must have that $\dim(\mathcal{K}_{t+1}(\mathbf{A}, \mathbf{v})) \le t$. It easily follows that: $\mathbf{A}^{t+1}\mathbf{v} + \sum_{i=0}^{t-1}\alpha_i\mathbf{A}^{i+1}\mathbf{v} = \mathbf{A}\left(\mathbf{A}^t\mathbf{v} + \sum_{i=0}^{t-1}\alpha_i\mathbf{A}^i\mathbf{v}\right) = \mathbf{0}$ In other words, the vectors $\mathbf{A}\mathbf{v}, \ldots, \mathbf{A}^{t+1}\mathbf{v}$ will also be linearly dependent, which in turn implies that $\mathbf{v}, \mathbf{A}\mathbf{v}, \ldots, \mathbf{A}^{t+1}\mathbf{v}$ are linearly dependent. So we must have that $\dim(\mathcal{K}_{t+1}(\mathbf{A}, \mathbf{v})) \le t$. We can continue this way, hence we have $\dim(\mathcal{K}_k(\mathbf{A}, \mathbf{v})) \le t, \forall k \ge t$.

Since t is the smallest integer with such property, for any $k < t$, we have $\mathbf{A}^k\mathbf{v} + \sum_{i=0}^{k-1}\alpha_i\mathbf{A}^i\mathbf{v} \ne \mathbf{0}$ for **all** $\alpha_i, i = 0, \ldots, k-1$. This implies that all the vectors $\mathbf{v}, \mathbf{A}\mathbf{v}, \mathbf{A}^2\mathbf{v}, \ldots, \mathbf{A}^k\mathbf{v}$ are linearly independent. Indeed, consider any $\alpha_i, i = 0, \ldots, k$ with $\alpha_k \ne 0$. From the above assumption, we have: $\alpha_k\mathbf{A}^k\mathbf{v} + \sum_{i=0}^{k-1}\alpha_i\mathbf{A}^i\mathbf{v} = \mathbf{A}^k\mathbf{v} + \sum_{i=0}^{k-1}\frac{\alpha_i}{\alpha_k}\mathbf{A}^i\mathbf{v} \ne \mathbf{0}$ Now consider the case where $\alpha_k = 0$ and suppose $\sum_{i=0}^{k-1}\alpha_i\mathbf{A}^i\mathbf{v} = \mathbf{0}$ for some $\alpha_i, i = 0, \ldots, k-1$ that are not all zero. Let i be the largest index with non-zero α_i. We have $\mathbf{A}^i\mathbf{v} = \sum_{\ell=0}^{i}\left(\frac{\alpha_\ell}{\alpha_i}\right)\mathbf{A}^\ell\mathbf{v}$ which contradicts the assumption on t. So $\dim(\mathcal{K}_k(\mathbf{A}, \mathbf{v})) = k \; \forall k \le t$. ■

Corollary 1. $t = \min\{k \mid \mathbf{A}^{-1}\mathbf{v} \in \mathcal{K}_k(\mathbf{A}, \mathbf{v})\}$

Proof. Recall that an application of the Cayley-Hamilton theorem implied that: $\mathbf{A}^{-1}\mathbf{v} = \sum_{i=0}^{n-1}\alpha_i\mathbf{A}^i\mathbf{v}$ But since $\mathcal{K}_k(\mathbf{A}, \mathbf{v}) = \mathcal{K}_{K+1}(\mathbf{A}, \mathbf{v}), k \ge t$, we can write: $\mathbf{A}^{-1}\mathbf{v} = \sum_{i=0}^{t-1}\beta_i\mathbf{A}^i\mathbf{v}$ So $\mathbf{A}^{-1}\mathbf{v} \in \mathcal{K}_k(\mathbf{A}, \mathbf{v}), k \ge t$. Now suppose this also holds for $k = t-1$, i.e. $\mathbf{A}^{-1}\mathbf{v} = \sum_{i=0}^{t-2}\gamma_i\mathbf{A}^i\mathbf{v}$. But then this gives $\mathbf{v} = \sum_{i=0}^{t-2}\gamma_i\mathbf{A}^{i+1}\mathbf{v}$. In other words, $\{\mathbf{v}, \mathbf{A}\mathbf{v}, \ldots, \mathbf{A}^{t-1}\mathbf{v}\}$ are linearly dependent, which implies $\dim(\mathcal{K}_t(\mathbf{A}, \mathbf{v})) < t$ which is a contradiction. ■

Corollary 2. *For any $\mathbf{x}_0$, we have $\mathbf{x}^\star \in \mathbf{x}_0 + \mathcal{K}_t(\mathbf{A}, \mathbf{r}_0)$ where $\mathbf{r}_0 = b - \mathbf{A}\mathbf{x}_0$ and t is the grade of $\mathbf{r}_0$ with respect to $\mathbf{A}$.*

Theorem 46. *Assume the Arnoldi process does not terminate before k steps. Then the vectors $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_k\}$ form an orthonormal basis for $\mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)$.*

Proof. First note that $\mathcal{K}_k(\mathbf{A}, \mathbf{r}_0) = \mathcal{K}_k(\mathbf{A}, \mathbf{q}_1)$. Orthonormality is clear from the construction. For $j = 1$, we trivially have $\mathbf{q}_1 = p_0(\mathbf{A})\mathbf{q}_1$, where $p_{i-1}(t)$ $p_0(\mathbf{A}) = \mathbf{1}$. Suppose for all $i \le j$ we have $\mathbf{q}_i = p_{i-1}(\mathbf{A})\mathbf{q}_1$, where $p_{i-1}(t)$ is a polynomial of degree $i-1$. For $j+1$, it follows that $h_{j+1,j}\mathbf{q}_{i+1} = \mathbf{A}\mathbf{q}_j - \sum_{i=1}^{j}h_{ij}\mathbf{q}_i = \mathbf{A}p_{j-1}(\mathbf{A})\mathbf{q}_1 - \sum_{i=1}^{j}h_{ij}p_{i-1}(\mathbf{A})\mathbf{q}_i$ so we have $\mathbf{q}_{j+1} = p_j(\mathbf{A})\mathbf{q}_1$. In other words, each column of $\mathbf{Q}_k$ can be written as linear combination of vectors $\left\{\mathbf{q}_1, \mathbf{A}\mathbf{q}_1, \ldots, \mathbf{A}^{k-1}\mathbf{q}_1\right\}$, and since $\mathbf{q}_j$'s are independent, they must span the same space, i.e., $\mathcal{K}_k(\mathbf{A}, \mathbf{q}_1)$. ■

Theorem 47. *The Arnoldi process breaks down at step j, i.e. $h_{j+1,j} = 0$ if and only if the grade of $\mathbf{r}_0$ with respect to $\mathbf{A}$ is j, i.e. $t(\mathbf{r}_0, \mathbf{A}) = j$.*

Proof. ($\Longleftarrow$) First, note that $t(\mathbf{r}_0, \mathbf{A}) = t(\mathbf{q}_1, \mathbf{A})$. Suppose $t(\mathbf{q}_1, \mathbf{A}) = j$ which implies $\dim(\mathcal{K}_{j+1}(\mathbf{A}, \mathbf{q}_1)) = j$. Hence, we must have $\mathbf{A}\mathbf{q}_j - \sum_{i=1}^{j}h_{ij}\mathbf{q}_i = \mathbf{0}$. Otherwise $\mathbf{q}_{i+1}$ could be defined, which in turn implies that $\dim(\mathcal{K}_{j+1}(\mathbf{A}, \mathbf{q}_1)) = \dim(\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_{j+1}\}) = j+1$, which is a contradiction. Hence we get $h_{j+1,j} = \left\|\mathbf{A}\mathbf{q}_j - \sum_{i=1}^{j}h_{ij}\mathbf{q}_i\right\| = 0$.

($\Longrightarrow$) To prove the converse, suppose $h_{j+1,j} = 0$, which means $\mathrm{A}\mathrm{q}_j - \sum_{i=1}^{j}h_{ij}\mathbf{q}_i = 0$. Now since by previous theorem, $\mathrm{Span}\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_j\} = \mathcal{K}_j(\mathbf{A}, \mathbf{q}_1)$, we have $\mathbf{A}\mathbf{q}_j \in \mathcal{K}_j(\mathbf{A}, \mathbf{q}_1)$. But similar to the proof of the previous theorem, we can get that $\mathbf{A}\mathbf{q}_j = p_j(\mathbf{A})\mathbf{q}_1$, where $p_j(\mathbf{A})$ is a matrix polynomial of degree exactly j. This in particular implies $\mathbf{A}^j\mathbf{q}_1 \in \mathcal{K}_j(\mathbf{A}, \mathbf{q}_1)$. Hence, we must have $t(\mathbf{q}_1, \mathbf{A}) \le j$. However, we cannot have $t(\mathbf{q}_1, \mathbf{A}) < j$, as otherwise by the first part of the proof, the algorithms would have already stopped. ■

Proposition 5. *The matrix $\mathbf{L}^\top\mathbf{A}\mathbf{K}$ is non-singular if either:*

1. *$\mathbf{A} \succ \mathbf{0}$ and $\mathcal{L} = \mathcal{K}$ or;*
2. *$\det(\mathbf{A}) \ne 0$ and $\mathcal{L} = \mathbf{A}\mathcal{K}$.*

Proof.

1. Since $\mathcal{L} = \mathcal{K}$, any basis of $\mathcal{L}$ is also a basis for $\mathcal{K}$. In fact, we can write $\mathbf{L} = \mathbf{K}\mathbf{B}$ where $\mathbf{B} \in \mathbb{R}^{k\times k}$ is non-singular. Now, we have $\mathbf{L}^\top\mathbf{A}\mathbf{K} = \mathbf{B}^\top\mathbf{K}^\top\mathbf{A}\mathbf{K}$ and since $\mathbf{A} \succ 0$, we have $\mathbf{K}^\top\mathbf{A}\mathbf{K} \succ \mathbf{0}$ and hence the entire product is non-singular.
2. Since $\mathcal{L} = \mathbf{A}\mathcal{K}$, we can write $\mathbf{L} = \mathbf{A}\mathbf{K}\mathbf{B}$ where $\mathbf{B} \in \mathbb{R}^{k\times k}$ is non-singular. Now we have $\mathbf{L}^\top\mathbf{A}\mathbf{K} = \mathbf{B}^\top\mathbf{K}^\top\mathbf{A}^\top\mathbf{A}\mathbf{K}$ and since $\mathbf{A}$ is non-singular, we have $\mathbf{A}^\top\mathbf{A} \succ \mathbf{0}$, which as above, implies that the entire product is non-singular.

■

Theorem 48. *The case where $\mathbf{A} \succ \mathbf{0}$ and $\mathcal{L}_k = \mathcal{K}_k$ is equivalent to $\mathbf{x}_k = \underset{\mathbf{x}\in\mathbf{x}_0+\mathcal{K}_k}{\arg\min}\|\mathbf{x} - \mathbf{x}^\star\|_{\mathbf{A}} = \underset{\mathbf{x}\in\mathbf{x}_0+\mathcal{K}_k}{\arg\min}\frac{1}{2}\langle\mathbf{x}, \mathbf{A}\mathbf{x}\rangle - \langle\mathbf{b}, \mathbf{x}\rangle$*

Proof. Let $\mathbf{x} = \mathbf{x}_0 + \mathbf{K}_k\mathbf{y}$ for $\mathbf{y} \in \mathbb{R}^k$, where $\mathbf{K}_k$ is a basis matrix for $\mathcal{K}_k$. So:

$$\begin{aligned} \mathbf{x}_k &= \underset{\mathbf{x}\in\mathbf{x}_0+\mathcal{K}_k}{\arg\min}\frac{1}{2}\langle\mathbf{x}, \mathbf{A}\mathbf{x}\rangle - \langle\mathbf{b}, \mathbf{x}\rangle \\ &= \underset{\mathbf{y}\in\mathbb{R}^k}{\arg\min}\frac{1}{2}\langle\mathbf{x}_0 + \mathbf{K}_k\mathbf{y}, \mathbf{A}(\mathbf{x}_0 + \mathbf{K}_k\mathbf{y})\rangle - \langle\mathbf{b}, \mathbf{x}_0 + \mathbf{K} \\ &= \underset{\mathbf{y}\in\mathbb{R}^k}{\arg\min}\frac{1}{2}\langle\mathbf{K}_k\mathbf{y}, \mathbf{A}\mathbf{K}_k\mathbf{y}\rangle - \langle\mathbf{b} - \mathbf{A}\mathbf{x}_0, \mathbf{K}\mathbf{y}\rangle \end{aligned}$$

Since $\mathbf{A} \succ \mathbf{0}$, it is necessary and sufficient for the optimal $\mathbf{y}_k$ to satisfy $\mathbf{K}_k^\top\mathbf{A}\mathbf{K}_k\mathbf{y}_k - \mathbf{K}_k^\top(\mathbf{b} - \mathbf{A}\mathbf{x}_0) = \mathbf{0}$, which is the same as $\mathbf{K}_k^\top(\mathbf{A}\mathbf{x}_k - \mathbf{b}) = \mathbf{0}$, i.e., $\mathbf{r}_k \perp \mathcal{K}_k$ ■

Theorem 49. *The case where $\det(\mathbf{A}) \ne 0$ and $\mathcal{L}_k = \mathbf{A}\mathcal{K}_k$ is equivalent to $\mathbf{x}_k = \underset{\mathbf{x}\in\mathbf{x}_0+\mathcal{K}_k}{\arg\min}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2$*

Proof. Similarly as above, let $\mathbf{x} = \mathbf{x}_0 + \mathbf{K}_k\mathbf{y}$ for $\mathbf{y} \in \mathbb{R}^k$, where $\mathbf{K}_k$ is a basis matrix for $\mathcal{K}_k$:

$$\begin{aligned} \mathbf{x}_k &= \underset{\mathbf{x}\in\mathbf{x}_0+\mathcal{K}_k}{\arg\min}\frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 \\ &= \underset{\mathbf{y}\in\mathbb{R}^k}{\arg\min}\frac{1}{2}\|\mathbf{A}(\mathbf{x}_0 + \mathbf{K}_k\mathbf{y}) - \mathbf{b}\|^2 \end{aligned}$$

Now, since A is non-singular, it is necessary and sufficient for the optimal y_k to satisfy $\mathbf{K}_k^\top\mathbf{A}^\top\mathbf{A}\mathbf{K}\mathbf{y}_k = \mathbf{K}_k^\top\mathbf{A}^\top(\mathbf{b} - \mathbf{A}\mathbf{x}_0)$ which is the same as $\mathbf{K}_k^\top\mathbf{A}^\top(\mathbf{A}\mathbf{x}_k - \mathbf{b}) = \mathbf{0}$, i.e., $\mathbf{r}_k \perp \mathbf{A}\mathcal{K}_k$ ■

Proposition 6. *If Arnoldi (or Lanczos) Process breaks down at step $t = t(\mathbf{A}, \mathbf{r}_0)$, then $\mathbf{x}_t$ from any projection method onto $\mathcal{K}_t(\mathbf{A}, \mathbf{r}_0)$ or $\mathbf{A} \cdot \mathcal{K}_t(\mathbf{A}, \mathbf{r}_0)$ would be exact.*

Proof. We show the proof for Arnoldi, as that for Lancsoz is identical. First consider the projection method onto $\mathcal{K}_t(\mathbf{A}, \mathbf{r}_0)$, i.e., $\mathcal{L}_t = \mathcal{K}_t$. Recall again that $\mathbf{x}_t = \mathbf{x}_0 + \mathbf{Q}_t\mathbf{y}$, $\mathrm{Range}(\mathbf{Q}_t) = \mathcal{K}_t(\mathbf{A}, \mathbf{r}_0), \mathbf{y} \in \mathbb{R}^t$. Since $\mathbf{r}_0 \in \mathrm{Range}(\mathbf{Q}_t)$, we have $\mathbf{Q}_t\mathbf{Q}_t^\top\mathbf{r}_0 = \mathbf{r}_0$. Also, since $h_{t+1,t} = 0$, we have $\mathbf{A}\mathbf{Q}_t = \mathbf{Q}_{t+1}\mathbf{H}_{t+1,t} = \mathbf{Q}_t\mathbf{H}_t$. It follows that $\mathbf{Q}_t\mathbf{Q}_t^\top(\mathbf{A}\mathbf{Q}_t\mathbf{y}) = \mathbf{Q}_t\mathbf{H}_t\mathbf{y} = \mathbf{A}\mathbf{Q}_t\mathbf{y}$. Hence, we have:

$$\begin{aligned} \mathbf{0} = \mathbf{Q}_t^\top(\mathbf{r}_0 - \mathbf{A}\mathbf{Q}_t\mathbf{y}) \Longleftrightarrow \mathbf{0} &= \mathbf{Q}_t\mathbf{Q}_t^\top(\mathbf{r}_0 - \mathbf{A}\mathbf{Q}_t\mathbf{y}) \\ &= \mathbf{r}_0 - \mathbf{A}\mathbf{Q}_t\mathbf{y} \\ &= \mathbf{b} - \mathbf{A}\mathbf{x}_0 - \mathbf{A}\mathbf{Q}_t\mathbf{y} \\ &= \mathbf{b} - \mathbf{A}\mathbf{x}_t \end{aligned}$$

In other words, $\mathbf{x}_t$ is the exact solution. Now consider the projection method onto $\mathbf{A} \cdot \mathcal{K}_t(\mathbf{A}, \mathbf{r}_0)$. Just as before, we get:

$$\begin{aligned} \mathbf{y} &= \underset{\mathbf{y}\in\mathbb{R}^t}{\arg\min}\frac{1}{2}\left\|\mathbf{H}_{t+1,t}\mathbf{y} - \mathbf{Q}_{t+1}^\top\mathbf{r}_0\right\|^2 \\ &= \underset{\mathbf{y}\in\mathbb{R}^t}{\arg\min}\frac{1}{2}\left\|\mathbf{H}_t\mathbf{y} - \mathbf{Q}_t^\top\mathbf{r}_0\right\|^2 \\ &= \mathbf{H}_t^{-1}\mathbf{Q}_t^\top\mathbf{r}_0 \end{aligned}$$

where the last equality follows since H_t is an invertible square matrix. Again, noting that $\mathbf{r}_0 \in \mathrm{Range}(\mathbf{Q}_t)$, we have

$$\begin{aligned} \mathbf{H}_t\mathbf{y} = \mathbf{Q}_t^\top\mathbf{r}_0 &\Longleftrightarrow \mathbf{Q}_t\mathbf{H}_t\mathbf{y} = \mathbf{Q}_t\mathbf{Q}_t^\top\mathbf{r}_0 = \mathbf{r}_0 \Longleftrightarrow \mathbf{A}\mathbf{Q}_t\mathbf{y} = \\ &\Longleftrightarrow \mathbf{A}(\mathbf{x}_0 + \mathbf{Q}_t\mathbf{y}) = \mathbf{b} \Longleftrightarrow \mathbf{A}\mathbf{x}_t = \mathbf{b} \end{aligned}$$

In other words, $\mathbf{x}_t$ is the exact solution. ■