MATH262 Cheat Sheet

Oğul Can Yurdakul

Last Updated: 01/07/2021

These are the summaries I made from when was taking the MATH262 Linear Algebra II course from Ali Ulaş Özgür Kişisel, during Spring 2020-2021 term.

Contents

1 Uniqueness of the Determinant

1.1 Multilinear Algebra

Free Vector Space: Let F be a vector space and S any set. The subset $\mathcal{F}(S)$ of the vector space $\text{Fun}(S, F)$ will be defined as follow

$$\mathcal{F}(S) = \{f \in \text{Fun}(S, F) : \text{there exists finitely many } s \in S \text{ such that } f(s) \neq 0\}$$

$\mathcal{F}(S)$ is called as "the set of functions from S to F with finite support".

Subspace: $\mathcal{F}(S)$ is a subspace of $\mathrm{Fun}(S, F)$, called the free vector space over S with F-coefficients.

An alternative notation for the elements of $\mathcal{F}(S)$: Suppose $f \in \mathcal{F}(S)$ and f is non-zero only at $\{\boldsymbol{s}_1, \dots, \boldsymbol{s}_m\}$, and that $f(\boldsymbol{s}_1) = c_1, \dots, f(\boldsymbol{s}_m) = c_m$. Then denote f by

$$f = c_1\boldsymbol{s}_1 + c_2\boldsymbol{s}_2 + \cdots + c_m\boldsymbol{s}_m$$

and remember that this is just a formal expression (sum)! For example, if

$$g = d_1\boldsymbol{s}_1 + \cdots + d_n\boldsymbol{s}_n$$

then

$$f + g = (c_1 + d_1)\boldsymbol{s}_1 + \cdots + (c_m + d_m)\boldsymbol{s}_m + d_{m+1}\boldsymbol{s}_{m+1} + \cdots + d_n\boldsymbol{s}_n$$

et cetera.

Basis: Say F is a field, S is a set, $\mathcal{F}(S)$ is the free vector space as above. Then

$$\mathcal{B} = \{1\boldsymbol{s}\}_{\boldsymbol{s}\in S}$$

is a basis for $\mathcal{F}(S)$.

Universal Property: Let F be a field, S any set and V a vector space on F. Suppose $\varphi : S \to V$ is any function. Then there exists a unique, linear transformation $T_\varphi : \mathcal{F}(S) \to V$ such that

$$T_\varphi(1\boldsymbol{s}) = \varphi(S) \quad \forall \boldsymbol{s} \in S$$

Tensor Product: Let V, W be vector spaces over a field F, with $\mathcal{F}(V \times W)$ the free vector space of their Cartesian product. Let $\mathcal{U}$ be the subspace of $\mathcal{F}(V \times W)$ spanned by all the elements of the following form:

i) $1(\boldsymbol{v}_1 + \boldsymbol{v}_2, \boldsymbol{w}) - 1(\boldsymbol{v}_1, \boldsymbol{w}) - 1(\boldsymbol{v}_2, \boldsymbol{w})$
ii) $1(\boldsymbol{v}, \boldsymbol{w}_1 + \boldsymbol{w}_2) - 1(\boldsymbol{v}, \boldsymbol{w}_1) - 1(\boldsymbol{v}, \boldsymbol{w}_2)$
iii) $1(c\boldsymbol{v}, \boldsymbol{w}) - c(\boldsymbol{v}, \boldsymbol{w})$
iv) $1(\boldsymbol{v}, c\boldsymbol{w}) - c(\boldsymbol{v}, \boldsymbol{w})$

Then the quotient space

$$V \otimes W = {}^{\mathcal{F}(V\times W)}/_{\mathcal{U}}$$

is called the tensor product of the vector spaces V and W. A typical element in the tensor product is denoted as follows:

$$\sum_i c_i\boldsymbol{v}_i \otimes \boldsymbol{w}_i := \sum_i c_i(\boldsymbol{v}_i, \boldsymbol{w}_i)$$

Properties: The elements that are "quotiened-out" make sure the following properties hold:

i) $(\boldsymbol{v}_1 + \boldsymbol{v}_2) \otimes \boldsymbol{w} = \boldsymbol{v}_1 \otimes \boldsymbol{w} + \boldsymbol{v}_2 \otimes \boldsymbol{w}$
ii) $\boldsymbol{v} \otimes (\boldsymbol{w}_1 + \boldsymbol{w}_2) = \boldsymbol{v} \otimes \boldsymbol{w}_1 + \boldsymbol{v} \otimes \boldsymbol{w}_2$
iii) $(c\boldsymbol{v}) \otimes \boldsymbol{w} = c(\boldsymbol{v} \otimes \boldsymbol{w})$
iv) $\boldsymbol{v} \otimes (c\boldsymbol{w}) = c(\boldsymbol{v} \otimes \boldsymbol{w})$

so the elements of the tensor product satisfy linearity in both of their components!

Dimension and Basis: Say V and W are finite dimensional vector spaces over some field F and their bases are $\{v_1, \dots, v_n\}$ and $\{w_1, \dots, w_m\}$. Then

$$\mathcal{B} = \{v_i \otimes w_j\}_{1\le i,j\le n}$$

is a basis for $V \otimes W$, and hence

$$\dim(V \otimes W) = nm = \dim(V)\dim(W)$$

Bilinear Map: Say V, W and Z are vector spaces over some field F. Then a map $\psi : V \times W \to Z$ is called bilinear if it satisfies linearity in both of its arguments, meaning

$$\psi(c_1v_1 + c_2v_2, w) = c_1\psi(v_1, w) + c_2\psi(v_2, w)$$
$$\psi(v, c_1w_1 + c_2w_2) = c_1\psi(v, w_1) + c_2\psi(v, w_2)$$

The map defined by $\psi(v, w) = v \otimes w$ is naturally bilinear.

Universal Property of the Tensor Product: Say V, W and Z are vector spaces over some field F and $\psi : V \times W \to Z$ is a bilinear map. Then, there exists a unique linear transformation $T_\psi : V \otimes W \to Z$ such that

$$T_\psi(v \otimes w) = \psi(v, w) \quad \forall v \in V \forall w \in W$$

Symmetric Power: First we work with the symmetric square, then move on to higher powers.

Symmetric Square: Let V be a vector field over some field F. The second symmetric power of V, or its symmetric square is defined as

$$\mathrm{Sym}^2(V) = {}^{V\otimes V}/_{\mathcal{U}}$$

where $\mathcal{U}$ is the subspace of $V \otimes V$ spanned by elements of the form

$$v \otimes w - w \otimes v$$

This ensures that in this quotient space, we have

$$v \otimes w + \mathcal{U} = w \otimes v + \mathcal{U}$$

where the equivalence class $v \otimes w + \mathcal{U}$ is denoted by vw.

Dimension and Basis: For $V = \mathrm{Span}(\{v_1, v_2, \dots, v_n\})$ meaning $\dim(V) = n$, the basis for $\mathrm{Sym}^2(V)$ is given by

$$\mathcal{B} = \{v_iv_j\}_{1\leq i\leq j\leq n}$$

because remember that $v_iv_j = v_jv_i$, we cannot include both of them, and one way to get rid of the unnecessary elements is to impose a partial ordering on the indices. And so, we obtain

$$\dim(V) = \frac{n(n+1)}{2}$$

Universal Property: Say V and Z are vector spaces over some field F and $\psi : V^2 \to Z$ is a bilinear map such that $\psi(v, w) = \psi(w, v)$ for all $v, w \in W$. Then, there exists a unique linear transformation $T_\psi : \mathrm{Sym}^2(V) \to Z$ such that

$$T_\psi(vw) = \psi(v, w) \quad \forall v, w \in V$$

Higher Symmetric Powers: Let V be a vector field over some field F. The k-th symmetric power of V is defined as

$$\mathrm{Sym}^k(V) = {}^{V^{\otimes k}}/_{\mathcal{U}}$$

where

$$V^{\otimes k} = \underbrace{V \otimes V \otimes \cdots \otimes V}_{k \text{ times}}$$

(the order of tensor products is not important, i.e. we consider upto isomorphism) and $\mathcal{U}$ is the subspace of $V^{\otimes k}$ spanned by elements of the form

$$u \otimes \cdots \otimes \underbrace{(v \otimes w - w \otimes v)}_{\text{this part can be anywhere}} \otimes \cdots \otimes r$$

This ensures that in this quotient space, we have

$$v_1 \otimes \cdots \otimes v_i \otimes v_{i+1} \otimes \cdots \otimes v_k + \mathcal{U} = v_1 \otimes \cdots \otimes v_{(i+1)} \otimes v_i \otimes \cdots \otimes v_k + \mathcal{U}$$

where the equivalence class $v_1 \otimes \cdots \otimes v_i \otimes v_{(i+1)} \otimes \cdots \otimes v_k + \mathcal{U}$ is denoted by $v_1 v_2 \dots v_k$. Notice that because we satisfy equality for adjacent transpositions, we satisfy equality for all permutations as any permutation can be written as a composition of adjacent transpositions.

Dimension and Basis: For $V = \mathrm{Span}(\{v_1, v_2, \dots, v_n\})$ meaning $\dim(V) = n$, the basis for $\mathrm{Sym}^k(V)$ is given by

$$\mathcal{B} = \{v_{i_1} v_{i_2} \dots v_{i_k}\}_{1 \le i_1 \le i_2 \le \cdots \le i_k \le n}$$

And so, we obtain

$$\dim(\mathrm{Sym}^k(V)) = \frac{n(n+1)\dots(n+k-1)}{k!} = \binom{n+k-1}{k}$$

which is the total number of unordered repeated k-selections of n objects, and that makes sense.

Exterior Power: Again, first the exterior square, then higher exterior powers.

Exterior Square: Say V is a vector space over some field F. The second exterior power (or the second alternating power) is defined as

$$\mathrm{Alt}^2(V) = {}^{V \otimes V}/_{\mathcal{U}}$$

where $\mathcal{U}$ is the subspace of $V \otimes V$ spanned by elements of the form

$$v \otimes w + w \otimes v$$

This ensures that in this quotient space, we have

$$v \otimes w + \mathcal{U} = -(w \otimes v) + \mathcal{U}$$

where the equivalence class $v \otimes w + \mathcal{U}$ is denoted by $v \wedge w$.

Dimension and Basis: For $V = \mathrm{Span}(\{v_1, v_2, \dots, v_n\})$ meaning $\dim(V) = n$, the basis for $\mathrm{Alt}^2(V)$ is given by

$$\mathcal{B} = \{v_i v_j\}_{1 \le i < j \le n}$$

because remember that $v_i \wedge v_j = -v_j \wedge v_i$ so we cannot include both of them. Also, $v_i \wedge v_i = -v_i \wedge v_i \;\Rightarrow\; v_i \wedge v_i = 0$, so we don't have the "wedge squares" of the elements as well. One way to get rid of the unnecessary elements is to impose a strict ordering on the indices. And so, we obtain

$$\dim(V) = \frac{n(n-1)}{2} = \binom{n}{2}$$

Universal Property: Say V and Z are vector spaces over some field F and $\psi : V^2 \to Z$ is a bilinear map such that $\psi(v,w) = -\psi(w,v)$ for all $v, w \in W$. Then, there exists a unique linear transformation $T_\psi : \mathrm{Alt}^2(V) \to Z$ such that

$$T_\psi(v \wedge w) = \psi(v,w) \quad \forall v, w \in V$$

Higher Exterior Powers: Let V be a vector field over some field F. The k-th exterior power of V is defined as

$$\mathrm{Alt}^k(V) = {}^{V^{\otimes k}}\!/\mathcal{U}$$

where

$$V^{\otimes k} = \underbrace{V \otimes V \otimes \cdots \otimes V}_{k \text{ times}}$$

and $\mathcal{U}$ is the subspace of $V^{\otimes k}$ spanned by elements of the form

$$u \otimes \cdots \otimes \underbrace{(v \otimes w + w \otimes v)}_{\text{this part can be anywhere}} \otimes \cdots \otimes r$$

This ensures that in this quotient space, we have

$$v_1 \otimes \cdots \otimes v_i \otimes v_{i+1} \otimes \cdots \otimes v_k + \mathcal{U} = -(v_1 \otimes \cdots \otimes v_{(i+1)} \otimes v_i \otimes \cdots \otimes v_k) + \mathcal{U}$$

where the equivalence class $v_1 \otimes \cdots \otimes v_i \otimes v_{(i+1)} \otimes \cdots \otimes v_k + \mathcal{U}$ is denoted by $v_1 \wedge v_2 \wedge \cdots \wedge v_k$. Notice that each adjacent transposition flips the sign once, and so we obtain a general formula using the sign of the permutations as

$$v_1 \wedge v_2 \wedge \cdots \wedge v_k = \mathrm{sgn}(\sigma) v_{\sigma(1)} \wedge v_{\sigma(2)} \wedge \cdots \wedge v_{\sigma(k)}$$

where $\sigma \in \mathrm{Sym}(k)$ is a permutation, and its sign $\mathrm{sgn}(\sigma)$ is given by

$$\mathrm{sgn}(\sigma) = \begin{cases} 1, & \text{if } \sigma \text{ is a composition of an even number of transpositions} \\ -1, & \text{if } \sigma \text{ is a composition of an odd number of transpositions} \end{cases}$$

Dimension and Basis: For $V = \mathrm{Span}(\{v_1, v_2, \ldots, v_n\})$ meaning $\dim(V) = n$, the basis for $\mathrm{Alt}^k(V)$ is given by

$$\mathcal{B} = \{v_{i_1} \wedge v_{i_2} \wedge \cdots \wedge v_{i_k}\}_{1 \le i_1 < i_2 < \cdots < i_k \le n}$$

And so, we obtain

$$\dim(\mathrm{Alt}^k(V)) = \frac{n(n-1)\ldots(n-k+1)}{k!} = \binom{n}{k}$$

which is the total number of unordered unrepeated k-selections of n objects, and that makes sense.

One interesting thing to note is that

i) We cannot arbitrarily increase the dimension of an exterior power of a finite dimensional space, as it follows the binomial coefficients. For $\dim(V) = n$, the exterior power $\mathrm{Alt}^k(V)$ reduces to the null vector space for $k > n$ as we necessarily repeat at least one basis element of V in all of the basis elements of $\mathrm{Alt}^k(V)$, so all the basis elements must equal to the zero vector. This is not the case for the symmetric power $\mathrm{Sym}^k(V)$, as its dimensionality always increases with k.

ii) Because the dimensionality of $\mathrm{Alt}^k(V)$ follows the binomial coefficients $\binom{n}{k}$, it decreases after some point and eventually becomes 1 again.

1.2 The True Nature of the Determinant

Determinant: Let F be a field and $n \in \mathbb{Z}^+$. Then, the determinant is a function over matrices $\det : M_{n\times n}(F) \to F$ given for each $A = (a_{ij}) \in M_{n\times n}(F)$ as

$$\det(A) = \sum_{\sigma \in \mathrm{Sym}(n)} \mathrm{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{n\sigma(n)}$$

Determinants of the Elementary Matrices:

Row multiplication: The only non-zero permutation is the identity permutation.

$$(cR_i \to R_i) \;\Rightarrow\; E = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & c & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} \;\Rightarrow\; \det(E) = c$$

Row switching: The only non-zero permutation is the single relevant transposition ($\sigma = (i\ j)$), which has a negative sign.

$$(R_i \leftrightarrow R_j) \;\Rightarrow\; E = \begin{bmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 0 & & \cdots & & 1 & \\ & & & 1 & & & & \\ & & \vdots & & \ddots & & \vdots & \\ & & & & & 1 & & \\ & & 1 & & \cdots & & 0 & \\ & & & & & & & \ddots & \\ & & & & & & & & 1 \end{bmatrix} \;\Rightarrow\; \det(E) = -1$$

Row sum: The only non-zero permutation is the identity permutation, all others that possibly incorporate the c entry have zero multiplicand.

$$(cR_i + R_j \to R_j) \;\Rightarrow\; E = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & \vdots & \ddots & & \\ & & c & \dots & 1 & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} \;\Rightarrow\; \det(E) = 1$$

Properties of Determinant: Recall that applying an elementary row operation on a matrix is equivalent to multiplying it from the left with the relevant elementary matrix.
All the properties of the determinant that are skipped before are proven using this definition and the elementary matrices above.

- Transpose:

$$\det(A) = \det(A^T)$$

- Multiplicativity:
$$\det(AB) = \det(A)\det(B)$$
- Row/column expansion formula:
$$\det(A) = \sum_{i=1}^{n}(-1)^{(i+j)}a_{ij}\det(A_{ij}) = \sum_{j=1}^{n}(-1)^{(i+j)}a_{ij}\det(A_{ij})$$

Determinant and the Exterior Power: Write $A = \begin{bmatrix} R_1^T & R_2^T & \cdots & R_n^T \end{bmatrix}^T$ as a stack of its rows. Then we can consider det as an n-ary function of these rows, and it satisfies the following properties:

- The determinant is multilinear on the rows:
$$\begin{vmatrix} R_1 \\ \vdots \\ c_1 R_i + c_2 \hat{R}_i \\ \vdots \\ R_n \end{vmatrix} = c_1 \begin{vmatrix} R_1 \\ \vdots \\ R_i \\ \vdots \\ R_n \end{vmatrix} + c_2 \begin{vmatrix} R_1 \\ \vdots \\ \hat{R}_i \\ \vdots \\ R_n \end{vmatrix}$$
- The determinant is alternating on the rows:
$$\begin{vmatrix} R_1 \\ \vdots \\ R_i \\ \vdots \\ R_j \\ \vdots \\ R_n \end{vmatrix} = - \begin{vmatrix} R_1 \\ \vdots \\ R_j \\ \vdots \\ R_i \\ \vdots \\ R_n \end{vmatrix}$$
- $\det(I) = 1$

The universal property of the exterior product ensures that the only function that satisfies all three of these conditions must be the determinant function, which is kinda really cool.

2 Structure of Linear Operators in Finite Vector Spaces

2.1 Polynomial Rings in 1 Variable Over a Field

Polynomial Ring: Say F is a field. Then $F[x]$ denotes the set of all polynomials in 1 variable (x) with coefficients from F. $F[x]$ has polynomial addition and polynomial multiplication on it, and these two operations satisfy the axioms of a commutative ring with identity.

Degree: Let $p \in F[x]$. Then $p(x) = c_0 x^n + c_1 x^{n-1} + \cdots + c_n$ with $c_0 \neq 0$. Then $\deg(p) = n$.

Units: An element of a ring is called a unit if it has a multiplicative inverse.

Units in $F[x]$: The only units in $F[x]$ are the constant (i.e. of degree 0) non-zero polynomials, i.e. $F \backslash \{0\}$.

Polynomial Division: Say $f, g \in F[x]$ and $g \neq 0$. Then, there exist unique polynomials $q, r \in F[x]$ such that $\deg(r) < \deg(g)$ and

$$f = q * g + r$$

The Euclidean Algorithm: Take some non-zero $f, g \in F[x]$ and apply the division algorithm repeatedly:

$$\begin{aligned} f &= q_1 * g + r_1 && \deg(r_1) < \deg(g) \\ g &= q_2 * r_1 + r_2 && \deg(r_2) < \deg(r_1) \\ r_1 &= q_3 * r_2 + r_3 && \deg(r_3) < \deg(r_2) \\ &\vdots \\ r_{m-2} &= q_m * r_{m-1} + r_m && \deg(r_m) < \deg(r_{m-1}) \\ r_{m-1} &= q_{m+1} * r_m + 0 && \deg(r_3) < \deg(r_2) \end{aligned}$$

The process must end as at each step we "consume" the degree of the dividend by taking it as the previous step's remainder. Notice that all these q_i and r_j polynomials are uniquely determined by f and g, and the r_m obtained

- divides both f and g and
- it divides any $h \in F[x]$ such that h divides both f and g. This r_m

Greatest Common Divisor: This r_m polynomial is called the greatest common divisor of f and g. For this greatest common divisor, there exist unique polynomials $a, b \in F[x]$ such that

$$r_m = a * f + b * g$$

Irreducibility: A polynomial $f \in F[x]$ is said to be irreducible if it is impossible to write $f = g.h$ with $\deg(g), \deg(h) < \deg(f)$.

Prime: A polynomial $f \in F[x]$ is said to be prime if

$$f | g.h \;\Rightarrow\; f | g \text{ or } f | h \qquad \forall g, h \in F[x]$$

Unique Factorization: Any polynomial $f \in F[x]$ with $\deg(f) > 0$ can be expressed as a product or irreducible (i.e. prime) polynomials

$$f = q_1 q_2 \cdots q_m$$

which is unique up to reordering and multiplication by units. This makes $F[x]$ a Unique Factorization Domain (UFD).

Ideals in $F[x]$: Let R be a commutative ring with identity. A subset $I \subseteq R$ is said to be an (double-sided) ideal of R if

i) $\forall a, b \in I, a + b \in I$
ii) $\forall a \in I \forall r \in R, ar \in I$

A principle ideal generated by some $r \in R$ is

$$(r) = \{p \in R \;:\; p = rq \text{ for some } q \in R\}$$

Ideals in $F[x]$: Any ideal of the ring $F[x]$ is a principle ideal. This makes $F[x]$ a Principle Ideal Domain (PID).

Irreducible Polynomials in...

$\mathbb{C}[x]$ **or the Fundamental Theorem of Algebra:** Other than linear polynomials, there aren't any. Equivalently, any non-constant polynomial in $\mathbb{C}[x]$ has a root.

$\mathbb{R}[x]$**:** There exists no irreducible polynomial of odd degree in $\mathbb{R}[x]$.
Any non-constant polynomial in $\mathbb{R}[x]$ can be written as a product of linear and/or quadratic irreducible factors.

$F[x]$ **for** $F = \mathbb{Q}$ **or Finite** F**:** There exist irreducible polynomials of any degree in such a case.

2.2 Eigenvectors, Eigenvalues & Beyond

Eigenvectors & Eigenvalues: Let V be a vector space over a field F and $T : V \to V$ be a linear operator. A vector $v \in V$ is called an eigenvector of T if

i) $v \neq \vec{0}$
ii) $Tv = \lambda v$ for some $\lambda \in F$

λ is called the eigenvalue associated to v.
Let A_T be the matrix representation of the operator T in some (any) basis. Then the eigenvalues of T are found by solving

$$\det(A_T - \lambda I) = 0$$

for λ in the field. Once you have the eigenvalues, solve the system

$$(A_T - \lambda I)[v]_B = \vec{0}$$

for each eigenvalue λ to find the eigenvectors.

Linear Independence of Eigenvectors: The eigenvectors corresponding to distinct eigenvalues are linearly independent, so the sum of eigenspaces (corresponding to different eigenvalues) is really a direct sum:

$$W_{\lambda_1} + W_{\lambda_2} + \cdots + W_{\lambda_k} = W_{\lambda_1} \oplus W_{\lambda_2} \oplus \cdots \oplus W_{\lambda_k}$$

meaning

$$W_{\lambda_i} \cap \underbrace{\left(W_{\lambda_1} + \cdots + W_{\lambda_{i-1}} + W_{\lambda_{i+1}} + \cdots + W_{\lambda_k}\right)}_{\text{the sum without } W_{\lambda_i}} = \{\vec{0}\} \qquad \forall i = 1, \ldots, k$$

Characteristic Polynomial: The polynomial

$$\Delta_T(x) = \det(xI - A_T)$$

is called the characteristic polynomial of T.
Properties:

- $\deg(\Delta_T(x)) = n = \dim(V)$
- $\Delta_T(x)$ is a monic polynomial.
- $\Delta_T(x)$ does not depend on the choice of basis.
- Eigenvalues are the roots of the characteristic polynomial.

- Two coefficients, the constant and the first (except the leading coefficient which is always 1) of the characteristic polynomial are very easy to compute:

$$\Delta_T(x) = x^n - \mathrm{Tr}(A_T)x^{n-1} + \cdots + (-1)^n \det(A_T)$$

If the characteristic polynomial splits into linear factors as

$$\Delta_T(x) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_k)^{m_k}$$

where $\lambda_i \in F$ are distinct eigenvalues with multiplicity $m_i \in \mathbb{N}^+$, then

$$\mathrm{Tr}(A_T) = -m_1\lambda_1 + \cdots + m_k\lambda_k \qquad \det(A_T) = \lambda_1^{m_1} \cdots \lambda_k^{m_k}$$

Eigenspaces: Let $T : V \to V$ be a linear operator and λ an eigenvalue of T. Then

$$W_\lambda = \ker(T - \lambda I) = \{\text{Eigenvectors of } T \text{ with eigenvalue } \lambda\} \cup \{\vec{0}\}$$

is called the eigenspace of T associated to the eigenvalue λ, which is a subspace of V. For any eigenvalue λ, $\dim(W_\lambda) \geq 1$.

Polynomial of Operators: Let $T : V \to V$ be a linear operator, and $p(x) \in F[x]$ with

$$p(x) = c_m x^m + c_{m-1}x^{m-1} + \cdots + c_1 x + c_0$$

We can then define an operator $p(T) : V \to V$ by

$$p(T) = c_m T^m + c_{m-1}T^{m-1} + \cdots + c_1 T + c_0$$

where $T^k = \underbrace{T \circ \cdots \circ T}_{k \text{ times}}$, which is again a linear operator.

Properties:

- If A_T is the matrix representation of T, then of course the matrix representation of $p(T)$ becomes $p(A_T)$.
- If $p(x) = q_1(x)q_1(x)$ for some $q_1(x), q_2(x) \in F[x]$, then $p(T) = q_1(T) \circ q_2(T)$.
- If λ is an eigenvalue of T with eigenvector v, then $p(\lambda)$ is an eigenvalue of $p(T)$ with eigenvector v.

Invariant Subspaces: Let $T : V \to V$ be a linear operator. A subspace W of V is said to be T-invariant or called an invariant subspace of T if for all $w \in W$, $Tw \in W$.
$W = V$ and $W = \{\vec{0}\}$ are trivial invariant subspaces for all linear operators. An eigenspace of an operator T is T-invariant.

Restrictions to and Quotient by Invariant Subspaces: Let V be a vector space, $T : V \to V$ a linear operator and W a T-invariant subspace of V. Then T induces two linear operators: Its restriction onto T

$$T|_W : W \to W$$

and the one on the space with W quotiened-out:

$$\begin{aligned} \overline{T} : {}^{V}\!/\!_{W} &\to {}^{V}\!/\!_{W} \\ \overline{T}(v + W) &\mapsto T(v) + W \end{aligned}$$

Matrix Representations: Let V be an n-dimensional vector space, $T : V \to V$ a linear operator and W a T-invariant subspace of V, with a basis $\{w_1, \cdots, w_k\}$ ($0 < k < n$ for non-trivial subspaces). Complete this basis to a basis of the whole V to obtain

$\mathcal{B} = \{w_1, \cdots, w_k, v_{k+1}, \cdots, v_n\}$. Then under this basis, the matrix representation of T, namely $[T]_{\mathcal{B}}^{\mathcal{B}}$ has the following form:

$$[T]_{\mathcal{B}_V}^{\mathcal{B}_V} = \left[\begin{array}{c|c} A_{k\times k} & C_{k\times(n-k)} \\ \hline 0_{(n-k)\times k} & B_{(n-k)\times(n-k)} \end{array}\right]$$

According to this form, then,

- the matrix representation of $T|_W$ with respect to the basis $\{w_1, \cdots, w_k\}$ is A, and
- the matrix representation of $\overline{T}$ with respect to the basis $\{v_{k+1} + W, \cdots, v_n + W\}$ is B.

Characteristic Polynomial: Suppose we have a matrix (hence a linear operator)

$$D = \left[\begin{array}{c|c} A_{k\times k} & C_{k\times(n-k)} \\ \hline 0_{(n-k)\times k} & B_{(n-k)\times(n-k)} \end{array}\right]$$

with respect to some basis $\mathcal{B}$, so the subspace generated by the first k basis vectors is invariant under this matrix. Then

$$\Delta_D(x) = \Delta_A(x)\Delta_B(x)$$

This implies, according to our restriction and quotient operators as defined above, that

$$\Delta_T(x) = \Delta_{T|_W}(x)\Delta_{\overline{T}}(x)$$

Dimension of Eigenspaces: Say for a linear operator T on an n-dimensional vector space V, λ is an eigenvalue. Then write

$$\Delta_T(x) = (x-\lambda)^m q(x)$$

where $m > 0$ and $q(\lambda) \neq 0$ so m is the maximal exponent of $(x - \lambda)$. Then

$$1 \leq \dim(W_\lambda) \leq m$$

Diagonalizability: Let T be a linear operator on an n-dimensional vector space V, and write

$$\Delta_T(x) = (x-\lambda_1)^{m_1} \cdots (x-\lambda_k)^{m_k} q(x)$$

where $q(x)$ contains no linear factors, and the linear factor decomposition is maximal. Upon this setup, there are several equivalent definitions of diagonalizability. T is said to be diagonalizable if

i) $n = m_1 + \cdots + m_k$ and $\dim(W_{\lambda_i}) = m_i$ for all $i = 1, \ldots, k$.
ii) $V = W_{\lambda_1} \oplus \cdots \oplus W_{\lambda_k}$
iii) There exists a basis $\mathcal{B}$ for V such that $[T]_{\mathcal{B}}^{\mathcal{B}}$ is a diagonal matrix.
iv) There exists a set $\{v_1, \ldots, v_n\}$ of n linearly independent eigenvectors of T.

Matrix Representations: Diagonalizability for a matrix A (corresponding to a linear operator) is defined as being similar to a diagonal matrix, i.e. there exists a diagonal matrix D and an invertible matrix P such that

$$A = PDP^{-1}$$

In this case D contains the eigenvalues of A along its diagonal, and P is a matrix of (column) eigenvectors, with matching orderings (i.e. d_{ii} is the eigenvector of the ith column of P).

Integer Powers and Roots of Diagonalizable Matrices: If A is a diagonalizable matrix, similar to a diagonal matrix D through the invertible matrix P, then we can calculate its integer powers easily by computing

$$A^n = (PDP^{-1})^n = PD^nP^{-1}$$

which is easy as taking the powers of a diagonal matrix is easy.
For the integer roots, the same logic applies but only in reverse: This time, we need to find a diagonal matrix $\widetilde{D}$ such that $(\widetilde{D})^n = D$ which amounts to computing the integer roots of the diagonal entries of D, so that we can set $A^{1/n} = P\widetilde{D}P^{-1}$ to have

$$(A^{1/n})^n = (P\widetilde{D}P^{-1})^n = P(\widetilde{D})^nP^{-1} = PDP^{-1} = A$$

Simultaneous Diagonalization: Let V be a finite-dimensional vector space over a field F. Say $\mathcal{A}$ is a set whose elements are commuting diagonalizable linear operators $T_i : V \to V$ such that $T_iT_j = T_jT_i$ for all i, j. Then, there exists a basis $\mathcal{B}$ of V such that all of the matrix representations $[T]_{\mathcal{B}}^{\mathcal{B}}$ are diagonal matrices. Then the operators in $\mathcal{A}$ are said to be simultaneously diagonalizable, and the basis vectors are eigenvectors for all T_i operators at the same time.

Not being Diagonalizable:

Sufficient Conditions: There are two reasons why an operator may not be diagonalizable, one of which is circumventable:

1. The characteristic polynomial $\Delta_T(x)$ may contain irreducible factors other than the linear ones. This is circumvented by viewing T in the vector space of the algebraic closure of the original vector space.
   More dangerously,
2. Even if $\Delta_T(x)$ splits into linear factors, say

$$\Delta_T(x) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_k)^{m_k}$$

   one of its eigenspaces may be deficient in dimension, i.e. there may exist some W_{λ_i} such that $\dim(W_{\lambda_i}) < m_i$.

Necessary Conditions: Say T is a linear, non-diagonalizable operator on an n-dimensional vector space with

$$\Delta_T(x) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_k)^{m_k}$$

Then, there exists some $m_i \gneq 1$, meaning T has at least one repeated eigenvalue.

Cayley-Hamilton Theorem: Say T is a linear operator on an n-dimensional vector space with characteristic polynomial $\Delta_T(x)$. Then

$$\Delta_T(T) = 0$$

"Every linear operator on a finite dimensional vector space satisfies its own characteristic polynomial."
Equivalently, if A_T is the matrix representation of T, we obtain

$$\Delta_T(A_T) = 0_{n\times n}$$

Minimal Polynomial: Say T is a linear operator on an n-dimensional vector space. Then define $J_T \subset F[x]$ as

$$J_T = \{p(x) \in F[x] : p(T) = 0\}$$

then J_T is an ideal of $F[x]$, and as $F[x]$ is a PID, J_T is a principle ideal. This means there exists a unique monic polynomial, say $\delta_T(x)$, generating J_T:

$$J_T = \big(\delta_T(x)\big)$$

where $\delta_T(x)$ is the polynomial of least degree for which $\delta_T(T) = 0$.

Properties:

- As $\Delta_T(x) \in J_T$ by Cayley-Hamilton Theorem, we have

$$\delta_T(x)|\Delta_T(x)$$

- In turn, we also have

$$\Delta_T(x)|\left(\delta_T(x)\right)^n$$

  meaning every irreducible factor of $\Delta_T(x)$ appears at least once in $\delta_T(x)$.

Minimal Polynomial & Diagonalizability: An operator T is diagonalizable if and only if its minimal polynomial is a product of distinct linear factors.

3 Inner Product

3.1 Elements of a Vector Space

Inner Product on a Real Vector Space: Let V be a vector space over $\mathbb{R}$. An inner product on V is a function $V \times V \to \mathbb{R}$ taking a pair (v, w) to a real number $\langle v, w\rangle$, that satisfies the following conditions:

(1) $\langle c_1v_1 + c_2v_2, w\rangle = c_1 \langle v_1, w\rangle + c_2 \langle v_2, w\rangle$ for all $c_1, c_2 \in \mathbb{R}$ and $v_1, v_2, w \in V$.
(2) $\langle v, w\rangle = \langle w, v\rangle$ for all $v, w \in V$.
(3) $\langle v, v\rangle \geq 0$ for all $v \in V$, with equality happening if and only if $v = \vec{0}$.

Norm/Length: Let $(V, \langle,\rangle)$ be a real inner product space. Then

$$\|v\| = \sqrt{\langle v, v\rangle}$$

is a valid norm definition.

Unit Vector: A vector of norm 1 is called a unit vector.

Orthogonality: Let $(V, \langle,\rangle)$ be a real inner product space. Two vectors $v, w \in V$ are said to be orthogonal or perpendicular, denoted $v \perp w$, if

$$\langle v, w\rangle = 0$$

Orthogonal Complement: Let $(V, \langle,\rangle)$ be a real inner product space, and say $S \subset V$ is a non-empty subset. The orthogonal complement of S is then defined as

$$S^\perp = \{v \in V : \forall s \in S \, \langle v, s\rangle = 0\}$$

The orthogonal complement is always a subspace of V.
For a finite dimensional vector space V and a subspace W of V, we have

$$V = W \oplus W^\perp$$

(Proved using orthogonal projections.)

Orthogonal Projection onto a Vector: Let $(V, \langle , \rangle)$ be a real inner product space, $v, w \in V$ and $w \neq \vec{0}$. Then there exists a unique vector $\mathrm{proj}_w(v)$ such that

i) $\mathrm{proj}_w(v)$ is a scalar multiple of w and

ii) $v - \mathrm{proj}_w(v)$ is orthogonal to v.

This vector is given by the formula

$$\mathrm{proj}_w(v) = \frac{\langle v, w\rangle}{\langle w, w\rangle} w = \frac{\langle v, w\rangle}{\|w\|^2} w = \left\langle v, \frac{w}{\|w\|} \right\rangle \frac{w}{\|w\|}$$

Cauchy-Schwarz Inequality: Let $(V, \langle , \rangle)$ be a real inner product space and $v, w \in V$. Then

$$|\langle v, w\rangle| \leq \|v\|\|w\|$$

with equality happening if and only if v and w are linearly dependent, i.e. one is the scalar multiple of the other.

Triangle Inequality: Let $(V, \langle , \rangle)$ be a real inner product space and $v, w \in V$. Then

$$\|v + w\| \leq \|v\| + \|w\|$$

with equality happening if and only if one is a non-negative scalar multiple of the other.

Pythagorean Theorem: Say $(V, \langle , \rangle)$ is a real inner product space. If $v, w \in V$ are orthogonal vectors, then

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2$$

Angle: Let $(V, \langle , \rangle)$ be a real inner product space and $v, w \in V$. The angle between v and w is defined by the relation

$$\cos\theta = \frac{\langle v, w\rangle}{\|v\|\|w\|}$$

Complex Hermitian Inner Product: Let V be a vector space over $\mathbb{C}$. An inner product on V is a function $V \times V \to \mathbb{C}$ taking a pair (v, w) to a complex number $\langle v, w\rangle$, that satisfies the following conditions:

(1) $\langle c_1 v_1 + c_2 v_2, w\rangle = c_1 \langle v_1, w\rangle + c_2 \langle v_2, w\rangle$ for all $c_1, c_2 \in \mathbb{R}$ and $v_1, v_2, w \in V$.

(2) $\langle v, w\rangle = \overline{\langle w, v\rangle}$ for all $v, w \in V$.

(3) $\langle v, v\rangle$ is real and non-negative for all $v \in V$, with equality happening if and only if $v = \vec{0}$.

The same definitions of norm/length and orthogonality hold for a Hermitian inner product as well.

Orthogonal and Orthonormal Sets: A subset $S \subset V$ where V is an inner product space (either real or complex Hermitian) is called an orthogonal set if for all distinct $v, w \in S$, we have $\langle v, w\rangle = 0$. If further all the vectors in S are unit vectors, then S is called an orthonormal set.

Linear Independence: An orthogonal set is linearly independent.

Basis and Expansions: If $\mathcal{B} = \{v_1, \ldots, v_n\}$ is an orthogonal basis for V, then for any $v \in V$ we have

$$v = \frac{\langle v, v_1\rangle}{\langle v_1, v_1\rangle} v_1 + \cdots + \frac{\langle v, v_n\rangle}{\langle v_n, v_n\rangle} v_n$$

Gram-Schmidt Orthogonalization: Say $(V, \langle , \rangle)$ is a real or complex Hermitian inner product space, and $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$ is an ordered basis for V. Define w_i for $i = 1, \ldots, n$ as follows:

$$w_1 = v_1 \qquad \& \qquad w_i = v_i - \sum_{k=1}^{i-1} \frac{\langle v_i, w_k \rangle}{\langle w_k, w_k \rangle} w_k \quad \text{for } 1 < i \leq n$$

Then $\mathcal{B}' = \{w_1, \ldots, w_n\}$ is an orthogonal basis for V.

Orthogonal Projection onto a Subspace: Say $(V, \langle , \rangle)$ is a real or complex Hermitian inner product space, and W is a subspace of V. An orthogonal projection of a vector v along W is a vector $u = \mathrm{proj}_W(v)$ such that

i) $u \in W$
ii) $v - u \perp W$ or equivalently $v - u \in W^{\perp}$

If W is a finite-dimensional subspace of V, with an orthogonal basis $\mathcal{B} = \{w_1, \ldots, w_m\}$, such an orthogonal projection of any vector v along W then exists, is unique, and is given by

$$\mathrm{proj}_W(v) = \sum_{i=1}^{m} \frac{\langle v, w_i \rangle}{\langle w_i, w_i \rangle} w_i = \frac{\langle v, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 + \cdots + \frac{\langle v, w_m \rangle}{\langle w_m, w_m \rangle} w_m$$

This projection satisfies, for any $w \in W$

$$\|v - \mathrm{proj}_W(v)\| \leq \|v - w\|$$

with equality occurring if and only if $w = \mathrm{proj}_W(v)$.

Bessel's Inequality: Let $(V, \langle , \rangle)$ be a real or complex Hermitian inner product space. Say $\{v_1, \ldots, v_n\}$ is an orthogonal set in V. Then for any $v \in V$,

$$\|v\|^2 \geq \sum_{k=1}^{n} \frac{|\langle v, v_k \rangle|^2}{\|v_k\|^2}$$

Furthermore, the equality holds if and only if

$$v = \sum_{k=1}^{n} \frac{|\langle v, v_k \rangle|^2}{\|v_k\|^2} v_k$$

i.e. $v \in \mathrm{Span}(v_1, \ldots, v_n)$.
Two **remarks:**

i) For the $n = 1$ case, the Bessel's inequality becomes the Cauchy-Schwarz inequality:

$$\|v\|^2 \geq \frac{|\langle v, v_1 \rangle|}{\|v_1\|^2} \iff \|v\|^2 \|v_1\|^2 \geq |\langle v, v_1 \rangle|$$

ii) If the set given is further ortho*normal*, then Bessel's inequality becomes

$$\|v\|^2 \geq \sum_{k=1}^{n} |\langle v, v_k \rangle|^2$$

3.2 Two Applications

Method of Least Squares: Say we have a $m \times n$ real or complex linear system of equations, where m, the number of equations, is typically much larger than n, the number of variables.

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

or in another notation,

$$\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} x_1 + \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} x_2 + \cdots + \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} x_n = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

or yet another, more compact notation

$$\vec{a_1}x_1 + \vec{a_2}x_2 + \cdots + \vec{a_n}x_n = \underbrace{\begin{bmatrix} \vec{a_1} & \vec{a_2} & \dots & \vec{a_n} \end{bmatrix}}_{A} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A\vec{x} = \vec{b}$$

For $m > n$ we don't expect the system to have a solution unless

$$\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \in \operatorname{Span}(\vec{a_1}, \vec{a_2}, \dots, \vec{a_n}) = W$$

In such a case (or even if $\vec{b} \in \operatorname{Span}(\vec{a_1}, \vec{a_2}, \dots, \vec{a_n})$) we can look for an approximate solution. Consider $\mathbb{R}^n$ or $\mathbb{C}^n$ with its standard (Hermitian) inner product. We know that

$$\|\vec{b} - w\| \geq \|\vec{b} - \operatorname{proj}_W(\vec{b})\|$$

Then the solution to

$$\vec{a_1}x_1 + \vec{a_2}x_2 + \cdots + \vec{a_n}x_n = \operatorname{proj}_W(\vec{b})$$

which surely exits as $\operatorname{proj}_W(\vec{b}) \in W$ by definition, would give us the approximate solution we are looking for.
So, in short, instead of solving

$$A\vec{x} = \vec{b}$$

we solve

$$A\vec{x} = \operatorname{proj}_W(b)$$

where W is the vector space spanned by the columns of the matrix A.

Fourier Series: Let $V = \mathcal{C}[a, b]$ be the set of continuous real-valued functions on $[a, b]$, with the inner product given by

$$\langle f, g \rangle = \int_a^b f(x)g(x)dx$$

Our goal is to find a "large" subset of V that is orthogonal. The proposition of the Fourier Series is that we can do this using trigonometric functions. Without loss of generality, take $V = [-\pi, \pi]$ and set

$$\mathcal{S} = \{\sin(nx)\}_{n=1,2,\dots} \cup \{\cos(nx)\}_{n=0,1,2,\dots}$$

One can easily check that under the given inner product,

- $\mathcal{S}$ is an orthogonal set,
- $\|\sin(nx)\| = \|\cos(nx)\| = \sqrt{\pi}$
- $\|1\| = \sqrt{2\pi}$

Then the Fourier series of $f \in \mathcal{C}[-\pi, \pi]$ is defined as

$$\mathrm{proj}_{\overline{\mathrm{Span}(\mathcal{S})}}(f)$$

where we remark that Span($\mathcal{S}$) is infinite dimensional.

Fourier Convergence Theorem: For any choice of $f \in \mathcal{C}[-\pi, \pi]$,

- $\mathrm{proj}_{\overline{\mathrm{Span}(\mathcal{S})}}(f)$ converges at all points $x \in [-\pi, \pi]$.
- $f = \mathrm{proj}_{\overline{\mathrm{Span}(\mathcal{S})}}(f)$, i.e. $\overline{\mathrm{Span}(\mathcal{S})} = \mathcal{C}[-\pi, \pi]$

If we wish to work with complex-valued functions $f \in [-\pi, \pi] \to \mathbb{C}$, the same arguments work with the standard Hermitian inner product in $V = \mathcal{C}[-\pi, \pi]$ given by

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(z)\overline{g(z)}dz$$

and the set

$$\mathcal{S} = \{e^{inz}\{n \in \mathbb{Z}\}$$

The set $\mathcal{S}$ is again orthogonal with $\|e^{inz}\| = \sqrt{2\pi}$ for all $n \in \mathbb{Z}$, and the Fourier Convergence Theorem has an analogue for this case as well.

3.3 Operators on a Vector Space

Riesz Representation Theorem: Let $(V, \langle,\rangle)$ be a finite-dimensional vector space over $\mathbb{R}$ of $\mathbb{C}$ with a real or complex Hermitian inner product, and φ be a linear functional on V. Then there exists a unique $u \in V$ such that $\forall v \in V$,

$$\varphi(v) = \langle v, u \rangle$$

Adjoint of an Operator: Let $(V, \langle,\rangle)$ be a real or complex Hermitian, finite-dimensional inner product space, and $T : V \to V$ is a linear operator. Then there exists a unique linear operator $T^* : V \to V$ such that

$$\langle Tv, w \rangle = \langle v, T^*w \rangle$$

for all $v, w \in V$ called the adjoint of T.

Double Adjoint: Taking adjoints is an involution, meaning $(T^*)^* = T$ for all linear operators T.

Matrix Representation: If the matrix representation of a linear operator T in some basis is given by the matrix A, then the matrix representation of its adjoint in the same basis is

- A^T if $(V, \langle , \rangle)$ is a real inner product space, and
- $A^* := \overline{A^T}$ if $(V, \langle , \rangle)$ is a complex Hermitian inner product space.

Changing Orders: For linear operators S, T over an inner product space V, we have $(ST)^* = T^* S^*$.

Self-adjoint Operators: A linear operator $T : V \to V$ is said to be self-adjoint if $T = T^*$.

Matrix Representation: The matrix representation of a self-adjoint operator in a real inner product space is a symmetric matrix, i.e. $A = A^T$. Similarly, the matrix representation of a self-adjoint operator in a complex Hermitian inner product space is given by a Hermitian matrix, i.e. $A = A^* = \overline{A^T}$.

Orthogonal and Unitary Operators: Let $(V, \langle , \rangle)$ be an inner product case and T a linear operator on it. Then if T preserves the inner product, meaning

$$\langle Tv, Tw \rangle = \langle v, w \rangle$$

for all $v, w \in V$, then T is said to be

- orthogonal if V is a real inner product space, or
- unitary if V is a complex Hermitian inner product space.

Equivalent Descriptions: For an operator T in a real or complex Hermitian inner product space, the followings are equivalent:

i) T is orthogonal/unitary.
ii) T preserves the lengths of all vectors.
iii) $TT^* = T^*T = I$
iv) T takes any orthonormal basis of V to another orthonormal basis.

Matrix Representations: An $n \times n$ real matrix is called orthogonal if $AA^T = A^T A = I_n$. Similarly, an $n \times n$ complex matrix is called unitary if $AA^* = A^* A = I_n$.

Orthogonal Group: The $n \times n$ orthogonal/unitary matrices form a group under matrix multiplication, called the orthogonal group, denoted $\mathcal{O}(n)$ or $\mathcal{O}(n, \mathbb{R})/\mathcal{O}(n, \mathbb{C})$

Normal Operators: Let $(V, \langle , \rangle)$ is a real or complex Hermitian inner product space, and T be a linear operator on it. T is said to be a normal operator if

$$TT^* = T^*T$$

Self-adjoint, orthogonal/unitary operators are all normal operators as well.

Polynomial of a Normal Operator: For any $p(x) \in \mathbb{R}[x]$ or $p(x) \in \mathbb{C}[x]$, if T is a normal operator, then so is $p(T)$.

Nilpotency: If a normal operator T is also nilpotent (i.e. $T^k = 0$ for some $k \in \mathbb{N}$), then $T = 0$.

Diagonalizability: Every normal operator T on a finite-dimensional real or complex Hermitian inner product space is diagonalizable over $\mathbb{C}$.
As for a normal operator T, T^* is also normal and hence diagonalizable *and* T and T^* commute by the definition of normality, this means that T and T^* are simultaneously diagonalizable.

- **Eigenvalues and Eigenvectors:** For any normal operator T on a finite-dimensional inner product space, if v is an eigenvector of T with eigenvalue λ, then v is an eigenvector of T^* with eigenvalue $\overline{\lambda}$.
- **Eigenspaces:** For any normal operator T, the eigenspaces of T are all mutually orthogonal. That is, eigenvectors with different eigenvalues are orthogonal to one another.
- **Orthogonal Basis of Eigenvectors:** For any normal operator T on an inner product space V, there exists an orthonormal basis $\mathcal{B}$ of eigenvectors of T for V.

Complexification: To be able to state the last fact about a normal operator inducing an orthogonal basis of its eigenvectors on its domain vector space, we need the complexification concept, because even if the vector space is real, the eigenvalues and eigenvectors of an operator on it may indeed be complex.
For a real vector space V, its complexification is defined as

$$V_{\mathbb{C}} := V \otimes_{\mathbb{R}} \mathbb{C}$$

$V_{\mathbb{C}}$ is a complex vector space in a natural way, under the following scalar multiplication:

$$c\left(\sum v_i \otimes a_i\right) := \sum v_i \otimes (ca_i)$$

And so the previous theorem can be stated as follows: If $T : V \to V$ is a normal operator on a real and finite-dimensional inner product space, then $T : V_{\mathbb{C}} \to V_{\mathbb{C}}$ defined by

$$T\left(\sum v_i \otimes a_i\right) := \sum T(v_i) \otimes a_i$$

is diagonalizable and has an orthonormal basis of eigenvectors.

3.4 Spectral Theorem & Its Consequences

Orthogonal Projection: Let $(V, \langle , \rangle)$ be a real or complex Hermitian inner product space. A linear operator $P : V \to V$ is called an orthogonal projection if

i) $P^2 = P$, so P is a projection,
ii) $P^* = P$, so P is self-adjoint.

Then it is easy to prove that for all $v \in V$, we have $Pv \perp v - Pv$ due to the self-adjointness, hence the name "orthogonal" projection.

Spectral Theorem: Let $(V, \langle , \rangle)$ be a real or complex Hermitian finite-dimensional inner product space, and $T : V \to V$ a normal operator on it with distinct eigenvalues $\lambda_1, \dots, \lambda_k$ and corresponding eigenspaces $W_{\lambda_1}, \dots, W_{\lambda_k}$. Then, there exists orthogonal projections $P_1, \dots, P_k$ with the following properties:

i) $\mathrm{Im}(P_i) = W_{\lambda_i}$ for all i,
ii) $T = \lambda_1 P_1 + \cdots + \lambda_k P_k$,
iii) $I = P_1 + \cdots + P_k$,
iv) $P_i P_j = 0$ whenever $i \neq j$.

Spectral Characterization of Operators: The Spectral Theorem provides further means of classification for the operator types we know from before. Let $(V, \langle , \rangle)$ be a real or complex Hermitian finite-dimensional inner product space, and $T : V \to V$ a normal operator on it.

Self-Adjoint Operators: T is self-adjoint if and only if all its eigenvalues are real.

Diagonalizability/Matrix Representations: This means that symmetric/Hermitian $n \times n$ matrix A is similar to a real diagonal matrix D, say via P, so we have $A = PDP^{-1}$. Because we also have an orthonormal basis of eigenvectors, we can construct this P matrix in such a way that its columns are these orthonormal basis, which yields $P^*P = I$, yielding $P^{-1} = P^*$ and making P unitary. Thus we conclude that for a symmetric/Hermitian matrix A, there exists a unitary matrix P and a real diagonal matrix D such that

$$A = PDP^* = PDP^{-1}$$

Unitary & Orthogonal Operators: T is unitary if and only if all its eigenvalues have modulus 1, i.e. they lie on the unit circle in the complex plane.

Diagonalizability/Matrix Representations: With a similar reasoning as above, if A is a unitary matrix, then there exist a unitary matrix P and a diagonal matrix D of unit modulus entries such that

$$A = PDP^* = PDP^{-1}$$

Real Orthogonal Matrices: There exists a basis $\mathcal{B}$ of $\mathbb{R}^n$ such that

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_k \end{bmatrix}$$

where each A_j is one of the following 1×1 or 2×2 matrices:

$$[1] \quad \text{OR} \quad [-1] \quad \text{OR} \quad \begin{bmatrix} \cos\theta_j & -\sin\theta_j \\ \sin\theta_j & \cos\theta_j \end{bmatrix} \text{ for some } \theta_j$$

Rotations of $\mathbb{R}^3$: Every rotation of $\mathbb{R}^3$ has an axis (along which there is no change occurs, so the eigenvalue is 1 for the unit vector (eigenvector) of this direction).

4 Quadratic Forms

We will be taking F to be a field of characteristic different than 2.

Quadratic Form: Let $n \geq 1$ is an integer. A quadratic form q in n-variables is a degree 2 homogenous polynomial in n variables, say $x_1, \ldots, x_n$:

$$q(x_1, \ldots, x_n) = \sum_{i=1}^{n} a_{ij} x_i x_j$$

where $a_{ij} \in F$. We can represent each quadratic form with a vector-matrix multiplication. Define a variable vector $\vec{x}_{n\times 1} = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix}^T$ and write

$$q(\vec{x}) = \vec{x}^T B \vec{x}$$

where $B \in M_{n\times n}(F)$ with $B^T = B$. From the polynomial expression, we can extract the entries of the matrix B as

$$b_{ij} = \begin{cases} a_{ij} & \text{if } i = j \\ \frac{a_{ij}}{2} & \text{if } i \neq j \end{cases}$$

so there exists a one-to-one correspondence between quadratic forms and symmetric matrices in F.

Diagonal Quadratic Form: A diagonal quadratic form is one that constitutes only of x_i^2 terms. Equivalently, a quadratic form is diagonal if its corresponding matrix B is diagonal.

Equivalence of Quadratic Forms: Say $q_1(\vec{x}) = \vec{x}^T A_1 \vec{x}$ and $q_2(\vec{y}) = \vec{y}^T A_2 \vec{y}$ are two quadratic forms. We say that they are equivalent if there exists an invertible matrix $P \in M_{n\times n}(F)$ such that

$$A_2 = P^T A_1 P$$

Notice that in this case A_1 and A_2 are *not* similar. However, the two equivalent forms being equivalent means that one acts like the other under a change of basis ($\vec{x} = P\vec{y}$). This relation is indeed an equivalence relation.

Diagonalization of Quadratic Forms: Any quadratic form $q(\vec{x})$ is equivalent to a diagonal quadratic form.

Principle Axis Theorem: Take the inner product space $(\mathbb{R}^n, \langle , \rangle)$ with its standard inner product. Then, for any quadratic form $q(\vec{x}) = \vec{x}^T A\vec{x}$, there exists an orthogonal matrix P such that $P^T AP$ is diagonal. This follows very easily from the Spectral Theorem.

4.1 Classification of Real Quadratic Forms up to Equivalence

The goal is to find which features of a quadratic form are preserved under the equivalence relation. In other words, we want to understand the equivalence classes of the quadratic form equivalence relation.

Rank: Say $q_1(\vec{x}) = \vec{x}^T A_1 \vec{x}$ and $q_2(\vec{y}) = \vec{y}^T A_2 \vec{y}$ are two equivalent quadratic forms. Then

$$\text{rank}(A_1) = \text{rank}(A_2)$$

Sylvester's Law of Inertia: Suppose $A \in M_{n\times n}(\mathbb{R})$ is a symmetric matrix, and P_1 and P_2 are two invertible matrices such that $P_1^T AP_1 = D_1$ and $P_2^T AP_2 = D_2$ are both diagonal matrices. Then the number of positive/negative entries of D_1 and that of D_2 are the same.

Signature: The signature of a real quadratic form q defined by a symmetric matrix A is defined as the number of positive eigenvalues minus that of negative eigenvalues. Sylvester's theorem implies that two equivalent real quadratic forms have the same signature.
Notice that the rank of A is exactly the sum of these two quantities, so given the rank and signature of A, we can solve for these two quantities with ease.

Classification of Real Quadratic Forms up to Equivalence: Two real quadratic forms are equivalent if and only if they have the same rank and the same signature.
So, say we fix a dimension n and consider the quadratic forms on $\mathbb{R}^n$. Each equivalence class necessarily has different number of positive, negative, and null eigenvalues. Say we denote these by n_+, n_- and n_0 respectively. We conclude, via also the rank-nullity theorem, that the number of equivalence classes of quadratic forms is the number of positive integer solutions to the equation

$$n_+ + n_- + n_0 = n$$

5 Singular Value Decomposition & Polar Decomposition

5.1 Positive and Non-negative Operators

Positive and Non-negative Operators: Let $(V, \langle , \rangle)$ be a real or complex Hermitian inner product space with finite dimension. A self-adjoint operator $T : V \to V$ is called

- positive if $\langle Tv, v\rangle > 0$ for all non-zero vectors $v \in V$.
- non-negative if $\langle Tv, v\rangle \geq 0$ for all $v \in V$.

Eigenvalues: Being self-adjoint, such a T is diagonalizable with real eigenvalues. T is

- positive if and only if all of its eigenvalues are positive.
- non-negative if and only if all of its eigenvalues are non-negative.

Matrices: A real symmetric matrix A is

- positive-definite if $v^T Av > 0$ for all non-zero column vectors.
- non-negative-definite if $v^T Av \geq 0$ for all column vectors.

The eigenvalue characterization is the same for matrices.

Automatically Positive/Non-negative Matrices: Let A be any $n \times m$ matrix. Then A^*A and AA^* are always non-negative-definite. If further A has rank m (necessitating $n \geq m$), then A^*A is positive-definite.

Positive Inverses: The inverse of a positive-definite matrix is again positive definite.

Positive Roots: Every positive-definite matrix P has a positive-definite n-th root for any positive integer $n \in \mathbb{N}^+$, meaning there exists some Q such that $Q^n = P$.

Product of Positive Matrices: The product of two positive-definite matrices need not be positive itself (as we may lose self-adjointness), but it has positive eigenvalues.

5.2 Singular Value Decomposition & Polar Decomposition

Singular Value Decomposition: Let A be a real/complex $n \times m$ matrix. Then there exist an $n \times n$ orthogonal/unitary matrix U, an $m \times m$ orthogonal/unitary matrix V and a real diagonal $n \times m$ matrix Σ with non-negative entries such that

$$A = U\Sigma V^*$$

The entries of Σ are called the singular values of A.

Computation: Whether A has full column rank or not, we start by computing the eigenvalues and eigenvectors of A^*A, which we know is diagonalizable as it is self-adjoint, with positive/non-negative eigenvalues. Say λ_i are the eigenvalues of A^*A corresponding to eigenvectors v_i. As A^*A is self-adjoint, we know the vectors v_i form an orthogonal, further orthonormal (after normalization) basis for V. Then set

$$\sigma_i = \sqrt{\lambda_i} \qquad u_i = \frac{1}{\sigma_i} Av_i$$

If necessary, complete the sets $\{v_i\}$ and $\{u_i\}$ to orthonormal bases of appropriate dimensions. And then set

$$U_{n\times n} = \begin{bmatrix} u_1 & \dots & u_n \end{bmatrix} \qquad \Sigma_{n\times m} = \operatorname{diag}(\sigma_1, \dots, \sigma_r) \qquad V_{m\times m} = \begin{bmatrix} v_1 & \dots & v_m \end{bmatrix}$$

Polar Decomposition: This decomposition is inspired by the polar representation of complex numbers, i.e. with one number that is non-negative, the other with unit norm.

Existence: Let A be any $n \times n$ complex matrix. Then there exist a unitary matrix U and a non-negative matrix P such that

$$A = UP$$

If A is real, then both U and P can be taken to be real as well (in which case U is an orthogonal matrix.)

Uniqueness: If further this A is invertible, then the polar decomposition is unique. Meaning there exist unique unitary matrix U and positive-definite matrix P such that $A = UP$.

Computation: These matrices P and U are computed from the singular value decomposition of $A = V\Sigma W^*$ as

$$P = W\Sigma W^* = W\Sigma W^{-1} \qquad \& \qquad U = VW^* = VW^{-1}$$

so we have

$$UP = V \underbrace{W^* W}_{I} \Sigma W^* = V\Sigma W^* = A$$

6 Canonical Forms

How can we classify linear operators up to a change of basis?

An Equivalence of Linear Operators: Let V be a finite-dimensional vector space over a field F. We declare two linear operators $T_1, T_2 : V \to V$ to be equivalent if they become the same operator after a change of basis. In other words, they are equivalent if there exist bases $\mathcal{B}_1$ and $\mathcal{B}_2$ of V such that

$$[T_1]_{\mathcal{B}_1}^{\mathcal{B}_1} = [T_2]_{\mathcal{B}_2}^{\mathcal{B}_2}$$

For matrices, as they represent linear operators, this amounts to declaring them equivalent if and only if they are similar.

Some Part of Our Answer: Two diagonalizable operators are equivalent (similar) if and only if they have the same characteristic polynomial (and hence eigenvalues).

Dealing with Non-diagonalizability: As we have seen before, there are two problems that cause non-diagonalizability:

1) The characteristic polynomial may not split into linear factors. But then, a fact from field theory states that every field can be enlarged in such a way that all polynomials on it splits into linear factors (algebraic closure), e.g. treating an operator on $\mathbb{R}^n$ as if it is in $\mathbb{C}^n$. So this problem is easy to deal with.
2) The characteristic polynomial may split into linear factors but some eigenspaces may be dimension-deficit. This is the interesting case, and we will assume to be in this situation from now on.

6.1 Jordan-Chevalley Decomposition

Semisimple $\equiv$ Diagonalizable. Let V be a finite-dimensional vector space vector space over F, $T : V \to V$ a linear operator that is not diagonalizable but $\Delta_T(x)$ splits into linear factors.

Direct Sum of Eigenspace-ish Subspaces: Say the characteristic polynomial of T is expressed as

$$\Delta_T(x) = (x - \lambda_1)^{\mu_1}(x - \lambda_2)^{\mu_2} \dots (x - \lambda_k)^{\mu_k}$$

where $\lambda_1, \lambda_2, \dots, \lambda_k \in F$ are the unique eigenvalues of T. Define for each i

$$T_i := (T - \lambda_i I)^{\mu_i} \qquad \text{and} \qquad C_i := \ker\left((T - \lambda_i I)^{\mu_i}\right) = \ker(T_i)$$

Then we have

$$V = C_1 \oplus C_2 \oplus \cdots \oplus C_k$$

Invariance: Each of these C_i subspaces are T-invariant.

Jordan-Chevalley Decomposition: Say V is a finite-dimensional vector space over a field F, $T : V \to V$ a linear operator whose characteristic polynomial $\Delta_T(x)$ splits into linear factors. Then there exist unique semisimple (diagonalizable) operator S and nilpotent operator N such that

$$T = S + N$$

and any two of T, S and N commute.

Defining S and N: Define S using the direct sum decomposition $V = C_1 \oplus C_2 \oplus \cdots \oplus C_k$ as

$$S(v_i) = \lambda_i v_i \text{ on each } C_i \text{ and extend it to } V \text{ linearly } \Rightarrow S|_{C_i} = \lambda_i I$$

$$N = T - S$$

Invariance of C_i's: The C_i subspaces from before are also S and N-invariant.

Implications for Matrix Representations: Each of these C_i subspaces are T-invariant, so we can select a basis $\mathcal{B}_i$ for each C_i and define $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k$ as a basis for V. Thus we find

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_k \end{bmatrix}$$

so $[T]_{\mathcal{B}}^{\mathcal{B}}$ is in block diagonal form an $A_i = [T|_{C_i}]_{\mathcal{B}_i}^{\mathcal{B}_i}$. On the other hand,

$$[S]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} \lambda_1 I & & & 0 \\ & \lambda_2 I & & \\ & & \ddots & \\ 0 & & & \lambda_k I \end{bmatrix}$$

and thus

$$[N]_{\mathcal{B}}^{\mathcal{B}} = [T]_{\mathcal{B}}^{\mathcal{B}} - [S]_{\mathcal{B}}^{\mathcal{B}}$$

But we see that this representation is not necessarily useful.

A Careful Choice of Basis: There exists a basis for each C_i such that $T|_{C_i}]_{\mathcal{B}_i}^{\mathcal{B}_i}$ is upper triangular with λ_i on the diagonal:

$$[T|_{C_i}]_{\mathcal{B}_i}^{\mathcal{B}_i} = \begin{bmatrix} \lambda_i & \star & \star \\ & \ddots & \star \\ 0 & & \lambda_i \end{bmatrix}$$

which implies

$$[S|_{C_i}]_{\mathcal{B}_i}^{\mathcal{B}_i} = \begin{bmatrix} \lambda_i & & 0 \\ & \ddots & \\ 0 & & \lambda_i \end{bmatrix} \qquad [N|_{C_i}]_{\mathcal{B}_i}^{\mathcal{B}_i} = \begin{bmatrix} 0 & \star & \star \\ & \ddots & \star \\ 0 & & 0 \end{bmatrix}$$

6.2 Jordan Form

An even more careful choice of basis, which in a sense generalizes the concept of diagonalizability.

Jordan Block: Say F is any field. A matrix $J_\lambda \in M_{m\times m}(F)$ for some $\lambda \in F$ is said to be a Jordan block if it is of the following form:

$$J_\lambda = \begin{bmatrix} \lambda & 1 & & 0 \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{bmatrix}$$

meaning $(J_\lambda)_{i,i} = \lambda$, $(J_\lambda)_{i,i+1} = 1$ and all other elements are 0.

Jordan Form: A matrix $J \in\in M_{m\times m}(F)$ is said to be in Jordan form if it is in block-diagonal form with each block being a Jordan block:

$$J = \begin{bmatrix} J_{\lambda_1} & & & 0 \\ & J_{\lambda_2} & & \\ & & \ddots & \\ 0 & & & J_{\lambda_k} \end{bmatrix}$$

Existence of Jordan Form: Say V is a finite-dimensional vector space over a field F. Let $T : V \to V$ be a linear operator such that $\Delta_T(x)$ splits into linear factors in $F[x]$. Then there exists a basis $\mathcal{B}$ for V such that $[T]_{\mathcal{B}}^{\mathcal{B}}$ is in Jordan form.

Multiple Jordan Blocks: While writing the Jordan form of an operator/matrix, one might need to use more than one Jordan block for each C_i.

Eigenvalues & Eigenvectors of a Nilpotent Operator: The only (unique) eigenvalue of a nilpotent operator is 0. Also, say we have written out the Jordan form of a nilpotent operator as

$$[N]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} 0 & 1 & & 0 & & & & & & & & & \\ & 0 & \ddots & & & & & & & & & & \\ & & \ddots & 1 & & & & & & & & & \\ 0 & & & 0 & & & & & & & & & \\ & & & & 0 & 1 & & 0 & & & & & \\ & & & & & 0 & \ddots & & & & & & \\ & & & & & & \ddots & 1 & & & & & \\ & & & & 0 & & & 0 & & & & & \\ & & & & & & & & \ddots & & & & \\ & & & & & & & & & 0 & 1 & & 0 \\ & & & & & & & & & & 0 & \ddots & \\ & & & & & & & & & & & \ddots & 1 \\ & & & & & & & & & 0 & & & 0 \end{bmatrix}$$

Then the basis elements that coincide with the zero columns of this matrix representation are the eigenvectors of this nilpotent operator.

Implications for Matrices: Because each matrix $A \in M_{n\times n}(F)$ represents a linear operator, the existence of Jordan form implies that every matrix is similar to a matrix Jordan form through a change of basis matrix P. This matrix J is called the matrix canonical form of A.

Uniqueness of Jordan Form: This J matrix is unique up to a permutation of the Jordan blocks, while the change of basis matrix P is certainly not unique.

Computing the Jordan Form: While computing the Jordan form of a matrix, we often (almost always) make use of the following facts:

Combinatorial Possibilities: For a single eigenvalue λ_i with multiplicity $\mu_i \in \mathbb{N}^+$ and its corresponding subspace C_i, the number of possible Jordan forms of this subspace only is equal to the number of partitions of the positive natural number μ_i.

Minimal Polynomial: Because A and its Jordan form J are similar, their minimal polynomial is the same. We can use this to our benefit as follows: Say the minimal polynomial of A, namely $\delta_A(x)$ contains the factor $(x - \lambda_i)$ with multiplicity $1 \le \eta_i \le \mu_i$. Then, the maximum size of the Jordan block corresponding to the subspace C_i can be this η_i.

Number of Jordan Blocks: Even after finding the minimum polynomial and hence the maximum size of a Jordan block, we might have some partitions to choose from. These partitions will likely be distinguished by their number of components. By noticing that the number of components in a partition is equal to the dimension of the eigenspace W_{λ_i}, we can find the correct partition, and hence the Jordan block for the C_i in question.

Consistent Solutions for Basis Vectors: After finding the Jordan form J of a matrix A, we also need to find the change of basis matrix P that relates the two, meaning we need this invertible P to satisfy

$$A = PJP^{-1} \quad \text{so} \quad AP = PJ$$

While solving this equation, we set

$$P = \left[\ v_1 \mid v_2 \mid \dots \mid v_n\ \right]$$

and thus we solve

$$A\left[\ v_1 \mid v_2 \mid \ldots \mid v_n\ \right] = \left[\ Av_1 \mid Av_2 \mid \ldots \mid Av_n\ \right] = \left[\ v_1 \mid v_2 \mid \ldots \mid v_n\ \right] J$$

For Jordan blocks of size 1, the problem reduces to finding eigenvalues. However for the larger Jordan blocks, one needs to solve equations of the form

$$Av_{i+1} = v_i + \lambda v_{i+1} \ \Rightarrow\ (A - \lambda I)v_{i+1} = v_i$$

where v_i is a non-zero vector that either is an eigenvector itself (2×2 Jordan block) or comes from a chain of vector selections that itself originates from an eigenvector (even larger Jordan block). Either way, the v_i vector must be selected in such a way that the equation $(A-\lambda I)v_{i+1} = v_i$ is consistent. Thus, while solving these equations, the best practice is to

1. leave the initial eigenvectors in the most general form,
2. propagate to the "deepest-level" equation that requires consistency in its equation while keeping tabs on the consistency requirements,
3. select vectors that solve these deepest-level equations, and
4. propagate these selections back up to the eigenvector-level until all basis vectors are selected.

One must always remember that *all* selected basis vectors must be linearly independent to have an invertible P.

So we found that any matrix is similar to a unique Jordan form, which enables us to explicitly write out all the Jordan forms in a given finite-dimensional vector space.

6.3 Two Applications

Computing the Jordan-Chevalley decomposition, or even better the Jordan form of a matrix eases certain computations greatly. Say in a k-dimensional vector space V we have a linear operator $T : V \to V$, with a characteristic polynomial

$$\Delta_T(x) = (x - \lambda_1)^{\mu_1} \ldots (x - \lambda_k)^{\mu_k}$$

where λ_i's are distinct eigenvalues. Then we can decompose $T = S + N$ where S is diagonalizable and N is nilpotent, with any two of T, S and N commuting. In the basis $\mathcal{B}$ found for the Jordan form, $[S]_{\mathcal{B}}^{\mathcal{B}}$ is diagonal and $[N]_{\mathcal{B}}^{\mathcal{B}}$ is strictly upper triangular.

Matrix Powers: Because any two of T, S and N commute, for $m \in \mathbb{N}^+$ we can write

$$T^m = (S + N)^m = \sum_{k=0}^{n} \binom{m}{k} S^{m-k} N^k$$

For matrix representations, because the Jordan blocks in a Jordan form do not interact in multiplication, we can only consider a single Jordan block. So, say J_λ is a single Jordan block with eigenvalue λ:

$$\underbrace{\begin{bmatrix} \lambda & 1 & & 0 \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{bmatrix}}_{J_\lambda} = \underbrace{\begin{bmatrix} \lambda & 0 & & 0 \\ & \lambda & \ddots & \\ & & \ddots & 0 \\ 0 & & & \lambda \end{bmatrix}}_{[S]_{\mathcal{B}}^{\mathcal{B}}} + \underbrace{\begin{bmatrix} 0 & 1 & & 0 \\ & 0 & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{bmatrix}}_{[N]_{\mathcal{B}}^{\mathcal{B}}}$$

The exponentiation of $[S]_{\mathcal{B}}^{\mathcal{B}}$ is easy as it is a diagonal matrix, and the exponentiation of $[N]_{\mathcal{B}}^{\mathcal{B}}$ shifts the ones on the diagonal "one up" until all the ones are "outside the matrix bounds":

$$\left([N]_{\mathcal{B}}^{\mathcal{B}}\right)^m = \begin{cases} \begin{bmatrix} 0 & 0 & \cdots & 1 & \cdots & 0 \\ & 0 & 0 & & \ddots & \vdots \\ & & \ddots & \ddots & & 1 \\ & & & \ddots & 0 & \vdots \\ & & & & 0 & 0 \\ 0 & & & & & 0 \end{bmatrix} (m+1)\text{-th diagonal} & \text{if } m \leq k-1 \\ 0_{k\times k} & \text{otherwise} \end{cases}$$

Therefore, using the binomial expansion formula above, we obtain the following result:

$$\begin{aligned} (J_\lambda)^m = \left([S]_{\mathcal{B}}^{\mathcal{B}} + [N]_{\mathcal{B}}^{\mathcal{B}}\right) &= \sum_{k=0}^{n} \binom{m}{k} \left([S]_{\mathcal{B}}^{\mathcal{B}}\right)^{m-k} \left([N]_{\mathcal{B}}^{\mathcal{B}}\right)^k \\ &= \begin{bmatrix} \lambda^m & \binom{m}{1}\lambda^{m-1} & \binom{m}{2}\lambda^{m-2} & \cdots & \\ & & & & \vdots \\ & & & & \binom{m}{2}\lambda^{m-2} \\ & & & & \binom{m}{1}\lambda^{m-1} \\ 0 & & & & \lambda^m \end{bmatrix} \end{aligned}$$

where the elements in the upper triangle are repeated along the diagonals and the strictly lower triangle is all zeroes. And so, we conclude

$$[T]_{\mathcal{B}}^{\mathcal{B}} = P \begin{bmatrix} J_{\lambda_1} & & & \\ & J_{\lambda_2} & & \\ & & \ddots & \\ & & & J_{\lambda_k} \end{bmatrix} P^{-1} \Rightarrow \left([T]_{\mathcal{B}}^{\mathcal{B}}\right)^m = P \begin{bmatrix} (J_{\lambda_1})^m & & & \\ & (J_{\lambda_2})^m & & \\ & & \ddots & \\ & & & (J_{\lambda_k})^m \end{bmatrix} P^{-1}$$

Matrix Exponentials: Because we can now compute matrix powers, we can technically compute any function of a matrix with a Taylor series expansion (given that we have the related convergence issues resolved). We'll deal with the exponential function here:

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$$

For a square matrix A with a similar Jordan form $A = PJP^{-1}$, we then expect to have

$$\begin{aligned} \exp(A) &= I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \ldots \\ &= PIP^{-1} + PJP^{-1} + \frac{(PJP^{-1})^2}{2!} + \frac{(PJP^{-1})^3}{3!} + \ldots \\ &= P\left(I + J + \frac{J^2}{2!} + \frac{J^3}{3!} + \ldots\right) P^{-1} = P\exp(J)P^{-1} \end{aligned}$$

So, we just need the exponential of the Jordan form of A to compute its exponential. But again, because the Jordan blocks do not interact with one another, the exponential of a single Jordan block is enough to derive and the rest follows easily. So, let J_λ be a $k \times k$ Jordan block with eigenvalue λ. Then, we can find that

$$\exp(J_\lambda) = e^\lambda \begin{bmatrix} 1 & 1 & 1/2! & 1/3! & \dots & 1/(k-1)! \\ & & & & & \vdots \\ & & & & & 1/3! \\ & & & & & 1/2! \\ & & & & & 1 \\ 0 & & & & & 1 \end{bmatrix}$$

where again the elements in the upper triangle are repeated along the diagonals and the strictly lower triangle is all zeroes. And so, the exponential of the total Jordan form J becomes

$$J = \begin{bmatrix} J_{\lambda_1} & & & \\ & J_{\lambda_2} & & \\ & & \ddots & \\ & & & J_{\lambda_k} \end{bmatrix} \Rightarrow \exp(J) = \begin{bmatrix} \exp(J_{\lambda_1}) & & & \\ & \exp(J_{\lambda_2}) & & \\ & & \ddots & \\ & & & \exp(J_{\lambda_k}) \end{bmatrix}$$

and so we are done:

$$\exp(A) = P \exp(J) P^{-1}$$