MATH2501/2601 Revision Sheet

UNSW Mathematics Society: Bruce Chen, Lachlan Tobin, Sai Nair

We would like to preface this document by saying that this resource is first and foremost meant to be used as a reference and should NOT be used as a replacement for the course resources or lecture recordings. Said resources provided on Moodle are wonderfully written and contain an abundance of fully worked solutions and in depth explanations. Studying for this course using *only* this revision sheet would not be sufficient.

In addition, although the authors have tried their best to include everything essential taught in the course, it was ultimately up to their discretion on whether or not to include results/theorems/definitions etc. Anything that is missing is most definitely a conscious choice made by the authors.

Finally, any and all errors found within this document are most certainly our own. If you have found an error, please contact us via our Facebook page, or give us an email.

Contents

Groups and Fields

Groups

Definition

A group G is a **non-empty** set with a binary operation $*$ defined on it. It satisfies the following properties:

1. *Closure:* for all $a, b \in G, a * b \in G$
2. *Associativity:* for all $a, b, c \in G, a*(b*c) = (a*b)*c$
3. *Identity:* there exists an element $e \in G$ such that, for all $a \in G$, $a * e = e * a = a$
4. *Inverse:* for all $a \in G$, there is $a' \in G$ such that $a * a' = a' * a = e$.

The group is the pair of the set and the operation, often denoted $(G, *)$.

A group is called *abelian* if it also satisfies *commutativity:* $a * b = b * a$ for all $a, b \in G$.

Properties of Groups

The group definition leads to some other useful properties:

- The identity element is unique
- The inverse of any given element is always unique
- $(a^{-1})^{-1} = a$ i.e. the inverse of an inverse is the original element
- For all elements $a, b \in G$, $(a * b)^{-1} = b^{-1} * a^{-1}$
- Elements can be 'cancelled' on both sides of an equation i.e. if $a, b, c \in G$ and $a * b = a * c$, then $b = c$. This also holds true if $b * a = c * a$.

Permutation Groups

Let $\Omega_n = \{1, 2, 3, \ldots, n\}$. There are $n!$ permutations of Ω_n, and each of these map the elements of Ω_n to other elements in Ω_n, forming a bijection. The set of these permutation maps, with the operation of composition, the *symmetric group*, denoted as S_n.

Fields

Definition

A field $(\mathbb{F}, +, \times)$ is a set $\mathbb{F}$ with two binary operations, $+, \times$ defined on it (not necessarily the obvious addition and multiplication). It satisfies the following properties:

1. $(\mathbb{F}, +)$ is an abelian group with its identity element being denoted 0 and the inverse element of $a \in \mathbb{F}$ being $-a$
2. $(\mathbb{F}*, \times)$ is an abelian group, where $\mathbb{F}* = \mathbb{F} \setminus 0$. Its identity element is denoted 1 and the inverse element of $a \in \mathbb{F}$ being a^{-1}
3. Distributive laws from obvious addition and multiplication hold with the two binary operations i.e. if $a, b, c \in \mathbb{F}, a \times (b + c) = a \times b + a \times c$ and $(a + b) \times c = a \times c + b \times c$.

Note that some extra notation is often used: $a - b = a + (-b)$ and $a/b = a \times b^{-1}$.

Properties of Fields

The field definition leads to some other useful properties:

- $a \times 0 = 0$
- $a \times (-b) = -(a \times b)$
- If $a \times b = 0$, $a = 0$ or $b = 0$.

Subgroups and Subfields

Definition

Let $(G, *)$ be a group, and H be a non-empty subset of G. If $(H, *)$ is a group, then it is called a *subgroup* of G, denoted as $H \leq G$. We say that H *inherits* the group structure of G.

Let $(\mathbb{F}, +, \times)$ be a field, and $\mathbb{E}$ be a non-empty subset of $\mathbb{F}$. If $(\mathbb{E}, +, \times)$ is a field, then it is called a *subfield* of $\mathbb{F}$, denoted $\mathbb{E} \leq \mathbb{F}$.

Subgroup Lemma

A non-empty subset H of a group $(G, *)$ is a subgroup of G if and only if, for all $a, b \in H$:

1. $a * b \in H$ (closure)
2. $a^{-1} \in H$ (inverse).

Subfield Lemma

A non-empty and non-trivial subset $\mathbb{E}$ of a field $\mathbb{F}$ is a subfield of $\mathbb{F}$ if and only if:

1. $a + b \in \mathbb{E}$ (closure under addition)
2. $-a \in \mathbb{E}$ (inverse under addition)
3. $a \times b \in \mathbb{E}$ (closure under multiplication)
4. $b^{-1} \in \mathbb{E}$ for $b \neq 0$ (inverse under multiplication).

Morphisms

Definition

A homomorphism is a map $\phi : G \to H$ (where $(G, *)$ and $(H, \circ)$ are groups), with the special property that $\phi(a * b) = \phi(a) \circ \phi(b)$ for all $a, b \in G$.
If ϕ is bijective, then it is called an isomorphism, with G, H being called isomorphic (to each other). These are essentially identical groups with regard to group theory.

Properties of Homomorphisms

The homomorphism definition leads to some other useful properties:

- ϕ maps the identity of G to the identity of H
- ϕ maps inverses in G to inverses in H i.e. $\phi(a^{-1}) = (\phi(a))^{-1}$
- Isomorphisms have an inverse which are also isomorphisms.

Kernel and Image of a Homomorphism

For a homomorphism $\phi : G \to H$, with H having an identity element e':

- Its kernel is the set $\ker(\phi) = \{g \in G : \phi(g) = e'\}$, with $\ker(\phi) \leq G$
- Its image is the set $\operatorname{im}(\phi) = \{h \in H : h = \phi(g)$ for some $g \in G\}$, with $\operatorname{im}(\phi) \leq H$.

ϕ is injective (one-to-one) if and only if $\ker(\phi) = \{e\}$. If ϕ is injective, we also have that $\operatorname{im}(\phi)$ is isomorphic to G (the converse is not necessarily true).

Vector Spaces

Definition

Consider an abelian group $(V, +)$ with identity element $\mathbf{0}$ and a field $\mathbb{F}$, with a function from $\mathbb{F} \times V$ to V (scalar multiplication, denoted $\alpha\mathbf{v}$). V is then called a vector space over $\mathbb{F}$ if it satisfies the following properties for all $\alpha, \beta \in \mathbb{F}, \mathbf{u}, \mathbf{v} \in V$:

1. $\alpha(\beta\mathbf{v}) = (\alpha\beta)\mathbf{v}$ (associativity)
2. $1\mathbf{v} = \mathbf{v}$ (identity element of $\mathbb{F}$ acts as an identity map)
3. $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$ (distributivity of scalar multiplication over vector addition)
4. $(\alpha + \beta)\mathbf{u} = \alpha\mathbf{u} + \beta\mathbf{u}$ (distributivity of scalar addition over scalar multiplication).

Note that there are 6 axioms involved in proving a vector space - these come from the abelian group and proving closure for scalar multiplication.

Properties

The vector space definition leads to some other useful properties:

- $0\mathbf{v} = \mathbf{0}$ and $\lambda\mathbf{0} = \mathbf{0}$
- $\lambda\mathbf{v} = \mathbf{0}$ implies $\lambda = 0$ or $\mathbf{v} = \mathbf{0}$
- $(-1)\mathbf{v} = -\mathbf{v}$ (the inverse of $\mathbf{v}$)
- $\lambda\mathbf{v} = \lambda\mathbf{v}$ with $\lambda \neq 0$ means that $\mathbf{v} = \mathbf{w}$

Subspaces

Definition

If V is a vector space over $\mathbb{F}$ and $U \subseteq V$, then U is a *subspace* of V, denoted $U \leq V$, if it is a vector space over $\mathbb{F}$ with the same vector addition and scalar multiplication.

The trivial subspace ($\{\mathbf{0}\}$) and V itself are always subspaces.

Subspace Test Lemma

If V is a vector space over $\mathbb{F}$ and U is a non-empty subset, then it is a subspace of V if and only if for all $\mathbf{u}, \mathbf{v} \in U$ and $\alpha \in \mathbb{F}$, $\alpha\mathbf{u} + \mathbf{v} \in U$.

Linear Combinations, Spans and Independence

Linear Combinations

For a vector space V over $\mathbb{F}$, a (finite) linear combination of vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \in V$ is any vector which can be expressed as

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n,$$

where $\alpha_i \in \mathbb{F}$ for $1 \leq i \leq n$.

Span

If S is a subset of V, then the *span* of S is the set of all finite linear combinations of vectors in S, denoted $\mathrm{span}(S)$.

If $\mathrm{span}(S) = V$, then we say S *spans* V, or S is a *spanning set* of V.

Note that $\mathrm{span}(S) \leq V$ if S is a non-empty subset of V.

Linear Independence

A non-empty (and finite) subset S of V is linearly independent if, for all vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \in S$,

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n = \mathbf{0},$$

with $\alpha_i \in \mathbb{F}$ for $1 \leq i \leq n$, implies $\alpha_i = 0$ for $1 \leq i \leq n$. A set which isn't linearly independent is called linearly dependent.

Linear Dependence

If S is an arbitrarily ordered set of n linearly independent vectors, then at least one of the vectors in the set can be written as a linear combination of previous vectors.

Properties of Linear Independence

- Any subset of a linearly independent set is also linearly independent
- If $\mathbf{v} \in \mathrm{span}(S)$ and $\mathbf{v} \notin S$, $S \cup \{\mathbf{v}\}$ is linearly dependent. Similarly, if S is linearly independent and $S \cup \{\mathbf{v}\}$ is not, then $v \in \mathrm{span}(S)$
- If $S_1 \subseteq S_2$, then $\mathrm{span}(S_1) \subseteq \mathrm{span}(S_2)$
- $\mathrm{span}(S \cup \{\mathbf{v}\}) = \mathrm{span}(S)$ if and only if $\mathbf{v} \in \mathrm{span}(S)$
- If S is linearly dependent, then there exists $\mathbf{v} \in S$ such that $\mathrm{span}(S \setminus \{\mathbf{v}\}) = \mathrm{span}(S)$
- For an invertible matrix $P \in \mathrm{GL}(p, \mathbb{F})$ and a linearly independent set $\{\mathbf{v}_i\}$, $\{P\mathbf{v}_i\}$ is also linearly independent (related to linear transformations)

Bases

Definition

A set $S \subseteq V$ is a *basis* for V if and only if it is a linearly independent spanning set of V. An example of this is the standard basis of $\mathbb{F}^n$, which has a basis $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ where $\mathbf{e}_i$ has 1 at the ith position and 0 everywhere else.

An equivalent result is that S is a basis if and only if every vector in V can be *uniquely* expressed as a linear combination of the vectors in S.

A useful result to know is that any linearly independent set can be extended to a basis by adding more vectors (implying that linearly independent sets must have at most the same number of vectors as a basis)

Dimension

Definition

If a vector space V can have a finite spanning set, it can have a finite basis, and all bases will contain the same number of elements, which is called the dimension of V, or $\dim(V)$. This can be infinite (though we mainly deal with finite dimensional vector spaces).

Properties from Spans, Linear Independence and Bases

For an n-dimensional vector space V:

- The number of elements in a spanning set of V must be at least n
- The number of elements in a linearly independent set of V must be at most n
- If $\mathrm{span}(S) = V$ and $|S| = n$ then S is a basis
- If S is linearly independent and $|S| = n$ then S is a basis.

Another useful property is that if $U \leq V$, then $dim(U) = dim(V)$ with equality if and only if $U = V$.

Coordinates in Bases

Definition

Consider a vector v, an element of vector space V, and an ordered basis $\mathcal{B}$ of V with vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$. If $v = \alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n$ (a unique representation), then

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

is called the coordinate vector of $\mathbf{v}$ with respect to an ordered basis of V, called $\mathcal{B}$. This is often denoted as $\boldsymbol{\alpha} = [\mathbf{v}]_{\mathcal{B}}$.

Addition and Scalar Multiplication in Different Bases

- $\mathbf{u} = \mathbf{v}$ if and only if $[\mathbf{u}]_{\mathcal{B}} = [\mathbf{v}]_{\mathcal{B}}$ for every basis $\mathcal{B}$ of V.
- $[\mathbf{u} + \mathbf{v}]_{\mathcal{B}} = [\mathbf{u}]_{\mathcal{B}} + [\mathbf{v}]_{\mathcal{B}}$
- $[\lambda \mathbf{v}]_{\mathcal{B}} = \lambda [\mathbf{v}]_{\mathcal{B}}$.

This is useful as any vector space can have calculations done on it by forming coordinates in $\mathbb{F}^n$.

Sums and Direct Sums of Vector Spaces

Definition

If you have two vector spaces S, T over the same field $\mathbb{F}$, the sum $S + T$ is defined as

$$S + T = \{\mathbf{a} + \mathbf{b} : \mathbf{a} \in S, \mathbf{b} \in T\}.$$

If $S \cap T = \{\mathbf{0}\}$, then this sum is called a *direct sum*, denoted as $S \oplus T$.

Properties

A sum $S + T$ is direct if and only if any vector in $\mathbf{x} \in S + T$ can be written *uniquely* as a sum of two vectors, one from S and one from T.

$\dim(S) + \dim(T) = \dim(S + T) + \dim(S \cap T)$. For a direct sum, this can be simplified to $\dim(S) + \dim(T) = \dim(S \oplus T)$

Complementary Subspaces

For any vector space $X \leq V$, there exists a (not necessarily unique) subspace Y such that $X \oplus Y = V$, called the complementary subspace.

External Direct Sums

The above are called internal direct sums. External direct sums are the Cartesian product of two vector spaces, and turning that into its own vector space. Addition and scalar multiplication are done elementwise. This external direct sum is also denoted with the symbol $\oplus$.

Linear Transformations

Linear Transformations

Definition

Let V, W be vector spaces over a field F. A linear transformation or linear map $T : V \to W$ is a function that satisfies

- $T(\mathbf{v} + \mathbf{u}) = T(\mathbf{v}) + T(\mathbf{u})$
- $T(\lambda \mathbf{v}) = \lambda T(\mathbf{v})$

for all $\mathbf{v}, \mathbf{u} \in V$ and $\lambda \in \mathbb{F}$.

Linearity Test Lemma

A function $T : V \to W$ for vector spaces V, W over $\mathbb{F}$ is linear if and only if

$$T(\lambda \mathbf{v} + \mathbf{u}) = \lambda T(\mathbf{v}) + T(\mathbf{u})$$

for all $\mathbf{v}, \mathbf{u} \in V$ and $\lambda \in \mathbb{F}$.

Properties

Let V, W, X be vector spaces over $\mathbb{F}$ and let $T : V \to W$ and $S : W \to X$ be linear transformations between these spaces.

(a) The identity map $\mathrm{id} : V \to V$ defined by $\mathrm{id}(\mathbf{v}) = \mathbf{v}$ is linear.

(b) $T(\mathbf{0}) = \mathbf{0}$ and $T(-\mathbf{v}) = -\mathbf{v}$.

(c) The map $S \circ T : V \to X$ is linear.

(d) If T is invertible, then the map $T^{-1} : W \to V$ is also linear.

(e) Let $\mathcal{B} = \{\mathbf{v}_1, ..., \mathbf{v}_p\}$ be a basis for V. Then $T : V \to \mathbb{F}^p$ defined by $T(\mathbf{x}) = [\mathbf{x}]_{\mathcal{B}}$ is linear. In other words, taking coordinates is a linear map.

Kernel and Image

Kernel

The kernel or nullspace of a linear map $T : V \to W$ is the set of all vectors in V which map to $\mathbf{0}$, or equivalently,

$$\ker(T) = \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}\}.$$

Image

The image of a particular $U \leq V$ is the set of all possible outputs under the linear map $T : V \to W$, or equivalently $T(U)$ where

$$T(U) = \{T(\mathbf{u}) \colon \mathbf{u} \in U\}.$$

The image of T, denoted $\mathrm{im}(T)$ is the image of V under T; that is, $\mathrm{im}(T) = T(V)$

Properties

The kernel and image are subspaces. Let $T : V \to W$ be a linear transformation and suppose $U \leq V$. Then:

(a) $\ker T$ is a subspace of V.

(b) $\operatorname{im} T$ is a subspace of W.

(c) If U is finite dimensional then so is $T(U)$.

Rank and Nullity

Define *nullity* to be the dimension of the kernel, and *rank* to be the dimension of the image. That is, if T is a linear map, then

$$\text{nullity}(T) = \dim(\ker T), \quad \text{rank}(T) = \dim(\operatorname{im} T).$$

For a linear map T to be injective (one-to-one) then $\ker(T) = \{\mathbf{0}\}$ or equivalently $\text{nullity}(T) = 0$

Rank-Nullity Theorem

If V is a finite dimensional vector space over a field $\mathbb{F}$ and $T : V \to W$ is a linear transformation, then

$$\text{rank}(T) + \text{nullity}(T) = \dim(V).$$

As a corollary suppose that $\dim(V) = \dim(W)$. Then, all the following statements are equivalent:

(i) T is invertible (bijective).

(ii) T is one-to-one (injective) so $\text{nullity}(T) = 0$.

(iii) T is onto (surjective) so $\text{rank}(T) = \dim(V)$.

Isomorphisms

An invertible linear map between two vector spaces V and W is an isomorphism. Then, V and W are *isomorphic* if such a map exists. Finite dimensional vector spaces are isomorphic if and only if their dimensions are equal.

A p dimensional vector space is isomorphic to $\mathbb{F}^p$, and the isomorphism is taking coordinates with respect to a certain basis.

Matrices as Linear Maps

Theorem

Suppose V, W are finite dimensional vector spaces over $\mathbb{F}$. Let $\dim(V) = q$ and $\dim W = p$, and let $\mathcal{B}$ be a basis for V with $\mathcal{C}$ being a basis for W.

If $T : V \to W$ is a linear map then there exists a unique matrix $A \in M_{p,q}(\mathbb{F})$ such that

$$[T(\mathbf{v})]_{\mathcal{C}} = A[\mathbf{v}]_{\mathcal{B}}$$

Conversely, for any matrix $A \in M_{p,q}(\mathbb{F})$, the above equation defines a unique linear map from V to W. A is referred to as the *matrix of the linear map*, and often denoted as $[T]_{\mathcal{C}}^{\mathcal{B}}$.

Intuition: applying linear maps is represented by matrix multiplication of coordinate vectors. The matrix of the linear map, $[T]_{\mathcal{C}}^{\mathcal{B}}$, is found by applying T to each vector in $\mathcal{B}$. Each column will be this vector expressed with respect to the basis $\mathcal{C}$.

Properties

(a) Let $T : V \to W$ and $S : W \to X$ be linear maps, where V, W, X have bases $\mathcal{A}, \mathcal{B}, \mathcal{C}$ respectively. Then the matrix representing the linear map $S \circ T$ is the product of the matrices representing S and T. That is,

$$[S \circ T]_{\mathcal{C}}^{\mathcal{A}} = [S]_{\mathcal{C}}^{\mathcal{B}}[T]_{\mathcal{B}}^{\mathcal{A}}$$

Composition of linear maps corresponds to matrix multiplication

(b) For an invertible linear map T with corresponding matrix M, the inverse map T^{-1} os represented by the inverse matrix M^{-1}

Change of Basis

Let $\mathcal{B} = \{\mathbf{v}_1, ..., \mathbf{v}_p\}$ be a basis for a vector space V with each $\mathbf{v}_i$ written with respect to the standard basis. Then the matrix with columns $\mathbf{v}_1, ..., \mathbf{v}_n$ corresponds to transforming coordinates in $\mathcal{B}$ to the standard basis and is denoted $[\text{id}]_{\mathcal{S}}^{\mathcal{B}}$.

Change of Basis Matrix

Consider two bases $\mathcal{B}$ and $\mathcal{C}$. The matrix $[\text{id}]_{\mathcal{C}}^{\mathcal{B}}$ is the *change of basis* matrix between $\mathcal{B}$ and $\mathcal{C}$.

$$[\mathbf{v}]_{\mathcal{C}} = [\text{id}]_{\mathcal{C}}^{\mathcal{B}}[\mathbf{v}]_{\mathcal{B}}.$$

Commutative Diagram

The following diagram summarises the change of basis operations for a linear map T from V to W.

$$\begin{tikzcd}
\mathcal{S} & V \arrow[r, "A"] & W & \mathcal{S}' \\
\mathcal{B} & V \arrow[u, "P"] \arrow[r, "M"'] & W \arrow[u, "Q"'] & \mathcal{C}
\end{tikzcd}$$

Each element of the diagram is defined as:

- The matrix of T with respect to the standard basis in the domain ($\mathcal{S}$) and standard basis in the codomain ($\mathcal{S}'$) is A.
- P is the change of basis from $\mathcal{B}$ to $\mathcal{S}$ in V, while Q is the change of basis from $\mathcal{C}$ to $\mathcal{S}'$ in W.
- M is the matrix of T with respect to $\mathcal{B}$ in the domain and $\mathcal{C}$ in the codomain.

Thus $M = Q^{-1}AP$.

Normal Form

Invariant Subspaces

Let T be a linear transformation from V to V. If $X \leq V$ and $T(X) \leq X$ then X is an *invariant* subspace of V.

Theorem

Suppose $V = X \oplus Y$ for invariant X, Y, under a linear map T, with dimensions p, q respectively. Then there exists a basis $\mathcal{B}$ so that the matrix of T takes the form

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & B \end{pmatrix}$$

where A is $p \times p$ and B is $q \times q$. This matrix is referred to as the direct sum of A and B, denoted $A \oplus B$.

Normal Form

Let $T : V \to W$ be linear with $\dim V = p$, $\dim W = q$, $\text{rank} T = r$. Then there exist bases $\mathcal{B}$ and $\mathcal{C}$ in V, W respectively such that the matrix of T takes the form

$$N_{q,p;r} = \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \in M_{q,p}(\mathbb{F})$$

I_r is the $r \times r$ identity matrix and each zero matrix is appropriately sized. This is the normal form of the linear map T.

Similarity

$p \times p$ matrices A, B are similar if there exists a $p \times p$ invertible matrix P such that $B = P^{-1}AP$.

Similarity Theorem

Matrices are similar if and only if they represent the same linear map with respect to two different bases.

Similarity Invariants

Similarity invariants are properties that remain the same for all similar matrices. Examples are:

- Rank
- Nullity
- Determinant
- Trace

Multilinear Maps

Let V_1, V_2, W be three vector spaces over $\mathbb{F}$ and define a map $T : V_1 \times V_2 \to W$. T is bilinear if it is linear in each argument, that is,

$$T(\mathbf{v}_1 + \lambda \mathbf{u}_1, \mathbf{v}_2) = T(\mathbf{v}_1, \mathbf{v}_2) + \lambda T(\mathbf{u}_1, \mathbf{v}_2),$$

and the same for the second argument. This can be extended to trilinear maps which are linear in three arguments.

Inner Product Spaces

Inner Products

Generalising the real and complex dot product, we define other 'inner products' which maintain some properties from the real and complex dot products.

Definition

Given a vector space V over $\mathbb{F}$, an inner product is a complex valued function $\langle , \rangle : V \times V \to \mathbb{F}$ which, for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, satisfies:

1. $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$ (additive part of sesquilinearity)
2. $\langle \mathbf{u}, \alpha\mathbf{v} \rangle = \alpha\langle \mathbf{u}, \mathbf{v} \rangle$ (multiplicative part of sesquilinearity)
3. $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$ (conjugate symmetry)
4. $\langle \mathbf{v}, \mathbf{v} \rangle \in \mathbb{R}$ and $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ with equality if and only if $\mathbf{v} = \mathbf{0}$ (positive definiteness).

V paired with $\langle , \rangle$ is called an 'inner product space'. The corresponding norm is then $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$.

Properties

1. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
2. $\langle \alpha\mathbf{u}, \mathbf{v} \rangle = \overline{\alpha}\langle \mathbf{u}, \mathbf{v} \rangle$
3. $\|\alpha\mathbf{v}\| = |\alpha|\|\mathbf{v}\|$
4. $\langle \mathbf{x}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$ for all $\mathbf{v} \in V$
5. $|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|$ (Cauchy-Schwarz inequality)
6. $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ (Triangle inequality).

Generalising Orthogonality and Orthonormality

Definition

For an inner product space V, two non-zero vectors $\mathbf{u}, \mathbf{v}$ are orthogonal if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. This is denoted as $\mathbf{u} \perp \mathbf{v}$.
A set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is orthogonal if each pair of distinct vectors is orthogonal.
We say that it is orthonormal instead if the norm of each vector is 1 as well. This is equivalent to saying it is orthonormal if

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

for all $1 \leq i, j \leq n$.
Note that all orthogonal and orthonormal sets are linearly independent.

Projection

Definition

For $\mathbf{v} \neq \mathbf{0}$ in an inner product space V, the projection of $\mathbf{u}$ onto $\mathbf{v}$ is defined as

$$\mathrm{proj}_{\mathbf{v}}(\mathbf{u}) = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v}.$$

This projection is the component of $\mathbf{u}$ in the direction of $\mathbf{v}$ - subtracting it from $\mathbf{u}$ will give a vector orthogonal to $\mathbf{v}$.

Properties

- If $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a set of orthogonal vectors, and $\mathbf{v} \in \mathrm{span}(S)$, then $\mathbf{v}$ is equal to
$$\sum_{i=1}^{k} \mathrm{proj}_{\mathbf{v}_i} \mathbf{v},$$
i.e. the sum of the projection of $\mathbf{v}$ onto all vectors in $\mathbf{S}$.
- Specifically, if S is orthonormal, then $\langle \mathbf{v}_i, \mathbf{v}_i \rangle = 1$, so
$$v = \sum_{i=1}^{k} \langle \mathbf{v}_i, \mathbf{v} \rangle \mathbf{v}_i.$$

Gram-Schmidt Process

The Gram-Schmidt Process enables us to form an orthonormal basis for a finite dimensional inner product space, by taking a basis, orthogonalising it and then normalising it.

Construction

Given a basis $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, we define a new set of orthogonal vectors $S' = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ where:

$$\begin{aligned}
\mathbf{w}_1 &= \mathbf{v}_1 \\
\mathbf{w}_2 &= \mathbf{v}_2 - \mathrm{proj}_{\mathbf{w}_1}(\mathbf{v}_2) \\
\mathbf{w}_3 &= \mathbf{v}_3 - \mathrm{proj}_{\mathbf{w}_1}(\mathbf{v}_3) - \mathrm{proj}_{\mathbf{w}_2}(\mathbf{v}_3) \\
&\vdots \\
\mathbf{w}_n &= \mathbf{v}_n - \sum_{i=1}^{n-1} \mathrm{proj}_{\mathbf{w}_i}(\mathbf{v}_n).
\end{aligned}$$

From here, we normalise our vectors by defining $\mathbf{e}_i = \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|}$ for $1 \leq i \leq n$, to get our orthonormal basis $T = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$.

Orthonormal Complement

Definition

For a inner product space V and a subspace X, the space $Y = \{\mathbf{y} \in V : \langle \mathbf{y}, \mathbf{x} \rangle = 0 \text{ for all } \mathbf{x} \in X\}$ is called the orthogonal complement to X, denoted $X^{\perp}$.

Properties

- $V = W \oplus W^\perp$
- $\left(W^\perp\right)^\perp = W$
- If $\mathbf{v} \in V$ then $\mathbf{v} - \operatorname{proj}_W(\mathbf{v})$ is in $W^\perp$, where $\operatorname{proj}_W(\mathbf{v})$ is defined as the component of which is part of W (which must exist from point 1)
- For $\mathbf{w} \in W, \operatorname{proj}_W(\mathbf{w}) = \mathbf{w}$
- $\|\operatorname{proj}_W(\mathbf{v})\| \leq \|\mathbf{v}\|$ for all $\mathbf{v}$.
- $\operatorname{proj}_W(\mathbf{v}) + \operatorname{proj}_{W^\perp}(\mathbf{v}) = \mathbf{v}$.
- $\|\mathbf{v} - \mathbf{w}\| \geq \|\mathbf{v} - \operatorname{proj}_W(\mathbf{v})\|$ with equality if and only if $\mathbf{w} = \operatorname{proj}_W(\mathbf{v})$ (i.e. $\mathbf{v} - \mathbf{w} \in W^\perp$).

Adjoints

Covectors

Covectors are linear maps from a vector space V to the field it lies on, $\mathbb{F}$. The set of all covectors in V is called the dual space, V^*.

All covectors can be represented as a dot product with a constant vector (which is unique), and this leads to an isomorphism between V and V^*.

Going from the map to this vector is called 'raising the index,' denoted $\mathbf{t} = T^\sharp$. Going in the opposite direction is called 'lowering the index,' denoted $T = \mathbf{t}^\flat$.

Defining the Adjoint

For a linear map $T : V \to W$ where V, W are finite dimensional, there is a unique linear map $T* : W \to V$, called the adjoint, with

$$\langle \mathbf{w}, T(\mathbf{v}) \rangle = \langle T^*(\mathbf{w}), \mathbf{v} \rangle$$

for all $\mathbf{v} \in V, w \in W$.

Specifically for matrices, we find that the matrix representation can also have an adjoint, which is defined as $A^* = \overline{A^T}$

Properties

For linear transforms S, T mapping from V to W,

- $(S + T)^* = S^* + T^*$
- $(\alpha T)^* = \overline{\alpha} T^*$
- $(T^*)^* = T$
- For linear $U : W \to X$, $(U \circ T)^* = T^* \circ U^*$
- If T has a matrix representation A, then T^* has a matrix representation $A^* = \overline{A^T}$.

Maps with Special Adjoints

Definitions

For a linear map $T : V \to V$ where V is a finite dimensional inner product space. T is:

- unitary if $T^* = T^{-1}$
- an isometry if $\|T(\mathbf{v})\| = \|\mathbf{v}\|$ for all $\mathbf{v} \in V$
- self-adjoint or Hermitian if $T^* = T$

These definitions can also be applied to the corresponding matrices - A is unitary if $A^* = A^{-1}$ and Hermitian if $A = A^*$.

Properties

For a linear map from a finite dimensional inner product space V to itself, the following are equivalent

- T is an isometry
- T^* is an isometry
- T is unitary
- T^* is unitary
- $\langle T(\mathbf{v}), T(\mathbf{w}) \rangle = \langle \mathbf{v}, \mathbf{w} \rangle$ for all $\mathbf{v}, w$
- If $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is an orthonormal basis for V, so is $\{T(\mathbf{e}_1), \ldots, T(\mathbf{e}_n)\}$.
- The corresponding matrix of T has columns (and rows) which form an orthonormal basis of $\mathbb{C}^p$

QR Factorisation

Construction

We aim to decompose $A \in M_{p,q}(\mathbb{F})$ with rank q (meaning $p \geq q$) into a product QR where $Q \in M_{p,q}(\mathbb{F})$ with orthonormal columns, and $R \in GL_q(\mathbb{F})$ which is upper triangular.

To do this, we use the Gram-Schmidt process to generate an orthogonal basis from the columns of A (the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$), generating the set $\{\mathbf{w}_1, \ldots, \mathbf{w}_q\}$ with a corresponding normalised set of vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_q\}$.

We then define

$$Q = \begin{pmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{pmatrix}$$

and

$$R = \begin{pmatrix} \|\mathbf{w}_1\| & \langle \mathbf{q}_1, \mathbf{v}_2 \rangle & \cdots & \langle \mathbf{q}_1, \mathbf{v}_q \rangle \\ & \|\mathbf{w}_2\| & \cdots & \langle \mathbf{q}_2, \mathbf{v}_q \rangle \\ & & \ddots & \vdots \\ & & & \|\mathbf{w}_q\| \end{pmatrix}$$

.

Alternative Construction

If A has $p > q$ (i.e. isn't square), then we can also write $A = \tilde{Q}\tilde{R}$, with $\tilde{Q} \in GL_p(\mathbb{F})$ being unitary and $\tilde{R} \in M_{p,q}(\mathbb{F})$ of rank q and in echelon form.

We form $\tilde{Q}$ by completing the columns of Q to an orthonormal basis, and create $\tilde{R}$ by adding $p - q$ rows of zeroes to the bottom of R.

Method of Least Squares

Definition

When solving a system of linear equations $A\mathbf{x} = \mathbf{b}$, we often can't find the best solution. The best we can get is done by minimising the sum of the squares of the errors, which is equivalent to minimising $\|A\mathbf{x} - \mathbf{b}\|$ with the norm defined based on the standard inner product. Such a solution always exists, and is unique provided the columns of A are independent.

Equivalent Problem

Any least squares solution to $A\mathbf{x} = \mathbf{b}$ will be a solution to the equations $A^*A\mathbf{x} = A^*\mathbf{b}$, known as the normal equations.

If A has independent columns, then (A^*A) is invertible, meaning the least squares solution is $\mathbf{x} = (A^*A)^{-1}A^*\mathbf{b}$.

Determinants

Determinants from the Definition

Permutations

Recall $\mathcal{S}_n$ is group of permutations on n objects. Let $\Omega_n = \{1, 2, ..., n\}$. Then a shorthand notation for a permutation on Ω_n is $[p_1, p_2, ..., p_n]$ where 1 is mapped to p_1, 2 to p_2 and so forth.

Inversions

An inversion is when a permutation causes a larger number to preceed a smaller number, for example 4 before 2. For a permutation $\sigma = [p_1, p_2, ..., p_n]$ which contains k inversions, define

$$\operatorname{sign}(\sigma) = (-1)^k$$

A transposition (swap) has an odd number of inversions, hence its sign is negative.

Determinant Definition

The determinant of an $n \times n$ matrix $A = (a)_{ij}$ is

$$\det A = \sum_{\sigma \in \mathcal{S}_n} \operatorname{sign}(\sigma) a_{1\sigma(1)} a_{2\sigma 2} \cdots a_{n\sigma(n)}$$

Key Properties

Let A be an $n \times n$ matrix

(i) $\det A = \sum_{\sigma \in \mathcal{S}_n} \operatorname{sign}(\sigma) a_{\sigma(1)1} a_{\sigma 22} \cdots a_{\sigma(n)n}$. That is, we can order by columns or rows.

(ii) $\det A = \det A^T$ and $\det A^* = \overline{\det A}$.

(iii) If there is a 0 row or column, $\det(A) = 0$.

(iv) If the rows or columns of A are permuted, the determinant is multiplied by the sign of the permutation - a consequence is if rows are transposed then the determinant is multiplied by -1.

(v) If A has two equal columns or rows the determinant is 0.

(vi) Adding a multiple of a row or column to another row or column does not affect the determinant.

(vii) The determinant is a multilinear and alternating map from the rows and columns to a single number.

(viii) If B is also $n \times n$ then $\det(AB) = \det A \det B$.

(ix) A is invertible if and only if $\det A \neq 0$

(x) If A is invertible then $\det A^{-1} = \frac{1}{\det A}$

Calculating Determinants

Minors

If A is a $p \times p$ matrix, the (i, j)-minor of A, denoted A_{ij}, is the matrix obtained by deleting the i^{th} row and j^{th} column from A.

If A has row i entirely 0 except for entry a_{ij} then

$$\det A = (-1)^{i+j} a_{ij} \det A_{ij}$$

Likewise goes for columns. A method to calculate determinants is then

1. Add multiples of rows/columns until one row/column is zero except for one entry.
2. Reduce the determinant to one that is one row and one column smaller using the above result.
3. Repeat this process until the determinant is 2×2 and can be easily calculated.

Row Operations and Elementary Matrices

The determinant of an upper or lower triangular matrix is the product of the diagonal elements.

Defintion: *Elementary* row operations are one of the following:

- Swapping two rows.
- Multiplying rows by a non-zero scalar.
- Adding multiples of one row to another.

An elementary matrix is formed from applying elementary row operations to the identity matrix.

Given an elementary matrix corresponding to a particular row operation E and a matrix A, applying the row operation on A is equivalent to finding EA.

(a) If E is a swap of two rows, then $\det E = -1$.

(b) If E is multiplying a row by a non-zero scalar λ, then $\det E = \lambda$.

(c) If E is adding multiples of one row to another, then $\det E = 1$.

Lemma: If A is invertible there is a sequence of elementary matrices $E_1, E_2, ..., E_k$ such that $A = E_1 E_2 \cdots E_k$ and

$$\det A = \prod_{i=1}^{k} \det E_i$$

Thus a method for determining $\det A$ is to perform row operations until A is an upper triangular matrix. Then, find the product of the diagonal elements and multiply by the corresponding determinants of the elementary row operations required.

Cofactor Expansion

In $A \in M_{p,p}(\mathbb{F})$, the cofactor of element a_{ij} is the number $c_{ij} = (-1)^{i+j} \det A_{ij}$. Then

$$\det A = \sum_{i=1}^{p} a_{ij} c_{ij}$$

The *adjugate* of A is the matrix $\mathrm{adj}(A)$, which is the transpose of the matrix of cofactors, $\mathrm{adj}(A)_{ij} = c_{ji}$. An alternative way to express the inverse matrix A^{-1} is

$$A^{-1} = \frac{\mathrm{adj}(A)}{\det A}$$

In general, cofactor expansion and the adjugate-inverse formula are only useful in 2×2 case. The most useful method is using row-reduction.

Eigenvalues and Eigenvectors

Definition

Let T be a linear map from V to V. Then we have the following definitions:

1. If $T(\mathbf{v}) = \lambda\mathbf{v}$, for $\lambda \in \mathbb{F}$ and $\mathbf{v} \neq \mathbf{0}$, then λ is an *eigenvalue* of T and $\mathbf{v}$ is its corresponding *eigenvector*.
2. If λ is an eigenvalue of T then the *eigenspace* of T is
$$E_\lambda(T) = \{\mathbf{v} \in V : T(\mathbf{v}) = \lambda\mathbf{v}\}$$
3. The set of all eigenvalues is called the *spectrum* of T.

A key distinction is that eigenvectors are never $\mathbf{0}$.

Key Properties

(i) $E_\lambda(T) = \ker(\lambda\mathrm{id} - T)$.

(ii) If $\lambda_1, ..., \lambda_k$ are distinct eigenvalues with corresponding eigenvectors $\mathbf{v}_1, ..., \mathbf{v}_k$ then $\mathbf{v}_1, ..., \mathbf{v}_k$ are linearly independent.

(iii) If $\lambda \neq \mu$ then $E_\lambda(T) \cap E_\mu(T) = \{\mathbf{0}\}$.

Diagonalisation

$A \in M_{p,p}(\mathbb{F})$ is diagonalisable on $\mathbb{F}$ if it is similar to a diagonal matrix (only non-zero entries along the main diagonal). That is,

$$D = P^{-1}AP$$

where P is an invertible $p \times p$ matrix and D is diagonal. A linear map $T : V \to V$ is diagonalisable if there is a basis of V in which the matrix representing T is diagonal.

We can write $A = PDP^{-1}$ where D is a diagonal matrix of the eigenvalues of A and P is a matrix formed by the corresponding eigenvectors as columns.

Theorem

If $T : V \to V$ is linear, for a finite dimensional vector space V over $\mathbb{F}$, T is diagonalisable if and only if V has a basis whose elements are all eigenvectors of T.

The $p \times p$ matrix A over $\mathbb{F}$ is likewise diagonalisable if and only if $\mathbb{F}^p$ has a basis consisting of eigenvectors of A.

Special Case: A $p \times p$ matrix with p distinct eigenvalues is diagonalisable.

Characteristic Polynomial

The characteristic polynomial of the $p \times p$ matrix A is defined as:

$$\mathrm{cp}_A(t) = \det(tI - A)$$

Its properties are:

(a) It is a degree p polynomial in t.

(b) The zeroes are eigenvalues of A, i.e. the set of zeroes of cp_A is the spectrum of A.

(c) It is monic.

(d) It is a similarity invariant.

In practice, eigenvalues are usually determined by solving $\det(A - tI) = (-1)^p \mathrm{cp}_A(t) = 0$ instead.

Note that as the characteristic polynomial is a similarity invariant, if A represents the linear map $T : V \to V$ then $\mathrm{cp}_T(t) = \mathrm{cp}_A(t)$. This gives further properties:

(e) λ is an eigenvalue of T if and only if $\mathrm{cp}_T(\lambda) = 0$.

(f) $W \leq E_\lambda(T)$ implies W is an invariant subspace.

(g) λ is an eigenvalue if and only if $\mathrm{nullity}(T - \lambda\mathrm{id}) > 0$

The problem of finding eigenvalues and eigenvectors for a linear map can be reduced to finding them for a matrix representing the map with respect to a certain basis.

Multiplicity

Diagonalisation of $T : V \to V$ relies on whether V (n dimensional vector space) has a basis of eigenvectors of T, that is, whether $\sum_\lambda \dim E_\lambda(T) = n$. This is contigent ont the *multiplicities* of the eigenvalues.

Definition

Let $T : V \to V$ be linear, with eigenvalue λ, hence $(t - \lambda)$ must be a factor of $\mathrm{cp}_T(t)$.

- The *geometric multiplicity* (gm) of λ is $\dim E_\lambda(T)$.
- The *algebraic multiplicity* (am) of λ is the multiplicity of $(t - \lambda)$ in $\mathrm{cp}_T(t)$.

Relationship to Determinant and Trace

NOTE: To get p roots from $\mathrm{cp}_T(t)$, the field should be $\mathbb{C}$, which is algebraically closed.

For $A \in M_{p,p}(\mathbb{C})$, A will have p eigenvalues $\lambda_1, ..., \lambda_p$ accounting for algebraic multiplicities. Further,

$$\det A = \prod_{i=1}^{p} \lambda_i \quad \text{and} \quad \mathrm{tr}(A) = \sum_{i=1}^{p} \lambda_i$$

Properties

Let $T : V \to V$ be a linear map on a finite dimensional vector space T. Then

$$1 \leq \mathrm{gm}(\lambda) \leq \mathrm{am}(\lambda)$$

Further, the four statements are equivalent:

(a) T is diagonalisable.

(b) There is a basis for V consisting of eigenvectors of T.

(c) $V = E_{\lambda_1}(T) \oplus E_{\lambda_2}(T) \oplus \cdots \oplus E_{\lambda_n}(T)$ for the distinct eigenvalues $\lambda_1, ..., \lambda_n$.

(d) $\sum_{j=1}^{k} \dim E_{\lambda_j}(T) = \dim V$

Normal Operators

A linear transformation on an inner product space V is *normal* if and only if $T^* \circ T = T \circ T^*$, that is, the linear map commutes with its adjoint.

Unitary, Hermitian, orthogonal and symmetric matrices are normal.

Properties

Let T be a normal linear map on an inner product space V with eigenvalues λ, μ.

(i) For all $\mathbf{v} \in V$, $||T(\mathbf{v})|| = ||T^*(\mathbf{v})||$.

(ii) For a scalar α, $(T - \alpha\mathrm{id})$ is also normal.

(iii) If λ is an eigenvalue of T, then $\overline{\lambda}$ is an eigenvalue of T^*.

(iv) $E_\lambda(T) = E_{\overline{\lambda}}(T^*)$.

(v) Distinct eigenvalues have orthogonal eigenspaces, so if $\lambda \neq \mu$ then $E_\lambda(T) \perp E_\mu(T)$.

(vi) Geometric multiplicity of λ equals algebraic multiplicity of λ.

Spectral Theorem

Let V be a finite dimensional inner product space over $\mathbb{C}$. Let $T : V \to V$ be a normal linear map. Then V has an orthonormal basis consisting of eigenvectors for T. Then, if A is a $p \times p$ normal matrix then there exists a unitary matrix P such that

$$P^{-1}AP = P^*AP$$

is diagonal.

Conversely, if $[T]_\mathcal{B}^\mathcal{B} = A = PDP^{-1}$ for unitary P and diagonal D (making $\mathcal{B}$ an orthonormal basis) then A and thus T are normal.

Key idea: normal maps are unitarily diagonalisable. They can be decomposed in terms of unitary matrices because the eigenvectors form an orthonormal basis, hence the usual P matrix is unitary.

Self-Adjoint Maps

Let $T : V \to V$ be a linear map over finite dimensional vector space V and field $\mathbb{F}$. Suppose T is self-adjoint (Hermitian if $\mathbb{F} = \mathbb{C}$ and symmetric if $\mathbb{F} = \mathbb{R}$). Then,

(a) The eigenvalues of T are real.

(b) There is an orthonormal basis for V consisting of eigenvectors for T.

If you can rewrite a curve in the form $\mathbf{x}^T A\mathbf{x} = c$ for some constant c, then to classify the curve, only the signs of the eigenvalues of A are necessary. For example, if both are positive then the curve is an ellipse. If one is negative, one is positive, then the curve is a hyperbola.

Unitary Maps

Suppose V is a p-dimensional inner product space over $\mathbb{F}$ with unitary map $T : V \to V$. The eigenvalues of T lie on the unit circle in $\mathbb{C}$, so are of the form $e^{i\alpha_k}$ for real α_k. Further, V has a unitary (orthonormal) basis of eigenvectors of T.

Further, if T is an isometry on V, then its characteristic polynomial is of the form

$$\mathrm{cp}_T(t) = (t-1)^a(t+1)^b \prod_{j=1}^{k} (t - e^{i\alpha_j})(t - e^{-i\alpha_j})$$

where $a + b + 2k = \dim(V)$ and $\alpha_j \in (0, \pi)$. There is an orthonormal basis for V in which the matrix for T takes the form $I_a \oplus I_b \bigoplus_{j=1}^{k} R(\alpha_j)$, where R is a rotation matrix by the angle α_j.

Singular Value Decomposition

Let A be a $p \times q$ matrix over $\mathbb{C}$. A singular value decomposition (SVD) of A is the factorisation

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

where U, V are square unitary matrices. Σ is the $p \times q$ singular value matrix, consisting of *singular values* σ_i in the diagonal entries and 0 elsewhere, ordered as $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_q \geq 0$. The columns of U and V are the left and right *singular vectors* accordingly.

For any matrix A the SVD exists and the singular values are unique.

Construction

To find the SVD of a $p \times q$ matrix A,

1. Find the eigenvalues and corresponding eigenvectors of the $q \times q$ matrix A^*A.
2. Find the unitary diagonalising matrix V of A^*A.
3. Order the eigenvalues in decreasing order, then set $\sigma_i = \sqrt{\lambda_i}$ so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_q \geq 0$. These are singular values.
4. Set $\widehat{\Sigma}$ to be the matrix with diagonal entries as non-zero singular values, then set $\Sigma = \widehat{\Sigma} \oplus O$ to be the $p \times q$ matrix formed by filling $\widehat{\Sigma}$ with zeroes.
5. Let $\mathbf{v}_i$ be the ith column of V and then the vectors $\mathbf{u}_i = {\sigma_i}^{-1}A\mathbf{v}_i$ extended to an orthonormal basis for $\mathbb{C}^p$ form the columns of the $p \times p$ unitary matrix U.

Reduced SVD

If $A \in M_{p,q}(\mathbb{C})$ has rank k, the reduced SVD of A is

$$A = \widehat{U}\widehat{\Sigma}\widehat{V}^*$$

for $\widehat{U}, \widehat{V}$ with orthonormal columns and $k \times k$, invertible and diagonal $\widehat{\Sigma}$. Its construction for a matrix A involves

1. Find $\widehat{\Sigma}$, U and V from the previous construction of a SVD for A.
2. Delete the last $q - k$ columns from V to form $\widehat{V}$.
3. Delete the last $p - k$ columns from U to form $\widehat{U}$.

Properties

For any $p \times p$ matrix of rank k:

1. The last $q - k$ right singular vectors are an orthonormal basis for $\ker(A)$.
2. The first k left singular vectors are an orthonormal basis for $\mathrm{im}(A)$.
3. $\sum_{i=1}^{k} \sigma_i = \mathrm{tr}(A^*A)$

Pseudoinverse

For A (not necessarily square) with SVD $A = \widehat{U}\widehat{\Sigma}\widehat{V}^*$, the pseudoinverse is given by

$$A^+ = \widehat{V}\widehat{\Sigma}^{-1}\widehat{U}^*$$

The least squares solution to $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = A^+\mathbf{b}$.

Canonical Forms

Canonical forms extend the process of diagonalisation to matrices that cannot be undiagonalised. It reduces a matrix A to a simple matrix that it is similar to (i.e. a matrix B such that $A = P^{-1}BP$).

Generalised Eigenspaces

Definition

Given an eigenvalue λ, the generalised eigenspace of A in the vector space V is defined as
$GE_\lambda(T) = \{\mathbf{x} \in V : (T - \lambda \mathrm{id})^k \mathbf{x} \text{ for some } k \in \mathbb{Z}\}$.

Height of an eigenvalue

In a finite dimensional space V, the **height** of an eigenvalue λ is the least integer such that
$GE_\lambda(T) = \ker (T - \lambda \mathrm{id})^h$.

We use $V_k(\lambda)$ to denote $\ker (T - \lambda \mathrm{id})^k$.

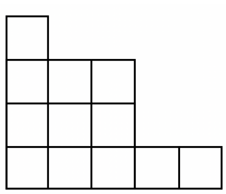
- $V_k(\lambda) = V_{k+1}(\lambda) = \ldots = GE_\lambda(T)$
- $V_1(\lambda) = E_\lambda(T)$
- $0 = \dim V_0(\lambda) < \dim V_1(\lambda) < ... < \dim V_{h-1}(\lambda) < \dim V_h(\lambda) = \dim V_{h+1}(\lambda) = \dim V_{h+2}(\lambda) = \ldots$

We can use $Z_k(\lambda)$ to denote spaces $Z_k(\lambda)$ such that $V_k(\lambda) = V_{k-1}(\lambda) \oplus Z_k(\lambda)$. They are not unique.

- $Z_1(\lambda) \oplus Z_2(\lambda) \oplus \ldots \oplus Z_k(\lambda) = V_k(\lambda)$
- $\dim Z_k(\lambda) = \dim V_k(\lambda) - \dim V_{k-1}(\lambda)$
- $\dim Z_1(\lambda) \geq \dim Z_2(\lambda) \geq \ldots \geq \dim Z_h(\lambda)$. That is, the dimensions of $V_k(\lambda)$ increase at a decreasing rate.

Block diagrams

We can represent the structures of generalised eigenspaces with a Jordan block diagram:



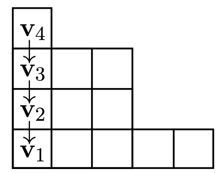
From bottom up, the rows represent the dimensions of $Z_k(\lambda)$ as k increases. The dimensions of $Z_k(\lambda)$ must be non-increasing, since there must be fewer blocks on each row as you go up.

- The entire structure (all the rows) represents the generalised eigenspace.
- The first k rows of the structure (bottom up) represent $V_k(\lambda)$.

In this case, $\dim V_1(\lambda) = 5$, $\dim V_2(\lambda) = 8$, $\dim V_3(\lambda) = 11$, $\dim V_4(\lambda) = 12$.

Step down theorem

If $\mathbf{v} \in Z_k\ \{\mathbf{0}\}$, then $(T - \lambda \mathrm{id})(\mathbf{v}) \in V_{k-1}(\lambda)\ V_{k-2}(\lambda)$. That is, applying $(T - \lambda \mathrm{id})$ to a vector moves it down one level in the block diagram.



Jordan chains

Definition: A Jordan chain of length k is an ordered set of non-zero vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ such that $T(\mathbf{v_1}) = \lambda \mathbf{v}_1$ and $T(\mathbf{v}_i) = \lambda \mathbf{v}_i + \mathbf{v}_{i-1}$.

- One can construct a Jordan chain by taking a vector from the kth level of the block diagram as $\mathbf{v}_k$, and then take $\mathbf{v}_{i-1} = (T - \lambda \mathrm{id})(\mathbf{v}_i)$ for $i = k, k-1, \ldots, 2$.
- The columns of the block diagram represent the maximal (length) Jordan chains. For instance, $\{\mathbf{v}_1, \ldots, \mathbf{v}_4\}$ in the above diagram represents a Jordan chain.

Spaces spanned by Jordan chains

A Jordan chain $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly independent.

- The space spanned by a Jordan chain is **invariant** under T, that is, $T(U) \subseteq U$.
- It is also **indecomposable** under T, which means that it cannot further decomposed into invariant non-trivial subspaces.

Jordan blocks

For $U = \mathrm{span}\{\mathbf{v}_1, \ldots \mathbf{v}_k\}$, the matrix of $T_{|U}$ with respect to the basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ has form

$$\begin{bmatrix} \lambda & 1 & \ldots & 0 & 1 \\ 0 & \lambda & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & \lambda & 1 \\ 0 & 0 & \ldots & 0 & \lambda \end{bmatrix} := J_k(\lambda).$$

A matrix of this form is called a **Jordan block**.

Jordan Forms

Jordan basis

For any linear transformation T on a finite-dimensional vector space V, there exists a basis of the entire space made up of the concatenation of Jordan chains of the eigenvalues of T.

Decomposition lemma

For any linear transformation T on a vector space V of non-zero finite dimension, for some integer $m \geq 1$, there exist T-invariant and indecomposable subspaces $U_1, \ldots, U_m$ such that $V = U_1 \oplus \ldots \oplus U_m$.

- Each of the subspaces is spanned by a maximal length Jordan chain (i.e. a column of one of the Jordan blocks).
- **Theorem**: The number of splitting spaces in the decomposition lemma is equal to the number of independent eigenvectors and the number of maximal Jordan chains.

Jordan matrices and canonical forms

A **Jordan matrix** is a matrix consisting of a direct sum of Jordan blocks.

- **Theorem**: For any linear transformation T on a non-zero, finite-dimensional complex vector space V, there exists a basis of V with respect to which the matrix of T is a Jordan matrix. This is called a **Jordan canonical form**.
- The **numbers** and **lengths** of the maximal Jordan chains are similarity invariants. That is, the Jordan form of a matrix is unique up to reordering of the individual Jordan blocks.
- Matrices are similar if and only if they have the same Jordan form.

Finding a Jordan form and the corresponding Jordan basis

The process is as follows:

- Find the eigenvalues of the matrix A.
- For each eigenvalue λ_i:
  - Calculate the nullities of $(A - \lambda_i I)^k$ for $k = 1, 2, \ldots$ until the nullity no longer increases (which means you've found the height h_i of the eigenvalue λ_i).
  - Construct the Jordan block diagram.
  - Find a vector in $\ker(A - \lambda_i I)^{h_i} \ \ker(A - \lambda_i I)^{(h_i - 1)})$ to begin the first Jordan chain.
  - Form the Jordan chain by successively applying $(A - \lambda_i I)$ to the vector until you reach the bottom of the chain.
  - Do the same for each of the other columns. Note that all the columns must be linearly independent of the others.
- For each Jordan chain, construct the corresponding Jordan matrix.
- Take the direct sum of the Jordan matrices, then concatenate the corresponding Jordan chains **in the same order**.

Miscellaneous Properties

Space decompositions

Miscellaneous properties of Jordan chains

Useful properties of Jordan chains. Many of them are quite intuitive:

- The geometric multiplicity of λ_i is the number of maximal Jordan chains for the eigenvalue.
- The algebraic multiplicity of λ_i is the sum of the sizes of the maximal Jordan chains for the eigenvalue.
- The height h_i of eigenvalue λ_i is the size of the largest Jordan block for λ_i.
- V is the direct sum of its generalised eigenspaces. Each of the generalised eigenspaces is the direct sum of splitting spaces for that eigenvalue.

Cayley-Hamilton theorem

For any matrix $A \in M_{p,p}(\mathbb{C})$, $cp_A(A) = 0$.

- **Corollary**: The space spanned by $\{I, A, A^2, \ldots\}$ is at most p-dimensional.
- The **minimal polynomial** of A is the lowest degree polynomial such that $mp_A(A) = 0$. It is given by
  $$\prod_{i=1}^{m} (t - \lambda_i)^{h_i}.$$
  Note thate $mp_A(A)$ divides $cp_A(A)$.

Functions of Matrices and Systems of ODEs

Powers of Matrices

Matrix expansions

- If $M = A \oplus B$, then $M^n = A \oplus B$.
- if $A = PBP^{-1}$, then $A^n = PB^nP^{-1}$.
  - It follows that if $A = PBP^{-1}$, then $f(A) = Pf(B)P^{-1}$ for any polynomial f.

Binomial theorem for commuting matrices

Let A and B be $p \times p$ matrices for which $AB = BA$. Then for any integer $n \geq 0$,

$$(A+B)^n = A^n + \binom{n}{1}A^{n-1}B + \binom{n}{2}A^{n-1}B^2 + \ldots + B^n.$$

Powers of Jordan blocks

For any Jordan block

$$J_k(\lambda) = \begin{bmatrix} \lambda & 1 & \ldots & 0 & 0 \\ 0 & \lambda & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & \lambda & 1 \\ 0 & 0 & \ldots & 0 & \lambda \end{bmatrix} := J_k(\lambda),$$

we have that $J_k(\lambda) = \lambda I + N$, with

$$N = \begin{bmatrix} 0 & 1 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & \ldots & 0 & 0 \end{bmatrix}.$$

- The powers of N are given by shifting the diagonal up for each subsequent power, e.g.

$$N^2 = \begin{bmatrix} 0 & 0 & 1 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 0 & 1 \\ 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 0 & 0 \end{bmatrix}.$$

- This makes it easy to calculate powers of Jordan matrices, by using the binomial theorem on $(\lambda I + N)^n$.
- Since every matrix can be written as $A = PJP^{-1}$ for some direct sum of Jordan blocks J, this can be extended to calculate powers of any matrix.

Entrywise convergence

A matrix $A^{(k)}$ converges entrywise to A as $k \to \infty$ iff all its entries converge to the corresponding entry of A as $k \to \infty$.

Norms

- The ∞-norm of A is defined as

$$\|A\|_\infty = \max\{|a_{ij}| : 1 \leq i, j \leq p\}.$$

- The operator norm (or 2-norm) of A is defined as

$$\|A\|_{op} = \max\{\|A\mathbf{v}\| : \mathbf{v} \in \mathbb{C}^p \text{ and } \|\mathbf{v}\| = 1\}.$$

- The Frobenius norm of A is defined as

$$\|A\|_F = \sqrt{\operatorname{tr}(A^*A)}.$$

Norm properties

- If a matrix $A \in M_{p,p}(\mathbb{C})$ has non-zero singular values
  $$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k \geq 0,$$
  then
  $$\|A\|_{op} = \sigma_1$$
  and
  $$\|A\|_F = \sqrt{\sigma_1^2 + \ldots + \sigma_k^2}.$$
- For any matrix $A \in M_{p,p}(\mathbb{C})$, $\|A\|_F \geq \|A\|_\infty \geq \frac{1}{p}\|A\|_F$.

Convergence with norms

For any norm N on $M_{p,p}((C))$, and any sequence $A^{(k)} \in M_{p,p}(\mathbb{C})$, $A^{(k)}$ converges to A in the norm N if $N(A^{(k)} - A) \to 0$ as $k \to \infty$.

Properties of convergence

- Convergence in the Frobenius norm, the ∞-norm, the operator norm, and entrywise are equivalent, that is, any one of them implies the other three.
- If $A^{(k)} \in M_{p,p}(\mathbb{C})$ converges to A and P is an invertible $p \times p$ matrix, then
  $$B^{(k)} = P^{-1}A^{(k)}P$$
  converges to $P^{-1}AP$.

Matrix power series

Theorem: If $f : \mathbb{C} \to \mathbb{C}$ has a power series expansion $f(t) = \sum_{k=0}^{\infty} a_k t^k$ with a radius of convergence R, then if $A \in M_{p,p}(\mathbb{C})$, then $\sum_{k=0}^{\infty} a_k A^k$ converges provided that $p||A||_\infty < R$. The limit is $f(A)$.

This is a matrix power series.

Many of the properties of matrix polynomial expansions still apply. For a function $f : \mathbb{C} \to \mathbb{C}$ with a power series expansion:

- If $f(A)$ is defined and $B = PAP^{-1}$, then $f(B) = Pf(B)P^{-1}$.
- If $M = A \oplus B$ and $f(A)$ and $f(B)$ are defined, then $f(M) = f(A) \oplus f(B)$.
- If $f(A)$ is defined and $A\mathbf{v} = \lambda\mathbf{v}$, then $f(A)\mathbf{v} = f(\lambda)\mathbf{v}$.

Matrix exponentials

Definition: $e^A = \exp(A) = I + A + \frac{1}{2!}A^2 + \dots$

Some properties of the matrix exponential:

- For any matrix $A = M_{p,p}(\mathbb{C})$, $\exp(tA)$ is differentiable with respect to t, and

$$\frac{d}{dt}\exp(tA) = A\exp(tA) = \exp(tA)A.$$

- e^{tA} is always invertible and has inverse e^{-tA}.
- If $AB = BA$ for $A, B \in M_{p,p}(\mathbb{C})$, then $\exp(A + B) = \exp(A)\exp(B)$.

Matrix Exponentials and Differential Equations

IVP solution

For $A \in M_{p,p}(\mathbb{C})$, the set of solutions of $\mathbf{y}' = A\mathbf{y}$ is a vector space of dimension p.

The columns of e^{tA} form a basis for this vector space. This means all such solutions can be expressed in the form $e^{tA}\mathbf{c}$ for some $\mathbf{c} \in \mathbb{C}^p$.

It follows that the initial value problem $\mathbf{y}' = A\mathbf{y}$, $\mathbf{y}(\mathbf{0}) = \mathbf{c}$ has a unique solution $\mathbf{y} = e^{tA}\mathbf{c}$.

The column method

The **column method** provides a simple way to evaluate $e^{tA}\mathbf{c}$ without actually calculating e^{tA}.

- Use the basis of generalised eigenvectors to write $\mathbf{c} = \sum_{k=1}^{t} a_k\mathbf{v}_k$ for generalised eigenvectors $\mathbf{v}_k$.
- Write

$$\begin{aligned} e^{tA}\mathbf{c} &= e^{tA}\left(\sum_{k=1}^{t} a_k\mathbf{v}_k\right) \\ &= \sum_{k=1}^{t} a_k e^{tA}\mathbf{v}_k. \end{aligned}$$

- Say that the generalised eigenvector $\mathbf{v}_k$ has corresponding eigenvalue λ_k. Then write

$$e^{tA}\mathbf{c} = \sum_{k=1}^{t} a_k e^{\lambda_k t} e^{t(A-\lambda_k I)}\mathbf{v}_k.$$

- Simplify the exponentials by using the fact that

$$e^{t(A-\lambda I)}\mathbf{v} = \mathbf{v} + t(A - \lambda I)\mathbf{v} + \frac{1}{2!}t^2(A - \lambda I)^2 + \dots$$

  - Find the first power n_k such that $(A - \lambda_k I)^{n_k}\mathbf{v} = \mathbf{0}$.
  - Then you can reduce the infinite sum in the exponential term to a finite sum, that is,

$$\begin{aligned} e^{(t-\lambda_k I)}\mathbf{v}_k = \mathbf{v}_k + t(A - \lambda_k I)\mathbf{v}_k + \\ \frac{1}{2!}t^2(A - \lambda_k I)^2 + \dots + \\ \frac{1}{(n_k - 1)!}t^{n_k-1}(A - \lambda_k I)^{n_k-1}. \end{aligned}$$

Fundamental matrices

A **fundamental matrix** is any square matrix whose columns are independent and each solutions of $\mathbf{y}' = A\mathbf{y}$. e^{tA} is one such example.

For any basis $\{\mathbf{v}_1, \mathbf{v}_2, \dots \mathbf{v}_n\}$ of generalised eigenvectors,

$$\Phi(t) = \begin{bmatrix}\exp(tA)\mathbf{v}_1 & \dots & \exp(tA)\mathbf{v}_n\end{bmatrix}$$

is a fundamental matrix.

Inhomogeneous systems of ODEs

Take the inhomogeneous ODE $\mathbf{y}' = A\mathbf{y} + \mathbf{b}(t)$. If you take any fundamental matrix $\Phi(t)$ for the homogeneous equation, then the solution is $\mathbf{y}(t) = \Phi(t)\mathbf{c} + \Phi(t)\int_0^t \Phi^{-1}(s)\mathbf{b}(s)ds$.

This is most commonly done with $\Phi(t) = e^{At}$.

Recall that the general solution to any inhomogeneous equation is $\mathbf{y} = \mathbf{y}_h + \mathbf{y}_p$, where $\mathbf{y}_h$ is the general solution to the corresponding homogeneous equation, and $\mathbf{y}_p$ is any solution to the inhomogeneous equation.

- Here, $\Phi(t)\mathbf{c} = \mathbf{y}_h$.
- $\Phi(t)\int_0^t \Phi^{-1}(s)\mathbf{b}(s)ds = \mathbf{y}_p$.