Notes on algebra

John Alan McDonald (palisades dot lakes at gmail dot com)

draft of 2021-03-19

Contents

1 Mathematical Structures

A mathematical structure [9] consists of a few sets, and functions between those sets. Types of structure:

Algebraic Characterized by operations (functions of 2 arguments, sometimes more) that have given properties, like commutativity: $(+\,a\,b) = (+\,b\,a)$.

Order and equivalence A set plus a one or more relations. See **??**

Topology A primary set plus a family of 'open' subsets of the primary set. The subsets have to obey certain properties which enable to be used to define what is 'connected' to what in the primary set.

Metric A set plus a *distance* function, mapping each pairs of elements to non-negative real number, having certain properties. Metric/distance induces a topology:

$$\mathcal{B}(x_0, r) = \left\{x \in \mathcal{X} \mid \text{distance}\,(x_0, x) < r\right\}$$

See section 3.2.

Measure A set plus a function that assigns a non-negative real value to some subsets of the primary set. Examples are length, area, volume.

Geometry **TODO:**

For $\mathbb{R}$: Order and Metric induce Topology. Order and Algebraic structure lead to ordered field. Algebraic structure and topology make Lie group.

There is usually a primary set, whose elements are the 'elements' of the structure:

Example 1.1: Linear Space

A *linear space* $\mathbb{V}$ is:

- a set of vectors $\mathcal{V}$,
- a field of scalars $\mathbb{F}$,
- a linear combination operation/function:

$$(\text{linear-combination } a_0\, v_0\, a_1\, v_1) \;=\; a_0 * v_0 + a_1 * v_1 \;\rightarrow\; v_2 \in \mathcal{V} \tag{1.1}$$

for $v_0, v_1 \in \mathcal{V}$ and $a_0, a_1 \in \mathbb{F}$. Linear combination is often defined in terms of 2 other operations: scalar multiplication $a * v \in \mathcal{V}$, and vector addition $v_0 + v_1 \in \mathcal{V}$

Usually the distinction between $\mathcal{V}$ and $\mathbb{V} = [\mathcal{V}, \mathbb{F}, \text{linear-combination}]$ is ignored; I prefer to define in terms of linear-combination because it naturally extends to affineCombination and **convex-combination**.

It's tempting to identify a mathematical structure with a class, but that won't work, because you will need to support multiple representations (dense and sparse vectors, `double` and `BigFraction`) and the functions operate on multiple sets (vectors and scalars) each with multiple representations. Interfaces also don't work, because operations are typically functions of more than one argument, with multiple representations for each.

Implementation is easier with *generic functions* (aka "multimethods"). Or context object determining what $(+\,a\,b)$ means for any acceptable implementation of a and b.

2 Algebraic structures

TODO: redo as Universal Algebras? Implies more operations, typically binary, unary, and nullary (nullary function equivalent to element of set).

An algebraic structure [7, 9, 10, 12–14] consists of:

- A primary set — the ***elements*** of the structure.
- zero or a few auxilliary sets.
- functions (called ***operations***) that take a small number of arguments from one or more of the sets and return elements of the sets.

The type of an algebraic structure corresponds to identities that the operations satisfy. (**TODO:** Examples of operation identities?)

Unfortunately, the names for algebraic structures are, as a rule, not very informative.

2.1 One set, one operation

2.1.1 Monoid

Definition 2.1: Monoid

Set $\mathcal{S}$ and operation $\diamond$ such that, if $a, b, c \in \mathcal{S}$, then
Closed $a \diamond b = \diamond\,(a, b) = (\diamond\ a\ b) \in \mathcal{S}$
Associative $a \diamond b \diamond c = (a \diamond b) \diamond c = a \diamond (b \diamond c)$
Identity There is an $i \in \mathcal{S}$ such that $i \diamond a = a\ \forall a \in \mathcal{S}$.

Exercise 2.1: Monoid identity is unique.

Show that i is unique.

Example 2.2: Function composition monoid

$\mathcal{S}$ the functions from some domain $\mathcal{X}$ to itself. Operation $\diamond$ is function composition: $\left(f \diamond g\right)(x) = f\left(g\,(x)\right)$.

2.1.2 Group

Monoid $[\mathcal{S}, \diamond]$ such that

Inverse For every $a \in \mathcal{S}$ there exists an a^{-1} such that $a^{-1} \diamond a = i$. Exercise: show that this implies that $a \diamond a^{-1} = i$.

2.1.3 Commutative group

(aka abelian group.)

A group $[\mathcal{S}, \diamond]$ where

Commutative $a \diamond b = b \diamond a\ \forall a, b \in \mathcal{S}$.

Example: $[\mathbb{Z}, *_{\mathbb{Z}}]$

2.2 One set, two operations

2.2.1 Semiring

A set and 2 operations: $[\mathcal{S}, +, *]$ where

Addition $+$ is commutative, associative, and has an identity element (0): $a + b = b + a$; $a + (b + c) = (a + b) + c$; and $0 + a = a$, for all $a, b, c \in \mathcal{S}$.

Multiplication $*$ is associative, and has an identity element (1): $a * (b * c) = (a * b) * c$ and $1 * a = a$, for all $a, b, c \in \mathcal{S}$.]

Distributive $a * (b + c) = (a * b) + (a * c)$

Example:

The ***natural numbers***: $\mathbb{N} = \{i \in \mathbb{Z} \mid i \geq 0\}$. (sec **??**).

2.2.2 Ring

[11]

A set and 2 operations: $[\mathcal{S}, +, *]$ where

Addition group $[\mathcal{S}, +]$ is a commutative group (with the identity written 0).

Multiplication monoid $[\mathcal{S}, *]$ is a monoid (with the identity written 1). Note this means $*$ may not commute. If it does, then this is a ***commutative ring***.

Distributive $a * (b + c) = (a * b) + (a * c)$

Example:

The ***integers***: $\mathbb{Z} = \{\cdots, -2, -1, 0, 1, 2, \cdots\}$.

2.2.3 Field

[8]

A ring $[\mathcal{S}, +, *]$ where $*$ is commutative and the nonzero elements of $\mathcal{S}$ have multiplicative inverses.

Without refering to other structures: A ***field*** is set and 2 operations: $[\mathcal{S}, +, *]$ where

Additive closure $a + b \in \mathcal{S}\ \forall\, a, b \in \mathcal{S}$

Additive associativity $(a + b) + c = a + (b + c)$

Additive commutativity $a + b = b + a$

Additive identity $\exists\, 0 \in \mathcal{S}$ s.t. $a + 0 = 0 + a = a\ \forall\, a \in \mathcal{S}$

Additive inverse $\forall a \in \mathcal{S}\ \exists -a \in \mathcal{S}$ s.t. $a + (-a) = (-a) + a = 0$

Multiplicative closure $a * b \in \mathcal{S}\ \forall\, a, b \in \mathcal{S}$

Multiplicative associativity $(a * b) * c = a * (b * c)$

Multiplicative commutativity $a * b = b * a$

Multiplicative identity $\exists 1 \in \mathcal{S}$ s.t. $a * 1 = 1 * a = a\ \forall\, a \in \mathcal{S}$

Multiplicative inverse $\forall a \neq 0 \in \mathcal{S} \exists a^{-1} \in \mathcal{S}$ s.t. $a * a^{-1} = a^{-1} * a = 1$ (Note the restriction to elements other than the additive identity.)

Distributive $a * (b + c) = (a * b) + (a * c)$

Examples: $\mathbb{Q}$, $\mathbb{R}$.

2.3 Two sets, one operation

2.3.1 Category

A base set, $\mathcal{O}$, whose elements are ususally refered to as ***objects***. A second set, $\mathcal{M}$, of ***morphisms***, each of which depends on an ordered pair $\big(\text{source}, \text{target}\big)$ of elements of $\mathcal{O}$. A ***composition*** operation, $\circ$, on the subset of the pairs of morphisms $f, g \in \mathcal{M}$ where $\text{source}(f) = \text{target}(g)$, satisfying:

Associativity: $f \circ g \circ g \overset{\text{def}}{=} \big(f \circ g\big) \circ h = f \circ \big(g \circ h\big)$

Identity: For each $x \in \mathcal{O}$, there exists an element $1_x \in \mathcal{M}$ such that $x = \text{source}(1_x) = \text{target}(1_x)$, and $1_x \circ f = f$ (when $x = \text{target}(f)$) and $f \circ 1_x = f$ (when $x = \text{source}(f)$).

Example 2.3: The Set category

$\mathcal{O}$ = some collection of sets; $\mathcal{M}$ = the functions between those sets; $\circ$ = function composition.

3 Spaces

3.1 Topological spaces

A generalization of Spivak (1965) *Calculus on Manifolds*, chapter 1.

Open sets, ***neighborhoods***.

Finite intersection of open sets is open. Arbitrary union is is open.

Interior: elements in set with a neighborhood contained in the set.

Exterior: elements not in set with a neighborhood not intersecting the set.

Boundary: elements where every neighborhood intersects both the interior and the exterior.

Open cover.

Compact set has finite open cover.

A set is ***closed*** if its complement is open.

Connectivity: set is connected if no disjoint open sets contain whole set.

Limits.

Continuity of functions: (See Spivak (1965) *Calculus on Manifolds*, Theorem 1-8.) $f : \mathcal{D} \mapsto \mathcal{C}$, for topological spaces $\mathcal{D}$ and $\mathcal{C}$, is ***continuous*** if for any open set $\mathcal{O}_{\mathcal{C}} \subset \mathcal{C}$, $f^{-1}(\mathcal{O}_{\mathcal{C}}) = \mathcal{O}_{\mathcal{D}}$, an open set $\subset \mathcal{D}$. (**TODO:** Is this assuming domain and codomain are open? Is this generalization of Spivak 1.8 really correct?)

Example: open intervals in $\mathbb{R}$, open balls in $\mathbb{R}^n$ with l_1, l_2, l_∞ distances (what is required to work?). Figures for open balls with various metrics.

3.2 Metric spaces

Open sets generated by distance function.

3.3 Linear spaces

Quote 3.1: MacLane (1954) "Of course and courses"

> Throughout these courses the infusion of a geometrical point of view is of paramount importance. A vector is geometrical; it is an element of a vector space, defined by suitable axioms—whether the scalars be real numbers or elements of a general field. A vector is not an n-tuple of numbers until a coordinate system has been chosen. Any teacher and any text book which starts with the idea that vectors are n-tuples is committing a crime for which the proper punishment is ridicule. The n-tuple idea is not 'easier,' it is harder; it is not clearer, it is more misleading. By the same token, linear transformations are basic and matrices are their representations...

My approach to linear (aka vector) spaces is largely based on the texts I used as a college freshman for linear algebra and multivariate calculus: Halmos Halmos (1958) *Finite-dimensional Vector Spaces* and Spivak Spivak (1965) *Calculus on Manifolds*.

Definition 3.1: Linear space

A ***linear space*** $\mathbb{V} = [\mathcal{V}, \mathbb{K}, \text{linear-combination}]$ is:

- a set of ***vectors*** $\mathcal{V}$,
- a field of scalars $\mathbb{K}$,

- a linear combination function:

$$(\text{linear-combination } a_0\, v_0\, a_1\, v_1) \;\rightarrow\; v_2 \in \mathcal{V} \tag{3.1}$$

for $v_0, v_1 \in \mathcal{V}$ and $a_0, a_1 \in \mathbb{K}$. Linear combination is often defined in terms of 2 binary operations: scalar multiplication $a * v \in \mathcal{V}$, and vector addition $v_0 + v_1 \in \mathcal{V}$:

$$(\text{linear-combination } a_0\, v_0\, a_1\, v_1) \;=\; a_0 * v_0 + a_1 * v_1 \tag{3.2}$$

TODO: required identities for $+$ and $*$ from Spivak or Halmos.

Usually the distinction between $\mathcal{V}$ and $\mathbb{V}$ is ignored, and we will say, for example, $v \in \mathbb{V}$.

Definition 3.2: Linear dependence

Suppose $\mathbb{V}$ is a linear space and $v_0 \dots v_{n-1} \in \mathbb{V}$. If there exists $a_0 \dots a_{n-1}$ such that $0 = \sum a_i v_i$ then the $\{v_i\}$ are *linearly dependent*. Halmos (1958) *Finite-dimensional Vector Spaces*, section 5
Otherwise they are *linearly independent*.

Lemma 3.3: The set of non-zero vectors $v_0 \dots v_{n-1}$ is linearly dependent iff some v_k, $1 \leq n-1$, is a linear combination of the preceding ones Halmos (1958) *Finite-dimensional Vector Spaces*, Section 6.

Proof: Assume $v_0 \dots v_{n-1}$ are linearly dependent. Consider the smallest k such that $v_0 \dots v_k$ is linearly dependent. By definition, there exists non-zero $a_0 \dots a_k$ such that

$$0 = \sum_0^k a_i v_i \tag{3.3}$$

which implies that

$$v_k = \sum_0^{k-1} \frac{a_i}{-a_k} v_i \tag{3.4}$$

Theorem 3.4: The set of non-zero vectors $v_0 \dots v_{n-1}$ is linearly dependent iff some v_k, $1 \leq n-1$, is a linear combination of the preceding ones Halmos (1958) *Finite-dimensional Vector Spaces*, Section 6.

Proof: Assume $v_0 \dots v_{n-1}$ are linearly dependent. Consider the smallest k such that $v_0 \dots v_k$ is linearly dependent. By definition, there exists non-zero $a_0 \dots a_k$ such that

$$0 = \sum_0^k a_i v_i \tag{3.5}$$

which implies that

$$v_k = \sum_0^{k-1} \frac{a_i}{-a_k} v_i \tag{3.6}$$

Corollary 3.5: The set of non-zero vectors $v_0 \dots v_{n-1}$ is linearly dependent iff some v_k, $1 \leq n-1$, is a linear combination of the preceding ones Halmos (1958) *Finite-dimensional Vector Spaces*, Section 6.

Proof: Assume $v_0 \dots v_{n-1}$ are linearly dependent. Consider the smallest k such that $v_0 \dots v_k$ is linearly dependent. By definition, there exists non-zero $a_0 \dots a_k$ such that

$$0 = \sum_0^k a_i v_i \tag{3.7}$$

which implies that

$$v_k = \sum_0^{k-1} \frac{a_i}{-a_k} v_i \tag{3.8}$$

Examples:

Example 3.6: $\mathbb{K}^n$

Where $\mathbb{K}$ is any field.

- vectors: $\mathcal{V} = \mathbb{K}^n = \{x = [x_0 \dots x_{n-1}]\}$, tuples of n elements $x_i \in \mathbb{K}$.
- scalars: $\mathbb{K}$
- scalar multiplication: $a *_{\mathbb{K}^n} [x_0 \dots x_{n-1}] = [\dots, a *_{\mathbb{K}} x_i, \dots]$
- vector addition: $x +_{\mathbb{K}^n} y = [\dots, (x_i +_{\mathbb{K}} y_i), \dots]$

Example 3.7: $\mathbb{R}^n$

See Spivak (1965) *Calculus on Manifolds*, chapter 1 and Halmos (1958) *Finite-dimensional Vector Spaces*, chapter 1.

- vectors: $\mathcal{V} = \mathbb{R}^n = \{x = [x_0 \dots x_{n-1}]\}$, tuples of n elements $x_i \in \mathbb{R}$.
- scalars: $\mathbb{R}$
- scalar multiplication: $a * [x_0 \dots x_{n-1}] = [\dots, (a * x_i), \dots]$
- vector addition: $x + y = [\dots, (x_i + y_i), \dots]$

TODO: DANGER: Apple-orange mistakes resulting from using $\mathbb{R}^n$ in problems where coordinates don't mean the same thing, so canonical inner product and l2 distance aren't correct.

Homogeneous problems often don't have meaningful coordinates.

Can only approximate $\mathbb{R}^n$ with tuples of `double`, which is fundamentally different due to lack of associativity, which leads to accumulation of rounding error.

TODO: Exact float arithmetic as an alternative?

Example 3.8: $\mathbb{Q}^n$

Like $\mathbb{R}^n$, only over rational rather than real numbers. Has the advantage that it can be implemented accurately using arbitrary precision fractions, though at considerable space-time cost.
TODO: measure cost compared to `double` approximation to $\mathbb{R}^n$

Example 3.9: The functions from any domain to some linear space.

The functions from any domain to some linear space. **TODO:** lisp notation for clarity below.

- vectors: $\mathcal{F}$ = any function on $\mathcal{D}$ that returns values in the linear space $\mathbb{V}$.
- scalars: $\mathbb{K}$, the same scalar field used by $\mathbb{V} = [\mathcal{V}, \mathbb{K}, +, *]$.
- scalar multiplication: $(a * f) : \mathcal{D} \to \mathbb{V}$ is the function defined by $(a * f)(x) = a * (f(x))$
- vector addition: $(f + g)$ is the function defined by $(f + g)(x) = f(x) + g(x)$

Canonical coordinates: $f(d) \ \ \forall d \in \mathcal{D}$

Example 3.10: $\mathbb{R}^n$ as a function space

We can identify Halmos (1958) *Finite-dimensional Vector Spaces*, sec. 24 $\mathbb{R}^n$ and $\mathbb{F}[\{0, 1, \dots, n-1\}, \mathbb{R}]$ by $x \Leftrightarrow f_x$ where $f_x(i) = x_i$ for $x \in \mathbb{R}^n$ and $i \in \{0, 1, \dots, n-1\}$

Definition 3.11: Linear dependence

Suppose $\mathbb{V}$ is a linear space and $v_0 \dots v_{n-1} \in \mathbb{V}$. If there exists $a_0 \dots a_{n-1}$ such that $0 = \sum a_i v_i$ then the $\{v_i\}$ are ***linearly dependent***. Halmos (1958) *Finite-dimensional Vector Spaces*, section 5
Otherwise they are ***linearly independent***.

Theorem 3.12: The set of non-zero vectors $v_0 \dots v_{n-1}$ is linearly dependent iff some v_k, $1 \leq n - 1$, is a linear combination of the preceding ones Halmos (1958) *Finite-dimensional Vector Spaces*, Section 6.

Proof: Assume $v_0 \dots v_{n-1}$ are linearly dependent. Consider the smallest k such that $v_0 \dots v_k$ is linearly dependent. By definition, there exists non-zero $a_0 \dots a_k$ such that

$$0 = \sum_0^k a_i v_i \tag{3.9}$$

which implies that

$$v_k = \sum_0^{k-1} \frac{a_i}{-a_k} v_i \tag{3.10}$$

3.3.1 Bases

3.3.2 Dimension

3.3.3 Subspaces

A subset which is also a linear space with the same scalars and operations.

Example 3.13: Canonical subspaces of a function space

Let $\mathcal{D}_0 \subset \mathcal{D}_1$ be a proper subset. Consider $\mathbb{F}_0 = \mathbb{F}\left[\mathcal{D}_0, \mathbb{V}\right]$ and $\mathbb{F}_1 = \mathbb{F}\left[\mathcal{D}_1, \mathbb{V}\right]$.

Example 3.14: Canonical subspaces of $\mathbb{R}^n$

Finite index set of integers, a real value for each. Relationships between index sets define super/sub space relations intersections.

3.3.4 Normed linear spaces

3.3.5 Inner product (linear) spaces

Let $\mathbb{V}$ be an n-dimensional real inner product space. Let $v, w \in \mathbb{V}$.

- The inner (dot) product on $\mathbb{R}^n$:

$$v \bullet w \;\equiv\; \sum_{i=0}^{n-1} v_i w_i \tag{3.11}$$

- The euclidean (l_2) norm:

$$\|v\|^2 \;\equiv\; v \bullet v \tag{3.12}$$

- $\theta(v, w)$ is the angle between v and w and is defined by:

$$\begin{aligned} v \bullet w \;&=\; \|v\|\|w\| \cos(\theta(v, w)) \\ \theta(v, w) \;&\equiv\; \cos^{-1}\left(\frac{v \bullet w}{\|v\|\|w\|}\right) \end{aligned} \tag{3.13}$$

- The tensor (outer) product:

  Let $v, u \in \mathbb{V}$, $w \in \mathbb{W}$. $w \otimes v$ is a rank 1 linear map from $\mathbb{V}$ to $\mathbb{W}$, defined by:

$$(w \otimes v)(u) \;\equiv\; w(v \bullet u) \tag{3.14}$$

  Note: this is an abuse of the usual definition of tensor product $\otimes$. This operation, which takes a pair of vectors and returns a linear map, is more conventionally referred to as the 'outer product', and written $wv^\dagger$. However, because I am working in spaces other than $\mathbb{R}^n$ (eg. $\mathbb{L}(\mathbb{V}, \mathbb{W})$, the space of linear maps between 2 vector spaces), I want to avoid notations that suggest thinking in terms of 'row' and 'column' vectors.

  The following is a useful identity. If $t \in \mathbb{T}$, $u, v \in \mathbb{V}$, and $w \in \mathbb{W}$. then

$$(t \otimes u)(v \otimes w)(u) = (u \bullet v)(t \otimes w) \tag{3.15}$$

- Elementary orthogonal projection:

$$\Pi_w v \;\equiv\; \left(\frac{w}{\|w\|} \otimes \frac{w}{\|w\|}\right) v \;=\; \left(\frac{w}{\|w\|} \bullet v\right) \frac{w}{\|w\|} \tag{3.16}$$

- Orthogonal complement:

$$\perp_w v \;\equiv\; v \perp w \;\equiv\; v \;-\; \Pi_w v \;=\; v \;-\; \left(\frac{w}{\|w\|} \bullet v\right) \frac{w}{\|w\|} \tag{3.17}$$

3.4 Affine spaces

3.4.1 Euclidean space

3.5 Projective spaces

3.5.1 Oriented projective spaces

Stolfi (1991) *Oriented projective geometry : a framework for geometric computations*

3.6 Barycentric (convex) spaces and functions

3.7 Spherical spaces

3.8 Manifolds

3.9 Functions between spaces

In general, the functions discussed here map between real inner product spaces: $f : \mathbb{V} \mapsto \mathbb{W}$, where $\mathbb{V}$ is the ***domain*** and $\mathbb{W}$ is the ***codomain***. The real inner product spaces are almost derived from some $\mathbb{R}^n$.

The ***range*** of f, $\mathrm{range}(f)$, is the set $f(\mathbb{V})$, which may be a proper subset of its codomain $\mathbb{W}$. The ***kernel*** of f, $\mathrm{kernel}(f)$, is the set $\mathrm{kernel}(f) = \{v \in \mathbb{V} : f(v) = 0\}$.

When I want to distinguish between real- and vector-valued functions, I may use 'function' for vector-valued functions and 'functional' for real-valued ones.

I use $\mathbb{U}, \mathbb{V}, \mathbb{W}$ for generic linear spaces, u, v, w, etc., for elements of linear spaces, usually called ***vectors*** and f, g, h for vector-valued functions. I generally do not distinguish $\mathbb{R}$, the real numbers, and $\mathbb{R}^1$, or any other 1-dimensional real linear space. I sometimes use f, g, h for extra clarity in the special case of real-valued functions.

The domains of many interesting functions, such as those that depend on vertex positions, are direct sum of inner product spaces. The ***direct sum*** $\mathbb{V} \oplus \mathbb{W}$ is the inner product space consisting of the ordered pairs $\{(v, w) : v \in \mathbb{V}, w \in \mathbb{W}\}$ inheriting the inner product space operations in the obvious way: $(v_0, w_0) \bullet (v_1, w_1) = (v_0 \bullet v_1) + (w_0 \bullet w_1)$. I will usually write an element of $\bigoplus^n \mathbb{V}$ as $(v_0, \dots, v_{n-1})$ and use $f(v_0, v_1, \dots, v_{n-1})$ for a function that depends on n vectors.

3.9.1 Linear functions

A function $L(v) : \mathbb{V} \mapsto \mathbb{W}$ is ***linear*** iff $L(a_0 v_0 + a_1 v_1) = a_0 L(v_0) + a_1 L(v_1)$. I will often write $Lv \equiv L(v)$.

It is not hard to see that, for a linear function, the range and kernel are linear subspaces of the codomain and domain, respectively. Thus any linear function between inner product spaces divides its domain and codomain each into 2 orthogonal subspaces. The domain is divided into $\mathbb{V} = \mathrm{kernel}(L) \oplus \mathrm{kernel}^{\perp}(L)$, and the codomain is divided into $\mathbb{W} = \mathrm{range}(L) \oplus \mathrm{range}^{\perp}(L)$.

The most common representation for linear functions is the ***matrix:*** Let $L(v) : \mathbb{V} \mapsto \mathbb{W}$ be linear, $\{e_0^{\mathbb{V}} \dots e_{m-1}^{\mathbb{V}}\}$ an orthonormal basis for $\mathbb{V}$, and $\{e_0^{\mathbb{W}} \dots e_{n-1}^{\mathbb{W}}\}$ an orthonormal basis for $\mathbb{W}$ Then L can be expressed as

$$L = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} L_{ij} (e_i^{\mathbb{W}} \otimes e_j^{\mathbb{V}}) \tag{3.18}$$

(L_{ij}) is the matrix representation of L with respect to the two bases Halmos (1958) *Finite-dimensional Vector Spaces*.

It is important to note that there are many useful representations for linear functions other than matrices McDonald (1989) "Object-oriented programming for linear algebra". Sometimes other representations are used for convenience, or to enforce some constraint like symmetry. In some cases, a non-matrix representation must be used, because a particular linear transformation cannot be accurately represented by a matrix of floating point numbers.

Examples:

- Column-wise: $L = \sum_{j=0}^{n-1} (c_j^L \otimes e_j^{\mathbb{V}})$

  $c_j^L \in \mathbb{W}$ are the 'columns' of L. $\mathrm{span}\,\{c_0^L \dots c_{n-1}^L\} = \mathrm{range}(L)$ (see section 3.9.3).
- Row-wise: $L = \sum_{i=0}^{m-1} (e_i^{\mathbb{W}} \otimes r_i^L)$

  $r_i^L \in \mathbb{V}$ are the 'rows' of L. $\mathrm{span}\,\{r_0^L \dots r_{m-1}^L\} = \mathrm{kernel}(L)^{\perp}$ (see section 3.9.3).
- Householder: $h_v = I_{\mathbb{V}} - \frac{2}{\|v\|^2}(v \otimes v)$

  Householder maps are usually chosen to zero the elements of a vector, or a row or column of a matrix, for a contiguous range of indices, say, $[i_0, \dots, i_n)$.

3.9.2 Affine functions

A function $\mathbf{A}(v) : \mathbb{V} \mapsto \mathbb{W}$ is ***affine*** if distributes over affine combinations: $\mathbf{A}(\sum_{i=0}^{n-1} a_i v_i) = \sum_{i=0}^{n-1} a_i \mathbf{A}(v_i)$ for all $\{a_i\}$ such that $1 = \sum_{i=0}^{n-1} a_i$. (Note that I am describing affine functions on vector (linear) spaces, rather than the slightly more general notion of affine functions on affine spaces.) Any linear function between linear spaces is automatically affine. The other major class of affine functions on linear spaces are the translations. A ***translation,*** $\mathbf{T}_t$, $\mathbb{V} \mapsto \mathbb{V}$, simply adds a vector (t) to its argument: $\mathbf{T}_t v = v + t$. It's not hard to see that any affine function between two linear spaces can be represented as the sum of a linear function and a translation. A typical representation for a general affine function $\mathbf{A} : \mathbb{V} \mapsto \mathbb{W}$ is as a pair (L, t) where $\mathsf{L} : \mathbb{V} \mapsto \mathbb{W}$ is linear, $\mathsf{t} \in \mathbb{W}$, and $\mathbf{A}(\mathsf{v}) = \mathsf{L}(\mathsf{v}) + \mathsf{t}$.

3.9.3 Spans and projections

Let $\mathbb{V}$ be an n-dimensional inner product space.

The ***linear span*** of a set of m vectors in $\mathbb{V}$ is the set of linear combinations of those vectors:

$$\mathsf{span}\left\{\mathsf{v}_0 \ldots \mathsf{v}_{m-1}\right\} = \left\{\mathsf{v} \in \mathbb{V} : \mathsf{v} = \sum_{i=0}^{m-1} a_i \mathsf{v}_i\right\} \tag{3.19}$$

$\mathsf{span}\left\{\mathsf{v}_0 \ldots \mathsf{v}_{m-1}\right\}$ is a linear subspace of $\mathbb{V}$.

The ***projection*** $\Pi_{\mathcal{S}}\mathsf{v}$ of a vector $\mathsf{v} \in \mathbb{V}$ onto an arbitrary subset $\mathcal{S} \subset \mathbb{V}$ is the closest point in $\mathcal{S}$ to v. Projection onto a linear subspace is a linear function and can be computed by summing elementary orthogonal projections onto an orthonormal basis for the subspace.

An orthonormal basis for $\mathsf{span}\left\{\mathsf{v}_0 \ldots \mathsf{v}_{m-1}\right\}$ (and $\mathsf{span}\left\{\mathsf{v}_0 \ldots \mathsf{v}_{m-1}\right\}^{\perp}$) can be computed using the QR decomposition of the function $\mathsf{V} = \sum_{i=0}^{m-1} \mathsf{v}_i \otimes \mathsf{e}_i$, (the $n \times m$ matrix whose columns are the v_i) (see Golub and Van Loan (2012) *Matrix computations*, sec. 5.2).

The ***affine span*** of a set of $m+1$ vectors in $\mathbb{V}$ is the set of affine combinations of those vectors:

$$\mathbf{aspan}\left\{\mathbf{p}_0 \ldots \mathbf{p}_m\right\} = \left\{\mathsf{v} \in \mathbb{V} : \mathsf{v} = \sum_{i=0}^{m} b_i \mathbf{p}_i ; 1 = \sum_{i=0}^{m} b_i\right\}. \tag{3.20}$$

$\mathbf{aspan}\left\{\mathbf{p}_0 \ldots \mathbf{p}_m\right\}$ is an affine subspace of $\mathbb{V}$. $\mathsf{b} = (b_0 \ldots b_m)$ are ***barycentric coordinates*** for v with respect to $\left\{\mathbf{p}_0 \ldots \mathbf{p}_m\right\}$. The barycentric coordinates are unique if $\left\{\mathbf{p}_0 \ldots \mathbf{p}_m\right\}$ are affinely independent.

Any affine subspace, $\mathbb{A}$, of a linear space, $\mathbb{V}$ can be represented as as a translation of a linear subspace of $\mathbb{V}$: $\mathbb{A} = \mathbb{T}(\mathbb{A}) + \mathsf{t}$, $\mathbb{T}(\mathbb{A})$ is the set of differences of elements of $\mathbb{A}$, a linear subspace of $\mathbb{V}$. If t is any element of $\mathbb{A}$. then projection onto $\mathbb{A}$ can be computed as a translation of an orthogonal projection onto $\mathbb{T}(\mathbb{A})$: $\Pi_{\mathbb{A}}(\mathbf{p}) = \mathsf{t} + \Pi_{\mathbb{T}(\mathbb{A})}(\mathbf{p} - \mathsf{t})$. Typically, we pick t to be the smallest element of $\mathbb{A}$. Projection onto an affine space is clearly an affine function.

We can represent the affine span of a set of $m + 1$ vectors as a translation of a linear span:

$$\mathbf{aspan}\left\{\mathbf{p}_0 \ldots \mathbf{p}_m\right\} = \mathbf{p}_m + \mathsf{span}\left\{\mathsf{v}_0 \ldots \mathsf{v}_{m-1}\right\} \tag{3.21}$$

where $\mathsf{v}_i = \mathbf{p}_i - \mathbf{p}_m$, which allows us to compute the projection onto $\mathbf{aspan}\left\{\mathbf{p}_0 \ldots \mathbf{p}_m\right\}$ again using the QR decomposition of $\mathsf{V} = \sum_{i=0}^{m-1} \mathsf{v}_i \otimes \mathsf{e}_i$.

3.9.4 Linear inverses and pseudo-inverses

A convenient definition for the ***true inverse*** of a function $\mathsf{f}(\mathsf{v}) : \mathbb{V} \mapsto \mathbb{W}$ is $\mathsf{f}^{-1}(\mathsf{w}) = \{\mathsf{v} : \mathsf{f}(\mathsf{v}) = \mathsf{w}\}$. The usual definition of inverse treats f^{-1} as a function from $\mathbb{W} \mapsto \mathbb{V}$, which is undefined where the value of the true inverse is not a set containing a single point.

For functions between inner product spaces, the ***pseudo-inverse***, f^{-}, is a function $\mathbb{W} \mapsto \mathbb{V}$ defined everywhere on $\mathbb{W}$. Let $\hat{\mathsf{w}}$ be an element of $\mathbb{W}$ closest to w such that $\mathsf{f}^{-1}(\mathsf{w})$ is not empty. Let $\hat{\mathsf{v}}$ be a minimum norm element of $\mathsf{f}^{-1}(\hat{\mathsf{w}})$. Then $\mathsf{f}^{-}(\mathsf{w}) = \hat{\mathsf{v}}$.

If L is linear, then it's not hard to see that $\hat{\mathsf{w}} = \pi_{\mathsf{range}(\mathsf{L})}\mathsf{w}$, the projection of w on the range of L and $\hat{\mathsf{v}}$ is the unique element of $\mathsf{kernel}^{\perp}(\mathsf{L})$ such that $\mathsf{L}(\hat{\mathsf{v}}) = \hat{\mathsf{w}}$.

The pseudo-inverse of a linear function can be characterized by the four Moore-Penrose conditions Golub and Van Loan (2012) *Matrix computations*, sec. 5.5.2:

1. $\mathsf{L}\mathsf{L}^{-}\mathsf{L} = \mathsf{L}$
2. $\mathsf{L}^{-}\mathsf{L}\mathsf{L}^{-} = \mathsf{L}^{-}$
3. $(\mathsf{L}\mathsf{L}^{-})^{\dagger} = \mathsf{L}\mathsf{L}^{-}$
4. $(\mathsf{L}^{-}\mathsf{L})^{\dagger} = \mathsf{L}^{-}\mathsf{L}$

When the 'columns' of L, $\mathsf{r}_j^{\mathsf{L}}$ $(\mathsf{L} = \sum_{j=0}^{n-1}(\mathsf{L}_j^{\mathbb{W}} \otimes \mathsf{e}_j^{\mathbb{V}}))$ are linearly independent, then a useful identity is:

$$\mathsf{L}^{-} = \left(\mathsf{L}^{\dagger}\mathsf{L}\right)^{-1}\mathsf{L}^{\dagger} \tag{3.22}$$

The pseudoinverse can be computed using standard matrix decompositions such as the QR and SVD Golub and Van Loan (2012) *Matrix computations*. The pseudoinverse is an example of a linear transformation which should ***not*** be represented by a matrix McDonald (1989) "Object-oriented programming for linear algebra".

If A is affine, let $\mathsf{A} = \mathsf{L} + \mathsf{t}$, where L is linear, and t is an element of $\mathsf{range}(\mathsf{A})$. Then $\mathsf{A}^{-}(\mathsf{w}) = \mathsf{L}^{-}(\mathsf{w} - \mathsf{t})$.

A Plots, Tables, etc.

B Algorithms, Code, Pseudo-code, etc.

C Definitions, Theorems, etc.

D Notes

E Exercises

F Quotes

G Glossary

A | D | F | I | N | R | S

A

a generic space ($\mathbb{S}$) a generic space.

D

doubles ($\mathbb{D}$) the IEEE 754 64 bit floating point numbers.

F

floats ($\mathbb{F}$) the IEEE 754 32 bit floating point numbers.

I

integers ($\mathbb{Z}$) the integers. 3, 4

- **positive integers** ($\mathbb{Z}_+$) $\{i \in \mathbb{Z} \mid i > 0\}$.

N

natural numbers ($\mathbb{N}$) $\{i \in \mathbb{Z} \mid i \geq 0\}$. 4

R

rational numbers ($\mathbb{Q}$) $\left\{i/j \mid i \in \mathbb{Z}, j \in \mathbb{Z}_+\right\}$. 4,

real numbers ($\mathbb{R}$) the real numbers. 2, 4,

S

Sets

- **elementOf** ($\in$) $x \in \mathcal{S}$ means x is an element of the set $\mathcal{S}$. 4,
- **Set** ($\mathcal{S}$) a generic *set*.

Spaces

- **Homogeneous Space** a *space* where every point looks the same.

H References

[1] Gene H. Golub and Charles F. Van Loan.
Matrix computations.
4th.
Johns Hopkins U Press, 2012
(Cited on page 10).

[2] Paul R. Halmos.
Finite-dimensional Vector Spaces.
Van Nostrand, 1958
(Cited on pages 5–7, 9, 14).

[3] Saunders MacLane.
"Of course and courses".
In: *American Mathematical Monthly* 61 (1954), pages 151–157
(Cited on pages 5, 17).

[4] John Alan McDonald.
"Object-oriented programming for linear algebra".
In: *SIGPLAN Notices (Proceedings OOPSLA'89)* 24.10 (1989), pages 175–184
(Cited on pages 9, 10).

[5] Michael Spivak.
Calculus on Manifolds.
Addison-Wesley, 1965
(Cited on pages 5, 7).

[6] Jorge Stolfi.
Oriented projective geometry : a framework for geometric computations.
Boston, MA: Academic Press, 1991.
url: https://www.elsevier.com/books/oriented-projective-geometry/stolfi/978-0-12-672025-9
(Cited on page 8).

[7] Wikipedia.
Algebraic structure.
url: https://en.wikipedia.org/w/index.php?title=Algebraic_structure
(Cited on page 3).

[8] Wikipedia.
Field (mathematics).
url: https://en.wikipedia.org/w/index.php?title=Field_(mathematics)
(Cited on page 4).

[9] Wikipedia.
Mathematical structure.
url: https://en.wikipedia.org/w/index.php?title=Mathematical_structure
(Cited on pages 2, 3).

[10] Wikipedia.
Outline of algebraic structures.
URL: https://en.wikipedia.org/w/index.php?title=Outline_of_algebraic_structures
(Cited on page 3).

[11] Wikipedia.
Ring (mathematics).
URL: https://en.wikipedia.org/w/index.php?title=Ring_(mathematics)
(Cited on page 4).

[12] Brandon Williams.
Algebraic structure and protocols.
An example of the problem that results from combining types statically.
2017.
URL: https://medium.com/@mbrandonw/algebraic-structure-and-protocols-2bd48eb3e083
(Cited on page 3).

[13] Brandon Williams.
Semirings and predicates.
2017.
URL: https://medium.com/@mbrandonw/semirings-and-predicates-4474cfa2ac81
(Cited on page 3).

[14] Brandon Williams.
The algebra of predicates and sorting functions.
2017.
URL: https://medium.com/@mbrandonw/the-algebra-of-predicates-and-sorting-functions-b0282b17ef71
(Cited on page 3).

I Index