FORMAL RULES

1. Algebraic structures

1.1. Typical structures

A vector space has been introduced as a 4-tuple $\mathcal{V} = (V, S, +, \cdot)$ with specific behavior of the vector addition and scaling operations. Arithmetic operations between scalars were implicitly assumed to be similar to those of the real numbers, but also must be specified to obtain a complete definition of a vector space. Algebra defines various structures that specify the behavior operations with objects. Knowledge of these structures is useful not only in linear algebra, but also in other mathematical approaches to data analysis such as topology or geometry.

Groups. A group is a 2-tuple $\mathcal{G} = (G, +)$ containing a set G and an operation $+$ with properties from Table 2. If $\forall a, b \in G$, $a + b = b + a$, the group is said to be commutative. Besides the familiar example of integers under addition $(\mathbb{Z}, +)$, symmetry groups that specify spatial or functional relations are of particular interest. The rotations by $0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$ or vertices of a square form a group.

Addition rules	
$a + b \in G$	Closure
$a + (b + c) = (a + b) + c$	Associativity
$0 + a = a$	Identity element
$a + (-a) = 0$	Inverse element

Table 1. Group $\mathcal{G} = (G, +)$ properties, for $\forall a, b, c \in G$

Rings. A ring is a 3-tuple $\mathcal{R} = (R, +, \cdot)$ containing a set R and two operations $+, \cdot$ with properties from Table 1. As is often the case, a ring is more complex structure built up from simpler algebraic structures. With respect to addition a ring has the properties of a commutative group. Only associativity and existence of an identity element is imposed for multiplication. Matrix addition and multiplication has the structure of ring $(\mathbb{R}^{m \times m}, +, \cdot)$.

Addition rules	
$(R, +)$ is a commutative (Abelian) group	
Multiplication rules	
$a \cdot b \in R$	Closure
$(a \cdot b) \cdot c = a \cdot (b \cdot c)$	Associativity
$a \cdot 1 = 1 \cdot a = a$	Identity element
Distributivity	
$a \cdot (b + c) = (a \cdot b) + (a \cdot c)$	on the left
$(a + b) \cdot c = (a \cdot c) + (b \cdot c)$	on the right

Table 2. Ring $\mathcal{R} = (R, +, \cdot)$ properties, for $\forall a, b, c \in R$.

Fields. A ring is a 3-tuple $\mathcal{F} = (F, +, \cdot)$ containing a set F and two operations $+, \cdot$, each with properties of a commutative group, but with special behavior for the inverse of the null element. The multiplicative inverse is denoted as a^{-1}. Scalars S in the definition of a vector space must satisfy the properties of a field. Since the operations are often understood from context a field might be referred to as the full $3 -$ tuple, or, more concisely just through the set of elements as in the definition of a vector space.

Addition rules	
$(F, +)$ is a commutative (Abelian) group	
Multiplication rules	
$(F, \cdot)$ is a commutative group except	
that 0^{-1} does not exist	
Distributivity	
$a \cdot (b + c) = (a \cdot b) + (a \cdot c)$	

Table 3. Field $\mathcal{R} = (F, +, \cdot)$ properties, for $\forall a, b, c \in F$.

Using the above definitions, a vector space $\mathcal{V} = (V, S, +, \cdot)$ can be described as a commutative group $(V, +)$ combined with a field S that satisfies the scaling properties $a\boldsymbol{u} \in V$, $a(\boldsymbol{u} + \boldsymbol{v}) = a\boldsymbol{u} + b\boldsymbol{v}$, $(a + b)\boldsymbol{u} = a\boldsymbol{u} + b\boldsymbol{u}$, $a(b\boldsymbol{u}) = (ab)\boldsymbol{u}$, $1\boldsymbol{u} = \boldsymbol{u}$, for $\forall a, b \in S$, $\forall \boldsymbol{u}, \boldsymbol{v} \in V$.

1.2. Vector subspaces

A central interest in data science is to seek simple description of complex objects. A typical situation is that many instances of some object of interest are initially given as an m-tuple $\boldsymbol{v} \in \mathbb{R}^m$ with large m. Assuming that addition and scaling of such objects can cogently be defined, a vector space is obtained, say over the field of reals with an Euclidean

distance, E_m. Examples include for instance recordings of medical data (electroencephalograms, electrocardiograms), sound recordings, or images, for which m can easily reach in to the millions. A natural question to ask is whether all the m real numbers are actually needed to describe the observed objects, or perhaps there is some intrinsic description that requires a much smaller number of descriptive parameters, that still preserves the useful idea of linear combination. The mathematical transcription of this idea is a vector subspace.

DEFINITION. (VECTOR SUBSPACE). *$\mathcal{U} = (U, S, +, \cdot)$, $U \neq \emptyset$, is a vector subspace of vector space $\mathcal{V} = (V, S, +, \cdot)$ over the same field of scalars S, denoted by $\mathcal{U} \leq \mathcal{V}$, if $U \subseteq V$ and $\forall a, b \in S$, $\forall u, v \in U$, the linear combination $au + bv \in U$.*

The above states a vector subspace must be closed under linear combination, and have the same vector addition and scaling operations as the enclosing vector space. The simplest vector subspace of a vector space is the null subspace that only contains the null element, $U = \{0\}$. In fact any subspace must contain the null element 0, or otherwise closure would not be verified for the particular linear combination $u + (-u) = 0$. If $U \subset V$, then $\mathcal{U}$ is said to be a *proper subspace* of $\mathcal{V}$, denoted by $\mathcal{U} < \mathcal{V}$.

Setting $n - m$ components equal to zero in the real space $\mathcal{R}_m$ defines a proper subspace whose elements can be placed into a one-to-one correspondence with the vectors within $\mathcal{R}_n$. For example, setting component m of $x \in \mathbb{R}^m$ equal to zero gives $x = [\, x_1 \ x_2 \ \ldots \ x_{m-1} \ 0 \,]^T$ that while not a member of $\mathbb{R}^{m-1}$, is in a one-to-one relation with $x' = [\, x_1 \ x_2 \ \ldots \ x_{m-1} \,]^T \in \mathbb{R}^{m-1}$. Dropping the last component of $y \in \mathbb{R}^m$, $y = [\, y_1 \ y_2 \ \ldots \ y_{m-1} \ y_m \,]^T$ gives vector $y' = [\, y_1 \ y_2 \ \ldots \ y_{m-1} \,] \in \mathbb{R}^{m-1}$, but this is no longer a one-to-one correspondence since for some given y', the last component y_m could take any value.

```
octave] m=3; x=[1; 2; 0]; xp=x(1:2); disp(xp)

   1
   2
octave] y=[1; 2; 3]; yp=y(1:2); disp(yp)

   1
   2
octave]
```

Vector subspaces arise in decomposition of a vector space. The converse, composition of vector spaces $\mathcal{U} = (U, S, +, \cdot)$ $\mathcal{V} = (V, S, +, \cdot)$ is also defined in terms of linear combination. A vector $x \in \mathbb{R}^3$ can be obtained as the linear combination

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ x_2 \\ x_3 \end{bmatrix},$$

but also as

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 - a \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ a \\ x_3 \end{bmatrix},$$

for some arbitrary $a \in \mathbb{R}$. In the first case, x is obtained as a unique linear combination of a vector from the set $U = \{[\, x_1 \ 0 \ 0 \,]^T | x_1 \in \mathbb{R}\}$ with a vector from $V = \{[\, 0 \ x_2 \ x_3 \,]^T | x_2, x_3 \in \mathbb{R}\}$. In the second case, there is an infinity of linear combinations of a vector from V with another from $W = \{[\, x_1 \ x_2 \ 0 \,]^T | x_1, x_2 \in \mathbb{R}\}$ to the vector x. This is captured by a pair of definitions to describe vector space composition.

DEFINITION. *Given two vector subspaces $\mathcal{U} = (U, S, +, \cdot)$, $\mathcal{V} = (V, S, +, \cdot)$ of the space $\mathcal{W} = (W, S, +, \cdot)$, the sum is the vector space $\mathcal{U} + \mathcal{V} = (U + V, S, +, \cdot)$, where the sum of the two sets of vectors U, V is $U + V = \{u + v \,|\, u \in U, v \in V\}$.*

DEFINITION. *Given two vector subspaces $\mathcal{U} = (U, S, +, \cdot)$, $\mathcal{V} = (V, S, +, \cdot)$ of the space $\mathcal{W} = (W, S, +, \cdot)$, the direct sum is the vector space $\mathcal{U} \oplus \mathcal{V} = (U \oplus V, S, +, \cdot)$, where the direct sum of the two sets of vectors U, V is $U \oplus V = \{\boldsymbol{u} + \boldsymbol{v} \mid \exists! \boldsymbol{u} \in U, \exists! \boldsymbol{v} \in V\}$. (unique decomposition)*

Since the same scalar field, vector addition, and scaling is used , it is more convenient to refer to vector space sums simply by the sum of the vector sets $U + V$, or $U \oplus V$, instead of specifying the full tuplet for each space. This shall be adopted henceforth to simplify the notation.

```
octave] u=[1; 0; 0]; v=[0; 2; 3]; vp=[0; 1; 3]; w=[1; 1; 0]; disp([u+v vp+w])
   1   1
   2   2
   3   3
octave]
```

In the previous example, the essential difference between the two ways to express $\boldsymbol{x} \in \mathbb{R}^3$ is that $U \cap V = \{\boldsymbol{0}\}$, but $V \cap W = \{[\, 0 \ a \ 0 \,]^T \mid a \in \mathbb{R}\} \neq \{\boldsymbol{0}\}$, and in general if the zero vector is the only common element of two vector spaces then the sum of the vector spaces becomes a direct sum. In practice, the most important procedure to construct direct sums or check when an intersection of two vector subspaces reduces to the zero vector is through an inner product.

DEFINITION. *Two vector subspaces U, V of the real vector space $\mathbb{R}^m$ are orthogonal, denoted as $U \perp V$ if $\boldsymbol{u}^T \boldsymbol{v} = 0$ for any $\boldsymbol{u} \in U, \boldsymbol{v} \in V$.*

DEFINITION. *Two vector subspaces U, V of $U + V$ are orthogonal complements, denoted $U = V^\perp$, $V = U^\perp$ if they are orthogonal subspaces, $U \perp V$, and $U \cap V = \{\boldsymbol{0}\}$, i.e., the null vector is the only common element of both subspaces.*

```
octave] disp([u'*v vp'*w])
   0   1
octave]
```

The above concept of orthogonality can be extended to other vector subspaces, such as spaces of functions. It can also be extended to other choices of an inner product, in which case the term conjugate vector spaces is sometimes used.

The concepts of sum and direct sum of vector spaces used linear combinations of the form $\boldsymbol{u} + \boldsymbol{v}$. This notion can be extended to arbitrary linear combinations.

DEFINITION. *In vector space $\mathcal{V} = (V, S, +, \cdot)$, the span of vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n \in V$, is the set of vectors reachable by linear combination*

$$\text{span}\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\} = \{\boldsymbol{b} \in V \mid \exists x_1, \ldots, x_n \in S \text{ such that } \boldsymbol{b} = x_1 \boldsymbol{a}_1 + \ldots + x_n \boldsymbol{a}_n\}.$$

Note that for real vector spaces a member of the span of the vectors $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\}$ is the vector $\boldsymbol{b}$ obtained from the matrix vector multiplication

$$\boldsymbol{b} = A\boldsymbol{x} = [\, \boldsymbol{a}_1 \ \boldsymbol{a}_2 \ \ldots \ \boldsymbol{a}_n \,] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

From the above, the span is a subset of the co-domain of the linear mapping $f(x) = Ax$.

2. Vector subspaces of a linear mapping

The wide-ranging utility of linear algebra essentially results a complete characterization of the behavior of a linear mapping between vector spaces $f: U \rightarrow V, f(au + bv) = af(u) + bf(v)$. For some given linear mapping the questions that arise are:

1. Can any vector within V be obtained by evaluation of f?

2. Is there a single way that a vector within V can be obtained by evaluation of f?

Linear mappings between real vector spaces $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, have been seen to be completely specified by a matrix $A \in \mathbb{R}^{m \times n}$. It is common to frame the above questions about the behavior of the linear mapping $f(x) = Ax$ through sets associated with the matrix A. To frame an answer to the first question, a set of reachable vectors is first defined.

DEFINITION. *The column space (or range) of matrix $A \in \mathbb{R}^{m \times n}$ is the set of vectors reachable by linear combination of the matrix column vectors*

$$C(A) = \text{range}(A) = \{b \in \mathbb{R}^m | \exists x \in \mathbb{R}^n \text{ such that } b = Ax\}.$$

By definition, the column space is included in the co-domain of the function $f(x) = Ax$, $C(A) \subseteq \mathbb{R}^m$, and is readily seen to be a vector subspace of $\mathbb{R}^m$. The question that arises is whether the column space is the entire co-domain $C(A) = \mathbb{R}^m$ that would signify that any vector can be reached by linear combination. If this is not the case then the column space would be a proper subset, $C(A) \subset \mathbb{R}^m$, and the question is to determine what part of the co-domain cannot be reached by linear combination of columns of A. Consider the orthogonal complement of $C(A)$ defined as the set vectors orthogonal to all of the column vectors of A, expressed through inner products as

$$a_1^T y = 0, a_2^T y = 0, \ldots, a_n^T y = 0.$$

This can be expressed more concisely through the transpose operation

$$A = [\begin{array}{cccc} a_1 & a_2 & \ldots & a_n \end{array}], A^T y = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix} = \begin{bmatrix} a_1^T y \\ a_2^T y \\ \vdots \\ a_n^T y \end{bmatrix},$$

and leads to the definition of a set of vectors for which $A^T y = 0$

DEFINITION. *The left null space (or cokernel) of a matrix $A \in \mathbb{R}^{m \times n}$ is the set*

$$N(A^T) = \text{null}(A^T) = \{y \in \mathbb{R}^m | A^T y = 0\}.$$

Note that the left null space is also a vector subspace of the co-domain of $f(x) = Ax$, $N(A^T) \subseteq \mathbb{R}^m$. The above definitions suggest that both the matrix and its transpose play a role in characterizing the behavior of the linear mapping $f = Ax$, so analagous sets are define for the transpose A^T.

DEFINITION. *The row space (or corange) of a matrix $A \in \mathbb{R}^{m \times n}$ is the set*

$$R(A) = C(A^T) = \text{range}(A^T) = \{c \in \mathbb{R}^n | \exists y \in \mathbb{R}^m c = A^T y\} \subseteq \mathbb{R}^n$$

DEFINITION. *The null space of a matrix $A \in \mathbb{R}^{m \times n}$ is the set*

$$N(A) = \text{null}(A) = \{x \in \mathbb{R}^n | Ax = 0\} \subseteq \mathbb{R}^n$$

Examples. Consider a linear mapping between real spaces $f : \mathbb{R}^n \to \mathbb{R}^m$, defined by $y = f(x) = Ax = [\ y_1\ \ldots\ y_n\ ]^T$, with $A \in \mathbb{R}^{m \times n}$.

1. For $n = 1$, $m = 3$,

$$A = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, A^T = [\ 1\ 0\ 0\ ],$$

the column space $C(A)$ is the y_1-axis, and the left null space $N(A^T)$ is the $y_2 y_3$-plane. Vectors that span these spaces are returned by the Octave `orth` and `null` functions.

```octave
octave] A=[1; 0; 0];
        disp(orth(A));
        disp('-----');
        disp(null(A'))

   -1
   -0
   -0
  -----
        0    0
        1    0
        0    1
octave]
```

2. For $n = 2$, $m = 3$,

$$A = \begin{bmatrix} 1 & -1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = [\ a_1\ a_2\ ], A^T = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix},$$

the columns of A are colinear, $a_2 = -a_1$, and the column space $C(A)$ is the y_1-axis, and the left null space $N(A^T)$ is the $y_2 y_3$-plane, as before.

```octave
octave] A=[1 -1; 0 0; 0 0];
        disp(orth(A));
        disp('-----');
        disp(null(A'))

   -1.00000
   -0.00000
   -0.00000
  -----
        0    0
        1    0
        0    1
```

```octave
octave]
```

3. For $n = 2$, $m = 3$,

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, A^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

the column space $C(A)$ is the $y_1 y_2$-plane, and the left null space $N(A^T)$ is the y_3-axis.

```octave
octave] A=[1 0; 0 1; 0 0];
        disp(orth(A));
        disp('-----');
        disp(null(A'))

   -1    -0
   -0    -1
   -0    -0
  -----
        0
        0
        1
octave]
```

4. For $n = 2$, $m = 3$,

$$A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 0 \end{bmatrix}, A^T = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \end{bmatrix},$$

the same $C(A)$, $N(A^T)$ are obtained, albeit with a different set of spanning vectors returned by `orth`.

```octave
octave] A=[1 1; 1 -1; 0 0];
        disp(orth(A));
        disp('-----');
        disp(null(A'))

    0.70711     0.70711
    0.70711    -0.70711
   -0.00000    -0.00000
  -----
        0
        0
        1
```

5. For $n = 3$, $m = 3$,

$$A = \begin{bmatrix} 1 & 1 & 3 \\ 1 & -1 & -1 \\ 1 & 1 & 3 \end{bmatrix} = [\, \boldsymbol{a}_1 \ \boldsymbol{a}_2 \ \boldsymbol{a}_3 \,],$$

$$A^T = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 3 & -1 & 3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1^T \\ \boldsymbol{a}_2^T \\ \boldsymbol{a}_3^T \end{bmatrix}, A^T \boldsymbol{y} = \begin{bmatrix} \boldsymbol{a}_1^T \boldsymbol{y} \\ \boldsymbol{a}_2^T \boldsymbol{y} \\ \boldsymbol{a}_3^T \boldsymbol{y} \end{bmatrix}$$

since $\boldsymbol{a}_3 = \boldsymbol{a}_1 + 2\boldsymbol{a}_2$, the orthogonality condition $A^T \boldsymbol{y} = \boldsymbol{0}$ is satisfied by vectors of form $\boldsymbol{y} = [\, a \ 0 \ -a \,]$, $a \in \mathbb{R}$.

```octave
octave] A=[1 1 3; 1 -1 -1; 1 1 3];
         disp(orth(A));
         disp('-----');
         disp(null(A'))
   0.69157    0.14741
  -0.20847    0.97803
   0.69157    0.14741
-----
   0.70711
   0.00000
  -0.70711
octave]
```

The above low dimensional examples are useful to gain initial insight into the significance of the spaces $C(A), N(A^T)$. Further appreciation can be gained by applying the same concepts to processing of images. A gray-scale image of size p_x by p_y pixels can be represented as a vector with $m = p_x p_y$ components, $\boldsymbol{b} \in [0,1]^m \subset \mathbb{R}^m$. Even for a small image with $p_x = p_y = 128 = 2^7$ pixels along each direction, the vector $\boldsymbol{b}$ would have $m = 2^{14}$ components. An image can be specified as a linear combination of the columns of the identity matrix

$$\boldsymbol{b} = I\boldsymbol{b} = [\, \boldsymbol{e}_1 \ \boldsymbol{e}_2 \ \dots \ \boldsymbol{e}_m \,] \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix},$$

with b_i the gray-level intensity in pixel i. Similar to the inclined plane example from §1, an alternative description as a linear combination of another set of vectors $\boldsymbol{a}_1, \dots, \boldsymbol{a}_m$ might be more relevant. One choice of greater utility for image processing mimics the behavior of the set $\{1, \cos t, \cos 2t, \dots, \sin t, \sin 2t, \dots\}$ that extends the second example in §1, would be for $m = 4$

$$A = [\, \boldsymbol{a}_1 \ \boldsymbol{a}_2 \ \boldsymbol{a}_3 \ \boldsymbol{a}_4 \,] = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$