FORMAL RULES

1. Algebraic structures

1.1. Typical structures

A vector space has been introduced as a 4-tuple $\mathscr{V} = (V, S, +, \cdot)$ 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 $\mathscr{G} = (G, +)$ 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 $\mathscr{G} = (G, +)$ properties, for $\forall a, b, c \in G$

Rings. A ring is a 3-tuple $\mathscr{R} = (R, +, \cdot)$ 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 $\mathscr{R} = (R, +, \cdot)$ properties, for $\forall a, b, c \in R$.

Fields. A ring is a 3-tuple $\mathscr{F} = (F, +, \cdot)$ 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 $\mathscr{R} = (F, +, \cdot)$ properties, for $\forall a, b, c \in F$.

Using the above definitions, a vector space $\mathscr{V} = (V, S, +, \cdot)$ 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) . *$\mathscr{U} = (U, S, +, \cdot)$, $U \neq \emptyset$, is a vector subspace of vector space $\mathscr{V} = (V, S, +, \cdot)$ over the same field of scalars S, denoted by $\mathscr{U} \leq \mathscr{V}$, if $U \subseteq V$ and $\forall a, b \in S$, $\forall \boldsymbol{u}, \boldsymbol{v} \in U$, the linear combination $a\boldsymbol{u} + b\boldsymbol{v} \in U$.*

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, $U = \{\mathbf{0}\}$. In fact any subspace must contain the null element $\mathbf{0}$, or otherwise closure would not be verified for the particular linear combination $\boldsymbol{u} + (-\boldsymbol{u}) = \mathbf{0}$. If $U \subset V$, then $\mathscr{U}$ is said to be a *proper subspace* of $\mathscr{V}$, denoted by $\mathscr{U} < \mathscr{V}$.

Setting $n - m$ components equal to zero in the real space $\mathscr{R}_m$ defines a proper subspace whose elements can be placed into a one-to-one correspondence with the vectors within $\mathscr{R}_n$. For example, setting component m of $\boldsymbol{x} \in \mathbb{R}^m$ equal to zero gives $\boldsymbol{x} = [\ x_1\ \ x_2\ \ \dots\ \ x_{m-1}\ \ 0\ ]^T$ that while not a member of $\mathbb{R}^{m-1}$, is in a one-to-one relation with $\boldsymbol{x}' = [\ x_1\ \ x_2\ \ \dots\ \ x_{m-1}\ ]^T \in \mathbb{R}^{m-1}$. Dropping the last component of $\boldsymbol{y} \in \mathbb{R}^m$, $\boldsymbol{y} = [\ y_1\ \ y_2\ \ \dots\ \ y_{m-1}\ \ y_m\ ]^T$ gives vector $\boldsymbol{y}' = [\ y_1\ \ y_2\ \ \dots\ \ y_{m-1}\ ] \in \mathbb{R}^{m-1}$, but this is no longer a one-to-one correspondence since for some given $\boldsymbol{y}'$, 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 $\mathscr{U} = (U, S, +, \cdot)$ $\mathscr{V} = (V, S, +, \cdot)$ is also defined in terms of linear combination. A vector $\boldsymbol{x} \in \mathbb{R}^3$ can be obtained as the linear combination

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

but also as

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

for some arbitrary $a \in \mathbb{R}$. In the first case, $\boldsymbol{x}$ 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 $\boldsymbol{x}$. This is captured by a pair of definitions to describe vector space composition.

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

DEFINITION. *Given two vector subspaces $\mathscr{U} = (U, S, +, \cdot)$, $\mathscr{V} = (V, S, +, \cdot)$ of the space $\mathscr{W} = (W, S, +, \cdot)$, the direct sum is the vector space $\mathscr{U} \oplus \mathscr{V} = (U \oplus V, S, +, \cdot)$, where the direct sum of the two sets of vectors U, V is $U \oplus V = \{\boldsymbol{u} + \boldsymbol{v} | \exists! \boldsymbol{u} \in U, \exists! \boldsymbol{v} \in V\}$. (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 $V \cap W = \{[\ 0\ \ a\ \ 0\ ]^T | a \in \mathbb{R}\} \neq \{\boldsymbol{0}\}$, 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 $\mathscr{V} = (V, S, +, \cdot)$, 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*

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

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

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

From the above, the span is a subset of the co-domain of the linear mapping $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$.

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 $\boldsymbol{f}: U \to V, \boldsymbol{f}(a\boldsymbol{u} + b\boldsymbol{v}) = a\boldsymbol{f}(\boldsymbol{u}) + b\boldsymbol{f}(\boldsymbol{v})$. For some given linear mapping the questions that arise are:

1. Can any vector within V be obtained by evaluation of $\boldsymbol{f}$?
2. Is there a single way that a vector within V can be obtained by evaluation of $\boldsymbol{f}$?

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

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

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

By definition, the column space is included in the co-domain of the function $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$, $C(\boldsymbol{A}) \subseteq \mathbb{R}^m$, 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 $C(\boldsymbol{A}) = \mathbb{R}^m$ 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, $C(\boldsymbol{A}) \subset \mathbb{R}^m$, and the question is to determine what part of the co-domain cannot be reached by linear combination of columns of $\boldsymbol{A}$. Consider the orthogonal complement of $C(\boldsymbol{A})$ defined as the set vectors orthogonal to all of the column vectors of $\boldsymbol{A}$, expressed through inner products as

$$\boldsymbol{a}_1^T \boldsymbol{y} = 0, \boldsymbol{a}_2^T \boldsymbol{y} = 0, \ldots, \boldsymbol{a}_n^T \boldsymbol{y} = 0.$$

This can be expressed more concisely through the transpose operation

$$\boldsymbol{A} = [\ \boldsymbol{a}_1\ \ \boldsymbol{a}_2\ \ \ldots\ \ \boldsymbol{a}_n\ ], \boldsymbol{A}^T\boldsymbol{y} = \begin{bmatrix} \boldsymbol{a}_1^T \\ \boldsymbol{a}_2^T \\ \vdots \\ \boldsymbol{a}_n^T \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1^T \boldsymbol{y} \\ \boldsymbol{a}_2^T \boldsymbol{y} \\ \vdots \\ \boldsymbol{a}_n^T \boldsymbol{y} \end{bmatrix},$$

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

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

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

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

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

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

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

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

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

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

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

the column space $C(\boldsymbol{A})$ is the y_1-axis, and the left null space $N(\boldsymbol{A}^T)$ is the y_2y_3-plane. Vectors that span these spaces are returned by the Octave `orth` and `null` functions.

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

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

$$\boldsymbol{A} = \begin{bmatrix} 1 & -1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = [\ \boldsymbol{a}_1\ \ \boldsymbol{a}_2\ ], \boldsymbol{A}^T = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix},$$

the columns of $\boldsymbol{A}$ are colinear, $\boldsymbol{a}_2 = -\boldsymbol{a}_1$, and the column space $C(\boldsymbol{A})$ is the y_1-axis, and the left null space $N(\boldsymbol{A}^T)$ is the y_2y_3-plane, as before.

```
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]
```

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

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

the column space $C(\boldsymbol{A})$ is the y_1y_2-plane, and the left null space $N(\boldsymbol{A}^T)$ is the y_3-axis.

```
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 $n = 2$, $m = 3$,

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

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

```
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
```

```
octave]
```

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

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

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

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

```
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 $C(\boldsymbol{A}),N(\boldsymbol{A}^T)$. 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 $m=p_xp_y$ 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

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

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 $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_m$ might be more relevant. One choice of greater utility for image processing mimics the behavior of the set $\{1,\cos t,\cos 2t,\ldots,\sin t,\sin 2t,\ldots\}$ that extends the second example in §1, would be for $m=4$

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

DATA REDUNDANCY

1. Linear dependence

For the simple scalar mapping $f:\mathbb{R}\to\mathbb{R}$, $f(x)=ax$, the condition $f(x)=0$ implies either that $a=0$ or $x=0$. Note that $a=0$ can be understood as defining a zero mapping $f(x)=0$. Linear mappings between vector spaces, $\boldsymbol{f}:U\to V$, can exhibit different behavior, and the condtion $\boldsymbol{f}(\boldsymbol{x})=\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$, might be satisfied for both $\boldsymbol{x}\neq\boldsymbol{0}$, and $\boldsymbol{A}\neq\boldsymbol{0}$. Analogous to the scalar case, $\boldsymbol{A}=\boldsymbol{0}$ can be understood as defining a zero mapping, $\boldsymbol{f}(\boldsymbol{x})=\boldsymbol{0}$.

In vector space $\mathscr{V}=(V,S,+,\cdot)$, vectors $\boldsymbol{u},\boldsymbol{v}\in V$ related by a scaling operation, $\boldsymbol{v}=a\boldsymbol{u}$, $a\in S$, are said to be colinear, and are considered to contain redundant data. This can be restated as $\boldsymbol{v}\in\text{span}\{\boldsymbol{u}\}$, from which it results that $\text{span}\{\boldsymbol{u}\}=\text{span}\{\boldsymbol{u},\boldsymbol{v}\}$. Colinearity can be expressed only in terms of vector scaling, but other types of redundancy arise when also considering vector addition as expressed by the span of a vector set. Assuming that $\boldsymbol{v}\notin\text{span}\{\boldsymbol{u}\}$, then the strict inclusion relation $\text{span}\{\boldsymbol{u}\}\subset\text{span}\{\boldsymbol{u},\boldsymbol{v}\}$ holds. This strict inclusion expressed in terms of set concepts can be transcribed into an algebraic condition.

DEFINITION. *The vectors* $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n \in V$, *are linearly dependent if there exist n scalars,* $x_1, \ldots, x_n \in S$, *at least one of which is different from zero such that*

$$x_1 \boldsymbol{a}_1 + \ldots + x_n \boldsymbol{a}_n = \boldsymbol{0}.$$

Introducing a matrix representation of the vectors

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

allows restating linear dependence as the existence of a non-zero vector, $\exists \boldsymbol{x} \neq \boldsymbol{0}$, such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$. Linear dependence can also be written as $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \nRightarrow \boldsymbol{x} = \boldsymbol{0}$, or that one cannot deduce from the fact that the linear mapping $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ attains a zero value that the argument itself is zero. The converse of this statement would be that the only way to ensure $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ is for $\boldsymbol{x} = \boldsymbol{0}$, or $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \Rightarrow \boldsymbol{x} = \boldsymbol{0}$, leading to the concept of linear independence.

DEFINITION. *The vectors* $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n \in V$, *are linearly independent if the only n scalars,* $x_1, \ldots, x_n \in S$, *that satisfy*

$$x_1 \boldsymbol{a}_1 + \ldots + x_n \boldsymbol{a}_n = \boldsymbol{0}, \tag{1}$$

are $x_1 = 0$, $x_2 = 0, \ldots, x_n = 0$.

2. Basis and dimension

Vector spaces are closed under linear combination, and the span of a vector set $\mathcal{B} = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots\}$ defines a vector subspace. If the entire set of vectors can be obtained by a spanning set, $V = \text{span}\,\mathcal{B}$, extending $\mathcal{B}$ by an additional element $\mathcal{C} = \mathcal{B} \cup \{\boldsymbol{b}\}$ would be redundant since $\text{span}\,\mathcal{B} = \text{span}\,\mathcal{C}$. This is recognized by the concept of a basis, and also allows leads to a characterization of the size of a vector space by the cardinality of a basis set.

DEFINITION. *A set of vectors* $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n \in V$ *is a basis for vector space* $\mathcal{V} = (V, S, +, \cdot)$ *if*

1. $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ *are linearly independent;*
2. $\text{span}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\} = V$.

DEFINITION. *The number of vectors* $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n \in V$ *within a basis is the dimension of the vector space* $\mathcal{V} = (V, S, +, \cdot)$.

3. Dimension of matrix spaces

The domain and co-domain of the linear mapping $\boldsymbol{f}: U \to V, \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$, are decomposed by the spaces associated with the matrix $\boldsymbol{A}$. When $U = \mathbb{R}^n$, $V = \mathbb{R}^m$, the following vector subspaces associated with the matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ have been defined:

- $C(\boldsymbol{A})$ the column space of $\boldsymbol{A}$
- $C(\boldsymbol{A}^T)$ the row space of $\boldsymbol{A}$
- $N(\boldsymbol{A})$ the null space of $\boldsymbol{A}$
- $N(\boldsymbol{A}^T)$ the left null space of $\boldsymbol{A}$, or null space of $\boldsymbol{A}^T$

DEFINITION. *The rank of a matrix* $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ *is the dimension of its column space and is equal to the dimension of its row space.*

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

DATA INFORMATION

1. Partition of linear mapping domain and codomain

A partition of a set S has been introduced as a collection of subsets $P = \{S_i | S_i \subset P, S_i \neq \emptyset\}$ such that any given element $x \in S$ belongs to only one set in the partition. This is modified when applied to subspaces of a vector space, and a partition of a set of vectors is understood as a collection of subsets such that any vector except $\mathbf{0}$ belongs to only one member of the partition.

Linear mappings between vector spaces $\boldsymbol{f}: U \to V$ can be represented by matrices $\boldsymbol{A}$ with columns that are images of the columns of a basis $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots\}$ of U

$$\boldsymbol{A} = [\, \boldsymbol{f}(\boldsymbol{u}_1) \;\; \boldsymbol{f}(\boldsymbol{u}_2) \;\; \dots \,].$$

Consider the case of real finite-dimensional domain and co-domain, $\boldsymbol{f}: \mathbb{R}^n \to \mathbb{R}^m$, in which case $\boldsymbol{A} \in \mathbb{R}^{m \times n}$,

$$\boldsymbol{A} = [\, \boldsymbol{f}(\boldsymbol{e}_1) \;\; \boldsymbol{f}(\boldsymbol{e}_2) \;\; \dots \;\; \boldsymbol{f}(\boldsymbol{e}_n) \,] = [\, \boldsymbol{a}_1 \;\; \boldsymbol{a}_2 \;\; \dots \;\; \boldsymbol{a}_n \,].$$

The column space of $\boldsymbol{A}$ is a vector subspace of the codomain, $C(\boldsymbol{A}) \leq \mathbb{R}^m$, but according to the definition of dimension if $n < m$ there remain non-zero vectors within the codomain that are outside the range of $\boldsymbol{A}$,

$$n < m \Rightarrow \exists \boldsymbol{v} \in \mathbb{R}^m, \boldsymbol{v} \neq \mathbf{0}, \boldsymbol{v} \notin C(\boldsymbol{A}).$$

All of the non-zero vectors in $N(\boldsymbol{A}^T)$, namely the set of vectors orthogonal to all columns in $\boldsymbol{A}$ fall into this category. The above considerations can be stated as

$$C(\boldsymbol{A}) \leq \mathbb{R}^m, \;\; N(\boldsymbol{A}^T) \leq \mathbb{R}^m, \;\; C(\boldsymbol{A}) \perp N(\boldsymbol{A}^T) \;\; C(\boldsymbol{A}) + N(\boldsymbol{A}^T) \leq \mathbb{R}^m .$$

The question that arises is whether there remain any non-zero vectors in the codomain that are not part of $C(\boldsymbol{A})$ or $N(\boldsymbol{A}^T)$. The fundamental theorem of linear algebra states that there no such vectors, that $C(\boldsymbol{A})$ is the orthogonal complement of $N(\boldsymbol{A}^T)$, and their direct sum covers the entire codomain $C(\boldsymbol{A}) \oplus N(\boldsymbol{A}^T) = \mathbb{R}^m$.

LEMMA 1. *Let $\mathcal{U}, \mathcal{V}$, be subspaces of vector space $\mathcal{W}$. Then $\mathcal{W} = \mathcal{U} \oplus \mathcal{V}$ if and only if*

i. $\mathcal{W} = \mathcal{U} + \mathcal{V}$, *and*

ii. $\mathcal{U} \cap \mathcal{V} = \{\mathbf{0}\}$.

Proof. *$\mathcal{W} = \mathcal{U} \oplus \mathcal{V} \Rightarrow \mathcal{W} = \mathcal{U} + \mathcal{V}$ by definition of direct sum, sum of vector subspaces. To prove that $\mathcal{W} = \mathcal{U} \oplus \mathcal{V} \Rightarrow \mathcal{U} \cap \mathcal{V} = \{\mathbf{0}\}$, consider $\boldsymbol{w} \in \mathcal{U} \cap \mathcal{V}$. Since $\boldsymbol{w} \in \mathcal{U}$ and $\boldsymbol{w} \in \mathcal{V}$ write*

$$\boldsymbol{w} = \boldsymbol{w} + \mathbf{0} \;\; (\boldsymbol{w} \in \mathcal{U}, \mathbf{0} \in \mathcal{V}), \;\; \boldsymbol{w} = \mathbf{0} + \boldsymbol{w} \;\; (\mathbf{0} \in \mathcal{U}, \boldsymbol{w} \in \mathcal{V}),$$

and since expression $\boldsymbol{w} = \boldsymbol{u} + \boldsymbol{v}$ is unique, it results that $\boldsymbol{w} = \mathbf{0}$. Now assume (i),(ii) and establish an unique decomposition. Assume there might be two decompositions of $\boldsymbol{w} \in \mathcal{W}$, $\boldsymbol{w} = \boldsymbol{u}_1 + \boldsymbol{v}_1$, $\boldsymbol{w} = \boldsymbol{u}_2 + \boldsymbol{v}_2$, with $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathcal{U}$, $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathcal{V}$. Obtain $\boldsymbol{u}_1 + \boldsymbol{v}_1 = \boldsymbol{u}_2 + \boldsymbol{v}_2$, or $\boldsymbol{x} = \boldsymbol{u}_1 - \boldsymbol{u}_2 = \boldsymbol{v}_2 - \boldsymbol{v}_1$. Since $\boldsymbol{x} \in \mathcal{U}$ and $\boldsymbol{x} \in \mathcal{V}$ it results that $\boldsymbol{x} = \mathbf{0}$, and $\boldsymbol{u}_1 = \boldsymbol{u}_2$, $\boldsymbol{v}_1 = \boldsymbol{v}_2$, i.e., the decomposition is unique. □

In the vector space $U + V$ the subspaces U, V are said to be orthogonal complements is $U \perp V$, and $U \cap V = \{\mathbf{0}\}$. When $U \leq \mathbb{R}^m$, the orthogonal complement of U is denoted as $U^\perp$, $U \oplus U^\perp = \mathbb{R}^m$.

THEOREM. *Given the linear mapping associated with matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ we have:*

1. *$C(\boldsymbol{A}) \oplus N(\boldsymbol{A}^T) = \mathbb{R}^m$, the direct sum of the column space and left null space is the codomain of the mapping*
2. *$C(\boldsymbol{A}^T) \oplus N(\boldsymbol{A}) = \mathbb{R}^n$, the direct sum of the row space and null space is the domain of the mapping*

3. $C(\boldsymbol{A}) \perp N(\boldsymbol{A}^T)$ *and* $C(\boldsymbol{A}) \cap N(\boldsymbol{A}^T) = \{\boldsymbol{0}\}$*, the column space is orthogonal to the left null space, and they are orthogonal complements of one another,*

$$C(\boldsymbol{A}) = N(\boldsymbol{A}^T)^{\perp},\ N(\boldsymbol{A}^T) = C(\boldsymbol{A})^{\perp}\ .$$

4. $C(\boldsymbol{A}^T) \perp N(\boldsymbol{A})$ *and* $C(\boldsymbol{A}^T) \cap N(\boldsymbol{A}) = \{\boldsymbol{0}\}$*, the row space is orthogonal to the null space, and they are orthogonal complements of one another,*

$$C(\boldsymbol{A}^T) = N(\boldsymbol{A})^{\perp},\ N(\boldsymbol{A}) = C(\boldsymbol{A}^T)^{\perp}\ .$$

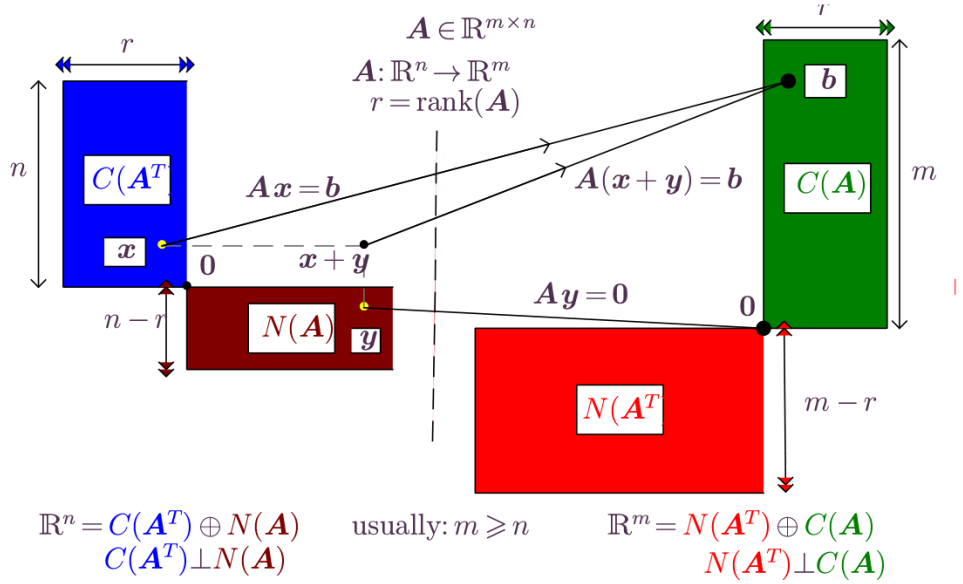

Figure 1. Graphical represenation of the Fundamental Theorem of Linear Algebra, Gil Strang, *Amer. Math. Monthly* **100**, 848-855, 1993.

Consideration of equality between sets arises in proving the above theorem. A standard technique to show set equality $A = B$, is by double inclusion, $A \subseteq B \wedge B \subseteq A \Rightarrow A = B$. This is shown for the statements giving the decomposition of the codomain $\mathbb{R}^m$. A similar approach can be used to decomposition of $\mathbb{R}^n$.

i. $C(\boldsymbol{A}) \perp N(\boldsymbol{A}^T)$ (column space is orthogonal to left null space).

Proof. Consider arbitrary $\boldsymbol{u} \in C(\boldsymbol{A}), \boldsymbol{v} \in N(\boldsymbol{A}^T)$. By definition of $C(\boldsymbol{A})$, $\exists \boldsymbol{x} \in \mathbb{R}^n$ such that $\boldsymbol{u} = \boldsymbol{A}\boldsymbol{x}$, and by definition of $N(\boldsymbol{A}^T)$, $\boldsymbol{A}^T\boldsymbol{v} = \boldsymbol{0}$. Compute $\boldsymbol{u}^T\boldsymbol{v} = (\boldsymbol{A}\boldsymbol{x})^T\boldsymbol{v} = \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{v} = \boldsymbol{x}^T(\boldsymbol{A}^T\boldsymbol{v}) = \boldsymbol{x}^T\boldsymbol{0} = 0$, hence $\boldsymbol{u} \perp \boldsymbol{v}$ for arbitrary $\boldsymbol{u}$, $\boldsymbol{v}$, and $C(\boldsymbol{A}) \perp N(\boldsymbol{A}^T)$. □

ii. $C(\boldsymbol{A}) \cap N(\boldsymbol{A}^T) = \{\boldsymbol{0}\}$ ($\boldsymbol{0}$ is the only vector both in $C(\boldsymbol{A})$ and $N(\boldsymbol{A}^T)$).

Proof. (By contradiction, *reductio ad absurdum*). Assume there might be $\boldsymbol{b} \in C(\boldsymbol{A})$ and $b \in N(\boldsymbol{A}^T)$ and $\boldsymbol{b} \neq \boldsymbol{0}$. Since $\boldsymbol{b} \in C(\boldsymbol{A})$, $\exists \boldsymbol{x} \in \mathbb{R}^n$ such that $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}$. Since $\boldsymbol{b} \in N(\boldsymbol{A}^T)$, $\boldsymbol{A}^T\boldsymbol{b} = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{0}$. Note that $\boldsymbol{x} \neq 0$ since $\boldsymbol{x} = 0 \Rightarrow \boldsymbol{b} = 0$, contradicting assumptions. Multiply equality $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ on left by $\boldsymbol{x}^T$,

$$\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \Rightarrow (\boldsymbol{A}\boldsymbol{x})^T(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{b}^T\boldsymbol{b} = \|\boldsymbol{b}\|^2 = 0,$$

thereby obtaining $\boldsymbol{b} = 0$, using norm property 3. Contradiction.

□

iii. $C(\boldsymbol{A}) \oplus N(\boldsymbol{A}^T) = \mathbb{R}^m$

Proof. (iii) and (iv) have established that $C(\boldsymbol{A}), N(\boldsymbol{A}^T)$ are orthogonal complements

$$C(\boldsymbol{A}) = N(\boldsymbol{A}^T)^{\perp}, N(\boldsymbol{A}^T) = C(\boldsymbol{A})^{\perp}.$$

By Lemma 2 it results that $C(\boldsymbol{A}) \oplus N(\boldsymbol{A}^T) = \mathbb{R}^m$. □

The remainder of the FTLA is established by considering $\boldsymbol{B} = \boldsymbol{A}^T$, e.g., since it has been established in (v) that $C(\boldsymbol{B}) \oplus N(\boldsymbol{A}^T) = \mathbb{R}^n$, replacing $\boldsymbol{B} = \boldsymbol{A}^T$ yields $C(\boldsymbol{A}^T) \oplus N(\boldsymbol{A}) = \mathbb{R}^m$, etc.

DATA PARTITIONING

1. Mappings as data

1.1. Vector spaces of mappings and matrix representations

A vector space $\mathscr{L}$ can be formed from all linear mappings from the vector space $\mathscr{U} = (U, S, +, \cdot)$ to another vector space $\mathscr{V} = (V, S, +, \cdot)$

$$\mathscr{L} = \{L, S, +, \cdot\}, \;\; L = \{\boldsymbol{f} | \boldsymbol{f} : U \to V, \boldsymbol{f}(a\boldsymbol{u} + b\boldsymbol{v}) = af(\boldsymbol{u}) + bf(\boldsymbol{v})\},$$

with addition and scaling of linear mappings defined by $(\boldsymbol{f} + \boldsymbol{g})(\boldsymbol{u}) = \boldsymbol{f}(\boldsymbol{u}) + \boldsymbol{g}(\boldsymbol{u})$ and $(a\boldsymbol{f})(\boldsymbol{u}) = a\boldsymbol{f}(\boldsymbol{u})$. Let $B = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots\}$ denote a basis for the domain U of linear mappings within $\mathscr{L}$, such that the linear mapping $\boldsymbol{f} \in \mathscr{L}$ is represented by the matrix

$$\boldsymbol{A} = [\, \boldsymbol{f}(\boldsymbol{u}_1) \;\; \boldsymbol{f}(\boldsymbol{u}_2) \;\; \ldots \,].$$

When the domain and codomain are the real vector spaces $U = \mathbb{R}^n$, $V = \mathbb{R}^m$, the above is a standard matrix of real numbers, $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. For linear mappings between infinite dimensional vector spaces the matrix is understood in a generalized sense to contain an infinite number of columns that are elements of the codomain V. For example, the indefinite integral is a linear mapping between the vector space of functions that allow differentiation to any order,

$$\int : \mathscr{C}^\infty \to \mathscr{C}^\infty \;\; v(x) = \int u(x)\, \mathrm{d}x$$

and for the monomial basis $B = \{1, x, x^2, \ldots\}$, is represented by the generalized matrix

$$\boldsymbol{A} = \left[\; x \;\; \tfrac{1}{2}x^2 \;\; \tfrac{1}{3}x^3 \;\; \ldots \; \right].$$

Truncation of the basis expansion $u(x) = \sum_{j=1}^{\infty} u_j x^j$ where $u_j \in \mathbb{R}$ to n terms, and sampling of $u \in \mathscr{C}^\infty$ at points $x_1, \ldots, x_m$, forms a standard matrix of real numbers

$$\boldsymbol{A} = \left[\; \boldsymbol{x} \;\; \tfrac{1}{2}\boldsymbol{x}^2 \;\; \tfrac{1}{3}\boldsymbol{x}^3 \;\; \ldots \; \right] \in \mathbb{R}^{m \times n}, \;\; \boldsymbol{x}^j = \begin{bmatrix} x_1^j \\ \vdots \\ x_m^j \end{bmatrix}.$$

As to be expected, matrices can also be organized as vector space $\mathscr{M}$, which is essentially the representation of the associated vector space of linear mappings,

$$\mathscr{M} = (M, S, +, \cdot) \;\; M = \{\boldsymbol{A} | \boldsymbol{A} = [\, \boldsymbol{f}(\boldsymbol{u}_1) \;\; \boldsymbol{f}(\boldsymbol{u}_2) \;\; \ldots \,]\}\,.$$

The addition $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$ and scaling $\boldsymbol{S} = a\boldsymbol{R}$ of matrices is given in terms of the matrix components by

$$c_{ij} = a_{ij} + b_{ij}, s_{ij} = a r_{\mathrm{ij}}\,.$$

1.2. Measurement of mappings

From the above it is apparent that linear mappings and matrices can also be considered as data, and a first step in analysis of such data is definition of functionals that would attach a single scalar label to each linear mapping of matrix. Of particular interest is the definition of a norm functional that characterizes in an appropriate sense the size of a linear mapping.

Consider first the case of finite matrices with real components $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ that represent linear mappings between real vector spaces $\boldsymbol{f}\colon \mathbb{R}^m \to \mathbb{R}^n$. The columns $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ could be placed into a single column vector $\boldsymbol{c}$ with mn components

$$\boldsymbol{c} = \begin{bmatrix} \boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix}.$$

Subsequently the norm of the matrix $\boldsymbol{A}$ could be defined as the norm of the vector $\boldsymbol{c}$. An example of this approach is the Frobenius norm

$$\|\boldsymbol{A}\|_F = \|\boldsymbol{c}\|_2 = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}.$$

A drawback of the above approach is that the structure of the matrix and its close relationship to a linear mapping is lost. A more useful characterization of the size of a mapping is to consider the amplification behavior of linear mapping. The motivation is readily understood starting from linear mappings between the reals $f\colon \mathbb{R} \to \mathbb{R}$, that are of the form $f(x) = ax$. When given an argument of unit magnitude $|x| = 1$, the mapping returns a real number with magnitude $|a|$. For mappings $\boldsymbol{f}\colon \mathbb{R}^2 \to \mathbb{R}^2$ within the plane, arguments that satisfy $\|\boldsymbol{x}\|_2 = 1$ are on the unit circle with components $\boldsymbol{x} = [\ \cos\theta \ \ \sin\theta\ ]$ have images through $\boldsymbol{f}$ given analytically by

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} = [\ \boldsymbol{a}_1 \ \ \boldsymbol{a}_2\ ] \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} = \cos\theta\, \boldsymbol{a}_1 + \sin\theta\, \boldsymbol{a}_2,$$

and correspond to ellipses.

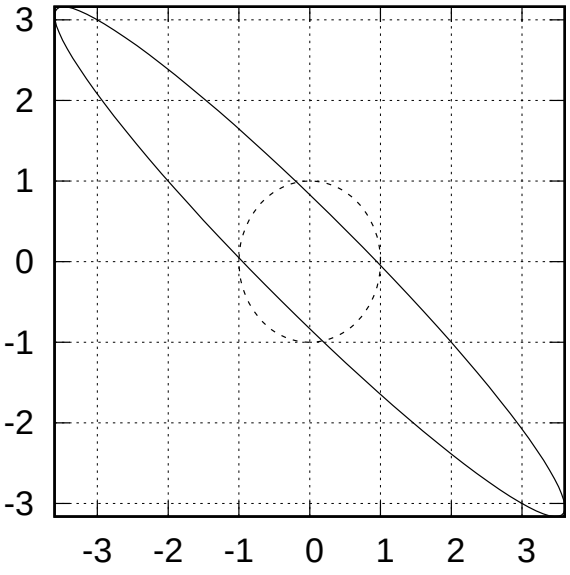

Figure 1. Mapping of unit circle by $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{A} = \begin{bmatrix} 2 & 3 \\ -1 & -3 \end{bmatrix}$.

From the above the mapping associated $\boldsymbol{A}$ amplifies some directions more than others. This suggests a definition of the size of a matrix or a mapping by the maximal amplification unit norm vectors within the domain.

DEFINITION. *For vector spaces U, V with norms $\|\ \|_U\colon U \to \mathbb{R}_+$, $\|\ \|_V\colon V \to \mathbb{R}_+$, the induced norm of $\boldsymbol{f}\colon U \to V$ is*

$$\|\boldsymbol{f}\| = \sup_{\|\boldsymbol{x}\|_U = 1} \|\boldsymbol{f}(\boldsymbol{x})\|_V.$$

DEFINITION. *For vector spaces $\mathbb{R}^n, \mathbb{R}^m$ with norms $\|\ \|^{(n)}\colon U \to \mathbb{R}_+$, $\|\ \|^{(m)}\colon V \to \mathbb{R}_+$, the induced norm of matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is*

$$\|\boldsymbol{A}\| = \sup_{\|\boldsymbol{x}\|^{(n)} = 1} \|\boldsymbol{A}\boldsymbol{x}\|^{(m)}.$$

In the above, any vector norm can be used within the domain and codomain.

2. The Singular Value Decomposition (SVD)

The fundamental theorem of linear algebra partitions the domain and codomain of a linear mapping $\boldsymbol{f}: U \to V$. For real vectors spaces $U = \mathbb{R}^n$, $V = \mathbb{R}^m$ the partition properties are stated in terms of spaces of the associated matrix $\boldsymbol{A}$ as

$$C(\boldsymbol{A}) \oplus N(\boldsymbol{A}^T) = \mathbb{R}^m \;\; C(\boldsymbol{A}) \perp N(\boldsymbol{A}^T) \;\; C(\boldsymbol{A}^T) \oplus N(\boldsymbol{A}) = \mathbb{R}^n \;\; C(\boldsymbol{A}^T) \perp N(\boldsymbol{A}) \; .$$

The dimension of the column and row spaces $r = \dim C(\boldsymbol{A}) = \dim C(\boldsymbol{A}^T)$ is the rank of the matrix, $n - r$ is the nullity of $\boldsymbol{A}$, and $m - r$ is the nullity of A^T. A infinite number of bases could be defined for the domain and codomain. It is of great theoretical and practical interest bases with properties that faciliatate insight or computation.

2.1. Orthogonal matrices

The above partitions of the domain and codomain are orthogonal, and suggest searching for orthogonal bases within these subspaces. Introduce a matrix representation for the bases

$$\boldsymbol{U} = [\; \boldsymbol{u}_1 \;\; \boldsymbol{u}_2 \;\; \ldots \;\; \boldsymbol{u}_m \;] \in \mathbb{R}^{m \times m}, \boldsymbol{V} = [\; \boldsymbol{v}_1 \;\; \boldsymbol{v}_2 \;\; \ldots \;\; \boldsymbol{v}_n \;] \in \mathbb{R}^{n \times n},$$

with $C(\boldsymbol{U}) = \mathbb{R}^m$ and $C(\boldsymbol{V}) = \mathbb{R}^n$. Orthogonality between columns $\boldsymbol{u}_i, \boldsymbol{u}_j$ for $i \neq j$ is expressed as $\boldsymbol{u}_i^T \boldsymbol{u}_j = 0$. For $i = j$, the inner product is positive $\boldsymbol{u}_i^T \boldsymbol{u}_i > 0$, and since scaling of the columns of $\boldsymbol{U}$ preserves the spanning property $C(\boldsymbol{U}) = \mathbb{R}^m$, it is convenient to impose $\boldsymbol{u}_i^T \boldsymbol{u}_i = 1$. Such behavior is concisely expressed as a matrix product

$$\boldsymbol{U}^T \boldsymbol{U} = \boldsymbol{I}_m,$$

with $\boldsymbol{I}_m$ the identity matrix in $\mathbb{R}^m$. Expanded in terms of the column vectors of $\boldsymbol{U}$ the first equality is

$$[\; \boldsymbol{u}_1 \;\; \boldsymbol{u}_2 \;\; \ldots \;\; \boldsymbol{u}_m \;]^T [\; \boldsymbol{u}_1 \;\; \boldsymbol{u}_2 \;\; \ldots \;\; \boldsymbol{u}_m \;] = \begin{bmatrix} \boldsymbol{u}_1^T \\ \boldsymbol{u}_2^T \\ \vdots \\ \boldsymbol{u}_m^T \end{bmatrix} [\; \boldsymbol{u}_1 \;\; \boldsymbol{u}_2 \;\; \ldots \;\; \boldsymbol{u}_m \;] = \begin{bmatrix} \boldsymbol{u}_1^T \boldsymbol{u}_1 & \boldsymbol{u}_1^T \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_1^T \boldsymbol{u}_m \\ \boldsymbol{u}_2^T \boldsymbol{u}_1 & \boldsymbol{u}_2^T \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_2^T \boldsymbol{u}_m \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{u}_m^T \boldsymbol{u}_1 & \boldsymbol{u}_m^T \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_m^T \boldsymbol{u}_m \end{bmatrix} = \boldsymbol{I}_m.$$

It is useful to determine if a matrix $\boldsymbol{X}$ exists such that $\boldsymbol{U}\boldsymbol{X} = \boldsymbol{I}_m$, or

$$\boldsymbol{U}\boldsymbol{X} = \boldsymbol{U} [\; \boldsymbol{x}_1 \;\; \boldsymbol{x}_2 \;\; \ldots \;\; \boldsymbol{x}_m \;] = [\; \boldsymbol{e}_1 \;\; \boldsymbol{e}_2 \;\; \ldots \;\; \boldsymbol{e}_m \;].$$

The columns of $\boldsymbol{X}$ are the coordinates of the column vectors of $\boldsymbol{I}_m$ in the basis $\boldsymbol{U}$, and can readily be determined

$$\boldsymbol{U}\boldsymbol{x}_j = \boldsymbol{e}_j \Rightarrow \boldsymbol{U}^T \boldsymbol{U}\boldsymbol{x}_j = \boldsymbol{U}^T \boldsymbol{e}_j \Rightarrow \boldsymbol{I}_m \boldsymbol{x}_j = \begin{bmatrix} \boldsymbol{u}_1^T \\ \boldsymbol{u}_2^T \\ \vdots \\ \boldsymbol{u}_m^T \end{bmatrix} \boldsymbol{e}_j \Rightarrow \boldsymbol{x}_j = (\boldsymbol{U}^T)_j,$$

where $(\boldsymbol{U}^T)_j$ is the j^{th} column of $\boldsymbol{U}^T$, hence $\boldsymbol{X} = \boldsymbol{U}^T$, leading to

$$\boldsymbol{U}^T \boldsymbol{U} = \boldsymbol{I} = \boldsymbol{U}\boldsymbol{U}^T.$$

Note that the second equality

$$[\; \boldsymbol{u}_1 \;\; \boldsymbol{u}_2 \;\; \ldots \;\; \boldsymbol{u}_m \;][\; \boldsymbol{u}_1 \;\; \boldsymbol{u}_2 \;\; \ldots \;\; \boldsymbol{u}_m \;]^T = [\; \boldsymbol{u}_1 \;\; \boldsymbol{u}_2 \;\; \ldots \;\; \boldsymbol{u}_m \;] \begin{bmatrix} \boldsymbol{u}_1^T \\ \boldsymbol{u}_2^T \\ \vdots \\ \boldsymbol{u}_m^T \end{bmatrix} = \boldsymbol{u}_1 \boldsymbol{u}_1^T + \boldsymbol{u}_2 \boldsymbol{u}_2^T + \cdots + \boldsymbol{u}_m \boldsymbol{u}_m^T = \boldsymbol{I}$$

acts as normalization condition on the matrices $\boldsymbol{U}_j = \boldsymbol{u}_j \boldsymbol{u}_j^T$.

DEFINITION. *A square matrix $\boldsymbol{U}$ is said to be orthogonal if $\boldsymbol{U}^T \boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{I}$.*

2.2. Intrinsic basis of a linear mapping

Given a linear mapping $\boldsymbol{f}: U \to V$, expressed as $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$, the simplest description of the action of $\boldsymbol{A}$ would be a simple scaling, as exemplified by $\boldsymbol{g}(\boldsymbol{x}) = a\boldsymbol{x}$ that has as its associated matrix $a\boldsymbol{I}$. Recall that specification of a vector is typically done in terms of the identity matrix $\boldsymbol{b} = \boldsymbol{I}\boldsymbol{b}$, but may be more insightfully given in some other basis $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{I}\boldsymbol{b}$. This suggests that especially useful bases for the domain and codomain would reduce the action of a linear mapping to scaling along orthogonal directions, and evaluate $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ by first re-expressing $\boldsymbol{y}$ in another basis $\boldsymbol{U}$, $\boldsymbol{U}\boldsymbol{s} = \boldsymbol{I}\boldsymbol{y}$ and re-expressing $\boldsymbol{x}$ in another basis $\boldsymbol{V}$, $\boldsymbol{V}\boldsymbol{r} = \boldsymbol{I}\boldsymbol{x}$. The condition that the linear operator reduces to simple scaling in these new bases is expressed as $s_i = \sigma_i r_i$ for $i = 1, \ldots, \min(m, n)$, with σ_i the scaling coefficients along each direction which can be expressed as a matrix vector product $\boldsymbol{s} = \boldsymbol{\Sigma}\boldsymbol{r}$, where $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times n}$ is of the same dimensions as $\boldsymbol{A}$ and given by

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & \sigma_2 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_r & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \end{bmatrix}.$$

Imposing the condition that $\boldsymbol{U}, \boldsymbol{V}$ are orthogonal leads to

$$\boldsymbol{U}\boldsymbol{s} = \boldsymbol{y} \Rightarrow \boldsymbol{s} = \boldsymbol{U}^T\boldsymbol{y}, \boldsymbol{V}\boldsymbol{r} = \boldsymbol{x} \Rightarrow \boldsymbol{r} = \boldsymbol{V}^T\boldsymbol{x},$$

which can be replaced into $\boldsymbol{s} = \boldsymbol{\Sigma}\boldsymbol{r}$ to obtain

$$\boldsymbol{U}^T\boldsymbol{y} = \boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{x} \Rightarrow \boldsymbol{y} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{x}.$$

From the above the orthogonal bases $\boldsymbol{U}, \boldsymbol{V}$ and scaling coefficients $\boldsymbol{\Sigma}$ that are sought must satisfy $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$.

THEOREM. *Every matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ has a singular value decomposition (SVD)*

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T,$$

with properties:

1. *$\boldsymbol{U} \in \mathbb{R}^{m \times m}$ is an orthogonal matrix, $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_m$;*
2. *$\boldsymbol{V} \in \mathbb{R}^{m \times m}$ is an orthogonal matrix, $\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}_n$;*
3. *$\boldsymbol{\Sigma} \in \mathbb{R}^{m \times n}$ is diagonal, $\boldsymbol{\Sigma} = \text{diag}(\sigma_1, \ldots, \sigma_p)$, $p = \min(m, n)$, and $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_p \geqslant 0$.*

Proof. *The proof of the SVD makes use of properties of the norm, concepts from analysis and complete induction. Adopting the 2-norm set $\sigma_1 = \|A\|_2$,*

$$\sigma_1 = \sup_{\|\boldsymbol{x}\|_2 = 1} \|\boldsymbol{A}\boldsymbol{x}\|_2.$$

The domain $\|\boldsymbol{x}\|_2 = 1$ is compact (closed and bounded), and the extreme value theorem implies that $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ attains its maxima and minima, hence there must exist some vectors $\boldsymbol{u}_1, \boldsymbol{v}_1$ of unit norm such that $\sigma_1\boldsymbol{u}_1 = \boldsymbol{A}\boldsymbol{v}_1 \Rightarrow \sigma_1 = \boldsymbol{u}_1^T\boldsymbol{A}\boldsymbol{v}_1$. Introduce orthogonal bases $\boldsymbol{U}_1$, $\boldsymbol{V}_1$ for $\mathbb{R}^m, \mathbb{R}^n$ whose first column vectors are $\boldsymbol{u}_1, \boldsymbol{v}_1$, and compute

$$\boldsymbol{U}_1{}^T\boldsymbol{A}\boldsymbol{V}_1 = \begin{bmatrix} \boldsymbol{u}_1^T \\ \vdots \\ \boldsymbol{u}_m^T \end{bmatrix} [\, \boldsymbol{A}\boldsymbol{v}_1 \;\; \ldots \;\; \boldsymbol{A}\boldsymbol{v}_n \,] = \begin{bmatrix} \sigma_1 & \boldsymbol{w}^T \\ \boldsymbol{0} & \boldsymbol{B} \end{bmatrix} = \boldsymbol{C}.$$

In the above $\boldsymbol{w}^T$ is a row vector with $n-1$ components $\boldsymbol{u}_1^T\boldsymbol{A}\boldsymbol{v}_j$, $j = 2, \ldots, n$, and $\boldsymbol{u}_i^T\boldsymbol{A}\boldsymbol{v}_1$ must be zero for $\boldsymbol{u}_1$ to be the direction along which the maximum norm $\|\boldsymbol{A}\boldsymbol{v}_1\|$ is obtained. Introduce vectors

$$\boldsymbol{y} = \begin{bmatrix} \sigma_1 \\ \boldsymbol{w} \end{bmatrix}, \boldsymbol{z} = \boldsymbol{C}\boldsymbol{y} = \begin{bmatrix} \sigma_1^2 + \boldsymbol{w}^T\boldsymbol{w} \\ \boldsymbol{B}\boldsymbol{w} \end{bmatrix},$$

and note that $\|\boldsymbol{z}\|_2 \geqslant \|\boldsymbol{y}\|_2^2 = \sigma_1^2 + \boldsymbol{w}^T\boldsymbol{w}$. From $\|\boldsymbol{U}_1{}^T\boldsymbol{A}\boldsymbol{V}_1\| = \|\boldsymbol{A}\| = \sigma_1 = \|\boldsymbol{C}\| \geqslant \sigma_1^2 + \boldsymbol{w}^T\boldsymbol{w}$ it results that $\boldsymbol{w} = \boldsymbol{0}$. By induction, assume that $\boldsymbol{B}$ has a singular value decomposition, $\boldsymbol{B} = \boldsymbol{U}_2\boldsymbol{\Sigma}_2\boldsymbol{V}_2^T$, such that

$$\boldsymbol{U}_1{}^T\boldsymbol{A}\boldsymbol{V}_1 = \begin{bmatrix} \sigma_1 & \boldsymbol{0}^T \\ \boldsymbol{0} & \boldsymbol{U}_2\boldsymbol{\Sigma}_2\boldsymbol{V}_2^T \end{bmatrix} = \begin{bmatrix} 1 & \boldsymbol{0}^T \\ \boldsymbol{0} & \boldsymbol{U}_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & \boldsymbol{0}^T \\ \boldsymbol{0} & \boldsymbol{\Sigma}_2 \end{bmatrix} \begin{bmatrix} 1 & \boldsymbol{0}^T \\ \boldsymbol{0} & \boldsymbol{V}_2^T \end{bmatrix},$$

and the orthogonal matrices arising in the singular value decomposition of $\boldsymbol{A}$ *are*

$$\boldsymbol{U} = \boldsymbol{U}_1 \begin{bmatrix} 1 & \boldsymbol{0}^T \\ \boldsymbol{0} & \boldsymbol{U}_2 \end{bmatrix}, \boldsymbol{V}^T = \begin{bmatrix} 1 & \boldsymbol{0}^T \\ \boldsymbol{0} & \boldsymbol{V}_2^T \end{bmatrix} \boldsymbol{V}_1^T.$$

□

The scaling coefficients σ_j are called the *singular values* of $\boldsymbol{A}$. The columns of $\boldsymbol{U}$ are called the *left singular vectors*, and those of $\boldsymbol{V}$ are called the *right singular vectors*.
The fact that the scaling coefficients are norms of $\boldsymbol{A}$ and submatrices of $\boldsymbol{A}$, $\sigma_1 = \|\boldsymbol{A}\|$, is crucial importance in applications. Carrying out computation of the matrix products

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \dots & \boldsymbol{u}_r & \boldsymbol{u}_{r+1} & \dots & \boldsymbol{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_r & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1^T \\ \boldsymbol{v}_2^T \\ \vdots \\ \boldsymbol{v}_r^T \\ \vdots \\ \boldsymbol{v}_n^T \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \dots & \boldsymbol{u}_r & \boldsymbol{u}_{r+1} & \dots & \boldsymbol{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 \boldsymbol{v}_1^T \\ \sigma_2 \boldsymbol{v}_2^T \\ \vdots \\ \sigma_r \boldsymbol{v}_r^T \\ \vdots \\ 0 \end{bmatrix}$$

leads to a representation of $\boldsymbol{A}$ as a sum

$$\boldsymbol{A} = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T, r \leqslant \min(m, n).$$

Each product $\boldsymbol{u}_i \boldsymbol{v}_i^T$ is a matrix of rank one, and is called a rank-one update. Truncation of the above sum to p terms leads to an approximation of $\boldsymbol{A}$

$$\boldsymbol{A} \cong \boldsymbol{A}_p = \sum_{i=1}^{p} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T.$$

In very many cases the singular values exhibit rapid, exponential decay, $\sigma_1 \gg \sigma_2 \gg \cdots$, such that the approximation above is an accurate representation of the matrix $\boldsymbol{A}$.

Figure 2. Successive SVD approximations of Frida Kahlo's (1907-1954) painting, *Portrait of a Lady in White* (1929), with $k = 10, 20, 40$ rank-one updates.

2.3. SVD solution of linear algebra problems

The SVD can be used to solve common problems within linear algebra.

Change of coordinates. To change from vector coordinates $\boldsymbol{b}$ in the canonical basis $\boldsymbol{I} \in \mathbb{R}^{m \times m}$ to coordinates $\boldsymbol{x}$ in some other basis $\boldsymbol{A} \in \mathbb{R}^{m \times m}$, a solution to the equation $\boldsymbol{I}\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}$ can be found by the following steps.

1. Compute the SVD, $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T = \boldsymbol{A}$;

2. Find the coordinates of $\boldsymbol{b}$ in the orthogonal basis $\boldsymbol{U}$, $\boldsymbol{c} = \boldsymbol{U}^T \boldsymbol{b}$;
3. Scale the coordinates of $\boldsymbol{c}$ by the inverse of the singular values $y_i = c_i / \sigma_i$, $i = 1,...,m$, such that $\boldsymbol{\Sigma y} = \boldsymbol{c}$ is satisfied;
4. Find the coordinates of $\boldsymbol{y}$ in basis $\boldsymbol{V}^T$, $\boldsymbol{x} = \boldsymbol{V y}$.

Best 2-norm approximation. In the above $\boldsymbol{A}$ was assumed to be a basis, hence $r = \text{rank}(\boldsymbol{A}) = m$. If columns of $\boldsymbol{A}$ do not form a basis, $r < m$, then $\boldsymbol{b} \in \mathbb{R}^m$ might not be reachable by linear combinations within $C(\boldsymbol{A})$. The closest vector to $\boldsymbol{b}$ in the norm is however found by the same steps, with the simple modification that in Step 3, the scaling is carried out only for non-zero singular values, $y_i = c_i / \sigma_i$, $i = 1, \ldots, r$.

The pseudo-inverse. From the above, finding either the solution of $\boldsymbol{Ax} = \boldsymbol{Ib}$ or the best approximation possible if $\boldsymbol{A}$ is not of full rank, can be written as a sequence of matrix multiplications using the SVD

$$(\boldsymbol{U \Sigma V}^T)\boldsymbol{x} = \boldsymbol{b} \Rightarrow \boldsymbol{U}\,(\boldsymbol{\Sigma V}^T \boldsymbol{x}) = \boldsymbol{b} \Rightarrow (\boldsymbol{\Sigma V}^T \boldsymbol{x}) = \boldsymbol{U}^T \boldsymbol{b} \Rightarrow \boldsymbol{V}^T \boldsymbol{x} = \boldsymbol{\Sigma}^+ \boldsymbol{U}^T \boldsymbol{b} \Rightarrow \boldsymbol{x} = \boldsymbol{V \Sigma}^+ \boldsymbol{U}^T \boldsymbol{b},$$

where the matrix $\boldsymbol{\Sigma}^+ \in \mathbb{R}^{n \times m}$ (notice the inversion of dimensions) is defined as a matrix with elements σ_i^{-1} on the diagonal, and is called the pseudo-inverse of $\boldsymbol{\Sigma}$. Similarly the matrix

$$\boldsymbol{A}^+ = \boldsymbol{V \Sigma}^+ \boldsymbol{U}^T$$

that allows stating the solution of $\boldsymbol{Ax} = \boldsymbol{b}$ simply as $\boldsymbol{x} = \boldsymbol{A}^+ \boldsymbol{b}$ is called the *pseudo-inverse* of $\boldsymbol{A}$. Note that in practice $\boldsymbol{A}^+$ is not explicitly formed. Rather the notation $\boldsymbol{A}^+$ is simply a concise reference to carrying out steps 1-4 above.