Linear Algebra 2/2

Vector Subspace

Let $V = (\mathcal{V}, +, \cdot)$ be a vector space and $\mathcal{U} \subseteq \mathcal{V}, \mathcal{U} \neq \emptyset$.

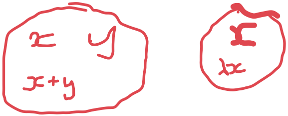
- Then $U = (\mathcal{U}, +, \cdot)$ is a vector subspace of V if U is vector space with the vector space operations $+$ and $\cdot$ restricted to $\mathcal{U} \times \mathcal{U}$ and $\mathbb{R} \times \mathcal{U}$.

The notation $U \subseteq V$ will be used to denote that U is a subspace of the vector space V.

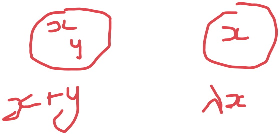
$(\mathcal{U}, +, \cdot)$ is a subspace of V if and only if

1. $\mathcal{U} \neq \emptyset$, particularly $\mathbf{0} \in \mathcal{U}$
2. Closure of U
   - (outer operation): $\forall \lambda \in \mathbb{R}, \forall \mathbf{x} \in \mathcal{U} : \lambda \mathbf{x} \in \mathcal{U}$
   - (inner operation): $\forall \mathbf{x}, \mathbf{y} \in \mathcal{U} : \mathbf{x} + \mathbf{y} \in \mathcal{U}$

Closure

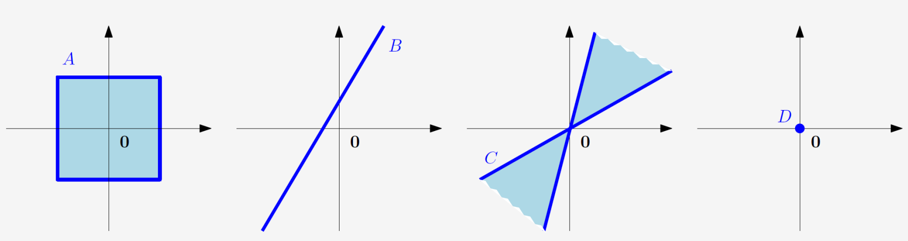


Not Closure



Vector Subspaces

Of the four figure below, which is a subspace of $\mathbb{R}^2$? $V = (\mathcal{V}; + ; \cdot)$



Source: M.P. Deisenroth *et al*, Mathematics for Machine Learning (First Edition)

$\bar{0}$

Closure $+$ $\cdot$

An interesting example is that

- The solution set of a homogeneous system of linear equations $\mathbf{Ax} = \mathbf{0}$, with n unknowns is a subspace of $\mathbb{R}^n$
  - Can you show this?
  - Recall you just need to verify that $\mathcal{U} \neq \emptyset$ and the two closures!
- The intersection of arbitrarily many subspaces is a subspace itself.
  - Can you show this?
  - Try for $\underline{\bigcap_{i=1}^{n} V_n}$ when V_i is a subspace of the vector space V and each contain the subset $\mathcal{V}_i$ and the same inner and outer operations.

$V_1 \cup V_2$ is not a subspace

Linear Independence

Linear Combination

Consider the vector space V and a finite number of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k \in V$.

- Then all $\mathbf{v} \in V$ that we can construct as

$$\mathbf{v} = \lambda_1 \mathbf{x}_1 + \cdots + \lambda_k \mathbf{x}_k = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i \tag{1}$$

  where all $\lambda_i \in \mathbb{R}$, is a *linear combination* of the vector $\mathbf{x}_1, \ldots, \mathbf{x}_k$

Linear Independence

Linear (In)dependence

Consider the vector space V and a finite number of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k \in V$.

- If there is a **non-trivial** linear combination, such that

$$\mathbf{0} = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i \tag{2}$$

  then the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are *linearly dependent*.
- If only the **trivial** solution exists the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are *linearly independent*.

A **trivial** linear combination is one where all $\lambda_i = 0$

Linear Independence

One of the main intuitions to draw from a set of linear dependent vectors is that

- from the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ you can construct **at least one** $\mathbf{x}_j$ from the others, namely there exists λ_is such that

$$\mathbf{x_j} = \sum_{i=1, i \neq j}^{k} \lambda_i \mathbf{x}_i \tag{3}$$

- This fact will be revisited later one but keep it in your mind as we proceed.

Linear Independence

Checking for linear independence or dependence of a set of vectors

- If $\mathbf{0} \in S$ then S is linearly dependent.
- If the elements of S are not unique then S is linearly dependent
- A more general complete test is:
  - With $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are all the elements of S
  - Construct the matrix $\mathbf{A}$ by using each vector $\mathbf{x}_j$ as a column vector of $\mathbf{A}$.
  - Get the matrix $\mathbf{A}$ into reduced row echelon (for example using Gaussian elimination), let's call the resultant matrix $\hat{\mathbf{A}}$
  - All column vectors of $\hat{\mathbf{A}}$ are **linearly independent** if and only if all columns are pivot columns
    - ★ If there is at least one non-pivot column, the columns are **linearly dependent**

Linear Independence:

For example

$$\hat{\mathbf{A}} = \begin{bmatrix} 1 & 2 & 0 & -1 \\ 0 & 0 & 0 & -15 \\ 0 & 0 & 1 & -18 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

has two no pivot columns and therefore it's columns vector as linearly dependent.

Linear Independence:

Use this approach to check if

$$\begin{bmatrix}1\\2\\0\end{bmatrix}, \quad \begin{bmatrix}1\\1\\-1\end{bmatrix}, \quad \begin{bmatrix}1\\4\\2\end{bmatrix}$$

are linear independent or dependent at home.

Linear Independence: Moving toward transformations

Consider the following slightly more complex set up.

- You have k linearly independent vectors $\mathbf{b}_1, \ldots, \mathbf{b}_k$ from V a vector space.
- You also have m vectors constructed using the k linearly independent vectors

$$\mathbf{x}_1 = \sum_{i=1}^{k} \lambda_{i1} \mathbf{b}_i$$

$$\vdots$$

$$\mathbf{x}_m = \sum_{i=1}^{k} \lambda_{im} \mathbf{b}_i$$

- What is the relationship between the linear independence of $\mathbf{x}_1, \ldots, \mathbf{x}_m$ and all the λ_{ij}s?

Linear Independence: Moving toward transformations

Let $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_k]$, we can then re-write each $\mathbf{x}_j$ as

$$\mathbf{x}_j = \mathbf{B}\boldsymbol{\lambda}_j, \ \boldsymbol{\lambda}_j = \begin{bmatrix} \lambda_{1j} \\ \vdots \\ \lambda_{kj} \end{bmatrix}, \ j = 1, \ldots, m \tag{4}$$

- We can now test if $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ are linearly independent by checking if we can solve $\sum_{j=1}^{m} \psi_j \mathbf{x}_j = \mathbf{0}$ with a nontrivial solution.
- Using equation 4 we obtain

$\psi B\lambda_j = B\psi\lambda$

$B\lambda_1\psi_1 + B\lambda_2\psi_2 = B(\lambda_1\psi_1 + \lambda_2\psi_2)$

$$\sum_{j=1}^{m} \psi_j \mathbf{x}_j = \sum_{j=1}^{m} \psi_j \mathbf{B}\boldsymbol{\lambda}_j = \mathbf{B}\sum_{j=1}^{m} \psi_j \boldsymbol{\lambda}_j \tag{5}$$

This means that $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ are linearly independent if and only if the column vectors are $\{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m\}$ linearly independent.

Generating Set and Basis

Generating Set and Span

Consider a vector space $V = (\mathcal{V}, +, \cdot)$ and a set of vectors $\mathcal{A} = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \subseteq \mathcal{V}$.

- If every vector $\mathbf{v} \in \mathcal{V}$ can be expressed as a linear combination of $\mathbf{x}_1, \ldots, \mathbf{x}_k$, we call $\mathcal{A}$ a *generating set* of V.
- The set of all linear combinations of vectors in $\mathcal{A}$ is called the span of $\mathcal{A}$.
- If $\mathcal{A}$ spans the vector space V, we write the $V = span[\mathcal{A}]$ or equivalently $V = span[\mathbf{x}_1, \ldots, \mathbf{x}_k]$.

Generating Set and Basis

Basis

Consider a vector space $V = (\mathcal{V}, +, \cdot)$ and a set of vectors $\mathcal{A} = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \subseteq \mathcal{V}$.

- A generating set $\mathcal{A}$ of V is called *minimal* if there exists no smaller set $\hat{\mathcal{A}} \subset \mathcal{A} \subseteq \mathcal{V}$
- Every linearly independent generating set of V is minimal and is called a *basis* of V.

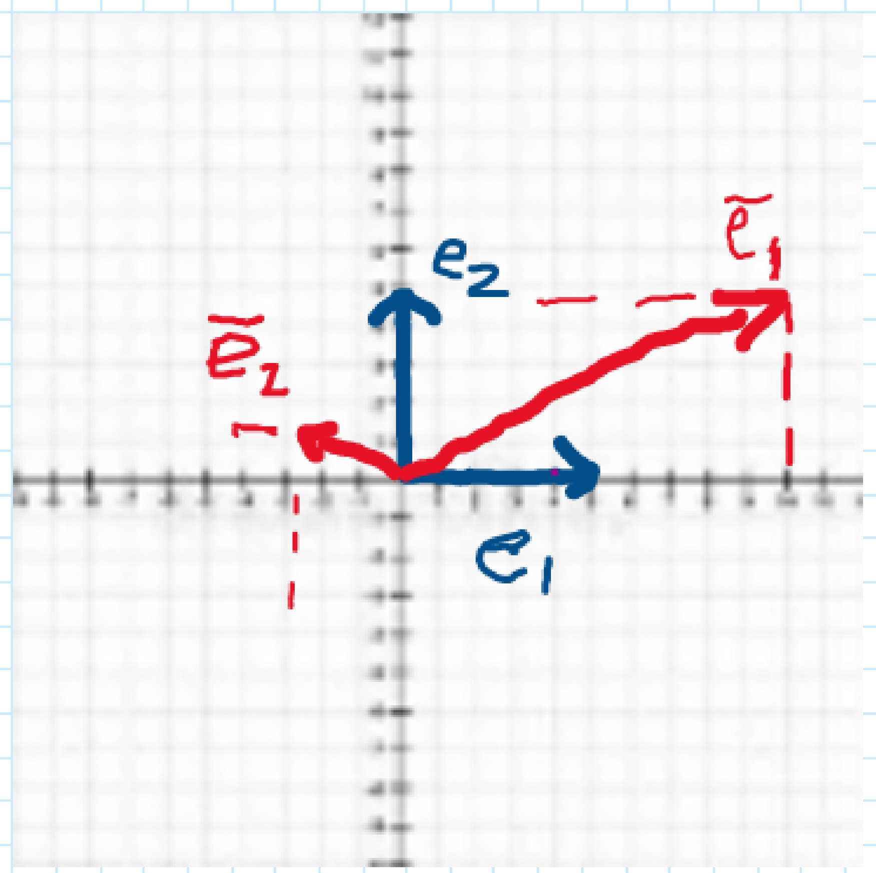
Generating Set and Basis: Properties

Let $V = (\mathcal{V}, +, \cdot)$ be a vector space and $\mathcal{B} \subseteq \mathcal{V}, \mathcal{B} \neq \emptyset$. Then, the following statements are equivalent:

1. $\mathcal{B}$ is a basis for V.
2. $\mathcal{B}$ is a minimal generating set.
3. $\mathcal{B}$ is a maximal linearly independent set of vectors in V
4. Every vector $\mathbf{v} \in V$ is a linear combination of vectors from $\mathcal{B}$, and every linear combination is unique.

Change of Basis

Thursday, March 3, 2022 11:22 PM



Old Basis $\{e_1; e_2\}$ Forward

New Basis $\{\tilde{e}_1; \tilde{e}_2\}$ Backward

$\tilde{e}_1 = 2e_1 + 1e_2$

$\tilde{e}_2 = -\frac{1}{2}e_1 + \frac{1}{4}e_2$

$F = \begin{bmatrix} 2 & -\frac{1}{2} \\ 1 & \frac{1}{4} \end{bmatrix}$

$[\tilde{e}_1 \ \tilde{e}_2] = [e_1 \ e_2]\begin{bmatrix} 2 & -\frac{1}{2} \\ 1 & \frac{1}{4} \end{bmatrix}$

$\tilde{e}_j = \sum_{k=1}^{n} F_{kj} e_k$

$e_1 = \frac{1}{4}\tilde{e}_1 + -1\tilde{e}_2$

$e_2 = \frac{1}{2}\tilde{e}_1 + 2\tilde{e}_2$

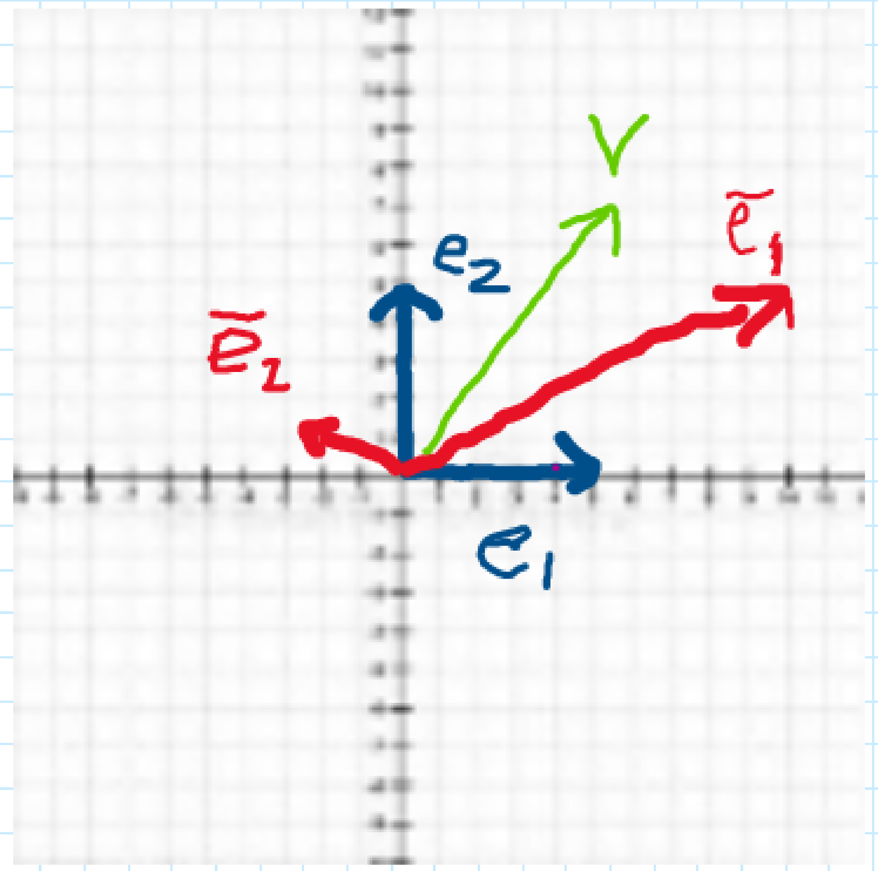
$B = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} \\ -1 & 2 \end{bmatrix}$

$[e_1 \ e_2] = [\tilde{e}_1 \ \tilde{e}_2]\begin{bmatrix} \frac{1}{4} & \frac{1}{2} \\ -1 & 2 \end{bmatrix}$

$e_i = \sum_{j=1}^{n} B_{ji} \tilde{e}_j$

$FB = \begin{bmatrix} 2 & 1 \\ -\frac{1}{2} & \frac{1}{4} \end{bmatrix}\begin{bmatrix} \frac{1}{4} & -1 \\ \frac{1}{2} & 2 \end{bmatrix}$

$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \therefore B = F^{-1}$



Old Basis $\{e_1; e_2\}$

New Basis $\{\tilde{e}_1; \tilde{e}_2\}$

$V = 1e_1 + 1\frac{1}{2}e_2 \quad \therefore V = \begin{bmatrix} 1 \\ 1\frac{1}{2} \end{bmatrix}_{e_i}$

$V = 1\tilde{e}_1 + 2\tilde{e}_2 \quad \therefore V = \begin{bmatrix} 1 \\ 2 \end{bmatrix}_{\tilde{e}_i}$

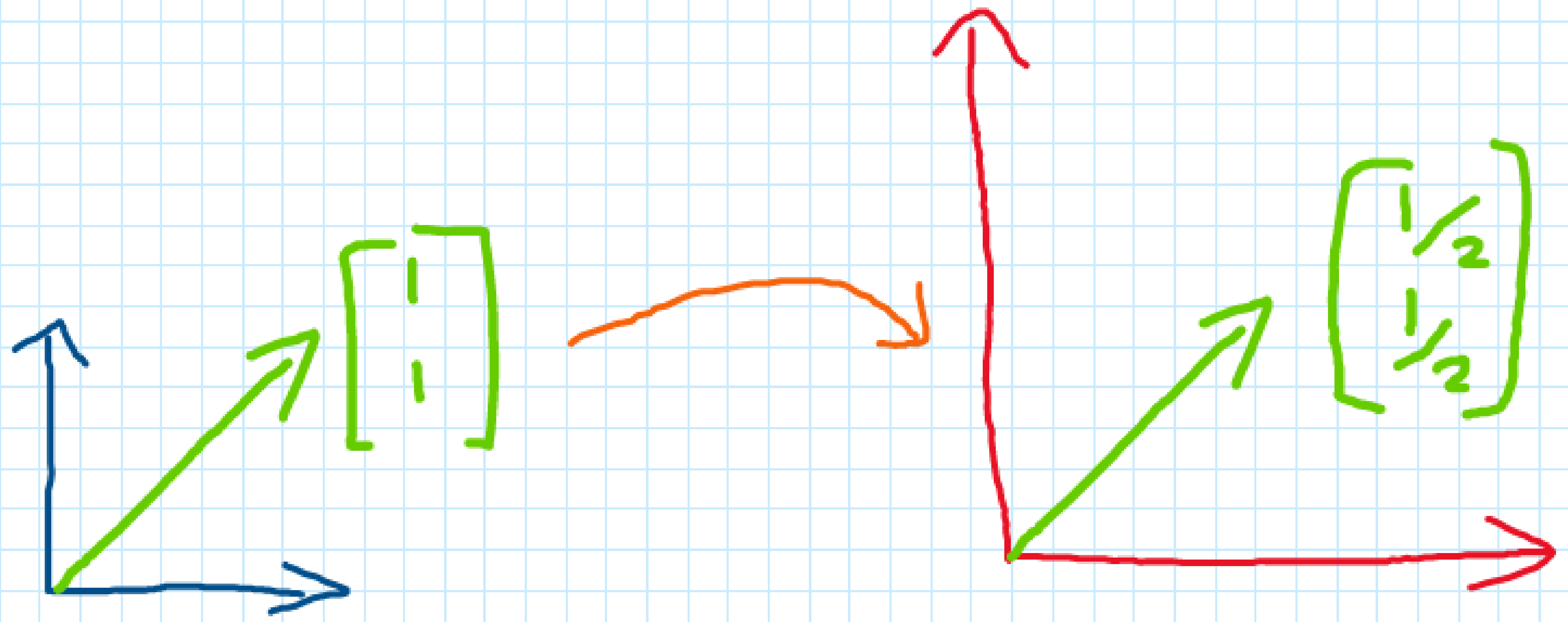
$V = [e_1 \ e_2]\begin{bmatrix} 1 \\ 1\frac{1}{2} \end{bmatrix}_{e_i} = [\tilde{e}_1 \ \tilde{e}_2]\begin{bmatrix} \frac{1}{4} & \frac{1}{2} \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 1\frac{1}{2} \end{bmatrix}_{e_i} = [\tilde{e}_1 \ \tilde{e}_2]\begin{bmatrix} 1 \\ 2 \end{bmatrix}_{\tilde{e}_i}$

$V = [\tilde{e}_1 \ \tilde{e}_2]\begin{bmatrix} 1 \\ 2 \end{bmatrix}_{\tilde{e}_i} = [e_1 \ e_2]\begin{bmatrix} 2 & -\frac{1}{2} \\ 1 & \frac{1}{4} \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix}_{\tilde{e}_i} = [e_1 \ e_2]\begin{bmatrix} 1 \\ 1\frac{1}{2} \end{bmatrix}_{e_i}$

Old Basis $\{e_1; e_2\}$ Forward

New Basis $\{\tilde{e}_1; \tilde{e}_2\}$ Backward

So coefficients transform opposite to the basis vectors!



We can speak about "1000 meters" or "1 kilometer"

Generating Set and Basis: Properties

$V = \mathbb{R}^3$

Every vector space V possesses a basis $\mathcal{B}$

- This basis is not necessarily unique!
- For example

$$\mathcal{B}_0 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}, \qquad \mathcal{B}_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

$$\mathcal{B}_2 = \left\{ \begin{bmatrix} 0.5 \\ 0.8 \\ 0.4 \end{bmatrix}, \begin{bmatrix} 1.8 \\ 0.30 \\ 0.3 \end{bmatrix}, \begin{bmatrix} -2.2 \\ -1.3 \\ 3.5 \end{bmatrix} \right\}$$

Are each a basis for $\mathbb{R}^3$.

Generating Set and Basis: Properties

In the case of a finite-dimensional vector spaces V.

- $dim(V)$ is the dimension of V, which is the number of basis vectors of V.
- If $U \subseteq V$ is a subspace of V, then $dim(U) \leq dim(V)$
  - Equality $dim(U) = dim(V)$, occurs if and only if $U = V$.

Determining a Basis

A basis of a subspace $U = span[\mathbf{x}_1, \ldots, \mathbf{x}_m] \subseteq \mathbb{R}^n$ can be found by executing the following steps:

1. Write the spanning vectors as columns of a matrix $\mathbf{A}$.
2. Determine the row-echelon form of $\mathbf{A}$.
3. The spanning vectors associated with the pivot columns are a basis of U.

Determining a Basis: Example

See example 2.17, but also try and determine a basis from the vectors below that spans $U \subseteq \mathbb{R}^4$

$$\mathbf{a} = \begin{bmatrix} 1 \\ 2 \\ -3 \\ 4 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 1 \\ -1 \\ 2 \\ 1 \end{bmatrix}, \mathbf{c} = \begin{bmatrix} 3 \\ 0 \\ 1 \\ 6 \end{bmatrix}, \mathbf{d} = \begin{bmatrix} 1 \\ 1 \\ -2 \\ 1 \end{bmatrix}, \mathbf{e} = \begin{bmatrix} 6 \\ -1 \\ 1 \\ 3 \end{bmatrix}$$

Think about the following:

- How do we tell if there is a unique basis derived from the above set?
- Can we have infinity many or no basis from the above set?

Rank

If $\mathbf{A} \in \mathbb{R}^{m \times n}$ then the rank of $\mathbf{A}$, denoted as $rk(\mathbf{A})$, is number of linearly independent columns of $\mathbf{A}$.

Importantly

- $rk(\mathbf{A}) = rk(\mathbf{A}^T)$, which means the rank of matrix $\mathbf{A}$ is also the number of linearly independent rows of $\mathbf{A}$.

Rank: Properties

The rank of a matrix has some useful properties:

- The columns of $\mathbf{A} \in \mathbb{R}^{m \times n}$ span a subspace $U \subseteq \mathbb{R}^m$ with $dim(U) = rk(A)$.
- The rows of $\mathbf{A} \in \mathbb{R}^{m \times n}$ span a subspace $W \subseteq \mathbb{R}^n$ with $dim(W) = rk(A)$.
- For all $\mathbf{A} \in \mathbb{R}^{n \times n}$ it holds that $\mathbf{A}$ is invertible if and only if $rk(\mathbf{A}) = n$.

$[A|b] \rightsquigarrow [I|A^{-1}b]$ only if n lin-indep vectors

Rank: Properties

The rank of a matrix has some useful properties:

- For all $\mathbf{A} \in \mathbb{R}^{m \times n}$ and all $\mathbf{b} \in \mathbb{R}^m$ it holds that the linear equation system $\mathbf{Ax} = \mathbf{b}$ can be solved if and only if $rk(\mathbf{A}) = rk(\mathbf{A}|\mathbf{b})$.
  - Think back to example 2.6 with $a \neq -1$
- For $\mathbf{A} \in \mathbb{R}^{m \times n}$ the subspace of solutions for $\mathbf{Ax} = \mathbf{0}$ possesses dimension $n - rk(A)$.
- The matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ has *full rank* if $rk(A) = min(m, n)$
  - If $rk(A) < min(m, n)$ we say that $\mathbf{A}$ *rank deficient*

Linear Mappings

Linear Mapping

For vector spaces V, W, a mapping $\Phi : V \to W$ is called a linear mapping if

$$\forall \mathbf{x}, \mathbf{y} \in V \; \forall \lambda, \psi \in \mathbb{R} :$$
$$\Phi(\lambda \mathbf{x} + \psi \mathbf{y}) = \lambda \Phi(\mathbf{x}) + \psi \Phi(\mathbf{y}) \tag{6}$$

- Also called a vector space homomorphism
- We can represent linear transformations as matrices! (when our vectors spaces are of finite dimension)

1 $\phi(\lambda x) = \lambda \phi(x)$

2 $\phi(x+y) = \phi(x)+\phi(y)$

Can show the properties separately

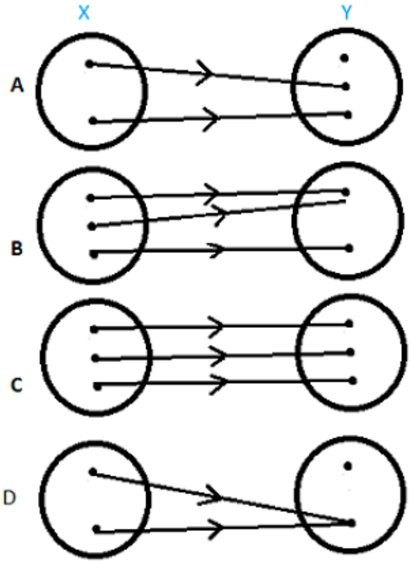
Special Fundamental Mappings

Injective, Surjective, Bijective

Consider a mapping $\Phi : \mathcal{V} \to \mathcal{W}$, where $\mathcal{V}$, $\mathcal{W}$ are arbitrary sets. Then Φ is called

- *Injective* if $\forall \mathbf{x}, \mathbf{y} \in \mathcal{V} :\ \Phi(\mathbf{x}) = \Phi(\mathbf{y}) \implies \mathbf{x} = \mathbf{y}$
- *Surjective* if $\Phi(V) = W$.
- *Bijective* if it is injective and surjective.

Special Fundamental Mappings



Injective

Surjective

invertibility

Surjective and Injective

Neither $\bar{0}x = 0$

Special Fundamental Mappings

A bijective Φ can be "undone", i.e., there exists a mapping Ψ so that $\Psi \circ \Phi$.

- This mapping Ψ is then called the inverse of Φ and normally denoted by Φ^{-1}.

Let V and W be vector spaces:

- Isomorphism: $\Phi : V \to W$ linear and bijective
- Endomorphism: $\Phi : V \to V$ linear
- Automorphism: $\Phi : V \to V$ linear and bijective
- We define $id_V : V \to V, \underbrace{\mathbf{x} \mapsto \mathbf{x}}_{x_B \to x_C}$ as the identity mapping or identity automorphism in V.

where B and C are different bases

Isomorphic Vector Spaces

Theorem 3.59 in Axler (2015)

Finite-dimensional vector spaces V and W are isomorphic if and only if $dim(V) = dim(W)$.

This theorem states that **there exists** a *linear*, *bijective* mapping between two vector spaces of the **same** dimension.

- It is possible to transform from V to W without incurring any loss!

Axler, Sheldon. 2015. Linear Algebra Done Right. Springer.

Composition and Inverse

If V, W, X are vector spaces then

- For linear mappings $\Phi : V \to W$ and $\Psi : W \to X$, the mapping $\Psi \circ \Phi : V \to X$ is also linear
- If $\Phi : V \to W$ is an isomorphism, then $\Phi^{-1} : W \to V$ is an isomorphism, too.
- If $\Phi : V \to W$, $\Psi : V \to W$ are linear, then $\Phi + \Psi$ and $\lambda\Phi, \lambda \in \mathbb{R}$, are linear, too

Matrix Representation of Linear Mappings: Coordinates

Coordinates

Consider a vector space V and an ordered basis $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ of V.

- For any $\mathbf{x} \in V$ we obtain a unique representation

$$\mathbf{x} = \alpha_1 \mathbf{b}_1 + \cdots + \alpha_n \mathbf{b}_n \tag{7}$$

- Then $\alpha_1, \ldots, \alpha_n$ are the *coordinates* of $\mathbf{x}$ with respect to the ordered basis B,
- and the vector

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{R}^n \tag{8}$$

  is the *coordinate vector/coordinate representation* of $\mathbf{x}$ with respect to the ordered basis B.

Matrix Representation of Linear Mappings: Coordinates

$e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$; $e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

The most familiar coordinate system is the *Cartesian coordinate system*

- Where, in the case of $\mathbb{R}^2$, is spanned by $\mathbf{e}_1$ and $\mathbf{e}_2$, and the ordered basis is therefore $B = (\mathbf{e}_1, \mathbf{e}_2)$
- If the vector $\mathbf{x}$ was represented as $[3, 1]^T$ using B how would it be represented using $B_1 = (\mathbf{e}_2, \mathbf{e}_1)$?
  - Simply as $[1, 3]^T$
  - There are many valid coordinate systems, we have just gotten most accustomed to the standard Cartesian coordinate system where the basis vectors are simply $\mathbf{e_j}$ and are arranged in from lowest index to the highest.

Matrix Representation of Linear Mappings: Coordinates

For an n-dimensional vector space V and an ordered basis B of V, the mapping

$$\Phi : \mathbb{R}^n \to V,$$
$$\Phi(\mathbf{e}_i) = \mathbf{b_i},\ i = 1, \ldots, n$$

is **linear**, where $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ is the standard basis of $\mathbb{R}^n$

- From THM Theorem 2.17 Φ is also an isomorphism.

In essence this implies that we can always find a mapping from each $\mathbf{e}_j$ to $\mathbf{b}_j$ and vice versa.

Matrix Representation of Linear Mappings

Transformation Matrix

Consider vector spaces V, W with corresponding (ordered) bases $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ and $C = (\mathbf{c}_1, \ldots, \mathbf{c}_m)$, and the linear mapping $\Phi : V \to W$

- For $j \in \{1, \ldots, n\}$,

$$\Phi(\mathbf{b}_j) = \alpha_{1j}\mathbf{c}_1 + \cdots + \alpha_{mj}\mathbf{c}_m = \sum_{i=1}^{m} \alpha_{ij}\mathbf{c}_i \tag{9}$$

  is the unique representation of $\Phi(\mathbf{b}_j)$ with respect to C.

- Then we call the $m \times n$-matrix $\mathbf{A}_\Phi$, whose elements are given by

$$A_\Phi(i, j) = \alpha_{ij} \tag{10}$$

  the transformation matrix of Φ with respect to the ordered bases B of V and and C of W.

Matrix Representation of Linear Mappings

The coordinates of $\Phi(\mathbf{b}_j)$ with respect to the ordered basis C of W are the j-th column of $\mathbf{A}_\Phi$.

Transforming

Consider (finite-dimensional) vector spaces V, W with ordered bases B, C and a linear mapping $\Phi : V \to W$ with transformation matrix $\mathbf{A}_\Phi$

- If $\hat{\mathbf{x}}$ is the coordinate vector of $\mathbf{x} \in V$ with respect to B and
- $\hat{\mathbf{y}}$ is the coordinate vector of $\mathbf{y} = \Phi(\mathbf{x}) \in W$ with respect to C, then

$$\hat{\mathbf{y}} = \mathbf{A}_\Phi \hat{\mathbf{x}} \tag{11}$$

This means that the transformation matrix can be used to map coordinates with respect to an ordered basis in V to coordinates with respect to an ordered basis in W.

Transformation example

Consider homomorphism $\Phi : V \to W$ and ordered bases $B = (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ of V and $C = (\mathbf{c}_1, \ldots, \mathbf{c}_4)$ of W. With

$$\begin{aligned} \Phi(\mathbf{b}_1) =& \mathbf{c}_1 - \mathbf{c}_2 + 3\mathbf{c}_3 - \mathbf{c}_4 \\ \Phi(\mathbf{b}_2) =& 2\mathbf{c}_1 + \mathbf{c}_2 + 7\mathbf{c}_3 + 2\mathbf{c}_4 \\ \Phi(\mathbf{b}_3) =& 3\mathbf{c}_2 + \mathbf{c}_3 + \mathbf{c}_4 \end{aligned} \tag{12}$$

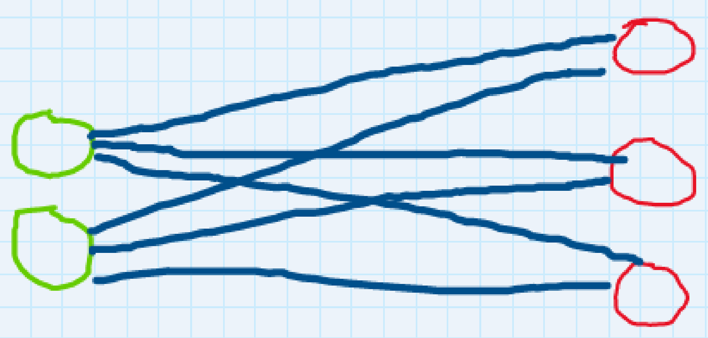
the transformation matrix $\mathbf{A}_\Phi$ with respect to B and C satisfies $\Phi(\mathbf{b}_k) = \sum_{i=1}^{4} \alpha_{ik}\mathbf{c}_i$ for $k = 1, 2, 3$ and is given as

$$\mathbf{A}_\Phi = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3] = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 1 & 3 \\ 3 & 7 & 1 \\ -1 & 2 & 4 \end{bmatrix} \tag{13}$$

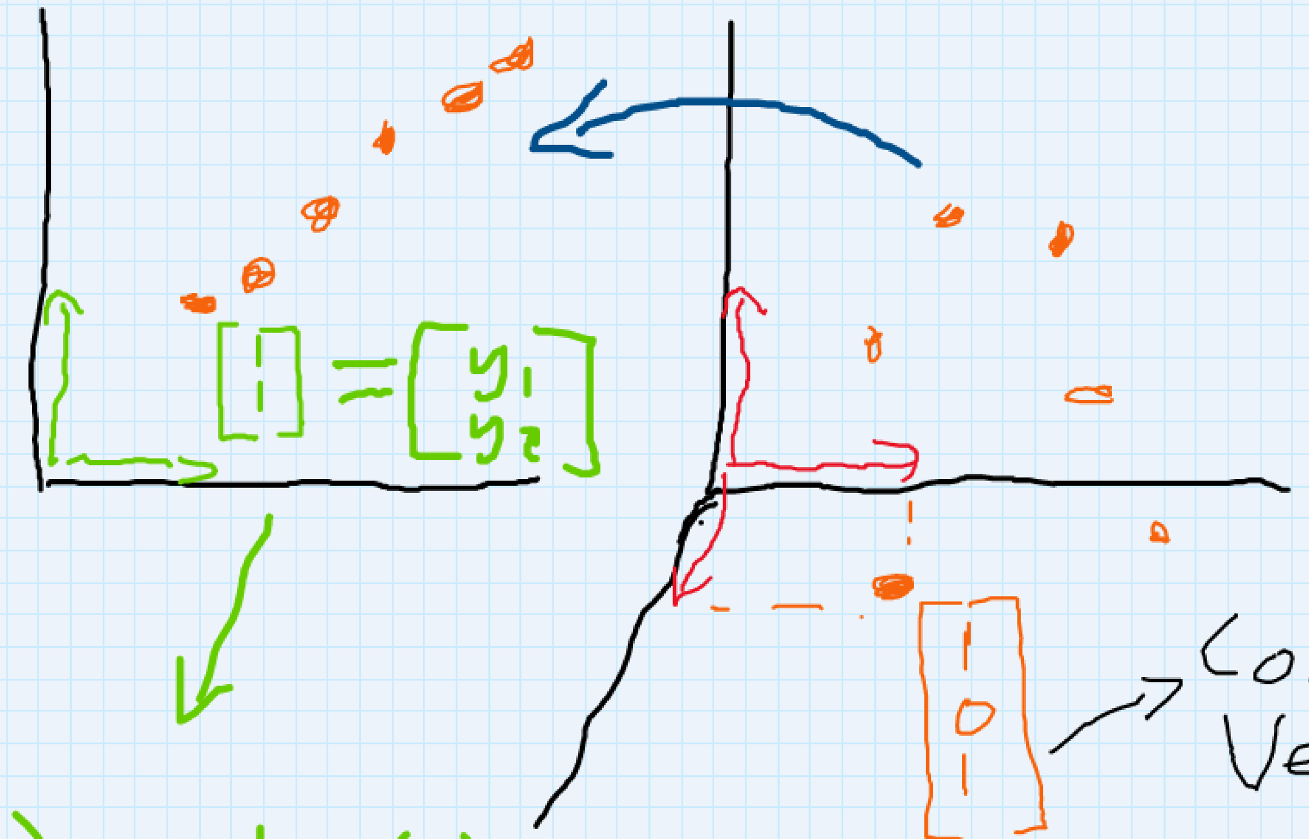
where the $\boldsymbol{\alpha}_j$, $j = 1, 2, 3$ are the coordinate vectors of $\Phi(\mathbf{b}_j)$ with respect to C

Neural Network

Wednesday, March 2, 2022 11:16 PM



$$\begin{bmatrix} y_1\\ y_2\end{bmatrix} = \begin{bmatrix} W_{11} & W_{12} & W_{13}\\ W_{21} & W_{22} & W_{23}\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix}$$

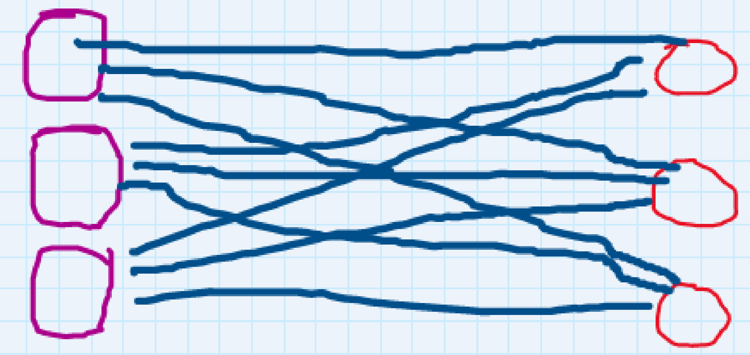


Coordinate Vector

$$y_1(\uparrow) + y_2(\rightarrow) = y_1 b_1 + y_2 b_2 = 1\begin{bmatrix}0\\1\end{bmatrix} + 1\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix}$$

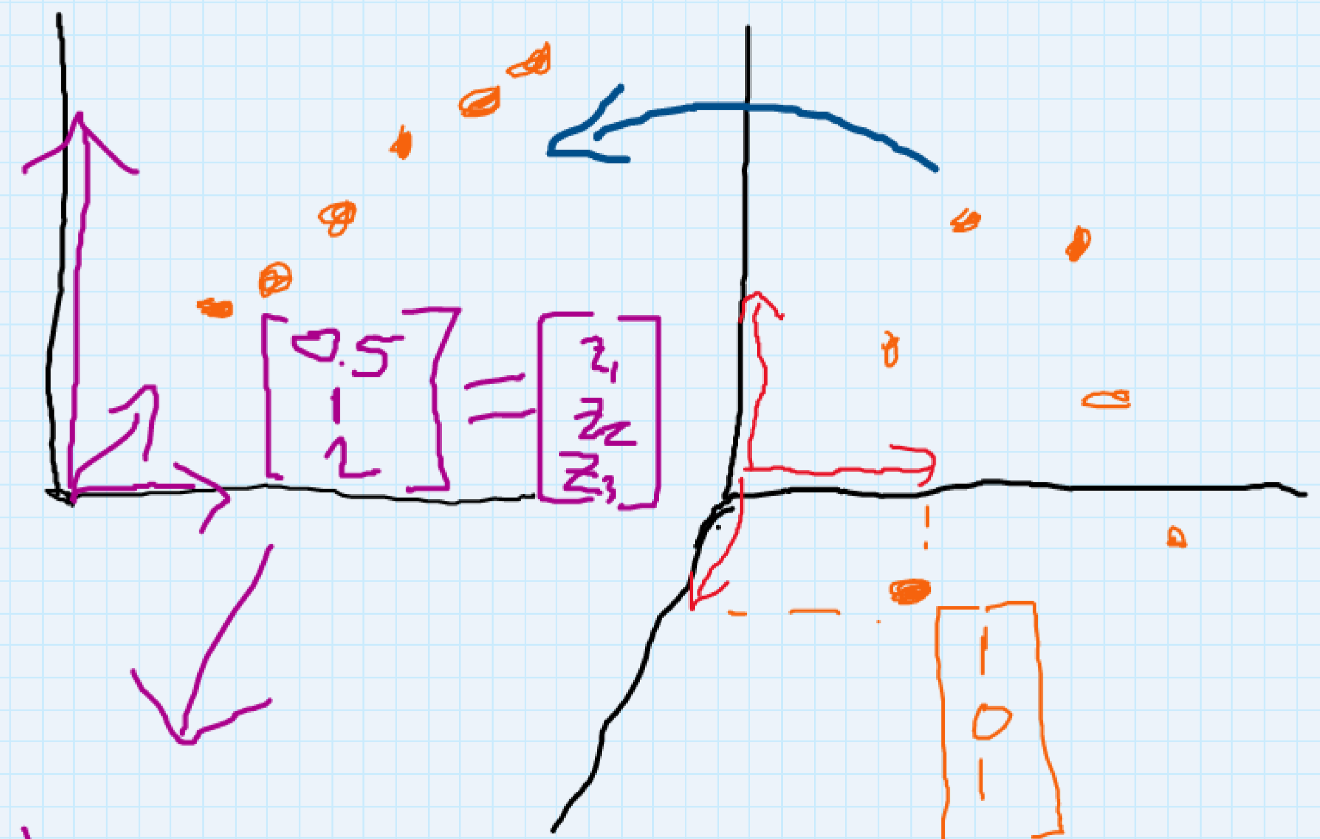
↳ We only know the input and output bases

↳ Eg: pixel space and "dogness" space



$$\begin{bmatrix} z_1\\ z_2\\ z_3\end{bmatrix} = \begin{bmatrix} W_{11} & W_{12} & W_{13}\\ W_{21} & W_{22} & W_{23}\\ W_{31} & W_{32} & W_{33}\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix}$$

Note dim is $\mathbb{R}^2$ even though we have 3 neurons (dim is num basis vectors needed)



$$z_1(\uparrow) + z_2(\rightarrow) + z_3(\nearrow) = z_1\hat{b}_1 + z_2\hat{b}_2 + z_3\hat{b}_3 = \frac{1}{4}\begin{bmatrix}0\\2\end{bmatrix} + \frac{1}{2}\begin{bmatrix}1\\0\end{bmatrix} + 1\begin{bmatrix}0.5\\0.5\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix}$$

$\Longrightarrow$ We have a useless neuron

$b_1 \Rightarrow$ "How big the dogs ears are"

$b_2 \Rightarrow$ "How big the dogs tail is"

$b_3 \Rightarrow$ "How big is the dogs tail and ears"

} The data gives our basis meaning

*Usually this kind of redundancy leads to "overfitting"

↳ The network has to also learn the nullspace

↳ b_2 was the same for both spaces but the coefficient changed because of b_3

We still can tell if Z is linearly independent just by looking at W if we know x is lin indep

Basis Change

Consider the two ordered bases of V

$$B = (\mathbf{b}_1, \ldots, \mathbf{b}_n),\ \tilde{B} = (\tilde{\mathbf{b}}_1, \ldots, \tilde{\mathbf{b}}_n)$$

and the two ordered bases of W

$$C = (\mathbf{c}_1, \ldots, \mathbf{c}_n),\ \tilde{C} = (\tilde{\mathbf{c}}_1, \ldots, \tilde{\mathbf{c}}_n)$$

And you currently have the transformation matrix, $\mathbf{A}_\Phi \in \mathbb{R}^{m \times n}$, of the linear mapping $\Phi : V \to W$ with respect to the bases B and C.

- How can we find $\tilde{\mathbf{A}}_\Phi \in \mathbb{R}^{m \times n}$ to achieve the corresponding transformation with respect to the bases $\tilde{B}$ and $\tilde{C}$.

Basis Change

Theorem 2.20 (Basis Change)

For a linear mapping $\Phi : V \to W$, ordered bases

$$B = (\mathbf{b}_1, \ldots, \mathbf{b}_n),\ \tilde{B} = (\tilde{\mathbf{b}}_1, \ldots, \tilde{\mathbf{b}}_n)$$

of V and

$$C = (\mathbf{c}_1, \ldots, \mathbf{c}_m),\ \tilde{C} = (\tilde{\mathbf{c}}_1, \ldots, \tilde{\mathbf{c}}_m)$$

of W, and a transformation matrix, $\mathbf{A}_\Phi$ of Φ with respect to B and C, the corresponding transformation matrix $\tilde{\mathbf{A}}_\Phi$ with respect to the bases $\tilde{B}$ and $\tilde{C}$ is given as

$$\tilde{\mathbf{A}}_\Phi = \mathbf{T}^{-1}\mathbf{A}_\Phi\mathbf{S} \tag{14}$$

Basis Change

Theorem 2.20 (Basis Change) continued

- Here, $\mathbf{S} \in \mathbb{R}^{n \times n}$ is the transformation matrix of id_V that maps coordinates with respect to $\tilde{\mathbf{B}}$ onto coordinates with respect to $\mathbf{B}$, and
- $\mathbf{T} \in \mathbb{R}^{m \times m}$ is the transformation matrix of id_W that maps coordinates with respect to $\tilde{\mathbf{C}}$ onto coordinates with respect to $\mathbf{C}$

Think of using the transformation as doing the following

- We start with a vector from V represent using basis $\tilde{B}$
- We use id_V to transform it to the basis B
- Now we are in the correct form to use $\mathbf{A}_\Phi$ directly
- After using $\mathbf{A}_\Phi$ our vector is unfortunately represented using basis C
- We now apply id_W^{-1} to transform it to the basis $\tilde{C}$
- The concatenation/product of all of these transformation is our desired transformation $\tilde{\mathbf{A}}_\Phi$

Basis Change

Vector spaces

Ordered bases

$$\begin{tikzcd}
V \arrow[r, "\Phi"] & W \\
B \arrow[r, "\Phi_{CB}", "\color{red}\boldsymbol{A}_\Phi"'] & C \\
\tilde{B} \arrow[u, "\Psi_{B\tilde{B}}", "\color{red}\boldsymbol{S}"'] \arrow[r, "\color{red}\tilde{\boldsymbol{A}}_\Phi", "\Phi_{\tilde{C}\tilde{B}}"'] & \tilde{C} \arrow[u, "\color{red}\boldsymbol{T}", "\Xi_{C\tilde{C}}"']
\end{tikzcd}
\qquad
\begin{tikzcd}
V \arrow[r, "\Phi"] & W \\
B \arrow[r, "\Phi_{CB}", "\color{red}\boldsymbol{A}_\Phi"'] & C \arrow[d, "\color{red}\boldsymbol{T}^{-1}"', "\Xi_{\tilde{C}C} = \Xi^{-1}_{C\tilde{C}}"] \\
\tilde{B} \arrow[u, "\Psi_{B\tilde{B}}", "\color{red}\boldsymbol{S}"'] \arrow[r, "\color{red}\tilde{\boldsymbol{A}}_\Phi", "\Phi_{\tilde{C}\tilde{B}}"'] & \tilde{C}
\end{tikzcd}$$

Source: M.P. Deisenroth *et al*, Mathematics for Machine Learning (First Edition)

where $\Psi_{B\tilde{B}} = id_V$ and $\Xi_{C\tilde{C}} = id_W$.

Basis Change

invertible

Equivalence

Two matrices $\mathbf{A}, \tilde{\mathbf{A}} \in \mathbb{R}^{m \times n}$ are *equivalent* if there exists regular matrices $\mathbf{S} \in \mathbb{R}^{n \times n}$ and $\mathbf{T} \in \mathbb{R}^{m \times m}$, such that $\tilde{\mathbf{A}} = \mathbf{T}^{-1}\mathbf{A}\mathbf{S}$.

Similar

Two matrices $\mathbf{A}, \tilde{\mathbf{A}} \in \mathbb{R}^{n \times n}$ are *similar* if there exists regular matrices $\mathbf{S} \in \mathbb{R}^{n \times n}$, such that $\tilde{\mathbf{A}} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$. $V \longrightarrow V$

- Similar matrices are always equivalent. However, equivalent matrices are not necessarily similar.

Homework: Complete example 2.24 (fill in all the missing calculations)

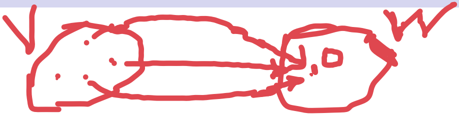


Image and Kernel

For $\Phi : V \to W$, we define the *kernel/null space*

$$ker\,(\Phi) := \Phi^{-1}(\mathbf{0}_W) = \{\mathbf{v} \in V \ : \Phi(v) = \mathbf{0}_W\} \tag{15}$$

and the *image/range*

$$Im(\Phi) := \Phi(V) = \{\mathbf{w} \in W \mid \exists \mathbf{v} \in V : \Phi(\mathbf{v}) = \mathbf{w}\} \tag{16}$$

We also call V and W the *domain* and *codomain* of Φ, respectively.

Image and Kernel: Linear mappings

Consider a linear mapping $\Phi : V \to W$, where V, W are vector spaces.

- It always holds that $\Phi(\mathbf{0}_V) = \mathbf{0}_W$ and, therefore, $\mathbf{0}_V \in Ker(\Phi)$
  - The null space/kernel of a **linear mapping** is never empty
- $Im(\Phi) \subseteq W$ is a subspace of W, and $ker(\Phi) \subseteq V$ is a subspace of V
- Φ is injective (one-to-one) if and only if $ker(\Phi) = \{0\}$

Image and Kernel: Null Space and Column Space

Let us consider $\mathbf{A} \in \mathbb{R}^{m\times n}$ and a linear mapping $\Phi : \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{x} \mapsto \mathbf{Ax}$.

- For $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_n]$, where $\mathbf{a}_i$ are the columns of $\mathbf{A}$, we obtain

$$Im(\Phi) = \underline{\{\mathbf{Ax} : \mathbf{x} \in \mathbb{R}^n\}} = \left\{ \sum_{i=1}^{n} x_i \mathbf{a}_i : x_1, \dots, x_n \in \mathbb{R} \right\} \tag{17}$$

$$= span[\mathbf{a}_1, \dots, \mathbf{a}_n] \subseteq \mathbb{R}^m \tag{18}$$

  - The image is the span of the columns of $\mathbf{A}$, also called the *column space*.

Let us consider $\mathbf{A} \in \mathbb{R}^{m \times n}$ and a linear mapping $\Phi : \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{x} \mapsto \mathbf{Ax}$.

span[A]

- $rk(A) = dim(Im(\Phi))$
- The kernel/null space $ker(\Phi)$ is the general solution to the homogeneous system of linear equations $\mathbf{Ax} = \mathbf{0}$.

Image and Kernel: Rank-Nullity Theorem

$$\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Rank-Nullity Theorem

For vector spaces V, W and a linear mapping $\Phi : V \to W$ it holds that

$$dim(ker(\Phi)) + dim(Im(\phi)) = dim(V) \tag{19}$$

The rank-nullity theorem is also referred to as the *fundamental theorem of linear mappings*

$\dim V = n$

Let: $K = \{v_1, \ldots, v_k\} \subset Ker(\Phi)$

$S = \{w_1, \ldots, w_{n-k}\} \subset V \setminus Ker(\Phi)$

$B = K \cup S = \{v_1, \ldots, v_k, w_1, \ldots, w_{n-k}\}$

$Im(\Phi) = span(B) = span\{\Phi(v_1) \ldots \Phi(w_{n-k})\}$

$= span\{\Phi(w_1), \ldots, \Phi(w_{n-k})\}$

$\therefore \Phi(S)$ is a gen set of $Im(\Phi)$. Prove its a basis (so min gen set)

$\Rightarrow Rank(\Phi) + Nullity(\Phi)$

$= \dim Im(\Phi) + \dim Ker(\Phi)$

$\therefore n - k + k = n = \dim V$

Image and Kernel: Rank-Nullity Theorem

The following important results follow directly from the Rank-Nullity theorem

- If $dim(Im(\Phi)) < dim(V)$, then $ker(\Phi)$ is non-trivial.
- If $\mathbf{A}_\Phi$ is the transformation matrix of Φ with respect to an ordered basis and $dim(Im(\Phi)) < dim(V)$, then the system of linear equations $\mathbf{A}_\Phi \mathbf{x} = \mathbf{0}$ has infinitely many solutions
- If $dim(V) = dim(W)$, then the following three-way equivalence holds:
  - Φ is injective
  - Φ is surjective
  - Φ is bijective

  since $Im(\Phi)$ is a subspace of W.

Assume $\phi(S)$ is not min span

$\sum^{n-k} \alpha_j \phi(s_j) = 0_W$

$\Rightarrow \phi\left(\sum^{n-k} \alpha_j s_j\right) = 0_W \in Ker(\phi)$

Contradiction

$\therefore \phi(S)$ is a minimal span set (ie: a basis for $Im(\phi)$)

$dim(S) = dim(\phi(S))$

Equation Example

Wednesday, March 9, 2022 12:44 AM

Desk takes 3 wood and 4 metal
Chair takes 2 wood and 3 metal
There is 4 wood and 20 metal available

$$\begin{bmatrix}3\\4\end{bmatrix} * \underbrace{\text{num desk}}_{x_1} + \begin{bmatrix}2\\3\end{bmatrix} * \underbrace{\text{num chair}}_{x_2} = \begin{bmatrix}4\\20\end{bmatrix} \begin{matrix}\leftarrow \text{Wood}\\ \leftarrow \text{Metal}\end{matrix}$$

So what we are trying to optimize is i.t.o. Wood and Metal

$$\begin{bmatrix}3 & 2\\4 & 3\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}4\\20\end{bmatrix}$$

$$\left[\begin{array}{cc|c}3 & 2 & 4\\4 & 3 & 20\end{array}\right] \quad \text{swap } R_1 \text{ and } R_2$$

$$\rightsquigarrow \left[\begin{array}{cc|c}4 & 3 & 20\\3 & 2 & 4\end{array}\right] \quad R_1 = R_1 - R_2$$

$$\rightsquigarrow \left[\begin{array}{cc|c}1 & 1 & 16\\3 & 2 & 4\end{array}\right] \quad R_2 = R_2 - 3R_1$$

$$\rightsquigarrow \left[\begin{array}{cc|c}1 & 1 & 16\\0 & -1 & -44\end{array}\right] \quad R_2 = -1R_2$$

$$\rightsquigarrow \left[\begin{array}{cc|c}1 & 1 & 16\\0 & 1 & 44\end{array}\right] \quad R_1 = R_1 - R_2$$

$$\rightsquigarrow \left[\begin{array}{cc|c}1 & 0 & -28\\0 & 1 & 44\end{array}\right]$$

$$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} = \underbrace{\begin{bmatrix}3 & -2\\-4 & 3\end{bmatrix}}_{A^{-1}}\begin{bmatrix}4\\20\end{bmatrix}$$

$$\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}-28\\44\end{bmatrix} \begin{matrix}\leftarrow \text{Desks}\\ \leftarrow \text{Chairs}\end{matrix}$$

We are changing basis from units of resources to units of products

$$B = \left\{\begin{bmatrix}3\\4\end{bmatrix}, \begin{bmatrix}2\\3\end{bmatrix}\right\} \quad \text{Forward}$$

$$\tilde{B} = \left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\} \quad \text{Backward}$$

$$\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}3\\4\end{bmatrix}3 + \begin{bmatrix}2\\3\end{bmatrix}(-4)$$

$$\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}3\\4\end{bmatrix}(-2) + \begin{bmatrix}2\\3\end{bmatrix}3$$

$$F = \begin{bmatrix}3 & -2\\-4 & 3\end{bmatrix}$$

$$\begin{bmatrix}3\\4\end{bmatrix} = 3\begin{bmatrix}1\\0\end{bmatrix} + 4\begin{bmatrix}0\\1\end{bmatrix}$$

$$\begin{bmatrix}2\\3\end{bmatrix} = 2\begin{bmatrix}1\\0\end{bmatrix} + 3\begin{bmatrix}0\\1\end{bmatrix}$$

$$\boxed{F' = \begin{bmatrix}3 & -2\\-4 & 3\end{bmatrix}}$$

$$V = \begin{bmatrix}B_1 & B_2\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}_B = \begin{bmatrix}\tilde{B}_1 & \tilde{B}_2\end{bmatrix}F^{-1}\begin{bmatrix}x_1\\x_2\end{bmatrix}_{\tilde{B}}$$

$$V = F^{-1}\begin{bmatrix}x_1\\x_2\end{bmatrix}$$

$$FV = \begin{bmatrix}x_1\\x_2\end{bmatrix}$$

$$\begin{bmatrix}3 & -2\\-4 & 3\end{bmatrix}\begin{bmatrix}4\\20\end{bmatrix} = \begin{bmatrix}x_1\\x_2\end{bmatrix}$$

$$\begin{bmatrix}-28\\44\end{bmatrix} = \begin{bmatrix}x_1\\x_2\end{bmatrix}$$