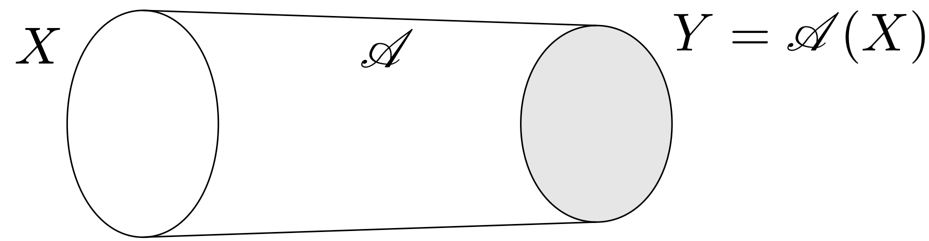
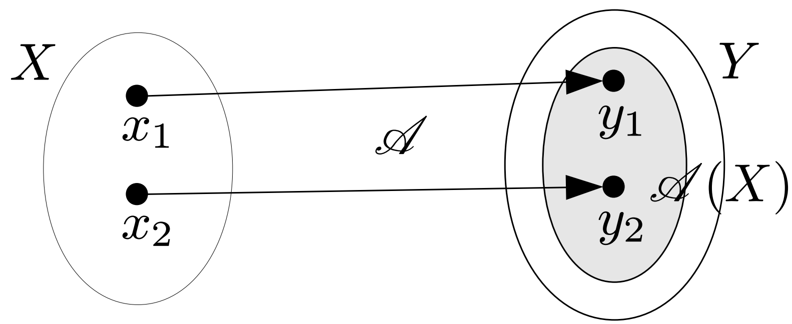
3.1 Mapping

- Surjection（全射）, injection（単射）and bijection（全単射）
  - A mapping $\mathscr{A}$ is said to be a **surjection** or a onto-mapping（上への写像）, when $\mathscr{A}(X) = Y$



  - A mapping $\mathscr{A}$ is said to be an **injection** or a one-to-one-mapping（1対1写像）, when $x_1 \neq x_2 \Rightarrow \mathscr{A}x_1 \neq \mathscr{A}x_2$ holds for $\forall x_1, x_2 \in X$, i.e., for $\forall y \in \mathscr{A}(X) \subset Y$, there is only one x that satisfies $y = \mathscr{A}x$



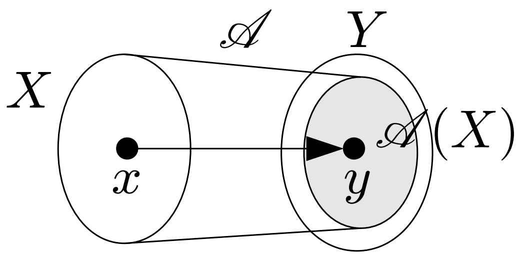
  - When $\mathscr{A}$ is a surjection and an injection, it is said to be a **bijection**

3.2 Amount of information transferred by a mapping

- How much of information is transferred by a mapping
  - A mapping $\mathscr{A} : X \to Y$ transfers information contained in set X to set Y.
    There is no external source of information. Therefore, the mapping does not increase the amount of information (it reduces it or keeps it the same)

3.2 Amount of information transferred by a mapping

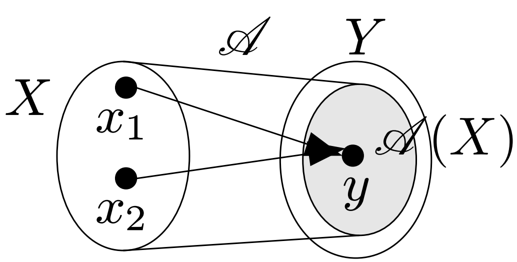
– Injection



The amount of information in X
= the amount of information in $\mathscr{A}(X)$
≤ the amount of information in Y

(For each and every $y \in \mathscr{A}(X)$,
there is one and only one x that is related by $y = \mathscr{A}x$.)

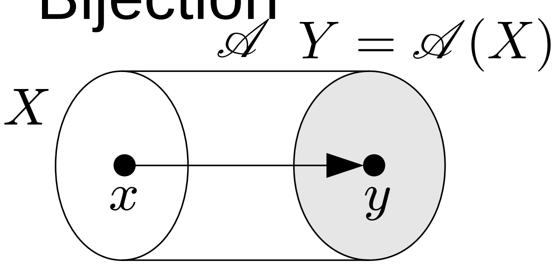
– Non-injection



The amount of information in X
> the amount of information in $\mathscr{A}(X)$

(Different x_1 and x_2 can be mapped to an identical y.
Information about the difference $x_1 - x_2$ is lost by mapping $\mathscr{A}$.)

– Bijection



The amount of information in X
= the amount of information in $\mathscr{A}(X)$
= the amount of information in Y

4. Amount of information contained in linear spaces

- In the previous slide, we considered whether a mapping $\mathscr{A}$ can fully transfer information from X into Y
- Then, let us now consider how much information is contained in X, in the first place
- In short, the amount of information contained in a linear space is represented by the maximum number of linearly independent（線形独立な） vectors in the space

4.1 Linearly independent vectors

- Linearly independent vectors
  - A set of vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$ is a linearly independent set

$\Longleftrightarrow$ Its linear combination（線形結合）

$$c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \cdots + c_n\boldsymbol{x}_n,\ c_i \in F,\ \boldsymbol{x}_i \in X$$

is $\mathbf{0}$ (zero vector), if and only if

$$c_1 = c_2 = \cdots c_n = 0$$

holds.

4.1 Linearly independent vectors

- Representing a vector by a linear combination of other vectors

  Suppose that $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$ is linearly independent.

  Any vector $\boldsymbol{x}_j, 1 \leq j \leq n$ cannot be represented by a linear combination of the other vectors $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{j-1}, \boldsymbol{x}_{j+1}, \cdots, \boldsymbol{x}_n$.

  Makes use of all the operations (addition and multiplication by a scalar) defined in linear spaces.

Assume that a vector $\boldsymbol{x}_j$ can be represented by a linear combination, then

$$\boldsymbol{x}_j = d_1\boldsymbol{x}_1 + \cdots + d_{j-1}\boldsymbol{x}_{j-1} + d_{j+1}\boldsymbol{x}_{j+1} + \cdots + d_n\boldsymbol{x}_n$$

holds, which is equivalent to

$$d_1\boldsymbol{x}_1 + \cdots + d_{j-1}\boldsymbol{x}_{j-1} + (-1)\boldsymbol{x}_j + d_{j+1}\boldsymbol{x}_{j+1} + \cdots + d_n\boldsymbol{x}_n = \boldsymbol{0}.$$

This contradicts the fact the the set of vectors is linearly independent since the coefficient on $\boldsymbol{x}_j$ is -1 and is not zero.

4.2 Amount of information in linear spaces

- When $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$ is linearly independent, any of the n vectors in this set contains its **inherent** information that cannot be represented by the other vectors even if any possible operations are used. So the number of linearly independent vectors corresponds to the amount of information.

  Note that it might be represented by using other operations such as square and square root which are not defined in linear spaces.

In a given linear space, there are many linearly independent subsets of vectors with different numbers of vectors in the subsets.

$$\{\boldsymbol{x}_1, \boldsymbol{x}_2\},\ \{\boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{x}_6\},\ \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_4\},\ \{\boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{x}_6, \boldsymbol{x}_7\}, \cdots$$

$$\Downarrow$$

The <u>maximum</u> number of linearly independent vectors contained in a linear space represents the amount of information contained in the linear space.

4.2 Amount of information in linear spaces

- Basis（基底）

Suppose that the maximum number of linearly independent vectors in X is n, i.e., $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$ is linearly independent.

$\Downarrow$

Any $\boldsymbol{x} \in X$ can be represented as a linear combination of the vectors in $\underbrace{\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}}_{\text{Basis}}$.

n is the <u>maximum</u> number of linearly independent vectors in X.

$\Leftrightarrow$ For $\forall \boldsymbol{x} \in X$, the set of $n+1$ vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n, \boldsymbol{x}\}$ is not linearly independent.

$\Leftrightarrow$ There exist $c_1, \cdots, c_n, c$ such that $c_1\boldsymbol{x}_1 + \cdots + c_n\boldsymbol{x}_n + c\boldsymbol{x} = \boldsymbol{0}, \ c \neq 0$.

$\Leftrightarrow$ $\boldsymbol{x}$ can be represented as $\boldsymbol{x} = -\frac{c_1}{c}\boldsymbol{x}_1 - \cdots - \frac{c_n}{c}\boldsymbol{x}_n$.

4.2 Amount of information in linear spaces

– Example of basis

$\forall \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3$ can be represented as

$\boldsymbol{x} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + x_3\boldsymbol{e}_3$ by $\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, $\boldsymbol{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

Thus $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ is the basis of $\mathbb{R}^3$.

4.2 Amount of information in linear spaces

- Dimension（次元）

Suppose $\boldsymbol{x}$ is represented by a basis $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$as

$$\boldsymbol{x} = \xi_1 \boldsymbol{x}_1 + \cdots + \xi_n \boldsymbol{x}_n = [\boldsymbol{x}_1 \ \cdots \boldsymbol{x}_n] \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix},$$

$\xi_1, \cdots, \xi_n$ is called coordinates（座標）or components（成分）of $\boldsymbol{x}$ with respect to basis $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$.

The number of vectors contained in the basis of X is called dimension of X, $\mathrm{dim} X$.

$\mathrm{dim} X$ can be considered to represent the amount of information contained in the linear space X.

5. Linear mappings

- We observed matrices are closely related to linearity and mappings. Also we reviewed linear spaces
- Now let us have a look at the relation between matrices and linear mappings

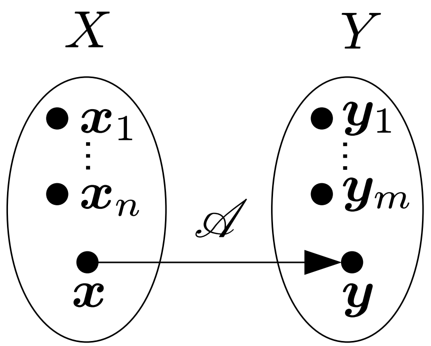
5.1 Linear mappings and matrices

Let X and Y be linear spaces.

$\dim X = n$, basis of X is $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$, $\boldsymbol{x}_i \in \mathbb{R}^n,\ i = 1, \ldots, n$.

$\dim Y = m$, basis of Y is $\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_m\}$, $\boldsymbol{y}_i \in \mathbb{R}^m,\ i = 1, \ldots, m$.

Let $\mathscr{A} : X \to Y$ be a linear mapping.



$\forall \boldsymbol{x} \in X$ can be represented by the basis of X as

$$\boldsymbol{x} = \xi_1 \boldsymbol{x}_1 + \cdots + \xi_n \boldsymbol{x}_n = [\boldsymbol{x}_1 \ \cdots \boldsymbol{x}_n] \boxed{\begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}}$$

Coordinates of $\boldsymbol{x}$

Let $\boldsymbol{y} = \mathscr{A}\boldsymbol{x}$, and $\boldsymbol{y}$ can be represented by the basis of Y as

$$\boldsymbol{y} = \eta_1 \boldsymbol{y}_1 + \cdots + \eta_m \boldsymbol{y}_m = [\boldsymbol{y}_1 \ \cdots \boldsymbol{y}_m] \boxed{\begin{bmatrix} \eta_1 \\ \vdots \\ \eta_m \end{bmatrix}} .$$

Coordinates of $\boldsymbol{y}$

5.1 Linear mappings and matrices

Principle of superposition holds, and therefore

$$\boldsymbol{y} = \mathscr{A}\boldsymbol{x} = \mathscr{A}\left(\xi_1\boldsymbol{x}_1 + \cdots + \xi_n\boldsymbol{x}_n\right) = \mathscr{A}\xi_1\boldsymbol{x}_1 + \cdots + \mathscr{A}\xi_n\boldsymbol{x}_n$$

$$= \xi_1\mathscr{A}\boldsymbol{x}_1 + \cdots + \xi_n\mathscr{A}\boldsymbol{x}_n$$

$\mathscr{A}\boldsymbol{x}_i \in Y$ can be represented by a linear combination of the basis of Y, $\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_m\}$ as

$$\mathscr{A}\boldsymbol{x}_i = a_{1i}\boldsymbol{y}_1 + \cdots + a_{mi}\boldsymbol{y}_m = \underbrace{[\boldsymbol{y}_1 \ \cdots \boldsymbol{y}_m]}_{m \times m \text{ matrix}} \underbrace{\begin{bmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{bmatrix}}_{m \text{ vector}} .$$

$$\therefore \boldsymbol{y} = \xi_1[\boldsymbol{y}_1 \cdots \boldsymbol{y}_m]\begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + \cdots + \xi_n[\boldsymbol{y}_1 \cdots \boldsymbol{y}_m]\begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} .$$

5.1 Linear mappings and matrices

$$
\begin{aligned}
\boldsymbol{y} &= \xi_1 [\boldsymbol{y}_1 \cdots \boldsymbol{y}_m] \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + \cdots + \xi_n [\boldsymbol{y}_1 \cdots \boldsymbol{y}_m] \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} \\
&= \left[\; [\boldsymbol{y}_1 \ \cdots \boldsymbol{y}_m] \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} \ \cdots \ [\boldsymbol{y}_1 \ \cdots \boldsymbol{y}_m] \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} \; \right] \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix} \\
&= [\boldsymbol{y}_1 \ \cdots \boldsymbol{y}_m] \boxed{\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}} .
\end{aligned}
$$

Turn out to be equal

Recall that $\boldsymbol{y}$ is also represented by

$$
\boldsymbol{y} = \eta_1 \boldsymbol{y}_1 + \cdots + \eta_m \boldsymbol{y}_m = [\boldsymbol{y}_1 \ \cdots \boldsymbol{y}_m] \boxed{\begin{bmatrix} \eta_1 \\ \vdots \\ \eta_m \end{bmatrix}} .
$$

5.1 Linear mappings and matrices

A linear mapping $\boldsymbol{y} = \mathscr{A}\boldsymbol{x}$ is represented as

$$\begin{bmatrix} \eta_1 \\ \vdots \\ \eta_m \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix} = A \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}.$$

Coordinates of $\boldsymbol{y}$ with respective the basis in Y

A Matrix!

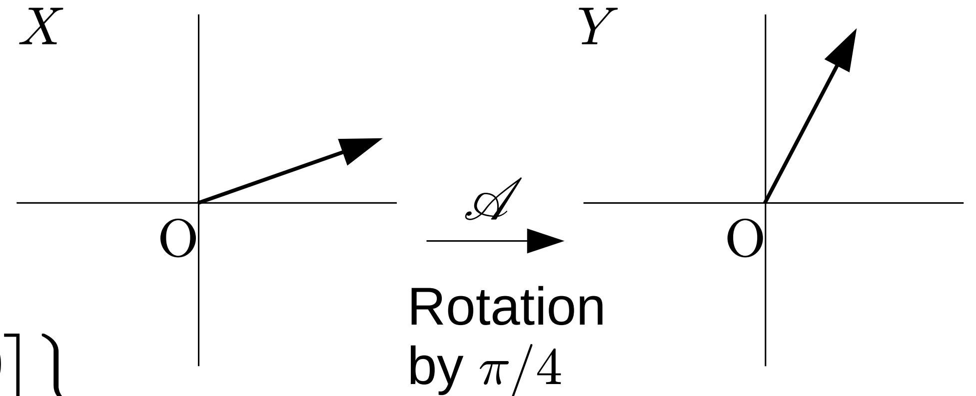
Coordinates of $\boldsymbol{x}$ with respective the basis in X

Once bases are given,
a matrix represents a linear mapping.

The elements of the matrix, a_{ij} are derived by representing $\mathscr{A}\boldsymbol{x}_j$ using the basis of Y.

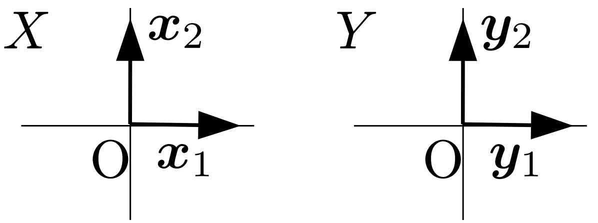
5.1 Linear mappings and matrices

- Examples
  - Rotation of a vector by $\pi/4$



Basis of X : $\{\boldsymbol{x}_1, \boldsymbol{x}_2\} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$

Basis of Y: $\{\boldsymbol{y}_1, \boldsymbol{y}_2\} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$



$$\mathscr{A}\boldsymbol{x}_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \underbrace{\frac{\sqrt{2}}{2}}_{a_{11}}\boldsymbol{y}_1 + \underbrace{\frac{\sqrt{2}}{2}}_{a_{21}}\boldsymbol{y}_2, \ \mathscr{A}\boldsymbol{x}_2 = \begin{bmatrix} \frac{-\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \underbrace{-\frac{\sqrt{2}}{2}}_{a_{12}}\boldsymbol{y}_1 + \underbrace{\frac{\sqrt{2}}{2}}_{a_{22}}\boldsymbol{y}_2$$

The mapping is represented by a matrix

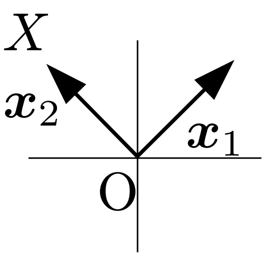
$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}.$$

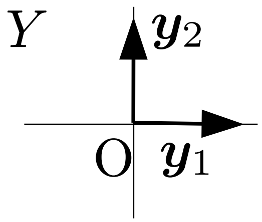
5.1 Linear mappings and matrices

– Rotation of a vector by $\pi/4$ (with a different basis of X)

Basis of X : $\{\boldsymbol{x}_1, \boldsymbol{x}_2\} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$

Basis of Y: $\{\boldsymbol{y}_1, \boldsymbol{y}_2\} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$





$$\mathscr{A}\boldsymbol{x}_1 = \begin{bmatrix} 0 \\ \sqrt{2} \end{bmatrix} = \underbrace{0}_{a_{11}} \cdot \boldsymbol{y}_1 + \underbrace{\sqrt{2}}_{a_{21}}\boldsymbol{y}_2, \ \mathscr{A}\boldsymbol{x}_2 = \begin{bmatrix} -\sqrt{2} \\ 0 \end{bmatrix} = \underbrace{-\sqrt{2}}_{a_{12}}\boldsymbol{y}_1 + \underbrace{0}_{a_{22}} \cdot \boldsymbol{y}_2$$

The mapping is represented by a matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 0 & -\sqrt{2} \\ \sqrt{2} & 0 \end{bmatrix}.$$

5.1 Linear mappings and matrices

- To summarize
  - Given a basis $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$, a vector $\boldsymbol{x}$ is represented as

$$\boldsymbol{x} = \xi_1 \boldsymbol{x}_1 + \cdots + \xi_n \boldsymbol{x}_n = [\boldsymbol{x}_1 \ \cdots \boldsymbol{x}_n] \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix} = [\boldsymbol{x}_1 \ \cdots \boldsymbol{x}_n] \boldsymbol{\xi}.$$

  - A linear mapping $\mathscr{A} : \boldsymbol{x} \mapsto \boldsymbol{y}$ is represented as $\boldsymbol{\eta} = A\boldsymbol{\xi}$ using a matrix A, given coordinates of the vectors:

$$\boldsymbol{\xi} = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}, \ \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_m \end{bmatrix}.$$

Given bases, a vector $\boldsymbol{x}$ can be represented by its coordinate $\boldsymbol{\xi}$, and a mapping $\mathscr{A}$ can be represented by its matrix representation A.

5.1 Linear mappings and matrices

- In Section 6 and thereafter, we will use the same symbol $\boldsymbol{x}$ for a vector and its coordinates, and represent a mapping using its matrix representation, $\boldsymbol{y} = A\boldsymbol{x}$

5.2 Composite mappings and multiplication of matrices

- Let us consider two linear mappings

  $$\mathscr{A} : \boldsymbol{x} \mapsto \boldsymbol{y},\ \mathscr{B} : \boldsymbol{y} \mapsto \boldsymbol{z}$$

  where

  $$\boldsymbol{x} \in \mathbb{R}^\ell,\ \boldsymbol{y} \in \mathbb{R}^m,\ \boldsymbol{z} \in \mathbb{R}^n.$$

  Also let the following are their respective matrix representations

  $$\boldsymbol{\eta} = A\boldsymbol{\xi},\ \boldsymbol{\zeta} = B\boldsymbol{\eta}$$

  where $\boldsymbol{\xi},\ \boldsymbol{\eta}$ and $\boldsymbol{\zeta}$ are coordinates of $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$, respectively

  The composite mapping $\mathscr{B} \circ \mathscr{A}$ maps $\boldsymbol{x}$ to $\boldsymbol{z}$ via $\boldsymbol{y}$, and its matrix representation is $\boldsymbol{\zeta} = B\boldsymbol{\eta} = B(A\boldsymbol{\xi}) = BA\boldsymbol{\xi}$.

  Therefore the product of matrices BA represents a composite mapping $\mathscr{B} \circ \mathscr{A}$.

5.2 Composite mappings and multiplication of matrices

- How should the product BA be naturally defined?

Let $\boldsymbol{\xi} = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_\ell \end{bmatrix}$, $\boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_m \end{bmatrix}$, $\boldsymbol{\zeta} = \begin{bmatrix} \zeta_1 \\ \vdots \\ \zeta_n \end{bmatrix}$ and

$$A = \begin{bmatrix} a_{11} & \dots & a_{1\ell} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{m\ell} \end{bmatrix},\ B = \begin{bmatrix} b_{11} & \dots & b_{1m} \\ \vdots & & \vdots \\ b_{n1} & \dots & b_{nm} \end{bmatrix}.$$

Then $\boldsymbol{\eta} = A\boldsymbol{\xi} = \begin{bmatrix} a_{11}\xi_1 + \cdots + a_{1\ell}\xi_\ell \\ \vdots \\ a_{m1}\xi_1 + \cdots + a_{m\ell}\xi_\ell \end{bmatrix}$ and

$$\boldsymbol{\zeta} = B\boldsymbol{\eta} = \begin{bmatrix} b_{11}(a_{11}\xi_1 + \cdots + a_{1\ell}\xi_\ell) + \cdots + b_{1m}(a_{m1}\xi_1 + \cdots + a_{m\ell}\xi_\ell) \\ \vdots \\ b_{n1}(a_{11}\xi_1 + \cdots + a_{1\ell}\xi_\ell) + \cdots + b_{nm}(a_{m1}\xi_1 + \cdots + a_{m\ell}\xi_\ell) \end{bmatrix}.$$

5.2 Composite mappings and multiplication of matrices

$$\boldsymbol{\zeta} = B\boldsymbol{\eta} = \begin{bmatrix} b_{11}(a_{11}\xi_1 + \cdots + a_{1\ell}\xi_\ell) + \cdots + b_{1m}(a_{m1}\xi_1 + \cdots + a_{m\ell}\xi_\ell) \\ \vdots \\ b_{n1}(a_{11}\xi_1 + \cdots + a_{1\ell}\xi_\ell) + \cdots + b_{nm}(a_{m1}\xi_1 + \cdots + a_{m\ell}\xi_\ell) \end{bmatrix}$$

$$= \begin{bmatrix} (b_{11}a_{11} + \cdots + b_{1m}a_{m1})\xi_1 + \cdots + (b_{11}a_{1\ell} + \cdots + b_{1m}a_{m\ell})\xi_\ell \\ \vdots \\ (b_{n1}a_{11} + \cdots + b_{nm}a_{m1})\xi_1 + \cdots + (b_{n1}a_{1\ell} + \cdots + b_{nm}a_{m\ell})\xi_\ell \end{bmatrix}$$

$$= \begin{bmatrix} b_{11}a_{11} + \cdots + b_{1m}a_{m1} & \ldots & b_{11}a_{1\ell} + \cdots + b_{1m}a_{m\ell} \\ \vdots & & \vdots \\ b_{n1}a_{11} + \cdots + b_{nm}a_{m1} & \ldots & b_{n1}a_{1\ell} + \cdots + b_{nm}a_{m\ell} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_\ell \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} \sum_{j=1}^m b_{1j}a_{j1} & \ldots & \sum_{j=1}^m b_{1j}a_{j\ell} \\ \vdots & & \vdots \\ \sum_{j=1}^m b_{nj}a_{j1} & \ldots & \sum_{j=1}^m b_{nj}a_{j\ell} \end{bmatrix}}_{\text{The product of matrices } B \text{ and } A} \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_\ell \end{bmatrix} = BA \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_\ell \end{bmatrix}.$$

6. Rank of a matrix
6.1 Rank

- Rank（階数）of a matrix
  - Column rank（列階数）

Let us express an $m \times n$ matrix A as

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = [\boldsymbol{a}_1 \ \cdots \boldsymbol{a}_n],$$

$$\boldsymbol{a}_j = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix}$$: the j-th column vector of A.

Column rank of matrix A : the maximum number of linearly independent vectors in $\{\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n\}$, $\leq n$

If the column rank of matrix A is equal to n, then A is said to be of **full column rank**（最大列階数）.

6.1 Rank

– Row rank（行階数）

Let us express an $m \times n$ matrix A as

$$A = \begin{bmatrix} \boxed{a_{11} \quad \cdots \quad a_{1n}} \\ \vdots \qquad\qquad \vdots \\ \boxed{a_{m1} \quad \cdots \quad a_{mn}} \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{a}}_1^T \\ \vdots \\ \tilde{\boldsymbol{a}}_m^T \end{bmatrix},$$

$\tilde{\boldsymbol{a}}_i^T = [a_{i1} \ \cdots \ a_{in}]$: the i-th row vector of A.

Row rank of matrix A : the maximum number of linearly independent vectors in $\{\tilde{\boldsymbol{a}}_1, \cdots, \tilde{\boldsymbol{a}}_m\}$, $\leq m$

If the row rank of matrix A is equal to m,
then A is said to be of **full row rank**（最大行階数）.

6.1 Rank

– Rank

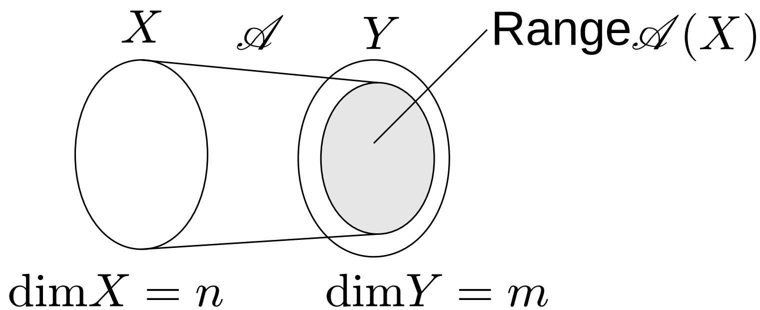
The column rank and the row rank of a matrix take the same value which is called **rank** of the matrix and expressed as $\mathrm{rank} A$.

$$\mathrm{rank} A \leq \min(m, n)$$

6.2 Amount of information a mapping transfers

- A matrix represents a linear mapping

$$\mathscr{A} : \boldsymbol{x} \mapsto \boldsymbol{y} \qquad \Leftrightarrow \qquad \boldsymbol{y} = A\boldsymbol{x}$$



$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

The domain X contains information whose amount is $\mathrm{dim}X$.

Linear mapping $\mathscr{A}$ transfers information in the domain X into the range $\mathscr{A}(X)$.

Then how much information does the linear mapping $\mathscr{A}$ transfer into the range $\mathscr{A}(X)$? In other words what is $\mathrm{dim}(\mathscr{A}(X))$?

The answer is $\mathrm{dim}(\mathscr{A}(X)) = \mathrm{rank}A$.

6.2 Amount of information a mapping transfers

$\dim(\mathscr{A}(X)) = \mathrm{rank} A$

Let us suppose $\mathrm{rank} A = r$ and column vectors $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_r$ are linearly independent. ($A = [\boldsymbol{a}_1 \ \ldots \ \boldsymbol{a}_n]$)

Each of the other column vectors of A, $\boldsymbol{a}_{r+1}, \cdots, \boldsymbol{a}_n$ is represented by a linear combination of $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_r$ as

$$\boldsymbol{a}_k = c_{k1}\boldsymbol{a}_1 + \cdots + c_{kr}\boldsymbol{a}_r, \ k = r+1, \cdots, n.$$

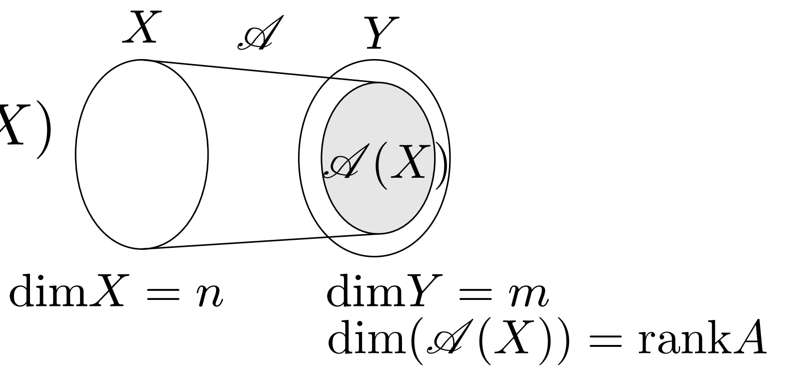
Then

$$\boldsymbol{y} = A\boldsymbol{x} = [\boldsymbol{a}_1 \ \ldots \ \boldsymbol{a}_n]\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\boldsymbol{a}_1 + \cdots + x_n\boldsymbol{a}_n$$

$$\begin{aligned} &= x_1\boldsymbol{a}_1 + \cdots + x_r\boldsymbol{a}_r \\ &\quad + x_{r+1}\{c_{r+1,1}\boldsymbol{a}_1 + \cdots + c_{r+1,r}\boldsymbol{a}_r\} + \cdots + x_n\{c_{n1}\boldsymbol{a}_1 + \cdots + c_{nr}\boldsymbol{a}_r\} \\ &= (x_1 + x_{r+1}c_{r+1,1} + \cdots + x_n c_{n1})\boldsymbol{a}_1 + \cdots \\ &\quad + (x_r + x_{r+1}c_{r+1,r} + \cdots + x_n c_{nr})\boldsymbol{a}_r. \end{aligned}$$

Thus, any $\boldsymbol{y} \in \mathscr{A}(X)$ is represented by a linear combination of r vectors $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_r$.

6.2 Amount of information a mapping transfers



Therefore the amount of information that a linear mapping $\mathscr{A}$ transfers into its range $\mathscr{A}(X)$ is $\mathrm{rank}\ A$.

When $\dim(\mathscr{A}(X)) = \mathrm{rank} A = n = \dim X$, the mapping $\mathscr{A}$ can transfer all the information contained in X.

In contrast, when $\dim(\mathscr{A}(X)) = \mathrm{rank} A < n = \dim \mathrm{X}$, $\mathscr{A}$ transfers NOT all the information. Then, where does the lost information go to?

For every $\boldsymbol{x} \in X$ there is a corresponding $\boldsymbol{y} \in \mathscr{A}(X)$. Still $\mathscr{A}(X)$ has a lower dimension than X.

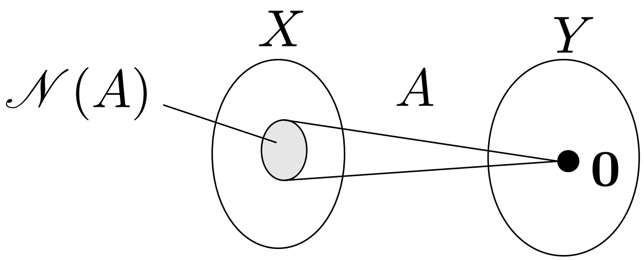
What happens here is that for some $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ where $\boldsymbol{x}_i \neq \boldsymbol{x}_j$, $A\boldsymbol{x}_i = A\boldsymbol{x}_j$.

6.3 Null space and range

Suppose matrix A represents a linear mapping.

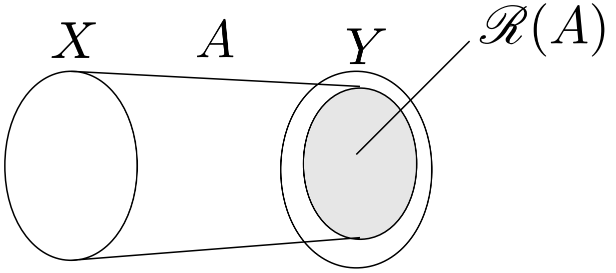
Null space（零空間）of A, $\mathscr{N}(A)$: $\{\boldsymbol{x} | A\boldsymbol{x} = \boldsymbol{0}\}$

aka kernel（核）of A, $\mathrm{Ker}\ A$



Range（値域）of A, $\mathscr{R}(A)$: $\{\boldsymbol{y} | \boldsymbol{y} = A\boldsymbol{x}, \boldsymbol{x} \in X\}$

aka image（像）of A, $\mathrm{Im}\ A$



Note that $\mathscr{R}(A)$ and $\mathscr{A}(X)$ represent the same thing.

Both null space and range are linear spaces.

6.3 Null space and range

The following holds regarding the null space and the range of an $m \times n$ matrix A:

$$\boxed{\dim \mathscr{N}(A) + \dim \mathscr{R}(A) = n = \dim X.}$$

Suppose that $\dim \mathscr{N}(A) = l < n$ and that $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_l\}$ is the basis of $\mathscr{N}(A)$.

We can add $n-l$ vectors $\boldsymbol{x}_{l+1}, \ldots, \boldsymbol{x}_n$ to form a basis $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_l, \boldsymbol{x}_{l+1}, \ldots, \boldsymbol{x}_n\}$ of the n dimensional linear space X.

Then, showing that $\{A\boldsymbol{x}_{l+1}, \ldots, A\boldsymbol{x}_n\}$ is a basis of $\mathscr{R}(A)$ proves the above.

For any $\boldsymbol{y} \in \mathscr{R}(A)$, there exists an $\boldsymbol{x} \in X$ that satisfies $\boldsymbol{y} = A\boldsymbol{x}$.

The vector $\boldsymbol{x}$ can be expressed in terms of the basis of X, $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, as,

$$\boldsymbol{x} = c_1\boldsymbol{x}_1 + \cdots + c_l\boldsymbol{x}_l + c_{l+1}\boldsymbol{x}_{l+1} + \cdots + c_n\boldsymbol{x}_n.$$

Then, since $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_l \in \mathscr{N}(A)$ and therefore $A\boldsymbol{x}_1 = \cdots = A\boldsymbol{x}_l = \boldsymbol{0}$,

$$\begin{aligned}\boldsymbol{y} = A\boldsymbol{x} &= c_1A\boldsymbol{x}_1 + \cdots + c_lA\boldsymbol{x}_l + c_{l+1}A\boldsymbol{x}_{l+1} + \cdots + c_nA\boldsymbol{x}_n \\ &= c_{l+1}A\boldsymbol{x}_{l+1} + \cdots + c_nA\boldsymbol{x}_n,\end{aligned}$$

which indicates that $\boldsymbol{y}$ is expressed as a linear combination of $A\boldsymbol{x}_{l+1}, \ldots, A\boldsymbol{x}_n$.

We will show, on the next page, that $A\boldsymbol{x}_{l+1}, \ldots, A\boldsymbol{x}_n$ are linearly independent.

6.3 Null space and range

Let $\alpha_{l+1}(A\boldsymbol{x}_{l+1}) + \cdots + \alpha_n(A\boldsymbol{x}_n) = \mathbf{0}$.

Then we have $A(\alpha_{l+1}\boldsymbol{x}_{l+1} + \cdots + \alpha_n\boldsymbol{x}_n) = \mathbf{0}$, which implies

$$\alpha_{l+1}\boldsymbol{x}_{l+1} + \cdots + \alpha_n\boldsymbol{x}_n \in \mathscr{N}(A).$$

So, this vector is expressed using the basis $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_l\}$ of $\mathscr{N}(A)$ as,

$$\alpha_{l+1}\boldsymbol{x}_{l+1} + \cdots + \alpha_n\boldsymbol{x}_n = \beta_1\boldsymbol{x}_1 + \cdots + \beta_l\boldsymbol{x}_l.$$

Therefore, we have,

$$-\beta_1\boldsymbol{x}_1 - \cdots - \beta_l\boldsymbol{x}_l + \alpha_{l+1}\boldsymbol{x}_{l+1} + \cdots + \alpha_n\boldsymbol{x}_n = \mathbf{0}.$$

Because $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ is a basis of X, these vectors are linearly independent with each other and thus

$$\alpha_{l+1} = \cdots = \alpha_n = \beta_1 = \cdots = \beta_l = 0,$$

which proves that $A\boldsymbol{x}_{l+1}, \ldots, A\boldsymbol{x}_n$ are linearly independent with each other.

6.3 Null space and range

$$\boxed{\dim\mathscr{N}(A) + \dim\mathscr{R}(A) = n = \dim X.}$$

Here, $\dim X$: the amount of information that X originally has;

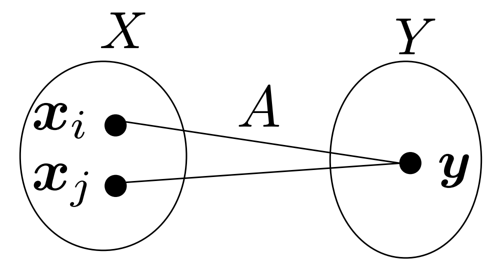
$\dim\mathscr{R}(A)$: the amount of information that the mapping A can transfer into $\mathscr{R}(A) \subseteq Y$.

$\Downarrow$

$\dim\mathscr{N}(X) = \dim X - \dim\mathscr{R}(A)$ is the amount of information that A fails to transfer.

In other words, the information that is not transferred by the mapping A is in the null space $\mathscr{N}(A)$.

6.3 Null space and range



As discussed on page 64, when

$$\dim \mathscr{R}(A) = \dim(\mathscr{A}(X)) = \mathrm{rank} A < n = \dim X,$$

there are distinct $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ that satisfy $A\boldsymbol{x}_i = A\boldsymbol{x}_j = \boldsymbol{y}$.

$\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are different from each other, but mapped by A to the identical $\boldsymbol{y}$.
The difference between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, namely $\boldsymbol{x}_i - \boldsymbol{x}_j$, is not transferred into Y.

Then where does it go to? In the first place, where does it come from?
When we subtract the left hand sides of the equations

$$A\boldsymbol{x}_i = \boldsymbol{y}, \quad A\boldsymbol{x}_j = \boldsymbol{y}$$

from each other and the right hand sides from each other, we get

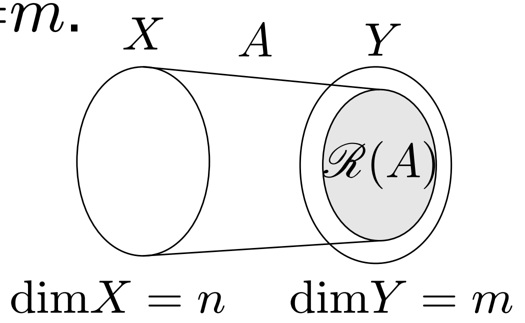
$$A\boldsymbol{x}_i - A\boldsymbol{x}_j = \boldsymbol{0}$$

or $A(\boldsymbol{x}_i - \boldsymbol{x}_j) = \boldsymbol{0}$, which implies $\boldsymbol{x}_i - \boldsymbol{x}_j \in \mathscr{N}(A)$.

That is, $\boldsymbol{x}_i - \boldsymbol{x}_j$ goes to $\boldsymbol{0}$ in Y, and it comes from $\mathscr{N}(A) \subseteq X$.

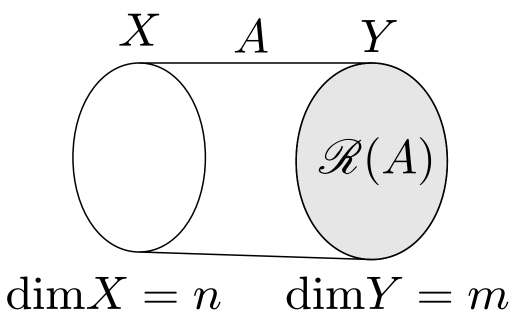
6.4 Surjection/injection and rank

Let A be an $m\times n$ matrix. Then $\dim X=n$, $\dim Y=m$.



When the linear mapping $\mathscr{A}$ represented by A is a <u>surjection</u>, $\mathscr{A}(X) = \mathscr{R}(A) = Y$ and therefore

$$\dim \mathscr{A}(X) = \dim \mathscr{R}(A) = \dim Y = m.$$



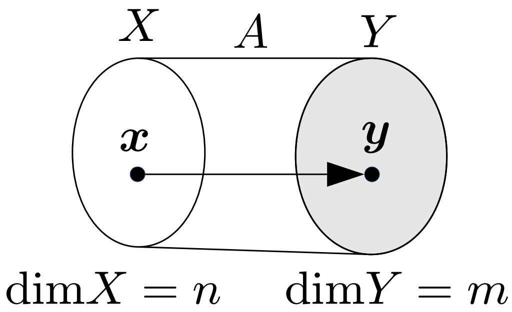
Now recall that $\dim \mathscr{A}(X) = \mathrm{rank} A$.

The above implies $\mathrm{rank} A = m$, i.e., A is of full row rank, and therefore all of its m row-vectors are linearly independent with each other.

6.4 Surjection/injection and rank

$\mathscr{A}$ is a surjection.

$\Longleftrightarrow$ Any $\boldsymbol{y} \in Y$ can be represented as $\boldsymbol{y} = A\boldsymbol{x}$ by choosing $\boldsymbol{x} \in X$ appropriately.



Suppose that Y is our destination world and that X is the origin world.

For example, imagine that Y is the surface of the moon and that X is the surface of the planet earth. In addition, imagine that the mapping A represents how a rocket travels from the earth to the moon. $\boldsymbol{y} = A\boldsymbol{x}$ implies that the rocket starts from $\boldsymbol{x}$ and reaches $\boldsymbol{y}$.

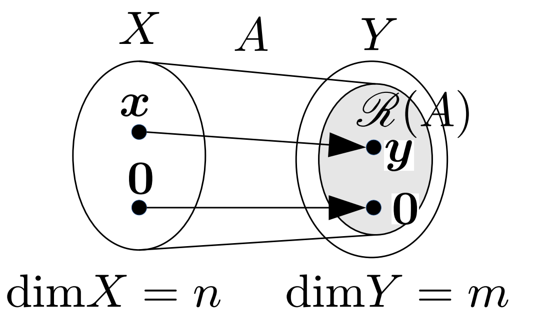
If A is a surjection, it implies that for any point $\boldsymbol{y}$ on the moon Y, we can launch a rocket that reaches $\boldsymbol{y}$ by appropriately choosing the launching point $\boldsymbol{x}$ on the earth X.

Surjection or full row rank is closely related to control problems.

6.4 Surjection/injection and rank

- Example
  - When Ms A and Mr B have a date, she pays two thirds of the meal bill and she pays the remaining one third, and for the other expenses they go Dutch.
    - Is it possible to choose the meal bill x_1 yen and other expenses x_2 yen so that Ms A pays 5,000 yen in total and Mr B pays 3,000 yen in total?
    - Is it possible to choose the meal bill x_1 yen and other expenses x_2 yen so that Ms A pays y_1 yen in total and Mr B pays y_2 yen in total for arbitrarily given values of y_1 and y_2?
    - What if she pays two thirds of all the expenses and he pays the remaining one third?
    - What if she pays two thirds of the meal bill, three fifths of the transport costs and one half of the other expenses and he pays the remaining?

6.4 Surjection/injection and rank



- When the linear mapping $\mathscr{A}$ represented by A is an <u>injection</u>, for any $\boldsymbol{y} \in \mathscr{R}(A)$ there is only one $\boldsymbol{x} \in X$ such that $\boldsymbol{y} = A\boldsymbol{x}$.
  This is the case for $\boldsymbol{y} = \mathbf{0}$ as well, and in this case the only $\boldsymbol{x}$ is $\boldsymbol{x} = \mathbf{0}$.
  This implies that $\mathscr{N}(A) = \{\mathbf{0}\}$ and therefore that $\dim \mathscr{N}(A) = 0$

  Recall that $\dim \mathscr{N}(A) + \dim \mathscr{R}(A) = n = \dim X.$

  The above implies

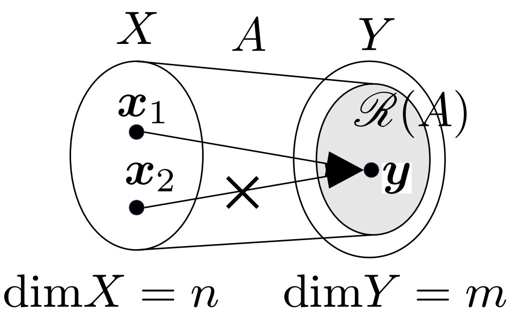
$$\mathrm{rank} A = \dim \mathscr{A}(X) = \dim \mathscr{R}(A) = n - \dim \mathscr{N}(A) = n - 0 = n,$$

  i.e., A is of full column rank, and therefore all of its n column-vectors are linearly independent with each other.

6.4 Surjection/injection and rank

$\mathscr{A}$ is an injection.

$\Longleftrightarrow$ For any $\boldsymbol{y} \in Y$, there is only one $\boldsymbol{x} \in X$ such that $\boldsymbol{y} = A\boldsymbol{x}$.



Suppose that X is the set of original signals and that Y is the set of measured signals with noises.

The mapping A represents how an original signal $\boldsymbol{x}$ is converted (and contaminated by a noise) to the measured signal $\boldsymbol{y}$.

If A is an injection, it implies that for any measured signals $\boldsymbol{y}$ with noises, we can find uniquely its original signal $\boldsymbol{x}$ without the noises.

Injection or full column rank is closely related to estimation problems.

6.4 Surjection/injection and rank

- Example
  - When Ms A and Mr B have a date, she pays two thirds of meal bills and he pays the remaining one third, and for the other expenses they go Dutch.
    - On a date, she payed 3,000 yen in total and he payed 2,000 yen in total. Then how much was the meal bill and how much were the other expenses?
    - What if she pays two thirds of meal bills, three fifths of the transport costs and one half of the other expenses and he pays the remaining?

6.4 Surjection/injection and rank

- When the linear mapping $\mathscr{A}$ represented by A is a <u>bijection</u>,
  $\mathrm{rank}A = m$ and A is of full raw rank (surjection),
  and also
  $\mathrm{rank}A = n$ and A is of full column rank (injection).

  The above implies $\mathrm{rank}A = m = n$, and therefore A is a square matrix（正方行列）.
  Besides, all the row-vectors are linearly independent with each other, and all the column-vectors are linearly independent with each other.