Chapter 1 Systems of Linear Equations

Math 39, San Jose State University

Prof. Guangliang Chen

Fall 2022

Covered sections from Chapter 1:

1.1 Systems of linear equations

1.2 Row reduction and Echelon forms

1.3 Vector equations

1.4 The matrix equation $\mathbf{Ax} = \mathbf{b}$

1.5 Solution sets of linear systems

1.7 Linear independence

1.8 Introduction to linear transformations

1.9 The matrix of a linear transformation

Systems of linear equations

A **linear equation** is one of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

where $a_1, a_2, \ldots, a_n$ are given **coefficients**, b is another given number, and $x_1, x_2, \ldots, x_n$ are the **variables** (unknowns).

For example,

$$x_1 - 2x_2 + 4x_3 = -1.$$

Note that in a linear equation, there are no terms like $\sqrt{x_i}, e^{x_i}, x_i^2, \log(x_i)$.

A **system of linear equations**, or a **linear system** in short, is a collection of linear equations involving the same variables, e.g.,

$$\begin{cases} 2x_1 - x_2 + 5x_3 & = 1 \\ x_1 - 2x_2 + 4x_3 & = -1 \\ 3x_1 + x_2 + 6x_3 & = 1 \end{cases}$$

Two fundamental questions about a linear system:

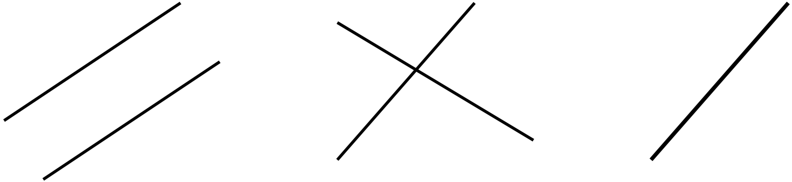
- **Existence**: Is there at least one solution?
- **Uniqueness**: If a solution exists, is it unique (i.e., the only one)?

In general, the number of solutions to a linear system could be 0, or 1, or infinity.

In the first case (no solution), the linear system is said to be inconsistent; otherwise, it is said to be consistent.

For example, in a linear system of two equations in two unknowns:

- Solutions to each single equation follow a line in the plane.
- The solution of the system are the points of intersection of the two lines.



To solve a linear system is to find its **solution set**, i.e., the set of all possible solutions of the linear system.

Two linear systems are called **equivalent** if they have the same solution set.

The basic strategy is to reduce the given linear system to an easy yet equivalent linear system through a sequence of **elementary equation operations** (which always preserve the solution set):

- Interchange two equations in the linear system
- Multiply an equation by a nonzero number
- Add a scalar multiple of one equation to another equation (to replace that equation)

Example 0.1. Solve the following linear system:

$$\begin{cases} 2x_1 - x_2 + 5x_3 & = 1 \\ x_1 - 2x_2 + 4x_3 & = -1 \\ 3x_1 + x_2 + 8x_3 & = 1 \end{cases}$$

The matrix way of solving a linear system

For the linear system in the previous example, if we let

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 5 \\ 1 & -2 & 4 \\ 3 & 1 & 8 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

then the above system can be symbolically represented as follows:

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

We call

- $\mathbf{A}$: the **coefficient matrix** of the system, and
- $\mathbf{x}$: the **vector of unkowns**.

One can further form an **augmented matrix** by concatenating $\mathbf{A}$ and $\mathbf{b}$

$$[\mathbf{A} \mid \mathbf{b}] = \begin{bmatrix} 2 & -1 & 5 & | & 1 \\ 1 & -2 & 4 & | & -1 \\ 3 & 1 & 8 & | & 1 \end{bmatrix}$$

and use it to represent the linear system instead (no loss of information).

As a result, solving a system of linear equations (based on elementary equation operations) is equivalent to performing three kinds of elementary row operations on the augmented matrix:

- *(switching) Interchange two rows*
- *(scaling) Multiply all entries in a row by a nonzero scalar*
- *(replacement) Add a nonzero multiple of a row to another row to replace it*

A few remarks:

Row operations can be applied to any matrix, not merely to one that arises as the augmented matrix of a linear system.

Row operations are *reversible*.

Two matrices are called **row equivalent** if there is a sequence of elementary row operations that transforms one matrix into the other.

If the augmented matrices of two linear systems are row equivalent, then the two systems have the same solution set (and thus are also equivalent).

Example 0.2. Solve the same linear system using the matrix method

$$\begin{cases} 2x_1 - x_2 + 5x_3 & = 1 \\ x_1 - 2x_2 + 4x_3 & = -1 \\ 3x_1 + x_2 + 8x_3 & = 1 \end{cases}$$

Row reduction and Echelon forms

When performing elementary row operations on the augmented matrix, what row reduced forms do we seek in order to solve the linear system easily?

One standard reduction for the augmented matrix is to transform it into a **row echelon form (REF)**:

$$[\mathbf{A} \mid \mathbf{b}] = \begin{bmatrix} 2 & -1 & 5 & | & 1 \\ 1 & -2 & 4 & | & -1 \\ 3 & 1 & 8 & | & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} \boxed{1} & -2 & 4 & | & -1 \\ 0 & \boxed{1} & -1 & | & 1 \\ 0 & 0 & \boxed{3} & | & -3 \end{bmatrix},$$

One can then use **back substitution** to solve the system.

The leading nonzeros of the rows, which are boxed, are called **pivots**.

Formal definition of REF: We say that a matrix is in a REF if

- The pivots always move further to the right (this implies that all the entries preceding and under any pivot must be zero).
- Zero rows (if any) are all at the bottom of the matrix.

$$\begin{bmatrix} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & \blacksquare & * & * & * & * & * & * & * & * \\ 0 & 0 & 0 & \blacksquare & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & \blacksquare & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & * \end{bmatrix}$$

$\blacksquare$: leading nonzeros (pivots); $*$: any value (including zero).

The REF of a matrix is not unique.

Another reduced form is **reduced row echelon form (RREF)**:

$$\left[\begin{array}{ccc|c} \boxed{1} & -2 & 4 & -1 \\ 0 & \boxed{1} & -1 & 1 \\ 0 & 0 & \boxed{3} & -3 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} \boxed{1} & 0 & 0 & 3 \\ 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & \boxed{1} & -1 \end{array}\right]$$

It is obtained from the REF by applying more elementary row operations:

- **Row scaling**: to transform all pivots to 1;
- **Row replacement**: to create zeros above the pivots.

The RREF is easiest for solving the linear system.

Formal definition of RREF: We say that a matrix is in the RREF if it is already in a REF, and

- The pivots are all 1.
- The entries above any pivot are all zero (thus, each pivot is the only nonzero entry in the column containing it).

$$\begin{bmatrix} 1 & 0 & * & * \\ 0 & 1 & * & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & * & 0 & 0 & 0 & * & * & 0 & * \\ 0 & 0 & 0 & 1 & 0 & 0 & * & * & 0 & * \\ 0 & 0 & 0 & 0 & 1 & 0 & * & * & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 1 & * & * & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & * \end{bmatrix}$$

The RREF of a matrix must be unique.

The pivot locations in any REF for a given matrix are fixed:

- In a REF, the pivots can change values but the locations are fixed.
- In a RREF, the pivots and their locations are all fixed (recall that the RREF of a matrix is unique).

The rows containing the pivot locations are the only nonzero rows in the REF.

The columns containing the pivot locations are called **pivot columns**.

The variables of a linear system corresponding to pivot columns in the augmented matrix are called **basic variables**, and the other variables are called **free variables**.

Example 0.3. Solve the following modified linear system:

$$\begin{cases} 2x_1 - 4x_2 + 5x_3 & = 1 \\ x_1 - 2x_2 + 4x_3 & = -1 \\ 3x_1 - 6x_2 + 8x_3 & = 1 \end{cases}$$

Existence of solution of systems of linear equations.

A linear system has no solution if and only if a REF of the augmented matrix has a row of the form

$$\left[\begin{array}{ccc|c} 0 & \cdots & 0 & \boxed{b} \end{array}\right] \quad (b \neq 0)$$

i.e., the augmented column is a pivot column.

For example,

$$\begin{cases} 2x_1 - 4x_2 + 5x_3 & = 1 \\ x_1 - 2x_2 + 4x_3 & = -1 \\ 3x_1 - 6x_2 + 8x_3 & = 4 \end{cases} \longrightarrow \left[\begin{array}{ccc|c} \boxed{1} & -2 & 4 & -1 \\ 0 & 0 & \boxed{1} & -1 \\ 0 & 0 & 0 & \boxed{3} \end{array}\right]$$

General Procedure for determining existence and uniqueness of solution of systems of linear equations

Given a linear system, first form the augmented matrix and reduce it to a REF:

- If the augmented column contains a pivot, then the system is inconsistent (i.e., has no solution)
- If the augmented column contains no pivot, then the linear system is consistent (i.e., has at least one solution).
  - If there is no free variable, then the system has a unique solution;
  - if there is at least one free variable, then the system has infinitely many solutions.

Vectors

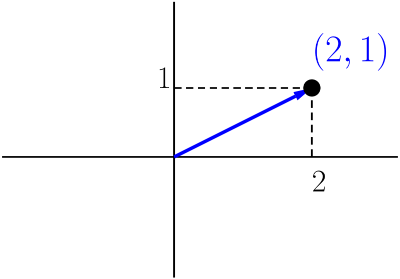
Vectors in $\mathbb{R}^2$ are ordered pairs of numbers arranged in column form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$$

They can also be regarded as matrices with a single column.

Geometrically, they are identified with points in the two dimensional plane:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \Longleftrightarrow \quad (x_1, x_2)$$



In the above figure:

vector: $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$; point: $(2, 1)$.

Vectors are associated with two kinds of operations (both are componentwise):

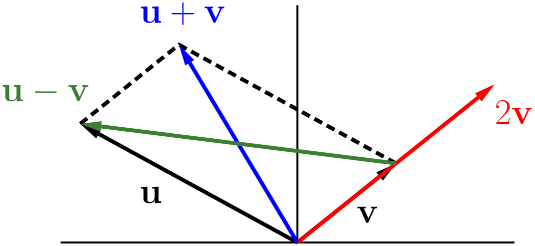
- **vector addition**: $\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix}$
- **scalar multiplication**: $c \cdot \mathbf{x} = \begin{bmatrix} cx_1 \\ cx_2 \end{bmatrix}$

Vector subtraction is defined as $\mathbf{x} - \mathbf{y} = \mathbf{x} + (-1) \cdot \mathbf{y} = \begin{bmatrix} x_1 - y_1 \\ x_2 - y_2 \end{bmatrix}$

For example,

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}, \quad 5 \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 10 \end{bmatrix}$$

Vector addition and subtraction in $\mathbb{R}^2$ follow the Parallelogram Rule:



Vectors in $\mathbb{R}^n$ are ordered n-tuples of numbers arranged in column form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$$

They are generalizations of vectors in $\mathbb{R}^2$, associated with same operations, **addition** and **scalar multiplication**, in the same componentwise fashion.

Two special vectors: $\mathbf{0}$ (all components are 0), $\mathbf{1}$ (all components are 1).

Operations among vectors of the same length satisfy the following laws:

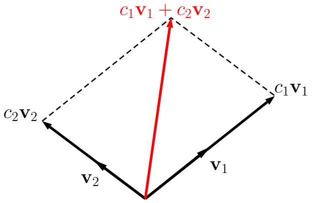
- Commutativity: $\mathbf{x}+\mathbf{y} = \mathbf{y}+\mathbf{x}$. In particular, $\mathbf{x}+\mathbf{0} = \mathbf{0}+\mathbf{x} = \mathbf{x}$.
- Associativity: $(\mathbf{x}+\mathbf{y})+\mathbf{z} = \mathbf{x}+(\mathbf{y}+\mathbf{z})$, and $a(b\mathbf{x}) = (ab)\mathbf{x}$
- Distributivity: $(a+b)\mathbf{x} = a\mathbf{x}+b\mathbf{x}$ and $a(\mathbf{x}+\mathbf{y}) = a\mathbf{x}+a\mathbf{y}$.

Linear combinations

Given k vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$, a **linear combination** of them is a vector of the form

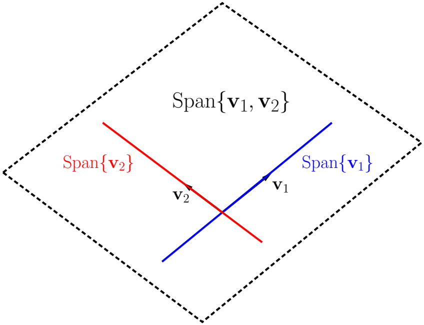
$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \sum_{i=1}^{k} c_i\mathbf{v}_i$$

where $c_1, \ldots, c_k$ are given real numbers (including zero), called **weights**.



The set of all linear combinations of these vectors is called **the set spanned by them**, or sometimes, simply their **span**:

$$\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \{\mathbf{v} = \sum_{i=1}^{k} c_i \mathbf{v}_i \mid c_1, \ldots, c_k \text{ are arbitrary real numbers}\}$$



Matrix-vector multiplication

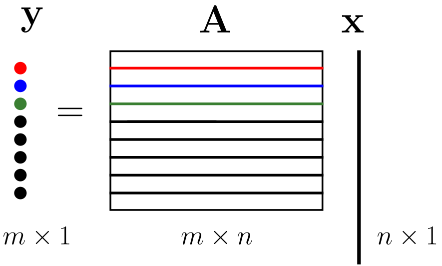
We may also multiply a matrix and a vector of "proper" length, such as

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \cdot (-1) + 2 \cdot 0 + 3 \cdot 1 \\ 4 \cdot (-1) + 5 \cdot 0 + 6 \cdot 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

In the example, the matrix has size 2×3 and the vector has size 3×1 so that they are compatible for multiplication, yielding a 2×1 product vector.

The product between each row of the matrix and the vector yields the corresponding entry of the product vector.

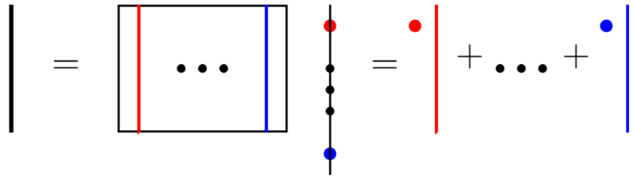
More generally, for any matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and vector $\mathbf{x} \in \mathbb{R}^n$, we can multiple them to get a vector $\mathbf{y} \in \mathbb{R}^m$, by taking the product between each row of $\mathbf{A}$ and the vector $\mathbf{x}$.



This method is called the **row-wise matrix-vector multiplication**.

Alternatively, one can multiply $\mathbf{A}$ and $\mathbf{x}$ in a **column-wise** fashion:

$$\mathbf{A}\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n$$



For example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = (-1) \cdot \begin{bmatrix} 1 \\ 4 \end{bmatrix} + 0 \cdot \begin{bmatrix} 2 \\ 5 \end{bmatrix} + 1 \cdot \begin{bmatrix} 3 \\ 6 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

These two kinds of matrix-vector multiplications are always equivalent.

Example 0.4. Let $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \end{bmatrix}$. Find $\mathbf{A1}$ using both rowwise and columnwise methods. What does the product vector represent?

Properties of matrix-vector multiplication:

- $\mathbf{A}(\mathbf{x}+\mathbf{y}) = \mathbf{A}\mathbf{x} + \mathbf{A}\mathbf{y}$
- $\mathbf{A}(c\mathbf{x}) = c(\mathbf{A}\mathbf{x})$
- $\mathbf{I}\mathbf{x} = \mathbf{x}$ and $\mathbf{O}\mathbf{x} = \mathbf{0}$ for any vector $\mathbf{x} \in \mathbb{R}^n$. Here, $\mathbf{I}$ is the **identity matrix** of size $n \times n$ and $\mathbf{O}$ is the **zero matrix** of size $n \times n$:

$$\mathbf{I} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}, \quad \mathbf{O} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The last property is easy to verify; we will prove the first two when we get to Chapter 2 Matrix Algebra.

Remark. Another special matrix is the matrix of ones, denoted by

$$\mathbf{J} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & 1 \end{bmatrix} \in \mathbb{R}^{n\times n}$$

It can be shown that

$$\mathbf{J}\mathbf{x} = \left(\sum_{i=1}^{n} x_i\right)\mathbf{1}$$

for any vector $\mathbf{x} \in \mathbb{R}^n$.

Vector equations

Previously when adopting the notation $\mathbf{A}\mathbf{x} = \mathbf{b}$ for a system of linear equations, we have already used the row-wise multiplication rule:

$$\begin{cases} 2x_1 - x_2 + 5x_3 &= 1 \\ x_1 - 2x_2 + 4x_3 &= -1 \\ 3x_1 + x_2 + 8x_3 &= 1 \end{cases} \qquad \Longleftrightarrow \qquad \mathbf{A}\mathbf{x} = \mathbf{b}$$

where

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 5 \\ 1 & -2 & 4 \\ 3 & 1 & 8 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

We can now apply the columnwise multiplication rule to rewrite the same linear system:

$$\mathbf{b} = \mathbf{A}\mathbf{x} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \mathbf{a}_3] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3.$$

More specifically,

$$\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = x_1 \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix} + x_3 \begin{bmatrix} 5 \\ 4 \\ 8 \end{bmatrix}$$

This is called the vector equation of the system, which has the same solution set.

Remark. The vector equation has the geometric interpretation that vector $\mathbf{b}$ is a linear combination of the columns of $\mathbf{A}$, if the linear system is consistent.

Theorem 0.1. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^n$. The equation $\mathbf{Ax} = \mathbf{b}$ is consistent if and only if $\mathbf{b}$ lives in the span of the columns of $\mathbf{A}$.

Proof. This is due to the following equation:

$$\mathbf{b} = \mathbf{Ax} \equiv \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \equiv \underbrace{x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n}_{\in \mathrm{span}\{\mathbf{a}_1,\ldots,\mathbf{a}_n\}}$$

$\square$

To solve a vector equation, i.e.,

$$\mathbf{b} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$$

we still write it into matrix form

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \qquad \text{where} \quad \mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

and work on the augmented matrix:

$$\left[\begin{array}{cccc|c} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n & \mathbf{b} \end{array}\right]$$

Example 0.5. Let

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}$$

Determine if $\mathbf{b} \in \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$, i.e., if $\mathbf{b}$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2$.

Solution sets of linear systems

We classify a linear system $\mathbf{Ax} = \mathbf{b}$ into one of the following two categories, based on the vector $\mathbf{b}$:

- **Homogeneous**: $\mathbf{b} = \mathbf{0}$
- **Nonhomogeneous**: $\mathbf{b} \neq \mathbf{0}$

and study their solution sets separately.

Homogeneous linear systems ($\mathbf{Ax} = \mathbf{0}$)

For example, the following is a homogeneous system of linear equations

$$\begin{cases} 2x_1 - x_2 + 5x_3 & = 0 \\ x_1 - 2x_2 + 4x_3 & = 0 \\ 3x_1 + 0x_2 + 6x_3 & = 0 \end{cases}$$

What can we say about the existence and uniqueness of solutions of such systems?

Existence (✓): A trivial solution $\mathbf{x} = \mathbf{0}$ (always) exists.

Uniqueness: Does the system have a nontrivial (i.e., nonzero) solution as well?

Answer: It depends. In the preceding example,

$$\left[\begin{array}{ccc|c} 2 & -1 & 5 & 0 \\ 1 & -2 & 4 & 0 \\ 3 & 0 & 6 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} \boxed{1} & 0 & 2 & 0 \\ 0 & \boxed{1} & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Due to the free variable (x_3), there are infinitely many solutions:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2x_3 \\ x_3 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \longleftarrow \text{parametric equation of a line}$$

Can you give an example of a homogeneous linear system that has only the trivial solution $\mathbf{x} = \mathbf{0}$?

Summary: Homogeneous linear systems ($\mathbf{Ax} = \mathbf{0}$) are always consistent (with a trivial solution $\mathbf{x} = \mathbf{0}$).

To determine if the system have more solutions (which are nonzero), take the coefficient matrix $\mathbf{A}$ and reduce it to a REF:

- If there are no free variables (i.e., every column has a pivot), then the system only has the zero solution;
- If there is at least a free variable, then the system has infinitely many solutions.

Solution sets of nonhomogeneous systems

Consider the following nonhomogeneous linear system:

$$\begin{cases} 2x_1 - x_2 + 5x_3 & = 0 \\ x_1 - 2x_2 + 4x_3 & = -1 \\ 3x_1 + 0x_2 + 6x_3 & = 1 \end{cases}$$

The system is consistent with the following general solution

$$x_1 = -2x_3 + \frac{1}{3}, \quad x_2 = x_3 + \frac{2}{3}, \quad x_3 \text{ is free}$$

The parametric form of the general solution of the nonhomogeneous system is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2x_3 + \frac{1}{3} \\ x_3 + \frac{2}{3} \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ 0 \end{bmatrix}$$

How is it different from the general solution of the corresponding homogeneous system?

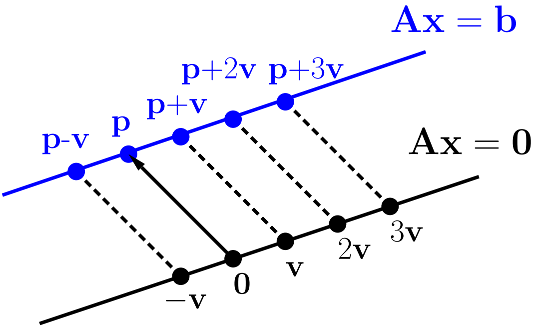
They differ only by a particular solution of the nonhomogeneous system!

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2x_3 + \frac{1}{3} \\ x_3 + \frac{2}{3} \\ x_3 \end{bmatrix} = \underbrace{x_3 \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}}_{\text{general solution of } \mathbf{Ax}=\mathbf{0}} + \underbrace{\begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ 0 \end{bmatrix}}_{\text{particular solution}}$$

That is, the solution set of $\mathbf{Ax} = \mathbf{b}$ is equal to the solution set of $\mathbf{Ax} = \mathbf{0}$ plus a particular solution of $\mathbf{Ax} = \mathbf{b}$.

Denote $\mathbf{v} = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{p} = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ 0 \end{bmatrix}$. Then $\mathbf{x} = x_3\mathbf{v} + \mathbf{p}$ (x_3 is a free parameter).

Geometric interpretation: The solution set of $\mathbf{Ax} = \mathbf{b}$ is that of $\mathbf{Ax} = \mathbf{0}$ shifted along the vector $\mathbf{p}$!



Summary: Suppose we are given a nonhomogeneous linear system $\mathbf{Ax} = \mathbf{b}$.

If it is consistent and has a particular solution $\mathbf{p}$, then the solution set of the nonhomogeneous linear system $(\mathbf{Ax} = \mathbf{b})$ is the solution set of the corresponding homogeneous linear system $(\mathbf{Ax} = \mathbf{0})$ shifted by the vector $\mathbf{p}$.

Two follow-up questions:

- Is the choice of $\mathbf{p}$ unique?
- What if the nonhomogeneous linear system $\mathbf{Ax} = \mathbf{b}$ is inconsistent?

Another question: How to find a particular solution $\mathbf{p}$ for the (consistent) nonhomogeneous system $\mathbf{A}\mathbf{x} = \mathbf{b}$ easily, when given the solution set of the homogeneous system $\mathbf{A}\mathbf{x} = \mathbf{0}$?

Example 0.6. Describe the solution sets of the homogeneous system (of only one equation)

$$x_1 - 3x_2 + 2x_3 = 0$$

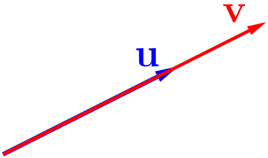
and the nonhomogeneous system

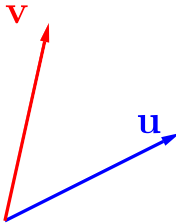
$$x_1 - 3x_2 + 2x_3 = 1$$

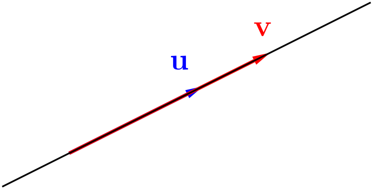
as well as their relationship.

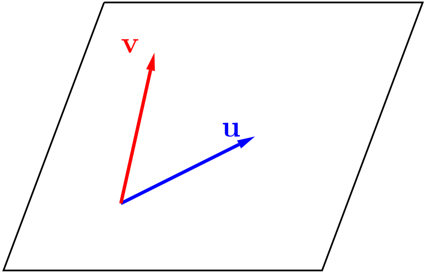
Linear independence

Consider the span of the two vectors in each case:





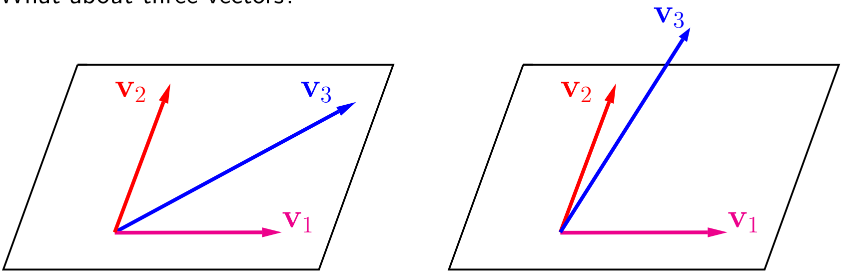




In the first case (colinear), the two vectors are said to be **linearly dependent** (because $\mathbf{v} = c\mathbf{u}$ for some number c), while in the second case, the two vectors are said to be **linearly independent** ($\mathbf{v} \neq c\mathbf{u}$ for any c).

Note also that $\mathbf{v} = c\mathbf{u}$ can be rewritten as $c\mathbf{u} - \mathbf{v} = 0$, or $c \cdot \mathbf{u} + (-1) \cdot \mathbf{v} = \mathbf{0}$, indicatin that $\mathbf{u}, \mathbf{v}$ can be linearly combined to become the zero vector.

What about three vectors?



Left (co-planar): Linearly dependent, as

$$\mathbf{v}_3 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2, \qquad \text{or equivalently,} \quad c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + (-1)\cdot\mathbf{v}_3 = \mathbf{0}.$$

Right: linearly independent.

Def 0.1. A set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ are said to be **linearly dependent** if there exist scalars $c_1, \ldots, c_k$ that are not all zero such that

$$[\mathbf{v}_1 \ldots \mathbf{v}_k] \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}.$$

Otherwise, if the above equation only has the trivial solution $\mathbf{c} = \mathbf{0}$, then the vectors are said to be **linearly independent**.

Remark. When a nontrivial solution exists (in which case, the vectors are linearly dependent), at least one of the coefficients is nonzero. Suppose $c_1 \neq 0$. Then

$$\mathbf{v}_1 = -\frac{c_2}{c_1}\mathbf{v}_2 - \cdots - \frac{c_k}{c_1}\mathbf{v}_k.$$

This shows that $\mathbf{v}_1$ is a linear combination of the other vectors.

How to check for linear independence

Algorithm: Given vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$, solve the equation for $c_1, \ldots, c_k$:

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = [\mathbf{v}_1 \ldots \mathbf{v}_k] \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix} = \mathbf{0}$$

There are two cases:

- The system has only the trivial solution (i.e., $\mathbf{c} = \mathbf{0}$): linear independence;
- There are nontrivial solutions: linear dependence.

Example 0.7. Determine if the following vectors are linearly independent:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$$

Some quick facts about linear (in)dependence

Consider a set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subset \mathbb{R}^n$. If the set contains the zero vector (i.e., certain $\mathbf{v}_i = \mathbf{0}$), then such a set must be linearly dependent:

$$0 \cdot \mathbf{v}_1 + \cdots + 1 \cdot \mathbf{v}_i + \cdots + 0 \cdot \mathbf{v}_k = \mathbf{0}.$$

Assuming that they are all nonzero:

(1) If $k = 1$ (single-vector set $\{\mathbf{v}_1\}$): The set is linearly independent.

(1) If $k = 2$ (i.e., $\{\mathbf{v}_1, \mathbf{v}_2\}$): The set is linearly dependent when $\mathbf{v}_1, \mathbf{v}_2$ are scalar multiples of each other, and linearly independent when they have a nonzero angle.

(3) If $k > n$ (e.g., 3 vectors in $\mathbb{R}^2$, 5 vectors in $\mathbb{R}^3$): The set must be linearly dependent.

Example 0.8. Determine in each case if the vectors are linearly independent.

(1) $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} 7 \\ 8 \end{bmatrix}$

(2) $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

(3) $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 4 \\ 6 \\ 7 \end{bmatrix}$

(4) $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

Linear transformations

Recall that a system of m linear equations in n unknowns may be viewed as

- a matrix equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ (a solution exists if and only if the last column of the augmented matrix $\begin{bmatrix} \mathbf{A} & | & \mathbf{b} \end{bmatrix}$ contains no pivot), or
- a vector equation $x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$ (a solution exists if and only if $\mathbf{b} \in \mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$, i.e., the vector $\mathbf{b}$ lives in the span of the columns of $\mathbf{A}$).

Next, we look at the linear system from a **transformation** point of view:

$$\mathbf{A} : \mathbf{x} \in \mathbb{R}^n \quad \longmapsto \quad \mathbf{A}\mathbf{x} \in \mathbb{R}^m.$$

(This is a generalization of functions $f : \mathbb{R} \mapsto \mathbb{R}$)

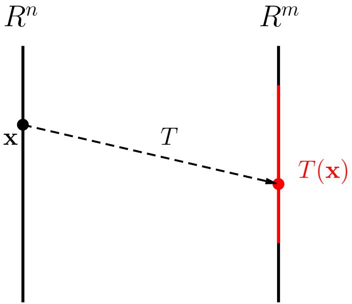
Def 0.2. A transformation (or function, or mapping) T from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a rule that assigns to each vector $\mathbf{x}$ in $\mathbb{R}^n$ a vector $T(\mathbf{x})$ in $\mathbb{R}^m$:

$$T : \mathbf{x} \in \mathbb{R}^n \quad \longmapsto \quad T(\mathbf{x}) \in \mathbb{R}^m.$$

In the above,

- $\mathbb{R}^n$ is called the **domain** of T,
- $\mathbb{R}^m$ is called the **co-domain**, or **target space**, of T
- $T(\mathbf{x}) \in \mathbb{R}^m$ is called the **image** of $\mathbf{x} \in \mathbb{R}^n$ (under T). The set of images of all points in the domain is called the range of T:

$$\mathrm{Range}(T) = \{T(\mathbf{x}) \in \mathbb{R}^m \mid \mathbf{x} \in \mathbb{R}^n\}$$



(A transformation is like a rocket)

Example 0.9. Consider the transformation

$$T : \mathbb{R} \mapsto \mathbb{R}, \qquad \text{with} \quad T(x) = x^2.$$

Determine the domain, co-domain (target space), and range of T.

Example 0.10. Let $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & -1 \end{bmatrix}$. Then the matrix $\mathbf{A}$ may be used to define a transformation

$$T : \mathbb{R}^3 \longmapsto \mathbb{R}^2, \qquad \text{with} \quad T(\mathbf{x}) = \mathbf{A}\mathbf{x}.$$

Answer the following questions:

- What are the domain and codomain of T?
- What is the image of $\mathbf{x} = \begin{bmatrix} 0 \\ 2 \\ -1 \end{bmatrix}$?
- Which points in $\mathbb{R}^3$ have an image of $\mathbf{0} \in \mathbb{R}^2$?
- What is the range of T?

Def 0.3. A transformation $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ is said to be **linear** if

- $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$
- $T(c \cdot \mathbf{x}) = c \cdot T(\mathbf{x})$ for all scalars $c \in \mathbb{R}$ and vectors $\mathbf{x} \in \mathbb{R}^n$.

Remark. These two conditions imply the following:

- $T(\mathbf{0}) = \mathbf{0}$;
- For all scalars $c, d \in \mathbb{R}$ and vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$T(c \cdot \mathbf{x} + d \cdot \mathbf{y}) = c \cdot T(\mathbf{x}) + d \cdot T(\mathbf{y}).$$

In fact, the bottom statement also implies the two conditions of the defintion, thus they are fully equivalent.

A very important result

Theorem 0.2. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be any matrix (with m rows and n columns). Then the transformation

$$T : \mathbb{R}^n \longmapsto \mathbb{R}^m, \qquad \text{with} \quad T(\mathbf{x}) = \mathbf{A}\mathbf{x}$$

is always linear.

Proof. It suffices to verify the two conditions:

$$\begin{aligned} T(\mathbf{x}+\mathbf{y}) &= \mathbf{A}(\mathbf{x}+\mathbf{y}) = \mathbf{A}\mathbf{x} + \mathbf{A}\mathbf{y} = T(\mathbf{x}) + T(\mathbf{y}) \\ T(c\cdot\mathbf{x}) &= \mathbf{A}(c\cdot\mathbf{x}) = c\cdot(\mathbf{A}\mathbf{x}) = c\cdot T(\mathbf{x}). \end{aligned}$$

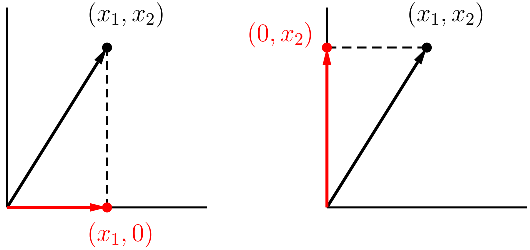
□

Example 0.11. From the theorem, we know the following transformations

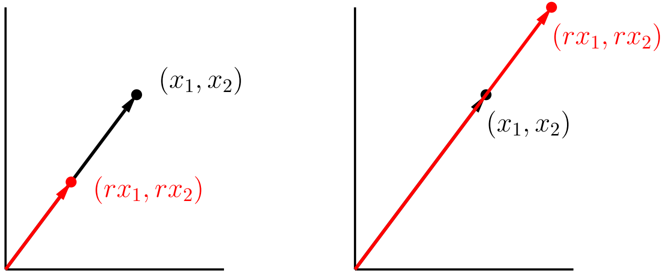
$$T : \mathbb{R}^2 \longmapsto \mathbb{R}^2, \qquad \text{with} \quad T(\mathbf{x}) = \mathbf{A}\mathbf{x}.$$

are all linear:

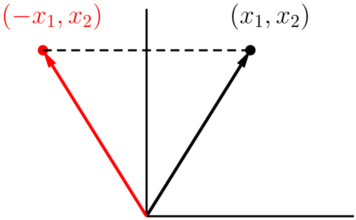
- Projections: $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ (horizontal), $\mathbf{A} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ (vertical)

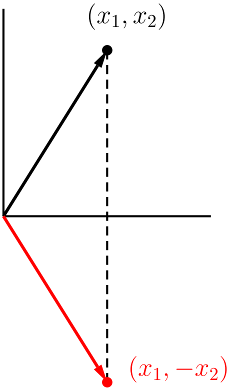


- Contractions: $\mathbf{A} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}$ (where $0 < r < 1$)
- Dilations: $\mathbf{A} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}$ (where $r > 1$)

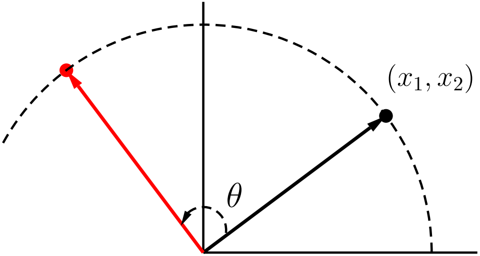


- Reflections: $\mathbf{A} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$ (horizontal), $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ (vertical)

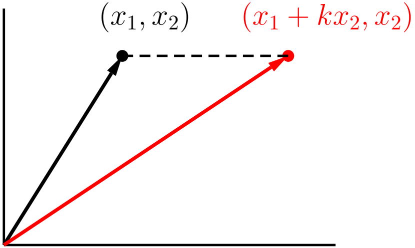


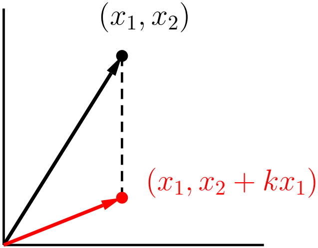


- Rotations: $\mathbf{A} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$ (counterclockwise by θ)



- Shear transformations: $\mathbf{A} = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$ (horizontal), $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$ (vertical)





We have showed that **all transformations of the form $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$ are linear**.

A natural question then arises: Is this all?

In other words, are there other kinds of linear transformations?

We answer this question during the next few slides.

First, we define the following vectors in $\mathbb{R}^n$ (called standard vectors):

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots \quad \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

What we know about these vectors:

- $\mathbf{I} = [\mathbf{e}_1\, \mathbf{e}_2 \ldots \mathbf{e}_n]$ is called the identity matrix.
- $\mathbf{I}\mathbf{v} = \mathbf{v}$ for any $\mathbf{v} \in \mathbb{R}^n$.
- The matrix equation $\mathbf{I}\mathbf{x} = \mathbf{b}$ (for any $\mathbf{b}$) has a unique solution $\mathbf{x} = \mathbf{b}$.
- The vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are linearly independent.

Another important result

Theorem 0.3. Let $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a linear transformation. Then there exists a unique matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ such that

$$T(\mathbf{x}) = \mathbf{A}\mathbf{x}, \qquad \text{for all vectors} \quad \mathbf{x} \in \mathbb{R}^n.$$

In fact, the matrix $\mathbf{A}$ must have the form

$$\mathbf{A} = [T(\mathbf{e}_1) \ldots T(\mathbf{e}_n)] \longleftarrow \text{standard matrix for } T$$

Remark. This theorem shows that $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$ are the only linear transformations.

Proof. For any $\mathbf{x} \in \mathbb{R}^n$, write

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ x_n \end{bmatrix} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n.$$

Then by linearity of T,

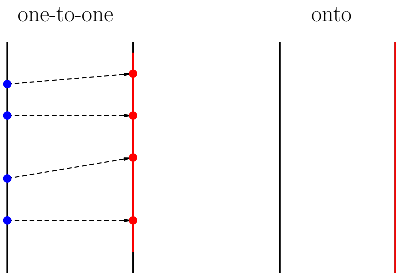
$$T(\mathbf{x}) = x_1 T(\mathbf{e}_1) + \cdots + x_n T(\mathbf{e}_n) = [T(\mathbf{e}_1) \ldots T(\mathbf{e}_n)] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{A}\mathbf{x}.$$

This completes the proof. $\square$

Example 0.12. Determine the linear transformation that maps the points $(2, 0), (1, 1)$ in $\mathbb{R}^2$ to $(-1, 0), (0, -1)$ in $\mathbb{R}^2$, respectively.

Def 0.4. Let $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ be any mapping. We say that

- T is **one-to-one** if no two vectors in the domain are mapped to the same vector in the co-domain;
- T is **onto** if every vector in the co-domain is an image of some vector(s) in the domain.



Remark. The one-to-one definition is equivalent to "Whenever $T(\mathbf{x}) = T(\mathbf{y})$, we must have $\mathbf{x} = \mathbf{y}$", while the onto definition is equivalent to "$T(\mathbf{x}) = \mathbf{b}$ (for any $\mathbf{b}$) always has a solution".

Theorem 0.4. Let $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a linear transformation, with standard matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then

(a) T is one-to-one if and only if the equation $\mathbf{A}\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(b) T is onto if and only if the equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a solution for any vector $\mathbf{b} \in \mathbb{R}^m$.

The following corollary follows directly from the preceding theorem.

Corollary 0.5. Let $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a linear transformation with standard matrix $\mathbf{A}$ (i.e., $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$). Then

- T is one-to-one if and only if the columns of $\mathbf{A}$ are linearly independent.
- T is onto if and only if the columns of $\mathbf{A}$ span $\mathbb{R}^m$.

Example 0.13. Determine in each case, if the linear transformation $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$ is one-to-one, or onto, or both.

- $\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$
- $\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 5 & 7 \\ 1 & 3 \end{bmatrix}$
- $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$