Mathematical Foundations of Machine Learning

Lecture 1

Welcome to CS 556

- Today…
  - Course overview and logistics
  - What will we do in this course?
  - Linear Algebra

Overview (approximate)

- Linear Algebra (Vector and Matrices)
- Dimensionality Reduction
- Probability and Distributions
- Vector Calculus
- Optimization
- Linear/Logistic Regression/Decision Trees/SVM

Machine Learning

- Machine Learning is about designing algorithms that automatically extract valuable information from data
- Three Main Blocks
  - Data
  - Model
  - Learning

Why Math for ML?

- To understand **fundamental principles** upon which more complicated ML systems are built
- Help creating **new** machine learning **solutions**
- Helps in **understanding** and **debugging** existing approaches
- Helps in understanding the **assumptions** and **limitations** of the methodologies we are working with

Linear Algebra

- Linear Algebra is the study of vectors and certain rules to manipulate vectors.
- We represent numerical data as vectors

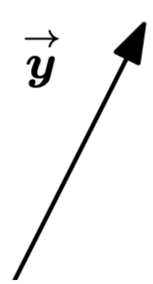
Vectors

- An algebraic vector is ordered list of elements, where the number of elements determine the dimensionally of the vector.

  Example $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \in \mathbb{R}^2$

- A geometric vector is a straight line with some length and some direction.

  Example:



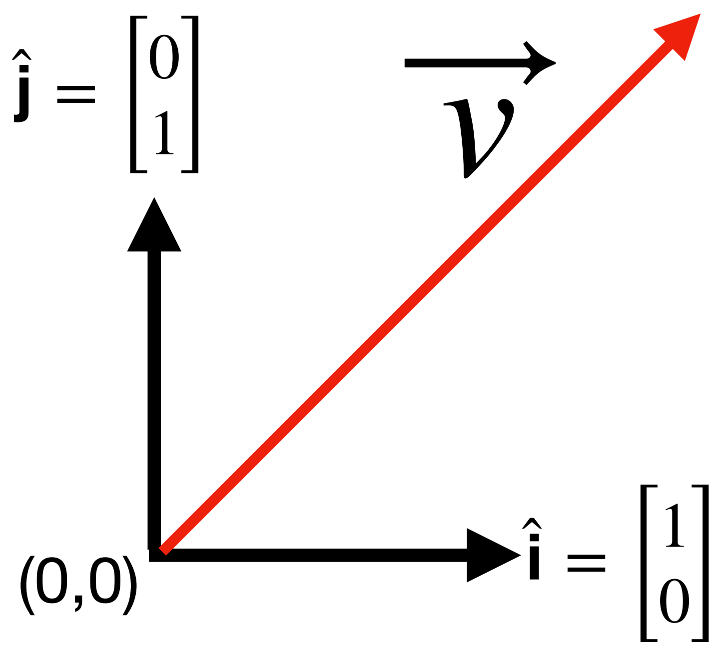
Vectors

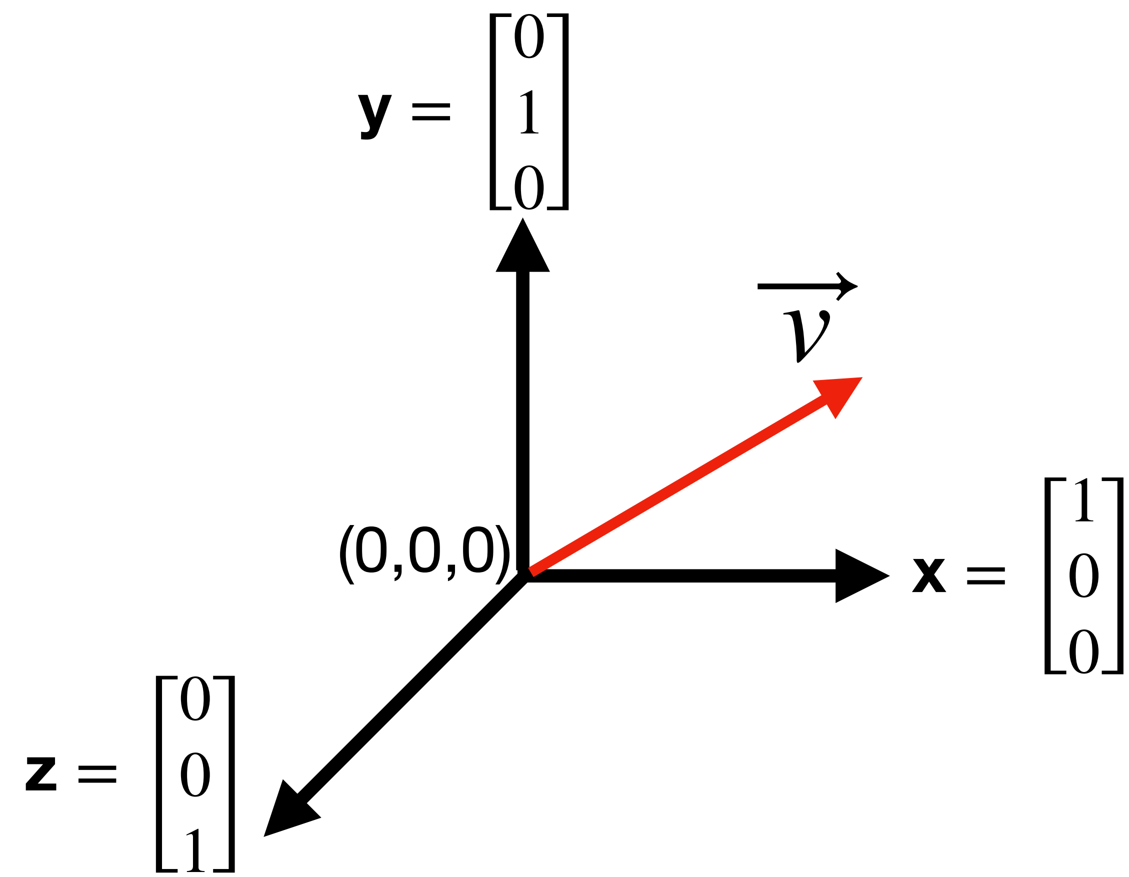
- Vectors - tuples *n* of real numbers $\mathbb{R}^n$

3 bedrooms
2 bathrooms
2 floors,
Year 2008
Price 150K

$$\longrightarrow \quad a = \begin{bmatrix} 3 \\ 2 \\ 2008 \\ 150 \end{bmatrix}$$

How to express vectors?





$$\mathbf{v} = a\hat{i} + b\hat{j} = \begin{bmatrix} a \\ b \end{bmatrix} \in \mathbb{R}^2$$

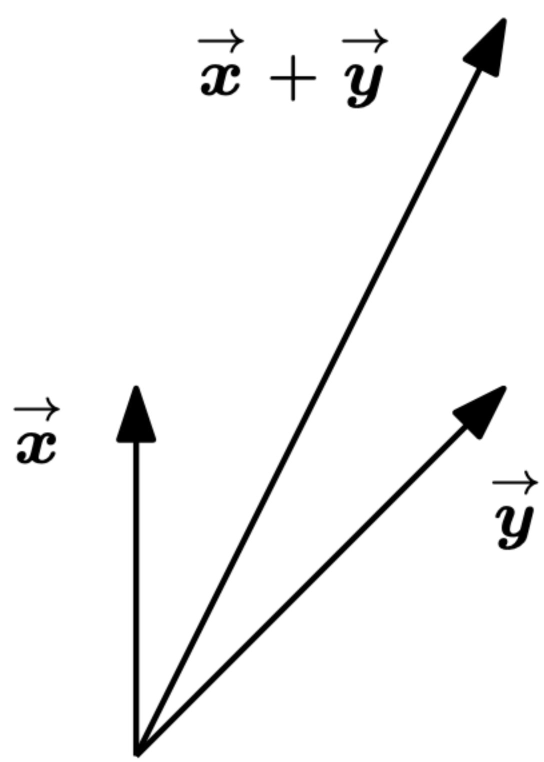
$$\mathbf{v} = ax + by + cz = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \in \mathbb{R}^3$$

Vectors Operations

- Addition/Subtraction
- Scalar Multiplication

Addition/Subtraction

- Add/Subtract elements across corresponding dimensions.
- Put the tail of one vector at the head of the other vector.



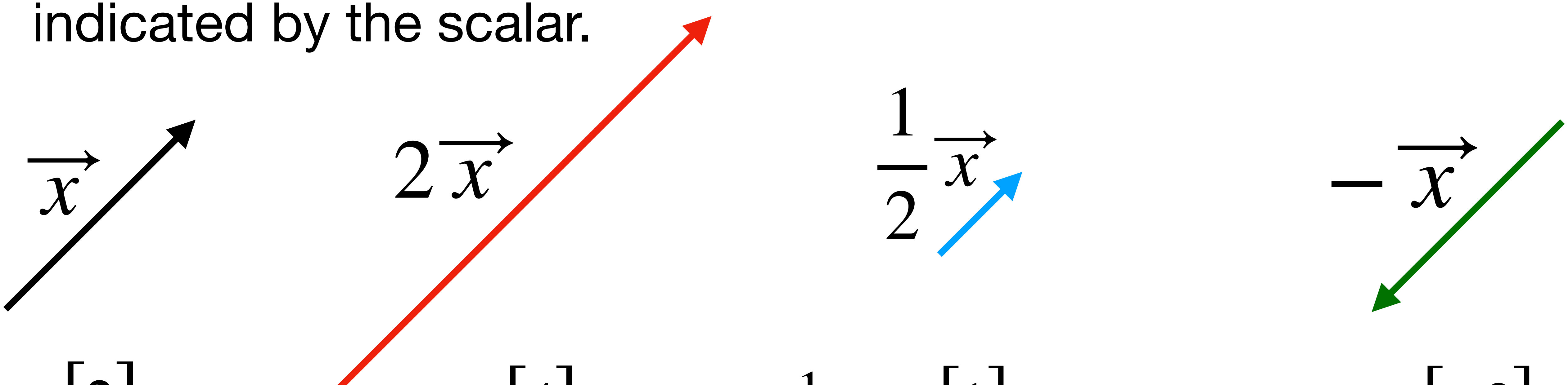
$$\mathbf{x} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

$$\mathbf{x} - \mathbf{y} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} + \begin{bmatrix} -2 \\ -2 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

Scalar Multiplication

- Scalar: A number, represented by a lower case greek letter such as α, β, λ
- Algebraic: $\lambda\mathbf{x}$ multiply each element of the vector by the scalar
- Geometric: Stretch or shrink the vector by the amount indicated by the scalar.



$$x = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \qquad 2x = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \qquad \frac{1}{2}x = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad -x = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$

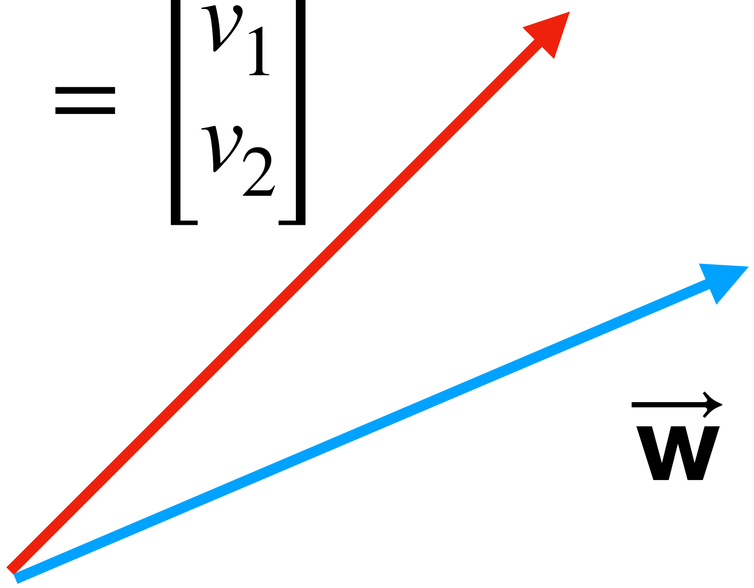
Dot Product

Dot product or **scalar product** is an algebraic operation that takes two equal-length sequences of numbers and returns a single number.

$$\mathbf{x} \cdot \mathbf{y} = < \mathbf{x}, \mathbf{y} > = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^{n} x_i y_i$$

Dot Product - Example

$$\vec{\mathbf{v}} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

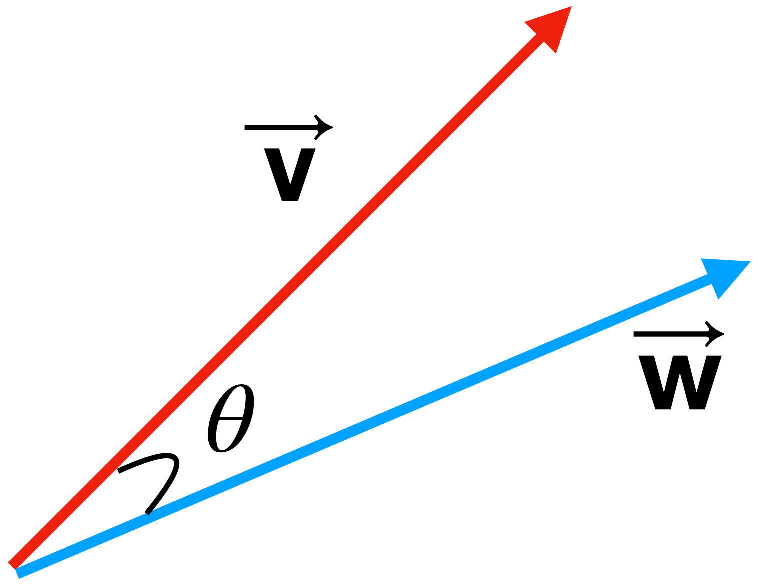


$$\vec{\mathbf{w}} = \begin{bmatrix} 4 \\ 3 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$v \cdot w = v_1 w_1 + v_2 w_2 = 3 \times 4 + 4 \times 3 = 24$$

Dot Product

Cosine of the angle between the vectors scaled by the product of the lengths of these vectors.

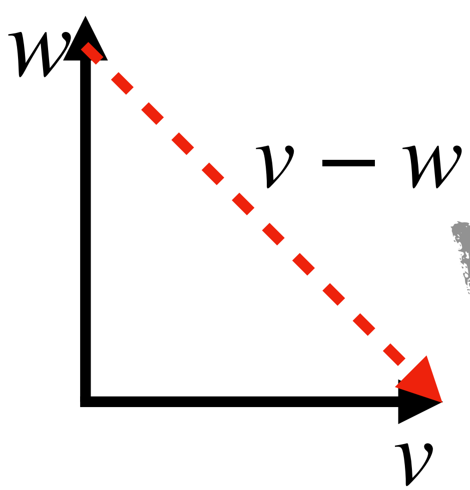


$$v \cdot w = v^T w = ||v|| \; ||w|| \; cos(\theta)$$

Angle between two vectors

The dot product is $v \cdot w = 0$ when v is perpendicular to w.

Proof: When v and w are perpendicular, they form the sides of a right triangle. The hypotenuse is $v - w$.



Why length = v - w?

$$||v||^2 + ||w||^2 = ||v - w||^2$$

$$(v_1^2 + v_2^2) + (w_1^2 + w_2^2) = (v_1 - w_1)^2 + (v_2 - w_2)^2$$

$$-2v_1w_1 - 2v_2w_2 = 0$$

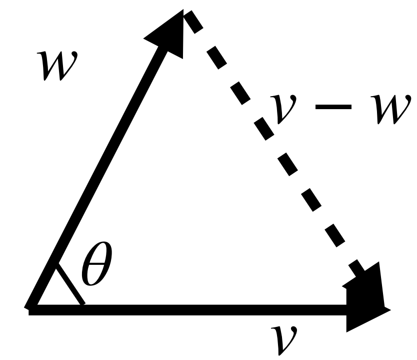
$$v_1w_1 + v_2w_2 = 0$$

$$v \cdot w = 0$$

Cosine Formula for Dot Product

Let v, w be two non-zero vectors in $\mathbb{R}^n$, then:

$$v \cdot w = v^T w = ||v|| \; ||w|| \; cos(\theta)$$



Proof:

$$||v - w||^2 = ||v||^2 + ||w||^2 - 2||v|| \; ||w|| cos(\theta)$$

Cosine Law

$$||v - w||^2 = (v - w) \cdot (v - w) = v \cdot v - 2(v \cdot w) + w \cdot w$$
$$= ||v||^2 - 2(v \cdot w) + ||w||^2$$

$$||v||^2 - 2(v \cdot w) + ||w||^2 = ||v||^2 + ||w||^2 - 2||v|| \; ||w|| cos(\theta)$$

$$v \cdot w = ||v|| \; ||w|| \; cos(\theta)$$

Properties of Dot Product

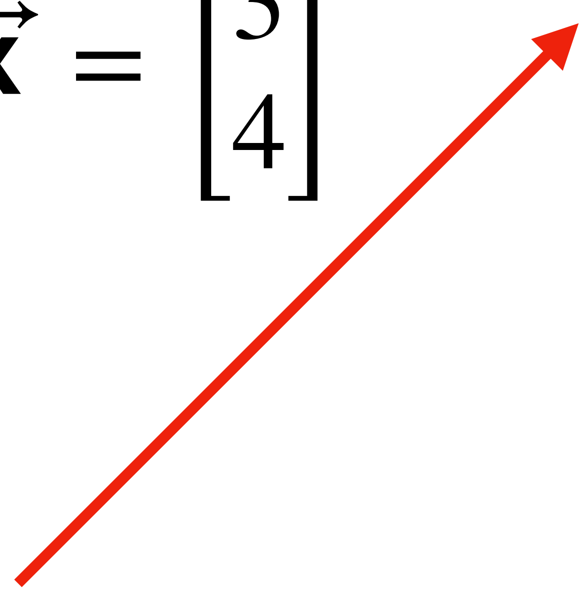
- Distributive $a^T(b+c) = a^Tb + a^Tc$
- Not Associative: $a^T(b^Tc) \neq (a^Tb)^Tc$
- Commutative: $a^Tb = b^Ta$

Why Not Associative?

Vector Length/Magnitude/Norm

$$||\mathbf{x}|| = \sqrt{\mathbf{x}^T\mathbf{x}} = \sum_{i=1}^{n} x_i^2$$

$$\vec{\mathbf{x}} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$



$$\mathbf{x} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

$$||\mathbf{x}|| = \sqrt{3^2 + 4^2} = 5$$

Unit Vectors

Unit Vector: Vector with length of 1

$$\mu\mathbf{x} \;\; s.t. \;\; ||\mu\mathbf{x}|| = 1$$

How to choose μ?

$$\mu = \frac{1}{||\mathbf{x}||}$$

Vector Properties

Operations of special interest in linear algebra are:

1. Vector Addition

2. Scalar Multiplication

If **u**, **v**, **w** are three vectors, some important properties of vector addition are…

Tan, P.-N., Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining*. Pearson.

Vector Addition Properties

1. Commutativity: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
2. Associativity: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
3. Existence of an **Identity** Element: $\mathbf{u} + \mathbf{0} = \mathbf{u}$
4. Existence of Additive Inverse: $\mathbf{u} + (\mathbf{-u}) = \mathbf{0}$

The zero vector

An inverse is some entity (a vector in the current context) when *added* to the original vector results in the '**Identity**'

Tan, P.-N., Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining*. Pearson.

Properties of Vectors Multiplied by a Scalar

Additionally let us consider 2 scalars α, β

1. Associativity: $\alpha(\beta\mathbf{u}) = (\alpha\beta)\mathbf{u}$

2. Existence of Scalar Identity: $\alpha = 1$

3. Distributivity:

   A. Scalar Multiplication over Vector Addition: $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$

   B. Scalar Addition over Multiplication of Scalar:
   $(\alpha + \beta)\mathbf{u} = \alpha\mathbf{u} + \beta\mathbf{u}$

Tan, P.-N., Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining*. Pearson.

Vector Space

A *vector space* is defined as a set of all vectors that can be created by taking linear combinations of some vectors or a set of vectors.

Formally, a vector space is the set of all points that satisfy the following conditions (in addition to previously stated vector properties):

What does it mean to be "closed"?

1. Must be **closed** under addition and scalar multiplication

2. Must contain the zero vector

$$\forall x, y \in \mathbf{V}, \forall \lambda,\ \mu \in \mathbb{R};\ \ \lambda \mathbf{x} + \mu \mathbf{y} \in \mathbf{V}$$

Linear Combinations

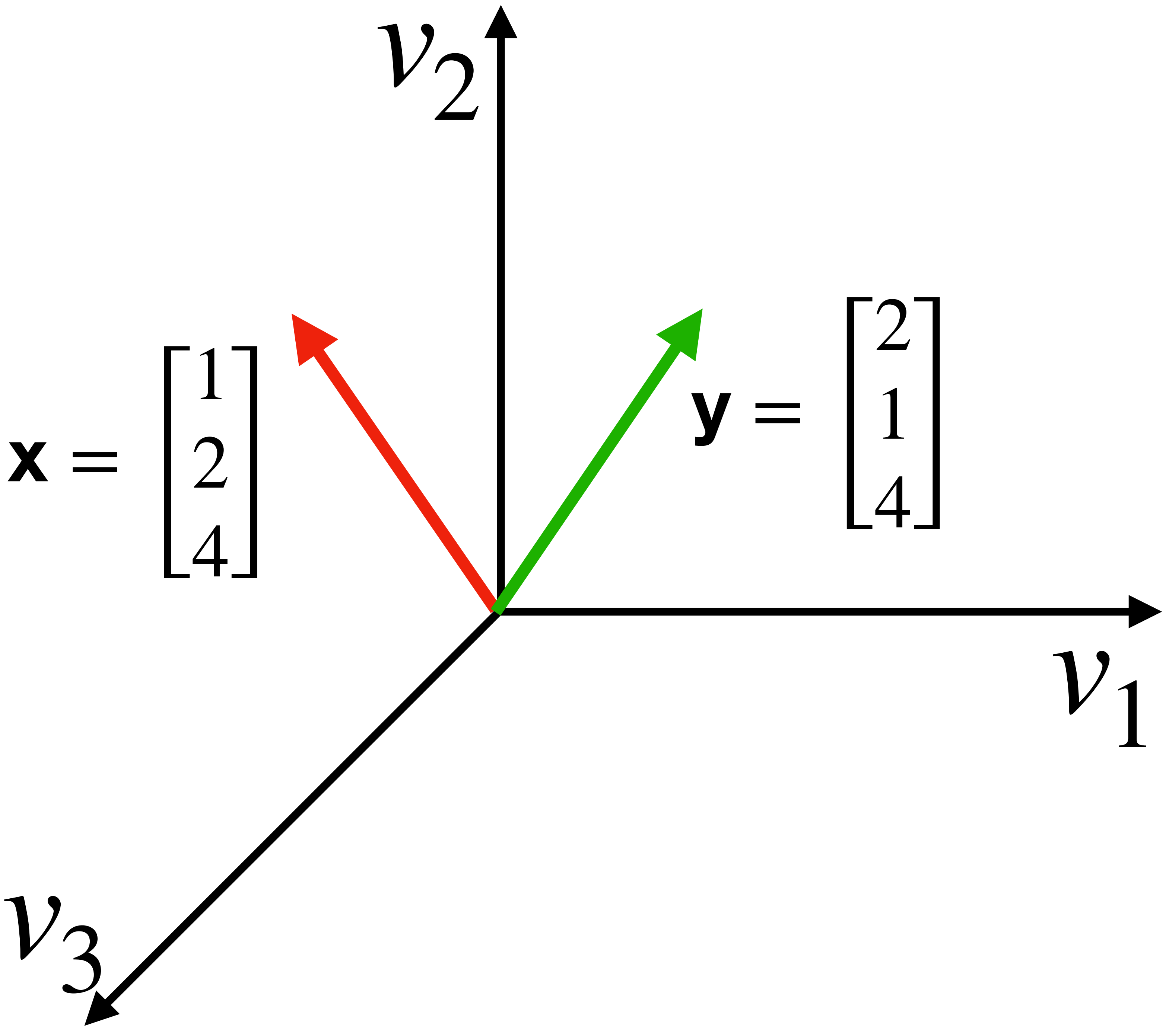
- Linear combinations of vectors are created by combining addition with scalar multiplication.

- For instance, assume we have two vectors $\mathbf{v}$ and $\mathbf{w}$ and c and d are two scalars. The sum of $c\mathbf{v}$ and $d\mathbf{w}$ is a linear combination $c\mathbf{v} + d\mathbf{w}$.

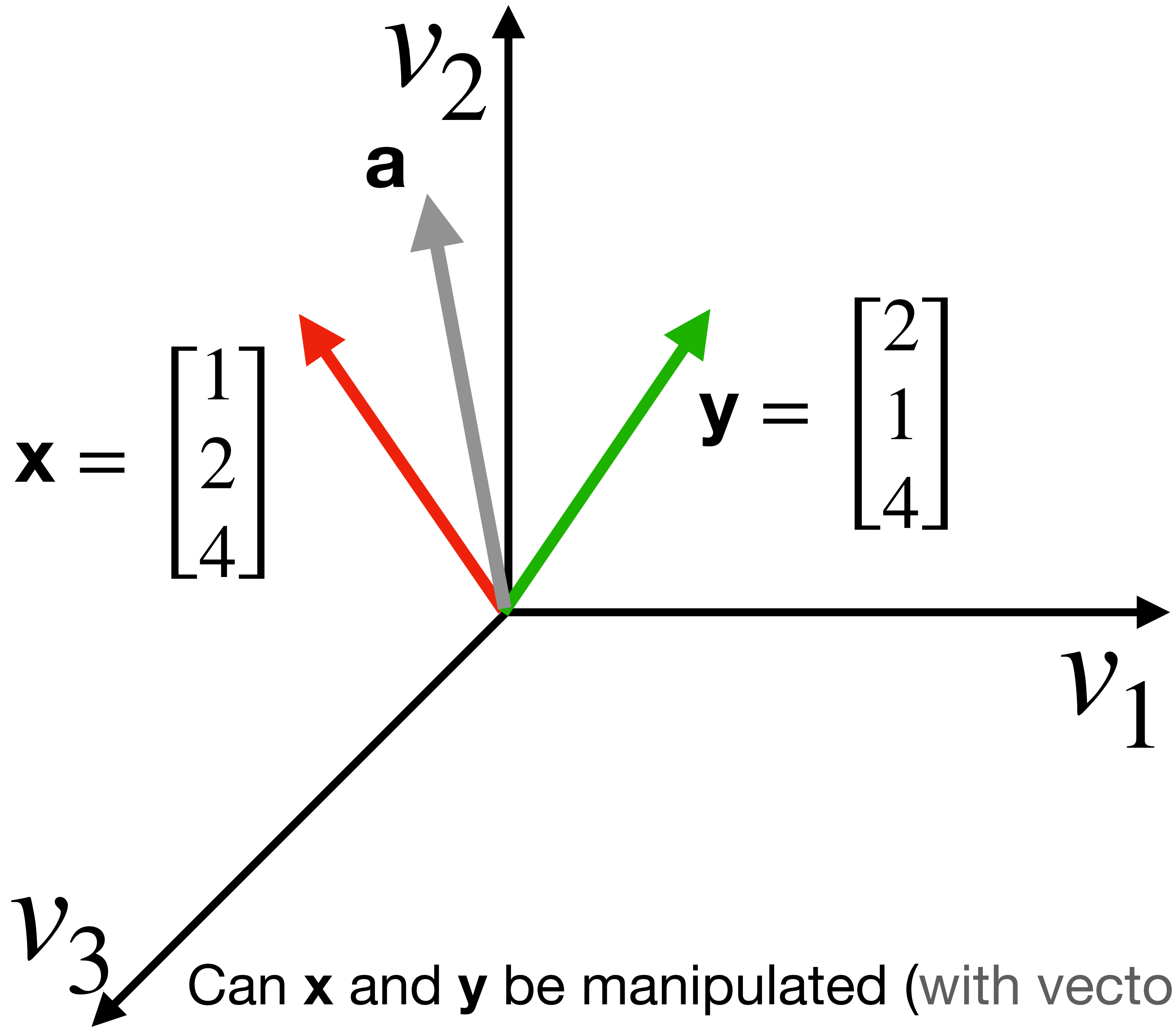
Span

Span of a set of vectors *S (in a vector space V)* is defined as all possible vectors in *V* that can be 'reached' using only the vectors from *S*.

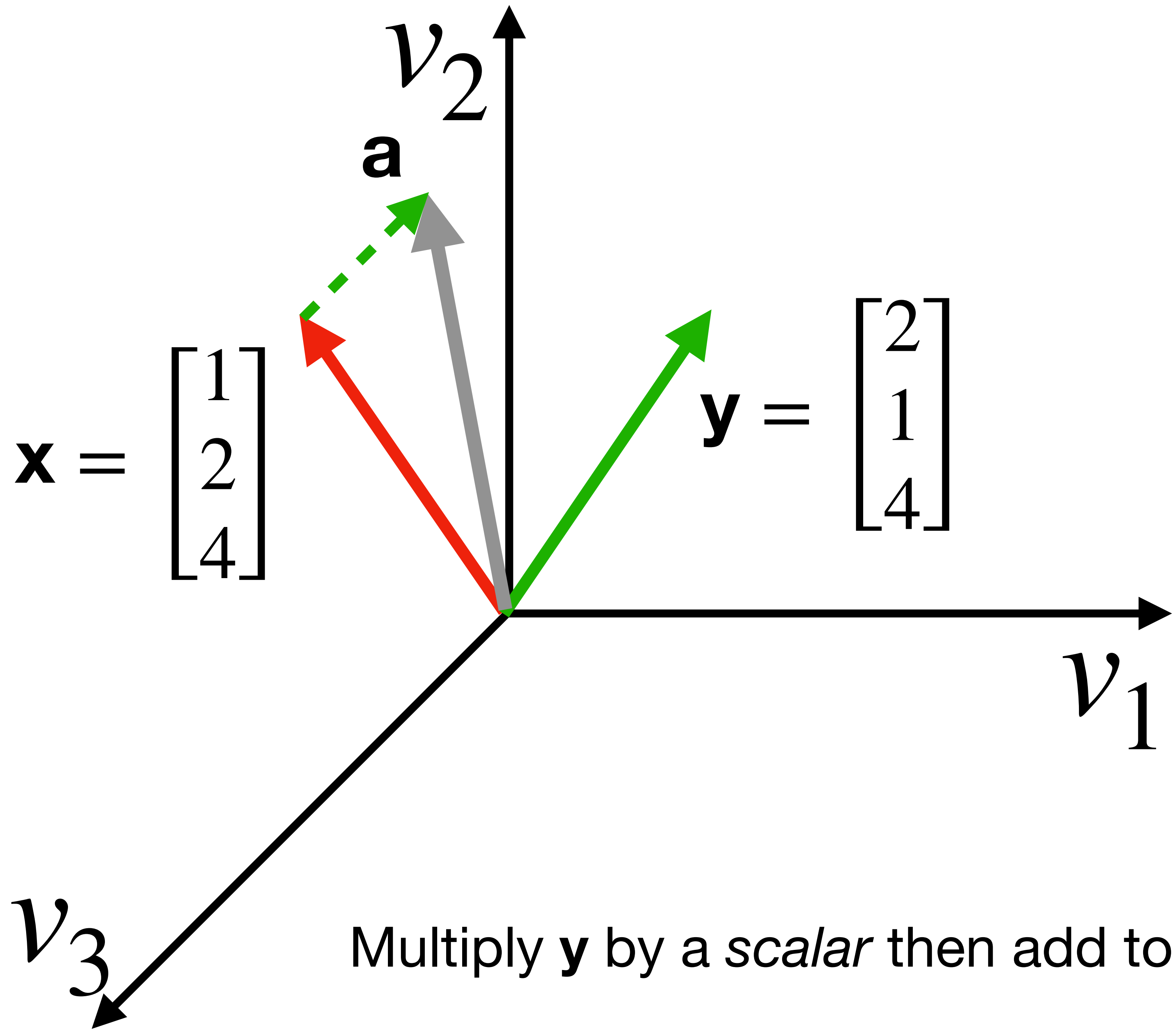
$$span(\{v_1, v_2, \ldots, v_n\}) = \underbrace{\alpha_1 v_1 + \ldots + \alpha_n v_n}_{\textbf{Linear Combination}},\ \alpha_i \in \mathbb{R}$$



v_2
v_1
v_3
$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$
$\mathbf{y} = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}$

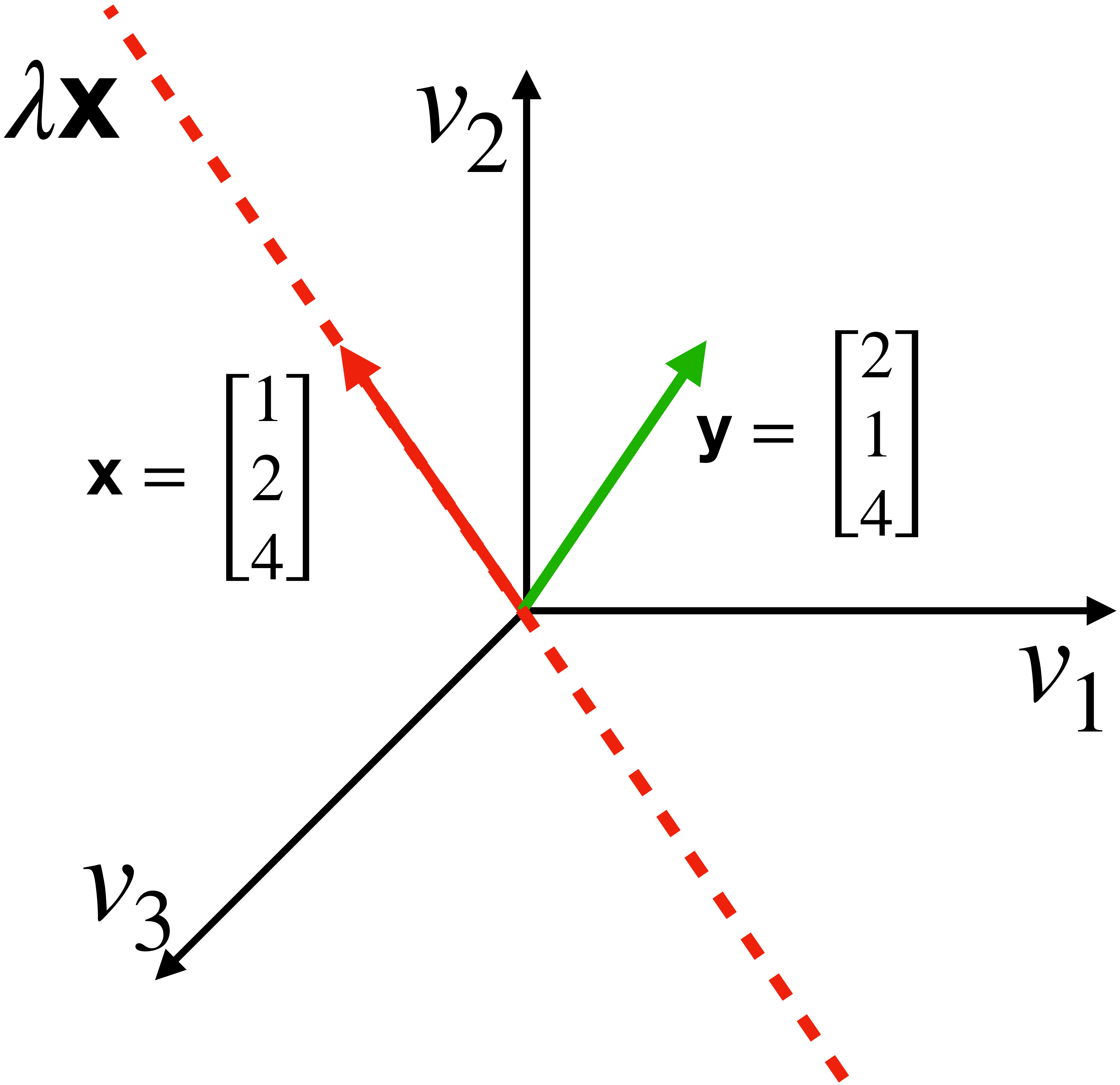


Can **x** and **y** be manipulated (with vector addition, scalar multiplication) to produce **a** ?

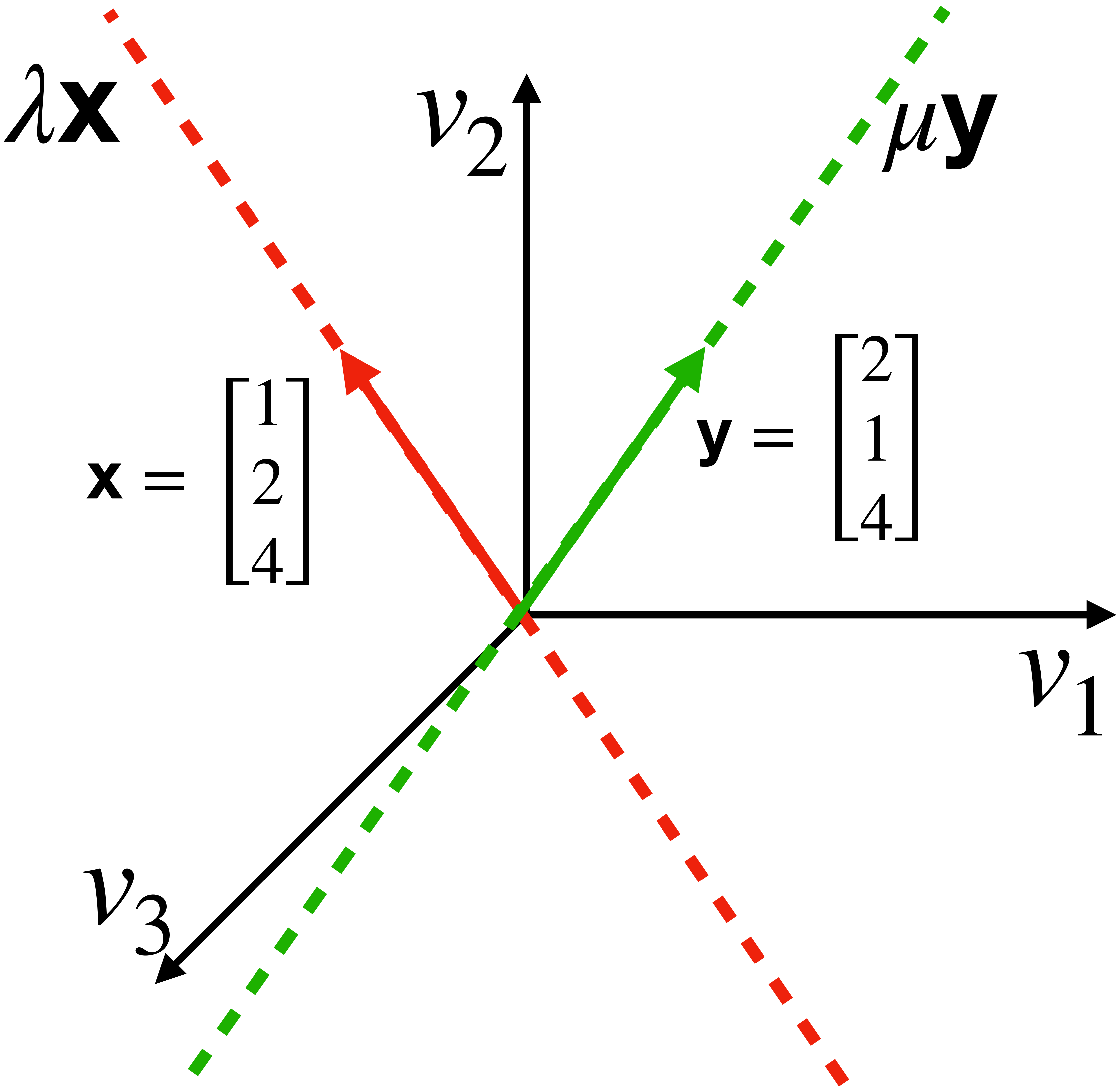


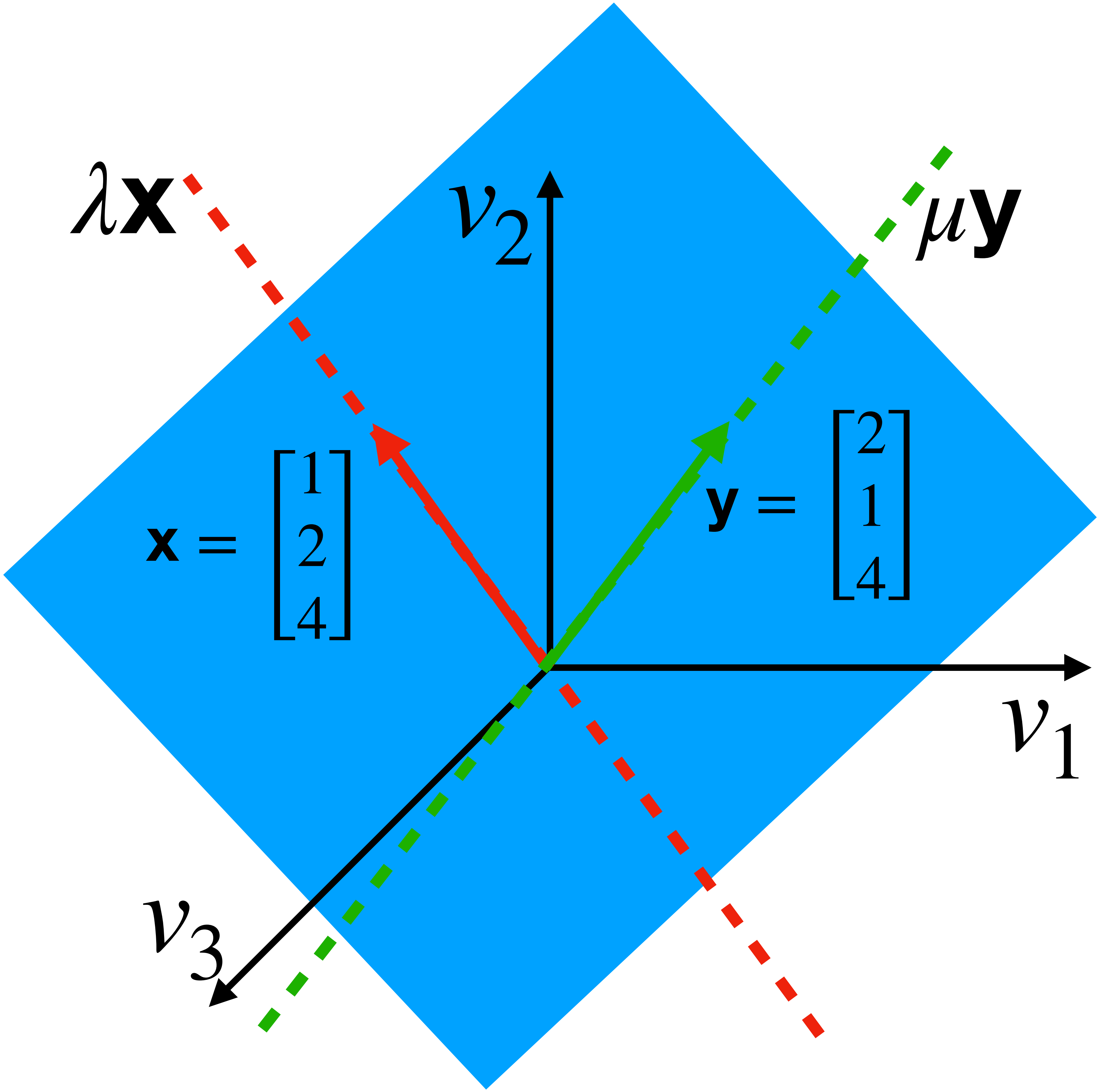
Multiply **y** by a *scalar* then add to **x** ?

Hence a is in the *span* of S = {**x, y**}



$\lambda \mathbf{x}$
v_2
v_1
v_3
$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$
$\mathbf{y} = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}$


$\lambda\mathbf{x}$
v_2
$\mu\mathbf{y}$
$\mathbf{x} = \begin{bmatrix}1\\2\\4\end{bmatrix}$
$\mathbf{y} = \begin{bmatrix}2\\1\\4\end{bmatrix}$
v_1
v_3



For two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ the linear combinations are $\lambda\boldsymbol{x} + \mu\boldsymbol{y}$.

All combinations $\lambda\boldsymbol{x} + \mu\boldsymbol{y}$ of two typical nonzero vectors fill a plane through (0,0,0).

Span

To determine if a vector **v** is in the span of a set S we need to check whether **v** can be expressed as a linear combination of vectors in S.

$\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$, $\mathbf{S} = \{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 7 \\ 0 \end{bmatrix} \}$. Check if $\mathbf{v} \in \mathbf{S}$

How can we check this?

Span

To determine if a vector $\mathbf{v}$ is in the span of a set S we need to check whether $\mathbf{v}$ can be expressed as a linear combination of vectors in S.

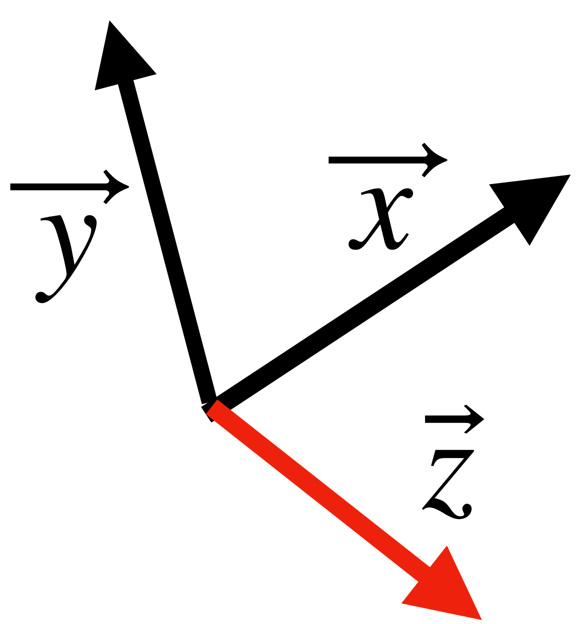
$\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$, $\mathbf{S} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 7 \\ 0 \end{bmatrix} \right\}$. Check if $\mathbf{v} \in \mathbf{S}$

$\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \frac{5}{6} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \frac{1}{6} \begin{bmatrix} 1 \\ 7 \\ 0 \end{bmatrix}$ Yes, $\mathbf{v} \in \mathbf{S}$

Linear Independence

Definition 2.12 (Linear (In)dependence). Let us consider a vector space V with $k \in \mathbb{N}$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in V$. If there is a non-trivial linear combination, such that $\mathbf{0} = \sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i$ with at least one $\lambda_i \neq 0$, the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ are *linearly dependent*. If only the trivial solution exists, i.e., $\lambda_1 = \ldots = \lambda_k = 0$ the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ are *linearly independent*.

Geometric Intuition: *A set of vectors is linearly independent if each vector point in a geometric dimension is not reachable using other vectors in the set.*



$$z \neq \alpha x + \beta y$$

z can not be express as a linear combination of *x* and *y*

Linear Independence (Proof)

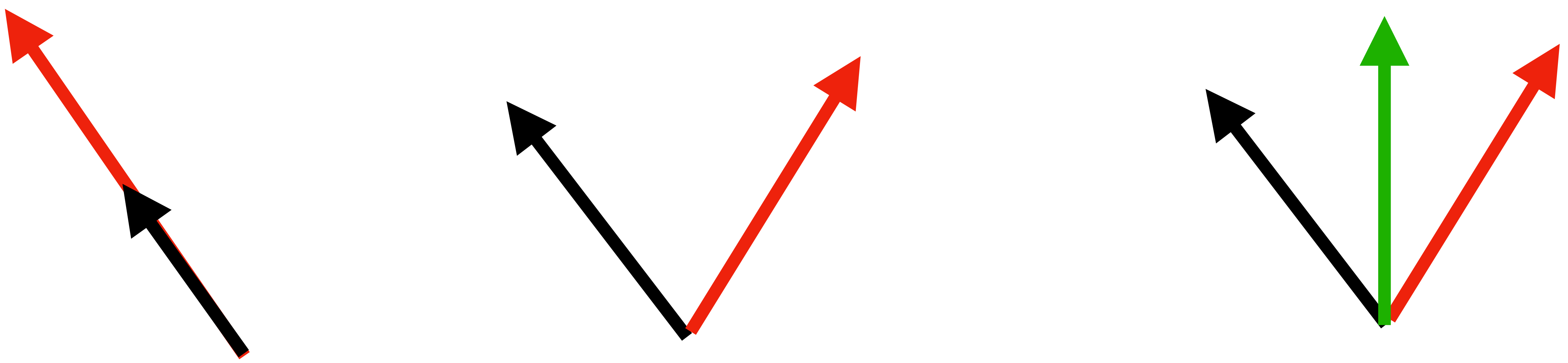
Given a subset of vectors $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ for $k \in \mathbb{N}$, of a vector space V, prove that S is linearly independent iff a linear combination of elements in *S* with non-zero coefficients does not yield **0**.

Linear Independence (Proof)

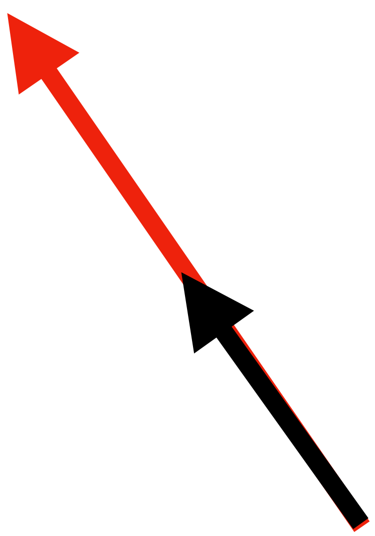
Given a subset of vectors $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ for $k \in \mathbb{N}$, of a vector space V, prove that S is linearly independent iff a linear combination of elements in *S* with non-zero coefficients does not yield **0**.

Hint: To prove *iff* statements i.e., A **iff** B ($A \iff B$), first prove $A \to B$, then $B \to A$.

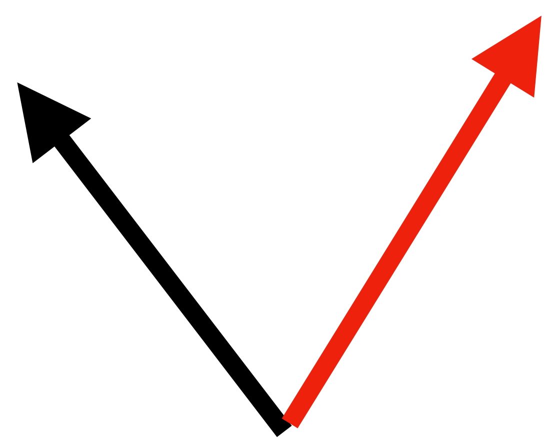
Are these sets of vectors linearly independent?



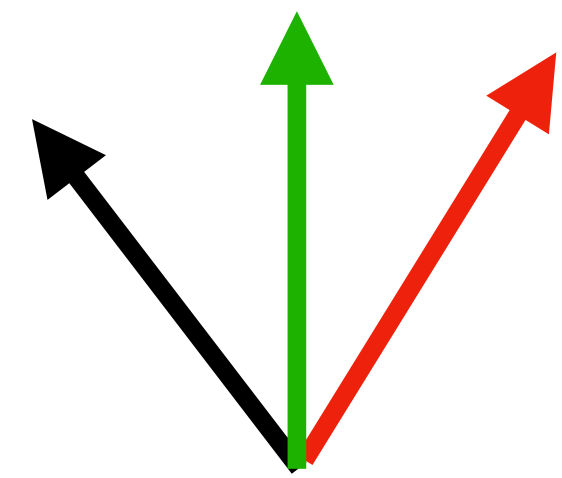
Are these sets of vectors linearly independent?



NO

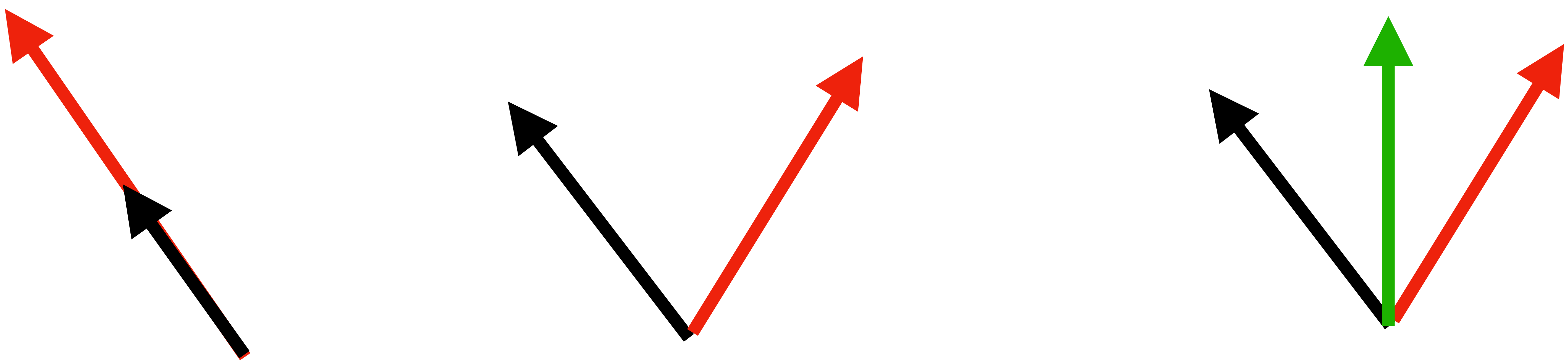


YES



NO

Are these sets of vectors linearly independent?



NO **YES** **NO**

Theorem

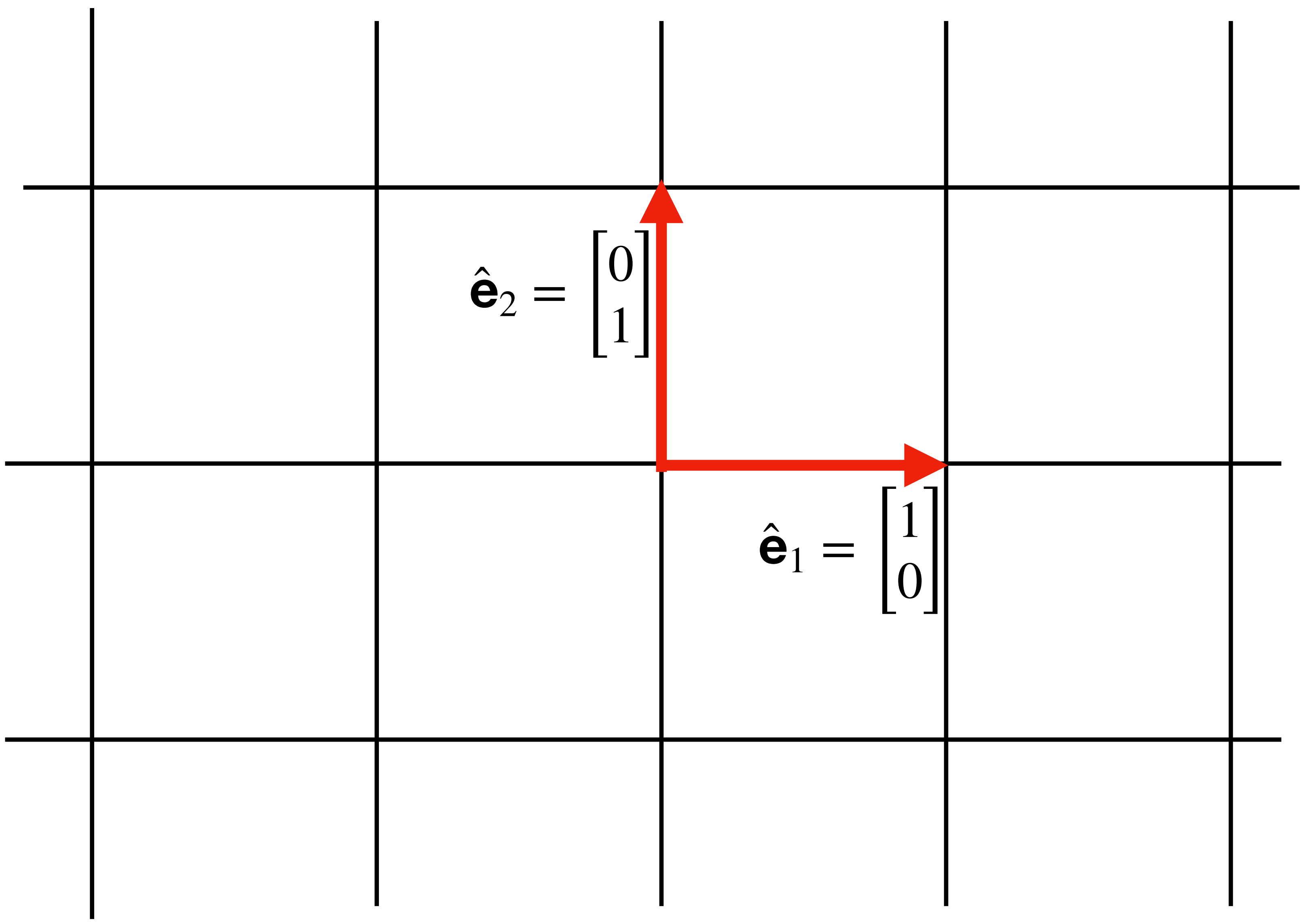
There are a maximum of N independent vectors in $\mathbb{R}^N$.

Basis - Definition

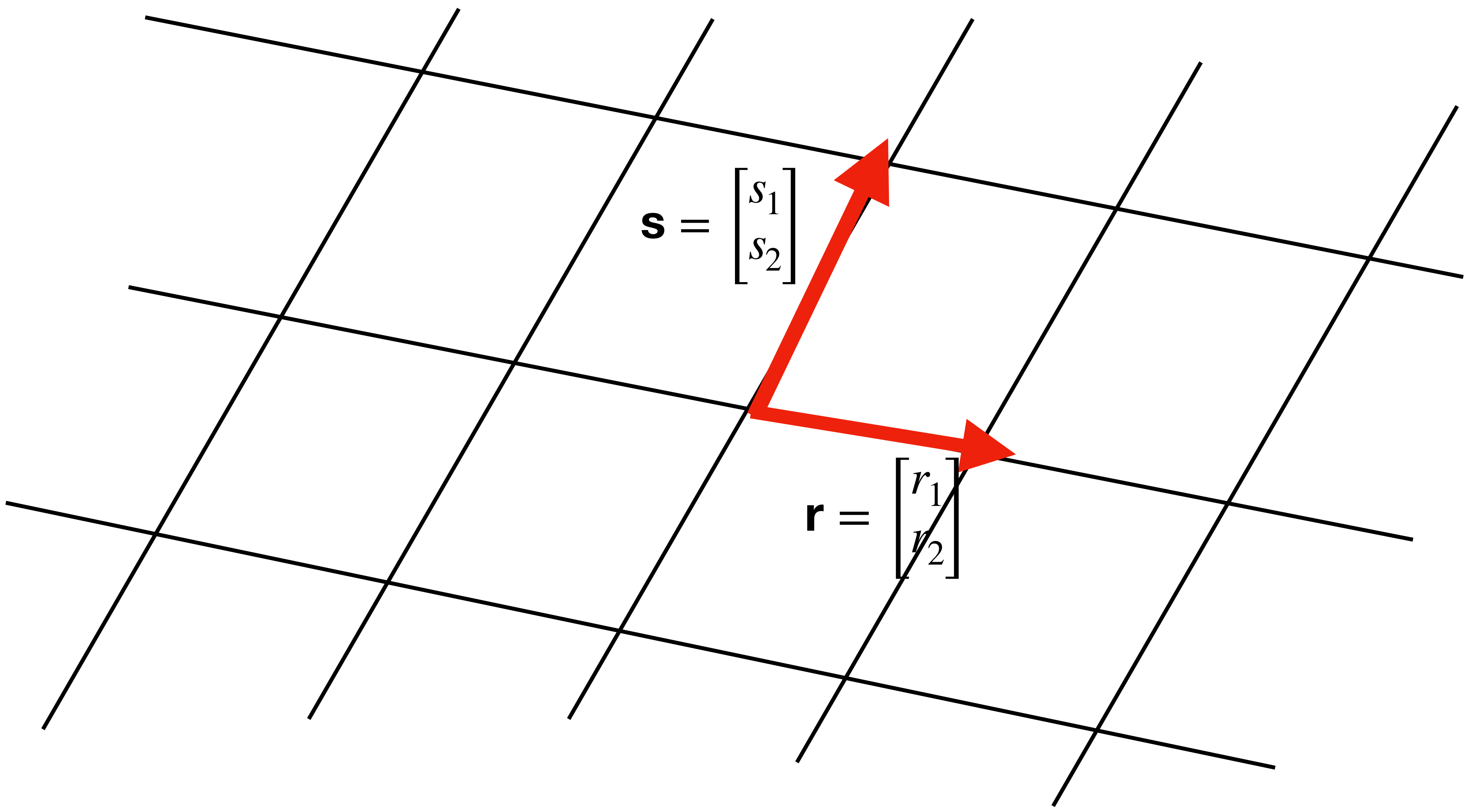
Basis is a set of n vectors that:

- are linearly independent
  - Are not linear combinations of each other
- span the space
- the space is then n-dimensional

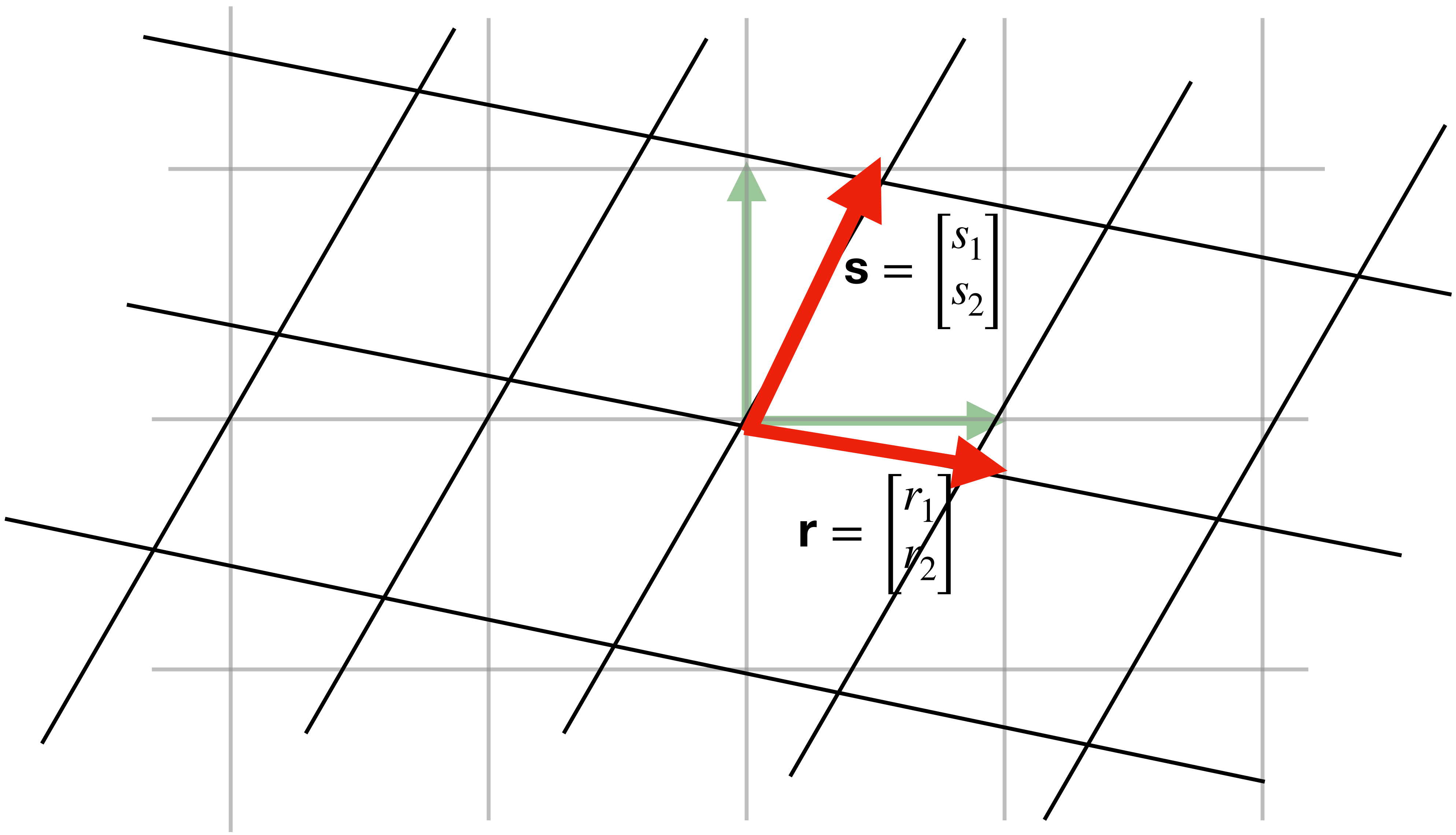
Natural Basis



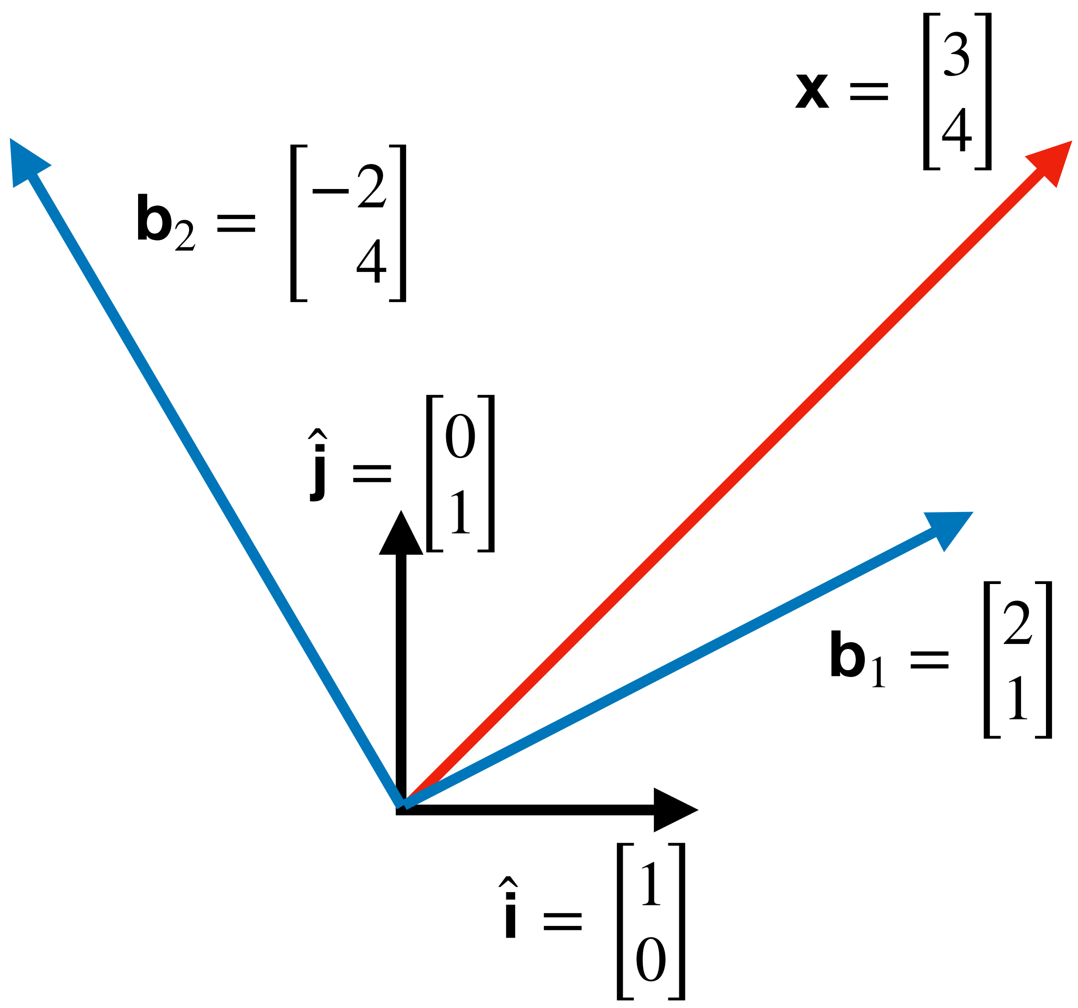
Another Basis

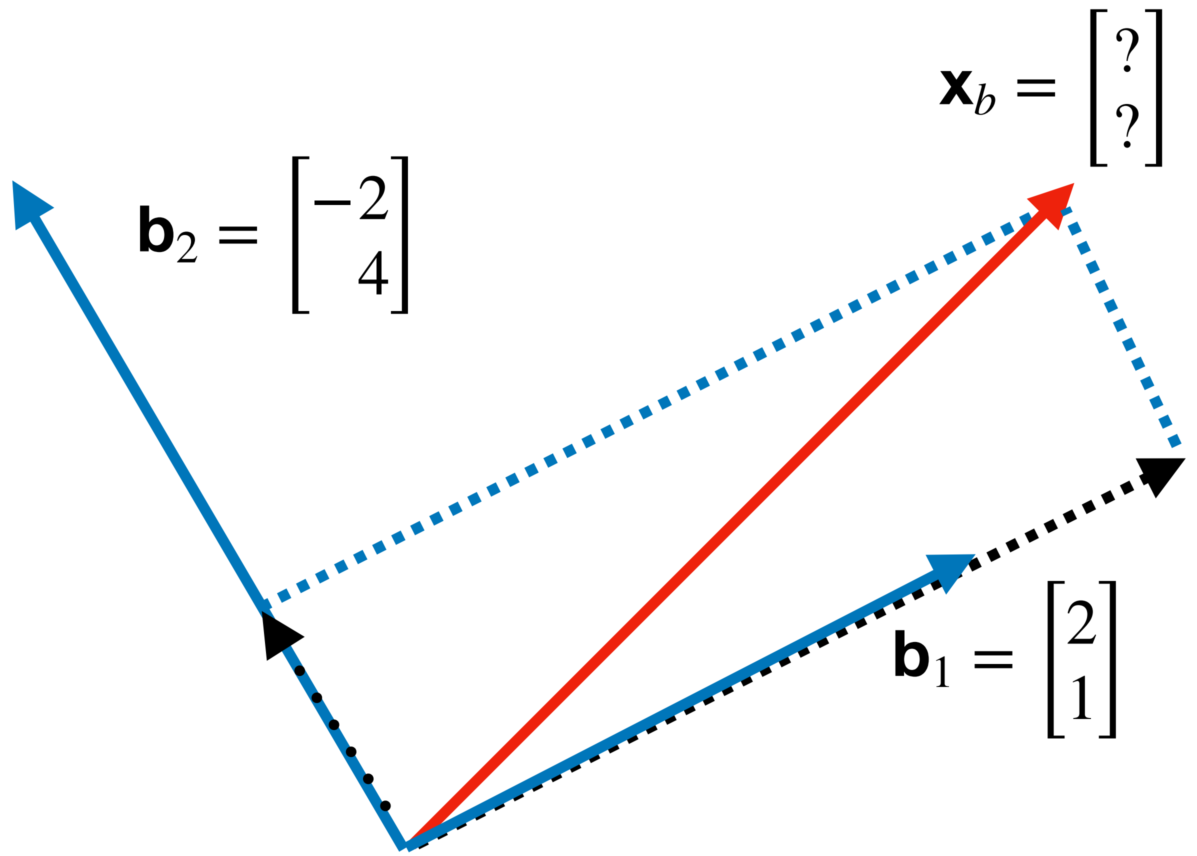


Change of Basis (Intuition)

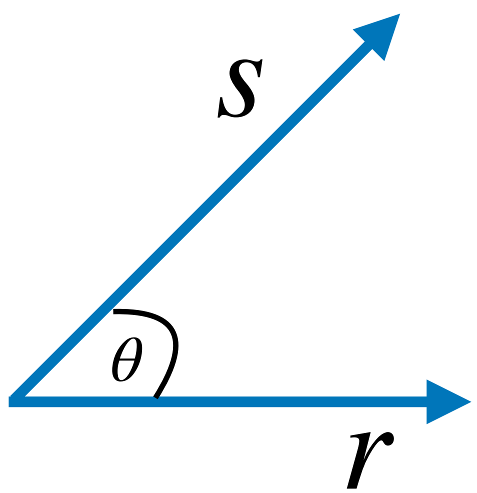


Changing basis



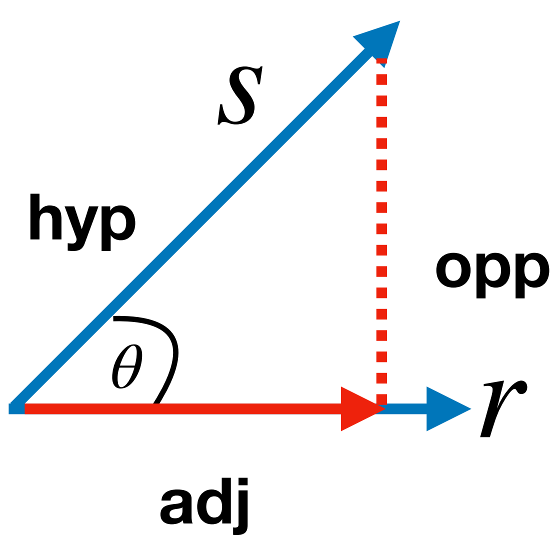


Vector Projections



$$r \cdot s = ||r||\ ||s||cos\theta$$ From Cosine Rule

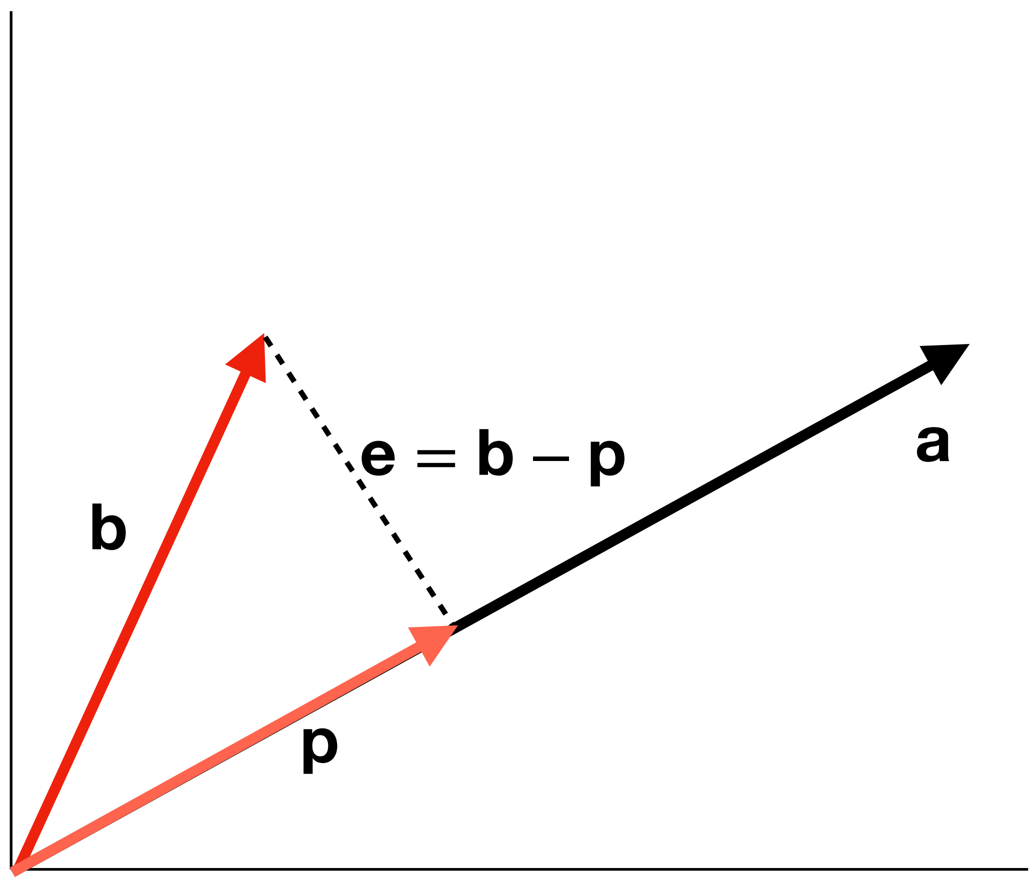
$$cos\theta = \frac{\text{adj}}{\text{hypotenuse}} = \frac{\text{adj}}{||s||}$$



$$\boxed{\text{adj} = \frac{r \cdot s}{||r||}}$$ Scalar Projection

$$\boxed{\overrightarrow{\text{adj}} = \frac{r \cdot s}{||r||\ ||r||}r}$$ Vector Projection

Projection onto a Line



$$\mathbf{p} = \hat{x}\mathbf{a}, \quad \mathbf{a} \perp (\mathbf{b} - \mathbf{p})$$

$$\mathbf{a} \cdot (\mathbf{b} - \hat{x}\mathbf{a}) = 0$$

$$\mathbf{a} \cdot \mathbf{b} - \hat{x}\mathbf{a} \cdot \mathbf{a} = 0$$

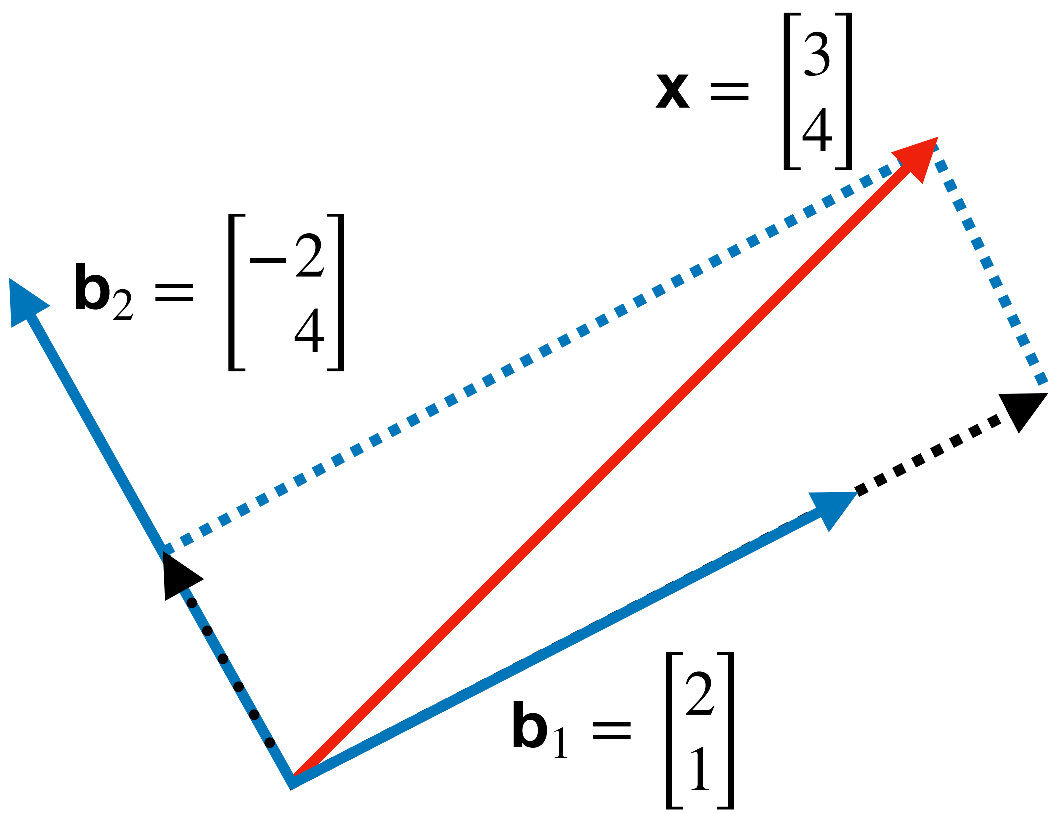
$$\mathbf{a}^T\mathbf{b} - \hat{x}\mathbf{a}^T\mathbf{a} = 0$$

$$\hat{x} = \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}$$

$$\mathbf{p} = \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}\mathbf{a}$$

$\hat{x}$

Perpendicular Basis Example



1. Check if the new basis perpendicular

$$cos\theta = \frac{b_1 \,.\, b_2}{||b_1||\,||b_2||} = \frac{2 * (-2) + 1 * 4}{\sqrt{2^2 + 1^2} * \sqrt{(-2)^2 + 4^2}} = 0$$

$$x_{b_1} = \frac{x \,.\, b_1}{||b_1||^2} = \frac{3 * 2 + 4 * 1}{\sqrt{2^2 + 1^2}^2} = \frac{10}{5} = 2$$

$$x_{b_2} = \frac{x \,.\, b_2}{||b_2||^2} = \frac{3 * (-2) + 4 * 4}{\sqrt{(-2)^2 + 4^2}^2} = \frac{10}{20} = \frac{1}{2}$$

$$\mathbf{x}_b = \begin{bmatrix} 2 \\ 1/2 \end{bmatrix}$$