Machine Learning

CMPT 726

Mo Chen
SFU School of Computing Science
2021-09-13

Why Machine Learning?

What is Artificial Intelligence (AI)?

- The study of how to engineer *intelligent* systems/machines.
- What is *intelligence*? Anything that humans can do that machines can’t do easily.
  - The ability to see and interpret visual input
  - The ability to read and understand language
  - The ability to move and interact with the world
  - The ability to reason and perform logical deduction

What is Artificial Intelligence (AI)?

- The study of how to engineer *intelligent* systems/machines.
- What is *intelligence*? Anything that humans can do that machines can't do easily.
  - The ability to see and interpret visual input — Computer Vision
  - The ability to read and understand language — Natural Language Processing (NLP)
  - The ability to move and interact with the world — Robotics
  - The ability to reason and perform logical deduction — Traditional AI

Successes of AI

Computer Vision

Natural Language Processing (NLP)

Google
Translate

Robotics

Traditional AI

Why is AI hard?

- Suppose we want to write a program that recognizes chairs.

- First attempt:
  - A chair is something that has a seat, a back and four legs.

Images taken from ShapeNet

Why is AI hard?

- Suppose we want to write a program that recognizes chairs.

- Second attempt:
    - A chair is something that has a seat, a back and multiple legs.

Images taken from ShapeNet

Why is AI hard?

- Suppose we want to write a program that recognizes chairs.

- Third attempt:
  - A chair is something that has a seat, a back and a frame.

Images taken from ShapeNet

Why is AI hard?

- Suppose we want to write a program that recognizes chairs.

- Fourth attempt:
  - A chair is something that has a seat and a back.

Images taken from ShapeNet

Why is AI hard?

- Suppose we want to write a program that recognizes chairs.

- Fifth attempt:
    - A chair is something that has a seat.

Images taken from ShapeNet

Why is AI hard?

- Suppose we want to write a program that recognizes chairs.

- Why is this not a chair?
- There are exceptions to every rule, and exceptions to every exception.

Images taken from ShapeNet

Why is AI hard?

- Problem: The inner workings of our brain are not well understood.
- We don’t know *how* our brain converts input to output, so we can’t write a program to do so.
- This problem perplexed early computer scientists:

“The Analytical Engine[1] has no pretensions to *originate* anything. It can do *whatever we know how to order it* to perform. ”

Ada Lovelace

[1]The Analytical Engine was the first conception of a general-purpose (i.e. Turing complete) computer.

Learning Machines

- Alan Turing proposed the concept of a *learning machine* in 1950 (in the same paper that proposed the Turing test).
- Idea: Divide the problem into two parts:
  - A machine that simulates a child's brain (analogous to a blank notebook: should function by simple mechanisms and have lots of blank sheets)
  - A way of teaching the child machine (should be simple since we know how to teach a human child)
- Teacher rewards good behaviour and penalizes bad behaviour.

Learning Machines

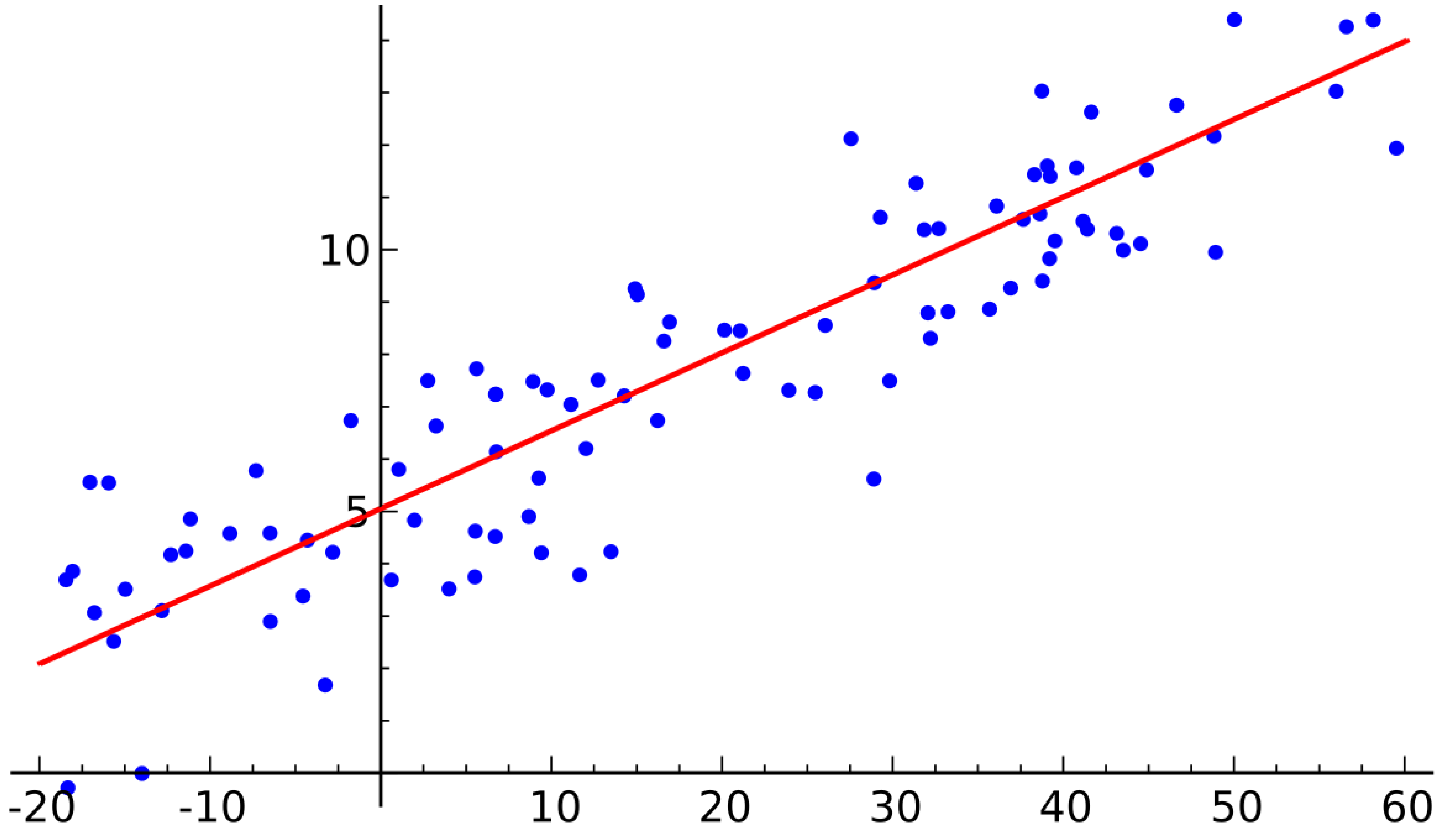
"An important feature of a learning machine is that its teacher will often be very largely ignorant of quite what is going on inside"

Alan Turing

- While we don't know *how* our brain converts input to output, we know what the output should be for every input.
- We can use this knowledge to teach the machine.

Machine Learning

- In modern terms:
  - Child machine: *Model*
  - Blank sheets: *Model parameters*
  - Teacher: *Loss function*



Input

Predicted Output

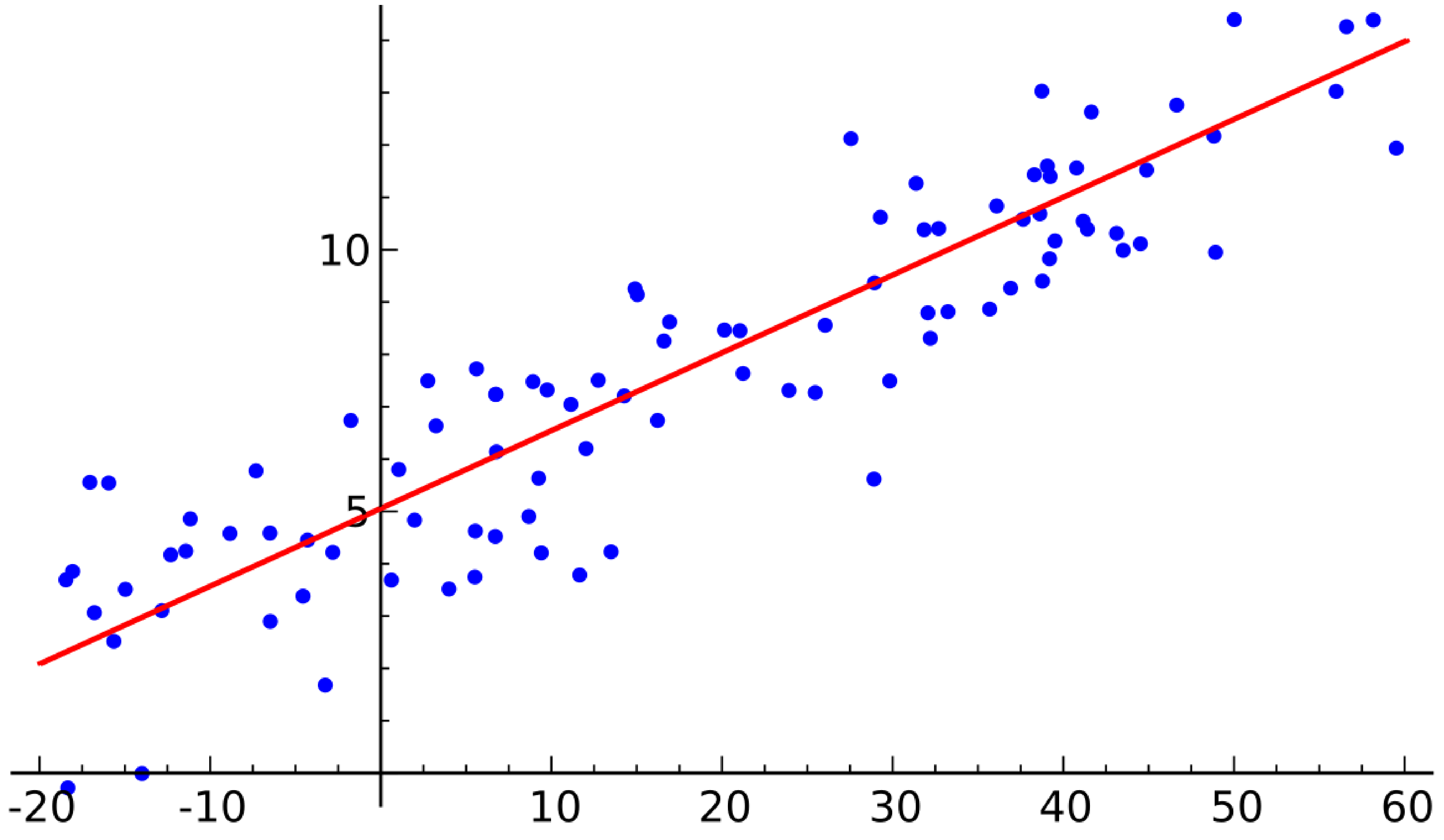
$$\hat{y} = wx + b$$

$$L = (y - \hat{y})^2$$

Desired Output

Machine Learning

- In modern terms:
  - Child machine: *Model*
  - Blank sheets: *Model parameters*
  - Teacher: *Loss function*



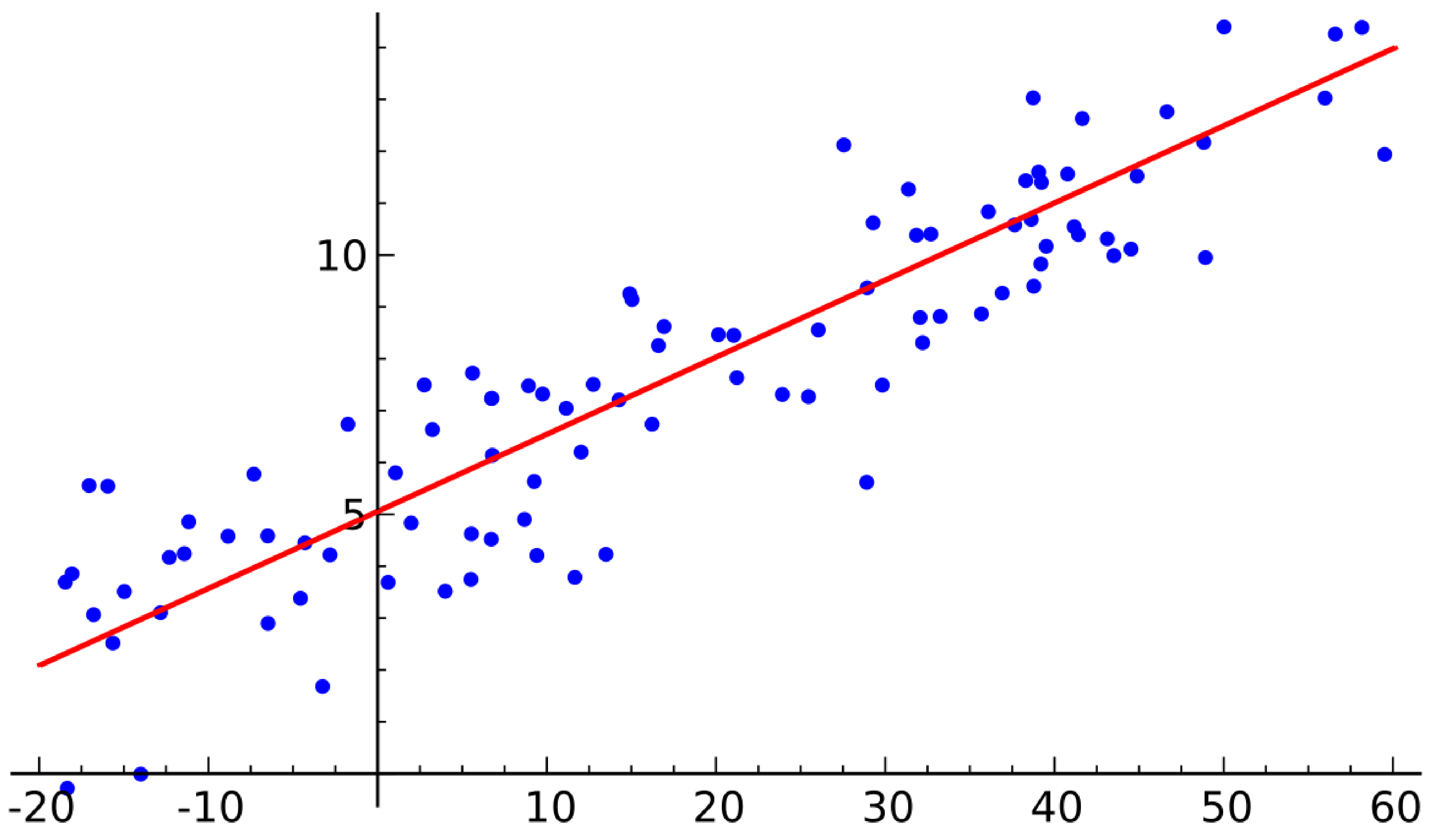
Parameters

Model $\boxed{\hat{y} = wx + b}$

Loss Function $L = (y - \hat{y})^2$

Machine Learning

- In modern terms:
  - Child machine: *Model*
  - Blank sheets: *Model parameters*
  - Teacher: *Loss function*



We want to find the parameter values that minimize the loss:

$$w^*, b^* = \arg\min_{w,b} L$$

Linear Algebra and Calculus Review

Vectors

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \in \mathbb{R}^4$$

Addition

$$\vec{x} + \vec{y} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}$$

Scaling

$$c\vec{x} = \begin{pmatrix} cx_1 \\ cx_2 \\ cx_3 \\ cx_4 \end{pmatrix}$$

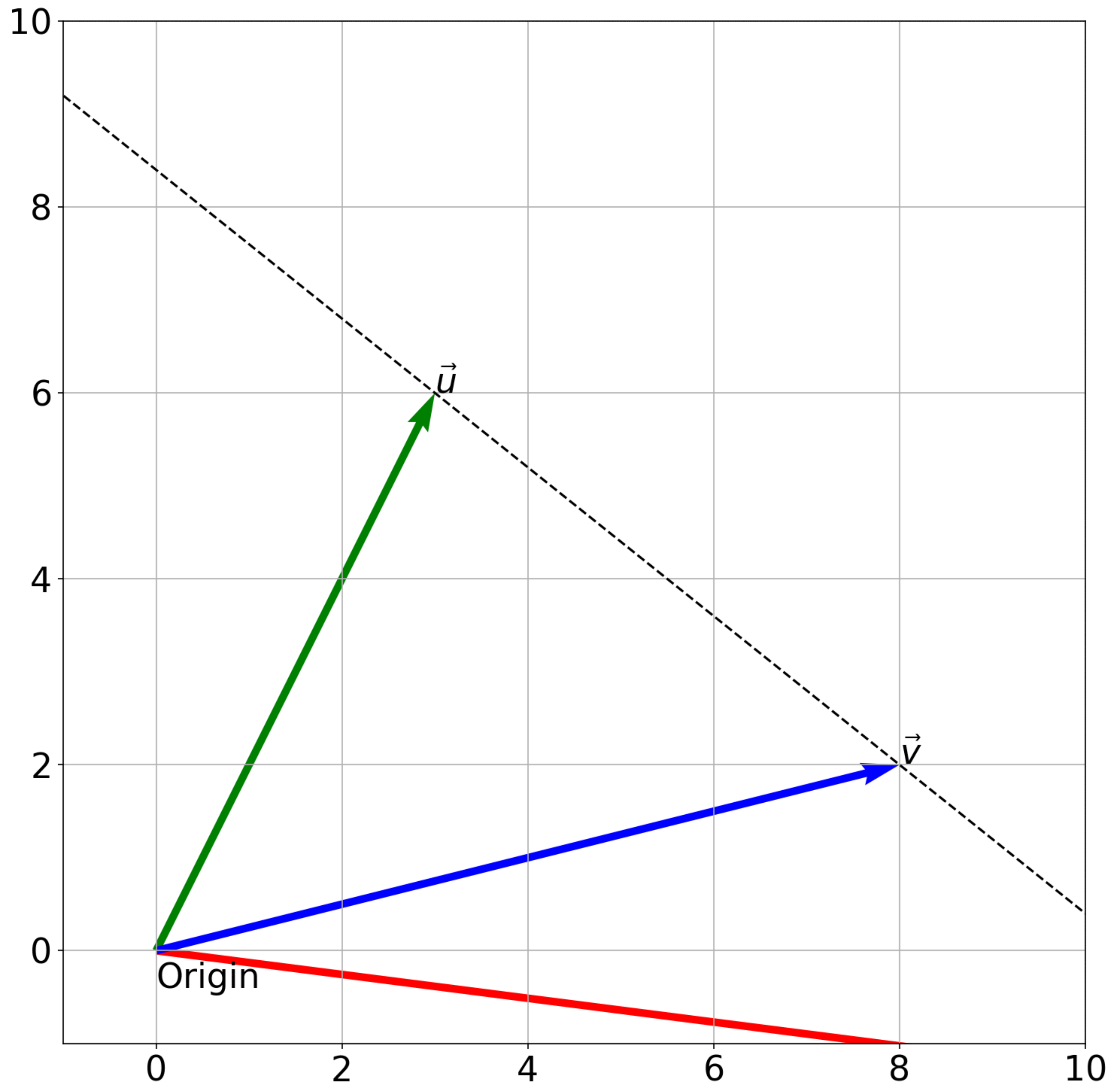
Linear Combinations

$$\vec{x} = \alpha_1 \vec{v}_1 + \cdots + \alpha_n \vec{v}_n = \sum_{i=1}^{N} \alpha_i \vec{v}_i$$

Special Case:

$$\sum_{i=1}^{N} \alpha_i = 1$$

Called "Affine Combination"



$\alpha = -0.9$

$\beta = 1.9$

subject to $\alpha + \beta = 1$

$\vec{x} = \alpha\vec{u} + \beta\vec{v}$

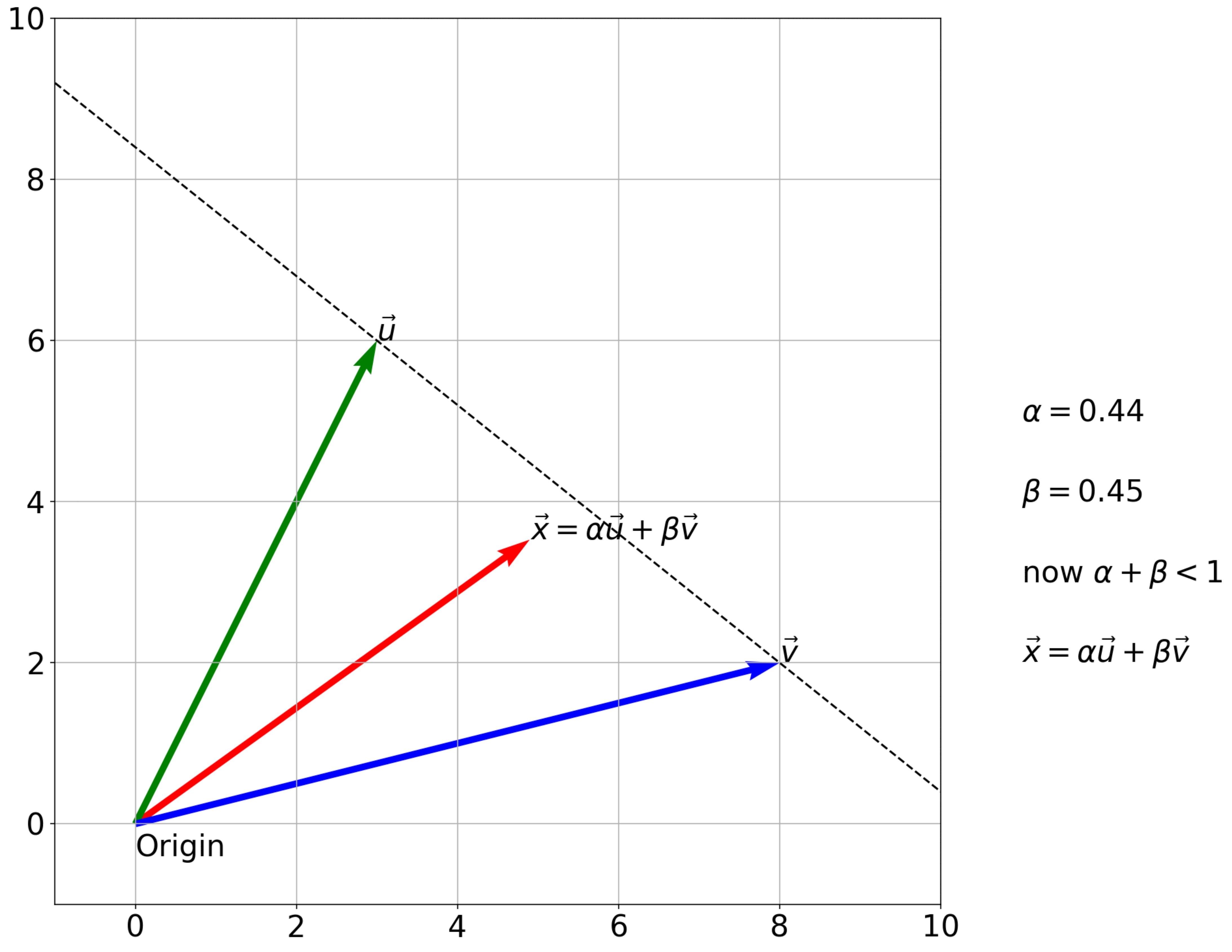
$\vec{x} = \alpha\vec{u} + \beta\vec{v}$

Linear Combinations

$$\vec{x} = \alpha_1 \vec{v}_1 + \cdots + \alpha_n \vec{v}_n = \sum_{i=1}^{N} \alpha_i \vec{v}_i$$

Special Case:

$$\sum_{i=1}^{N} \alpha_i < 1$$

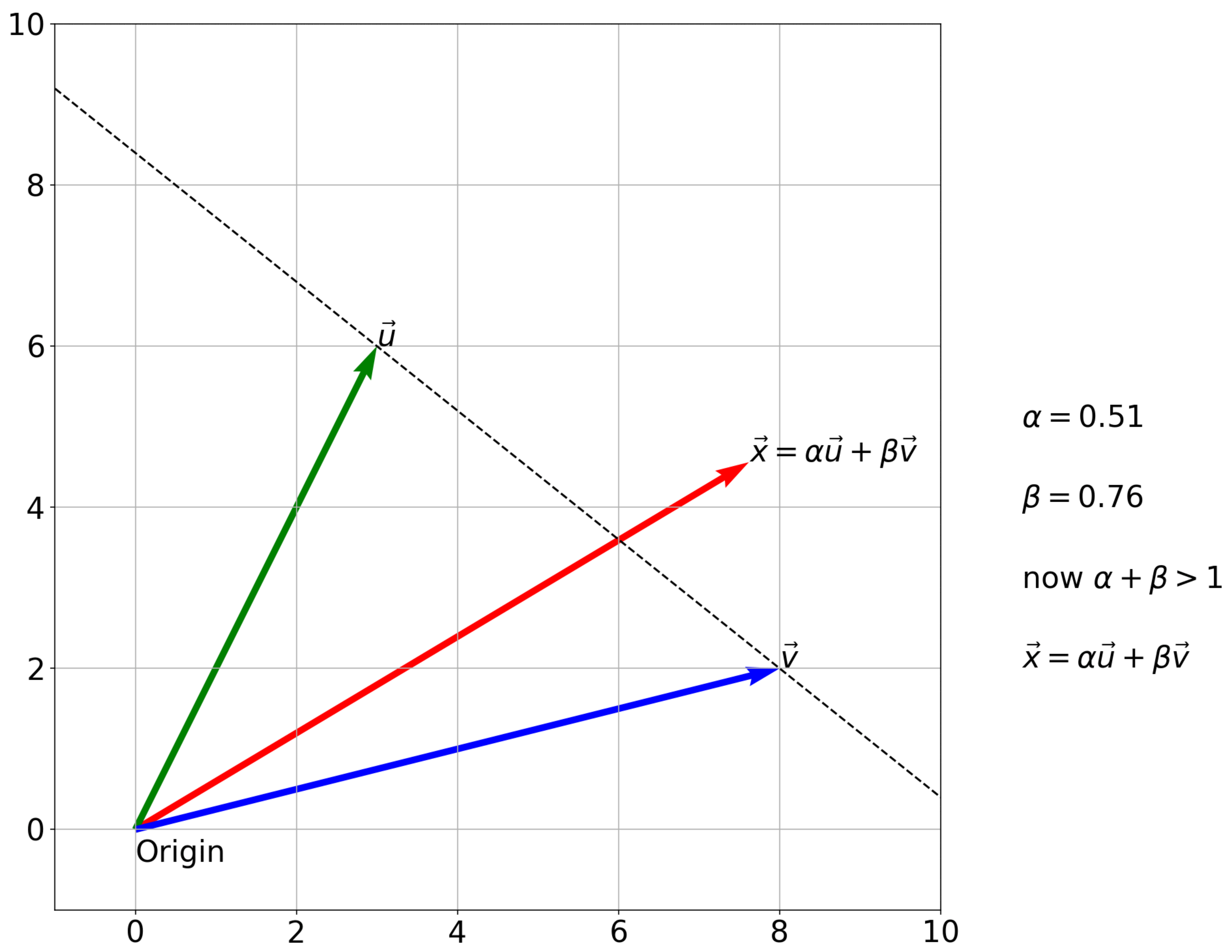


Linear Combinations

$$\vec{x} = \alpha_1 \vec{v}_1 + \cdots + \alpha_n \vec{v}_n = \sum_{i=1}^{N} \alpha_i \vec{v}_i$$

Special Case:

$$\sum_{i=1}^{N} \alpha_i > 1$$



Linear Combinations

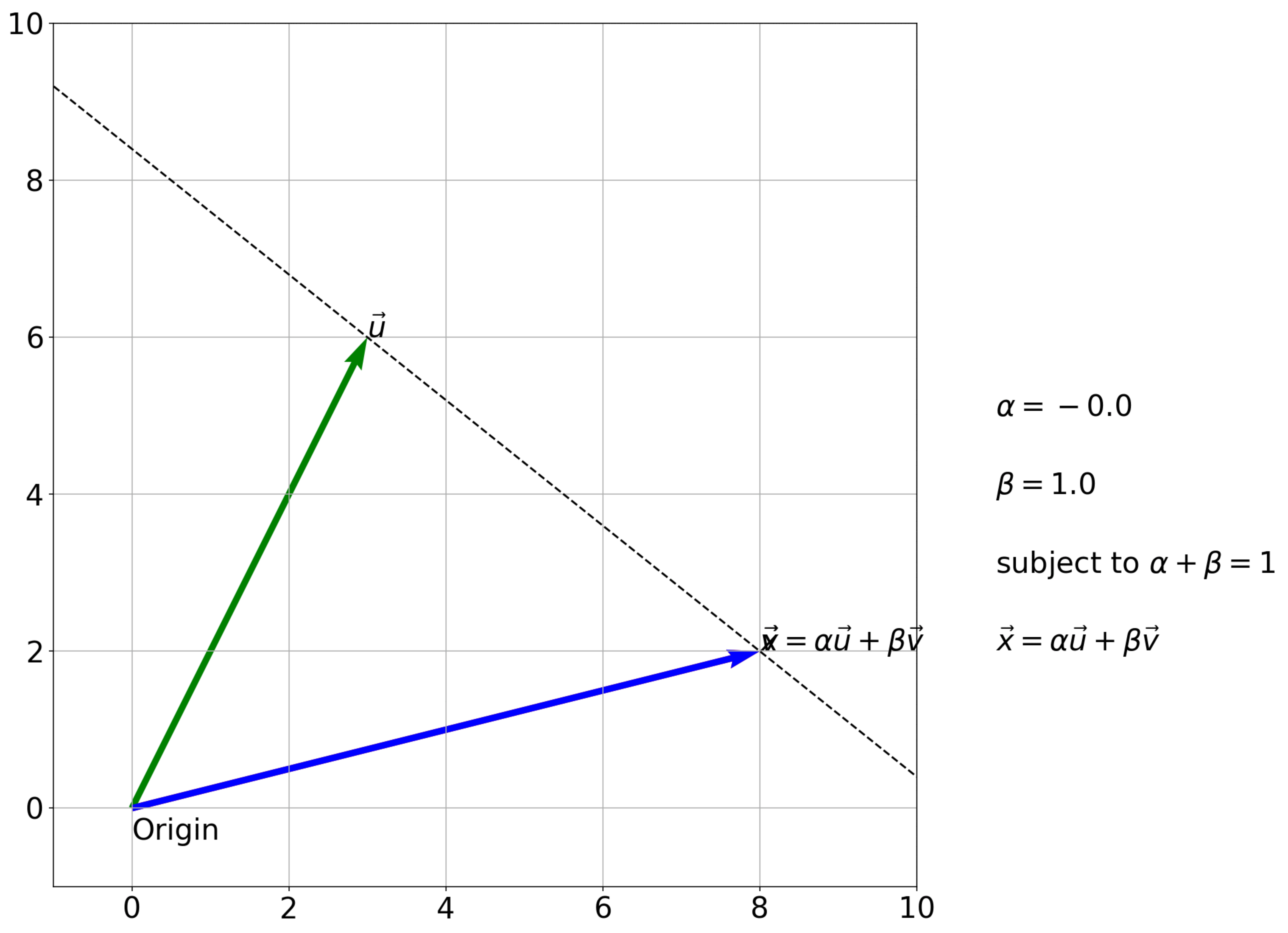
$$\vec{x} = \alpha_1 \vec{v}_1 + \cdots + \alpha_n \vec{v}_n = \sum_{i=1}^{N} \alpha_i \vec{v}_i$$

Special Case:

$$\sum_{i=1}^{N} \alpha_i = 1$$

$$\alpha_i \geq 0 \ \forall i$$

Called “Convex Combination”



Linear Combinations

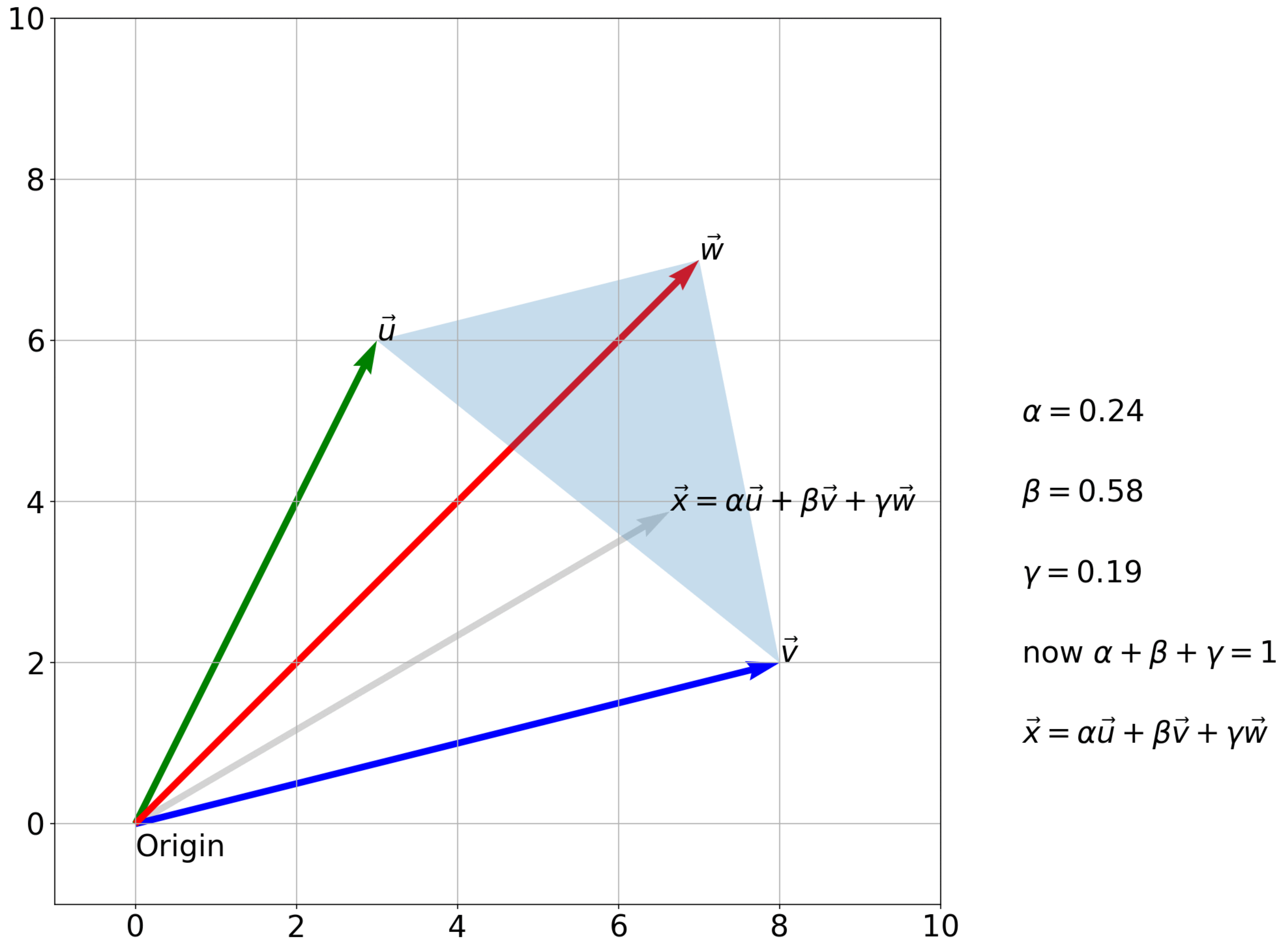
$$\vec{x} = \alpha_1 \vec{v}_1 + \cdots + \alpha_n \vec{v}_n = \sum_{i=1}^{N} \alpha_i \vec{v}_i$$

Special Case:

$$\sum_{i=1}^{N} \alpha_i = 1$$

$$\alpha_i \geq 0 \ \forall i$$

Called “Convex Combination”



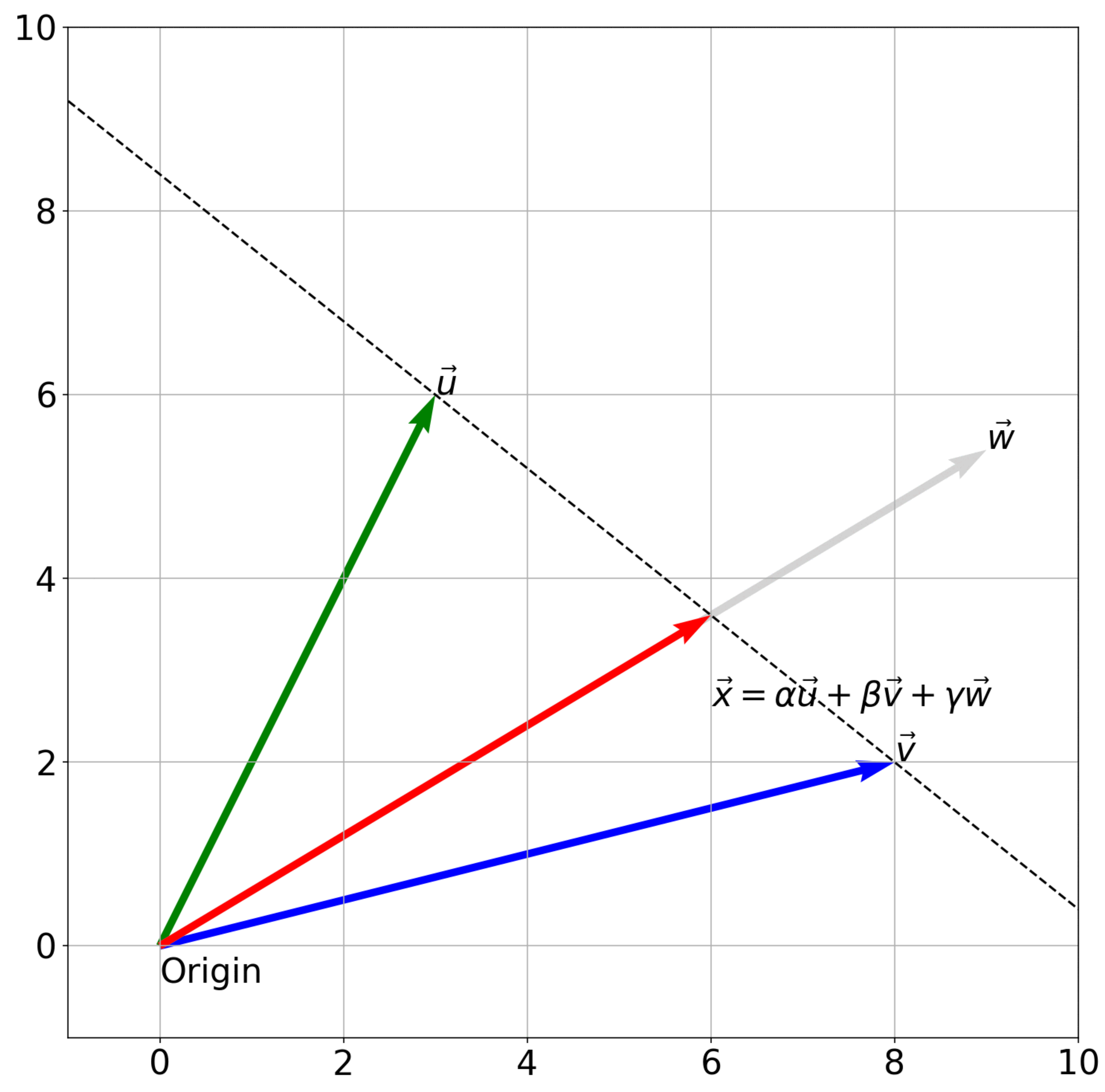
Span and Linear Independence

$$\text{span}\{\vec{v}_1, \dots, \vec{v}_n\} = \left\{ \sum_{i=1}^{N} \alpha_i \vec{v}_i \,|\alpha_1, \dots, \alpha_n \in \mathbb{R} \right\}$$

A set of vectors is *linearly independent* if no vector is in the span of the other vectors.

$$\sum_{i=1}^{N} \alpha_i \vec{v}_i = \vec{0} \Longrightarrow \alpha_1, \dots, \alpha_n = 0$$

Otherwise, they are *linearly dependent*.



$\vec{w}$ can be represnted

by combination of $\vec{u}$ and $\vec{v}$

starting from $\alpha + \beta = 1$

and adjust proportionally

$\alpha = 0.4$

$\beta = 0.6$

$\vec{x} = \alpha\vec{u} + \beta\vec{v}$

Inner Products*

$$\langle \vec{x}, \vec{y} \rangle = \sum_{i=1}^{N} x_i y_i$$

For the special case of real vectors:

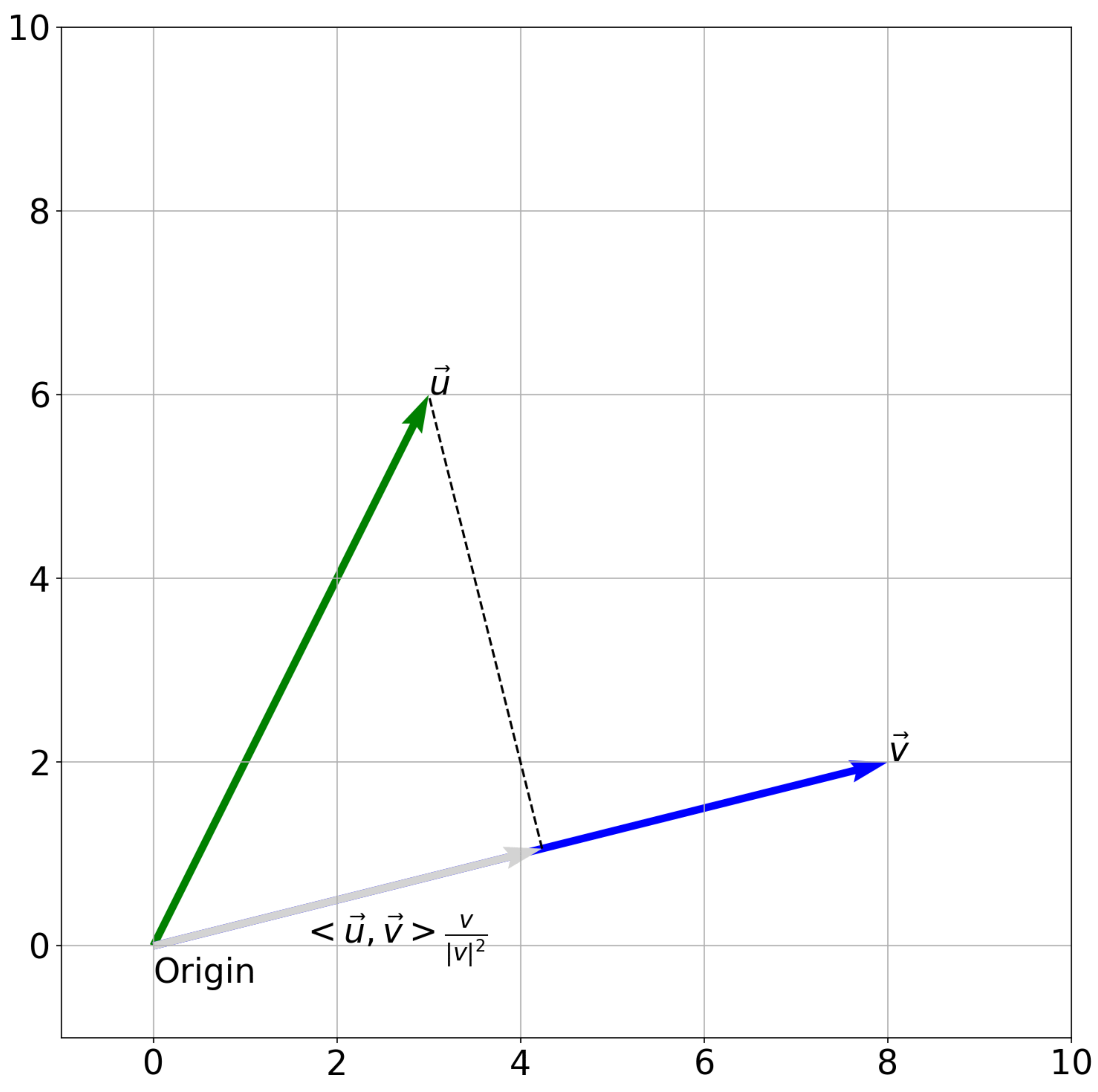
Symmetry: $\langle \vec{x}, \vec{y} \rangle = \langle \vec{y}, \vec{x} \rangle$

Bilinearity:

$$\langle \alpha\vec{x} + \beta\vec{y}, \vec{z} \rangle = \alpha\langle \vec{x}, \vec{z} \rangle + \beta\langle \vec{y}, \vec{z} \rangle$$

$$\langle \vec{x}, \alpha\vec{y} + \beta\vec{z} \rangle = \alpha\langle \vec{x}, \vec{y} \rangle + \beta\langle \vec{x}, \vec{z} \rangle$$

Nonnegativity: $\langle \vec{x}, \vec{x} \rangle \geq 0 \;\; \forall \vec{x}$



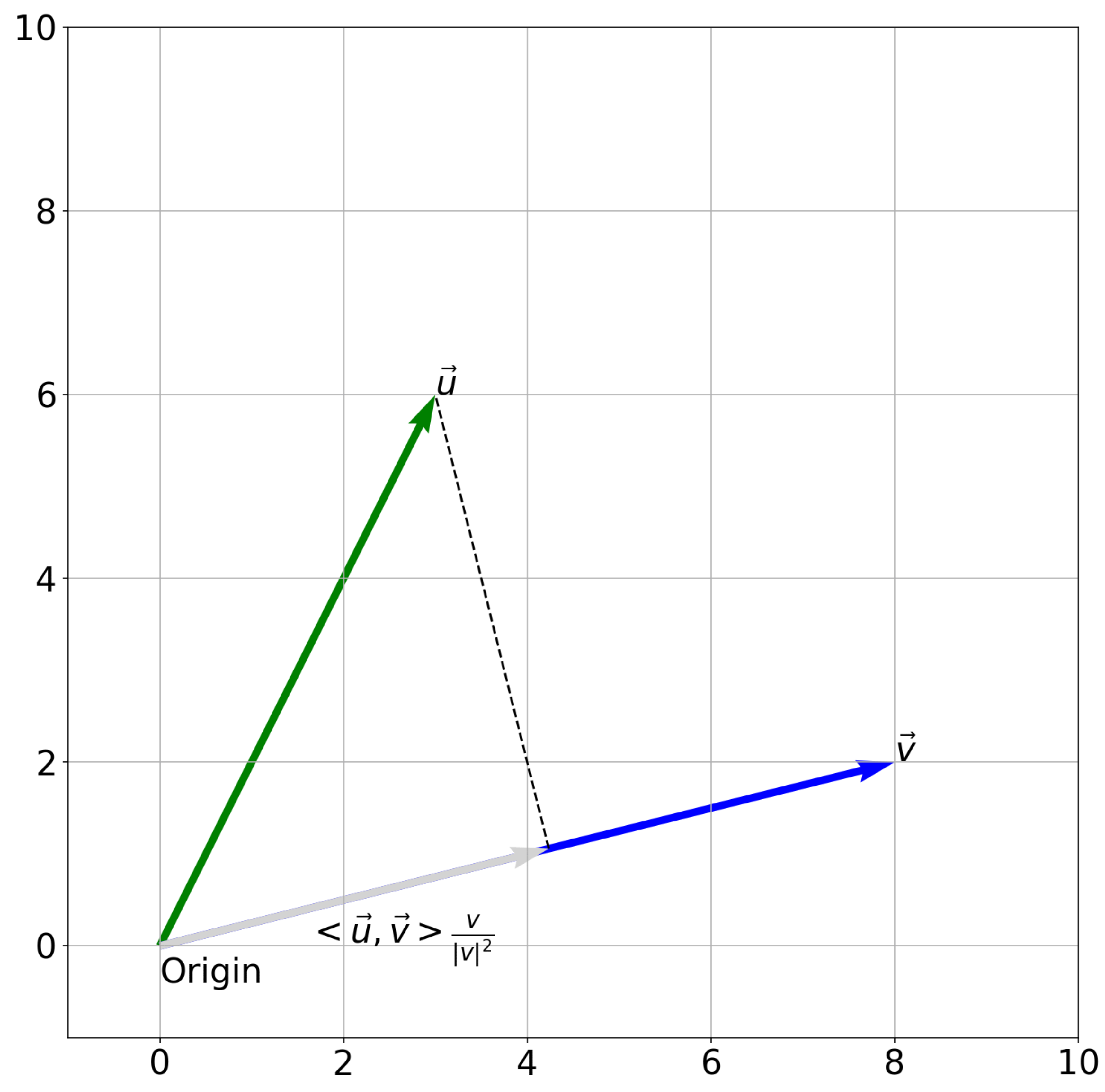
*Strictly speaking, this is the special case of standard inner products

Norms

Euclidean norm:

$$\|\vec{x}\|_2 = \sqrt{\langle \vec{x}, \vec{x} \rangle} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

$$\|\vec{x}\|_2 \geq 0 \ \forall \vec{x}$$



Unit vector: vector with norm 1, often referred to as "unit norm"

Normalizing a vector $\vec{x}$: $\dfrac{\vec{x}}{\|\vec{x}\|_2}$

Projecting a vector $\vec{x}$ onto $\vec{y}$:

$$\left\langle \vec{x}, \frac{\vec{y}}{\|\vec{y}\|_2} \right\rangle \frac{\vec{y}}{\|\vec{y}\|_2}$$

Sign of $\langle \vec{x}, \vec{y} \rangle$ is positive if angle is less than 90 degrees, and negative otherwise.

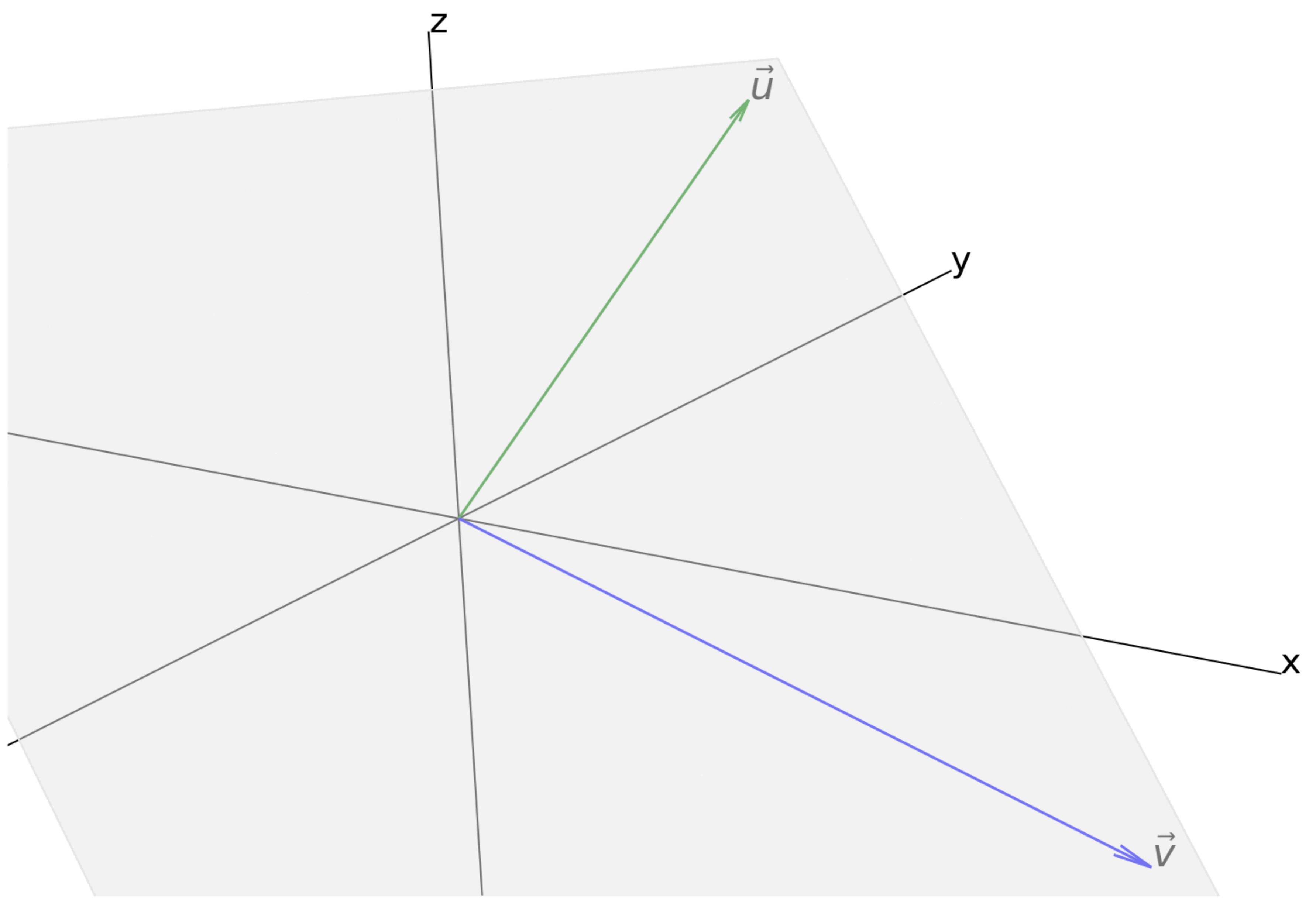
Linear Subspace

A lower-dimensional slice of the vector space that passes through the origin.

Examples:

A plane through the origin in 3D Euclidean space

A line through the origin in 3D Euclidean space

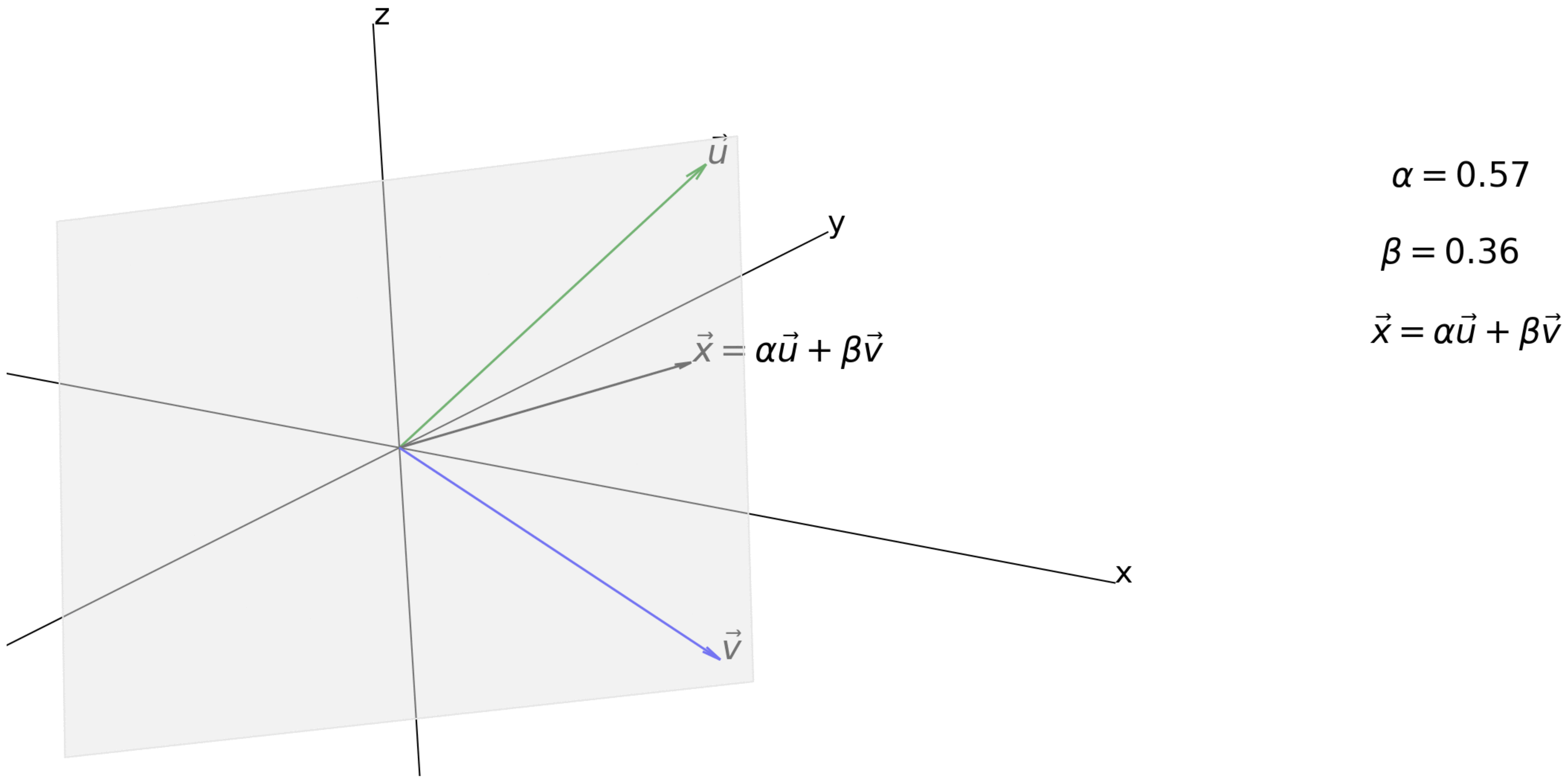


Basis

A linearly independent set of vectors that spans the full (sub)space.

Any vector in the (sub)space can be written as a linear combination of the basis vectors.

$\vec{u}, \vec{v}$ are a basis for the subspace

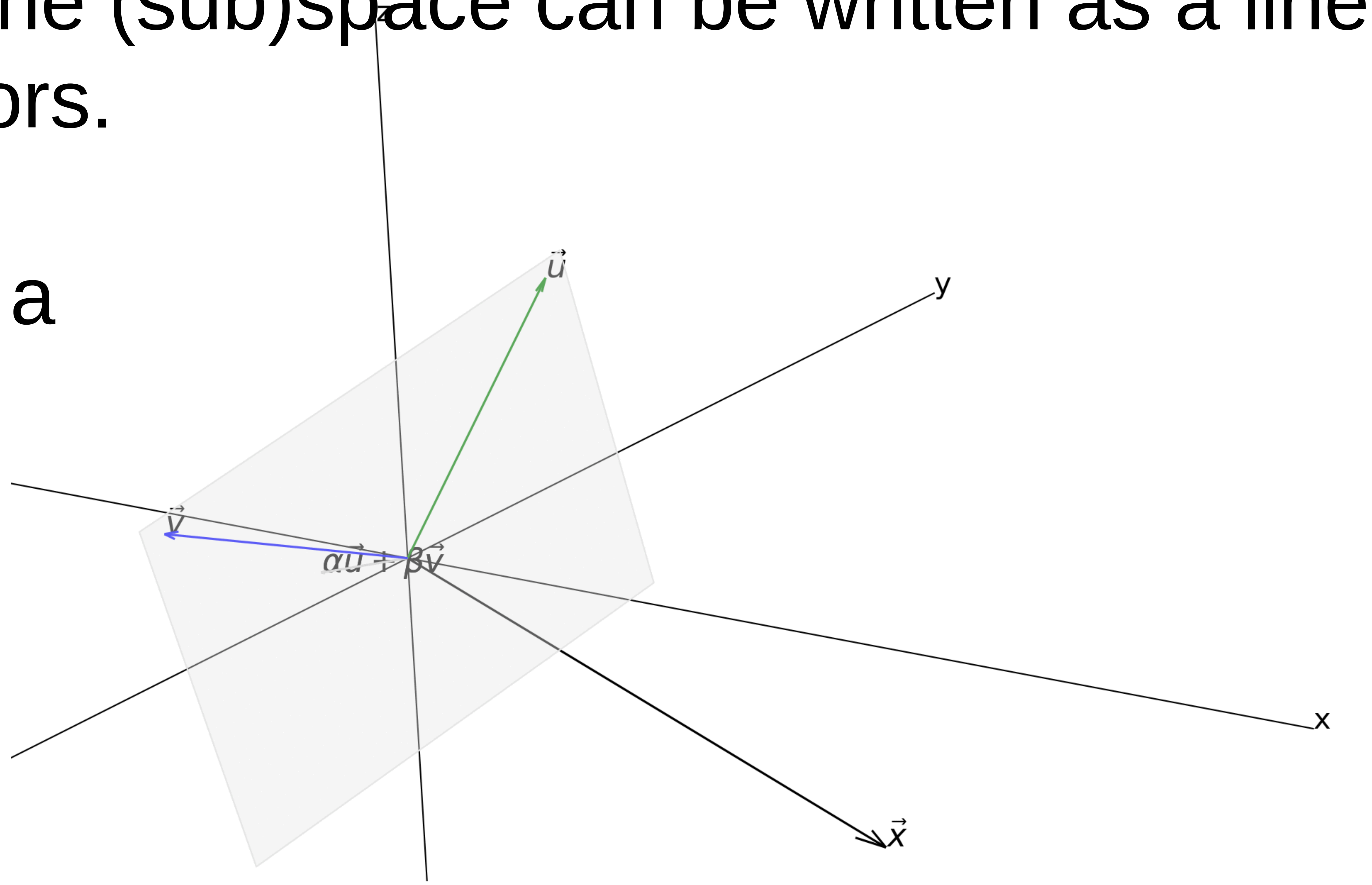


Basis

A linearly independent set of vectors that spans the full (sub)space.

Any vector in the (sub)space can be written as a linear combination of the basis vectors.

$\vec{u}, \vec{v}$ are not a basis for $\mathbb{R}^3$



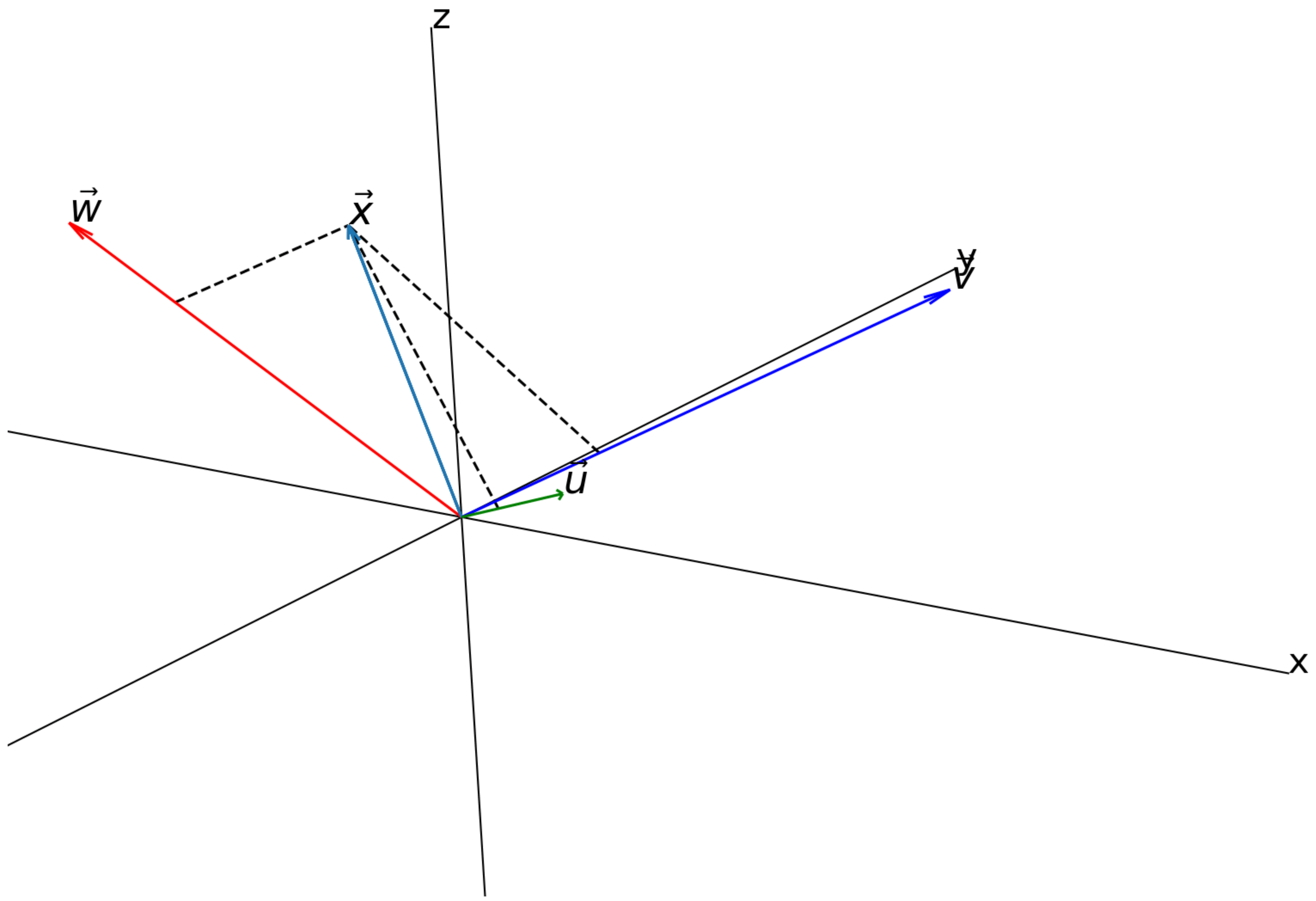
$\nexists \alpha, \beta, s.t.$

$\vec{x} = \alpha\vec{u} + \beta\vec{v}$

Orthogonal Basis

Two vectors $\vec{x}, \vec{y}$ are orthogonal if and only if $\langle \vec{x}, \vec{y} \rangle = 0$.

An orthogonal basis is a basis whose vectors are orthogonal to one another.

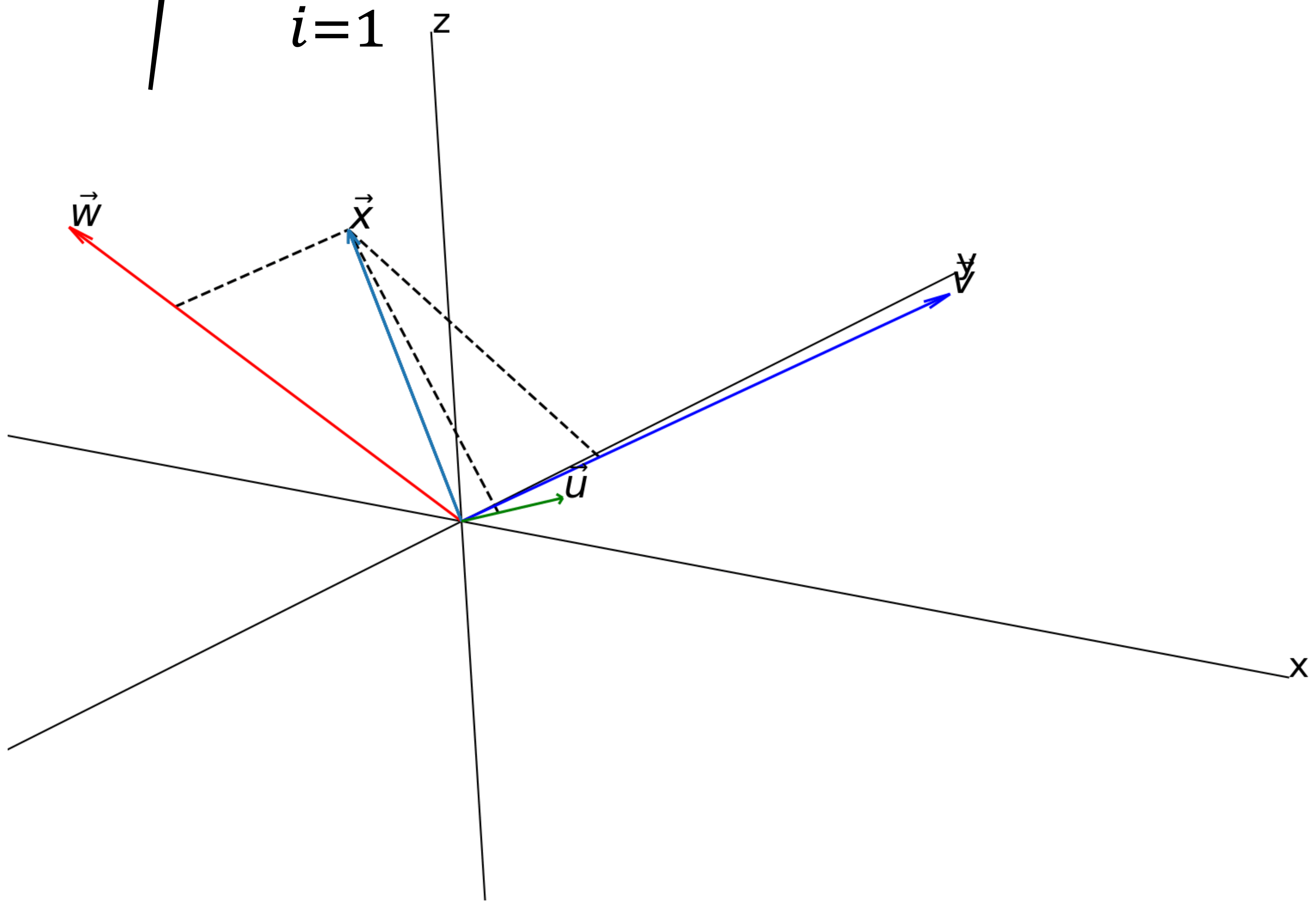


$\vec{x}$ is an arbitray vector

Orthogonal Basis

We like an orthogonal basis because it is easy to compute the coordinates with respect to the basis.

$$\langle \vec{x}, \vec{v}_j \rangle = \left\langle \sum_{i=1}^{N} \alpha_i \vec{v}_i \, , \vec{v}_j \right\rangle = \sum_{i=1}^{n} \alpha_i \langle \vec{v}_i, \vec{v}_j \rangle = \alpha_j \langle \vec{v}_j, \vec{v}_j \rangle = \alpha_j \left\| \vec{v}_j \right\|_2^2 \Longrightarrow \alpha_j = \frac{\langle \vec{x}, \vec{v}_j \rangle}{\left\| \vec{v}_j \right\|_2^2}$$

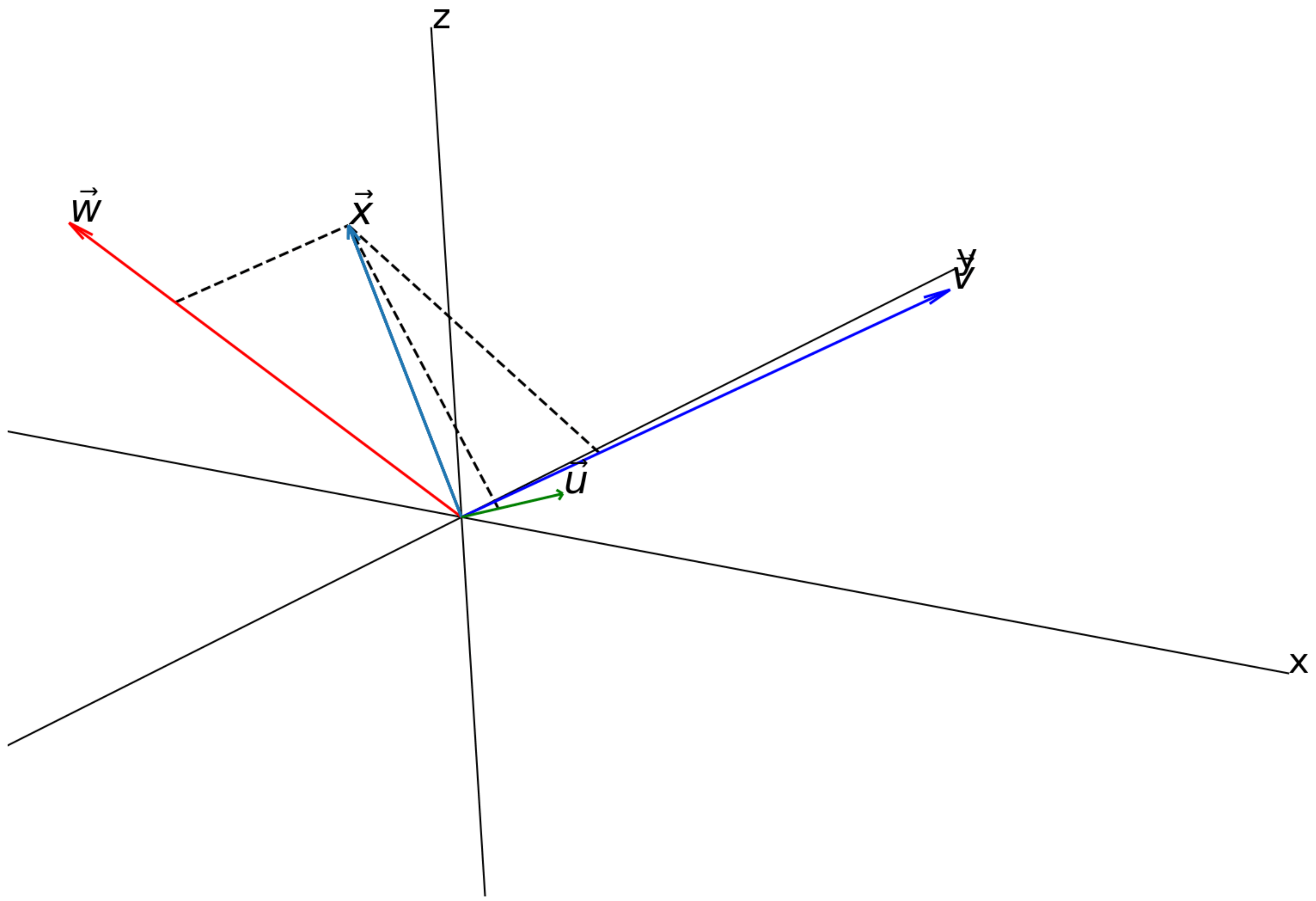


$\vec{x}$ is an arbitray vector

Orthonormal Basis

An orthonormal basis is a special case of an orthogonal basis with unit vectors.

$$\langle \vec{x}, \vec{v}_j \rangle = \left\langle \sum_{i=1}^{n} \alpha_i \vec{v}_i \, , \vec{v}_j \right\rangle = \sum_{i=1}^{n} \alpha_i \langle \vec{v}_i, \vec{v}_j \rangle = \alpha_i \langle \vec{v}_i, \vec{v}_j \rangle = \alpha_j \left\| \vec{v}_j \right\|_2^2 \Longrightarrow \alpha_j = \frac{\langle \vec{x}, \vec{v}_j \rangle}{\left\| \vec{v}_j \right\|_2^2}$$



$$\left\| \vec{v}_j \right\|_2 = 1 \Longrightarrow \alpha_j = \langle \vec{x}, \vec{v}_j \rangle$$

$\vec{x}$ is an arbitray vector

Finding the coordinates gets even easier!

Standard Basis

The standard basis is an orthonormal basis with the following vectors:

$$\begin{pmatrix}1\\0\\0\\\vdots\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\\\vdots\\0\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\\\vdots\\0\\0\end{pmatrix}, \cdots, \begin{pmatrix}0\\0\\0\\\vdots\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\0\\\vdots\\0\\1\end{pmatrix}$$

Usually denoted as:

$$\vec{e}_1, \vec{e}_2, \vec{e}_3, \dots, \vec{e}_{n-1}, \vec{e}_n$$

Known as the standard basis vectors.

Matrices

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \\ a_{4,1} & a_{4,2} & a_{4,3} \end{pmatrix} \in \mathbb{R}^{4\times 3}$$

Short-hand:

$$A = \left(a_{ij}\right) \in \mathbb{R}^{m\times n}$$

$$A + B = \left(a_{ij} + b_{ij}\right)$$

$$cA = \left(ca_{ij}\right)$$

Multiplying a matrix with a vector:

$$\vec{y} = A\vec{x}$$

$$\begin{pmatrix} ?? \\ ?? \\ ?? \\ ?? \end{pmatrix} = \begin{pmatrix} 3 & 6 & 9 \\ 1 & 1 & -1 \\ 2 & 0 & 4 \\ -2 & 0 & -4 \end{pmatrix} \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}$$

Multiplying a matrix with a matrix:

$$AX = A\begin{pmatrix} x_{11} & \cdots & x_{1n} \\ x_{21} & \cdots & x_{2n} \\ x_{31} & \cdots & x_{3n} \end{pmatrix}$$

$$= A(\vec{x}_{\cdot 1} \quad \cdots \quad \vec{x}_{\cdot n})$$

$$= (A\vec{x}_{\cdot 1} \quad \cdots \quad A\vec{x}_{\cdot n})$$

$(AB)C = A(BC)$, but $AB \neq BA$

Matrices

Transpose: $A^\top$

If $A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \\ a_{4,1} & a_{4,2} & a_{4,3} \end{pmatrix}$

then $A^\top = \begin{pmatrix} a_{1,1} & a_{2,1} & a_{3,1} & a_{4,1} \\ a_{1,2} & a_{2,2} & a_{3,2} & a_{4,2} \\ a_{1,3} & a_{2,3} & a_{3,3} & a_{4,3} \end{pmatrix}$

Alternative inner product notation:

$$\vec{x}^\top \vec{y} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$= \sum_{i=1}^{n} x_i y_i = \langle \vec{x}, \vec{y} \rangle$$

Will treat vectors as column vectors:

$$\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

Outer product:

$$\vec{x}\vec{y}^\top = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \begin{pmatrix} y_1 & \cdots & y_n \end{pmatrix}$$

$$= \begin{pmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_n y_1 & x_n y_2 & \cdots & x_n y_n \end{pmatrix}$$

Vector transpose denotes a row vector:

$$\vec{x}^\top = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix}$$

Properties of Transpose:

$$(A^\top)^\top = A$$

$$(A + B)^\top = A^\top + B^\top$$

$$(cA)^\top = cA^\top$$

$$(AB)^\top = B^\top A^\top$$

Matrices

Rank: Number of linearly independent columns, or equivalently the number of linearly independent rows, denoted as $\mathrm{rank}(A)$

$$A = \begin{pmatrix} -1 & 2 \\ 0 & 0 \\ 2 & -4 \end{pmatrix}$$ $\mathrm{rank}(A) = 1$

Rank-deficient

$$B = \begin{pmatrix} -1 & 2 \\ 0 & 1 \\ 2 & -4 \end{pmatrix}$$ $\mathrm{rank}(B) = 2$

Full-rank

Identity Matrix:

$$I = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

Property: $I\vec{x} = \vec{x}$

Inverse:

A matrix A^{-1} such that $A^{-1}A = I$

Or equivalently,

A matrix A^{-1} such that $AA^{-1} = I$

Not all matrices are invertible! Only those that are square and full-rank are.

Interpreting Matrices

- Each matrix represents a linear transformation, that is, a linear function that maps a vector to a vector.

$$A = \begin{pmatrix} \vec{a}_{\cdot 1} & \cdots & \vec{a}_{\cdot n} \end{pmatrix}$$

$$A\vec{x} = \begin{pmatrix} \vec{a}_{\cdot 1} & \cdots & \vec{a}_{\cdot n} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1\vec{a}_{\cdot 1} + \cdots + x_n\vec{a}_{\cdot n}$$

- This is a linear combination of $\vec{a}_{\cdot 1}, \dots, \vec{a}_{\cdot n}$, where the coefficients are given by $\vec{x}$

  - Before: $\vec{x} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n$

  - After: $A\vec{x} = x_1\vec{a}_{\cdot 1} + \cdots + x_n\vec{a}_{\cdot n}$

Essentially replacing the original basis vectors with the columns of A

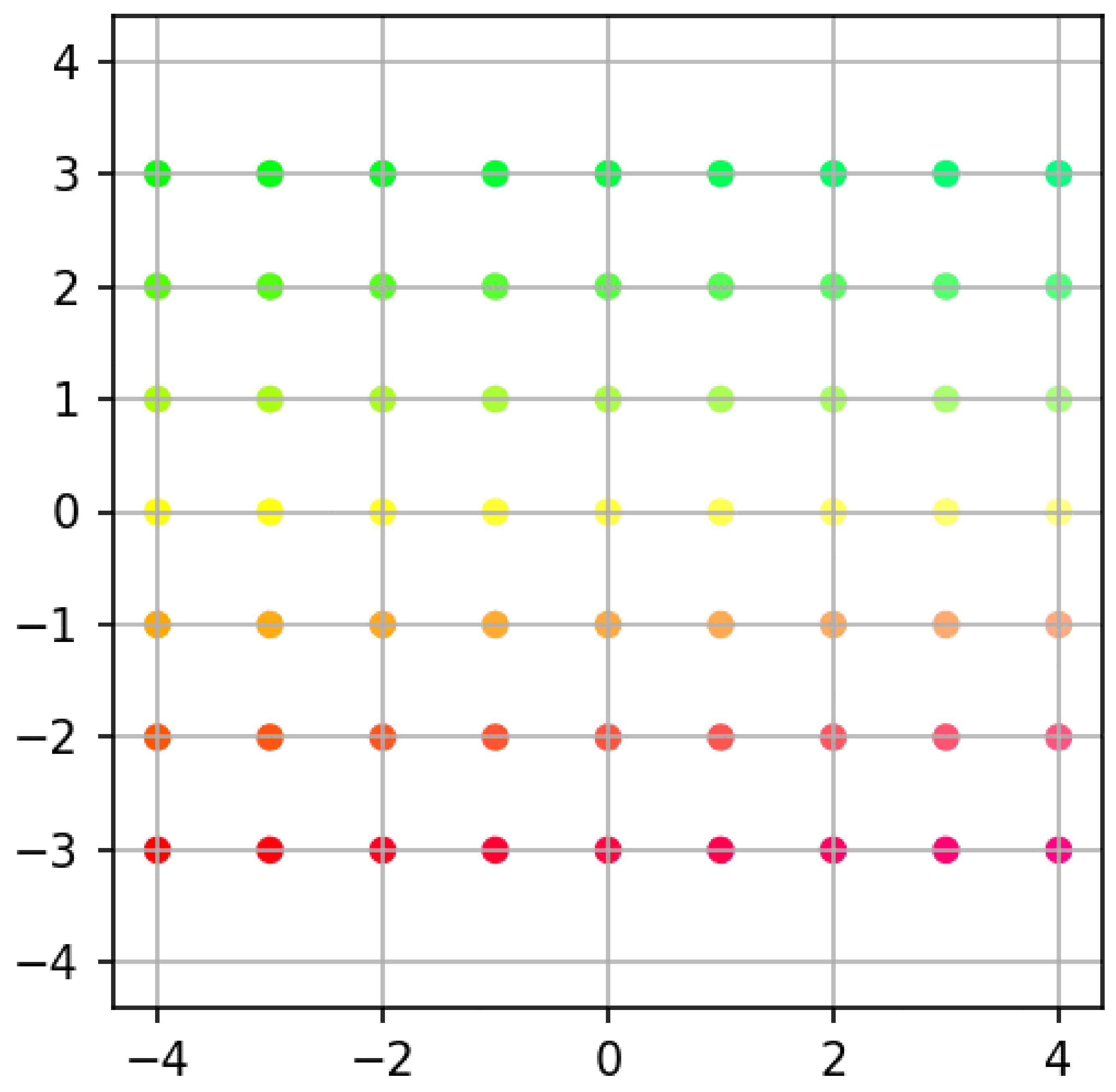
Diagonal Matrices

$$A = \begin{pmatrix} a_{1,1} & 0 \\ 0 & a_{2,2} \end{pmatrix}$$

$$A\vec{x} = A\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{1,1}x_1 \\ a_{2,2}x_2 \end{pmatrix}$$

In the case of diagonal matrices, rank is the number of non-zero entries.

Full-rank: $A = \begin{pmatrix} a_{1,1} & 0 \\ 0 & a_{2,2} \end{pmatrix}$



0.9 0.0

0.0 0.5

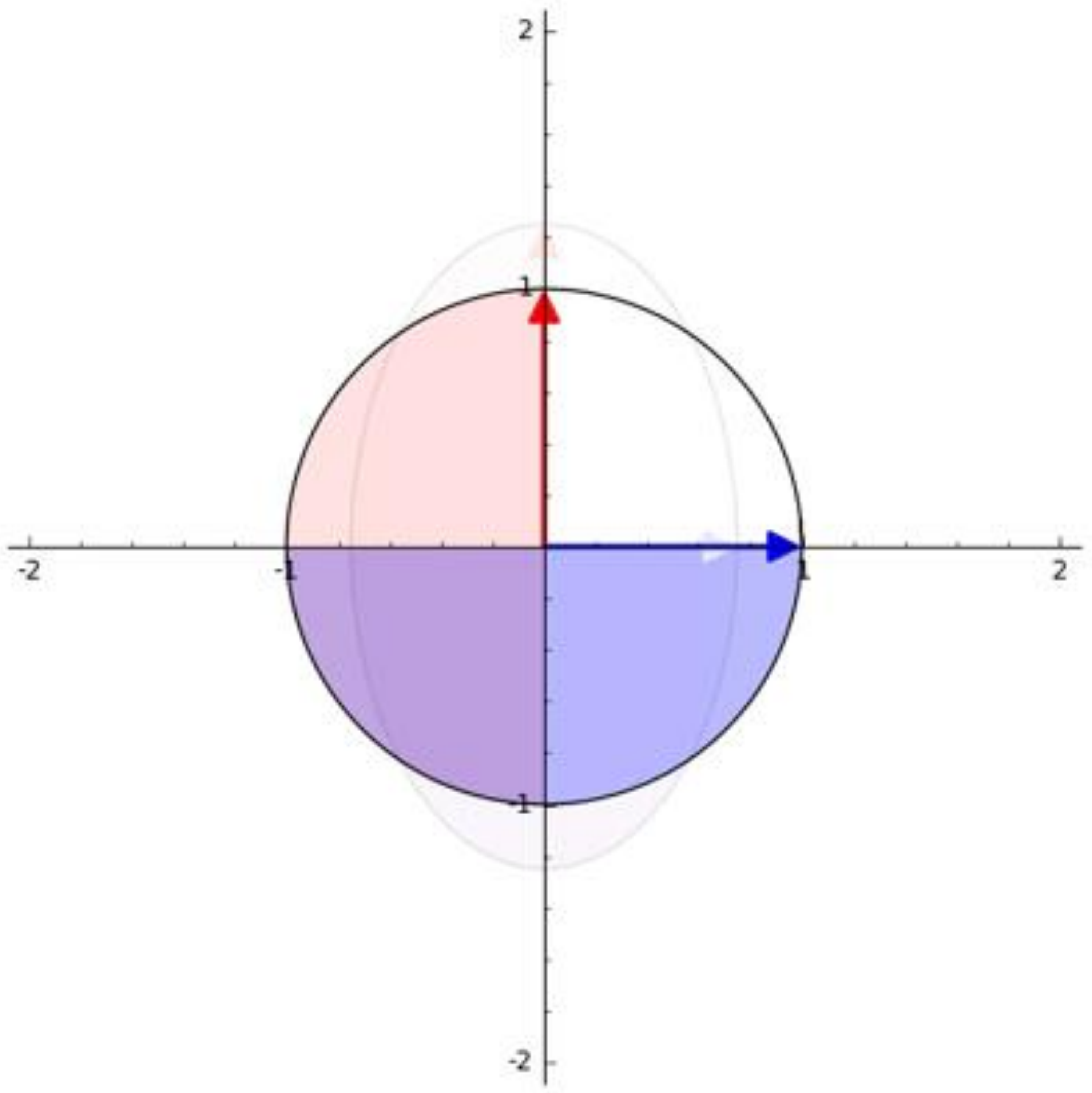
Diagonal Matrices

$$A = \begin{pmatrix} a_{1,1} & 0 \\ 0 & a_{2,2} \end{pmatrix}$$

$$A\vec{x} = A\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{1,1}x_1 \\ a_{2,2}x_2 \end{pmatrix}$$

In the case of diagonal matrices, rank is the number of non-zero entries.

Full-rank: $A = \begin{pmatrix} a_{1,1} & 0 \\ 0 & a_{2,2} \end{pmatrix}$



Credit: Ryan Holbrook

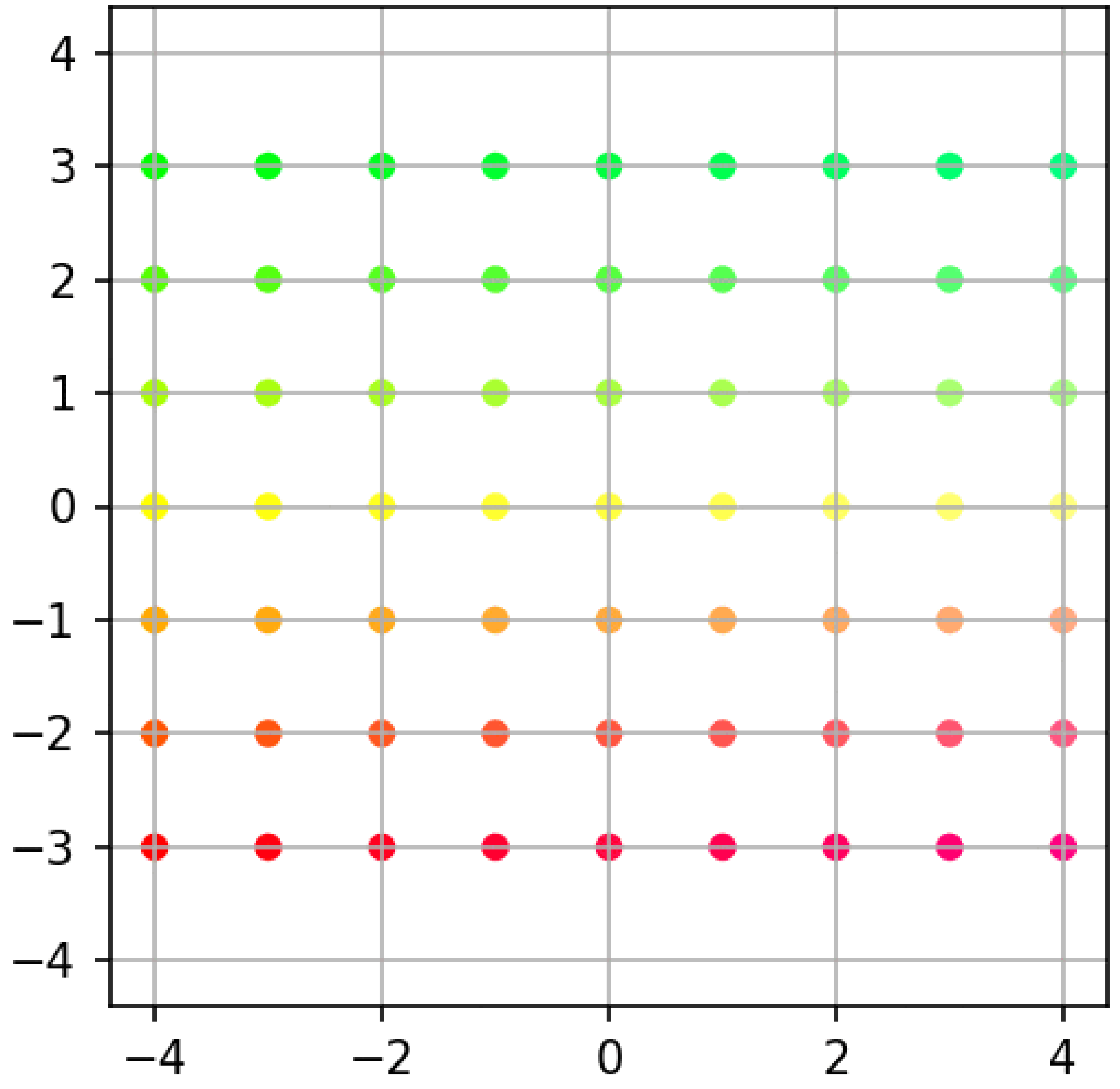
Diagonal Matrices

$$A = \begin{pmatrix} a_{1,1} & 0 \\ 0 & a_{2,2} \end{pmatrix}$$

$$A\vec{x} = A\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{1,1}x_1 \\ a_{2,2}x_2 \end{pmatrix}$$

In the case of diagonal matrices, rank is the numbe of non-zero entries.

Rank-deficient: $A = \begin{pmatrix} 0 & 0 \\ 0 & a_{2,2} \end{pmatrix}$



0.0	0.0
0.0	0.5

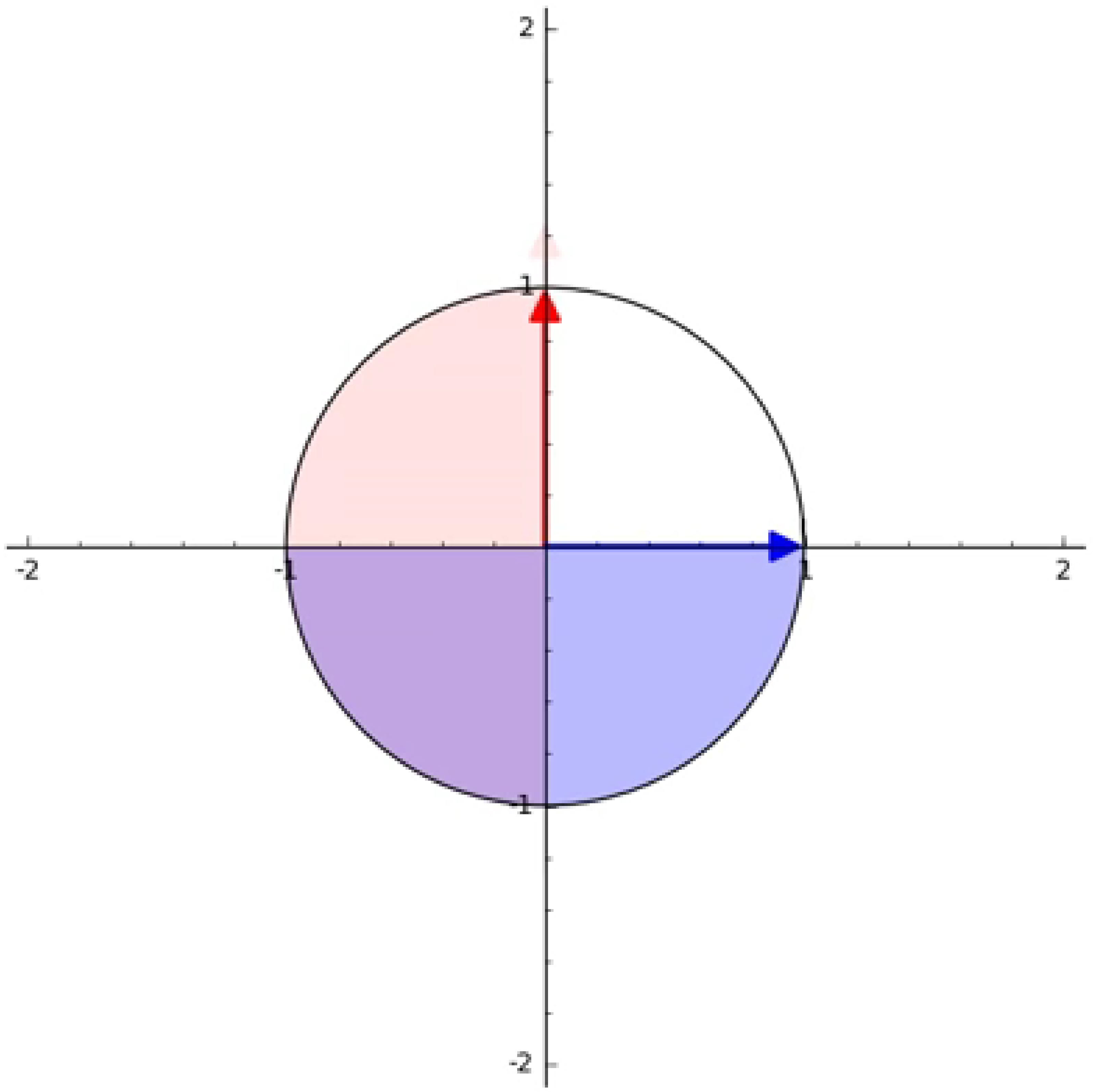
Diagonal Matrices

$$A = \begin{pmatrix} a_{1,1} & 0 \\ 0 & a_{2,2} \end{pmatrix}$$

$$A\vec{x} = A\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{1,1}x_1 \\ a_{2,2}x_2 \end{pmatrix}$$

In the case of diagonal matrices, rank is the number of non-zero entries.

Rank-deficient: $A = \begin{pmatrix} 0 & 0 \\ 0 & a_{2,2} \end{pmatrix}$



Credit: Ryan Holbrook

Orthogonal Matrices

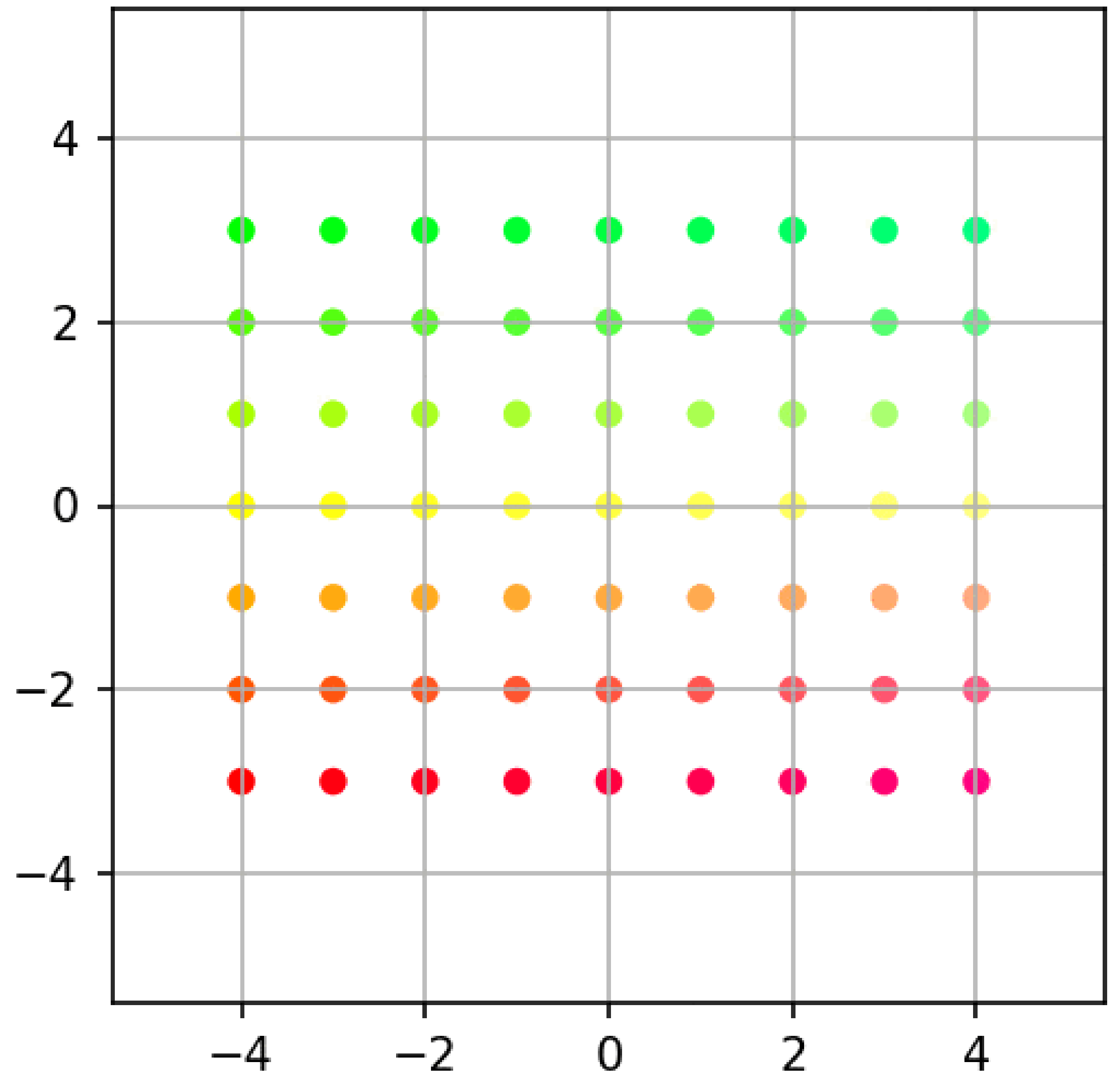
A square matrix A such that $AA^\top = I$ and $A^\top A = I$ (implies $A^{-1} = A^\top$)

$$A^\top A = \begin{pmatrix} \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot n} \\ \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot n} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{a}_{\cdot n}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot n}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot n}^\top \vec{a}_{\cdot n} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

The columns of A form an orthonormal basis

Determinant of 1: $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$



$$\begin{matrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{matrix}$$

Orthogonal Matrices

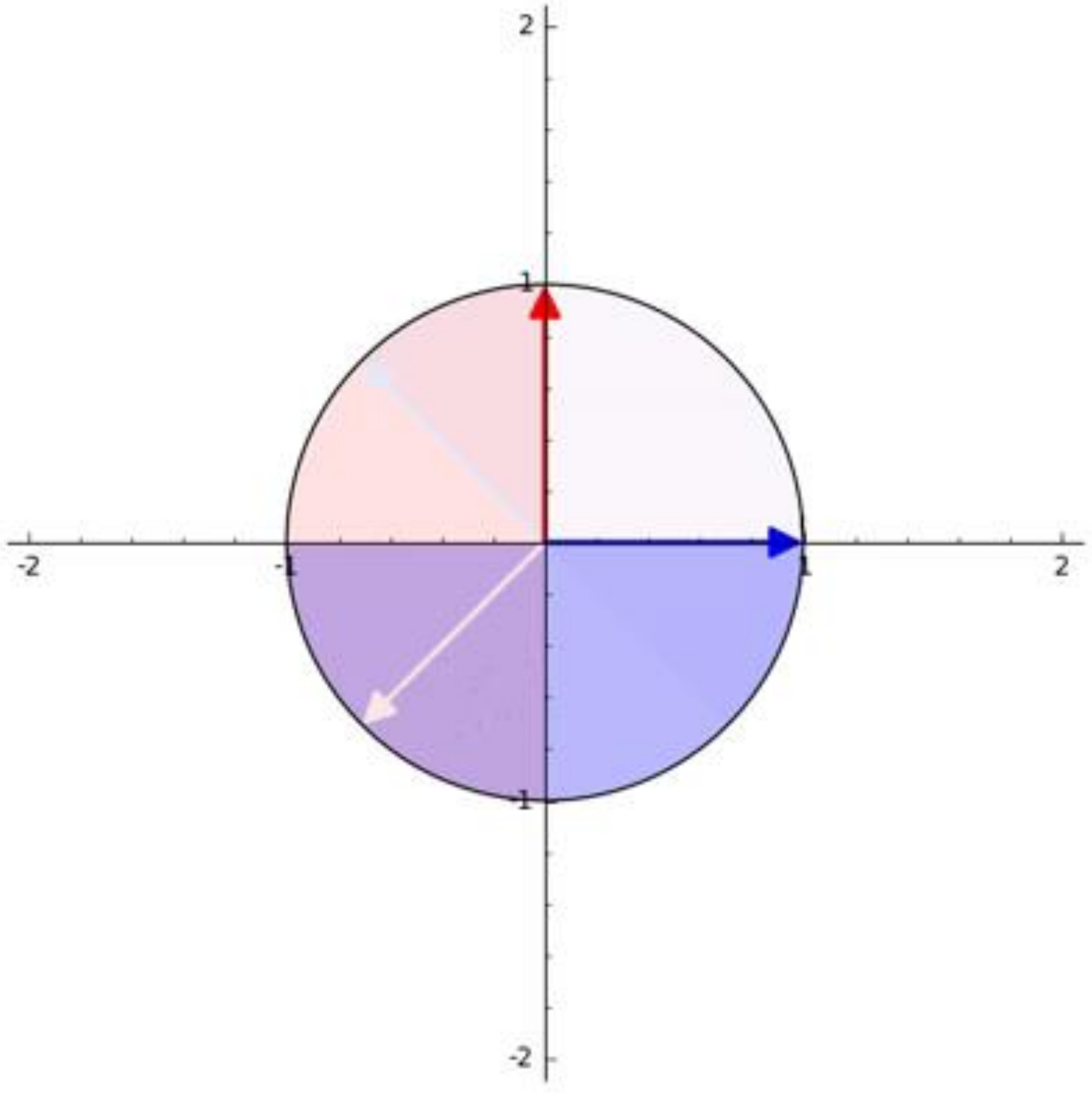
A square matrix A such that $AA^\top = I$
and $A^\top A = I$ (implies $A^{-1} = A^\top$)

$$A^\top A = \begin{pmatrix} \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot n} \\ \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot n} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{a}_{\cdot n}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot n}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot n}^\top \vec{a}_{\cdot n} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

The columns of A form an orthonormal basis

Determinant of 1: $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$



Credit: Ryan Holbrook

Orthogonal Matrices

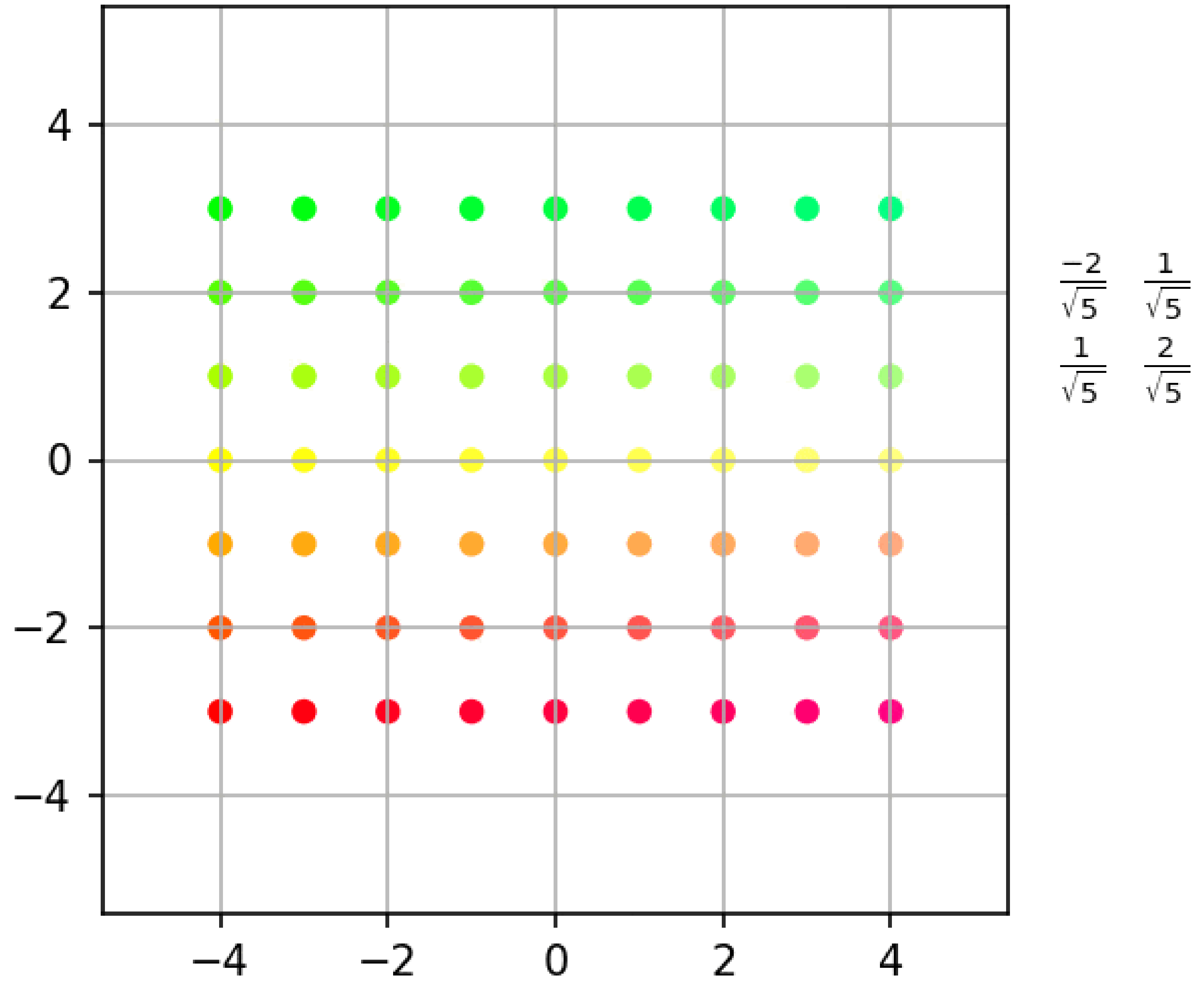
A square matrix A such that $AA^\top = I$ and $A^\top A = I$ (implies $A^{-1} = A^\top$)

$$A^\top A = \begin{pmatrix} \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot n} \\ \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot n} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{a}_{\cdot n}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot n}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot n}^\top \vec{a}_{\cdot n} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

The columns of A form an orthonormal basis

Determinant of -1: $A = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$



Orthogonal Matrices

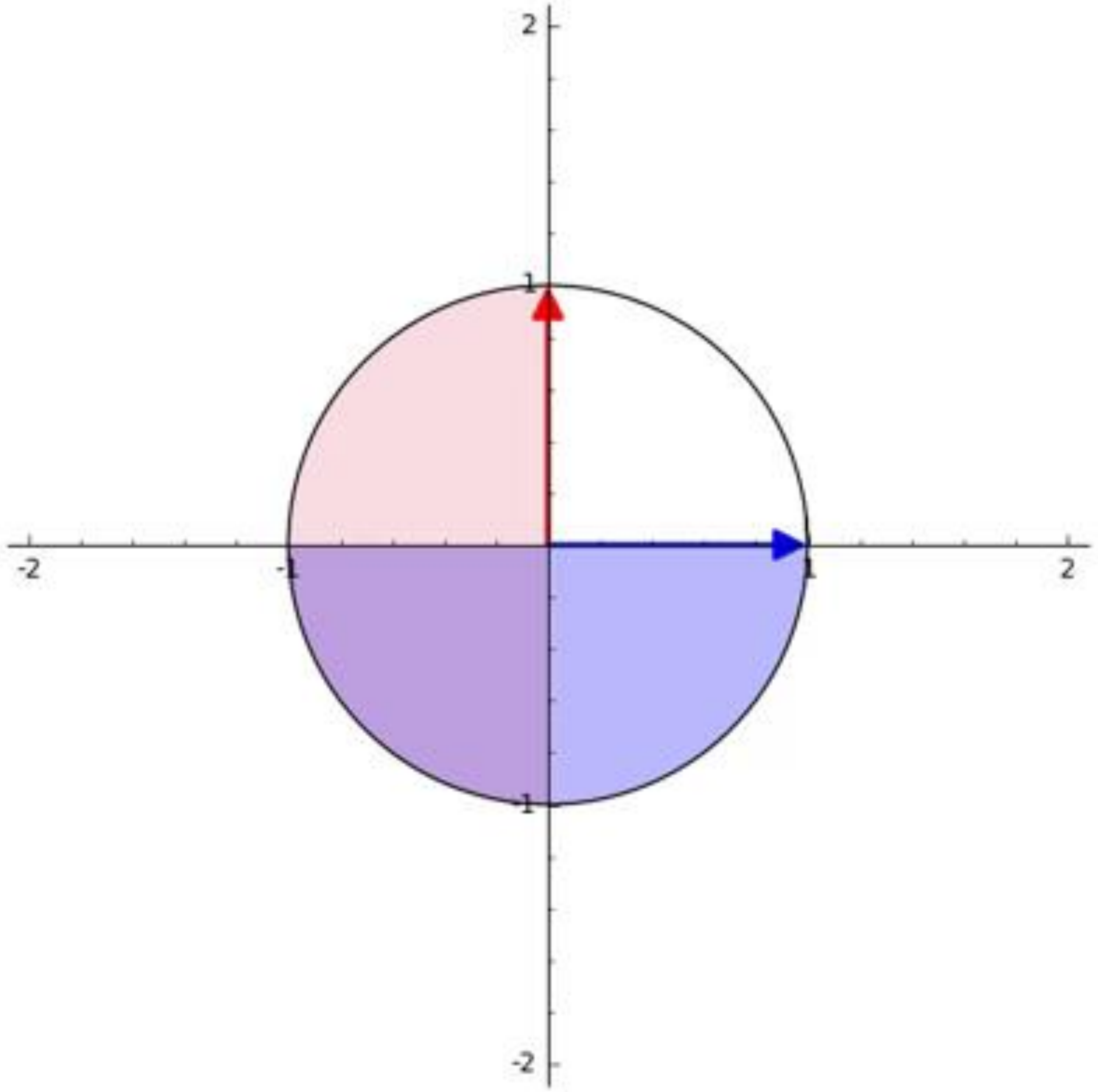
A square matrix A such that $AA^\top = I$ and $A^\top A = I$ (implies $A^{-1} = A^\top$)

$$A^\top A = \begin{pmatrix} \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot n} \\ \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot n} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{a}_{\cdot n}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot n}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot n}^\top \vec{a}_{\cdot n} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

The columns of A form an orthonormal basis

Determinant of -1: $A = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$



Credit: Ryan Holbrook

Singular Value Decomposition (SVD)

All matrices can be decomposed into a sequence of:

1. Orthogonal matrix (rotation/reflection)

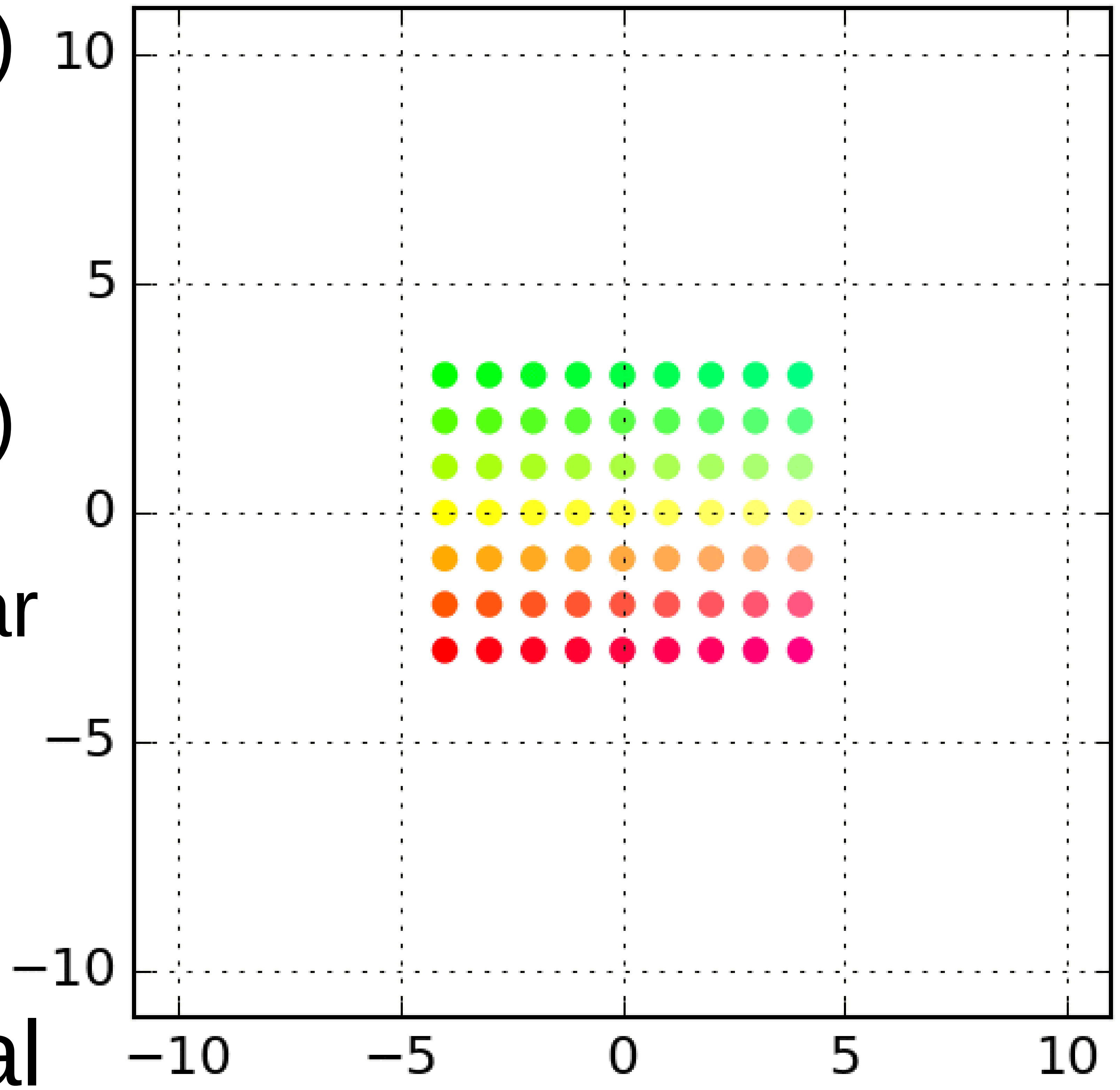
2. Diagonal matrix (scaling along axes)

3. Orthogonal matrix (rotation/reflection)

This decomposition is known as singular value decomposition (SVD):

$$A = U\Sigma V^{\top}$$

U, V are orthogonal, Σ is diagonal with real non-negative entries



$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$$

Singular Value Decomposition (SVD)

All matrices can be decomposed into a sequence of:

1. Orthogonal matrix (rotation/reflection)

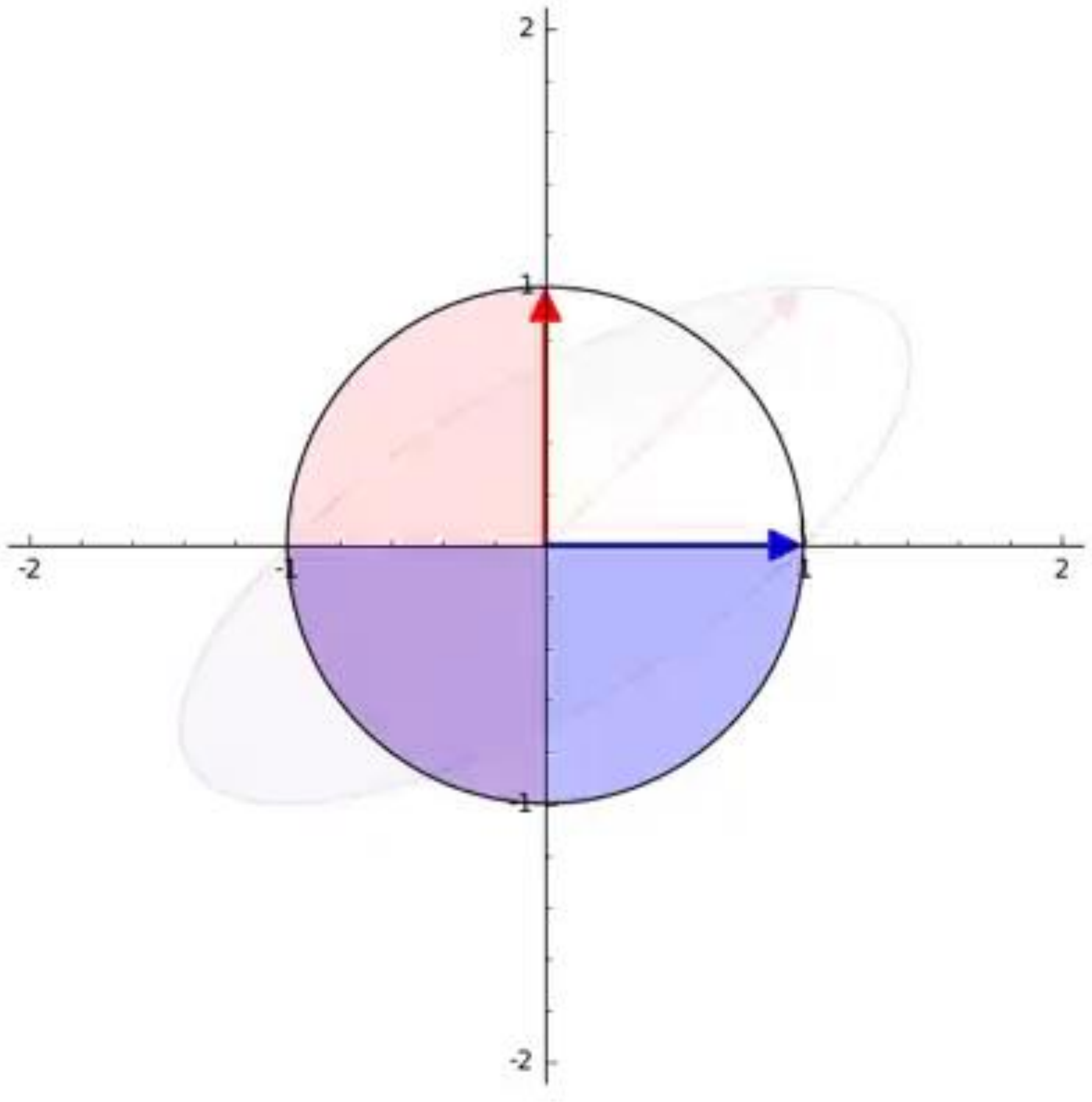
2. Diagonal matrix (scaling along axes)

3. Orthogonal matrix (rotation/reflection)

This decomposition is known as singular value decomposition (SVD):

$$A = U\Sigma V^\top$$

U, V are orthogonal, Σ is diagonal with real non-negative entries



Credit: Ryan Holbrook