Linear Algebra Background

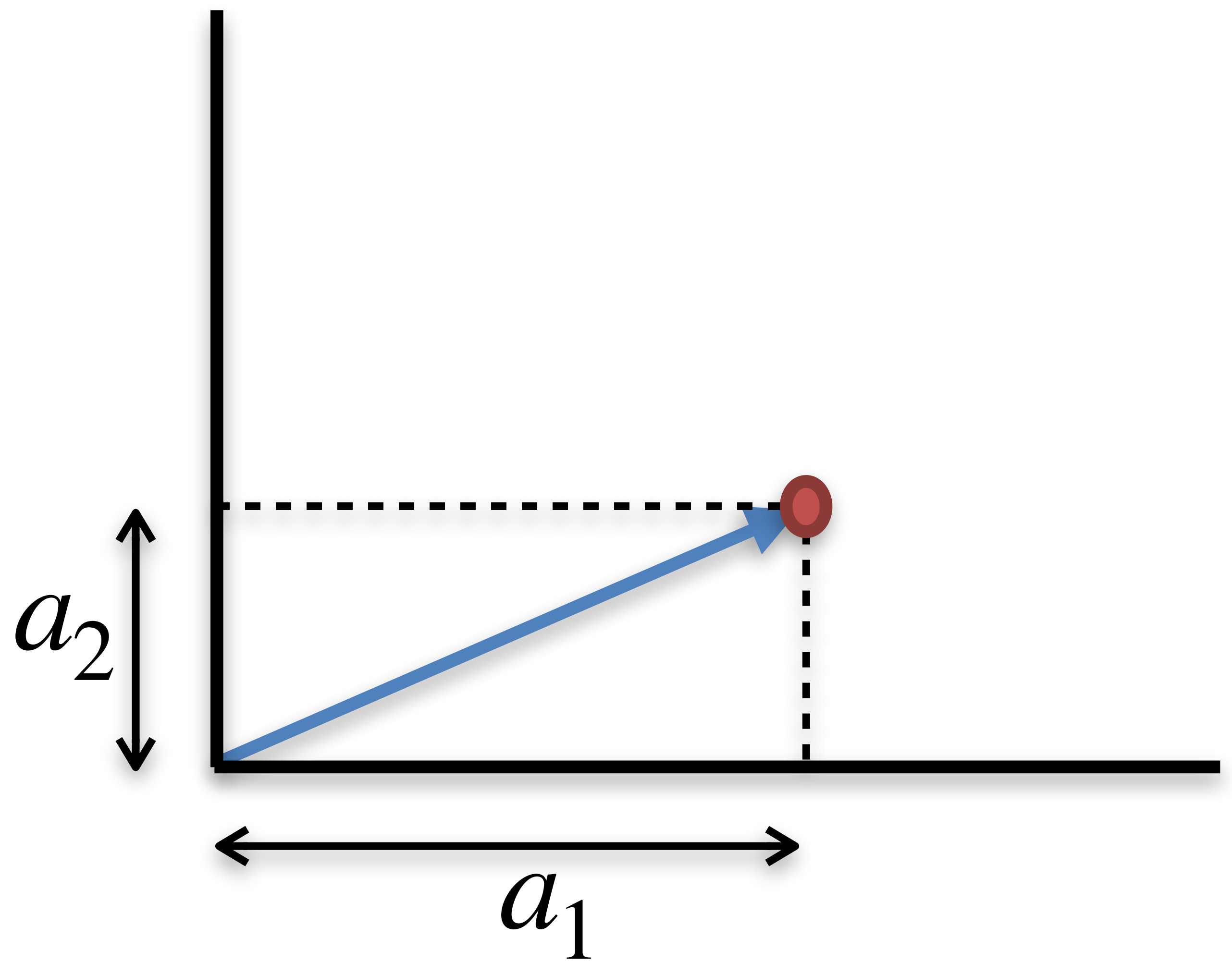
Debapriyo Majumdar

Indian Statistical Institute Kolkata

Scalars and Vectors

- A scalar is a number $a \in \mathbb{R}$
- A vector is a tuple of n numbers $(a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$
- Interpretation of vector:

1. A point in the n dimensional space
2. A magnitude and a direction (e.g. velocity)
3. An object represented by n features (attributes), with the values of the features being $a_1, \ldots, a_n$



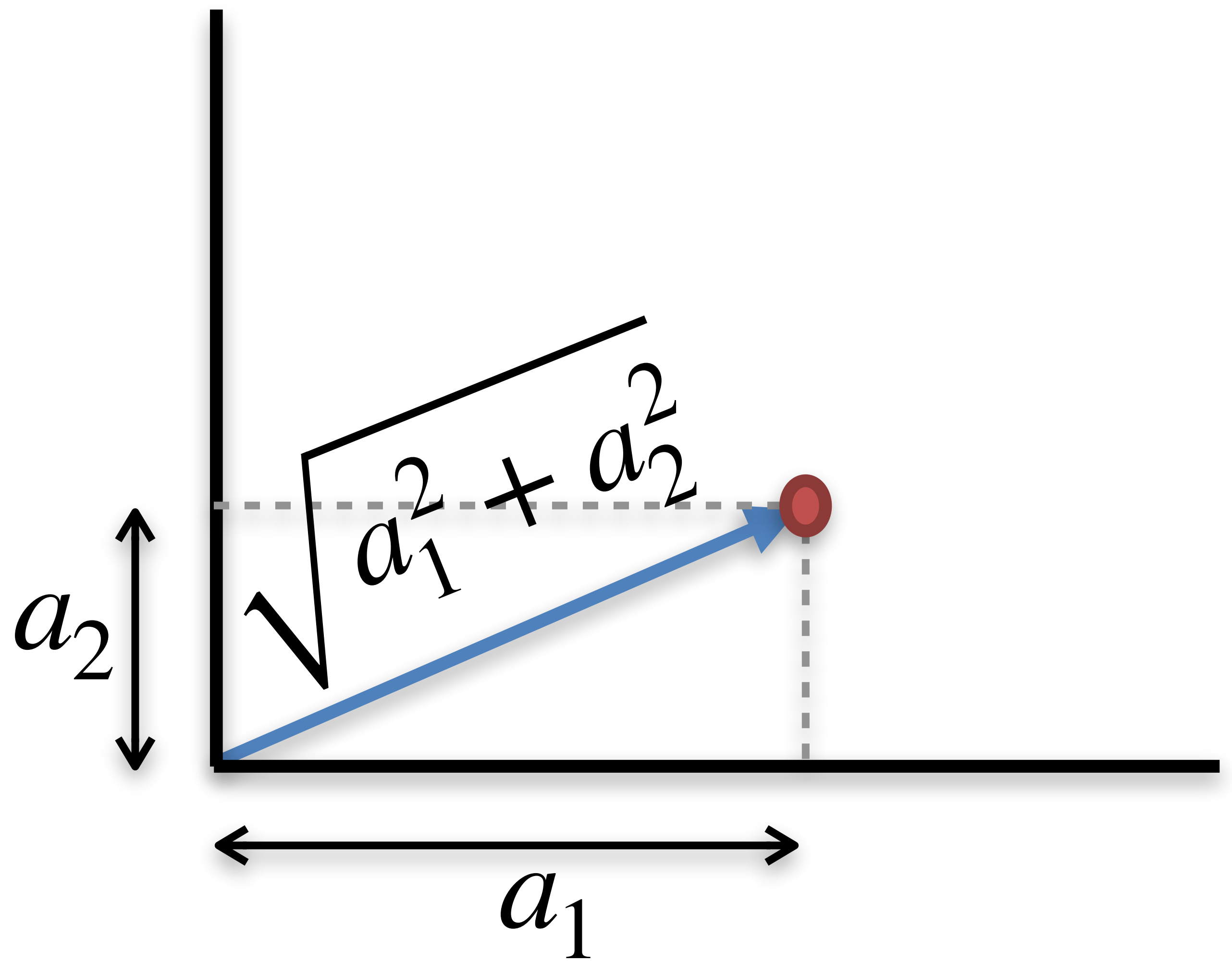
- We would usually interpret vectors as points in the n dimensional space (equivalently, objects specified by the n features)
- In numpy, we can create a vector using the 1-D array to start with

```
import numpy as np

# Creating an array from a list
a = np.array([1,2,0,1,3])
```

Norm of a vector

- Norm of a vector: distance between the origin (0,…,0) and the point according to the corresponding distance measure
- Euclidean (L_2) norm: $\| \mathbf{a} \|_2 = \sqrt{a_1^2 + \ldots + a_n^2}$
  - Generally called the Frobenius norm (for vectors and matrices)



- In general L_p norm: $\| \mathbf{a} \|_p = (a_1^p + \ldots + a_n^p)^{\frac{1}{p}}$
- For p = 1, the L_1 norm: $\| \mathbf{a} \|_1 = \sum_{i=1}^{n} |a_i|$
  - Also called the Manhattan (taxicab) norm. Why?
- As $p \to \infty$, the L_∞ norm: $\| \mathbf{a} \|_\infty = \max_{i=1}^{n} |a_i|$

```
from numpy import linalg as LA

# The 2-norm  (default)
print(LA.norm(a))
# The 1-norm
print(LA.norm(a, 1))
# inf-norm
print(LA.norm(a, np.inf))
```

Basic vector operations

Let $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$

Addition of vectors: $\mathbf{a} + \mathbf{b} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{bmatrix}$

Multiplication by a scalar: $x\mathbf{a} = \begin{bmatrix} xa_1 \\ xa_2 \\ \vdots \\ xa_n \end{bmatrix}$

```
print("a:", a)
print("b:", b)
print("a+b:", a + b)
print("a-b:", a-b)
print("3a:", 3 * a)
```

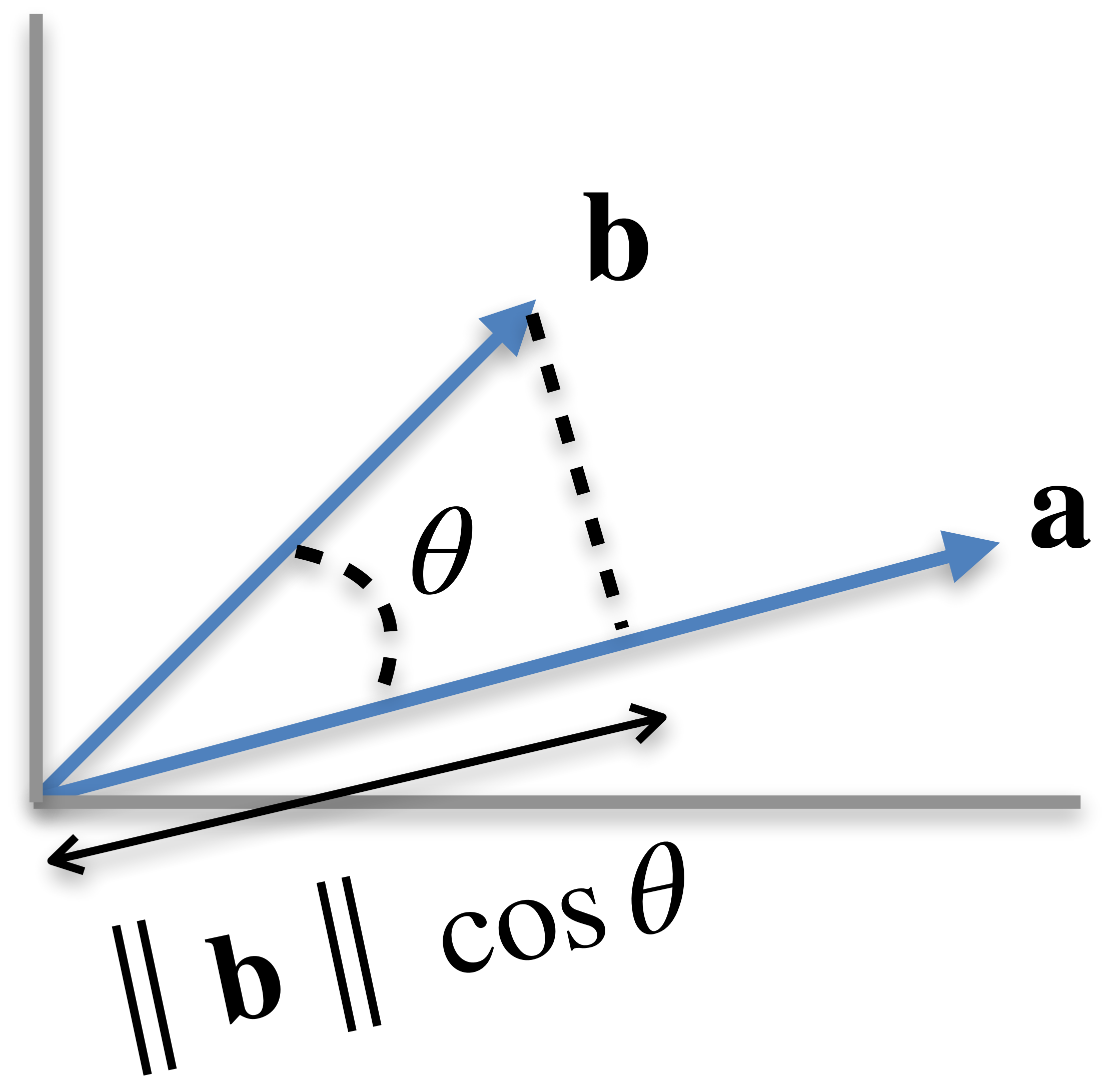
```
a: [1 2 0 1 3]
b: [ 3 -1  2 -3  1]
a+b: [ 4  1  2 -2  4]
a-b: [-2  3 -2  4  2]
3a: [3 6 0 3 9]
```

Dot product / Scalar product / Inner product

- Dot product: $\mathbf{a} \cdot \mathbf{b} = < \mathbf{a}, \mathbf{b} > = a_1 b_1 + \ldots + a_n b_n$
- In the Euclidean space: $\mathbf{a} \cdot \mathbf{b} = \| \mathbf{a} \| \, \| \mathbf{b} \| \cos \theta$
- The norm of a vector its dot product with itself
- Intuition: how much of **b** is in the direction of **a**?
  - Projection of **b** onto **a**: $\| \mathbf{b} \| \cos \theta$
  - Product of that with the length of **a**

$$\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

Convention: we usually mean **column vectors**



Vector space

- *Definition*: A vector space over real numbers ($\mathbb{R}$) is a set V of objects, called *vectors*, satisfying the following properties:
  - The set V is closed under an operation called vector addition: for all $\mathbf{v}, \mathbf{w} \in V$, we have $\mathbf{v} + \mathbf{w} \in V$.
  - There is a unique vector $\mathbf{0} \in V$ such that $\mathbf{0} + \mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in V$.
  - For each $\mathbf{v} \in V$, there exists a unique vector $-\mathbf{v} \in V$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.
  - The set V is closed under an operation called scalar multiplication: for all $\mathbf{v} \in V$ and $\alpha \in \mathbb{R}$, we have $\alpha\mathbf{v} \in V$. Also, the following hold for any $\alpha, \beta \in \mathbb{R}$ and $\mathbf{v}, \mathbf{w} \in V$:
    - $1\mathbf{v} = \mathbf{v}$
    - $(\alpha\beta)\mathbf{v} = \alpha(\beta\mathbf{v})$
    - $\alpha(\mathbf{v} + \mathbf{w}) = \alpha\mathbf{v} + \alpha\mathbf{w}$
    - $(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$

The set of *n-tuples* $\mathbb{R}^n$ is a vector space over real numbers ($\mathbb{R}$) for any $n \geq 1$. In most of our life in data mining, we will be happy dealing with this vector space.

Subspaces

- *Definition*: Let V be a vector space over $\mathbb{R}$. A subspace of V is a *subset* $W \subseteq V$ such that W itself is a vector space over $\mathbb{R}$ with the same vector addition and scalar multiplication.
- Examples: Let $V = \mathbb{R}^2$, the set of points on the 2-dimensional plane. What are some of its subspaces?

Is $\{(0,0)\}$ a subspace of $\mathbb{R}^2$?

Yes. All required properties are satisfied.

Is $\{(0,0), (0,1)\}$ a subspace of $\mathbb{R}^2$?

No. Not closed under vector addition or scalar multiplication.

Is the x axis $\{(x,0) : x \in \mathbb{R}\}$ a subspace of $\mathbb{R}^2$?

Yes. All required properties are satisfied.

Is the line $\{(x, y) \in \mathbb{R}^2 : 3x + 2y = 5\}$ a subspace of $\mathbb{R}^2$?

No. *Why?*

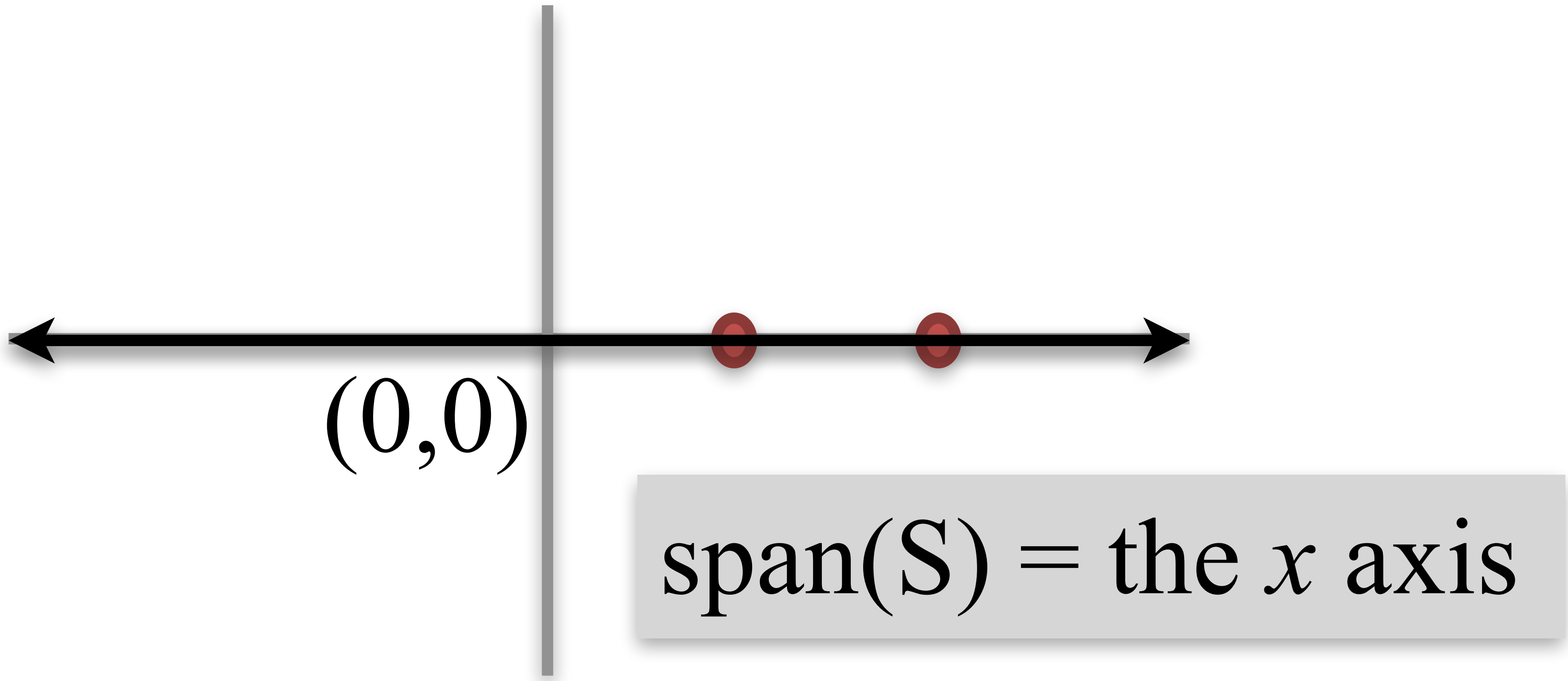
Span (Linear span / Linear hull)

- Let V be a vector space over $\mathbb{R}$. Let $S \subseteq V$ be a subset (finite or infinite) of V.
- *Definition*: The span of S is defined as the set of all finite linear combinations of vectors in S.

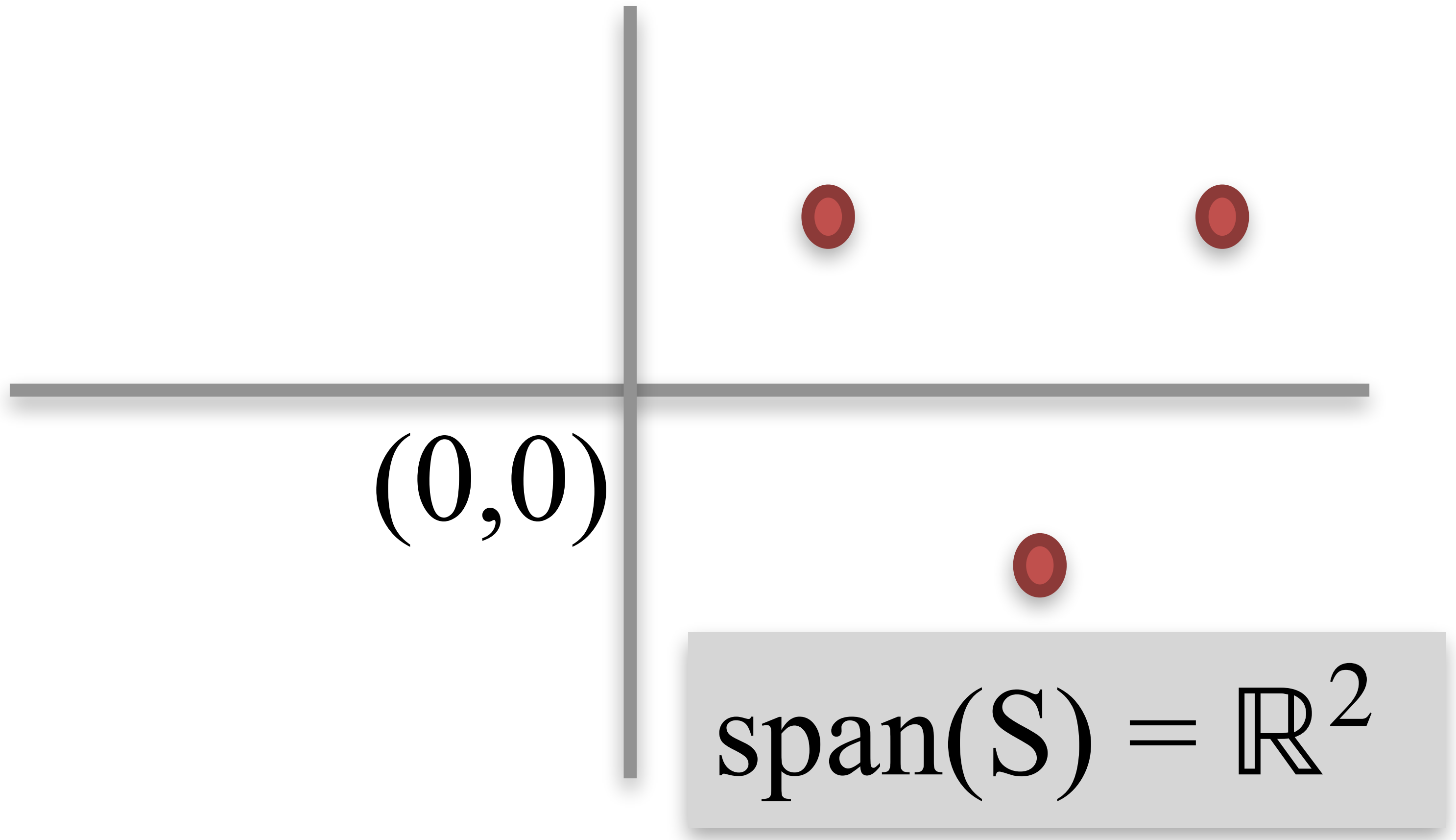
$$\text{span}(S) = \left\{ \sum_{i=1}^{k} \alpha_i v_i : k \in \mathbb{N}, v_i \in S, \alpha_i \in \mathbb{R} \right\}$$

- Examples: Let $V = \mathbb{R}^2$.

$S = \{(1,0), (2,0)\}.$



$S = \{(1,1), (2,-1), (3,1)\}.$

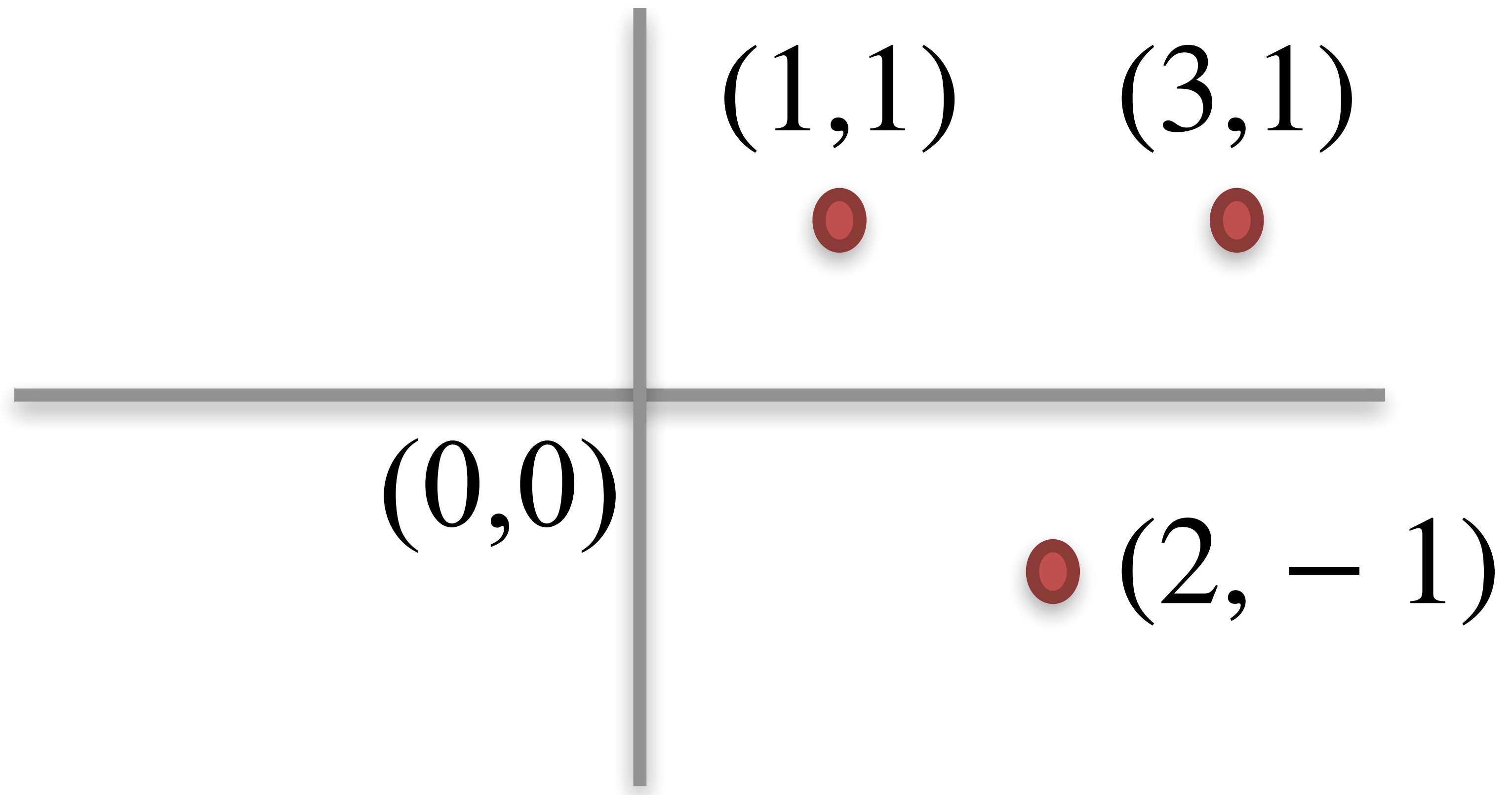


Basis

- *Definition*: A set of vectors $B \subseteq V$ is called a basis for the vector space V if every $\mathbf{v} \in V$ can be written as a unique finite linear combination of vectors in *B*

$$\mathbf{v} = \sum_{i=1}^{k} \alpha_i \mathbf{b}_i \text{ where } \alpha_i \in \mathbb{R} \text{ and } \mathbf{b}_i \in B \text{ (unique)}$$

- Clearly, a basis *B* is a subset for which span(*B*) = V
- Is that all? Consider $B = \{(1,1), (2,-1), (3,1)\}$
- span(*B*) = $\mathbb{R}^2$. Is *B* a basis for $\mathbb{R}^2$?
- Take $(3,0) \in \mathbb{R}^2$.
- Write: $(3,0) = (1,1) + (2,-1)$
- Also: $(3,0) = -1.5 \times (1,1) + 1.5 \times (3,1)$
- Uniqueness not satisfied. Why?



Linearly (in)dependent

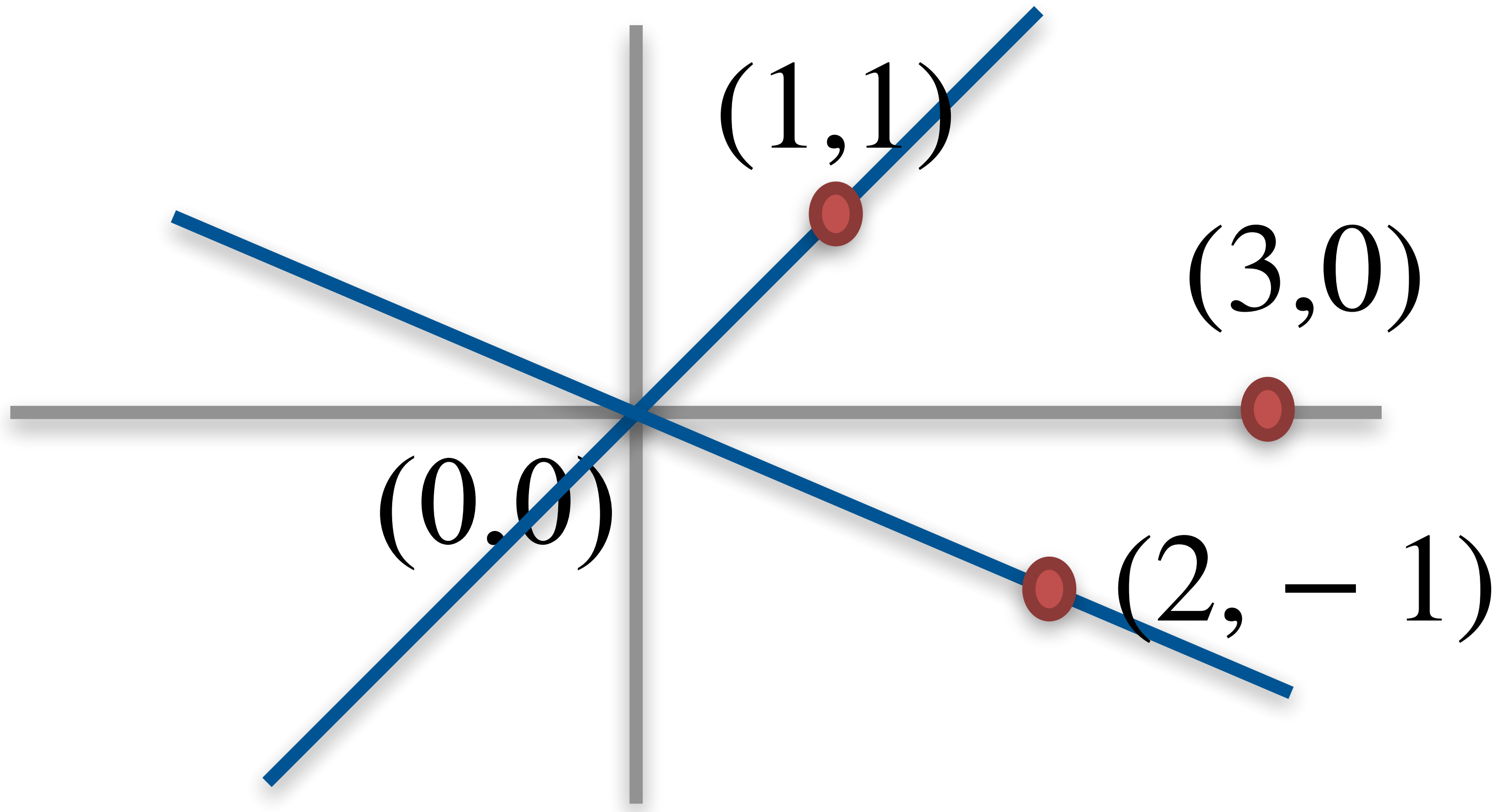
- A set of vectors is called linearly dependent if at least one of the vectors can be written as a linear combination of the others
- Linearly independent: a set of vectors which are not linearly dependent
- Examples:

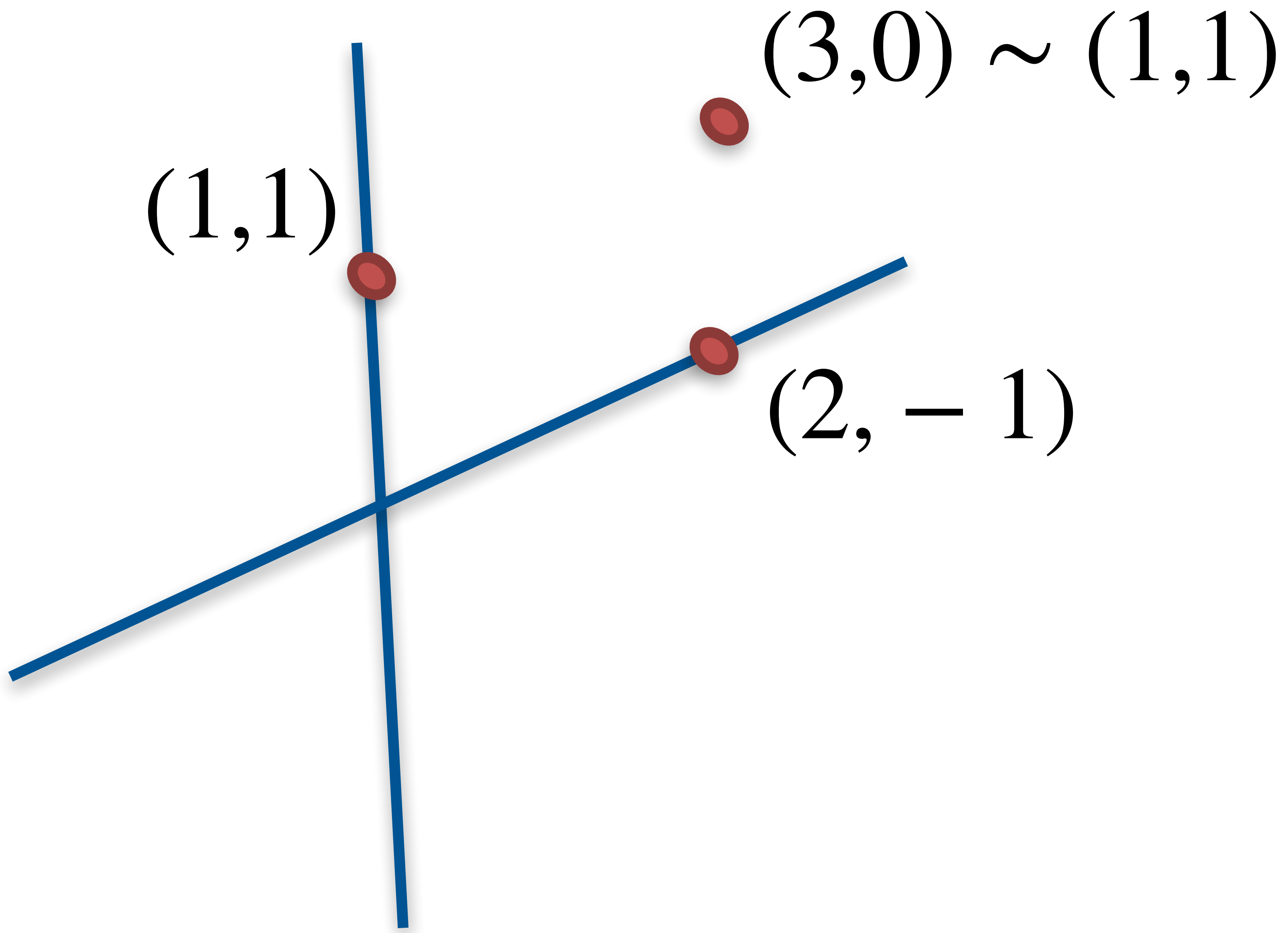
$\{(0,1),(1,0)\}$	$\{(0,1),(0,3)\}$	$\{(1,1),(-1,-2),(0,-1)\}$	$\{(12,13),(-21,2),(0,15)\}$
Linearly independent	Linearly dependent	Linearly dependent	Linearly dependent. Why?

A basis is a linearly independent subset that spans the vector space

Basis: the frame of reference

- We are used to interpreting points as (x_1, x_2) in $\mathbb{R}^2$ with the set of vectors $\{(1,0),(0,1)\}$ as the basis.
  - This is called the canonical basis
- The vector $(3,0)$ is interpreted as $3 \times (1,0) + 0 \times (0,1)$
- Change of basis: let us take the basis $\{(1,1),(2,-1)\}$
- The vector (3, 0) then becomes $1 \times (1,1) + 1 \times (2,-1)$
- More like (1,1) in the new frame of reference





Basis: frame of reference

- Let us write our canonical basis as two vectors, stacked as columns

$$B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- The vector $\mathbf{v} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$ is viewed as

$$Bv = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

- A different basis gives a different frame of reference

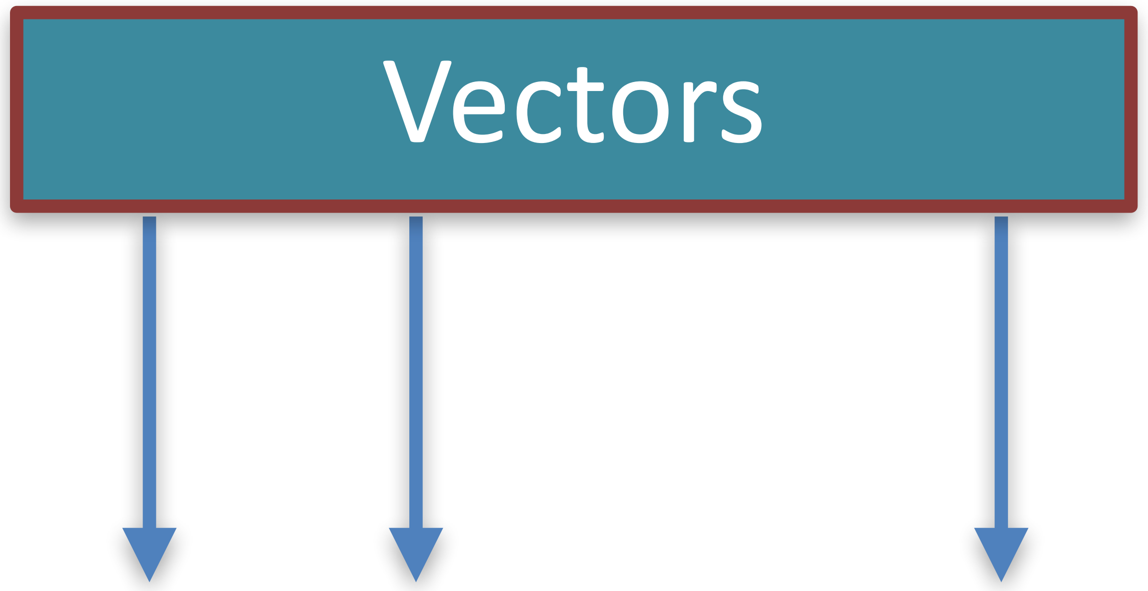
$$B_2 = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}$$

- The vector $\mathbf{v} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$ is viewed as

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} B_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

Matrix

- A matrix is a 2-dimensional $m \times n$ array (a_{ij}), for $i = 1,\ldots,m$ and $j = 1,\ldots,n$
- *Usually* a collection of vectors, stacked as columns
  - Used for many different purposes
- Interpretation for matrices
  - A basis for a vector space
  - A system of linear equations
  - Collection of data points (vectors)
  - A linear transformation
  - etc



$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Matrix Addition (element wise)

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{bmatrix}$$

Matrix Multiplication

$$a_{11}b_{11} + a_{12}b_{21} + \ldots + a_{1n}b_{n1}$$

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1p} \\ b_{21} & b_{22} & \ldots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \ldots & b_{np} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \ldots & c_{1p} \\ c_{21} & a_{22} & \ldots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mp} \end{bmatrix}$$

$$AB = C = (c_{ik}) \text{ where } c_{ik} = \sum_{j=1}^{n} a_{ij}b_{jk}$$

The Identity Matrix

- The $n \times n$ matrix

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

- For any $n \times n$ matrix A, we have $AI = IA = A$
- The 1×1 identity matrix is the multiplicative scalar identity 1

A System of Linear Equations

- A system of linear equations with n variables $x_1, \ldots, x_n$

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\
&\ldots \\
&\ldots \\
a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m
\end{aligned}
$$

- In matrix form

$$
\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}
$$

- We write $A\mathbf{x} = \mathbf{b}$

Linear Map (Transformation)

- *Definition*: Let *V* and *W* be vector spaces over $\mathbb{R}$. A function $f : V \to W$ is said to be a linear map if for all $\mathbf{u}, \mathbf{v} \in V$ and real number $c \in \mathbb{R}$,

$$f(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v})$$
$$f(c\mathbf{v}) = cf(\mathbf{v})$$

- Intuition

- Any linear map $f : \mathbb{R}^n \to \mathbb{R}^m$ is essentially multiplication by some $m \times n$ matrix *A*

For any $\mathbf{v} \in \mathbb{R}^n$, $f(\mathbf{v}) = A\mathbf{v}$

Basis: frame of reference

- Let us write our canonical basis as two vectors, stacked as columns

$$B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- The vector $\mathbf{v} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$ is viewed as

$$Bv = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

- A different basis gives a different frame of reference

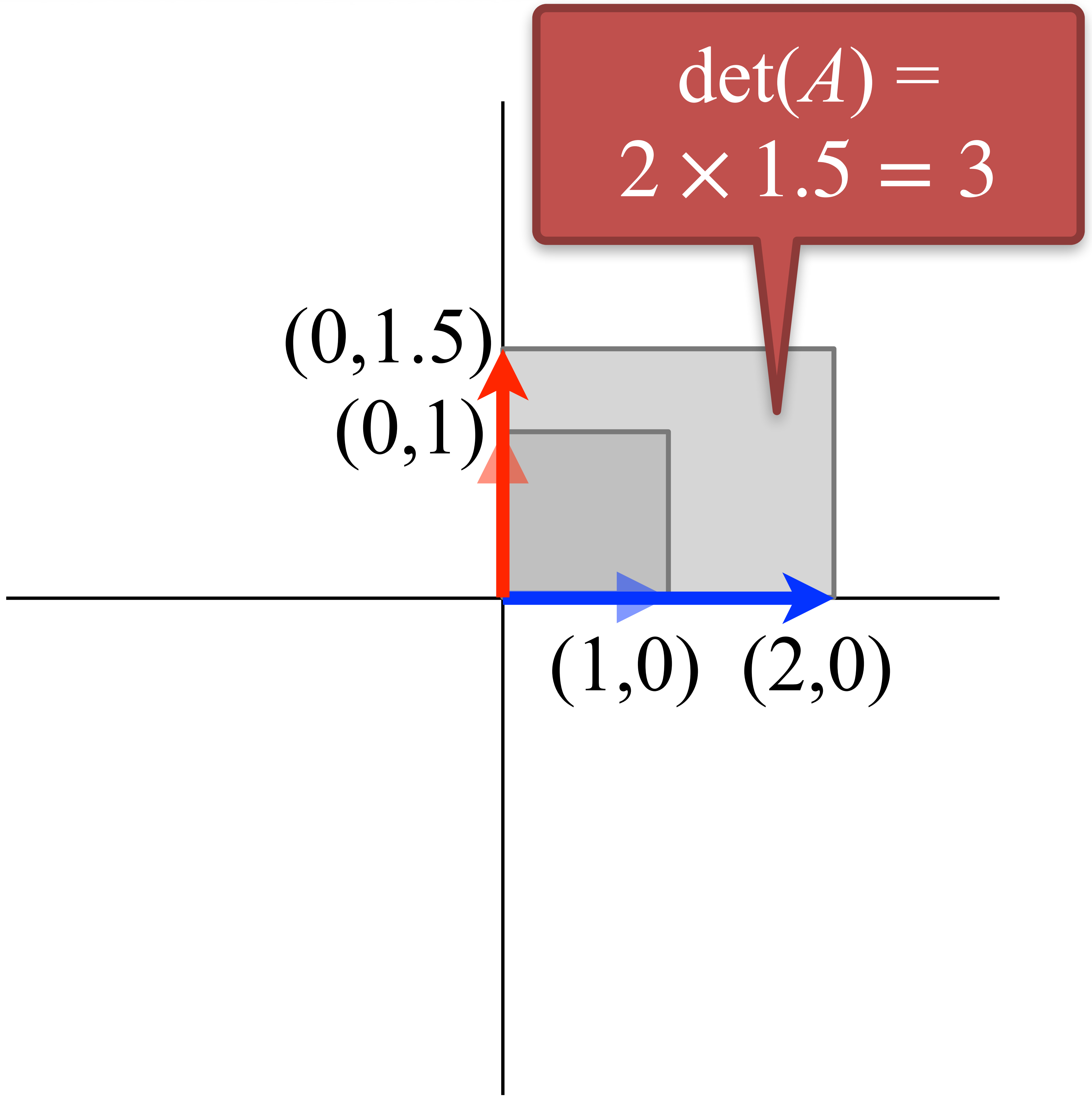
$$B_2 = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}$$

- The vector $\mathbf{v} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$ is viewed as

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} B_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

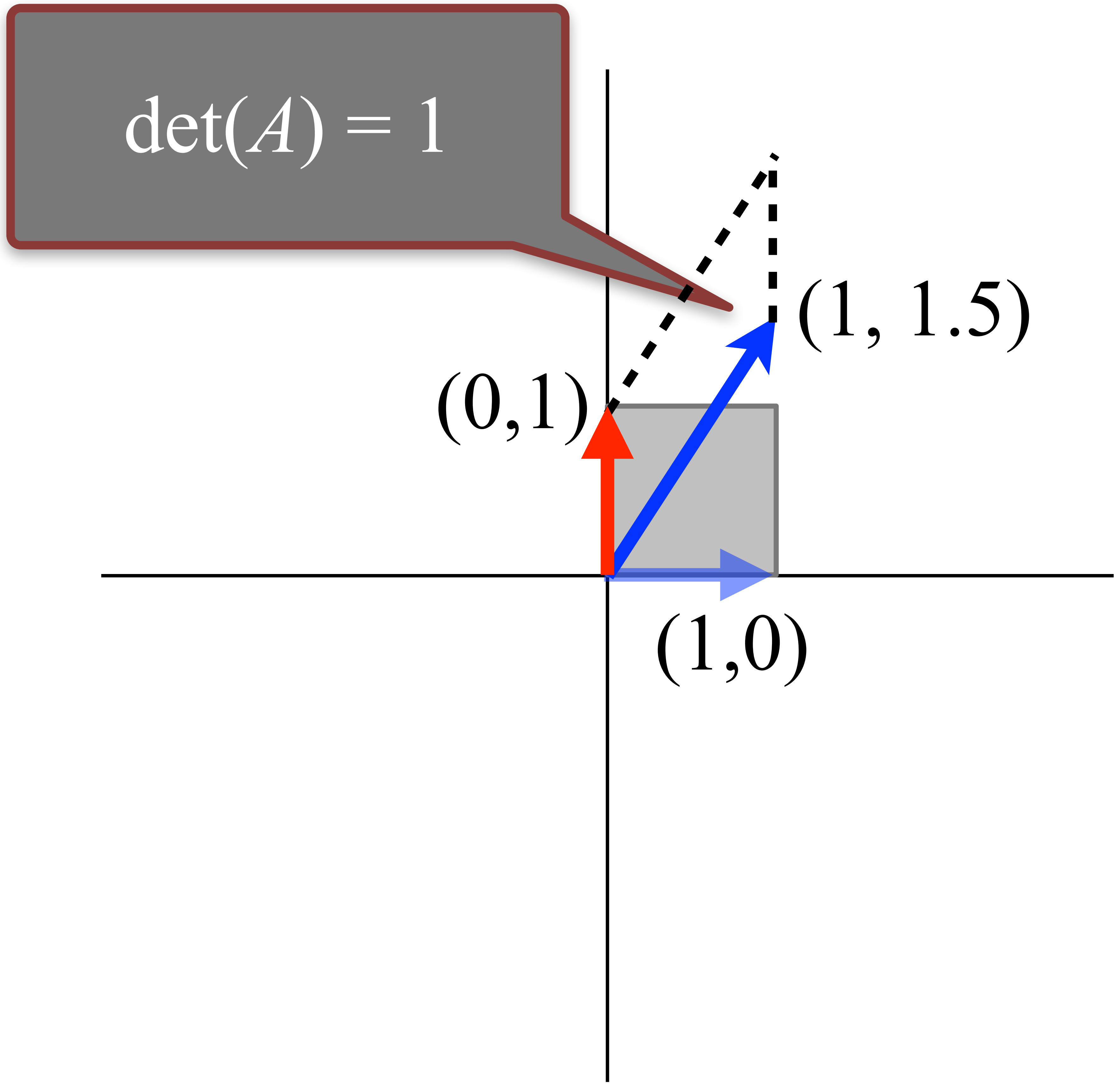
The Volume of a Linear Transformation

- Recall: A linear transformation is determined by how it transforms the basis
- Consider the canonical basis $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
- The basis vectors form the square of area 1
- Consider a transformation $A = \begin{bmatrix} 2 & 0 \\ 0 & 1.5 \end{bmatrix}$
- The basis vectors are transformed
- Determinant is the volume of the transformation — in other words, the transformed volume of the unit hypercube (area of the quadrilateral in 2-D)



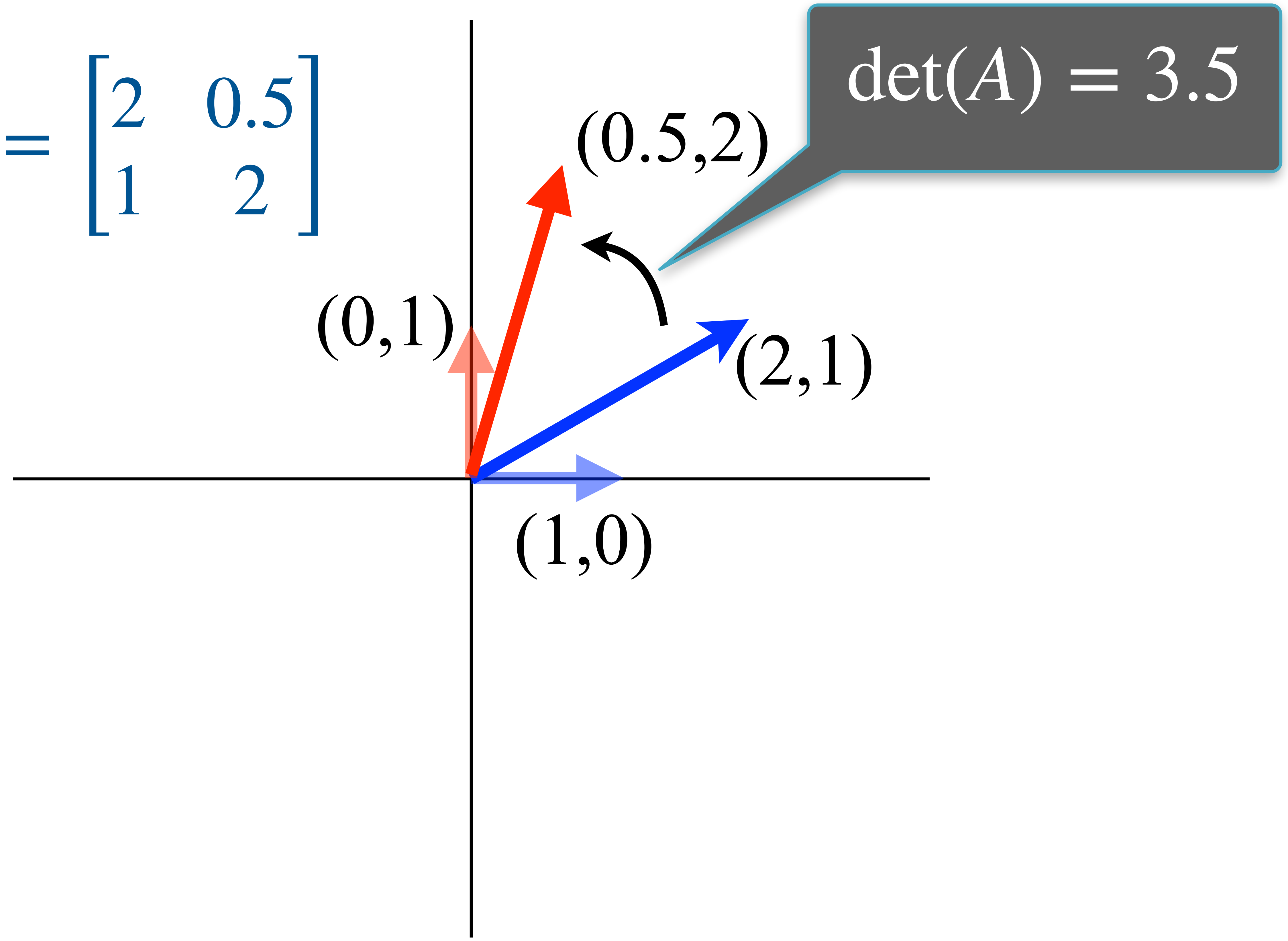
Determinants of 2×2 matrices

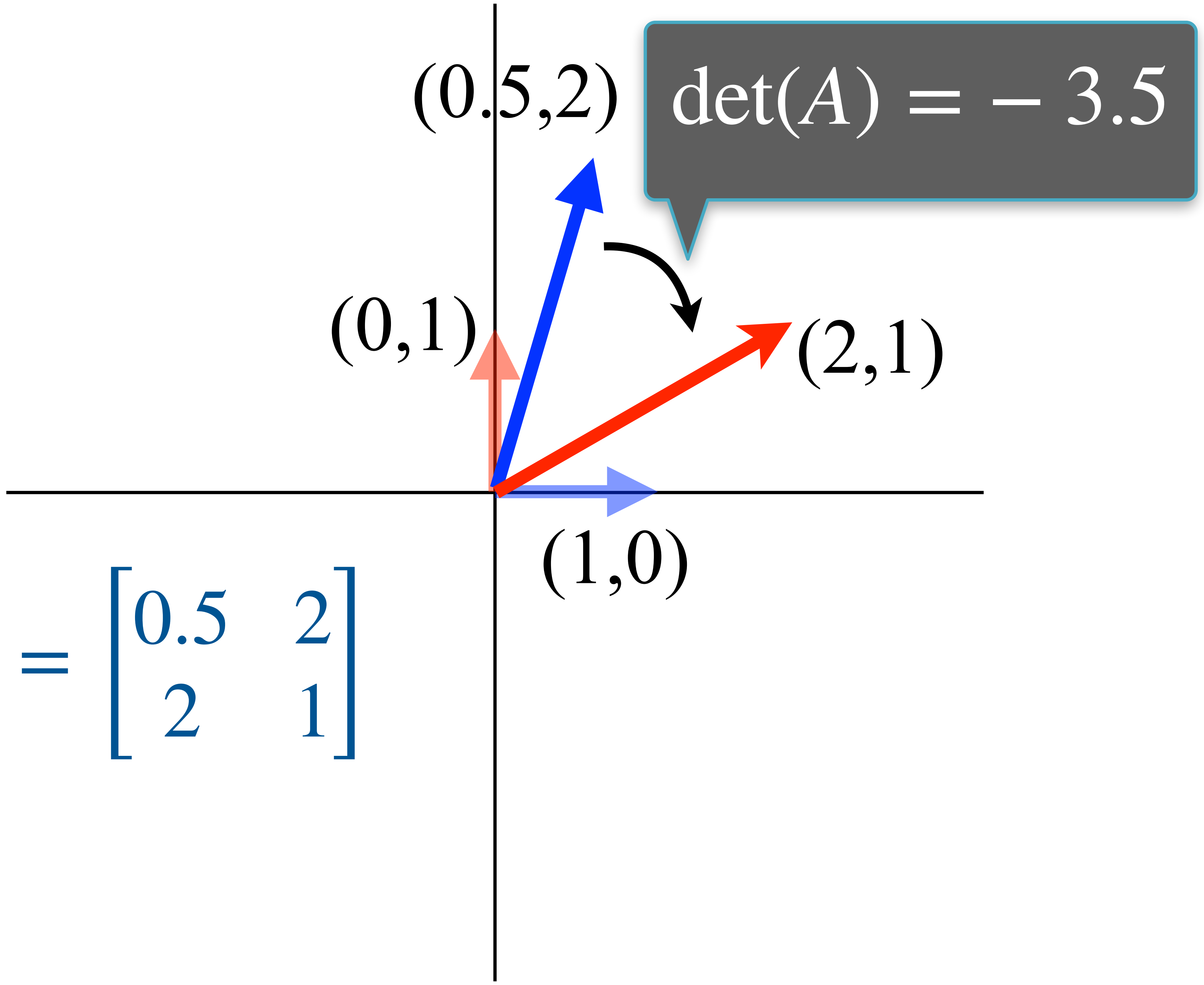
- Another example: sheer
  $A = \begin{bmatrix} 1 & 0 \\ 1.5 & 1 \end{bmatrix}$
- Area of the transformed unit square = 1 = det(*A*)
- **General formula** (2×2)**:**
  If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $\det(A) = ad - bc$



Positive and Negative Determinant

$$A = \begin{bmatrix} 2 & 0.5 \\ 1 & 2 \end{bmatrix}$$





$$A = \begin{bmatrix} 0.5 & 2 \\ 2 & 1 \end{bmatrix}$$

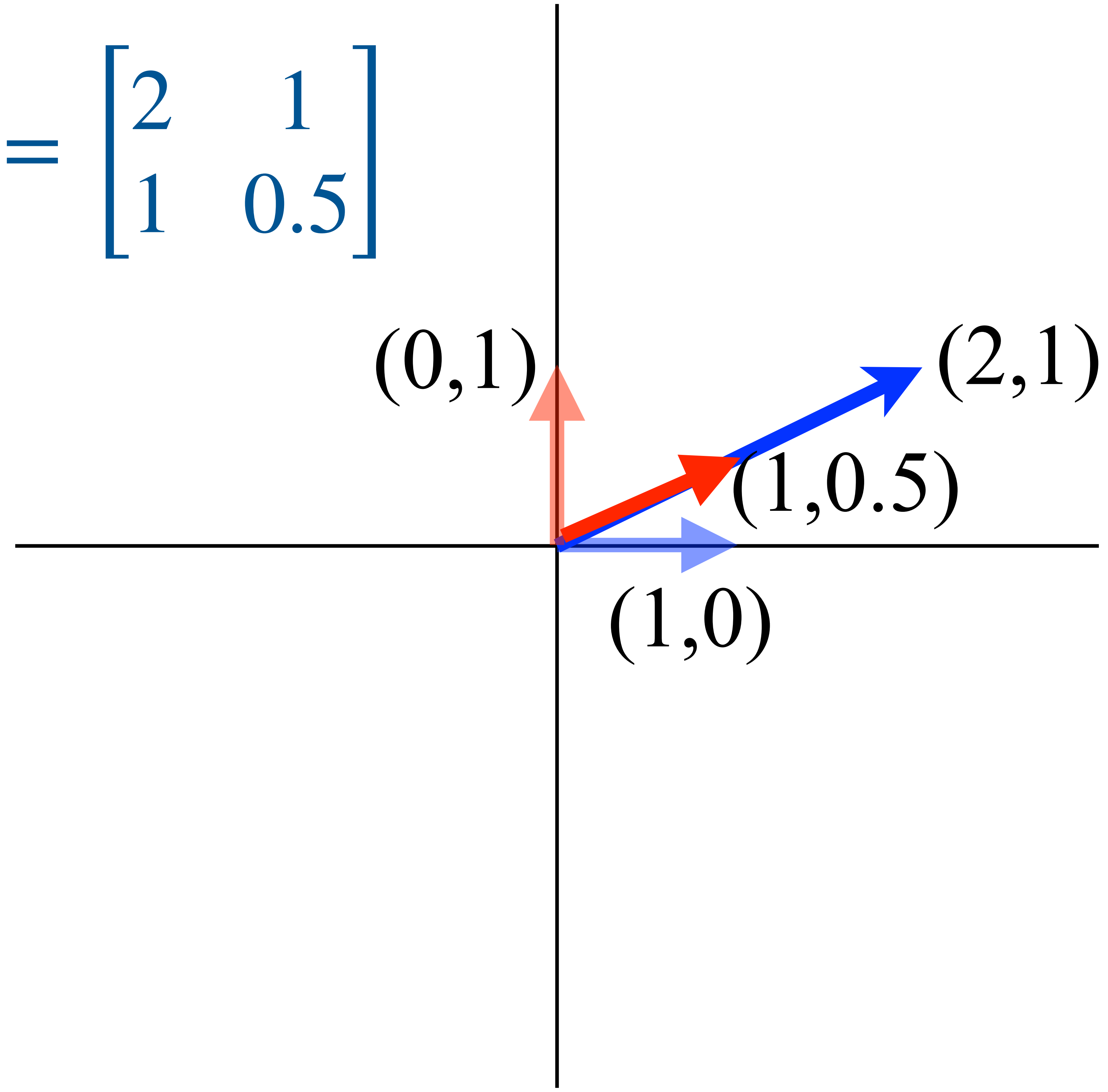
If $\det(A)$ is negative, then it means that as a transformation, A flips the orientation

Determinant and Inverse (2×2)

- The transformation A sends the unit hypercube to a hypercube of volume $\det(A)$
- The inverse transformation must send a hypercube of volume $\det(A)$ to a hypercube of volume 1
- Hence: $\det(A^{-1}) = \dfrac{1}{\det(A)}$
- If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then consider the matrix $B = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$
- We have $AB = \begin{bmatrix} ad - bc & 0 \\ 0 & ad - bc \end{bmatrix}$
- We can multiply AB by $\dfrac{1}{\det(A)}$, then we get the identity
- But we can do that only if $\det(A) = ad - bc \neq 0$

What if $\det(A) = 0$?

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 0.5 \end{bmatrix}$$



Suppose the columns of A are not linearly independent

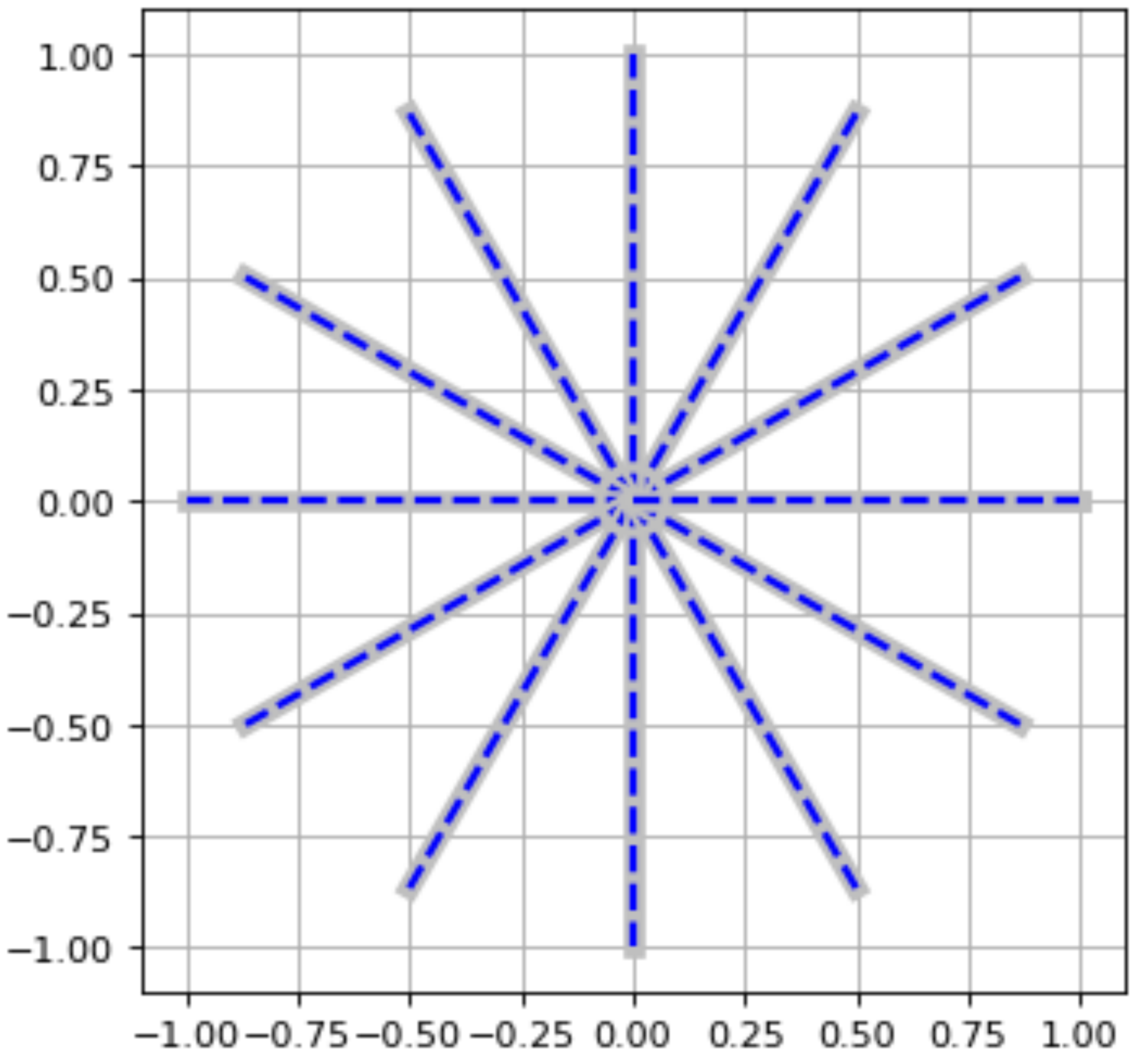
$\implies$ Both basis vectors (1,0) and (0,1) gets mapped to the same line

$\implies$ The unit square is collapsed onto the line, area becomes 0 (for higher dimension, the unit hypercube collapses onto a lower dimensional space $\sim$ volume 0)

$\implies$ There cannot be an inverse map

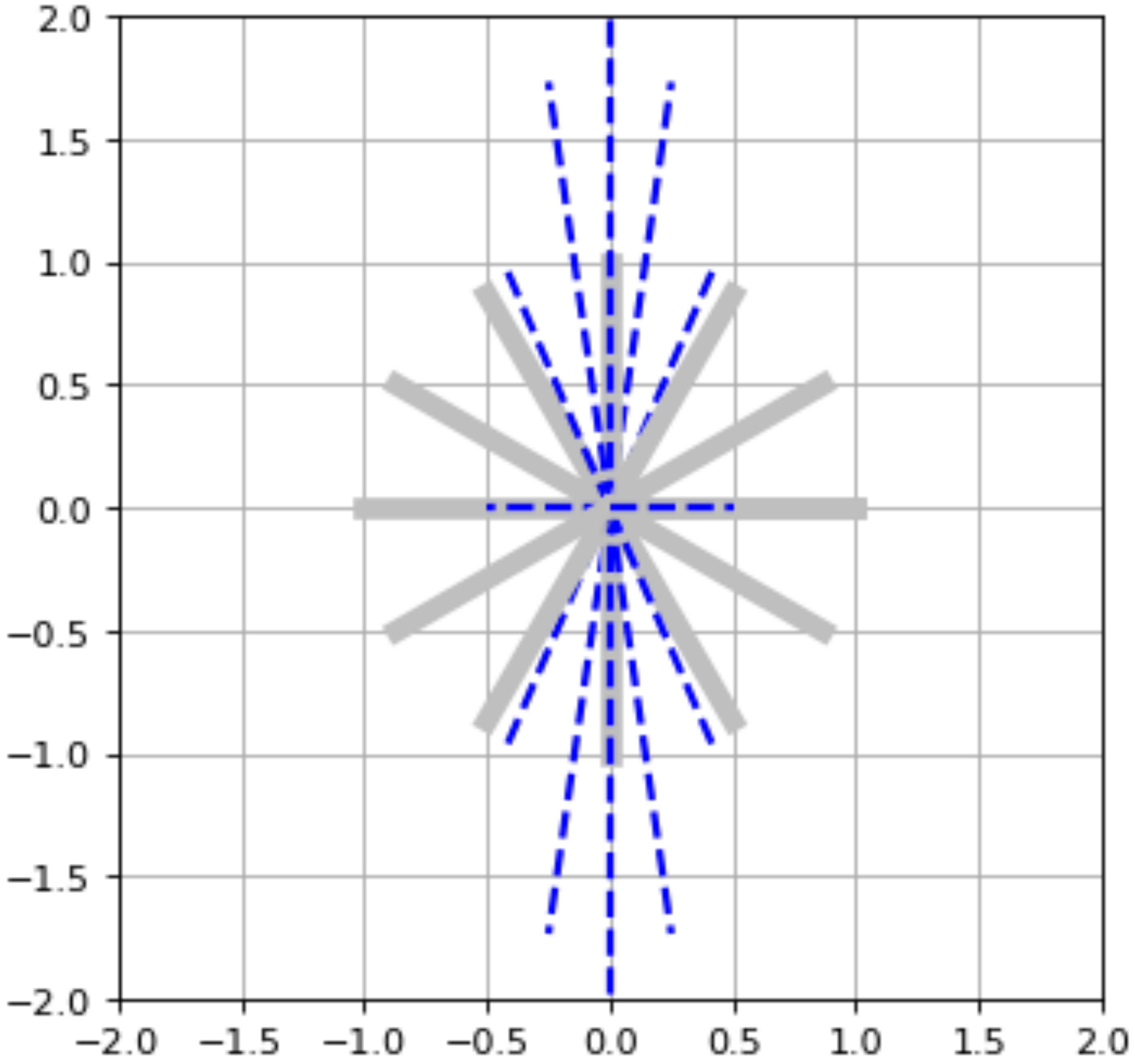
How does a matrix transform all vectors?

- Let us draw unit vectors in *all* directions
- The identity matrix performs the identity transformation
- How many vectors are *retaining* their directions?
  - All!
- The identity matrix works as scaling (with a factor 1) for all vectors



Scaling with a diagonal matrix

- Let us take a diagonal matrix that has different diagonal entries
$$A = \begin{bmatrix} 0.5 & 0 \\ 0 & 2 \end{bmatrix}$$
and observe how it transforms the unit vectors
- Which vectors are scaled along the same line?
- The vectors along (1,0) and (0,1)
- All other vectors change their directions

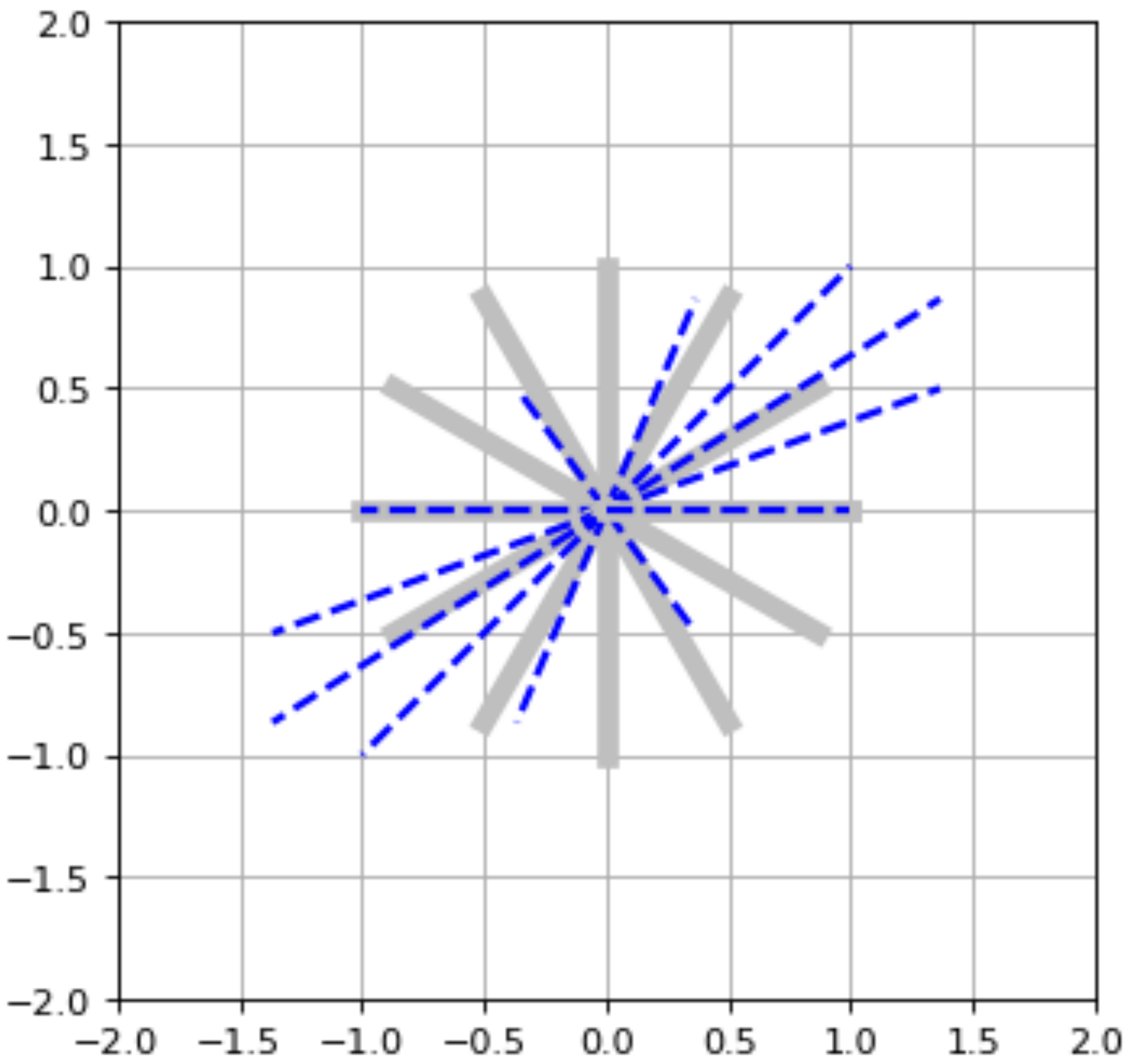


Eigenvectors and Eigenvalues

- *Definition*: An eigenvector $\mathbf{x}$ of a linear transformation T is a *non-zero* vector that changes only by a scalar factor when T is applied to it.
- In other words, $T\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ
- For a matrix A, $T\mathbf{x} = \lambda\mathbf{x}$
- *Eigen* means *characteristic* in German
- The scalar λ is called the corresponding eigenvalue

Example: sheer

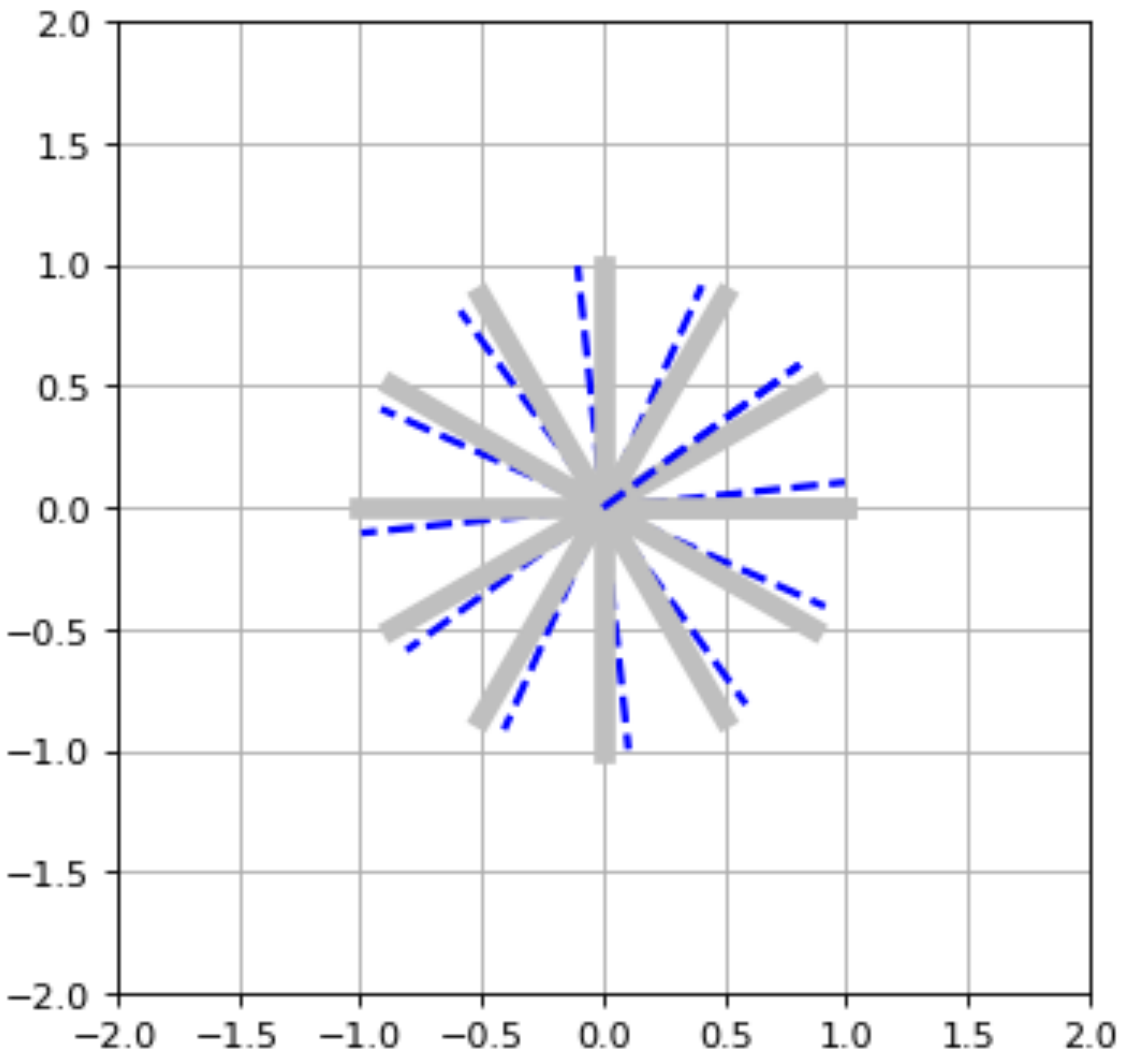
- What are the eigenvectors of the matrix $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$?
- Which vectors are scaled along the same line?
- Only vectors along (1,0)
- All other vectors change their directions



Example: rotation

- What are the eigenvectors of the matrix $A = \begin{bmatrix} \cos(\pi/5) & -\sin(\pi/5) \\ \sin(\pi/5) & \cos(\pi/5) \end{bmatrix}$?

- Which vectors are scaled along the same line?
- No vector at all



How to find eigenvectors?

- We have $A\mathbf{x} = \lambda\mathbf{x}$
- We can write: $(\lambda I - A)\mathbf{x} = 0$
- Since we are not interested in $\mathbf{x}$ being zero, we have $\det(\lambda I - A) = 0$
- This is a polynomial of degree *n* (for an $n \times n$ matrix *A*), called the *characteristic polynomial* of *A*
- The eigenvalues are the roots of this polynomial
- Fundamental theorem of algebra: the characteristic polynomial can be written as $(\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_n - \lambda)$
- The λ_is are complex numbers in general, there are *n* complex roots, may not be all real

Eigen-decomposition

- Suppose A is an $n \times n$ matrix with n linearly independent eigenvectors $\mathbf{q}_1, \ldots, \mathbf{q}_n$ and let the corresponding eigenvalues be $\lambda_1, \ldots, \lambda_n$
- Then, A can be factorized as

$$A = Q\Lambda Q^{-1}$$

where $Q = \begin{bmatrix} \vdots & \vdots & \vdots \\ \mathbf{q}_1 & \vdots & \mathbf{q}_n \\ \vdots & \vdots & \vdots \end{bmatrix}$ is the matrix with columns being the eigenvectors of A

and $\Lambda = \begin{bmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_n \end{bmatrix}$ is the diagonal matrix with diagonal entries as the eigenvalues of A

Proof

- For each eigenvector $\mathbf{q}_i$, we have $A\mathbf{q}_i = \lambda_i \mathbf{q}_i$
- Stacking all the eigenvectors into a matrix Q, we can write $AQ = Q\Lambda$
- Since Q have linearly independent vectors, Q has an inverse Q^{-1}
- So, $Q\Lambda Q^{-1} = AQQ^{-1} = A$

Intuition

$$A = Q\Lambda Q^{-1}$$

- Think of all the matrices as transformations
- Then, transformation by A is equivalent to combination of 3 transformations, one of which is diagonal
- What is Q?
- Set of linearly independent vectors
- Can be considered as a change of basis
- However, not all matrices are *diagonalizable* this way

Why is this good?

- Suppose, we need to apply some transformation to some vector repeatedly
- A vector $\mathbf{v}$ $\rightarrow A\mathbf{v} \rightarrow A(A\mathbf{v}) = A^2\mathbf{v} \rightarrow \cdots A^k\mathbf{v}$
- High complexity

- Now, consider $A = Q\Lambda Q^{-1}$
- Then, $A^2 = Q\Lambda Q^{-1}Q\Lambda Q^{-1} = A = Q\Lambda^2 Q^{-1}$
- Or, $A^k = Q\Lambda^k Q^{-1}$

$$\begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix}^k = \begin{bmatrix} \lambda_1^k & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n^k \end{bmatrix}$$

- It is easy to compute power of a diagonal matrix
- So, the transformation $A^k\mathbf{v}$ can be considered as a change of basis, then one transformation by Λ^k, and changing the basis back