ARE211, Fall 2009

LINALGEBRA3: THU, OCT 8, 2009 PRINTED: AUGUST 27, 2009 (LEC# 13)

Contents

3. Linear Algebra (cont)

3.9. **Linear Functions**

A function with domain X and range Y is a *rule* that assigns a *unique* point in the range to *every* point in the domain.

Notation: $f : X \to Y$.

The *image* of f, denoted $f(X)$, is the set of points in the range that are reached from some point in the domain, i.e., $f(X) = \{f(x) \in Y : x \in X\}$.

Note: it's not required that *every* point in the range of a function be reached from some point in the domain. This means that the image of a function is not the same as the range. Indeed, if Y is *any* superset of $f(X)$, then we can write $f : X \to Y$.

There is a lot of variation in language concerning the names that are assigned to $f(X)$ vs Y.

- Some books refer to Y as the "target space" (S&B) or the "co-domain"
- Some books even use the word "range" to refer to $f(X)$.
- I follow the language used by the classical Analysis texts (e.g., Rudin) and will use the above terminology consistently.

The *graph* of $f : X \to Y$ is defined as:

$$\text{graph} f = \{(\mathbf{x}, y) \in X \times Y : \quad y = f(\mathbf{x})\} \tag{1}$$

The symbol $\times$ indicates the *Cartesian product* of the two sets. The result is a set of vectors made by pairing elements of the first set and elements of the second. Formally:

$$X \times Y = \{(\mathbf{x}, \mathbf{y}) : \mathbf{x} \in X, \mathbf{y} \in Y\} \tag{2}$$

A <u>linear function</u> is a function that satisfies additivity and proportionality, that is, $f : X \to Y$ is a linear function if for all $\mathbf{x}, \mathbf{y} \in X$ and all $\alpha \in \mathbb{R}$,

- $f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$ Additivity
- $f(\alpha\mathbf{x}) = \alpha f(\mathbf{x})$ Proportionality

Examples:

- any function whose graph is a straight line? Ans: no (In fact, these are properly called <u>af</u>fine functions).
- $f(x) = 1 + x$ where x is a scalar. Ans: no.
- $f(x) = ax$ where a and x are scalars. Ans: yes.
- $f(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x}$ where $\mathbf{x}$ and $\mathbf{a}$ are vectors. Ans: yes.

FACTS:

- *Any* linear function from $\mathbb{R}^1$ to $\mathbb{R}^1$ can be written in the form $f(x) = ax$, for some scalar a.
- *Any* linear function from $\mathbb{R}^n$ to $\mathbb{R}^1$ can be written in the form $f(\mathbf{x}) = \mathbf{a}\cdot\mathbf{x}$, for some n-vector a.
- *Any* linear function from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be written in the form $f(\mathbf{x}) = A\mathbf{x}$, where A is a matrix with m rows and n columns.

Note that the definition of a linear function *implies* that both the domain X and the image $f(X)$ of f are vector spaces. To see this, note that

(1) if f is defined at $x^1, x^2 \in X$, then by applying the properties of linearity, f is also defined at $\alpha x^1 + \beta x^2$, and hence $\alpha x^1 + \beta x^2 \in X$.
(2) similarly, suppose that for $i = 1, 2$, $y^i \in f(X)$, i.e., for some x^i, $y^i = f(x^i)$. In this case, for any $\alpha, \beta \in \mathbb{R}$, $\alpha y^1 + \beta y^2 = f(\alpha x^1 + \beta x^2)$, and hence $\alpha y^1 + \beta y^2 \in f(X)$.

So we've established that a necessary condition for f to be linear is that its domain and image are both vector spaces.

We can go a bit further than this.

Fact A function f is linear if and only if its graph is a vector space.
It's a good exercise to try to prove this.

It is *not* necessarily the case that the *range* of a linear function is a vector space. Recall that the range of a function is a very sloppy term. In particular, the range is not uniquely defined: it can be *any* set that contains the image of the function. Here are two examples.

(1) consider the perfectly good linear function, $f : \mathbb{R} \to Y$, defined by $f(x) = 0$ for all $x \in \mathbb{R}$. $f(\mathbb{R})$ is the zero-dimensional vector space $\{0\}$. The range can be *any* subset of $\mathbb{R}$ that contains zero.

(2) now consider the linear function $y = Ax$, where $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. This matrix maps $\mathbb{R}^2$ to a one-dimensional subspace of $\mathbb{R}^2$, specifically the 45° line. The "natural" range of this function is $\mathbb{R}^2$ but it could equally well be *any* subset of $\mathbb{R}^2$ that contains the 45° line.

3.10. **The "graph" of a linear function from $\mathbb{R}^2$ to $\mathbb{R}^2$**

If we have a linear function from $\mathbb{R}^2$ to $\mathbb{R}$, $y = \mathbf{a} \cdot \mathbf{x}$, we can get a good intuitive sense of the properties of this function by looking at its graph, i.e., it is a plane in $\mathbb{R}^3$. Similarly, we'd like to get a look at the graph of the linear function from $\mathbb{R}^2$ to $\mathbb{R}^2$, $\mathbf{y} = A\mathbf{x}$, but this graph is difficult to envisage because it is in $\mathbb{R}^4$.

We can, however, get a pretty good sense of what the graph would looks like by asking the question *what does the matrix A "do" to the unit circle?* (For a function mapping $\mathbb{R}^3$ to $\mathbb{R}^3$, we would look at what the function would do to the unit *sphere*. This way we can simulate looking into $\mathbb{R}^6$.). Technically, we will be investigating *the image of the unit circle under A*. Once we know what happens to the unit circle, we will have a good sense of what the entire graph of the function $A\mathbf{x}$ would look like. We will also know a great deal about:

- determinant of A
- the rank of A
- the eigenvectors of A
- the eigenvalues of A
- linear difference equations of the form $\mathbf{x}^t = A\mathbf{x}^{t-1}$
- and much, much more.

For the purposes of this lecture we are going to stick to *symmetric* matrices. Reason is that later on we are going to be talking about eigenvalues, eigenvectors and definiteness and for symmetric matrices, these relationships are very clearcut. The waters are substantially muddied with non-symmetric matrices. For our purposes, we are only interested definiteness of matrices, when the matrices are Hessians, and these are always symmetric, i.e., the cross-partials are identical. So the restriction is harmless.

The *the image of the unit circle under* A is defined as the following set:

$$\{\mathbf{b} : \mathbf{b} = A\mathbf{x}, \text{ for some } \mathbf{x} \text{ whose norm is unity}\}$$

Example: Let $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.

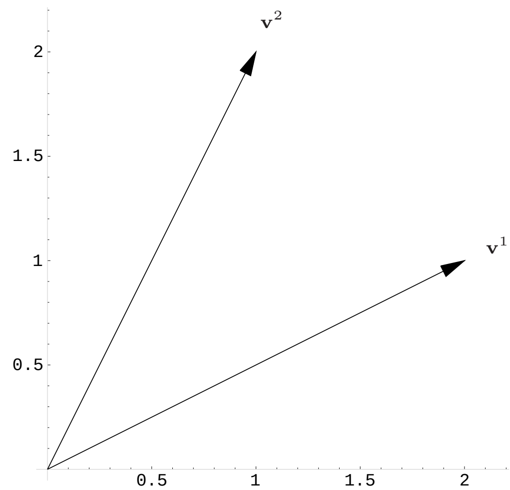
Look at what this matrix "does" to selected elements of the unit circle:

- set $\mathbf{x}^1 = (1, 0)$: A maps this vector to the first column of A
- set $\mathbf{x}^2 = (0, 1)$: A maps this vector to the second column of A
- set $\mathbf{x}^3 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$: A maps this vector to a scalar multiple of itself.
- set $\mathbf{x}^4 = \left(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ A maps this vector to a scalar multiple of itself (what is the scalar, in this case?).

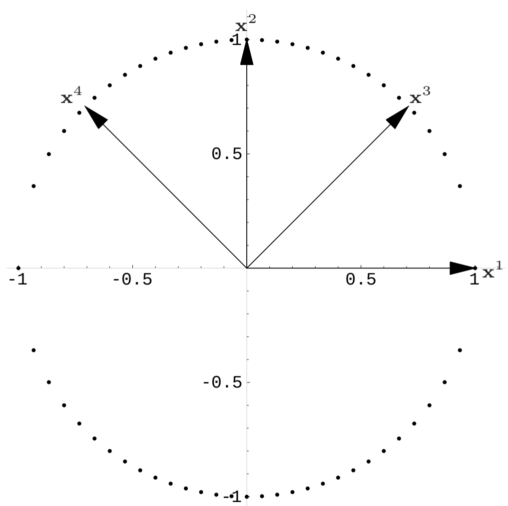
Observe the ellipse. What can we learn about A from the picture of

$$\{\mathbf{b} : \mathbf{b} = A\mathbf{x}, \text{ for some } \mathbf{x} \text{ whose norm is unity}\}$$

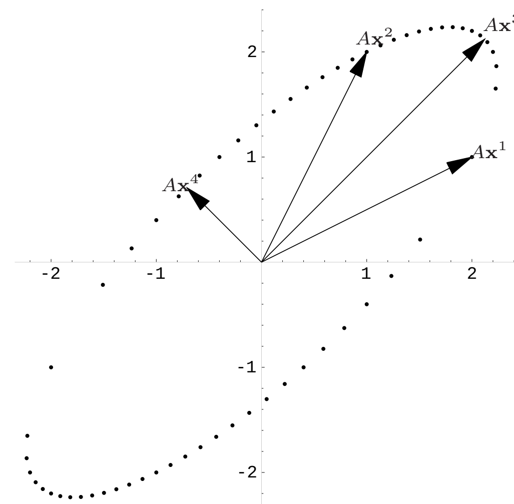
- the *eigenvectors* of A are the vectors that A i.e., *maps in the same direction (or reverse).* In this case the two vectors are $\mathbf{x}^3 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ and $\mathbf{x}^4 = \left(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$.
  - Note that there are a lot of vectors with this property.
  - Important fact: Symmetric matrices have always a set of eigenvectors that are pairwise orthogonal.
  - Note that the four unit eigenvectors split up the unit circle into 4 equal segments.



(a) The columns of A



(b) Selected elements of the unit circle



(c) The image of the circle under A

FIGURE 1. What the matrix A does to the unit circle

- Observe how any arrow in the unit circle gets swivelled by no more than 90 degrees, i.e., for *any* vector $\mathbf{x}$, the inner product of $A\mathbf{x}$ and $\mathbf{x}$ is positive. Called a *positive* definite matrix.
- To see why this must be, and what the relationship is between positive definiteness and the eigenvalues, note what A does to any vector $\mathbf{b}$ that lies in the *nonnegative cone* defined by our two eigenvectors, $\mathbf{x}^3$ and $\mathbf{x}^4$:
  * $A\mathbf{b}$ has to lie in the nonnegative cone defined by the vectors $A\mathbf{x}^3$ and $A\mathbf{x}^4$:

  - ∗ Why? Because $A\mathbf{x}$ is a linear function, i.e., if $\mathbf{b} = \alpha\mathbf{x}^3 + \beta\mathbf{x}^4$, then $A\mathbf{b} = A(\alpha\mathbf{x}^3 + \beta\mathbf{x}^4) = \alpha A\mathbf{x}^3 + \beta A\mathbf{x}^4$, i.e., $A\mathbf{b}$ is necessarily a nonnegative linear combination of $A\mathbf{x}^3$ and $A\mathbf{x}^4$, i.e., in the nonnegative cone that these vectors define.
  - ∗ but this cone is *the same* cone as the one defined by $\mathbf{x}^3$ and $\mathbf{x}^4$!
  - ∗ so, we've established that $\mathbf{b}$ and $A\mathbf{b}$ live in the same cone which has an arc of exactly 90 degrees.
  - ∗ Example in Fig. 1: look at what happens to the vector $\mathbf{x}^2$.

- The *eigenvalues* of A: each eigenvector has a corresponding eigenvalue: this value is a scalar that measures the size and sign of the magnification of the eigenvectors. That is, for any eigenvector of A, $A\mathbf{x}$ and $\mathbf{x}$ are collinear, i.e., $A\mathbf{x}$ is a scalar multiple of $\mathbf{x}$: the eigenvalue tells you by how much the vector is stretched or shrunk.

Defn: A vector $\mathbf{x}$ is an *eigenvector* of a matrix A if there exists a scalar $\lambda \in \mathbb{R}$ such that $A\mathbf{x} = \lambda\mathbf{x}$. In this case, λ is referred to as the *eigenvalue corresponding to* $\mathbf{x}$. (The word "*eigen*" in German means "belonging to," which is appropriate since a (nonzero) vector v is an eigenvector of A if the image of that vector under A, i.e., Av *belongs to* (i.e., is collinear with) the single-dimensional vector space spanned by v.)

Defn: An $n \times n$ matrix A is *positive definite* (*positive semidefinite*) if for every $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \neq 0$ implies $\mathbf{x}'A\mathbf{x} > (\geq)0$.

Fact: Every symmetric $n \times n$ matrix A has *at least* n, pairwise orthogonal eigenvectors.

That is, any two of the n vectors identified above make a right angle with each other.

Actually, every symmetric $n \times n$ matrix A has many more than n eigenvectors. For this reason we typically focus on *unit* eigenvectors, i.e., vectors with unit norm. Still we have more than n: in fact, every symmetric $n \times n$ matrix A has *at least* $2n$ eigenvectors, the n identified above plus the negatives of these. The point is that we *only need* n pair-wise orthogonal ones to "build" the image of the unit circle/sphere/hyper-sphere. We don't care about the rest. Hence the wording of the above fact.

Fact: A *symmetric* matrix A is positive definite (positive semidefinite) if and only if all of its eigenvalues are positive (nonnegative).

Now consider the negative of matrix A, which has both arrows pointing into the negative orthant. Let $B = \begin{bmatrix} -2 & -1 \\ -1 & -2 \end{bmatrix}$.

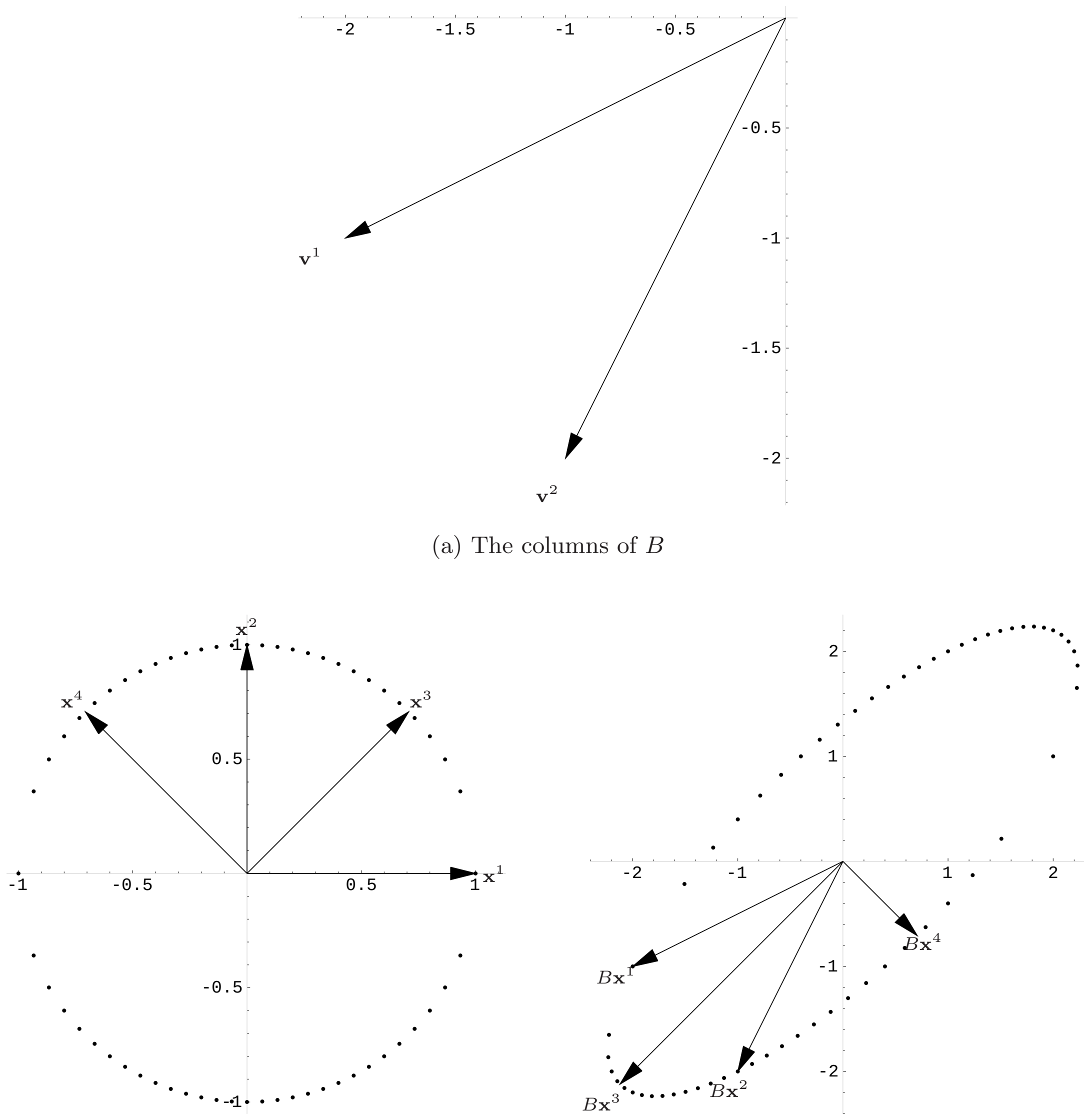
- Note that the two eigenvectors get flipped by 180 degrees. Look at the cone defined by the two eigenvectors $\mathbf{x}^3$ and $\mathbf{x}^4$: the whole cone gets flipped over.
- Conclude that every vector gets swivelled by more than 90 degrees, i.e., for *every* vector $\mathbf{b}$, the inner product of $B\mathbf{b}$ and $\mathbf{b}$ is negative. Called a *negative* definite matrix.
- Example in Fig. 1: look at what happens to the vector $\mathbf{x}^2$.

Defn: An $n \times n$ matrix A is *negative definite* (*negative semidefinite*) if for every $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \neq 0$ implies $\mathbf{x}A\mathbf{x} < (\leq)0$.

Fact: A *symmetric* matrix A is negative definite (negative semidefinite) if and only if all of its eigenvalues are negative (nonpositive).

Finally, we construct an *indefinite matrix* C with the property that for some vectors $\mathbf{x}$, $\mathbf{x}$ and $C\mathbf{x}$ make an acute angle with each other, and for for some other vectors $\mathbf{y}$, $\mathbf{y}$ and $C\mathbf{y}$ make an obtuse angle with each other.

Let $C = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. Note that C is obtained by flipping the order of the column vectors in A. Look at what happens to the image of the unit circle under C:



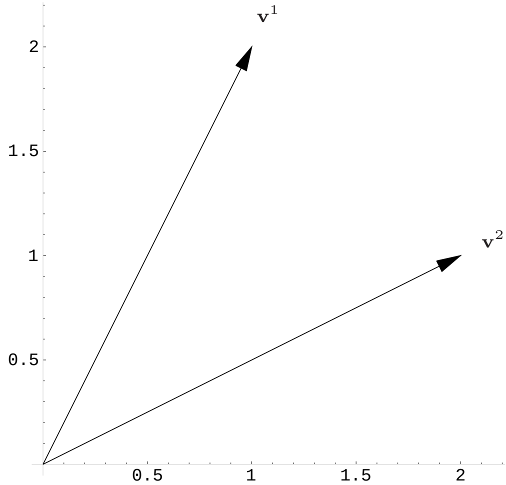
(a) The columns of B

(b) Selected elements of the unit circle

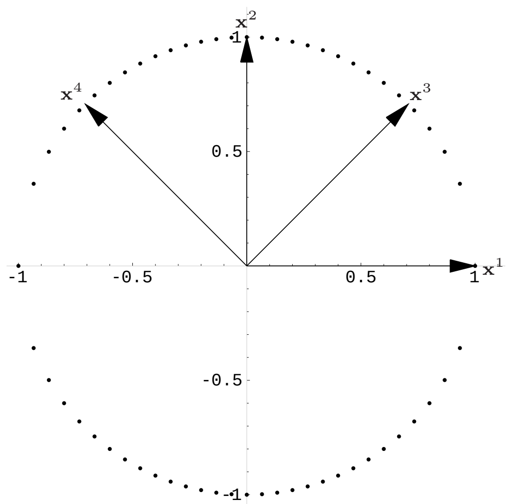
(c) The image of the circle under B

FIGURE 2. What the matrix B does to the unit circle

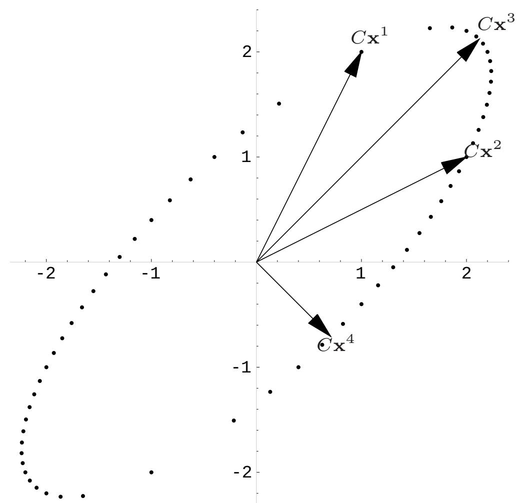
- The ellipse for C looks exactly the same as for matrix A: but in fact it is a mirror image of what happens to A: a vector that would have gotten mapped to one side of the long axis now gets mapped to the other side.
- Which vectors are going to make an obtuse angle with their images under C and which ones make an acute angle? ones that are close to the eigenvector with a negative eigenvalue will end up making an obtuse angle, etc.



(a) The columns of C



(b) Selected elements of the unit circle



(c) The image of the circle under C

FIGURE 3. What the matrix C does to the unit circle

Defn: An $n \times n$ matrix A is *indefinite* if there exists $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ such that the product of $\mathbf{x}'A\mathbf{x}$ is positive and $\mathbf{y}'A\mathbf{y}$ is negative (that is, they are neither both positive nor both negative).

Notice a striking difference between the figures for matrix A and for C. The matrices map the unit circle into identical ellipses, but the "orientation" of the ellipses is different. Matrix A preserves orientation in the sense that for any two unit vectors v and w, if vector w in the unit circle is reached by moving counter-clockwise from v, then Aw in the ellipse is reached by moving counter-clockwise from Av. On the other hand, C reverses orientation in the sense that Cw is reached by moving *clockwise* from Cv. As we shall see in the next lecture, whether a matrix preserves orientation or reverses it is reflected in the signs of the determinants of the two matrices, i.e., the determinants of A and C are equal in absolute magnitude, but A's has a positive sign, while C's is negative.

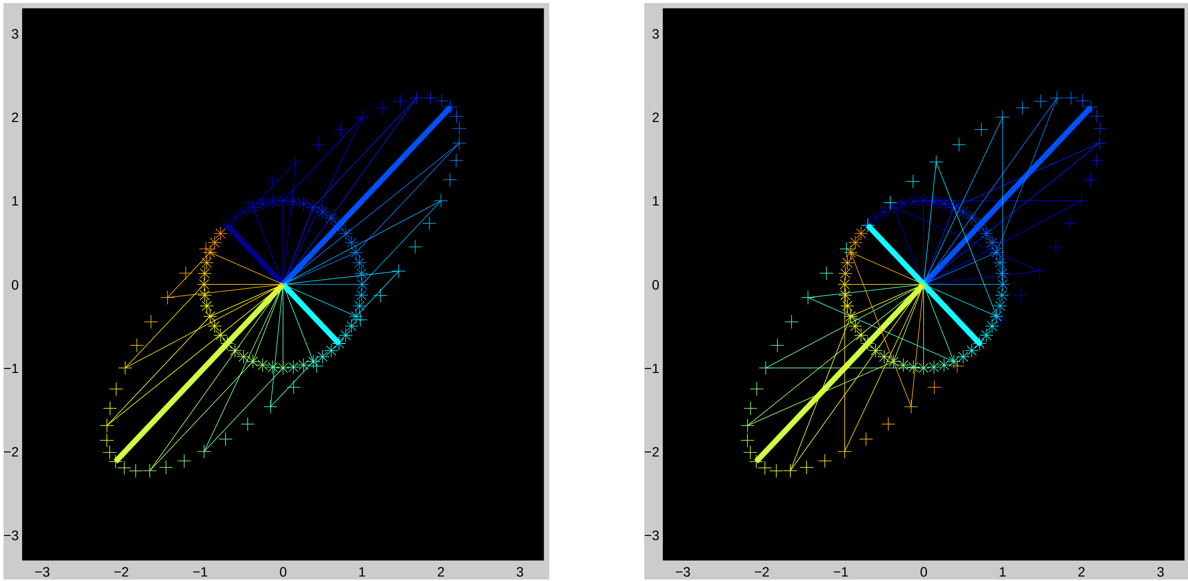


FIGURE 4. Relationship between orientation and determinant