ARE201-Simon, Fall2015

Contents

1. Linear Algebra (cont)

1.1. Linear Functions

A function with domain X and codomain Y is a *rule* that assigns a *unique* point in the codomain 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 codomain 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 codomain 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 codomain. 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 "codomain" to refer to $f(X)$.

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

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

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 linear function 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

We can combine these two properties into one and obtain
Theorem: $f : X \to Y$ is a linear function iff $\forall \mathbf{x}, \mathbf{y} \in X$, $\forall \alpha, \beta \in \mathbb{R}$, $f(\alpha \mathbf{x} + \beta \mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y})$

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

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 *codomain* of a linear function is a vector space. Recall that the codomain of a function is a very sloppy term. In particular, the codomain 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 codomain 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" codomain of this function is $\mathbb{R}^2$ but it could equally well be *any* subset of $\mathbb{R}^2$ that contains the 45° line.

An affine function has all the properties of a linear function—straight line graph, goes on forever—*except* that its graph needn't pass through the origin.

Definition: $f : X \to Y$ is *affine* if there exists a linear function $g : X \to Y$ and $\mathbf{a} \in X$ s.t. $f(\cdot) = \mathbf{a} + g(\cdot)$.

Theorem: (econ201) $f : X \to Y$ is affine iff $\forall \mathbf{x}, \mathbf{y} \in X$, $\forall \alpha \in \mathbb{R}$, $f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y}) = \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y})$

Note the difference between this and the corresponding statement for a *linear* function

- for linear functions, the condition is in terms of *linear* combinations
- for affine functions, the condition is in terms of *convex* combinations

1.2. Baby Riesz

Some linear functions:

- any function whose graph is a straight line? Ans: no (In fact, these are properly called affine 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.

Theorem: (Riesz Representation (baby version))

- *Any* linear function from $\mathbb{R}^1$ to $\mathbb{R}^1$ can be written in the form $f(x) = ax$, for some $a \in \mathbb{R}$.
- *Any* linear function from $\mathbb{R}^n$ to $\mathbb{R}^1$ can be written as $f(\mathbf{x})=\mathbf{a} \cdot \mathbf{x}$, for some $\mathbf{a} \in \mathbb{R}^n$
- *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.

It's obvious that any $a \in \mathbb{R}$ defines a linear function $f(x) = ax$. What's definitely not obvious is that given any function $f : \begin{cases} \mathbb{R} \to \mathbb{R} \\ \mathbb{R}^n \to \mathbb{R} \\ \mathbb{R}^n \to \mathbb{R}^m \end{cases}$ satisfying additivity and proportionality, there exists $\begin{cases} a \in \mathbb{R} \\ \mathbf{a} \in \mathbb{R}^n \\ A \in \mathbb{R}^{m \times n} \end{cases}$ such that for all $\begin{cases} x \in \mathbb{R},\ f(x) = ax \\ \mathbf{x} \in \mathbb{R}^n,\ f(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x} \\ \mathbf{x} \in \mathbb{R}^n,\ f(\mathbf{x}) = A \cdot \mathbf{x} \end{cases}$. It's this reverse implication that is the real content of Riesz's representation theorem; the Riesz result has enormous implications for everything we do as economists. In particular

- Theorem: (Taylor-Young) if $f : \mathbb{R}^n \to \mathbb{R}$, is twice continuously differentiable, then for each $\mathbf{x} \in \mathbb{R}^n$ there exists *a unique linear* function $g^{\mathbf{x}} : \mathbb{R}^n \to \mathbb{R}$ such that if $\mathbf{dx}$ is sufficiently small and $g^{\mathbf{x}}(\mathbf{dx}) \neq 0$, then $\text{sgn}\big(f(\mathbf{x} + \mathbf{dx}) - f(\mathbf{x})\big) = \text{sgn}\big(g^{\mathbf{x}}(\mathbf{dx})\big)$.
- Useful thing to know (indeed fantastically helpful, the basis of *all* comparative statics), *especially* if we can figure out what function $g^{\mathbf{x}}(\cdot)$ is
- Reisz tells us that for any linear function g from $\mathbb{R}^n$ to $\mathbb{R}$, there exists a unique vector $\mathbf{a}$ such that $g(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x}$. So to nail down the function that Taylor-Young is talking about, all we have to do is find $\mathbf{a}$.
- It turns out that $\mathbf{a} = \bigtriangledown f(\mathbf{x})$, the vector of partial derivatives of f, evaluated at $\mathbf{x}$.
- That is: $g^{\mathbf{x}}(\cdot)$ is the function defined by, for $\mathbf{dx} \in \mathbb{R}^n$, $g^{\mathbf{x}}(\mathbf{dx}) = \bigtriangledown f(\mathbf{x}) \cdot \mathbf{dx}$.
- The above result is useful, but limitedly so: what can we say about $\text{sgn}\big(f(\mathbf{x} + \mathbf{dx}) - f(\mathbf{x})\big)$ for the case in which our unique linear function $g^{\mathbf{x}}(\mathbf{dx})$ is identically zero? This situation crops up occasionally, since $g^{\mathbf{x}}(\mathbf{dx}) = \bigtriangledown f(\mathbf{x}) = 0$ happens to be a necessary conditions for an unconstrained optimum, and we as economists are often interested in such things.
  - In this case, we turn to a different, also uniquely defined linear function $G^{\mathbf{x}} : \mathbb{R}^n \to \mathbb{R}^n$.
  - Theorem: (Taylor-Young (again)) if $f : \mathbb{R}^n \to \mathbb{R}$, is thrice continuously differentiable and $\bigtriangledown f(\mathbf{x}) = 0$, there exists *a unique linear* function $G^{\mathbf{x}} : \mathbb{R}^n \to \mathbb{R}^n$ such that if $\mathbf{dx}$ is sufficiently small and $G^{\mathbf{x}}(\mathbf{dx}) \neq 0$, then $\text{sgn}\big(f(\mathbf{x} + \mathbf{dx}) - f(\mathbf{x})\big) = \text{sgn}\big(\mathbf{dx} \cdot G^{\mathbf{x}}(\mathbf{dx})\big)$.
  - This is also useful for us, because it enables us to distinguish between local maxes, local mins, and local nothings. Provided we can figure out what $G^{\mathbf{x}}$ is.

  - Reisz tells us that for any linear function G from $\mathbb{R}^n$ to $\mathbb{R}^n$, there exists a unique $n \times n$ matrix A such that $G(\mathbf{x}) = A \cdot \mathbf{x}$. So to nail down the function that Taylor-Young is talking about, all we have to do is find A.
  - It turns out that $G^{\mathbf{x}}(\mathbf{dx}) = \mathrm{H}\mathbf{f}(\mathbf{x}) \cdot \mathbf{dx}$, where $\mathrm{H}\mathbf{f}(\mathbf{x})$ is the matrix of second partials of f, evaluated at $\mathbf{x}$.

1.3. 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, which is a plane in $\mathbb{R}^3$. Similarly, it would be nice if we could 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 $\mathbb{R}^4$.

It turns out, however, that there *is* a way to visualize the graph of a linear function from $\mathbb{R}^n$ to $\mathbb{R}^m$, provided both m and n are not greater than 3, i.e., the graph of the function $f(\mathbf{x}) = A\mathbf{x}$, where A is $m \times n$. We do this by asking the question "*what does the matrix A "do" to the unit circle (sphere)?*" Formally, we will be investigating *the image of the unit circle under A*. Once we know what this image looks like, we'll know everything there is to know about the entire graph of the function $A\mathbf{x}$. Note that this only works for *linear* functions.

- consider $f : \mathbb{R}^n \to \mathbb{R}^m = A\mathbf{x}$ and let C denote the unit sphere in $\mathbb{R}^n$, i.e., $\{\mathbf{x} \in \mathbb{R}^n : \sum_{k=1}^n x_i = 1\}$.
- for any $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \in \alpha C$, where $\alpha = \sum_{k=1}^n x_i$.
- by linearity, $f(\alpha C) = \alpha f(C)$.
- if we know what $f(C)$ looks like, we know what $\alpha f(C)$ looks like.

In particular, studying the image of the unit circle under A tells us *almost* all there is to know about

- the determinant of A
- the rank of A
- the eigenvectors of A
- the eigenvalues of A
- definiteness of A (positive vs negative definiteness; indefiniteness)
- linear difference equations of the form $\mathbf{x}^t = A\mathbf{x}^{t-1}$
- and much, much more.

We'll focus particularly on the concept of *definiteness* of a matrix, which you will recall is the cornerstone of second order conditions.

(1) recall that if your first order conditions for a maximum are satisfied, it doesn't mean much, i.e., you could have a max, min, or neither
(2) specifically, for $\mathbf{x}, \mathbf{dx} \in \mathbb{R}^n$,

$$f(\bar{\mathbf{x}} + \mathbf{dx}) - f(\bar{\mathbf{x}}) \quad \approx \quad \bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx} + \frac{1}{2}\mathbf{dx}'\mathrm{H}\mathbf{f}(\bar{\mathbf{x}})\mathbf{dx} \tag{3}$$

(3) if $\mathbf{dx}$ satisfies the FOC for an unconstrained maximum then $\bigtriangledown f(\bar{\mathbf{x}}) = 0$.
(4) whether $\bar{\mathbf{x}}$ is a local max, local min or neither depends, then, on the sign of $\mathbf{dx}'\mathrm{H}\mathbf{f}(\bar{\mathbf{x}})\mathbf{dx}$
(5) it will be a local
   (a) *max* if $\mathbf{dx}'\mathrm{H}\mathbf{f}(\bar{\mathbf{x}})\mathbf{dx} < 0$, for all $\mathbf{dx}$, i.e., if $\mathrm{H}\mathbf{f}(\bar{\mathbf{x}})$ is *a negative definite* matrix.
   (b) *min* if $\mathbf{dx}'\mathrm{H}\mathbf{f}(\bar{\mathbf{x}})\mathbf{dx} > 0$, for all $\mathbf{dx}$, i.e., if $\mathrm{H}\mathbf{f}(\bar{\mathbf{x}})$ is *a positive definite* matrix.

(c) *neither* if there exists $\mathbf{dy}, \mathbf{dz} \in \mathbb{R}^n$ such that $\mathbf{dy}'\mathrm{H}\mathbf{f}(\bar{\mathbf{x}})\mathbf{dy} > 0 > \mathbf{dz}'\mathrm{H}\mathbf{f}(\bar{\mathbf{x}})\mathbf{dz}$ i.e., if $\mathrm{H}\mathbf{f}(\bar{\mathbf{x}})$ is *an indefinite* matrix.

(6) in the following we're going to get some graphical intuition for what these definite concepts really mean, and their relationship to the *eigen-vectors* and *eigen-values* of the matrix H.

For now, we are going to focus on the image of the circle under *symmetric* matrices. Reason is that later on we are going to be talking about eigenvalues, eigenvectors and definiteness. For symmetric matrices, these relationships are very clearcut. The waters are substantially muddied with nonsymmetric matrices. For our purposes, we are only interested the definiteness of matrices that are the Hessians of some twice differentiable function; these are always symmetric, i.e., the cross-partials are identical. So the restriction is harmless for current purposes. (For the study of difference equations, it's important to look at non-symmetric matrices.)

The *the image of the unit circle under* $A \subset \mathbb{R}^{2\times 2}$ is defined as the following set:

$$\{\mathbf{b} \in \mathbb{R}^2 : \mathbf{b} = A\mathbf{x}, \text{ for some } \mathbf{x} \text{ whose norm is unity}\}$$

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

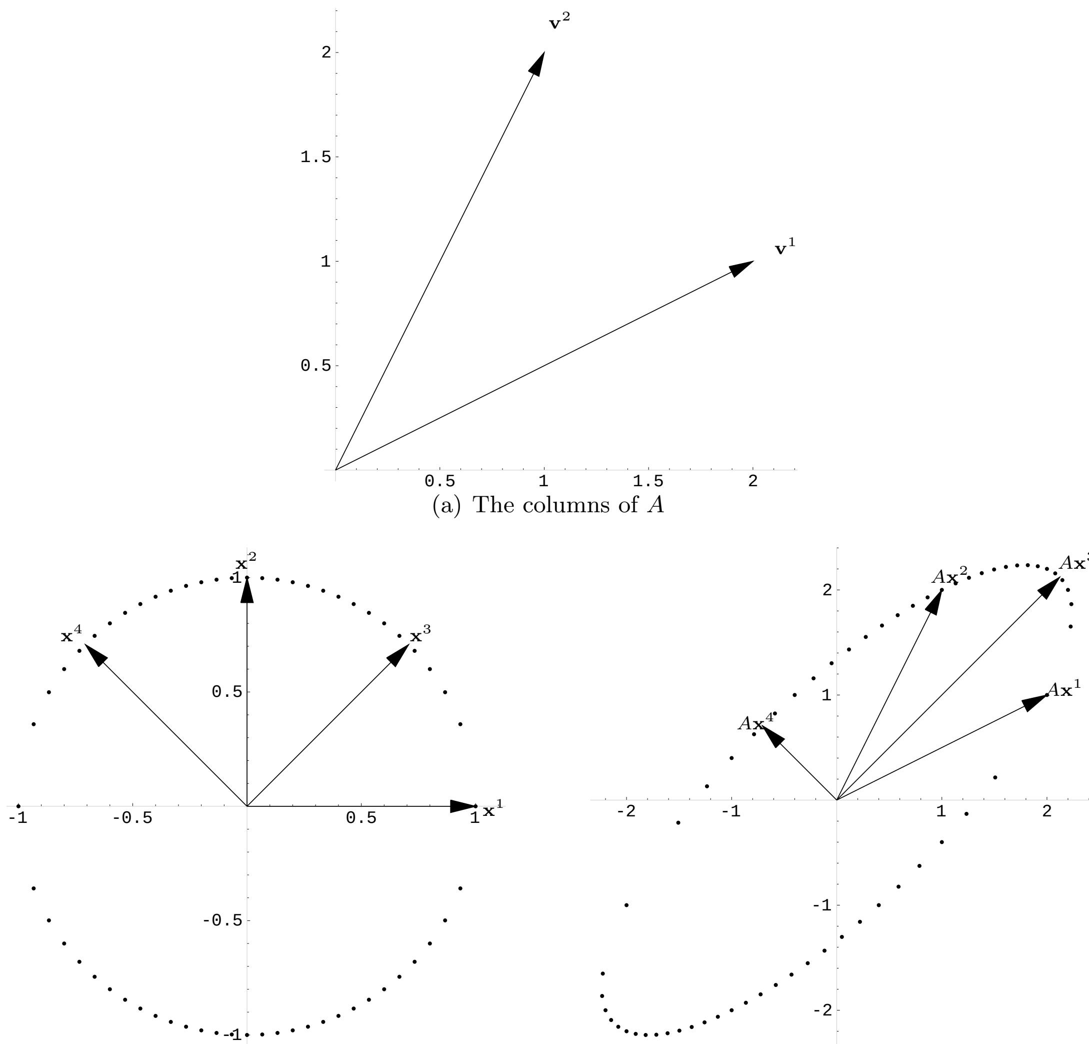
Look at what this matrix "does" to selected elements of the unit circle:

- 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

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

- a vector is said to be an *eigenvector* of a matrix M if the vector and its image under M are *collinear*, i.e., $\mathbf{v}$ is an eigenvector of M if $\mathbf{v}$ and $M\mathbf{v}$ point either in the same direction or opposite directions. In the case of our matrix A, two eigenvectors 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.
  - The four unit eigenvectors—$\pm\mathbf{x}^1$ and $\pm\mathbf{x}^2$—split up the unit circle into 4 equal segments, each having an arc of 90 degrees, in the following sense:
    * Given any vector v, once we know what a symmetric matrix M does to the two eigenvectors on either side of it, we have a lot of information about what M's going to do to v.
    * For example, for $\alpha, \beta \geq 0$, let $v = \alpha\mathbf{x}^3 + \beta\mathbf{x}^4$ be a vector that's in the nonnegative cone defined by $\mathbf{x}^3$ and $\mathbf{x}^4$
    * by linearity, $Av = \alpha A\mathbf{x}^3 + \beta A\mathbf{x}^4$, i.e., Av is in the nonnegative cone defined by $A\mathbf{x}^3$ and $A\mathbf{x}^3$.



(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

  * The same is true for a positive linear combination of, say, $-\mathbf{x}^3$ and $-\mathbf{x}^4$.
- Now consider the particular matrix A. 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.)

For every symmetric $n \times n$ matrix A, there are infinitely many vectors that satisfy the above definition: if $\mathbf{v}$ is an eigenvector, then so also is $\gamma\mathbf{v}$, for arbitrary $\gamma \in \mathbb{R}$. 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* $2n$ unit 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.

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

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

Theorem: If v^1 and v^2 are eigen-values of a symmetric $n \times n$ matrix A, with eigen-values λ_1 and λ_2, if $\lambda_1 \neq \lambda_2$, then v^1 and v^2 are orthogonal

Proof: If v^i is an eigenvector of A then $Av^i = \lambda_i v^i$. Therefore

(1) $(v^2)'Av^1 = (v^2)'\lambda_1 v^1 = \lambda_1 (v^2)'v^1$.
(2) $(v^1)'Av^2 = (v^1)'\lambda_2 v^2 = \lambda_2 (v^1)'v^2$.
(3) since A is symmetric, $(v^2)'Av^1 = (v^1)'Av^2$
(4) hence $\lambda_1 (v^2)'v^1 = \lambda_2 (v^1)'v^2$
(5) Now since $(v^2)'v^1 = (v^1)'v^2$, we can factor out this common term, to obtain.

$$(\lambda_1 - \lambda_2)(v^1)'v^2 = 0$$

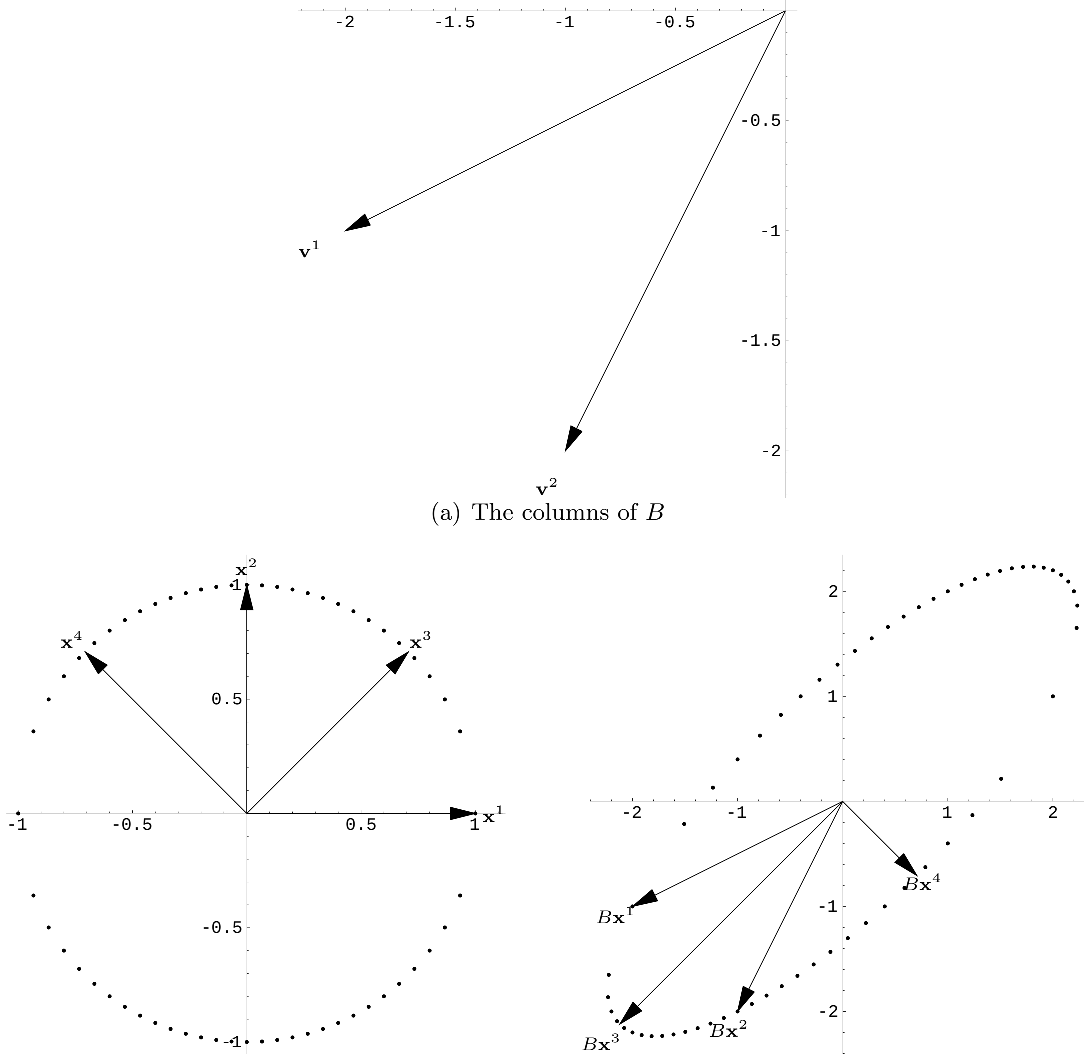
Since by assumption $(\lambda_1 - \lambda_2) \neq 0$, it follows that $(v^1)'v^2 = 0$, i.e., v^1 and v^2 are orthogonal.

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

Question: Under what conditions will a symmetric matrix have an infinite number of unit eigenvectors?

Answer: Iff at least two distinct eigenvectors have *eigenvalues* that are equal to each other

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 $\mathbf{x}'A\mathbf{x} > (\geq) 0$.



(a) The columns of B

(b) Selected elements of the unit circle

(c) The image of the circle under B

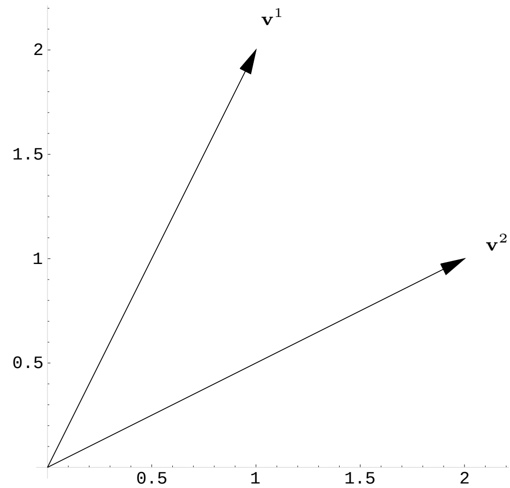
FIGURE 2. What the matrix B does to the unit circle

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

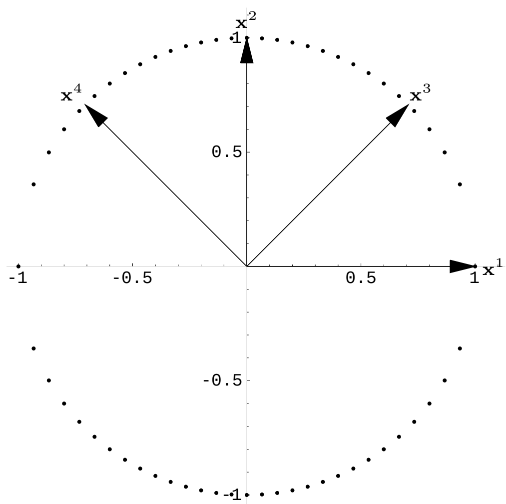
Now consider the negative of matrix A, which has both arrows pointing into the negative orthant. Let $B = \begin{bmatrix} -2 & -1 \\ -1 & -2 \end{bmatrix}$.

- 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 $\mathbf{x}^2$.

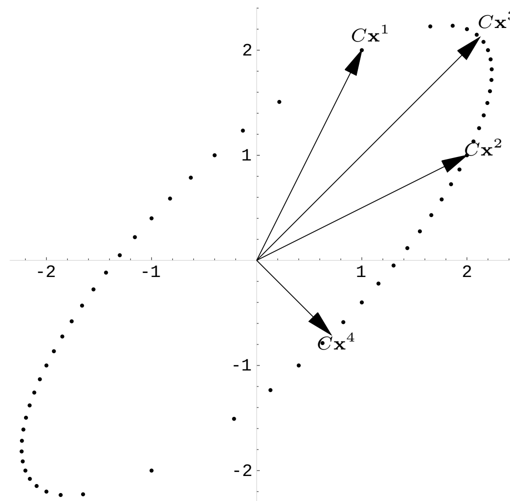
Defn: An $n \times n$ matrix A is *negative definite* (*negative semidefinite*) if for every $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \neq 0$ implies $\mathbf{x}A\mathbf{x} < (\leq) 0$.



(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

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 there is some vector $\mathbf{x}$ such that $\mathbf{x}$ and $C\mathbf{x}$ make an acute angle with each other, and some other vector $\mathbf{y}$ such that $\mathbf{y}$ 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:

- 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.

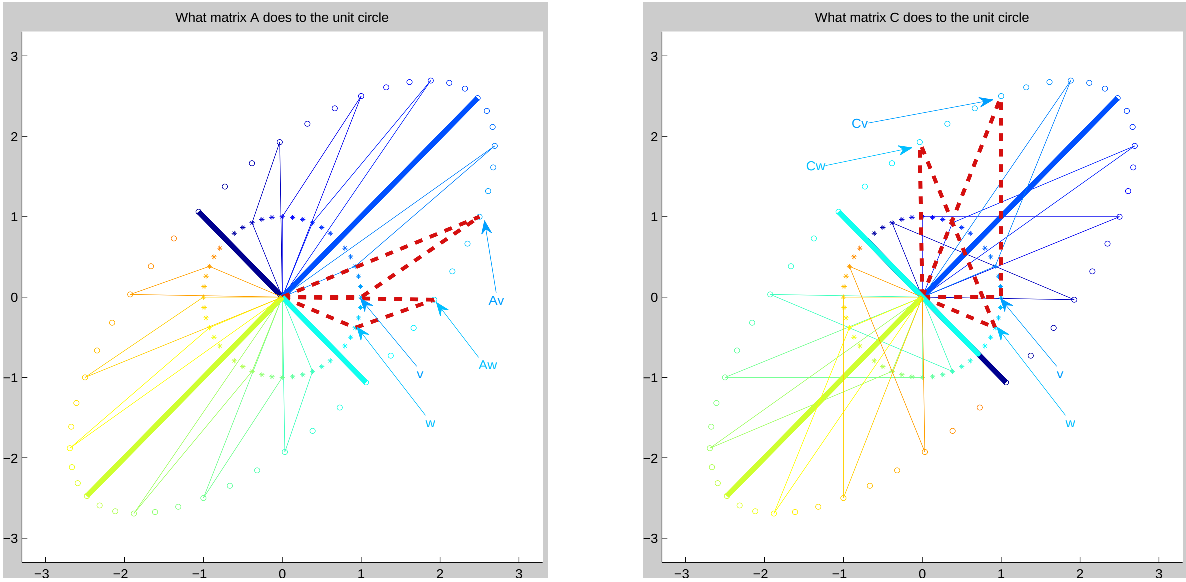


FIGURE 4. Relationship between orientation and determinant

- 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.

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. This is illustrated in Fig. 4. The left panel superimposes the image of the unit circle under A on the circle itself. The figure consists of a number of triangles. For each triangle in the figure:

(1) vertex #1 is the origin
(2) vertex #2 is a point on the unit circle
(3) vertex #3 is the point on the ellipse to which vertex #2 is mapped by A.

The right panel does the same for the matrix C. Matrix A preserves orientation in the sense that for any two unit vectors $\mathbf{v}$ and $\mathbf{w}$, if vector $\mathbf{w}$ in the unit circle is reached by moving counter-clockwise from $\mathbf{v}$, then $A\mathbf{w}$ in the ellipse is reached by moving counter-clockwise from $A\mathbf{v}$. On the other hand, C reverses orientation in the sense that $C\mathbf{w}$ is reached by moving *clockwise* from $C\mathbf{v}$. 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.