2. Linear Algebra (cont) 1

2.9. Linear Functions 1

2.10. The “graph” of a linear function from $\mathbb{R}^2$ to $\mathbb{R}^2$ 3

2. LINEAR ALGEBRA (CONT)

2.9. 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., $f(X) = \{f(x) : x \in X\}$.

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 $f : X \rightarrow Y$ is defined as:

$$\text{graph } f = \{ (x, y) \in X \times Y : y = f(x) \}$$  \hfill (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:

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

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

- $f(x + y) = f(x) + f(y)$ \hspace{1cm} Additivity
- $f(\alpha x) = \alpha f(x)$ \hspace{1cm} Proportionality

Examples:

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

Note that the definition of a linear function doesn’t make any sense unless both the domain $X$ and the image, $I = f(X)$ are vector spaces. To see this note recall that we require for all $x, y \in X$ and all $\alpha \in \mathbb{R}$,

- $f(x + y) = f(x) + f(y)$
- $f(\alpha x) = \alpha f(x)$

but if $X$ is not a vector space, then it is possible that $x, y \in X$ but either $x + y$ or $\alpha x$ isn’t in $X$, in which case $f$ won’t be defined for these values. Similarly, if both $f(x), f(y) \in I$ but $f(x) + f(y)$ isn’t, then $I$ is not in fact the image of $f$, i.e., the function has been improperly specified.

Note also that the requirement of linearity doesn’t impose any restriction on the range of $f$. from now on we’ll assume that both the domain and the image of every linear function $f$ are vector spaces.

Fact A function $f$ is linear if and only if its graph is a vector space.

2.10. **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 $\mathbb{R}$, $y = a \cdot x$, 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 $\mathbb{R}^2$, $y = Ax$, 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 $Ax$ 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 $x' = Ax'$
• 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 nonsymmetric 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-partial are identical. So the restriction is harmless.

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

\[ \{ b : b = Ax, \text{ for some } x \text{ whose norm is unity} \} \]

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

\[ \{ b : b = Ax, \text{ for some } x \text{ whose norm is unity} \} \]

• 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 $x^3 = \left( \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}} \right)$ and $x^4 = \left( \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right)$.
  – 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.
- Observe how any arrow in the unit circle gets swivelled by no more than 90 degrees, i.e., for any vector $x$, the inner product of $Ax$ and $x$ 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 $b$ that lies in the nonnegative cone defined by our two eigenvectors, $x^3$ and $x^4$:
  * $Ab$ has to lie in the nonnegative cone defined by the vectors $Ax^3$ and $Ax^4$:  

![Diagrams showing the columns of matrix A, selected elements of the unit circle, and the image of the circle under A.](image-url)
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 colinear, 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, $\lambda$ 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 $\mathbf{v}$ is an eigenvector of $A$ if the image of that vector under $A$, i.e., $A\mathbf{v}$ belongs to (i.e., is colinear with) the single-dimensional vector space spanned by $\mathbf{v}$.)

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: 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 $x \in \mathbb{R}^n$, $x \neq 0$ implies $x^T A x < (\leq) 0$.

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 $x, x$ and $C x$ make an
acute angle with each other, and for for some other vectors $y$, $y$ and $Cy$ 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.
- 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 $x, y \in \mathbb{R}^n$ such that the product of $x'Ax$ is positive and $y'Ay$ 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 clockwise from $v$, then $Aw$ in the ellipse is reached by moving clockwise from $Av$. On the other hand, $C$ reverses orientation in the sense that $Cw$ is reached by moving *counter-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.
Matrix A has positive determinant

Matrix C has negative determinant

**Figure 4.** Relationship between orientation and determinant