Linear Transformations and Matrix Algebra

A. Havens

Department of Mathematics
University of Massachusetts, Amherst

February 10-16, 2018

Outline

Components Revisited

Observe that any $\mathbf{x} \in \mathbb{R}^2$ can be written as a linear combination of vectors along the standard rectangular coordinate axes using their components relative to this standard rectangular coordinate system:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

These two vectors along the coordinate axes will form *the standard basis for* $\mathbb{R}^2$.

Elementary Vectors

Definition

The vectors along the standard rectangular coordinate axes of $\mathbb{R}^2$ are denoted

$$\mathbf{e}_1 := \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \mathbf{e}_2 := \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

They are called *elementary vectors* (hence the notation $\mathbf{e}_i$, $i = 1, 2$), and the ordered list $(\mathbf{e}_1, \mathbf{e}_2)$ is called the *standard basis of* $\mathbb{R}^2$.

Observe that $\mathrm{Span}\,\{\mathbf{e}_1, \mathbf{e}_2\} = \mathbb{R}^2$.

Elementary Vectors

We can also define elementary vectors and a standard basis in $\mathbb{R}^n$, by taking the unit vectors along the n different coordinate axes of the standard rectangular coordinate system:

Definition

The n vectors

$$\mathbf{e}_1 := \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{e}_2 := \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \mathbf{e}_{n-1} := \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{e}_n := \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

are called *elementary vectors* for the standard rectangular coordinate system on $\mathbb{R}^n$.

A Remark on Notation

Remark

Note that we use the symbols $\mathbf{e}_1, \mathbf{e}_2, \ldots$ to respectively represent the first, second, etc elementary vectors in whatever real vector space we are working in, indexed with respect to the order of our coordinate axes.

The number of components necessary to represent a given $\mathbf{e}_i$ depends on the particular $\mathbb{R}^n$ with which we are working, and will be clear from context.

Thus, the notation $\mathbf{e}_i$ always refers to a vector with a 1 in the ith component, but the vector may have however many zeroes we need.

The Standard Basis of $\mathbb{R}^n$

Observation

Clearly, the set $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\} \subset \mathbb{R}^n$ is linearly independent, as the matrix $[\mathbf{e}_1 \ \ldots \ \mathbf{e}_n]$ has precisely n columns and n pivots.

Definition

The ordered n-tuple of vectors $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ is called *the standard basis of* $\mathbb{R}^n$.

Observe that $\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{e}_i \implies \mathrm{Span}\{\mathbf{e}_1, \ldots, \mathbf{e}_n\} = \mathbb{R}^n$.

Therefore in analogy to the case of $\mathbb{R}^2$, the n-tuple $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ earns the title of a basis because it is an ordered, linearly independent collection of vectors that spans the whole of $\mathbb{R}^n$.

RREF and the Standard Basis

Observation

Given a collection of n linearly independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$, the matrix with these vectors as columns has

$$\mathrm{RREF}[\mathbf{v}_1 \ \ldots \ \mathbf{v}_n] = [\mathbf{e}_1 \ \ldots \ \mathbf{e}_n] = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & 1 \end{bmatrix}.$$

Thus, any $n \times n$ matrix whose columns are linearly independent is row equivalent to the matrix whose columns are the standard basis.

Matrix Representations

Definition

Given a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$, we will say that an $m \times n$ matrix A is a *matrix representing the linear transformation* T if the image of a vector $\mathbf{x}$ in $\mathbb{R}^n$ is given by the matrix vector product

$$T(\mathbf{x}) = \mathrm{A}\mathbf{x}\,.$$

Our aim is to find out how to find a matrix A representing a linear transformation T. In particular, we will see that the columns of A come directly from examining the action of T on the standard basis vectors.

Before we state the formal result, let us consider a simple two dimensional reflection, and try to represent it as a matrix-vector product.

An Example

Example

Find the matrix representing the map $M : \mathbb{R}^2 \to \mathbb{R}^2$ that reflects a vector $\mathbf{x}$ through the line $\mathrm{Span}\left\{\mathbf{e}_1 + \mathbf{e}_2\right\}$.

First, note that

$$\mathbf{e}_1 + \mathbf{e}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

so $\mathrm{Span}\left\{\mathbf{e}_1 + \mathbf{e}_2\right\}$ is the solution set of the linear equation $x_1 - x_2 = 0$, i.e., it is the line $x_1 = x_2$. You should convince yourself that reflection through this line swaps the vectors $\mathbf{e}_1$ and $\mathbf{e}_2$, and in general acts on a vector by swapping its components.

An Example

Example

So $M(\mathbf{e}_1) = \mathbf{e}_2$ and $M(\mathbf{e}_2) = \mathbf{e}_1$. If we consider an arbitrary 2-vector $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2$, it is easy to check that because M is linear and swaps the elementary vectors, M must swap the components of $\mathbf{x}$.

Indeed by linearity
$M(\mathbf{x}) = M(x_1\mathbf{e}_1 + x_2\mathbf{e}_2) = x_1 M(\mathbf{e}_1) + x_2 M(\mathbf{e}_2) = x_1\mathbf{e}_2 + x_2\mathbf{e}_1$.

An Example

Example

Comparing the image

$$M(\mathbf{x}) = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}$$

with

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{bmatrix}$$

we see that

$$\begin{bmatrix} x_2 \\ x_1 \end{bmatrix} = \begin{bmatrix} 0x_1 + 1x_2 \\ 1x_1 + 0x_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

An Example

Example

So we can represent the reflection map $\mathbf{x} \mapsto M(x)$ by the matrix-vector product map

$$M(\mathbf{x}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{x} = [\mathbf{e}_2\ \mathbf{e}_1]\mathbf{x}\,.$$

It is not a coincidence that the matrix of M is $[\mathbf{e}_2\ \mathbf{e}_1] = [M(\mathbf{e}_1)\ M(\mathbf{e}_2)]$!

Indeed, consider the matrix vector product $\mathrm{A}\mathbf{e}_i$ for an arbitrary $m \times n$ matrix A and $\mathbf{e}_i$ the ith elementary vector of the standard basis of $\mathbb{R}^n$. What is the vector $\mathrm{A}\mathbf{e}_i$?

Selecting the Matrix Columns

Since $\mathbf{e}_i$ has a one in the ith coordinate, and zeroes in all other coordinates, we deduce that $\mathrm{A}\mathbf{e}_i$ is the linear combination

$$0\mathbf{a}_1 + \ldots + 0\mathbf{a}_{i-1} + 1\mathbf{a}_1 + 0\mathbf{a}_{i+1} + \ldots + 0\mathbf{a}_n = \mathbf{a}_i\,,$$

that is, $\mathrm{A}\mathbf{e}_i$ is just the ith column of A.

If T is some linear map, and A is a matrix representing it, then we can deduce that the image of an elementary vector $\mathbf{e}_i$ under the map T is $T(\mathbf{e}_i) = \mathbf{a}_i$, so the columns of the matrix are precisely the images of the standard basis by the map T!

The Theorem

Theorem

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ *may be uniquely represented as a matrix-vector product* $T(\mathbf{x}) = \mathrm{A}\mathbf{x}$ *for the* $m \times n$ *matrix* A *whose columns are the images of the standard basis* $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ *of* $\mathbb{R}^n$ *by the transformation* T*. Specifically, the* i*th column of* A *is the vector* $T(\mathbf{e}_i) \in \mathbb{R}^m$ *and*

$$T(\mathbf{x}) = \mathrm{A}\mathbf{x} = \left[\begin{array}{cccc} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \ldots & T(\mathbf{e}_n) \end{array}\right] \mathbf{x}\,.$$

A Satisfying, Simple Proof

Proof.

The result is a consequence of the calculation

$$T(\mathbf{x}) = T\left(\sum_{i=1}^n x_i\mathbf{e}_i\right) = \sum_{i=1}^n x_i T(\mathbf{e}_i)$$
$$= \left[\ T(\mathbf{e}_1) \quad T(\mathbf{e}_2) \quad \dots \quad T(\mathbf{e}_n)\ \right]\mathbf{x} =: \mathrm{A}\mathbf{x}\,,$$

where the first equality follows from the representation of $\mathbf{x}$ in the standard basis, the second equality follows from properties of linearity, and the third equality follows from the definition of the matrix vector product $\mathrm{A}\mathbf{x}$ as being the linear combination of column vectors of A taking the components x_i as the weights. □

Using this Result

There are two ways in which this result is useful:

- Given a linear map described geometrically, one can examine its effect on basis elements $\mathbf{e}_i$ and then describe the matrix representing the map,
- Given a matrix, one can try to understand the geometry of the map $\mathbf{x} \mapsto A\mathbf{x}$ by examining the columns, and understanding how the matrix acts on *the frame* $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$.

The first use is quite practical and appears in many applications. The second use, while occasionally practical, is better viewed as a conceptual framework for understanding the geometry of linear maps when given their matrices; it is typically quite impractical to actually grasp the meaning of a linear map from its representing matrix whenever the matrix is large.

Envisioning the Effect of a Matrix

To envision how the matrix acts on the frame $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$, imagine the basis vectors like a rigid collection of unit length rods, making right angles with each other and aligned with coordinate axes, emanating from the origin of $\mathbb{R}^n$.

The map $\mathbf{x} \mapsto \mathrm{A}\mathbf{x}$ then contorts, bends, rotates, collapses and/or shoves this frame onto a new collection of vectors, the columns of A, sitting in $\mathbb{R}^m$.

The rigidity of the condition that the map is linear means that linear combinations, built via the frame, must map to the correctly weighted linear combinations of the frame vectors' images. This is conceptually why specifying just the image of this standard basis frame determines the effect on arbitrary vectors.

Some Useful Exercises

- Go back through the examples of linear transformations, such as rotations, projections, and similarity transformations, given in the previous lectures on linear maps. For these examples, try to use the theorem to justify any given matrix representations via geometry, to find general matrix representations when they were not given, and to understand the geometry of transformations given by matrices previously encountered in recent lectures.
- Pay particular attention to rotations and reflections in two dimensions. You should become comfortable recognizing matrices that accomplish these transformations, and you should be able to construct a rotation or reflection matrix given sufficient information (such as an angle of rotation, and either a vector or an angle specifying the line of reflection).

Existence Questions and Images

Given a system $\mathrm{A}\mathbf{x} = \mathbf{b}$, what is the connection between the existence of a solution $\mathbf{x}$ and the linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ given by $T(\mathbf{x}) = \mathrm{A}\mathbf{x}$?

If a solution exists for some $\mathbf{b}$, that means there is some $\mathbf{x} \in \mathbb{R}^n$ such that $T(\mathbf{x}) = \mathbf{b}$. Thus, we can rephrase the question of existence of solutions as follows: given a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ represented by an $m \times n$ matrix A, and a vector $\mathbf{b} \in \mathbb{R}^m$, is $\mathbf{b} \in T(\mathbb{R}^n)$, i.e., is $\mathbf{b}$ in the image of the map T?

An affirmative answer implies there exists at least one $\mathbf{x}$, a *pre-image of* $\mathbf{b}$, which solves the system $\mathrm{A}\mathbf{x} = \mathbf{b}$.

A negative answer implies the system $\mathrm{A}\mathbf{x} = \mathbf{b}$ is inconsistent.

Existence and Pre-Images

Formally, define the *pre-image of a vector* $\mathbf{b} \in \mathbb{R}^m$ *under a map* $T : \mathbb{R}^n \to \mathbb{R}^m$ by

$$T^{-1}(\mathbf{b}) = \{\mathbf{x} \in \mathbb{R}^n | T(\mathbf{x}) = \mathbf{b}\} .$$

Let $T(\mathbf{x}) = A\mathbf{x}$ be a linear transformation. Then a solution to $A\mathbf{x} = \mathbf{b}$ always exists (for any $\mathbf{b}$) if and only if for every $\mathbf{b} \in \mathbb{R}^m$, the cardinality of the pre-image of $\mathbf{b}$ is at least 1: $|T^{-1}(\mathbf{b})| \geq 1$.

Uniqueness Questions And Kernels

Suppose for a given $m \times n$ matrix A and a given vector $\mathbf{b} \in \mathbb{R}^m$, system $\mathrm{A}\mathbf{x} = \mathbf{b}$ is consistent, so $\mathbf{b} \in T(\mathbf{x})$ where $T(\mathbf{x}) = \mathrm{A}\mathbf{x}$ is the corresponding linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$.

The uniqueness question concerns whether there is only one solution $\mathbf{x}$ such that $\mathrm{A}\mathbf{x} = \mathbf{b}$, or infinitely many. This corresponds to asking whether the *pre-image of* $\mathbf{b}$ contains more than one element.

For example, if $\mathbf{b} = \mathbf{0}$, then the system is homogenous. Any nontrivial solution would imply that the solution is not unique.

Uniqueness Questions And Kernels

Recall the notion of the kernel of the map $\mathbf{x} \mapsto \mathrm{A}\mathbf{x}$: the kernel is precisely the pre-image of the zero vector. For any linear map $T : \mathbb{R}^n \to \mathbb{R}^m$, let

$$\ker T := T^{-1}(\mathbf{0}) = \{\mathbf{x} \in \mathbb{R}^n | T(\mathbf{x}) = \mathbf{0}\}.$$

The kernel is called *trivial* if it contains only the zero vector, i.e., $\ker T = \{\mathbf{0}\}$.

Then observe that the homogeneous system $T(\mathbf{x}) = \mathrm{A}\mathbf{x} = \mathbf{0}$ has nontrivial solutions if and only if if the kernel is nontrivial.

Kernels and Inhomogeneous Systems

But, the solution to an inhomogeneous system is constructed by finding a particular solution, and then adding solutions of the homogeneous system, i.e., adding elements of the kernel of the map $\mathbf{x} \mapsto \mathrm{A}\mathbf{x}$.

It thus follows that if $\ker T$ is nontrivial, the corresponding system $\mathrm{A}\mathbf{x} = \mathbf{b}$ has nontrivial solutions provided $\mathbf{b}$ is in the image of T.

So the question of uniqueness may be rephrased in terms of the cardinality of the pre-image $T^{-1}(\mathbf{b})$ of the vector $\mathbf{b}$ by the map $T(\mathbf{x}) = \mathrm{A}\mathbf{x}$. And this in turn is equivalent to the question of whether the kernel of T is trivial.

We recall two ideas about functions, before collecting answers to our existence and uniqueness questions.

Defining Surjectivity

Definition

A function $f : X \to Y$ (not necessarily linear) is called *surjective* or *onto* if for every $y \in Y$ there exists at least one $x \in X$ such that $y = f(x)$.
Equivalently, the function f is surjective if and only if the cardinalities of all pre-images are at least 1, i.e. for every $y \in Y$, $|f^{-1}(y)| \geq 1$.

Intuitively, an onto map *covers* the codomain, i.e., the whole codomain is the image: $f(X) = Y$.

Defining Injectivity

Definition

A function $f : X \to Y$ (not necessarily linear) is called *injective* or *one-to-one* if and only if whenever two images $f(x_1)$, $f(x_2)$ are equal, the corresponding inputs x_1 and x_2 are also equal. Equivalently, f is injective/one-to-one if and only if whenever two domain elements x_1 and x_2 are distinct, the corresponding images $f(x_1)$ and $f(x_2)$ are distinct.

Injectivity and Pre-Images

Intuitively, a one-to-one function is one which "never repeats an output".

That is, distinct inputs always produce distinct outputs for an injective function. Another way to understand injectivity is to consider pre-images: all of the pre-images of an injective function contain at most one point.

Thus $f : X \to Y$ is injective if and only if $x_1 \neq x_2 \implies f(x_1) \neq f(x_2)$, if and only if for every $y \in Y$, the cardinality of the pre-image $f^{-1}(y) = \{x \in X | y = f(x)\}$ satisfies $|f^{-1}(y)| \leq 1$.

Existence and Surjectivity

Theorem

A linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ given by $T(\mathbf{x}) = \mathrm{A}\mathbf{x}$ is surjective if and only if the columns of A span $\mathbb{R}^m$.

The system $\mathrm{A}\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b}$ in $\mathbb{R}^m$ if and only if the map $T(\mathbf{x}) = \mathrm{A}\mathbf{x}$ is surjective.

Uniqueness and Injectivity

Theorem

A linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ is injective if and only if $T(\mathbf{x}) = \mathbf{0}$ only admits the trivial solution, i.e., if and only if $\ker T = \{\mathbf{0}\}$.

If $T(\mathbf{x}) = \mathrm{A}\mathbf{x}$, the system $\mathrm{A}\mathbf{x} = \mathbf{b}$ has a unique solution if and only if T is injective, i.e., if and only if the kernel is trivial.

The kernel of T is trivial if and only if the columns of the representative matrix A are linearly independent.

An Example

Example

Consider the map $T : \mathbb{R}^3 \to \mathbb{R}^3$ given by $T(\mathbf{x}) = \mathrm{A}\mathbf{x}$, where

$$\mathrm{A} = \begin{bmatrix} 0 & e & \pi \\ \sqrt{2} & 0 & 1 \\ 0 & -e & \pi \end{bmatrix}.$$

Is the map surjective? What does this say about solutions to $\mathrm{A}\mathbf{x} = \mathbf{b}$ for arbitrary $\mathbf{b} \in \mathbb{R}^3$?

Is this map injective? What does this say about uniqueness of solutions to $\mathrm{A}\mathbf{x} = \mathbf{b}$?

An Example

Example

The map is both injective and surjective: we previously showed that $\mathrm{RREF}(\mathrm{A}) = \mathrm{I}_3$, the 3×3 identity matrix.

Since there are three pivots, one in each column, the columns of A are linearly independent, so they are not coplanar. They thus span $\mathbb{R}^3$, which proves surjectivity.

Since the map T is surjective, any $\mathbf{b} \in \mathbb{R}^3$ is in the image of T, whence, $\mathrm{A}\mathbf{x} = \mathbf{b}$ is solvable.

Since the columns are linearly independent, $\ker T$ is trivial, and any system $\mathrm{A}\mathbf{x} = \mathbf{b}$ is in fact uniquely solved.

Composition

Suppose we wanted to compose a pair of linear maps induced by matrix multiplication:

$$\mathbb{R}^k \xrightarrow{T_{\mathrm{B}}} \mathbb{R}^n \xrightarrow{T_{\mathrm{A}}} \mathbb{R}^m,$$

where B is the $n \times k$ matrix representing T_{B} and A is the $m \times n$ matrix representing T_{A}. Let $T_{\mathrm{AB}} = T_{\mathrm{A}} \circ T_{\mathrm{B}}$ denote the composition obtained by first applying T_{B} and then applying T_{A}.

You should work out that this composition is indeed also a linear map.

Composition

We know that we should be able to represent this composition by a matrix map. Our theorem for building such matrices representing linear transformations tells us that if we track what happens to the standard basis through the two maps building the decomposition, we will know the columns of the matrix representing the decomposition.

It turns out we can, and the corresponding matrix can be though of as a *matrix product* of A and B. Let us do an example before defining this product in full generality.

An Example

Example

$$\text{Let } \mathrm{A} = \begin{bmatrix} 3 & 2 & 1 \\ 6 & 5 & 4 \end{bmatrix}, \text{ and } \mathrm{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}.$$

Thus, $T_{\mathrm{A}} : \mathbb{R}^3 \to \mathbb{R}^2$ is given by $T_{\mathrm{A}}\mathbf{y} = \mathrm{A}\mathbf{y}$ and $T_{\mathrm{B}} : \mathbb{R}^2 \to \mathbb{R}^3$ is given by $T_{\mathrm{B}}\mathbf{x} = \mathrm{B}\mathbf{x}$.

Given $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$, the map $T_{\mathrm{AB}} : \mathbb{R}^2 \to \mathbb{R}^2$ sends $\mathbf{x}$ to $\mathrm{A}(\mathrm{B}\mathbf{x})$. Let $\mathbf{y} = \mathrm{B}\mathbf{x}$.

An Example

Example

Then $\mathbf{y} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 \\ 3x_1 + 4x_2 \\ 5x_1 + 6x_2 \end{bmatrix}$.

We can then compute $T_{\mathrm{AB}}\mathbf{x} = \mathrm{A}\mathbf{y}$:

$$\mathrm{A}\mathbf{y} = \mathrm{A}(\mathrm{B}\mathbf{x}) = \begin{bmatrix} 3 & 2 & 1 \\ 6 & 5 & 4 \end{bmatrix} \begin{bmatrix} x_1 + 2x_2 \\ 3x_1 + 4x_2 \\ 5x_1 + 6x_2 \end{bmatrix}$$

$$= \begin{bmatrix} 3(x_1 + 2x_2) + 2(3x_1 + 4x_2) + (5x_1 + 6x_2) \\ 6(x_1 + 2x_2) + 5(3x_1 + 4x_2) + 4(5x_1 + 6x_2) \end{bmatrix}$$

An Example

Example

Carrying on the computation, we find that the composition is given by

$$T_{AB}\mathbf{x} = \begin{bmatrix} \big(3(1)+2(3)+1(5)\big)x_1 + \big(3(2)+2(4)+1(6)\big)x_2 \\ \big(6(1)+5(3)+4(5)\big)x_1 + \big(6(2)+5(4)+4(6)\big)x_2 \end{bmatrix}$$
$$= \begin{bmatrix} 3(1)+2(3)+1(5) & 3(2)+2(4)+1(6) \\ 6(1)+5(3)+4(5) & 6(2)+5(4)+4(6) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$
$$= \begin{bmatrix} 14 & 20 \\ 41 & 56 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 14x_1 + 20x_2 \\ 41x_1 + 56x_2 \end{bmatrix}.$$

An Example

Example

Observe that the matrix in the penultimate line above is obtained by forming dot products from the row vectors of A with the column vectors of B to obtain each entry. This is how we will define matrix multiplication in general: we treat the columns of the second matrix as vectors, and compute matrix-vector products in order to obtain new column vectors.

Defining the Matrix Product

Definition

Suppose we have linear maps

$$\mathbb{R}^k \xrightarrow{T_{\mathrm{B}}} \mathbb{R}^n \xrightarrow{T_{\mathrm{A}}} \mathbb{R}^m \,,$$

represented respectively by a $n \times k$ matrix B and an $m \times n$ matrix A.

Let $T_{\mathrm{AB}} = T_{\mathrm{A}} \circ T_{\mathrm{B}} : \mathbb{R}^k \to \mathbb{R}^m$ denote the composition obtained by first applying T_{B} and then applying T_{A}.

Then there is an $m \times k$ matrix M such that $T_{AB}\mathbf{x} = \mathrm{M}\mathbf{x}$ for any $\mathbf{x} \in \mathbb{R}^k$, and we define the *matrix product of* A *and* B to be the matrix $\mathrm{AB} := \mathrm{M}$.

Defining the Matrix Product

Definition

In particular, $\mathrm{M} = (m_{ij})$ is the matrix whose entries are given by the formula

$$m_{ij} = \sum_{l=1}^{n} a_{il} b_{lj} \,,$$

where a_{il} is the ith element of the lth column $\mathbf{a}_l$ of A (which is the lth element of the ith row of A), and b_{lj} is the lth element of the jth column $\mathbf{b}_j$ of B.

Thus, the jth column of $\mathrm{M} = \mathrm{AB}$ is precisely the matrix-vector product $\mathrm{A}\mathbf{b}_j$ where $\mathbf{b}_j$ is the jth column of B:

$$\mathrm{AB} = \begin{bmatrix} \mathrm{A}\mathbf{b}_1 & \dots & \mathrm{A}\mathbf{b}_k \end{bmatrix}.$$

Compatibility

Observation

For the product to be defined, the number of columns of the first matrix must match the number rows of the second matrix. In particular, if A is an $m \times n$ matrix and B is an $n \times k$, then AB is well defined, but BA is well defined if and only if $k = m$.

The final size has the same number of rows (m) as A and the same number of columns (k) as B.

Non-Commutativity

Remark

The above observation implies that there are pairs matrices A and B such that AB is defined while the product in reverse order BA is not defined.

If both products are defined, they need not be equal, and indeed, may even have different sizes. E.g., if A is a 2×3 matrix and B is a 3×2 matrix, then AB is a 2×2 matrix, but BA is a 3×3 matrix!

For square matrices of the same dimensions, the product is defined in either order, and returns a square matrix of the same size.

But the results of such a product still need not be equal, as the following example shows.

An example of Non-Commuting Square matrices

Example

Consider the following matrices:

$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

We compute the products in each order:

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

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

Identity

There is a distinguished $n \times n$ *identity matrix* I_n such that for any $m \times n$ matrix A, $\mathrm{A}\mathrm{I}_n = \mathrm{A}$ and for any $n \times k$ matrix B, $\mathrm{I}_n\mathrm{B} = \mathrm{B}$.

This matrix consists of entries δ_{ij} which are 1 if $i = j$ and 0 if $i \neq j$:

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

Clearly, for any vector $\mathbf{x} \in \mathbb{R}^n$, $\mathrm{I}_n\mathbf{x} = \mathbf{x}$, whence it also acts as an identity for matrix-vector multiplication, when products are defined.

Associativity

Remark

Matrix multiplication of real matrices *is associative*. In particular, if A, B and C are matrices for which the products $\mathrm{A}(\mathrm{BC})$ and $(\mathrm{AB})\mathrm{C}$ are defined, then in fact these are the same and thus without ambiguity we have

$$\mathrm{A}(\mathrm{BC}) = \mathrm{ABC} = (\mathrm{AB})\mathrm{C}\,.$$

This follows generally from the associativity of function composition, but can also be proven in a "hands-on" (albeit, tedious) way using the formula for the entries of a matrix product, some indicial manipulations, and the associativity of real addition.

Elementary Matrices and Row Operations

A useful fact for when we study *inverses of square matrices* is that the elementary row operations performed during row reduction can be represented by matrix products.

Can you find a matrix which swaps the ith and jth rows of an $m \times n$ matrix A, and leaves all other rows unchanged? Should it multiply A from the right, or from the left?

A hint is to consider elementary vectors, and how they can "pick out" columns of a matrix. How can you pick out rows?

Can you find a matrix that scales the ith row of an $m \times n$ matrix A by a scalar s, but leaves the remaining rows unchanged? Should it multiply A from the right, or from the left?

Elementary Matrices and Row Operations

Can you identify a matrix which acts on an $m \times n$ matrix A by replacing the ith row with the sum of the ith row and s times the jth row, for some scalar $s \in \mathbb{R}$?

You can also manipulate *columns* by matrix products. How can you accomplish the column operation analogs of the above elementary row operations, using matrix multiplication?

Row and column manipulations by matrix multiplication aid in many programming applications, and play a prominent role in linear coding theory and digital signal processing.

Adding Matrices

Definition

Given two $m \times n$ matrices A and B, the sum $\mathrm{A}+\mathrm{B}$ is defined to be the matrix such that for any $\mathbf{x} \in \mathbb{R}^n$, $(\mathrm{A}+\mathrm{B})\mathbf{x} = \mathrm{A}\mathbf{x} + \mathrm{B}\mathbf{x}$.

Using the indicial notation for entries, we have then that the ith entry of $(\mathrm{A}+\mathrm{B})\mathbf{x}$ is

$$\sum_{j=1}^{n} a_{ij}x_j + \sum_{j=1}^{n} b_{ij}x_j = \sum_{j=1}^{n}(a_{ij}+b_{ij})x_j\,,$$

which implies that $\mathrm{A}+\mathrm{B}$ is obtained by adding corresponding entries of A and B.

Adding Matrices

Example

What is the sum of the matrices

$$A = \begin{bmatrix} 11 & 6 & 2 \\ 3 & -7 & 9 \end{bmatrix}, B = \begin{bmatrix} -4 & 2 & 3 \\ 7 & 8 & -5 \end{bmatrix}?$$

By adding the components, we obtain

$$A + B = \begin{bmatrix} 11-4 & 6+2 & 2+3 \\ 3+7 & -7+8 & 9-5 \end{bmatrix} = \begin{bmatrix} 7 & 8 & 5 \\ 10 & 1 & 4 \end{bmatrix}.$$

Scaling Matrices

Matrices can also be scaled, by simply scaling all the entries: $s\mathrm{A} = (sa_{ij})$ for any $s \in \mathbb{R}$.

In particular, we may also subtract matrices, and each matrix has an additive inverse, equal to -1 times the original matrix.

There's also a unique zero matrix of any given size, consisting of all zero entries. We denote the $m \times n$ zero matrix by $\mathbf{0}_{m\times n}$. If there's no risk of confusion, we may omit the subscript indicating the dimensions.

Spaces of Matrices

It is easy to check that since matrices can be scaled, added, and have an identity, all in analogy to vectors, they satisfy the same eight fundamental properties we described for real vectors.

In fact, we can consider the set of all $m \times n$ matrices as being equivalent to the set of all mn-component vectors, as a real vector space. There is of course not a unique way to identify these spaces.

One can write $\mathbb{R}^{m\times n}$ to denote the space of all $m \times n$ matrices.

The vector arithmetic on the space of matrices $\mathbb{R}^{m\times n}$ is in a sense equivalent to that of the vector space $\mathbb{R}^{mn}$.

We'll clarify this later when we study vector spaces in chapter 4.

The Matrix Transposed

If $\mathrm{A} \in \mathbb{R}^{m\times n}$, then we can define a new matrix called its transpose, which lives in $\mathbb{R}^{n\times m}$:

Definition

The matrix $\mathrm{A} = (a_{ij})$ has transpose $\mathrm{A}^{\mathrm{t}} = (a_{ji})$, in other words, the transpose matrix is the matrix obtained by exchanging the rows of A for columns.

Example

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}^{\mathrm{t}} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

Observation

The dot product of two vectors is equivalent to a matrix vector product where one of the vectors has been transposed:

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^{t}\mathbf{v} = \mathbf{v}^{t}\mathbf{u}\,.$$

For a given 3-vector $\mathbf{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$, we describe a plane equation for the plane $ax + by + cz = 0$ perpendicular to $\mathbf{n}$ through the origin $\mathbf{0} \in \mathbb{R}^3$ as

$$\mathbf{n} \cdot \mathbf{x} = \mathbf{0} \iff \mathbf{n}^{t}\mathbf{x} = [0] \in \mathbb{R}^{1\times 1} \simeq \mathbb{R}.$$

Properties of Transposition

The transpose of a product is the product of the separate transposition, taken in opposite order:

$$(\mathrm{AB})^{\mathrm{t}} = \mathrm{B}^{\mathrm{t}}\mathrm{A}^{\mathrm{t}}\,.$$

One can view transposition as a map from the space $\mathbb{R}^{m\times n}$ to the space $\mathbb{R}^{n\times m}$. It turns out this map is linear!

Indeed, you should verify that transposition commutes with scaling, and distributes over sums:

$$(s\mathrm{A})^{\mathrm{t}} = s(\mathrm{A}^{\mathrm{t}})\,, \quad (\mathrm{A}+\mathrm{B})^{\mathrm{t}} = \mathrm{A}^{\mathrm{t}} + \mathrm{B}^{\mathrm{t}}\,.$$

Homework

- I recommend reading sections 1.7, and 1.8 for Monday 2/12, 1.9 by Wednesday 2/14 (if not by Monday), and 2.1 by Friday 2/16.
- The MyMathLab assignment on 1.7 (linear independence) is due 2/13, and 1.8 (linear transformations) is due 2/15.
- The first exam is coming up! Our section, math 235-04, meets in Hasbrouck Laboratory Addition (HASA) 124 on Tuesday night, February 27th, 7 - 9 pm.
- The first exam covers the material of sections 1.1, 1.2, 1.3, 1.4, 1.5, 1.7, 1.8, 1.9, and 2.1 in the text.