Unit 2F: Linear transformations

Prof.Dr.P.M.Bajracharya

School of Mathematical Sciences
T.U., Kirtipur

March 6, 2024

Summary

1. **Some important theorems**
2. **Standard basis vectors**
3. **Some special matrices**
4. **Examples of matrices as transformations**
   - Random Walk Matrix

Some important theorems

Here are two important theorems that we won't prove but that the above discussion suggests.

Theorem.

Let A be an $m \times n$ matrix. Then, A defines a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ by matrix multiplication: $T(v) = Av$, where v is a column vector.

Theorem.

Every linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is given by an $m \times n$ matrix, call it A. The functional form is given by $T(v) = Av$, i.e., a matrix -vector multiplication, where th jth column of A is $T(e_j)$..

The two theorems given above relate matrices and linear transformations more precisely.

The two theorems given above relate matrices and linear transformations more precisely. Due to them, we will sometimes, but not always, distinguish between a linear transformation T and its associated matrix A.

The two theorems given above relate matrices and linear transformations more precisely. Due to them, we will sometimes, but not always, distinguish between a linear transformation T and its associated matrix A.
The second theorem is powerful and surprising.

The two theorems given above relate matrices and linear transformations more precisely. Due to them, we will sometimes, but not always, distinguish between a linear transformation T and its associated matrix A.
The second theorem is powerful and surprising. It says not just that every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is given by a matrix; it also says that one can construct the matrix by seeing how the transformation acts on the standard basis vectors. This is rather remarkable.

For completeness, we note the following results, which states that if we have a linear transformation corresponding to a matrix, then the inverse linear transformation corresponds to the inverse matrix.

Theorem.

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is invertible iff the $m \times n$ matrix A associated with it is invertible, and $T^{-1} = A^{-1}$.

Note that only square matrices can have inverses. Thus for a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ to be invertible, we must have $m = n$.

Standard basis vectors

Standard basis vectors

The **standard basis vectors** (or **standard unit vectors**) in $\mathbb{R}^n$, denoted e_k, have n entries, with a 1 in the kth position and a 0 in all the other positions.

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

*k*th position

Figure: kth standard basis vector

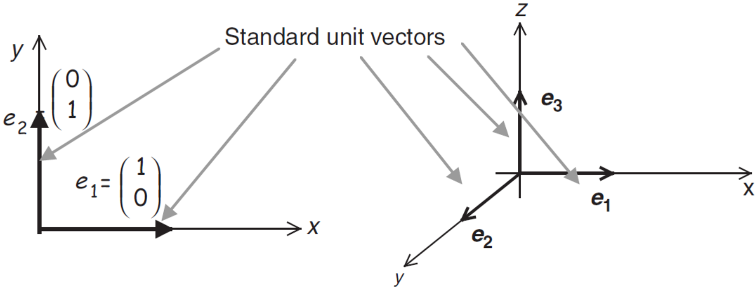


Figure: Standard basis vectors in the plane and in the space.

Question: Why are these standard basis vectors important?

Question: Why are these standard basis vectors important?
Ans. Every vector in the space can be expressed as a linear combination of the basis vectors of the space. It is easier to work with linear combinations.

Consider a vector $x = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \in \mathbb{R}^2$.

Consider a vector $x = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \in \mathbb{R}^2$. It can be written as

$$\begin{pmatrix} 2 \\ 3 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 3 \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Consider a vector $x = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \in \mathbb{R}^2$. It can be written as

$$\begin{pmatrix} 2 \\ 3 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 3 \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

This gives

$$x = x_1 e_1 + x_2 e_2,$$

where $x_1 = 2$ and $x_2 = 3$.

Thus, we have expressed x as a *linear combination* of the basis vectors e_1 and e_2, Thus, we have expressed x as a *linear combination* of the basis vectors e_1 and e_2, and the coefficients involved in the linear combination are the components of the vector x.

Further, consider a linear combination:

$$x_1 e_1 + x_2 e_2,$$

Further, consider a linear combination:

$$x_1e_1 + x_2e_2,$$

Now the question is:

Under what conditions is this linear combination zero?

that is,

When does the equality $x_1e_1 + x_2e_2 = 0$ hold?

To answer to this question, put

$$x = x_1 e_1 + x_2 e_2.$$

To answer to this question, put

$$x = x_1 e_1 + x_2 e_2.$$

Then

$$x = x_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

To answer to this question, put

$$x = x_1 e_1 + x_2 e_2.$$

Then

$$x = x_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$
$$= \begin{pmatrix} x_1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ x_2 \end{pmatrix}$$

To answer to this question, put

$$x = x_1 e_1 + x_2 e_2.$$

Then

$$\begin{aligned} x &= x_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} x_1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ x_2 \end{pmatrix} \\ &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \end{aligned}$$

If $x_1e_1 + x_2e_2 = 0$, then

$$x = 0 \Rightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0 \Rightarrow x_1 = x_2 = 0.$$

If $x_1e_1 + x_2e_2 = 0$, then

$$x = 0 \Rightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0 \Rightarrow x_1 = x_2 = 0.$$

On the otherhand,

$$x_1 = x_2 = 0 \Rightarrow x_1e_1 + x_2e_2 = 0.$$

Therefore,

$$x_1e_1 + x_2e_2 = 0 \Leftrightarrow x_1 = x_2 = 0.$$

Problem

1. Show that any vector in $\mathbb{R}^3$ can be expressed as a linear combination of the three unit basis vectors in $\mathbb{R}^3$. Also, show that a linear combination of the three unit basis vectors in $\mathbb{R}^3$ equals to 0 if and only if all coefficients in the linear combination are zeros.
2. Do the above problem for $\mathbb{R}^n$.

Multiplying a matrix by a standard basis vector

Multiplying a matrix A by the standard basis vector e_i selects out the ith column of A.

This is shown in the following example.

Example

Show that the second column of $A = \begin{pmatrix} 3 & -2 & 0 \\ 2 & 1 & 2 \\ 0 & 4 & 4 \\ 1 & 0 & 2 \end{pmatrix}$ is Ae_2.

Example

Show that the second column of $A = \begin{pmatrix} 3 & -2 & 0 \\ 2 & 1 & 2 \\ 0 & 4 & 4 \\ 1 & 0 & 2 \end{pmatrix}$ is Ae_2.

Solution. We have

$$Ae_2 = \begin{pmatrix} 3 & -2 & 0 \\ 2 & 1 & 2 \\ 0 & 4 & 4 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \\ 4 \\ 0 \end{pmatrix} = A_{:2}.$$

$\square$

Generalizing it we get

jth column of a matrix

The jth column of a matrix A is Ae_j, that is,

$$Ae_j = A_{:j}.$$

jth column of the product AB

The jth column of the product AB is the product of A and the jth column of B.

Example

The second column of the product AB is the product of A and the second column of B.

Example

The second column of the product AB is the product of A and the second column of B.

To show it, let

$$A = \begin{pmatrix} 2 & -1 \\ 3 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 4 & -2 \\ 3 & 0 & 2 \end{pmatrix}.$$

Then

$$AB = \begin{pmatrix} 2 & -1 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} 1 & 4 & -2 \\ 3 & 0 & 2 \end{pmatrix} = \begin{pmatrix} -1 & 8 & -6 \\ 9 & 12 & -6 \end{pmatrix}.$$

Now, we find the product $AB_{:2}$.

$$AB_{:2} = \begin{pmatrix} 2 & -1 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} 4 \\ 0 \end{pmatrix} = \begin{pmatrix} 8 \\ 12 \end{pmatrix} = (AB)_{:2}.$$

The next theorem is a key result for application of matrix operations in data science.

Theorem.

Let $S : \mathbb{R}^\ell \to \mathbb{R}^m$ and $T : \mathbb{R}^n \to \mathbb{R}^\ell$ be linear transformations, given by matrices A and B, respectively. Then, the composition $S \circ T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation and is given by AB.

Scheme of proof:

- A composition of linear transformations is linear.
- Each column of the matrix representing the compositon coincides with corresponding column of AB.

Proof.

Let $v, w \in \mathbb{R}^n$ and $a, b \in \mathbb{R}$. Then

$$\begin{aligned}(S \circ T)(av + bw) &= S(T(av + bw)) \\ &= S(aT(v) + bT(w)) \\ &= aS(T(v)) + bS(T(w)) \\ &= a(S \circ T)(v) + b(S \circ T)(w).\end{aligned}$$

This shows that $S \circ T$ is linear. Then it can be represented by a matrix, say, C.

Proof ...

For any basis vector e_i of $\mathbb{R}^n$ we then have

$$\begin{aligned} Ce_i &= (S \circ T)(e_i) = S(T(e_i)) \\ &= S(Be_i) = A(Be_i) \\ &= (AB)e_i. \end{aligned}$$

This implies that each column of C is equal to the corresponding column of AB and so,

$$C = AB.$$

Therefore, $S \circ T$ is given by AB. $\square$

Some special matrices

Symmetric matrices are important in general but are very important in data science.

Symmetric matrices are important in general but are very important in data science.

Symmetric matrix

A symmetric matrix is a matrix that equals its transpose.

This means, a matrix $A = (A_{ij})$ is symmetric iff

$$A_{ij} = A_{ji}.$$

For example, the matrix

$$\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$$

is a symmetric matrix, while the matrix

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

is not a symmetric matrix.

Triangular matrices

An **upper triangular matrix** is a square matrix with non-zero entries only on or above the diagonal.

An **lower triangular matrix** is a square matrix with non-zero entries only on or below the diagonal.

For example, the matrix

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

is neither upper-triangular nor lower-triangular,
the matrix

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 5 & 6 \\ 0 & 0 & 9 \end{pmatrix}$$

is upper triangular but not lower triangular.

The matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 9 \end{pmatrix}$$

is both upper triangular and lower triangular.

Diagonal matrices.

A **diagonal matrix** is a square matrix with nonzero entries (if any) only on the main diagonal.

Diagonal matrices.

A **diagonal matrix** is a square matrix with nonzero entries (if any) only on the main diagonal.

For example, the matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 9 \end{pmatrix}$$

is a diagonal matrix.

Note that a diagonal matrix is both upper triangular and lower triangular.

Diagonal matrices have many nice properties. For example,

$$\begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}^k = \begin{pmatrix} A_{11}^k & 0 \\ 0 & A_{22}^k \end{pmatrix}.$$

Diagonal matrices have many nice properties. For example,

$$\begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}^k = \begin{pmatrix} A_{11}^k & 0 \\ 0 & A_{22}^k \end{pmatrix}.$$

Due to this property, if $0 < A_{22} < 1$, then we obtain

$$\begin{pmatrix} 1 & 0 \\ 0 & A_{22} \end{pmatrix}^k = \begin{pmatrix} 1 & 0 \\ 0 & A_{22}^k \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{as } k \to \infty.$$

Examples of matrices as transformations

We saw that every matrix represents a linear transformation, and vice versa. In spite of that, it can be difficult to tell what exactly an arbitrary matrix is "doing" when it is just presented as an array of numbers.

- Fortunately, there are some basic examples of matrices that are relatively simple to think about, in the sense that their associated transformations are relatively simple to understand.

- Fortunately, there are some basic examples of matrices that are relatively simple to think about, in the sense that their associated transformations are relatively simple to understand.
- In some cases, these basic examples form the "building blocks" of arbitrary matrices.

- Fortunately, there are some basic examples of matrices that are relatively simple to think about, in the sense that their associated transformations are relatively simple to understand.
- In some cases, these basic examples form the "building blocks" of arbitrary matrices.
- When that happens, one can express the arbitrary matrix in terms of those building blocks; these are often called **matrix decompositions**.

Matrix decompositions are useful for two things: They can help

- To understand structure in data (by describing what a matrix is "doing," and thus what is happening in data that are being represented by the matrix, in terms of simpler operations).
- To perform computations faster (by describing a difficult-to-work with matrix in terms of simpler easier-to-implement parts).

Now, we will give several examples.

1. Identity: The identity transformation $I : \mathbb{R}^n \to \mathbb{R}^n$ is linear and is given by the matrix

$$I_n = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

This is the trivial transformation that doesn't do anything.

2. Scaling: The transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ that takes any input vector and multiplies it by $a \in \mathbb{R}$ is given by

$$A = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$$

2. Scaling: The transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ that takes any input vector and multiplies it by $a \in \mathbb{R}$ is given by

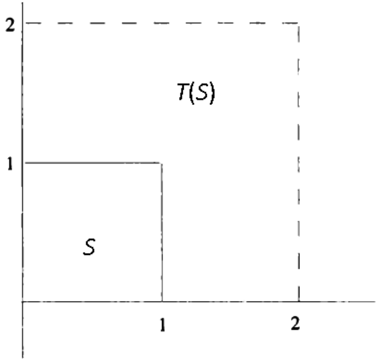
$$A = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$$

For example,

$$Ae_1 = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ 0 \end{pmatrix},$$

$$Ae_2 = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ a \end{pmatrix}.$$

The following figure shows the result of applying T represented by the matrix $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ to the unit square S.



Note that

$$Ax = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 \\ ax_2 \end{pmatrix} = a \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = ax.$$

3. Stretching: The transformation T given by

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \quad \text{with } a \neq b$$

is like scaling, except that the scaling is by a different amount in each direction.

3. Stretching: The transformation T given by

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \quad \text{with } a \neq b$$

is like scaling, except that the scaling is by a different amount in each direction. For example, since

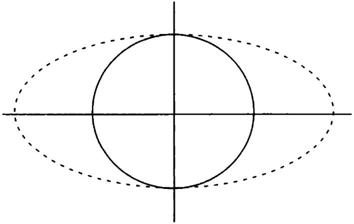
$$Ae_1 = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ 0 \end{pmatrix},$$

$$Ae_2 = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ b \end{pmatrix},$$

we have that

$$Ax = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 \\ bx_2 \end{pmatrix}.$$

The linear transformation given by the matrix $\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ stretches the unit circle into an ellipse.



<u>Note that</u> although we are calling this stretching, if one of the entries is negative, it might be a reflection.

Note that although we are calling this stretching, if one of the entries is negative, it might be a reflection. For example, if $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, then

$$Ax = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ -x_2 \end{pmatrix}.$$

This is a reflection about the line $x_1 = 0$.

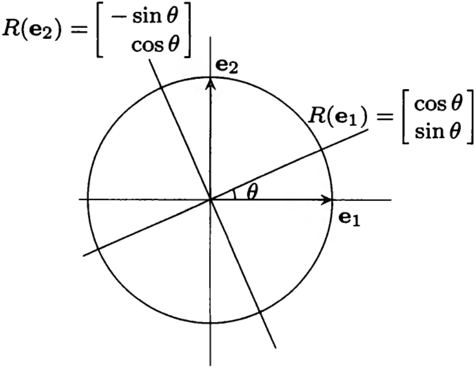
4. Rotation: The transformation given by the matrix $R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ involves rotating by an angle of θ.

4. Rotation: The transformation given by the matrix $R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ involves rotating by an angle of θ. In particular,

$$Re_1 = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix},$$

$$Re_2 = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$$

See the figure given below.



$R(\mathbf{e}_2) = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$
$\mathbf{e}_2$
$R(\mathbf{e}_1) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$
θ
$\mathbf{e}_1$

Consider an adjacency matrix:

$$\begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}.$$

Divide each (vertical) column by the sum of the entries in that column.

$$\begin{pmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1 & 0 & 0 \end{pmatrix}.$$

The resulting matrix is called the **random walk matrix**.

In this example, we look for the **meaning of the matrix product from a transformation perspective**. Let S and T be matrix transformations defined by

$$S(y) = Ay \text{ and } T(x) = Bx,$$

where

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 3 & 0 \\ 5 & -2 \\ 0 & 1 \end{pmatrix}.$$

In this example, we look for the **meaning of the matrix product from a transformation perspective**. Let S and T be matrix transformations defined by

$$S(y) = Ay \text{ and } T(x) = Bx,$$

where

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 3 & 0 \\ 5 & -2 \\ 0 & 1 \end{pmatrix}.$$

(a) What are the domains and codomains of S and T?

In this example, we look for the **meaning of the matrix product from a transformation perspective**. Let S and T be matrix transformations defined by

$$S(y) = Ay \text{ and } T(x) = Bx,$$

where

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 3 & 0 \\ 5 & -2 \\ 0 & 1 \end{pmatrix}.$$

(a) What are the domains and codomains of S and T? Why is the composite transformation $S \circ T$ defined?

In this example, we look for the **meaning of the matrix product from a transformation perspective**. Let S and T be matrix transformations defined by

$$S(y) = Ay \text{ and } T(x) = Bx,$$

where

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 3 & 0 \\ 5 & -2 \\ 0 & 1 \end{pmatrix}.$$

(a) What are the domains and codomains of S and T? Why is the composite transformation $S \circ T$ defined? What are the domain and the codomain of $S \circ T$?

(b) Let $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. Determine $T(x)$.

(b) Let $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. Determine $T(x)$.

(c) Find $(S \circ T)(x)$.

(b) Let $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. Determine $T(x)$.

(c) Find $(S \circ T)(x)$.

(d) Find a matrix C so that $(S \circ T)(x) = Cx$.

(b) Let $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. Determine $T(x)$.

(c) Find $(S \circ T)(x)$.

(d) Find a matrix C so that $(S \circ T)(x) = Cx$.

(e) Why is it reasonable to define AB to be the matrix C. Does the matrix C agree with the AB.

(b) Let $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. Determine $T(x)$.

(c) Find $(S \circ T)(x)$.

(d) Find a matrix C so that $(S \circ T)(x) = Cx$.

(e) Why is it reasonable to define AB to be the matrix C. Does the matrix C agree with the AB.

(f) Show that S, T and $S \circ T$ are linear.