## Matrix-Vector Multiplication

The phrase matrix-vector multiplication is shorthand for one of two different operations including:

Matrix-column-vector multiplication: Multiply a matrix $A \in \mathbb{R}^{m \times n}$ on the right by a column vector $\mathbf{x} \in \mathbb{R}^{n \times 1}$ to produce column vector

$$
A \mathbf{x}=\mathbf{b} \in \mathbb{R}^{m \times 1}
$$

Row-vector-matrix multiplication: Multiply a matrix $A \in \mathbb{R}^{m \times n}$ on the left by a row vector $\mathbf{x}^{T}$ where $\mathbf{x} \in \mathbb{R}^{m \times 1}$ to produce row vector

$$
\mathbf{x}^{T} A=\mathbf{b} \in \mathbb{R}^{1 \times n}
$$

We think about generating our desired output vector $\mathbf{b}$ via two different paradigms:
Vectorize the data: generate the output to our matrix-vector product as a linear combination of vectors and chunk the arithmetic operations in terms of vector-valued output.

Scalarize the data: generate the output to our matrix-vector product as a set of individual scalars by focusing on arithmetic operations that produce scalar-valued output.

In other words, we develop two different approaches for each of our two operation. This yields a total of four different definitions of matrix-vector multiplication. Using the definitions, we explore a number of algebraic properties for matrix-vector multiplication. We also explore the connections between matrix-column-vector and row-vector-matrix products that result from using the transpose operator.

We might think of each matrix-column-vector multiplication as a function from the the set of column vectors in $\mathbb{R}^{n \times 1}$ to the set of column vectors in $\mathbb{R}^{m \times 1}$. "Given" any matrix $A \in \mathbb{R}^{m \times n}$, we can define a function

$$
f(\mathbf{x})=A \mathbf{x}
$$

where $f: \mathbb{R}^{n \times 1} \rightarrow \mathbb{R}^{m \times 1}$. Notice that the domain of this function is $\mathbb{R}^{n}$ and the codomain is $\mathbb{R}^{m}$.

## Definition 4.1

## Matrix-column-vector multiplication via linear combinations

Let $A \in \mathbb{R}^{m \times n}$ be a matrix and $\mathbf{x} \in \mathbb{R}^{n \times 1}$. Then the linear combination version of the matrix-column-vector product $A \mathbf{x}=\mathbf{b} \in \mathbb{R}^{m \times 1}$ is given by

$$
A \mathbf{x}=x_{1} A(:, 1)+x_{2} A(:, 2)+\cdots+x_{n} A(:, n)=\sum_{k=1}^{n} x_{k} A(:, k)=\mathbf{b}
$$

This version of the matrix-column-vector product $A \mathrm{x}$ creates a linear combination of the columns of $A$ with scalar weights defined by the coefficients stored in the entries of vector $\mathbf{x}$. Let's take a look at a toy example to explore how this operation works in practice.

## EXAMPLE 4.1.1

Let's use the linear combination version of matrix-column-vector multiplication to find $\mathbf{b}=A \mathbf{x}$ where

$$
A=\left[\begin{array}{rrr}
-3 & 4 & -3 \\
-1 & 7 & 6 \\
0 & 1 & 2 \\
2 & -5 & -2
\end{array}\right] \quad \text { and } \quad \mathbf{x}=\left[\begin{array}{r}
5 \\
2 \\
-2
\end{array}\right]
$$

This version of matrix-column-vector multiplication dictates that we calculate the output "all at once." In other words, we're not thinking of the vector bin terms of it's individual entries. Instead, we're looking at $\mathbf{b}$ as a data structure whose value comes from a sequence arithmetic operations on vectors (rather than on scalars). For this example, the matrix $A$ has $m=4$ rows and $n=3$ columns. Using our definition of matrix-column-vector multiplication via linear combinations, we see

$$
\begin{aligned}
\mathbf{b}=A \mathbf{x} & =x_{1} A(:, 1)+x_{2} A(:, 2)+x_{3} A(:, 3) \\
& =(5)\left[\begin{array}{r}
-3 \\
-1 \\
0 \\
2
\end{array}\right]+(2)\left[\begin{array}{r}
4 \\
7 \\
1 \\
-5
\end{array}\right]+(-2)\left[\begin{array}{r}
-3 \\
6 \\
2 \\
-2
\end{array}\right]=\left[\begin{array}{r}
-1 \\
-3 \\
-2 \\
4
\end{array}\right]
\end{aligned}
$$

In your early stages of exploring this definition, I encourage you to explicitly write out the linear combination before you calculate the values for each individual entry of the output. This habit helps you develop a visceral sense of creating the output vector $\mathbf{b}$ as a linear combination of the columns of the matrix $A$ scaled by the entries in the vector $\mathbf{x}$. Later in our studies, we will think a lot about how to transform a modeling problem into an equivalent problem that is easier to solve by using as sequence of strategic steps. During those processes, it may be very helpful to view each matrix-vector product from the standpoint of linear combinations.

One of the major themes of our approach to learning linear algebra is to develop multiple representations for every mathematics idea that we study. We have just seen how to vectorize our data and use linear combinations to achieve matrix-column-vector multiplication. However, we can approach this operations from a scalar perspective. Instead of thinking about the output as a vector, we can slice the output into individual scalar entries and calculate each of those using different tools from our toolbox.

## Definition 4.2

## Matrix-column-vector multiplication via dot products

Let $A \in \mathbb{R}^{m \times n}$ and $\mathbf{x} \in \mathbb{R}^{n \times 1}$. We can define an equivalent method to find the matrix-column-vector product $A \mathbf{x}=\mathbf{b} \in \mathbb{R}^{m \times 1}$ using dot products. The $i$ th entry of the product $\mathbf{b}=A \mathbf{x}$ can be calculated by a dot product between the $i$ th row of $A$ and the vector $\mathbf{x}$ :

$$
b_{i}=\operatorname{entry}_{(i, 1)}(A \mathbf{x})=(A(i,:))^{T} \cdot \mathbf{x}=\sum_{k=1}^{n} a_{i k} x_{k}
$$

for row index $i \in\{1,2, \ldots, m\}$.

With this version of the definition, we view the product $A \mathbf{x}$ as the set of individual dot products of the rows of $A$ with the column vector $\mathbf{x}$. While it is tempting to see this second definition as "easier" since there is less to remember, I highly suggest that you use this only to check your work. The column partition version of Matrix-Vector multiplication is a much more powerful way to think about this problem. The row version is helpful only when you are actually trying to calculate by hand (and a select few other applications).

## EXAMPLE 4.1.2

Let's use the dot product version of matrix-column-vector multiplication to find $\mathbf{b}=A \mathbf{x}$ where

$$
A=\left[\begin{array}{rrr}
-3 & 4 & -3 \\
-1 & 7 & 6 \\
0 & 1 & 2 \\
2 & -5 & -2
\end{array}\right] \quad \text { and } \quad \mathbf{x}=\left[\begin{array}{r}
5 \\
2 \\
-2
\end{array}\right]
$$

This version of matrix-column-vector multiplication dictates that we calculate the output via individual entires. Said a different way, we don't visualize b as a vector but instead as a list of $m$ separate individual entries stacked on top of each other. The output is determined by a sequence of scalar arithmetic operations. The matrix
$A$ is identical to Example 4.1 .1 with $m=4$ rows and $n=3$ columns. Using our definition of matrix-column-vector multiplication via dot products combinations, we see

$$
\begin{aligned}
& b_{1}=[A(1,:)]^{T} \cdot \mathbf{x}=\left[\begin{array}{r}
-3 \\
4 \\
-3
\end{array}\right] \cdot\left[\begin{array}{r}
5 \\
2 \\
-2
\end{array}\right]=(-3)(5)+(4)(2)+(-3)(-2)=-1, \\
& b_{2}=[A(2,:)]^{T} \cdot \mathbf{x}=\left[\begin{array}{r}
-1 \\
7 \\
6
\end{array}\right] \cdot\left[\begin{array}{r}
5 \\
2 \\
-2
\end{array}\right]=(-1)(5)+(7)(2)+(6)(-2)=-3, \\
& b_{3}=[A(3,:)]^{T} \cdot \mathbf{x}=\left[\begin{array}{l}
0 \\
1 \\
2
\end{array}\right] \cdot\left[\begin{array}{r}
5 \\
2 \\
-2
\end{array}\right]=(0)(5)+(1)(2)+(2)(-2)=-2, \\
& b_{4}=[A(4,:)]^{T} \cdot \mathbf{x}=\left[\begin{array}{r}
2 \\
-5 \\
-2
\end{array}\right] \cdot\left[\begin{array}{r}
5 \\
2 \\
-2
\end{array}\right]=(2)(5)+(-5)(2)+(2)(-2)=4 .
\end{aligned}
$$

Combining the four individual scalar-valued calculations together, we form the matrix $\mathbf{b}$ with

$$
\mathbf{b}=\left[\begin{array}{l}
b_{1} \\
b_{2} \\
b_{3} \\
b_{4}
\end{array}\right]=\left[\begin{array}{r}
-1 \\
-3 \\
-2 \\
4
\end{array}\right]
$$

We notice that this is the same vector we found in Example 4.1.1. We expect this result since two different procedures to calculate the same value should produce identical outputs.

We now have two different ways to think about matrix-column-vector multiplication. Both techniques result in the same output and yet the approach differs. Let's use these definition to explore the algebraic properties of this operation.

## Theorem 1

## Algebraic properties of matrix-column-vector multiplication

Let $A \in \mathbb{R}^{m \times n}$, let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{m}$ and let $\alpha, \beta \in \mathbb{R}$. Then, we have the following:
i. Distributivity: $A(\mathbf{x}+\mathbf{y})=A \mathbf{x}+A \mathbf{y}$
ii. Scalar Multiplication: $A\left(c_{1} \mathbf{x}\right)=c_{1}(A \mathbf{x})$
iii. Linearity: $A(\alpha \mathbf{x}+\beta \mathbf{y})=\alpha A \mathbf{x}+\beta A \mathbf{y}$

The theorem statement above actually encodes three different conditional statements in the form $P \Rightarrow Q$ which are written as follows:

Distributivity: If $\underbrace{A \in \mathbb{R}^{m \times n} \text { and } \mathbf{x}, \mathbf{y} \in \mathbb{R}^{n}}_{P}$, then $\underbrace{A(\mathbf{x}+\mathbf{y})=A \mathbf{x}+A \mathbf{y}}_{Q}$

Scalar Mult: If $\underbrace{A \in \mathbb{R}^{m \times n}, \mathbf{x} \in \mathbb{R}^{n} \text {, and } \alpha \in \mathbb{R}}_{P}$, then $\underbrace{A(\alpha \mathbf{x})=\alpha A \mathbf{x}}_{Q}$

Linearity: If $\underbrace{A \in \mathbb{R}^{m \times n}, \mathbf{x}, \mathbf{y} \in \mathbb{R}^{n} \text {, and } \alpha, \beta \in \mathbb{R}}_{P}$, then $\underbrace{A(\alpha \mathbf{x}+\beta \mathbf{y})=\alpha A \mathbf{x}+\beta A \mathbf{y}}_{Q}$
For each of these statements, we might develop a formal mathematical proof using either the linear combination or dot product of matrix-column-vector multiplication. In other words, there are at least six different proofs to be explored and discovered. Later in this chapter, we'll see even more ways to prove these properties. Remember that the purpose of all the investigations we do is to help push keep you in the sweet spot. As long as you feel challenged, keep up your effort. The moment that this starts to feel easy, you're ready to move on.

Let's begin with a direct proof of distributivity using the definition of matrix-column-vector multiplication via linear combinations.

Proof. Assume that $A \in \mathbb{R}^{m \times n}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{n}$. Consider

$$
\begin{aligned}
A(\mathbf{x}+\mathbf{y}) & =[A(:, 1)|A(:, 2)| \cdots \mid A(:, n)]\left(\left[\begin{array}{c}
x_{1}+y_{1} \\
x_{2}+y_{2} \\
\vdots \\
x_{n}+y_{n}
\end{array}\right]\right) \\
& =\left(x_{1}+y_{1}\right) A(:, 1)+\left(x_{2}+y_{2}\right) A(:, 2)+\cdots+\left(x_{n}+y_{n}\right) A(:, n) \\
& =\left(x_{1} A(:, 1)+x_{2} A(:, 2)+\cdots+x_{n} A(:, n)\right) \\
& +\left(y_{1} A(:, 1)+y_{2} A(:, 2)+\cdots+y_{n} A(:, n)\right) \\
& =[A(:, 1)|A(:, 2)| \cdots \mid A(:, n)]\left[\begin{array}{c}
x_{1} \\
x_{2} \\
\vdots \\
x_{n}
\end{array}\right]+[A(:, 1)|A(:, 2)| \cdots \mid A(:, n)]\left[\begin{array}{c}
y_{1} \\
y_{2} \\
\vdots \\
y_{n}
\end{array}\right] \\
& =A \mathbf{x}+A \mathbf{y} .
\end{aligned}
$$

Let's continue with a direct proof of the scalar multiplication property using the dot product version of our definition of matrix-column-vector multiplication.

Proof. Suppose that $A \in \mathbb{R}^{m \times n}, \mathbf{x} \in \mathbb{R}^{n}$, and $\alpha \in \mathbb{R}$. For any $i=1,2, \ldots, m$, we want to show that $\operatorname{entry}_{(i, 1)}(A(\alpha \mathbf{x}))=\operatorname{entry}_{(i, 1)}(\alpha(A \mathbf{x}))$. To this end, consider

$$
\begin{aligned}
\operatorname{entry}_{(i, 1)}(A(\alpha \mathbf{x})) & =(A(i,:))^{T} \cdot(\alpha \mathbf{x}) \\
{\left[\begin{array}{c}
a_{i 1} \\
a_{i 2} \\
\vdots \\
a_{i n}
\end{array}\right] \cdot\left[\begin{array}{c}
\alpha x_{1} \\
\alpha x_{2} \\
\vdots \\
\alpha x_{n}
\end{array}\right] } & \\
& =a_{i 1}\left(\alpha x_{1}\right)+a_{i 2}\left(\alpha x_{2}\right)+\cdots+a_{i n}\left(\alpha x_{n}\right) \\
& =\alpha\left(a_{i 1} x_{1}+a_{i 2} x_{2}+\cdots+a_{i n} x_{n}\right) \\
& =\alpha\left((A(i,:))^{T} \cdot \mathbf{x}\right) \\
& =\operatorname{entry}(i, 1)(\alpha(A \mathbf{x}))
\end{aligned}
$$

This is exactly what we wanted to show.

## Definition 4.3

## Row-vector-matrix multiplication via linear combinations

Let $A \in \mathbb{R}^{m \times n}$ be a matrix and $\mathbf{x} \in \mathbb{R}^{m \times 1}$. Then the linear-combination version of the row-vector-matrix product $\mathbf{x}^{T} A=\mathbf{b} \in \mathbb{R}^{1 \times n}$ is given by the linear combination

$$
\mathbf{x}^{T} A=x_{1} A(1,:)+x_{2} A(2,:)+\cdots+x_{m} A(m,:)=\sum_{i=1}^{m} x_{i} A(i,:)=\mathbf{b}
$$

This version of the matrix-column-vector product $A \mathbf{x}$ creates a linear combination of the columns of $A$ with scalar weights defined by the coefficients stored in the entries of vector $\mathbf{x}$. Let's take a look at a toy example to explore how this operation works in practice.

## EXAMPLE 4.1.3

Let's use the linear combination version of row-vector-matrix multiplication to find $\mathbf{b}=\mathbf{x}^{T} A$ where

$$
A=\left[\begin{array}{rrr}
-3 & 4 & -3 \\
-1 & 7 & 6 \\
0 & 1 & 2 \\
2 & -5 & -2
\end{array}\right] \quad \text { and } \quad \mathbf{x}=\left[\begin{array}{r}
2 \\
-1 \\
5 \\
3
\end{array}\right]
$$

In this version of row-vector-matrix multiplication, we calculate the output using vector-valued arithmetic. Specifically, we don't think of the output vector bin terms of individual entries but instead generate this output data using a sequence arithmetic operations on row vectors. For this example, since matrix $A$ has $m=4$ rows and $n=3$ columns, we see

$$
\begin{aligned}
\mathbf{b}=\mathbf{x}^{T} A & =x_{1} A(1,:)+x_{2} A(2,:)+x_{3} A(3,:)+x_{4} A(4,:) \\
& =(2)\left[\begin{array}{rrr}
-3 & 4 & -3
\end{array}\right] \\
& +(-1)\left[\begin{array}{rrr}
-1 & 7 & 6
\end{array}\right] \\
& +(5)\left[\begin{array}{rrr}
0 & 1 & 2
\end{array}\right] \\
& +(3)\left[\begin{array}{rrr}
2 & -5 & -2
\end{array}\right] \\
& =\left[\begin{array}{rrr}
1 & -9 & -8
\end{array}\right]
\end{aligned}
$$

## Definition 4.4

## Row-vector-matrix multiplication via dot products

Let $A \in \mathbb{R}^{m \times n}$ and $\mathbf{x} \in \mathbb{R}^{m \times 1}$. We the define row-vector-matrix product $\mathbf{x}^{T} A=\mathbf{b} \in \mathbb{R}^{1 \times n}$ via dot products using an entry-by-entry approach. Specifically, the $k$ th entry of the product $\mathbf{b}=\mathbf{x}^{T} A$ can be calculated by a dot product between the $k$ th column of $A$ and the vector $\mathbf{x}$ :

$$
b_{k}=\operatorname{entry}_{(1, k)}\left(\mathbf{x}^{T} A\right)=\mathbf{x} \cdot(A(:, k))=\sum_{i=1}^{m} a_{i k} x_{k}
$$

for column index $k \in\{1,2, \ldots, n\}$.

## EXAMPLE 4.1.4

Let's use our definition of row-vector-matrix multiplication via dot products to find $\mathbf{b}=\mathbf{x}^{T} A$ where

$$
A=\left[\begin{array}{rrr}
-3 & 4 & -3 \\
-1 & 7 & 6 \\
0 & 1 & 2 \\
2 & -5 & -2
\end{array}\right] \quad \text { and } \quad \mathbf{x}=\left[\begin{array}{r}
2 \\
-1 \\
5 \\
3
\end{array}\right]
$$

In this case, we'll generate $\mathbf{b}$ entry-by-entry by using the dot product operation. The matrix $A$ is identical to Example 4.1.1 with $m=4$ rows and $n=3$ columns. Using our definition, we see

$$
\begin{aligned}
& b_{1}=\mathbf{x} \cdot A(:, 1)=\left[\begin{array}{r}
2 \\
-1 \\
5 \\
3
\end{array}\right] \cdot\left[\begin{array}{r}
-3 \\
-1 \\
0 \\
2
\end{array}\right]=(2)(-3)+(-1)(-1)+(5)(0)+(3)(2)=1 \\
& b_{2}=\mathbf{x} \cdot A(:, 2)=\left[\begin{array}{r}
2 \\
-1 \\
5 \\
3
\end{array}\right] \cdot\left[\begin{array}{r}
4 \\
7 \\
1 \\
-5
\end{array}\right]=(2)(4)+(-1)(7)+(5)(1)+(3)(-5)=-9 \\
& b_{3}=\mathbf{x} \cdot A(:, 3)=\left[\begin{array}{r}
2 \\
-1 \\
5 \\
3
\end{array}\right] \cdot\left[\begin{array}{r}
-3 \\
6 \\
2 \\
-2
\end{array}\right]=(2)(-3)+(-1)(6)+(5)(2)+(3)(-2)=-8
\end{aligned}
$$

Placing each of these into the appropriate column, we form the output vector

$$
\mathbf{b}=\left[\begin{array}{lll}
b_{1} & b_{2} & b_{3}
\end{array}\right]=\left[\begin{array}{lll}
1 & -9 & -8
\end{array}\right]
$$

This is the same output we found in Example 4.1 .3 verifying that the two different procedures to calculate the row-vector-matrix multiplication result in identical outputs.

Compare and contrast our two definitions for both matrix-column-vector multiplication and row-vector-matrix multiplication. Notice that we've developed to different approaches for both operations that yield the same result but do so via different underlying approaches. Let's develop algebraic properties for row-vectormatrix multiplication and use our definitions to explore.

## Theorem 2

## Algebraic properties of row-vector-matrix multiplication

Let $A \in \mathbb{R}^{m \times n}$, let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{m \times 1}$ and let $\alpha, \beta \in \mathbb{R}$. Then, each of the following laws holds true:
i. Distributivity: $(\mathbf{x}+\mathbf{y})^{T} A=\mathbf{x}^{T} A+\mathbf{y}^{T} A$
ii. Scalar Multiplication: $(\alpha \mathbf{x})^{T} A=\alpha\left(\mathbf{x}^{T} A\right)$
iii. Linearity: $(\alpha \mathbf{x}+\beta \mathbf{y})^{T} A=\alpha \mathbf{x}^{T} A+\beta \mathbf{y}^{T} A$

Once again, there are actually three distinct propositions encoded in this statement. For each of those propositions, we might generate a proof based on either the dot product or linear combination definition for row-vector-matrix multiplication. This yields a least six different proofs available for discovery. Since we've already seen the basic structure of these proof in our work with Theorem 1, let's develop an alternate approach for proving these properties are true.

To do this, let's make a conjecture. Specifically, perhaps we notice that the properties are interrelated and make the following claims:


Conjecture 2: If $\underbrace{\text { linearity }}_{P}$, then $\underbrace{\text { distributivity and scalar mult. }}_{Q}$
Let's show that conjecture 1 is true.
Proof. Let $A \in \mathbb{R}^{m \times n}$, let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{m \times 1}$ and let $\alpha, \beta \in \mathbb{R}$. Assume that properties i. and ii. hold for any vectors and scalars. Now consider

$$
\begin{aligned}
(\alpha \mathbf{x}+\beta \mathbf{y})^{T} A & =((\alpha \mathbf{x})+(\beta \mathbf{y}))^{T} A \\
& =(\alpha \mathbf{x})^{T} A+(\beta \mathbf{y})^{T} A \\
& =\alpha\left(\mathbf{x}^{T} A\right)+\beta\left(\mathbf{y}^{T} A\right)
\end{aligned}
$$

This is what we wanted to show.

Next let's show that conjecture 2 is true.
Proof. Let $A \in \mathbb{R}^{m \times n}$, let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{m \times 1}$ and let $\alpha, \beta \in \mathbb{R}$. Assume that propery iii. holds for all vectors and scalars. We'll show that properties i. and ii. follow. To this end, consider

$$
\begin{array}{rlr}
(\mathbf{x}+\mathbf{y})^{T} A & =(\alpha \mathbf{x}+\beta \mathbf{y})^{T} A & \text { where } \alpha=1=\beta \\
& =\alpha\left(\mathbf{x}^{T} A\right)+\beta\left(\mathbf{y}^{T} A\right) & \\
& =\mathbf{x}^{T} A+\mathbf{y}^{T} A . &
\end{array}
$$

This establishes property i. follows from property iii. Let's show property ii. must also follow. Consider

$$
\begin{array}{rlrl}
(\alpha \mathbf{x})^{T} A & =(\alpha \mathbf{x}+\beta \mathbf{y})^{T} A & \text { where } \beta=0 \\
& =\alpha\left(\mathbf{x}^{T} A\right)+\beta\left(\mathbf{y}^{T} A\right) & \\
& =\alpha\left(\mathbf{x}^{T} A\right) & &
\end{array}
$$

These proofs show that if we establish linearity we get the other two properties for free and vice versa. To end this exploration, let's prove linearity based on the linear combination definition of row-vector matrix multiplication.

Proof. Let $A \in \mathbb{R}^{m \times n}$, let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{m \times 1}$ and let $\alpha, \beta \in \mathbb{R}$. Consider

$$
\begin{aligned}
(\alpha \mathbf{x}+\beta \mathbf{y})^{T} A & =\left[\alpha x_{1}+\beta y_{1}\left|\alpha x_{2}+\beta y_{2}\right| \cdots \mid \alpha x_{m}+\beta y_{m}\right]\left[\begin{array}{c}
A(1,:) \\
A(2,:) \\
\vdots \\
A(m,:)
\end{array}\right] \\
& =\left(\alpha x_{1}+\beta y_{1}\right) A(1,:)+\left(\alpha x_{2}+\beta y_{2}\right) A(2,:)+\cdots+\left(\alpha x_{m}+\beta y_{m}\right) A(m,:) \\
& =\alpha\left(x_{1} A(1,:)+x_{2} A(2,:)+\cdots+x_{m} A(m,:)\right) \\
& +\beta\left(y_{1} A(1,:)+y_{2} A(2,:)+\cdots+y_{m} A(m,:)\right) \\
& =\alpha\left[x_{1}\left|x_{2}\right| \cdots \mid x_{m}\right]\left[\begin{array}{c}
A(1,:) \\
A(2,:) \\
\vdots \\
A(m,:)
\end{array}\right]+\beta\left[y_{1}\left|y_{2}\right| \cdots \mid y_{m}\right]\left[\begin{array}{c}
A(1,:) \\
A(2,:) \\
\vdots \\
A(m,:)
\end{array}\right] \\
& =\alpha \mathbf{x}^{T} A+\beta \mathbf{y}^{T} A .
\end{aligned}
$$

This is exactly what we wanted to show.

## Theorem 3

## The Transpose of a Matrix-Vector Product

Let $A \in \mathbb{R}^{m \times n}$, let $\mathbf{x} \in \mathbb{R}^{n \times 1}$, and let $\mathbf{y} \in \mathbb{R}^{m \times 1}$. Then, each of the following holds:
i. $(A \mathbf{x})^{T}=\mathbf{x}^{T} A^{T}$
ii. $\left(\mathbf{y}^{T} A\right)^{T}=A^{T} \mathbf{y}$

As we've noted previously, formal theorem statements found in math textbooks often present a number of statements in the form $P \Rightarrow Q$. A best practice for building deep understanding of mathematics theorems is to identify every conditional statement you can find in a particular theorem statement separately. Below we do exactly that:

$$
\text { Proposition 1: If } \underbrace{A \in \mathbb{R}^{m \times n} \text { and } \mathbf{x} \in \mathbb{R}^{n}}_{P} \text {, then } \underbrace{(A \mathbf{x})^{T}=\mathbf{x}^{T} A^{T}}_{Q}
$$

$$
\text { Proposition 2: If } \underbrace{A \in \mathbb{R}^{m \times n} \text { and } \mathbf{y} \in \mathbb{R}^{m}}_{P} \text {, then } \underbrace{\left(\mathbf{y}^{T} A\right)^{T}=A^{T} \mathbf{y}}_{Q}
$$

In this work, let's develop two different proofs for proposition 1. Our first proof relies on the definition of matrix-column-vector multiplication via dot products while the second proof leverages the definition of matrix-column-vector multiplication via linear combinations. Since this is the last theorem of the section, we present the general proofs without creating base-case examples.

Proof. Let $A \in \mathbb{R}^{m \times n}$ and $\mathbf{x} \in \mathbb{R}^{n \times 1}$. Now define two different vectors

$$
\mathbf{b}=A \mathbf{x} \quad \text { and } \mathbf{r} \quad=\mathbf{x}^{T} A^{T}
$$

We want to prove that $\mathbf{b}^{T}=\mathbf{r}$. In this proof, we establish this fact via an entry-byentry approach. In other words, we show that for $k=1,2, \ldots, m$, we have $b_{k}=r_{k}$.

To this end, recall that our definition of matrix-column-vector multiplication via dot products gives us that

$$
\begin{aligned}
b_{k}=\operatorname{entry}_{(1, k)}\left(\mathbf{b}^{T}\right) & =\operatorname{entry}_{(k, 1)}(\mathbf{b}) \\
& =\operatorname{entry}_{(k, 1)}(A \mathbf{x}) \\
& \left.=\operatorname{(row}_{k}(A)\right)^{T} \cdot \mathbf{x} \\
& =(A(k,:))^{T} \cdot \mathbf{x} \\
& =\sum_{j=1}^{n} a_{k j} x_{j}
\end{aligned}
$$

In this summation, each individual term is the product between two scalars. By our knowledge of algebra, we know that this product is commutative and we can switch the order in which we execute the multiplication in each term. In other words, we have

$$
\begin{aligned}
b_{k} & =\sum_{j=1}^{n} x_{j} a_{k j} \\
& =\mathbf{x} \cdot(A(k,:))^{T} \\
& =\mathbf{x} \cdot\left(A^{T}(:, k)\right) \\
& =\operatorname{entry}_{(1, k)}\left(\mathbf{x}^{T} A^{T}\right)=r_{k}
\end{aligned}
$$

This is exactly what we wanted to show and we conclude that $(A \mathbf{x})^{T}=\mathbf{x}^{T} A^{T}$.
Our first proof above is based on the entry-by-entry view of matrix-columnvector multiplication. I like to say that one of the best ways to engage in deep learning of mathematical concepts is to search for, practice, and develop multiple representations of every mathematical idea we study. In this case, let's explore another proof from the vector perspective by using the linear combination version of our definition.

Proof. Let $A \in \mathbb{R}^{m \times n}$ and $\mathbf{x} \in \mathbb{R}^{n \times 1}$. Now define two different vectors

$$
\mathbf{b}=A \mathbf{x} \quad \text { and } \mathbf{r} \quad=\mathbf{x}^{T} A^{T}
$$

We want to prove that $\mathbf{b}^{T}=\mathbf{r}$. In this proof, we establish this fact by relying on vector operations. To this end, recall that our definition of matrix-column-vector multiplication via dot products gives us that

$$
\begin{aligned}
\mathbf{b}^{T}=(A \mathbf{x})^{T} & =\left(\sum_{k=1}^{n} x_{k} A(:, k)\right)^{T} \\
& =\sum_{k=1}^{n}\left(x_{k} A(:, k)\right)^{T} \\
& =\sum_{k=1}^{n} x_{k}(A(:, k))^{T} \\
& =\sum_{k=1}^{n} x_{k} A^{T}(k,:) \\
& =\mathbf{x}^{T} A^{T}
\end{aligned}
$$

This is exactly what we wanted to show and we conclude that $(A \mathbf{x})^{T}=\mathbf{x}^{T} A^{T}$.

## Matrix-Vector Multiplication Problem Set

1. Let $\mathbf{e}_{i} \in \mathbb{R}^{4}$ be the $i$ th elementary basis vector with 4 rows and 1 column, for $i \in\{1,2,3,4\}$. For example, $\mathbf{e}_{2}=\left[\begin{array}{llll}0 & 1 & 0 & 0\end{array}\right]^{T}$. Let

$$
A=\left[\begin{array}{rrrr}
1 & 0 & -1 & 2 \\
2 & 1 & -3 & 4 \\
0 & 2 & -2 & 3 \\
1 & 1 & -4 & -2
\end{array}\right]
$$

A. Calculate the matrix-column-vector product $A \mathbf{e}_{1}, A \mathbf{e}_{2}, A \mathbf{e}_{3}$ and $A \mathbf{e}_{4}$. For each product, use both the linear combination AND dot product version of matrix-column-vector multiplication.
B. Make sense of the output you get from problem 1A above. What patterns do you notice? Describe these patterns in abuelita language and explain why this must be true.
C. Calculate the row-vector-matrix product $\mathbf{e}_{1}^{T} A, \mathbf{e}_{2}^{T} A, \mathbf{e}_{3}^{T} A$, and $\mathbf{e}_{4}^{T} A$ For each product, use both the linear combination AND dot product version of row-vector-matrix multiplication.
D. Make sense of the output you get from problem 1C above. What patterns do you notice? Compare and contrast this to your work on problems 1A and 1B above? Describe these patterns in abuelita language.
E. Using results from part (A) through (D) above, create a vector $\mathbf{x}$ such that $A \mathbf{x}=A(:, 1)-A(:, 4)$.
2. Let $A \in \mathbb{R}^{4 \times 5}$. Multiply $A$ on the right by a column vector $\mathbf{x} \in \mathbb{R}^{5}$ to achieve each of the operations below. In each case, specifically state the entry-by-entry definition of the column vector $\mathbf{x}$ used to accomplish these operations.
A. Select column 3
B. Triple column 5
C. Subtract column 2 from column 4
D. Scale column 1 by $c=\frac{1}{3}$ and add this to column 3
E. Find scalar $c$ such that the sum of $c$ times column 4 plus column 1 has a zero in the first entry.
3. Let $A \in \mathbb{R}^{4 \times 5}$. Multiply $A$ on the left by a row vector $\mathbf{x}^{T} \in \mathbb{R}^{1 \times 4}$ to achieve each of the operations below. In each case, specifically state the entry-by-entry definition of the row vector $\mathbf{x}$ used to accomplish these operations.
A. Multiply row 3 by $c=-4$
B. Select row 2
C. Subtract row 2 from row 4
D. Scale row 2 by $c=\frac{5}{2}$ and add this to row 3
E. Find scalar $c$ such that the sum of $c$ times row 1 plus row 4 has a zero in the first entry.
4. How many different proofs for theorems 1,2 , or 3 can you create? If needed, create specific examples to help get insights into the general proofs. Adapt your work on the specific cases to create general proofs. How is each proof related to our various methods for performing matrix-vector multiplication?
5. Let $C \in \mathbb{R}^{4 \times 4}$ be a diagonal matrix. Suppose that $\mathbf{x} \in \mathbb{R}^{4 \times 1}$. Consider the quadratic form

$$
q(\mathbf{x})=\mathbf{x}^{T} \cdot C \cdot \mathbf{x}
$$

A. Identify the domain and codomain of the function $q(\mathbf{x})$
B. Write a scalar-based algorithm to find the outputs of $q(\mathbf{x})$ (hint: this definition should be in terms of entries of $\mathbf{x}$ and entries of $C$ )
C. If you know that $c_{i i} \geq 0$ for all $i \in[4]$, what can you say about the range of $q(\mathbf{x})$ ?
D. If you know that $c_{i i} \leq 0$ for all $i \in[4]$, what can you say about the range of $q(\mathbf{x})$ ?
E. Consider the so-called "Raleigh Quotient" function

$$
R(\mathbf{x})=\frac{\mathbf{x}^{T} \cdot C \cdot \mathbf{x}}{\mathbf{x}^{T} \cdot \mathbf{x}}
$$

What is the maximum value of $R(\mathbf{x})$ ? What is the minimum value of $R(\mathbf{x})$ ? How are these optimum values related to the entries of $C$ ?

