I.1 Multiplication $Ax$ Using Columns of $A$

We hope you already know some linear algebra. It is a beautiful subject—more useful to more people than calculus (in our quiet opinion). But even old-style linear algebra courses miss basic and important facts. This first section of the book is about matrix-vector multiplication $Ax$ and the column space of a matrix and the rank.

We always use examples to make our point clear.

**Example 1** Multiply $A$ times $x$ using the three rows of $A$. Then use the two columns:

\[
\begin{bmatrix}
2 & 3 \\
2 & 4 \\
3 & 7 \\
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
\end{bmatrix}
= 
\begin{bmatrix}
x_1 + 3x_2 \\
x_1 + 4x_2 \\
3x_1 + 7x_2 \\
\end{bmatrix}
\]

**By rows**

\[
\begin{bmatrix}
2 & 3 \\
2 & 4 \\
3 & 7 \\
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
\end{bmatrix}
= 
(x_1, x_2)
\]

**By columns**

\[
\begin{bmatrix}
2 & 3 \\
2 & 4 \\
3 & 7 \\
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
\end{bmatrix}
= 
\begin{bmatrix}
x_1 \\
x_2 \\
\end{bmatrix}
\begin{bmatrix}
2 \\
3 \\
\end{bmatrix}
+ 
\begin{bmatrix}
x_1 \\
x_2 \\
\end{bmatrix}
\begin{bmatrix}
4 \\
7 \\
\end{bmatrix}
= 
\begin{bmatrix}
2x_1 + 3x_2 \\
2x_1 + 4x_2 \\
3x_1 + 7x_2 \\
\end{bmatrix}
\]

You see that both ways give the same result. The first way (a row at a time) produces three inner products. Those are also known as “dot products” because of the dot notation:

\[
\text{row} \cdot \text{column} = (2, 3) \cdot (x_1, x_2) = 2x_1 + 3x_2
\]

This is the way to find the three separate components of $Ax$. We use this for computing—but not for understanding. It is low level. Understanding is higher level, using vectors.

The vector approach sees $Ax$ as a “linear combination” of $a_1$ and $a_2$. This is the fundamental operation of linear algebra! A linear combination of $a_1$ and $a_2$ includes two steps:

1. Multiply the columns $a_1$ and $a_2$ by “scalars” $x_1$ and $x_2$

2. Add vectors $x_1a_1 + x_2a_2 = Ax$.

Thus $Ax$ is a linear combination of the columns of $A$. This is fundamental.

This thinking leads us to the column space of $A$. The key idea is to take all combinations of the columns. All real numbers $x_1$ and $x_2$ are allowed—the space includes $Ax$ for all vectors $x$. In this way we get infinitely many output vectors $Ax$. And we can see those outputs geometrically.

In our example, each $Ax$ is a vector in 3-dimensional space. That 3D space is called $\mathbb{R}^3$. (The $\mathbb{R}$ indicates real numbers. Vectors with three complex components lie in the space $\mathbb{C}^3$.) We stay with real vectors and we ask this key question:

All combinations $Ax = x_1a_1 + x_2a_2$ produce what part of the full 3D space?

Answer: Those vectors produce a plane. The plane contains the complete line in the direction of $a_1 = (2, 2, 3)$, since every vector $x_1a_1$ is included. The plane also includes the line of all vectors $x_2a_2$ in the direction of $a_2$. And it includes the sum of any vector on one line plus any vector on the other line. **This addition fills out an infinite plane containing the two lines.** But it does not fill out the whole 3-dimensional space $\mathbb{R}^3$. 
I.1. Multiplication $Ax$ Using Columns of $A$

Definition  The combinations of the columns fill out the column space of $A$.

Here the column space is a plane. That plane includes the zero point $(0, 0, 0)$ which is produced when $x_1 = x_2 = 0$. The plane includes $(5, 6, 10) = a_1 + a_2$ and $(-1, -2, -4) = a_1 - a_2$. Every combination $x_1a_1 + x_2a_2$ is in this column space. With probability 1 it does not include the random point rand$(3, 1)$! Which points are in the plane?

\[ b = (b_1, b_2, b_3) \] is in the column space of $A$ exactly when $Ax = b$ has a solution $(x_1, x_2)$.

When you see that truth, you understand the column space $\mathbf{C}(A)$: The solution $x$ shows how to express the right side $b$ as a combination $x_1a_1 + x_2a_2$ of the columns. For some $b$ this is impossible—they are not in the column space.

Example 2  $b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is not in $\mathbf{C}(A)$. $Ax = \begin{bmatrix} 2x_1 + 3x_2 \\ 2x_1 + 4x_2 \\ 3x_1 + 7x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is unsolvable.

The first two equations force $x_1 = \frac{1}{2}$ and $x_2 = 0$. Then equation 3 fails: $3\left(\frac{1}{2}\right) + 7(0) = 1.5$ (not 1). This means that $b = (1, 1, 1)$ is not in the column space—the plane of $a_1$ and $a_2$.

Example 3  What are the column spaces of $A_2 = \begin{bmatrix} 2 & 3 & 5 \\ 2 & 4 & 6 \\ 3 & 7 & 10 \end{bmatrix}$ and $A_3 = \begin{bmatrix} 2 & 3 & 1 \\ 2 & 4 & 1 \\ 3 & 7 & 1 \end{bmatrix}$?

Solution.  The column space of $A_2$ is the same plane as before. The new column $(5, 6, 10)$ is the sum of column 1 + column 2. So $a_3 = \text{column 3}$ is already in the plane and adds nothing new. By including this “dependent” column we don’t go beyond the original plane.

The column space of $A_3$ is the whole 3D space $\mathbb{R}^3$. Example 2 showed us that the new third column $(1, 1, 1)$ is not in the plane $\mathbf{C}(A)$. Our column space $\mathbf{C}(A_3)$ has grown bigger. But there is nowhere to stop between a plane and the full 3D space. Visualize the $x-y$ plane and a third vector $(x_3, y_3, z_3)$ out of the plane (meaning that $z_3 \neq 0$). They combine to give every vector in $\mathbb{R}^3$.

Here is a total list of all possible column spaces inside $\mathbb{R}^3$. Dimensions 0, 1, 2, 3:

Subspaces of $\mathbb{R}^3$
- The zero vector $(0, 0, 0)$ by itself
- A line of all vectors $x_1a_1$
- A plane of all vectors $x_1a_1 + x_2a_2$
- The whole $\mathbb{R}^3$ with all vectors $x_1a_1 + x_2a_2 + x_3a_3$

In that list we need the vectors $a_1, a_2, a_3$ to be “independent”. The only combination that gives the zero vector is $0a_1 + 0a_2 + 0a_3$. So $a_1$ by itself gives a line, $a_1$ and $a_2$ give a plane, $a_1$ and $a_2$ and $a_3$ give every vector $b$ in $\mathbb{R}^3$. The zero vector is in every subspace! In linear algebra language:

- Three independent columns in $\mathbb{R}^3$ produce an invertible matrix: $AA^{-1} = A^{-1}A = I$.
- $Ax = 0$ requires $x = (0, 0, 0)$. Then $Ax = b$ has exactly one solution $x = A^{-1}b$.

You see the picture for the columns of an $n$ by $n$ invertible matrix. Their combinations fill its column space: all of $\mathbb{R}^n$. We needed those ideas and that language to go further.
Independent Columns and the Rank of $A$

After writing those words, I thought this short section was complete. Wrong. With just a small effort, we can find a basis for the column space of $A$, we can factor $A$ into $C$ times $R$, and we can prove the first great theorem in linear algebra. You will see the rank of a matrix and the dimension of a subspace.

All this comes with an understanding of independence. The goal is to create a matrix $C$ whose columns come directly from $A$—but not to include any column that is a combination of previous columns. The columns of $C$ (as many as possible) will be “independent”. Here is a natural construction of $C$ from the $n$ columns of $A$:

- If column 1 of $A$ is not all zero, put it into the matrix $C$.
- If column 2 of $A$ is not a multiple of column 1, put it into $C$.
- If column 3 of $A$ is not a combination of columns 1 and 2, put it into $C$. Continue.

At the end $C$ will have $r$ columns ($r \leq n$).
They will be a “basis” for the column space of $A$.

The left out columns are combinations of those basic columns in $C$.

A basis for a subspace is a full set of independent vectors: All vectors in the space are combinations of the basis vectors. Examples will make the point.

Example 4  If $A = \begin{bmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{bmatrix}$ then $C = \begin{bmatrix} 1 & 3 \\ 1 & 2 \\ 0 & 1 \end{bmatrix}$, $n = 3$ columns in $A$, $r = 2$ columns in $C$.

Column 3 of $A$ is $2$ (column 1) + $2$ (column 2). Leave it out of the basis in $C$.

Example 5  If $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$ then $C = A$, $n = 3$ columns in $A$, $r = 3$ columns in $C$.

This matrix $A$ is invertible. Its column space is all of $\mathbb{R}^3$. Keep all 3 columns.

Example 6  If $A = \begin{bmatrix} 1 & 2 & 5 \\ 1 & 2 & 5 \\ 1 & 2 & 5 \end{bmatrix}$ then $C = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $n = 3$ columns in $A$, $r = 1$ column in $C$.

The number $r$ is the “rank” of $A$. It is also the rank of $C$. It counts independent columns. Admittedly we could have moved from right to left in $A$, starting with its last column. This would not change the final count $r$. Different basis, but always the same number of vectors. That number $r$ is the “dimension” of the column space of $A$ and $C$ (same space).

The rank of a matrix is the dimension of its column space.
The matrix $C$ connects to $A$ by a third matrix $R$: $A = CR$. Their shapes are $(m \text{ by } n) = (m \text{ by } r)(r \text{ by } n)$. I can show this “factorization of $A$” in Example 4 above:

$$A = \begin{bmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \end{bmatrix} = CR \quad (2)$$

When $C$ multiplies the first column $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ of $R$, this produces column 1 of $C$ and $A$.

When $C$ multiplies the second column $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ of $R$, we get column 2 of $C$ and $A$.

When $C$ multiplies the third column $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$ of $R$, we get $2(\text{column 1}) + 2(\text{column 2})$.

This matches column 3 of $A$. All we are doing is to put the right numbers in $R$. Combinations of the columns of $C$ produce the columns of $A$. Then $A = CR$ stores this information as a matrix multiplication. Actually $R$ is a famous matrix in linear algebra:

$$R = \text{rref}(A) = \text{row-reduced echelon form of } A \text{ (without zero rows)}.$$

Example 5 has $C = A$ and then $R = I$ (identity matrix). Example 6 has only one column in $C$, so it has one row in $R$:

$$A = \begin{bmatrix} 1 & 2 & 5 \\ 1 & 2 & 5 \\ 1 & 2 & 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 5 \end{bmatrix} = CR \quad \text{All three matrices have rank } r = 1 \quad \text{Column Rank} = \text{Row Rank}

The number of independent columns equals the number of independent rows

This rank theorem is true for every matrix. Always columns and rows in linear algebra! The $m$ rows contain the same numbers $a_{ij}$ as the $n$ columns. But different vectors.

The theorem is proved by $A = CR$. Look at that differently—by rows instead of columns. The matrix $R$ has $r$ rows. **Multiplying by $C$ takes combinations of those rows.** Since $A = CR$, we get every row of $A$ from the $r$ rows of $R$. And those $r$ rows are independent, so they are a **basis for the row space of $A$**. The column space and row space of $A$ both have dimension $r$, with $r$ basis vectors—columns of $C$ and rows of $R$.

One minute: Why does $R$ have independent rows? Look again at Example 4.

$$A = \begin{bmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \end{bmatrix} \leftarrow \text{independent rows of } R$$

It is those ones and zeros in $R$ that tell me: No row is a combination of the other rows.

The big factorization for data science is the “SVD” of $A$—when the first factor $C$ has $r$ orthogonal columns and the second factor $R$ has $r$ orthogonal rows.
Problem Set I.1

1. Give an example where a combination of three nonzero vectors in \( \mathbb{R}^4 \) is the zero vector. Then write your example in the form \( Ax = 0 \). What are the shapes of \( A \) and \( x \) and \( 0 \)?

2. Suppose a combination of the columns of \( A \) equals a different combination of those columns. Write that as \( Ax = Ay \). Find two combinations of the columns of \( A \) that equal the zero vector (in matrix language, find two solutions to \( Az = 0 \)).

3. (Practice with subscripts) The vectors \( a_1, a_2, \ldots, a_n \) are in \( m \)-dimensional space \( \mathbb{R}^m \), and a combination \( c_1 a_1 + \cdots + c_n a_n \) is the zero vector. That statement is at the vector level.
   
   (1) Write that statement at the matrix level. Use the matrix \( A \) with the \( a \)'s in its columns and use the column vector \( c = (c_1, \ldots, c_n) \).
   
   (2) Write that statement at the scalar level. Use subscripts and sigma notation to add up numbers. The column vector \( a_j \) has components \( a_{1j}, a_{2j}, \ldots, a_{mj} \).

4. Suppose \( A \) is the 3 by 3 matrix \( \text{ones}(3,3) \) of all ones. Find two independent vectors \( x \) and \( y \) that solve \( Ax = 0 \) and \( Ay = 0 \). Write that first equation \( Ax = 0 \) (with numbers) as a combination of the columns of \( A \). Why don’t I ask for a third independent vector with \( Az = 0 \)?

5. The linear combinations of \( v = (1, 1, 0) \) and \( w = (0, 1, 1) \) fill a plane in \( \mathbb{R}^3 \).
   
   (a) Find a vector \( z \) that is perpendicular to \( v \) and \( w \). Then \( z \) is perpendicular to every vector \( cv + dw \) on the plane: \( (cv + dw)^T z = cv^T z + dw^T z = 0 + 0 \).
   
   (b) Find a vector \( u \) that is not on the plane. Check that \( u^T z \neq 0 \).

6. If three corners of a parallelogram are \((1, 1), (4, 2), \) and \((1, 3)\), what are all three of the possible fourth corners? Draw two of them.

7. Describe the column space of \( A = [v \ w \ v + 2w] \). Describe the nullspace of \( A \): all vectors \( x = (x_1, x_2, x_3) \) that solve \( Ax = 0 \). Add the “dimensions” of that plane (the column space of \( A \)) and that line (the nullspace of \( A \)):

   \[
   \text{dimension of column space} + \text{dimension of nullspace} = \text{number of columns}
   \]

8. \( A = CR \) is a representation of the columns of \( A \) in the basis formed by the columns of \( C \) with coefficients in \( R \). If \( A_{ij} = j^2 \) is 3 by 3, write down \( A \) and \( C \) and \( R \).

9. Suppose the column space of an \( m \) by \( n \) matrix is all of \( \mathbb{R}^3 \). What can you say about \( m \)? What can you say about \( n \)? What can you say about the rank \( r \)?
10. Find the matrices $C_1$ and $C_2$ containing independent columns of $A_1$ and $A_2$:

$$A_1 = \begin{bmatrix} 1 & 3 & -2 \\ 3 & 9 & -6 \\ 2 & 6 & -4 \end{bmatrix} \quad A_2 = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

11. Factor each of those matrices into $A = CR$. The matrix $R$ will contain the numbers that multiply columns of $C$ to recover columns of $A$.

This is one way to look at matrix multiplication: $C$ times each column of $R$.

12. Produce a basis for the column spaces of $A_1$ and $A_2$. What are the dimensions of those column spaces—the number of independent vectors? What are the ranks of $A_1$ and $A_2$? How many independent rows in $A_1$ and $A_2$?

13. Create a 4 by 4 matrix $A$ of rank 2. What shapes are $C$ and $R$?

14. Suppose two matrices $A$ and $B$ have the same column space.

(a) Show that their row spaces can be different.

(b) Show that the matrices $C$ (basic columns) can be different.

(c) What number will be the same for $A$ and $B$?

15. If $A = CR$, the first row of $A$ is a combination of the rows of $R$. Which part of which matrix holds the coefficients in that combination—the numbers that multiply the rows of $R$ to produce row 1 of $A$?

16. The rows of $R$ are a basis for the row space of $A$. What does that sentence mean?

17. For these matrices with square blocks, find $A = CR$. What ranks?

$$A_1 = \begin{bmatrix} \text{zeros} & \text{ones} \\ \text{ones} & \text{ones} \end{bmatrix}_{4 \times 4} \quad A_2 = \begin{bmatrix} A_1 \\ A_1 \end{bmatrix}_{8 \times 4} \quad A_3 = \begin{bmatrix} A_1 & A_1 \\ A_1 & A_1 \end{bmatrix}_{8 \times 8}$$

18. If $A = CR$, what are the $CR$ factors of the matrix $\begin{bmatrix} 0 & A \\ 0 & A \end{bmatrix}$?

19. “Elimination” subtracts a number $\ell_{ij}$ times row $j$ from row $i$: a “row operation.” Show how those steps can reduce the matrix $A$ in Example 4 to $R$ (except that this row echelon form $R$ has a row of zeros). The rank won’t change!

$$A = \begin{bmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{bmatrix} \rightarrow \rightarrow R = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} = \text{rref}(A).$$
This page is about the factorization $A = CR$ and its close relative $A = CMR$. $C$ has the same $r$ independent columns taken from $A$. The new matrix $R$ has $r$ independent rows, also taken directly from $A$. The $r$ by $r$ “mixing matrix” is $M$. This invertible matrix makes $A = CMR$ a true equation. Here is an example:

$$
\begin{bmatrix}
1 & 2 & 3 \\
2 & 4 & 6 \\
3 & 5 & 8
\end{bmatrix}
= 
\begin{bmatrix}
1 & 2 \\
2 & 4 \\
3 & 5
\end{bmatrix}
\begin{bmatrix}
-5 & 2 \\
3 & -1 \\
3 & 5
\end{bmatrix}
= CMR
$$

How did we find that mixing matrix $M$? We realized that the matrix $\begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}$ is in both $C$ and $R$. It is the overlap of the independent columns 1, 2 and independent rows 1, 3. Then the correct mixing matrix $M$ is the inverse of this 2 by 2 overlap matrix $M^{-1}$:

$$
MM^{-1} = 
\begin{bmatrix}
-5 & 2 \\
3 & -1
\end{bmatrix}
\begin{bmatrix}
1 & 2 \\
3 & 5
\end{bmatrix}
= 
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}.
$$

Here are extra problems to give practice with all these rectangular matrices of rank $r$.

20 Find $A = CR$ ($R$ contains $I$) and also $A = CMR$ for these matrices.

$$
A = 
\begin{bmatrix}
2 & 4 \\
3 & 6
\end{bmatrix}
(M \text{ is 1 by 1})
A = 
\begin{bmatrix}
1 & 2 & 1 \\
2 & 4 & 4 \\
3 & 6 & 5
\end{bmatrix}
(M \text{ is 2 by 2})
$$

21 To find a general formula for $M$, multiply $A = CMR$ by $C^T$ on the left and $R^T$ on the right. Then multiply by $(C^T C)^{-1}$ on the left and $(R R^T)^{-1}$ on the right. This leaves the formula for $M$ that was in earlier printings of this book.

**Inverse of a 2 by 2 matrix**

No inverse if $ad = bc$

$$
\begin{bmatrix}
a & b \\
c & d
\end{bmatrix}^{-1}
= 
\frac{1}{ad - bc}
\begin{bmatrix}
d & -b \\
-c & a
\end{bmatrix}
$$

(∗∗)

22 Show that this formula (∗∗) breaks down if

$$
\begin{bmatrix}
b \\
d
\end{bmatrix}
= m
\begin{bmatrix}
a \\
c
\end{bmatrix}:	ext{ dependent columns.}
$$

The reason for this page is that the factorizations $A = CR$ and $A = CMR$ have jumped forward in importance for large matrices. When $C$ takes columns directly from $A$, and $R$ takes rows directly from $A$, those matrices preserve properties that are lost in the more famous $QR$ and SVD factorizations. Where $A = QR$ and $A = U \Sigma V^T$ involve orthogonalizing the vectors, $C$ and $R$ keep the original data:

If $A$ is nonnegative, so are $C$ and $R$. If $A$ is sparse, so are $C$ and $R$. 