Origins of Matrix/Vector and Matrix/Matrix Multiplication

R. Scott McIntire

Dec 5, 2024

1 Algebraic properties of Matrices

Why is matrix/vector and matrix/matrix multiplication defined the way it is? One motivation is to try to extend the (albeit trivial) solution of linear equations in 1-dimension to n dimensions. To do this, we go through the solution of the 1-dimensional case in careful detail.

The scalar problem is:

$$a\,x = b \tag{1}$$

Here is a very explicit solution keeping an eye towards generalization.

$$\begin{aligned}
a\,x &= b && && (2)\\
a^{-1}(a\,x) &= a^{-1}b && \text{(Multiply by Inverse)} && (3)\\
(a^{-1}a)x &= a^{-1}b && \text{(Use associativity of multiplication)} && (4)\\
1\,x &= a^{-1}b && \text{(What Inverse multiplication does)} && (5)\\
x &= a^{-1}b && \text{(What identity multiplication does)} && (6)
\end{aligned}$$

The multi-dimensional problem has a lot more variables and coefficients. To start, we need to be more systematic about the naming of these coefficients. With this in mind, the multi-dimensional linear problem can be written:

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1\\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2\\
\cdots &= \cdots\\
a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n
\end{aligned}$$

Here, a_{ij} is the coefficient in the i^{th} row and j^{th} column.

To make this look like the 1-dimensional case, we need to think of the b's as a single unit. Our single unit will be the *vector* of the b's.

The multi-dimensional case can be rewritten in vector terms as:

$$\begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \tag{7}$$

Two things need to be done: treat the x's as a unit – as we did with the b's – and separate the a coefficients from the x's. This must be done formally and yet have the same meaning as the original problem formulation. This is done in the most natural of ways:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \tag{8}$$

Taking the rectangular collection (matrix) of a's to be A; the (vector) collection of x's to be $\mathbf{x}$; and the collection of the b's to be $\mathbf{b}$; we may write (8) as:

$$A\,\mathbf{x} = \mathbf{b} \tag{9}$$

This now looks like the scalar problem, (1).

We proceed to try solving using the solution procedure used above for the scalar case. In the process, we will need to:

- Vectorize the input variable x and the outputs b. – **Done**.
- Define an, a, matrix. – **Done**.
- Define matrix-vector multiplication.
- Define matrix-matrix multiplication.
- Define the identity matrix.
- Define the inverse matrix.

How do we make sense of this matrix/vector syntax? It should have the proper meaning; that is, (8) should have the same meaning as (7).

Examining (8) and the left hand side of (7), gives us our definition of matrix/vector multiplication:

$$[A\,\mathbf{x}]_i \equiv \sum_{j=1}^{n} A_{ij}x_j \tag{10}$$

That is, A acts on a vector, $\mathbf{x}$, to create a new vector, $A\,\mathbf{x}$, whose i^{th} entry is the i^{th} entry of the left hand side of (7).

Another way to write this is:

$$A\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{A}^j \tag{11}$$

Here, $\mathbf{A}^j$ is the j^{th} column vector of A. Applying A to the special vectors, $\mathbf{e}_j$ – which are 0 everywhere except at j where they are 1 – we see that $A\,\mathbf{e}_j = \mathbf{A}^j$. Consequently, this matrix/vector multiplication determines A uniquely – if there is another matrix, its columns would have to match A.

We note the following for future reference.

$$[\mathbf{A}^j]_i = A_{i,j} \tag{12}$$

That is, the i^{th} entry of the column vector, $\mathbf{A}^j$, is the i^{th} row, j^{th} column of the matrix A.

Before moving on with our outline, we note one very important property of matrix/vector multiplication: It is *linear*. By that we mean the following:[1]

$$A(a\mathbf{x} + b\mathbf{y}) = a\,A\mathbf{x} + b\,A\mathbf{y} \tag{13}$$

One can show this from our definition of matrix/vector multiplication which we leave to the reader. What this means in practice is that for a given vector, $\mathbf{z} = a\mathbf{x} + b\mathbf{y}$, we can compute $A\mathbf{z}$ by computing A on the "components" of **zl** and then multiplying by scalars and adding the resulting vectors. This can be an easier way to compute the action of a matrix on vector in some cases. The result above can be extended to arbitrary sums as:

$$A\left(\sum_{j=1}^{n} c_j \mathbf{x}_j\right) = \sum_{j=1}^{n} c_j A\mathbf{x}_j \tag{14}$$

Continuing with our solution outline, matrix/matrix multiplication is needed to go from equation (3) to (4). How must this be defined? Well, we need to make sense of:

$$(A^{-1}A)\mathbf{x} = A^{-1}(A\,\mathbf{x}) \tag{15}$$

This involves the inverse of the matrix A, which we also need to define. What we are imposing on matrix multiplication is that it be associative in the specific case of a matrix and its inverse. But this is not what happens in the scalar case; associativity works not just for a special case of multiplication, but for all numbers. We will forgo what the inverse of a matrix might mean and focus now on imposing the condition that matrix/matrix multiplication be associative – just as multiplication is for numbers. Specifically, the condition we impose is:

$$(B\,A)\mathbf{x} = B(A\,\mathbf{x}) \quad \forall \mathbf{x} \in R^n \tag{16}$$

Notice that we are imposing what matrix/matrix multiplication is by saying how the new matrix formed, $B\,A$, acts on an *arbitrary* vector, $\mathbf{x}$. And we specify that action by the right hand side of (16), which involves only matrix/vector multiplication – something we already know.

[1] Here, we assume the reader has a passing knowledge of vector addition.

We want to emphasize that this is a strong condition to impose; meaning, that the requirement that this is true for *all* $\mathbf{x}$ gives us a lot to work with. For instance, if one had two $n \times n$ matrices, C and D, what could you conclude if someone told you that $C\mathbf{x} = D\mathbf{x}$ for *some* n vector, $\mathbf{x}$? Well, the answer is: not much. However, if I told you that $C\mathbf{x} = D\mathbf{x}$ *for all* $\mathbf{x}$, what could you say? I claim that tells us that $C \equiv D$; meaning that every entry in C is the same as the corresponding entry in D. We can see this by applying C and D to each of the $\mathbf{e}_i$, ($i \in [1, n]$) vectors. For each i, $C\mathbf{e}_j$ and $D\mathbf{e}_j$ pick off the j^{th} column of C and D respectively. So we see that each column of C and D must match; consequently, $C \equiv D$.

We apply this idea to (16) to find a formula for matrix/matrix multiplication. Since (16) must be true *for all* $\mathbf{x}$; in particular, it must be true for the vectors $\{\mathbf{e}\}_{\mathbf{i=1}}^{\mathbf{n}}$. This gives us the following equations:

$$(B\,A)\mathbf{e}_j \quad = \quad B(A\,\mathbf{e}_j) \tag{17}$$

Or,

$$(\mathbf{B}\,\mathbf{A})^j \quad = \quad B\mathbf{A}^j \tag{18}$$

For any given, i, the i^{th} entry of the left and right hand side is (after using the definition of matrix/vector multiplication, (10) on the right hand side)

$$\left[(\mathbf{B}\,\mathbf{A})^j\right]_i \quad = \quad \sum_{k=1}^{n} B_{i,k}\left[\mathbf{A}^j\right]_k \quad \forall i \in [1, n] \tag{19}$$

Using (12) on the left and right hand sides of the previous equation we have

$$(B\,A)_{i,j} \quad = \quad \sum_{k=1}^{n} B_{i,k}A_{k,j} \tag{20}$$

And we are done, we have shown what the new matrix, $B\,A$ is by showing what every entry, i, j of the new matrix is. So using this set of vectors, $\{\mathbf{e}_j\}_{j=1}^n$, we have completely determined what the matrix multiplication of $B\,A$ is in order that matrix/matrix multiplication be associative when using these vectors. Does this mean that associativity works for *arbitrary* vectors? The answer is yes and this is because of the linearity of matrix/vector multiplication. Any given $\mathbf{x}$ can be written as a *linear combination* of the vectors $\{\mathbf{e}_j\}_{j=1}^n$: $\mathbf{x} = a_1\mathbf{e}_1 + a_2\mathbf{e}_\mathbf{2} + \cdots + a_n\mathbf{e}_n$. Matrix multiply the left and right hand sides by the matrix, $B\,A$.

This gives

$$
\begin{aligned}
(B\,A)\mathbf{x} &= (B\,A)\left(\sum_{j=1}^{n} a_j \mathbf{e}_j\right) \qquad \text{Replace } \mathbf{x} \text{ with its components.} \\
&= \sum_{j=1}^{n} a_j (B\,A)\mathbf{e}_j \qquad \textit{Linearity} \text{ of matrix, } (B\,A). \\
&= \sum_{j=1}^{n} a_j B(A\mathbf{e}_j) \qquad \textit{Associativity} \text{ of matrix multiplication on } \mathbf{e}_j. \\
&= B\left(\sum_{j=1}^{n} a_j A\mathbf{e}_j\right) \qquad \textit{Linearity} \text{ of matrix } B. \\
&= B\left(A\left(\sum_{j=1}^{n} a_j \mathbf{e}_j\right)\right) \qquad \textit{Linearity} \text{ of matrix } A. \\
&= B(A\mathbf{x}) \qquad \text{Replace components with } \mathbf{x}.
\end{aligned} \tag{21}
$$

In the above calculations, we applied our new matrix, $(B\,A)$ to an arbitrary $\mathbf{x}$, and used the *linearity* of matrix/vector multiplication and the *associativity* of matrix/matrix multiplication on the components, to show that matrix/matrix multiplication, as we have defined it, is associative – irrespective of the input.

The next order of business is to identify an $n \times n$ matrix which serves as an identity (in terms of matrix/vector multiplication). Let us suppose that we have such a matrix and let's call it I. Then we must have $I\,\mathbf{x} = \mathbf{x} \quad \forall \mathbf{x} \in R^n$. Then, it is not hard to see that $I\,\mathbf{e}_j = \mathbf{e}_j \quad \forall j \in [1, n]$. However, we know that $I\,\mathbf{e}_j = I^j$. Therefore, I must have the property that $\mathbf{I^j} = \mathbf{e}_j$. Consequently I must have the form:

$$
I = \begin{bmatrix}
1 & 0 & 0 & 0 & \cdots & 0 \\
0 & 1 & 0 & 0 & \cdots & 0 \\
0 & 0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 0 & 0 & \ddots & 0 \\
0 & 0 & 0 & 0 & \cdots & 1
\end{bmatrix} \tag{22}
$$

That is $I_{ij} = \delta_{ij}$.[2] We have shown that the only candidate for the identity matrix with respect to matrix/vector multiplication is the matrix I. That is, if there is an identity matrix, it must be I. Does it satisfy the property of being an identity matrix (again, in the matrix/vector multiplication world)? That is, do we have $I\,\mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \in R^n$? As before, we can express $\mathbf{x}$ as: $\mathbf{x} = \sum_{j=1}^{n} c_j \mathbf{e}_j$. By the *linearity* of

[2] $\delta_{ij} = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{otherwise.} \end{cases}$

matrix/vector multiplication, we have

$$
\begin{aligned}
I\mathbf{x} &= I\left(\sum_{j=1}^{n} c_j \mathbf{e}_j\right) && \text{Replace } \mathbf{x} \text{ with components.} \\
&= \sum_{j=1}^{n} c_j I \mathbf{e}_j && \textit{Linearity} \text{ of matrix/vector multiplication.} \\
&= \sum_{j=1}^{n} c_j \mathbf{e}_j && \textit{Identity} \text{ of } I \text{ on components.} \\
&= \mathbf{x} && \text{Replace components with } \mathbf{x}.
\end{aligned} \tag{23}
$$

One can show that this identity matrix, I, is also the identity operator for matrix/matrix multiplication.

The only thing left is to know when a matrix inverse exists and how to compute it. We do not attempt to do this in this paper. In the next section we continue with a qualitative comparison of the solution to the scalar problem, $a\,x = b$, and its vectorized cousin.

2 Qualitative Features of Solutions

We can view the multiplication of two numbers, a and x, as just that. Or, we can think of a being fixed and letting x "run-through" all numbers. Here we see two cases: if $a \neq 0$, then letting x run through all of the numbers in R will produce all of the numbers in R. We could think of a as an "operator" and call the set of all possible outputs, the range of a and denote it: $\mathcal{R}(a)$. There is another case, a could be zero. In this case it's range is the set $\{0\}$. In the first case, with non-zero a it is clear that we can find an x to "hit" a given value b. That is, we can solve $a\,x = b$.

One can define the same concept for a matrix, A. Using this language of ranges, here is what we can say about the scalar problem: $a\,x = b$.

Unique Solution: If a^{-1} exists ($a \neq 0$), b is any number (that is: ($b \in \mathcal{R}(a)$) then there is a **unique** solution.

No Solution: If a^{-1} does not exist (i.e., $a = 0$) *AND* b is **not** in the range of a (that is: $b \neq 0$), then there is **no** solution.

Infinite Solutions: If a^{-1} does not exist (i.e., $a = 0$) *BUT* b is in the range of a (that is: $b = 0$), then there are an **infinite** number of solutions.

Here is the analog of this solution categorization for the multi-dimensional case: $A\,\mathbf{x} = \mathbf{b}$.

Unique Solution: If A^{-1} exists, $\mathbf{b}$ is in the range of A ($\mathbf{b} \in R(A)$) then there is a **unique** solution.

No Solution: If A^{-1} does not exist *AND* $\mathbf{b}$ is not in the range of A (that is $\mathbf{b} \notin \mathcal{R}(A)$), then there is **no** solution.

Infinite Solutions: If A^{-1} does not exist *BUT* $\mathbf{b}$ is in the range of A (that is: $\mathbf{b} \in \mathcal{R}(A)$), then there are an **infinite** number of solutions.