# Origins of Matrix/Vector and Matrix/Matrix Multiplication

R. Scott McIntire

Feb 26, 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.

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

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 $i^{\text{th}}$ row and $j^{\text{th}}$ 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 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 $i^{\text{th}}$ entry is the $i^{\text{th}}$ entry of the left hand side of (7).

Another way to write this is:

$$A\mathbf{x} \quad = \quad \sum_{i=1}^{n} x_i \mathbf{A}^i \tag{11}$$

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

To complete the solution outline, matrix/matrix multiplication is needed via (4). How must this be defined? Well, we need to make sense of:

$$(A^{-1}A)\mathbf{x} \quad = \quad A^{-1}(A\,\mathbf{x}) \tag{12}$$

This involves the inverse of the matrix $A$, which we also need to define. Once we have the inverse we then need to do matrix-matrix multiplication. Let us start with matrix-matrix multiplication. This will be determined by the fact that we (just as in the scalar case) demand that the multiplication be *associative*. [1]

$$(B\,A)\mathbf{x} \quad \equiv \quad B(A\,\mathbf{x}) \quad \forall x \in R^n \tag{13}$$

This definition would mean that $B\,A$ is a new $n \times n$ matrix whose $i^{\text{th}}$ entry – when acting on an arbitrary vector $\mathbf{x}$ – is (using (10) twice):

$$[(B\,A)\mathbf{x}]_i \quad = \quad [B(A\,\mathbf{x})]_i = \sum_{k=1}^{n} B_{ik} \left( \sum_{j=1}^{n} A_{kj} x_j \right) \quad \forall x \in R^n, \forall i \in [1, n] \tag{14}$$

Using (10) on the left hand side yields:

$$\sum_{j=1}^{n} (B\,A)_{ij} x_j \quad = \quad \sum_{k=1}^{n} B_{ik} \left( \sum_{j=1}^{n} A_{kj} x_j \right) \tag{15}$$

Or,

$$\sum_{j=1}^{n} (B\,A)_{ij} x_j \quad = \quad \sum_{k=1}^{n} \sum_{j=1}^{n} B_{ik} A_{kj} x_j \tag{16}$$

Changing the order of summation on the right hand side, this is:

$$\sum_{j=1}^{n} (B\,A)_{ij} x_j \quad = \quad \sum_{j=1}^{n} \left( \sum_{k=1}^{n} B_{ik} A_{kj} \right) x_j \tag{17}$$

Or,

$$\sum_{j=1}^{n} \left[ (B\,A)_{ij} - \left( \sum_{k=1}^{n} B_{ik} A_{kj} \right) \right] x_j \quad = \quad 0 \tag{18}$$

[1] We know from above that defining how a given matrix acts on all vector via matrix/vector multiplication uniquely determines the matrix. So, this is a proper definition of matrix/matrix multiplication.

This suggests that the $i^{\text{th}}, j^{\text{th}}$ entry of the multiplication of $B$ and $A$ is:

$$(B\,A)_{ij} \quad = \quad \sum_{k=1}^{n} B_{ik} A_{kj} \quad \forall i \in R^n, \forall j \in R^n \tag{19}$$

To see that this follows, notice that (18) must hold for all vectors $\mathbf{x}$. Setting $\mathbf{x}$ to the successive $\mathbf{e}_i$ vectors defined above yields (19).

By (6) we need to identify a $n \times n$ 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 $I\,\mathbf{e}_i = \mathbf{e}_i \quad \forall i \in [1, n]$. However, we know that $I\,\mathbf{e}_i = I^i$. Therefore, $I$ must have the property that $\mathbf{I^i} = \mathbf{e}_i$. Consequently $I$ must have the form:

$$I \quad = \quad \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{20}$$

That is $I_{ij} = \delta_{ij}$.[2] We have shown that the only candidate matrix that has the identity property, 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)? Do we have $I\,\mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \in R^n$? Using the definition of matrix/vector multiplication, (10), we have for any given, $i$:

$$\begin{aligned} [I\mathbf{x}]_i &= \sum_{i=1}^{n} I_{ij} x_j \\ &= \sum_{i=1}^{n} \delta_{ij} x_j \\ &= x_i \\ &= [\mathbf{x}]_i \end{aligned} \tag{21}$$

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.

# 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

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

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.