Linear algebra for numerical analysis

Douglas Wilhelm Harder

April 7, 2025

Not unexpectedly, a number of students forget most of what they learned in their linear algebra course following a one-year hiatus between learning the material and using that material.

Please note, despite being given these notes, topics that are used from linear algebra will never-the-less be summarized and described in class. Thus, while you should read these now, it is not expected that you have mastered all this material for the course lectures.

1 Vectors

An n-dimensional vector is a column of real or complex numbers:

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

If the entries are real, it is called a *real* n-dimensional vector, and if the entries are complex, it is called a *complex* n-dimensional vector. For this course, we will generally focus on real vectors. If a vector is being represented by a letter, that letter will be a lower-case boldface letter and the corresponding entries of the vector will be that same letter but lower-case italicized with a subscript indicating which of the n entries that entry corresponds to. Thus, if $\mathbf{u}$ is a real n-dimensional vector, then u_k is a real number (a scalar) and k is an integer from 1 to n.

The collection of all real n-dimensional vectors is represented by $\mathbf{R}^n$. If $\mathbf{u}$ is a real n-dimensional vectors, we may write $\mathbf{u} \in \mathbf{R}^n$.

Two n-dimensional vectors are "equal" if their corresponding entries are the same. Thus $\mathbf{u} = \mathbf{v}$ if and only if all of $u_1 = v_1$, $u_2 = v_2$, all the way up to $u_n = v_n$.

Given an n-dimensional vector $\mathbf{u}$, we can multiply that vector by a scalar α, written as $\alpha\mathbf{u}$ by multiplying each entry in that vector by α, so

$$\alpha\mathbf{u} = \begin{pmatrix} \alpha u_1 \\ \alpha u_2 \\ \vdots \\ \alpha u_n \end{pmatrix}.$$

For example, $-3.2 \begin{pmatrix} 1.3 \\ -2.1 \\ 3.0 \end{pmatrix} = \begin{pmatrix} -4.16 \\ 6.72 \\ -9.60 \end{pmatrix}$.

Given two n-dimensional vectors $\mathbf{u}$ and $\mathbf{v}$, the sum of the two vectors, written $\mathbf{u} + \mathbf{v}$, is that n-dimensional vector with entries equal to the sum of the corresponding entries of $\mathbf{u}$ and $\mathbf{v}$, so

$$\mathbf{u} + \mathbf{v} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{pmatrix}.$$

For example, $\begin{pmatrix} 5.62 \\ 6.65 \\ -4.06 \\ -7.83 \end{pmatrix} + \begin{pmatrix} 7.31 \\ -4.84 \\ 8.89 \\ -0.67 \end{pmatrix} = \begin{pmatrix} 12.93 \\ 1.81 \\ 4.83 \\ -8.50 \end{pmatrix}$.

Given N vectors $\mathbf{u}_1, \ldots, \mathbf{u}_N$, and N scalars $\alpha_1, \ldots, \alpha_N$, a "linear combination" of those N vectors is $\alpha_1\mathbf{u}_1 + \cdots + \alpha_N\mathbf{u}_N$. If you are calculating such a linear combination, once you perform all of the scalar multiplications, it doesn't matter in which order you add them, just like calculating the sum of n real numbers: it doesn't matter what order you add them up.

The "zero vector" is that vector that has all entries equal to zero, and is written as a boldface zero with a subscript indicating the dimension of the vector. For example,

$$\mathbf{0}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

The "one vector" is that vector that has all entries equal to one, and is written as a boldface one with a subscript indicating the dimension of the vector. For example,

$$\mathbf{1}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

The "dot product" or "inner product" of two real n-dimensional vectors is the sum of the products of the corresponding entries. Thus, given $\mathbf{u}$ and $\mathbf{v}$,

$$\mathbf{u} \cdot \mathbf{v} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{k=1}^{n} u_k v_k.$$

Two n-dimensional vectors $\mathbf{u}$ and $\mathbf{v}$ are at "right angles" or "orthogonal" if and only if the dot product is zero. Every vector is orthogonal to the zero vector, and the zero vector is the only vector that is orthogonal to all vectors.

If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal, then so are all scalar multiples of those vectors. Thus, if $\mathbf{u} \cdot \mathbf{v} = 0$ and if α and β are real numbers, then $(\alpha\mathbf{u}) \cdot (\beta\mathbf{v}) = 0$.

Note that in $\mathbf{R}^n$, the only vector that is orthogonal to all other vectors $\mathbf{u} \in \mathbf{R}^n$ is the zero vector $\mathbf{0}_n$. That is, if $\mathbf{u} \cdot \mathbf{v} = 0$ for all vectors $\mathbf{u} \in \mathbf{R}^n$, then $\mathbf{v} = \mathbf{0}_n$. This is because, if nothing else, any non-zero vector is not orthogonal to itself, for if $\mathbf{u}$ is not the zero vector, then there must be at least one entry u_k that is non-zero, and therefore $u_k^2 > 0$. Thus $\mathbf{u} \cdot \mathbf{u} = u_1 u_1 + u_2 u_2 + \cdots + u_n u_n = u_1^2 + u_2^2 + \cdots + u_n^2$. All the terms in this sum are non-negative, and therefore the sum must be greater than or equal to the one entry we did identify as being positive: $\mathbf{u} \cdot \mathbf{u} \geq u_k^2 > 0$. Thus, no non-zero vector is orthogonal to itself, and therefore the only vector that is orthogonal to all vectors is the zero vector.

1.1 Length or 2-norm of a vector

The "length" or "2-norm" of a real n-dimensional vector is defined as

$$\|\mathbf{u}\|_2 = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2} = \sqrt{\sum_{k=1}^{n} u_k^2}.$$

For example, if $\mathbf{u} = \begin{pmatrix} 3 \\ -3.75 \\ -4 \end{pmatrix}$, then $\|\mathbf{u}\|_2 = \sqrt{9 + 14.0625 + 16} = 6.25$.

Two important properties are:

1. Only the zero vector has a length or 2-norm equal to zero. All other vectors have a positive 2-norm. Thus, $\|\mathbf{u}\|_2 \geq 0$ and $\|\mathbf{u}\|_2 = 0$ if and only if $\mathbf{u} = \mathbf{0}_n$.
2. If $\mathbf{u}$ is multiplied by a scalar α, then the length or 2-norm of that vector is stretched by $|\alpha|$. Thus, $\|\alpha\mathbf{u}\|_2 = |\alpha|\|\mathbf{u}\|_2$.

Notice that the two-norm $\|\mathbf{u}\|_2 = \sqrt{\mathbf{u} \cdot \mathbf{u}}$; that is, the 2-norm is the norm "induced" by the inner product.

The "angle" between two vectors may be calculated as being $\cos^{-1}\left(\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|_2\|\mathbf{v}\|_2}\right)$. If the dot product is zero, then $\cos^{-1}(0) = \frac{\pi}{2}$, so the two vectors are at right angles.

Just like two real numbers x and y are "close" to each other if $|x - y| < \epsilon$, and two complex numbers w and z are "close" if $|w - z| < \epsilon$, two vectors $\mathbf{u}$ and $\mathbf{v}$ may be said to be "close" if $\|\mathbf{u} - \mathbf{v}\|_2 < \epsilon$.

1.2 Normalized or unit vectors

A vector is "normalized" or described as a "unit vector" if its length or 2-norm equals 1. A non-unit non-zero vector can be normalized by dividing it by its norm. For example, $\begin{pmatrix} -0.6 \\ 0.8 \end{pmatrix}$ is a unit vector, while

$\begin{pmatrix} -1 \\ 4 \\ 9.5 \\ -7 \end{pmatrix}$, it can be divided by its norm of 12.5 to get the unit vector $\begin{pmatrix} -0.08 \\ 0.32 \\ 0.76 \\ -0.56 \end{pmatrix}$. If a vector is a unit vector, it is usually drawn with a "hat", so for example, $\hat{\mathbf{u}}$.

The most familiar unit vectors are the "canonical basis vectors". If we are dealing with vectors in $\mathbf{R}^n$, then $\mathbf{e}_k$ is the vector of all zeros except for the k^{th} entry. For example, if we are dealing with vectors in $\mathbf{R}^4$, then $\mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$.

2 Matrices

An $m \times n$ matrix is a grid of real or complex numbers:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix}.$$

The first number is the number of rows, and the second is the number of columns. If you need a mnemonic, think of "rum and coke" for "rows and columns". Similarly, the index of each entry stores the row first, and then the column, so $a_{5,3}$ is the entry in Row 5 and Column 3.

We will see that it is often useful to think of as an ordered collection of column vectors, so $A = (\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_n)$, where $\mathbf{a}_j = \begin{pmatrix} a_{1,j} \\ a_{2,j} \\ \vdots \\ a_{m,j} \end{pmatrix}$ for each $j = 1, 2, \ldots, n$.

For example, the 3×4 matrix

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{pmatrix} = (\mathbf{a}_1\ \mathbf{a}_2\ \mathbf{a}_3\ \mathbf{a}_4)$$

where

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ 5 \\ 9 \end{pmatrix}, \quad \mathbf{a}_2 = \begin{pmatrix} 2 \\ 6 \\ 10 \end{pmatrix}, \quad \mathbf{a}_3 = \begin{pmatrix} 3 \\ 7 \\ 11 \end{pmatrix}, \text{ and } \mathbf{a}_4 = \begin{pmatrix} 4 \\ 8 \\ 12 \end{pmatrix}.$$

If the entries are real, it is called a *real* $m \times n$ matrix, and if the entries are complex, it is called a *complex* $m \times n$ matrix. For this course, we will generally focus on real matrices. If a matrix is being represented by a letter, that letter will be an upper-case italicized letter and the corresponding entries of the matrix will be that same letter but lower-case italicized with a pair of subscript indicating which of the $m \times n$ entries that entry corresponds to. Thus, if A is a real $m \times n$ matrix, then $a_{i,j}$ is a real number (a scalar), i is an integer from 1 to m, and j is an integer from 1 to n.

Two $m \times n$ matrices A and B are considered to be equal if and only if all their corresponding entries are equal. Thus, $A = B$ if and only if $a_{i,j} = b_{i,j}$ for all values $i = 1, \ldots, m$ and $j = 1, \ldots n$.

An $m \times n$ matrix is called "square" if $m = n$.

The collection of all real $m \times n$ matrices is represented by $\mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$. If $m = n$, we may write this as $\mathcal{L}(\mathbf{R}^n)$. If A is a real $m \times n$ matrix, we may write $A \in \mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$, and if A is a real $n \times n$ matrix, we may write $A \in \mathcal{L}(\mathbf{R}^n)$.

Given an $m \times n$ matrix A, we can multiply that matrix by a scalar α, written as αA by multiplying each entry in that matrix by α, so

$$\alpha A = \begin{pmatrix} \alpha a_{1,1} & \alpha a_{1,2} & \cdots & \alpha a_{1,n} \\ \alpha a_{2,1} & \alpha a_{2,2} & \cdots & \alpha a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha a_{m,1} & \alpha a_{m,2} & \cdots & \alpha a_{m,n} \end{pmatrix}.$$

Given two $m \times n$ matrices A and B, the sum of the two matrices, written $A + B$, is that $m \times n$ matrix with entries equal to the sum of the corresponding entries of A and B, so

$$A+B = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} + \begin{pmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,n} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m,1} & b_{m,2} & \cdots & b_{m,n} \end{pmatrix}$$
$$= \begin{pmatrix} a_{1,1}+b_{1,1} & a_{1,2}+b_{1,2} & \cdots & a_{1,n}+b_{1,n} \\ a_{2,1}+b_{2,1} & a_{2,2}+b_{2,2} & \cdots & a_{2,n}+b_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1}+b_{m,1} & a_{m,2}+b_{m,2} & \cdots & a_{m,n}+b_{m,n} \end{pmatrix}.$$

The zero $m \times n$ matrix is that $m \times n$ matrix that has all entries equal to zero, and is written as an italicized "O" with subscripts indicating the dimensions of the matrix, or $O_{m,n}$. For example,

$$O_{3,4} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Given an $m \times n$ matrix A, the zero matrix $O_{m,n}$ is the only matrix such that $A + O_{m,n} = O_{m,n} + A = A$.

3 Matrix-vector multiplication

Given an $m \times n$ matrix A and an n-dimensional vector $\mathbf{u}$, the matrix-vector product, represented as $A\mathbf{u}$, is that m-dimensional vector where the k^{th} entry is computed by multiplying each entry in Row k of the matrix A with the corresponding entry of $\mathbf{u}$, and summing the results. Thus,

$$v_k = a_{k,1}u_1 + \cdots + a_{k,n}u_n = \sum_{j=1}^{n} a_{k,j}u_j.$$

For example,

$$\begin{pmatrix} 3 & 2 & 5 & -9 \\ 6 & 10 & 0 & 7 \\ \mathbf{4} & \mathbf{1} & \mathbf{11} & \mathbf{8} \end{pmatrix} \begin{pmatrix} 2.7 \\ 3.1 \\ -0.4 \\ 5.6 \end{pmatrix} = \begin{pmatrix} 3\cdot 2.7 + 2\cdot 3.1 + 5\cdot(-0.4) - 9\cdot 5.6 \\ 6\cdot 2.7 + 10\cdot 3.1 + 0\cdot(-0.4) + 7\cdot 5.6 \\ \mathbf{4}\cdot 2.7 + \mathbf{1}\cdot 3.1 + \mathbf{11}\cdot(-0.4) + \mathbf{9}\cdot 5.6 \end{pmatrix} = \begin{pmatrix} -38.1 \\ 86.4 \\ \mathbf{54.3} \end{pmatrix}.$$

Thus, an $m \times n$ matrix A represents a mapping from n-dimensional vectors to m-dimensional vectors, and thus we write $A : \mathbf{R}^n \to \mathbf{R}^m$.

We will describe a property that of matrix-vector multiplication where, for example, $A(3.5\mathbf{u}_1 - 7.1\mathbf{u}_2) = 3.5A\mathbf{u}_1 - 7.1A\mathbf{u}_2$, as "linearity", making the matrix a "linear mapping". You are aware that $\sin(3.5x - 7.1y) \neq 3.5\sin(x) - 7.1\sin(y)$, but you should also be aware that

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(3.5f(x) - 7.1g(x)\right) = 3.5\frac{\mathrm{d}}{\mathrm{d}x}f(x) - 7.1\frac{\mathrm{d}}{\mathrm{d}x}g(x),$$

and more generally

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\alpha f(x) + \beta g(x)\right) = \alpha\frac{\mathrm{d}}{\mathrm{d}x}f(x) + \beta\frac{\mathrm{d}}{\mathrm{d}x}g(x),$$

in other words, the derivative applied to a linear combination of functions is that same linear combination of the derivative applied to each of the functions.

Similarly, if A is an $m \times n$ matrix and $\mathbf{u}_1$ and $\mathbf{u}_2$ are n-dimensional vectors, then

$$A(\alpha\mathbf{u}_1 + \beta\mathbf{u}_2) = \alpha A\mathbf{u}_1 + \beta A\mathbf{u}_2.$$

In other words, multiplying a linear combination of vectors by a matrix is the same as taking the same linear combination of the matrix multiplied by the corresponding vectors. This property is why we describe matrix-vector multiplication as a "linear mapping".

Note: Some textbooks call a "linear mapping" a "linear transformation". This is also acceptable terminology, but it's easier to think of a matrix as a mapping or function.

As an aside, integration is also linear:

$$\int_a^b \alpha f(x) + \beta g(x)\mathrm{d}x = \alpha\int_a^b f(x)\mathrm{d}x + \beta\int_a^b \beta g(x)\mathrm{d}x.$$

Why this is relevant is that you will be learning about the Laplace and Fourier transforms, and these are defined in terms of integrals, and thus, these, too, are linear mappings: if the Laplace transform of the function $f(t)$ is $\mathscr{L}(f)(s)$, then $\mathscr{L}(\alpha f + \beta g)(s) = \alpha\mathscr{L}(f)(s) + \beta\mathscr{L}(g)(s)$.

There is exactly one $n \times n$ matrix called the "identity matrix" or I_n such that $I_n \mathbf{u} = \mathbf{u}$ for all n-dimensional vectors $\mathbf{u}$, and this is the $n \times n$ matrix where if the indices are equal, the entry is 1, otherwise the entry is zero. For example, $I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. For obvious reasons, the three entries containing a 1 are called the "diagonal entries". This is called the "identity matrix" because, like multiplying a complex number by $1 + 0j$, it leaves what is being multiplied unchanged. You will note that $I_n = (\mathbf{e}_1 \cdots \mathbf{e}_n)$ where you will recall that $\mathbf{e}_j$ is the j^{th} n-dimensional canonical basis vector.

3.1 Matrix-vector multiplication as a linear combination of vectors

Multiplying a matrix and a vector is actually the same as taking a linear combination of the columns of the matrix where the coefficients are the scalars of the corresponding vectors. Thus, for example,

$$\begin{pmatrix} -3 & 2 & 5 & 8 \\ 1 & 0 & -8 & 7 \\ -6 & 9 & 4 & 3 \end{pmatrix} \begin{pmatrix} 5.4 \\ -7.9 \\ -6.1 \\ 2.8 \end{pmatrix} = 5.4 \begin{pmatrix} -3 \\ 1 \\ -6 \end{pmatrix} - 7.9 \begin{pmatrix} 2 \\ 0 \\ 9 \end{pmatrix} - 6.1 \begin{pmatrix} 5 \\ -8 \\ 4 \end{pmatrix} + 2.8 \begin{pmatrix} 8 \\ 7 \\ 3 \end{pmatrix}$$

Thus, it is often useful to think of an $m \times n$ matrix as a collection of ordered m-dimensional vectors, so we may write $A = (\mathbf{a}_1 \ \mathbf{a}_2 \ \mathbf{a}_3 \ \cdots \ \mathbf{a}_n)$ where each $\mathbf{a}_k$ is an m-dimensional vector, so that $A\mathbf{u} = u_1 \mathbf{a}_1 + u_2 \mathbf{a}_2 + u_3 \mathbf{a}_3 + \cdots + u_n \mathbf{a}_n$ for an n-dimensional vector $\mathbf{u}$.

3.2 Matrix-vector multiplication as a system of linear equations

A system of m linear equations in n unknowns may be written as

$$\begin{array}{ccccccccccc} a_{1,1}u_1 & + & a_{1,2}u_2 & + & a_{1,3}u_3 & + & \cdots & + & a_{1,n}u_n & = & v_1 \\ a_{2,1}u_1 & + & a_{2,2}u_2 & + & a_{2,3}u_3 & + & \cdots & + & a_{2,n}u_n & = & v_2 \\ a_{3,1}u_1 & + & a_{3,2}u_2 & + & a_{3,3}u_3 & + & \cdots & + & a_{3,n}u_n & = & v_3 \\ \vdots & & \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m,1}u_1 & + & a_{m,2}u_2 & + & a_{m,3}u_3 & + & \cdots & + & a_{m,n}u_n & = & v_m \end{array}$$

Observe that the left-hand side of this system consists of linear combinations of the unknowns $u_1, \ldots, u_n$, with the coefficients given by the entries of the matrix. This is precisely what is computed when a matrix multiplies a vector. Thus, the system can be rewritten compactly as a matrix-vector equation: we seek all vectors $\mathbf{u} \in \mathbf{R}^n$ such that the matrix of coefficients times $\mathbf{u}$ equals the target vector $\mathbf{v} \in \mathbf{R}^m$:

$$\begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots & a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & a_{m,3} & \cdots & a_{m,n} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \\ \vdots \\ \\ u_n \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_m \end{pmatrix}$$

For example, these are equivalent problems: a solution (if any) to one is a solution to the other:

$$\begin{aligned} 5x + 6y - 3z &= 8 \\ -2x + 9y + 4z &= -1 \end{aligned} \quad \text{and} \quad \begin{pmatrix} 5 & 6 & -3 \\ -2 & 9 & 4 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 8 \\ -1 \end{pmatrix}.$$

Thus, rather than focusing on algorithms to solve systems of linear equations, we will instead focus on the question: Given a matrix $A : \mathbf{R}^n \to \mathbf{R}^m$ and a target vector $\mathbf{v} \in \mathbf{R}^m$, what vectors $\mathbf{u} \in \mathbf{R}^n$, if any equal this target vector when multiplied by A, or in other words, we are solving $A\mathbf{u} = \mathbf{v}$. It turns out there are three general possibilities:

1. There is one unique solution.
2. There are infinitely many solutions.
3. There are no solutions.

In your course on linear algebra, in the case where there were infinitely many solutions, you found a basis for the null space, and then you found one particular solution to $A\mathbf{u} = \mathbf{v}$, and to this particular solution, you could add an linear combination of the basis vectors. We will focus on situations where the matrix is square and when a solution exists.

3.3 Matrix-vector multiplication in practice

While it is best to think of matrix-vector multiplication $A\mathbf{u}$ as defining a linear combination of the columns of the matrix A with the coefficients prescribed by the corresponding entries in the vector, this can be more difficult to calculate. Instead, the easiest way to calculate a matrix-vector product is if $\mathbf{v} = A\mathbf{u}$, then the k^{th} entry of $\mathbf{v}$ is assigned the sum of the element-wise products of Row k of the matrix A and the vector $\mathbf{u}$. For example, in this example, the third entry of the product, 49.6, is the sum of the element-wise products of Row 3 of the matrix with the corresponding entries of the 4-dimensional vector:

$$\begin{pmatrix} 5 & 2 & 3 & 1 \\ -4 & 6 & 0 & -5 \\ \mathbf{7} & \mathbf{-3} & \mathbf{9} & \mathbf{8} \end{pmatrix} \begin{pmatrix} 1.5 \\ -3.7 \\ 2.4 \\ 0.8 \end{pmatrix} = \begin{pmatrix} 5 \cdot 1.5 & + & 2 \cdot (-3.7) & + & 3 \cdot 2.4 & + & 1 \cdot 0.8 \\ -4 \cdot 1.5 & + & 6 \cdot (-3.7) & + & 0 \cdot 2.4 & + & (-5) \cdot 0.8 \\ \mathbf{7 \cdot 1.5} & + & (\mathbf{-3}) \cdot (\mathbf{-3.7}) & + & \mathbf{9 \cdot 2.4} & + & \mathbf{8 \cdot 0.8} \end{pmatrix}$$
$$= \begin{pmatrix} 8.1 \\ -32.2 \\ \mathbf{49.6} \end{pmatrix}.$$

3.4 One-to-one and onto

In the review of calculus, we discussed the concepts of functions being one-to-one and onto, as well as invertability. Because an $m \times n$ matrix A maps $A : \mathbf{R}^n \to \mathbf{R}^m$, we can think of A as a function (or map) from the domain $\mathbf{R}^n$ to the co-domain $\mathbf{R}^m$.

An $m \times n$ matrix $A : \mathbf{R}^n \to \mathbf{R}^m$ is "onto" if and only if for every vector $\mathbf{v}$ in the co-domain, there is a vector $\mathbf{u}$ in the domain such that $A\mathbf{u} = \mathbf{v}$. Such a matrix is one-to-one if there are never two different vectors $\mathbf{u}_1 \neq \mathbf{u}_2$ that map to the same vector.

For example, the matrix $A = \begin{pmatrix} -1 & 2 & -1 \\ 1 & -1 & -2 \\ 1 & -2 & 1 \end{pmatrix}$ is neither one-to-one nor is it onto. For example, there is no 3-dimensional vector $\mathbf{u}$ such that $A\mathbf{u} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ and while $A\begin{pmatrix} 2 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ -3 \end{pmatrix}$, so does $A\begin{pmatrix} -3 \\ -1 \\ -2 \end{pmatrix}$ and so does $A\left(\begin{pmatrix} 2 \\ 2 \\ -1 \end{pmatrix} + \alpha \begin{pmatrix} 5 \\ 3 \\ 1 \end{pmatrix}\right)$ for any scalar α.

A non-square matrix can never be both one-to-one and onto because:

1. If $A : \mathbf{R}^n \to \mathbf{R}^m$ and $n < m$, then A can never be onto, although it may be one-to-one.
2. If $A : \mathbf{R}^n \to \mathbf{R}^m$ and $n > m$, then A can never be one-to-one, although it may be onto.

An $n \times n$ square matrix A is either both one-to-one and onto or it is neither one-to-one nor onto. Those square matrices that are both one-to-one and onto are the only invertible square matrices. A non-square matrix can never be invertible. If $A : \mathbf{R}^n \to \mathbf{R}^n$ is a square matrix that is both one-to-one and onto, we will define $A^{-1} : \mathbf{R}^n \to \mathbf{R}^n$ such that $A^{-1}\mathbf{v} = \mathbf{u}$ when $\mathbf{v} = A\mathbf{u}$.

Consequently, if you are solving a system of n linear equations in n unknowns of the form $A\mathbf{u} = \mathbf{v}$ and A is invertible (so A^{-1} exists and $AA^{-1} = A^{-1}A = I_n$), then instead of using Gaussian elimination and backward substitution, you can solve this directly with the matrix-vector multiplication $\mathbf{u} = A^{-1}\mathbf{v}$.

3.5 Diagonal matrices

The "diagonal entries" of a matrix are those where the indices are equal, so given a $m \times n$ matrix A, then entries $a_{1,1}$ $a_{2,2}, ...,$. All other entries are called "off-diagonal entries". The entries immediately above or to the right of the diagonal entries are called the "super-diagonal entries", and the entries immediately below or to the left of the diagonal entries are called the "sub-diagonal entries".

An $m \times n$ matrix D is a "diagonal matrix" if all off-diagonal entries are zero, so $d_{i,j} = 0$ if $i \neq j$. For example, the following three matrices are diagonal:

$$\begin{pmatrix} 5 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix}, \begin{pmatrix} 7 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -9 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \end{pmatrix}, \begin{pmatrix} -14 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 8 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

A square diagonal matrix D is invertible if and only if all the diagonal entries are non-zero. To find inverse of such a matrix, D^{-1}, just take the reciprocal of the diagonal entries, so if

$$D = \begin{pmatrix} 5 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 2.5 \end{pmatrix},$$

then

$$D^{-1} = \begin{pmatrix} 0.2 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0.4 \end{pmatrix},$$

This should be intuitive, because if you were told:

$$5u_1 = v_1, 10u_2 = v_2, -u_3 = v_3, 2.5u_4 = v_4,$$

then this system of linear equations is described by the matrix-vector product:

$$D\mathbf{u} = \begin{pmatrix} 5 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 2.5 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} = \mathbf{v}.$$

It should be obvious that the above system of linear equations is equivalent to:

$$0.2v_1 = u_1, 0.1v_2 = u_2, -v_3 = u_3, 0.4v_4 = u_4,$$

This is the system of linear equations described by the matrix:

$$D^{-1}\mathbf{v} = \begin{pmatrix} 0.2 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0.4 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix} = \mathbf{u}.$$

4 Upper- and lower-triangular matrices

A matrix A is "upper triangular" if all entries below or to the left of the diagonal are zero. For example, the following three matrices are upper triangular:

$$\begin{pmatrix} 5 & 3 & 8 & 2 \\ 0 & 9 & 7 & 0 \\ 0 & 0 & -1 & 4 \\ 0 & 0 & 0 & 5 \end{pmatrix}, \begin{pmatrix} 7 & 4 & -3 & 1 & 3 \\ 0 & 0 & 7 & 9 & -1 \\ 0 & 0 & -9 & 0 & 8 \\ 0 & 0 & 0 & 2 & 4 \end{pmatrix}, \begin{pmatrix} -14 & 9 & -4 \\ 0 & 3 & -5 \\ 0 & 0 & 8 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

A matrix A is "lower triangular" if all entries above or to the right of the diagonal are zero. For example, the following three matrices are lower triangular:

$$\begin{pmatrix} 5 & 0 & 0 & 0 \\ 4 & 9 & 0 & 0 \\ 7 & 0 & -1 & 0 \\ 6 & -3 & 8 & 5 \end{pmatrix}, \begin{pmatrix} 7 & 0 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 \\ -3 & 2 & -9 & 0 & 0 \\ 4 & 1 & 8 & 2 & 0 \end{pmatrix}, \begin{pmatrix} -14 & 0 & 0 \\ 9 & 3 & 0 \\ -8 & 7 & 8 \\ 12 & 8 & 3 \\ -1 & 3 & 1 \end{pmatrix}$$

A square upper- or lower-triangular matrix is invertible if and only if all the diagonal entries are non-zero. However, it is easy enough to solve a system of linear equations defined by an upper- or lower-triangular matrix by using backward or forward substitution, respectively.

If A is an $n \times n$ upper-triangular invertible square matrix and we are attempting to solve $A\mathbf{u} = \mathbf{v}$, then we use the following algorithm of backward substitution:

1. Starting with $i \leftarrow n$.
2. Assign $u_i \leftarrow v_i$.
3. Starting with $j \leftarrow i + 1$:
   (a) If $j > n$, go to Step 4.
   (b) Otherwise:
      i. Subtract $a_{i,j}u_j$ from u_i.
      ii. Increment j (add one to j).
4. Divide u_i by $a_{i,i}$, so $u_i \leftarrow \frac{u_i}{a_{i,i}}$, recalling that we required that the diagonal entries were non-zero.

5. If $i = 1$, we are done.
6. Otherwise, decrement i (subtract one from i) and return to Step 2.

For example, if

$$A = \begin{pmatrix} 3.5 & -2.7 & 1.3 \\ 0 & 0.9 & -2.4 \\ 0 & 0 & 12.7 \end{pmatrix} \mathbf{u} = \begin{pmatrix} 5.82 \\ 1.08 \\ 7.62 \end{pmatrix},$$

We follow the steps:

1. $u_3 \leftarrow 7.62$.
2. $u_3 \leftarrow \frac{u_3}{12.7} = 0.6$.
3. $u_2 \leftarrow 1.08$.
4. $u_2 \leftarrow u_2 - (-2.4) \cdot 0.6 = 2.52$.
5. $u_2 \leftarrow \frac{u_2}{0.9} = 2.8$.
6. $u_1 \leftarrow 5.82$.
7. $u_1 \leftarrow u_1 - (-2.7) \cdot 2.8 = 13.38$.
8. $u_1 \leftarrow u_1 - 3.5 \cdot 0.6 = 12.6$.
9. $u_1 \leftarrow \frac{u_1}{3.5} = 3.6$.

Thus, the solution is the vector $\mathbf{u} = \begin{pmatrix} 3.6 \\ 2.8 \\ 0.6 \end{pmatrix}$. The runtime of this algorithm is $O(n^2)$ requiring $\frac{n(n-1)}{2}$ floating-point additions and multiplications and n floating-point divisions, for a total of $2n^2 - n$ FLOPs.

If A is an $n \times n$ lower-triangular invertible square matrix and we are attempting to solve $A\mathbf{u} = \mathbf{v}$, then we use the following algorithm of forward substitution:

1. Starting with $i \leftarrow 1$.
2. Assign $u_i \leftarrow v_i$.
3. Starting with $j \leftarrow 1$:
   (a) If $j = i$, go to Step 4.
   (b) Otherwise:
      i. Subtract $a_{i,j}u_j$ from u_i.
      ii. Increment j (add one to j).
4. Divide u_i by $a_{i,i}$, so $u_i \leftarrow \frac{u_i}{a_{i,i}}$, recalling that we required that the diagonal entries were non-zero.
5. If $i = n$, we are done.
6. Otherwise, increment i (add one from i) and return to Step 2.

For example, if

$$A = \begin{pmatrix} 13.9 & 0 & 0 \\ 0.6 & -2.4 & 0 \\ -3.9 & 0.8 & -5.2 \end{pmatrix} \mathbf{u} = \begin{pmatrix} 5.56 \\ 6.72 \\ -12.04 \end{pmatrix},$$

We follow the steps:

1. $u_1 \leftarrow 5.56$.
2. $u_1 \leftarrow \frac{u_1}{13.9} = 0.4$.
3. $u_2 \leftarrow 6.72$.
4. $u_2 \leftarrow u_2 - 0.6 \cdot 0.4 = 6.48$.
5. $u_2 \leftarrow \frac{u_2}{-2.4} = -2.7$.
6. $u_3 \leftarrow -12.04$.
7. $u_3 \leftarrow u_3 - (-3.9) \cdot 0.4 = -10.48$.
8. $u_3 \leftarrow u_3 - 0.8 \cdot (-2.7) = -8.32$.
9. $u_3 \leftarrow \frac{u_3}{-5.2} = 1.6$.

Thus, the solution is the vector $\mathbf{u} = \begin{pmatrix} 0.4 \\ -2.7 \\ 1.6 \end{pmatrix}$. The runtime of this algorithm is $O(n^2)$ requiring the same number of floating-point multiplications, additions and divisions as the previous backward substitution algorithm.

5 Gaussian elimination

We will describe the more specific Gaussian elimination algorithm that finds the unique solution of a system of n linear equations in n unknowns when the corresponding matrix is one-to-one and onto, or invertible. In this case, there is exactly one unique solution to the system of linear equations defined by $A\mathbf{u} = \mathbf{v}$ given the $n \times n$ square matrix A and the given target vector $\mathbf{v}$:

1. Create the $n \times (n+1)$ augmented matrix $(A|\mathbf{b})$.
2. Initialize $j \leftarrow 1$.
3. If $a_{j,j} = 0$, find any Row k of the augmented matrix such that $a_{k,j} \neq 0$ and swap Rows j and k in that matrix.
4. Starting with $i \leftarrow j+1$:
   (a) Add $-\frac{a_{i,j}}{a_{j,j}}$ times Row j onto Row i in the augmented matrix.
   (b) If $i = n$, we are done, go to Step 5.
   (c) Otherwise, increment i (add one to i) and return to Step 4(a).
5. If $j = n$, we are done.
6. Otherwise, increment j (add one to j) and return to Step 3.

The first n columns of the augmented matrix should now be in row-echelon form, and also therefore should form an $n \times n$ upper-triangular matrix. Then solve the resulting system using backward substitution, where the last column of the augmented matrix becomes the new target vector.

For example, if we are trying to solve $\begin{pmatrix} 6 & 4 \\ -9 & -1 \end{pmatrix} \mathbf{u} = \begin{pmatrix} 8 \\ 13 \end{pmatrix}$, we create the augmented matrix

$$\left(\begin{array}{cc|c} 6 & 4 & 8 \\ -9 & -1 & 13 \end{array}\right)$$

and perform Gaussian elimination:

1. Add $-\frac{-9}{6}$ times Row 1 onto Row 2 to get $\left(\begin{array}{cc|c} 6 & 4 & 8 \\ 0 & 5 & 25 \end{array}\right)$.

This is in row-echelon form, so we may now do backward substitution to get $u_2 = 5$ and then $u_1 = -2$. Thus, the solution $\mathbf{u} = \begin{pmatrix} -2 \\ 5 \end{pmatrix}$.

Next, if we are trying to solve $\begin{pmatrix} 0 & -5 & 3 \\ 4 & 2 & 9 \\ 12 & -4 & 27 \end{pmatrix} \mathbf{u} = \begin{pmatrix} 12 \\ -11 \\ -3 \end{pmatrix}$, we create the augmented matrix

$$\left(\begin{array}{ccc|c} 0 & -5 & 3 & 12 \\ 4 & 2 & 9 & -11 \\ 12 & -4 & 27 & -3 \end{array}\right)$$

and perform Gaussian elimination:

1. Because entry (1, 1) is zero, swap Rows 1 and 2 to get $\left(\begin{array}{ccc|c} 4 & 2 & 9 & -11 \\ 0 & -5 & 3 & 12 \\ 12 & -4 & 27 & -3 \end{array}\right)$.
2. Add $-\frac{12}{4} = -3$ times Row 1 onto Row 3 to get $\left(\begin{array}{ccc|c} 4 & 2 & 9 & -11 \\ 0 & -5 & 3 & 12 \\ 0 & -10 & 0 & 30 \end{array}\right)$.
3. Add $-\frac{-10}{-5} = -2$ times Row 2 onto Row 3 to get $\left(\begin{array}{ccc|c} 4 & 2 & 9 & -11 \\ 0 & -5 & 3 & 12 \\ 0 & 0 & -6 & 6 \end{array}\right)$, and we are done.

We may now perform backward substitution to get $u_3 = -1$, then $u_2 = -3$ and finally $u_1 = 1$, so the solution is $\mathbf{u} = \begin{pmatrix} 1 \\ -3 \\ -1 \end{pmatrix}$.

Here is a third system of linear equations to be solved: $\begin{pmatrix} 5 & 2 & -4 & 1 \\ 5 & 2 & 1 & 0 \\ 10 & 8 & -10 & -1 \\ -15 & 6 & 16 & -11 \end{pmatrix} \mathbf{u} = \begin{pmatrix} -2 \\ 5 \\ 8 \\ 50 \end{pmatrix}$, we create the augmented matrix

$$\left(\begin{array}{cccc|c} 5 & 2 & -4 & 1 & -2 \\ 5 & 2 & 1 & 0 & 5 \\ 10 & 8 & -10 & -1 & 8 \\ -15 & 6 & 16 & -11 & 50 \end{array}\right)$$

and perform Gaussian elimination:

1. Add $-\frac{5}{5} = -1$ times Row 1 onto Row 2, add $-\frac{10}{5} = -2$ times Row 1 onto Row 3, and add $-\frac{-15}{5} = -3$ times Row 1 onto Row 4 to get $\left(\begin{array}{cccc|c} 5 & 2 & -4 & 1 & -2 \\ 0 & 0 & 5 & -1 & 7 \\ 0 & 4 & -2 & -3 & 12 \\ 0 & 12 & 4 & -8 & 44 \end{array}\right)$
2. Going to the next column, the entry at $(2, 2)$ is zero but $(3, 2)$ is non-zero, so we swap Rows 2 and 3 to get $\left(\begin{array}{cccc|c} 5 & 2 & -4 & 1 & -2 \\ 0 & 4 & -2 & -3 & 12 \\ 0 & 0 & 5 & -1 & 7 \\ 0 & 12 & 4 & -8 & 44 \end{array}\right)$
3. Add $-\frac{12}{4} = -3$ times Row 2 onto Row 4 to get $\left(\begin{array}{cccc|c} 5 & 2 & -4 & 1 & -2 \\ 0 & 4 & -2 & -3 & 12 \\ 0 & 0 & 5 & -1 & 7 \\ 0 & 0 & 10 & 1 & 8 \end{array}\right)$
4. Finally, we add $-\frac{10}{5} = -2$ times Row 3 onto Row 4 to get $\left(\begin{array}{cccc|c} 5 & 2 & -4 & 1 & -2 \\ 0 & 4 & -2 & -3 & 12 \\ 0 & 0 & 5 & -1 & 7 \\ 0 & 0 & 0 & 3 & -6 \end{array}\right)$

We may now perform backward substitution to get $u_4 = -2$, $u_3 = 1$, then $u_2 = 2$ and finally $u_1 = 0$, so the solution is $\mathbf{u} = \begin{pmatrix} 0 \\ 2 \\ 1 \\ -2 \end{pmatrix}$.

6 The transpose

Given an $m \times n$ matrix A, the "transpose" of this matrix is the $n \times m$ matrix $A^\top$ that has entries $a'_{i,j}$ where $a'_{i,j} = a_{j,i}$. That is, the diagonal of the matrix $A^\top$ is the same as the diagonal of A, but each other entry in $A^\top$ equals the entry of A that is "reflected" across the diagonal.

For example, here are three matrices and their transposes:

$$\text{If } A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{pmatrix}, \text{ then } A^\top = \begin{pmatrix} 1 & 5 & 9 & 13 \\ 2 & 6 & 10 & 14 \\ 3 & 7 & 11 & 15 \\ 4 & 8 & 12 & 16 \end{pmatrix}$$

$$\text{If } B = \begin{pmatrix} 0 & 3 & 6 & 1 & -3 \\ 8 & -5 & 7 & -9 & -1 \\ -7 & -6 & -2 & 0 & -8 \\ 5 & 9 & -4 & 2 & 4 \end{pmatrix}, \text{ then } B^\top = \begin{pmatrix} 0 & 8 & -7 & 5 \\ 3 & -5 & -6 & 9 \\ 6 & 7 & -2 & -4 \\ 1 & -9 & 0 & 2 \\ -3 & -1 & -8 & 4 \end{pmatrix}$$

$$\text{If } C = \begin{pmatrix} -4 & 9 & -4 \\ 2 & 3 & -5 \\ 5 & 1 & 8 \\ -6 & 7 & 6 \\ -1 & -8 & -2 \end{pmatrix}, \text{ then } C^\top = \begin{pmatrix} -4 & 2 & 5 & -6 & -1 \\ 9 & 3 & 1 & 7 & -8 \\ -4 & -5 & 8 & 6 & -2 \end{pmatrix}$$

You will notice that Row i of A becomes Column i of $A^\top$, and Column j of A becomes Row j of $A^\top$.

A square matrix is "symmetric" of $A = A^\top$. For example, the following matrices are symmetric:

$$\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{pmatrix}, \begin{pmatrix} 3.7 & 8.2 & -4.9 & 1.5 \\ 8.2 & 0.6 & -7.8 & 9.0 \\ -4.9 & -7.8 & 2.1 & -5.4 \\ 1.5 & 9.0 & -5.4 & 6.3 \end{pmatrix}.$$

If $A : \mathbf{R}^n \to \mathbf{R}^m$, then $A^\top : \mathbf{R}^m \to \mathbf{R}^n$. The most important property about the transpose is that if $\mathbf{u} \in \mathbf{R}^n$ and $\mathbf{v} \in \mathbf{R}^m$, then $A\mathbf{u} \in \mathbf{R}^m$, so we can calculate the dot product $(A\mathbf{u}) \cdot \mathbf{v}$, and not only is $A^\top \mathbf{v} \in \mathbf{R}^n$, it is always true that

$$(A\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (A^\top \mathbf{v}).$$

7 Matrix-matrix multiplication

If we have a polynomial $p(x)$, such as $p(x) = x^2 + 1$, then we can evaluate $p(1) = 2$. Given a second polynomial $q(x) = x^3 - 2x + 1$, then we can calculate $q(p(1)) = q(2) = 5$. We may ask: can we find a single polynomial called $q \circ p$ such that $(q \circ p)(1) = 5$, and more generally, $(q \circ p)(x) = q(p(x))$ for all real x? The answer is yes: we can expand $q(p(x)) = (x^2+1)^3 - 2(x^2+1) + 1 = x^6 + 3x^4 + x^2$, define this to be $(q \circ p)(x)$, and and we note that $(q \circ p)(1) = 1 + 3 + 1 = 5$. To find the polynomial $q \circ p$ is said to find the "composition" of the polynomials q and p.

If we have one matrix $A : \mathbf{R}^n \to \mathbf{R}^m$, and a second matrix $B : \mathbf{R}^m \to \mathbf{R}^\ell$, then if we start with a vector $\mathbf{u} \in \mathbf{R}^n$, then $A\mathbf{u} \in \mathbf{R}^m$; that is $A\mathbf{u}$ must be an m-dimensional vector. In that case, we can multiply an m-dimensional vector by B to get an ℓ-dimensional vector $B(A\mathbf{u})$. Can we find a single matrix, we'll call it BA, such that $(BA)\mathbf{u} = B(A\mathbf{u})$ for all vectors $\mathbf{u} \in \mathbf{R}^n$? In a process similar to compositing polynomials as described above, can we "compose" or "multiply" two matrices B and A to find a composition or product matrix BA?

Note: We use $q \circ p$ to represent the composition of q and p because qp generally represents the product of the two polynomials; however, because we do not generally use the element-wise product of two matrices of the same dimension, we simply juxtapose BA to represent the composition of B and A.

For example, if $A = \begin{pmatrix} 2 & 1 \\ -3 & 4 \end{pmatrix}$ and $B = \begin{pmatrix} 3 & -2 \\ 5 & 0 \end{pmatrix}$, we see that $A\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ and that $B\begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 15 \end{pmatrix}$. Your first guess might be to just multiply the entries element-wise, to get $\begin{pmatrix} 2 \cdot 3 & 1 \cdot (-2) \\ -3 \cdot 5 & 4 \cdot 0 \end{pmatrix} = \begin{pmatrix} 6 & -2 \\ -15 & 0 \end{pmatrix}$, but $\begin{pmatrix} 6 & -2 \\ -15 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ -15 \end{pmatrix}$, and therefore this approach does not work.

To find the matrix BA, the easiest way to do this is to consider $A = (\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \mathbf{a}_n)$, that is, an ordered collection of n m-dimensional column vectors $\mathbf{a}_1$ through $\mathbf{a}_n$. In this case, then,

$$A\mathbf{u} = A\mathbf{u} = u_1\mathbf{a}_1 + \cdots u_n\mathbf{a}_n.$$

Thus,

$$\begin{aligned} B(A\mathbf{u}) &= B\left(A\mathbf{u} = u_1\mathbf{a}_1 + \cdots u_n\mathbf{a}_n\right) \\ &= u_1 B\mathbf{a}_1 + \cdots u_n B\mathbf{a}_n \end{aligned}$$

because matrix-vector multiplication is linear: $B(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2) = \alpha_1 B\mathbf{v}_1 + \alpha_2 B\mathbf{v}_2$. However, look at that final sum of products: it is just the matrix with n columns $B\mathbf{a}_1$ through $B\mathbf{a}_n$ multiplied by the vector $\mathbf{u}$, so

$$B(A\mathbf{u}) = (B\mathbf{a}_1 \ \cdots \ B\mathbf{a}_n)\mathbf{u}.$$

Thus, we may calculate $BA = (B\mathbf{a}_1 \ B\mathbf{a}_2 \ \cdots B\mathbf{a}_n)$, or in other words, multiply each column of A by the matrix B. Using the above example, we have that $B\mathbf{a}_1 = \begin{pmatrix} 3 & -2 \\ 5 & 0 \end{pmatrix}\begin{pmatrix} 2 \\ -3 \end{pmatrix} = \begin{pmatrix} 12 \\ 10 \end{pmatrix}$ and $B\mathbf{a}_2 = \begin{pmatrix} 3 & -2 \\ 5 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 4 \end{pmatrix} = \begin{pmatrix} -5 \\ 5 \end{pmatrix}$, and therefore $BA = \begin{pmatrix} 12 & -5 \\ 10 & 5 \end{pmatrix}$, and we see that $(BA)\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 15 \end{pmatrix} = B(A\mathbf{u})$.

As another example, if $A = \begin{pmatrix} 3 & -1 & 2 \\ 4 & 3 & -1 \end{pmatrix}$ and $B = \begin{pmatrix} 2 & -4 \\ 1 & 3 \\ 2 & 1 \\ 0 & 5 \end{pmatrix}$, then we see $A : \mathbf{R}^3 \to \mathbf{R}^2$, and $B : \mathbf{R}^2 \to \mathbf{R}^4$, and so if we calculate $B(A\mathbf{u})$, this should take a 3-dimensional vector $\mathbf{u}$ and return a 4-dimensional vector. For example, $A\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 & -1 & 2 \\ 4 & 3 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 6 \end{pmatrix}$ and $B\begin{pmatrix} 4 \\ 6 \end{pmatrix} = \begin{pmatrix} 2 & -4 \\ 1 & 3 \\ 2 & 1 \\ 0 & 5 \end{pmatrix}\begin{pmatrix} 4 \\ 6 \end{pmatrix} = \begin{pmatrix} -16 \\ 22 \\ 14 \\ 30 \end{pmatrix}$. Can we find a single 4×3 matrix BA such that $(BA)\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -16 \\ 22 \\ 14 \\ 30 \end{pmatrix}$?

Because the matrices are not the same dimensions, it should be clear that we cannot multiply the entries element-wise. However, our previous algorithm still works: consider A as a matrix of vectors $A = (\mathbf{a}_1 \ \mathbf{a}_2 \ \mathbf{a}_3)$

and then let $BA = (B\mathbf{a}_1 \; B\mathbf{a}_2 \; B\mathbf{a}_3)$:

$$B\mathbf{a}_1 = \begin{pmatrix} 2 & -4 \\ 1 & 3 \\ 2 & 1 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} -10 \\ 15 \\ 10 \\ 20 \end{pmatrix}$$

$$B\mathbf{a}_2 = \begin{pmatrix} 2 & -4 \\ 1 & 3 \\ 2 & 1 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} -1 \\ 3 \end{pmatrix} = \begin{pmatrix} -14 \\ 8 \\ 1 \\ 15 \end{pmatrix}$$

$$B\mathbf{a}_3 = \begin{pmatrix} 2 & -4 \\ 1 & 3 \\ 2 & 1 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 8 \\ -1 \\ 3 \\ -5 \end{pmatrix}$$

Therefore, $BA = \begin{pmatrix} -10 & -14 & 8 \\ 15 & 8 & -1 \\ 10 & 1 & 3 \\ 20 & 15 & -5 \end{pmatrix}$ and we observe that $(BA)\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -16 \\ 22 \\ 14 \\ 30 \end{pmatrix}$, and this is also true for all other vectors $\mathbf{u} \in \mathbf{R}^3$.

The identity matrix I_m is the only matrix that has the property that $I_m A = A$ for all $m \times n$ matrices A, and $BI_m = B$ for all $\ell \times m$ matrices B.

If a square $n \times n$ matrix A is invertible, then if we calculate the following matrix-matrix products, we find: $AA^{-1} = A^{-1}A = I_n$, the $n \times n$ identity matrix.

Some properties of matrix-matrix multiplication parallel multiplication of real and complex numbers: $(B + C)A = BA + CA$, but matrix-matrix multiplication is not commutative. If A and B are both $n \times n$ square matrices, in general $AB \neq BA$. If A is $m \times n$ and B is $\ell \times m$ with at least one matrix not being square, then while BA is $\ell \times n$, either AB may not even be defined or it may have different dimensions from BA.

For example, observe that

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix} \quad \text{but} \quad \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 23 & 34 \\ 31 & 46 \end{pmatrix}.$$

7.1 Matrix-matrix multiplication in practice

While it is best to think of matrix-matrix multiplication BA as defining a new matrix were the j^{th} column is the matrix-vector product $B\mathbf{a}_j$ where $\mathbf{a}_j$ is Column j of A, this is a quicker way to calculate an individual entry: If B is $\ell \times m$ and A is $m \times n$, then $C = BA$ is $\ell \times n$ and the entry $c_{i,j}$ is assigned the sum of the element-wise products of Row i of the matrix B and Column j of the matrix A. For example, in this example, the entry $(3, 2)$ of the product, 45.4, is the sum of the element-wise products of Row 3 of the left-hand matrix B with the corresponding entries of Column 2 of the right-hand matrix A:

$$\begin{pmatrix} 5 & 2 & 3 & 1 \\ \mathbf{-4} & \mathbf{6} & \mathbf{0} & \mathbf{-5} \\ 7 & -3 & 9 & 8 \end{pmatrix} \begin{pmatrix} 1.5 & \mathbf{3.0} \\ -3.7 & \mathbf{4.9} \\ 2.4 & \mathbf{-1.7} \\ 0.8 & \mathbf{-5.6} \end{pmatrix}$$

$$= \begin{pmatrix} 5 \cdot 1.5 + 2 \cdot (-3.7) + 3 \cdot 2.4 + 1 \cdot 0.8 & 5 \cdot 3.0 + 2 \cdot 4.9 + 3 \cdot (-1.7) + 1 \cdot (-5.6) \\ -4 \cdot 1.5 + 6 \cdot (-3.7) + 0 \cdot 2.4 + (-5) \cdot 0.8 & \mathbf{-4} \cdot \mathbf{3.0} + \mathbf{6} \cdot \mathbf{4.9} + \mathbf{0} \cdot (\mathbf{-1.7}) + (\mathbf{-5}) \cdot (\mathbf{-5.6}) \\ 7 \cdot 1.5 + (-3) \cdot (-3.7) + 9 \cdot 2.4 + 8 \cdot 0.8 & 7 \cdot 3.0 + (-3) \cdot 4.9 + 9 \cdot (-1.7) + 8 \cdot (-5.6) \end{pmatrix}$$

$$= \begin{pmatrix} 8.1 & 14.1 \\ -32.2 & \mathbf{45.4} \\ 49.6 & -53.8 \end{pmatrix}.$$

7.2 Properties of the transpose

If A and B are $m \times n$ matrices numbers, then $A^{\top\top} = A$ and $(A + B)^\top = A^\top + B^\top$. If A is $m \times n$ and B is $\ell \times m$, then BA is $\ell \times n$ and $(BA)^\top = A^\top B^\top$. Also, for a square matrix A, $A = A^\top$ if and only if A is symmetric (where all eigenvalues are real), $A = -A^\top$ if and only if A is "skew-symmetric" (where all eigenvalues are imaginary), and $A^\top = A^{-1}$ if and only if A is orthogonal.

You will see that many of the properties of the transpose have analogous properties with respect to the complex conjugate of a complex number: $(w + z)^* = w^* + z^*$, $(wz)^* = w^* z^*$ (complex multiplication is commutative, matrix multiplication is not), $z = z^*$ if and only if z is real, $z = -z^*$ if and only if z is imaginary, and $z^* = z^{-1}$ if and only if z lies on the unit circle.

8 Orthogonal matrices

If the columns of an $n \times n$ square matrix form a collection of n mutually orthogonal unit vectors, so each column has a length or 2-norm of one, and each column is orthogonal to every other column, then we call that matrix "orthogonal".

An $n \times n$ matrix A is orthogonal if and only if $A^\top = A^{-1}$. Thus, calculating the inverse of an orthogonal matrix is as trivial as calculating the transpose. The other important property of orthogonal matrices is that they are the only matrices for which $\|A\mathbf{u}\|_2 = \|\mathbf{u}\|_2$ for all vectors $\mathbf{u} \in \mathbf{R}^n$; that is, multiplying a vector $\mathbf{u}$ by an orthogonal matrix does not change its length, only possibly its direction. This is true because

$$\begin{aligned}
\|A\mathbf{u}\|_2\|_2^2 &= (A\mathbf{u}) \cdot (A\mathbf{u}) \\
&= \mathbf{u} \cdot (A^\top(A\mathbf{u})) \qquad \text{because } (A\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (A^\top \mathbf{v})\text{, only here } v = A\mathbf{u}. \\
&= \mathbf{u} \cdot (A^{-1}(A\mathbf{u})) \qquad \text{because } A^\top = A^{-1}. \\
&= \mathbf{u} \cdot ((A^{-1}A)\mathbf{u}) \qquad \text{because } B(A\mathbf{u}) = (BA)\mathbf{u} \\
&= \mathbf{u} \cdot (I_n\mathbf{u}) \qquad \text{because } A^{-1}A = I_n. \\
&= \mathbf{u} \cdot \mathbf{u} \qquad \text{because } I_n\mathbf{u} = \mathbf{u}. \\
&= \|\mathbf{u}\|_2
\end{aligned}$$

All five of these statements are equivalent, so if one is true, all are true:

1. An $n \times n$ matrix A is orthogonal.
2. The columns of an $n \times n$ matrix A form a collection of n mutually orthogonal unit vectors.
3. The rows of an $n \times n$ matrix A form a collection of n mutually orthogonal unit vectors.
4. The transpose $A^\top$ of an $n \times n$ matrix A is orthogonal.
5. For an $n \times n$ matrix A, $A^{-1} = A^\top$.
6. For an $n \times n$ matrix A, $\|A\mathbf{u}\|_2 = \|\mathbf{u}\|_2$ for all vectors $\mathbf{u} \in \mathbf{R}^n$.

For example,

$$\text{If } A = \begin{pmatrix} 0.8 & 0.6 \\ -0.6 & 0.8 \end{pmatrix}, \text{ then } A^{-1} = A^\top = \begin{pmatrix} 0.8 & -0.6 \\ 0.6 & 0.8 \end{pmatrix}$$

$$\text{If } B = \begin{pmatrix} 0.48 & 0.6 & -0.64 \\ 0.8 & 0 & 0.6 \\ 0.36 & -0.8 & -0.48 \end{pmatrix}, \text{ then } B^{-1} = B^\top = \begin{pmatrix} 0.48 & 0.8 & 0.36 \\ 0.6 & 0 & -0.8 \\ -0.64 & 0.6 & -0.48 \end{pmatrix}$$

$$\text{If } C = \begin{pmatrix} 0.52 & 0.56 & -0.32 & -0.56 \\ 0.4 & 0.2 & -0.4 & 0.8 \\ 0.64 & -0.08 & 0.76 & 0.08 \\ -0.4 & 0.8 & 0.4 & 0.2 \end{pmatrix}, \text{ then } C^{-1} = C^\top = \begin{pmatrix} 0.52 & 0.4 & 0.64 & -0.4 \\ 0.56 & 0.2 & -0.08 & 0.8 \\ -0.32 & -0.4 & 0.76 & 0.4 \\ -0.56 & 0.8 & 0.08 & 0.2 \end{pmatrix}$$

If you examine these matrices carefully, you will see that all columns and rows are unit vectors, and that all columns or rows within one matrix are orthogonal. For example, looking at the second and third columns of $C^\top$, $0.4^2 + 0.2^2 + (-0.4)^2 + 0.8^2 = 0.16 + 0.04 + 0.16 + 0.64 = 1$ and $\sqrt{1} = 1$, while $0.4 \cdot 0.64 + 0.2 \cdot (-0.08) + (-0.4) \cdot 0.76 + 0.8 \cdot 0.08 = 0.256 - 0.016 - 0.304 + 0.064 = 0$. Unlike calculating the inverse of a diagonal matrix, calculating the inverse of an orthogonal matrix requires no floating-point calculations.

In the last case, note that if $\mathbf{u} = \begin{pmatrix} -7 \\ 2 \\ -10 \\ 4 \end{pmatrix}$, then $\|\mathbf{u}\|_2 = 13$, and $C\mathbf{u} = \begin{pmatrix} -1.56 \\ 4.8 \\ -11.92 \\ 1.2 \end{pmatrix}$ where $\|C\mathbf{u}\|_2 = 13$, and $C^\top\mathbf{u} = \begin{pmatrix} -10.84 \\ 0.48 \\ -4.56 \\ 5.52 \end{pmatrix}$ where $\|C^\top\mathbf{u}\|_2 = 13$, as well.

A "permutation matrix" is a matrix that has one "1" per column, and one "1" per row, and all other entries are zero. Every permutation matrix is an orthogonal matrix.

9 Eigenvalues and eigenvectors

A matrix A has an eigenvalue λ if there is a vector $\mathbf{u}$ such that $A\mathbf{u} = \lambda\mathbf{u}$; that is, the matrix A simply scales a vector, leaving it pointing in the same direction if $\lambda > 0$ or the opposite direction if $\lambda < 0$.

If $\mathbf{u}$ is an eigenvector corresponding to an eigenvalue λ, then all scalar multiples of $\mathbf{u}$ are also eigenvectors. Thus, it is common, though not necessary, to normalize eigenvectors.

If $A\mathbf{u} = \mathbf{u}$, then $\mathbf{u}$ is an eigenvector corresponding to the eigenvalue 1.

If A is an $n \times n$ square matrix and $A\mathbf{u} = \lambda\mathbf{u}$, then because $I_n\mathbf{u} = \mathbf{u}$, we may rewrite this as $A\mathbf{u} = \lambda I_n\mathbf{u}$. We can now collect all terms to the side, so $\lambda I_n\mathbf{u} - A\mathbf{u} = \mathbf{0}_n$. We can now use the property that $(A+B)\mathbf{u} = A\mathbf{u} + B\mathbf{u}$, so $(\lambda I_n - A)\mathbf{u} = \mathbf{0}_n$.

There is always the trivial solution to such a homogeneous system, because any matrix times the zero vector equals the zero vector. However, we want values of λ such that there is a non-zero vector $\mathbf{u}$ such that $(\lambda I_n - A)\mathbf{u} = \mathbf{0}_n$. The only way for this to happen is if we find a λ such that $\lambda I_n - A$ is not invertible. The only way you know how to do this is to calculate the determinant $\det(\lambda I_n - A)$, and you may recall that this creates a polynomial of degree n with real coefficients (this is called the "characteristic polynomial of the matrix A), and in our review of complex numbers, such a polynomial must have, counting multiplicity, n real or complex roots, and if a complex number z is a root, then its complex conjugate must also be a root. These n roots are called the eigenvalues of A.

Fortunately, if A is symmetric, all the roots of the characteristic polynomial are real, meaning that all the eigenvalues are also real. If you take any courses in quantum mechanics, this will be relevant.

One aside: Note that for a matrix, $A(\alpha\mathbf{u}_1 + \beta\mathbf{u}_2) = \alpha A\mathbf{u}_1 + \beta A\mathbf{u}_2$, but similarly, $\frac{\mathrm{d}}{\mathrm{d}x}(\alpha f(x) + \beta g(x)) = \alpha\frac{\mathrm{d}}{\mathrm{d}x}f(x) + \beta\frac{\mathrm{d}}{\mathrm{d}x}g(x)$, and so the derivative seems to behave very much like a matrix, just on functions. However, you can add two functions and you can multiply a function by a scalar, and there is a zero function, so functions seem to behave just like vectors.

Now, observe that $\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^{\lambda x} = \lambda\mathrm{e}^{\lambda x}$, that is, for an exponential function, given any λ, the derivative of $\mathrm{e}^{\lambda x}$ is a scalar multiple of itself, meaning that every real λ is an eigenvalue of the derivative, and the eigen*functions* associated with one specific eigenvalue λ are all scalar multiples $c\mathrm{e}^{\lambda x}$ where c is a real number.

10 Acknowledgments

Many thanks to Zarra Sarker for finding various mistakes and typos, as well as making valuable suggestions to make the material clearer for a second-year engineering student.