Representing Translation With Matrices

by

Alan G. Herron, Ph. D.

2012m02d04

http://www.GreatBlue.ca/

Introduction

Matrix algebra is commonly used to represent rotation and scaling of a reference frame. However, it is not commonly known that it can also be used to represent translation or frame offset. I first discovered the possibility while programming for a surveying (Geomatics) application, but the process was being used "blindly", without knowing how the representation was derived.

This paper corrects that deficiency for problems in one, two and three geometric dimensions, and illustrates the general process.

Given the prior use one might assume the theory has been developed formally somewhere else. This paper thus should be regarded as one written for convenience, with no intent to plagiarize the original work. I can't cite the original work because I don't know where it is. I do know that I went through my entire formal education, with some supplementary reading in matrix theory, without realizing the possibility existed, so the technique is not commonly known even among many matrix theorists.

One Dimensional Theory

A one dimensional transformation consists of a single linear equation, as shown in equation (1)

$$[u]=[a][x] \tag{1}$$

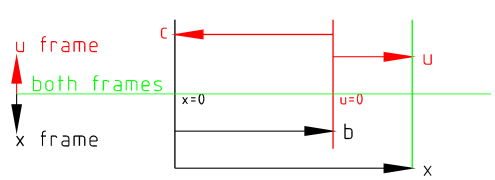
I have used matrix notation in order to foreshadow the subsequent development. The inverse is

$$[x]=[a]^{-1}[u] \tag{2}$$

The coefficient "matrix" here actually consists of a single number, a scale change. The matrix inverse is simply the reciprocal of the number, but the pattern will be found to generalize as we proceed.

Now let us add a translation term to (1) in accordance with Figure 1.

Figure 1



One Dimensional Transformation

$$[u]=[a][-b+x] \tag{3}$$

Note that the scale factor, [a], between the x and u frames is applied to both x and the offset, b. This is done because both x and b are measured in the units of the x frame. The scale factor must correct the displacement to the units of the u frame, and thus must be applied to both x and b, as shown.

In order to invert (3) we could simply solve for x, but there is another way. We could consider the offset, b, as a second variable in the original frame which must be transformed to an equivalent in the new frame. That is, we add a second equation to express the frame offset in the new frame, like this

$$\begin{aligned} c &= -a\,b + 0\,x \\ u &= -a\,b + a\,x \end{aligned} \tag{4}$$

In matrix form the two equations (4) become

$$\begin{bmatrix} c \\ u \end{bmatrix} = \begin{bmatrix} -a & 0 \\ -a & a \end{bmatrix} \begin{bmatrix} b \\ x \end{bmatrix} \tag{5}$$

By standard matrix processes we can get the inverse transformation, shown in (6)

$$\begin{bmatrix} b \\ x \end{bmatrix} = \frac{1}{-a^2} \begin{bmatrix} a & 0 \\ a & -a \end{bmatrix} \begin{bmatrix} c \\ u \end{bmatrix} \tag{6}$$

The offset in the u frame, c, is actually negative, as may be seen from (5) and from Figure 1. With this consideration the values for x and b from (6) are intuitively correct.

Note that if we were to take the minus sign from the determinant into the transformation matrix in (6) the result would be the original transformation matrix in (5). Also, the determinant turned out to be minus the scaling factor squared. Both are facts to which we will return later.

This illustrates the general process. We had to double the number of equations and adjust the usual form of (3), *viz.* u = a x + b, to recognize the scope of the scaling and the sign of the offset when seen as a transformation instead of as a simple linear equation.

Two Dimensional Theory

The simple two dimensional transformation, corresponding to (1) for one dimension, is shown in (7)

$$\begin{aligned} u_1 &= a_{11} x_1 + a_{12} x_2 \\ u_2 &= a_{21} x_1 + a_{22} x_2 \end{aligned} \tag{7}$$

In matrix form (7) is

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{8}$$

Two dimensional systems add the possibility of rotations to the transformation so the elements of [A] now contain both scale and rotation information. Alternately, one might consider the elements as axis-by-axis scale factors, without regard to rotational significance.

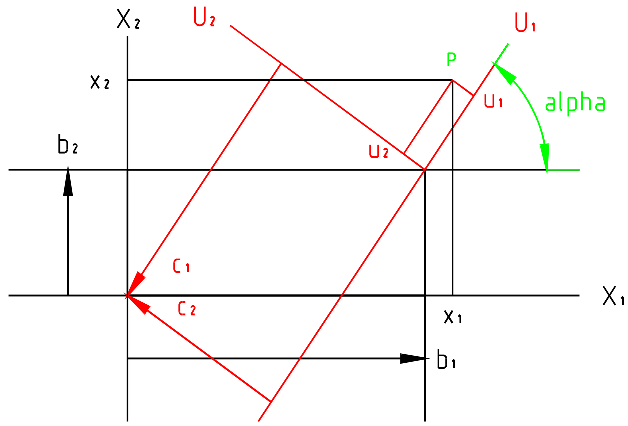
The inverse transformation is shown in (9)

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{a_{11}a_{22} - a_{21}a_{12}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{9}$$

Once again we add a translation term in accordance with Figure 2, as shown in (10)

Figure 2

\



Two Dimensional Transformation

$$\begin{aligned} u_1 &= a_{11}(-b_1 + x_1) + a_{12}(-b_2 + x_2) \\ u_2 &= a_{21}(-b_1 + x_1) + a_{22}(-b_2 + x_2) \end{aligned} \tag{10}$$

Grouping the variables and the offsets and adding the offset equations gives (11)

$$\begin{aligned} c_1 &= -a_{11}b_1 - a_{12}b_2 + 0x_1 + 0x_2 \\ c_2 &= -a_{21}b_1 - a_{22}b_2 + 0x_1 + 0x_2 \\ u_1 &= -a_{11}b_1 - a_{12}b_2 + a_{11}x_1 + a_{12}x_2 \\ u_2 &= -a_{21}b_1 - a_{22}b_2 + a_{21}x_1 + a_{22}x_2 \end{aligned} \tag{11}$$

In matrix form this becomes (12)

$$\begin{bmatrix} c_1 \\ c_2 \\ u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} -a_{11} & -a_{12} & 0 & 0 \\ -a_{21} & -a_{22} & 0 & 0 \\ -a_{11} & -a_{12} & a_{11} & a_{12} \\ -a_{21} & -a_{22} & a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ x_1 \\ x_2 \end{bmatrix} \tag{12}$$

The elements of the transformation matrix contain the rotational information corresponding to angle alpha in Figure 2 and any scaling, such as that resulting from unit changes.

At this point it becomes apparent that the transformation matrix actually consists of four two-by-two sub-matrices; three copies of the original in (8) with signs adjusted, and a zero. Conceptually the x vector and the b vector are transformed by the same sub-matrix, so the extended system shows the redundancy. The key feature is the zeroes in the upper right corner, which expresses the lack of dependence of the offsets on the variable part of the input vector. This structure in the system enables us to simplify the determination of the inverse from the inverse of the basic sub-matrix.

The inverse is, in fact, the matrix shown in (13)

$$\begin{bmatrix} b_1 \\ b_2 \\ x_1 \\ x_2 \end{bmatrix} = \frac{1}{det[T]} \begin{bmatrix} -a_{22} & a_{12} & 0 & 0 \\ a_{21} & -a_{11} & 0 & 0 \\ -a_{22} & a_{12} & a_{22} & -a_{12} \\ a_{21} & -a_{11} & -a_{21} & a_{11} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ u_1 \\ u_2 \end{bmatrix} \tag{13}$$

Where [T] is the 4x4 matrix in (12). To confirm this: The product of the transformation matrix [T] and its inverse in (13) (neglecting the determinant) is

$$\begin{matrix} a_{11}a_{22}-a_{12}a_{21} & 0 & 0 & 0 \\ 0 & a_{11}a_{22}-a_{12}a_{21} & 0 & 0 \\ 0 & 0 & a_{11}a_{22}-a_{12}a_{21} & 0 \\ 0 & 0 & 0 & a_{11}a_{22}-a_{12}a_{21} \end{matrix} \tag{14}$$

By expansion across the first row of (12) the determinant is

$$\begin{aligned} det[T] &= a_{11}(a_{11}a_{22}^2 - a_{22}a_{12}a_{21}) - a_{21}(a_{11}a_{22}a_{12} - a_{21}a_{12}^2) - 0 + 0 \\ &= a_{11}^2 a_{22}^2 - a_{11}a_{22}a_{12}a_{21} - a_{11}a_{22}a_{12}a_{21} + a_{12}^2 a_{21}^2 \\ &= (a_{11}a_{22} - a_{12}a_{21})^2 \end{aligned} \tag{15}$$

Applying this to (14) we get

$$[T]^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{16}$$

The determinant of the sub-matrix is 1 if the axes are perpendicular in both the old and new frames, and there is no scaling, for then

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} \cos(alpha) & \sin(alpha) \\ -\sin(alpha) & \cos(alpha) \end{bmatrix} \tag{17}$$

The square in det[T] allows for non-orthogonal axes. If this created a degenerate situation the square would force the inverse to go singular despite the cancellation of one power.

An Alternate Perspective

The process just discussed, while valid, tends to obscure the essence of the perspective that is the primary point of this paper. But we need the foregoing to support the needed abstractions.

Consider (12). If we partition the vectors into their variable and constant parts, and partition the transformation matrix to correspond, the following pattern emerges:

$$\begin{bmatrix}[C]\\ [U]\end{bmatrix}=\begin{bmatrix}-[A] & [0]\\ -[A] & [A]\end{bmatrix}\begin{bmatrix}[B]\\ [X]\end{bmatrix} \tag{18}$$

In this form the fundamental pattern of the process emerges more clearly. The sub-matrix, [A], is repeated in the overall transformation. It is used

- in the 11 position, to transform the offset components in the input frame to the offset components in the output frame
- in the 21 position, to transform the origin (tail) of the vector [X]
- in the 22 position, to transform the variable components (tip) of the input vector, [X]

The upper right (12) partition, denoted here as [0], is used to exclude the influence of the input variable [X] on the offset expressed in the output frame, [C]. We need the transformation of offsets, isolated from the variables, in order to facilitate inversion, thus completing the symmetry of the process.
In this form the dimensionality of the basic vectors becomes irrelevant: one, two, three dimensions...it doesn't matter. The form of the partitioned matrix system still applies.

What is also important is that matrix calculations, such as determinants and inverses, can be performed on the partition [A], and combined <u>after</u> calculation to derive the overall equivalents using the known form of the overall transformation. The supporting argument runs as follows:

- The first partitioned equation, expressing [C] as a function of [B], can stand alone, since it is independent of [X] for all values thereof. This implies that the inverse equation, expressing [B] as a function of [C], involves only the inverse of [A] in the 11 position, and is unaffected by the other partitioned equation.
- The offset vectors must transform the same way in the second partitioned equation as they do in the first, implying the 21 partition in the inverse is the same as that in the 11 position.
- The variable part of the first partitioned equation must transform by the same equation as the offset part, since the latter is just a particular example of the former.

Together these constraints imply that

$$\begin{bmatrix}[B]\\ [X]\end{bmatrix}=\begin{bmatrix}(-[A])^{-1} & [0]\\ (-[A])^{-1} & [A]^{-1}\end{bmatrix}\begin{bmatrix}[C]\\ [U]\end{bmatrix} \tag{19}$$

In other words, we can calculate the inverse of the four row system simply by entering the inverse of the two row system in the system transformation with appropriate signs, as shown.

In computing the inverse of [A] we divide by the determinant. When we combine the two partitioned equations this means we get one copy of the determinant for each partition row. If we factor the full determinant out, we need to square the determinant of [A] to get the determinant of [T].

Three Dimensional Theory

The basic transformation with translation in three dimensions, analogous to (12) in two dimensions, is

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} -a_{11} & -a_{12} & -a_{13} & 0 & 0 & 0 \\ -a_{21} & -a_{22} & -a_{23} & 0 & 0 & 0 \\ -a_{31} & -a_{32} & -a_{33} & 0 & 0 & 0 \\ -a_{11} & -a_{12} & -a_{13} & a_{11} & a_{12} & a_{13} \\ -a_{21} & -a_{22} & -a_{23} & a_{21} & a_{22} & a_{23} \\ -a_{31} & -a_{32} & -a_{33} & a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{20}$$

That is, using our partitioned notation

$$[A] = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{21}$$

The sub-determinant and determinant are

$$det[A] = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12} \tag{22}$$

$$det[T] = -det^2[A] \tag{23}$$

The sub-inverse is

$$[A]^{-1} = \frac{1}{det[A]} \begin{bmatrix} a_{22}a_{33} - a_{32}a_{23} & a_{13}a_{32} - a_{12}a_{33} & a_{12}a_{23} - a_{22}a_{13} \\ a_{31}a_{23} - a_{21}a_{33} & a_{11}a_{33} - a_{31}a_{13} & a_{21}a_{13} - a_{11}a_{23} \\ a_{21}a_{32} - a_{22}a_{31} & a_{12}a_{31} - a_{11}a_{32} & a_{11}a_{22} - a_{21}a_{12} \end{bmatrix} \tag{24}$$

With these definitions we have the three dimensional system by simple substitutions in the general partitioned forms (18) and (19).

An Example: Celsius From / To Fahrenheit

The conversion between Celsius and Fahrenheit temperatures serves as a convenient example of a one dimensional transformation involving translation or offset. Let us set up the transformation regarding Fahrenheit as the input frame and Celsius as the output.

The freezing point of water is 0 C and 32 F, so the offset is 32. That is, the origin of the Celsius frame has an offset of 32 degrees in Fahrenheit units (degrees) in the Fahrenheit frame.

The interval between the freezing and boiling points of water is 100 C units and 180 F units. The scale factor which must be applied to temperatures measured in F is, therefore, 100 / 180 or 1 / 1.8.

The conversion in matrix form thus becomes (25)

$$\begin{bmatrix} -32/1.8 \\ C \end{bmatrix} = \begin{bmatrix} -1/1.8 & 0 \\ -1/1.8 & 1/1.8 \end{bmatrix} \begin{bmatrix} 32 \\ F \end{bmatrix} \tag{25}$$

The inverse of this, obtained and verified by matrix algebra, is (26)

$$\begin{bmatrix} 32 \\ F \end{bmatrix} = \begin{bmatrix} -1.8 & 0 \\ -1.8 & 1.8 \end{bmatrix} \begin{bmatrix} -32/1.8 \\ C \end{bmatrix} \tag{26}$$

Conclusion

In summary: translation may be handled in matrix form with the equations indicated in (18) and (19), where the dimensions of the sub-matrices and sub-vectors reflect the dimensionality inherent in the problem at hand. Singular transformations are avoided by recognizing that the offset information must be preserved as well as the information about the tip of the variable vector.

If we regard the vector which specifies the offset as simply a general vector in a <u>single</u> frame the process becomes a means to describe the addition or subtraction of two vectors. In fact, there is nothing preventing us adding more rows and columns to the partitioned matrix to obtain sums or differences of several vectors, not just two. We simply place sub-matrices into the general partitioned transformation with signs and positions appropriate to the sums or differences desired.

The ability to handle translation or vector addition with matrix algebra, especially by a standardized procedure, is potentially useful in many disciplines. It is actually surprising the technique is not more widely known. I trust this exposition will help alleviate that.

If someone is able to provide a citation to the original exposition of the technique I would appreciate being informed of same.

Finally, I wish to thank my companion, Tannis Ewing, for her patience and suggestions while reading the drafts. Familiarity, unfortunately, tends to bias the meaning of “obvious”, a problem her comments help to correct.