Closed-Form Solution of Absolute Orientation Using Orthonormal Matrices.

Berthold K.P. Horn*,
Department of Electrical Engineering,
University of Hawaii at Manoa,
Honolulu, Hawaii 96822.

Hugh M. Hilden,
Department of Mathematics,
University of Hawaii at Manoa,
Honolulu, Hawaii 96822.

Shahriar Negahdaripour**,
Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139.

*On leave from: Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

**Now at: Department of Electrical Engineering, University of Hawaii at Manoa, Honolulu, Hawaii 96822.

Closed-Form Solution of Absolute Orientation Using Orthonormal Matrices.

B.K.P. Horn, H.M. Hilden & S. Negahdaripour

Abstract: Finding the relationship between two coordinate systems using pairs of measurements of the coordinates of a number of points in both systems is a classic photogrammetric task. It finds applications in stereophotogrammetry and in robotics. We present here a closed-form solution to the least-squares problem for three or more points. Currently, various empirical, graphical and numerical iterative methods are in use. Derivation of a closed-form solution can be simplified by using unit quaternions to represent rotation, as was shown in an earlier paper[1]. Since orthonormal matrices are more widely used to represent rotation, we now present a solution using 3×3 matrices. Our method requires the computation of the square-root of a symmetric matrix. We compare the new result with an alternate method where orthonormality is not directly enforced. In this other method a best fit linear transformation is found and then the nearest orthonormal matrix chosen for the rotation.

We note that the best translational offset is the difference between the centroid of the coordinates in one system and the rotated and scaled centroid of the coordinates in the other system. The best scale is equal to the ratio of the root-mean-square deviations of the coordinates in the two systems from their respective centroids. These exact results are to be preferred to approximate methods based on measurements of a few selected points.

1. Origin of the Problem.

Suppose that we are given the coordinates of a number of points as measured in two different Cartesian coordinate systems (see Fig. 1). The photogrammetric problem of recovering the transformation between the two systems from these measurements is referred to as that of *absolute orientation*[2,3,4]. It occurs in several contexts, foremost in relating a stereo model developed from pairs of aerial photographs to a geodetic coordinate system. It also is of importance in robotics, where measurements in a camera coordinate system must be related to coordinates in a system attached to a mechanical manipulator. Here one speaks of the determination of the "hand-eye" transform[5].

1.A Previous Work.

The problem of absolute orientation is usually treated in an empirical, graphic or numerical, iterative fashion[2,3,4]. Thompson[6] gives a solution to this problem when exactly three points are measured. His method, as well as the simpler one of Schut[7], depends on selective neglect of the extra constraints available when all coordinates of three points are known, as dicussed later. Schut uses unit quaternions and arrives at a set of linear equations. A simpler solution, that does not require the solution of a system of linear equations, was presented in a precursor of this paper[1]. These three methods all suffer from the defect that they cannot handle more than three points. Perhaps more importantly, they do not even use all of the information available from the three points.

Oswal and Balasubramanian[8] developed a least-squares method that can handle more than three points, *but* their method does not enforce the orthonormality of the rotation matrix. Instead, they simply find the best-fit linear transform. An iterative method is then used to "square up" the result—bringing it closer to being orthonormal. Their method for doing this is iterative (and without mathematical justification). In addition, the result obtained is *not* the solution of the original least-squares problem.

We study their approach in section 4 using a closed-form solution for the nearest orthonormal matrix derived in section 3.F. This is apparently not entirely novel, since an equivalent problem has been treated in the psychological literature[15−21] (as recently pointed out to us by Takeo Kanade). The existing methods, however, cannot deal with a singular matrix. We extend our method to deal with the case where the rank deficiency of the matrix is one. This is an important extension, since the matrix will be singular when either of the sets of measurements are coplanar, as will always happen when there are only three measurements.

The main result presented here, however, is the closed-form solution to the least-squares problem of absolute orientation. Our new result *can* be applied in the special case when one or the other of the sets of measurements happens to be coplanar. This is important, because sometimes only three points are available, and three points are, of course, always coplanar. The solution we present differs from the schemes discussed at the beginning of this section in that it does not selectively neglect information provided by the measurements—it uses all of it.

We should point out that a version of this problem has been solved by Farrel & Stuelpnagel[9] (as C.J. Standish pointed out to us after reading a draft of our paper). Their solution only applies, however, when neither of the sets of measurements is coplanar. We also learned recently that S. Arun, T. Huang and S.D. Blostein independently developed a solution to an equivalent problem[10]. They use singular value decomposition of an arbitrary matrix instead of the eigenvalue-eigenvector decomposition of a symmetric matrix inherent in our approach.

1.B Minimum Number of Points.

The transformation between two Cartesian coordinate systems can be thought of as the result of a rigid-body motion and can thus be decomposed into a rotation and a translation. In stereophotogrammetry, in addition, the scale may not be known. There are obviously three degrees of freedom to translation. Rotation has another three (direction of the axis about which the rotation takes place plus the angle of rotation about this axis). Scaling adds one more degree of freedom. Three points known in both coordinate systems provide nine constraints (three coordinates each), more than enough to allow determination of the seven unknowns, as shown, for example, in reference 1. By discarding two of the constraints, seven equations in seven unknowns can be developed that allow one to recover the parameters.

1.C Least Sum of Squares of Errors.

In practice, measurements are not exact, and so greater accuracy in determining the transformation parameters will be sought after by using more than three points. We no longer expect to be able to find a transformation that maps the measured coordinates of points in one system *exactly* into the measured coordinates of these points in the other. Instead, we minimize the sum of the squares of the residual errors. Finding the best set of transformation parameters is not easy. In practice, various empirical, graphical and numerical procedures are in use[2,3,4]. These are all iterative in nature. That is, given an approximate solution, such a method is applied repeatedly until the remaining error becomes negligible.

At times information is available that permits one to obtain so good an initial guess of the transformation parameters, that a single step of the iteration brings one close enough to the true solution of the least-squares problem for all practical purposes; but this is rare.

1.D Closed-Form Solution.

In this paper we present a closed-form solution to the least-squares problem of absolute orientation, one that does not require iteration. One advantage of a closed-form solution is that it provides us in one step with the best possible tranformation, given the measurements of the points in the two coordinate systems. Another advantage is that one need not find a good initial guess, as one does when an iterative method is used.

A solution of this problem was presented previously that uses unit quaternions to represent rotations[1]. The solution for the desired quaternion was shown to be the eigenvector of a symmetric 4×4 matrix associated with the largest positive eigenvalue. The elements of this matrix are simple combinations of sums of products of coordinates of the points. To find the eigenvalues, a quartic equation has to be solved whose coefficients are sums of products of elements of the matrix. It was shown that this quartic is particularly simple, since one of its coefficients is zero. It simplifies even more when only three points are used.

1.E Orthonormal Matrices.

While unit quaternions constitute an elegant representation for rotation, most of us are more familiar with the use of proper orthonormal matrices for this purpose. Working directly with matrices is difficult, because of the need to deal with six nonlinear constraints that ensure that the matrix is orthonormal. We nevertheless are able to derive a solution for the rotation matrix using direct manipulation of 3×3 matrices. This closed-form solution requires the computation of the positive semi-definite square-root of a positive semi-definite matrix. We show in section 3.C how such a square-root can be found, once the eigenvalues and eigenvectors of the matrix are available. Finding the eigenvalues requires the solution of a cubic equation.

The method discussed here finds the same solution as does the method presented earlier that uses unit quaternions to represent rotation, since it minimizes the same error sum[1]. We present the new method only because the use of orthonormal matrices is so widespread. We actually consider the solution using unit quaternions to be more elegant.

1.F Symmetry of the Solution.

Let us call the two coordinate systems "left" and "right." A desirable property of a solution method is that when applied to the problem of finding the best transformation from the left to the right system, it gives the exact inverse of the best transformation from the left to the right system. It was shown in reference 1 that the scale factor has to be treated in a particular way to guarantee that this happens. The method we develop here for directly computing the rotation matrix gives two apparently different results when applied to the problem of finding the best transformation from left to right and the problem of finding the best transformation from right to left. We show that these two results are in fact different forms of the same solution and that our method does indeed have the sought after symmetry property.

1.G The Nearest Orthonormal Matrix.

Since the constraint of orthonormality leads to difficulties, some in the past have chosen to find a 3×3 matrix that fits the data best in a least-squares sense *without* constraint on its elements[8]. The result will typically not be orthonormal. If the data is fairly accurate, the matrix may be almost orthonormal. In this case, we might wish to find the "nearest" orthonormal matrix. That is, we wish to minimize the sum of the squares of differences between the elements of the matrix obtained from the measurements and an ideal orthonormal matrix. Iterative methods exist for finding the nearest orthonormal matrix.

A closed-form solution, shown in section 3.F, again involves square-roots of 3×3 symmetric matrices. The answer obtained this way is different, however, from the solution that minimizes the sum of the squares of the residual errors. In particular, it does not have the highly desirable symmetry property discussed above, and it requires the accumulation of a larger number of sums of products of coordinates of measured points.

2. Solution Methods

As we shall see, the translation and the scale factor are easy to determine once the rotation is known. The difficult part of the problem is finding the rotation. Given three noncollinear points, we can easily construct a useful triad in each of the left and the right coordinate systems[1] (see

Fig. 2). Take the line from the first to the second point to be the direction of the new x-axis. Place the new y-axis at right angles to the new x-axis in the plane formed by the three points. The new z-axis is then made orthogonal to the x- and y-axis with orientation chosen to satisfy the right-hand rule. This construction is carried out in both left and right systems. The rotation that takes one of these constructed triads into the other is also the rotation that relates the two underlying Cartesian coordinate systems. This rotation is easy to find, as is shown in reference 1.

This *ad hoc* method constitutes a "closed-form" solution for finding the rotation given three points. Note that it uses the information from the three points selectively. Indeed, if we renumber the points, we obtain a different rotation matrix (unless the data happens to be perfect). Also note that the method cannot be extended to deal with more than three points. Even with just three points we should really attack this problem by means of a least-squares method, since there are more constraints than unknown parameters. The least-squares solution for translation and scale will be given in subsections 2.B and 2.C. The optimum rotation will be found in section 4.

2.A Finding the Translation.

Let there by n points. The measured coordinates in the left and right coordinate system will be denoted by

$$\{\mathbf{r}_{l,i}\} \quad \text{and} \quad \{\mathbf{r}_{r,i}\}$$

respectively, where i ranges from 1 to n. We are looking for a transformation of the form

$$\mathbf{r}_r = s\,R(\mathbf{r}_l) + \mathbf{r}_0$$

from the left to the right coordinate system. Here s is a scale factor, $\mathbf{r}_0$ is the translational offset, and $R(\mathbf{r}_l)$ denotes the rotated version of the vector $\mathbf{r}_l$. We do not, for the moment, use any particular notation for rotation. We use only the fact that rotation is a linear operation and that it preserves lengths so that

$$\|R(\mathbf{r}_l)\|^2 = \|\mathbf{r}_l\|^2 ,$$

where $\|\mathbf{r}\|^2 = \mathbf{r}\cdot\mathbf{r}$ is the square of the length of the vector $\mathbf{r}$.

Unless the data are perfect, we will not be able to find a scale factor, a translation, and a rotation such that the transformation equation above is satisfied for each point. Instead there will be a residual error,

$$\mathbf{e}_i = \mathbf{r}_{r,i} - s\,R(\mathbf{r}_{l,i}) - \mathbf{r}_0 .$$

We will minimize the sum of the squares of these errors,

$$\sum_{i=1}^{n} \|\mathbf{e}_i\|^2 .$$

(It was shown in reference 1 that the measurements can be weighted without changing the basic solution method.)

We consider the variation of the total error first with translation, then with scale, and finally with respect to rotation.

2.B Centroids of the Sets of Measurements.

It turns out to be useful to refer all measurements to the centroids defined by

$$\bar{\mathbf{r}}_l = \frac{1}{n}\sum_{i=1}^{n} \mathbf{r}_{l,i} \quad \text{and} \quad \bar{\mathbf{r}}_r = \frac{1}{n}\sum_{i=1}^{n} \mathbf{r}_{r,i}.$$

Let us denote the new coordinates by

$$\mathbf{r}'_{l,i} = \mathbf{r}_{l,i} - \bar{\mathbf{r}}_l \quad \text{and} \quad \mathbf{r}'_{r,i} = \mathbf{r}_{r,i} - \bar{\mathbf{r}}_r.$$

Note that

$$\sum_{i=1}^{n} \mathbf{r}'_{l,i} = \mathbf{0} \quad \text{and} \quad \sum_{i=1}^{n} \mathbf{r}'_{r,i} = \mathbf{0}.$$

Now the error term can be rewritten in the form

$$\mathbf{e}_i = \mathbf{r}'_{r,i} - s\,R(\mathbf{r}'_{l,i}) - \mathbf{r}'_0,$$

where

$$\mathbf{r}'_0 = \mathbf{r}_0 - \bar{\mathbf{r}}_r + s\,R(\bar{\mathbf{r}}_l).$$

The sum of the squares of the errors becomes

$$\sum_{i=1}^{n} \left\|\mathbf{r}'_{r,i} - s\,R(\mathbf{r}'_{l,i}) - \mathbf{r}'_0\right\|^2 ,$$

or

$$\sum_{i=1}^{n} \left\|\mathbf{r}'_{r,i} - s\,R(\mathbf{r}'_{l,i})\right\|^2 - 2\,\mathbf{r}'_0 \cdot \sum_{i=1}^{n} \left(\mathbf{r}'_{r,i} - s\,R(\mathbf{r}'_{l,i})\right) + n\,\|\mathbf{r}'_0\|^2 .$$

Now the sum in the middle of this expression is zero, since the sum of the vectors $\{\mathbf{r}'_{l,i}\}$ and the sum of the vectors $\{\mathbf{r}'_{r,i}\}$ are zero, as mentioned above. As a result, we are left with the first and the third terms. The first does not depend on $\mathbf{r}_0'$, and the last cannot be negative. The total error is obviously minimized with $\mathbf{r}_0' = \mathbf{0}$, or

$$\mathbf{r}_0 = \bar{\mathbf{r}}_r - s\,R(\bar{\mathbf{r}}_l).$$

That is, the translation is just the difference of the right centroid and the scaled and rotated left centroid. We return to this equation to find the translational offset once we have found the scale and rotation. This method, based on all available information, is to be preferred to one that uses only measurements of one or a few selected points to estimate the translation.

At this point we note that the error term can be simplified to read

$$\mathbf{e}_i = \mathbf{r}_{r,i}' - s\,R(\mathbf{r}_{l,i}'),$$

since $\mathbf{r}_0' = \mathbf{0}$, and so the total error to be minimized is just

$$\sum_{i=1}^{n} \left\|\mathbf{r}_{r,i}' - s\,R(\mathbf{r}_{l,i}')\right\|^2 .$$

2.C Symmetry in Scale.

It is shown in reference 1 that the above formulation of the error term leads to an asymmetry in the determination of the optimal scale factor. That is, the "optimal" transformation from the left to the right coordinate system is then not the exact inverse of the "optimal" transformation from the right to the left coordinate system. The latter corresponds to use of the error term

$$\mathbf{e}_i = \mathbf{r}_{l,i}' - (1/s)R^T(\mathbf{r}_{r,i}'),$$

or

$$\mathbf{e}_i = -\big((1/s)(\mathbf{r}_{r,i}') - R(\mathbf{r}_{l,i}')\big),$$

and leads to a total error to be minimized of

$$\sum_{i=1}^{n} \left\|(1/s)(\mathbf{r}_{r,i}') - R(\mathbf{r}_{l,i}')\right\|^2 .$$

If the errors in both sets of measurements are similar, it is more reasonable to use a symmetrical expression for the error term:

$$\mathbf{e}_i = \frac{1}{\sqrt{s}}\,\mathbf{r}_{r,i}' - \sqrt{s}\,R(\mathbf{r}_{l,i}').$$

Then the total error becomes

$$\frac{1}{s}\sum_{i=1}^{n}\left\|\mathbf{r}'_{r,i}\right\|^2 - 2\sum_{i=1}^{n}\mathbf{r}'_{r,i}\cdot(R(\mathbf{r}'_{l,i})) + s\sum_{i=1}^{n}\left\|\mathbf{r}'_{l,i}\right\|^2,$$

or

$$\frac{1}{s}S_r - 2D + sS_l,$$

where

$$S_l = \sum_{i=1}^{n}\left\|\mathbf{r}'_{l,i}\right\|^2, \quad D = \sum_{i=1}^{n}\mathbf{r}'_{r,i}\cdot(R(\mathbf{r}'_{l,i})), \quad \text{and} \quad S_r = \sum_{i=1}^{n}\left\|\mathbf{r}'_{r,i}\right\|^2.$$

Completing the square in s, we obtain

$$\left(\sqrt{s}\sqrt{S_l} - \frac{1}{\sqrt{s}}\sqrt{S_r}\right)^2 + 2(\sqrt{S_l S_r} - D).$$

This is minimized with respect to scale s when the first term is zero or $s = \sqrt{S_r/S_l}$, that is,

$$s = \sqrt{\sum_{i=1}^{n}\left\|\mathbf{r}'_{r,i}\right\|^2 \bigg/ \sum_{i=1}^{n}\left\|\mathbf{r}'_{l,i}\right\|^2}.$$

One advantage of this symmetrical result is that it allows one to determine the scale without the need to know the rotation. Importantly, however, the determination of the rotation is *not* affected by the choice of the value of the scale factor. In each case the remaining error is minimized when D is as large as possible. That is, we have to choose the rotation that makes

$$\sum_{i=1}^{n}\mathbf{r}'_{r,i}\cdot(R(\mathbf{r}'_{l,i}))$$

as large as possible.

3. Dealing with Rotation.

There are many ways to represent rotation, including the following: Euler angles, Gibbs vector, Cayley-Klein parameters, Pauli spin matrices, axis and angle, orthonormal matrices, and Hamilton's quaternions[11,12]. Of these representations, orthonormal matrices have been used most often in photogrammetry, graphics and robotics. While unit quaternions have many advantages when used to represent rotation, few are familiar with their properties. That is why we present here a closed-form solution using orthonormal matrices that is similar to the closed-form solution obtained earlier using unit quaternions[1].

The new method, which we present next, depends on eigenvalue-eigenvector decomposition of a 3×3 matrix and so requires solution of a cubic equation. Well-known methods such as Ferrari's solution can be used[12,13,14]. When one or the other sets of measurements (left or right) is coplanar, the method simplifies, in that only a quadratic needs to be solved. It turns out, however, that much of the complexity of this approach stems from the need to deal with this and other special cases.

3.A Best Fit Orthonormal Matrix.

We have to find a matrix R that maximizes

$$\sum_{i=1}^{n} \mathbf{r}'_{r,i} \cdot \left(R(\mathbf{r}'_{l,i})\right) = \sum_{i=1}^{n} (\mathbf{r}'_{r,i})^T R\, \mathbf{r}'_{l,i}.$$

Now

$$\mathbf{a}^T R\, \mathbf{b} = \mathrm{Trace}(R^T \mathbf{a}\mathbf{b}^T),$$

so we can rewrite the above expression in the form

$$\mathrm{Trace}\left(R^T \sum_{i=1}^{n} \mathbf{r}'_{r,i}\, (\mathbf{r}'_{l,i})^T\right) = \mathrm{Trace}(R^T M),$$

where

$$M = \sum_{i=1}^{n} \mathbf{r}'_{r,i}\, (\mathbf{r}'_{l,i})^T,$$

that is,

$$M = \begin{pmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{pmatrix},$$

with

$$S_{xx} = \sum_{i=1}^{n} x_{r,i}\, x_{l,i}, \quad S_{xy} = \sum_{i=1}^{n} x_{r,i}\, y_{l,i}, \quad \ldots$$

and so on*.

To find the rotation that minimizes the residual error, we have to find the orthonormal matrix R that maximizes

$$\mathrm{Trace}(R^T M).$$

*We denote the elements of the matrix S_{xx}, S_{xy} …, rather than M_{xx}, M_{xy} …, in order to be consistent with reference 1.

3.B Product of Orthonormal and Symmetric Matrix.

It follows from Theorem 1 (p. 169) in [22] that a square matrix M can be decomposed into the product of an orthonormal matrix U and a positive semi-definite matrix S. The matrix S is always uniquely determined. The matrix U is uniquely determined when M is nonsingular. (We show in section 3.D that U can also be determined up to a two-way ambiguity when M is singular, but has rank deficiency one). When M is nonsingular, we can actually write directly

$$M = US,$$

where

$$S = (M^T M)^{1/2}$$

is the positive-definite square-root of the symmetric matrix $M^T M$, while

$$U = M(M^T M)^{-1/2}$$

an orthonormal matrix. It is easy to verify that $M = US$, $S^T = S$ and $U^T U = I$.

3.C Positive-Definite Square-Root of Positive-Definite Matrix.

The matrix $M^T M$ can be written in terms of the set of its eigenvalues $\{\lambda_i\}$ and the corresponding orthogonal set of unit eigenvectors $\{\hat{\mathbf{u}}_i\}$ as follows:

$$M^T M = \lambda_1 \hat{\mathbf{u}}_1 \hat{\mathbf{u}}_1^T + \lambda_2 \hat{\mathbf{u}}_2 \hat{\mathbf{u}}_2^T + \lambda_3 \hat{\mathbf{u}}_3 \hat{\mathbf{u}}_3^T.$$

(This can be seen by checking that the expression on the right hand side has eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\hat{\mathbf{u}}_i\}$).

Now $M^T M$ is positive definite, so the eigenvalues will be positive. Consequently the square-roots of the eigenvalues will be real and we can construct the symmetric matrix

$$S = \sqrt{\lambda_1}\hat{\mathbf{u}}_1 \hat{\mathbf{u}}_1^T + \sqrt{\lambda_2}\hat{\mathbf{u}}_2 \hat{\mathbf{u}}_2^T + \sqrt{\lambda_3}\hat{\mathbf{u}}_3 \hat{\mathbf{u}}_3^T.$$

It is easy to show that

$$S^2 = \lambda_1 \hat{\mathbf{u}}_1 \hat{\mathbf{u}}_1^T + \lambda_2 \hat{\mathbf{u}}_2 \hat{\mathbf{u}}_2^T + \lambda_3 \hat{\mathbf{u}}_3 \hat{\mathbf{u}}_3^T = M^T M,$$

using the fact that the eigenvectors are orthogonal. Also, for any nonzero vector $\mathbf{x}$,

$$\mathbf{x}^T S \mathbf{x} = \lambda_1 (\hat{\mathbf{u}}_1 \cdot \mathbf{x})^2 + \lambda_2 (\hat{\mathbf{u}}_2 \cdot \mathbf{x})^2 + \lambda_3 (\hat{\mathbf{u}}_3 \cdot \mathbf{x})^2 > 0.$$

We see that S is positive definite, since $\lambda_1 > 0$, $\lambda_2 > 0$, and $\lambda_3 > 0$. This construction of $S = (M^T M)^{1/2}$ applies even when some of the eigenvalues are zero; the result then is positive semi-definite (rather than positive definite).

3.D The Orthonormal Matrix in the Decomposition.

If all the eigenvalues are positive, then

$$S^{-1} = (M^T M)^{-1/2} = \frac{1}{\sqrt{\lambda_1}}\hat{\mathbf{u}}_1\hat{\mathbf{u}}_1^T + \frac{1}{\sqrt{\lambda_2}}\hat{\mathbf{u}}_2\hat{\mathbf{u}}_2^T + \frac{1}{\sqrt{\lambda_3}}\hat{\mathbf{u}}_3\hat{\mathbf{u}}_3^T,$$

as can be verified by multiplying by S. This expansion can be used to calculate the orthonormal matrix

$$U = MS^{-1} = M(M^T M)^{-1/2}.$$

The sign of $\det(U)$ is the same as the sign of $\det(M)$, because

$$\det(U) = \det(MS^{-1}) = \det(M)\det(S^{-1})$$

and $\det(S^{-1})$ is positive as all of its eigenvalues are positive. So U represents a rotation when $\det(M) > 0$ and a reflection when $\det(M) < 0$. (We expect to always obtain a rotation in our case. Only if the data is severely corrupted may a reflection provide a better fit).

When M only has rank two, the above method for constructing the orthonormal matrix breaks down. Instead we use

$$U = M\left(\frac{1}{\lambda_1}\hat{\mathbf{u}}_1\hat{\mathbf{u}}_1^T + \frac{1}{\lambda_2}\hat{\mathbf{u}}_2\hat{\mathbf{u}}_2^T\right) \pm \hat{\mathbf{u}}_3\hat{\mathbf{v}}_3^T,$$

or

$$U = MS^+ \pm \hat{\mathbf{u}}_3\hat{\mathbf{v}}_3^T,$$

where S^+ is the pseudo-inverse of S, that is,

$$S^+ = \frac{1}{\sqrt{\lambda_1}}\hat{\mathbf{u}}_1\hat{\mathbf{u}}_1^T + \frac{1}{\sqrt{\lambda_2}}\hat{\mathbf{u}}_2\hat{\mathbf{u}}_2^T,$$

and uu_3 and uv_3 are the third left and right singular vectors respectively (i.e. the third columns of U_0 and V_0, where U_0 and V_0 are the left and right factors in the singular value decomposition $MS^+ = U_0\Sigma_0V_0^T$ — this fix is due to Carlo Tomasi). The sign of the last term in the expression for U above is chosen to make the determinant of U positive. It is easy to show that the matrix constructed in this fashion is orthonormal and provides the desired decomposition $M = US$.

3.E Maximizing the Trace.

We have to maximize

$$\text{Trace}(R^T M) = \text{Trace}(R^T US),$$

where $M = US$ is the decomposition of M discussed above. From the expression for S in section 3.C, we see that

$$\begin{aligned}&\text{Trace}(R^T US)\\ &\quad = \sqrt{\lambda_1}\,\text{Trace}(R^T U\hat{\mathbf{u}}_1\hat{\mathbf{u}}_1^T) + \sqrt{\lambda_2}\,\text{Trace}(R^T U\hat{\mathbf{u}}_2\hat{\mathbf{u}}_2^T) + \sqrt{\lambda_3}\,\text{Trace}(R^T U\hat{\mathbf{u}}_3\hat{\mathbf{u}}_3^T).\end{aligned}$$

For any matrices X and Y, such that XY and YX are square, $\text{Trace}(XY) = \text{Trace}(YX)$. Therefore

$$\text{Trace}(R^T U\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^T) = \text{Trace}(\hat{\mathbf{u}}_i^T R^T U\hat{\mathbf{u}}_i) = \text{Trace}(R\hat{\mathbf{u}}_i \cdot U\hat{\mathbf{u}}_i) = (R\hat{\mathbf{u}}_i \cdot U\hat{\mathbf{u}}_i).$$

Since $\hat{\mathbf{u}}_i$ is a unit vector, and both U and R are orthonormal transformations, we have $(R\hat{\mathbf{u}}_i \cdot U\hat{\mathbf{u}}_i) \le 1$, with equality if and only if $R\hat{\mathbf{u}}_i = U\hat{\mathbf{u}}_i$. It follows that

$$\text{Trace}(R^T US) \le \sqrt{\lambda_1} + \sqrt{\lambda_2} + \sqrt{\lambda_3} = \text{Trace}(S).$$

and the maximum value of $\text{Trace}(R^T US)$ is attained when $R^T U = I$, or $R = U$. Thus the sought after orthonormal matrix is the one that occurs in the decomposition of M into the product of an orthonormal matrix and a symmetric one. If M is not singular, then

$$R = M(M^T M)^{-1/2}.$$

If M only has rank two, however, we have to resort to the second method discussed in the previous section to find R.

3.F Nearest Orthonormal Matrix.

Owe can now show that the nearest orthonormal matrix R to a given non-singular matrix M is the matrix U that occurs in the decomposition of M into the product of an orthonormal matrix and a positive-definite matrix. That is,

$$U = M(M^T M)^{-1/2}$$

We wish to find the matrix R that minimizes

$$\sum_{i=1}^{3}\sum_{j=1}^{3}(m_{i,j} - r_{i,j})^2 = \text{Trace}\big((M-R)^T(M-R)\big),$$

subject to the condition $R^T R = I$. That is, minimize

$$\mathrm{Trace}(M^T M) - 2\,\mathrm{Trace}(R^T M) + \mathrm{Trace}(R^T R).$$

Now $R^T R = I$, so we conclude that the first and third terms do not depend on R. The problem then is to maximize

$$\mathrm{Trace}(R^T M)$$

We conclude immediately, using the result of the previous section, that the nearest orthonormal matrix to the matrix M is the orthonormal matrix that occurs in the decomposition of M into the product of an orthonormal and a symmetric matrix.

Thus the orthonormal matrix that maximizes the residual error in our original least-squares problem is the orthonormal matrix nearest to the matrix

$$M = \sum_{i=1}^{n} \mathbf{r}'_{r,i}\,(\mathbf{r}'_{l,i})^T.$$

We note here that this orthonormal matrix can be found once an eigenvalue-eigenvector decomposition of the symmetric 3×3 matrix $M^T M$ has been obtained.

3.G Rank of the Matrix M.

It is clear that the rank of $M^T M$ is the same as the rank of M, since the two matrices have exactly the same eigenvectors with zero eigenvalue. The first method for finding the desired orthonormal matrix applies only when M, and hence $M^T M$, is nonsingular.

If, on the other hand,

$$M\mathbf{n}_l = \mathbf{0},$$

for any nonzero vector $\mathbf{n}_l$, then the matrix M, and hence $M^T M$, is singular. This happens when all of the left measurements lie in the same plane. That is, when

$$\mathbf{r}'_{l,i} \cdot \mathbf{n}_l = 0,$$

for $i = 1, 2, \ldots, n$, where $\mathbf{n}_l$ is a normal to the plane, since

$$M\mathbf{n}_l = \left(\sum_{i=1}^{n} \mathbf{r}'_{r,i}\,(\mathbf{r}'_{l,i})^T \right) \mathbf{n}_l = \sum_{i=1}^{n} \mathbf{r}'_{r,i}\,(\mathbf{r}'_{l,i} \cdot \mathbf{n}_l) = \mathbf{0}.$$

Similarly, if all of the right measurements lie in the same plane

$$\mathbf{r}'_{r,i} \cdot \mathbf{n}_r = 0,$$

where $\mathbf{n}_r$ is a normal to the plane and so $M^T\mathbf{n}_r = \mathbf{0}$. Now $\det(M^T) = \det(M)$, so this implies that M is singular also. As a result, we cannot use the simple expression,

$$U = M(M^TM)^{-1/2},$$

to find the orthonormal matrix when either of the two sets of measurements (left or right) are coplanar. This happens, for example, when there are only three points.

If one or both sets of measurements are coplanar, we have to use the second method for constructing U, given in section 3.D. This method requires that the matrix M have rank two (which will be the case unless the measurements happen to be collinear—in which case the absolute orientation problem does not have a unique solution). Note that the second method requires the solution of a quadratic equation to find the eigenvalues, whereas a cubic must be solved in the general case. One might, by the way, anticipate possible numerical problems when the matrix M is ill-conditioned, that is, when one of the eigenvalues is nearly zero. This will happen when one of the sets of measurements lies *almost* in a plane.

3.H Symmetry in the Transformation.

If, instead of finding the best transformation from the left to the right coordinate system, we decided instead to find the best transformation from the right to the left, we would have to maximize

$$\sum_{i=1}^{n} (\mathbf{r}'_{l,i})^T \overline{R}\, \mathbf{r}'_{r,i},$$

by choosing an orthonormal matrix $\overline{R}$. We can immediately write down the solution

$$\overline{R} = M^T(MM^T)^{-1/2},$$

since M becomes M^T when we interchange left and right. We would expect $\overline{R}^T$ to be equal to R, but, much to our surprise,

$$\overline{R}^T = (MM^T)^{-1/2}M.$$

This appears to be different from

$$R = M(M^TM)^{-1/2}.$$

but in fact they are equal. This is so because

$$\left(M^{-1}(MM^T)^{1/2}M\right)^2 = M^{-1}(MM^T)M = M^TM.$$

Taking inverses and square roots we obtain

$$M^{-1}(MM^T)^{-1/2}M = (M^TM)^{-1/2},$$

and, premultiplying by M, we find

$$\overline{R}^T = (MM^T)^{-1/2}M = M(M^TM)^{-1/2} = R.$$

3.I Finding the Eigenvalues and Eigenvectors.

We need to find the roots of the cubic in λ obtained by expanding

$$\det(M^TM - \lambda I) = 0,$$

where M^TM is

$$\begin{pmatrix} S_{xx}^2 + S_{yx}^2 + S_{zx}^2 & S_{xx}S_{xy} + S_{yx}S_{xy} + S_{zx}S_{zy} & S_{xx}S_{xz} + S_{yx}S_{yz} + S_{zx}S_{zz} \\ S_{xy}S_{xx} + S_{yy}S_{yx} + S_{zy}S_{zx} & S_{xy}^2 + S_{yy}^2 + S_{zy}^2 & S_{xy}S_{xz} + S_{yy}S_{yz} + S_{zy}S_{zz} \\ S_{xz}S_{xx} + S_{yz}S_{yx} + S_{zz}S_{zx} & S_{xz}S_{xy} + S_{yz}S_{yy} + S_{zz}S_{zy} & S_{xz}^2 + S_{yz}^2 + S_{zz}^2 \end{pmatrix}.$$

Having found the three solutions of the cubic, λ_i for $i = 1$, 2, and 3 (all real, and in fact positive), we then solve the homogeneous equations

$$(M^TM - \lambda_i I)\,\hat{\mathbf{u}}_i = \mathbf{0}$$

to find the three orthogonal eigenvectors $\hat{\mathbf{u}}_i$ for $i = 1$, 2, and 3.

3.J Coefficients of the Cubic.

Suppose that we write the matrix M^TM in the form

$$M^TM = \begin{pmatrix} a & d & f \\ d & b & e \\ f & e & c \end{pmatrix},$$

where $a = (S_{xx}^2 + S_{yy}^2 + S_{zz}^2)$ and so on. Then

$$\det(M^TM - \lambda I) = 0$$

can be expanded as

$$-\lambda^3 + d_2\lambda^2 + d_1\lambda + d_0 = 0,$$

where

$$d_2 = a + b + c,$$

$$d_1 = (e^2 - bc) + (f^2 - ac) + (d^2 - ab),$$

$$d_0 = abc + 2def - (ae^2 + bf^2 + cd^2).$$

We may note at this point that

$$d_2 = \mathrm{Trace}(M^T M)$$

so

$$d_2 = (S_{xx}^2 + S_{xy}^2 + S_{xz}^2) + (S_{yx}^2 + S_{yy}^2 + S_{yz}^2) + (S_{zx}^2 + S_{zy}^2 + S_{zz}^2),$$

while

$$d_0 = \det(M^T M) = \big(\det(M)\big)^2$$

or

$$\begin{aligned} d_0 = \big((&S_{xx}S_{yy}S_{zz} + S_{xy}S_{yz}S_{zx} + S_{yx}S_{xz}S_{zy}) \\ &- (S_{xx}S_{yz}S_{zy} + S_{yy}S_{zx}S_{xz} + S_{zz}S_{xy}S_{yx})\big)^2. \end{aligned}$$

4. Ignoring the Orthonormality.

Since it is so difficult to enforce the six nonlinear constraints that ensure that the matrix R is orthonormal, it is tempting to just find the best-fit linear transformation from left to right coordinate system. This is a straightforward least-squares problem. One can then try to find the "nearest" orthonormal matrix to the one obtained in this fashion. We show that this approach actually involves more work and does not produce the solution to the original least-squares problem. In fact, the result is asymmetric, in that the best-fit linear transform from left to right is not the inverse of the best-fit linear transform from right to left. Furthermore, at least *four* points need to be measured, whereas the method that enforces orthonormality requires only three. We discuss this approach next.

4.A Best Linear Transformation.

Here we have to find the matrix X that minimizes

$$\sum_{i=1}^{n} \left\|\mathbf{r}'_{r,i} - X\,\mathbf{r}'_{l,i}\right\|^2,$$

or

$$\sum_{i=1}^{n}\left(\left\|\mathbf{r}'_{r,i}\right\|^2 - 2\,\mathbf{r}'_{r,i}\cdot(X\,\mathbf{r}'_{l,i}) + \left\|X\mathbf{r}'_{l,i}\right\|^2\right).$$

Since X is not necessarily orthonormal, we cannot simply replace

$$\left\|X\,\mathbf{r}'_{l,i}\right\|^2 \quad \text{by} \quad \left\|\mathbf{r}'_{l,i}\right\|^2.$$

Note that $\|\mathbf{x}\|^2 = \mathbf{x}\cdot\mathbf{x}$ and that $\mathbf{x}\cdot\mathbf{y} = \mathrm{Trace}(\mathbf{x}\mathbf{y}^T)$. The sum above can be rewritten in the form

$$\sum_{i=1}^{n}\mathrm{Trace}\left(\mathbf{r}'_{r,i}(\mathbf{r}'_{r,i})^T - 2\,\mathbf{r}'_{r,i}(\mathbf{r}'_{l,i})^T X^T + X\,\mathbf{r}'_{l,i}(\mathbf{r}'_{l,i})^T X^T\right)$$
$$= \mathrm{Trace}(XA_lX^T - 2MX^T + A_r),$$

where

$$A_l = \sum_{i=1}^{n}\mathbf{r}'_{l,i}(\mathbf{r}'_{l,i})^T \quad \text{and} \quad A_r = \sum_{i=1}^{n}\mathbf{r}'_{r,i}(\mathbf{r}'_{r,i})^T$$

are symmetric 3×3 matrices obtained from the left and right sets of measurements respectively.

We can find the minimum essentially by completing the square. First of all, we use the fact that $\mathrm{Trace}(MX^T) = \mathrm{Trace}(X^TM)$ to rewrite the above in the form

$$\mathrm{Trace}\left(XA_lX^T - MX^T - XM^T + MA_l^{-1}M^T\right) + \mathrm{Trace}\left(A_r - MA_l^{-1}M^T\right).$$

The second term does not depend on X while the first can be written as the trace of a product:

$$\mathrm{Trace}\left((XA_l - M)(X - MA_l^{-1})^T\right).$$

Now it is easy to see that A_l is positive semi-definite. In fact, the matrix A_l is positive definite, provided that at least four measurments are available that are not collinear. This means that A_l has a positive-definite square-root and that this square-root has an inverse. As a result, we can then rewrite the above in the form

$$\mathrm{Trace}\left((XA_l^{1/2} - MA_l^{-1/2})(XA_l^{1/2} - MA_l^{-1/2})^T\right)$$
$$= \left\|XA_l^{1/2} - MA_l^{-1/2}\right\|^2.$$

This is zero when

$$XA_l^{1/2} = MA_l^{-1/2},$$

or

$$X = MA_l^{-1}.$$

4.B Asymmetry in Simple Linear Solution.

It is easy to find X by multiplying M by the inverse of A_l. Note, however, that we are using more information here than before. The method that does enforce orthonormality requires only the matrix M. Also note that A_l depends on the left measurements alone. This suggests an asymmetry. Indeed, if we minimize instead

$$\sum_{i=1}^{n} \left\|\mathbf{r}'_{l,i} - \overline{X}\,\mathbf{r}'_{r,i}\right\|^2,$$

we obtain

$$\overline{X} = M^T A_r^{-1}.$$

In general, $\overline{X}$ is not equal to X^{-1}, as one might expect.

Neither X nor $\overline{X}$ need by orthonormal. The nearest orthonormal matrix to X was shown in sections 3.F to be equal to

$$R = X(X^T X)^{-1/2} = (XX^T)^{-1/2}X,$$

while the one nearest to $\overline{X}$ is

$$\overline{R} = \overline{X}(\overline{X}^T\overline{X})^{-1/2} = (\overline{X}\,\overline{X}^T)^{-1/2}\overline{X}.$$

Typically $\overline{R}^T \neq R$.

4.C Relationship of Simple Linear Solution to Exact Solution.

We saw earlier that the solution of the original least-squares problem is the orthonormal matrix closest to M. The simple best-fit linear solution instead leads to the matrix MA_l^{-1}. The closest orthonormal matrix to MA_l^{-1} will in general not be equal to that closest to M. To see this, suppose that

$$M = US \quad \text{and} \quad MA_l^{-1} = U'S'$$

are the decompositions of M and MA_l^{-1} into orthonormal and positive-definite matrices. Then

$$US = U'(S'A_l).$$

For the solutions to be identical (that is $U = U'$) , we would need to have

$$S = S'A_l,$$

but the product of two symmetric matrices is, in general, not symmetric; so in general $U' \neq U$.

4.D Disadvantages of the Simple Linear Method.

The simple linear method does not lead to an orthonormal matrix. The closest orthonormal matrix can be found, but that is just as much work as that required for the exact solution of the original least-squares problem. In addition, the simple linear method requires that twice as much data be accumulated (A_l or A_r in addition to M). Furthermore, the linear transformation has more degrees of freedom (nine independent matrix elements) than an orthonormal matrix, so more constraint is required. Indeed, for A_l or A_r to be nonsingular, at least *four* points must be measured. This is a result of the fact that the vectors are taken relative to the centroid, and so three measurements do not provide three independent vectors. More seriously, this method does not produce the solution to the original least-squares problem.

5. Conclusion.

We presented here a closed-form solution of the least-squares problem of absolute orientation using orthonormal matrices to represent rotation. The method provides the best rigid-body transformation between two coordinate systems given measurements of the coordinates of a set of points that are not collinear. A closed-form solution using unit quaternions to represent rotation was given earlier[1]. In this paper we have derived an alternate method that uses manipulation of matrices and their eigenvalues-eigenvector decomposition. The description of this method may perhaps appear to be rather lengthy. This is the result of the need to deal with various special cases, such as that of coplanar sets of measurements.

We have shown that the best scale is the ratio of the root-mean-square deviations of the measurements from their respective centroids. The best translation is the difference between the centroid of one set of measurements and the scaled and rotated centroid of the other measurements. These exact results are to be preferred to ones based on measurements of one or two points only.

We contrast the exact solution of the absolute orientation problem to various approaches advocated in the past. The exact solution turns out to be easier to compute than one of these alternatives. The solution presented here may seem relatively complex. The ready availability of program packages for solving algebraic equations and finding eigenvalues and eigenvectors of symmetric matrices makes implementation straightforward, however. Methods for finding the eigenvectors efficiently were discusse in reference 1. It should also be noted that we are only dealing with 3×3 matrices.

Acknowledgments.

Jan T. Galkowski provided us with copies of a number of interesting papers on subjects relevant to our endeavour. He also put us in touch with C.J. Standish, who discovered a paper by Farrel & Stuelpnagel[9] that presents a closed-form solution to the least-squares problem of absolute orientation that applies when neither of the sets of measurements is coplanar. Thomas Poiker made several helpful suggestions. Takeo Kanade drew our attention to some references in the psychological literature that are relevant to the task of finding the nearest orthonormal matrix[15–21]. After we had arrived at our solution, Thomas Huang brought to our attention a solution of this problem that he, S. Arun, and S.D. Blostein had developed[10], using singular-value decomposition of an arbitrary matrix, instead of the eigenvalue-eigenvector decomposition of a symmetric matrix inherent in our approach. Elaine Shimabukuro uncomplainingly typed a number of drafts of this paper. We also wish to thank the reviewers for several helpful comments.

References.

1. B.K.P. Horn, "Closed-form Solution of Absolute Orientation using Unit Quaternions," *Journal of the Optical Society A*, **4**, 629–642 (April 1987).
2. C.C. Slama, C. Theurer & S.W. Henrikson, eds. *Manual of Photogrammetry* (American Society of Photogrammetry, Falls Church, Virginia 1980).
3. P.R. Wolf, *Elements of Photogrammetry* (McGraw Hill, New York 1974).
4. Sanjib K. Gosh, *Theory of Stereophotogrammetry* (Ohio State University Bookstores, Columbus, Ohio 1972).
5. B.K.P. Horn, *Robot Vision* (MIT Press, Cambridge Massachusetts & McGraw-Hill, New York 1986).
6. E.H. Thompson, "An Exact Linear Solution of the Problem of Absolute Orientation," *Photogrammetria*, **15**, 4, 163–179 (1958-1959).
7. G.H. Schut, "On Exact Linear Equations for the Computation of the Rotational Elements of Absolute Orientation," *Photogrammetria*, **16**, 1, 34–37 (1960).
8. H.L. Oswal & S. Balasubramanian, "An Exact Solution of Absolute Orientation," *Photogrammetric Engineering*, **34**, 1079–1083 (1968).
9. J. L. Farrell & J.C. Stuelpnagel, "Problem 65-1: A Least Squares Estimate of Satellite Attitude," *SIAM Review*, 384–386 (July 1966).
10. S. Arun, T.S. Huang & S.D. Blostein, "Least-Squares Fitting of two 3-D point Sets," Technical Note ISP-81, Coordinated Science Laboratory, University of Illinois (February 1986). Submitted to *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
11. J.H. Stuelpnagle, "On the Parameterization of the three-dimensional Rotation Group," *SIAM Review*, **6**, 4, 422–430 (October 1964).
12. Granino A. Korn & Theresa M. Korn, *Mathematical Handbook for Scientists and Engineers* (McGraw Hill, New York 1968).
13. G. Birkhoff & S. MacLane, *A Survey of Modern Algebra* (Macmillan, New York 1953).
14. P.H.Winston & B.K.P. Horn, *LISP* (Addison-Wesley, Reading, Massachusetts 1984).
15. Carl Eckart & Gale Young, "The Approximation of One Matrix by Another of Lower Rank," *Psychometrika*, **1**, 3, 211–218, (September 1936).

16. Bert F. Green, “The Orthogonal Approximation of an Oblique Structure in Factor Analysis,” *Psychometrika*, **1**, 4, 429–440 (December 1952).

17. W.A. Gibson, “On the Least-Squares Orthogonalization of an Oblique Transformation,” *Psychometrika*, **27**, 2, 193–195. (June 1962).

18. Richard M. Johnson, “On a Theorem Stated by Eckart and Young,” *Psychometrika*, **28**, 3, 259–263 (September 1963).

19. Peter H. Schönemann, “A Generalized Solution of the Orthogonal Procrustes Problem,” *Psychometrika*, **31**, 1, 1–10. (March 1966).

20. Norman Cliff, “Orthogonal Rotation to Congruence,” *Psychometrika*, **31**, 1, 33–42. (March 1966).

21. Peter H. Schönemann & Robert M. Carroll, “Fitting One Matrix to Another Under Choice of a Central Dilation and a Rigid Motion,” *Psychometrika*, **35**, (1970).

22. Paul R. Halmos, *Finite Dimensional Vector Spaces*, 2-nd edition, (Van Nostrand, Princeton, 1958).

Appendix: The Nearest Orthonormal Matrix.

In this Appendix we derive a method for finding the nearest orthonormal matrix R to a given matrix X*. We show that the desired matrix is the orthonormal matrix R that maximizes $\mathrm{Trace}(R^T X)$. This in turn is the orthonormal matrix U that appears in the decomposition $X = US$ of the matrix X into a product of an orthonormal matrix and a symmetric matrix. It is fairly straightforward to find U when X is nonsingular. It turns out, however, that there is a unique decomposition even when one of the eigenvalues of X is zero. We need this extension, because we want to be able to deal with the case when one of the sets of measurements is coplanar.

Lemma 1: If R is an orthonormal matrix, then for any vector $\mathbf{x}$,

$$(R\mathbf{x}) \cdot \mathbf{x} \leq \mathbf{x} \cdot \mathbf{x},$$

with equality holding only when $R\mathbf{x} = \mathbf{x}$.

Proof: First of all,

$$(R\mathbf{x}) \cdot (R\mathbf{x}) = (R\mathbf{x})^T (R\mathbf{x}) = \mathbf{x}^T R^T R\mathbf{x} = \mathbf{x}^T \mathbf{x} = \mathbf{x} \cdot \mathbf{x},$$

since $R^T R = I$. Now

$$\begin{aligned}(R\mathbf{x} - \mathbf{x}) \cdot (R\mathbf{x} - \mathbf{x}) &= (R\mathbf{x}) \cdot (R\mathbf{x}) - 2(R\mathbf{x}) \cdot \mathbf{x} + \mathbf{x} \cdot \mathbf{x} \\ &= 2\big(\mathbf{x} \cdot \mathbf{x} - (R\mathbf{x}) \cdot \mathbf{x}\big),\end{aligned}$$

but, since $(R\mathbf{x} - \mathbf{x}) \cdot (R\mathbf{x} - \mathbf{x}) \geq 0$ we must have

$$\mathbf{x} \cdot \mathbf{x} \geq (R\mathbf{x} \cdot \mathbf{x}).$$

Equality holds only when $(R\mathbf{x} - \mathbf{x}) \cdot (R\mathbf{x} - \mathbf{x}) = 0$, that is, when $R\mathbf{x} = \mathbf{x}$. ∎

Lemma 2: Any positive semi-definite $n \times n$ matrix S can be written in the form

$$S = \sum_{i=1}^{n} \mathbf{u}_i \mathbf{u}_i^T,$$

in terms of an orthogonal set of vectors $\{\mathbf{u}_i\}$.

Proof: Let the eigenvalues of S be $\{\lambda_i\}$ and a corresponding orthogonal set of unit eigenvectors $\{\hat{\mathbf{u}}_i\}$. Then we can write S in the form

$$\sum_{i=1}^{n} \lambda_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^T,$$

*In this appendix we deal with $n \times n$ matrices, although in the body of this paper we only need 3×3 matrices.

since the sum has exactly the same eigenvalues and eigenvectors as S. Now the eigenvalues of a positive semi-definite matrix are not negative, so have real square-roots. Thus we can write

$$S = \sum_{i=1}^{n} \mathbf{u}_i \mathbf{u}_i^T,$$

where

$$\mathbf{u}_i = \sqrt{\lambda_i}\,\hat{\mathbf{u}}_i.$$

■

Corollary 1: The identity matrix I can be written in terms of any orthogonal set of unit vectors as

$$I = \sum_{i=1}^{n} \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^T.$$

Lemma 3: If S is a positive semi-definite matrix, then for any orthonormal matrix R,

$$\text{Trace}(RS) \le \text{Trace}(S),$$

with equality holding only when $RS = S$.

Proof: Using Lemma 2, we can write S in the form

$$S = \sum_{i=1}^{n} \mathbf{u}_i \mathbf{u}_i^T.$$

Now $\text{Trace}(\mathbf{a}\mathbf{b}^T) = \mathbf{a} \cdot \mathbf{b}$, so

$$\text{Trace}(S) = \sum_{i=1}^{n} \mathbf{u}_i \cdot \mathbf{u}_i,$$

while

$$\text{Trace}(RS) = \sum_{i=1}^{n} (R\mathbf{u}_i) \cdot \mathbf{u}_i,$$

which by Lemma 1, is less than or equal to $\text{Trace}(S)$. Equality holds only when $R\mathbf{u}_i = \mathbf{u}_i$, for $i = 1, 2, \ldots, n$, that is, when $RS = S$, since $S\mathbf{u}_i = \mathbf{u}_i$. ■

Corollary 2: If S is a positive-definite matrix, then for any orthonormal matrix R

$$\text{Trace}(RS) \le \text{Trace}(S),$$

with equality holding only when $R = I$.

Lemma 4: If S is a real symmetric $n \times n$ matrix of rank $(n-1)$, R is an orthonormal rotation matrix and $RS = S$, then $R = I$.

Proof: Let the eigenvalues of S be $\{\lambda_i\}$ and let $\{\hat{\mathbf{u}}_i\}$ be a corresponding orthogonal set of unit eigenvectors. Without loss of generality, let us assume that $\lambda_n = 0$. Now $\hat{\mathbf{u}}_n$ is orthogonal to the other $(n-1)$ eigenvectors, or

$$\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_n = 0 \quad \text{for} \quad i \neq n.$$

There is a unique (up to sign) unit vector orthogonal to $(n-1)$ vectors, so $\hat{\mathbf{u}}_n$ can be found (up to sign) from the other eigenvectors. Now

$$S\hat{\mathbf{u}}_i = \lambda_i \hat{\mathbf{u}}_i,$$

from the definition of eigenvalues and eigenvectors. As a result

$$RS\hat{\mathbf{u}}_i = \lambda_i R\,\hat{\mathbf{u}}_i.$$

From $RS = S$ we can then conclude that

$$R\,\hat{\mathbf{u}}_i = \hat{\mathbf{u}}_i \quad \text{for} \quad i \neq n,$$

or

$$\hat{\mathbf{u}}_i = R^T \hat{\mathbf{u}}_i \quad \text{for} \quad i \neq n,$$

since $R^T R = I$. Now $\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_n = 0$ for $i \neq n$, so

$$(R^T \hat{\mathbf{u}}_i) \cdot \hat{\mathbf{u}}_n = 0 \quad \text{for} \quad i \neq n,$$

or, since $R^T R = I$,

$$(R\,\hat{\mathbf{u}}_n) \cdot \hat{\mathbf{u}}_i = 0 \quad \text{for} \quad i \neq n.$$

As a consequence $R\,\hat{\mathbf{u}}_n$ is perpendicular to the $(n-1)$ vectors $\{\hat{\mathbf{u}}_i\}$ for $i \neq n$, so $R\,\hat{\mathbf{u}}_n$ has to be equal to $+\hat{\mathbf{u}}_n$ or $-\hat{\mathbf{u}}_n$.

Now for any matrix U,

$$\det(R\,U) = \det(R)\det(U) = \det(U),$$

since R is a rotation matrix (that is, $\det(R) = +1$). Now let

$$U = [\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \ldots, \hat{\mathbf{u}}_n]$$

be the matrix obtained by adjoining the eigenvectors of S. The first $(n-1)$ columns of RU are just the same as the columns of U, so the last has to be $+\hat{\mathbf{u}}_n$ (not $-\hat{\mathbf{u}}_n$). That is,

$$R\,\hat{\mathbf{u}}_n = \hat{\mathbf{u}}_n.$$

Now construct the matrix

$$S + \hat{\mathbf{u}}_n \hat{\mathbf{u}}_n^T.$$

It is nonsingular, since the first $(n-1)$ eigenvalues are the same as those of S, and since $\lambda_n = 1$. Now

$$R(S + \hat{\mathbf{u}}_n\hat{\mathbf{u}}_n^T) = RS + R\,\hat{\mathbf{u}}_n\hat{\mathbf{u}}_n^T = (S + \hat{\mathbf{u}}_n\hat{\mathbf{u}}_n^T),$$

and so $R = I$. ■

Corollary 3: If S is a positive semi-definite $n \times n$ matrix, of rank at least $(n-1)$, and R an orthonormal rotation matrix, then

$$\mathrm{Trace}(RS) \le \mathrm{Trace}(S),$$

with equality holding only when $R = I$.

Lemma 5: The matrix

$$T = \sum_{i=1}^{n} \sqrt{\lambda_i}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^T$$

is the positive semi-definite square-root of the positive semi-definite matrix

$$S = \sum_{i=1}^{n} \lambda_i\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^T$$

where $\{\lambda_i\}$ are the eigenvalues and $\{\hat{\mathbf{u}}_i\}$ a corresponding set of orthogonal unit eigenvectors of S.

Proof: We have

$$T^2 = \left(\sum_{i=1}^{n} \sqrt{\lambda_i}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^T\right)\left(\sum_{j=1}^{n} \sqrt{\lambda_j}\hat{\mathbf{u}}_j\hat{\mathbf{u}}_j^T\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \sqrt{\lambda_i\lambda_j}\,(\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j)\,\hat{\mathbf{u}}_i\hat{\mathbf{u}}_j^T,$$

so

$$T^2 = \sum_{k=1}^{n} \lambda_k\,\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^T = S,$$

since $\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j = 0$ when $i \neq j$. Furthermore,

$$\mathbf{x}^T T\mathbf{x} = \sum_{i=1}^{n} \lambda_i(\hat{\mathbf{u}}_i \cdot \mathbf{x})^2 \ge 0,$$

since $\lambda_i \ge 0$ for $i = 1, 2, \ldots, n$ and so T is positive semi-definite. ■

Note: There are 2^n square-roots of S, because one can choose the n signs of the square-roots of the λ_i's independently. But only one of these square-roots of S is positive semi-definite.

Note: It is possible to show that the positive semi-definite square-root is unique, even when there are repeated eigenvalues.

Corollary 4: If S is positive definite, there exists a positive-definite square-root of S.

Corollary 5: The matrix

$$T^{-1} = \sum_{i=1}^{n} \frac{1}{\sqrt{\lambda_i}} \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^T$$

is the inverse of the positive-definite square-root of the positive-definite matrix

$$S = \sum_{i=1}^{n} \lambda_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^T.$$

Corollary 6: If S is a positive semi-definite matrix, then the pseudo-inverse of the positive semi-definite square-root of S is

$$T^{+} = \sum_{i \in P} \frac{1}{\sqrt{\lambda_i}} \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^T,$$

where P is the set of the integers for which $\lambda_i > 0$.

Note: The pseudo-inverse of a matrix T can be defined by the limit

$$T^{+} = \lim_{\delta \to 0} (T + \delta I)^{-1}.$$

Lemma 6: For any matrix X,

$$X^T X \quad \text{and} \quad X X^T$$

are positive semi-definite matrices.

Proof: First of all

$$(X^T X)^T = X^T (X^T)^T = X^T X,$$

so $X^T X$ is symmetric. For any vector $\mathbf{x}$,

$$\mathbf{x}^T (X^T X) \mathbf{x} = (\mathbf{x}^T X^T)(X\mathbf{x}) = (X\mathbf{x})^T (X\mathbf{x}) = (X\mathbf{x}) \cdot (X\mathbf{x}) \geq 0.$$

We conclude that $X^T X$ is positive semi-definite. Similar arguments apply to $X X^T$. ∎

Corollary 7: For any nonsingular square matrix X,

$$X^T X \quad \text{and} \quad X X^T$$

are positive-definite matrices.

Lemma 7: The eigenvectors of $X^T X$ with the zero eigenvalue are the same as the eigenvectors of X with zero eigenvalue. The eigenvectors of XX^T with zero eigenvalue are the same as the eigenvectors of X^T with zero eigenvalue.

Proof: Let $\hat{\mathbf{u}}_n$ be an eigenvector of X with zero eigenvalue. That is,

$$X\,\hat{\mathbf{u}}_n = \mathbf{0}.$$

Then certainly

$$X^T X\,\hat{\mathbf{u}}_n = \mathbf{0}.$$

Conversely, if $X^T X\,\hat{\mathbf{u}}_n = \mathbf{0}$ then

$$\hat{\mathbf{u}}_n^T X^T X\,\hat{\mathbf{u}}_n = 0$$

or

$$(X\hat{\mathbf{u}}_n)^T(X\hat{\mathbf{u}}_n) = (X\hat{\mathbf{u}}_n)\cdot(X\hat{\mathbf{u}}_n) = \|X\hat{\mathbf{u}}_n\|^2 = 0,$$

which implies that $X\,\hat{\mathbf{u}}_n = \mathbf{0}$. Similar arguments hold when X is replaced by X^T and $X^T X$ by XX^T. ∎

Lemma 8: Every nonsingular matrix X can by written in the form

$$X = US$$

where

$$U = X(X^T X)^{-1/2},$$

is an orthonormal matrix, while

$$S = (X^T X)^{1/2}$$

is a positive-definite matrix.

Proof: Since X is not singular, Corollary 7 tells us that $X^T X$ is positive-definite. We use Corollary 4 to give us the positive-definite square-root $(X^T X)^{1/2}$. The inverse can be constructed using Corollary 5. As a result S and U can be found, given X. Their product is clearly just

$$US = X(X^T X)^{-1/2}(X^T X)^{1/2} = X.$$

We still need to check that U is orthonormal. Now

$$U^T = (X^T X)^{-1/2} X^T,$$

so

$$U^T U = (X^T X)^{-1/2}(X^T X)(X^T X)^{-1/2}$$
$$= (X^T X)^{-1/2}(X^T X)^{1/2}(X^T X)^{1/2}(X^T X)^{-1/2} = I.$$

■

Note: It can be shown that the decomposition is unique.

Corollary 8: Every nonsingular matrix X can be written in the form

$$X = \overline{S}\,\overline{U},$$

where

$$\overline{U} = (XX^T)^{-1/2}X$$

is an orthonormal matrix, while

$$\overline{S} = (XX^T)^{1/2}$$

is a positive-definite matrix.

Note: The orthonormal matrix U appearing in the decomposition $X = US$ is the same as the orthonormal matrix $\overline{S}$ appearing in the decomposition $X = \overline{S}\,\overline{U}$.

Corollary 9: The orthonormal matrix U in the decomposition of a nonsingular matrix X into the product of an orthonormal matrix and a positive-definite matrix is a rotation or a reflection according to whether $\det(X) > 0$ or $\det(X) < 0$.

Lemma 9: If X is nonsingular then

$$X(X^T X)^{-1/2} = (XX^T)^{-1/2}X.$$

Proof: Let

$$U = X(X^T X)^{-1/2}.$$

Then

$$U(X^T X)^{1/2} = X,$$

and, since we showed in Lemma 8 that U is orthonormal,

$$(X^T X)^{1/2} = U^T X.$$

Squaring, we see that

$$X^T X = (U^T X)(U^T X),$$

and so

$$X(X^T X) = X(U^T X)(U^T X).$$

Since X is nonsingular, it has an inverse. Post-multiplying by this inverse, we obtain

$$XX^T = (XU^T)(XU^T).$$

We can find the positive-definite square-root of XX^T and write

$$(XX^T)^{1/2} = XU^T,$$

or

$$(XX^T)^{1/2}U = X,$$

so

$$U = (XX^T)^{-1/2}X.$$

Note: When X is singular, it defines an invertible transformation on the subspace that is the range of X. We can find a decomposition of this transformation into the product of an orthonormal and a symmetric part. The orthonormal part can be extended to the whole space by taking the direct sum of the orthonormal transformation in the subspace and any orthonormal transformation in its orthogonal complement.

In general, this extension will not be unique, so there will not be a unique decomposition of X into a product of an orthonormal part and a symmetric part. When the rank of X is $(n-1)$, however, the orthogonal complement has dimension one and so only two "orthogonal transformations" are possible in the orthogonal complement. As a result there are only two ways of decomposing X into the desired way. One gives us a rotation, the other a reflection.

Lemma 10: If the $n \times n$ matrix X has rank $(n-1)$ we can write it in the form

$$X = US$$

where

$$U = X\left(\sum_{i=1}^{n-1} \frac{1}{\sqrt{\lambda_i}} \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^T\right) \pm \hat{\mathbf{u}}_n \hat{\mathbf{u}}_n^T$$

is an orthonormal matrix and

$$S = (X^T X)^{1/2}$$

is a positive semi-definite matrix. Here we have arranged the eigenvalues $\{\lambda_i\}$ and the corresponding unit eigenvectors $\{\hat{\mathbf{u}}_i\}$ in such a way that $\hat{\mathbf{u}}_n$ is the eigenvector corresponding to the eigenvalue $\lambda_n = 0$.

Proof: From Lemma 5 we have

$$S = \sum_{i=1}^{n} \sqrt{\lambda_i}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^T = \sum_{i=1}^{n-1} \sqrt{\lambda_i}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^T,$$

since $\lambda_n = 0$, so

$$US = X\left(\sum_{i=1}^{n-1} \frac{1}{\sqrt{\lambda_i}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^T\right)\left(\sum_{j=1}^{n-1} \sqrt{\lambda_j}\hat{\mathbf{u}}_j\hat{\mathbf{u}}_j^T\right) \pm (\hat{\mathbf{u}}_n\hat{\mathbf{u}}_n^T)\left(\sum_{i=1}^{n-1} \sqrt{\lambda_i}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^T\right).$$

Now $\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j = 0$ for $i \neq j$, so

$$US = X\left(\sum_{i=1}^{n-1} \hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^T\right).$$

Now from Lemma 7 we conclude that $X\,\hat{\mathbf{u}}_n = \mathbf{0}$, so

$$US = X\left(\sum_{i=1}^{n-1} \hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^T + \hat{\mathbf{u}}_n\hat{\mathbf{u}}_n^T\right) = X\sum_{i=1}^{n} \hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^T = XI = X,$$

using Corollary 1.

We still need to show that U is orthonormal. Now

$$U^T = \left(\sum_{i=1}^{n-1} \frac{1}{\sqrt{\lambda_i}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^T\right)X^T \pm \hat{\mathbf{u}}_n\hat{\mathbf{u}}_n^T,$$

so

$$U^TU = \left(\sum_{i=1}^{n-1} \frac{1}{\sqrt{\lambda_i}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^T\right)X^TX\left(\sum_{j=1}^{n-1} \frac{1}{\sqrt{\lambda_j}}\hat{\mathbf{u}}_j\hat{\mathbf{u}}_j^T\right) + (\hat{\mathbf{u}}_n\hat{\mathbf{u}}_n^T)(\hat{\mathbf{u}}_n\hat{\mathbf{u}}_n^T),$$

where some terms dropped out, because $\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j = 0$ for $i \neq j$. Now

$$X^TX = \sum_{k=1}^{n} \lambda_k\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^T,$$

so

$$U^TU = \sum_{i=1}^{n-1} \hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^T + \hat{\mathbf{u}}_n\hat{\mathbf{u}}_n^T = \sum_{i=1}^{n} \hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^T = I$$

using Corollary 1. ∎

Note: We had to add $\pm\hat{\mathbf{u}}_n\hat{\mathbf{u}}_n^T$ to the product of X and the pseudo-inverse of the positive semi-definite square-root of X^TX to create an orthonormal matrix. Without this term the result would have been singular, since X is singular.

Note: We can choose to make U a rotation or a reflection by choosing the sign of the term $\hat{\mathbf{u}}_n\hat{\mathbf{u}}_n^T$. Thus, in this case there is a unique decomposition into the product of an orthonormal *rotation* matrix and a positive semi-definite matrix.

Note: To obtain a rotation matrix, we chose the sign to be the same as the sign of

$$\prod_{i=1}^{n-1} \overline{\lambda_i},$$

where the $\overline{\lambda_i}$ are the eigenvalues of the matrix X.

Theorem 1: If X is a nonsingular matrix, then the orthonormal matrix R that maximizes

$$\mathrm{Trace}(R^T X)$$

is the matrix U that occurs in the decomposition of X into the product of an orthonormal matrix and a positive-definite matrix. That is

$$U = X(X^T X)^{-1/2} = (XX^T)^{-1/2} X.$$

Proof: Let $X = US$ be the decomposition of X given in Lemma 8. We wish to maximize

$$\mathrm{Trace}(R^T X) = \mathrm{Trace}\big(R^T(US)\big) = \mathrm{Trace}\big((R^T U)S\big).$$

Now using Lemma 3, we see that

$$\mathrm{Trace}\big((R^T U)S\big) \leq \mathrm{Trace}(S),$$

with equality holding only when $R^T U = I$ by Corollary 2. That is, $R = U$. ■

Corollary 10: If the $n \times n$ matrix X has rank $(n-1)$, then the orthonormal matrix that maximizes

$$\mathrm{Trace}(R^T X)$$

is the matrix U in the decomposition of Lemma 10. That is,

$$U = X\left(\sum_{i=1}^{n-1} \frac{1}{\sqrt{\lambda_i}} \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^T\right) \pm \hat{\mathbf{u}}_n \hat{\mathbf{u}}_n^T$$

Theorem 2: The nearest orthonormal matrix R to a given nonsingular matrix X is the matrix U that occurs in the decomposition of X into the product of an orthonormal matrix and a positive-definite matrix. That is,

$$U = X(X^T X)^{-1/2} = (XX^T)^{-1/2} X.$$

Proof: We wish to find the matrix R that minimizes

$$\sum_{i=1}^{n} \sum_{j=1}^{n} (x_{i,j} - r_{i,j})^2 = \operatorname{Trace}\big((X - R)^T (X - R)\big),$$

subject to the condition $R^T R = I$. That is, minimize

$$\operatorname{Trace}(X^T X) - 2 \operatorname{Trace}(R^T X) + \operatorname{Trace}(R^T R).$$

Now $R^T R = I$, so we conclude that the first and third terms do not depend on R. The problem then is to maximize

$$\operatorname{Trace}(R^T X)$$

The result follows from Theorem 1. ■

Corollary 11: If the $n \times n$ matrix X has at least rank $(n - 1)$, then the nearest orthonormal matrix is the matrix U given in the decomposition $X = US$. Here U is given by Lemma 8 when X is nonsingular and by Lemma 10 if X is singular.

Figure Captions

Figure 1: The coordinates of a number of points is measured in two coordinate systems. The transformation between the two systems is to be found using these measurements.

Figure 2: Three points can be used to define a triad. Such a triad can be constructed using the left measurements. A second triad is then constructed from the right measurments. The required coordinate transformation can be estimated by finding the transformation that maps one triad into the other. This method does not use the information about each of the three points equally.