Closed-form and iterative weighted least squares solutions of Helmert transformation parameters

Research Article

L.E. Sjöberg

Division of Geodesy and Geoinformatics, Royal Institute of Technology (KTH), SE 10044, Stockholm, Sweden

Abstract:
The Helmert transformation is the most common transformation between different geodetic systems. In 2-D, in contrast to higher dimensions, it is a well-known procedure how to determine the 4 transformation parameters in a closed form. Here we derive the closed-form weighted least squares solution in $m$-dimensional space for an arbitrary number ($\geq m$) of coordinate set-ups in two related systems. The solution employs singular value decomposition (SVD) for the rotation matrix, while the translation vector and scale parameters are obtained in simpler ways. To avoid the SVD routine, we also present an iterative approach to solve for the rotation matrix. The paper is completed with a test procedure for detecting outlying coordinate pairs.

Keywords:
Helmert transformation • Procrustes problem • registration • singular value decomposition

© Versita sp. z o.o.

Received 11-09-2012; accepted 12-12-2012

1. Introduction

The Helmert transformation (Helmert 1924, p. 548) is a common transformation method in geodesy, surveying and photogrammetry between two coordinate systems, characterized in 2-D space by two translations, one rotation and a scale factor and in 3-D by three translations, three rotations and a scale factor. Importantly, the rotation matrix $R$ is orthonormal, implying that its inverse equals its transpose and its determinant is unity. In recent years the method has become important also in psychometrics, computer vision, image and laser scanning processing, robotics and several other disciplines, as part of the concept frequently called registration.

Basically the problem is a non-linear least squares problem, but in 2-D it can be converted to and solved as a linear problem. This is not the case in 3-D and higher dimensions, and geodesists usually solve the problem by linearization and iteration. However, this usually leads to that the property of the rotation matrix to be orthonormal is lost. Also, such a solution requires that an approximate solution is known, which may not always be the case. Mathematically, the problem and analysis to fit one matrix to another by a rotation matrix is called Procrustes problem and analysis, respectively, and there are numerous articles on this problem in the literature. See e.g., Schönemann (1966) for an original solution and Lissitz et al. (1976) for a generalized, weighted least squares solution.

A solution to the present problem was first provided by Schönemann and Carroll (1970), Horn (1987) used it to absolute orientation in photogrammetry, and Crosilla (2003), Grafarend and Awange (2003) (limited to 3-D) and Awange et al. (2010) studied it for geodetic applications. However, as demonstrated and corrected by Umeyama (1991), the rotation matrices for all these solutions may come out incorrect with reflection rather than rotation. The problem at hand can be described mathematically as follows. The Helmert transformation between the pairs of Cartesian coordinate vectors $x$ and $y$ in $m$-dimensional space is given by the matrix equation

$$
y_i = t + kRx; \quad i = 1, 2, \ldots, n \geq m,
$$

where $t$, $k$, $n$ and $R$ are the $m$-dimensional translation vector, the scale factor, the number of points and the $m \times m$ rotation matrix,
respectively. For $m \geq 3$ the problem is always over-determined, i.e. there are more observations than there are unknowns. The rotation matrix has the properties of being orthonormal and positive definite, i.e.

$$R^T R = I = \text{unit matrix and } \det(R) = 1. \quad (2)$$

The direct application of the transformation with known transformation parameters is straightforward, while the direct solution of the inverse problem, i.e. of determining the transformation parameters from a sufficient number of points with coordinates in both systems, is not evident. In general one should expect that the system of equations given by Eq. (1) is inconsistent, and each such relation is not necessarily of the same quality, which calls for some kind of weighted least squares solution to the problem. Here we will follow Umeyama’s (1991) approach, but the method is generalized to using a full weight matrix for the solution, and some error analyses and test procedure for outlier detection complete the paper. A key to solving the problem is that the solution for the rotation matrix does not depend on the scale factor and translation vector.

The general solution that follows in Sect. 2 relies on singular value decomposition (SVD) of a regular $m$-dimensional matrix, say $A$, by (e.g., Bjerhammar 1973, p. 345) $A = U \Sigma V^T$, where $U^T U = V^T V = I$ (orthonormal matrices) and $D = \text{diag} \left( \sigma_1, \sigma_2, \ldots, \sigma_m \right)$. Here $\sigma_i$ are the positive square-roots of eigen-values (called singular values) of $A^T A$. (SVD is a well-established mathematical procedure, common in most mathematical matrix software packages.)

2. The least squares problem and solution

We start this section by presenting a useful lemma for a rotation matrix, i.e. a matrix that satisfies the two conditions of Eq. (2).

**Lemma 1 (Procrustes problem):** Let $A$ and $B$ be fixed matrices, whose product $BA^T$ can be decomposed by SVD to $UDV^T$, where $U^T U = V^T V = I$, $D$ is a diagonal matrix with positive singular values and $R$ is a rotation matrix satisfying the conditions of Eq. (2). Then the minimum of the objective function

$$g = \text{tr} \left\{ (RA - B)(RA - B)^T \right\} \quad (3a)$$

is given by

$$R = U \Sigma V^T, \quad \Sigma = \text{diag} \{ 1, 1, \ldots, 1, \text{det}(U) \text{det}(V) \} = \text{diag} \{ 1, 1, \ldots, 1, \pm 1 \}. \quad (3b)$$

**Proof.** Considering the constraints of Eq. (2), the optimization problem can be augmented to finding the minimum of the function $g_1 = g + \text{tr} \left\{ G \left( R^T R - I \right) \right\} + 2\lambda \left[ \text{det}(R) - 1 \right]$, where $G$ is a symmetric matrix that together with the scalar $\lambda$ constitute Lagrange’s multipliers.

Taking the derivative of this function w.r.t. $R$ and equating to zero, one obtains the following equation for the minimum:

$$HR^T = AB^T = VDU^T, \quad (4a)$$

where

$$H = AA^T + G + \lambda I. \quad (4b)$$

As $G$, and therefore also $H$, is symmetric, and $R$ is a rotation matrix, Eq. (4a) yields

$$HR^T R = HH = VDV^T, \quad (5)$$

and it follows that

$$H = VMV^T; \quad \text{where } M^2 = D^2, \quad (6)$$

and therefore, from Eq. (4a),

$$VMV^T R^T = VDU^T \Rightarrow R^T = VM^{-1}DU^T, \quad (7)$$

which in view of the last part of Eq. (6) implies that

$$R^T = V \Sigma U^T, \quad \square$$

where $\Sigma$ was given in the lemma.

**Note 1:** The sign in the last element of $\Sigma$ is chosen such that the determinant of $R$ becomes +1.

**Note 2:** The solution is restricted to the case with a regular matrix $BA^T$. The singular case was treated by Myronenko and Song (2009).

We now consider the matrix form of the equation system of Eq. (1):

$$Y - \varepsilon = T + kRX \quad (8)$$

where $Y, \varepsilon, T$ and $X$ are all $(m \times n)$-dimensional matrices with $Y = (y_1, y_2, y_3, \ldots, y_n)$, $X = (x_1, x_2, x_3, \ldots, x_n)$, $\varepsilon$ is a residual matrix and

$$T = te^T; \quad e^T = (1, 1, 1, \ldots, 1). \quad (9)$$

Then the least squares problem can be formulated as follows.

**Problem:** Determine the transformation parameters of Eq. (8) that minimize the objective function

$$f = \text{tr} \{ F \}, \quad (10a)$$

where

$$F = e^T e + G \left( R^T R - I \right) + 2\lambda \text{det}(R). \quad (10b)$$
Here $P$ is a positive definite weight matrix, $G$ and $\lambda$ are a symmetric matrix and a scalar, respectively, of Lagrange's multipliers. Again, the objective function is augmented by the terms with Lagrange's multipliers to warrant that the solution for $R$ becomes a rotation matrix.

The solution is given in the theorem below. However, some preparatory steps, Lemmas 2 - 4, will help in formulating the proof of the theorem.

Note 3. The weight matrix has the same dimension as the number of observation points, i.e. the weighting is steered per point and not per individual coordinate.

**Lemma 2.** Given the rotation matrix and scale factor of Eq. (8), the translation matrix of the above problem is obtained by

$$T = \bar{Y} - kR\bar{X},$$

(11a)

with $\bar{Y}$ and $\bar{X}$ being the row mean values of $Y$ and $X$, i.e.

$$\bar{Y} = \frac{\text{Yee}}{n} \quad \text{and} \quad \bar{X} = Xee^T/n.$$  

(11b)

**Proof.** Let us first rewrite Eq. (8) as

$$\epsilon = \tilde{\epsilon} - \bar{T},$$

(12a)

where

$$\tilde{\epsilon} = \bar{Y} - kR\bar{X},$$

(12b)

$$\bar{Y} = Y - \bar{Y}, \quad \bar{X} = X - \bar{X}, \quad \bar{T} = T - \bar{Y} + kR\bar{X}. $$

(12c)

Then it follows from Eq. (12a) that

$$\epsilon P \epsilon^T = \tilde{\epsilon} P \tilde{\epsilon}^T + \bar{T}P\bar{T}^T,$$

(13a)

because

$$\bar{T}P\bar{T}^T = 0.$$  

(13b)

As $tr\left\{\bar{T}P\bar{T}^T\right\} \geq 0$, it follows that $\bar{T}$ must vanish when minimizing $\ell$, which proves the lemma.

**Lemma 3.** The least squares solution to the rotation matrix is independent of the translation vector and the scale factor, and it maximizes $tr\left\{R\bar{X}\bar{P}\bar{Y}^T\right\}$.

**Proof.** The independence on the translation vector was proved in Lemma 2. The remaining part of the proof follows from the objective function

$$l_1 = tr\left\{\tilde{\epsilon} P \tilde{\epsilon}^T\right\}/n = s_y^2 + k^2s_x^2 - 2ks_{xy},$$

(14a)

where

$$s_y^2 = tr\left\{\bar{Y}P\bar{Y}^T\right\}/n$$  

(14b)

$$s_x^2 = tr\left\{\bar{X}P\bar{X}^T R^T R\right\}/n = tr\left\{\bar{X}P\bar{X}^T\right\}/n$$  

(14c)

and

$$s_{xy} = tr\left\{R\bar{X}\bar{P}\bar{Y}^T\right\}/n.$$  

(14d)

It follows that the least squares solution for $R$ maximizes $s_{xy}$, which is independent of $T$ and $k$. □

**Lemma 4.** Given the rotation matrix, the least squares solution for the scale factor is

$$k = s_{xy}/s_x^2.$$

(15)

The proof follows directly by differentiating Eq. (14a) w.r.t. $k$ and equating to zero.

**Theorem 1.** The least squares transformation parameters of Eq. (8) that minimize the objective function of Eqs. (10a) and (10b) are $R = U\Sigma V^T$, where $U$, $\Sigma$ and $V$ are given by Lemma 1 with $BA^T = \bar{Y}_P\bar{X}^T = UD\bar{V}^T$, $T = \bar{Y} - kR\bar{X}$, and $k = s_{xy}/s_x^2$.

**Proof.** The proof follows directly from Lemmas 1 - 4.

**Corollary 1:** If $P = I$, $k = tr\left\{D\Sigma\right\}/(ns_x^2)$.

The proof follows directly by inserting $P = I$ into the solution for $k$ of the theorem and considering Eq. (14c).

□

2.1. Computational steps

**Given input data:**

$(x_i, y_i) ; i = 1, 2, ..., n$, $P$ and $\epsilon^T = (1, 1, 1, ..., 1)$

Step 1: $X = (x_1, x_2, x_3, ..., x_n)$ and $\epsilon^T = (y_1, y_2, y_3, ..., y_n)$

Step 2: $\bar{X} = Xee^T/n$ and $\bar{Y} = Yee^T/n$

Step 3: $\bar{X} = X - \bar{X}, \bar{Y} = Y - \bar{Y}$

Step 4: $s_y^2 = tr\left\{\bar{X}P\bar{X}^T\right\}/n$ and $s_x^2 = tr\left\{\bar{Y}P\bar{Y}^T\right\}/n$

Step 5: $C = \bar{Y}P\bar{X}$

Step 6: $C \Rightarrow U, \Sigma , V^T$ (SVD)

Step 7: $R = U\Sigma V^T$, where $\Sigma = diag(1, 1, ..., 1, \pm 1)$

Step 8: $s_{xy} = tr\left\{RC\right\}/n$

Step 9: $k = s_{xy}/s_x^2$

Step 10: $T = \bar{Y} - kR\bar{X}$

Step 11: $l = Te/n$

END

2.2. The rotation angels in 3-D

In $m$-dimensional space the rotation matrix is a rotation around the $m$ coordinate axes. In 3-D with the rotation angels $(\alpha, \beta, \gamma)$ around the three axes, the rotation matrix can be expressed

$$R = \begin{pmatrix}
1 & \cos \alpha & -\sin \alpha \\
\sin \alpha & \cos \alpha & \cos \beta \\
-\sin \beta & \cos \beta & \cos \gamma
\end{pmatrix},$$

(16)
which can be written as an explicit (3x3)-matrix. In particular one obtains the elements

\[ R_{11} = \cos \beta \cos \gamma, \quad R_{12} = -\cos \beta \sin \lambda, \quad R_{13} = -\sin \beta, \]

(17a)

and

\[ R_{23} = -\sin \alpha \cos \beta, \quad R_{33} = \cos \alpha \cos \beta, \]

(17b)

which yield

\[ \tan \alpha = \frac{R_{23}}{R_{33}}, \quad \tan \beta = \frac{R_{33}}{\sqrt{R_{11}^2 + R_{12}^2}}, \quad \text{and} \quad \tan \gamma = -\frac{R_{12}}{R_{11}}, \]

(18)

3. An iterative procedure for the rotation matrix

The closed-form solution is obtained at the prize of utilizing SVD. Späth (2003) and (2004) as well as Watson (2006) avoided this by presenting iterative solutions for the transformation parameters in 3-D for all observation pairs assumed being uncorrelated and weighted equally. However, both the scale factor and the translation vector were already solved by closed formulas in the previous section, so only the somewhat tedious solution by SVD for the rotation matrix may need to be simplified. Here we will outline such an iterative solution. The presentation is limited to 3-D, but could easily be extended to any dimension.

As stated in Lemma 3, the rotation matrix maximizes the objective function

\[ h(\alpha, \beta, \gamma) = tr\{RC\}, \quad C = \tilde{X}\tilde{Y}^T, \]

(19)

and the maximum is obtained by the three equations

\[ h_i = tr\left\{ \frac{dR\cdot C}{du_i} \right\} = 0; (u_1, u_2, u_3) = (\alpha, \beta, \gamma); \ i = 1, 2, 3. \]

(20)

By making a Taylor expansion to first order of each of these equations, one arrives at the (approximate) system of equations

\[ h_i = h_{i0} + h_{i\alpha} d\alpha + h_{i\beta} d\beta + h_{i\gamma} d\gamma, \quad i = 1, 2, 3, \]

(21)

or, in matrix form,

\[ \begin{bmatrix} h'_{i\alpha} & h'_{i\beta} & h'_{i\gamma} \\ h_{i\alpha} & h_{i\beta} & h_{i\gamma} \\ h_{i\alpha} & h_{i\beta} & h_{i\gamma} \end{bmatrix} \begin{bmatrix} d\alpha \\ d\beta \\ d\gamma \end{bmatrix} = \begin{bmatrix} h_{i0} \\ h_{i0} \\ h_{i0} \end{bmatrix}, \]

(22)

where \( h_{i0}, h'_{i\alpha}, h'_{i\beta}, \) and \( h'_{i\gamma} \) are approximations to \( h_i \), and its first derivatives, respectively, estimated with approximate values for \( (\alpha, \beta, \gamma) \). The solutions of the system are the corrections \( (d\alpha, d\beta, d\gamma) \) to the approximate angles. By correcting the angels and using them as input to all elements of the design matrix and right hand-side vector of Eq. (22), the solution can be iterated until convergence, which, of course, is attained only for initial approximations sufficiently close to the true values. In geodetic problems the rotation matrix is frequently close to unity, so that the starting values for the rotation angels could often be set to zero. However, in other cases the selection of suitable starting values could be a problem.

4. Some error aspects

Corollary 2: Let the coordinate matrices \( X \) and \( Y \) be unbiased, uncorrelated stochastic variables. Then Theorem 1 yields an unbiased estimate of the rotation matrix only if \( P \) is the unit matrix or deviates from it by a constant.

Proof. If \( X \) and \( Y \) are unbiased and uncorrelated, these properties hold also for \( \tilde{X} \) and \( \tilde{Y} \). Introducing the notations \( \xi \) and \( \eta \) for the true values and \( e_x \) and \( e_y \) for the error of the latter matrices, the expectation of the generating matrix of the rotation matrix becomes

\[ E\{ \tilde{X}\tilde{Y}^T \} = E\{ (\xi + e_x)P (\eta + e_y)^T \} = \xi P \eta^T, \]

(23)

which has the same eigenvalues as \( \xi \eta^T \) only if \( P \) is unit matrix times a constant.

If the rotation matrix is biased, it follows also that the scale factor and translation matrix are biased. However, for a fixed rotation matrix and scale factor the estimated translation parameters, given by the theorem, are always unbiased, because

\[ E\{ e_T \} = E\{ e_T - kR e_t \} = 0. \]

Finally, the mean square residual of the least squares fitting \( (s^2) \) follows from Eqs (14a) and (15):

\[ s^2 = s_{xy}^2 - s_{xy}^2/s_{xx}^2. \]

(25)

The residual vector for each observation point after a preliminary determination of the transformation parameters (see Eq. (1)):

\[ \hat{e}_i = y_i - t - kR e_i \]

(26)

can now be tested for outliers by the test parameter \( p_{ii} e_i^T (ms^2)^{-1} e_i \) against some tolerance level (say 2). Here \( p_{ii} \) is the i-th diagonal element of matrix \( P \). If the test parameter exceeds the tolerance level, the observations of that point are regarded as outliers for possible discarding, otherwise not, and the adjustment may be repeated and iterated until all outliers are removed.
5. Conclusions

Frequently the coordinate pairs, which are used for determining the Helmert transformation parameters between two coordinate systems, have variable qualities and distributions in the area of interest, and they may also be statistically correlated. Here a weight matrix is introduced to consider such non-homogeneities in the underlying data. A closed-form solution was derived for the least-squares problem, but the rotation matrix (conditioned to be orthonormal with determinant +1; the second condition frequently not considered) could only be achieved at the price of utilizing the somewhat tedious SVD technique. As an alternative, an iterative method was outlined for computing the rotation matrix. However, generally, the iterative solution needs starting values, which are frequently not available, and, although SVD is avoided, the computing time with iterations is not necessarily faster. We presented also a test procedure for discriminating outlying data.

Finally, it is a well-known fact among geodesists that the translation and rotation parameters between geocentric and local coordinate systems, when determined by standard Helmert transformation technique, become highly correlated. However, this correlation is reduced in our method, making modifications, such as the Molodensky-Badekas alternative, superfluous.

Acknowledgements

I appreciate the helpful recommendations on the preliminary manuscript by the editor, Dr. S. Claessens, and two unknown reviewers.

References


Grafarend E. W. and Awange J. L., 2003, Nonlinear analysis of the three-dimensional datum transformation conformal group C7 (3), J Geod. 77, 66-76

Helmert F.R., 1924, Ausgleichungsrechnung nach der Methode der kleinsten Quadrate mit Anwendungen auf die Geodäsie, die Physik und die Theorie der Messinstrumente, 3rd edition, Leipzig and Berlin


Lissitz, R. W., Schönemann P. H. and Lingoes J. C., 1976, A solution to the weighted Procrustes problem in which the transformation is in agreement with the loss function. Psychometrika 41, 4, 547-550


Schönemann P. H., 1966, Generalised solution of the orthogonal Procrustes problem, Psychometrika 31, 1-10


Späth H., 2003, Identifying spatial point sets, Math. Com. 8, 69-75

