A Purely Algebraic Justification of the Kabsch-Umeyama Algorithm

Jim Lawrence[1,2], Javier Bernal[1], Christoph Witzgall[1]

[1]*National Institute of Standards and Technology,*

Gaithersburg, MD 20899, USA

[2]*George Mason University,*

4400 University Dr, Fairfax, VA 22030, USA

{james.lawrence,javier.bernal}@nist.gov   lawrence@gmu.edu

Abstract

The constrained orthogonal Procrustes problem is the least-squares problem that calls for a rotation matrix that optimally aligns two matrices of the same order. Over past decades, the algorithm of choice for solving this problem has been the Kabsch-Umeyama algorithm which is essentially no more than the computation of the singular value decomposition of a particular matrix. Its justification as presented separately by Kabsch and Umeyama is not totally algebraic as it is based on solving the minimization problem via Lagrange multipliers. In order to provide a more transparent alternative, it is the main purpose of this paper to present a purely algebraic justification of the algorithm through the exclusive use of simple concepts from linear algebra. For the sake of completeness, a proof is also included of the well-known and widely-used fact that the orientation-preserving rigid motion problem, i.e., the least-squares problem that calls for an orientation-preserving rigid motion that optimally aligns two corresponding sets of points in d-dimensional Euclidean space, reduces to the constrained orthogonal Procrustes problem.

Key words. constrained; Frobenius; least-squares; orientation-preserving; orthogonal; Procrustes; rigid motion; rotation; singular value decomposition; trace

MSC. 15A18, 15A42, 65H17, 65K99, 93B60

1   Introduction

In the *orthogonal Procrustes problem* [2, 8], given real matrices P and Q of size $d \times n$, the problem is that of finding a $d \times d$ orthogonal matrix U that minimizes $\|UQ - P\|_F$, where $\| \cdot \|_F$ denotes the Frobenius norm of a matrix. On the other hand, in the *constrained orthogonal Procrustes problem* [4, 5, 10], the same function is minimized but U is constrained to be a rotation matrix, i.e., an orthogonal matrix of determinant 1. By letting p_i, q_i, $i = 1 \ldots, n$, be the vectors in $\mathbb{R}^d$ that are the columns from left to right of P and Q, respectively, since clearly $\|UQ - P\|_F^2 = \sum_{i=1}^n \|Uq_i - p_i\|^2$, where $\| \cdot \|$ denotes the $d-$dimensional Euclidean norm, then an alternative formulation of the two problems above is that of finding an orthogonal matrix U (of determinant 1 for the constrained problem) that minimizes $\sum_{i=1}^n \|Uq_i - p_i\|^2$. We note that minimizing matrices do exist for the two problems as the function being minimized is continuous and both the set of orthogonal matrices and the set of rotation matrices are compact (in some topology). Finally, in the same vein, another problem of interest is the *orientation-preserving rigid motion problem* which is that of finding an orientation-preserving rigid motion ϕ of $\mathbb{R}^d$ that minimizes $\sum_{i=1}^n \|\phi(q_i) - p_i\|^2$. An affine linear function ϕ, $\phi : \mathbb{R}^d \to \mathbb{R}^d$, is a rigid motion of $\mathbb{R}^d$ if it is of the form $\phi(q) = Uq + t$ for $q \in \mathbb{R}^d$, where U is a $d \times d$ orthogonal matrix, and t is a vector in $\mathbb{R}^d$. The rigid motion ϕ is orientation preserving if $\det(U) = 1$, i.e., the determinant of U equals 1. With $\bar{p}$, $\bar{q}$ denoting the centroids of $\{p_i\}$, $\{q_i\}$, respectively, as will be shown in Section 3 of this paper, this problem can be reduced to the constrained orthogonal Procrustes problem by translating $\{p_i\}$, $\{q_i\}$ to become $\{p_i - \bar{p}\}$, $\{q_i - \bar{q}\}$, respectively, so that the centroid of each set becomes $0 \in \mathbb{R}^d$.

With P, Q, p_i, q_i, $i = 1, \ldots, n$, as above, in this paper we focus our attention mostly on the constrained orthogonal Procrustes problem, and therefore wish to find a $d \times d$ rotation matrix U that minimizes $\sum_{i=1}^n \|Uq_i - p_i\|^2$.

With this purpose in mind, we rewrite $\sum_{i=1}^n \|Uq_i - p_i\|^2$ as follows, where given a square matrix R, $\mathrm{tr}(R)$ stands for the trace of R.

$$\sum_{i=1}^n ||Uq_i - p_i||^2 = \sum_{i=1}^n (Uq_i - p_i)^T (Uq_i - p_i) = \mathrm{tr}\big((UQ - P)^T (UQ - P)\big)$$

$$= \mathrm{tr}\big((Q^T U^T - P^T)(UQ - P)\big) = \mathrm{tr}(Q^T Q + P^T P - Q^T U^T P - P^T UQ)$$

$$= \mathrm{tr}(Q^T Q) + \mathrm{tr}(P^T P) - 2\mathrm{tr}(P^T UQ).$$

2

Since only the third term depends on U, it suffices to find a $d \times d$ rotation matrix U that maximizes $\text{tr}(P^T U Q)$. Since $\text{tr}(P^T U Q) = \text{tr}(U Q P^T)$ (note in general $\text{tr}(AB) = \text{tr}(BA)$, A an $n \times d$ matrix, B a $d \times n$ matrix), denoting the $d \times d$ matrix $Q P^T$ by M, this problem is equivalent to finding a $d \times d$ rotation matrix U that maximizes $\text{tr}(UM)$, and it is well known that one such U can be computed from the singular value decomposition of M [4, 5, 10]. This is done with the Kabsch-Umeyama algorithm [4, 5, 10] (see Algorithm Kabsch-Umeyama below, where $\text{diag}\{s_1, \ldots, s_d\}$ is the $d \times d$ diagonal matrix with numbers $s_1, \ldots, s_d$ as the elements of the diagonal, in that order running from the upper left to the lower right of the matrix). A *singular value decomposition* (SVD) [6] of M is a representation of the form $M = V S W^T$, where V and W are $d \times d$ orthogonal matrices and S is a $d \times d$ diagonal matrix with the singular values of M, which are non-negative real numbers, appearing in the diagonal of S in descending order, from the upper left to the lower right of S. Finally, note that any matrix, not necessarily square, has a singular value decomposition, not necessarily unique [6].

Algorithm Kabsch-Umeyama

Compute $d \times d$ matrix $M = Q P^T$.
Compute SVD of M, i.e., identify $d \times d$ matrices V, S, W,
so that $M = V S W^T$ in the SVD sense.
Set $s_1 = \ldots = s_{d-1} = 1$.
If $\det(VW) > 0$, then set $s_d = 1$, else set $s_d = -1$.
Set $\tilde{S} = \text{diag}\{s_1, \ldots, s_d\}$.
Return $d \times d$ rotation matrix $U = W \tilde{S} V^T$.

Algorithm Kabsch-Umeyama has existed for several decades [4, 5, 10], however the known justifications of the algorithm [4, 5, 10] are not totally algebraic as they are based on exploiting the optimization technique of Lagrange multipliers. It is the main purpose of this paper to justify the algorithm in a purely algebraic manner through the exclusive use of simple concepts from linear algebra. This is done in Section 2 of the paper. Finally, we note that applications of the algorithm can be found, in particular, in the field of functional and shape data analysis [1, 9].

2 Algebraic justification of the Kabsch-Umeyama algorithm

We justify Algorithm Kabsch-Umeyama using exclusively simple concepts from linear algebra, mostly in the proof of the following useful proposition. We note that most of the proof of the proposition is concerned with proving 3. of the proposition. Thus, it seems reasonable to say that any justification of the algorithm that requires the conclusion in 3. but lacks a proof for it, is not exactly complete. See page 47 of the otherwise excellent thesis in [7] for an example of this situation. See [3] for an outline of this dissertation.

Proposition 1: If $D = \text{diag}\{\sigma_1, \ldots, \sigma_d\}$, $\sigma_j \geq 0$, $j = 1, \ldots, d$, and W is a $d \times d$ orthogonal matrix, then
1. $\text{tr}(WD) \leq \sum_{j=1}^{d} \sigma_j$.
2. If B is a $d \times d$ orthogonal matrix, $S = B^T D B$, then $\text{tr}(WS) \leq \text{tr}(S)$.
3. If $\det(W) = -1$, $\sigma_d \leq \sigma_j$, $j = 1, \ldots, d-1$, then $\text{tr}(WD) \leq \sum_{j=1}^{d-1} \sigma_j - \sigma_d$.

Proof: Since W is orthogonal and if W_{kj}, $k, j = 1, \ldots, d$, are the entries of W, then, in particular, $W_{jj} \leq 1$, $j = 1, \ldots, d$, so that
$\text{tr}(WD) = \sum_{j=1}^{d} W_{jj}\sigma_j \leq \sum_{j=1}^{d} \sigma_j$, and therefore 1. holds.
Accordingly, assumming B is a $d \times d$ orthogonal matrix, since BWB^T is also orthogonal, it follows from 1. that
$\text{tr}(WS) = \text{tr}(WB^T DB) = \text{tr}(BWB^T D) \leq \sum_{j=1}^{d} \sigma_j = \text{tr}(D) = \text{tr}(S)$, and therefore 2. holds.
If $\det(W) = -1$, we show next that a $d \times d$ orthogonal matrix B can be identified so that with $\bar{W} = B^T W B$, then $\bar{W} = \begin{pmatrix} W_0 & O \\ O^T & -1 \end{pmatrix}$, W_0 interpreted as the upper leftmost $d-1 \times d-1$ entries of $\bar{W}$ and as a $d-1 \times d-1$ matrix as well; O interpreted as a vertical column or vector of $d-1$ zeroes.
With I as the $d \times d$ identity matrix, then $\det(W) = -1$ implies $\det(W+I) = -\det(W)\det(W+I) = -\det(W^T)\det(W+I) = -\det(I+W^T) = -\det(I+W)$ which implies $\det(W + I) = 0$ so that $x \neq 0$ exists in $\mathbb{R}^d$ with $Wx = -x$. It also follows then that $W^T W x = W^T(-x)$ which gives $x = -W^T x$ so that $W^T x = -x$ as well.
Letting $b_d = x$, vectors $b_1, \ldots, b_{d-1}$ can be obtained so that $b_1, \ldots, b_d$ form a basis of $\mathbb{R}^d$, and by the Gram-Schmidt process starting with b_d, we may assume $b_1, \ldots, b_d$ form an orthonormal basis of $\mathbb{R}^d$ with $Wb_d = W^T b_d = -b_d$.

4

Letting $B = (b_1, \ldots, b_d)$, interpreted as a $d \times d$ matrix with columns $b_1, \ldots, b_d$, in that order, it then follows that B is orthogonal, and with $\bar{W} = B^T W B$ and W_0, O as previously described, noting $B^T W b_d = B^T(-b_d) = \binom{O}{-1}$ and $b_d^T W B = (W^T b_d)^T B = (-b_d)^T B = (O^T - 1)$, then $\bar{W} = \begin{pmatrix} W_0 & O \\ O^T & -1 \end{pmatrix}$. Note $\bar{W}$ is orthogonal and therefore so is the $d - 1 \times d - 1$ matrix W_0.

Let $S = B^T D B$ and write $S = \begin{pmatrix} S_0 & a \\ b^T & \gamma \end{pmatrix}$, S_0 interpreted as the upper leftmost $d - 1 \times d - 1$ entries of S and as a $d - 1 \times d - 1$ matrix as well; a and b interpreted as vertical columns or vectors of $d - 1$ entries, and γ as a scalar. Note $\text{tr}(WD) = \text{tr}(B^T W D B) = \text{tr}(B^T W B B^T D B) = \text{tr}(\bar{W}S)$, so that $\bar{W}S = \begin{pmatrix} W_0 & O \\ O^T & -1 \end{pmatrix} \begin{pmatrix} S_0 & a \\ b^T & \gamma \end{pmatrix} = \begin{pmatrix} W_0 S_0 & W_0 a \\ -b^T & -\gamma \end{pmatrix}$ gives $\text{tr}(WD) = \text{tr}(W_0 S_0) - \gamma$.

We show $\text{tr}(W_0 S_0) \le \text{tr}(S_0)$. For this purpose let $\hat{W} = \begin{pmatrix} W_0 & O \\ O^T & 1 \end{pmatrix}$, W_0 and O as above. Since W_0 is orthogonal, then clearly $\hat{W}$ is a $d \times d$ orthogonal matrix, and by 2. $\text{tr}(\hat{W}S) \le \text{tr}(S)$ so that $\hat{W}S = \begin{pmatrix} W_0 & O \\ O^T & 1 \end{pmatrix} \begin{pmatrix} S_0 & a \\ b^T & \gamma \end{pmatrix} = \begin{pmatrix} W_0 S_0 & W_0 a \\ b^T & \gamma \end{pmatrix}$ gives $\text{tr}(W_0 S_0) + \gamma = \text{tr}(\hat{W}S) \le \text{tr}(S) = \text{tr}(S_0) + \gamma$. Thus, $\text{tr}(W_0 S_0) \le \text{tr}(S_0)$.

Note $\text{tr}(S_0) + \gamma = \text{tr}(S) = \text{tr}(D)$, and if B_{kj}, $k, j = 1, \ldots, d$ are the entries of B, then $\gamma = \sum_{k=1}^{d} B_{kd}^2 \sigma_k$, a convex combination of the σ_k's, so that $\gamma \ge \sigma_d$. It then follows that
$\text{tr}(WD) = \text{tr}(W_0 S_0) - \gamma \le \text{tr}(S_0) - \gamma = \text{tr}(D) - \gamma - \gamma \le \sum_{j=1}^{d-1} \sigma_j - \sigma_d$, and therefore 3. holds. $\square$

Finally, the following theorem, a consequence of Proposition 1, justifies the Kabsch-Umeyama algorithm.

Theorem: Given a $d \times d$ matrix M, let V, S, W be $d \times d$ matrices such that the singular value decomposition of M gives $M = VSW^T$. If $\det(VW) > 0$, then $U = WV^T$ maximizes $\text{tr}(UM)$ over all $d \times d$ rotation matrices U. Otherwise, with $\tilde{S} = \text{diag}\{s_1, \ldots, s_d\}$, $s_1 = \ldots = s_{d-1} = 1$, $s_d = -1$, then $U = W\tilde{S}V^T$ maximizes $\text{tr}(UM)$ over all $d \times d$ rotation matrices U.

Proof: Let σ_j, $j = 1, \ldots, d$, $\sigma_1 \ge \sigma_2 \ge \ldots \ge \sigma_d \ge 0$, be the singular values of M, so that $S = \text{diag}\{\sigma_1, \ldots, \sigma_d\}$.
Assume $\det(VW) > 0$. If U is any rotation matrix, then U is orthogonal. From 1. of Proposition 1 since $W^T U V$ is orthogonal, then
$\text{tr}(UM) = \text{tr}(UVSW^T) = \text{tr}(W^T UVS) \le \sum_{j=1}^{d} \sigma_j$.
On the other hand, if $U = WV^T$, then U is clearly orthogonal, $\det(U) = 1$,

and $\mathrm{tr}(UM) = \mathrm{tr}(WV^TVSW^T) = \mathrm{tr}(WSW^T) = \mathrm{tr}(S) = \sum_{j=1}^{d} \sigma_j$.

Thus, $U = WV^T$ maximizes $\mathrm{tr}(UM)$ over all $d \times d$ rotation matrices U. Finally, assume $\det(VW) < 0$. If U is any rotation matrix, then U is orthogonal and $\det(U) = 1$. From 3. of Proposition 1 since W^TUV is orthogonal and $\det(W^TUV) = -1$, then

$\mathrm{tr}(UM) = \mathrm{tr}(UVSW^T) = \mathrm{tr}(W^TUVS) \leq \sum_{j=1}^{d-1} \sigma_j - \sigma_d$.

On the other hand, if $U = W\tilde{S}V^T$, then U is clearly orthogonal, $\det(U) = 1$, and $\mathrm{tr}(UM) = \mathrm{tr}(W\tilde{S}V^TVSW^T) = \mathrm{tr}(W\tilde{S}SW^T) = \mathrm{tr}(\tilde{S}S) = \sum_{j=1}^{d-1} \sigma_j - \sigma_d$.

Thus, $U = W\tilde{S}V^T$ maximizes $\mathrm{tr}(UM)$ over all $d \times d$ rotation matrices U. $\square$

3 Reduction of the orientation-preserving rigid motion problem to the constrained orthogonal Procrustes problem

Although not exactly related to the main goal of this paper, for the sake of completeness, we show the orientation-preserving rigid motion problem reduces to the constrained orthogonal Procrustes problem. For this purpose, let $\bar{q}$ and $\bar{p}$ denote the centroids of the sets $\{q_i\}_{i=1}^{n}$ and $\{p_i\}_{i=1}^{n}$ in $\mathbb{R}^d$, respectively:

$$\bar{q} = \frac{1}{n} \sum_{i=1}^{n} q_i \quad \text{and} \quad \bar{p} = \frac{1}{n} \sum_{i=1}^{n} p_i \ .$$

First, we prove a proposition that shows, in particular, $\phi(\bar{q}) = \bar{p}$ if ϕ minimizes

$$\Delta(\phi) = \sum_{i=1}^{n} \|\phi(q_i) - p_i\|^2$$

over either the set of all rigid motions ϕ of $\mathbb{R}^d$ or the smaller set of rigid motions ϕ of $\mathbb{R}^d$ that are orientation preserving.

Proposition 2: Let ϕ be a rigid motion of $\mathbb{R}^d$ with $\phi(\bar{q}) \neq \bar{p}$ and define an affine linear function τ, $\tau : \mathbb{R}^d \to \mathbb{R}^d$, $\tau(q) = \phi(q) - \phi(\bar{q}) + \bar{p}$ for $q \in \mathbb{R}^d$. Then τ is a rigid motion of $\mathbb{R}^d$, $\tau(\bar{q}) = \bar{p}$, $\Delta(\tau) < \Delta(\phi)$, and if ϕ is orientation preserving, then so is τ.

Proof: Clearly $\tau(\bar{q}) = \bar{p}$. Let U be a $d \times d$ orthogonal matrix and $t \in R^d$ such that $\phi(q) = Uq + t$ for $q \in \mathbb{R}^d$. Then $\tau(q) = Uq - U\bar{q} + \bar{p}$ so that τ is

a rigid motion of $\mathbb{R}^d$, τ is orientation preserving if ϕ is, and for $1 \leq i \leq n$ we have

$$||\phi(q_i) - p_i||^2 - ||\tau(q_i) - p_i||^2 =$$

$$(Uq_i + t - p_i)^T(Uq_i + t - p_i) - (Uq_i - U\bar{q} + \bar{p} - p_i)^T(Uq_i - U\bar{q} + \bar{p} - p_i) =$$

$$\big((Uq_i - p_i)^T(Uq_i - p_i) + 2(Uq_i - p_i)^T t + t^T t\big)$$

$$-\big((Uq_i - p_i)^T(Uq_i - p_i) - 2(Uq_i - p_i)^T(U\bar{q} - \bar{p}) + (U\bar{q} - \bar{p})^T(U\bar{q} - \bar{p})\big)$$

$$= 2(Uq_i - p_i + t)^T(U\bar{q} - \bar{p} + t) - (U\bar{q} - \bar{p} + t)^T(U\bar{q} - \bar{p} + t).$$

It follows that

$$\Delta(\phi) - \Delta(\tau) = \sum_{i=1}^{n} \big(2(Uq_i - p_i + t)^T(U\bar{q} - \bar{p} + t) - (U\bar{q} - \bar{p} + t)^T(U\bar{q} - \bar{p} + t)\big)$$

$$= n||U\bar{q} - \bar{p} + t||^2 = n||\phi(\bar{q}) - \bar{p}||^2 > 0 \quad \text{as } \phi(\bar{q}) - \bar{p} \text{ is nonzero. } \square$$

Finally, the following corollary, a consequence of Proposition 2, shows that the problem of finding an orientation-preserving rigid motion ϕ of $\mathbb{R}^d$ that minimizes $\sum_{i=1}^{n} ||\phi(q_i) - p_i||^2$ can be reduced to a constrained orthogonal Procrustes problem which, of course, then can be solved with the Kabsch-Umeyama algorithm. Here $r_i = p_i - \bar{p}$, $s_i = q_i - \bar{q}$, for $i = 1, \ldots, n$, and if $\bar{r} = \frac{1}{n}\sum_{i=1}^{n} r_i$, $\bar{s} = \frac{1}{n}\sum_{i=1}^{n} s_i$, then clearly $\bar{r} = \bar{s} = 0$.

Corollary: Let $\hat{U}$ be such that $U = \hat{U}$ minimizes $\sum_{i=1}^{n} ||Us_i - r_i||^2$ over all $d \times d$ rotation matrices U. Let $\hat{t} = \bar{p} - \hat{U}\bar{q}$, and let $\hat{\phi}$ be given by $\hat{\phi}(q) = \hat{U}q + \hat{t}$ for $q \in \mathbb{R}^d$. Then $\phi = \hat{\phi}$ minimizes $\sum_{i=1}^{n} ||\phi(q_i) - p_i||^2$ over all orientation-preserving rigid motions ϕ of $\mathbb{R}^d$.

Proof: One such $\hat{U}$ can be computed with the Kabsch-Umeyama algorithm. By Proposition 2, in order to minimize $\sum_{i=1}^{n} ||\phi(q_i) - p_i||^2$ over all orientation-preserving rigid motions ϕ of $\mathbb{R}^d$, it suffices to do it over those for which $\phi(\bar{q}) = \bar{p}$. Therefore, it suffices to minimize $\sum_{i=1}^{n} ||Uq_i + t - p_i||^2$ with $t = \bar{p} - U\bar{q}$ over all $d \times d$ rotation matrices U, i.e., it suffices to minimize

$$\sum_{i=1}^{n} ||Uq_i + \bar{p} - U\bar{q} - p_i||^2 = \sum_{i=1}^{n} ||(U(q_i - \bar{q}) - (p_i - \bar{p})||^2$$

over all $d \times d$ rotation matrices U. But minimizing the last expression is equivalent to minimizing $\sum_{i=1}^{n} ||Us_i - r_i||^2$ over all $d \times d$ rotation matrices U.

Since $U = \hat{U}$ is a solution to this last problem, it then follows that $U = \hat{U}$ minimizes $\sum_{i=1}^{n} \|Uq_i + \bar{p} - U\bar{q} - p_i\|^2 = \sum_{i=1}^{n} \|Uq_i + t - p_i\|^2$ with $t = \bar{p} - U\bar{q}$ over all $d \times d$ rotation matrices U. Consequently, if $\hat{t} = \bar{p} - \hat{U}\bar{q}$, and $\hat{\phi}$ is given by $\hat{\phi}(q) = \hat{U}q + \hat{t}$ for $q \in \mathbb{R}^d$, then $\phi = \hat{\phi}$ clearly minimizes $\sum_{i=1}^{n} \|\phi(q_i) - p_i\|^2$ over all orientation-preserving rigid motions ϕ of $\mathbb{R}^d$. $\square$

References

[1] Dogan, G., Bernal, J., Hagwood, C. R.: FFT-based alignment of 2d closed curves with application to elastic shape analysis. Proceedings of the 1st DIFF-CV Workshop, British Machine Vision Conference. (2015)

[2] Gower, J. C., Dijksterhuis, G. B.: Procrustes Problems. Oxford University Press. (2004)

[3] Higham, N.: Matrix Procrustes Problems. (1994) http://www.ma.man.ac.uk/ higham/talks

[4] Kabsch, W.: A solution for the best rotation to relate two sets of vectors. Acta Crystallographica Section A: Crystal Physics. 32(5): 922-923 (1976).

[5] Kabsch, W.: A discussion of the solution for the best rotation to relate two sets of vectors. Acta Crystallographica Section A: Crystal Physics. 34(5): 827-828 (1978).

[6] Lay, D. C., Lay, S. R., McDonald, J. J.: Linear Algebra and its Applications, 5th edition. Boston: Pearson Education. (2016)

[7] Papadimitriou, P.: Parallel Solution of SVD-Related Problems, with Applications. Ph. D. Thesis, University of Manchester, England, 1993.

[8] Schönemann, P. H.: A generalized solution of the orthogonal Procrustes problem. Psychometrika. 31(1): 1-10 (1966)

[9] Srivastava, A., Klassen, E. P.: Functional and Shape Data Analysis. New York: Springer. (2016)

[10] Umeyama, S.: Least-Squares Estimation of Transformation Parameters Between Two Point Patterns. IEEE Trans. Pattern Analysis and Machine Intelligence. 13(4): 376-380 (1991).