The Canadian Journal of Statistics
Vol. 34, No 2, 2006, Pages ???–???
La revue canadienne de statistique

Regression and correlation for 3×3 rotation matrices

Louis-Paul RIVEST and Ted CHANG

Key words and phrases: Directional data; joint kinematics; multivariate statistics; quaternions; rotations.
MSC 2000: Primary 62H11; secondary 62H20.

Abstract: The authors investigate a regression model for orthogonal matrices introduced by Prentice (1989). They focus on the special case of 3×3 rotation matrices. The model under study expresses the dependent rotation matrix V as $A_1 U A_2^\top$ perturbed by experimental errors, where A_1 and A_2 are unknown 3×3 rotation matrices and U is an explanatory 3×3 rotation matrix. Several specifications for the errors in this regression model are proposed. The asymptotic distributions of the least squares estimators for A_1 and A_2 are derived when the sample size becomes large or as the experimental errors become small. A new algorithm for calculating the least squares estimates of A_1 and A_2 is presented. As the independence model is not a sub-model of Prentice's regression model, the independence between the U and the V sample cannot be tested when fitting Prentice's model. To overcome this difficulty, the authors investigate permutation tests of independence. They illustrate their methodology using examples dealing with postural variations of subjects performing a drilling task and with the calibration of a camera system for motion analysis using a magnetic tracking device.

Un modèle de régression et une mesure de corrélation pour des matrices de rotation 3×3

Résumé : Les auteurs étudient un modèle de régression pour des matrices de corrélation introduit par Prentice (1989). Ils s'intéressent plus particulièrement au cas des matrices de rotations 3×3. Ce modèle exprime la rotation dépendante V comme $A_1 U A_2^\top$ perturbée par des erreurs expérimentales, où A_1 et A_2 sont des matrices de rotation 3×3 inconnues et U est une matrice de rotation 3×3 explicative. Plusieurs spécifications pour les erreurs de ce modèle de régression sont mises de l'avant. Les distributions asymptotiques des estimateurs des moindres carrés de A_1 et A_2 sont calculées lorsque la taille d'échantillon tend vers l'infini ou que les erreurs expérimentales deviennent petites. Un nouvel algorithme de calcul pour les estimateurs des moindres carrés de A_1 et A_2 est présenté. Le modèle d'indépendance n'étant pas un cas particulier du modèle de Prentice, l'indépendance entre U et V ne peut pas être testée en ajustant ce modèle. Pour pallier ce problème, les auteurs construisent des tests d'indépendance de permutation. Ils illustrent leur méthodologie au moyen d'exemples portant sur la variation des postures lorsqu'un sujet manipule une perceuse et sur la calibration d'un système de caméras pour l'étude du mouvement utilisant une méthode de repérage magnétique.

1. INTRODUCTION

Camera systems that record, at a very high frequency, the three dimensional coordinates of markers attached to experimental subjects are powerful tools for the study of human kinematics. These machines give the time varying positions and orientations of markers attached to the body of experimental subjects. The orientation of a marker is reported as a 3×3 rotation matrix. The statistical treatment of markers orientation raises new problems in directional data analysis. For instance, Rancourt, Rivest & Asselin (2000) applied large concentration statistical procedures for the matrix Fisher–von Mises distribution of Downs (1972) to compare the postures, as characterized by the relative orientations of the three joints of the right arm, of experimental subjects performing various drilling tasks.

Regression and correlation problems for rotations arise when treating the data collected by camera systems studying human kinematics. When a magnetic tracking device is used to collect the orientations and the locations of markers, systematic deviations occur in the measurements that are taken far from the signal's source. Statistical techniques are needed to calibrate these machines; see Day, Dumas & Murdoch (1998) and Day, Murdoch & Dumas (2000). Calibrating the orientations involves a regression model for expressing the relationship between the observed and the true 3×3 rotation matrices characterizing the orientations of the device. Correlation problems occur when investigating the extent of the association between the orientations of two joints, say the wrist and the elbow, of an upper limb performing a task.

The regression model for matched pairs of orientations due to Prentice (1989) is the statistical tool for investigating these questions. It was introduced as a general regression model on Stiefel manifolds. It expresses a dependent orthogonal matrix V as $A_1 U A_2^\top$ perturbed by experimental errors, where U is a known orthogonal matrix having the same dimension as V, and A_1 and A_2 are unknown rotation matrices. Prentice (1989) derives statistical properties of the least squares estimators of A_1 and A_2 by adapting techniques developed by Chang (1986) for the spherical regression model. Unfortunately, many of the derivations reported in Prentice (1989) appear to be in error; see Chang & Rivest (2001) for details.

This paper is motivated by three-dimensional applications. Prentice's model is investigated when U, V, A_1 and A_2 belong to $SO(3)$, the set of 3×3 rotation matrices. As exemplified in Section 5.1, the errors in Prentice's regression model do not exhibit the high degree of symmetry that is assumed in Prentice (1989) and Chang & Rivest (2001) in many applications. Flexible error models for rotation matrices are first introduced and the asymptotic distribution of the least squares estimators of A_1 and A_2 are derived under these models. A test of independence in a bivariate sample of 3×3 rotation matrices is also proposed.

Quaternions and rotations in $SO(3)$ are reviewed in Section 2. Section 3 introduces a wide class of error distributions for Prentice's model, which contains the symmetric model of Prentice (1989) and Chang & Rivest (2001) as special cases. Section 4 gives a condition on the explanatory rotations insuring that the least squares estimates of A_1 and A_2 are uniquely defined. It provides an expansion of the residual sum of squares for Prentice's model around two arbitrary rotations A_1 and A_2. This expansion is then used to derive the asymptotic distributions of the least squares estimators of A_1 and A_2 and to construct a new scoring algorithm for calculating the least squares estimates of A_1 and A_2. The permutation test of independence is proposed in Section 5. Data analyses are presented in Section 6. The calibration of the orientations reported by a camera system based on a magnetic tracking device is investigated with Prentice's regression model. A correlation analysis of postural variations of the upper limb joints performing a drilling task is also presented.

2. SKEW-SYMMETRIC MATRICES, 3×3 ROTATION MATRICES, AND QUATERNIONS

Statistical models for rotation matrices involve non-linear functions of both the parameters and the experimental errors, see for instance Chang, Ko, Royer & Lu (2000). The set of 3×3 rotation matrices, $SO(3)$, has a rich mathematical structure. It is a Lie group and the set of 3×3 skew-symmetric matrices is the Lie algebra associated to this Lie group (see Warner 1983). Skew-symmetric matrices play a key role in the statistical models for rotation matrices. Given a vector

a in $\mathbb{R}^3$, one can construct a 3×3 skew-symmetric matrix as follows,

$$\mathcal{S}(a) = \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}.$$

The function $\mathcal{S}^{-1}$ maps a skew-symmetric matrix B into its 3×1 component vector b satisfying $\mathcal{S}(b) = B$. Rotation matrices can be represented as exponentials of skew-symmetric matrices. Let a be a non-null vector in $\mathbb{R}^3$ and define the angle $\theta = (a^\top a)^{1/2}$ and the unit vector $u = a/\theta$. The exponential of $\mathcal{S}(a)$ is given by

$$\begin{aligned} \exp\{\mathcal{S}(a)\} &= I + \mathcal{S}(a) + \frac{1}{2!}\{\mathcal{S}(a)\}^2 + \frac{1}{3!}\{\mathcal{S}(a)\}^3 + \cdots & (1) \\ &= \cos(\theta) I + \sin(\theta)\,\mathcal{S}(u) + \{1 - \cos(\theta)\} u u^\top. & (2) \end{aligned}$$

This is a rotation matrix of angle θ about unit vector u; it is denoted $R(\theta, u)$.

Representation (1) allows to linearize rotation matrices, thereby providing a parametrization for their infinitesimal neighbourhoods. Letting A_1 and A_2 be $SO(3)$ elements, rotation matrices close to A_1 and A_2 can be approximated by $A_1\{I + \mathcal{S}(\beta_1) + \mathcal{S}^2(\beta_1)/2\}$ and $A_2\{I + \mathcal{S}(\beta_2) + \mathcal{S}^2(\beta_2)/2\}$, where β_1 and β_2 are small $\mathbb{R}^3$ vectors. These two term expansions are used in Section 4. They arise when deriving the large sample distributions of estimators $\hat{A}_1$ and $\hat{A}_2$ of the true rotations A_1 and A_2. Typically one has $\hat{A}_1 = A_1\{I + \mathcal{S}(\hat{\beta}_1) + o_p(n^{-1/2})\}$, $\hat{A}_2 = A_2\{I + \mathcal{S}(\hat{\beta}_2) + o_p(n^{-1/2})\}$, where $(\hat{\beta}_1, \hat{\beta}_2)$ represents the first order contribution of the experimental errors to the estimators. Representation (2) also appears in the algorithm of Section 4.3 for calculating the least squares estimates for A_1 and A_2.

The following properties of $\mathcal{S}(\cdot)$ are used in the derivation of the second order expansion given in Proposition 3,

$$\begin{aligned} Q\mathcal{S}(a)Q^\top &= |Q|\mathcal{S}(Qa), \quad \text{for any orthogonal matrix } Q & (3) \\ \mathcal{S}(a)\mathcal{S}(b) &= ba^\top - a^\top \text{ for any } a, b \in \mathbb{R}^3, & (4) \end{aligned}$$

where $|Q|$ denotes the determinant of Q; it is equal to 1 if Q is a rotation and to -1 otherwise.

The above discussion highlights that $SO(3)$ is a three-dimensional submanifold of the Euclidean nine-dimensional space of 3×3 matrices. Representing rotations by the 9 entries of a 3×3 matrix is highly redundant. The *quaternions* are an alternative representation for $SO(3)$ rotations, convenient for statistical manipulations. The quaternion for $R(\theta, \mu)$ is the 4×1 unit vector defined by $q(\theta, \mu) = (\cos(\theta/2), \sin(\theta/2)\mu^\top)^\top$ (Hamilton 1969). This representation preserves dimensionality since the unit sphere in $\mathbb{R}^4$ is also three-dimensional. Observe that quaternions q and $-q$ represent the same rotation since $q(\theta, \mu) = -q(\theta + 2\pi, \mu)$. Statistical analysis for samples of rotations written as quaternions were introduced in Prentice (1986); see also Rancourt, Rivest & Asselin (2000).

The quaternion corresponding to the rotation product $R(\theta_1, \mu_1)R(\theta_2, \mu_2)$ can be calculated from the individual quaternions $q(\theta_1, \mu_1)$ and $q(\theta_2, \mu_2)$ as either $M_+(\theta_1, \mu_1)q(\theta_2, \mu_2)$ or $M_-(\theta_2, \mu_2)q(\theta_1, \mu_1)$, where M_+ and M_- are 4×4 rotation matrices defined by

$$M_+(\theta_1, \mu_1) = \begin{pmatrix} \cos(\theta_1/2) & -\sin(\theta_1/2)\,\mu_1^\top \\ \sin(\theta_1/2)\,\mu_1 & \cos(\theta_1/2)\,I + \sin(\theta_1/2)\,\mathcal{S}(\mu_1) \end{pmatrix} \qquad (5)$$

and

$$M_-(\theta_2, \mu_2) = \begin{pmatrix} \cos(\theta_2/2) & -\sin(\theta_2/2)\,\mu_2^\top \\ \sin(\theta_2/2)\,\mu_2 & \cos(\theta_2/2)\,I - \sin(\theta_2/2)\,\mathcal{S}(\mu_2) \end{pmatrix}. \qquad (6)$$

See McCarthy (1990, Chapter 5) for a presentation of these results which are used throughout the paper.

3. PRENTICE'S REGRESSION MODEL

The physical system associated with Prentice's model typically involves a global and a local reference frame. It is convenient to present this system in the context of the calibration experiment to be considered in Section 6.1. The global reference frame is defined in terms of the laboratory where the experiment takes place. The global x axes points up, the y and z axes go east and north respectively. The local reference frame is a characteristic of the marker whose orientation is recorded. Suppose that the tracking device consists of 4 coplanar diodes; then the local x and y-axes can be defined such that the $x-y$ plane coincides with the diodes plane. The local z-axis is then orthogonal to the diodes plane. The orientation of the marker is characterized by the 3×3 rotation matrix U whose columns give the directions of the local x, y, and z axes in the laboratory reference frame.

Suppose that a new laboratory system of axes whose orientation in the original system is given by 3×3 rotation matrix $A_1^\top$ is to be used. The three columns of $A_1^\top$ give the coordinates of the new x, y, and z axes in the old laboratory system of axes. If the orientation of a marker is U in the old reference frame, then it becomes $A_1 U$ in the new one. Suppose also that the tracking device reference frame is changed and that the rotation matrix $A_2^\top$ gives the orientation of the new frame in the old one. Matrix A_2 could, for instance, be associated to a relabeling of the axes in such a way that the new x-axis is orthogonal to the diodes plane. With these changes, the orientation of the rotation matrices is given by $A_1 U A_2^\top$. In Section 6.1, the rotations A_1 and A_2 are associated with distortions of the signal of the tracking device. The statistical problem is to estimate the rotation matrices A_1 and A_2 from a sample of pairs (U, V), where V is equal to $A_1 U A_2^\top$ perturbed by experimental errors. The specification of these errors is discussed next.

3.1. Modeling experimental errors.

Experimental errors can occur in the two reference frames involved in Prentice's model. Consider, as before, the experimental set-up underlying the example of Section 6.1 with the global and local z-axes respectively given by the northern direction and the perpendicular to the diodes plane. Suppose that the apparatus for changing the orientation of the marker is not fastened tightly to the floor of the laboratory and that small random rotations about the laboratory z-axis occur. This error occur in the global, or laboratory reference frame; it can be expressed as

$$E = \begin{pmatrix} \cos\epsilon & -\sin\epsilon & 0 \\ \sin\epsilon & \cos\epsilon & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{7}$$

where ϵ is the error's angle. In the presence of such an error, the recorded rotation V is either $A_1 E U A_2^\top$ or $E A_1 U A_2^\top$. In a similar way the binding between the arm of the apparatus for changing the marker's orientation and the marker can be loose and slight random rotations about the marker's z-axis can occur when changing its orientation. These errors would be associated with the local reference frame and enter the model as $A_1 U E A_2^\top$ or $A_1 U A_2^\top E$. They can be interpreted as errors in the explanatory rotation since their occurrence changes U into UE or $U A_2^\top E A_2$. Thus the proposed model can account for errors in both the explanatory and the dependent rotations. In practice the reference frame in which the errors occur is not known and not identifiable. For instance, writing $A_1 U E A_2^\top$ as $A_1 (U E U^\top) U A_2^\top$ one can interpret the local reference frame error E as a global reference frame error of $U E U^\top$. For $\Im$ to be an acceptable model for the error rotations, it must contain the densities for the four possible definitions.

Prentice (1989) and Chang & Rivest (2001) considered errors with symmetric densities belonging to the family $\Im_s = \{g(R) = f\{\mathrm{tr}(R)\}\}$, where $R \in SO(3)$ and $f(\cdot)$ is a function defined in $\mathbb{R}$. This is a situation where the errors do not have a preferred rotation axis. The four possible specifications of the errors all lead to the same density, namely $f\{\mathrm{tr}(A_2 U^\top A_1^\top V)\}$, for V under

$\Im_s$. An interesting special case is the symmetric density of Downs (1972), $g(R) = \exp\{\kappa \mathrm{tr}(R)\}/c_\kappa$, where κ is a positive shape parameter and c_κ is a normalizing constant.

As shown in Section 6.1, it may happen that the errors distribution cannot be modeled with $\Im_s$. A generalization of this class is given by

$$\Im_h = \{g(R) = f\{\mathrm{tr}(\wp\Psi\wp^\top R)\}/c_\Psi \mid \wp \in SO(3)\},$$

where Ψ is a 3×3 diagonal matrix with positive diagonal entries, $f(\cdot)$ is a function defined on $\mathbb{R}$ and c_Ψ is a normalization constant. When Ψ is proportional to the identity matrix, $\Im_h$ reduces to $\Im_s$. The special case $f(\cdot) = \exp(\cdot)$ corresponds to the matrix Fisher–von Mises distribution introduced by Downs (1972); see also Section 13.2.3 of Mardia & Jupp (2000) and Chikuse (2003). Observe that if error rotation $A_1^\top V A_2 U^\top$ has a density in $\Im_h$, with parameters $(\Psi, \wp)$, then the density of $U^\top A_1^\top V A_2$ is in $\Im_h$ with parameters $(\Psi, U^\top \wp)$. Thus, under $\Im_h$, the density of the errors may depend on the explanatory variable U; this suggests to let the error parameters Ψ and $\wp$ vary between sample units under $\Im_h$.

The differences between $\Im_s$ and $\Im_h$ are best seen by calculating moments of the rotation R and of its associated error vector $r = \mathcal{S}^{-1}(R - R^\top)/2$ appearing in the derivations of Section 4. From (2), r is equal to the rotation axis of R multiplied by the sine of its angle. Two rotation matrices R correspond to an error vector r, namely

$$R_1 = \sqrt{1 - r^\top r} I + \mathcal{S}(r) + \frac{1}{1 + \sqrt{1 - r^\top r}}\, rr^\top, \quad R_2 = -\sqrt{1 - r^\top r} I + \mathcal{S}(r) + \frac{1}{1 - \sqrt{1 - r^\top r}}\, rr^\top. \quad (8)$$

The rotation angle for R_1 belongs to $(0, \pi/2)$ while that for R_2 is in $(\pi/2, \pi)$. For $\Im_s$, the distribution of R depends only on its angle since, according to (2), $\mathrm{tr}(R)$ is independent of R's axis. Thus R's axis is uniformly distributed on the unit sphere in $\mathbb{R}^3$ independently of its angle. Therefore, $\mathrm{E}(R) = d_1 I$ and $\mathrm{E}(rr^\top) = c_1 I$ where c_1, d_1 are positive constants depending on $f(\cdot)$. Now if R's density is in $\Im_h$, with parameter $\wp$, the density of $S = \wp^\top R \wp$ is $f(\mathrm{tr}(\Psi S))$. Thus $\wp^\top R \wp$ and $D\wp^\top R \wp D$ have the same distribution, for any 3×3 diagonal matrix D of ± 1. This yields $\mathrm{E}(\wp^\top R \wp) = D_1$ and $\mathrm{E}(r_\wp r_\wp^\top) = C_1$ where $r_\wp$ is the residual vector calculated from $\wp^\top R \wp$ and D_1 and C_1 are 3×3 diagonal matrices depending on $f(\cdot)$ and Ψ. Applying (3) and combining these results lead to the following.

Proposition 1. *Let $r = \mathcal{S}^{-1}(R - R^\top)/2$ denote the error vector for a random rotation R. If the density of R belongs to $\Im_s$, then*

$$\mathrm{E}(R) = d_1 I \quad \textit{and} \quad \mathrm{E}(rr^\top) = c_1 I$$

where c_1, d_1 are positive constants depending on $f(\cdot)$. If the density of R belongs to $\Im_h$, with parameters $\wp$ and Ψ, then

$$\mathrm{E}(R) = \wp D_1 \wp^\top \quad \textit{and} \quad \mathrm{E}(rr^\top) = \wp C_1 \wp^\top,$$

where D_1 and C_1 are 3×3 diagonal matrices depending on $f(\cdot)$ and Ψ.

León, Massé & Rivest (2006) study a density in $\Im_s$,

$$f(R|\kappa) = \frac{\sqrt{\pi}\, \Gamma(\kappa + 2)}{2^{2\kappa}\Gamma(\kappa + 1/2)} \{1 + \mathrm{tr}(R)\}^\kappa, \quad R \in SO(3)$$

where $\Gamma(\cdot)$ is the gamma function and $\kappa > 0$ measure the concentration of the error. They prove that $d_1 = \kappa/(\kappa + 2)$ and $c_1 = (2\kappa + 1)/\{(\kappa + 2)(\kappa + 3)\}$ for this distribution. In addition, if θ is the rotation angle of R, then $(1 + \cos\theta)/2$ has a beta distribution with parameters $(\kappa + 1/2, 3/2)$. Thus it is straightforward to generate random rotation errors according to this distribution.

Under $\Im_s$, the error rotation axis is uniformly distributed on S^2, the unit sphere in $\mathbb{R}^3$. For distributions in $\Im_h$, the preferred direction of the rotation axis can be defined as the eigenvector corresponding to the largest eigenvalue of $\mathrm{E}(rr^\top)$. This largest eigenvalue is proportional to the the variance of the rotation angle about this preferred direction. For instance, error rotation E defined in (7) has error vector $r = (0, 0, \sin\epsilon)$ and the z-axis, $(0, 0, 1)^\top$, is the eigenvector for the largest eigenvalue of $\mathrm{E}(rr^\top)$.

For cosmetic reasons, the working model for V given U is written with laboratory reference frame errors as

$$V = A_{10} E U A_{20}^\top, \tag{9}$$

where A_{10} and A_{20} are the true values for the rotation matrices A_1 and A_2, E is a random rotation whose densities belong to either $\Im_s$ or $\Im_h$. If $g(R)$ is the density of E, then the density of V is $g(A_{10}^\top V A_{20} U^\top)$. In (9), E stands for the error in the laboratory reference frame while that in the object reference frame is $U^\top E U$. Furthermore if r is the error vector for E, that for an object reference frame error is $U^\top r$. This is used in Section 5.1 for investigating which reference frame contributes most to the error.

4. ESTIMATION OF THE UNKNOWN ROTATION MATRICES A_1 AND A_2

Let $\{(U_i, V_i);\ i = 1, \ldots, n\}$ be a bivariate sample of rotation matrices such that given U_i, V_i satisfies (9) for $i = 1, \ldots, n$. The error rotation E_i has a density in $\Im_h$, with the error parameters Ψ and $\wp$ depending on i for $i = 1, \ldots, n$. In principle the U_i's are error free; this is true in the example of Section 6.1, where they are the targeted orientations of the marker that are fed to the device for changing its orientation. In some other applications, such as that presented in Section 6.2, both U_i and V_i are the outcome of some measurement process and it is not possible to distinguish the explanatory from the dependent rotation. This can be accommodated within (9) since, as argued in Section 3.1, the distribution of V_i given U_i can be modeled by (9) even if U_i has measurement errors.

Prentice (1989) proposed least squares estimates for A_1 and A_2; these are defined as the values that minimize

$$\begin{aligned} SSR(A_1, A_2) &= \sum_{i=1}^{n} \mathrm{tr}\{(V_i - A_1 U_i A_2^\top)(V_i - A_1 U_i A_2^\top)^\top\} \qquad (10) \\ &= 6n - 2\sum_{i=1}^{n} \mathrm{tr}(A_1 U_i A_2^\top V_i^\top). \end{aligned}$$

The least squares estimates of A_1 and A_2 maximize $\sum \mathrm{tr}(A_1 U_i A_2^\top V_i^\top)/n$. Let $\hat\rho_P$ denote the maximum value of this sum; $\hat\rho_P$ is called Prentice's correlation. One has $0 < \hat\rho_P \leq 3$. The least squares estimators for A_1 and A_2 are the maximum likelihood estimators under Downs symmetric density discussed in Section 3.1. The maximum likelihood estimators corresponding to an arbitrary symmetric density in $\Im_s$ are considered in Chang & Rivest (2001).

4.1. The large sample distribution of $\hat A_1$ and $\hat A_2$.

Several aspects of the estimation of A_1 and A_2 are discussed in this section. The next proposition, whose proof appears in the appendix with that of Proposition 3, gives conditions on the matrices U_i and on the error distributions which insure that A_{10} and A_{20} are estimable.

PROPOSITION 2. *Let $\{(U_i, V_i); i = 1, \ldots, n\}$ be a bivariate sample of rotations such that the conditional distribution of V_i given U_i is given by (9) for $i = 1, \ldots, n$. The only rotation matrices A_1 and A_2 minimizing $\mathrm{E}\{SSR(A_1, A_2)\}$ are the true values A_{10} and A_{20} under the following hypotheses:*

1. $\mathrm{E}(E_i)$ *is a symmetric positive definite matrix possibly depending on i, for* $i = 1, \ldots, n$;

2. *The quaternions for the rotations* U_i, $i = 1, \ldots, n$ *do not lie on two orthogonal great circles of* S^3, *the unit sphere in* $\mathbb{R}^4$.

The minimal sample size for the second assumption of Proposition 2 to be met is $n = 3$. Hypothesis 2 for the uniqueness of the least squares estimator of A_1 and A_2 was first given in Shin (1999). Uniqueness and minimal sample sizes for spherical regression are considered in Shin, Takahara & Murdoch (2001). Theorem 1a of Prentice (1989) for the consistency $\hat{A}_1$ and $\hat{A}_2$ applies if the assumption of Proposition 2 holds for each finite n and if as n goes to infinity the number of U_i's not lying on two orthogonal great circles converge to infinity for any pair of orthogonal great circles.

The spectral decomposition of the quaternion cross-product matrix $\sum q_i q_i^\top/n$, where q_i is the quaternion for U_i is useful for determining whether the uniqueness condition holds. When it fails, the two great circles referred to in Proposition 2 are spanned by two disjoint pairs of eigenvectors of $\sum q_i q_i^\top/n$. Each data point has null loadings on the two unit vectors of one of these two pairs.

The asymptotic results presented in this section rely on a two term expansion of $SSR(A_1, A_2)$ given in the next proposition; this expansion is also used to investigate the algorithm proposed in Section 4.3.

Proposition 3. *Let* $\beta = (\beta_1^\top, \beta_2^\top)^\top$ *be a* 6×1 *vector, then for any rotations* A_1 *and* A_2,

$$SSR[A_1 \exp\{\mathcal{S}(\beta_1)\}, A_2 \exp\{\mathcal{S}(\beta_2)\}] = \sum_{i=1}^n \mathrm{tr}(S_i^2) + 2r^\top r$$
$$-4r^\top X\beta + \beta^\top(Z + 2X^\top X)\beta + 2\sum_{i=1}^n \beta_2^\top U_i^\top \mathcal{S}(r_i)\beta_1 + o(\beta^\top\beta)$$

where

1. r *is a* $3n \times 1$ *error vector whose three components for the ith data point are*
$$r_i = \mathcal{S}^{-1}\left(\frac{A_1^\top V_i A_2 U_i^\top - U_i A_2^\top V_i^\top A_1}{2}\right); \tag{11}$$

2. $S_i = \dfrac{A_1^\top V_i A_2 U_i^\top + U_i A_2^\top V_i^\top A_1}{2} - I$;

3. X *is a* $3n \times 6$ *design matrix whose 3 rows for data point i are* $X_i = [I; -U_i]$;

4. $Z = Z_1 + \cdots + Z_n$, *where* Z_i *is a* 6×6 *matrix given by* $Z_i = X_i^\top\{\mathrm{tr}(S_i)I - S_i\}X_i$.

Observe that when $\beta_1 = \beta_2 = 0$, $SSR(A_1, A_2) = \sum_i \mathrm{tr}(S_i^2) + 2r^\top r$. Note that the design matrix X satisfies,

$$X^\top X = \begin{pmatrix} nI & -\sum_i U_i \\ -\sum_i U_i^\top & nI \end{pmatrix}.$$

This matrix is singular when one can find vectors u_1 and u_2 such that $u_1 = \sum U_i u_2/n$ and $u_2 = \sum U_i^\top u_1/n$. A close look at the proof of Proposition 2 shows that this is weaker than the uniqueness condition of Proposition 2. Thus, $X^\top X$ is non singular when the assumptions of Proposition 2 are met.

The least squares estimators for A_1 and A_2 can be expressed in terms of the true values A_{10} and A_{20} as $\hat{A}_1 = A_{10}\exp\{\mathcal{S}(\hat{\beta}_1)\}$ and $\hat{A}_2 = A_{20}\exp\{\mathcal{S}(\hat{\beta}_2)\}$, where $\hat{\beta} = (\hat{\beta}_1^\top, \hat{\beta}_2^\top)^\top$ is such that $SSR(\hat{A}_1, \hat{A}_2)$ is minimal. Proposition 3 provides the first two terms of the Taylor series expansion of $SSR(\hat{A}_1, \hat{A}_2)$ about A_{10} and A_{20} needed to characterize the asymptotic distribution of $\hat{A}_1$ and $\hat{A}_2$; see Brown (1985).

Proposition 4. *Suppose that as $n \to \infty$, the following assumptions are met:*

1. $\hat{A}_1 = A_{10} \exp\{\mathcal{S}(\hat{\beta}_1)\}$ *and* $\hat{A}_2 = A_{20} \exp\{\mathcal{S}(\hat{\beta}_2)\}$ *are consistent estimators of* A_{10} *and* A_{20}*;*
2. *The assumptions of Proposition 2 are met for any finite n;*
3. $\{X^\top X + \mathrm{E}(Z)/2\}/n$ *converges to a* 6×6 *positive definite matrix.*

Then the asymptotic distribution of $\hat{\beta}$ is a six-variate normal distribution with 0 *mean vector and variance covariance matrix*

$$\mathrm{var}(\hat{\beta}) = \{X^\top X + \mathrm{E}(Z)/2\}^{-1} \sum_i \mathrm{E}(X_i^\top r_i r_i^\top X_i)\{X^\top X + \mathrm{E}(Z)/2\}^{-1},$$

where r_i, X_i, and Z are as defined in Proposition 3, with $A_1 = A_{10}$ and $A_2 = A_{20}$.

The variance-covariance matrix of Proposition 4 can be estimated using the observed residuals, $\hat{E}_i = \hat{A}_1^\top V_i \hat{A}_2 U_i^\top$, together with $\hat{r}_i = \mathcal{S}^{-1}\{(\hat{E}_i - \hat{E}_i^\top)/2\}$, and $\hat{S}_i = (\hat{E}_i + \hat{E}_i^\top)/2 - I$. Tests on A_1 and A_2 are easily carried out. Consider, for instance, testing the hypothesis $H_0 : A_2 = I$. Let $\hat{\beta}_2$ be the 3×1 vector such that $\hat{A}_2 = \exp\{\mathcal{S}(\hat{\beta}_2)\}$. The statistic $\hat{\beta}_2^\top v(\hat{\beta}_2)^{-1}\hat{\beta}_2$ whose asymptotic null distribution is χ^2_3 is appropriate for H_0, where $v(\hat{\beta}_2)$ is an estimate of the sub-matrix of $\mathrm{var}(\hat{\beta})$ pertaining to $\hat{A}_2$.

Proposition 1 can be used to evaluate $\mathrm{var}(\hat{\beta})$. Under symmetric error model $\Im_s$, $\mathrm{E}(Z_i) = 2(d_1 - 1)X_i^\top X_i$, $\mathrm{E}(r_i r_i^\top) = c_1 I$ and

$$\mathrm{var}(\hat{\beta}) \;=\; \frac{c_1}{d_1^2}\,(X^\top X)^{-1} = \frac{c_1}{d_1^2}\begin{pmatrix} nI & -\sum_i U_i \\ -\sum_i U_i^\top & nI \end{pmatrix}^{-1}.$$

This result was first derived in Proposition 6 of Chang & Rivest (2001); observe that Chang and Rivest's value for c_1 is equal to $2c_1$ as defined in Proposition 1. Note that $-\sum_i U_i$ represents the dependency between $\hat{A}_1$ and $\hat{A}_2$. If the U_i's are uniformly scattered in $SO(3)$, $\sum U_i \approx 0$ and the two estimators are approximately independent. On the other hand, there is a strong dependency between the two when the U_i's are clustered together.

When $\Im_s$ provides a good description of the experimental errors, the variance-covariance matrix of $\hat{\beta}$ can be estimated by taking

$$\hat{d}_1 = \frac{6n - SSR(\hat{A}_1, \hat{A}_2)}{6n} = \frac{\hat{\rho}_P}{3} \quad \text{and} \quad \hat{c}_1 = \sum_{i=1}^{n} \frac{\hat{r}_i^\top \hat{r}_i}{3n}\,.$$

The residuals $\hat{r}_i$ can help to characterize the distribution of the errors. Suppose, for instance, that E_i defined in (9) has for $i = 1, \ldots, n$ a distribution in $\Im_h$ with fixed $\wp$ and Γ. Thus, all the errors occur in the laboratory reference frame. An estimate of $\mathrm{E}(rr^\top)$ is given by $\sum \hat{r}_i \hat{r}_i^\top/n$; and the largest eigenvalue of this matrix estimates λ_1, the largest diagonal entry of C_1 defined in Proposition 1. Residuals in the object reference frame are $U_i^\top \hat{r}_i$ and their cross product matrix, $\sum \hat{U}_i^\top \hat{r}_i \hat{r}_i^\top U_i/n$, estimates $\sum U_i^\top \wp C_1 \wp^\top U_i/n$. The largest eigenvalue of this matrix is less than λ_1. This gives a basis to determine which reference frame, the laboratory or the object, contributes most to the errors. One simply picks the one for which the largest eigenvalue of the residual cross product matrix is maximal. This is correct since these two cross product matrices have the same trace. Model $\Im_s$ is appropriate when the eigenvalues of these two matrices are approximately equal. Formal test statistics for comparing them are given in Prentice (1986).

4.2. Large concentration distributional results.

This section assumes that the sample size n is fixed and that the errors in (9) are small. This means that the error rotations in (9) have most of their probability mass close to the identity matrix. Small errors occur in many applications of directional statistics and inference for large concentration is often more relevant than for large sample; see for instance Rivest (1999) and Chang, Ko, Royer & Lu (2000). Large concentration asymptotics highlight a local linear model underlying the estimation of the parameters and allows one to use standard linear model techniques to carry out statistical inference.

This section assumes that the error rotation matrices follow a distribution with concentration indexed by κ and that as $\kappa \to \infty$,

$$E_i = I + \mathcal{S}(r_i) + O_p(\kappa^{-1}), \quad i = 1, \dots, n. \tag{12}$$

Under the symmetric error density of Downs (1972), $f(R) = \exp\{\kappa \mathrm{tr}(R)\}/c_\kappa$, or under León, Massé & Rivest (2006) distribution with parameter $\kappa/2$, (12) holds and the limiting distribution of $\kappa^{1/2} r_i$ is a $\mathcal{N}_3(0, I_3)$, a standardized trivariate normal distribution.

Approximate sampling distributions for $\hat{A}_1$ and $\hat{A}_2$ are obtained from the expansion of Proposition 3 by considering that both, the residual vectors r_i and the components of β, are small or $O_p(\kappa^{-1/2})$. In a second order expansion, one only keeps the terms that are $O_p(\kappa^{-1})$. For instance since $S_i = (E_i + E_i^\top)/2 - I$ is $O_p(\kappa^{-1})$, S_i^2 and $\beta^\top Z_i \beta$ are negligible or $o_p(\kappa^{-1})$ in this setting. In a large concentration situation, Proposition 3 gives

$$SSR[A_{10}\exp\{\mathcal{S}(\beta_1)\}, A_{20}\exp\{\mathcal{S}(\beta_2)\}] = 2(r - X\beta)^\top (r - X\beta) + o_p(\kappa^{-1}),$$

as an expansion of $SSR(\hat{A}_1, \hat{A}_2)$ around the true values A_{10} and A_{20}. Fitting Prentice's model is therefore locally equivalent to fitting a linear model with design matrix X. This expansion is not tied to the normality assumption; it holds for any distribution of the error vectors r_i that has his probability mass concentrated around 0. The next proposition assumes that the limiting distributions of the error vectors are normal and applies the standard linear model theory to Prentice's regression model.

Proposition 5. *Let $\hat{\beta} = (\hat{\beta}_1^\top, \hat{\beta}_2^\top)^\top$ be a 6×1 vector such that $\hat{A}_1 = A_{10}\exp\{\mathcal{S}(\hat{\beta}_1)\}$ and $\hat{A}_2 = A_{20}\exp\{\mathcal{S}(\hat{\beta}_2)\}$ and r be the $3n \times 1$ vector of the errors as defined in (11) with $A_1 = A_{10}$ and $A_2 = A_{20}$. If the design matrix X satisfies the assumption of Proposition 2, and if, as $\kappa \to \infty$, $\kappa^{1/2} r$ converges in distribution to a $\mathcal{N}_{3n}(0, I_{3n})$, then the following results hold:*

1. *The limiting distribution of $\kappa^{1/2}\hat{\beta}$ is a $\mathcal{N}_6[0, (X^\top X)^{-1}]$;*
2. *The limiting distribution of $\kappa SSR(\hat{A}_1, \hat{A}_2)/2$ is chi-square with $3n - 6$ degrees of freedom;*
3. *$\kappa^{1/2}\hat{\beta}$ and $\kappa SSR(\hat{A}_1, \hat{A}_2)$ are asymptotically independent.*

Consider now testing a null hypothesis H_0 specifying that (A_1, A_2) belongs to some submanifold of $SO(3) \times SO(3)$ of dimension j. Such an hypothesis might specify that A_1 or A_2 is fixed, or that one of their axes is fixed. If $SSR(\hat{A}_{10}, \hat{A}_{20})$ is the minimal value of (10) under H_0, one can reject H_0 if

$$F_{obs} = \frac{\{SSR(\hat{A}_{10}, \hat{A}_{20}) - SSR(\hat{A}_1, \hat{A}_2)\}/(6 - j)}{SSR(\hat{A}_1, \hat{A}_2)/(3n - 6)}$$

is larger than the appropriate critical value for an F distribution with $6 - j$ and $3n - 6$ degrees of freedom. This test is valid as long as the hypothesis of Proposition 5 are met; see Rivest (1989) for a discussion of similar tests in the context of the spherical regression model. Simulation studies are easily carried out to investigate the accuracy of the large concentration asymptotics. Using $\hat{\kappa} = (6n-12)/SSR(\hat{A}_1, \hat{A}_2)$ as an estimate of κ, one can simulate data from a model with $\hat{A}_1$ and $\hat{A}_2$ as the true value of the two rotations and with errors distributed according to León, Massé & Rivest (2006) distribution with parameter $\hat{\kappa}/2$.

Large concentration distributional results can be derived when the errors do not have a symmetric distribution. Suppose that E_i follows the distribution of Downs (1972), so that for each $i = 1, \dots, n$, $f_i(R) = \exp\{\mathrm{tr}(\wp_i \Psi_i \wp_i^\top R)\}/c_{\Psi i}$. When the entries of Ψ_i are large, (12) holds with r_i approximately distributed as a $\mathcal{N}_3(0, \wp_i \Psi_i^{-1} \wp_i^\top)$. In this situation, $SSR(A_1, A_2)$ can still be approximated by a linear model similar to that underlying Proposition 5. However this linear model has heteroscedastic, possibly dependent, errors and the associated F-statistics do not have a simple distribution under H_0.

4.3. A scoring algorithm for minimizing $SSR(A_1, A_2)$.

This section investigates numerical methods for minimizing (10). Unfortunately there is no simple way to calculate the least squares estimates for A_1 and A_2. Prentice (1989) suggests an algorithm that repeatedly optimizes (10) for one rotation while keeping the other one fixed. An alternative algorithm that optimizes simultaneously for both rotations is presented here.

A Newton–Raphson algorithm can be constructed using the second order expansion given in Proposition 3 since it features a closed formed expression for the matrix of second order partial derivatives. In this matrix, the dominant term is usually $X^\top X$. This suggests considering the simpler scoring algorithm given by.

1. Calculate X, the $3n \times 6$ design matrix defined in Proposition 3 and $P = (X^\top X)^{-1} X^\top$.
2. Take $A_1 = A_2 = I$ as starting values for the algorithm.
3. Calculate the $3n \times 1$ residual vector r using (11) for components $3i-2$ to $3i$.
4. Let $b = (b_1^\top, b_2^\top)^\top = Pr$ be the coefficients of the regression of r on X and use $A_1 \exp\{\mathcal{S}(b_1)\}$ and $A_2 \exp\{\mathcal{S}(b_2)\}$ as updated estimates.
5. Test for convergence and go to 3, if it fails.

This algorithm is best carried out using the quaternion representation for rotations. Multiplication rules (5) and (6), together with the fact that the quaternion for $R(\theta, \mu)^\top$ is $q(-\theta, \mu)$, allow an easy evaluation for the quaternion s_i of $A_1^\top V_i A_2 U_i^\top$. As in Rancourt Rivest & Asselin (2000), residual (11) can be evaluated as $r_i = 2 s_{i1} (s_{i2}, s_{i3} s_{i4})^\top$.

When the current values of A_1 and A_2 are such that all residual vectors in Proposition 3 have angles smaller that $\pi/2$, all the residual rotations have the form R_1 in (8). This may happen in applications where (9) is true and where the true values of A_1 and A_2 are not too far from the identity matrix. In this situation the matrix of the second order terms in the expansion of Proposition 3 is likely to be positive definite, meaning that $SSR(A_1, A_2)$ should be convex. The algorithm then converges smoothly to a minimum. If the opposite is true and if all the residual rotations have the form R_2 in (8), with small residual vectors, then the matrix of the second order terms in the expansion of $SSR(A_1, A_2)$ can be shown to be negative definite. The performance of the scoring algorithm is questionable in this situation. One iteration may fail to decrease $SSR(A_1, A_2)$. Thus the scoring algorithm cannot be proved to converge to the minimum value of $SSR(A_1, A_2)$; its performance depends on the starting value. Still it should do better than Prentice's when $\sum U_i$ is far from 0, i.e., when the U_i's are not uniformly scattered over $SO(3)$ since it accounts for the correlation between $\hat{A}_1$ and $\hat{A}_2$. When numerical difficulties occur, the eigenvalues of the 6×6 matrix of second order partial derivatives at the current values for A_1 and A_2 as given in Proposition 3 are useful diagnostic tools.

4.4. Properties of $\hat{A}_1$ and $\hat{A}_2$ when (9) does not hold.

Suppose that $\{U_i\}$ and $\{V_i\}$ are independent samples from densities in $\Im_s$. Then, $\mathrm{E}\{SSR(A_1, A_2)\}$ is proportional to $\mathrm{tr}(A_1 A_2^\top)$ and any pair of rotation matrices (A_1, A_2) with $A_1 = A_2$ maximizes $\mathrm{E}\{SSR(A_1, A_2)\}$; A_1 and A_2 are not estimable. $\mathrm{E}(\hat{\rho}_P)$ is then related to the clustering of the U and the V samples; it cannot be used to assess whether the samples are independent. The expectation of the 6×6 matrix of the second term in the expansion presented in Proposition 3 can be shown to have rank 3 when $A_1 = A_2$. The function $SSR(A_1, A_2)$ has a plateau and the algorithm of Section 4.3 converges very slowly when the samples are independent.

If $\{(U_i, V_i);\ i = 1, \ldots, n\}$ follows some model, not necessarily (9), such that $\mathrm{E}\{SSR(A_1, A_2)\}$ has a unique maximum for n large enough, then $\hat{A}_1$ and $\hat{A}_2$ converge in probability to this maximum as $n \to \infty$ under suitable regularity conditions. Furthermore $\hat{A}_1$ and $\hat{A}_2$ have an asymptotically normal distribution with a sandwich variance covariance matrix similar to that given in Proposition 4. Note however that it can differ from $\mathrm{var}(\hat{\beta})$ given in Proposition 4 since the expression $\sum U_i^\top \mathcal{S}(r_i)$ appearing in the second order term of the expansion of Proposition 3 may not have a null expectation, for an arbitrary model. This suggests considering the following variance estimator, robust to model misspecifications,

$$v(\hat{\beta}) = \hat{F}^{-1} \left(\sum_{i=1}^{n} X_i^\top \hat{r}_i \hat{r}_i^\top X_i \right) \hat{F}^{-1},$$

where

$$\hat{F} = \frac{1}{2} \sum_{i=1}^{n} \begin{pmatrix} \mathrm{tr}(\hat{E}_i)I - (\hat{E}_i + \hat{E}_i^\top)/2 & \hat{E}_i^\top U_i - \mathrm{tr}(\hat{E}_i)U_i \\ U_i^\top \hat{E}_i - U_i^\top \mathrm{tr}(\hat{E}_i) & \mathrm{tr}(\hat{E}_i)I - (\hat{E}_i + \hat{E}_i^\top)/2 \end{pmatrix}$$

is proportional to the second order partial derivatives in the expansion of Proposition 3.

5. A TEST OF CORRELATION FOR CONCENTRATED SAMPLES

Prentice's regression model does not give any simple method to test whether the U-sample and the V-sample are independent since the independence model is not nested within Prentice's. One approach, adopted by Epp, Tukey & Watson (1971), is to carry out a permutation test. This involves repeatedly evaluating $SSR(\hat{A}_1, \hat{A}_2)$ for all possible permutations of the U-sample without changing the V-sample. The permutation P-value is then equal to the proportion of permuted samples for which $SSR(\hat{A}_1, \hat{A}_2)$ is smaller than the actual value of this statistic. Alternative independence tests for directions are presented in Jupp & Spurr (1985); permutation tests for directional data are discussed by Welner (1979).

The permutation test of independence is numerically demanding since the minimal value of SSR is obtained, for each permuted sample, via an iterative algorithm. This section derives an approximation to SSR, valid when the U and the V-samples are clustered around their respective modal values, whose calculation is simpler.

Let $\bar{U}$ and $\bar{V}$ be the mean rotation of the samples. They are defined as the rotations that are closest to the averages $\sum U_i/n$ and $\sum V_i/n$ in the least squares sense; see Rancourt, Rivest & Asselin (2000) for details. In concentrated samples, $\bar{U}^\top U_1$ is close to the identity matrix and the 3×1 vector $e_{u1} = \mathcal{S}^{-1}(\bar{U}^\top U_1 - U_1^\top \bar{U})/2$ has small $O_p(\kappa^{-1/2})$ components; thus $U_1 = \bar{U}\{I + \mathcal{S}(e_{u1})\} + o_p(\kappa^{-1/2})$. In a similar way, one can define residual vectors e_{ui} and e_{vi} such that

$$U_i = \bar{U}\{I + \mathcal{S}(e_{ui})\} + o_p(\kappa^{-1/2}) \quad \text{and} \quad V_i = \bar{V}\{I + \mathcal{S}(e_{vi})\} + o_p(\kappa^{-1/2}).$$

With these approximations, using (3), one has

$$\begin{aligned} V_i - A_1 U_i A_2^\top &= \bar{V}[I + \mathcal{S}(e_{vi}) - \bar{V}^\top A_1 \bar{U} A_2^\top \{I + \mathcal{S}(A_2 e_{ui})\}] + o_p(\kappa^{-1/2}) \\ &= \bar{V}[(I - \bar{V}^\top A_1 \bar{U} A_2^\top)\{I + \mathcal{S}(A_2 e_{ui})\} + \mathcal{S}(e_{vi} - A_2 e_{ui})] + o_p(\kappa^{-1/2}). \end{aligned}$$

The minimum for SSR is $O_p(\kappa^{-1})$; thus $I - \bar{V}^\top \hat{A}_1 \bar{U} \hat{A}_2^\top$ needs to be $O_p(\kappa^{-1/2})$. Since $\sum e_{vi} = \sum e_{ui} = 0$ (see Rancourt, Rivest & Asselin 2000), when A_1 and A_2 are in $O(\kappa^{-1/2})$ neighbourhoods of the least squares estimates one has

$$\begin{aligned} SSR = n\mathrm{tr}\{(I - \bar{V}^\top A_1 \bar{U} A_2^\top)(I - \bar{V}^\top A_1 \bar{U} A_2^\top)^\top\} \\ + \sum \mathrm{tr}\{\mathcal{S}(e_{vi} - A_2 e_{ui})\mathcal{S}(e_{vi} - A_2 e_{ui})^\top\} + o_p(\kappa^{-1}). \end{aligned}$$

Thus approximate least squares estimates are obtained by taking $\hat{A}_1 = \bar{V}\hat{A}_2\bar{U}^\top$ and $\hat{A}_2$ equal to the rotation matrix minimizing the second term of the above expression. This second term can be reexpressed as

$$2\sum(e_{vi}^\top e_{vi} + e_{ui}^\top e_{ui}) - 4\sum e_{vi}^\top A_2 e_{ui},$$

and the optimal $\hat{A}_2$ is obtained by Procrustes analysis (Dryden & Mardia 1998, Chapter 5). If $\sum e_{ui} e_{vi}^\top / n = P\ \mathrm{diag}(\gamma_1, \gamma_2, \gamma_3) Q^\top$ is a singular value decomposition, where P and Q are rotation matrices and where the singular values satisfy $\gamma_1 > \gamma_2 > |\gamma_3|$, then $\hat{A}_2 = QP^\top$. A simple Procrustes statistic to carry out a permutation test is given by $\gamma_1 + \gamma_2 + \gamma_3$.

6. DATA ANALYSES

6.1. The calibration of angular data from a magnetic tracking device.

Camera systems that records the time varying positions and orientations of markers attached on the body of an experimental subject are widely used to study human kinematics. Some systems use a magnetic tracking device to record the data. Unfortunately, when collected far from its source, the signal is distorted and the recorded orientations and positions have systematic errors. Day, Murdoch & Dumas (2000) discuss methods to calibrate such systems. This section uses Prentice's regression model to calibrate the orientations.

The (x, y, z) position of the tracking device in the laboratory system of axis is fixed; only its orientation varies. The data set consists of rotation matrices U_i giving the true orientations of the tracking device and of the orientations V_i reported by the camera system, for $i = 1, \ldots, n$. The U_i's are obtained by applying predetermined rotations to the tracking device. The discussion in Section 3.1 shows that parameters A_1 and A_2 are associated with deviations in the two reference frames, respectively that of the laboratory and that of the tracking device, involved in this experiment.

The calibration equation is $\hat{U}_i = \hat{A}_1^\top V_i \hat{A}_2$ where $\hat{A}_1$ and $\hat{A}_2$ are least squares estimates as defined in Section 3. Null hypotheses $H_0 : A_1 = I$ and $H_0 : A_2 = I$ are of special interest since they allow a reduction, from $n = 3$ to $n = 1$ of the minimal sample size required to estimate the calibration equation at a fixed location. If one, say $H_0 : A_2 = I$, is true, then $\hat{A}_1$ is the rotation maximizing $\mathrm{tr}\sum(A_1 U_i V_i^\top)/n$. This is the mean rotation of the $\{V_i U_i^\top\}$ sample; see Rancourt, Rivest & Asselin (2000). It can be estimated using only one pair (U_i, V_i).

Table 1 contains the quaternions for a sample of size $n = 12$ collected using the Fastrak machine in the Rehabilitation Centre at Université Laval. The data points have been collected at about two meters from the signal's source. One has $\hat{\rho}_P = 2.99231$. Since the residual $\hat{r}_2$, for the second data point, has large entries, consider the analysis for the sample of size $n = 11$, excluding this outlier. One has $\hat{\rho}_P = 2.99898$; $\hat{A}_1$ and $\hat{A}_2$ are rotations of respectively 40.29 degrees and 1.58

V-sample (Recorded Orientations)				U-sample (True Orientations)			
0.9483	0.0417	−0.1478	−0.2777	1	0	0	0
0.5766	−0.0807	−0.7985	−0.1531	0.7071	0	−0.7071	0
−0.2523	0.5396	0.3825	0.7063	0.5	−0.5	−0.5	−0.5
0.5353	0.2617	−0.7139	0.3679	0.5	0.5	−0.5	0.5
0.6853	−0.6413	0.1118	−0.3265	0.7071	−0.7071	0	0
0.5758	−0.672	−0.4007	0.2372	0.5	−0.5	−0.5	0.5
0.8858	−0.0733	−0.1105	0.4448	0.7071	0	0	0.7071
0.0744	0.8939	0.4312	0.0973	0	−0.7071	−0.7071	0
0.0295	−0.9385	0.3157	−0.137	0	−1	0	0
0.257	−0.7515	0.2075	0.5711	0	−0.7071	0	0.7071
0.7131	−0.3518	0.5537	0.2471	0.5	−0.5	0.5	0.5
0.3516	0.703	0.2551	−0.5631	0.5	0.5	0.5	−0.5

Table 1: The 12 pairs of quaternions for the Calibration of a Fastrak Camera System.

degrees about axis $(0.07, -0.40, -0.91)^\top$ and $(-0.90, 0.44, -0.04)^\top$. The χ^2_3 statistic for testing $H_0 : A_2 = I$, constructed with the sandwich variance estimator of Proposition 4, is 47.4 (P-value $= 0$). This suggests to reject H_0 despite A_2's small rotation angle. The maximum value of ρ_P calculated with the restriction that $A_2 = I$ is 2.99833. The $F_{3\,27}$-statistic for H_0 introduced in Section 3.3 is 5.81 (P-value $= 0.004$).

The largest eigenvalues of the cross product matrices of the residuals calculated in the laboratory and in the tracking device reference frames are useful for characterizing error distribution. These two eigenvalues, respectively equal to 6.42×10^{-4}, and 7.87×10^{-4}, represent 64% and 77% of the total residual variation. They suggest that most of the errors occur, according to model $\Im_h$, in the tracking device reference frame. The preferred rotation axis for the errors in this frame, as given by the eigenvector corresponding to the largest eigenvalue of $\sum \hat{U}_i^\top \hat{r}_i \hat{r}_i^\top U_i/11$, is close to $(0, 0, 1)^\top$. The rotation angle standard deviation for error rotation about this axis is 1.61 degrees. This characterization of the error distribution is useful for planning future studies.

The calibration experiment reported in Table 1 suggests that including rotations A_2 in Prentice's model improves slightly the calibration equation. A possible explanation for A_2's significant contribution is a systematic deviation, of A_2, between the recorded and the real orientation of the device. As an alternative to using Prentice's model for the calibration, one could attempt to remove the systematic deviation in the device's recordings. This would make a simple calibration equation, involving A_1 only, appropriate.

6.2. Investigating variations in drilling poses.

Rancourt, Rivest & Asselin (2000) studied the impact of the location of the drilling hole on drill pose, as characterized by the orientations of the three joints, the wrist, the elbow and the shoulder, of the right arm. There were eight subjects in this experiment, drilling at six different locations. This data set can be obtained from (`http://www.blackwellpublishers.co.uk/rss/`). Rotations giving the orientations of the three joints of a subject's drilling arm for five replicates at six drilling locations are the data points for analysis. The orientation for one joint, say the elbow, is calculated using the rotations R_f and R_a giving the orientation of the forearm and of the arm in the laboratory system of axis. The elbow orientation is then defined as $V = R_a^\top R_f$. Rotating by respectively $A_1^\top$ and $A_2^\top$ the arm's and the forearm's markers would yield $A_1 V A_2^\top$ as the elbow orientation. For a fixed subject drilling at a given location, the five orientations $\{V_i;\ i = 1, \ldots, 5\}$ obtained at one joint are perfectly correlated to the five orientations $\{U_i;\ i = 1, \ldots, 5\}$ at another joint if one can recover the V-sample from the U-sample by changing the orientations of the markers, that is if $V_i = A_1 U_i A_2^\top$ for $i = 1, \ldots, 5$. Thus Prentice correlation is a natural tool to investigate an association between limb orientations. This section investigates the correlation between the three joints for the $6 \times 8 = 48$ subject$\times$location treatments.

The randomization test described in Section 5 was performed repeatedly on bivariate samples of rotations of size 5. A wrist-elbow relationship was tested on 40 samples (8 of the possible 48 bivariate samples had missing values); a significant correlation, at the 5% level, was observed in 5 of the 40 tests. For the wrist-shoulder and elbow-shoulder correlations, 28 complete bivariate samples were available. Significant correlations were found in respectively 2 and 9 of them. The strongest relationship is between the elbow and the shoulder with 9 significant tests out of 28. This suggests that the elbow and the shoulder have compensatory roles when taking a drilling pose.

The two algorithms of Section 3.4 were tried on the data sets for the two examples of this section. For fitting Prentice's model to the data in Table 1, Prentice's algorithm did better than the new scoring algorithm requiring only one iteration, instead of 4. The opposite is true for the $n = 5$ samples for limb orientations. In many instances, Prentice's algorithm failed to converge after 50 iterations while the new scoring algorithm always converged in less than 25.

ACKNOWLEDGEMENTS

We are grateful to the Associate Editor and to the referees for their constructive comments, to Mireille Guay for programming the two algorithms for fitting Prentice's model, to François Comeau for his assistance with data collection, to the Natural Sciences and Engineering Research Council of Canada and to the Fonds québécois de la recherche sur la nature et les technologies for their financial assistance, and to Duncan J. Murdoch and Hwa Shin Shin for pointing out an error in an earlier version of Proposition 2.

APPENDIX PROOF OF PROPOSITIONS 2 AND 3

A.1. Proof of Proposition 2.

Let $\mathrm{E}(E_i) = P_i \mathrm{diag}(\lambda_i) P_i^\top$ denote the eigenvalue decomposition of $\mathrm{E}(E_i)$, where λ_i is the 3×1 vector of the eigenvalues of $\mathrm{E}(E_i)$, which are all positive. According to (9),

$$\mathrm{E}(V_i) = A_{10} P_i \mathrm{diag}(\lambda_i) P_i^\top U_i A_{20}^\top,$$

where A_{10} and A_{20} stand for the true values of A_1 and A_2. Let $R_1 = A_{10}^\top A_1$ and $R_2 = A_{20}^\top A_2$. Using this notation, one has

$$\begin{aligned} \mathrm{E}\left\{\sum_i \mathrm{tr}(A_1 U_i A_2^\top V_i^\top)\right\} &= \sum_i \mathrm{tr}\{A_1 U_i A_2^\top A_{20} U_i^\top P_i \mathrm{diag}(\lambda_i) P_i^\top A_{10}^\top\} \\ &= \sum_i \mathrm{tr}\{P_i^\top R_1 U_i R_2^\top U_i^\top P_i \mathrm{diag}(\lambda_i)\}. \end{aligned}$$

To prove Proposition 2 we need to show that $R_1 = R_2 = I$ uniquely maximize this expression. The maximum value of the ith term in this sum is $\mathrm{tr}\{\mathrm{diag}(\lambda_i)\}$; this maximum is obtained when

$$R_1 U_i R_2^\top U_i^\top = I.$$

Let q_1 and q_2 be the quaternions associated with R_1 and R_2 and let $M_+(U_i)$ and $M_-(U_i^\top)$ be 4×4 rotations (5) and (6) associated to the quaternions for U_i and $U_i^\top$. Written in terms of quaternions, the condition $R_1 U_i R_2^\top U_i^\top = I$ becomes $\{q_1^\top M_+(U_i) M_-(U_i^\top) q_2\}^2 = 1$, since quaternions $(1, 0, 0, 0)$ and $(-1, 0, 0, 0)$ correspond to the identity matrix. Elementary calculations show that

$$M_+(U_i) M_-(U_i^\top) = \begin{pmatrix} 1 & 0 \\ 0 & U_i \end{pmatrix}.$$

Thus, the solution to $\mathrm{E}\{SSR(A_1, A_2)\} = 0$ is uniquely defined provided that the only solution to

$$\left\{q_1^\top \begin{pmatrix} 1 & 0 \\ 0 & U_i \end{pmatrix} q_2\right\}^2 = 1 \quad \text{for } i = 1, \ldots, n \tag{13}$$

is $q_1 = q_2 = \pm(1, 0, 0, 0)^\top$.

Write, for $j = 1, 2$, $q_j = (\cos(\psi_j/2), \sin(\psi_j/2)u_j^\top)^\top$ where ψ_j and u_j are respectively the angle and the axis of the rotation corresponding to q_j. Condition (13) becomes

$$\{\cos(\psi_1/2)\cos(\psi_2/2) + \sin(\psi_1/2)\sin(\psi_2/2)u_1^\top U_i u_2\}^2 = 1 \quad \text{for } i = 1, \ldots, n.$$

If one cannot find unit vectors u_1 and u_2 such that $(u_1^\top U_i u_2)^2 = 1$, $i = 1, \ldots, n$ the only possible solution is $\psi_1 = \psi_2 = 0$. Let $M_{12} = (u_1 + u_2)(u_1 + u_2)^\top/(1 + u_1^\top u_2) - I$ be a rotation mapping u_1 into u_2; note that M_{12} is a rotation of angle π around axis $(u_1+u_2)/(2+2u_1^\top u_2)^{1/2}$. The uniqueness condition is not met if $(u_2^\top M_{12} U_i u_2)^2 = 1$ for some unit vector u_2. The case $u_2^\top M_{12} U_i u_2 = 1$ is only possible when the axis of the rotation $M_{12} U_i$ is u_2; the quaternion for $M_{12} U_i$ is then

$$q_i^m = \cos(\theta_i/2) \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \sin(\theta_i/2) \begin{pmatrix} 0 \\ u_2 \end{pmatrix},$$

where θ_i is the angle of $M_{12} U_i$. The second possibility, $u_2^\top M_{12} U_i u_2 = -1$, means that $M_{12} U_i$ is a rotation of angle π around an axis orthogonal to u_2. Its quaternion can be written in terms of u_3 and u_4, two 3×1 unit vectors which are a basis of the vector space orthogonal to u_2, as

$$q_i^m = \cos(\theta_i/2) \begin{pmatrix} 0 \\ u_3 \end{pmatrix} + \sin(\theta_i/2) \begin{pmatrix} 0 \\ u_4 \end{pmatrix},$$

where θ_i is related to the direction of the axis of $M_{12} U_i$ in the vector space orthogonal to u_2. Thus the uniqueness condition fails when the quaternions q_i^m belong to two orthogonal great circles on S_3, the four dimensional unit sphere. The quaternions for rotations fulfilling $u_2^\top M_{12} U_i u_2 = 1$ are on the great circle spanned by $(1, 0, 0, 0)^\top$ and $(0, u_2^\top)^\top$ while the great circle in the plane orthogonal to these two vectors contains the quaternions for the rotations satisfying $u_2^\top M_{12} U_i u_2 = -1$. Now the quaternions for U_i, can be expressed as $q_i = M_+(\pi, (u_1 + u_2)/(2 + 2u_1^\top u_2)^{1/2}) q_i^m$ where the rotation M_+ is defined by (5). This completes the proof of Proposition 2.

A.2. Proof of Proposition 3.

Using (3), one can express $SSR[A_1 \exp\{\mathcal{S}(\beta_1)\}, A_2 \exp\{\mathcal{S}(\beta_2)\}]$ as

$$\sum_{i=1}^n \mathrm{tr}\Big([A_1^\top V_i A_2 U_i^\top - \exp\{\mathcal{S}(\beta_1)\}\exp\{\mathcal{S}(-U_i\beta_2)\}] \\ [A_1^\top V_i A_2 U_i^\top - \exp\{\mathcal{S}(\beta_1)\}\exp\{\mathcal{S}(-U_i\beta_2)\}]^\top\Big).$$

Keeping only quadratic terms, and using (3) and (4) to evaluate some matrix products,

$$\begin{aligned} \exp\{\mathcal{S}(\beta_1)\}\exp\{\mathcal{S}(-U_i\beta_2)\} &= I + \mathcal{S}(\beta_1 - U_i\beta_2) - U_i\beta_2\beta_1^\top + \beta_1^\top U_i \beta_2 I \\ &\quad + \beta_1\beta_1^\top/2 - \beta_1^\top\beta_1 I/2 + U_i\beta_2\beta_2^\top U_i^\top/2 - \beta_2^\top\beta_2 I/2 + o(\beta^\top\beta). \end{aligned}$$

In the above expression, the quadratic terms can be written as $S(\beta_1, \beta_2) - U_i\beta_2\beta_1^\top$ where $S(\beta_1, \beta_2)$ is a symmetric matrix. Let r_i, the 3×1 residual vector for observation i be defined by (11). One can write $A_1^\top V_i A_2 U_i^\top - I = S_i + \mathcal{S}(r_i)$ where S_i is a symmetric matrix. Thus up to quadratic terms in β, $A_1^\top V_i A_2 U_i^\top - \exp\{\mathcal{S}(\beta_1)\}\exp\{\mathcal{S}(-U_i\beta_2)\}$ is equal to

$$S_i - S(\beta_1, \beta_2) + \mathcal{S}(r_i - \beta_1 + U_i\beta_2) + U_i\beta_2\beta_1^\top.$$

For any 3×3 symmetric matrix S, $\mathrm{tr}\{S\mathcal{S}(a)\} = 0$. Thus, up to quadratic terms, the trace of the product of $A_1^\top V_i A_2 U_i^\top - \exp\{\mathcal{S}(\beta_1)\}\exp\{\mathcal{S}(-U_i\beta_2)\}$ by its transpose is equal to

$$\begin{aligned}
& \mathrm{tr}(S_i^2) - 2\mathrm{tr}\{S_i S(\beta_1, \beta_2)\} + 2\mathrm{tr}(S_i\beta_1\beta_2^\top U_i^\top) \\
& \quad +2(r_i - \beta_1 + U_i\beta_2)^\top(r_i - \beta_1 + U_i\beta_2) + 2\mathrm{tr}\{\mathcal{S}(r_i)\beta_1\beta_2^\top U_i^\top\} \qquad (14) \\
= & \quad \mathrm{tr}(S_i^2) - \mathrm{tr}(S_i)(2\beta_1^\top U_i\beta_2 - \beta_1^\top\beta_1 - \beta_2^\top\beta_2) - \beta_1^\top S_i\beta_1 - \beta_2^\top U_i^\top S_i U_i\beta_2 \\
& \quad +2(r_i - X_i\beta)^\top(r_i - X_i\beta) + 2\beta_2^\top U_i^\top\{S_i + \mathcal{S}(r_i)\}\beta_1,
\end{aligned}$$

where X_i is the 3×6 component of the design matrix for data point i. Thus, in terms of the matrix Z_i defined in proposition 3, (14) reduces to

$$\mathrm{tr}(S_i^2) + 2(r_i - X_i\beta)^\top(r_i - X_i\beta) + \beta^\top Z_i\beta + 2\beta_2^\top U_i^\top\mathcal{S}(r_i)\beta_1.$$

Summing the contributions of the n data points to the expansion completes the proof of Proposition 3.

REFERENCES

B. M. Brown (1985). Multiparameter linearization theorems. *Journal of the Royal Statistical Society, Series B*, 47, 323–331.

T. Chang (1986). Spherical regression. *The Annals of Statistics*, 14, 907–924.

T. Chang, D. Ko, J.-Y. Royer & J. Lu (2000). Regression techniques in plate tectonics. *Statistical Science*, 15, 342–356.

T. Chang & L.-P. Rivest (2001). M-Estimation for location and regression models on Stiefel manifolds. *The Annals of Statistics*, 29, 784–814.

Y. Chikuse (2003). *Statistics on Special Manifolds*. Lecture Notes in Statistics, Volume 174. Springer, New York.

J. S. Day, G. S. Dumas & D. J. Murdoch (1998). Evaluation of a long range transmitter for use with a magnetic tracking device in motion analysis. *Journal of Biomechanics*, 31, 957–961.

J. S. Day, D. J. Murdoch & G. S. Dumas (2000). Calibration of position and angular data from a magnetic tracking device. *Journal of Biomechanics*, 33, 1039–1045.

T. D. Downs (1972). Orientation statistics. *Biometrika*, 59, 665–676.

I. L. Dryden & K. V. Mardia (1998). *Statistical Analysis of Shape*. Wiley, New York.

R. J. Epp, J. W. Tukey & G. S. Watson (1971). Testing unit vectors for correlation. *Journal of Geophysical Research*, 76, 8480–8483.

W. R. Hamilton (1969). *Elements of Quaternions*. Chelsea Press, Chelsea.

P. E. Jupp & B. D. Spurr (1985). Sobolev tests for independence of directions. *The Annals of Statistics*, 13, 1140–1155.

C. León, J.-C. Massé & L.-P. Rivest (2006). A statistical model for random rotations. *Journal of Multivariate Analysis*, 97, 412–430.

J. M. McCarthy (1990). *Introduction to Theoretical Kinematics*. The MIT Press, Cambridge, MA.

K. V. Mardia & P. E. Jupp (2000). *Directional Statistics*. Wiley, New York.

M. J. Prentice (1986). Orientation statistics without parametric assumptions. *Journal of the Royal Statistical Society, Series B*, 48, 214–222.

M. J. Prentice (1989). Spherical regression on matched pairs of orientation statistics. *Journal of the Royal Statistical Society*, *Series B*, 51, 241–248.

D. Rancourt, L.-P. Rivest & J. Asselin (2000). Using orientation statistics to investigate variations in human kinematics. *Applied Statistics*, 49, 81–94.

L.-P Rivest (1989). Spherical regression for concentrated Fisher–von Mises distributions. *The Annals of Statistics*, 17, 307–317.

H. S. H. Shin (1999). *Experimental Design for Orientation Models.* Doctoral dissertation, Queen's University, Kingston, Ontario, Canada.

H. S. H. Shin, G. K. Takahara & D. J. Murdoch (2001). Uniqueness, consistency and optimality in spherical regression experiments. *Statistics and Probability Letters*, 54, 61–65.

F. W. Warner (1983) *Foundations of Differentiable Manifolds and Lie Groups*. Springer, New York.

J. A. Wellner (1979). Permutation tests for directional data. *The Annals of Statistics*, 7, 929–943.

Received 3 March 2004
Accepted 5 December 2005

Louis-Paul RIVEST: `lpr@mat.ulaval.ca`
Département de mathématiques et de statistique
Université Laval, Québec, Canada G1K 7P4

Ted CHANG: `e-mail tcc8v@virginia.edu`
Department of Statistics, University of Virginia
Charlottesville, Virginia 22902-4135 U.S.A.