Measures of Agreement for Rotation Matrices with Application to Vectorcardiography Data

Haonan Wang
Department of Statistics
Colorado State University
Fort Collins, CO 80523

May 20, 2008

Abstract

In this paper, nonparametric methods are proposed for quantifying agreement and disagreement between different measurement methods when the results of the measurements are rotation matrices. Firstly, the expected squared distance is used to quantify the measurement agreement. Two choices of such distance are considered — the Frobenius distance and geodesic distance. Secondly, the notion of "concordance correlation coefficient", a commonly used measure of agreement, is extended to the space of rotation matrices. Such generalized concordance coefficient can be treated as a normalized expected squared distance. Since no two measurement systems can be expected to be in perfect agreement, it becomes necessary to define a notion of practical agreement. We define such a notion. Moreover, for both proposed methods, the percentile bootstrap procedure is implemented to provide a confidence interval to help make a decision concerning practical agreement/disagreement in real life applications. The methodology is illustrated using two data sets, one based on an application involving vectorcardiography data [1], and the other based on a synthetic data set.

Key words: Concordance correlation coefficient, Measurement disagreement, Object oriented data analysis, Orthogonal matrix, Vectorcardiography

1 Introduction

Most clinical measurements are not direct measurements but are estimates of quantities whose true values can never be known. Some disagreement between different methods of measurement of clinical parameters is inevitable and whenever a new method is introduced for measuring a particular clinical quantity there is a need to evaluate the adequacy of the new method. However, a direct comparison of the measured value to the true value is an impossibility. Instead one must resort to comparing the results from the new method to an established and accepted method and evaluate the degree of agreement between the two methods. Researchers would like to quantify the difference between competing methods. Practitioners would be interested in quantifying the level of agreement between two laboratories or two technicians, etc.

[2, 3, 4] have discussed this problem in great detail and have pointed out that the use of the Pearson correlation coefficient or the paired t-test, procedures commonly used by some practitioners, is an incorrect statistical approach to the method comparison problem. They proposed that the first step should be to plot the data and superimpose the unit slope line through the origin which will help visually evaluate the degree of agreement between the two methods. They noted that a better approach would be to plot the differences between the two values in each pair of measurements against their mean. Let $\bar{d}$ be the mean and s be the standard deviation of the differences. Then a large proportion of the measurement differences will be expected to lie in the interval $(\bar{d} - 2s, \bar{d} + 2s)$. Provided that this entire interval is *close enough to zero* for the differences to be clinically insignificant, the two methods may be deemed to be in agreement for practical purposes.

[5] proposed the use of the Concordance Correlation Coefficient (CCC) for comparing two methods. This suggestion has been followed in a number of pharmaceutical, medical and other applications by subsequent authors. See, for instance, [6], [7], [8]. The CCC assesses agreement between methods using the rescaled expected mean squared difference between the two results. The value one for CCC would correspond to perfect agreement. [9] generalized this idea to categorical data and continuous data. In a recent paper by [10], the concordance correlation coefficient for functional data has been proposed. A detailed review of various other methods and discussions relating to the strengths and weaknesses of these methods is provided by [11].

While any two measurement methods are bound to produce nonidentical results when measuring the same object, it is of practical interest to determine whether or not the disagreement between two measurements of the same quantity is within acceptable limits. Such an acceptance limit is prespecified by the regulatory agency in situations subject to government regulations. This is a well known issue in the field of bioequivalence.

In most applications, measurements are either scalars or vectors but this is not always the case. For instance, in the field of vectorcardiography, the QRS loop of a vectorcardiogram is described in terms of a rotation matrix. Thus, each measurement is actually a rotation matrix. Two commonly used measurement systems in vectorcardiography are the Frank system and the McFee system. One would therefore want to know if these systems may be considered equivalent for clinical applications. Even if only one measurement system is to be used, one would be interested in the magnitude of the differences in measurements from different laboratories or different technicians, etc. Measurement agreement studies, sometimes also called method comparison studies, are therefore of much practical importance especially in critical application areas such as vectorcardiography.

The next subsection explains the relevant details regarding a vectorcardiogram.

1.1 Some basics of vectorcardiogram

We begin with a brief description of a vectorcardiogram. During each cardiac cycle, the electrical currents pass through tissues surrounding the heart and generate electrical potentials, which can be recorded using electrodes placed on the skin. Such a recording is called an electrocardiogram (ECG). A typical ECG contains a P wave, a QRS loop/complex, and a T wave; see Figure 1. For diagnostic purposes the QRS loop is very important.

The electrical activity associated with the functioning of the heart during the QRS loop can be represented as a vector at any instant of time. The terminal point of the vector traces a closed

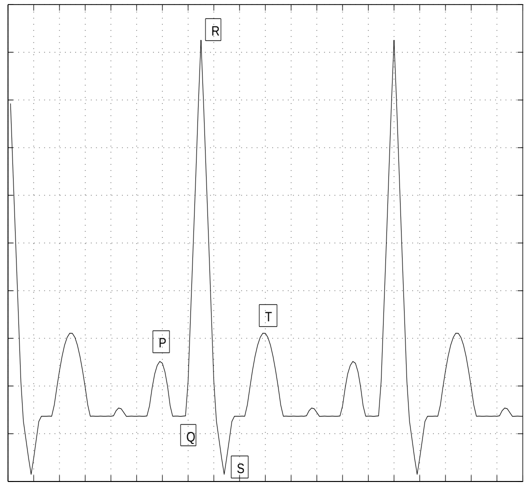

Figure 1: An illustrative example of ECG. The synthetic data is simulated from the matlab code provided by R. KARTHIK. The code is publicly available at the website of `MATLAB Central.`

path in the 3-dimensional space during each cardiac cycle. This path may be approximated by a planar closed curve roughly having the shape of a cardioid and is called a vectorcardiogram (VCG); see Figure 2. In a vectorcardiogram, the direction of the ray connecting the apex to the vertex is denoted by a unit vector $\boldsymbol{p}$. Let $\boldsymbol{q}$ be the unit vector that is orthogonal to the plane of the QRS loop and having a direction determined by the direction in which the curve is traced of the QRS loop. These two orthogonal vectors, $\boldsymbol{p}$ and $\boldsymbol{q}$, determine the orientation of the QRS loop. Let $\boldsymbol{r}$ denote the vector given by $\boldsymbol{r} = \boldsymbol{p} \times \boldsymbol{q}$. Then $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r}$ form a set of mutually orthogonal vectors. In the end, the orientation of a QRS loop may be described using the rotation matrix

$$\boldsymbol{M} = [\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r}].$$

By choice of $\boldsymbol{r}$, the determinant of M is $+1$. So M is an element of SO(3), the group of orthogonal matrices with determinant $+1$. SO(3) is in fact a mathematical object known as a Lie Group. See [12] for more details.

Methods for comparison of the Frank and McFee systems for vectorcardiography have been studied previously. [13] approached this problem by considering several scalar quantities describ-

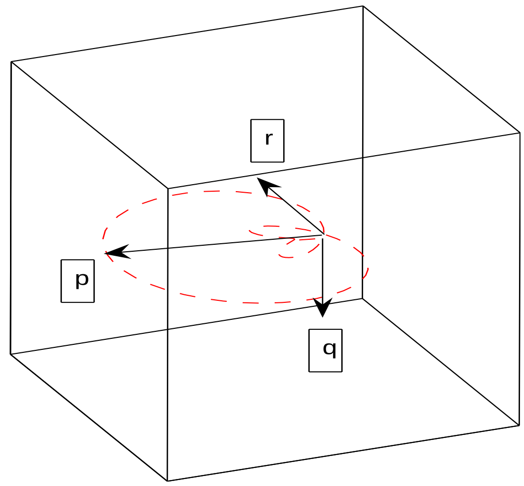

Figure 2: An illustrative example of VCG.

ing various features associated with a QRS loop. They were primarily interested in understanding the level of agreement or disagreement in a qualitative sense. They did not address the problem of equivalence of the two systems. [14] approached this problem by comparing the $\boldsymbol{p}$ vector with the $\boldsymbol{p}$ vector and the $\boldsymbol{q}$ vector with the $\boldsymbol{q}$ vector and studying the angles between these vectors.

In this paper we propose methods for comparing the two systems based on certain distance measures on the space of rotation matrices. For this we use the fact that each rotation matrix may be represented as a point in the 3-dimensional Euclidean space in a natural manner using the so called log map on SO(3). Two approaches for method comparison are proposed. One is based on measures of distance between pairs of rotation matrices, and the other is based on an extension of concordance correlation coefficient [5].

The paper is organized as follows. Section 2.1 gives some background information about rotation matrices, SO(3) group, the log map and distance measures. In Section 2.2, we define the notion of practical agreement between two measurements. In Section 2.3, a distance based measure of agreement and disagreement is proposed. Two natural distance measures are considered. One is the Euclidean distance and the other is the geodesic distance. In Section 2.4, we extend the notion of CCC for use with rotation matrices and show how it may be used to judge equivalence of the two vectorcardiography methods. All of the proposed tests for agreement or disagreement are based on the bootstrap principle. Section 3 includes an illustration using synthetic data to demonstrate the performance of the proposed methods for testing agreement. In Section 4, a case study involving vectorcardiography is discussed to illustrate how the computations are performed and what conclusions these lead to. Finally, Section 5 contains some summary remarks. The details related the principal logarithm of rotation matrices are included in the Appendix.

2 Methodology

The tests of agreement/disagreement we propose in this paper are based on distance measures defined on the space of square matrices. Hence we begin with a brief introduction to distances between matrices.

2.1 Matrix distances

A $p \times p$ matrix consists of p^2 elements and may be represented as a point in $\mathbb{R}^{p^2}$, the p^2-dimensional Euclidean space. We denote this mapping by E. Thus, for a given $p \times p$ matrix $A = (a_{ij})$,

$$E(A) = (a_{11}, a_{12}, \ldots, a_{1p}, \ldots, a_{p1}, \ldots, a_{pp})^T,$$

is a p^2-dimensional vector obtained by concatenating the rows of A. Hence an obvious distance measure between two matrices $A = (a_{ij})$ and $B = (b_{ij})$ is the standard Euclidean distance between $E(A)$ and $E(B)$. This is often called the Frobenius distance d_F (also called chordal distance) between A and B. Thus,

$$d_F(A, B) = ||A - B||_F = \left[\sum_{i=1}^{p} \sum_{j=1}^{p} (a_{ij} - b_{ij})^2 \right]^{1/2}.$$

While d_F seems to be a reasonable way of defining distance between matrices, other concepts of distance are sometimes more appropriate. In many problems, one is interested in a subclass of matrices with special properties. Some examples are: (a) class of symmetric matrices; (b) class of symmetric (perpendicular) projection matrices; (c) class of rotation matrices; (d) class of skew-symmetric matrices; and (e) class of square matrices with a specified rank. Such subclasses of matrices lie on lower dimensional surfaces embedded in $\mathbb{R}^{p^2}$. We give some illustrative examples.

Example 1: 2×2 **symmetric projection matrix**

Consider the set of 2×2 symmetric projection matrices of the form

$$P = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

P can be viewed as a point (a, b, b, c) in $\mathbb{R}^4$. All such points lie on a 3-dimensional hyperplane in $\mathbb{R}^4$. For convenience of visualization, it is sufficient to represent P by a point (a, b, c) in $\mathbb{R}^3$. Moreover, (a, b, c) will fall in one of three following cases: (i) $(a, b, c) = (0, 0, 0)$, (ii) $a^2 + b^2 = a$ and $a+c = 1$, (iii) $(a, b, c) = (1, 0, 1)$. Note that, these three cases correspond to three possibilities for the rank of matrix P : $0, 1$, and 2. Any symmetric rank-1 projection matrix (i.e. $P \neq 0$ or I_2) will correspond to a point on the intersection of the cylinder $a^2 + b^2 = a$ and plane $a + c = 1$ in $\mathbb{R}^3$. See Figure 3 for more detail.

The Frobenius distance between two rank-1 projection matrices

$$P_1 = \begin{pmatrix} a_1 & b_1 \\ b_1 & c_1 \end{pmatrix} \text{ and } P_2 = \begin{pmatrix} a_2 & b_2 \\ b_2 & c_2 \end{pmatrix}$$

is

$$d_F(P_1, P_2) = \sqrt{(a_1 - a_2)^2 + 2(b_1 - b_2)^2 + (c_1 - c_2)^2}.$$

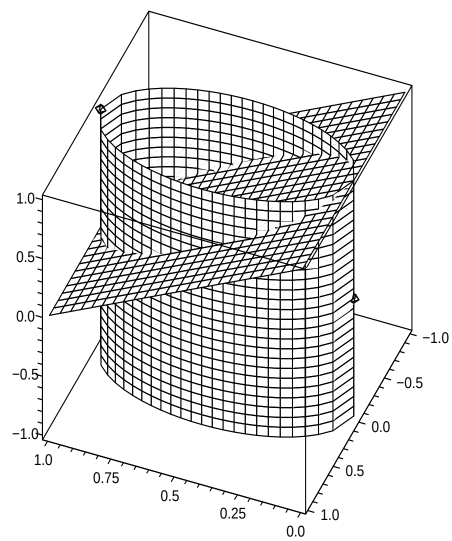

Figure 3: Graphical illustration of all 2×2 symmetric projection matrices: all rank-1 matrices lie on the intersection of the cylinder and the hyperplane, and two trivial projection matrices, 0 and I_2, are highlighted as circles.

Example 2: 2×2 **rotation matrix**

Any 2×2 rotation matrix can be represented as

$$P = \begin{pmatrix} p_1 & -p_2 \\ p_2 & p_1 \end{pmatrix}, \text{ with } p_1^2 + p_2^2 = 1.$$

If we denote p_1 as $\cos\theta$, then p_2 can be written as $\sin\theta$, for a unique $\theta \in [0, 2\pi)$. Again, matrix P can be viewed as a point on a 1-dimensional closed curve in 4-dimensional Euclidean space. The Frobenius distance between two rotation matrices P_1 and P_2, where

$$P_i = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{pmatrix}, \ i = 1 \text{ and } 2$$

is $d_F(P_1, P_2) = 2\sqrt{2}|\sin\frac{1}{2}(\theta_1 - \theta_2)|$.

Example 3: 3×3 **rotation matrix**

Our interest in this paper is restricted to the subclass of 3×3 rotation matrices that arise in vectorcardiography. Rotation matrices are a subclass of orthogonal matrices. Any orthogonal matrix P satisfies the following equations

$$\begin{array}{ll} x_{11}^2 + x_{21}^2 + x_{31}^2 = 1 & x_{11}x_{12} + x_{21}x_{22} + x_{31}x_{32} = 0 \\ x_{12}^2 + x_{22}^2 + x_{32}^2 = 1 & x_{11}x_{13} + x_{21}x_{23} + x_{31}x_{33} = 0 \\ x_{13}^2 + x_{23}^2 + x_{33}^2 = 1 & x_{12}x_{13} + x_{22}x_{23} + x_{32}x_{33} = 0 \end{array} \tag{2.1}$$

These equations are a consequence of the fact that the column vectors of P have unit norm and are pairwise orthogonal.

Also, the determinant of any orthogonal matrix is either $+1$ or -1. The set of 3×3 orthogonal matrices, although lying in a space with dimension 9, have only three degrees of freedom because of the 6 conditions in (2.1). So they lie on a 3-dimensional surface which is embedded in $\mathbb{R}^9$. The set of orthogonal matrices may be written as a disjoint union of sets O^+ and O^-, those with determinant $+1$ and those with determinant -1 respectively. The set O^+ is the set of rotation matrices, and denoted by $SO(3)$, the special orthogonal group of 3×3 rotation matrices. Since elements of $SO(3)$ lie on a 3-dimensional surface, it would be convenient to represent this set as a 3-parameter family. This is indeed possible and is in fact well known. This result is given in the following theorem.

Theorem 1 *Let $\mathcal{S}$ be the set of all 3×3 skew symmetric matrices of the form*

$$S = \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix}.$$

Then

$$\exp(S) = \sum_{k=0}^{\infty} \frac{S^k}{k!}$$

is a rotation matrix. Conversely, every rotation matrix may be written as $\exp(S)$ *for some* $S \in \mathcal{S}$.

The following remark states a well-known fact about 3×3 matrices. For more details, see [15].

Remark 1 *The mapping* $\mathbb{R}^3 \to SO(3)$

$$v = (v_1, v_2, v_3)^T \mapsto R(v) = \exp\left(\begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix}\right)$$

is an onto mapping but is not one-to-one.

Geometrically, for any two rotation matrices A and B, a natural distance would be the length of the shortest curve Γ connecting A and B such that each point of Γ is in fact a rotation matrix, i.e. Γ is a curve lying on the surface consisting of rotation matrices. Such a distance is called the geodesic distance between A and B.

More generally, When working with a subclass of matrices that lies on a connected surface, one can consider the geodesic distance on this surface rather than the Frobenius distance in $\mathbb{R}^{p^2}$. For more detail, see [16], and [17]. In general, the geodesic distance between rotation matrices A and B is given by

$$d_G(A, B) = ||\log(A^T B)||_F, \tag{2.2}$$

where log denotes the *principal logarithm* of a matrix. Some relevant facts regarding logarithm of a rotation matrix are given in the Appendix.

Let the principal logarithm of a rotation matrix A be

$$\begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix}.$$

We define a mapping $V : SO(3) \longrightarrow \mathbb{R}^3$, such that

$$V(A) = (v_1, v_2, v_3)^T.$$

In Example 1, the geodesic distance between two symmetric projection matrices is the length of the shortest segment along an ellipse in $\mathbb{R}^4$ connecting P_1 and P_2, and is equal to

$$\sqrt{2}|\cos^{-1}(1 - 2c_1) - \cos^{-1}(1 - 2c_2)|.$$

In Example 2, the geodesic distance between two 2×2 rotation matrices is

$$d_G(P_1, P_2) = \sqrt{2} \min_{k_0} |\theta_1 - \theta_2 + k_0 \pi|$$

where k_0 is an integer.

Computation of the geodesic distance between two 3×3 rotation matrices involves the calculation of the principal logarithm of rotation matrix; see Appendix for details. [17] provides a detailed discussion regarding the computation of geodesic distances in the space of $p \times p$ rotation matrices. If

$$P_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \text{ and } P_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \tag{2.3}$$

then the Frobenius distance between them is given by $d_F(P_1, P_2) = 2$. The geodesic distance between P_1 and P_2 is $d_G(P_1, P_2) = \sqrt{2}\pi$.

2.2 Measurement Disagreement

Let (X, Y) be a pair of random quantities representing outcomes from two different measurement methods, or measurement environments, e.g. two different laboratories. If X and Y are scalars, then $X - Y$ quantifies the disagreement of the measurements. If the two methods are in *perfect agreement*, then X and Y are equal almost surely, and hence the expectation $E(X - Y)^2$ is 0. Measures of agreement have been proposed in the literature based on $E(X - Y)^2$. This gives us the motivation to use the expectation of the squared distance between X and Y, when X and Y are not scalars, to quantify measurement disagreement.

In the vectorcardiography data, both X and Y are elements belonging to rotation matrix group. Whenever X and Y are equal we have $X^T Y = I_p$, where I_p is a $p \times p$ identity matrix. Thus, a measure of agreement may be defined based on the distance of $Q = X^T Y$ to I_p. In fact, it can be shown that

$$d_F(X, Y) = d_F(I_p, Q), \text{ and } d_G(X, Y) = d_G(I_p, Q).$$

Clearly, if there is no measurement disagreement between two methods, then the matrix Q is equal to I_p; that is

$$d_F(I_p, Q) = d_G(I_p, Q) = 0.$$

Analogous to the expected squared difference in the scalar case, we quantify the amount of disagreement by

$$\phi = \mathbb{E}\left[d^2(Q, I_p)\right].$$

If $\phi = 0$, then X and Y are in perfect agreement with probability 1.

Perfect agreement is too strong a requirement for applications. So we define the notion of *practical agreement* between two methods. If the expected squared distance between X and Y (choice of such distance may vary) is sufficiently small, i.e., smaller than a pre-specified constant C, then we say that X and Y are in *practical agreement*. This constant is chosen based on the given application scenario.

In the rest of this section, we consider two measures of agreement, both of which depend on squared distance. The difference is that one is an unnormalized quantity and the other is normalized.

2.3 Testing for Practical Agreement/Disagreement

In this section, we will discuss the issue of practical agreement between two measurement methods. We say the two methods are in *practical agreement* if $\phi \leq C$, for some prespecified constant C chosen by user for the particular application. The two methods are in practical disagreement if $\phi > C$.

A test of

$$H_0 : \phi > C \text{ versus } H_a : \phi \leq C$$

may be conducted and if H_0 is rejected then one can conclude practical agreement between the two methods. We construct such a test by first considering an upper $100(1-\alpha)$ percent confidence bound for ϕ. If this bound is no more than C we reject the null hypothesis and conclude practical agreement. (Note that, a test of $H_0 : \phi \leq C$ versus $H_a : \phi > C$ may be used to demonstrate practical disagreement.) Here, we construct such an upper bound using the bootstrap principle based on the estimator $\widehat{\phi}_n$ of the disagreement measure ϕ,

$$\widehat{\phi}_n = \frac{1}{n} \sum_{i=1}^{n} d^2(Q_i, I_p),$$

where $Q_i = X_i^T Y_i$, $i = 1, \ldots, n$.

The procedure is summarized as follows:

1. For subject i, $i = 1 \ldots, n$, compute
$$d_i = d(Q_i, I_p).$$
2. Generate B bootstrap samples from $\{d_1, \ldots, d_n\}$, and denote the k-th bootstrap sample by $S_k = \{d^*_{k,i} : 1 \leq i \leq n\}$.
3. For each bootstrap sample S_k, compute
$$\widehat{\phi}^*_{n,k} = \frac{1}{n} \sum_{i=1}^{n} d^{*2}_{k,i}.$$
4. Obtain the 100α and $100(1-\alpha)$ percentiles, denoted by D^α and $D^{1-\alpha}$ respectively, where α is a pre-specified level of significance. If $D^\alpha > C$, then we can conclude that there is practical disagreement between the two methods. On the other hand, if $D^{1-\alpha} < C$, then practical agreement is established. If neither $D^\alpha > C$ nor $D^{1-\alpha} < C$, then a decision is not available regarding agreement or disagreement.

The bootstrap procedure considered in this section is percentile bootstrap of [18] and [19]. It does not require any pivotal statistic. The procedure is valid not only for the distance measure d

used above but also for any general proximity measure. In the synthetic data example in Section 3 and the case study in Section 4, we implement this procedure with the Frobenius distance d_F and geodesic distance d_G.

2.4 Concordance Coefficient in $SO(3)$: a measure of agreement

The choice of C in the previous section requires very careful analysis of the practical consequences of type I and type II errors. In a regulatory framework, as in the context of bioequivalence, the US FDA selects C. However, for routine applications, the choice of C may not be very clear. This is because the criterion in the previous section is based on an absolute scale rather than a relative scale. To overcome this problem, in the case of scalar measurements, [5] proposed a relative measure, which is called concordance correlation coefficient (CCC), for paired observation (X, Y) from a bivariate population with mean $(\mu_1, \mu_2)^T$ and covariance matrix

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

The concordance correlation coefficient (CCC) is defined as

$$\rho_c = 1 - \frac{\mathbb{E}(X-Y)^2}{\mathbb{E}_0(X-Y)^2},$$

where $\mathbb{E}_0(X-Y)^2$ is the expected squared difference between X and Y calculated under the assumption of independence between X and Y. The maximum value of ρ_c is 1 which is achieved if and only if X and Y are almost surely equal. Many practitioners find it convenient to select a threshold value for ρ_c to conclude practical agreement. Commonly used such threshold values are 0.99, 0.95, ..., etc.

We extended this idea to more general situations such as when X and Y are rotation matrices by defining the generalized concordance coefficient (GCC) as

$$\rho_c = 1 - \frac{\mathbb{E}\left[d(X,Y)^2\right]}{\mathbb{E}_0\left[d(X,Y)^2\right]}, \tag{2.4}$$

where d is a distance measure between X and Y. Here, $\mathbb{E}_0\left[d(X,Y)^2\right]$ is calculated under the assumption of independence between X and Y. In this paper, we consider two choices for d: Frobenius distance d_F and geodesic distance d_G. The resulting concordance coefficients are denoted by $\rho_{c,F}$ and $\rho_{c,G}$ respectively. Clearly, both $\rho_{c,F}$ and $\rho_{c,G}$ are no more than 1, and are equal to 1 if and only if X and Y are almost surely equal.

In practice, for a sample of n independent paired observations, it is natural to use the corresponding sample counterparts. The expected squared distance in the numerator can be estimated by

$$\widehat{\phi} = \mathbb{E}\,[\widehat{d(X,Y)}^2] = \frac{1}{n}\sum_{i=1}^{n} d(X_i, Y_i)^2.$$

However, the denominator is somewhat complicated to compute directly. Here, we use the bootstrap principle to approximate this quantity in the following manner.

1. Generate B bootstrap samples from $\{X_1, \ldots, X_n\}$ and $\{Y_1, \ldots, Y_n\}$ independently. Then the paired observations for the k-th bootstrap are denoted by

$$S_k = \{(X^*_{k,i}, Y^*_{k,i}) : 1 \le i \le n\}.$$

2. For each bootstrap sample S_k, we can compute

$$\widehat{\phi}^*_{k,0} = \frac{1}{n}\sum_{i=1}^{n} d(X^*_{k,i}, Y^*_{k,i})^2.$$

   Here, $\phi^*_{k,0}$ is used as an estimate of $\mathbb{E}_0\left[d(X,Y)^2\right]$.

Overall, the sample GCC statistic is

$$\widehat{\rho}_c = 1 - \frac{\widehat{\phi}}{\frac{1}{B}\sum_{k=1}^{B}\widehat{\phi}^*_{k,0}}.$$

Next consider the problem of testing for practical agreement

$$H_0: \ X \text{ and } Y \text{ are in practical disagreement.}$$

$$\text{versus } H_a : X \text{ and } Y \text{ are in practical agreement.}$$

If the GCC $\rho_c > 1 - \eta$ for some prespecified small η, then we can conclude practical agreement. Usually, η is taken to be 0.01 or 0.05. Such a test can be conducted based on the estimated concordance coefficient $\widehat{\rho}_c$. In this paper, we implement the bootstrap principle to obtain the $100(1-\alpha)$ percent lower confidence bound ρ^α. If $\rho^\alpha > 1 - \eta$, practical agreement can be established. On the other hand, if the $100(1-\alpha)$ percent upper confidence bound $\rho^{1-\alpha} < 1 - \eta$, we can conclude practical disagreement.

3 A synthetic data example

In this section, we consider a situation where the vectorcardiograms of a group of $n = 200$ hypothetical patients are measured from two different laboratory environments using the Frank lead system. We created a synthetic data set to mimic this situation. First we generate 200 3×3 rotation matrices (denoted by W_i), which are treated as *true vectorcardiogram* for the 200 patients using the Frank lead system. Next, we considered the hypothetical scenario where the 200 subjects were asked to obtain the vectorcardiogram from two different laboratories and these rotation matrices are recorded by paired observations (X_i, Y_i). The resulting rotation matrices are generated as follows: for subject i, generate two directional vectors independently from the uniform distribution on the 3-dimensional sphere, denoted by $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$. Based on those two directional vectors, write

$$X_i = A_i W_i \text{ and } Y_i = W_i B_i,$$

where

$$A_i = R(\theta_1 \boldsymbol{u}_i) \text{ and } B_i = R(\theta_2 \boldsymbol{v}_i),$$

and θ_1, θ_2 are the distances of A_i and B_i to the identity matrix respectively. These can be considered as the measurement error levels of two laboratories. In this analysis, we choose

$\theta_1 = \theta_2 = 0.2$.

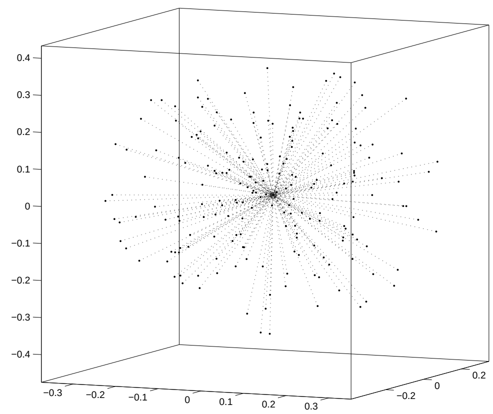

Figure 4: Scatterplot of the set of vectors, $V(\log(X^TY))$. The origin is marked as filled '$\circ$'.

Figure 4 depicts the log-transformed differences between X and Y, i.e. $V(\log(X^TY))$. Note that, the Euclidean distance between each point and the origin is exactly the geodesic distance between X and Y for each subject. All such vectors, $V(\log(X_i^TY_i))$, are contained within a small rectangular box. Again, this suggests that measurements X and Y would be in practical agreement with each other.

The estimates of ϕ, calculated based on both Frobenius and geodesic distances, are:

$$\widehat{\phi}_{n,F} = 0.1514 \text{ and } \widehat{\phi}_{n,G} = 0.1527.$$

Similarly, the sample generalized concordance coefficients are $\widehat{\rho}_{c,F} = 0.9724$ and $\widehat{\rho}_{c,G} = 0.9839$. We can see that, both $\widehat{\phi}_{n,F}$ and $\widehat{\phi}_{n,G}$ are small; on the other hand, both $\widehat{\rho}_{c,F}$ and $\widehat{\rho}_{c,G}$ are very close to 1. This surely suggests there is practical agreement between the measurements from two different laboratories. The 90% bootstrap confidence intervals for ϕ are given below

$$\text{Frobenius distance: } [0.1405, 0.1621]$$

and

$$\text{Geodesic distance: } [0.1417, 0.1636].$$

Similarly, the bootstrap confidence intervals, with Frobenius distance and geodesic distance, for ρ_c are

$$\text{Frobenius distance: } [0.9702, 0.9743]$$

and

$$\text{Geodesic distance: } [0.9827, 0.9851].$$

This suggests that there is a practical agreement between the measurements from two different laboratories.

In the next section, we will analyze a dataset of the vectorcardiograms from two different systems. In that case study, we are more interested in the disagreement between two measurement systems.

4 Vectorcardiography Data: a case study

In this section, the proposed methods will be implemented to analyze a data set of $n = 98$ matched pairs of vectorcardiogram orientations collected on 98 subjects using two different measurement systems. These data are described in [14], [1], and [20]. Each of the 98 pairs of orientations is obtained using two different lead systems, the *Frank system* (denoted by X_i) and the *McFee system* (denoted by Y_i), on the same patients. Interest naturally centers on any systematic difference between the orientations obtained on the same subject using different systems.

[1] considered the problem of parameter estimation using some parametric families of distributions on rotation matrices. [20] studied these data by transforming each orientation into a 4-dimensional axial data. Unlike [20], we consider the difference between X_i and Y_i within the set of 3×3 rotation matrices, $SO(3)$. In particular, denote

$$Q_i = X_i^T Y_i.$$

It is assumed that $\{Q_i : 1 \leq i \leq n\}$ form an i.i.d sample.

Remark 2 *An alternative approach is to define $Q_i^{'} = Y_i X_i^T$, and use $Q_i^{'}$ to quantify the difference between X_i and Y_i. Note that, Q_i and $Q_i^{'}$ may not be equal to each other since $p \times p$ rotation matrices are not commutative in general for $p > 2$. However, straightforward calculation shows that,*

$$d_F(Q_i, I_p) = d_F(Q_i^{'}, I_p) \text{ and } d_G(Q_i, I_p) = d_G(Q_i^{'}, I_p).$$

Thus, the proposed distance based measurement comparison methods will not be affected by the way we define matrix Q .

Figure 5 shows a scatter plot of $V(\log(Q_1)), \ldots, V(\log(Q_{98}))$. The Euclidean distance between each point and the origin is the geodesic distance between X_i and Y_i, i.e. the geodesic distance between Q_i and I_3. Unlike geodesic distance, it is in general impossible to visualize the pairwise Frobenius distances in $\mathbb{R}^3$.

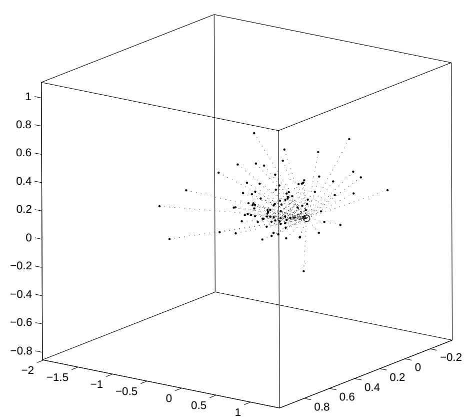

Figure 5: A graphical illustration of the log-transformed matrices, $\log(Q_i)$, in $\mathbb{R}^3$. Each transformed proximity matrix is represented as a black dot, and the origin (corresponds to I_3) is indicated as a circle.

In the left panel of Figure 6, a scatterplot of $d_G(Q_i, I_3)$ (x-axis) versus $d_F(Q_i, I_3)$ (y-axis) is provided. It can be seen that, the geodesic distance and the Frobenius distance are nearly linearly related, with Pearson correlation coefficient 0.9987 and the slope of the least-squares line through the origin 1.0853. This is not surprising because when X and Y are geodesically close to each other, $d_F(X,Y) \approx d_G(X,Y)$. Boxplots of those geodesic and Frobenius distances, shown in the right panel, suggest that there are five atypical values (outliers, shown as "+") in this sample. In fact, removal of those 5 atypical points will not affect the decision of both practical

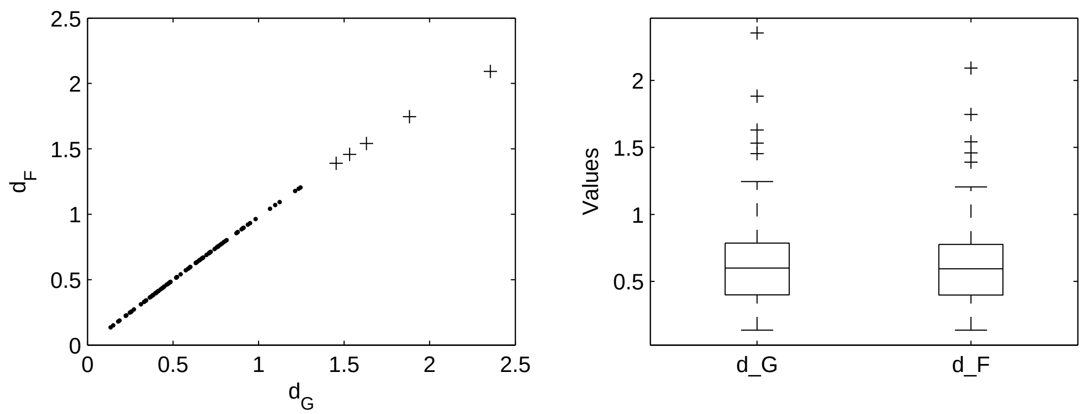

Figure 6: Scatterplot (left) and boxplots (right) of the pairwise distances, $d_F(Q_i, I_3)$ and $d_G(Q_i, I_3)$.

agreement and practical disagreement. Thus, the rest of data analysis will be conducted on the entire sample of 98 objects.

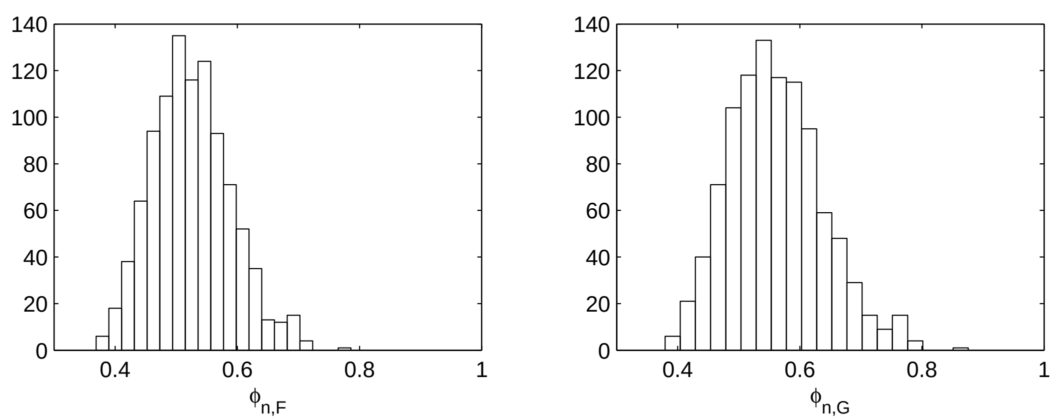

Figure 7: Histograms of $\hat{\phi}_n$ under Frobenius distance (Left) and geodesic distance (Right).

Next, we consider the problem of testing $H_0 : \phi > C$ as discussed in Section 2.3. The values of the estimate $\widehat{\phi}_n$ are $\widehat{\phi}_{n,F} = 0.5238$ and $\widehat{\phi}_{n,G} = 0.5590$, under the Frobenius distance and geodesic distance respectively. The proposed bootstrap procedure is conducted for $B = 1000$ bootstrap samples using both Frobenius distance and geodesic distance, as displayed in Figure 7. With α being taken to be 0.05, we get

$$\begin{aligned} \text{Frobenius} &: \quad D^{0.05} = 0.4249 \text{ and } D^{0.95} = 0.6344 \\ \text{Geodesic} &: \quad D^{0.05} = 0.4422 \text{ and } D^{0.95} = 0.6958 \end{aligned}$$

Thus, if a researcher plans to reject the null hypothesis

$$H_0 : X \text{ and } Y \text{ have practical measurement disagreement,}$$

the constant C has to be as large as 0.6344 or 0.6958 respectively. Otherwise, practical agreement can not be established.

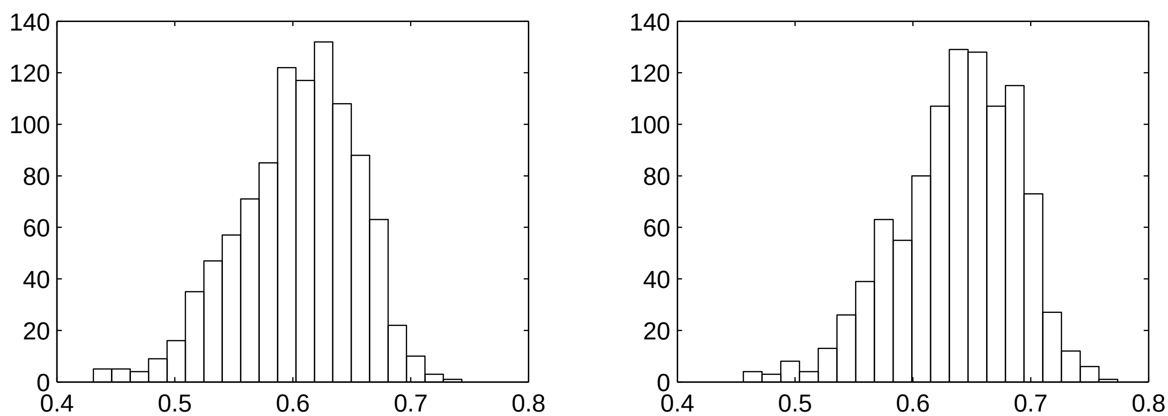

Figure 8: Histograms of $\hat{\rho}_c$ under Frobenius distance (Left) and geodesic distance (Right).

The sample generalized concordance coefficients in Equation (2.4) are $\widehat{\rho}_{c,F} = 0.6006$ and $\widehat{\rho}_{c,G} = 0.6320$. We can obtain a $100(1-\alpha)$ percent lower confidence bound and a $100(1-\alpha)$ percent upper confidence bound based on $B = 1000$ bootstrap samples. A histogram of those concordance correlation coefficients based on 1000 bootstrap samples are displayed in Figure 8, and ρ^{α} and $\rho^{1-\alpha}$ are found to be

$$\text{Frobenius} \quad : \quad \rho^{\alpha} = 0.5149 \text{ and } \rho^{1-\alpha} = 0.6757$$

$$\text{Geodesic} \quad : \quad \rho^{\alpha} = 0.5476 \text{ and } \rho^{1-\alpha} = 0.7088$$

respectively ($\alpha = 0.05$). This suggests that there is no practical agreement between two methods, but there is practical disagreement instead (with $\eta = 0.05$).

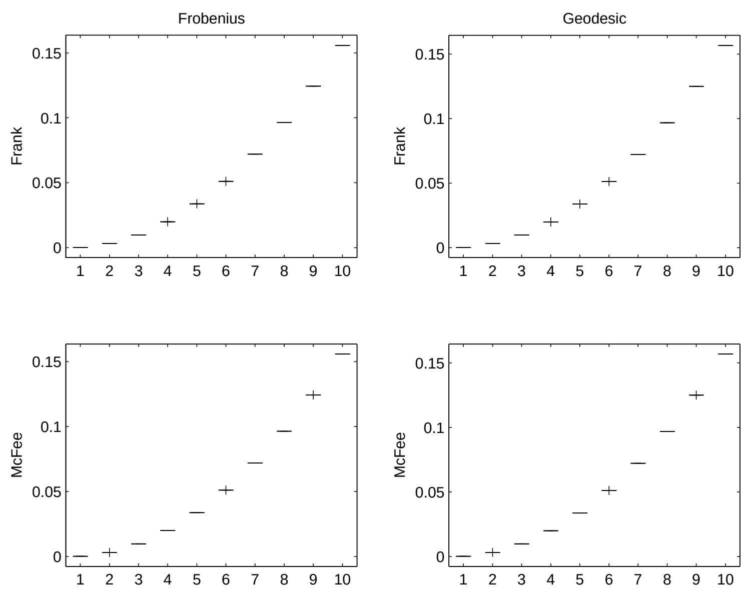

Figure 9: Boxplots of the estimates of ϕ as a function of ε using both Frobenius distance (Left Column) and geodesic distance (Right Column). In the first row, measurements from the Frank system have been selected to add a small perturbation. The second row, measurements from the McFee system have been perturbed.

To better understand how close to 1 that a practitioner should require GCC to be, the following exercise can provide valuable guidance. First, we select one measurement method, and perturb the rotation matrices with known amount of measurement disagreement. To be more specific, for subject i, a 3×1 vector

$$\boldsymbol{v}_i = (v_{i1}, v_{i2}, v_{i3})^T$$

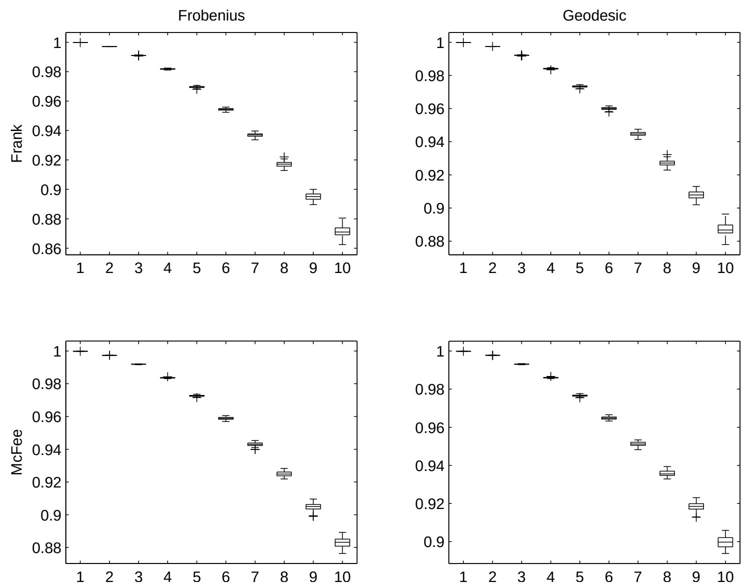

Figure 10: Boxplots of the estimates of ρ_c as a function of ε using both Frobenius distance (Left Column) and geodesic distance (Right Column). In the first row, measurements from the Frank system have been selected to add a small perturbation. The second row, measurements from the McFee system have been perturbed.

is generated from a multivariate normal distribution with mean 0 and covariance I_3. As discussed in Remark 1, let

$$\widetilde{X}_i = R(\varepsilon \boldsymbol{v}_i/||\boldsymbol{v}_i||)X_i.$$

Clearly, $\widetilde{X}_i$ is a rotation matrix, and $d_G(X_i, \widetilde{X}_i) = \varepsilon$. Then we calculate $\widehat{\phi}_{n,F}$ and $\widehat{\phi}_{n,G}$ based on $\{(X_i, \widetilde{X}_i) : 1 \leq i \leq n\}$. We repeat this process for $\varepsilon = 0.01, 0.04, \ldots, 0.28$, corresponding to the horizontal labels $1, 2, \ldots, 10$ in Figures 9 and 10, and for samples from both Frank and McFee systems. For each choice of ε, 100 independent realizations of $\widehat{\phi}_{n,F}$ and $\widehat{\phi}_{n,G}$ are obtained. Boxplots of $\widehat{\phi}_{n,F}$ and $\widehat{\phi}_{n,G}$, as functions of ε, are shown in Figure 9. It can be seen that, the values of ϕ increase as the level of error (i.e. ε) increases regardless of whether the Frobenius distance or geodesic distance is used. This helps the researchers choose the constant C by considering allowable value of ε for practical agreement. Same procedure can also be implemented to choose the threshold constant for GCC. In Figure 9, we can see that, for the disagreement level with

$\varepsilon = 0.1$, the threshold value of C can not exceed 0.025 to conclude practical agreement. At the same time, the threshold value of GCC has to be no less than 0.98; see Figure 10. Thus, in our vectorcardiography case study, we can conclude that, there is a strong disagreement between the Frank and McFee lead systems.

Although the Frank and McFee systems do not agree, it may be possible to calibrate one system with respect to the other system so that the measurements using one system may be transformed appropriately using the calibration equation in order to compare with measurements made using the other system. This problem has been considered by [21], where they tested the existence of an approximate relationship of the form $Y = AXB$ between Y and X. For detailed discussion, see [21] and [22].

5 Summary

Measurement agreement analysis is a very important topic in clinical measurements. New methods for measuring key clinical parameters need to be validated by comparing them to standard methods. Measurements made by one laboratory need to be compared with measurements made by a different laboratory. Statistical methods for comparing scalar measurements as well as vector measurements have been available for quite some time. However, in areas such as vectorcardiography, the measurements are actually rotation matrices and hence new approaches are needed for demonstrating agreement between methods. In this paper, we have proposed two approaches for tackling this situation. One approach uses a distance based statistic ϕ to quantify the level of agreement/disagreement and the other approach is a normalized version of ϕ, called generalized concordance coefficient (GCC). We have proposed statistical tests for demonstrating practical agreement/disagreement using the above measures and applying the bootstrap principle. The computations were illustrated using data from a case study involving vectorcardiograms. Practical guidelines are given to help practitioners in the choice of threshold constants for demonstrating practical agreement.

Acknowledgement

This work was partially supported by the National Science Foundation under grant DMS-0706761.

6 Appendix: Logarithm of a rotation matrix

The logarithm of a matrix S is a matrix T such that the matrix exponential of T is equal to S. A logarithm of a matrix exists if it is invertible, but this matrix logarithm may not be unique. Any rotation matrix has a logarithm. Moreover, if matrix S does not have any eigenvalues on the negative real axis, then it has a unique logarithm whose spectrum falls in an open strip $\{z \in \mathbb{C} : -\pi < \operatorname{Im} z < \pi\}$. This logarithm is called principal logarithm, denoted by $\log(S)$. In general, for any matrix S, if the spectral radius of $S - I$ is less than 1, $\log(S)$ is given by

$$\log(S) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} (S - I)^k. \tag{6.1}$$

For details about the matrix logarithm and exponential, see [23, 24].

In particular, for any 3×3 rotation matrix Q, except for the case where Q has eigenvalues equal to -1, the computation of the matrix $\log Q$ can be carried out through Rodrigues' rotation formula:

$$\log Q = \begin{cases} 0 & \text{if } \theta = 0 \\ \frac{\theta}{2\sin\theta}(Q - Q^T) & \text{if } \theta \neq 0 \end{cases}, \tag{6.2}$$

where $2\cos\theta + 1 = \operatorname{tr} Q$, and $\theta \in (-\pi, \pi)$. Moreover, it can be shown that

$$|| \log(Q) ||_F = \sqrt{2}|\theta|.$$

Even though, in the case of $\theta = \pi$, the logarithm is not unique, the above formula is still correct. For the two 3×3 rotation matrices in Equation (2.3), $P_1^T P_2$ indeed has a pair of eigenvalues of -1. Thus, $d_G(P_1, P_2) = \sqrt{2}\pi$.

References

[1] T. Downs. Orientation statistics. *Biometrika*, 59:665–676, 1972.

[2] D.G. Altman and J.M. Bland. Measurement in medicine: the analysis of method comparison studies. *The Statistician*, 32:307–317, 1983.

[3] J.M. Bland and D.G. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, i:307–310, 1986.

[4] J.M. Bland and D.G. Altman. Statistical methods for assessing agreement between measurement. *Biochimica Clinica*, 11:399–404, 1987.

[5] L. Lin. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45:255–268, 1989.

[6] D.J. Lai and S.Y.P.K. Shiao. Comparing two clinical measurements: A linear model approach. *Journal of Applied Statistics*, 32:855–860, 2005.

[7] J.L. Carrasco and L. Jover. Concordance correlation coefficient applied to discrete data. *Statistics in Medicine*, 24:4021–4034, 2005.

[8] T.S. King, V.M. Chinchilli, K.L. Wang, and J.L. Carrasco. A class of repeated measures concordance correlation coefficients. *Journal of Biopharmaceutical Statistics*, 17:653–672, 2007.

[9] T.S. King and V.M. Chinchilli. A generalized concordance correlation coefficient for continuous and categorical data. *Statist. Med.*, 20:2131–2147, 2001.

[10] R. Li and M. Chow. Evaluation of reproducibility for paired functional data. *Journal of Multivariate Analysis*, 93:81–101, 2005.

[11] L. Lin, A.S. Hedayat, B. Sinha, and M. Yang. Statistical methods in assessing agreement: models, issues and tools. *J. Amer. Statist. Assoc.*, 97:257–270, 2002.

[12] S.L. Altmann. *Rotations, Quaternions, And Double Groups.* Dover, 2005.

[13] J.S. Kan, J. Liebman, M.H. Lee, and A. Whitney. Quantification of the normal frank and mcfee-parungao orthogonal electrocardiogram at ages two to ten years. *Circulation*, 55:31:40, 1977.

[14] T. Downs, J. Liebman, and W. Mackay. Statistical methods for vectorcardiogram orientations. In I. Hoffman, R.I. Hamby, and E. Glassman, editors, *Vectorcardiography 2: proc. XIth International Symposium on Vectorcardiography*, pages 216–222. North-Holland, Amsterdam, 1971.

[15] S. Helgason. *Differential Geometry, Lie Groups, and Symmetric Spaces*, volume 34. American Mathematical Society, 2001.

[16] D. Barden and C. Thomas. *An Introduction to Differential Manifolds.* Imperial College Press, 2003.

[17] H. Wang. Object oriented data analysis: Vectorcardiography data. Manuscript, 2007.

[18] B. Efron. *The Jackknife, the bootstrap and other resampling plans.* SIAM, Philadelphia, 1982.

[19] P. Hall. *The bootstrap and Edgeworth expansion.* Springer, New York, 1992.

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

[21] T.D. Downs. Spherical regression. *Biometrika*, 90:655–668, 2003.

[22] T. Chang. Spherical regression. *The Annals of Statistics*, 15:907–924, 1986.

[23] Wikipedia. http://en.wikipedia.org/wiki/matrix_exponential. Internet site, 2007a.

[24] Wikipedia. http://en.wikipedia.org/wiki/logarithm_of_a_matrix. Internet site, 2007b.