MAXIMUM-LIKELIHOOD AND BEST INVARIANT ORIENTATION ESTIMATION

I. Vaughan L. Clarkson,[*] *Stephen D. Howard,*[†] *William Moran*[‡], *Douglas Cochran*[¶] *and Megan L. Dawson*[*]

[*] School of Information Technology & Electrical Engineering,The University of Queensland, Qld, 4072, AUSTRALIA

[†]Defence Science & Technology Organisation, P.O. Box 1500, Edinburgh,SA, 5111, AUSTRALIA

[‡] Department of Electrical & Electronic Engineering,The University of Melbourne, Vic, 3010, AUSTRALIA

[¶] School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ, USA.

ABSTRACT

Estimation of orientation or attitude is a longstanding problem in statistics and signal processing, with applications to crystallography, aeronautics and computer vision. In this paper, we consider the Mackenzie-Wahba estimator. This estimator is known to be maximum-likelihood but we further show that it is a Bayes estimator and also a best invariant estimator.

1. INTRODUCTION

Given a number of noisy directional measurements in a local reference frame to points (or distant objects) whose directions are known in a global reference frame, how can we best determine the rotation between the two frames? The problem arises not only in terrestrial navigation but in astronautics, crystallography, robotics, machine vision and even plate tectonics [1].

The orientation problem can be framed in two closely related ways. In one version, only a rigid-body rotation is allowed between the two frames. This makes sense when the application is to the determination of the attitude of a rigid body, such as a spacecraft. In a second version, rotation with reflection is allowed. This may be applicable in estimating the orientation of a crystal structure. In terms of group theory, the allowable transformation in the second version is an element of the *orthogonal group* $O(n)$ while in the first it is an element of the *special orthogonal group* $SO(n)$, where n is the dimension ($n = 2$ or $n = 3$ in crystallography and attitude estimation).

Detailed mathematical treatment of orientation estimation has a history of over 40 years, beginning with Mackenzie [2] in application to crystallography. He discovered an estimator which solves a certain least-squares formulation of the problem in $O(3)$. He showed that the estimator is also maximum-likelihood (ML) when directional errors have a Fisher distribution.

A closely related estimator was first presented as the solution to a puzzle posed by Wahba [3, 4] in application to the attitude of a satellite. Here, the formulation of the problem is in $SO(n)$ and is again a form of least squares.

Wahba's puzzle and its solution sparked a great deal of further research on the subject so that the problem now often bears her name. It has become a standard technique for spacecraft attitude estimation and is commonly used operationally. Efficient means for solving Wahba's problem, *i.e.*, computing the (weighted) least-squares estimate, in three dimensions using quaternions has been developed [5]. The connection with maximum likelihood was discovered (or rediscovered in the Wahba context) by Shuster [6] and the Fisher information examined. The problem has been extended and generalised in many different ways, *e.g.*, to dynamical modelling of attitude [7] and to generalised measurement paradigms [8].

Research interest has also been conspicuous in the statistics literature, where the problem is often known as *spherical regression* [1]. A Bayesian approach to spherical regression has been developed [9].

In this paper, we derive new properties of the Mackenzie-Wahba orientation estimator. Extending the Bayesian framework of Chang and Bingham [9], we show that the minimisation of Bayesian risk is the best invariant estimator and is equivalent to maximum-likelihood estimation under certain conditions.

The paper is set out as follows. In Section 2, we present the statistical model for orientation estimation and review the Mackenzie-Wahba estimator in the context of maximum likelihood. In Section 3, we present the main result of the paper, that the Bayesian estimator is identical to ML. In Section 4,we show that this ML/Bayes estimator is the best invariant estimator for estimation on $O(n)$ and $SO(n)$.

2. MAXIMUM-LIKELIHOOD ORIENTATION

Suppose we know the directions to N points specified as unit vectors $\mathbf{x}_j \in \mathbb{R}^n$, $j = 1, \ldots, N$, in a global or canonical coordinate frame. We measure directions $\mathbf{y}_j$ to the same points as unit vectors in a local coordinate frame. If the directions could be measured without error, we would have $\mathbf{x}_j = \mathbf{R}\mathbf{y}_j$ where $\mathbf{R}$ is an orthogonal matrix representing the rotation between the local and global coordinate frames.

Two different models may be distinguished governing the allowable values of $\mathbf{R}$. In one model, any orthogonal trans-

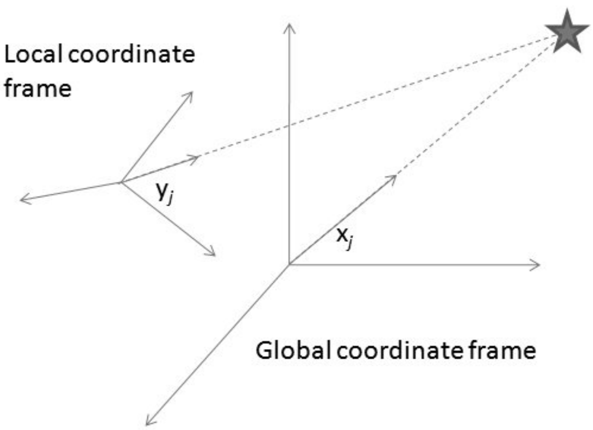

Fig. 1. Unit vectors x_j and y_j

formation is allowed so $\mathbf{R}$ represents an element of the group $O(n)$ and the matrix $\mathbf{R}$ has determinant ± 1. In the other model, only rigid-body rotations are allowed and so $\mathbf{R}$ represents an element of the group $SO(n)$ and $\det \mathbf{R} = 1$.

We will assume the directional measurements are made with errors. They are assumed to be instances of random variables $\mathbf{Y}_j$ from a Fisher distribution [10] with spherical means $\mathbf{R}^{-1}\mathbf{x}_j$, *i.e.*,

$$p_{\mathbf{Y}_j}(\mathbf{y}_j) = \frac{\kappa}{4\pi \sinh \kappa} \exp(\kappa \mathbf{x}_j^\mathsf{T} \mathbf{R} \mathbf{y}_j)$$

where κ is the concentration parameter of the distribution. Note that to simplify notation we will omit the subscripts from pdfs, $p_{\mathbf{Y}_j}$ will be written as $p(\mathbf{y}_j)$, $p_{\mathbf{R}|\mathbf{Y},\mathbf{X}}$ as $p(\mathbf{R}|\mathbf{Y},\mathbf{X})$.

With independent measurements and stacking the column vectors $\mathbf{x}_j$ to form the matrix $\mathbf{X}$ and column vectors $\mathbf{y}_j$ to form the matrix $\mathbf{Y}$, we find that the likelihood function for $\mathbf{R}$ is

$$L(\mathbf{R}) = \Big(\frac{\kappa}{4\pi \sinh \kappa}\Big)^N \exp \kappa \operatorname{Tr}(\mathbf{Y}\mathbf{X}^\mathsf{T}\mathbf{R}).$$

Define $\mathbf{P} = \mathbf{Y}\mathbf{X}^\mathsf{T}$ assume rank $\mathbf{P} = n$ if $\mathbf{R} \in O(n)$ and rank $\mathbf{P} \geq n-1$ if $\mathbf{R} \in SO(n)$. Full rank is not required if we restrict $\mathbf{R}$ to $SO(n)$ because, as will be seen in the proof of Theorem 2, the orientation of the final axis is determined by the others. The rank of $\mathbf{P}$ is dependent on the number (N) of chosen directions x_j, provided $N \geq n$, $\mathbf{P}$ will have full rank. The polar decomposition of $\mathbf{P}$ is $\mathbf{P} = \mathbf{Q}^\mathsf{T}\mathbf{A}$ where $\mathbf{A}$ is positive definite and $\mathbf{Q}$ is orthogonal. Write $\mathbf{P}$'s singular value decomposition (SVD) as $\mathbf{P} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\mathsf{T}$ where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices and $\boldsymbol{\Sigma}$ is a diagonal matrix with non-negative elements σ_i, the singular values, on the diagonal in descending order (note $\mathbf{Q} = \mathbf{V}\mathbf{U}^\mathsf{T}$ and $\mathbf{A} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^\mathsf{T}$).

We then have

$$L(\mathbf{R}) = \Big(\frac{\kappa}{4\pi \sinh \kappa}\Big)^N \exp \kappa \operatorname{Tr}(\boldsymbol{\Sigma}\mathbf{V}^\mathsf{T}\mathbf{R}\mathbf{U}). \tag{1}$$

We have the following theorem of Horn [11, 12]

Theorem 1. *The set D^+ of all diagonals of n-dimensional rotation matrices $SO(n)$ is equal to the convex hull of the set V^+ consisting of points $(\pm 1, \cdots, \pm 1)$ of which an even number (possibly 0) of the coordinates are -1.*

The convex hull of a set of points in n dimensions is the minimal convex set that encloses the points. Note that the points in V^+ are the diagonals of the diagonal rotation matrices. The orthogonal group $O(n)$ has two connected components, $SO(n)$ consisting of those orthogonal matrices with determinant $+1$, and $O^-(n)$ consisting of those othogonal matrices with determinant -1, i.e.,

$$O(n) = SO(n) \cup O^-(n)$$

A direct consequence of Theorem 1 is that the set D^- of diagonals of othogonal matrices with deteminant -1 is the convex hull of the set V^- consisting of those points $(\pm 1, \cdots, \pm 1)$ of which an odd number of the coordinates are -1.

Theorem 2 ([2, 3, 4, 13, 6])**.** *If $\mathbf{R} \in O(n)$ or $\det \mathbf{Q} = 1$ then the unique ML estimate of $\mathbf{R}$ is $\hat{\mathbf{R}}_{ML} = \mathbf{Q}$. Otherwise,*

$$\hat{\mathbf{R}}_{ML} = \mathbf{V} \operatorname{diag}(1, \cdots, 1, -1)\mathbf{U}^\mathsf{T}.$$

which is unique if the smallest singular value of $\mathbf{A}$ is non-degenerate.

Proof. From (1), the ML estimate is the maximiser of

$$\operatorname{Tr}(\boldsymbol{\Sigma}\mathbf{V}^\mathsf{T}\mathbf{R}\mathbf{U}) = \operatorname{Tr}(\boldsymbol{\Sigma}\mathbf{W}) = \mathbf{d}_W \boldsymbol{\sigma}^\mathsf{T} \tag{2}$$

where $\mathbf{W} = \mathbf{V}^\mathsf{T}\mathbf{R}\mathbf{U}$ is orthogonal and $\mathbf{d}_W$ and $\boldsymbol{\sigma}$ are the diagonals of $\mathbf{W}$ and $\boldsymbol{\Sigma}$ as vectors. We have three cases

1. If $\det \mathbf{Q} = 1$ and we are maximising (2) over $\mathbf{R} \in SO(n)$ then since D^+ is a convex hull
$$\mathbf{d}_{\hat{W}_{ML}} = \operatorname*{arg\,max}_{\mathbf{d}_W \in D^+} \mathbf{d}_W \boldsymbol{\sigma}^\mathsf{T} = \operatorname*{arg\,max}_{\mathbf{d}_W \in V^+} \mathbf{d}_W \boldsymbol{\sigma}^\mathsf{T}$$
Thus $\mathbf{d}_{\hat{W}_{ML}} = (1, 1, \cdots, 1)$ and so $\hat{\mathbf{W}}_{ML} = \mathbf{I}$, that is, $\hat{\mathbf{R}}_{ML} = \mathbf{Q}$.

2. If we are maximising (2) over $\mathbf{R} \in O(n)$ then
$$\mathbf{d}_{\hat{W}_{ML}} = \operatorname*{arg\,max}_{\mathbf{d}_W \in D^+ \cup D^-} \mathbf{d}_W \boldsymbol{\sigma}^\mathsf{T} = \operatorname*{arg\,max}_{\mathbf{d}_W \in V^+ \cup V^-} \mathbf{d}_W \boldsymbol{\sigma}^\mathsf{T}$$
Again $\mathbf{d}_{\hat{W}_{ML}} = (1, 1, \cdots, 1)$ and $\hat{\mathbf{R}}_{ML} = \mathbf{Q}$.

3. If $\det \mathbf{Q} = -1$ and we are maximizing (2) over $\mathbf{R} \in SO(n)$ then we must maximize $\mathbf{d}_W \boldsymbol{\sigma}^\mathsf{T}$ over $\mathbf{W} \in O^-(n)$, i.e.,
$$\mathbf{d}_{\hat{W}_{ML}} = \operatorname*{arg\,max}_{\mathbf{d}_W \in D^-} \mathbf{d}_W \boldsymbol{\sigma}^\mathsf{T} = \operatorname*{arg\,max}_{\mathbf{d}_W \in V^-} \mathbf{d}_W \boldsymbol{\sigma}^\mathsf{T}$$
This has a unique maximum $\mathbf{d}_{\hat{W}_{ML}} = (1, 1, \cdots, 1, -1)$ and so $\mathbf{R}_{ML} = \mathbf{V} \operatorname{diag}(1, \cdots, 1, -1)\mathbf{U}^\mathsf{T}$, if the smallest singular value of A is non-degenerate. If the smallest singular value is m-fold degenerate then we have the same maximum value if the single -1 in $\mathbf{d}_{\hat{W}_{ML}}$ occurs in any of the last m positions.

$\square$

3. BAYES ESTIMATOR

We take our loss function as the squared Hilbert-Schmidt distance between $\mathbf{R}$ and $\hat{\mathbf{R}}$, the choice of squared distance is motivated by the fact that it simplifies calculations and is equivalent to $\|\mathbf{R}-\hat{\mathbf{R}}\|_{HS}$.

$$\|\mathbf{R}-\hat{\mathbf{R}}\|_{HS}^2 = \mathrm{Tr}\left((\mathbf{R}-\hat{\mathbf{R}})(\mathbf{R}-\hat{\mathbf{R}})^\top\right).$$

Let G be the group, either $O(n)$ or $SO(n)$, in which $\mathbf{R}$ is an element. The expected loss for the likelihood (1) is

$$\rho(\hat{\mathbf{R}}) = \int_G \|\mathbf{R}-\hat{\mathbf{R}}\|_{HS}^2 p(\mathbf{R}|\mathbf{Y},\mathbf{X})d\mu(\mathbf{R}), \tag{3}$$

where $\mu(\mathbf{R})$ is the Haar measure on G. An application of Bayes theorem, with a uniform prior (with respect to the Haar measure) on G for the orientation, yields the following posterior distribution

$$p(\mathbf{R}|\mathbf{Y},\mathbf{X}) = \frac{1}{\alpha_0(\kappa A)}\exp\kappa\,\mathrm{Tr}(\mathbf{A}\mathbf{R}\mathbf{Q}^\top),$$

with the normalising factor $\alpha_0(\kappa A)$ given by

$$\begin{aligned}\alpha_0(\kappa\mathbf{A}) &= \int_G \exp\kappa\,\mathrm{Tr}(\mathbf{A}\mathbf{R}\mathbf{Q}^\top)d\mu(\mathbf{R})\\ &= \int_G \exp\kappa\,\mathrm{Tr}(\mathbf{A}\mathbf{S})d\mu(\mathbf{S})\end{aligned}$$

where $\mathbf{S}=\mathbf{R}\mathbf{Q}^\top$. We can choose to restrict $\mathbf{R}$ and $\mathbf{Q}$ to be elements of $SO(n)$ and so create a distribution with respect to the Haar measure on that group. It follows that

$$\begin{aligned}\rho(\hat{\mathbf{R}}) = {}& \frac{1}{\alpha_0(\kappa\mathbf{A})}\\ &\int_G [\mathrm{Tr}(\mathbf{R}^\top\mathbf{R}) - \mathrm{Tr}(\mathbf{R}^\top\hat{\mathbf{R}}) - \mathrm{Tr}(\hat{\mathbf{R}}^\top\mathbf{R}) + \mathrm{Tr}(\hat{\mathbf{R}}^\top\hat{\mathbf{R}})]\\ &\exp\kappa\,\mathrm{Tr}(\mathbf{A}\mathbf{R}\mathbf{Q}^\top)\,d\mu(\mathbf{R}).\end{aligned}$$

Given that $\mathrm{Tr}(\mathbf{R}^\top\mathbf{R}) = \mathrm{Tr}(\hat{\mathbf{R}}^\top\hat{\mathbf{R}}) = n$, we see

$$\begin{aligned}\rho(\hat{\mathbf{R}}) &= \frac{1}{\alpha_0(\kappa\mathbf{A})}\\ &\quad\int_G [2n - 2\,\mathrm{Tr}(\hat{\mathbf{R}}^\top\mathbf{R})]\exp\kappa\,\mathrm{Tr}(\mathbf{A}\mathbf{R}\mathbf{Q}^\top)\,d\mu(\mathbf{R})\\ &= 2n - \frac{2}{\alpha_0(\kappa\mathbf{A})}\int_G \mathrm{Tr}(\hat{\mathbf{R}}^\top\mathbf{S}\mathbf{Q})\exp\kappa\,\mathrm{Tr}(\mathbf{A}\mathbf{S})\,d\mu(\mathbf{R}).\end{aligned}$$

Then

$$\rho(\hat{\mathbf{R}}) = 2n - \frac{2\,\mathrm{Tr}(\mathbf{Q}\hat{\mathbf{R}}^\top\alpha_1(\kappa\mathbf{A}))}{\alpha_0(\kappa\mathbf{A})}, \tag{4}$$

where

$$\alpha_1(\kappa\mathbf{A}) = \int_G \mathbf{S}\exp\kappa\,\mathrm{Tr}(\mathbf{A}\mathbf{S})d\mu(\mathbf{S}). \tag{5}$$

Our aim is to now prove

Theorem 3. *If $\mathbf{R}\in O(n)$ or $\det\mathbf{Q}=1$ then the Bayesian estimate of $\mathbf{R}$ is $\hat{\mathbf{R}}_B=\mathbf{Q}$. Otherwise,*

$$\hat{\mathbf{R}}_B = \mathbf{V}\,\mathrm{diag}(1,...,1,-1)\mathbf{U}^\top.$$

To prove the theorem we require the following lemmata.

Lemma 4. *If $\mathbf{A}$ is positive definite with $\mathbf{A}=\mathbf{V}\mathbf{\Sigma}\mathbf{V}^\top$ then $\alpha_1(\kappa\mathbf{A}) = \mathbf{V}\alpha_1(\kappa\mathbf{\Sigma})\mathbf{V}^\top$.*

Proof. Accordingly, we write

$$\alpha_1(\kappa\mathbf{A}) = \int_G \mathbf{S}\exp\kappa\,\mathrm{Tr}(\mathbf{V}\mathbf{\Sigma}\mathbf{V}^\top\mathbf{S})d\mu(\mathbf{S}),$$

by a change of variable $\mathbf{T}=\mathbf{V}^\top\mathbf{S}\mathbf{V}$

$$\alpha_1(\kappa\mathbf{A}) = \mathbf{V}\int_G \mathbf{T}\exp\kappa\,\mathrm{Tr}(\mathbf{\Sigma}\mathbf{T})d\mu(\mathbf{T})\mathbf{V}^\top,$$

so

$$\alpha_1(\kappa\mathbf{A}) = \mathbf{V}\alpha_1(\kappa\mathbf{\Sigma})\mathbf{V}^\top.$$

□

Lemma 5. *Any matrix $\mathbf{U}\in SO(n)$ that commutes with $\mathbf{A}$ commutes with $\alpha_1(\kappa\mathbf{A})$.*

Proof. Consider

$$\begin{aligned}\mathbf{U}^{-1}\alpha_1(\kappa\mathbf{A})\mathbf{U} &= \int_{SO(n)} \mathbf{U}^{-1}\mathbf{S}\mathbf{U}\exp\kappa\,\mathrm{Tr}(\mathbf{A}\mathbf{S})d\mu(\mathbf{S})\\ &= \int_{SO(n)} \mathbf{S}\exp\kappa\,\mathrm{Tr}(\mathbf{A}\mathbf{U}\mathbf{S}\mathbf{U}^{-1})d\mu(\mathbf{S})\end{aligned}$$

by left and right invariance of the Haar measure on $SO(n)$. This equals

$$\begin{aligned}\mathbf{U}^{-1}\alpha_1(\kappa\mathbf{A})\mathbf{U} &= \int_{SO(n)} \mathbf{S}\exp\kappa\,\mathrm{Tr}(\mathbf{U}^{-1}\mathbf{A}\mathbf{U}\mathbf{S})d\mu(\mathbf{S})\\ &= \int_{SO(n)} \mathbf{S}\exp\kappa\,\mathrm{Tr}(\mathbf{A}\mathbf{S})d\mu(\mathbf{S}) = \alpha_1(\kappa\mathbf{A})\end{aligned}$$

because $\mathbf{U}$ commutes with $\mathbf{A}$. Hence $\mathbf{U}$ commutes with $\alpha_1(\kappa\mathbf{A})$. □

Lemma 6. *If $\mathbf{\Sigma}$ is diagonal then $\alpha_1(\kappa\mathbf{\Sigma})$ is diagonal also.*

Proof. By Lemma 5, every diagonal element of $SO(n)$ commutes with $\alpha_1(\kappa\mathbf{\Sigma})$. Diagonal elements of $SO(n)$ have ±1 on the diagonal with the number of -1's even. Also the identity matrix trivially commutes with $\alpha_1(\kappa\mathbf{\Sigma})$, and the commutation property is closed under linear combinations. Taking simple linear combinations, we see firstly that all diagonal matrices whose diagonals have two non-zero entries which are equal commute with $\alpha_1(\kappa\mathbf{\Sigma})$. Then, taking differences of these we obtain that each matrix whose diagonal is a permutation of the vector $(1,-1,0,...,0)$ commutes with $\alpha_1(\kappa\mathbf{\Sigma})$. Since vectors of the last form together with the vector $(1,1,...,1)$ form a basis of $\mathbb{R}^n$, it follows that all diagonal matrices commute with $\alpha_1(\kappa\mathbf{\Sigma})$ and hence $\alpha_1(\kappa\mathbf{\Sigma})$ is diagonal. □

Lemma 7. *The eigenvalues of $\alpha_1(\kappa\mathbf{\Sigma})$ are ordered in the same way as the eigenvalues of $\mathbf{\Sigma}$.*

Proof. Consider the integral

$$\alpha_1(\mathbf{A}) = \int_{SO(n)} \mathbf{R} \exp \operatorname{Tr}(\mathbf{AR}) d\mu(\mathbf{R})$$

Assume that A is diagonal and that its diagonal entries are $a_1 \geq a_2 \geq a_3 \cdots \geq a_n$. Our aim is to show that the diagonal entries of $\alpha_1(\mathbf{A})$, $z_1, z_2, \ldots, z_n$, are ordered in the same way as those of $\mathbf{A}$. In fact, it will be enough to show that $z_1 \geq z_2$, as we can permute the variables using a conjugation by a permutation matrix that will bring any pair of indicies i and j to the positions 1 and 2. We write

$$\tau(\tilde{\mathbf{r}}) = \tau(r_3, r_4, \ldots, r_n) = \exp \sum_{k=3}^{n} a_k r_{kk}$$

and $\kappa(r_{11}, r_{22}, \ldots, r_{nn})$ is the density of the marginalisation of the Haar measure to the diagonal so that

$$\begin{aligned} & \int_{SO(n)} f(r_{11}, r_{22}, \ldots, r_{nn}) d\mu(\mathbf{R}) \\ = \; & \int_{\mathbf{r} \in [-1,1]^n} f(r_{11}, r_{22}, \ldots, r_{nn}) \kappa(r_{11}, r_{22}, \ldots, r_{nn}) \\ & \qquad dr_{11}\, dr_{22}\, \ldots\, dr_{nn}. \end{aligned}$$

Notice that that κ is invariant with respect to the permutation of the elements in $\mathbf{R}$, since if $\mathbf{R} \in SO(n)$ has diagonal $\mathbf{r}$ then $\Pi\mathbf{R}$ has diagonal $\Pi\mathbf{r}$ where $\Pi\mathbf{R} \in SO(n)$. Hence $\tau(\tilde{\mathbf{r}})\kappa(\mathbf{r})$ is invariant under the permutation $r_{11} \leftrightarrow r_{22}$. Now write

$$z_i = \int_{\mathbf{W}} r_{ii} \exp(a_1 r_{11} + a_2 r_{22}) \tau(\tilde{\mathbf{r}}) \kappa(\mathbf{r}) dr_{11} dr_{22} d\tilde{\mathbf{r}} \qquad (i = 1, 2).$$

Consider

$$\begin{aligned} & z_1 - z_2 = \\ & \int_{\mathbf{W}} (r_{11} - r_{22}) \exp(a_1 r_{11} + a_2 r_{22}) \tau(\tilde{\mathbf{r}}) \kappa(\mathbf{r}) dr_{11} dr_{22} d\tilde{\mathbf{r}} \\ = & \int_{\mathbf{W}^+} + \int_{\mathbf{W}^-} r_{11} \exp(a_1 r_{11} + a_2 r_{22}) \tau(\tilde{\mathbf{r}}) \kappa(\mathbf{r}) dr_{11} dr_{22} d\tilde{\mathbf{r}} \\ - & \int_{\mathbf{W}^+} - \int_{\mathbf{W}^-} r_{22} \exp(a_1 r_{11} + a_2 r_{22}) \tau(\tilde{\mathbf{r}}) \kappa(\mathbf{r}) dr_{11} dr_{22} d\tilde{\mathbf{r}} \end{aligned}$$

where $\mathbf{W}^+ = \{\mathbf{r} : r_{11} > r_{22}\}$ and $\mathbf{W}^- = \{\mathbf{r} : r_{11} < r_{22}\}$. Reorganising this and permuting variables $r_{11} \leftrightarrow r_{22}$ we have

$$\begin{aligned} & z_1 - z_2 \\ = & \int_{\mathbf{W}^+} r_{11} \exp(a_1 r_{11} + a_2 r_{22}) \tau(\tilde{\mathbf{r}}) \kappa(\mathbf{r}) dr_{11} dr_{22} d\tilde{\mathbf{r}} \\ & - \int_{\mathbf{W}^+} r_{11} \exp(a_1 r_{22} + a_2 r_{11}) \tau(\tilde{\mathbf{r}}) \kappa(\mathbf{r}) dr_{11} dr_{22} d\tilde{\mathbf{r}} \\ & + \int_{\mathbf{W}^-} r_{11} \exp(a_1 r_{11} + a_2 r_{22}) \tau(\tilde{\mathbf{r}}) \kappa(\mathbf{r}) dr_{11} dr_{22} d\tilde{\mathbf{r}} \\ & - \int_{\mathbf{W}^-} r_{11} \exp(a_1 r_{22} + a_2 r_{11}) \tau(\tilde{\mathbf{r}}) \kappa(\mathbf{r}) dr_{11} dr_{22} d\tilde{\mathbf{r}} \end{aligned}$$

$$\begin{aligned} = & \int_{\mathbf{W}^+} r_{11} (\exp(a_1 r_{11} + a_2 r_{22}) - \exp(a_1 r_{22} + a_2 r_{11})) \\ & \qquad \tau(\tilde{\mathbf{r}}) \kappa(\mathbf{r}) dr_{11} dr_{22} d\tilde{\mathbf{r}} \\ & + \int_{\mathbf{W}^+} r_{22} (\exp(a_1 r_{22} + a_2 r_{11}) - \exp(a_1 r_{11} + a_2 r_{22})) \\ & \qquad \tau(\tilde{\mathbf{r}}) \kappa(\mathbf{r}) dr_{11} dr_{22} d\tilde{\mathbf{r}} \\ = & \int_{\mathbf{W}^+} (r_{11} - r_{22}) (\exp(a_1 r_{22} + a_2 r_{11}) - \exp(a_1 r_{11} + a_2 r_{22})) \\ & \qquad \tau(\tilde{\mathbf{r}}) \kappa(\mathbf{r}) dr_{11} dr_{22} d\tilde{\mathbf{r}}. \end{aligned}$$

Since the integral is over $\mathbf{W}^+$

$$\begin{aligned} & (a_1 r_{11} + a_2 r_{22}) - (a_1 r_{22} + a_2 r_{11}) \\ & = (a_1 - a_2)(r_{11} - r_{22}) \geq 0, \end{aligned}$$

the integrand is positive and therefore $z_1 \geq z_2$. ☐

We can now prove Theorem 3.

Proof. The expected loss (4) is minimised by maximising

$$\operatorname{Tr}(\mathbf{Q}\hat{\mathbf{R}}^\top \alpha_1(\kappa\mathbf{A})).$$

Applying Lemma 4 we have

$$\begin{aligned} \operatorname{Tr}(\mathbf{Q}\hat{\mathbf{R}}^\top \alpha_1(\kappa\mathbf{A})) &= \operatorname{Tr}(\mathbf{Q}\hat{\mathbf{R}}^\top \mathbf{V} \alpha_1(\kappa\mathbf{\Sigma}) \mathbf{V}^\top) \\ &= \operatorname{Tr}(\mathbf{V}^\top \mathbf{Q}\hat{\mathbf{R}}^\top \mathbf{V} \alpha_1(\kappa\mathbf{\Sigma})) \\ &= \operatorname{Tr}(\mathbf{W} \alpha_1(\kappa\mathbf{\Sigma})) \\ &= \mathbf{d}_W \boldsymbol{\alpha}^\top \end{aligned} \tag{6}$$

where $\mathbf{W} = \mathbf{V}^\top \mathbf{Q}\hat{\mathbf{R}}^\top \mathbf{V}$ is orthogonal, $\alpha_1(\kappa\mathbf{\Sigma})$ is diagonal from Lemma 6 and $\mathbf{d}_W$ and $\boldsymbol{\alpha}$ are the diagonals of $\mathbf{W}$ and $\alpha_1(\kappa\mathbf{\Sigma})$ as vectors.

By a similar argument to Theorem 2, we have three cases

1. If $\det Q = 1$ and we are maximising (6) over $\mathbf{R} \in SO(n)$ then since D^+ is a convex hull

$$\mathbf{d}_{\hat{W}_B} = \operatorname*{arg\,max}_{\mathbf{d}_W \in D^+} \mathbf{d}_W \boldsymbol{\alpha}^\top = \operatorname*{arg\,max}_{\mathbf{d}_W \in V^+} \mathbf{d}_W \boldsymbol{\alpha}^\top$$

   Thus $\mathbf{d}_{\hat{W}_B} = (1, 1, \cdots, 1)$ and so $\hat{\mathbf{W}}_B = \mathbf{I} = \mathbf{V}^\top \mathbf{Q} \hat{\mathbf{R}}_B^\top \mathbf{V}$, that is, $\hat{\mathbf{R}}_B = \mathbf{Q}$.

2. If we are maximising (6) over $\mathbf{R} \in O(n)$ then

$$\mathbf{d}_{\hat{W}_B} = \operatorname*{arg\,max}_{\mathbf{d}_W \in D^+ \cup D^-} \mathbf{d}_W \boldsymbol{\alpha}^\top = \operatorname*{arg\,max}_{\mathbf{d}_W \in V^+ \cup V^-} \mathbf{d}_W \boldsymbol{\alpha}^\top$$

   Again $\mathbf{d}_{\hat{W}_B} = (1, 1, \cdots, 1)$ and $\hat{\mathbf{R}}_B = \mathbf{Q}$.

3. If $\det \mathbf{Q} = -1$ and we are maximizing (6) over $\mathbf{R} \in SO(n)$ then we must maximize $\mathbf{d}_W \boldsymbol{\alpha}^\top$ over $\mathbf{W} \in O^-(n)$, i.e.,

$$\mathbf{d}_{\hat{W}_{ML}} = \operatorname*{arg\,max}_{\mathbf{d}_W \in D^-} \mathbf{d}_W \boldsymbol{\alpha}^\top = \operatorname*{arg\,max}_{\mathbf{d}_W \in V^-} \mathbf{d}_W \boldsymbol{\alpha}^\top$$

This has a unique maximum $\mathbf{d}_{\hat{W}_B} = (1, 1, \cdots, 1, -1)$ (under the same conditions mentioned in Theorem 2) as α has elements in decreasing order according to Lemma 7. The Bayes estimator if $\det \mathbf{Q} = -1$ or $\mathbf{R} \in SO(n)$ is $\hat{\mathbf{R}} = \mathbf{V} \operatorname{diag}(1, ..., 1, -1)\mathbf{V}^\mathsf{T}\mathbf{Q}$, where $\mathbf{Q} = \mathbf{V}\mathbf{U}^\mathsf{T}$ thus $\hat{\mathbf{R}}_\mathrm{B} = \mathbf{V} \operatorname{diag}(1, ..., 1, -1)\mathbf{U}^\mathsf{T}$.

□

4. THE BEST INVARIANT ESTIMATOR

The density $p(\mathbf{Y}|\mathbf{R}, \mathbf{X})$ is invariant under the tranformation

$$\begin{aligned} \mathbf{y}_j \to \mathbf{y}'_j &= \mathbf{S}\mathbf{y}_j, \quad j = 1, \cdots, N \\ \mathbf{R} \to \mathbf{R}' &= \mathbf{R}\mathbf{S}^\mathsf{T} \\ p(\mathbf{Y}|\mathbf{R}, \mathbf{X}) &= p(\mathbf{S}\mathbf{Y}|\mathbf{R}\mathbf{S}^\mathsf{T}, \mathbf{X}). \end{aligned} \tag{7}$$

With the loss function the Hilbert-Schmidt squared distance, as previously discussed, between $\mathbf{R}$ and $\hat{\mathbf{R}}$ the expected risk function $\mathcal{R}$ for the estimator $\hat{\mathbf{R}} : \mathbb{R}^{n\times N} \to SO(n)$ is

$$\mathcal{R}(\mathbf{R}, \hat{\mathbf{R}}) = \int_Y \|\mathbf{R} - \hat{\mathbf{R}}(\mathbf{Y})\|^2 p(\mathbf{Y}|\mathbf{R}, \mathbf{X}) d\mathbf{Y}. \tag{8}$$

An estimator $\hat{\mathbf{R}}$ is invariant if under the transformation (7) we have

$$\hat{\mathbf{R}}(\mathbf{Y}') = \hat{\mathbf{R}}(\mathbf{Y})\mathbf{S}^\mathsf{T}.$$

For an invariant estimator the risk function (8) is a constant function of $\mathbf{R}$ since

$$\begin{aligned} \mathcal{R}(\mathbf{R}, \hat{\mathbf{R}}) &= \int_Y \|\mathbf{R} - \hat{\mathbf{R}}(\mathbf{Y})\|^2 p(\mathbf{Y}|\mathbf{R}, \mathbf{X}) d\mathbf{Y} \\ &= \int_Y \|\mathbf{I} - \hat{\mathbf{R}}(\mathbf{Y})\mathbf{R}^\mathsf{T}\|^2 p(\mathbf{Y}|\mathbf{R}, \mathbf{X}) d\mathbf{Y} \\ &= \int_Y \|\mathbf{I} - \hat{\mathbf{R}}(\mathbf{R}\mathbf{Y})\|^2 p(\mathbf{Y}|\mathbf{R}, \mathbf{X}) d\mathbf{Y} \\ &= \int_Y \|\mathbf{I} - \hat{\mathbf{R}}(\mathbf{Y}')\|^2 p(\mathbf{R}^\mathsf{T}\mathbf{Y}'|\mathbf{R}, \mathbf{X}) d\mathbf{Y}' \\ &= \int_Y \|\mathbf{I} - \hat{\mathbf{R}}(\mathbf{Y}')\|^2 p(\mathbf{Y}'|\mathbf{I}, \mathbf{X}) d\mathbf{Y}' \\ &= \mathcal{R}(\mathbf{I}, \hat{\mathbf{R}}). \end{aligned}$$

Hence we can find the best invariant estimator by minimizing the constant risk

$$\begin{aligned} \mathcal{R}(\mathbf{R}, \hat{\mathbf{R}}) &= \frac{1}{8\pi^2} \int_{SO(3)} \mathcal{R}(\mathbf{R}, \hat{\mathbf{R}}) d\mu(\mathbf{R}) \\ &= \frac{1}{8\pi^2} \int_{SO(3)} \int_Y \|\mathbf{R} - \hat{\mathbf{R}}(\mathbf{y})\|^2 p(\mathbf{y}|\mathbf{R}, \mathbf{x}) d\mathbf{y} d\mu(\mathbf{R}) \\ &= \frac{1}{8\pi^2} \int_Y \int_{SO(3)} \|\mathbf{R} - \hat{\mathbf{R}}(\mathbf{y})\|^2 p(\mathbf{y}|\mathbf{R}, \mathbf{x}) d\mu(\mathbf{R}) d\mathbf{y} \\ &= \int_Y \alpha_0(\kappa \mathbf{A}(\mathbf{y})) \rho(\hat{\mathbf{R}}(\mathbf{y})) d\mathbf{y} \end{aligned}$$

.

We can exhange the order of the integrals above since they are over compact spaces, so that, by Bayes theorem $p(\mathbf{y}|\mathbf{R}, \mathbf{x}) = \alpha_0(\kappa \mathbf{A}(\mathbf{y})) p(\mathbf{R}|\mathbf{y}, \mathbf{x})$. Since $\alpha_0(\kappa \mathbf{A}(\mathbf{y}))$ is ≥ 0 and independent of our choice of estimator, the Bayes estimator (Theorem 3) which minimises the expected loss $\rho(\hat{\mathbf{R}})$ is also a best invariant estimator.

5. CONCLUSION

In this paper we have discussed the Mackenzie-Wahba estimator, a known ML orientation estimator, and have shown it to be both a Bayes estimator and a best invariant estimator. This was shown not only in 2 and 3 dimensions but generally for $O(n)$ and $SO(n)$.

6. REFERENCES

[1] Ted Chang, "Spherical regression," *Ann. Stat.*, vol. 14, no. 3, pp. 907–924, 1986.

[2] J. K. Mackenzie, "The estimation of an orientation relationship," *Acta Cryst.*, vol. 10, pp. 61–62, 1957.

[3] Grace Wahba, "Problem 65–1, a least squares estimate of satellite attitude," *SIAM Rev.*, vol. 7, no. 3, pp. 409, July 1965.

[4] J. L. Farrell, J. C. Stuelpnagel, R. H. Wessner, and J. R. Velman, "A least squares estimate of satellite attitude (Grace Wahba)," *SIAM Rev.*, vol. 8, no. 3, pp. 384–386, July 1966.

[5] J. Keat, "Analysis of least squares attitude determination routine (DOAOP)," Tech. Rep. CSC/TM-77/6034, Computer Sciences Corporation, 1977.

[6] Malcolm D. Shuster, "Maximum likelihood estimation of spacecraft attitude," *J. Astronautical Sci.*, vol. 37, no. 1, pp. 78–88, Jan.–Mar. 1989.

[7] E. J. Lefferts, F. L. Markley, and M. D. Shuster, "Kalman filtering for spacecraft attitude estimation," *J. Guidance Control Dynamics*, vol. 5, no. 5, pp. 417–429, Sep.–Oct. 1982.

[8] Malcolm D. Shuster, "The generalized Wahba problem," *J. Astronautical Sci.*, vol. 54, no. 1, pp. 113–123, Jan.–Mar. 2006.

[9] Ted Chang and Christopher Bingham, "Bayesian spherical regression," in *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*, Donald A. Berry, Kathryn M. Chaloner, and John Geweke, Eds., chapter 23. John Wiley & Sons, 1996.

[10] K. V. Mardia and P. E. Jupp, *Directional statistics*, John Wiley & Sons, 2nd edition, 2000.

[11] Alfred Horn, "Doubly stochastic matrices and the diagonal of a rotation matrix," *American Journal of Mathematics*, vol. 76, no. 3, pp. 620–630, July 1954.

[12] Alberto Borobia, "Diagonals of rotation matrices," *Linear Algebra and its Applications*, vol. 186, pp. 227 – 233, 1993.

[13] M. A. Stephens, "Vector correlation," *Biometrika*, vol. 66, no. 1, pp. 41–48, 1979.