An Information-Theoretic Approach to the Correspondence-Free AX=XB Sensor Calibration Problem

Martin Kendal Ackerman[†], Alexis Cheng[‡], Gregory Chirikjian[†]

Abstract

For the case of an exact set of compatible A's and B's with known correspondence, the AX=XB problem was solved decades ago. However, in many applications, data streams containing the A's and B's will often have different sampling rates or will be asynchronous. For these reasons and the fact that each stream may contain gaps in information, methods that require minimal a priori knowledge of the correspondence between A's and B's would be superior to the existing algorithms that require exact correspondence. We present an information-theoretic algorithm for recovering X from a set of A's and a set of B's that does not require a priori knowledge of correspondences. The algorithm views the problem in terms of distributions on the group $SE(3)$, and minimizing the Kullback-Leibler divergence of these distributions with respect to the unknown X. This minimization is performed by an efficient numerical procedure that reliably recovers an unknown X.

1. INTRODUCTION

The sensor calibration problem, stated mathematically as $AX = XB$, appears in many different guises in the fields of robotics and computer vision. Applications range from wrist-mounted sensors on manipulators, to cameras mounted on aerial vehicles, to ultrasound (US) probe calibration in medical robotics. The variables A, X, and B are each rigid-body motions (i.e., homogeneous transformations) with each pair of measurements (A,B) coming from sensors such as cameras, US probes, optical or electromagnetic pose tracking systems, etc., and X is the unknown rigid-body motion that is found as a result of solving $AX = XB$.

Any proper motion in Euclidean space can be described as a homogeneous transformation matrix of the form

$$H(R,\mathbf{t}) = \begin{pmatrix} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix}. \quad (1)$$

Here $\mathbf{t}$ is a translation vector and R is a rotation matrix. In the three-dimensional case $\mathbf{t} \in \mathbb{R}^3$ and R is an element of the group of 3×3 special orthogonal matrices, $SO(3)$. Matrices of the form in (1) are faithful matrix representations of the "Special Euclidean Group" $SE(3)$, which has matrix multiplication as its group operation.

Given multiple equations of the form

$$A_i X = X B_i \implies A_i = X B_i X^{-1} \quad (2)$$

for $(A_i,B_i) \in SE(3) \times SE(3)$ and $i \in \{1,...,n\}$, a solution $X \in SE(3)$ is sought in a wide variety of applications [1]- [7]. In fact, under some mild conditions, [8], the problem can be solved uniquely for $n = 2$. Constructive algorithms for solving this problem (in the case when correspondences are known a priori) date back to the 1980s [3]- [6].

In this paper we present a method to solve for an X wherein there does not need to be any a priori knowledge of the correspondence between the sets $\{A_i\}$ and $\{B_j\}$. That is not to say that such a correspondence does not exist between the continuous time trajectories $A(t)$ and $B(t)$. We assume that the correspondence exists, but is unknown, and is lost when these trajectories are "out of sync".

In typical applications involving the $AX = XB$ calibration problem, the sensors used to generate data will often have different internal delays and activation times that can cause the data streams to arrive at a central computor in an asynchronous fashion. It is also possible for the sensors to have different sampling rates that result in need for interpolation or may suffer errors that lead to gaps

This work was supported by NSF Grant RI-Medium: 1162095

[†]Martin Kendal Ackerman and Gregory Chirikjian are with the Department of Mechanical Engineering and Laboratory for Computational Sensing and Robotics, Johns Hopkins University, Baltimore, MD, USA, gregc at jhu.edu

[‡]Alexis Cheng is with the Department of Computer Science, and Laboratory for Computational Sensing and Robotics, Johns Hopkins University, Baltimore, MD, USA

in the data [9]. For all of these instances, loss of correspondence is manifested in a data stream as either 1) a number of A_i's that do not have corresponding B_j's (the data is missing completely due to error or need for interpolation); or, 2) the B_j for corresponding A_i is located at a different point in the stream (the time relationship of the data is not available).

The approach that we take is to view the two sets of reference frames $\{A_i\}$ and $\{B_j\}$ as histograms with corresponding probability densities $f_A(H)$ and $f_B(H)$ where $H \in SE(3)$. We view X also as a probability density on $SE(3)$ – a Dirac delta distribution shifted to have its peak at X, $\delta_X(H) = \delta(X^{-1}H)$. We then seek X as the solution to a minimization problem wherein the cost is

$$C(X) = D_{KL}(f_A \,\|\, \delta_X * f_B * \delta_{X^{-1}}) \tag{3}$$

where $*$ denotes convolution on $SE(3)$, as described in [10, 11] and $D_{KL}(\cdot\,\|\,\cdot)$ is the Kullback-Leibler divergence for $SE(3)$ [11–13].

The remainder of this paper is devoted to establishing the necessary mathematical machinery to define $f_A(H)$ and $f_B(H)$, explicitly computing $D_{KL}(\cdot\,\|\,\cdot)$ in closed form for special kinds of Gaussian functions on $SE(3)$, and performing the minimization required to find X.

Section 2 formulates the problem in terms of Gaussian distributions on $SE(3)$ using concepts of means and covariances of sets of reference frames. Section 3 solves the problem using our information-theoretic method. Section 4 demonstrates the performance of this new method with simulated and experimental data.

2. Probabilistic Preliminaries

The Lie group $SE(3)$ has the geometric structure of a Riemannian manifold, and hence there is a natural concept of integration. Here we use these properties, and the even more special characteristics of $SE(3)$ as a Lie group.

2.1. Convolution and Dirac Deltas

The convolution of two probability density functions on $SE(3)$ is defined as

$$(f_1 * f_2)(H') = \int_{SE(3)} f_1(H) f_2(H^{-1}H')dH \tag{4}$$

where in this equation $H, H' \in SE(3)$ with H serving as a dummy variable of integration for each value of H'. Here dH is the natural (bi-invariant) Riemannian volume element for $SE(3)$ [10, 11]. Such convolutions have been shown previously to arise in the characterization of manipulator workspaces [14, 15], pose probabilities in SLAM [16, 17], and even in the statistical mechanics of DNA [18]. The concept of $SE(3)$-convolution is also central to our new correspondence-free formulation of the sensor-calibration problem.

A Dirac delta distribution can be defined for $SE(3)$ to have the properties

$$\int_{SE(3)} \delta(H)dH = 1 \text{ and } (f * \delta)(H) = f(H).$$

A shifted Dirac delta distribution can be defined as $\delta_X(H) = \delta(X^{-1}H)$ which places the spike at $X \in SE(3)$.

The equation in (2) can be written as

$$\delta_{A_i}(H) = (\delta_X * \delta_{B_i} * \delta_{X^{-1}})(H). \tag{5}$$

This more complicated expression has the benefit that addition of functions, $f_1(H) + f_2(H)$, makes sense, whereas addition of homogeneous transformations does not. And since convolution is a linear operation on functions, we can write all n instances of (5) into a single equation of the form

$$f_A(H) = (\delta_X * f_B * \delta_{X^{-1}})(H) \tag{6}$$

where

$$f_A(H) = \frac{1}{n}\sum_{i=1}^{n} \delta(A_i^{-1}H) \text{ and } f_B(H) = \frac{1}{n}\sum_{i=1}^{n} \delta(B_i^{-1}H).$$

(In fact, in this formulation the value of n can be different for the A's and the B's, and we don't need knowledge of how the entries in these sets correspond, which is one of the strengths of the method.) When written in this way, it does not matter if we know the correspondence between each A_i and B_j or not. The above functions are normalized to be probability densities.

Of course, the functions $f_A(H)$ and $f_B(H)$ are not in $L^2(SE(3))$, but we can get around this technical difficulty if each discrete set is replaced with a Gaussian distribution with the same mean and covariance, as defined below.

2.2. Definitions of Mean and Covariance

In what follows, it will be convenient to define the subset $SE_<(3) \subset SE(3)$ which is a depleted version of $SE(3)$ in which all screw motions having a rotation angle of π have been removed. Since

this removal of a (five-dimensional) set of measure zero from the six-dimensional space $SE(3)$ has no effect on the value of integrals of continuous functions, integrals over $SE_<(3)$ and $SE(3)$ will be used interchangeably.

Let the mean and covariance of a probability density $f(H)$ be defined by the conditions [14,15]

$$\int_{SE(3)} \log(M^{-1}H) f(H) dH = \mathbb{O} \quad \text{and}$$

$$\Sigma = \int_{SE(3)} \log^\vee(M^{-1}H)[\log^\vee(M^{-1}H)]^T f(H) dH \quad (7)$$

where the $\vee$ operator that converts 4×4 infinitesimal screw matrices into 6-vectors, now standard in the field of robotics, is defined as in [11] (and references therein).

Explicit formulas defining the matrix exponential $\exp : se(3) \longrightarrow SE(3)$ and logarithm $\log : SE_<(3) \longrightarrow se(3)$, and $\log^\vee : SE_<(3) \to \mathbb{R}^6$ are given in [10,11,15], and are reviewed below. (The transpose on the second $\log^\vee(H)$ in the above definition ensures that Σ is a 6×6 symmetric matrix.) The operation $\log(H)$ takes any element in $H \in SE_<(3)$ (with rotational part that has an angle of rotation, θ, in the range $0 \le \theta < \pi$) into the the corresponding unique element in the Lie algebra $se(3)$ such that $\exp(\log(H)) = H$. That is, the logarithm map is not surjective, unless we consider the target space to be $se_<(3) \subset se(3)$, which can be thought of as the Cartesian product of the open solid ball of radius π with $\mathbb{R}^3$. Since $SO(3)$ can be viewed as a solid closed three-dimensional ball of radius π with antipodal points identified, the exclusion of the bounding sphere of radius π in $SE(3)$ defines a 5D set of measure zero that has no effect on the computation of Σ in the above equation.

With the above caveats, we can write

$$\log(H) = \begin{pmatrix} \Omega & \mathbf{v} \\ \mathbf{0}^T & 0 \end{pmatrix}$$

where $\Omega = -\Omega^T \in so(3)$. The map $\vee : se(3) \to \mathbb{R}^6$ is then composed with the log to give $\mathbf{z} = \log^\vee(H) = [\omega^T, \mathbf{v}^T]^T \in \mathbb{R}^6$ where $\omega \in \mathbb{R}^3$ is the vector corresponding to Ω such that $\Omega\mathbf{x} = \omega \times \mathbf{x}$ for any $\mathbf{x} \in \mathbb{R}^3$, where $\times$ is the vector cross product.

If $f(H)$ is a sum of Dirac deltas like $f_A(H)$ given above, then this has the effect of sampling the integrals, resulting in

$$\sum_{i=1}^{n} \log(M_A^{-1}A_i) = \mathbb{O} \quad \text{and}$$

$$\Sigma_A = \frac{1}{n}\sum_{i=1}^{n} \log^\vee(M_A^{-1}A_i)[\log^\vee(M_A^{-1}A_i)]^T. \quad (8)$$

An iterative procedure for computing M_A has been developed [15].

The adjoint matrix

$$Ad(H) = \begin{pmatrix} R & \mathbb{O} \\ \widehat{\mathbf{t}}R & R \end{pmatrix} \quad (9)$$

will be used heavily in the computations that follow since it has the property that

$$\log^\vee(H_1 H_2 H_1^{-1}) = Ad(H_1)\log^\vee(H_2).$$

In (9) we use the following notion. For any $\mathbf{a} \in \mathbb{R}^3$, $\widehat{\mathbf{a}}$ is the skew-symmetric matrix such that $\widehat{\mathbf{a}}\mathbf{b} = \mathbf{a}\times\mathbf{b}$. By a slight abuse of notation, we reuse $\vee$ as the reverse map which gives $(\widehat{\mathbf{a}})^\vee = \mathbf{a}$. The use of $\vee$ to denote maps from the Lie algebra $so(3)$ into $\mathbb{R}^3$ and from $se(3)$ into $\mathbb{R}^6$ should not be a source of confusion, as the version used is defined by the argument to which it is applied.

2.3. Constraints Imposed By Sensor Calibration

Evaluating the mean and covariance defined in (7) with the functions in (6) and using the bi-invariance of the integration measure gives

$$M_A = X M_B X^{-1} \quad \text{and} \quad (10)$$

$$\Sigma_A = Ad(X)\Sigma_B Ad^T(X). \quad (11)$$

This result is nonparametric in the sense that no assumptions were made about the underlying probability densities. Note that unlike in the covariance propagation problem addressed in [15], these formulas are exact and do not require any approximations. And taking the trace of both sides of (10) gives that the angle of rotation around the screw axes of M_A and M_B must be the same, $\theta_A = \theta_B$. This is one of the two invariants for $SE(3)$ [8,9]. The other one will not be used here.

One can imagine a number of approaches to solving the simultaneous equations (10) and (11). In particular, the search for an appropriate X can begin with (10), which can be rewritten as

$$\log^\vee(M_A) = Ad(X)\log^\vee(M_B). \quad (12)$$

In the case of general M_A and M_B (i.e., not degenerate cases in which the rotation angle [1] is outside of

[1] This angle is computed from the Frobenius norm $\theta_A = \|\frac{1}{2}\log R_A\| = \|\frac{1}{2}\log R_B\| = \theta_B$.

the range $(0,\pi)$, the solution space of all possible X's that satisfy this equation is known to be two dimensional [3,5]. This can be seen by defining

$$\log^{\vee}(M_A) = \begin{pmatrix} \theta_A \mathbf{n}_A \\ \mathbf{v}_A \end{pmatrix}.$$

Then (12) can be broken up into rotational and translational parts as

$$\mathbf{n}_A = R_X \mathbf{n}_B \quad \text{and} \tag{13}$$

$$\mathbf{v}_A = \theta_B \widehat{\mathbf{t}_X} R_X \mathbf{n}_B + R_X \mathbf{v}_B. \tag{14}$$

The first of these equations has a one-dimensional solution space of the form [5] $R_X = R(\mathbf{n}_A, \mathbf{n}_B) R(\mathbf{n}_B, \phi)$ where $\phi \in [0, 2\pi)$ is free and $R(\mathbf{n}_A, \mathbf{n}_B)$ is any rotation matrix that rotates the vector $\mathbf{n}_B$ into $\mathbf{n}_A$. In particular, we can choose

$$R(\mathbf{n}_A, \mathbf{n}_B) = \mathbb{I} + \widehat{\mathbf{n}_B \times \mathbf{n}_A} + \frac{(1 - \mathbf{n}_B \cdot \mathbf{n}_A)}{\|\mathbf{n}_B \times \mathbf{n}_A\|^2} \left(\widehat{\mathbf{n}_B \times \mathbf{n}_A}\right)^2. \tag{15}$$

The rotation $R(\mathbf{n}_B, \phi)$ is given by Euler's formula

$$R(\mathbf{n}_B, \phi) = \mathbb{I} + \sin\phi\, \widehat{\mathbf{n}_B} + (1 - \cos\phi)\, (\widehat{\mathbf{n}_B})^2.$$

Substituting $R_X = R(\mathbf{n}_A, \mathbf{n}_B) R(\mathbf{n}_B, \phi)$ into (14) and rearranging terms, we get

$$\frac{R(\mathbf{n}_A, \mathbf{n}_B) R(\mathbf{n}_B, \phi) \mathbf{v}_B - \mathbf{v}_A}{\theta_B} = \widehat{\mathbf{n}_A} \mathbf{t}_X. \tag{16}$$

The skew-symmetric matrix $\widehat{\mathbf{n}_A}$ has rank 2, so a free translational degree of freedom exists in $\mathbf{t}_X$ along the $\mathbf{n}_A$ direction. $\mathbf{t}_X$ can thus be described as

$$\mathbf{t}_X = \mathbf{t}(s) = s\,\mathbf{n}_A + a\,\mathbf{m}_A + b\,\mathbf{m}_A \times \mathbf{n}_A \tag{17}$$

where $s \in \mathbb{R}$ is a second free parameter, $\mathbf{m}_A$ and $\mathbf{m}_A \times \mathbf{n}_A$ are defined to be orthogonal to $\mathbf{n}_A$ by construction. If $\mathbf{n}_A = [n_1, n_2, n_3]^T$ and n_1, n_2 are not simultaneously zero, then we define[2]

$$\mathbf{m}_A \doteq \frac{1}{\sqrt{n_1^2 + n_2^2}} \begin{pmatrix} -n_2 \\ n_1 \\ 0 \end{pmatrix}.$$

The coefficients a and b are then computed by substituting (17) into (16) and using the fact that $\{\mathbf{n}_A, \mathbf{m}_A, \mathbf{n}_A \times \mathbf{m}_A\}$ is an orthonormal basis for $\mathbb{R}^3$. Explicitly,

$$a = -\left(\frac{R(\mathbf{n}_A, \mathbf{n}_B) R(\mathbf{n}_B, \phi) \mathbf{v}_B - \mathbf{v}_A}{\theta_B}\right) \cdot (\mathbf{m}_A \times \mathbf{n}_A) \quad \text{and}$$

$$b = \left(\frac{R(\mathbf{n}_A, \mathbf{n}_B) R(\mathbf{n}_B, \phi) \mathbf{v}_B - \mathbf{v}_A}{\theta_B}\right) \cdot \mathbf{m}_A.$$

This means that the feasible solutions can be completely parameterized as

$$X(\phi, s) = H(R(\mathbf{n}_A, \mathbf{n}_B) R(\mathbf{n}_B, \phi), \mathbf{t}(s)) \tag{18}$$

where $(\phi, s) \in [0, 2\pi) \times \mathbb{R}$ and $H(R, \mathbf{t})$ is the same as in (1).

3. INFORMATION-THEORETIC SOLUTION

Given that (10) constrains the possible solutions, $X(\phi, s)$, to a two-dimensional 'cylinder' defined by (18), the problem of solving for X reduces to that of solving (11) on this cylinder by determining the values (ϕ, s). There is therefore no need to search elsewhere in the 6D group $SE(3)$. Here we formulate and compare two approaches to finding (ϕ, s), thereby solving for X.

3.1. Approach 1: $\|\cdot\|_F^2$ Norm Minimization

In Approach 1 (which does not use any information theory), we simply back substitute $X = X(\phi, s)$ into (11) and minimize the cost function

$$C_1(\phi, s) = \|Ad([X(\phi, s)]^{-1}) \Sigma_A - \Sigma_B Ad^T(X(\phi, s))\|_F^2 \tag{19}$$

where $\|\cdot\|_F^2$ is the Frobenius norm. The reason for writing it this way (by pre-multiplying (11) by $[Ad(X)]^{-1} = Ad(X^{-1})$) before computing the norm, is that the parameter s then appears linearly inside the norm, and $C_1(\phi, s)$ is quadratic in s and can be written as[3]

$$C_1(\phi, s) = C_{10}(\phi) + C_{11}(\phi) s + \frac{1}{2} C_{12}(\phi) s^2.$$

The minimization over s can then be solved in closed form as

$$s = -\frac{C_{11}(\phi)}{C_{12}(\phi)}.$$

Back-substituting this into $C_1(\phi, s)$, an efficient 1D search over $\phi \in [0, 2\pi)$ can be performed.

[2] The special case when they are simultaneously zero is a set of measure zero, and hence is a rare event. Nevertheless, it is easy to handle, since in this case R_A is necessarily a rotation around $\mathbf{e}_3$.

[3] Neither the Frobenius norm, nor any other norm constructed from an inner product, can be Ad-invariant for $SE(3)$. Unlike for $SO(3)$ (and for that matter, all compact and all commutative Lie groups and their direct products) $\|Ad(H)\mathbf{v}\| \neq \|\mathbf{v}\|$ for generic $H \in SE(3)$.

3.2. Approach 2: Minimal KL Divergence

Given two populations of measured $\{A_i\}$ and $\{B_j\}$ for which means and covariances have been computed using (8), it is reasonable to fit parametric distributions to these data. As the amount of data becomes large, it is reasonable to assume that they behave as if they were drawn from Gaussian distributions. If this is not the case, it is possible to losslessly 'corral' the data so that it has Gaussian statistics, using a procedure described below.

A Gaussian on $SE(3)$ can be defined when the norm $\|\Sigma\|$ is small as [14,15]

$$\rho(H;M,\Sigma) = \frac{1}{(2\pi)^3|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}F(M^{-1}H)}$$

where $|\Sigma|$ denotes the determinant of Σ and

$$F(H) = [\log^{\vee}(H)]^T \Sigma^{-1} [\log^{\vee}(H)].$$

When H is parameterized with exponential coordinates, $H = \exp Z$, this means that $F(\exp Z) = \mathbf{z}^T\Sigma^{-1}\mathbf{z}$ where $\mathbf{z} = Z^{\vee}$ and $\rho(\exp Z; \mathbb{I}_4, \Sigma)$ becomes exactly a zero-mean Gaussian distribution on the Lie algebra $se(3)$, with covariance Σ, that is 'lifted up' to the Lie group $SE(3)$.

For the problem discussed in this paper, there is no loss of generality in assuming that $\|\Sigma_A\|$ and $\|\Sigma_B\|$ are small because the constraint equation (11) is linear in Σ_A and Σ_B, and so if they are not small, they can both be normalized resulting in $\Sigma'_A = \Sigma_A/(\|\Sigma_A\|)$ and likewise $\Sigma'_B = \Sigma_B/(\|\Sigma_A\|)$. Note that here we have normalized by the same quantity on both sides. We cannot use $\|\Sigma_A\|$ on one side of the equation and $\|\Sigma_B\|$ on the other because the Frobenius norm is not Ad-invariant for $SE(3)$.

Moreover, standard tests from multivariate statistical analysis such as q-q plots, as used recently in [22], can be used to assess whether or not the data are Gaussian. If they are not, they can be made Gaussian without loss of information or by introducing changes to the original mean and covariance in a simple way. Since $A_i = XB_iX^{-1}$, it follows that $A_i^p = XB_i^pX^{-1}$ for any power $p \in \mathbb{R}$. This means that each measured data point can be replaced with a continuum of equivalent data points parameterized by p. This neither adds nor destroys information content in the original data. Simply stated, Gaussians can be used in place of data even if the data are not Gaussian.

The search for X can then be reduced to finding the global minimum of the cost function

$$C_2(X) = D_{KL}(f_A \,\|\, \delta_X * f_B * \delta_{X^{-1}}) \qquad (20)$$

where $f_A(H) = \rho(H; M_A, \Sigma_A)$ and

$$(\delta_X * f_B * \delta_{X^{-1}})(H) = \rho(H; XM_BX^{-1}, Ad(X)\Sigma_B Ad^T(X)).$$

In general, the integral in this cost function cannot be solved in closed form because the log function is nonlinear, and in terms of exponential coordinates $dH = |J(\mathbf{z})|d\mathbf{z}$ where $|J(\mathbf{0})| = 1$, but this Jacobian is a nonlinear function of $\mathbf{z}$ as described in [10,11,15].

However, if a priori we limit the search for X to the cylinder defined in (18), then automatically $XM_BX^{-1} = M_A$. Then, we can define a new variable $K = M_A^{-1}H$ and using the property of invariance of integration under shifts, can write

$$C_2(X(\phi,s)) = D_{KL}(f'_A \,\|\, f'_B)$$

where

$$f'_A(K) = \rho(K; \mathbb{I}_4, \Sigma_A) \quad \text{and}$$

$$f'_B(K) = \rho(K; \mathbb{I}_4, Ad(X(\phi,s))\Sigma_B Ad^T(X(\phi,s))).$$

That is, when restricting to the cylinder, logarithms and exponentials cancel. Moreover, scaling covariances so that they are small ensures that the integral over $SE(3)$ reduces to a Gaussian integral over $se(3) \cong \mathbb{R}^6$ since $|J(\mathbf{0})| = 1$. The KL divergence of two distributions on $\mathbb{R}^n$ with means $\mathbf{m}_i$ and covariances Σ_i is

$$D_{KL}(f_1 \,\|\, f_2) =$$

$$\frac{1}{2}\left[\mathrm{tr}(\Sigma_2^{-1}\Sigma_1) + (\mathbf{m}_2 - \mathbf{m}_1)^T\Sigma_2^{-1}(\mathbf{m}_2 - \mathbf{m}_1) - n - \ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right)\right].$$

In our problem, $\mathbf{m}_2 - \mathbf{m}_1 = \mathbf{0}$, and since $SE(3)$ is unimodular, $|Ad(X)| = 1$ and so when evaluating $\Sigma_1 = \Sigma_A$ and $\Sigma_2 = Ad(X(\phi,s))\Sigma_B Ad^T(X(\phi,s))$, the final term in the above expression for $D_{KL}(f_1 \,\|\, f_2)$ is independent of X. Moreover, for our purposes additive and positive multiplicative constants can be ignored, so we can simply consider the first term in the KL-divergence, scaled by a factor of two:

$$C'_2(X(\phi,s)) = \mathrm{tr}(\Sigma_A^{-1}Ad(X(\phi,s))\Sigma_B Ad^T(X(\phi,s))),$$

minimized over $(\phi,s) \in [0,2\pi) \times \mathbb{R}$. Minimization over s can be done in closed form as in Section 3.1 since $C'_2(X(\phi,s))$ is also quadratic in s, and the result substituted back in for a 1D search over ϕ.

4. Results

We examined the efficacy of the algorithms both in simulation and in experiments. Simulated

data is used to show that the algorithms can perform accurately without knowledge of correspondence, validating their formulations. For experimental validation we chose US calibration as an example framework for data collection. For this case we judge the algorithms' performance based on the phantom model reconstruction given by the solved for X.

4.1. Validation in Simulation

To validate the proposed algorithms, we created sensor streams using simulated US probe trajectories with some a priori chosen X. We write the problem as $A^{ij}X = XB^{ij}$ where $B^{ij} = B_i^{-1}B_{i+1}$ are relative motions. B_i's are drawn from two sample "s-shaped" trajectories on a sphere. After forming relative motions, B^{ij}'s, we calculate $A^{ij} = XB^{ij}X^{-1}$, where X is the a priori chosen value.

The KL Batch and $\|\cdot\|_F^2$ Norm Batch algorithms were run with known correspondence in the data and then without knowledge of the correspondence to calculate X. To artificially remove knowledge of correspondence, we iteratively select each data point in one of the streams and then permutate this entry with one of its neighbors. The neighbor that is chosen is based on a value drawn from a normal distribution. In this manner, the highest likelihood is that the data point will not be permuted with any neighbor and will have a decreasing chance of being permuted with a neighbor, the farther away that neighbor is in the stream.

The calculated X was then compared to the a priori selected "true" X. The algorithms were unaffected by knowledge of correspondence and in each case performed with a high level of accuracy (see tab. 1). The results are the average of ten trials. The developed algorithms were compared against a commonly used $AX = XB$ algorithm which makes use of the Kronecker product and a least squares formulation [19, 20].

Table 1. Batch Method Results (Simulation)

	Correspondence is Known	
	Rotation Error(rad)	Translation Error(mm)
KL	$1.4 \cdot 10^{-4}$	$1.4 \cdot 10^{-3}$
$\|\cdot\|_F^2$	$7.3 \cdot 10^{-4}$	$7.4 \cdot 10^{-3}$
KP	0.0	2.46
	Correspondence is Unknown	
KL	$3.8 \cdot 10^{-4}$	$3.8 \cdot 10^{-3}$
$\|\cdot\|_F^2$	$7.3 \cdot 10^{-4}$	$7.4 \cdot 10^{-3}$
KP	$6.2 \cdot 10^{-2}$	2.65

Table 2. Batch Method Results (Experimental)

	Mean (mm)	Variance (mm)
KL	1.18	1.06
$\|\cdot\|_F^2$ Method	1.22	1.10

4.2. Experimental Results

To test real data, we chose a US calibration task as the realization of our $AX = XB$ problem. This problem requires a calibration phantom where any US image acquired from the phantom is unique. An example of such a phantom would be a planar Z-fiducial phantom [21]. We used a modified version of such a phantom. Data collection is the process of imaging the phantom with an US transducer in a series of poses. Each US image will intersect the Z-fiducials and the segmentation of these intersection points are used to generate our stream of A's. The stream of B's are the tracked poses of the US transducer, collected using an electromagnetic tracker. After removing correspondence, we used the two algorithms to calculate the calibration parameter for the US probe, X. To examine the accuracy of the computed X, we performed a reconstruction of the phantom model. Fig. 1 shows the results of plotting $B_i X^{-1}\mathbf{p}_i$ where $\mathbf{p}_i$ is the set of segmented points from each image. To quantify the difference, we fitted the phantom model to the figure and computed a normalized metric. For each point, we found its closest point on the model and computed the sum

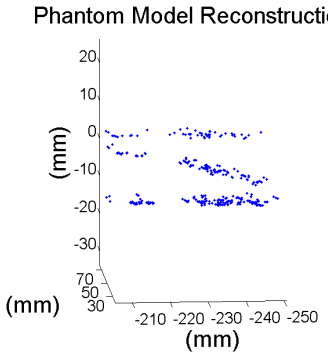

Figure 1. Reconstruction using the solved X

squared difference between them. Tab. 2 shows the mean and the standard deviation of this sum of squared differences and indicates that the error is reasonable. (The data that was used in the experimental results can be obtained at `https://sourceforge.net/p/jhu-axxb/`).

5. CONCLUSIONS AND FUTURE WORK

The AX=XB problem appears in many sensor calibration problems and has many solution methods. However, prior methods are not suitable for solving for X with data streams that contain gaps, are mis-aligned or, in general, when there is incomplete knowledge of the correspondence between A's and B's. We presented an information-theoretic algorithm (KL Batch) that constructs the data as distributions on the group $SE(3)$, and solves for X by minimizing the Kullback-Leibler divergence of these distributions with respect to the unknown X. In both simulation and experimentation, it was shown that this method reliably recovers an unknown X without the need for correspondence. In addition to performing slightly better in experimentation, the KL Batch was also determined to be superior to another new method presented here, the $\|\cdot\|_F^2$ norm batch, due to the simplified nature of the required calculation.

In future work we will further examine the proposed methods experimentally, for ultrasound calibration, as well as other contexts. Additionally we will work to improve our probability theoretic formulation by specifically accounting for sensor measurement noise, representing X by a mean and covariance, and not just a Dirac delta distribution.

References

[1] Bai, S., Teo, M.Y., "Kinematic Calibration and Pose Measurement of a Medical Parallel Manipulator by Optical Position Sensors," *Journal of Robotic Systems*, 20(4), 201-209, 2003.

[2] Fassi, I., Legnani, G., "Hand to Sensor Calibration: A Geometrical Interpretation of the Matrix Equation AX=XB," *Robotic Systems*, 22(9), 497-506, 2005.

[3] Shiu, Y.C., Ahmad, S.,"Calibration of Wrist-Mounted Robotic Sensors by Solving Homogeneous Transform Equations of the Form $AX = XB$," *IEEE Trans. on Robotics and Automation*, 5(1), 16-29, 1989.

[4] Chou, J.C.K., Kamel, M.,"Finding the Position and Orientation of a Sensor on a Robot Manipulator Using Quaternions," *IJRR*, 10(3), 240-254, 1991.

[5] Park, F.C., Martin, B.J.,"Robot Sensor Calibration: Solving $AX = XB$ on the Euclidean Group," *IEEE Trans. on Robotics and Automation*, 10(5), 717-721, 1994.

[6] Daniilidis, K., "Hand-Eye Calibration Using Dual Quaternions," *IJRR*, 18(3), 286-298, 1999.

[7] Kim, S.J., et. al., "Robot Head-Eye Calibration Using the Minimum Variance Method," *IEEE ROBIO*, 1446-1451, 2010.

[8] Chen, H.H."A Screw Motion Approach to Uniqueness Analysis of Head-Eye Geometry" *IEEE CVPR*, 145-151, 1991.

[9] Ackerman, M.K., Cheng, A., Shiffman, B., Boctor, E.M., Chirikjian, G.S., " Sensor Calibration with Unknown Correspondence: Solving AX=XB Using Euclidean-Group Invariants," *IROS*, 1308-13, 2013.

[10] Chirikjian, G.S., Kyatkin, A.B., *Engineering Applications of Noncommutative Harmonic Analysis*, CRC Press, Boca Raton, FL 2001.

[11] Chirikjian, G.S., *Stochastic Models, Information Theory, and Lie Groups: Volume 2 - Analytic Methods and Modern Applications*, Birkhäuser, Boston, Dec. 2011.

[12] Chirikjian, G.S., "Information-Theoretic Inequalities on Unimodular Lie Groups," *Journal of Geometric Mechanics*, 2(2), 119-158, 2010.

[13] Chirikjian, G. S., "Information Theory on Lie Groups and Mobile Robotics Applications," *IEEE ICRA*, 2751-2757, 2010.

[14] Wang Y., Chirikjian G.S.,"Error Propagation on the Euclidean Group with Applications to Manipulator Kinematics," *IEEE TRO* 591-602, 2006.

[15] Wang, Y., Chirikjian, G.S., "Nonparametric Second-Order Theory of Error Propagation on the Euclidean Group," *IJRR*, 22(4), 1258-1273, 2008.

[16] Long, A.W., Wolfe, K.C., Mashner, M.J., Chirikjian, G.S., "The Banana Distribution is Gaussian: A Localization Study with Exponential Coordinates," *RSS*, 2012.

[17] Long, A.W., Wolfe, K.C., Chirikjian, G.S., "Planar Uncertainty Propagation and a Probabilistic Algorithm for Interception," *WAFR*, 2012.

[18] Chirikjian, G.S., "Modeling Loop Entropy," *Methods in Enzymology*, 487, 99-132, 2011.

[19] Boctor, E.M., "Enabling Technologies For Ultrasound Imaging In Computer-Assisted Intervention," *Comp. Sci. Dep., Johns Hopkins*.Thesis, 2006.

[20] Andreff, N., Horaud, R., Espiau, B.,"Robot Hand-Eye Calibration Using Structure-from-Motion," *IJRR*, 20(3), 228-248, 2001.

[21] Pagoulatos, N., Haynor, D., Kim, Y., "A Fast Calibration Method for 3-D Tracking of Ultrasound Images Using a Spatial Localizer," *UMB*, 27(9), 1219-1229, 2001.

[22] Park, W., Midgett, C.R., Madden, D.R., Chirikjian, G.S., "A stochastic kinematic model of class averaging in single-particle electron microscopy," *IJRR*, 30(6), 730-754, 2011.