2016 IEEE International Conference on Robotics and Automation (ICRA)
Stockholm, Sweden, May 16-21, 2016

New Probabilistic Approaches to the $AX = XB$ Hand-Eye Calibration without Correspondence

Qianli Ma[1], Haiyuan Li[2] and Gregory S. Chirikjian[1]

***Abstract*— The hand-eye calibration problem was first formulated decades ago and is widely applied in robotics, image guided therapy, etc. It is usually cast as the "$AX = XB$" problem where the matrices A, B, and X are rigid body transformations in $SE(3)$. Many solvers have been proposed to recover X given data streams $\{A_i\}$ and $\{B_i\}$ with correspondence. However, exact correspondence might not be accessible in the real world due to the asynchronous sensors and missing data, etc. A probabilistic approach named "Batch method" was introduced in previous research of our lab, which doesn't require a prior knowledge of the correspondence between the two data streams $\{A_i\}$ and $\{B_j\}$. Analogous to non-probabilistic approaches which require data selection to filter out ill-conditioned data pairs, the Batch method has restrictions on the data set $\{A_i\}$ and $\{B_j\}$ that can be used. We propose two new probabilistic approaches built on top of the Batch method by giving new definitions of the mean on $SE(3)$, which alleviate the restrictions on the data set and significantly improve the calibration accuracy of X.**

I. Introduction

The Hand-eye calibration problem, or the "$AX = XB$" problem dates back to the 1980s when Shiu and Ahmad [1] first proposed a solution. Even after a quarter of a century, the topic remains active today in the areas such as aerial vehicle sensor calibration [2], image guided therapy (IGT) sensor calibration and endoscopic surgery [3].

Various $AX = XB$ solvers have been proposed so far for either same or different scenarios and applications. Tsai [4] proposed a method as well as several principles on data selection in 1989. A geometric view of $AX = XB$ is given by Fassi and Legnani [5] and the over-constrained and singular conditions are also discussed. Many other $AX = XB$ methods include but are not limited to the Euclidean Group method [6], the quaternion method [7], [8], the dual quaternion method [9], the Kronecker method [10], and the screw motion method [11]. Several new optimization methods emerged recently such as the convex optimization method [12] and the global optimization methods [13]. Most of the above methods are designed for off-line usage where a complete list of data pairs (A_i, B_i) where $i = 1, 2, ..., n$ has to be provided to recover X whereas online methods are more preferable in real time applications [10], [14] A common feature of these methods is that they demand the exact correspondence between the data streams $\{A_i\}$ and $\{B_i\}$. This might not always hold because of asynchronizicity of the sensors or missing data, etc. Therefore, a probabilistic method called the "Batch method" that deals with data pairs without correspondence was first brought up in [15]. An information-theoretic approach is proposed in [16] by viewing the problem as distributions on the special Euclidean group $SE(3)$ and minimizing the Kullback-Leibler divergence on the distributions.

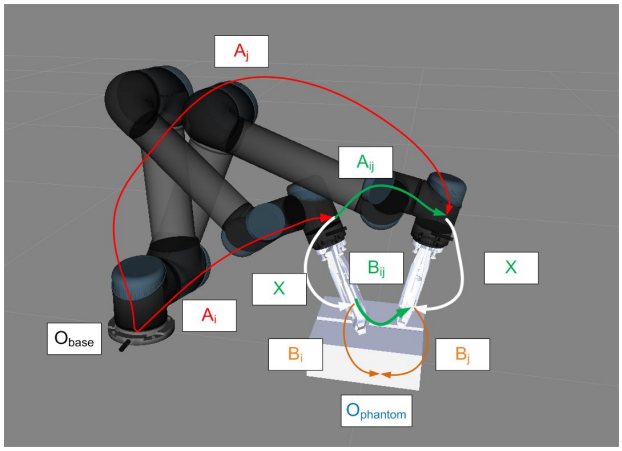

Fig. 1. Application of the $AX = XB$ problem in the extrinsic calibration of the ultrasound probe with respect to the UR5 robot arm. The white squarish object in the middle represents the ultrasound calibration phantom.

In the "$AX = XB$" formulation, A, X, and B are each homogeneous transformations or equivalently, elements in $SE(3)$. Data pairs (A_i, B_i) are measured by the robot and the sensors mounted at the end-effector such as cameras, ultrasound probes, etc. Fig .1 shows that a UR5 robot uses a phantom to calibrate the relative transformation between the ultrasound probe and the end-effector of the robot. X is the unknown rigid body transformation calculated by solving $AX = XB$. It is well known that for non-probabilistic methods, at least two exact data pairs (A_1, B_1) and (B_2, B_2) are required to recover a unique X. For probabilistic approaches such as the Batch method, no correspondence between A and B is needed and a comparison between the Batch method and Kronecker product method in [15] shows its extraordinary capability of handling data sets without correspondence. However, as claimed in [15], the Batch method has restrictions on the data set that can be used which limits its scope of application. In addition, the definition of mean on $SE(3)$ plays an important role in the Batch method, and the current definition can't give the desired "average" of

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

[1]Qianli Ma and Gregory S. Chirikjian are with the Department of Mechanical Engineering and Laboratory for Computational Sensing and Robotics, Johns Hopkins University, Baltimore, MD, USA, mqianli1@jhu.edu, gregc@jhu.edu

[2]Haiyuan Li is with the School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China haiyuanli@hotmail.com

a set rigid body transformations for the use of "$AX = XB$" problem. In this paper, we show that this restriction can be alleviated and by using new definitions, the accuracy of the calibrated X can be improved significantly.

The rest of the paper is organized as follows. In Section II, the probabilistic Batch method is introduced along with the definitions of the probability distribution function (PDF) and convolution on $SE(3)$. In Section III, we prove that the assumption of using highly focused distribution for the Batch method can be relaxed and the New Batch method can be built by using a new definition of the mean in $SE(3)$. In Section IV, simulation experiments are performed among the Kronecker product method, Batch method, and two new Batch methods to show both the effectiveness and the accuracy of the latter. In Section V, conclusions are drawn and possible future directions are noted.

II. Mathematical Background

In this section, the definition of PDF, Delta function and the convolution on $SE(3)$ will be introduced. Then the mathematical formulation of the Batch method will be presented explicitly and two key equations are obtained for solving X. The first equation describes the relationship between the mean of $\{A_i\}$ and the mean of $\{B_j\}$, and the second equations describes the relationship between their covariance matrices. Lastly, the disadvantages of the Batch method are pointed out, which serves as the motivation to build for the new Batch methods.

A. PDF, Delta function and Convolution on $SE(3)$

Let $\Sigma = \Sigma^T \in \mathbb{R}^{6\times 6}$ be a positive definite covariance matrix. Assuming the norm $\|\Sigma\|$ is small, we define a Gaussian probability distribution function on $SE(3)$ as:

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

where $H \in SE(3)$, $M \in SE(3)$ is the mean, $\Sigma \in \mathbb{R}^{6\times 6}$ is the covariance matrix, $\|\Sigma\|$ denotes the determinant of Σ and

$$F(H) = [\log^{\vee}(H)]^T \Sigma^{-1} [\log^{\vee}(H)]. \tag{2}$$

If $Z \in se(3)$ is the corresponding Lie algebra of $H \in SE(3)$, then $\exp(Z) = H$, $\log H = Z$ and $\log^{\vee}(H) = \mathbf{z}$ where $\mathbf{z} \in \mathbb{R}^{6\times 1}$ is the vectorization of Lie algebra.

The integration of a PDF on $SE(3)$ is defined as follows:

$$\int_{SE(3)} f(H)dH \doteq \int_{\mathbf{q}\in D} f(H(\mathbf{q}))|J(\mathbf{q})|d\mathbf{q}, \tag{3}$$

where $\mathbf{q} = [q_1, ..., q_6]^T$ is a global set of coordinates, $d\mathbf{q} = dq_1 dq_2 \cdots dq_6$ and $|J(\mathbf{q})|$ is the determinant of the Jacobian matrix as:

$$J(\mathbf{q}) = \left[\left(H^{-1}\frac{\partial H}{\partial q_1} \right)^{\vee}; \left(H^{-1}\frac{\partial H}{\partial q_2} \right)^{\vee}; \cdots \left(H^{-1}\frac{\partial H}{\partial q_6} \right)^{\vee} \right]. \tag{4}$$

Given two functions $f_1, f_2 \in (L^1 \cap L^2)(SE(3))$, their convolution is defined as:

$$(f_1 * f_2)(H) \doteq \int_{SE(3)} f_1(K) f_2(K^{-1}H) dK, \tag{5}$$

where $K \in SE(3)$ and $*$ denotes the convolution of functions on $SE(3)$.

The Dirac delta function on $SE(3)$ is defined as follows:

$$\int_{SE(3)} \delta(H)dH = 1 \tag{6a}$$

$$(f \star \delta)(H) = (\delta \star f)(H) = f(H). \tag{6b}$$

It can also be viewed informally as

$$\delta(H) = \begin{cases} +\infty, & H = I \\ 0, & H \neq I. \end{cases} \tag{7}$$

The shifted Dirac delta function is then defined as $\delta_X(H) = \delta(X^{-1}H)$.

B. Batch Method

Given a set of data pairs $(A_i, B_i) \in SE(3) \times SE(3)$ with correspondence, the following is true:

$$A_i X = X B_i, \tag{8}$$

where $i = 1, 2, ..., n$. If we use the probability theory on $SE(3)$, Eq. (8) can be converted into:

$$(\delta_{A_i} * \delta_X)(H) = (\delta_X * \delta_{B_i})(H). \tag{9}$$

Note that convolution is a linear operation on functions, therefore n instances of Eq.(9) can be added into a single equation as:

$$(f_A * \delta_X)(H) = (\delta_X * f_B)(H), \tag{10}$$

where

$$f_A(H) = \frac{1}{n}\sum_{i=1}^{n} \delta(A_i^{-1}H) \text{ and } f_B(H) = \frac{1}{n}\sum_{i=1}^{n} \delta(B_i^{-1}H). \tag{11}$$

The above functions can be normalized to be probability densities:

$$\int_{SE(3)} f_A(H)dH = \int_{SE(3)} f_B(H)dH = 1. \tag{12}$$

If we let the mean and covariance of a probability density $f(H)$ satisfy:

$$\int_{SE(3)} \log(M^{-1}H) f(H) dH = \mathbb{O} \tag{13a}$$

$$\Sigma = \int_{SE(3)} \log^{\vee}(M^{-1}H)[\log^{\vee}(M^{-1}H)]^T f(H) dH. \tag{13b}$$

Then for a PDF $f_A(H)$ as given in Eq.(11), the discrete version of the mean M_A and covariance Σ_A will be:

$$\sum_{i=1}^{n} \log(M_A^{-1}A_i) = \mathbb{O} \tag{14a}$$

$$\Sigma_A = \frac{1}{n}\sum_{i=1}^{n} \log^{\vee}(M_A^{-1}A_i)[\log^{\vee}(M_A^{-1}A_i)]^T. \tag{14b}$$

In [17], f_1 and f_2 are assumed to be highly focused functions so that the mean and covariance for the convolution can be closely approximated as:

$$M_{1*2} = M_1\, M_2 \tag{15a}$$

$$\Sigma_{1*2} = Ad(M_2^{-1})\,\Sigma_1\, Ad^T(M_2^{-1}) + \Sigma_2, \tag{15b}$$

where

$$Ad(H) = \begin{pmatrix} R & \mathbb{O} \\ \widehat{\mathbf{x}}R & R \end{pmatrix}. \tag{16}$$

The "hat" operator $\widehat{\ }$: $\mathbb{R}^{6\times 1} \to se(3)$ maps a 6 by 1 vector to its corresponding Lie algebra which satisfies the exponential map $\exp : se(3) \to SE(3)$.

Employing the fact that the mean of $\delta_X(H)$ is $M_X = X$ and the covariance $\Sigma_X = \mathbb{O}_{6\times 6}$ is a zero matrix, Eq.(15a) and Eq.(15b) give two key equations in the Batch method as:

$$M_A\, X = X\, M_B \tag{17}$$

and

$$Ad(X^{-1})\,\Sigma_A\, Ad^T(X^{-1}) = \Sigma_B. \tag{18}$$

[16] notes that there are two degrees of freedom in Eq.(17), so Σ_A and Σ_B are decomposed to provide the required constraints:

$$\Sigma_i = \begin{pmatrix} \Sigma_i^1 & \Sigma_i^2 \\ \Sigma_i^3 & \Sigma_i^4 \end{pmatrix}, \tag{19}$$

where $\Sigma_i^3 = (\Sigma_i^2)^T$ and the subscript i can be either A or B. We can extract the first two blocks of Eq. (18) and we can get:

$$\Sigma_{M_B}^1 = R_X^T \Sigma_{M_A}^1 R_X \tag{20a}$$

$$\Sigma_{M_B}^2 = R_X^T \Sigma_{M_A}^1 R_X \widehat{(R_X^T t_x)} + R_X^T \Sigma_{M_A}^2 R_X. \tag{20b}$$

As described in [16], a unique R_X can be obtained by calculating the eigendecomposition of $\Sigma_{M_A}^1$ and $\Sigma_{M_B}^1$ as $\Sigma_i = Q_i \Lambda Q_i^T$, where Q_i is the square matrix whose ith column is the eigenvector of Σ_i and Λ is the diagonal matrix with corresponding eigenvalues as diagonal entries. After further derivations, the mean of the rotation component of X can be written as:

$$R_X = Q_{M_A} Q Q_{M_B}^T, \tag{21}$$

where Q is a diagonal matrix and there exist four candidates of Q in total. The optimal solution of R_X can be picked by minimizing a cost function that contains the constraint information from Eq. (17). The translation component $\mathbf{t}_X$ can be uniquely determined using Eq. (20b).

C. Open Problems

To determine the mean and covariance of the convolution of two highly focused PDFs as in [17], the Baker-Campbell-Hausdorff (BCH) formula is used:

$$\log(e^X e^Y) = \underbrace{X + Y}_{\text{0th order}} + \underbrace{\frac{1}{2}[X, Y]}_{\text{1st order}} + h.o.t. \tag{22}$$

where $h.o.t.$ stands for the 2nd and higher order terms. If X and Y are further constrained to be small so that $\|X\| \ll 1$ and $\|Y\| \ll 1$, then the first order approximation of Eq. (22) can be used to derive Eq. (15a) and Eq. (15b). In the "$AX = XB$' context, e^X and e^Y represent A_i and B_i respectively. This derivation constrains the distribution function f_A and f_B to be highly focused. However, this condition is not necessary as proved later. Moreover, the current definition of mean on $SE(3)$ as in Eq.(14a) can't closely reflect the desired "average" of M_A and M_B in the probabilistic $AX = XB$ context. It will be shown that two new ways of defining the mean can be used to augment the existing Batch method which can even significantly improve the accuracy of X depending on the distributions of $\{A_i\}$ and $\{B_j\}$. In later sections, we will call the original Batch method as the "Batch method" and the two augmented Batch methods as the "Batch1" and "Batch2" methods.

III. New Batch Methods

In this section, we first prove that unlike in [15], Eq. (17) and Eq. (18), which is also referred to as the key equations, are exact (rather than approximate) expressions that do not depend on the smallness of Σ or any assumptions about the form of the functions f_A or f_B. Then two new definitions of the mean on $SE(3)$ will be given based on the 1st order and 2nd order approximation of of Eq. (13a). The new means for $\{A_i\}$ and $\{B_j\}$, and the corresponding covariance matrices Σ_A and Σ_B, can be directly incorporated into the Batch method to form the Batch1 and Batch2 methods.

A. Conditions for the Key Equations

Starting with Eq.(9), performing a convolution on the left of both sides of the equation with $\delta_{X^{-1}}(H)$, and using the associativity of convolution and the properties of delta functions, Eq.(9) can be replaced with:

$$(\delta_{X^{-1}} * \delta_{A_i} * \delta_X)(H) = \delta_{B_i}(H) \tag{23}$$

and summing both sides over i and dividing by n gives

$$(\delta_{X^{-1}} * f_A * \delta_X)(H) = f_B(H). \tag{24}$$

Let M_A be the mean of f_A and M_B be the mean of f_B as defined in Eq.(14a). After using the properties of the delta function and Eq.(24),

$$\int_{SE(3)} \log(M_B^{-1} H)(\delta_{X^{-1}} * f_A * \delta_X)(H) dH =$$

$$\int_{SE(3)} \log(M_B^{-1} H) f_A(X H X^{-1}) dH = \mathbb{O}.$$

Changing variables as $K = XHX^{-1}$ and using the invariance of integration,

$$\int_{SE(3)} \log(M_B^{-1} X^{-1} K X) f_A(K) dK = \mathbb{O}. \tag{25}$$

Multiplying on the left by X and right by X^{-1} and using the fact that $X[\log(M_B^{-1} X^{-1} K X)]X^{-1} = \log(X M_B^{-1} X^{-1} K)$ gives $X M_B^{-1} X^{-1} = M_A^{-1}$, which is the

same as Eq.(17). The proof for Eq.(18) follows in a similar way.

This result broadens the types of data sets $\{A_i\}$ and $\{B_j\}$ that can be used for the Batch method. However, the mean definition as in Eq.(13a) doesn't always provide the desired M_A and M_B in terms of minimizing $C_1(R_X) = ||R_{M_A}R_X - R_X R_{M_B}||$ or $C_2(\mathbf{t}_X) = ||R_{M_A}\mathbf{t}_X + \mathbf{t}_{M_A} - R_{M_X}\mathbf{t}_{M_B} - \mathbf{t}_{M_B}||$, which are simply the error metrics for the rotational and translational components of Eq. (17). To find a better definition of mean under the "$AX = XB$" context, we start by assuming $M^{-1}H$ is small such that $\|M^{-1}H - \mathbb{I}\| \ll 1$. This is different from the assumption of the Batch method that H is small enough such that $\|\log(H)\| \ll 1$. From there, the 1st order and 2nd order approximations of the mean in Eq.(13a) can be achieved and new definitions are proposed accordingly. As will be shown in section IV, the new definitions are more than just the approximations of the original mean.

Note that though Eq. (13a) and Eq. (13b) have a similar form, they are not necessarily bounded together. Equivalently, Eq. (13b) will be valid as long as the mean M is given, which does not have to be defined as in Eq. (13a). Therefore, it is appropriate to use a different mean while using the same definition of the covariance. One thing to be noted is that the proof of Eq. (15b) does not depend on the definition of mean, but only on the 1st order approximation of BCH. Therefore, Eq. (18) will still be valid when we use the new definitions of mean. As for Eq. (17), the 1st or 2nd order approximations can be made on $\log(XM_B^{-1}X^{-1}K)$ in the previous proof by treating $XM_B^{-1}X^{-1}$ as a whole. The same result can be obtained by following the similar procedure in the definitions as below, which will not be mentioned in detail here.

Batch1 and Batch2 methods simply replace the old means of $\{A_i\}$ and $\{B_j\}$ with new ones, obtain the corresponding covariances in terms of the new means, and use the two key equations to solve for X. For the two different kinds of distributions of $\{A_i\}$ and $\{B_j\}$ that are tested, Batch1 and Batch2 methods are able to recover a much more accurate R_X in a consistent manner, and recover $\mathbf{t}_X$ to a certain level of accuracy depending on the type of distribution.

B. Mean Based on the 1st Order Approximation

Consider the Taylor expansion of the matrix logarithm described as:

$$\log(\mathbb{I} + X) = X - \frac{1}{2}X^2 + \frac{1}{3}X^3 - ... \tag{26}$$

If $\|M^{-1}H - \mathbb{I}\| \ll 1$ and we retain the 1st order approximation of Eq.(26), it can be written as:

$$\log(M^{-1}H) = \log(\mathbb{I} + (M^{-1}H - \mathbb{I})) \approx \left(M^{-1}H - \mathbb{I}\right). \tag{27}$$

The first order approximation of Eq.(13a) is:

$$\int_{SE(3)} (M^{-1}H - \mathbb{I}) f(H) dH \approx \mathbb{O}. \tag{28}$$

Note that $f(H)$ is a normalized probability density such that $\int_{SE(3)} f(H)dH = 1$, so:

$$M^{-1}\int_{SE(3)} Hf(H)dH \approx \mathbb{I}. \tag{29}$$

Define the first order approximation of $M \in \mathbb{R}^{4\times4}$ as $\widehat{M}$:

$$\widehat{M} \doteq \int_{SE(3)} Hf(H)dH \notin SE(3). \tag{30}$$

Take $\{A_i\}$ for an example, the corresponding discrete version of $\widehat{M}$ will be:

$$\widehat{M_A} \doteq \sum_{i=1}^{n} A_i \left(\frac{1}{n}\sum_{j=1}^{n}\delta(A_j^{-1}A_i)\right) = \frac{1}{n}\sum_{i=1}^{n} A_i \notin SE(3). \tag{31}$$

Note that though Eq.(30) and Eq.(31) are obtained based on the 1st order approximation of the matrix logarithm, $\widehat{M}$ is not the 1st order approximation of the mean $M \in SE(3)$ because it is not necessary a group element in $SE(3)$ and elements in $SE(3)$ do not add. Therefore, we define a new mean based on Eq.(31) by projecting $\widehat{M}$ onto $SE(3)$, where singular value decomposition is performed on the "rotation component" $\widehat{R}_{\widehat{M_A}}$ of $\widehat{M_A}$:

$$\widehat{R}_{\widehat{M_A}} = U\Sigma V^T. \tag{32}$$

The rotation component $R_{M_A} \in SO(3)$ can be obtained according to [18] as:

$$R_{M_A} = UV^T. \tag{33}$$

After recovering the rotation component, the new mean based on the 1st order approximation becomes:

$$M_A^1 = \begin{pmatrix} R_{M_A} & \frac{1}{n}\sum_{i=1}^{n}\mathbf{t}_{A_i} \\ \mathbf{0}^T & 1 \end{pmatrix}. \tag{34}$$

Using this definition of mean, M_A^1 and M_B^1 can be calculated in a straight way. Σ_A^1 and Σ_B^1 can be obtained by Eq. (14b) afterwards. Batch1 method uses the above means and covariances for recovering X, and all the other procedures strictly follow the procedure described as in Eqs. (17-21).

C. Mean Based on the 2nd Order Approximation

In this section, we further define a mean on $SE(3)$ based on the second order approximation of Eq.(13a). Under the assumption that $\|M^{-1}H - \mathbb{I}\| \ll 1$ and using the Eq, (27), the 2nd order approximation of Eq.(13a) can be written as:

$$\int_{SE(3)} \left((M^{-1}H - \mathbb{I}) - \frac{1}{2}(M^{-1}H - \mathbb{I})^2\right) f(H)dH \approx \mathbb{O}. \tag{35}$$

Expand Eq.(35) and multiply M on both sides of Eq.(35):

$$\int_{SE(3)} \left(2H - \frac{1}{2}HM^{-1}H - \frac{3}{2}M\right) f(H)dH \approx \mathbb{O}. \tag{36}$$

Substituting Eq. (30) into Eq.(36), we have:

$$2\widehat{M} - \frac{1}{2}\int_{SE(3)} HM^{-1}Hf(H)dH - \frac{3}{2}M \approx \mathbb{O}. \tag{37}$$

Let the 2nd order approximation of $M \in \mathbb{R}^{4\times4}$ be denoted by $\overline{M}$:

$$2\widehat{M} - \frac{1}{2}\int_{SE(3)} H\overline{M}^{-1}Hf(H)dH - \frac{3}{2}\overline{M} = \mathbb{O}. \quad (38)$$

Take the data set $\{A_i\}$ for an example, the discrete version of Eq.(38) is:

$$\frac{2}{n}\sum_{i=1}^{n} A_i - \frac{1}{2n}\sum_{i=1}^{n} A_i\overline{M}_A^{-1}A_i - \frac{3}{2}\overline{M}_A = \mathbb{O}. \quad (39)$$

Similar to the 1st order case, $\overline{M}_A$ is the 2nd order approximation of $M_A \in \mathbb{R}^{4\times4}$, because the candidates that satisfy Eq.(39) are not likely to be elements of $SE(3)$. A straight forward way of calculating $\overline{M}_A$ is by using nonlinear optimization technique given an initial guess. It is possible to solve for $\overline{M}_A \in \mathbb{R}^{4\times4}$ and then project $\overline{M}_A$ back to $M_A^2 \in SE(3)$ as shown in [19], where M_A^2 denotes the new mean based on the 2nd order approximation. However, this highly nonlinear problem can be linearized by defining the update law as below:

$$M_A[j+1] = M_A[j]\exp(\Omega_A) \approx M_A[j](\mathbb{I} + \Omega_A), \quad (40)$$

where $M_A[j]$ is the value of $\overline{M}_A$ at the jth step, $\Omega_A \in se(3)$ is a Lie algebra element such that $\|\Omega_A\| \ll 1$. Under this assumption, the inverse of M_A can be updated as:

$$M_A^{-1}[j+1] = \exp^{-1}(\Omega_A)M_A^{-1}[j] \approx (\mathbb{I} - \Omega_A)M_A^{-1}[j]. \quad (41)$$

Substituting Eq. (40) and Eq. (41) into Eq. (39), the previous nonlinear problem on $\overline{M}_A$ is then converted into a linear one on Ω_A, where only Ω_A needs to be solved for given a known $M_A[j]$:

$$J \cdot vec(\Omega_A) = \mathbf{b}, \quad (42)$$

where

$$J = \frac{1}{2n}\sum_{i=1}^{n}\left((M_A[j]A_i)^T \otimes A_i\right) - \frac{3}{2}(\mathbb{I} \otimes M_A[j]) \quad (43a)$$

$$\mathbf{b} = vec(-\frac{2}{n}\sum_{i=1}^{n} A_i + \frac{1}{2n}\sum_{i=1}^{n}(A_iM_A[j]A_i) + \frac{3}{2}M_A[j]). \quad (43b)$$

$\otimes$ is the symbol for Kronecker product and $vec()$ denotes the vectorization of a matrix formed by stacking its columns into a single column vector. .

To start with, we use the result M_A^1 obtained by Batch1 as the initial guess such that $M_A[0] = M_A^1$. We keep solving for Ω_A and update $M_A[j+1]$ and $M_A^{-1}[j+1]$ until the matrix norm of left hand side of Eq. (39) falls below a small threshold. It is observed that $M_A[j]$ can converge quickly in approximately 4 steps to obtain $\overline{M}_A$, which will be projected onto $SE(3)$ to get M_A^2. Similar to the Batch1 method, Batch2 method uses the new means (M_A^2, M_B^2) and the corresponding covariances (Σ_A^2, Σ_B^2) to recover X.

IV. Numerical Simulations

In this section, numerical simulations are performed to show the advantages of using M^1 and M^2 as the means for $\{A_i\}$ and $\{B_j\}$ when solving for X. These Batch methods are compared with each other to show the accuracy in terms of recovering R_X and $\mathbf{t}_X$. They are also compared with the Kronecker product to show the effectiveness of probabilistic methods when dealing with data streams without correspondence.

A. Generation of $\{A_i\}$ and $\{B_j\}$ Using Different Distributions

First, we generate (A_i, B_i) data pairs which have correspondence given the ground truth of X_{true}. Then we scrambled the data in $\{A_i\}$ and $\{B_i\}$ up to a certain percentage rate to get $\{A_i\}$ and $\{B_j\}$ which have only partial or no correspondence. Finally, the scrambled data streams will be used in Batch, Batch1, Batch2 and Kronecker product methods to calculate the R_X and $\mathbf{t}_X$ which will be compared with $R_{X_{\mathbf{true}}}$ and $\mathbf{t}_{X_{\mathbf{true}}}$.

Eq. (44) and Eq. (45) show two ways to generate (A_i, B_i) data pairs for numerical experiments, both of which sample on B_i first and then given X_{true} obtain the corresponding A_i:

$$B_i = B_0\exp(\widehat{\delta_i})\exp(\widehat{\gamma_i}) \quad (44a)$$

$$\delta_i = (\mathbf{0}_{3\times1}^T, \sigma\mathbf{n}_1^T)^T \in \mathbb{R}^{6\times1} \quad (44b)$$

$$\gamma_i = (\sigma\mathbf{n}_2^T/\|\mathbf{n}_2\|, \mathbf{0}_{3\times1}^T)^T \in \mathbb{R}^{6\times1} \quad (44c)$$

$$\sigma \in \mathbb{R}, \ \ \mathbf{n}_1, \mathbf{n}_2 \in \mathcal{N}^{3\times1}(0,1) \in \mathbb{R}^{3\times1}, \quad (44d)$$

and

$$B_i = B_0\exp(\widehat{\delta_i}) \quad (45a)$$

$$\delta_i \in \mathcal{N}(\mathbf{0}; \Sigma) \subset \mathbb{R}^6 \quad (45b)$$

$$\Sigma = \sigma\mathbb{I}_{6\times6}, \sigma \in \mathbb{R}, \quad (45c)$$

where $B_0 \in SE(3)$ is an arbitrary "baseline" for generating the cloud of B_i, $\mathcal{N}^{3\times1}(0,1)$ denotes a 3 by 1 vector where each vector element follows a standard normal distribution $\mathcal{N}(0,1)$. $\mathcal{N}(\mathbf{0};\Sigma)$ is a zero mean multivariate Gaussian distribution. The major difference between Eq. (44) and Eq. (45) is that when perturbing B_0 to get $\{B_i\}$, the former samples on the rotational and translational component of $se(3)$ separately while the latter samples on each component simultaneously. After generating $\{B_i\}$, $\{A_i\}$ can be easily obtained as:

$$A_i = X_{true}B_iX_{true}^{-1}. \quad (46)$$

It turns out that the way of generating the data streams influences the effectiveness of different means. Equivalently speaking, each type of mean might be better at representing certain transformations in the $AX = XB$ context.

Next, we compare the performances of all the four methods in numerical simulation. Given an arbitrary B_0, 50 instances of B_i are generated using Eq. (44) and Eq. (45) respectively, with $\sigma = 0.9$. If we provide the ground truth X_{true}, 50 corresponding A_i can be calculated as in

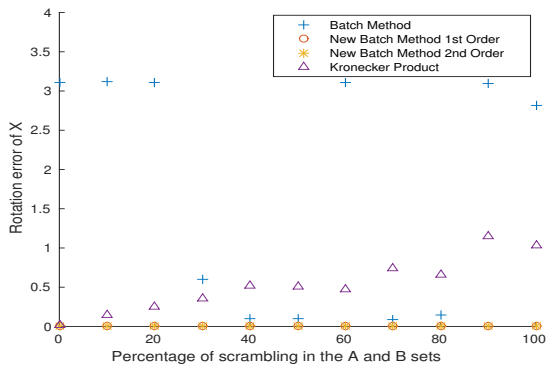

Fig. 2. Rotation error vs. percentage of scrambling in $\{A_i\}$ and $\{B_j\}$ for the Batch, Batch1, Batch2 and Kronecker product methods where $\{B_i\}$ is generated using Eq. (44).

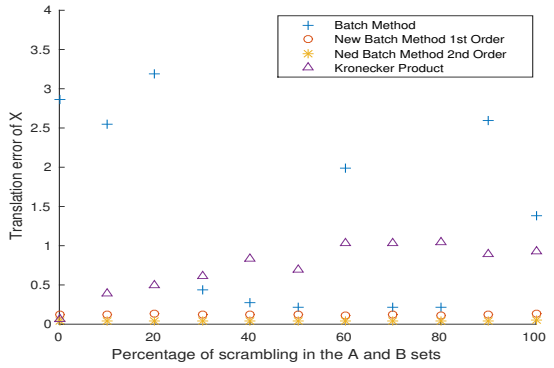

Fig. 3. Translation error vs. percentage of scrambling in $\{A_i\}$ and $\{B_j\}$ for the Batch, Batch1, Batch2 and Kronecker product methods where $\{B_i\}$ is generated using Eq. (44)

Eq. (46). In addition, different sets of $\{B_j\}$ are obtained by permuting the elements in $\{B_i\}$ by a percentage rate $r = 0, 10, 20, \ldots, 100$. Then the calculated R_X and $\mathbf{t}_X$ are compared with the ground truth X_{true} using the following error metrics:

$$error_R = ||\log^{\vee}(R_X^T R_{X_{\mathbf{true}}})||_2 \tag{47}$$

$$error_{\mathbf{t}} = \frac{||t_X - t_{X_{\mathbf{true}}}||_2}{||t_{X_{\mathbf{true}}}||_2}. \tag{48}$$

Eq. (48) has the advantage of eliminating the influence of using different units for the translation. 70 trials are performed for each scrambling rate r and the averages of $error_R$ and $error_{\mathbf{t}}$ are calculated for comparison.

B. Numerical Simulation Results

All the simulation results are plotted as in Figs. 2-7 where Fig. 2-4 use the data sampled from Eq. (44) and Figs. 5-7 use the data sampled from Eq. (45). In addition, Fig 4 and Fig. 7 are closer looks at the Batch methods in Fig. 2 and Fig. 4 respectively. Several observations can be made from the figures. First, Figs. 2-3 and Figs. 5-6 show that all the Batch methods are invariant on the scrambling rate r whereas the results of the Kronecker product deteriorate quickly as r

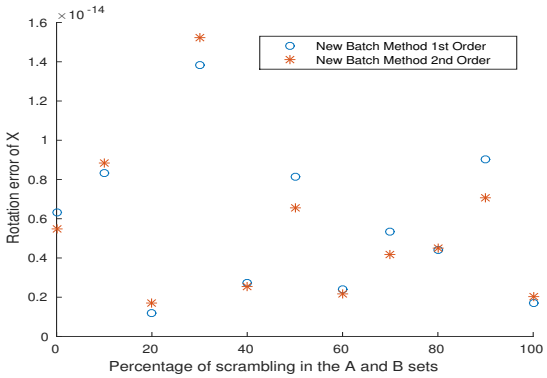

Fig. 4. A close look at Fig. 2 for the Batch1 and Batch2 methods which shows that the rotation errors from Batch1 and Batch2 methods are at the magnitude of 10^{-15}.

increases. Second, no matter which data sampling method is used, Batch1 and Batch2 methods show significant improvements over the Batch method on recovering R_X. The rotation error magnitude of Batch1 and Batch2 ranges from 10^{-14} to 10^{-15}, while the rotation error from Batch is at the level of 10^{-1} in Fig. 2 and 10^{-4} in Fig. 7. Moreover, the rotation error of Batch method can go beyond reasonable ranges occasionally as can be seen in Fig. 2. Third, Fig. 3 shows that the translational component $\mathbf{t}_X$ obtained by Batch1 and Batch2 are also better than Batch. However, Fig. 7 shows that the performance of Batch1 and Batch2 is not as good as that of Batch if Eq. (45) is used for data sampling. Lastly, in Fig. 3 and Fig. 6, the translational error from Batch2 is smaller than that of Batch1, which reflects the necessity of using the new mean M^2 based on the 2nd order approximation to solve for X.

In summary, for data sets $\{A_i\}$ and $\{B_j\}$ without correspondence, Batch1 and Batch2 are extremely good at recovering R_X in a precise and consistent manner no matter which sampling method is chosen. The accuracy of the calculated $\mathbf{t}_X$ is dependent on the sampled data, which is true for both the old and new Batch methods. One can be better than another if a different type of data samples is used. However, the Batch2 method is always better than Batch1 method in terms of getting a more accurate $\mathbf{t}_X$.

V. Conclusion

In this paper, we brought up two new probabilistic methods which can handle the "$AX = XB$" problem without a priori knowledge of the correspondence between $\{A_i\}$ and $\{B_j\}$. We build our new approaches on top of the Batch method which is a previous work of the lab, and show that an appropriate definition of mean for a set of rigid body transformations in $SE(3)$ can affect the effectiveness of the probabilistic methods to a large extent. The new definitions of means are derived from the $1st$ order and $2nd$ order approximations of the original definition. Rather than simple approximations, they reflect the "average" of a set of rigid body transformations from different perspectives, which significantly improves the accuracy and consistency for the

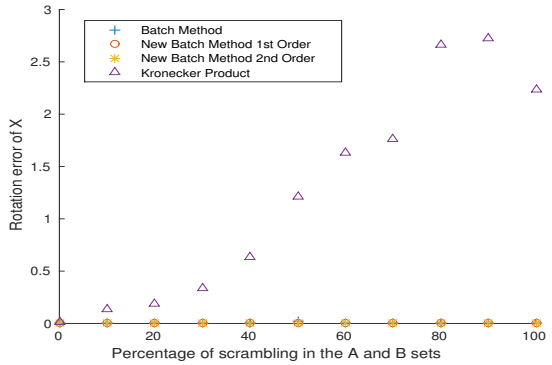

Fig. 5. Rotation error vs. percentage of scrambling in $\{A_i\}$ and $\{B_j\}$ for the Batch, Batch1, Batch2 and Kronecker product methods where $\{B_i\}$ is generated using Eq. (45).

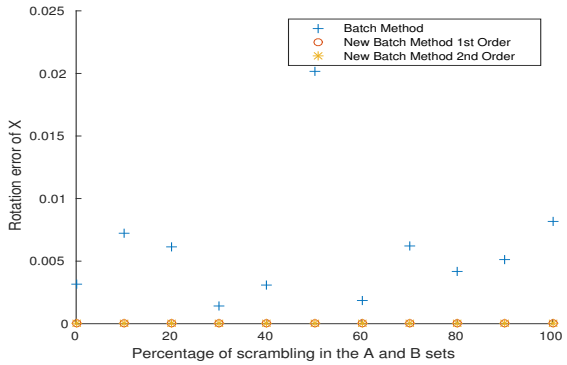

Fig. 7. A close look at Fig. 5 for the Batch, Batch1 and Batch2 methods. It shows that the rotation errors from Batch1 and Batch2 methods are at the magnitude of 10^{-15} (can't be discerned in the figure) which is significantly smaller than that of the Batch method.

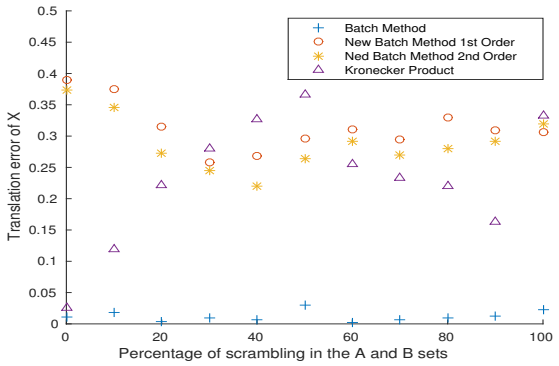

Fig. 6. Translation error vs. percentage of scrambling in $\{A_i\}$ and $\{B_j\}$ for the Batch, Batch1, Batch2 and Kronecker product methods where $\{B_i\}$ is generated using Eq. (45)

calibration of X given data sets without correspondence. Numerical simulations are performed to show the superiority of the Batch1 and Batch2 methods.

This work opens up many future directions. For the "$AX = XB$" problem which assumes data correspondence, how to design the trajectory of the robot arm to achieve the optimal calibration result is of critical importance to the non-probabilistic solvers. Similarly, planning a robot trajectory generated from the desired distribution of $\{B_i\}$ (or $\{A_i\}$) plays an important role in probabilistic approaches. Moreover, other potential definitions of means and how to match a given set of data to the appropriate mean definition can be explored for better performance.

REFERENCES

[1] Y. Shiu and S. Ahmad, "Finding the mounting position of a sensor by solving a homogeneous transform equation of the form AX = XB," in *IEEE International Conference on Robotics and Automation (ICRA)*, vol. 4. IEEE, 1987, pp. 1666–1671.

[2] E. Mair, M. Fleps, M. Suppa, and D. Burschka, "Spatio-temporal initialization for imu to camera registration," *IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 557–564.

[3] J. Schmidt, F. Vogt, and H. Niemann, "Robust hand–eye calibration of an endoscopic surgery robot using dual quaternions," in *Pattern Recognition*. Springer, 2003, pp. 548–556.

[4] R. Tsai and Y. Lenz, "A new technique for fully autonomous and efficient 3d robotics hand/eye calibration," *IEEE Transactions on Robotics and Automation*, vol. 5, no. 3, pp. 345–358, 1989.

[5] I. Fassi and G. Legnani, "Hand to sensor calibration: A geometrical interpretation of the matrix equation AX = XB," *Journal of Robotic Systems*, vol. 22, no. 9, pp. 497–506, 2005.

[6] F. Park and B. Martin, "Robot sensor calibration: solving AX = XB on the Euclidean group," *IEEE Transactions on Robotics and Automation*, vol. 10, no. 5, pp. 717–721, 1994.

[7] J. C. Chou and M. Kamel, "Quaternions approach to solve the kinematic equation of rotation, AX = XB, of a sensor-mounted robotic manipulator," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 1988, pp. 656–662.

[8] R. Horaud and F. Dornaika, "Hand-eye calibration," *The International Journal of Robotics Research*, vol. 14, no. 3, pp. 195–210, 1995.

[9] K. Daniilidis, "Hand-eye calibration using dual quaternions," *The International Journal of Robotics Research*, vol. 18, no. 3, pp. 286–298, 1999.

[10] N. Andreff, R. Horaud, and B. Espiau, "On-line hand-eye calibration," in *International Conference on 3-D Digital Imaging and Modeling*. IEEE, 1999, pp. 430–436.

[11] Z. Zhao and Y. Liu, "Hand-eye calibration based on screw motions," in *International Conference on Pattern Recognition (ICPR)*, vol. 3. IEEE, 2006, pp. 1022–1026.

[12] Z. Zhao, "Hand-eye calibration using convex optimization," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2011, pp. 2947–2952.

[13] J. Heller, D. Henrion, and T. Pajdla, "Hand-eye and robot-world calibration by global polynomial optimization," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 3157–3164.

[14] J. Angeles, G. Soucy, and F. P. Ferrie, "The online solution of the hand-eye problem," *IEEE Transactions on Robotics and Automation*, vol. 16, no. 6, pp. 720–731, 2000.

[15] M. Ackerman and G. S. Chirikjian, "A probabilistic solution to the AX=XB problem sensor calibration without correspondence," *Proceedings of the conference on the Geometric Science of Information*, 2013.

[16] M. K. Ackerman, A. Cheng, and G. Chirikjian, "An information-theoretic approach to the correspondence-free AX = XB sensor calibration problem," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 4893–4899.

[17] Y. Wang and G. S. Chirikjian, "Nonparametric second-order theory of error propagation on motion groups," *The International Journal of Robotics Research*, vol. 27, no. 11-12, pp. 1258–1273, 2008.

[18] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[19] C. Belta and V. Kumar, "Euclidean metrics for motion generation on se (3)," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 216, no. 1, pp. 47–60, 2002.