

New Definitions of Mean

Qianli Ma*

Gregory S. Chirikjian

Robot and Protein Kinematics Laboratory
Laboratory for Computational Sensing and Robotics
Department of Mechanical Engineering
The Johns Hopkins University
Baltimore, Maryland, 21218
Email: {mqianli1, gchirik1}@jhu.edu

Abstract to be added.

where

$$Ad(H) = \begin{pmatrix} R & \mathbb{O} \\ \hat{\mathbf{x}}R & R \end{pmatrix} \quad (4)$$

1 Introduction

We begin by defining a Gaussian probability distribution on $SE(3)$ (assuming the norm $\|\Sigma\|$ is small) as

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

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

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

Previously, in order to determine the mean of the convolution of two PDFs, Baker-Campbell-Hausdorff formula is used given the assumption that function f_1 and f_2 are both highly focused. If $X, Y \in se(3)$,

$$\log(e^X e^Y) = X + Y + \frac{1}{2}[X, Y] + \frac{1}{12}([X, [X, Y]] + [Y, [Y, X]]) + \dots \quad (1)$$

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.(1) can be written as:

$$\log(e^X e^Y) = X + Y + \frac{1}{2}[X, Y] \quad (2)$$

As shown in [?], the mean of covariance of the convolution of two highly focused functions are:

$$M_{1*2} = M_1 M_2 \text{ and } \Sigma_{1*2} = Ad(M_2^{-1}) \Sigma_1 Ad^T(M_2^{-1}) + \Sigma_2 \quad (3)$$

2 Mean and Covariance of A Single Distribution on $SE(3)$

2.1 Order Approximation of M

Though this approximation works well when the distribution of X is treated as a Delta function, it fails to extend to the case where its distribution is a general PDF $f(X)$. In an alternative to the first order approximation using Baker-Campbell-Hausdorff formula, it is possible to only assume $M^{-1}H$ is small such that $\|M^{-1}H - \mathbb{I}\| \ll 1$. Given the Taylor expansion of the matrix logarithm described as:

$$\log(\mathbb{I} + X) = X - \frac{1}{2}X^2 + \frac{1}{3}X^3 - \dots \quad (5)$$

Then it is straight forward to have:

$$\begin{aligned} \log(M^{-1}H) &= \log(\mathbb{I} + (M^{-1}H - \mathbb{I})) \\ &= (M^{-1}H - \mathbb{I}) - (M^{-1}H - \mathbb{I})^2/2 + (M^{-1}H - \mathbb{I})^3/3 - \dots \end{aligned} \quad (6)$$

Given the definition of the mean M of a probability density $f(H)$ as:

$$\int_{SE(3)} \log(M^{-1}H) f(H) dH = \mathbb{O} \quad (7)$$

The first order approximation of Eq.(7) is:

$$\int_{SE(3)} (M^{-1}H - \mathbb{I}) f(H) dH \approx \mathbb{O} \quad (8)$$

*Address all correspondence to this author.

$$M^{-1} \int_{SE(3)} Hf(H)dH \approx \mathbb{I} \quad (9) \quad \text{Multiply } M \text{ on both side of Eq.(16):}$$

$$\int_{SE(3)} \left(2M^{-1}H - \frac{1}{2}M^{-1}HM^{-1}H - \frac{3}{2}\mathbb{I} \right) f(H)dH \approx \mathbb{O} \quad (16)$$

Define the first order approximation of M as \hat{M} :

$$\hat{M} \doteq \int_{SE(3)} Hf(H)dH \quad (10) \quad \text{Substituting Eq.(10) into Eq.(17), we have:}$$

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

If $f(H)$ is of the form of $f_A(H)$ given above, then

$$\sum_{i=1}^n \log(M_A^{-1}A_i) = \mathbb{O} \text{ and} \quad (11)$$

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

Discrete version will be:

$$\hat{M}_A \doteq \sum_{i=1}^n A_i \left(\frac{1}{n} \sum_{j=1}^n \delta(A_j^{-1}A_i) dH \right) = \frac{1}{n} \sum_{i=1}^n A_i \quad (12)$$

Note that \hat{M} is generally not a group element in $SE(3)$, and the corresponding $SE(3)$ version can be obtained by projecting \hat{M} into $SE(3)$ using singular value decomposition (SVD) technique.:

$$R_{\hat{M}} = U \Sigma V^T \quad (13)$$

The rotation part of the projected \hat{M} (named as \hat{M}_{proj}) is:

$$R_{\hat{M}_{proj}} = V U^T \quad (14)$$

When testing with no-noise data, $\hat{A}X - X\hat{B}$ is pretty good in rotation, while the error in translation is significant. This will be rectified in the 2nd order approximation as introduced in the next section.

2.2 Second Order Approximation of M

Take the second order approximation in Eq.(7):

$$\int_{SE(3)} \left((M^{-1}H - \mathbb{I}) - \frac{1}{2}(M^{-1}H - \mathbb{I})^2 \right) f(H)dH \approx \mathbb{O} \quad (15)$$

$$\int_{SE(3)} \left(2M^{-1}H - \frac{1}{2}M^{-1}HM^{-1}H - \frac{3}{2}\mathbb{I} \right) f(H)dH \approx \mathbb{O} \quad (16)$$

$$\int_{SE(3)} \left(2H - \frac{1}{2}HM^{-1}H - \frac{3}{2}M \right) f(H)dH \approx \mathbb{O} \quad (17)$$

$$2\hat{M} - \frac{1}{2} \int_{SE(3)} HM^{-1}Hf(H)dH - \frac{3}{2}M \approx \mathbb{O} \quad (18)$$

The 2nd order approximation of M is denoted by \bar{M} defined as:

$$2\hat{M} - \frac{1}{2} \int_{SE(3)} H\bar{M}^{-1}Hf(H)dH - \frac{3}{2}\bar{M} = \mathbb{O} \quad (19)$$

$$2\hat{M} - \bar{M} \frac{1}{2} \int_{SE(3)} \bar{M}^{-1}H\bar{M}^{-1}Hf(H)dH - \frac{3}{2}\bar{M} = \mathbb{O} \quad (20)$$

$$2\hat{M} - \frac{1}{2n} \sum_{i=1}^n A_i \bar{M}^{-1} A_i - \frac{3}{2}\bar{M} = \mathbb{O} \quad (21)$$

$$\boxed{2\hat{M} - \frac{1}{2}\hat{M}\bar{M}^{-1}\hat{M} - \frac{3}{2}\bar{M} = \mathbb{O}} \quad (22)$$

Initial value of \bar{M} is important for the accuracy of its convergence. When $\bar{M}_0 = \hat{M}$, which means the initial guess of \bar{M} is obtained by another orthogonalization of \hat{M} , the result is much better.

Note: simulation shows that, given data streams without noise, if we minimize on the above cost function along all six directions of the Lie algebra, then the objective function is likely to fall into a local minimum that is not better than the 1st order approximation. Instead, we only minimize on the three translational directions defined by the corresponding Lie algebra and the result becomes better. The theoretical meaning behind this is still unknown.

Currently, the length of search steps is adjusted dynamically to get object function out of local minimums that are not small enough.

The second term of Eq.(20) is very similar to the definition of the covariance of $f(H)$, and maybe \bar{M} can be updated using the information of the covariance. Also, take a look at the cubness of variance in the first volume. The same technique as in Eq.(13) can be employed to project \hat{M} into $SE(3)$.

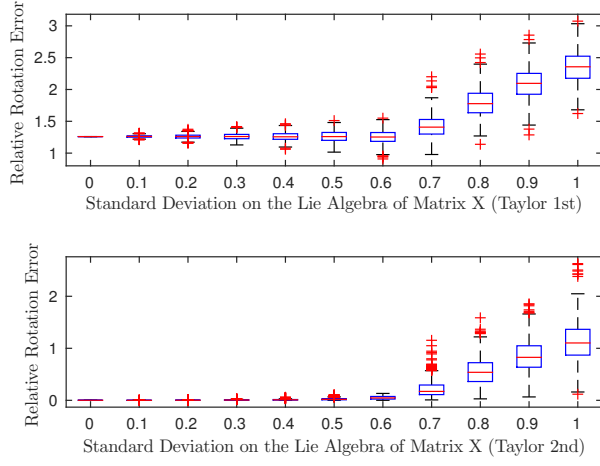


Fig. 1. Rotation Errors of 1st (up) and 2nd (down) Order Approximation With Respect to Different Noise Levels. Setup : (1) Number of Trials for Each Noise Level $n_t = 10$. (2) Number of Sample Points for Each Trial $n_s = 500$.

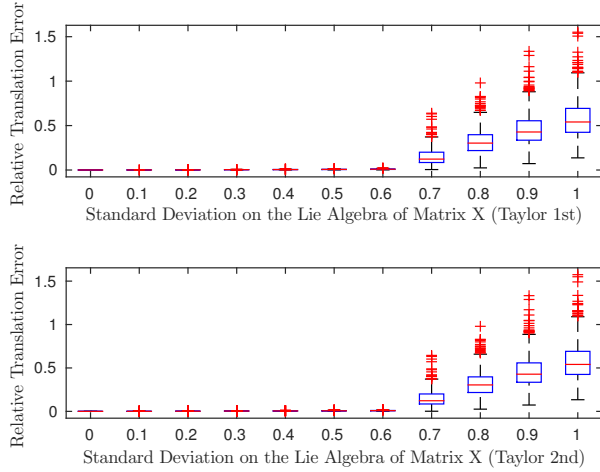


Fig. 2. Translation Errors of 1st (up) and 2nd (down) Order Approximation With Respect to Different Noise Levels. Setup : (1) Number of Trials for Each Noise Level $n_t = 10$. (2) Number of Sample Points for Each Trial $n_s = 500$.

2.3 Numerical Experiments for Validation of the Second Order Approximation of the Mean of a PDF

Let M_X , $MX1$ and $MX2$ be the means of X obtained using Eq.(11), Eq.(12 - 14) and Eq.(21) respectively.

Number of iterations for each noise level is 10 and the number of sample points for each trial is 500. The following metrics are used for error comparison:

$$error_R = \|R_X^T R_{X_{true}}\|_2 \quad (23)$$

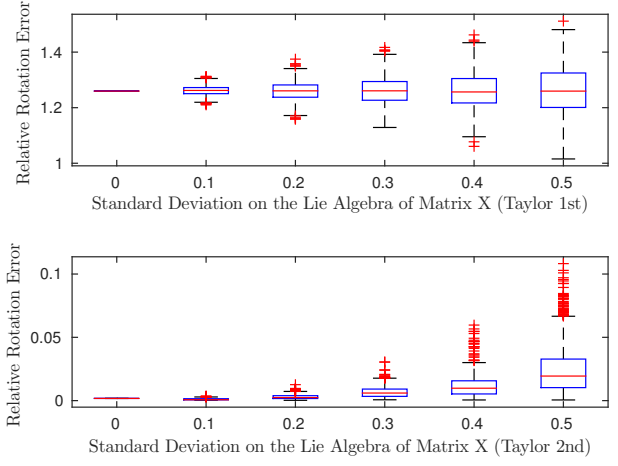


Fig. 3. A Closer Look at Fig.2.3 with $\sigma = 0, 0.1, \dots, 0.5$

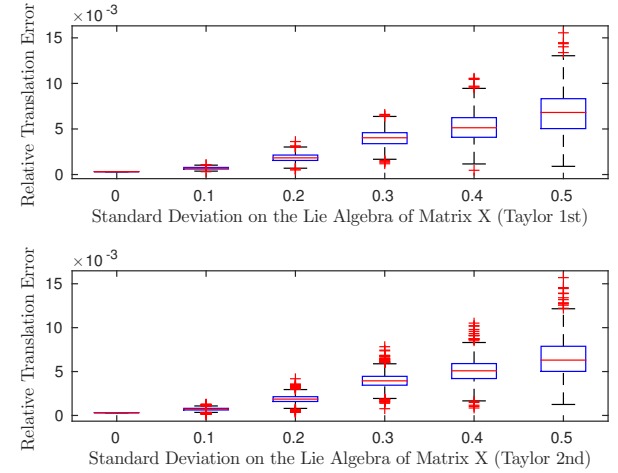


Fig. 4. A Closer Look at Fig.2.3 with $\sigma = 0, 0.1, \dots, 0.5$

$$error_t = \frac{\|t_X - t_{X_{true}}\|_2}{\|t_{X_{true}}\|_2} \quad (24)$$

Numerical simulation results are shown as in Fig.2.3, Fig.2.3, Fig.2.3 and Fig.2.3.

2.4 First Order Approximation of Σ

Given the definition of the covariance Σ of a PDF $f(H)$ as:

$$\Sigma = \int_{SE(3)} \log^\vee(M^{-1}H) [\log^\vee(M^{-1}H)]^T f(H) dH \quad (25)$$

Its first order approximation $\hat{\Sigma}$ can be written as:

$$\hat{\Sigma} \doteq \int_{SE(3)} (M^{-1}H - \mathbb{I})^\vee [(M^{-1}H - \mathbb{I})^\vee]^T f(H) dH \quad (26)$$

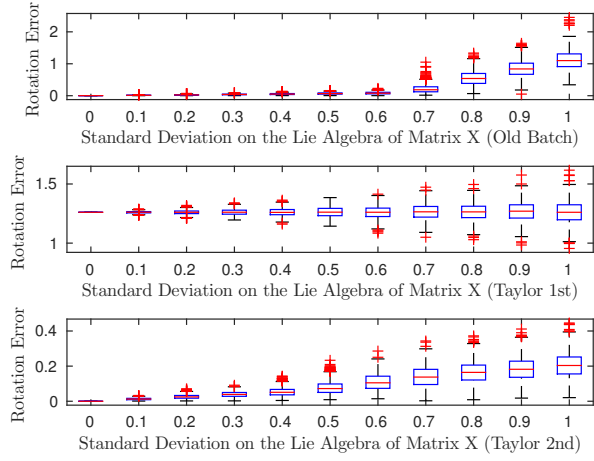


Fig. 5. Comparison of Rotation Errors With Respect to Different Noise Levels with Original Definition (up), 1st Taylor (mid) and 2nd Taylor (down). Setup : (1) Number of Trials for Each Noise Level $n_t = 10$. (2) Number of Sample Points for Each Trial $n_s = 500$.

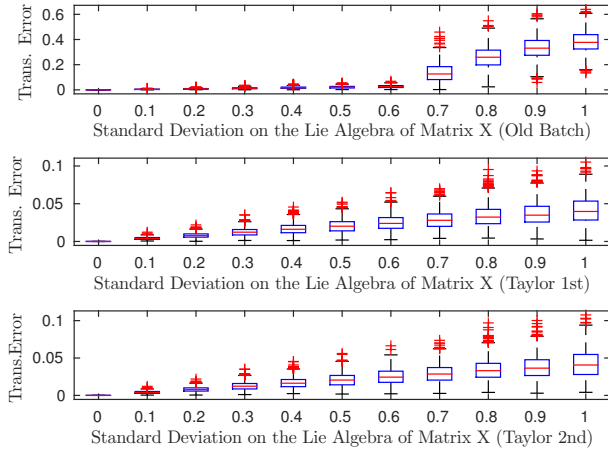


Fig. 6. Comparison of Translation Errors With Respect to Different Noise Levels with Original Definition (up), 1st Taylor (mid) and 2nd Taylor (down). Setup : (1) Number of Trials for Each Noise Level $n_t = 10$. (2) Number of Sample Points for Each Trial $n_s = 500$.

The discrete version for Σ_A will be:

$$\Sigma_A = \frac{1}{n} \sum_{i=1}^n (M_A^{-1} A_i - \mathbb{I})^\vee [(M_A^{-1} A_i - \mathbb{I})^\vee]^T. \quad (27)$$

By defining $Q = M^{-1}H$, Eq.(26) can be written as:

$$\hat{\Sigma} \doteq \int_{SE(3)} (Q - \mathbb{I})^\vee [(Q - \mathbb{I})^\vee]^T f(Q) dQ \quad (28)$$

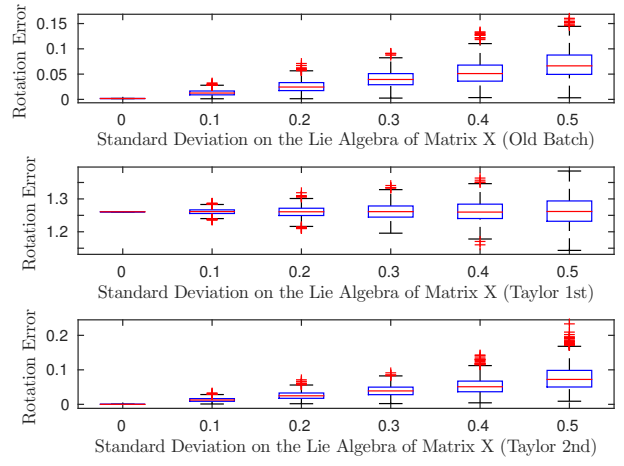


Fig. 7. A Close Look at Fig.5 with $\sigma = 0, 0.1, \dots, 0.5$

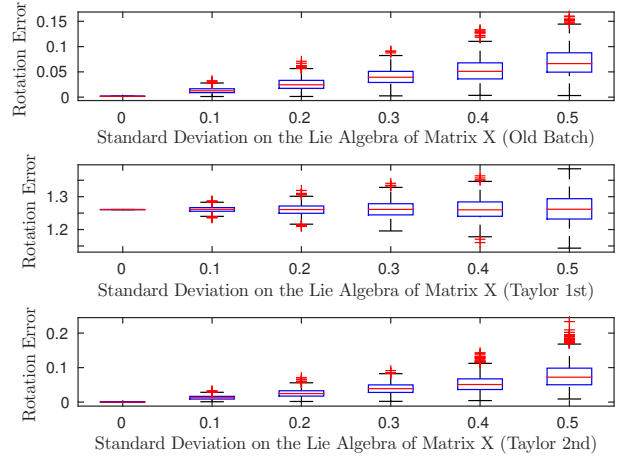


Fig. 8. A Closer Look at Fig.6 with $\sigma = 0, 0.1, \dots, 0.5$

A discrete version will be:

$$\Sigma = \frac{1}{n} \sum_{i=1}^n (Q_i - \mathbb{I})^\vee [(Q_i - \mathbb{I})^\vee]^T. \quad (29)$$

If $\|G - \mathbb{I}\| \ll 1$, then $\Sigma = \hat{\Sigma}$

$$(Q - \mathbb{I})^\vee = \begin{pmatrix} \frac{1}{2}(R - R^T) \\ \mathbf{t} \end{pmatrix} \quad (30)$$

2.5 Second Order Approximation of Σ

Let

$$P = (M^{-1}H - \mathbb{I}) - \frac{1}{2}(M^{-1}H - \mathbb{I})^2 \quad (31)$$

Similarly, the second order approximation of Σ is:

$$\bar{\Sigma} \doteq \int_{SE(3)} P^\vee [P^\vee]^T f(H) dH \quad (32)$$

which will be in a rather complex form. However, the calculation of the discrete version of this will be similar to how it is done in the 1st order approximation of the covariance.

3 Mean and Covariance of the Convolution between Two Distributions on $SE(3)$

3.1 First Order Approximation of M_{1*2}

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

$$(f_1 * f_2)(H) \doteq \int_{SE(3)} f_1(K) f_2(K^{-1}H) dK. \quad (33)$$

The corresponding mean M_{1*2} will be given as:

$$\int_{SE(3)} \log(M_{1*2}^{-1}H) (f_1 * f_2)(H) dH = \mathbb{O} \quad (34)$$

If both Σ_1 and Σ_2 are very small, then $M_{1*2} \approx M_1 M_2$ as is used in the old Batch method. Without using this approximation, take the assumption that $M^{-1}H$ is small and we will have:

$$\int_{SE(3)} \int_{SE(3)} (M_{1*2}^{-1}H - \mathbb{I}) f_1(K) f_2(K^{-1}H) dK dH \approx \mathbb{O} \quad (35)$$

Define $L = K^{-1}H$,

$$\int_{SE(3)} \int_{SE(3)} (M_{1*2}^{-1}KL - \mathbb{I}) f_1(K) f_2(L) dK dL \approx \mathbb{O} \quad (36)$$

(Note: when the integrating domain is the entire $SE(3)$, this is Ok. However, when formulating the discrete version of the equation, this can be problematic because the integral domain is only a subset of $SE(3)$ and $K^{-1}H$ can change the original integrating domain into a new one. The same problem can happen to Eq.(42).)

By using Eq.(10) twice, Eq.(36) becomes:

$$M_{1*2}^{-1} \widehat{M}_1 \widehat{M}_2 \approx \mathbb{I} \quad (37)$$

It is straight forward to define \widehat{M}_{1*2} as:

$$\boxed{\widehat{M}_{1*2} = \widehat{M}_1 \widehat{M}_2} \quad (38)$$

Note that though the definition of \widehat{M}_{1*2} is similar to that in the old Batch method, no assumption on Σ is made but $M^{-1}H$

is small. Moreover, in order to solve for the true M_{1*2} , projection of \widehat{M}_{1*2} back onto $SE(3)$ is still required. There are two ways to solve for M_{1*2}^{proj} .

$$M_{1*2}^{proj} = \begin{cases} (\widehat{M}_1 \widehat{M}_2)_{proj} \\ \widehat{M}_{1proj} \widehat{M}_{2proj} \end{cases} \quad (39)$$

The effectiveness of this approximation has been numerically simulated as shown in Fig.(9) and Fig.(10).

3.2 Second Order Approximation of M_{1*2}

For simplicity, we drop the domain of integral $SE(3)$,

$$\int \int \left(2M_{1*2}^{-1}H - \frac{1}{2}M_{1*2}^{-1}HM_{1*2}^{-1}H - \frac{3}{2}\mathbb{I} \right) f_1(K) f_2(K^{-1}H) dK dH \approx \mathbb{O} \quad (40)$$

Substitute $L = K^{-1}H$,

$$\int \int \left(2M_{1*2}^{-1}KL - \frac{1}{2}M_{1*2}^{-1}KLM_{1*2}^{-1}KL - \frac{3}{2}\mathbb{I} \right) f_1(K) f_2(L) dK dL \approx \mathbb{O} \quad (41)$$

Employing the definition of \widehat{M} ,

$$M_{1*2}^{-1} \widehat{M}_1 \widehat{M}_2 \approx \frac{1}{4} \int \int (M_{1*2}^{-1}KLM_{1*2}^{-1}KL) f_1(K) f_2(L) dK dL + \frac{3}{4}\mathbb{I} \quad (42)$$

The discrete version of Eq.(42) is:

$$\boxed{\bar{M}_{1*2}^{-1} \widehat{M}_1 \widehat{M}_2 = \frac{1}{4n^2} \sum_{i=1}^n \sum_{j=1}^n \bar{M}_{1*2}^{-1} K_i L_j \bar{M}_{1*2}^{-1} K_i L_j + \frac{3}{4}\mathbb{I}} \quad (43)$$

The effectiveness of this 2nd order approximation has been numerically simulated as shown in Fig.(9) and Fig.(10).

3.3 1st and 2nd Taylor Approximations of the Convolution of Two Distributions on $SE(3)$

4 Error Comparison between AX and XB

Given distributions on A , X , and B originating from the equation $AX = XB$, the difference between AX and XB are numerically simulated by calculating the means of $f_A * f_X$, $f_X * f_B$ respectively, as shown in Fig.(17) and Fig.(18).

5 Convolution of XBX^{-1}

Treating X as a probability distribution, then the following equation holds only when A and B have the same noise level (or similar distribution):

$$(f_A * f_X)(g) = (f_X * f_B)(g) \quad (44)$$

However, in the real world, the noise levels on A and B are different, and different forms of convolution should be chosen accordingly. For example, when the noise level on A is

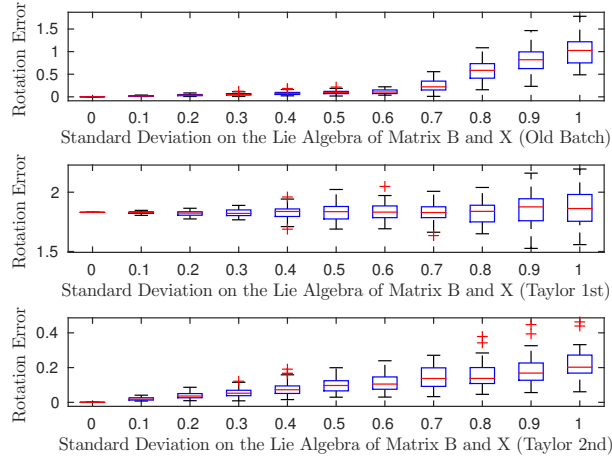


Fig. 9. Comparison of Translation Errors With Respect to Different Noise Levels Given One Pair of True X and True B

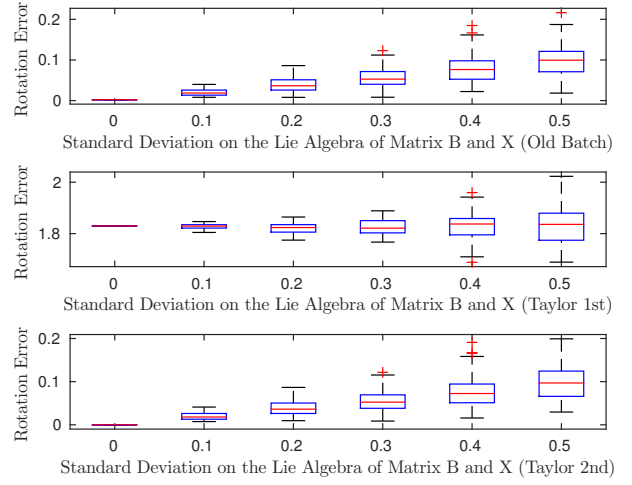


Fig. 11. Comparison of Translation Errors With Respect to Different Noise Levels Given One Pair of True X and True B

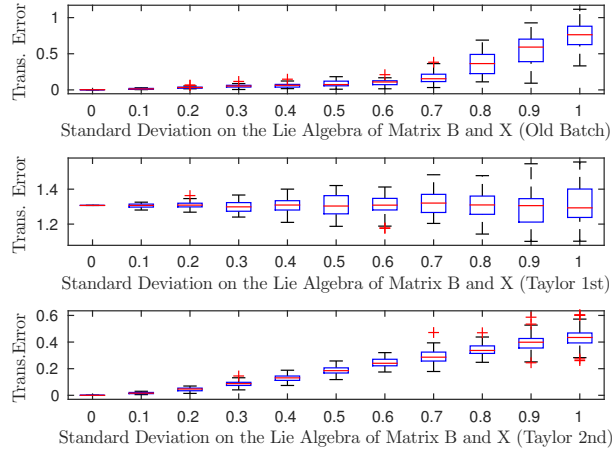


Fig. 10. Comparison of Translation Errors With Respect to Different Noise Levels Given One Pair of True X and True B

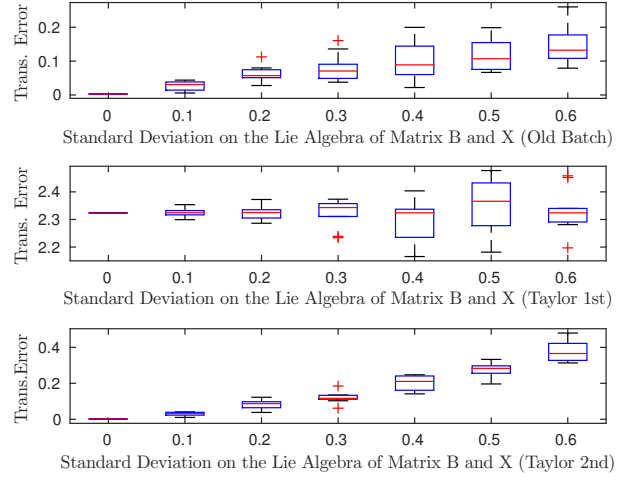


Fig. 12. Comparison of Translation Errors With Respect to Different Noise Levels Given One Pair of True X and True B

larger, the following equation can be used:

$$f_A(g) = (f_X * f_B * f_{X^{-1}})(g) \quad (45)$$

The goal becomes to compare M_A and $M_{X*B*X^{-1}}$ by using Eq.(43) twice. The first stage is just to verify the effectiveness of Eq.(45) given known data sets of A , B and X .

$$\bar{M}_{X*B}^{-1} \hat{M}_X \hat{M}_B = \frac{1}{4n^2} \sum_{i=1}^n \sum_{j=1}^n \bar{M}_{X*B}^{-1} X_i B_j \bar{M}_{X*B}^{-1} X_i B_j + \frac{3}{4} \mathbb{I} \quad (46)$$

$$\bar{M}_{X*B*X^{-1}}^{-1} \hat{M}_X \hat{M}_B \hat{M}_{X^{-1}} = \frac{1}{4n^2} \sum_{j=1}^n \bar{M}_{X*B*X^{-1}}^{-1} \bar{M}_{X*B} X_j^{-1} \bar{M}_{X*B*X^{-1}}^{-1} \bar{M}_{X*B} X_j^{-1} + \frac{3}{4} \mathbb{I} \quad (47)$$

An alternative form is:

$$\begin{aligned} & \bar{M}_{X*B*X^{-1}}^{-1} \hat{M}_X \hat{M}_B \hat{M}_{X^{-1}} \\ &= \frac{1}{4n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \bar{M}_{X*B*X^{-1}}^{-1} X_i B_j X_k^{-1} \bar{M}_{X*B*X^{-1}}^{-1} X_i B_j X_k^{-1} + \frac{3}{4} \mathbb{I} \end{aligned} \quad (48)$$

Error Metric 1:

$$error_R = ||\log(R_{M_A*X}^T R_{M_X*B})||_2 \quad (49)$$

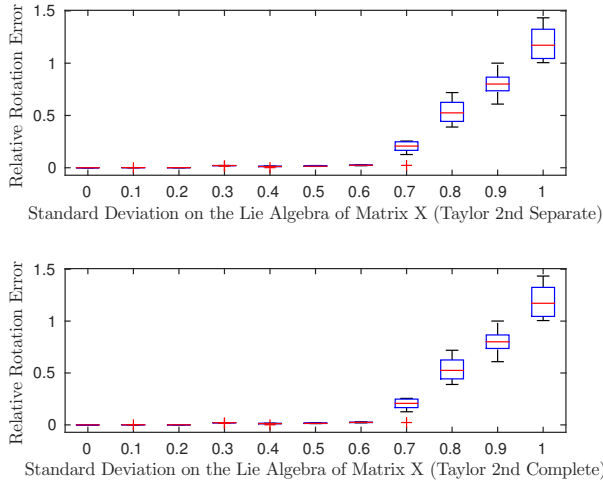


Fig. 13. Rotation Errors of 1st (up) and 2nd (down) Order Approximation With Respect to Different Noise Levels. Setup : (1) Number of Trials for Each Noise Level $n_t = 10$. (2) Number of Sample Points for Each Trial $n_s = 200$.

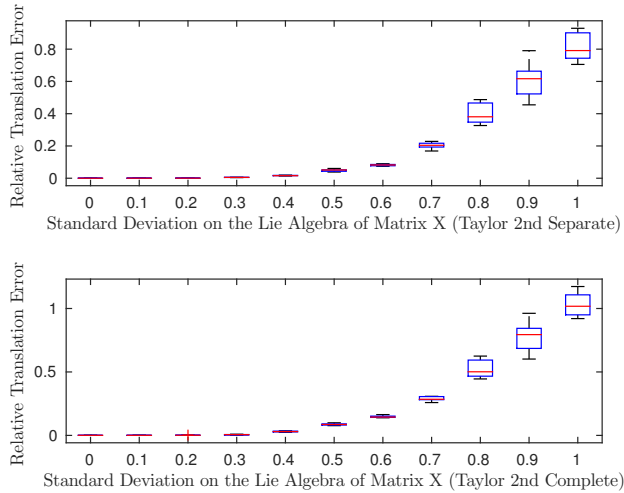


Fig. 14. Translation Errors of 1st (up) and 2nd (down) Order Approximation With Respect to Different Noise Levels. Setup : (1) Number of Trials for Each Noise Level $n_t = 10$. (2) Number of Sample Points for Each Trial $n_s = 200$.

$$error_t = \frac{\|t_{M_{A \times X}} - t_{M_{X \times B}}\|_2}{\|t_{M_{X \times B}}\|} \quad (50)$$

Error Metric 2:

$$error_R = \|\log(R_{M_A}^T R_{M_{X \times B \times X^{-1}}})\|_2 \quad (51)$$

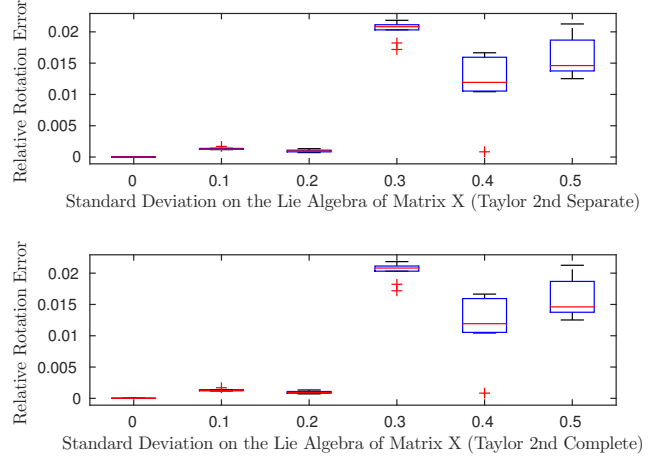


Fig. 15. A Closer Look at Fig.3.3 with $\sigma = 0, 0.1, \dots, 0.5$

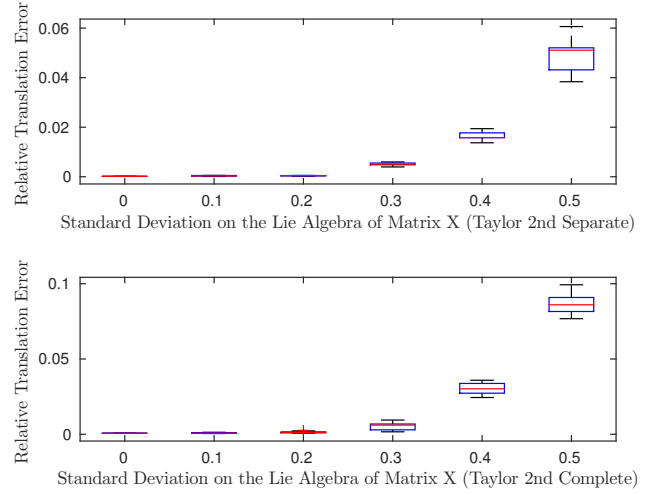


Fig. 16. A Closer Look at Fig.?? with $\sigma = 0, 0.1, \dots, 0.5$

$$error_t = \frac{\|t_{M_A} - t_{M_{X \times B \times X^{-1}}}\|_2}{\|t_{M_{X \times B \times X^{-1}}}\|} \quad (52)$$

where $M_{X \times B \times X^{-1}} = M_X M_B M_X^{-1}$ according to the old Batch method.

6 Problem

How to fit the information of covariance is still unknown, but the definition of the mean itself contains some information that is related to its covariance which might or might not have had a positive impact on the computation of the mean.

How to solve for X given $\{A\}$ and $\{B\}$ is still unclear. An initial guess of X might be needed for Eq.(43) to calculate $M_{1 \times 2}$ and as well as $M_{2 \times 3}$. Then $M_{A \times X} = M_{X \times B}$ or

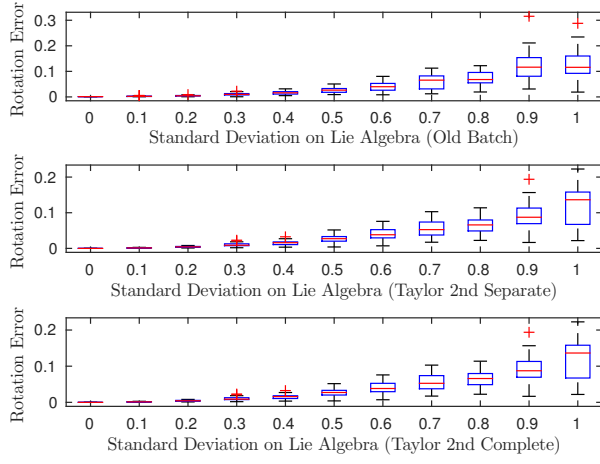


Fig. 17. AX and XB Error Comparison of Translation With Respect to Different Noise Levels

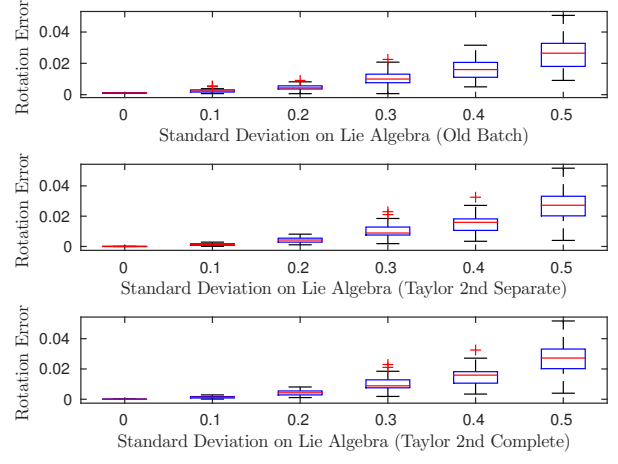


Fig. 19. AX and XB Error Comparison of Translation With Respect to Different Noise Levels

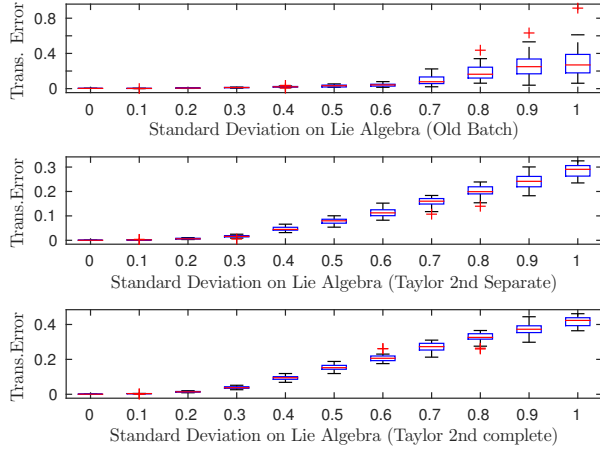


Fig. 18. AX and XB Error Comparison of Translation With Respect to Different Noise Levels

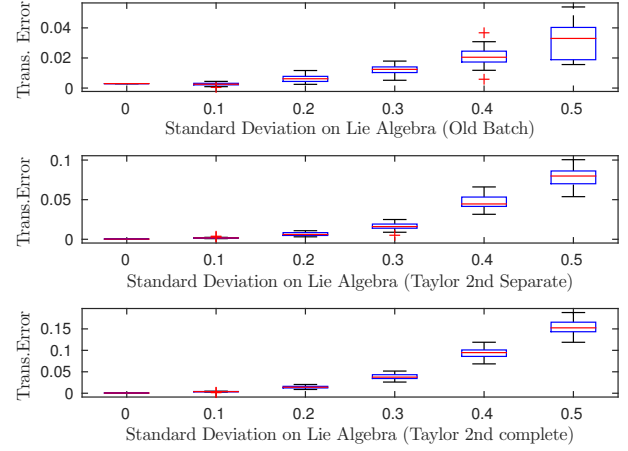


Fig. 20. AX and XB Error Comparison of Translation With Respect to Different Noise Levels

$M_A = M_{X*B*X^{-1}}$ is treated as an object function to optimize on X which is M_2 here. Moreover, though only 1st order approximations of A , B and X showed up in Eq.(43), it might be still possible to replace A and B with their 2nd order approximations to achieve a better X

7 Fourier Transform on Functions of SE(3)

Note: this is possible future work and will mostly likely be a separate paper.

For PDF version of $AX = XB$ as:

$$\rho_A(g) = (\rho_X * \rho_B * \rho_{X^{-1}})(g) \quad (53)$$

The equivalent form under Fourier transformation on both

sides is:

$$\hat{\rho}_A(g) = (\hat{\rho}_X * \hat{\rho}_B * \hat{\rho}_{X^{-1}})(g) \quad (54)$$

The above equation can be used given any clouds of A and B without the assumption of highly focused distribution.

Define a Gaussian by defining a diffusion equation and tuning its parameters. (Read Chapter 12 and 20 of the blue book. Tim Barfort Covariance Propagation. Corrected version of Wang's paper.)

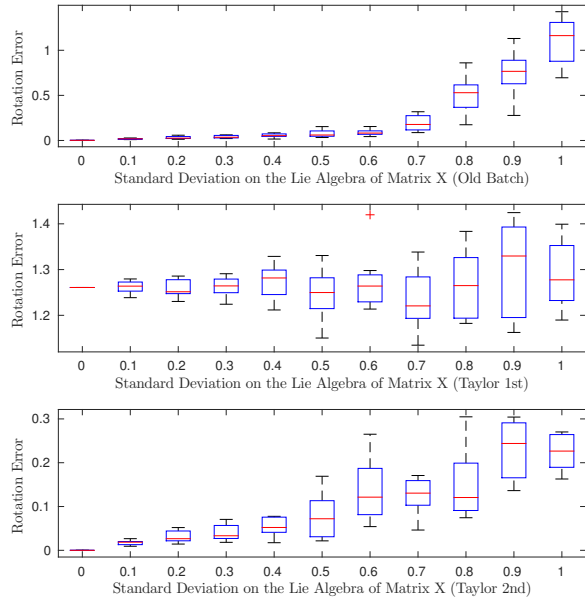


Fig. 21. A and XBX^{-1} Error Comparison of Translation With Respect to Different Noise Levels

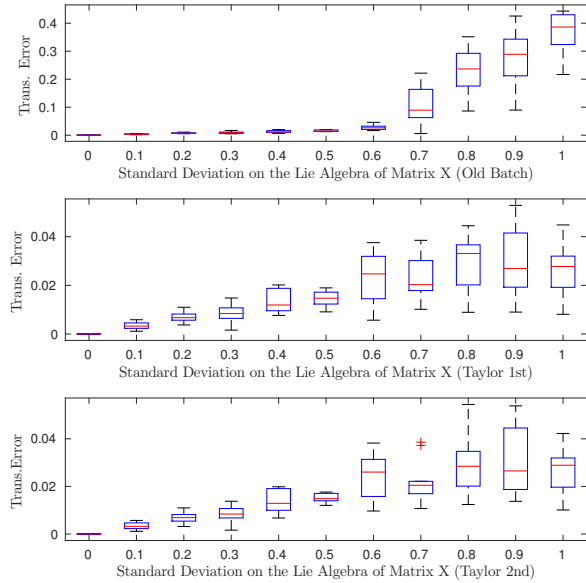


Fig. 22. A and XBX^{-1} Error Comparison of Translation With Respect to Different Noise Levels