Algorithms that will work + other patches

1
2
3
4
5
6

*** this is a completed version of the theorem and other material provided previously + new stuff. Please integrate. And somebody needs to code up the algorithms in Section 9 and 10. Let's all meet Monday morning and go over it. ***

Theorem 6.1. *If $f_B(H) \in (L^1 \cap L^2)(SE(3))$ is a symmetric function and $K \in SE(3)$ and $f_X \in (L^1 \cap L^2)(SE(3))$ are arbitrary, then:*
(a) $f_B(H)$ has its mean at the identity
(b) $(f_X * f_B * f_{X^{-1}})(H)$ has its mean at the identity
(c) $f_B(K^{-1}HK)$ has its mean at the identity.

Proof. (a) By definition,

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

and so

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

Letting $K = H^{-1}$, and using the same operations as in Theorem **??**, we arrive at

$$\int_{SE(3)} \log(M_B K) f_B(K)dK = \mathbb{O},$$

indicating that if M_B is unique, then $M_B = M_B^{-1}$ and so $M_B = \mathbb{I}$. Uniqueness can be guaranteed if Σ is sufficiently small, which in turn can be guaranteed in our application if all samples $\{A_i\}$ and $\{B_j\}$ are replaced by fractional powers $\{A_i^{1/n}\}$ and $\{B_j^{1/n}\}$.

(b) By definition, M, the mean of $(f_X * f_B * f_{X^{-1}})(H)$ must satisfy

$$\int_{SE(3)} \log(M^{-1}H)(f_X * f_B * f_{X^{-1}})(H)dH = \mathbb{O}.$$

By showing that $(f_X * f_B * f_{X^{-1}})(H)$ is symmetric, the fact that $M = \mathbb{I}$ will follow from (a). Using the definition of the convolution integral given in the Appendix, and expanding out both convolutions,

$$\begin{aligned}
&(f_X * f_B * f_{X^{-1}})(H) = \\
&\int_{SE(3)} \int_{SE(3)} f_X(J) f_B(J^{-1}K) f_{X^{-1}}(K^{-1}H) dJdK \\
= &\int_{SE(3)} \int_{SE(3)} f_{X^{-1}}(J^{-1}) f_B(K^{-1}J) f_X(H^{-1}K) dJdK
\end{aligned}$$

Here we have used the fact that f_B is symmetric. By the same logic,

$$\begin{aligned}
&(f_X * f_B * f_{X^{-1}})(H^{-1}) = \\
&\int_{SE(3)} \int_{SE(3)} f_X(J') f_B(J'^{-1}K') f_{X^{-1}}(K'^{-1}H^{-1}) dJ'dK'
\end{aligned}$$

We have the freedom to define $J' = H^{-1}K$ and $J^{-1} = K'^{-1}H^{-1}$. The proposition will then be true if the remaining term, $K^{-1}J = J'^{-1}K'$, holds. Indeed,

$$J'^{-1}K' = (H^{-1}K)^{-1}H^{-1}J = K^{-1}J,$$

completing the proof of (b).

Finally, (c) follows from (b) by letting $f_X(H) = \delta_K(H)$.

7 THE BATCH METHOD WITH NOISE IN A's and B's

If both sensors are reliable, the noise-free method developed in Section **??** is a reasonable approach. If one sensor has significantly more noise than the other, then the method developed in Section **??** can be used. Here we consider how to extend these methods to the case where there is noise in both the sets A and B and we don't know which sensor is more reliable. Therefore, we introduce a continuous power, $p \in [0,1]$, and write the original $AX = XB$ equation as

$$X^p A X^{-p} = X^{p-1} B X^{1-p}.$$

This leads to convolution equations of the form

$$(f_{(X,p)} * f_A * f_{(X^{-1},p)})(H) = (f_{(X^{-1},1-p)} * f_A * f_{(X,1-p)})(H).$$

Here we will seek not only M_X and Σ_X, but also p, which apportions the unknown amount of error to A or to B. When $p = 1/2$, both sensors have equivalent amounts of error, whereas when $p = 0$, this is the case discussed in the previous section.

We hypothesize that $f_{(X,p)}(H) = f_{X^p}(H)$ and $M_{X^p} = (M_X)^p$. This allows us to constrain $f_{(X,p)}(H)$, and hence $f_{(X,-p)}(H)$, by recursively solving $f_{(X,p)} * f_{(X,p)} = f_{(X,2p)}$ until $2p = 1$, thereby relating Σ_{X^p} and Σ_X. Note also that $f_{(X,-p)} * f_{(X,p)}$ and $f_{(X,p)} * f_{(X,-p)}$ are symmetric functions (and hence have means at the identity) and combining the results of * and **, they have covariances * and **.

8 Relating the Covariances of a PDF and Its Symmetrized Version

Recall that given an arbitrary pdf $f(H)$ with mean M and covariance Σ, a symmetrized version is defined as

$$\tilde{f}(H) = \frac{1}{2}(f(H) + f(H^{-1})).$$

Let $\tilde{M}$ and $\tilde{\Sigma}$ denote the mean and covariance of $\tilde{f}(H)$. Given that symmetrization of a pdf puts the mean of the result at the identity, we already know that $\tilde{M} = \mathbb{I}$. The remaining question to ask is what the relationship is between the covariances of the original and symmetrized versions of a pdf? This is answered here.

From the definition of covariance and the invariance of the integral over $SE(3)$ under inversions, and the fact that $\tilde{M} = \mathbb{I}$,

$$\tilde{\Sigma} = \int_{SE(3)} \log^{\vee}(H)[\log^{\vee}(H)]^T \tilde{f}(H)dH$$

simplifies as

$$\tilde{\Sigma} = \int_{SE(3)} \log^{\vee}(H)[\log^{\vee}(H)]^T f(H)dH.$$

This is not to be confused with

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

where

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

Though there appears to be no simple exact relationship between $\tilde{\Sigma}$ and Σ, in the case when $\|\Sigma\|$ and $\|M\|$ are both reasonably small, an approximate relationship can be constructed by using the Baker-Campbell-Hausdorf formula to expand out $\log^{\vee}(M^{-1}H)$. This was done in [**?**], and the result (modulo different notation) is

$$\tilde{\Sigma} = \Sigma + (\log^{\vee} M)(\log^{\vee} M)^T + \frac{1}{2}\left(\Sigma\, ad^T(\log M) + ad(\log M)\,\Sigma\right) \tag{1}$$

If (M, Σ) is known a priori, this means that $\tilde{\Sigma}$ can be computed from them. And since the computation of $\tilde{\Sigma}$ is exact and easy, (1) can be used as a consistency check on the accuracy of M and Σ by computing the norm of the difference of both sides.

9 An Algorithm Based on Incremental Linearization

By using the 'no-noise batch method' of Section **??**, we obtain an initial estimate for M_X, called M_{X_0}. This can then be substituted back into (**??**), which can be solved for an initial estimate of Σ_X, called Σ_{X_0}. Here we propose an update scheme that solves for small updates of the form

$$M_X = M_{X_0}(\mathbb{I} + Z) \text{ and } \Sigma_X = \Sigma_{X_0} + S \tag{2}$$

where $Z \in se(3)$ and $S = S^T \in \mathbb{R}^{6\times 6}$ are postulated to be small adjustments. Here

$$Z = \sum_{i=1}^{6} z_i E_i \text{ and } S = \sum_{k=1}^{21} s_i \mathcal{E}_i \tag{3}$$

where E_i are the natural unit basis elements for the Lie algebra $se(3)$ and $\{\mathcal{E}_i\}$ is a basis for the set of 6×6 real symmetric matrices. These include matrices that have a single 1 on the diagonal and zeros elsewhere, as well as those that have a pair of 1's symmetrically located off diagonal, and zeros elsewhere. Substituting (2) back into (**??**) and (**??**) and using the properties of $Ad(\cdot)$ respectively give[1]

$$M_A M_{X_0}(\mathbb{I} + Z) = M_{X_0}(\mathbb{I} + Z) M_B \tag{4}$$

and

$$\begin{aligned} &\Sigma_{X_0} + S + Ad(M_B^{-1})(\Sigma_{X_0} + S) Ad^T(M_B^{-1}) = \\ &(\mathbb{I} - ad(Z)) Ad(M_{X_0}^{-1}) \Sigma_A Ad^T(M_{X_0}^{-1})(\mathbb{I} - ad^T(Z)) - \Sigma_B. \end{aligned} \tag{5}$$

where

$$ad(Z) = \begin{pmatrix} \Omega & \mathbb{O} \\ V & \Omega \end{pmatrix} \text{ when } Z = \begin{pmatrix} \Omega & \mathbf{v} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

and $\mathbf{v} = V^{\vee}$. Ad and ad are related by the exact (and well-known) expression

$$Ad(\exp Z) = \exp(ad(Z)).$$

[1] If the A and B data are symmetrized, then (4) can be ignored, as it provides no constraint.

Upon substituting (3) into (4) and (5), the result can be rearranged into a system of linear equations of the form

$$J\begin{pmatrix}\mathbf{s}\\ \mathbf{z}\end{pmatrix} = \mathbf{c} \tag{6}$$

where the matrix J and the vector $\mathbf{c}$ are known. But typically J will have more columns than rows, or will not be full rank. This indicates that the problem is not fully constrained. As a result, we seek the solution of smallest magnitude. This can be done, for example, by using the SVD to invert J. When the noise levels are low, this can be done a single time rather than iteratively. However, this does not provide the flexibility to incorporate different weights. However, if J in (6) is full rank (or if both sides of (6) can be row-reduced to result in a full-rank sub-matrix J and corresponding reduced $\mathbf{c}$, then it is known from the field of redundant manipulator inverse kinematics [5, 6] that a solution of the form

$$\begin{pmatrix}\mathbf{s}\\ \mathbf{z}\end{pmatrix} = W^{-1}J^T(JW^{-1}J^T)^{-1}\mathbf{c}$$

will minimize the quadratic cost

$$\frac{1}{2}\begin{pmatrix}\mathbf{s}\\ \mathbf{z}\end{pmatrix}^T W \begin{pmatrix}\mathbf{s}\\ \mathbf{z}\end{pmatrix}$$

while exactly satisfying the linear constraint in (6).

10 Alternative Algorithms Based on the Lie-Group Structure of the Problem

The combination of (**??**) and (**??**) can be thought of abstractly as a system of matrix equations of the form

$$F(M_X, \Sigma_X) = \mathbb{O} \tag{7}$$

where M_X, Σ_X are unknown. We have already heavily used the fact that $M_X \in SE(3)$, which is a Lie group. The set of symmetric positive definite 6×6 matrices can be viewed as a homogeneous space on which $T \in GL(6) = GL(6,\mathbb{R})$ acts as $T\cdot\Sigma = T\Sigma T^T$. Then, starting with an initial guess (M_{X_0}, Σ_{X_0}), the problem of finding zeros of (7) can be reduced to one finding minima of a scalar cost function defined as the (squared) magnitude of this matrix equation. But since we have already observed that the equations can be degenerate, we seek solutions to

$$c(M_X, \Sigma_X) \doteq \|F(M_X,\Sigma_X)\|^2_{W_0} + \|M_X\|^2_{W_1} + \|\Sigma_X\|^2_{W_2} \tag{8}$$

where $\{W_i\}$ are an appropriate set of weighting matrices. By choosing W_1 and W_2 to be large enough, $c(M_X, \Sigma_X)$ will have a unique minimum. This problem can be formulated as a minimization on the product space of $SE(3)$ and $GL(6)/O(6)$ (the symmetric positive-definite matrices), or on the product group $SE(3)\times GL(6)$ by letting

$$(M_X, \Sigma_X) = (M_{X_0}H, T\Sigma_{X_0}T^T) \text{ where } (H,T)\in SE(3)\times GL(6)$$

are unknown. The benefit of this approach is that the added structure of being in a Lie group means that we can apply methods to $g = (H,T) \in SE(3)\times GL(6) = G$ that have been specificially formulated for minimization on Lie groups.

Alternatively, following ([1, 2]), the set of symmetric positive definite matrices can be endowed with an Abelian group operation

$$\Sigma_1 \circ \Sigma_2 \doteq \exp(\log\Sigma_1 + \log\Sigma_2).$$

** actually, I'm not 100 percent sure I believe their result, since it is not clear to me that the eigenvalues of $\exp(\log\Sigma_1 + \log\Sigma_2)$ will be positive when $[\Sigma_1, \Sigma_2] \neq \mathbb{O}$. **

Regardless, the problem can be posed as a minimization on a Lie group, G with group operation $\circ$. Let $\{E_i\}$ denote a basis for the corresponding Lie algebra, normalized so that $(E_i, E_j) = \delta_{ij}$. Then gradient descent can be formulated as follows: Update an initial value $g_0 \in G$ (e.g. the identity) as

$$g_1 = g_0 \circ \exp\left(-\varepsilon\sum_{i=1}^n (\tilde{E}_i c)(g)E_i\right) \tag{9}$$

where ε is a small update amount and

$$(\tilde{E}_i c)(g) = \frac{d}{dt}\, c(g\circ\exp(tE_i))\big|_{t=0} \approx \frac{c(g\circ\exp(\Delta E_i)) - c(g)}{\Delta}$$

where the right-hand-side is a finite-difference approximation with Δ a small positive real number. The trouble with gradient descent is that it requires a choice of step size, ε.

Alternatively, a Newton-like algorithm akin to those described in [**?**] can be used. But the added mathematical structure afforded by staying in the Lie-group setting makes computations a little easier than on a more general manifold. Following [**?**], the Taylor series of a smooth function on a Lie group can be expressed as

$$c(g_0\circ\exp(Z)) = c(g_0) + \sum_{i=1}^n z_i(\tilde{E}_i c)(g_0) + \frac{1}{2}\sum_{i,j=1}^n z_i z_j(\tilde{E}_i\tilde{E}_j c)(g_0) + \cdots$$

where $Z = \sum_{i=1}^n z_i E_i$. This can be written as

$$c(\mathbf{z} = c_0 + \mathbf{v}^T\mathbf{z} + \frac{1}{2}\mathbf{z}^T C\mathbf{z}$$

where $\mathbf{v}$ is the gradient vector and C is the Hessian matrix. Minimizing this quadratic cost with respect to $\mathbf{z}$ gives

$$\mathbf{z} = -C^{-1}\mathbf{v}.$$

The benefit of this approach over gradient descent is that no choice for ε is required, and when iterated, the convergence is supposed to be faster.

*** more new refs ***

References

[1] * GC will fill

[2] * GC will fill

[3] * cite theorems about sum of noncommuting logs

[4] Gwak, S., Kim, J., Park, F. C., "Numerical optimization on the Euclidean group with applications to camera calibration," *IEEE Transactions on Robotics and Automation*, 19(1):65-74, 2003.

[5] Klein, C. A., Huang, C. H., "Review of pseudoinverse control for use with kinematically redundant manipulators," *IEEE Transactions on Systems, Man and Cybernetics*, 2:245–250, 1983.

[6] Roberts, R. G., Maciejewski, A. A., "Repeatable generalized inverse control strategies for kinematically redundant manipulators," *IEEE Transactions on Automatic Control*, 38(5):689–699, 1993.