Note: pose measurements are Non-mutable.

> Given a set of relative pose "measurements", $\{T_{kl}, \Sigma_{kl}\}$, and a stationary frame 0, $\underrightarrow{\mathcal{F}}_0$. There are K poses in total, i.e., $k, l \in \mathbb{Z}$
$k, l \in [0, K]$.

We assume each relative pose measurement is drawn from a Gaussian (on SE(3)) that has a mean and variance of $\{\bar{T}_{kl}, \Sigma_{kl}\}$:

> So a pose measurement for the k^{th} frame relative to the l^{th} frame, T_{kl} can be thought of a random sample with.

$$T_{kl} = \underbrace{\exp(\xi_{kl}^{\wedge})}_{\text{noise}} \cdot \underbrace{\bar{T}_{kl}}_{\text{mean of Gaussian}}, \qquad (1)$$

Note: the T_{kl}'s are given values, we cannot change these measurements. The $\bar{T}_{kl}$ is the mean of underlying prob. dist., thus, we have no access to it.

which has a form of the mean is corrupted by some measurment noise, where

$$\xi_{kl} \sim \mathcal{N}(\vec{0}, \Sigma_{kl}).$$

> The problem will be set up as a batch Maximum Likelihood problem. For each measurement, we formulate an error term as:

$$\vec{e}_{kl}(\vec{x}) = \ln\left(\bar{T}_{kl}\,(T_k T_l^{-1})^{-1}\right)^{\vee} = \ln\left(\bar{T}_{kl}\, T_l \cdot T_k^{-1}\right)^{\vee}, \qquad (2)$$

where, $\vec{x} = \{T_1, \cdots, T_K\}$, and each T_k is a transformation matrix representing pose of $\underrightarrow{\mathcal{F}}_k$ relative to $\underrightarrow{\mathcal{F}}_0$.

> Note, in the above setting, each error term (or residual term) $\vec{e}_{kl}(\vec{x})$ is a 6×1 vector, and it can be thought of as with the form of (with a loose notation)

$$\text{estimated value} \boxminus \text{true value} \quad \overset{\longleftrightarrow}{\text{analogue}} \quad T_{truth} \cdot T_{estimated}^{-1},$$

one may think of the $\boxminus$ symbol as the "minus" operator (in scalar arithmetic). Since the direct element-wise minus operation on transformation matrices will not result in a valid transformation matrix, the $\boxminus$ is an analogue of such operation.

> By adopting the SE(3)-sensitive perturbation scheme, we have

$$T_k = \exp(\epsilon_k^\wedge)\, T_{op,k}, \qquad (3)$$

where $T_{op,k}$ is the operating point and ϵ_k is a 6x1 small perturbation. The <u>unknowns</u> that we are trying to find are the T_k's.

> By inserting (3) into (2) to substitute T_k and T_l:

$$\vec{e}_{kl}(\vec{x}) = \ln\left(\bar{T}_{kl} \left(\exp(\epsilon_l^\wedge) \cdot T_{op,l}\right) \left(\exp(\epsilon_k^\wedge) \cdot T_{op,k}\right)^{-1} \right)^\vee$$

$$= \ln\left(\bar{T}_{kl} \exp(\epsilon_l^\wedge)\, T_{op,l}\, T_{op,k}^{-1} \left(\exp(\epsilon_k^\wedge)\right)^{-1} \right)^\vee$$

Barfoot's SE(3) table (in CH 7)

$$= \ln\left(\bar{T}_{kl} \exp(\epsilon_l^\wedge) \cdot T_{op,l} \cdot T_{op,k}^{-1} \exp(-\epsilon_k^\wedge) \right)^\vee. \qquad (4)$$

> In order to continue to simplify (4), we first recall some properties of SE(3)

$$T \exp(\xi^\wedge) T^{-1} = \exp\left((\mathrm{Ad}(T) \cdot \xi)^\wedge \right) = \exp\left((\mathcal{T} \xi)^\wedge \right). \qquad (5)$$

> Let $A = T^{-1}$, which is also on SE(3), so (5) can also be applied on A:
(inversion of a transformation matrix is still a transformation)

$$A \exp(\xi^\wedge) A^{-1} = \exp\left((\mathrm{Ad}(A) \cdot \xi)^\wedge \right)$$

$$\Rightarrow T^{-1} \cdot \exp(\xi^\wedge)\, T = \exp\left((\mathrm{Ad}(T^{-1}) \cdot \xi)^\wedge \right)$$

$$\Rightarrow \begin{cases} T^{-1} \cdot \exp(\xi^\wedge) = \exp\left((\mathrm{Ad}(T^{-1})\, \xi)^\wedge \right) \cdot T^{-1} & \text{(i) obtained by multiplying } T^{-1} \text{ on right for both sides} \\ \exp(\xi^\wedge) \cdot T = T \cdot \exp\left((\mathrm{Ad}(T^{-1})\, \xi)^\wedge \right) & \text{(ii) obtained by multiplying } T \text{ on left for both sides} \end{cases}$$

(6)

Note:

$\mathcal{T}_{op,l} = Ad(T_{op,l})$

$\mathcal{T}_{op,k} = Ad(T_{op,k})$

P_3

> Back to (2) and substitutings (6) into (2):

$$\vec{e}_{kl}(\vec{x}) = \ln\left(\bar{T}_{kl} \cdot \exp(\epsilon_l^\wedge)\, T_{op,l}\, T_{op,k}^{-1} \cdot \exp(-\epsilon_k^\wedge)\right)^\vee$$

$$\overset{(ii)}{=} \ln\left(\bar{T}_{kl} \cdot T_{op,l} \cdot \exp\left((Ad(T_{op,l}^{-1}) \cdot \epsilon_l)^\wedge\right) \cdot T_{op,k}^{-1} \cdot \exp(-\epsilon_k^\wedge)\right)^\vee$$

$$\overset{(ii)}{=} \ln\left(\bar{T}_{kl} \cdot T_{op,l} \cdot T_{op,k}^{-1} \exp\left((Ad(T_{op,k}) Ad(T_{op,l}^{-1}) \cdot \epsilon_l)^\wedge\right) \cdot \exp(-\epsilon_k^\wedge)\right)^\vee$$

$$= \ln\left(\bar{T}_{kl}\, T_{op,l} \cdot T_{op,k}^{-1} \exp\left((\mathcal{T}_{op,k}\, \mathcal{T}_{op,l}^{-1} \cdot \epsilon_l)^\wedge\right) \cdot \exp(-\epsilon_k^\wedge)\right)^\vee, \quad (7)$$

where $\mathcal{T}_{op,k} = Ad(T_{op,k})$, $\mathcal{T}_{op,l}^{-1} = Ad(T_{op,l}^{-1})$. Based on Barfoot's SE(3) Table, we can approximate a transformation matrix as:

$$T = \exp(\xi^\wedge) \approx I_{4\times4} + \xi^\wedge \iff I_{4\times4} + \xi^\wedge \approx \exp(\xi^\wedge) = T \; (8)$$

> In this way, we can approximate (7) as:

$$\vec{e}_{kl}(\vec{x}) = \ln\left(\bar{T}_{kl}\, T_{op,l}\, T_{op,k}^{-1} \exp\left((\mathcal{T}_{op,k}\, \mathcal{T}_{op,l}^{-1} \epsilon_l)^\wedge\right) \cdot \exp(-\epsilon_k^\wedge)\right)^\vee$$

$$\approx \ln\left(\bar{T}_{kl} \cdot T_{op,l}\, T_{op,k}^{-1} \left(I_{4\times4} + (\mathcal{T}_{op,k}\, \mathcal{T}_{op,l}^{-1} \epsilon_l)^\wedge\right)\left(I - \epsilon_k^\wedge\right)\right)^\vee$$

$$= \ln\left(\bar{T}_{kl}\, T_{op,l}\, T_{op,k}^{-1} \left(I_{4\times4} + (\mathcal{T}_{op,k}\, \mathcal{T}_{op,l}^{-1} \epsilon_l)^\wedge - \epsilon_k^\wedge \underbrace{- (\mathcal{T}_{op,k}\, \mathcal{T}_{op,l}^{-1} \epsilon_l)^\wedge \epsilon_k^\wedge}_{\text{omit this product of two small terms}}\right)\right)^\vee$$

$$\approx \ln\left(\bar{T}_{kl}\, T_{op,l} \cdot T_{op,k}^{-1} \left(I_{4\times4} + (\mathcal{T}_{op,k}\, \mathcal{T}_{op,l}^{-1} \epsilon_l)^\wedge - \epsilon_k^\wedge\right)\right)^\vee$$

Note: $(\vec{x}+\vec{y})^\wedge = \vec{x}^\wedge + \vec{y}^\wedge$

$$= \ln\left(\bar{T}_{kl}\, T_{op,l}\, T_{op,k}^{-1} \left(I_{4\times4} + (\mathcal{T}_{op,k}\, \mathcal{T}_{op,l}^{-1} \epsilon_l - \epsilon_k)^\wedge\right)\right)^\vee$$

cont).

$$\vec{e}_{kl}(\vec{x}) \approx \ln\left(\bar{T}_{kl}\, T_{op,l}\, T_{op,k}^{-1}\left(I_{4\times 4} + (\mathcal{T}_{op,k}\mathcal{T}_{op,l}^{-1}\epsilon_l - \epsilon_k)^\wedge\right)\right)^\vee$$

$$\overset{(8)}{\approx} \ln\left(\bar{T}_{kl}\, T_{op,l}\, T_{op,k}^{-1} \exp\left((\mathcal{T}_{op,k}\mathcal{T}_{op,l}^{-1}\epsilon_l - \epsilon_k)^\wedge\right)\right)^\vee \qquad (9)$$

$\angle\, \vec{x}_{op} = \{T_{op,1}, \ldots, T_{op,K}\}$

> In (9), we can define $\exp(\vec{e}_{kl}(\vec{x}_{op})^\wedge) = \bar{T}_{kl}\, T_{op,l}\, T_{op,k}^{-1}$

$\Leftrightarrow \vec{e}_{kl}(\vec{x}_{op}) = \ln(\bar{T}_{kl}\, T_{op,l}\, T_{op,k}^{-1})^\vee$. Since $\bar{T}_{kl} T_{op,l} T_{op,k}^{-1}$ is a small term, we can further simplify (9) using (7.100a) in Barfoot's Book.

$$\vec{e}_{kl}(\vec{x}) \approx \ln\left(\exp\left((\vec{e}_{kl}(\vec{x}_{op}))^\wedge\right)\cdot \exp\left(\left(\underbrace{\mathcal{T}_{op,k}\mathcal{T}_{op,l}^{-1}\epsilon_l - \epsilon_k}_{\text{small term}}\right)^\wedge\right)\right)^\vee$$

$$\overset{(7.100a)}{\approx} \vec{e}_{kl}(\vec{x}_{op}) + \mathcal{J}(-\vec{e}_{kl}(\vec{x}_{op}))^{-1}\cdot(\mathcal{T}_{op,k}\mathcal{T}_{op,l}^{-1}\epsilon_l - \epsilon_k)$$

$$= \vec{e}_{kl}(\vec{x}_{op}) + \mathcal{J}(-\vec{e}_{kl}(\vec{x}_{op}))^{-1}\cdot \mathcal{T}_{op,k}\mathcal{T}_{op,l}^{-1}\epsilon_l - \mathcal{J}(-\vec{e}_{kl}(\vec{x}_{op}))^{-1}\epsilon_k$$

$$= \vec{e}_{kl}(\vec{x}_{op}) - G_{kl}\,\delta\vec{x}_{kl}, \qquad (10)$$

where,

$$G_{kl} = \left[-\mathcal{J}(-\vec{e}_{kl}(\vec{x}_{op}))^{-1}\mathcal{T}_{op,k}\mathcal{T}_{op,l}^{-1} \quad \mathcal{J}(-\vec{e}_{kl}(\vec{x}_{op}))^{-1}\right]_{6\times 12}$$

$$\delta\vec{x}_{kl} = \begin{bmatrix}\epsilon_l \\ \epsilon_k\end{bmatrix}_{12\times 1},$$

and $\mathcal{J} = \mathcal{J}_l$ is the left jacobian, and $\mathcal{J}(-\vec{e}_{kl}(\vec{x}_{op}))$ is the left Jacobian of $-\vec{e}_{kl}(\vec{x}_{op})$ w.r.t. ξ_{kl} (i.e., the lie algebra representation of the Transformation of k w.r.t l).

> By defining each individual error term, the overall objective function of the maximum likelihood problem is defined as: P5-

$$J(\vec{x}) = \frac{1}{2} \sum_{k,l} \vec{e}_{k,l}(\vec{x})^T \Sigma_{kl}^{-1} \vec{e}_{kl}(\vec{x}), \qquad (11)$$

where the total number of error terms involved in the summation is the number of relative pose measurements included in the pose graph. By substituting (10) into (11), we have

$$\underbrace{J(\vec{x})}_{[\text{scalar}]} \approx \frac{1}{2} \sum_{k,l} \left(\vec{e}_{kl}(\vec{x}_{op}) - G_{kl} \cdot \delta\vec{x}_{kl} \right)^T \Sigma_{kl}^{-1} \left(\vec{e}_{kl}(\vec{x}_{op}) - G_{kl} \cdot \delta\vec{x}_{kl} \right)$$

$$= \frac{1}{2} \sum_{k,l} \left(\vec{e}_{kl}(\vec{x}_{op})^T - \delta\vec{x}_{kl}^T \cdot G_{kl}^T \right) \Sigma_{kl}^{-1} \left(\vec{e}_{kl}(\vec{x}_{op}) - G_{kl} \cdot \delta\vec{x}_{kl} \right)$$

Note: Σ_{kl}^{-1} is symmetric in this setting

$$= \frac{1}{2} \sum_{kl} \left(\vec{e}_{kl}^T(\vec{x}_{op}) \Sigma_{kl}^{-1} - \delta\vec{x}_{kl}^T G_{kl}^T \Sigma_{kl}^{-1} \right) \left(\vec{e}_{kl}(\vec{x}_{op}) - G_{kl} \cdot \delta\vec{x}_{kl} \right)$$

$$= \frac{1}{2} \sum_{kl} \Big(\vec{e}_{kl}(\vec{x}_{op})^T \Sigma_{kl}^{-1} \vec{e}_{kl}(\vec{x}_{op}) - \vec{e}_{kl}(\vec{x}_{op})^T \Sigma_{kl}^{-1} G_{kl} \delta\vec{x}_{kl} - \delta\vec{x}_{kl}^T G_{kl}^T \Sigma_{kl}^{-1} \vec{e}_{kl}(\vec{x}_{op}) + \delta\vec{x}_{kl}^T G_{kl}^T \Sigma_{kl}^{-1} G_{kl} \cdot \delta\vec{x}_{kl} \Big)$$

$$= \underbrace{\frac{1}{2} \left(\sum_{kl} \vec{e}_{kl}(\vec{x}_{op})^T \Sigma_{kl}^{-1} \vec{e}_{kl}(\vec{x}_{op}) \right)}_{J(\vec{x}_{op})} - \frac{1}{2} \cdot 2 \sum_{k,l} \left(\vec{e}_{kl}^T(\vec{x}_{op}) \Sigma_{kl}^{-1} G_{kl} \delta\vec{x}_{kl} \right) + \frac{1}{2} \sum_{k,l} \left(\delta\vec{x}_{kl}^T G_{kl}^T \Sigma_{kl}^{-1} G_{kl} \delta\vec{x}_{kl} \right)$$

$$= J(\vec{x}_{op}) - \sum_{k,l} \left(\vec{e}^T \Sigma_{kl}^{-1} G_{kl} \delta\vec{x}_{kl} \right) + \frac{1}{2} \sum_{k,l} \left(\delta\vec{x}_{kl}^T G_{kl}^T \Sigma_{kl}^{-1} G_{kl} \delta\vec{x}_{kl} \right)$$

$$= J(\vec{x}_{op}) - \vec{b}^T \delta\vec{x} + \frac{1}{2} \delta\vec{x}^T A \cdot \delta\vec{x}, \qquad (12)$$

where

$$\vec{b} = \sum_{k,l} P_{kl}^T G_{kl}^T \Sigma_{kl}^{-1} \vec{e}_{kl}(\vec{x}_{op})$$

$$A = \sum_{kl} P_{kl}^T G_{kl}^T \Sigma_{kl}^{-1} G_{kl} P_{kl}$$

$$\delta\vec{x}_{kl} = P_{kl} \delta\vec{x},$$

(cont) and P_{kl} is a $[12 \times 6K]$ projection matrix to pick out the k-l P_6

12×1 perturbations, i.e., $\begin{bmatrix} \epsilon_l \\ \epsilon_k \end{bmatrix}$ from the full $[6K \times 1]$ perturbation vector,

$$\delta \vec{x}_{kl} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_K \end{bmatrix}_{6K\times 1}$$

> Note that each ϵ_k is a 6×1 vector., and each $\delta \vec{x}_{kl}$ consists of ϵ_k, ϵ_l, thus, having a dimension of 12×1. So, one may think of the P_{kl} matrix as two rows of blocks, where each block is an 6×6 matrix, and the l^{th} block in the first row is an identity matrix, while all other blocks in the first row are zeros. Similarly, the k^{th} block in the second block row of P_{kl} is an 6×6 identity matrix, where all other blocks are zero:

$$P_{kl} = \begin{bmatrix} O_{6\times 6} & \cdots & \overbrace{I_{6\times 6}}^{l^{th} \text{ block-column}} & \cdots & O_{6\times 6} & \cdots \\ O_{6\times 6} & \cdots & O_{6\times 6} & \cdots & \underbrace{I_{6\times 6}}_{k^{th} \text{ block-column}} & \cdots \end{bmatrix}_{12\times 6K} \qquad (13)$$

> Note that l does NOT necessarily be smaller the k. The example in (13) is just to show the form of the P matrix. However, $k \neq l$ is typical.

> Therefore, the $\vec{b}$ vector in (12) has a shape of $6K\times 1$; the A matrix in (12) is a $6K \times 6K$ matrix; and, the $\delta\vec{x}$ vector is a $6K\times 1$ vector; and the overall objective, $J(\vec{x})$, and $J(\vec{x}_{op})$ are both scalars.

> In this form, the approximate objective function is exactly quadratic is $\delta\vec{x}$, and we minimize the approximate objective (12) by taking the derivative:

$$\nabla_{\delta\vec{x}} J(\vec{x}) = \frac{\partial J(\vec{x})}{\partial \delta\vec{x}} = \underbrace{\frac{\partial J(\vec{x}_{op})}{\partial \delta\vec{x}}}_{\text{not a fn of } \delta\vec{x}} - \frac{\partial(\vec{b}^T \delta\vec{x})}{\partial \delta\vec{x}} + \frac{1}{2}\frac{\partial(\delta\vec{x}^T A \delta\vec{x})}{\partial \delta\vec{x}}$$

$$= 0 - \vec{b} + \left(\tfrac{1}{2}\times 2\right) A \cdot \delta\vec{x} \qquad (14)$$

> By setting (14) to zero to solve for the optimal perturbation $\delta\vec{x}^*$:

$$\vec{0} = -\vec{b} + A\,\delta\vec{x}^*$$

$$A\,\delta\vec{x}^* = \vec{b}. \qquad (15)$$

> The optimal perturbation will be a $6K \times 1$ vector as

$$\delta\vec{x}^* = \begin{bmatrix} \epsilon_1^* \\ \vdots \\ \epsilon_K^* \end{bmatrix},$$

where each ϵ_k^* is a 6×1 vector.

> The optimal perturbations will then be used to update the K poses w.r.t. the inertial stationary frame as

$$T_{op,k} \leftarrow \exp(\epsilon_k^{*\wedge})\, T_{op,k},$$

which will result in the updated pose is still belonging to $SE(3)$.

> The above updating process will iterate until certain stopping criterion is met.

> When the optimization is finished, we set $\hat{T}_{k0} = T_{op,k}$ to represent the k^{th} pose relative to pose 0, the stationary frame $\mathcal{F}_0$.

Notes from Implementation:

> In actual implementation, we may not follow exactly as the derived steps. One key reason is that, when there are many poses and observations, the Projection matrix P_{kl} is of the size $[12 \times 6K]$, and each measurement will require one of such matrix. However, since the majority of the entries in the P_{kl} matrix are zeros, and it is used to pick up the corresponding elements to perform addition operation. If we use the P_{kl} matrix in actual implementation, there will be many redundant operations of $0 + 0$, which do not contribute much to the final result. In addition, it can cause running out of memory if the problem is large.

> Therefore, in the implementation, we may calculate the

$$G_{kl}^T \Sigma_{kl}^{-1} \vec{e}_{kl}(\vec{x}_{op}) \qquad \text{for } \vec{b}$$

and

$$G_{kl}^T \Sigma_{kl}^{-1} G_{kl} \qquad \text{for } A$$

for the k-l^{th} measurement and use array index accessing to add these terms to the corresponding elements in the $\vec{b}$ vector and A matrix.