Supplementary Material to:
Direct Sparse Visual-Inertial Odometry using Dynamic Marginalization

Lukas von Stumberg, Vladyslav Usenko, Daniel Cremers

This is the supplementary material to the paper ”Direct Sparse Visual-Inertial Odometry using Dynamic Marginalization”. Here we provide parameter studies further evaluating the method (section 1), a way to measure the convergence of the scale (section 2) and details on how to compute the relative Jacobians (section 3).

1 Parameter Studies

In order to further evaluate the method we provide several parameter studies in this section. As in the result section of the main paper, we have run all variants 10 times for each sequence of the EuRoC-dataset and present either the accumulation of all results or the median of the 10 runs. First of all we show the dependence on the number of points in Fig. 1. The main takeaway of this should be that because of the addition of inertial data, we can significantly reduce the number of points without loosing tracking performance.

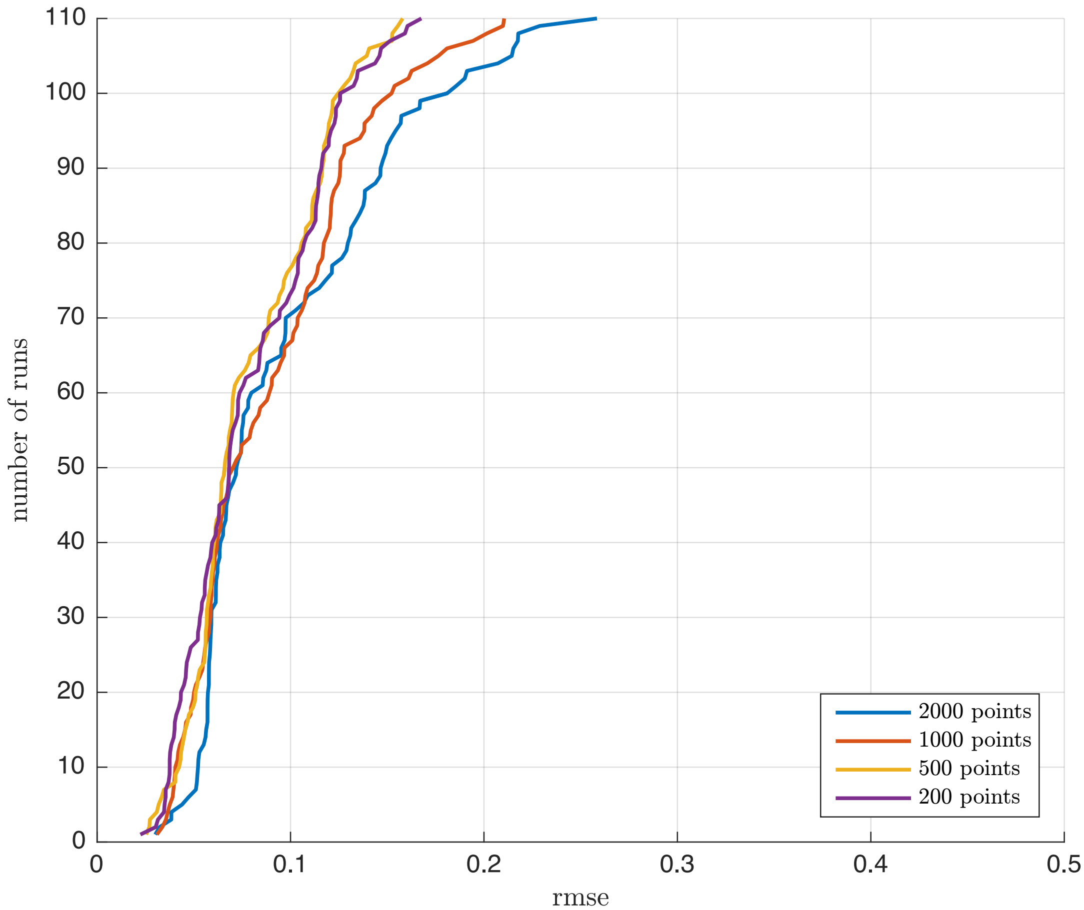


Fig. 1: RMSE when running the method (not in realtime) with different numbers of active points. Interestingly focusing on fewer (but more reliable points) actually improves the performance, both in precision and in runtime.

In order to evaluate the importance of the novel dynamic marginalization strategy we have replaced it with two alternatives (Fig. 2). For one example we have used normal marginalization instead where only one marginalization prior is used. Furthermore we have tried a simpler dynamic marginalization strategy, where we do not use M_{half}, but instead directly reset the marginalization prior with M_{visual}, as soon as the scale interval is exceeded. Clearly dynamic marginalization yields the most robust result. Especially on sequences with a large initial scale error, the other strategies do not work well. Fig. 3 shows the difference in the scale convergence. When using the simple marginalization strategy the marginalization prior gets reset directly to M_{visual} resulting in a slower scale convergence and oscillations (Fig. 3c and 3d). With a normal marginalization

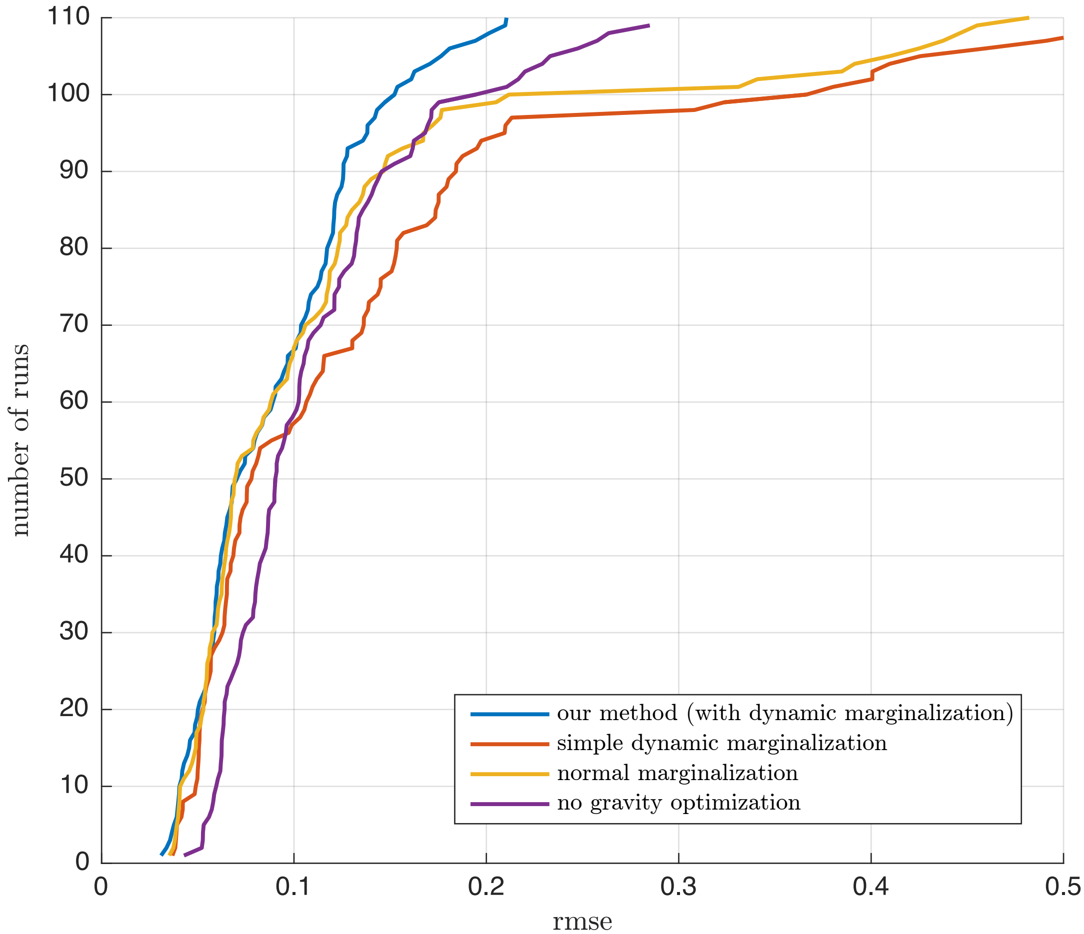


Fig. 2: The method run (in realtime) with different changes. "Simple dynamic marginalization" means that we have changed the marginalization to not use M_{half}, but rather replace M_{curr} directly with M_{visual}, when the scale interval is exceeded. For "normal marginalization" we have used the normal marginalization procedure with just one marginalization prior. For the pink line we have turned off the gravity direction optimization in the system (and only optimize for scale). This plot shows that the novelties presented in this paper, in particular the joint optimization with scale and gravity, and the dynamic marginalization procedure are important for the accuracy and robustness of the system.

the scale estimates overshoots and it takes a long time for the system to compensate for the factors with a wrong scaled that were initially marginalized (Fig. 3e and 3f). With our dynamic marginalization implementation however the scale converges much faster and with almost no overshoot or oscillations (Fig. 3a and 3b).

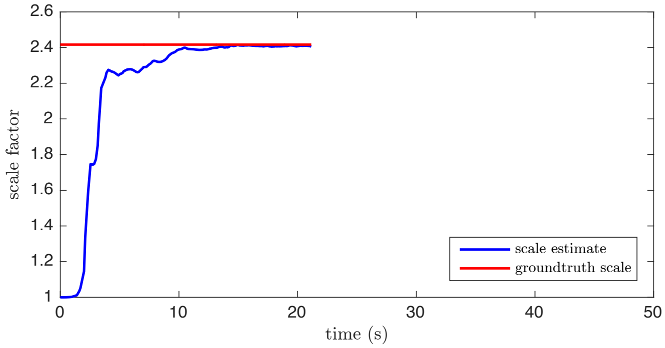
We have also disabled the joint optimization of gravity direction in the main system (Fig. 2). Clearly the simple initialization of gravity direction (described in section III-D) is not sufficient to work well, without adding gravity direction to the model.

2 Measuring Scale Convergence

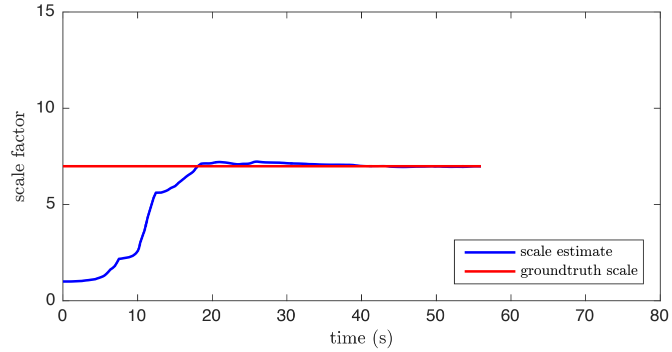
For many applications it is very useful to know when the scale has converged, especially as this can take an arbitrary amount of time in theory. The lack of this measure is also one of the main drawbacks of the initialization method described in [1]. Here we propose a simple but effective strategy. With s_i being the scale at time i and n being the maximum queue size (60 in all our experiments) we compute

$$c = \frac{\max_{j:=i-n+1}^{n} s_j}{\min_{j:=i-n+1}^{n} s_j} - 1 \tag{A.1}$$

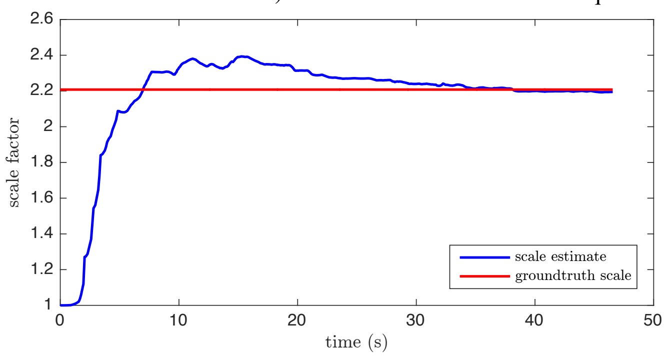
Fig. 4b shows a plot of this measure. When c is below a certain threshold c_{min} (0.005 in our experiments) we consider the scale converged and fix $\boldsymbol{\xi}_{m\_d}$. While this does not influence the accuracy of the system, fixing the scale might be useful for some applications.



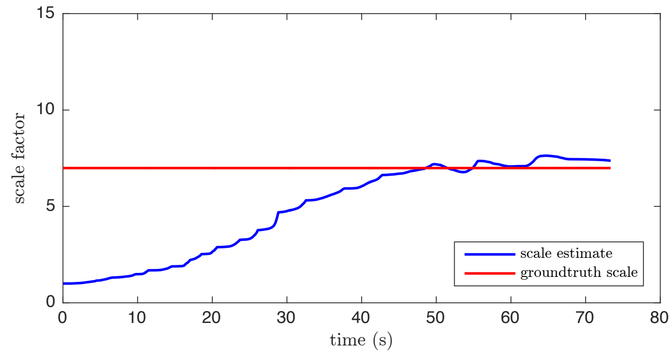
(a) Scale estimate for V2_03_difficult with our method. After 21 seconds our system considered the scale converged (using the measure defined in section 2) and fixed it for the rest of the sequence.



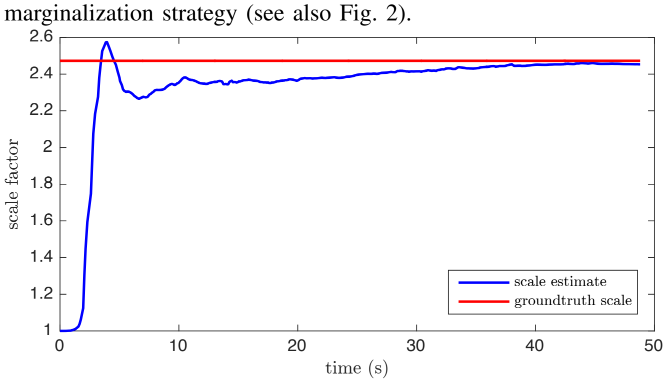
(b) Scale estimate for MH_04_difficult with our method



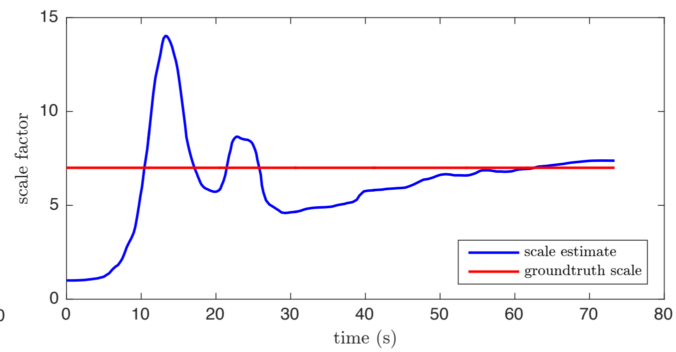
(c) Scale estimate for V2_03_difficult with the simple dynamic marginalization strategy (see also Fig. 2).



(d) Scale estimate for MH_04_difficult with the simple dynamic marginalization strategy.



(e) Scale estimate for V2_03_difficult with normal marginalization.



(f) Scale estimate for MH_04_difficult with normal marginalization.

Fig. 3: Scale estimates for the different marginalization strategies evaluated on V2_03_difficult (left) and MH_04_difficult (right). All figures show the median result of the 10 runs that were accumulated in Fig. 2.

3 Calculating the relative Jacobian

As mentioned in section III-F.1 we want to compute $\mathbf{J}_{\mathrm{rel}}$, such that (11) holds.

We convert between the poses using

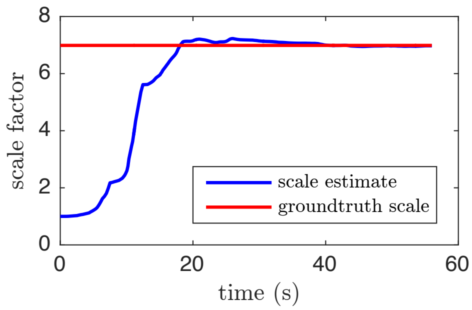
$$\boldsymbol{\xi}^{M}_{w\_imu} = \boldsymbol{\xi}_{m\_d} \boxplus \left(\boldsymbol{\xi}^{D}_{cam\_w}\right)^{-1} \boxplus \boldsymbol{\xi}^{-1}_{m\_d} \boxplus \boldsymbol{\xi}^{M}_{cam\_imu} =: \Psi(\boldsymbol{\xi}^{D}_{cam\_w}, \boldsymbol{\xi}_{m\_d}) \tag{A.2}$$

For any function $\Omega(\boldsymbol{\xi}) : \mathfrak{sim}(3) \rightarrow \mathfrak{sim}(3)$ we define the derivative $\frac{d\Omega(\boldsymbol{\xi} \boxplus \boldsymbol{\epsilon})}{d\boldsymbol{\epsilon}}$ implicitly using

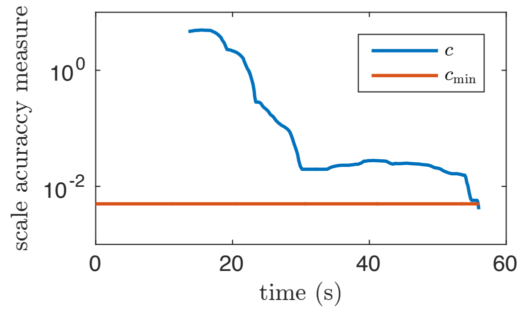
$$\Omega(\boldsymbol{\xi})^{-1} \boxplus \Omega\left(\boldsymbol{\xi} \boxplus \boldsymbol{\epsilon}\right) = \frac{d\Omega(\boldsymbol{\xi} \boxplus \boldsymbol{\epsilon})}{d\boldsymbol{\epsilon}} \cdot \boldsymbol{\epsilon} + \eta(\boldsymbol{\epsilon}) \cdot \boldsymbol{\epsilon} \tag{A.3}$$

where the error function $\eta(\boldsymbol{\epsilon})$ goes to 0 as $||\boldsymbol{\epsilon}||$ goes to 0.

Note that in principle there are three other ways to define this derivative (as you can place the increment with $\boldsymbol{\epsilon}$ as well as the multiplication with the inverse on either side). However it can be shown (see section 3.1) that only with this version



(a) Scale estimate



(b) Accuracy measure c.

Fig. 4: Scale estimate and scale accuracy measure c for MH_04_difficult (median result of 10 runs in terms of tracking accuracy).

the following chain rule holds for $f(\boldsymbol{\xi}) : \mathfrak{sim}(3) \to \mathbb{R}$

$$\frac{df\left(\Omega\left(\boldsymbol{\xi} \boxplus \boldsymbol{\epsilon}\right)\right)}{d\boldsymbol{\epsilon}} = \frac{df\left(\Omega\left(\boldsymbol{\xi}\right) \boxplus \boldsymbol{\delta}\right)}{\boldsymbol{\delta}} \cdot \frac{d\Omega\left(\boldsymbol{\xi} \boxplus \boldsymbol{\epsilon}\right)}{\boldsymbol{\epsilon}} \tag{A.4}$$

Using these definitions the relevant derivatives for Ψ can be computed (see section 3.2 and 3.3)

$$\frac{\partial \Psi(\boldsymbol{\xi}^D_{cam\_w} \boxplus \boldsymbol{\epsilon}, \boldsymbol{\xi}_{m\_d})}{\partial \boldsymbol{\epsilon}} = -\mathrm{Adj}\left(\mathbf{T}^{-1}_{cam\_imu} \cdot \mathbf{T}_{m\_d} \cdot \mathbf{T}_{cam\_w}\right) \tag{A.5}$$

$$\frac{\partial \Psi(\boldsymbol{\xi}^D_{cam\_w}, \boldsymbol{\xi}_{m\_d} \boxplus \boldsymbol{\epsilon})}{\partial \boldsymbol{\epsilon}} = \mathrm{Adj}(\mathbf{T}^{-1}_{cam\_imu} \cdot \mathbf{T}_{m\_d} \cdot \mathbf{T}_{cam\_w}) - \mathrm{Adj}(\mathbf{T}^{-1}_{cam\_imu} \cdot \mathbf{T}_{m\_d}) \tag{A.6}$$

Stacking these derivatives correctly we can compute a Jacobian $\mathbf{J}_{\text{rel}}$ such that

$$\mathbf{J}_{\text{imu}} = \mathbf{J}'_{\text{imu}} \cdot \mathbf{J}_{\text{rel}} \tag{A.7}$$

Using this we can finally compute

$$\mathbf{H}_{\text{imu}} = \mathbf{J}^T_{\text{rel}} \cdot \mathbf{H}'_{\text{imu}} \cdot \mathbf{J}_{\text{rel}} \text{ and } \boldsymbol{b}_{\text{imu}} = \mathbf{J}^T_{\text{rel}} \cdot \boldsymbol{b}'_{\text{imu}} \tag{A.8}$$

3.1 Proof of the chain rule

In this section we will proof the chain rule (A.4).

We define the multivariate derivatives implicitly:

$$f(\Omega(\boldsymbol{\xi}) \boxplus \boldsymbol{\delta}) - f(\Omega(\boldsymbol{\xi})) = \frac{df\left(\Omega\left(\boldsymbol{\xi}\right) \boxplus \boldsymbol{\delta}\right)}{\boldsymbol{\delta}} \cdot \boldsymbol{\delta} + \mu(\boldsymbol{\delta}) \cdot \boldsymbol{\delta} \tag{A.9}$$

where the error function $\mu(\boldsymbol{\delta})$ goes to 0 as $\boldsymbol{\delta}$ goes to zero. We can rewrite equation A.3 by multiplying $\Omega(\boldsymbol{\xi})$ from the left

$$\Omega\left(\boldsymbol{\xi} \boxplus \boldsymbol{\epsilon}\right) = \Omega(\boldsymbol{\xi}) \boxplus \left(\frac{d_l\Omega(\boldsymbol{\xi} \boxplus \boldsymbol{\epsilon})}{d\boldsymbol{\epsilon}} \cdot \boldsymbol{\epsilon} + \eta(\boldsymbol{\epsilon}) \cdot \boldsymbol{\epsilon}\right) \tag{A.10}$$

Using this we can compute

$$\begin{aligned}
&f(\Omega(\boldsymbol{\xi} \boxplus \boldsymbol{\epsilon})) - f\left(\Omega(\boldsymbol{\xi})\right) \overset{A.10}{=} f\left(\Omega(\boldsymbol{\xi}) \boxplus \underbrace{\left(\frac{d_l\Omega(\boldsymbol{\xi} \boxplus \boldsymbol{\epsilon})}{d\boldsymbol{\epsilon}} \cdot \boldsymbol{\epsilon} + \eta(\boldsymbol{\epsilon}) \cdot \boldsymbol{\epsilon}\right)}_{=:\boldsymbol{\delta}_{\boldsymbol{\epsilon}}}\right) - f(\Omega(\boldsymbol{\xi}))\\
&= f\left(\Omega(\boldsymbol{\xi}) \boxplus \boldsymbol{\delta}_{\boldsymbol{\epsilon}}\right) - f(\Omega(\boldsymbol{\xi})) \overset{A.9}{=} \frac{df\left(\Omega\left(\boldsymbol{\xi}\right) \boxplus \boldsymbol{\delta}\right)}{\boldsymbol{\delta}} \cdot \boldsymbol{\delta}_{\boldsymbol{\epsilon}} + \mu(\boldsymbol{\delta}_{\boldsymbol{\epsilon}}) \cdot \boldsymbol{\delta}_{\boldsymbol{\epsilon}}\\
&= \frac{df\left(\Omega\left(\boldsymbol{\xi}\right) \boxplus \boldsymbol{\delta}\right)}{\boldsymbol{\delta}} \cdot \left(\frac{d_l\Omega(\boldsymbol{\xi} \boxplus \boldsymbol{\epsilon})}{d\boldsymbol{\epsilon}} \cdot \boldsymbol{\epsilon} + \eta(\boldsymbol{\epsilon}) \cdot \boldsymbol{\epsilon}\right) + \mu(\boldsymbol{\delta}_{\boldsymbol{\epsilon}}) \cdot \left(\frac{d_l\Omega(\boldsymbol{\xi} \boxplus \boldsymbol{\epsilon})}{d\boldsymbol{\epsilon}} \cdot \boldsymbol{\epsilon} + \eta(\boldsymbol{\epsilon}) \cdot \boldsymbol{\epsilon}\right)\\
&= \frac{df\left(\Omega\left(\boldsymbol{\xi}\right) \boxplus \boldsymbol{\delta}\right)}{\boldsymbol{\delta}} \cdot \frac{d_l\Omega(\boldsymbol{\xi} \boxplus \boldsymbol{\epsilon})}{d\boldsymbol{\epsilon}} \cdot \boldsymbol{\epsilon} + \frac{df\left(\Omega\left(\boldsymbol{\xi}\right) \boxplus \boldsymbol{\delta}\right)}{\boldsymbol{\delta}} \cdot \eta(\boldsymbol{\epsilon}) \cdot \boldsymbol{\epsilon} + \mu(\boldsymbol{\delta}_{\boldsymbol{\epsilon}}) \cdot \frac{d_l\Omega(\boldsymbol{\xi} \boxplus \boldsymbol{\epsilon})}{d\boldsymbol{\epsilon}} \cdot \boldsymbol{\epsilon} + \mu(\boldsymbol{\delta}_{\boldsymbol{\epsilon}}) \cdot \eta(\boldsymbol{\epsilon}) \cdot \boldsymbol{\epsilon}\\
&= \frac{df\left(\Omega\left(\boldsymbol{\xi}\right) \boxplus \boldsymbol{\delta}\right)}{\boldsymbol{\delta}} \cdot \frac{d_l\Omega(\boldsymbol{\xi} \boxplus \boldsymbol{\epsilon})}{d\boldsymbol{\epsilon}} \cdot \boldsymbol{\epsilon} + \underbrace{\left(\frac{df\left(\Omega\left(\boldsymbol{\xi}\right) \boxplus \boldsymbol{\delta}\right)}{\boldsymbol{\delta}} \cdot \eta(\boldsymbol{\epsilon}) + \mu(\boldsymbol{\delta}_{\boldsymbol{\epsilon}}) \cdot \frac{d_l\Omega(\boldsymbol{\xi} \boxplus \boldsymbol{\epsilon})}{d\boldsymbol{\epsilon}} + \mu(\boldsymbol{\delta}_{\boldsymbol{\epsilon}}) \cdot \eta(\boldsymbol{\epsilon})\right)}_{=:\gamma(\boldsymbol{\epsilon})} \cdot \boldsymbol{\epsilon}
\end{aligned} \tag{A.11}$$

When $\boldsymbol{\epsilon}$ goes to 0, then $\boldsymbol{\delta_\epsilon}$ also goes to 0 (as can be seen from its definition). Using this it follows by the definition of the derivative that $\eta(\boldsymbol{\epsilon})$ and $\mu(\boldsymbol{\delta_\epsilon})$ go to 0 as well. This shows that $\gamma(\boldsymbol{\epsilon})$ goes to 0 when $\boldsymbol{\epsilon}$ goes to 0. Therefore the last line of equation A.11 is in line with our definition of the derivative and

$$\frac{df\left(\Omega\left(\boldsymbol{\xi} \boxplus \boldsymbol{\epsilon}\right)\right)}{d\boldsymbol{\epsilon}} = \frac{df\left(\Omega\left(\boldsymbol{\xi}\right) \boxplus \boldsymbol{\delta}\right)}{\boldsymbol{\delta}} \cdot \frac{d_l\Omega\left(\boldsymbol{\xi} \boxplus \boldsymbol{\epsilon}\right)}{\boldsymbol{\epsilon}} \tag{A.12}$$

□

3.2 Derivation of the Jacobian with respect to pose in Equation (A.5)

In this section we will show how to derive the Jacobians $\frac{\partial\Psi(\boldsymbol{\xi}^D_{cam\_w} \boxplus \boldsymbol{\epsilon}, \boldsymbol{\xi}_{m\_d})}{\partial\boldsymbol{\epsilon}}$ using the implicit definition of the derivative shown in Equation (A.3).

In order to do this we need the definition of the adjoint.

$$\mathbf{T} \cdot \exp(\boldsymbol{\epsilon}) = \exp(\mathrm{Adj}_{\mathbf{T}} \cdot \boldsymbol{\epsilon}) \cdot \mathbf{T} \tag{A.13}$$

with $\mathbf{T} \in \mathbf{SIM}(3)$. It follows that

$$\log\left(\mathbf{T} \cdot \exp(\boldsymbol{\epsilon}) \cdot \mathbf{T}^{-1}\right) = \mathrm{Adj}_{\mathbf{T}} \cdot \boldsymbol{\epsilon} \tag{A.14}$$

Using this we can compute

$$\begin{aligned}
&\Psi(\boldsymbol{\xi}^D_{cam_w}, \boldsymbol{\xi}_{m_d})^{-1} \boxplus \Psi(\boldsymbol{\xi}^D_{cam_w} \boxplus \boldsymbol{\epsilon}, \boldsymbol{\xi}_{m_d}) \overset{A.2}{=} \\
&\left(\boldsymbol{\xi}_{m_d} \boxplus \left(\boldsymbol{\xi}^D_{cam_w}\right)^{-1} \boxplus \boldsymbol{\xi}^{-1}_{m_d} \boxplus \boldsymbol{\xi}^M_{cam_imu}\right)^{-1} \boxplus \left(\boldsymbol{\xi}_{m_d} \boxplus \left(\boldsymbol{\xi}^D_{cam_w} \boxplus \boldsymbol{\epsilon}\right)^{-1} \boxplus \boldsymbol{\xi}^{-1}_{m_d} \boxplus \boldsymbol{\xi}^M_{cam_imu}\right) \\
&\quad = \left(\boldsymbol{\xi}^M_{cam_imu}\right)^{-1} \boxplus \boldsymbol{\xi}_{m_d} \boxplus \boldsymbol{\xi}^D_{cam_w} \boxplus \underbrace{\boldsymbol{\xi}^{-1}_{m_d} \boxplus \boldsymbol{\xi}_{m_d}}_{=0} \boxplus \boldsymbol{\epsilon}^{-1} \boxplus \left(\boldsymbol{\xi}^D_{cam_w}\right)^{-1} \boxplus \boldsymbol{\xi}^{-1}_{m_d} \boxplus \boldsymbol{\xi}^M_{cam_imu} \\
&\quad = \log\left(\left(\mathbf{T}^M_{cam_imu}\right)^{-1} \cdot \mathbf{T}_{m_d} \cdot \mathbf{T}^D_{cam_w} \cdot \exp\left(\boldsymbol{\epsilon}\right)^{-1} \cdot \left(\mathbf{T}^D_{cam_w}\right)^{-1} \cdot \mathbf{T}^{-1}_{m_d} \cdot \mathbf{T}^M_{cam_imu}\right) \overset{A.14}{=} \\
&\quad \mathrm{Adj}\left(\left(\mathbf{T}^M_{cam_imu}\right)^{-1} \cdot \mathbf{T}_{m_d} \cdot \mathbf{T}^D_{cam_w}\right) \cdot (-\boldsymbol{\epsilon})
\end{aligned} \tag{A.15}$$

After moving the minus sign to the left this is in line with our definition of the derivative in Equation (A.3) and it follows that

$$\frac{\partial\Psi(\boldsymbol{\xi}^D_{cam\_w} \boxplus \boldsymbol{\epsilon}, \boldsymbol{\xi}_{m\_d})}{\partial\boldsymbol{\epsilon}} = -\mathrm{Adj}\left(\left(\mathbf{T}^M_{cam\_imu}\right)^{-1} \cdot \mathbf{T}_{m\_d} \cdot \mathbf{T}^D_{cam\_w}\right) \tag{A.16}$$

3.3 Derivation of the Jacobian with respect to scale and gravity direction in Equation (A.6)

In order to derive the Jacobian $\frac{\partial\Psi(\boldsymbol{\xi}^D_{cam\_w}, \boldsymbol{\xi}_{m\_d} \boxplus \boldsymbol{\epsilon})}{\partial\boldsymbol{\epsilon}}$ we need the Baker-Campbell-Hausdorff formula:

Let $\boldsymbol{a}, \boldsymbol{b} \in \mathfrak{sim}(3)$, then

$$\log\left(\exp(\boldsymbol{a}) \cdot \exp(\boldsymbol{b})\right) = \boldsymbol{a} + \boldsymbol{b} + \frac{1}{2}[\boldsymbol{a}, \boldsymbol{b}] + \frac{1}{12}\left([\boldsymbol{a}, [\boldsymbol{a}, \boldsymbol{b}]] + [\boldsymbol{b}, [\boldsymbol{b}, \boldsymbol{a}]]\right) + \frac{1}{48}([\boldsymbol{b}, [\boldsymbol{a}, [\boldsymbol{b}, \boldsymbol{a}]]] + [\boldsymbol{a}, [\boldsymbol{b}, [\boldsymbol{b}, \boldsymbol{a}]]]) + \ldots \tag{A.17}$$

Here $[\boldsymbol{a}, \boldsymbol{b}] := \boldsymbol{ab} - \boldsymbol{ba}$ denotes the Lie bracket.

In this section we will omit the superscripts to simplify the notation.

$$\begin{aligned}
&\Psi(\boldsymbol{\xi}_{cam_w}, \boldsymbol{\xi}_{m_d})^{-1} \boxplus \Psi(\boldsymbol{\xi}_{cam_w}, \boldsymbol{\xi}_{m_d} \boxplus \boldsymbol{\epsilon}) \overset{A.2}{=} \\
&\left(\boldsymbol{\xi}_{m_d} \boxplus \boldsymbol{\xi}^{-1}_{cam_w} \boxplus \boldsymbol{\xi}^{-1}_{m_d} \boxplus \boldsymbol{\xi}_{cam_imu}\right)^{-1} \boxplus \left(\boldsymbol{\xi}_{m_d} \boxplus \boldsymbol{\epsilon} \boxplus \boldsymbol{\xi}^{-1}_{cam_w} \boxplus \left(\boldsymbol{\xi}_{m_d} \boxplus \boldsymbol{\epsilon}\right)^{-1} \boxplus \boldsymbol{\xi}_{cam_imu}\right) \\
&\quad = \boldsymbol{\xi}^{-1}_{cam_imu} \boxplus \boldsymbol{\xi}_{m_d} \boxplus \boldsymbol{\xi}_{cam_w} \boxplus \boldsymbol{\xi}^{-1}_{m_d} \boxplus \boldsymbol{\xi}_{m_d} \boxplus \boldsymbol{\epsilon} \boxplus \boldsymbol{\xi}^{-1}_{cam_w} \boxplus \boldsymbol{\epsilon}^{-1} \boxplus \boldsymbol{\xi}^{-1}_{m_d} \boxplus \boldsymbol{\xi}_{cam_imu} \\
&\quad = \boldsymbol{\xi}^{-1}_{cam_imu} \boxplus \boldsymbol{\xi}_{m_d} \boxplus \boldsymbol{\xi}_{cam_w} \boxplus \boldsymbol{\epsilon} \boxplus \boldsymbol{\xi}^{-1}_{cam_w} \boxplus \underbrace{\left(\boldsymbol{\xi}^{-1}_{cam_imu} \boxplus \boldsymbol{\xi}_{m_d}\right)^{-1} \boxplus \left(\boldsymbol{\xi}^{-1}_{cam_imu} \boxplus \boldsymbol{\xi}_{m_d}\right)}_{=0} \boxplus \boldsymbol{\epsilon}^{-1} \boxplus \boldsymbol{\xi}^{-1}_{m_d} \boxplus \boldsymbol{\xi}_{cam_imu} \\
&\quad = \underbrace{\boldsymbol{\xi}^{-1}_{cam_imu} \boxplus \boldsymbol{\xi}_{m_d} \boxplus \boldsymbol{\xi}_{cam_w} \boxplus \boldsymbol{\epsilon} \boxplus \boldsymbol{\xi}^{-1}_{cam_w} \boxplus \boldsymbol{\xi}^{-1}_{m_d} \boxplus \boldsymbol{\xi}_{cam_imu}}_{:=\boldsymbol{a}} \boxplus \underbrace{\boldsymbol{\xi}^{-1}_{cam_imu} \boxplus \boldsymbol{\xi}_{m_d} \boxplus \boldsymbol{\epsilon}^{-1} \boxplus \boldsymbol{\xi}^{-1}_{m_d} \boxplus \boldsymbol{\xi}_{cam_imu}}_{:=\boldsymbol{b}}
\end{aligned} \tag{A.18}$$

Using Equation A.14 we can compute

$$\boldsymbol{a} = \mathrm{Adj}\left(\mathbf{T}^{-1}_{cam\_imu} \cdot \mathbf{T}_{m\_d} \cdot \mathbf{T}_{cam\_w}\right) \cdot \boldsymbol{\epsilon} \tag{A.19}$$

and

$$\boldsymbol{b} = -\mathrm{Adj}\left(\mathbf{T}^{-1}_{cam\_imu} \cdot \mathbf{T}_{m\_d}\right) \cdot \boldsymbol{\epsilon} \tag{A.20}$$

Now we have to prove that all terms of the Baker-Campbell-Hausdorff formula which contain a Lie bracket can be written as $\mu(\boldsymbol{\epsilon}) \cdot \boldsymbol{\epsilon}$, where $\mu(\boldsymbol{\epsilon})$ goes to zero, when $\boldsymbol{\epsilon}$ goes to zero.

We can compute

$$\boldsymbol{a} \cdot \boldsymbol{b} = \underbrace{\text{Adj}\left(\mathbf{T}_{cam\_imu}^{-1} \cdot \mathbf{T}_{m\_d} \cdot \mathbf{T}_{cam\_w}\right) \cdot \boldsymbol{\epsilon} \cdot \left(-\text{Adj}\left(\mathbf{T}_{cam\_imu}^{-1} \cdot \mathbf{T}_{m\_d}\right)\right)}_{\mu_1(\boldsymbol{\epsilon})} \cdot \boldsymbol{\epsilon} \tag{A.21}$$

$$\boldsymbol{b} \cdot \boldsymbol{a} = \underbrace{-\text{Adj}\left(\mathbf{T}_{cam\_imu}^{-1} \cdot \mathbf{T}_{m\_d}\right) \cdot \boldsymbol{\epsilon} \cdot \text{Adj}\left(\mathbf{T}_{cam\_imu}^{-1} \cdot \mathbf{T}_{m\_d} \cdot \mathbf{T}_{cam\_w}\right)}_{\mu_2(\boldsymbol{\epsilon})} \cdot \boldsymbol{\epsilon} \tag{A.22}$$

$$[\boldsymbol{a}, \boldsymbol{b}] = \boldsymbol{ab} - \boldsymbol{ba} = (\mu_1(\boldsymbol{\epsilon}) + \mu_2(\boldsymbol{\epsilon})) \cdot \boldsymbol{\epsilon} \tag{A.23}$$

Obviously $\mu_1(\boldsymbol{\epsilon})$ and $\mu_2(\boldsymbol{\epsilon})$ go to zero when $\boldsymbol{\epsilon}$ goes to zero. For the remaining summands of Equation (A.17) the same argumentation can be used. It follows that:

$$\begin{aligned}(A.18) &= a + b + \mu(\boldsymbol{\epsilon}) \cdot \boldsymbol{\epsilon} \\ &= \left(\text{Adj}\left(\mathbf{T}_{cam\_imu}^{-1} \cdot \mathbf{T}_{m\_d} \cdot \mathbf{T}_{cam\_w}\right) - \text{Adj}\left(\mathbf{T}_{cam\_imu}^{-1} \cdot \mathbf{T}_{m\_d}\right) \cdot \boldsymbol{\epsilon}\right) \cdot \boldsymbol{\epsilon} + \mu(\boldsymbol{\epsilon}) \cdot \boldsymbol{\epsilon}\end{aligned} \tag{A.24}$$

where $\mu(\boldsymbol{\epsilon})$ goes to zero when $\boldsymbol{\epsilon}$ goes to zero. According to (A.3) this means that

$$\frac{\partial \Psi(\boldsymbol{\xi}_{cam\_w}^{D}, \boldsymbol{\xi}_{m\_d} \boxplus \boldsymbol{\epsilon})}{\partial \boldsymbol{\epsilon}} = \text{Adj}(\mathbf{T}_{cam\_imu}^{-1} \cdot \mathbf{T}_{m\_d} \cdot \mathbf{T}_{cam\_w}) - \text{Adj}(\mathbf{T}_{cam\_imu}^{-1} \cdot \mathbf{T}_{m\_d}) \tag{A.25}$$

References

[1] R. Mur-Artal and J. D. Tardós, "Visual-inertial monocular slam with map reuse," *IEEE Robot. and Autom. Lett.*, vol. 2, no. 2, 2017.