Supplementary Material

1 Appendix A

1.1 Stochastic robust approximation

The *stochastic robust approximation* problem [1, 2] considers an optimization objective that accounts for uncertainty or variations in the data. For example, given a basic quadratic cost function $\|f(x,b)\|^2$, where x represents the parameter to be estimated, and the data measurements are given by $b = \bar{b} + \delta b$, with $\bar{b}$ as the mean and δb representing the uncertainty. A natural idea to incorporate such uncertainty is to use expectation of the $\|f(x,b)\|^2$ as the final cost function, and formulate the optimization problem as

$$\min_x \mathbb{E}_b \left(\|f(x,b)\|^2 \right) \tag{1}$$

Since the expectation $\mathbb{E}_b \left(|f(x,b)|^2 \right)$ is generally intractable, the optimization problem is typically addressed by approximating the cost function. A commonly-adopted method to approximate this cost function is to linearize it at $\bar{b}$ using a first-order Taylor expansion as

$$f(x,b) \approx f(x,\bar{b}) + \frac{\partial f}{\partial b}(b - \bar{b}) \tag{2}$$

Substituting this approximation into the cost function and taking the expectation, we obtain:

$$\begin{aligned} \mathbb{E}_b \left(\|f(x,b)\|^2 \right) &\approx \mathbb{E}_b(\|f(x,\bar{b})\|^2) + \mathbb{E}_b \left(\frac{\partial f}{\partial b}(b-\bar{b})(b-\bar{b})^\top \frac{\partial f}{\partial b}^\top \right) \\ &= \|f(x,\bar{b})\|^2 + \text{trace}\left(\frac{\partial f}{\partial b} \mathbf{Q}_b \frac{\partial f}{\partial b}^\top \right) \end{aligned} \tag{3}$$

Through this approximation, we transform the stochastic robust approximation problem into a standard optimization problem, which can be efficiently solved using iterative methods such as Gauss-Newton or Levenberg-Marquardt.

1.2 Approximation of the cost function

Recall the cost function:

$$\min_{{}^{G}\mathbf{p}_a,\gamma} \sum_{k=1}^{m} \mathbb{E}_{\mathbf{x}_{I_k}}(\|d_k - h(\mathbf{x}_{I_k}, \mathbf{x}_U)\|^2) \tag{4}$$

where $\mathbf{x}_U = \left[{}^{G}\mathbf{p}_a^\top, \gamma\right]^\top$, and the above expectation $\mathbb{E}(\cdot)$ is taken over the instantiations of all possible robot (IMU) state $\mathbf{x}_{I_k}$. To simplify the notation, we omit the subscript $\mathbf{x}_{I_k}$ from the expectation term $\mathbb{E}_{\mathbf{x}_{I_k}}(\cdot)$ in the subsequent derivations. We can linearize the cost function in (4) at the current IMU state estimate $\hat{\mathbf{x}}_{I_k}$ as

$$\sum_{k=1}^{m} \mathbb{E}(\|d_k - h(\mathbf{x}_{I_k}, \mathbf{x}_U)\|^2) \approx \sum_{k=1}^{m} \mathbb{E}(\|d_k - h(\hat{\mathbf{x}}_{I_k}, \mathbf{x}_U) - \mathbf{H}_{I_k}(\mathbf{x}_{I_k} - \hat{\mathbf{x}}_{I_k})\|^2) \tag{5}$$

where $\mathbf{H}_{I_k}$ denotes the corresponding jacobians. By defining the measurement residual as:

$$\delta_k = d_k - h(\hat{\mathbf{x}}_{I_k}, \mathbf{x}_U), \tag{6}$$

and substituting this into the expectation term, we rewrite the original expression as:

$$\begin{aligned}
&\sum_{k=1}^{m} \mathbb{E}\left(\|d_k - h(\mathbf{x}_{I_k}, \mathbf{x}_U)\|^2\right) = \sum_{k=1}^{m} \mathbb{E}\left(\|\delta_k - \mathbf{H}_{I_k}(\mathbf{x}_{I_k} - \hat{\mathbf{x}}_{I_k})\|^2\right) \\
&= \sum_{k=1}^{m} \mathbb{E}\left(\delta_k^2 - 2\delta_k \mathbf{H}_{I_k}(\mathbf{x}_{I_k} - \hat{\mathbf{x}}_{I_k}) + (\mathbf{x}_{I_k} - \hat{\mathbf{x}}_{I_k})^\top \mathbf{H}_{I_k}^\top \mathbf{H}_{I_k}(\mathbf{x}_{I_k} - \hat{\mathbf{x}}_{I_k})\right) \\
&= \sum_{k=1}^{m} \mathbb{E}(\delta_k^2) - 2\mathbb{E}\left(\delta_k \mathbf{H}_{I_k}(\mathbf{x}_{I_k} - \hat{\mathbf{x}}_{I_k})\right) + \mathbb{E}\left((\mathbf{x}_{I_k} - \hat{\mathbf{x}}_{I_k})^\top \mathbf{H}_{I_k}^\top \mathbf{H}_{I_k}(\mathbf{x}_{I_k} - \hat{\mathbf{x}}_{I_k})\right)
\end{aligned} \tag{7}$$

If $\hat{\mathbf{x}}_{I_k}$ is an unbiased estimate of $\mathbf{x}_{I_k}$, i.e., $\mathbb{E}[\mathbf{x}_{I_k} - \hat{\mathbf{x}}_{I_k}] = 0$, then one has

$$\begin{aligned}
\sum_{k=1}^{m} \mathbb{E}\left(\|d_k - h(\mathbf{x}_{I_k}, \mathbf{x}_U)\|^2\right) &\approx \sum_{k=1}^{m} \mathbb{E}(\delta_k^2) + \mathbb{E}\left((\mathbf{x}_{I_k} - \hat{\mathbf{x}}_{I_k})^\top \mathbf{H}_{I_k}^\top \mathbf{H}_{I_k}(\mathbf{x}_{I_k} - \hat{\mathbf{x}}_{I_k})\right) \\
&= \sum_{k=1}^{m} \mathbb{E}(\delta_k^2) + \mathrm{trace}(\mathbf{H}_{I_k}\mathbf{P}_{I_k}\mathbf{H}_{I_k}^\top) \\
&= \sum_{k=1}^{m} \|d_k - h(\hat{\mathbf{x}}_{I_k}, \mathbf{x}_U)\|^2 + \mathrm{trace}(\mathbf{H}_{I_k}\mathbf{P}_{I_k}\mathbf{H}_{I_k}^\top)
\end{aligned} \tag{8}$$

Compared to the cost of deterministic optimization, given by $\sum_{k=1}^{m} \|d_k - h(\hat{\mathbf{x}}_{I_k}, \mathbf{x}_U)\|^2$, the derived cost has an additional term $\mathrm{trace}(\mathbf{H}_{I_k}\mathbf{P}_{I_k}\mathbf{H}_{I_k}^\top)$. It is important to note that this term is also a function of the UWB state $\mathbf{X}_U$, since the jacobian $\mathbf{H}_{I_k}$ depends on the state $\mathbf{x}_{U_k}$. By incorporating this term into the cost function, the uncertainty of the UAV's pose is explicitly considered in the optimization process, leading to a more robust initialization.

2 Appendix B

Recall the formulation of the likelihood function:

$$\mathbf{p}(\mathbf{d};\mathbf{x}_U) = \frac{1}{(2\pi)^{\frac{N}{2}}\det(\Sigma)^{\frac{1}{2}}}\exp\Big((\mathbf{d}-\mathbf{h})^\top \mathbf{\Sigma}^{-1}(\mathbf{d}-\mathbf{h})\Big), \tag{9}$$

and its corresponding log-likelihood function:

$$\ell(\mathbf{d};\mathbf{x}_U) = -\frac{1}{2}(\mathbf{d}-\mathbf{h})^\top \mathbf{\Sigma}^{-1}(\mathbf{d}-\mathbf{h}) \tag{10}$$

Given the log-likelihood function in (10), the (i,j)-th element of the FIM can be computed as

$$\mathbf{F}_{ij} = \mathbb{E}\left[\left(\frac{\partial \ell(\mathbf{d};\mathbf{x}_U^i)}{\partial \mathbf{x}_{U^i}}\right)\left(\frac{\partial \ell(\mathbf{d};\mathbf{x}_U^j)}{\partial \mathbf{x}_U^j}\right)^\top\right], \tag{11}$$

where $\mathbf{x}_U^i$ and $\mathbf{x}_U^j$ denotes the i-th and j-th element of the state vector $\mathbf{x}_U$, respectively. As previously mentioned, the key distinction between our method and existing approaches lies in how the uncertainty term $\mathbf{\Sigma}$ is handled. To provide a comprehensive analysis of its impact on the computation of the overall Fisher Information Matrix (FIM) and the initialization performance, we separately compute the FIM for both cases and conduct a detailed analysis:

FIM with UAV's uncertainty: Since our method explicitly incorporates the UAV's uncertainty during the UWB initialization, the covariance term not only includes the the UWB measurement covariance $\mathbf{Q}_d$, but also contains the uncertainty $\mathbf{P}_{II_k}$ arising from the UAV's localization. In this case, the covariance matrix $\mathbf{\Sigma} = \mathrm{diag}(\Sigma_0, \cdots, \Sigma_k, \cdots \Sigma_m)$ should account for both the UWB measurement noise covariance and the UAV's uncertainty. Each block Σ_k is given by:

$$\Sigma_k = \mathbf{Q}_d + \mathbf{H}_{I_k}\mathbf{P}_{II_k}\mathbf{H}_{I_k}^\top \tag{12}$$

It is important to note that the covariance $\mathbf{\Sigma}$ is also a function of the UWB estimate $\mathbf{x}_U$, as the value of the jacobian $\mathbf{H}_{I_k}$ depends on the UWB state $\mathbf{x}_U$, as derived in Sec. IV-B of the paper. Thus, the distribution of $\mathbf{d}$ can be written as $\mathbf{d} \sim \mathcal{N}(\mathbf{h}(\mathbf{x}_U), \mathbf{\Sigma}(\mathbf{x}_U))$, which is referred to as the *General Gaussian Distribution* [3, Section 3.9] This ensures that the estimation process properly considers the impact of pose uncertainty, leading to a more robust initialization. According to [3, Section 3.9], (11) can be computed as

$$\mathbf{F}_{ij} = \left(\frac{\partial \mathbf{h}(\mathbf{x}_U)}{\partial \mathbf{x}_U^i}\right)^\top \mathbf{\Sigma}(\mathbf{x}_U^i)^{-1}\left(\frac{\partial \mathbf{h}(\mathbf{x}_U)}{\partial \mathbf{x}_U^j}\right) + \frac{1}{2}\mathbf{trace}\left(\mathbf{\Sigma}(\mathbf{x}_U^i)^{-1}\frac{\partial \mathbf{h}(\mathbf{x}_U)}{\partial \mathbf{x}_U^i}\mathbf{\Sigma}(\mathbf{x}_U^i)^{-1}\frac{\partial \mathbf{h}(\mathbf{x}_U)}{\partial \mathbf{x}_U^j}\right) \tag{13}$$

FIM without UAV's uncertainty: If the UAV's uncertainty is not taken into account, the covariance Σ_k in (12) simplifies to a constant matrix:

$$\Sigma_k = \mathbf{Q}_d \tag{14}$$

which consists solely of the measurement noise covariance. Since $\mathbf{\Sigma}$ is not longer a function of the UWB state $\mathbf{x}_U$, the FIM in (11) is simplified to [4]:

$$\mathbf{F}_{ij} = \left(\frac{\partial \mathbf{h}(\mathbf{x}_U)}{\partial \mathbf{x}_U^i}\right)^\top \mathbf{\Sigma}(\mathbf{x}_U^i)^{-1} \left(\frac{\partial \mathbf{h}(\mathbf{x}_U)}{\partial \mathbf{x}_U^j}\right) \tag{15}$$

We observe that, compared to the previous case, this FIM includes only the first term, while the second term vanishes. This indicates that when the UAV's uncertainty is not considered, its contribution to the overall information gain is omitted, leading to a potentially less informative estimation process.

References

[1] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, 2004. [Online]. Available: https://web.stanford.edu/ boyd/cvxbook/

[2] L. A. Hannah, "Stochastic optimization," Columbia University, Tech. Rep., April 2014. [Online]. Available: https://sites.stat.columbia.edu/liam/teaching/compstat-spr14/lauren-notes.pdf

[3] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[4] T. H. Nguyen and L. Xie, "Estimating odometry scale and uwb anchor location based on semidefinite programming optimization," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7359–7366, 2022.