

I. FACTOR GRAPH FORMULATION

We formulate the state estimation problem as a Maximum-A-Posteriori (MAP) estimation over a factor graph. We utilize an Incremental Fixed-Lag Smoother to estimate the state over a moving window of recent measurements, marginalizing out older states to ensure bounded computation for real-time performance. The bipartite graph consists of variable nodes representing the robot's state at discrete time steps i , connected by factor nodes representing sensor measurements.

A. State and Frame Definition

We define the state at time i as the tuple $\mathcal{S}_i = \{\mathbf{x}_i, \mathbf{v}_i, \mathbf{b}_i\}$, where $\mathbf{x}_i \in SE(3)$ denotes the pose (rotation \mathbf{R}_i , position \mathbf{p}_i) of the vehicle Body frame B relative to the World frame W (ENU), $\mathbf{v}_i \in \mathbb{R}^3$ is the linear velocity of the vehicle in the World frame (ENU), and $\mathbf{b}_i \in \mathbb{R}^6$ represents the IMU accelerometer and gyroscope biases.

To simplify the measurement models and Jacobians for velocity integration, we define the vehicle Body frame B to be coincident with the DVL sensor frame. For all other sensors S (IMU, Depth, GPS, Heading), we define a static extrinsic rigid-body transformation $\mathbf{T}_{bs} = [\mathbf{R}_{bs}, \mathbf{t}_{bs}] \in SE(3)$ representing the lever arm from the Body B to the sensor S .

B. Measurement Factors

1) *Depth Factor* (f_i^z): The depth factor constrains the vertical position of the robot. We account for the physical offset \mathbf{t}_{bs} of the pressure sensor to ensure accurate depth tracking during pitch and roll maneuvers.

Given a depth measurement z_{meas} , the residual $r_{d,i}$ transforms the body pose \mathbf{x}_i to the sensor frame and extracts the vertical component:

$$r_{d,i} = (\mathbf{p}_i + \mathbf{R}_i \mathbf{t}_{bs}) \cdot \mathbf{e}_z - z_{meas} \quad (1)$$

where $\mathbf{e}_z = [0, 0, 1]^\top$.

2) *GPS Factor* (f_i^{xy}): When available, GPS measurements constrain the global position. Similar to the depth factor, we incorporate the lever arm \mathbf{t}_{bs} from the DVL/Body frame to the GPS antenna. The residual $\mathbf{r}_{gps,i}$ is defined as:

$$\mathbf{r}_{gps,i} = (\mathbf{p}_i + \mathbf{R}_i \mathbf{t}_{bs}) - \mathbf{p}_{meas} \quad (2)$$

While the GPS factor constrains the full 3D position, vertical constraints are often relaxed (via high covariance) in favor of the more accurate depth sensor.

3) *Heading Factor* (f_i^ψ): The heading factor constrains the vehicle's orientation using magnetometer data. We explicitly model the mounting rotation \mathbf{R}_{bs} to align the sensor reading with the Body frame. The residual $\mathbf{r}_{\psi,i}$ is computed on the $SO(3)$ manifold:

$$\mathbf{r}_{\psi,i} = \text{Log}((\mathbf{R}_{meas} \mathbf{R}_{bs}^{-1})^{-1} \mathbf{R}_i) \quad (3)$$

Similar to GPS, the heading factor mathematically constrains the full 3D orientation (roll, pitch, yaw). Practical application often discards roll and pitch information via the noise model, relying on the IMU gravity vector instead.

4) *DVL Velocity Factor* (f_i^v): The DVL velocity factor relates the estimated world-frame velocity \mathbf{v}_i to the measured body-frame velocity $\mathbf{v}_{b,meas}$.

By defining the Body frame to coincide with the DVL frame, we avoid complex lever-arm terms involving angular velocity cross-products. The residual simply rotates the world velocity into the Body frame:

$$\mathbf{r}_{dvl,i} = \mathbf{R}_i^\top \mathbf{v}_i - \mathbf{v}_{b,meas} \quad (4)$$

This naturally couples the pose \mathbf{x}_i and velocity \mathbf{v}_i .

5) *Preintegrated DVL Factor* (f_{ij}^D): To constrain relative motion during GPS outages, we utilize a DVL preintegration factor. We accumulate high-frequency DVL velocity measurements $\mathbf{v}_{b,k}$ between keyframes i and j . Using the method proposed in TURTLMAP, we integrate these velocities using the high-frequency IMU orientation $\Delta \tilde{\mathbf{R}}_{ik}$ to decouple rotation estimation from translation:

$$\Delta \tilde{\mathbf{p}}_{ij} = \sum_{k=i}^{j-1} \Delta \tilde{\mathbf{R}}_{ik} \mathbf{v}_{b,k} \Delta t \quad (5)$$

To strictly enforce keyframe timestamps, we extrapolate the last DVL reading over the small time residual δt . While this is not theoretically ideal (as it reuses a stochastic measurement), it allows for precise time alignment. We provide this factor for comparison purposes only; it is not enabled in our primary solution.

This integrated measurement $\Delta \tilde{\mathbf{p}}_{ij}$ represents the relative position change in the frame of keyframe i (i.e., $\mathbf{R}_i^\top (\mathbf{p}_j - \mathbf{p}_i)$). The residual is:

$$\mathbf{r}_{D_{ij}} = (\mathbf{x}_i^{-1} \mathbf{x}_j)_{trans} - \Delta \tilde{\mathbf{p}}_{ij} \quad (6)$$

where $(\cdot)_{trans}$ extracts the translation component of the relative pose.

In this implementation, velocity bias is assumed to be handled externally or negligible.

6) *Preintegrated IMU Factor* (f_{ij}^T): We utilize the GTSAM CombinedImuFactor to integrate high-frequency inertial measurements. The factor internally handles the lever arm \mathbf{t}_{bs} (tangential and centripetal accelerations), allowing state estimation at the DVL/Body frame while using disjointly mounted IMU measurements.