

# Closed-form Preintegration Methods for Graph-based Visual-Inertial Navigation

Journal Title  
XX(X):1–22  
©The Author(s) 0000  
Reprints and permission:  
[sagepub.co.uk/journalsPermissions.nav](http://sagepub.co.uk/journalsPermissions.nav)  
DOI: 10.1177/ToBeAssigned  
[www.sagepub.com/](http://www.sagepub.com/)



Kevin Eckenhoff, Patrick Geneva, and Guoquan Huang

## Abstract

In this paper we propose a new analytical preintegration theory for graph-based sensor fusion with an inertial measurement unit (IMU) and a camera (or other aiding sensors). Rather than using discrete sampling of the measurement dynamics as in current methods, we derive the closed-form solutions to the preintegration equations, yielding improved accuracy in state estimation. We advocate two new different inertial models for preintegration: (i) the model that assumes piecewise constant measurements, and (ii) the model that assumes piecewise constant local true acceleration. We show through extensive Monte-Carlo simulations the effect that the choice of preintegration model has on estimation performance. To validate the proposed preintegration theory, we develop both direct and indirect visual-inertial navigation systems (VINS) that leverage our preintegration. In the first, within a tightly-coupled, sliding-window optimization framework, we jointly estimate the features in the window and the IMU states while performing marginalization to bound the computational cost. In the second, we loosely-couple the IMU preintegration with a direct image alignment that estimates relative camera motion by minimizing the photometric errors (i.e., image intensity difference), allowing for efficient and informative loop closures. Both systems are extensively validated in real-world experiments and are shown to offer competitive performance to state-of-the-art methods.

## Keywords

Preintegration, visual-inertial odometry, SLAM, optimization, state estimation, Monte-Carlo analysis

## 1 Introduction

Accurate localization for autonomous systems is a prerequisite in many robotic applications such as planetary exploration (Mourikis, Trawny, Roumeliotis, Johnson and Matthies 2007), search and rescue (Ellekilde 2007), and autonomous driving (Geiger, Lenz and Urtasun 2012). In many of these scenarios, access to global information such as from a Global Positioning System (GPS), motion capture system, or a prior map of the environment is unavailable. Instead, one has to estimate the robot state and its surroundings based on noisy, local measurements from onboard sensors, by performing simultaneous localization and mapping (SLAM), which has witnessed significant research efforts in the past three decades (Cadena, Carlone, Carrillo, Latif, Scaramuzza, Neira, Reid and Leonard 2016).

Of many possible sensors used in SLAM, micro-electro-mechanical-system (MEMS) inertial measurement units (IMUs) have become ubiquitous. These low-cost and light-weight sensors typically provide local linear acceleration and angular velocity readings, and are well suited for many applications such as micro aerial vehicles (MAVs) (Ling, Liu and Shen 2016) and mobile devices (Wu, Ahmed, Georgiou and Roumeliotis 2015). IMUs provide information only about the derivatives of the kinematic states, so estimation must be performed by integrating over these noisy measurements. This may lead to large drifts over long periods of time, making the use of a low-cost IMU alone an unreliable solution. However, IMU readings are highly informative about short-term motion which is ideal for fusion with measurements from exteroceptive aiding sensors,

such as LiDAR and cameras. These sensors compensate for the drift issue inherent in inertial navigation, while high-rate inertial measurements are useful in tracking aggressive motion which may be difficult for exteroceptive low-rate sensors alone.

One canonical way of fusing IMU measurements in aided inertial navigation is to use an extended Kalman filter (EKF) (e.g., see Mourikis and Roumeliotis (2007)). In this method, the inertial measurements are used to predict to the next time instance, while measurements from exteroceptive sensors are used to update the state estimate. More recently, the development of preintegration has allowed for the efficient inclusion of high-rate IMU measurements in graph-based SLAM (Lupton and Sukkarieh 2012; Forster, Carlone, Dellaert and Scaramuzza 2015, 2017). In this paper, building upon our prior conference publication (Eckenhoff, Geneva and Huang 2016a), we investigate in-depth the optimal use of preintegration by providing models and their closed-form solutions for the preintegrated measurement dynamics, allowing for more accurate computation of the inertial factors for use in graph optimization of visual-inertial navigation systems (VINS).

---

The authors are with the Dept. of Mechanical Engineering, and Computer and Information Sciences, University of Delaware, Newark, DE 19716, USA.

### Corresponding author:

Kevin Eckenhoff, Department of Mechanical Engineering,  
126 Spencer Lab, Newark, DE 19716, USA.  
Email: keck@udel.edu

In particular, the main contributions of this work include:

- We advocate two new preintegration models (i.e., piecewise constant measurements and piecewise constant local true acceleration, instead of piecewise constant global acceleration as assumed in existing methods) to better capture the underlying motion dynamics and offer the analytical solutions to the preintegration equations. We have open sourced the proposed preintegration to better contribute to our research community.\*
- Using the proposed closed-form preintegration, we develop an indirect, tightly-coupled, sliding-window optimization based visual-inertial odometry (VIO), which marginalizes out features from the state vector when moving to the next time window to enable real-time performance of bounded computational cost.
- With the proposed closed-form IMU preintegration, we further develop a loosely-coupled, direct VINS, which fuses preintegrated inertial measurements with direct image alignment results.
- We conduct thorough Monte-Carlo simulation analysis of different preintegration models by varying motion dynamics and IMU sampling rates. We also perform extensive real-world experiments to validate the proposed VINS using our preintegration by comparing with a state-of-the-art method.

The remainder of the paper is organized as follows: After a brief overview of related work in the next section and estimation preliminaries in Section 3, we present in detail the proposed continuous preintegration in Section 4. The direct and indirect VINS that use the proposed preintegration are described in Section 5. In Sections 6 and 7, we validate the proposed VINS algorithms through both simulations and experiments. Finally, Section 8 concludes the work in this paper, as well as the possible future research directions.

## 2 Related Work

### 2.1 Visual-Inertial Navigation

Mourikis and Roumeliotis (2007) proposed one of the earliest successful VINS algorithms, known as the multi-state constraint Kalman filter (MSCKF). This filtering approach used quaternion-based inertial dynamics (Trawny and Roumeliotis 2005) for state propagation coupled with a novel EKF update step. Rather than adding features seen in the camera images to the state vector, their visual measurements were projected onto the nullspace of the feature Jacobian matrix (akin to feature marginalization (Yang, Maley and Huang 2017)), thereby retaining motion constraints that only related to the stochastically cloned camera poses in the state vector (Roumeliotis and Burdick 2002). While reducing the computational cost by removing the need to co-estimate features, this nullspace projection prevents the relinearization of the processed features’ nonlinear measurements at later time steps.

The standard MSCKF recently has been extended in various directions. For example, Hesch, Kottas, Bowman and Roumeliotis (2013); Huang, Kaess and Leonard (2014) improved the filter consistency by enforcing the correct observability properties of the linearized EKF VINS. Guo

and Roumeliotis (2013) showed that the inclusion of plane features increases the estimation accuracy. Guo, Kottas, DuToit, Ahmed, Li and Roumeliotis (2014) extended to the case of rolling-shutter cameras with inaccurate time synchronization. Recently, Wu, Ahmed, Georgiou and Roumeliotis (2015) further reformulated the VINS problem within a square-root inverse filtering framework for improved computational efficiency and numerical stability without sacrificing estimation accuracy. While these MSCKF-based methods have shown to exhibit accurate state estimation, they theoretically suffer from a limitation – that is, nonlinear measurements must have a *one-time* linearization before processing, possibly introducing large linearization errors into the estimator.

Batch optimization methods, by contrast, solve a nonlinear least-squares or bundle adjustment (BA) problem over a set of measurements, allowing for the reduction of error through relinearization (Kummerle, Grisetti, Strasdat, Konolige and Burgard 2011). The incorporation of *tightly-coupled* VINS in batch optimization methods requires overcoming the high frequency nature and computational complexity of the inertial measurements.

Leutenegger, Lynen, Bosse, Siegwart and Furgale (2015) introduced a keyframe-based VINS approach (i.e., OKVIS), whereby a set of non-sequential past camera poses and a series of recent inertial states, connected with inertial measurements, was used in nonlinear optimization for accurate trajectory estimation. These inertial factors took the form of a state prediction: every time that the linearization point for the starting inertial state threshold, it is required to reintegrate the IMU dynamics. This presents inefficiencies in the inertial processing, while the authors demonstrated the feasibility of such a scheme for a small number of inertial factors in a sliding window estimator. It should be noted that the well-known open-source implementation of OKVIS<sup>†</sup> in fact employs the method of inertial preintegration, described in detail later, while only triggering full reintroduction if the linearization point changes sufficiently and thus improving the efficiency.

### 2.2 Visual Processing

A key component to any VINS algorithm is the visual processing pipeline, responsible for transforming dense imagery data to motion constraints that can be incorporated into the estimation problem. Seen as the classical technique, indirect methods of visual SLAM extract and track features in the environment, while using geometric reprojection constraints during estimation. An example of state-of-the-art indirect visual-SLAM methods is ORB-SLAM2 (Mur-Artal and Tardós 2017), which performs graph-based optimization of camera poses using information from 3D feature point correspondences.

In contrast, direct methods utilize pixel intensities in their formulation and allow for inclusion of a larger percentage of the available image information. LSD-SLAM is an example of state-of-the-art direct visual-SLAM methods

---

\*The open source of the proposed continuous preintegration is available at: <https://github.com/rpng/cpi>

<sup>†</sup><https://github.com/ethz-asl/okvis>

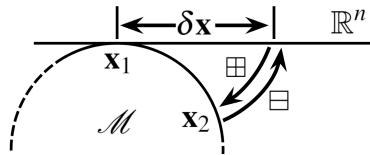
which optimizes the transformation between pairs of camera keyframes based on minimizing their intensity error (Engel, Schöps and Cremers 2014). Note that this approach also optimizes a separate graph containing keyframe constraints to allow for the incorporation of highly informative loop-closures to correct drift over long trajectories. This work was later extended from a monocular sensor to stereo and omnidirectional cameras for improved accuracy (Engel, Stückler and Cremers 2015; Caruso, Engel and Cremers 2015). Other popular direct methods include the work by Engel, Koltun and Cremers (2018) and Wang, Schwörer and Cremers (2017) which estimated keyframe depths along with the camera poses in a tightly-coupled manner, offering low-drift performance.

Application of direct methods to the visual-inertial problem has seen recent attention due to their ability to robustly track dynamic motion even in low-texture environments. For example, Bloesch, Omari, Hutter and Siegwart (2015); Bloesch, Burri, Omari, Hutter and Siegwart (2017) used a patch-based direct method to provide updates with an iterated EKF; Usenko, Engel, Stückler and Cremers (2016) introduced a sliding-window VINS based on the discrete preintegration and direct image alignment; Ling, Liu and Shen (2016) employed loosely-coupled direct alignment with preintegration factors for tracking aggressive quadrotor motions. While these methods have shown the feasibility of incorporating IMU measurements with direct methods, they employed the *discrete* form of inertial preintegration.

### 2.3 Inertial Preintegration

First introduced by Lupton and Sukkarieh (2012), inertial preintegration is a computationally efficient alternative to the standard inertial measurement integration, e.g., as performed in EKF propagation. The authors employed the discrete integration of the inertial measurement dynamics in a *local* frame of reference, preventing the need to reintegrate the state dynamics at each optimization step. While this addresses the computational complexity issue, this method suffers from singularities due to the use of Euler angles in the orientation representation. To improve the stability of this preintegration, an on-manifold representation was introduced by Forster, Carlone, Dellaert and Scaramuzza (2015, 2017) which presents a singularity-free orientation representation on the  $SO(3)$  manifold, incorporating the IMU preintegration into an efficient graph-based VINS algorithm.

While Shen, Michael and Kumar (2015) introduced preintegration in the continuous form, they still discretely sampled the measurement dynamics without offering closed-form solutions. This left a significant gap in the theoretical completeness of preintegration theory from a continuous-time perspective. Albeit, Qin, Li and Shen (2017) later extended to a robust tightly-coupled monocular visual-inertial localization system. As compared to the discrete approximation of the preintegrated measurement and covariance calculations used in previous methods, in our prior work (Eckenhoff, Geneva and Huang 2016a), we have derived the closed-form solutions to both the measurement and covariance preintegration equations and showed that these solutions offer improved accuracy over the discrete methods, especially in the case of highly dynamic motion.



**Figure 1.** Illustration of the state update operations on a manifold. The  $\boxplus$  operation maps  $x_1 \in \mathcal{M}$  and a vector  $\delta x \in \mathbb{R}^n$  to a new element  $x_2 \in \mathcal{M}$ , while the  $\boxminus$  operation maps  $x_1$  and  $x_2$  to the vector  $\delta x$ .

In this work, based on our preliminary results (Eckenhoff, Geneva and Huang 2016a, 2017), we provide a solid theoretical foundation for closed-form preintegration and show that it can be easily incorporated into different graph-based sensor fusion methods. We investigate the improved accuracy afforded by two different models of closed-form preintegration and scenarios in which they exhibit superior performance. We further develop both indirect and direct graph-based VINS and demonstrate their competitive performance to state-of-the-art methods.

## 3 Estimation Preliminaries

The IMU state of an aided inertial navigation system at time step  $k$  is given by (Mourikis and Roumeliotis 2007):

$$\mathbf{x}_k = [{}^k_G\bar{q}^\top \quad \mathbf{b}_{\omega_k}^\top \quad {}^G\mathbf{v}_k^\top \quad \mathbf{b}_{a_k}^\top \quad {}^G\mathbf{p}_k^\top]^\top \quad (1)$$

where  ${}^k_G\bar{q}$  is the unit quaternion of JPL form parameterizing the rotation  ${}^k_G\mathbf{R}$  from the global frame  $\{G\}$  to the current local frame  $\{k\}$  (Trawny and Roumeliotis 2005),  $\mathbf{b}_{\omega_k}$  and  $\mathbf{b}_{a_k}$  are the gyroscope and accelerometer biases, and  ${}^G\mathbf{v}_k$  and  ${}^G\mathbf{p}_k$  are the velocity and position of the IMU expressed in the global frame, respectively.

Note that while the state vector (1) contains 16 variables, there are only 15 degrees of freedom (DOF), due to the constraint that the quaternion  ${}^k_G\bar{q}$  must have unit length. In fact, the state lies on the manifold defined by the product of the unit quaternions  $\mathbb{H}$  with the vector space  $\mathbb{R}^{12}$  (i.e.,  $\mathcal{M} = \mathbb{H} \times \mathbb{R}^{12}$ ). In order to represent the estimation problem on manifold, we employ the “boxplus” update operation,  $\boxplus$ , which maps an element from a manifold,  $\mathbf{x} \in \mathcal{M}$ , and an error vector  $\delta \mathbf{x}$  into a new element on  $\mathcal{M}$  (Hertzberg, Wagner, Frese and Schröder 2013). As illustrated in Figure 1, for a manifold of dimension  $n$ , we can define the following operation:

$$\boxplus : \mathcal{M} \times \mathbb{R}^n \rightarrow \mathcal{M} \quad (2)$$

$$\mathbf{x}_1 \boxplus \delta \mathbf{x} = \mathbf{x}_2 \quad (3)$$

Similarly, the inverse “boxminus” operation  $\boxminus$  is given by:

$$\boxminus : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}^n \quad (4)$$

$$\mathbf{x}_2 \boxminus \mathbf{x}_1 = \delta \mathbf{x} \quad (5)$$

In the case of a state in a vector space,  $\mathbf{v} \in \mathbb{R}^n$ , these operations are the standard addition and subtraction:

$$\mathbf{v}_1 \boxplus \delta \mathbf{v} \triangleq \mathbf{v}_1 + \delta \mathbf{v} = \mathbf{v}_2 \quad (6)$$

$$\mathbf{v}_2 \boxminus \mathbf{v}_1 \triangleq \mathbf{v}_2 - \mathbf{v}_1 = \delta \mathbf{v} \quad (7)$$

In the case of a unit quaternion expressed using the JPL convention,  $\bar{q}$ , we have (Trawny and Roumeliotis 2005):

$$\bar{q}_1 \boxplus \delta\theta \triangleq \begin{bmatrix} \frac{\delta\theta}{2} \\ 1 \end{bmatrix} \otimes \bar{q}_1 \simeq \bar{q}_2 \quad (8)$$

$$\bar{q}_2 \boxminus \bar{q}_1 \triangleq 2\text{vec}(\bar{q}_2 \otimes \bar{q}_1^{-1}) = \delta\theta \quad (9)$$

where  $\text{vec}(\bar{q})$  refers to the vector portion of the quaternion argument (i.e.,  $\text{vec}([\mathbf{q}^\top q_4]^\top) = \mathbf{q}$ ). The quaternion multiplication,  $\otimes$ , is given by:

$$\bar{q} \otimes \bar{p} \triangleq \mathcal{R}(\bar{p}) \bar{q} = \mathcal{L}(\bar{q}) \bar{p} \quad (10)$$

$$\mathcal{R}(\bar{q}) = \begin{bmatrix} q_4 \mathbf{I} + [\mathbf{q}] & \mathbf{q} \\ -\mathbf{q}^\top & q_4 \end{bmatrix} \quad (11)$$

$$\mathcal{L}(\bar{p}) = \begin{bmatrix} p_4 \mathbf{I} - [\mathbf{p}] & \mathbf{p} \\ -\mathbf{p}^\top & p_4 \end{bmatrix} \quad (12)$$

where for  $\mathbf{q} = [q_x \ q_y \ q_z]^\top$ :

$$[\mathbf{q}] = \begin{bmatrix} 0 & -q_z & q_y \\ q_z & 0 & -q_x \\ -q_y & q_x & 0 \end{bmatrix} \quad (13)$$

In state estimation, these operations allow us to model the state on manifold using a Gaussian distribution on its *error state* vector. In particular, the random variable  $\mathbf{x}$  with mean value  $\hat{\mathbf{x}}$  takes the form:

$$\mathbf{x} = \hat{\mathbf{x}} \boxplus \delta\mathbf{x} \quad (14)$$

$$\delta\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \Sigma) \quad (15)$$

where  $\Sigma$  is the covariance of the zero-mean error state. The error state corresponding to (1) is thus given by:

$$\delta\mathbf{x}_k = [{}^k \delta\theta_G^\top \ \delta\mathbf{b}_{\omega_k}^\top \ {}^G \delta\mathbf{v}_k^\top \ \delta\mathbf{b}_{a_k}^\top \ {}^G \delta\mathbf{p}_k^\top]^\top \quad (16)$$

### 3.1 Batch Optimization

In the case of graph SLAM (Grisetti, Kummerle, Stachniss and Burgard 2010), the graph nodes can correspond to historical robot states and features in the environment, while the edges represent collected measurements from sensors which relate the incident nodes. As an example, a robot measuring a feature would add an edge between the feature and the robot state node. Using this graph formulation and under the assumption of independent zero-mean Gaussian noise, we can find a maximum a posteriori (MAP) estimate of all states by solving the following nonlinear least-squares problem (Kummerle, Grisetti, Strasdat, Konolige and Burgard 2011):

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \sum_i \frac{1}{2} \|\mathbf{e}_i(\mathbf{x})\|_{\Lambda_i}^2 \quad (17)$$

where  $\mathbf{e}_i$  is the error/residual of the  $i$ -th measurement,  $\Lambda_i$  is the associated information matrix (inverse covariance), and  $\|\mathbf{v}\|_\Lambda^2 = \mathbf{v}^\top \Lambda \mathbf{v}$  represents the squared energy norm. Note that as a common practice, a (Huber or Cauchy) robust cost function of Equation (17) is often used to compensate for outliers, in particular when fusing visual measurements (Hartley and Zisserman 2000). Optimization is typically performed iteratively, e.g., through a Gauss-Newton or Levenberg–Marquardt method, by linearizing the

nonlinear measurements about the current estimate,  $\hat{\mathbf{x}}$ , and defining a new weighted linear least squares problem in terms of the error state  $\delta\mathbf{x}$ :

$$\delta\hat{\mathbf{x}} = \underset{\delta\mathbf{x}}{\operatorname{argmin}} \sum_i \frac{1}{2} \|\mathbf{e}_i(\hat{\mathbf{x}}) + \mathbf{J}_i \delta\mathbf{x}\|_{\Lambda_i}^2 \quad (18)$$

$$\mathbf{J}_i = \left. \frac{\partial \mathbf{e}_i(\hat{\mathbf{x}} \boxplus \delta\mathbf{x})}{\partial \delta\mathbf{x}} \right|_{\delta\mathbf{x}=0} \quad (19)$$

We can see that the original optimization problem has been converted into finding the optimal *correction* vector,  $\delta\mathbf{x}$ , to the current state estimate. The optimal solution can be found by solving the following normal equation:

$$\left( \sum_i \mathbf{J}_i^\top \Lambda_i \mathbf{J}_i \right) \delta\hat{\mathbf{x}} = - \sum_i \mathbf{J}_i^\top \Lambda_i \mathbf{e}_i(\hat{\mathbf{x}}) \quad (20)$$

$$\iff \boldsymbol{\Lambda} \delta\hat{\mathbf{x}} = -\mathbf{g} \quad (21)$$

After obtaining the optimal correction,  $\delta\hat{\mathbf{x}}$ , we update our current estimate at the  $k$ -th iteration as:  $\hat{\mathbf{x}}^{(k+1)} = \hat{\mathbf{x}}^{(k)} \boxplus \delta\hat{\mathbf{x}}$ , and repeat the optimization process. After convergence, we will be left with the following distribution:

$$\mathbf{x} = \hat{\mathbf{x}} \boxplus \delta\mathbf{x} \quad (22)$$

$$\delta\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \Sigma) \quad (23)$$

$$\Sigma = \left( \sum_i \mathbf{J}_i^\top \Lambda_i \mathbf{J}_i \right)^{-1} \quad (24)$$

where the measurement Jacobians,  $\mathbf{J}_i$ , are evaluated at the final state estimate.

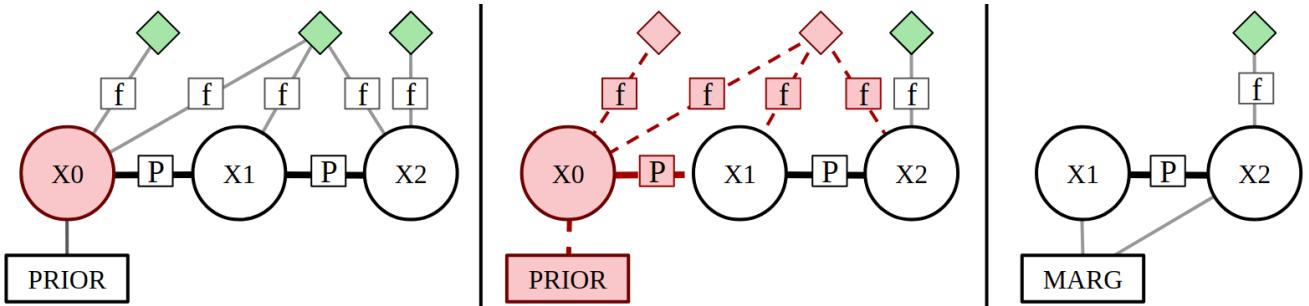
### 3.2 Marginalization

In a naive graph SLAM formulation, nodes are continuously added to the graph as time progresses without consideration to the computational burden. For example, as a robot moves through an unknown environment we would add robot state nodes at *every* measurement time. This becomes a problem due to the high computational complexity,  $O(n^3)$  with  $n = \dim(\mathbf{x})$ , of batch optimization, in the worst case. In order to bound the computational complexity of the system, marginalization is often performed to remove a set of nodes, called marginalized states, from the graph, while retaining the information contained in their incident edges (see Figure 2 for an example) (Huang, Kaess and Leonard 2013; Eckenhoff, Paull and Huang 2016b). Partitioning the optimization variables into states remaining after marginalization,  $\mathbf{x}_r$ , and the to-be marginalized states,  $\mathbf{x}_m$ , we can write (17) as the solution of the following minimization (Huang, Mourikis and Roumeliotis 2011):

$$\{\hat{\mathbf{x}}_r, \hat{\mathbf{x}}_m\} = \underset{\mathbf{x}_r, \mathbf{x}_m}{\operatorname{argmin}} \left( c_r(\mathbf{x}_r) + c_m(\mathbf{x}_m, \mathbf{x}_r) \right) \quad (25)$$

The second subcost,  $c_m(\mathbf{x}_m, \mathbf{x}_r)$ , is associated with the measurements incident to the marginalized states, and is a function of both these states and the remaining ones. The first,  $c_r(\mathbf{x}_r)$ , refers to all other edges in the graph. The optimal estimate for the remaining nodes can be written as:

$$\hat{\mathbf{x}}_r = \underset{\mathbf{x}_r}{\operatorname{argmin}} \left( c_r(\mathbf{x}_r) + \min_{\mathbf{x}_m} c_m(\mathbf{x}_m, \mathbf{x}_r) \right) \quad (26)$$



**Figure 2.** During graph optimization of VINS, IMU states (shown in circles) and 3D features (diamonds) are included in the graph. Image projection measurements connect features and the IMU state corresponding to the time that the image was recorded. Subsequent IMU states are connected with preintegrated factors, while a prior factor connects to the oldest IMU state. During marginalization, we first select the states to be marginalized, e.g., the oldest IMU state in the window and its associated features (in red). With these measurements we perform marginalization to form a new marginal measurement for future optimization.

That is, minimizing  $c_m(\mathbf{x}_m, \mathbf{x}_r)$  with respect to  $\mathbf{x}_m$  yields a cost that is a function *only* of the remaining states. This minimization is performed as in (21), where we write out the linear system for only the measurements involved in  $c_m$ :

$$\begin{bmatrix} \Lambda_{rr} & \Lambda_{rm} \\ \Lambda_{mr} & \Lambda_{mm} \end{bmatrix} \begin{bmatrix} \delta \mathbf{x}_r \\ \delta \mathbf{x}_m \end{bmatrix} = \begin{bmatrix} -\mathbf{g}_r \\ -\mathbf{g}_m \end{bmatrix} \quad (27)$$

The optimal subcost  $c_m$ , up to an irrelevant constant, is given by (Nerurkar, Wu and Roumeliotis 2014):<sup>‡</sup>

$$c_{marg}(\mathbf{x}_r) = \frac{1}{2} \|\mathbf{x}_r \boxminus \check{\mathbf{x}}_r\|_{\Lambda_{marg}}^2 + \mathbf{g}_{marg}^\top (\mathbf{x}_r \boxminus \check{\mathbf{x}}_r) \quad (28)$$

where  $\check{\mathbf{x}}_r$  is the linearization point used to build the system (in practice, the current state estimate at the time of marginalization), and  $\Lambda_{marg} = \Lambda_{rr} - \Lambda_{rm}\Lambda_{mm}^{-1}\Lambda_{mr}$  and  $\mathbf{g}_{marg} = \mathbf{g}_r - \Lambda_{rm}\Lambda_{mm}^{-1}\mathbf{g}_m$  are the marginalized Hessian and gradient, respectively.

In future optimization, this marginalization creates both a new quadratic *and* linear cost in terms of the error between the remaining states and their linearization points. This then replaces the marginalized measurements in the original graph, and we can write this new cost (28) up to a constant in the form of (17):

$$c_{marg}(\mathbf{x}_r) = \frac{1}{2} \|\mathbf{A}_m(\mathbf{x}_r \boxminus \check{\mathbf{x}}_r) + \mathbf{b}_m\|_2^2 \quad (29)$$

$$\text{with } \mathbf{A}_m^\top \mathbf{A}_m = \Lambda_{marg} \quad (30)$$

$$\mathbf{A}_m^\top \mathbf{b}_m = \mathbf{g}_{marg} \quad (31)$$

This cost yields the following residual and Jacobian for use in optimization (see (18) and (19)):

$$\mathbf{e}_{marg}(\hat{\mathbf{x}}) = \mathbf{A}_m(\hat{\mathbf{x}}_r \boxminus \check{\mathbf{x}}_r) + \mathbf{b}_m \quad (32)$$

$$\mathbf{J}_{marg} = \mathbf{A}_m \frac{\partial((\hat{\mathbf{x}}_r \boxplus \delta \mathbf{x}_r) \boxminus \check{\mathbf{x}}_r)}{\partial \delta \mathbf{x}_r} \Big|_{\delta \mathbf{x}_r=0} \quad (33)$$

where for the Jacobian of a vector (i.e., if  $\mathbf{x}_r = \mathbf{v}$ ):

$$\frac{\partial((\hat{\mathbf{v}} \boxplus \delta \mathbf{v}) \boxminus \check{\mathbf{v}})}{\partial \delta \mathbf{v}} = \frac{\partial(\hat{\mathbf{v}} + \delta \mathbf{v} - \check{\mathbf{v}})}{\partial \delta \mathbf{v}} = \mathbf{I} \quad (34)$$

and for a quaternion  $\tilde{q}$ , with  $\tilde{q} = \hat{q} \otimes \check{q}^{-1}$ , we have:

$$\begin{aligned} \frac{\partial((\hat{q} \boxplus \delta \theta) \boxminus \check{q})}{\partial \delta \theta} &= \frac{\partial 2\text{vec} \left( \begin{bmatrix} \frac{\delta \theta}{2} \\ 1 \end{bmatrix} \otimes \hat{q} \otimes \check{q}^{-1} \right)}{\partial \delta \theta} \\ &= \frac{\partial 2\text{vec} \left( \mathcal{R}(\tilde{q}) \begin{bmatrix} \frac{\delta \theta}{2} \\ 1 \end{bmatrix} \right)}{\partial \delta \theta} \\ &= \tilde{q}_4 \mathbf{I} + [\tilde{q}] \end{aligned} \quad (35)$$

## 4 Closed-form Preintegration

In this section, we present in detail the proposed closed-form IMU preintegration based on two different realistic inertial models, which is expected to be readily used in any graph-based aided inertial navigation, thus providing an essential building block for visual-inertial state estimation.

An IMU attached to the robot collects inertial readings of the underlying state dynamics. In particular, the sensor receives angular velocity  $\omega_m$  and local linear acceleration  $\mathbf{a}_m$  measurements which relate to the corresponding true values  $\omega$  and  $\mathbf{a}$  as follows:

$$\omega_m = \omega + \mathbf{b}_\omega + \mathbf{n}_\omega \quad (36)$$

$$\mathbf{a}_m = \mathbf{a} + {}^I_G \mathbf{R}^G \mathbf{g} + \mathbf{b}_a + \mathbf{n}_a \quad (37)$$

where  ${}^G \mathbf{g} = [0 \ 0 \ 9.81]^\top$  is the global gravity<sup>§</sup> and  ${}^I_G \mathbf{R}$  is the rotation from the global frame to the instantaneous local inertial frame. The measurements are corrupted both by the time-varying biases  $\mathbf{b}_\omega$  and  $\mathbf{b}_a$  (which must be co-estimated with the state), and the zero-mean white Gaussian noises  $\mathbf{n}_\omega$  and  $\mathbf{n}_a$ . The standard dynamics of the IMU state is given by

<sup>‡</sup>Throughout the paper, we reserve the symbol  $\hat{x}$  to denote the current estimate of state variable  $x$  in optimization, while  $\check{x}$  refers to the (inferred) measurement mean value.

<sup>§</sup>Note that gravity is slightly different in different parts of the globe.

(Chatfield 1997):

$${}_G^I \dot{\bar{q}} = \frac{1}{2} \Omega (\omega_m - \mathbf{b}_\omega - \mathbf{n}_\omega) {}_G^I \bar{q} \quad (38)$$

$$\dot{\mathbf{b}}_\omega = \mathbf{n}_{\omega b} \quad (39)$$

$${}^G \dot{\mathbf{v}}_I = {}_I^G \mathbf{R} (\mathbf{a}_m - \mathbf{b}_a - \mathbf{n}_a) - {}^G \mathbf{g} \quad (40)$$

$$\dot{\mathbf{b}}_a = \mathbf{n}_{ab} \quad (41)$$

$${}^G \dot{\mathbf{p}}_I = {}^G \mathbf{v}_I \quad (42)$$

where

$$\Omega(\omega) = \begin{bmatrix} -[\omega] & \omega \\ -\omega^\top & 0 \end{bmatrix} \quad (43)$$

#### 4.1 Standard IMU Processing

Given a series of IMU measurements,  $\mathcal{I}$ , collected over a time interval  $[t_k, t_{k+1}]$ , the standard (graph-based) IMU processing considers the following propagation function:

$$\mathbf{x}_{k+1} = \mathbf{g}(\mathbf{x}_k, \mathcal{I}, \mathbf{n}) \quad (44)$$

That is, the future state at time step  $k + 1$  is a function of the current state at step  $k$ , the IMU measurements  $\mathcal{I}$ , and the corresponding measurement noise  $\mathbf{n}$ . Conditioning on the current state, the expected value of the next state is found by evaluating the propagation function with zero noise:

$$\check{\mathbf{x}}_{k+1} = \mathbf{g}(\mathbf{x}_k, \mathcal{I}, \mathbf{0}) \quad (45)$$

which implies that we perform integration of the state dynamics in the absence of noise.

The residual for use in batch optimization of this propagation now constrains the start and end states of the interval and is given by (see Equation (17)):

$$c_{IMU}(\mathbf{x}) = \frac{1}{2} \|\mathbf{x}_{k+1} \ominus \check{\mathbf{x}}_{k+1}\|_{\mathbf{Q}_k^{-1}}^2 \quad (46)$$

$$= \frac{1}{2} \|\mathbf{x}_{k+1} \ominus \mathbf{g}(\mathbf{x}_k, \mathcal{I}, \mathbf{0})\|_{\mathbf{Q}_k^{-1}}^2 \quad (47)$$

where  $\mathbf{Q}_k$  is the linearized, discrete-time noise covariance computed from the IMU noise characterization and is a *function of the state*. This noise covariance matrix and the propagation function can be found by the integration of Equations (38)-(42) and their associated error state dynamics, to which we refer the reader to (Trawny and Roumeliotis 2005; Mourikis and Roumeliotis 2007). It is clear from (45) that ideally we need constantly re-evaluate the propagation function  $\mathbf{g}(\cdot)$  and the residual covariance  $\mathbf{Q}_k$  whenever the linearization point (state estimate) changes. However, the high frequency nature of the IMU sensors and the complexity of the propagation function and the noise covariance, can make direct incorporation of IMU data in real-time graph-based SLAM prohibitively expensive. This motivates the development of inertial preintegration.

#### 4.2 Model 1: Piecewise Constant Measurements

IMU preintegration seeks to directly reduce the computational complexity of incorporating inertial measurements by removing the need to re-integrate the propagation function and noise covariance. This is achieved by processing IMU

measurements in a *local* frame of reference, yielding measurements that are, in contrast to Equation (45), independent of the state (Lupton and Sukkarieh 2012).

Specifically, by denoting  $\Delta T = t_{k+1} - t_k$ , we have the following relationship between a series of IMU measurements, the start state, and the resulting end state (Eckenhoff, Geneva and Huang 2016a):

$${}^G \mathbf{p}_{k+1} = {}^G \mathbf{p}_k + {}^G \mathbf{v}_k \Delta T - \frac{1}{2} {}^G \mathbf{g} \Delta T^2 + {}_k^G \mathbf{R} \int_{t_k}^{t_{k+1}} \int_{t_k}^s {}_u^k \mathbf{R} (\mathbf{a}_m - \mathbf{b}_a - \mathbf{n}_a) du ds \quad (48)$$

$${}^G \mathbf{v}_{k+1} = {}^G \mathbf{v}_k - {}^G \mathbf{g} \Delta T + {}_k^G \mathbf{R} \int_{t_k}^{t_{k+1}} {}_u^k \mathbf{R} (\mathbf{a}_m - \mathbf{b}_a - \mathbf{n}_a) du \quad (49)$$

$${}^k \mathbf{R} = {}_k^{k+1} \mathbf{R} {}_G^k \mathbf{R} \quad (50)$$

$$\mathbf{b}_{\omega_{k+1}} = \mathbf{b}_{\omega_k} + \int_{t_k}^{t_{k+1}} \mathbf{n}_{\omega b} du \quad (51)$$

$$\mathbf{b}_{a_{k+1}} = \mathbf{b}_{a_k} + \int_{t_k}^{t_{k+1}} \mathbf{n}_{ab} du \quad (52)$$

where  $u$  and  $s$  are dummy variables in the integration. From the above, we define the following preintegrated IMU measurements:<sup>¶</sup>

$${}^k \boldsymbol{\alpha}_{k+1} = \int_{t_k}^{t_{k+1}} \int_{t_k}^s {}_u^k \mathbf{R} (\mathbf{a}_m - \mathbf{b}_a - \mathbf{n}_a) du ds \quad (53)$$

$${}^k \boldsymbol{\beta}_{k+1} = \int_{t_k}^{t_{k+1}} {}_u^k \mathbf{R} (\mathbf{a}_m - \mathbf{b}_a - \mathbf{n}_a) du \quad (54)$$

To remove the dependencies of the above preintegrated measurements on the true biases, we linearize about the current bias estimates at time step  $t_k$ ,  $\mathbf{b}_{a_k}^*$  and  $\mathbf{b}_{\omega_k}^*$ . Defining  $\Delta \mathbf{b} = \mathbf{b} - \mathbf{b}^*$ , we have (noting that time indices are occasionally omitted to keep expressions concise, which however can be easily inferred from the context):

$${}^k {}_G \mathbf{R} \left( {}^G \mathbf{p}_{k+1} - {}^G \mathbf{p}_k - {}^G \mathbf{v}_k \Delta T + \frac{1}{2} {}^G \mathbf{g} \Delta T^2 \right) \simeq \quad (55)$$

$${}^k \boldsymbol{\alpha}_{k+1} (\mathbf{b}_{\omega_k}^*, \mathbf{b}_{a_k}^*) + \left. \frac{\partial \boldsymbol{\alpha}}{\partial \mathbf{b}_\omega} \right|_{\mathbf{b}_{\omega_k}^*} \Delta \mathbf{b}_\omega + \left. \frac{\partial \boldsymbol{\alpha}}{\partial \mathbf{b}_a} \right|_{\mathbf{b}_{a_k}^*} \Delta \mathbf{b}_a$$

$${}^k {}_G \mathbf{R} ({}^G \mathbf{v}_{k+1} - {}^G \mathbf{v}_k + {}^G \mathbf{g} \Delta T) \simeq \quad (56)$$

$${}^k \boldsymbol{\beta}_{k+1} (\mathbf{b}_{\omega_k}^*, \mathbf{b}_{a_k}^*) + \left. \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{b}_\omega} \right|_{\mathbf{b}_{\omega_k}^*} \Delta \mathbf{b}_\omega + \left. \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{b}_a} \right|_{\mathbf{b}_{a_k}^*} \Delta \mathbf{b}_a$$

$${}^k {}_G \mathbf{R} {}_G^k \mathbf{R}^\top \simeq \mathbf{R} \left( \left. \frac{\partial \mathbf{R}}{\partial \mathbf{b}_\omega} \right|_{\mathbf{b}_{\omega_k}^*} \Delta \mathbf{b}_\omega \right) {}_k^{k+1} \mathbf{R} (\mathbf{b}_{\omega_k}^*) \quad (57)$$

Note that Equations (55) and (56) are simple Taylor series expansions for our  ${}^k \boldsymbol{\alpha}_{k+1}$  and  ${}^k \boldsymbol{\beta}_{k+1}$  measurements, while Equation (57) models an additional rotation induced due to a change of the linearization point (estimate) of the gyro bias (Forster, Carbone, Dellaert and Scaramuzza 2015; Eckenhoff, Geneva and Huang 2016a).

<sup>¶</sup>Note that along with the preintegrated inertial measurements in Equations (53) and (54), the preintegrated relative-orientation measurement  ${}^k \bar{q}$  (or  ${}^k \mathbf{R}$ ) can be obtained from the integration of the gyro measurements.

The preintegrated measurement's mean values,  ${}^k\check{\alpha}_{k+1}$ ,  ${}^k\check{\beta}_{k+1}$ , and  ${}^k\check{q}$ , must be computed for use in graph optimization. It is important to note that current preintegration methods (Lupton and Sukkarieh 2012; Forster, Carbone, Dellaert and Scaramuzza 2015; Ling, Liu and Shen 2016) are all based on discrete integration of the measurement dynamics through Euler or midpoint integration. In particular, the discrete approximation used by Forster, Carbone, Dellaert and Scaramuzza (2015) in fact corresponds to a piecewise constant *global acceleration* model (expressed in the fixed global frame of reference), which may be easily violated in realistic navigation. By contrast, we here offer *closed-form* solutions for the measurement means under the assumptions of piecewise constant (local) *measurements* and piecewise constant *local acceleration* (expressed in local coordinates) which will be presented later in Section 4.3.

**4.2.1 Computing preintegration mean:** Between two image times,  $t_k$  and  $t_{k+1}$ , the IMU receives a series of inertial measurements. We denote  $\tau$  as the step at which an IMU measurement is received, and  $\tau + 1$  as the step of the *next* IMU reading. The time associated with each of these steps is given by  $t_\tau$  and  $t_{\tau+1}$ , respectively. The relative orientation between the interval,  ${}^k\check{q}$ , can be found using successive applications of the zeroth order quaternion integrator (Trawny and Roumeliotis 2005). Based on the definitions of  ${}^k\alpha_{k+1}$  and  ${}^k\beta_{k+1}$  (see Equations (53) and (54)), we have the following continuous-time dynamics at every step  $u$  with  $t_u \in [t_\tau, t_{\tau+1}]$ :

$${}^k\dot{\alpha}_u = {}^k\beta_u \quad (58)$$

$${}^k\dot{\beta}_u = {}^k_u\mathbf{R}(\mathbf{a}_m - \mathbf{b}_a - \mathbf{n}_a) \quad (59)$$

From these governing differential equations, we formulate the following *linear* system that describes the evolution of the measurements by taking the expectation operation:

$$\begin{bmatrix} {}^k\dot{\alpha}_u \\ {}^k\dot{\beta}_u \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} {}^k\check{\alpha}_u \\ {}^k\check{\beta}_u \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ {}^k_u\check{\mathbf{R}} \end{bmatrix} (\mathbf{a}_m - \mathbf{b}_{a_k}^*) \quad (60)$$

Given  $\mathbf{a}_m$  and  $\omega_m$  sampled at time  $t_\tau$  and assuming that these local IMU measurements are *piecewise constant* during  $[t_\tau, t_{\tau+1}]$ , we analytically solve the above linear time-varying (LTV) system to obtain the updated preintegration mean values, which are computed as follows (Eckenhoff, Geneva and Huang 2018):

$$\begin{bmatrix} {}^k\check{\alpha}_{\tau+1} \\ {}^k\check{\beta}_{\tau+1} \end{bmatrix} = \begin{bmatrix} {}^k\check{\alpha}_\tau + {}^k\check{\beta}_\tau \Delta t + \mathbf{A}_\tau \hat{\mathbf{a}} \\ {}^k\check{\beta}_\tau + \mathbf{B}_\tau \hat{\mathbf{a}} \end{bmatrix} \quad (61)$$

$$\begin{aligned} \mathbf{A}_\tau &= {}^k_{\tau+1}\check{\mathbf{R}} \left( \frac{\Delta t^2}{2} \mathbf{I}_{3 \times 3} + \frac{|\hat{\omega}| \Delta t \cos(|\hat{\omega}| \Delta t) - \sin(|\hat{\omega}| \Delta t)}{|\hat{\omega}|^3} [\hat{\omega}] \right. \\ &\quad \left. + \frac{(|\hat{\omega}| \Delta t)^2 - 2 \cos(|\hat{\omega}| \Delta t) - 2(|\hat{\omega}| \Delta t) \sin(|\hat{\omega}| \Delta t) + 2}{2|\hat{\omega}|^4} [\hat{\omega}]^2 \right) \end{aligned} \quad (62)$$

$$\begin{aligned} \mathbf{B}_\tau &= {}^k_{\tau+1}\check{\mathbf{R}} \left( \Delta t \mathbf{I}_{3 \times 3} - \frac{1 - \cos(|\hat{\omega}| (\Delta t))}{|\hat{\omega}|^2} [\hat{\omega}] \right. \\ &\quad \left. + \frac{(|\hat{\omega}| \Delta t) - \sin(|\hat{\omega}| \Delta t)}{|\hat{\omega}|^3} [\hat{\omega}]^2 \right) \end{aligned} \quad (63)$$

where we have employed the definitions:  $\hat{\omega} = \omega_m - \mathbf{b}_{a_k}^*$ ,  $\hat{\mathbf{a}} = \mathbf{a}_m - \mathbf{b}_{a_k}^*$ , and  $\Delta t = t_{\tau+1} - t_\tau$ . Clearly, these *closed-form* expressions reveal the higher order affect of the angular

velocity on the preintegrated measurements due to the evolution of the orientation over the IMU sampling interval.

**4.2.2 Computing preintegration covariance:** In order to derive the preintegrated measurement covariance, we first write the linearized *measurement* error system as follows (Eckenhoff, Geneva and Huang 2018):

$$\begin{bmatrix} {}^u\delta\theta_k \\ \tilde{\mathbf{b}}_w \\ {}^k\delta\beta_u \\ \tilde{\mathbf{b}}_a \\ {}^k\delta\alpha_u \end{bmatrix} = \begin{bmatrix} -[\hat{\omega}] & -\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -{}^k_u\check{\mathbf{R}}[\hat{\mathbf{a}}] & \mathbf{0} & \mathbf{0} & -{}^k_u\check{\mathbf{R}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} {}^u\delta\theta_k \\ \tilde{\mathbf{b}}_w \\ {}^k\delta\beta_u \\ \tilde{\mathbf{b}}_a \\ {}^k\delta\alpha_u \end{bmatrix} + \begin{bmatrix} -\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -{}^k_u\check{\mathbf{R}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{n}_\omega \\ \mathbf{n}_{\omega b} \\ \mathbf{n}_a \\ \mathbf{n}_{ab} \end{bmatrix} \quad (64)$$

$$\iff \dot{\mathbf{r}} = \mathbf{F}\mathbf{r} + \mathbf{G}\mathbf{n} \quad (65)$$

which is akin to the standard VINS error state propagation equations in a local frame of reference (Mourikis and Roumeliotis 2007).

It is important to note that in contrast to our previous work (Eckenhoff, Geneva and Huang 2016a, 2017), we here couple the preintegration bias and measurement evolution for improved accuracy. Note also that the bias error terms in Equation (64),  $\tilde{\mathbf{b}}_w$  and  $\tilde{\mathbf{b}}_a$ , describe the deviation of the bias over the interval due to the random-walk drift, rather than the error of the current bias estimate. The discrete state transition matrix  $\Phi(t_{\tau+1}, t_\tau)$  can be computed either analytically in closed-form or numerically using Runge-Kutta methods based on the following continuous-time differential equation (see Hesch, Kottas, Bowman and Roumeliotis (2013); Trawny and Roumeliotis (2005)):

$$\dot{\Phi}(t_u, t_\tau) = \mathbf{F}(u) \Phi(t_u, t_\tau) \quad (66)$$

$$\Phi(t_\tau, t_\tau) = \mathbf{I} \quad (67)$$

The propagation of the measurement covariance,  $\mathbf{P}$ , over the time interval  $t_\tau \in [t_k, t_{k+1}]$ , takes the following form:

$$\mathbf{P}_k = \mathbf{0} \quad (68)$$

$$\mathbf{P}_{\tau+1} = \Phi(t_{\tau+1}, t_\tau) \mathbf{P}_\tau \Phi(t_{\tau+1}, t_\tau)^\top + \mathbf{Q}_\tau \quad (69)$$

$$\mathbf{Q}_\tau = \int_{t_\tau}^{t_{\tau+1}} \Phi(t_{\tau+1}, u) \mathbf{G}(u) \mathbf{Q}_c \mathbf{G}(u)^\top \Phi(t_{\tau+1}, u)^\top du \quad (70)$$

where  $\mathbf{Q}_c$  is the continuous-time IMU noise covariance. To keep presentation concise, the discrete-time noise covariance  $\mathbf{Q}_\tau$ , can be computed similarly as in (Trawny and Roumeliotis 2005).

**4.2.3 Preintegration measurement residuals and Jacobians:** For use in optimization, we form the associated preintegration measurement cost and residual as follows:

$$c_{IMU}(\mathbf{x}) = \frac{1}{2} \|\mathbf{e}_{IMU}(\mathbf{x})\|_{\mathbf{P}_{k+1}^{-1}}^2 \quad (71)$$

$$\mathbf{e}_{IMU}(\mathbf{x}) = \begin{bmatrix} 2\text{vec} \left( {}_G^k \bar{q} \otimes {}_G^k \bar{q}^{-1} \otimes {}_k^{k+1} \check{q}^{-1} \otimes \bar{q}_b \right) \\ \mathbf{b}_{\omega_{k+1}} - \mathbf{b}_{\omega_k} \\ {}_G^k \mathbf{R} \left( {}^G \mathbf{v}_{k+1} - {}^G \mathbf{v}_k + {}^G \mathbf{g} \Delta T \right) \\ -\mathbf{J}_\beta (\mathbf{b}_{\omega_k} - \mathbf{b}_{\omega_k}^*) - \mathbf{H}_\beta (\mathbf{b}_{a_k} - \mathbf{b}_{a_k}^*) - {}^k \check{\mathbf{b}}_{k+1} \\ \mathbf{b}_{a_{k+1}} - \mathbf{b}_{a_k} \\ \left( {}_G^k \mathbf{R} \left( {}^G \mathbf{p}_{k+1} - {}^G \mathbf{p}_k - {}^G \mathbf{v}_k \Delta T + \frac{1}{2} {}^G \mathbf{g} \Delta T^2 \right) \right. \\ \left. - \mathbf{J}_\alpha (\mathbf{b}_{\omega_k} - \mathbf{b}_{\omega_k}^*) - \mathbf{H}_\alpha (\mathbf{b}_{a_k} - \mathbf{b}_{a_k}^*) - {}^k \check{\mathbf{a}}_{k+1} \right) \end{bmatrix} \quad (72)$$

where we have employed  $\bar{q}_b = \begin{bmatrix} \frac{\theta}{\|\theta\|} \sin\left(\frac{\|\theta\|}{2}\right) \\ \cos\left(\frac{\|\theta\|}{2}\right) \end{bmatrix}$  and  $\theta =$

$\mathbf{J}_q (\mathbf{b}_{\omega_k} - \mathbf{b}_{\omega_k}^*)$ . In the above expressions,  $\mathbf{J}_q$ ,  $\mathbf{J}_\alpha$ ,  $\mathbf{J}_\beta$ ,  $\mathbf{H}_\alpha$ , and  $\mathbf{H}_\beta$ , are the Jacobian matrices of the pertinent residuals with respect to the biases, which are used to correct the measurements due to a change in the initial bias estimate  $\mathbf{b}^*$ , thus compensating for the fact that preintegrated measurements have been linearized about  $\mathbf{b}_{\omega_k}^*$  and  $\mathbf{b}_{a_k}^*$  without having to recompute the required integrals whenever the bias estimates change (see Equations (55) and (56)). In particular, using the fact that our preintegrated measurement means are linear in the acceleration bias  $\mathbf{b}_a$  (see Equation (61)), we have the following dynamics of its Jacobians (see Equations (62) and (63)):

$$\begin{bmatrix} \frac{\partial \mathbf{a}}{\partial \mathbf{b}_a} \\ \frac{\partial \mathbf{a}}{\partial \mathbf{b}_a} \end{bmatrix} =: \begin{bmatrix} \mathbf{H}_\alpha (\tau + 1) \\ \mathbf{H}_\beta (\tau + 1) \end{bmatrix} \\ = \begin{bmatrix} \mathbf{H}_\alpha (\tau) + \mathbf{H}_\beta (\tau) \Delta t - \mathbf{A}_\tau \\ \mathbf{H}_\beta (\tau) - \mathbf{B}_\tau \end{bmatrix} \quad (73)$$

Similarly, for the gyroscope bias Jacobians, we have:

$$\begin{bmatrix} \frac{\partial \mathbf{a}}{\partial \mathbf{b}_\omega} \\ \frac{\partial \mathbf{a}}{\partial \mathbf{b}_\omega} \end{bmatrix} =: \begin{bmatrix} \mathbf{J}_\alpha (\tau + 1) \\ \mathbf{J}_\beta (\tau + 1) \end{bmatrix} \\ = \begin{bmatrix} \mathbf{J}_\alpha (\tau) + \mathbf{J}_\beta (\tau) \Delta t + \frac{\partial \mathbf{A}_\tau \dot{\mathbf{a}}}{\partial \mathbf{b}_\omega} \\ \mathbf{J}_\beta (\tau) + \frac{\partial \mathbf{B}_\tau \dot{\mathbf{a}}}{\partial \mathbf{b}_\omega} \end{bmatrix} \quad (74)$$

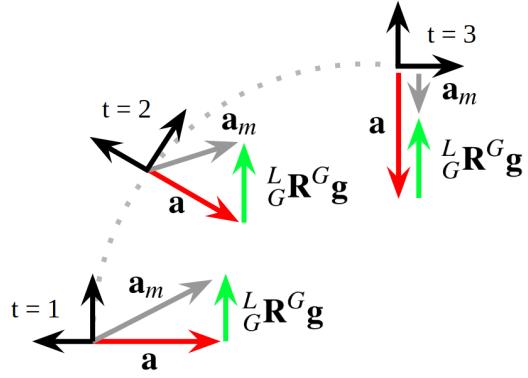
Finally, the orientation Jacobian with respect to gyroscope bias can be found incrementally as:

$$\mathbf{J}_q(\tau + 1) = {}^\tau \check{\mathbf{R}} \mathbf{J}_q(\tau) + \mathbf{J}_r(\hat{\omega} \Delta t) \Delta t \quad (75)$$

where  $\mathbf{J}_r(\cdot)$  is the right Jacobian of  $SO(3)$  and is defined as (Chirikjian 2011):

$$\mathbf{J}_r(\phi) = \mathbf{I}_{3 \times 3} - \frac{1 - \cos(\|\phi\|)}{\|\phi\|^2} [\phi] + \frac{\|\phi\| - \sin(\|\phi\|)}{\|\phi\|^3} [\phi]^2 \quad (76)$$

Moreover, the measurement Jacobians of these preintegrated measurements with respect to the error state (16), can also be analytically computed as shown in Appendix B.1, which are essential for batch optimization. For the detailed derivations and closed-form expressions of the preintegrated measurements and Jacobians, the reader is referred to our companion technical report (Eckenhoff, Geneva and Huang 2018).



**Figure 3.** An example of an IMU rotating about the gravity. It can be seen that the true local acceleration  $\mathbf{a}$  (red) remains constant, while its local measurement  $\mathbf{a}_m$  (grey) changes continuously due to the effect of gravity (green).

#### 4.3 Model 2: Piecewise Constant Local Acceleration

The previous preintegration (Model 1) assumes that noiseless IMU measurements can be approximated as remaining constant over a sampling interval, which, however, might not always be a good approximation (see Figure 3). In this section, we propose a new preintegration model that instead assumes piecewise constant *true* local acceleration during the sampling time interval, which may better approximate motion dynamics in practice. To this end, we first rewrite Equations (48) and (49) as:

$${}^G \mathbf{p}_{k+1} = {}^G \mathbf{p}_k + {}^G \mathbf{v}_k \Delta T + {}^G \mathbf{R} \int_{t_k}^{t_{k+1}} \int_{t_k}^s {}_u^k \mathbf{R} \mathbf{a} du ds \quad (77)$$

$${}^G \mathbf{v}_{k+1} = {}^G \mathbf{v}_k + {}^G \mathbf{R} \int_{t_k}^{t_{k+1}} {}_u^k \mathbf{R} \mathbf{a} du \quad (78)$$

Note that we have moved the effect of gravity back inside the integrals. We then define the following vectors:

$$\Delta \mathbf{p} = \int_{t_k}^{t_{k+1}} \int_{t_k}^s {}_u^k \mathbf{R} \mathbf{a} du ds \quad (79)$$

$$\Delta \mathbf{v} = \int_{t_k}^{t_{k+1}} {}_u^k \mathbf{R} \mathbf{a} du \quad (80)$$

which essentially are the true local position displacement and velocity change during  $[t_k, t_{k+1}]$ , and yields:

$$\Delta \dot{\mathbf{p}} = \Delta \mathbf{v} \quad (81)$$

$$\Delta \dot{\mathbf{v}} = {}_u^k \mathbf{R} \mathbf{a} \quad (82)$$

In particular, between two IMU measurement times inside the preintegration interval,  $[t_\tau, t_{\tau+1}] \subset [t_k, t_{k+1}]$ , we assume that the *local* acceleration will be constant:

$$\forall t_u \in [t_\tau, t_{\tau+1}], \quad \mathbf{a}(t_u) = \mathbf{a}(t_\tau) \quad (83)$$

Using this sampling model we can rewrite (82) as:

$$\Delta \dot{\mathbf{v}} = {}_u^k \mathbf{R} (\mathbf{a}_m - \mathbf{b}_a - \mathbf{n}_a - {}_k^T \mathbf{R} {}_G^k \mathbf{R}^G \mathbf{g}) \quad (84)$$

We now write the relationship of the states at the beginning and end of the interval as (see Equations (77) and (78)):

$${}^G \mathbf{R} ({}^G \mathbf{p}_{k+1} - {}^G \mathbf{p}_k - {}^G \mathbf{v}_k \Delta T) = \Delta \mathbf{p} \quad (85)$$

$${}^G \mathbf{R} ({}^G \mathbf{v}_{k+1} - {}^G \mathbf{v}_k) = \Delta \mathbf{v} \quad (86)$$

It is important to note that, since  $\Delta p$  and  $\Delta v$  are functions of both the biases *and* the initial orientation, we perform the following linearization with respect to these states:

$$\begin{aligned} {}^k_R \left( {}^G \mathbf{p}_{k+1} - {}^G \mathbf{p}_k - {}^G \mathbf{v}_k \Delta T \right) &\simeq \Delta p (\mathbf{b}_{\omega_k}^*, \mathbf{b}_{a_k}^*, {}^k_G \bar{q}^*) \\ + \frac{\partial \Delta p}{\partial \mathbf{b}_\omega} \Big|_{\mathbf{b}_{\omega_k}^*} \Delta \mathbf{b}_\omega + \frac{\partial \Delta p}{\partial \mathbf{b}_a} \Big|_{\mathbf{b}_{a_k}^*} \Delta \mathbf{b}_a + \frac{\partial \Delta p}{\partial \Delta \theta_k} \Big|_{{}^k_G \bar{q}^*} \Delta \theta_k \end{aligned}$$

$${}^k_R \left( {}^G \mathbf{v}_{k+1} - {}^G \mathbf{v}_k \right) \simeq \Delta v (\mathbf{b}_{\omega_k}^*, \mathbf{b}_{a_k}^*, {}^k_G \bar{q}^*) \quad (87)$$

$$+ \frac{\partial \Delta v}{\partial \mathbf{b}_\omega} \Big|_{\mathbf{b}_{\omega_k}^*} \Delta \mathbf{b}_\omega + \frac{\partial \Delta v}{\partial \mathbf{b}_a} \Big|_{\mathbf{b}_{a_k}^*} \Delta \mathbf{b}_a + \frac{\partial \Delta v}{\partial \Delta \theta_k} \Big|_{{}^k_G \bar{q}^*} \Delta \theta_k \quad (88)$$

where  $\Delta \theta_k = 2 \text{vec}({}^k_G \bar{q} \otimes {}^k_G \bar{q}^{*-1})$  is the rotation angle change associated with the change of the linearization point of quaternion  ${}^k_G \bar{q}$ .

**4.3.1 Computing preintegration mean:** To compute the new preintegrated measurement mean values, we first determine the continuous-time dynamics of the expected preintegration vectors by taking expectations of Equations (81) and (84), given by:

$$\dot{\Delta \tilde{p}} = \Delta \tilde{v} \quad (89)$$

$$\dot{\Delta \tilde{v}} = {}^k \check{\mathbf{R}} \left( \mathbf{a}_m - \mathbf{b}_{a_k}^* - {}^k \check{\mathbf{R}}_G^k \mathbf{R}^{*G} \mathbf{g} \right) \quad (90)$$

As in the case of Model 1 (see Section 4.2.1), we can formulate a linear system of the new preintegration measurement vectors and find the closed-form solutions. Specifically, we can integrate these differential equations and obtain the solution similar to Equation (61), while using the new definition:  $\hat{\mathbf{a}} = \mathbf{a}_m - \mathbf{b}_{a_k}^* - {}^k \check{\mathbf{R}}_G^k \mathbf{R}^{*G} \mathbf{g}$ , which serves as the estimate for the piecewise constant local acceleration over the sampling interval.

**4.3.2 Computing preintegration covariance:** To compute the new preintegration measurement covariance, we first determine the differential equations for the corresponding preintegration measurement errors (see Equations (81), (84), (89) and (90)):

$$\dot{\Delta \tilde{p}} = \Delta v - \Delta \tilde{v} = \Delta \tilde{v} \quad (91)$$

$$\begin{aligned} \dot{\Delta \tilde{v}} &= {}^k \check{\mathbf{R}} (\mathbf{I} + \lfloor {}^u \delta \theta_k \rfloor) \left( \mathbf{a}_m - \mathbf{b}_{a_k}^* - \tilde{\mathbf{b}} \right. \\ &\quad \left. - (\mathbf{I} - \lfloor {}^\tau \delta \theta_k \rfloor) {}^k \check{\mathbf{R}}_G^k \mathbf{R}^{*G} \mathbf{g} - \mathbf{n}_a \right) \\ &\quad - {}^k \check{\mathbf{R}} \left( \mathbf{a}_m - \mathbf{b}_{a_k}^* - {}^k \check{\mathbf{R}}_G^k \mathbf{R}^{*G} \mathbf{g} \right) \\ &= - {}^k \check{\mathbf{R}} \lfloor \hat{\mathbf{a}} \rfloor {}^u \delta \theta_k - {}^k \check{\mathbf{R}} \tilde{\mathbf{b}} \\ &\quad - {}^k \check{\mathbf{R}} \lfloor {}^\tau \tilde{\mathbf{g}} \rfloor {}^\tau \delta \theta_k - {}^k \check{\mathbf{R}} \mathbf{n}_a \end{aligned} \quad (92)$$

where  ${}^\tau \tilde{\mathbf{g}}$  represents the estimate for gravity in the sampled  $\tau$  frame. It is important to notice that, in the above expressions, we have used two angle errors: (i)  ${}^u \delta \theta_k$  corresponds to the active local IMU orientation error, and (ii)  ${}^\tau \delta \theta_k$  corresponds to the cloned orientation error at the sampling time  $t_\tau$ . In addition, the bias errors  $\tilde{\mathbf{b}}$  describe the deviation of the bias from the starting value over the interval due to bias drift. With this, we have the following time evolution of the full

preintegrated measurement errors:

$$\begin{bmatrix} {}^u \delta \dot{\theta}_k \\ \dot{\tilde{\mathbf{b}}}_\omega \\ \Delta \tilde{v} \\ \dot{\tilde{\mathbf{b}}}_a \\ \Delta \tilde{p} \\ {}^\tau \delta \theta_k \end{bmatrix} = \mathbf{F} \begin{bmatrix} {}^u \delta \theta_k \\ \tilde{\mathbf{b}}_\omega \\ \Delta \tilde{v} \\ \tilde{\mathbf{b}}_a \\ \Delta \tilde{p} \\ {}^\tau \delta \theta_k \end{bmatrix} + \begin{bmatrix} -\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -{}^k \check{\mathbf{R}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{n}_\omega \\ \mathbf{n}_{\omega b} \\ \mathbf{n}_a \\ \mathbf{n}_{ab} \end{bmatrix} \iff \dot{\mathbf{r}} = \mathbf{Fr} + \mathbf{Gn} \quad (93)$$

where

$$\mathbf{F} = \begin{bmatrix} -[\dot{\omega}] & -\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -{}^k \check{\mathbf{R}} [\dot{\mathbf{a}}] & \mathbf{0} & \mathbf{0} & -{}^k \check{\mathbf{R}} & \mathbf{0} & -{}^k \check{\mathbf{R}} [{}^\tau \tilde{\mathbf{g}}] \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (94)$$

In analogy to Equations (66), (67), and (70), we can determine the new state-transition matrix  $\Phi(t_{\tau+1}, t_\tau)$  and the new discrete noise covariance  $\mathbf{Q}_\tau$ . With that, we now propagate the measurement covariance over the time interval  $t_\tau \in [t_k, t_{k+1}]$  as follows:

$$\mathbf{P}_k = \mathbf{0} \quad (95)$$

$$\mathbf{P}_{\tau+1} = \Phi(t_{\tau+1}, t_\tau) \mathbf{P}_\tau \Phi(t_{\tau+1}, t_\tau)^\top + \mathbf{Q}_\tau \quad (96)$$

$$\mathbf{P}_{\tau+1} = \Gamma \mathbf{P}_{\tau+1} \Gamma^\top \quad (97)$$

where  $\Gamma$  is the permutation matrix that allows us to replace the previous static orientation error  ${}^\tau \delta \theta_k$  by the new one  ${}^{\tau+1} \delta \theta_k$  simply by cloning the current local orientation error  ${}^u \delta \theta_k$  at the end of current sampling interval  $t_u = t_{\tau+1}$  when moving from the current measurement time interval  $[t_\tau, t_{\tau+1}]$  to the next one, and is given by:

$$\Gamma = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (98)$$

The resulting preintegrated measurement covariance is then extracted from the top left  $15 \times 15$  block of  $\mathbf{P}_{k+1}$  after the propagation with Equations (95)-(97) over the entire preintegration interval  $[t_k, t_{k+1}]$ .

**4.3.3 Preintegration measurement residuals and Jacobians:** As we linearize this preintegration with respect to the IMU biases and the initial orientation, it is important to compute the Jacobians with respect to these quantities. In particular, we note that the solution to the preintegration equation for Model 2 can be expressed as:

$$\Delta \tilde{p}_{\tau+1} = \Delta \tilde{p}_\tau + \Delta \tilde{v}_\tau \Delta t + \mathbf{A}_\tau \left( \mathbf{a}_m - \mathbf{b}_a^* - {}^k \check{\mathbf{R}}_G^k \mathbf{R}^{*G} \mathbf{g} \right)$$

$$\Delta \tilde{v}_{\tau+1} = \Delta \tilde{v}_\tau + \mathbf{B}_\tau \left( \mathbf{a}_m - \mathbf{b}_a^* - {}^k \check{\mathbf{R}}_G^k \mathbf{R}^{*G} \mathbf{g} \right) \quad (99)$$

where  $\mathbf{A}_\tau$  and  $\mathbf{B}_\tau$  are defined the same as in Equations (62) and (63). Letting  $\mathbf{O}_\alpha$  and  $\mathbf{O}_\beta$  denote the Jacobians of the position and velocity preintegrated measurements with

respect to the initial orientation, we have:

$$\begin{bmatrix} \mathbf{J}_\alpha(\tau+1) \\ \mathbf{J}_\beta(\tau+1) \end{bmatrix} = \begin{bmatrix} \mathbf{J}_\alpha(\tau) + \mathbf{J}_\beta(\tau) \Delta t \\ \mathbf{J}_\beta(\tau) \end{bmatrix} + \begin{bmatrix} \frac{\partial \mathbf{A}_\alpha \hat{\mathbf{a}}}{\partial \mathbf{b}_\omega} + \mathbf{A}_\tau [\mathbf{R}_k^\top \mathbf{R}_G^* \mathbf{g}] \mathbf{J}_q(\tau) \\ \frac{\partial \mathbf{B}_\alpha \hat{\mathbf{a}}}{\partial \mathbf{b}_\omega} + \mathbf{B}_\tau [\mathbf{R}_k^\top \mathbf{R}_G^* \mathbf{g}] \mathbf{J}_q(\tau) \end{bmatrix} \quad (100)$$

$$\begin{bmatrix} \mathbf{H}_\alpha(\tau+1) \\ \mathbf{H}_\beta(\tau+1) \end{bmatrix} = \begin{bmatrix} \mathbf{H}_\alpha(\tau) + \mathbf{H}_\beta(\tau) \Delta t - \mathbf{A}_\tau \\ \mathbf{H}_\beta(\tau) - \mathbf{B}_\tau \end{bmatrix} \quad (101)$$

$$\begin{bmatrix} \mathbf{O}_\alpha(\tau+1) \\ \mathbf{O}_\beta(\tau+1) \end{bmatrix} = \begin{bmatrix} \mathbf{O}_\alpha(\tau) + \mathbf{O}_\beta(\tau) \Delta t \\ \mathbf{O}_\beta(\tau) \end{bmatrix} - \begin{bmatrix} \mathbf{A}_\tau \mathbf{R}_k^\top \mathbf{R}_G^* \mathbf{g} \\ \mathbf{B}_\tau \mathbf{R}_k^\top \mathbf{R}_G^* \mathbf{g} \end{bmatrix} \quad (102)$$

We note that Equation (102) reveals that only changes in the initial orientation *perpendicular* to local gravity ( ${}^k\mathbf{g}$ ) will cause a change in the preintegrated measurement. As these directions of orientation are observable and thus are expected to have small errors, this highlights the fact that our linearization scheme about the initial orientation is appropriate. At this point, using these Jacobians, we can write the residual associated with the new preintegrated IMU measurement as follows:

$$\mathbf{e}_{IMU}(\mathbf{x}) = \begin{bmatrix} 2\text{vec}({}^{k+1}\bar{q} \otimes {}^k\bar{q}^{-1} \otimes {}^{k+1}\check{q}^{-1} \otimes \bar{q}_b) \\ \mathbf{b}_{\omega_{k+1}} - \mathbf{b}_{\omega_k} \\ \left( {}_G^k \mathbf{R} ({}^G \mathbf{v}_{k+1} - {}^G \mathbf{v}_k) - \mathbf{J}_\beta (\mathbf{b}_{\omega_k} - \mathbf{b}_{\omega_k}^*) - \mathbf{H}_\beta (\mathbf{b}_{a_k} - \mathbf{b}_{a_k}^*) - \mathbf{O}_\beta 2\text{vec}({}^k\bar{q} \otimes {}^k\bar{q}^{*-1}) - \Delta \check{v} \right) \\ \mathbf{b}_{a_{k+1}} - \mathbf{b}_{a_k} \\ \left( {}_G^k \mathbf{R} ({}^G \mathbf{p}_{k+1} - {}^G \mathbf{p}_k - {}^G \mathbf{v}_k \Delta T) - \mathbf{J}_\alpha (\mathbf{b}_{\omega_k} - \mathbf{b}_{\omega_k}^*) - \mathbf{H}_\alpha (\mathbf{b}_{a_k} - \mathbf{b}_{a_k}^*) - \mathbf{O}_\alpha 2\text{vec}({}^k\bar{q} \otimes {}^k\bar{q}^{*-1}) - \Delta \check{p} \right) \end{bmatrix} \quad (103)$$

The resulting measurement Jacobians are necessary for an iterative solver, which we analytically compute as shown in Appendix B.2.

## 5 Visual-Inertial Navigation

To demonstrate the applicability of the proposed closed-form preintegration (CPI) theory presented in the preceding section, in this section, we develop two sliding-window optimization-based sensor fusion schemes for visual-inertial navigation systems (VINS) that utilize our inertial preintegration.

### 5.1 Tightly-Coupled Indirect VIO

As an IMU-camera sensor suite moves through an unknown environment, visual feature keypoints can be extracted and tracked from the images to provide motion information about

the platform. In particular, the measurement function that maps the 3D position,  ${}^G \mathbf{p}_f$ , of a feature into the normalized uv-coordinates on the  $j$ -th camera's image plane at time step  $k$  takes the following form:

$$\mathbf{z}_{fjk} = \Pi \left( {}_I^C \mathbf{R} {}_G^k \mathbf{R} ({}^G \mathbf{p}_f - {}^G \mathbf{p}_k) + {}_C^j \mathbf{p}_I \right) + \mathbf{n}_f \quad (104)$$

where  ${}_I^C \mathbf{R}$  and  ${}_C^j \mathbf{p}_I$  are the rigid IMU-to-camera extrinsic calibration parameters,  $\mathbf{n}_f \sim \mathcal{N}(\mathbf{0}, \Lambda_{fjk}^{-1})$ , and  $\Pi(\cdot)$  is the perspective projection function given by (Hartley and Zisserman 2000):

$$\Pi \left( \begin{bmatrix} x \\ y \\ z \end{bmatrix} \right) = \begin{bmatrix} x/z \\ y/z \end{bmatrix} \quad (105)$$

The error (or residual) associated with this visual measurement is given by:

$$\mathbf{e}_{fjk}(\mathbf{x}) = \Pi \left( {}_I^C \mathbf{R} {}_G^k \mathbf{R} ({}^G \mathbf{p}_f - {}^G \mathbf{p}_k) + {}_C^j \mathbf{p}_I \right) - \mathbf{z}_{fjk} \quad (106)$$

Using all these visual measurements available in a sliding window along with the preintegrated IMU measurements and marginalization prior, we solve the following optimization problem that tightly couples all available measurement residuals:

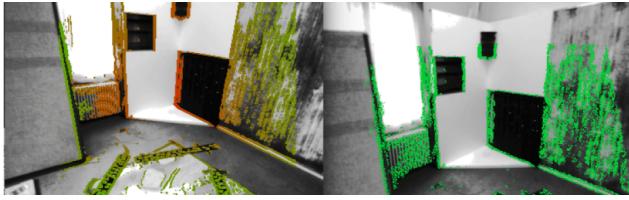
$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left( \|\mathbf{e}_{marg}(\mathbf{x})\|_2^2 + \sum_{p \in \mathcal{P}} \|\mathbf{e}_{IMU}(\mathbf{x})\|_{\mathbf{P}_p^{-1}}^2 + \sum_{(f,j,k) \in \mathcal{C}} \|\mathbf{e}_{fjk}(\mathbf{x})\|_{\Lambda_{fjk}}^2 \right) \quad (107)$$

where  $\mathcal{C}$  and  $\mathcal{P}$  are the set of feature and preintegrated measurements, respectively, while  $\mathbf{e}_{marg}(\mathbf{x})$  is the residual of the marginal prior (see Equation (32)). We want to point out again that in practice we instead employ a (Huber or Cauchy) robust cost function on the last visual error term in Equation (107), while we here omit the detailed derivations of this standard treatment to keep presentation concise, we do have a similar treatment in our ensuing loosely-coupled direct VINS (see Equation (112)).

**5.1.1 Inverse-depth representation:** A well-known disadvantage of the above representation for features is that points at infinity are difficult to utilize. To mitigate this issue, we instead employ an *inverse-depth* representation (Civera, Davison and Montiel 2008). In particular, we represent a feature using the inverse coordinates in the camera frame where it was first observed. Denoting  $\{C_{a,i}\}$  the frame of reference of the “anchoring” camera, which is associated with the  $i$ -th camera frame and the anchoring time  $a$ , we have the following inverse-depth representation (see Mourikis and Roumeliotis (2007)):

$${}^{C_{a,i}} \mathbf{m}_f = \begin{bmatrix} \alpha \\ \beta \\ \rho \end{bmatrix} \Rightarrow {}^{C_{a,i}} \mathbf{p}_f = \frac{1}{\rho} \begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix} \quad (108)$$

where we also show the relationship between the inverse-depth representation of the feature  ${}^{C_{a,i}} \mathbf{m}_f$  and the corresponding 3D position in the anchor frame  ${}^{C_{a,i}} \mathbf{p}_f$ . The



**Figure 4.** Visualization of selected depth map pixels with a large intensity gradient (left). Keyframe pixels are projected onto the query frame as a result of the optimized direct alignment of the frame-to-frame relative transformation (right).

feature position in the  $j$ -th camera frame at time step  $k$  can be computed as follows:

$${}_{C_{k,j}}\mathbf{p}_f = {}_{C_{a,i}}\mathbf{R} \frac{1}{\rho} \begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix} + {}_{C_{k,j}}\mathbf{p}_{C_{a,i}} \quad (109)$$

$$\begin{aligned} {}_{C_{a,i}}\mathbf{R} &= {}_I^C\mathbf{R} {}_G^k\mathbf{R} {}_G^a\mathbf{R} {}_I^{C_i}\mathbf{R}^\top \\ {}_{C_{k,j}}\mathbf{p}_{C_{a,i}} &= {}_I^C\mathbf{R} {}_G^k\mathbf{R} ({}^G\mathbf{p}_a + {}_G^a\mathbf{R}^\top {}_I\mathbf{p}_{C_i} - {}^G\mathbf{p}_k) + {}_I^C\mathbf{p}_I \end{aligned}$$

Note that due to the projective geometry of the perspective projection (105),  $\Pi(\mathbf{x}) = \Pi(\rho\mathbf{x})$ , we can multiply both sides of Equation (109) by  $\rho$  and have the equivalent measurement model:

$$\mathbf{z}_{fjk} = \Pi(\mathbf{h}) + \mathbf{n}_f \quad (110)$$

$$\mathbf{h} = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} := \rho {}_{C_{k,j}}\mathbf{p}_f = {}_{C_{a,i}}\mathbf{R} \begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix} + \rho {}_{C_{k,j}}\mathbf{p}_{C_{a,i}} \quad (111)$$

The measurement Jacobians of this inverse-depth model can be found in Appendix C. Note that this measurement model is numerically stable and can handle points at infinity, thus allowing for the gain of feature direction information from these far-off feature points.

## 5.2 Loosely-Coupled Direct VINS

To further validate the proposed closed-form preintegration theory, in the following, by leveraging our prior work (Eckenhoff, Geneva and Huang 2017), we develop a *loosely-coupled* VINS algorithm based on direct image alignment and IMU preintegration. In particular, we estimate the relative frame-to-frame motion through direct alignment of image pixels. These relative-motion constraints then allow us to efficiently perform loop closure without explicitly detecting/tracking (or matching) features.

Consider the case where we wish to directly align a current frame  $C_2$  against a keyframe  $C_1$  (see Figure 4). Finding the optimal transformation can be formulated as an optimization problem over the total (warped) pixel intensity difference (i.e., photometric error):

$$\begin{aligned} {}_{C_1}^C\check{\mathbf{T}} &= \underset{{}_{C_1}^C\mathbf{T}}{\operatorname{argmin}} \sum_f \gamma \left( \frac{1}{\sigma_r^2} \underbrace{\left( I_{C_2} \left( {}_{C_1}^C\mathbf{T} {}_{C_1}\mathbf{p}_f \right) - I_{C_1} \left( {}_{C_1}\mathbf{p}_f \right) \right)}_{e_f}^2 \underbrace{\right)}_{v_f}^2 \quad (112) \end{aligned}$$

where  ${}_{C_1}^C\mathbf{T}$  is the transformation between the two camera frames parameterized by the relative quaternion  ${}_{C_1}^C\bar{q}$  and relative position  ${}_{C_1}^C\mathbf{p}_{C_2}$ , while  $I_{C_i}(\cdot)$  returns the intensity of a given point projected into the image frame, and  $\gamma(\cdot)$  is the Huber cost. The pixel's position in the keyframe,  ${}_{C_1}\mathbf{p}_f$  can be found via an online or stereo pair depth map computation. This position is treated as a noisy parameter in the residual allowing for computation of the residual sigma,  $\sigma_r$ , with the summation being over all pixels  $f$  with valid depth estimates and high gradients along the epipolar line. The Huber cost function  $\gamma(\cdot)$  with parameter  $k$  is defined as (Eade 2013):

$$\gamma(r) = \begin{cases} r, & \text{if } r < k^2 \\ 2k\sqrt{r} - k^2, & \text{otherwise} \end{cases} \quad (113)$$

The purpose of the Huber cost is to down-weight large residuals which occur naturally in image alignment due to occlusions, and has been used extensively in the literature (e.g., Engel, Schöps and Cremers (2014)).

Note that the covariance of each residual  $\sigma_r^2$  encodes the uncertainty due to errors in the intensity measurements as well as the disparity map:

$$\sigma_r^2 = 2\sigma_{int}^2 + \left( \frac{\partial e_f}{\partial d} \right)^2 \sigma_d^2 \quad (114)$$

where  $\sigma_{int}^2$  denotes the covariance of the intensity reading,  $\frac{\partial e_f}{\partial d}$  is the Jacobian of the residual  $e_f$  (112) with respect to the measured disparity  $d$ , and  $\sigma_d^2$  is the covariance of the disparity measurement. In the case of a depth map computed from a stereo pair as considered in this work, we define  $\mathbf{t}$  as the pixel coordinates,  $z$  as the pixel depth, and  $b$  as the baseline between the stereo pair. The Jacobian  $\frac{\partial e_f}{\partial d}$  can be calculated using the chain rule of differentiation as follows (see Equation (112)):

$$\begin{aligned} \frac{\partial e_f}{\partial d} &= \frac{\partial I_{C_2}}{\partial \mathbf{t}} \frac{\partial \mathbf{t}}{\partial {}_{C_2}\mathbf{p}_f} \frac{\partial {}_{C_2}\mathbf{p}_f}{\partial {}_{C_1}\mathbf{p}_f} \frac{\partial {}_{C_1}\mathbf{p}_f}{\partial z} \frac{\partial z}{\partial d} \\ &= [I_{C_{2x}} \quad I_{C_{2y}}] \begin{bmatrix} \frac{f_x}{{}_{C_2}\mathbf{p}_{f_j}(3)} & 0 & -\frac{f_x {}_{C_2}\mathbf{p}_{f_j}(1)}{{}_{C_2}\mathbf{p}_{f_j}(3)^2} \\ 0 & \frac{f_y}{{}_{C_2}\mathbf{p}_{f_j}(3)} & -\frac{f_y {}_{C_2}\mathbf{p}_{f_j}(2)}{{}_{C_2}\mathbf{p}_{f_j}(3)^2} \end{bmatrix} \\ &\quad \times \frac{{}_{C_2}\mathbf{R}}{z} \frac{-f_x b}{d^2} \end{aligned} \quad (115)$$

where  $I_{C_{2x}}$  and  $I_{C_{2y}}$  are the image gradients in the  $x$  and  $y$  directions respectively, while  $f_x$  and  $f_y$  are the focal lengths of the camera.

The covariance of the pixel disparity,  $\sigma_d^2$ , is obtained based on the observation that this disparity is the maximum likelihood estimate for a single measurement graph, with the residual being the difference in intensity between the pixel in the left,  $I_{C_1L}$ , and right,  $I_{C_1R}$ , images in the keyframe stereo pair, which can be formulated as follows:

$$\check{d} = \underset{d}{\operatorname{argmin}} \frac{1}{\sigma_{rd}^2} \left( \underbrace{I_{C_1L}(v, u) - I_{C_1R}(v, u - d)}_{e_d} \right)^2 \quad (116)$$

where the covariance associated with this residual can be found as  $\sigma_{rd}^2 = 2\sigma_{int}^2$ , and comes from uncertainty in the intensity readings. The covariance on our disparity estimate

can then be approximated as:

$$\sigma_d^2 = \left( \frac{\partial e_d}{\partial d} \frac{1}{\sigma_{rd}^2} \right)^{-1} = \sigma_{rd}^2 \left( \frac{1}{I_{C_1 R_x}} \right)^2 \quad (117)$$

where  $I_{C_1 R_x}$  is the  $x$ -gradient of the pixel in the right image which is selected as the match.

Once we have determined the photometric error covariance  $\sigma_r^2$ , we now solve the direct alignment problem (112) using the Levenberg-Marquadt method. In particular, at each iteration we solve the following normal equation:

$$\begin{aligned} & \left( \left( \sum w_f \mathbf{J}_f^\top \mathbf{J}_f \right) + \lambda \text{diag} \left( \sum w_f \mathbf{J}_f^\top \mathbf{J}_f \right) \right) \delta \overset{C_2}{\check{\mathbf{T}}} \\ &= - \sum w_f \mathbf{J}_f^\top e_f \left( \overset{C_2}{\check{\mathbf{T}}} \right) \end{aligned} \quad (118)$$

where  $\lambda$  is the dampening parameter, and  $e_f(\overset{C_2}{\check{\mathbf{T}}})$  is the residual due to the  $f$ -th pixel in the alignment, evaluated at the current estimate (linearization point) for the relative transformation,  $\overset{C_2}{\check{\mathbf{T}}}$ . The weight  $w_f$  is computed at each iteration as follows:

$$w_f = \frac{\partial \gamma(v_f)}{\partial v_f} \frac{1}{\sigma_f^2} \quad (119)$$

$$\frac{\partial \gamma(v_f)}{\partial v_f} = \begin{cases} 1, & \text{if } v_f < k^2 \\ \frac{k}{\sqrt{v_f}}, & \text{otherwise} \end{cases} \quad (120)$$

where  $v_f$  is the raw cost fed into the Huber norm (see Equation (112)), and  $k$  is a design parameter. The Jacobian matrix  $\mathbf{J}_f$  is of the direct alignment measurement residual with respect to the state, computed as:

$$\begin{aligned} \mathbf{J}_f = & \begin{bmatrix} I_{C_2 x} & I_{C_2 y} \end{bmatrix} \begin{bmatrix} \frac{f_x}{\overset{C_2}{\mathbf{p}_f}(3)} & 0 & -\frac{f_x \overset{C_2}{\mathbf{p}_f}(1)}{(\overset{C_2}{\mathbf{p}_f}(3))^2} \\ 0 & \frac{f_y}{\overset{C_2}{\mathbf{p}_f}(3)} & -\frac{f_y \overset{C_2}{\mathbf{p}_f}(2)}{(\overset{C_2}{\mathbf{p}_f}(3))^2} \end{bmatrix} \\ & \times \begin{bmatrix} \overset{C_2}{\mathbf{p}_f} & -\overset{C_2}{\mathbf{R}} \end{bmatrix} \end{aligned} \quad (121)$$

After optimization, we will be left with a Gaussian distribution on our estimated relative *camera* pose. We can then transform this into a distribution on the relative *IMU* pose (denoted  $k$  and  $j$  for the keyframe and query frame IMU states respectively) using covariance propagation:

$$\overset{C_2}{\mathbf{T}} = \overset{C_2}{\check{\mathbf{T}}} \boxplus \overset{C_2}{\delta \mathbf{T}}, \text{ where } \overset{C_2}{\delta \mathbf{T}} \sim \mathcal{N}(\mathbf{0}, \Sigma_c) \quad (122)$$

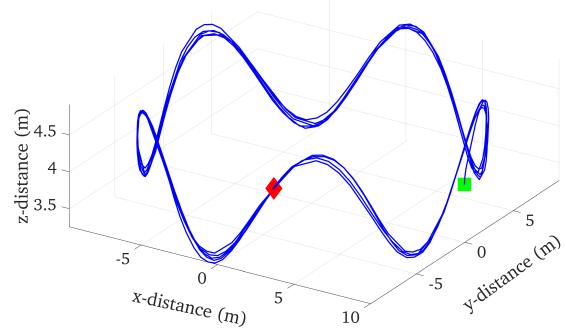
$$\overset{j}{\mathbf{T}} = \overset{j}{\check{\mathbf{T}}} \boxplus \overset{j}{\delta \mathbf{T}}, \text{ where } \overset{j}{\delta \mathbf{T}} \sim \mathcal{N}(\mathbf{0}, \Sigma_i) \quad (123)$$

$$\Sigma_i = \frac{\partial \overset{j}{\delta \mathbf{T}}}{\partial \overset{C_2}{\delta \mathbf{T}}} \Sigma_c \frac{\partial \overset{j}{\delta \mathbf{T}}}{\partial \overset{C_2}{\delta \mathbf{T}}}^\top \quad (124)$$

where  $\Sigma_c = (\sum w_f \mathbf{J}_f^\top \mathbf{J}_f)^{-1}$  is the covariance of the zero-mean alignment error. From this, the relative pose measurement that connects the IMU keyframe and query frame has the following residual:

$$\mathbf{e}_d(\mathbf{x}) = \begin{bmatrix} 2 \text{vec} \left( \overset{j}{G} \bar{\mathbf{q}} \otimes \overset{k}{G} \bar{\mathbf{q}}^{-1} \otimes \overset{j}{k} \bar{\mathbf{q}}^{-1} \right) \\ \overset{k}{G} \mathbf{R} \left( \overset{G}{\mathbf{p}_j} - \overset{G}{\mathbf{p}_k} \right) - \overset{k}{\check{\mathbf{p}}_j} \end{bmatrix} \quad (125)$$

whose Jacobians with respect to the state are provided in Appendix D, which will be used during graph optimization.



**Figure 5.** The ground truth trajectory of a MAV flying in a circle sinusoidal path generated in the Gazebo simulator. The total trajectory length is 307 meters with an average velocity of 6.13 m/s. Start and end positions are denoted with a green square and red diamond, respectively.

Using this visual measurement residual, along with the preintegrated IMU measurements, we have the following optimization problem for the loosely-coupled direct VINS, which can be solved analogously as in Equation (107):

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\text{argmin}} \sum_{d \in \mathcal{D}} \|\mathbf{e}_d(\mathbf{x})\|_{\Sigma_i^{-1}}^2 + \sum_{p \in \mathcal{P}} \|\mathbf{e}_{IMU}(\mathbf{x})\|_{\mathbf{P}_p^{-1}}^2 \quad (126)$$

where  $\mathcal{D}$  and  $\mathcal{P}$  are the set of direct alignment relative pose and preintegrated measurements, respectively. Note that this direct image alignment allows for computationally efficient incorporation of large-scale loop closures due to the direct compression of intensity residuals into a single informative relative motion measurement.

## 6 Monte-Carlo Simulation Analysis

To validate the proposed closed-form preintegration theory, we first perform extensive Monte-Carlo simulations in various conditions in terms of sampling rates and motion dynamics. In particular, to better model the motion dynamics of a physical system, we leverage the open-source Gazebo simulator of a micro air vehicle (MAV) (Koenig and Howard 2004) which allows for direct realistic simulation and collection of true inertial and pose data (constrained by the physical MAV motion). The simulated datasets were generated as follows: (i) the MAV was commanded to follow a series of waypoints after takeoff, (ii) the ground truth of 100 Hz inertial and pose information was recorded, (iii) 80 synthetic stereo visual feature measurements (uv-coordinates) were created for each camera frame using the true pose information at a static 10Hz frequency. Following the commonly-used IMU model (Trawny and Roumeliotis 2005), the true inertial measurements were corrupted with an additive discrete bias and white noise using the noise parameters from the VI-Sensor (Nikolic, Rehder, Burri, Gohl, Leutenegger, Furgale and Siegwart 2014), while the features' uv-coordinates were corrupted with an additive white noise to each axis with one pixel standard deviation.

In our tests, we used the popular GTSAM (Dellaert 2012) framework to construct, optimize, and marginalize our graph using the included fixed-lag smoother. To ensure a fair comparison, we evaluate our preintegration methods

**Table 1.** Analysis of the effect of different IMU frequencies on estimation accuracy. Note that each frequency run has a slightly different trajectory, and thus only the relative spread within a given frequency should be considered.

	<b>MODEL-1</b>		<b>MODEL-2</b>		<b>DISCRETE</b>	
Units	m	deg	m	deg	m	deg
100 Hz	0.096	0.327	<b>0.093</b>	<b>0.300</b>	0.107	0.328
200 Hz	0.051	0.204	<b>0.049</b>	<b>0.179</b>	0.058	0.207
400 Hz	0.033	0.107	<b>0.033</b>	<b>0.101</b>	0.035	0.109
800 Hz	<b>0.030</b>	0.085	0.030	<b>0.085</b>	0.031	0.086

against the state-of-art **discrete** preintegration (Forster, Carbone, Dellaert and Scaramuzza 2015, 2017), by using the on-manifold preintegrator class within the GTSAM implementation to compute the required measurement means, bias Jacobians, and covariances. We constructed all graphs side by side, ensuring that the measurements inserted are exactly the same, and thus fair to all methods. For simplicity we used the tightly-coupled indirect features in the graph, in which features are automatically marginalized out after three seconds (that is, no map was created for loop closures, and thus, the system is a VIO system). Note also that we initialized all systems to the ground-truth pose, and with zero bias. Figure 5 shows one example of the true simulated trajectory generated using Gazebo in our simulations.

The Monte-Carlo simulation comparison results of root mean squared error (RMSE) averaged over 50 runs are shown in Figure 6. Evidently, the proposed preintegration using piecewise constant local acceleration model (i.e., **Model 2**) is slightly better with the RMSE (averaged over all time steps and all runs) of 0.093 meters and 0.300 degrees than that using the piecewise constant measurement model (i.e., **Model 1**) with the RMSE of 0.096 meters and 0.327 degrees. More importantly, both methods are shown to outperform the discrete state-of-the-art method (Forster, Carbone, Dellaert and Scaramuzza 2015, 2017), which has the RMSE of 0.107 meters and 0.328 degrees. It is important to point out that the superior performance (though by a small margin in this MAV test, with larger improvement margins expected for higher dynamics not constrained by MAV motion) endowed by the proposed closed-form preintegration using the new inertial models over the discrete one does not incur extra computational overhead during graph-based VINS optimization.

Furthermore, we investigate the effect that the IMU frequency has on the relative performance of the preintegration methods under consideration. Using the same simulation setup, the MAV was commanded to follow the trajectory with different Gazebo simulation frequencies. It is important to note that since we are using a physical simulation, the true trajectory will vary from frequency to frequency since the controller will perform differently, however, this is acceptable since we are looking at the relative performance within a given frequency. Table 1 shows the averaged RMSE results of different IMU frequencies. It can be seen that the proposed closed-form preintegration methods have greater impact when the frequency of the IMU is lower (i.e., significantly better performance); while at higher frequencies,

the above methods become less distinguishable. This implies that the proposed closed-form preintegration methods are better suited for applications with limited IMU frequency, which is often the case for low-cost MEMS sensors.

## 7 Real-World Experimental Validations

### 7.1 Tightly-Coupled Indirect VIO

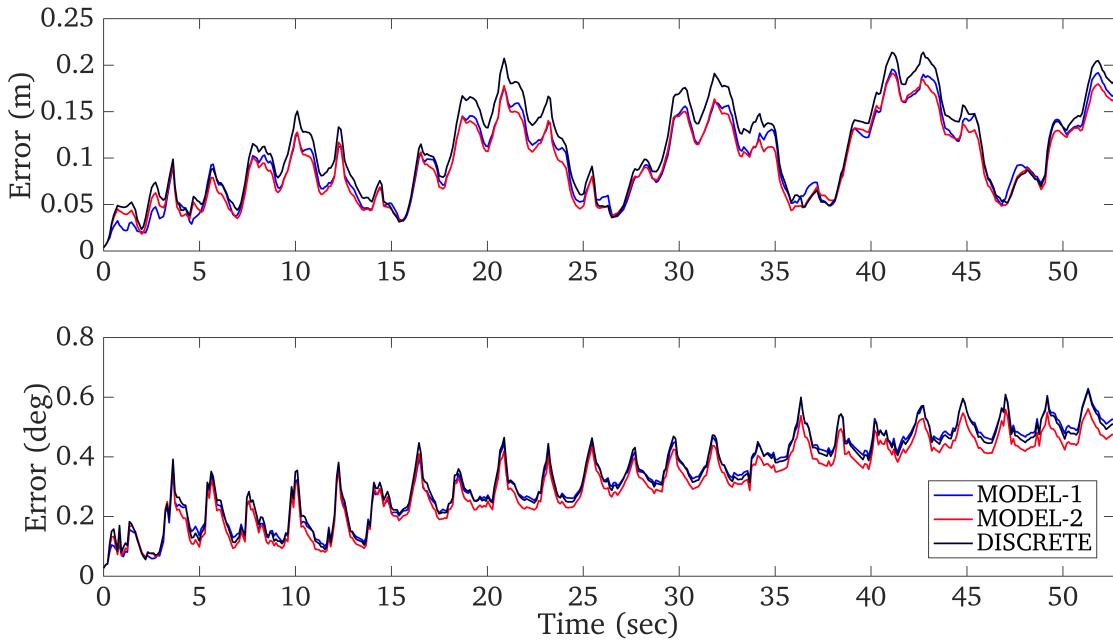
In our tightly-coupled VIO system, we use stereo vision due to its superior estimation performance as compared to monocular systems (Paul, Wu, Hesch, Nerurkar and Roumeliotis 2017). Stereo correspondences allow for accurate triangulation of features regardless of vehicle motion, making them robust to maneuvers troubling monocular systems such as hovering. In addition, stereo allows for a direct reading of the scale, which is highly informative to the estimator.

When a pair of stereo images arrive, we perform KLT tracking (Baker and Matthews 2004) of FAST (Rosten, Porter and Drummond 2010) features that have been extracted in an uniform grid over the image. Stereo correspondence information is known by initializing new features in the left image and KLT tracking them into the right. The set of stereo tracks from the current image is then tracked temporally forward at each future time step, while also ensuring to initialize new feature tracks if the number of active tracks falls under our desired active feature threshold. To reject outliers we perform 8-point RANSAC between both the temporal and stereo left-to-right matches. We have found that this frontend provides a good balance between track longevity, computational speed, and accuracy. If an active feature is successfully tracked, the normalized image coordinates are added as measurements associated with that feature. To robustify our system to outliers, we utilized the Cauchy loss function for all image measurements.

Inspired by Leutenegger, Lynen, Bosse, Siegwart and Furgale (2015), we maintain a sliding window of IMU states in the estimator that consists of two sub-windows. The first, denoted as the inertial window, contains the full 15 DOF IMU state and refers to the most recent imaging times. The second window, called the pose window, contains a set of pose-only clones (that is, only the orientation and position are maintained). At every imaging time we create a new corresponding IMU node. The IMU readings collected over the interval are preintegrated to both predict the new state and to form a preintegrated IMU measurement between the previous and new state.

After tracking, we formulate the sliding-window batch optimization (i.e., BA) problem using all features with a sufficient number of tracks as well as all nodes in the inertial and pose windows. The measurements contained in this graph are: (i) the prior, (ii) the visual measurements for the active features, and (iii) the preintegration factors between the inertial window states (see Equation (107)). We use the Ceres Solver with an elimination ordering that takes advantage of the sparsity of the problem through the Schur Complement (Agarwal, Mierle and Others 2018; Kummerle, Grisetti, Strasdat, Konolige and Burgard 2011).

If the inertial window has reached its maximum length, we flag the oldest state's velocity and biases for marginalization. If the pose window also reaches its maximum length, we



**Figure 6.** Monte-Carlo simulation results averaged over 50 runs: (top) position RMSE, and (bottom) orientation RMSE. In this test, physically-realistic synthetic data was generated using a Gazebo MAV simulator. It is clear that the proposed closed-form preintegration outperforms the state-of-the-art discrete approach (Forster, Carbone, Dellaert and Scaramuzza 2015, 2017).

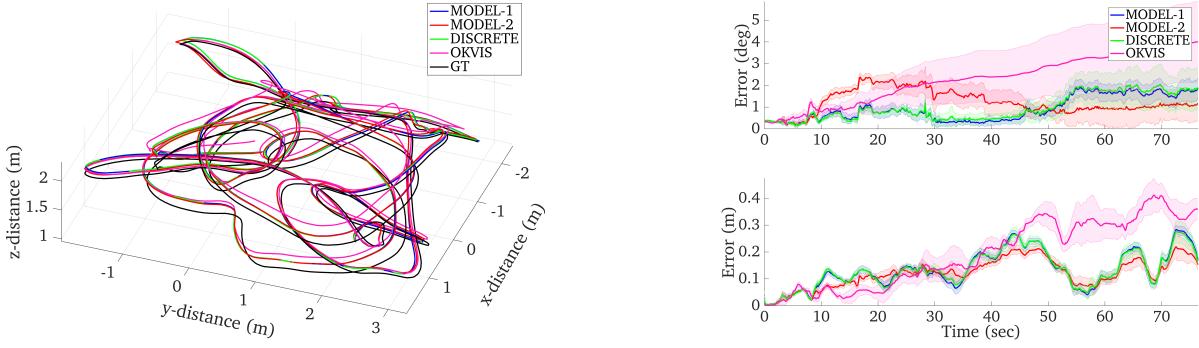
add both the oldest pose and all features it has seen into the marginal state list. Performing marginalization yields a new prior factor that has absorbed the old prior, the marginalized feature measurements, and the oldest preintegration factor. The oldest IMU state whose velocity and biases have been marginalized is then moved into the pose window.

**7.1.1 EuRoC MAV dataset:** We compared our tightly-coupled indirect VIO system with a state-of-the-art open-source VINS – that is, the Open Keyframe-based Visual-Inertial SLAM (OKVIS) (Leutenegger, Lynen, Bosse, Siegwart and Furgale 2015), although several different VINS methods were recently introduced (e.g., Bloesch, Burri, Omari, Hutter and Siegwart (2017)). We performed this comparison on the EuRoC MAV dataset (Burri, Nikolic, Gohl, Schneider, Rehder, Omari, Achterlik and Siegwart 2016), which has become the standard method for evaluating VINS algorithms and provides 20hz stereo pairs with a 200hz MEMS ADIS16448 IMU. Our tightly-coupled preintegration-based system was run with inertial and pose sliding windows of 6 and 8 with a maximum of 300 extracted features. Stereo-OKVIS was run with 4 and 6 inertial and keyframes with 300 features. These parameters were selected to ensure real-time performance with both systems having minimal dropped frames. It should be noted that depending on the tuning parameters used in the VINS algorithms, their performance may vary (e.g., see Delmerico and Scaramuzza (2018)). Note also that our VIO system uses a sliding window of poses as well as the inertial window connected with preintegrated measurements but does *not* keep any kind of map (to allow intra-window loop closures), while OKVIS employs a set of keyframes where mapped points are maintained. Nevertheless, to provide a direct comparison, we initialize both systems with the true orientation, biases, velocity, and position such that no post-processing yaw alignment is needed. Note also that due to some randomness

that may occur during the visual tracking frontend (e.g., RANSAC-based outlier rejection), variations in the VINS results can be observed even if running the same algorithm on the same sequences. To limit this variability of the algorithm, we perform 10 runs on the real-world sequences and average the results.

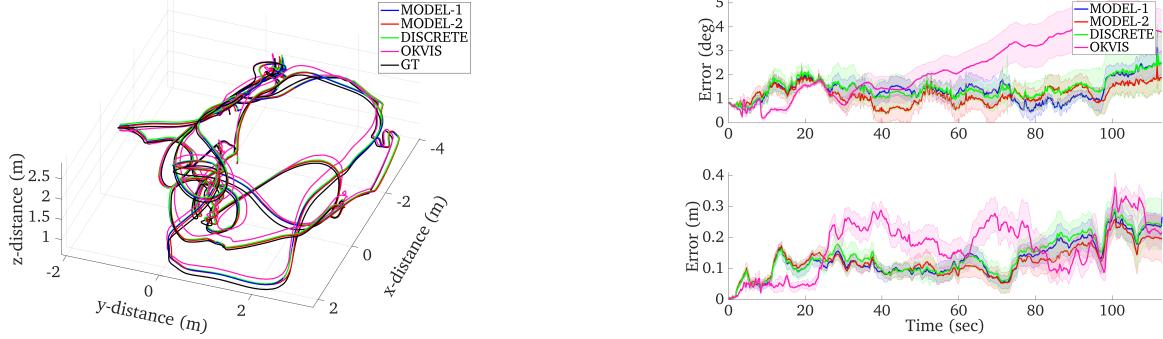
The “V1\_02\_med” and “V2\_02\_med” average trajectories can be seen in Figures 7a and 7c where we plot the estimated trajectories of our VIO and OKVIS along with the ground truth. Note that it is understood that the performance of VINS algorithms may vary even if re-running on the same sequences due to some randomness in visual tracking (e.g., RANSAC); and thus, we repeated the test for 10 times and averaged the results in order to better evaluate the relative performance of the compared approaches. Figures 7b and 7d show the averaged RMSE results of our VIO algorithm based on the proposed closed-form preintegration with the two models as compared to OKVIS, which were computed at every time step and then averaged over all the runs; while the averaged RMSE results are shown in Table 2. To show the repeatability/variability between runs, we also plot the standard deviation of the runs, noting that this should not be confused with the estimator uncertainty bounds commonly found in the literature.

We additionally evaluated the trajectories of the proposed models using the odometry error metric (Zhang and Scaramuzza 2018). As compared to the absolute RMSE value, this metric splits the trajectory into small segments of predetermined lengths, aligns the start of each segment to the ground truth, and then computes the error of the ending pose of the segment in respect to the ground truth. This allows for insights of how drift is a function of distance. Following the method proposed by Zhang and Scaramuzza (2018), each of the ten runs performed by each model on the EuRoC MAV sequences were evaluated and the total odometric error over all sequences was computed. Figure 8 and Table 3,



(a) Average trajectory estimates for "V1\_02\_med".

(b) Average position and orientation RMSE for "V1\_02\_med".



(c) Average trajectory estimates for "V2\_02\_med".

(d) Average position and orientation RMSE for "V2\_02\_med".

**Figure 7.** Average trajectory and RMSE error over ten runs for the “V1\_02.med” (top) and “V2\_02.med” (bottom) sequences of the proposed tightly-coupled indirect VIO system. The one-sigma bound on the mean error is also shown and can be interpreted as the repeatability of the system (due to some randomness occurred in visual tracking). Note that this is not the same as estimator uncertainty and instead shows the variance of the VIO systems. The total trajectory lengths are 80 and 88 meters, respectively.

**Table 2.** Average absolute RMSE results of the tightly-coupled indirect VINS for the EuRoC MAV sequences averaged over 10 runs. All systems were initialized with the ground truth state. The smallest position and orientation errors have been highlighted.

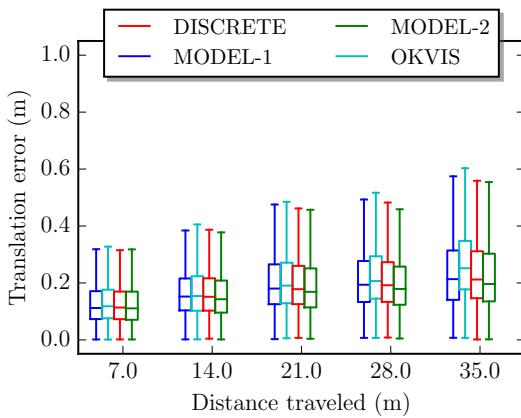
	MODEL-1		MODEL-2		DISCRETE		OKVIS	
Units	m	deg	m	deg	m	deg	m	deg
V1_01_easy	0.2522	2.749	<b>0.2160</b>	2.503	0.2547	2.781	0.2356	<b>2.458</b>
V1_02_med	0.1342	<b>0.942</b>	<b>0.1214</b>	1.215	0.1344	1.001	0.1996	2.321
V1_03_diff	0.1101	0.880	<b>0.0953</b>	<b>0.809</b>	0.1012	0.830	0.1830	3.498
V2_01_easy	0.1429	1.069	<b>0.1426</b>	1.148	<b>0.1426</b>	1.118	0.1806	<b>0.973</b>
V2_02_med	0.1297	1.390	<b>0.1223</b>	<b>1.135</b>	0.1375	1.450	0.1695	2.334
V2_03_diff	0.2982	2.159	<b>0.2800</b>	<b>1.769</b>	0.3055	2.052	0.3483	8.327
MH_01_easy	0.1817	1.398	<b>0.1653</b>	1.761	0.2050	1.321	0.2523	<b>0.728</b>
MH_02_easy	0.1533	0.691	<b>0.1498</b>	<b>0.525</b>	0.1564	0.599	0.2523	0.728
MH_03_med	0.2993	1.024	<b>0.2627</b>	0.968	0.2800	<b>0.840</b>	0.3193	1.903
MH_04_diff	0.3312	<b>0.849</b>	0.3515	0.974	0.3488	0.852	<b>0.2145</b>	1.022
MH_05_diff	0.3939	<b>0.692</b>	0.3971	0.715	<b>0.3835</b>	0.809	0.5432	0.738

show the resulting odometric error for trajectory segments of  $\{7, 14, 21, 28, 35\}$  meters.

These results clearly demonstrate that our VIO system can offer competitive performance to OKVIS; that is, we see instances where our method outperforms OKVIS, while in others OKVIS is superior. Between the two proposed preintegration models, for these experiments, Model 2 offers the best performance. We note that the discrete preintegration method tends to perform with lower accuracy compared to

the two proposed models (although not in all cases), thereby validating the proposed preintegration models.

**7.1.2 UD indoor datasets:** We further performed relatively large-scale (as compared to the EuRoC MAV dataset) indoor experiments in two buildings at the University of Delaware (UD) using our hand-held VI-Sensor with an IMU frequency of 400 Hz. In these experiments, because no ground truth was available, we initialized the system by keeping the device stationary for a short period of time (e.g., 2 seconds) so that the initial orientation and biases could be



**Figure 8.** Boxplot of the odometric translation error statistics for the tightly-coupled indirect system evaluated over all of the EuRoC MAV sequences. Errors were computed using the odometry metric over trajectory segments of {7,14,21,28,35} meters in length. The middle box spans the first and third quartiles, while the whiskers are the upper and lower limits.

**Table 3.** Mean odometric translation errors for the tightly-coupled indirect system evaluated over all of the EuRoC MAV sequences. Errors were evaluated over trajectory segments of {7,14,21,28,35} meters in length. All errors are in meters.

	MODEL-1	MODEL-2	DISCRETE	OKVIS
7 m	0.137	<b>0.135</b>	0.136	0.142
14 m	0.177	<b>0.169</b>	0.177	0.185
21 m	0.220	<b>0.209</b>	0.217	0.226
28 m	0.229	<b>0.216</b>	0.226	0.245
35 m	0.247	<b>0.238</b>	0.246	0.287

found, while the position and velocities were initialized as zero. To account for poor calibration of the sensor suite, both the IMU-to-camera spatial calibration parameters as well as the camera *intrinsics* were estimated online. This was done by adding these quantities into the state and using the *raw* image coordinates as measurements, while expressing these as a function of the normalized pixel coordinates (104) as well as the camera intrinsics (Li, Yu, Zheng and Mourikis 2014). The first indoor experiment was performed in the UD Gore Hall, in which the trajectory starts on the first floor, traverses the staircase to the third floor, and returns to the bottom floor, making a loop on each level. To evaluate the estimation performance the trajectory returns to the original starting location. The 3D trajectory estimate is shown in Figure 9a while its projection onto the building floor plan is shown in Figure 9b. We ran each preintegration model ten times across the 228 meter long dataset and averaged the results. Model 1 had an ending error of 0.763 m (0.33% of the path), Model 2 had an ending error 0.747 m (0.33%), discrete preintegration achieved 0.765 (0.34%), and OKVIS achieved an ending error of 0.762 m (0.33%) showing the improvement due to closed-form preintegration.

The second indoor experiment was conducted in the UD Smith Hall. Starting on the second floor, we traversed along a rectangular wall before descending the stairs, looping around the first floor, then returning up the stairs, looping one and a half times around the upper level before returning to the

starting position. Model 1 had an ending error of 0.632 m (0.28% of the path), Model 2 had an ending error 0.788 m (0.35%), discrete preintegration achieved 0.768 m (0.34%), and OKVIS achieved 1.699 m (0.75%) over the 230 meter trajectory. Note that this scenario was more challenging than the first experiment, primarily due to the fact that during this test, there were people walking around, lighting conditions were varying, and some parts of the environment lacked good features to detect and track (see Figure 10a).

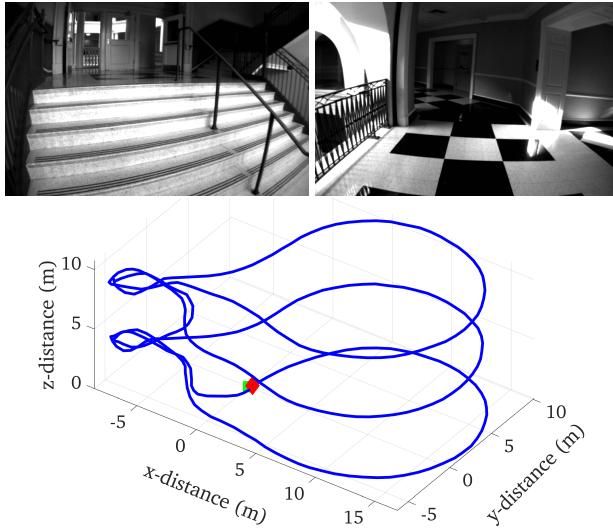
The 3D trajectory estimate and its projection onto the floor plan are show in Figures 10b and 10c, respectively. These results clearly demonstrate that our VIO systems using the proposed closed-form preintegration are able to perform accurate 3D motion tracking in relatively large-scale complex environments.

## 7.2 Loosely-Coupled Direct VINS

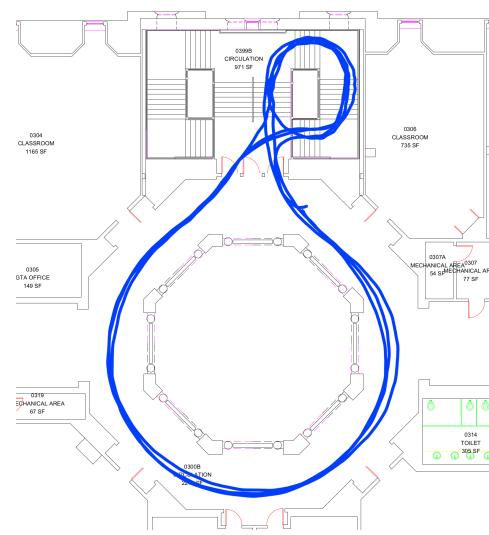
When a stereo pair arrives, as in the preceding indirect VIO, we perform the proposed closed-form preintegration from the previous IMU state to the current state. We then check the list of stored keyframes for a suitable candidate for direct image alignment, based on a field-of-view constraint between the candidate and the current image. If no such acceptable candidate is found, a new keyframe is created from the previous image pair and its depth map is computed using the OpenCV function StereoSGBM. In particular, in order to perform course-to-fine alignment, the depth map is computed for multiple image pyramid levels. Starting at the coarsest image level, we perform iterative image alignment, using the larger levels to further refine the coarse image alignment transform. This image alignment optimization was implemented in a CUDA kernel for GPU acceleration, thus allowing for the system to achieve real-time performance. After convergence, we recover the relative-pose constraint and add it as a factor to our direct-VINS graph. Note that as compared to our indirect VIO method, we do not perform marginalization and thus allow later incorporation of loop closures. To handle this increase of computational complexity and allow for real-time performance, we leverage the iSAM2 incremental smoothing implementation within the GTSAM framework (Kaess, Johannsson, Roberts, Ila, Leonard and Dellaert 2012; Dellaert 2012). However, the proposed framework is by no means optimal and can be further refined, for example, by more intelligently selecting keyframes.

**7.2.1 EurocMav dataset:** To validate our direct VINS approach, we perform tests on the same EurocMav sequences as before, which allow for direct comparison to a ground-truth trajectory (Burri, Nikolic, Gohl, Schneider, Rehder, Omari, Achtelik and Siegwart 2016). The results of the proposed direct VINS using two different preintegration models are shown in Table 4. Clearly, in scenarios in which a large amount of loop closures are present (e.g., “V1\_03\_diff”), this system can outperform the tightly-coupled VIO system (see Section 5.1 and Table 2). However, when such loop closures are not available, the loosely-coupled systems suffer from larger drifts, as can be seen from the result of “V2\_01\_easy”.

In these experiments, both proposed models tended to offer improved performance as compared to discrete



(a) Example images (top) and 3D trajectory (bottom).



(b) Projection of the estimated trajectory onto the floor plan.

**Figure 9.** The trajectory estimates of the indoor experiment performed in the UD Gore Hall. Two example images from the dataset can be seen in (a), while the starting and ending locations are shown by a green square and red diamond in the plot, respectively. Note that the three floors have similar layouts, and thus only one floor plan is shown in plot (b).

**Table 4.** Average absolute RMSE results for the EurocMav sequences over ten runs using the proposed direct VINS algorithm. All systems were initialized with the ground truth state. The smallest position and orientation errors have been highlighted.

Units	MODEL-1		MODEL-2		DISCRETE	
	m	deg	m	deg	m	deg
V1_01_easy	<b>0.2445</b>	<b>2.218</b>	0.2482	2.246	0.2530	2.223
V1_02_med	0.1598	1.767	<b>0.1309</b>	<b>1.483</b>	0.1763	1.899
V1_03_diff	<b>0.0990</b>	<b>1.180</b>	0.1030	1.279	0.1030	1.234
V2_01_easy	<b>0.1627</b>	2.089	0.1940	1.956	0.1664	<b>1.533</b>
V2_02_med	0.1809	2.530	<b>0.1665</b>	2.527	0.1688	<b>2.309</b>
V2_03_diff	0.9337	6.187	<b>0.8927</b>	5.425	1.0137	<b>4.998</b>
MH_01_easy	<b>0.2947</b>	2.270	0.3277	<b>2.148</b>	0.3217	2.226
MH_02_easy	<b>0.1882</b>	1.650	0.2136	1.582	0.2008	<b>1.483</b>
MH_03_med	0.2330	2.096	0.2295	2.121	<b>0.2288</b>	<b>2.092</b>
MH_04_diff	0.4792	<b>2.513</b>	0.4867	2.627	<b>0.4724</b>	2.562
MH_05_diff	<b>0.2884</b>	<b>1.664</b>	0.3014	1.722	0.2946	1.700

preintegration (although not in all cases), while providing similar levels of performance to each other, with each having trajectories where they outperform the other. In addition, the proposed direct VINS is sensitive to the tuning parameters, which in this experiment were chosen as *identical* across all sequences, rather than finding an optimal set per scenario. This led to situations such as “V2\_03\_diff”, in which some of the runs yield incorrect loop closures despite our system attempting to reject these, greatly corrupting the resulting trajectory estimates. However, as the purpose of this work is to show the accuracy of the proposed preintegration, instead of the robustness of the utilized front-ends, these results along with the previous simulation results strongly suggest that our preintegration models can be, and should be, used when designing graph-based VINS.

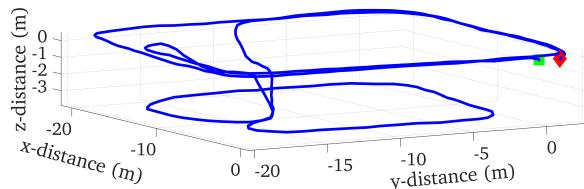
## 8 Conclusions and Future Work

In this paper, we have analytically derived closed-form inertial preintegration and successfully applied it to graph-based

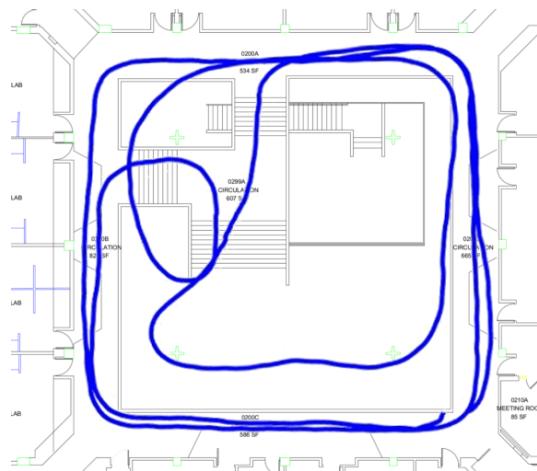
visual-inertial navigation systems (VINS). In particular, we advocate two new preintegration models for the evolution of IMU measurements across sampling intervals. In the first, we assume that the inertial measurements remain piecewise constant; while in the second, we incorporate a piecewise constant *local* acceleration model into the preintegration framework. We have validated through extensive Monte-Carlo simulations that both models outperform the state-of-the-art *discrete* preintegration. Furthermore, we have utilized this closed-form preintegration theory and developed two different VINS algorithms primarily to show the advantages of the proposed preintegration. In the first, we formulated an indirect (feature-based), tightly-coupled, sliding-window optimization based VIO system that offers competitive (if not better) performance to a state-of-the-art graph-based VINS algorithm. The second VINS method was developed instead based on loosely-coupled direct image alignment with the proposed preintegrations, allowing for efficient incorporation of informative loop closures.



(a) Two sample images seen during the experiment.



(b) View of the estimated 3D trajectory.



(c) Projection of the estimated trajectory onto the floor plan.

**Figure 10.** The results of the indoor experiment performed in the UD Smith Hall. Two example left camera images are shown in (a). In plot (b), the starting and ending locations of the trajectory estimates are shown as a green square and red diamond, respectively. Note that the trajectory shown in (c) occurs on both the second and first floors.

In the future, we will integrate the proposed closed-form preintegrations to aided inertial navigation systems with other aiding sources (e.g., LiDAR). We also seek to further robustify our VINS to handle more challenging scenarios (e.g., ultra-fast motion and highly-dynamic scenes).

## Declaration of Conflicting Interests

The authors declare that there is no conflict of interest.

## Funding

This work was partially supported by the University of Delaware College of Engineering, the NSF [grant number IIS-1566129], the DTRA [grant number HDTRA1-16-1-0039], and Google Daydream.

## References

- Agarwal S, Mierle K and Others (2018) Ceres solver. <http://ceres-solver.org>.
- Baker S and Matthews I (2004) Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision* 56: 221–255.
- Bloesch M, Burri M, Omari S, Hutter M and Siegwart R (2017) Iterated extended kalman filter based visual-inertial odometry using direct photometric feedback. *The International Journal of Robotics Research* 36(10): 1053–1072.
- Bloesch M, Omari S, Hutter M and Siegwart R (2015) Robust visual inertial odometry using a direct ekf-based approach. In: *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 298–304.
- Burri M, Nikolic J, Gohl P, Schneider T, Rehder J, Omari S, Achtelik MW and Siegwart R (2016) The euroc micro aerial vehicle datasets. *The International Journal of Robotics Research* 35(10): 1157–1163.
- Cadena C, Carlone L, Carrillo H, Latif Y, Scaramuzza D, Neira J, Reid ID and Leonard JJ (2016) Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics* 32(6): 1309–1332.
- Caruso D, Engel J and Cremers D (2015) Large-scale direct slam for omnidirectional cameras. In: *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 141–148.
- Chatfield A (1997) *Fundamentals of High Accuracy Inertial Navigation*. Reston, VA: American Institute of Aeronautics and Astronautics, Inc.
- Chirikjian GS (2011) *Stochastic Models, Information Theory, and Lie Groups, Volume 2: Analytic Methods and Modern Applications*, volume 2. Springer Science & Business Media.
- Civera J, Davison A and Montiel J (2008) Inverse depth parametrization for monocular SLAM. *IEEE Transactions on Robotics* 24(5): 932–945.
- Dellaert F (2012) Factor graphs and gtsam: A hands-on introduction. Technical report, Georgia Institute of Technology.
- Delmerico J and Scaramuzza D (2018) A benchmark comparison of monocular visual-inertial odometry algorithms for flying robots. In: *Proc. of the IEEE International Conference on Robotics and Automation*. Brisbane, Australia.
- Eade E (2013) Gauss-newton / levenberg-marquardt optimization. *Technical Report*.
- Eckenhoff K, Geneva P and Huang G (2016a) High-accuracy preintegration for visual-inertial navigation. In: *Proc. of the International Workshop on the Algorithmic Foundations of Robotics*. San Francisco, CA.
- Eckenhoff K, Geneva P and Huang G (2017) Direct visual-inertial navigation with analytical preintegration. In: *Proc. of the IEEE International Conference on Robotics and Automation*. Singapore.
- Eckenhoff K, Geneva P and Huang G (2018) Continuous preintegration theory for visual-inertial navigation. Technical Report RPNG-2018-CPI, University of Delaware. Available: [http://udel.edu/~ghuang/papers/tr\\_cpi.pdf](http://udel.edu/~ghuang/papers/tr_cpi.pdf).
- Eckenhoff K, Paull L and Huang G (2016b) Decoupled, consistent node removal and edge sparsification for graph-based SLAM. In: *Proc. IEEE/RSJ International Conference on Intelligent*

- Robots and Systems*. Daejeon, Korea, pp. 3275–3282.
- Ellekilde L (2007) Dense 3D Map Construction for Indoor Search and Rescue. *Journal of Field Robotics* 24(1-2): 71.
- Engel J, Koltun V and Cremers D (2018) Direct sparse odometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40(3): 611–625.
- Engel J, Schöps T and Cremers D (2014) Lsd-slam: Large-scale direct monocular slam. In: *European Conference on Computer Vision*. Springer, pp. 834–849.
- Engel J, Stückler J and Cremers D (2015) Large-scale direct slam with stereo cameras. In: *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 1935–1942.
- Forster C, Carlone L, Dellaert F and Scaramuzza D (2015) Imu preintegration on manifold for efficient visual-inertial maximum-a-posteriori estimation. In: *Robotics: Science and Systems*. Georgia Institute of Technology.
- Forster C, Carlone L, Dellaert F and Scaramuzza D (2017) On-manifold preintegration for real-time visual-inertial odometry. *IEEE Transactions on Robotics* 33(1): 1–21.
- Geiger A, Lenz P and Urtasun R (2012) Are we ready for autonomous driving? the kitti vision benchmark suite. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 3354–3361.
- Grisetti G, Kummerle R, Stachniss C and Burgard W (2010) A tutorial on graph-based SLAM. *IEEE Intelligent Transportation Systems Magazine* 2(4): 31–43.
- Guo C, Kottas D, DuToit R, Ahmed A, Li R and Roumeliotis S (2014) Efficient visual-inertial navigation using a rolling-shutter camera with inaccurate timestamps. In: *Robotics: Science and Systems*. Berkeley, USA.
- Guo C and Roumeliotis S (2013) Imu-rgbd camera navigation using point and plane features. In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 3164–3171.
- Hartley R and Zisserman A (2000) *Multiple View Geometry in Computer Vision*. Cambridge, U. K.: Cambridge University Press.
- Hertzberg C, Wagner R, Frese U and Schröder L (2013) Integrating generic sensor fusion algorithms with sound state representations through encapsulation of manifolds. *Information Fusion* 14(1): 57–77.
- Hesch J, Kottas D, Bowman S and Roumeliotis S (2013) Consistency analysis and improvement of vision-aided inertial navigation. *IEEE Transactions on Robotics* PP(99): 1–19.
- Huang G, Kaess M and Leonard J (2013) Consistent sparsification for graph optimization. In: *Proc. of the European Conference on Mobile Robots*. Barcelona, Spain, pp. 150–157.
- Huang G, Kaess M and Leonard J (2014) Towards consistent visual-inertial navigation. In: *Proc. of the IEEE International Conference on Robotics and Automation*. Hong Kong, China, pp. 4926–4933.
- Huang G, Mourikis AI and Roumeliotis SI (2011) An observability constrained sliding window filter for SLAM. In: *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. San Francisco, CA, pp. 65–72.
- Kaess M, Johannsson H, Roberts R, Ila V, Leonard J and Dellaert F (2012) iSAM2: Incremental smoothing and mapping using the Bayes tree. *International Journal of Robotics Research* 31: 217–236.
- Koenig N and Howard A (2004) Design and use paradigms for gazebo, an open-source multi-robot simulator. In: *Proc. 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004)*, volume 3. IEEE, pp. 2149–2154.
- Kummerle R, Grisetti G, Strasdat H, Konolige K and Burgard W (2011) g2o: A general framework for graph optimization. In: *Proc. of the IEEE International Conference on Robotics and Automation*. Shanghai, China, pp. 3607–3613.
- Leutenegger S, Lynen S, Bosse M, Siegwart R and Furgale P (2015) Keyframe-based visual-inertial odometry using nonlinear optimization. *The International Journal of Robotics Research* 34(3): 314–334.
- Li M, Yu H, Zheng X and Mourikis AI (2014) High-fidelity sensor modeling and calibration in vision-aided inertial navigation. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. Hong Kong, pp. 409–416.
- Ling Y, Liu T and Shen S (2016) Aggressive quadrotor flight using dense visual-inertial fusion. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 1499–1506.
- Lupton T and Sukkarieh S (2012) Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions. *IEEE Transactions on Robotics* 28(1): 61–76.
- Mourikis A and Roumeliotis S (2007) A multi-state constraint Kalman filter for vision-aided inertial navigation. In: *Proceedings of the IEEE International Conference on Robotics and Automation*. Rome, Italy, pp. 3565–3572.
- Mourikis A, Trawny N, Roumeliotis S, Johnson AE and Matthies LH (2007) Vision-aided inertial navigation for precise planetary landing: Analysis and experiments. In: *Robotics: Science and Systems*. Atlanta, GA, USA.
- Mur-Artal R and Tardós JD (2017) Orb-slam2: An open-source slam system for monocular, stereo, and rgbd cameras. *IEEE Transactions on Robotics* 33(5): 1255–1262.
- Nerurkar E, Wu K and Roumeliotis S (2014) C-KLAM: Constrained keyframe-based localization and mapping. In: *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. pp. 3638–3643.
- Nikolic J, Rehder J, Burri M, Gohl P, Leutenegger S, Furgale PT and Siegwart R (2014) A synchronized visual-inertial sensor system with fpga pre-processing for accurate real-time slam. In: *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 431–437.
- Paul MK, Wu K, Hesch JA, Nerurkar ED and Roumeliotis SI (2017) A comparative analysis of tightly-coupled monocular, binocular, and stereo vins. In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 165–172.
- Qin T, Li P and Shen S (2017) Vins-mono: A robust and versatile monocular visual-inertial state estimator. *CoRR* abs/1708.03852.
- Rosten E, Porter R and Drummond T (2010) Faster and better: A machine learning approach to corner detection. *IEEE transactions on pattern analysis and machine intelligence* 32(1): 105–119.
- Roumeliotis S and Burdick J (2002) Stochastic cloning: A generalized framework for processing relative state measurements. In: *Proceedings of the IEEE International Conference on Robotics and Automation*. Washington, DC, pp. 1788–1795.

- Shen S, Michael N and Kumar V (2015) Tightly-coupled monocular visual-inertial fusion for autonomous flight of rotorcraft mavs. In: *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 5303–5310.
- Trawny N and Roumeliotis S (2005) Indirect Kalman filter for 3D attitude estimation. Technical report, University of Minnesota, Dept. of Comp. Sci. & Eng.
- Usenko V, Engel J, Stückler J and Cremers D (2016) Direct visual-inertial odometry with stereo cameras. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 1885–1892.
- Wang R, Schwörer M and Cremers D (2017) Stereo dso: Large-scale direct sparse visual odometry with stereo cameras. In: *International Conference on Computer Vision (ICCV), Venice, Italy*.
- Wu KJ, Ahmed AM, Georgiou GA and Roumeliotis SI (2015) A square root inverse filter for efficient vision-aided inertial navigation on mobile devices. In: *2015 Robotics: Science and Systems Conference, RSS 2015*. MIT Press Journals.
- Yang Y, Maley J and Huang G (2017) Null-space-based marginalization: Analysis and algorithm. In: *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*. Vancouver, Canada.
- Zhang Z and Scaramuzza D (2018) A tutorial on quantitative trajectory evaluation for visual(-inertial) odometry. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

## Appendix A: Index to multimedia Extensions

Archives of IJRR multimedia extensions published prior to 2014 can be found at <http://www.ijrr.org>, after 2014 all videos are available on the IJRR YouTube channel at <http://www.youtube.com/user/ijrrmultimedia>

### Table of Multimedia Extension

Extension	Media type	Description
1	Video	Both tightly-coupled indirect and loosely-coupled direct VINS with Model 1 preintegration are tested on the “V1_03_difficult” EurocMav dataset. The estimated trajectories are plotted in red.
2	Video	The indirect system with Model 1 preintegration is demonstrated on a 240 meter large-scale indoor trajectory, on a collected dataset at the University of Delaware Smith Hall.

## Appendix B: Preintegration Measurement Jacobians

### B.1. Model 1 Measurement Jacobians

We first partition the preintegrated measurement residual as follows:

$$\mathbf{e}_{IMU}(\mathbf{x}) = [\mathbf{e}_\theta^\top \quad \mathbf{e}_{b_\omega}^\top \quad \mathbf{e}_v^\top \quad \mathbf{e}_{b_a}^\top \quad \mathbf{e}_p^\top]^\top \quad (127)$$

The measurement Jacobian with respect to each element of the error state vector can be found by perturbing the measurement function for the corresponding element. For example, the relative-rotation measurement residual  $\mathbf{e}_\theta$  is perturbed by a small change in gyro bias around the current estimate, i.e.,  $\mathbf{b}_{\omega_k} - \mathbf{b}_{\omega_k}^* = \hat{\mathbf{b}}_{\omega_k} + \delta\mathbf{b}_{\omega_k} - \mathbf{b}_{\omega_k}^*$ , which yields (see Equation (72)):

$$\begin{aligned} \mathbf{e}_\theta &= 2\text{vec} \left( \overset{k+1}{G} \hat{\bar{q}} \otimes \overset{k}{G} \hat{\bar{q}}^{-1} \otimes \overset{k+1}{\bar{q}} \hat{\bar{q}}^{-1} \otimes \begin{bmatrix} \frac{\mathbf{J}_q(\hat{\mathbf{b}}_{\omega_k} + \delta\mathbf{b}_{\omega_k} - \mathbf{b}_{\omega_k}^*)}{2} \\ 1 \end{bmatrix} \right) \\ &=: 2\text{vec} \left( \hat{\bar{q}}_r \otimes \begin{bmatrix} \frac{\mathbf{J}_q(\hat{\mathbf{b}}_{\omega_k} + \delta\mathbf{b}_{\omega_k} - \mathbf{b}_{\omega_k}^*)}{2} \\ 1 \end{bmatrix} \right) \\ &= 2\text{vec} \left( \mathcal{L}(\hat{\bar{q}}_r) \begin{bmatrix} \frac{\mathbf{J}_q(\hat{\mathbf{b}}_{\omega_k} + \delta\mathbf{b}_{\omega_k} - \mathbf{b}_{\omega_k}^*)}{2} \\ 1 \end{bmatrix} \right) \\ &= 2\text{vec} \left( \begin{bmatrix} \hat{q}_{r,4} \mathbf{I}_{3 \times 3} - \lfloor \hat{\mathbf{q}}_r \rfloor & \hat{\mathbf{q}}_r \\ -\hat{\mathbf{q}}_r^\top & \hat{q}_{r,4} \end{bmatrix} \begin{bmatrix} \frac{\mathbf{J}_q(\hat{\mathbf{b}}_{\omega_k} + \delta\mathbf{b}_{\omega_k} - \mathbf{b}_{\omega_k}^*)}{2} \\ 1 \end{bmatrix} \right) \\ &= (\hat{q}_{r,4} \mathbf{I}_{3 \times 3} - \lfloor \hat{\mathbf{q}}_r \rfloor) \mathbf{J}_q(\hat{\mathbf{b}}_{\omega_k} + \delta\mathbf{b}_{\omega_k} - \mathbf{b}_{\omega_k}^*) + \text{other terms} \end{aligned} \quad (128)$$

As a result, the Jacobian with respect to a perturbation in bias can be read out as:

$$\frac{\partial \mathbf{e}_\theta}{\partial \delta \mathbf{b}_{\omega_k}} = (\hat{q}_{r,4} \mathbf{I}_{3 \times 3} - \lfloor \hat{\mathbf{q}}_r \rfloor) \mathbf{J}_q \quad (129)$$

Proceeding analogously, the Jacobian with respect to  $\overset{k+1}{\delta\theta_G}$  can be found as follows:

$$\begin{aligned} \mathbf{e}_\theta &= 2\text{vec} \left( \begin{bmatrix} \frac{\overset{k+1}{\delta\theta_G}}{2} \\ 1 \end{bmatrix} \otimes \overset{k}{G} \hat{\bar{q}} \otimes \overset{k}{G} \hat{\bar{q}}^{-1} \otimes \overset{k+1}{\bar{q}} \hat{\bar{q}}^{-1} \otimes \hat{\bar{q}}_b \right) \\ &=: 2\text{vec} \left( \begin{bmatrix} \frac{\overset{k+1}{\delta\theta_G}}{2} \\ 1 \end{bmatrix} \otimes \hat{\bar{q}}_{rb} \right) \\ &= 2\text{vec} \left( \mathcal{R}(\hat{\bar{q}}_{rb}) \begin{bmatrix} \frac{\overset{k+1}{\delta\theta_G}}{2} \\ 1 \end{bmatrix} \right) \\ &= 2\text{vec} \left( \begin{bmatrix} \hat{q}_{rb,4} \mathbf{I}_{3 \times 3} - \lfloor \hat{\mathbf{q}}_{rb} \rfloor & \hat{\mathbf{q}}_{rb} \\ -\hat{\mathbf{q}}_{rb}^\top & \hat{q}_{rb,4} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \overset{k+1}{\delta\theta_G} \\ 1 \end{bmatrix} \right) \\ &= (\hat{q}_{rb,4} \mathbf{I}_{3 \times 3} + \lfloor \hat{\mathbf{q}}_{rb} \rfloor)^{k+1} \delta\theta_G + \text{other terms} \\ &\Rightarrow \frac{\partial \mathbf{e}_\theta}{\partial \overset{k+1}{\delta\theta_G}} = \hat{q}_{rb,4} \mathbf{I}_{3 \times 3} + \lfloor \hat{\mathbf{q}}_{rb} \rfloor \end{aligned} \quad (130)$$

Similarly, the Jacobian with respect to  $\overset{k}{\delta\theta_G}$  is computed by:

$$\begin{aligned} \mathbf{e}_\theta &= 2\text{vec} \left( \overset{k+1}{G} \hat{\bar{q}} \otimes \overset{k}{G} \hat{\bar{q}}^{-1} \otimes \begin{bmatrix} -\frac{\overset{k}{\delta\theta_G}}{2} \\ 1 \end{bmatrix} \otimes \overset{k+1}{\bar{q}} \hat{\bar{q}}^{-1} \otimes \hat{\bar{q}}_b \right) \\ &=: 2\text{vec} \left( \hat{\bar{q}}_n \otimes \begin{bmatrix} -\frac{\overset{k}{\delta\theta_G}}{2} \\ 1 \end{bmatrix} \otimes \hat{\bar{q}}_{mb}^{-1} \right) \\ &= 2\text{vec} \left( \mathcal{L}(\hat{\bar{q}}_n) \mathcal{R}(\hat{\bar{q}}_{mb}^{-1}) \begin{bmatrix} -\frac{\overset{k}{\delta\theta_G}}{2} \\ 1 \end{bmatrix} \right) \\ &= 2\text{vec} \left( \begin{bmatrix} \hat{q}_{n,4} \mathbf{I}_{3 \times 3} - \lfloor \hat{\mathbf{q}}_n \rfloor & \hat{\mathbf{q}}_n \\ -\hat{\mathbf{q}}_n^\top & \hat{q}_{n,4} \end{bmatrix} \times \right. \\ &\quad \left. \begin{bmatrix} \bar{q}_{mb,4} \mathbf{I}_{3 \times 3} - \lfloor \bar{\mathbf{q}}_{mb} \rfloor & -\mathbf{q}_{mb} \\ \bar{\mathbf{q}}_{mb}^\top & \bar{q}_{mb,4} \end{bmatrix} \begin{bmatrix} -\frac{\overset{k}{\delta\theta_G}}{2} \\ 1 \end{bmatrix} \right) \\ &= -((\hat{q}_{n,4} \mathbf{I}_{3 \times 3} - \lfloor \hat{\mathbf{q}}_n \rfloor)(\bar{q}_{mb,4} \mathbf{I}_{3 \times 3} - \lfloor \bar{\mathbf{q}}_{mb} \rfloor) \\ &\quad + \hat{\mathbf{q}}_n \bar{\mathbf{q}}_{mb}^\top)^k \delta\theta_G + \text{other terms} \\ &\Rightarrow \frac{\partial \mathbf{e}_\theta}{\partial \overset{k}{\delta\theta_G}} = \end{aligned}$$

$$-((\hat{q}_{n,4}\mathbf{I}_{3 \times 3} - [\hat{\mathbf{q}}_n])(\bar{q}_{mb,4}\mathbf{I}_{3 \times 3} - [\mathbf{q}_{mb}]) + \hat{\mathbf{q}}_n\bar{\mathbf{q}}_{mb}^\top) \quad (131)$$

where we have defined several intermediate quaternions,  $\hat{q}_r, \hat{q}_{rb}, \hat{q}_n$ , and  $\hat{q}_{mb}$ , for ease of notation. We compute the Jacobians of the remaining preintegrated measurements as follows (Eckenhoff, Geneva and Huang 2018):

$$\frac{\partial \mathbf{e}_{b_\omega}}{\partial \delta \mathbf{b}_{\omega_k}} = -\mathbf{I} \quad (132)$$

$$\frac{\partial \mathbf{e}_{b_\omega}}{\partial \delta \mathbf{b}_{\omega_{k+1}}} = \mathbf{I} \quad (133)$$

$$\frac{\partial \mathbf{e}_v}{\partial^k \delta \boldsymbol{\theta}_G} = \left[ {}_G^k \hat{\mathbf{R}} ({}^G \hat{\mathbf{v}}_{k+1} - {}^G \hat{\mathbf{v}}_k + {}^G \mathbf{g} \Delta t) \right] \quad (134)$$

$$\frac{\partial \mathbf{e}_v}{\partial \delta \mathbf{b}_{\omega_k}} = -\mathbf{J}_\beta \quad (135)$$

$$\frac{\partial \mathbf{e}_v}{\partial {}^G \delta \mathbf{v}_k} = -{}_G^k \hat{\mathbf{R}} \quad (136)$$

$$\frac{\partial \mathbf{e}_v}{\partial {}^G \delta \mathbf{v}_{k+1}} = {}_G^k \hat{\mathbf{R}} \quad (137)$$

$$\frac{\partial \mathbf{e}_v}{\partial \delta \mathbf{b}_a} = -\mathbf{H}_\beta \quad (138)$$

$$\frac{\partial \mathbf{e}_{b_a}}{\partial \delta \mathbf{b}_{\omega_k}} = -\mathbf{I} \quad (139)$$

$$\frac{\partial \mathbf{e}_{b_a}}{\partial \delta \mathbf{b}_{\omega_{k+1}}} = \mathbf{I} \quad (140)$$

$$\frac{\partial \mathbf{e}_p}{\partial^k \delta \boldsymbol{\theta}_G} = \left[ {}_G^k \hat{\mathbf{R}} \left( {}^G \hat{\mathbf{p}}_{k+1} - {}^G \hat{\mathbf{p}}_k - {}^G \hat{\mathbf{v}}_k \Delta t + \frac{1}{2} {}^G \mathbf{g} \Delta t^2 \right) \right] \quad (141)$$

$$\frac{\partial \mathbf{e}_p}{\partial {}^G \delta \mathbf{v}_k} = -{}_G^k \hat{\mathbf{R}} \Delta t \quad (142)$$

$$\frac{\partial \mathbf{e}_p}{\partial {}^G \delta \mathbf{p}_k} = -\mathbf{H}_\alpha \quad (143)$$

$$\frac{\partial \mathbf{e}_p}{\partial {}^G \delta \mathbf{p}_{k+1}} = -{}_G^k \hat{\mathbf{R}} \quad (144)$$

$$\frac{\partial \mathbf{e}_p}{\partial \delta \mathbf{b}_{\omega_{k+1}}} = {}_G^k \hat{\mathbf{R}} \quad (145)$$

$$(146)$$

## B.2. Model 2 Measurement Jacobians

For Model 2, the orientation measurement Jacobians remain the same as in Model 1. For the remaining measurement Jacobians, we compute them in the same way as in Model 1 and are given by (see Equation (103)):

$$\frac{\partial \mathbf{e}_{b_\omega}}{\partial \delta \mathbf{b}_{\omega_k}} = -\mathbf{I} \quad (147)$$

$$\frac{\partial \mathbf{e}_{b_\omega}}{\partial \delta \mathbf{b}_{\omega_{k+1}}} = \mathbf{I} \quad (148)$$

$$\frac{\partial \mathbf{e}_v}{\partial^k \delta \boldsymbol{\theta}_G} = [{}_G^k \mathbf{R} ({}^G \hat{\mathbf{v}}_{k+1} - {}^G \hat{\mathbf{v}}_k)] - \mathbf{O}_\beta (\tilde{q}_4 \mathbf{I} + [\tilde{\mathbf{q}}]) \quad (149)$$

$$\frac{\partial \mathbf{e}_v}{\partial \delta \mathbf{b}_{\omega_k}} = -\mathbf{J}_\beta \quad (150)$$

$$\frac{\partial \mathbf{e}_v}{\partial {}^G \delta \mathbf{v}_k} = -{}_G^k \hat{\mathbf{R}} \quad (151)$$

$$\frac{\partial \mathbf{e}_v}{\partial {}^G \delta \mathbf{v}_{k+1}} = {}_G^k \hat{\mathbf{R}} \quad (152)$$

$$\frac{\partial \mathbf{e}_v}{\partial \delta \mathbf{b}_{\omega_k}} = -\mathbf{H}_\beta \quad (153)$$

$$\frac{\partial \mathbf{e}_{b_a}}{\partial \delta \mathbf{b}_{\omega_{k+1}}} = \mathbf{I} \quad (154)$$

$$\frac{\partial \mathbf{e}_{b_a}}{\partial \delta \mathbf{b}_{\omega_k}} = -\mathbf{I} \quad (155)$$

$$\begin{aligned} \frac{\partial \mathbf{e}_p}{\partial^k \delta \boldsymbol{\theta}_G} &= [{}_G^k \mathbf{R} ({}^G \hat{\mathbf{p}}_{k+1} - {}^G \hat{\mathbf{p}}_k - {}^G \hat{\mathbf{v}}_k \Delta T)] \\ &\quad - \mathbf{O}_\alpha (\tilde{q}_4 \mathbf{I} + [\tilde{\mathbf{q}}]) \end{aligned} \quad (156)$$

$$\frac{\partial \mathbf{e}_p}{\partial \delta \mathbf{b}_{\omega_k}} = -\mathbf{J}_\alpha \quad (157)$$

$$\frac{\partial \mathbf{e}_p}{\partial {}^G \delta \mathbf{v}_k} = -{}_G^k \hat{\mathbf{R}} \Delta T \quad (158)$$

$$\frac{\partial \mathbf{e}_p}{\partial \delta \mathbf{b}_{\omega_k}} = -\mathbf{H}_\alpha \quad (159)$$

$$\frac{\partial \mathbf{e}_p}{\partial {}^G \delta \mathbf{p}_k} = -{}_G^k \hat{\mathbf{R}} \quad (160)$$

$$\frac{\partial \mathbf{e}_p}{\partial {}^G \delta \mathbf{p}_{k+1}} = {}_G^k \hat{\mathbf{R}} \quad (161)$$

where  $[\tilde{\mathbf{q}}^\top \tilde{q}_4]^\top = {}_G^k \hat{\mathbf{q}} \otimes {}_G^k \bar{\mathbf{q}}^{*-1}$

## Appendix C: Inverse-Depth Measurement Jacobians

We denote  $a$  and  $i$  the anchoring time step and the associated anchoring camera frame, respectively. Consider the case where we receive an image of the same feature at step  $k$  from camera  $j$ . This measurement can be divided into three categories: (i) when the measurement refers to both the anchoring time and camera that the inverse depth is being represented in; (ii) when the measurement refers to the same anchoring time, but a different camera; (iii) when the anchoring time and measurement time are distinct.

In case (i), we have (see Equation (111)):

$$\mathbf{h} = \begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix} \quad (162)$$

Then the measurement Jacobians are computed by (see Equations (106), (110) and (111)):

$$\frac{\partial \mathbf{e}_{fjk}}{\partial \alpha} = \mathbf{H}_{proj}(0, 0, 2, 1) \quad (163)$$

$$\frac{\partial \mathbf{e}_{fjk}}{\partial \beta} = \mathbf{H}_{proj}(0, 1, 2, 1) \quad (164)$$

$$\frac{\partial \mathbf{e}_{fjk}}{\partial \rho} = \mathbf{0} \quad (165)$$

$$\mathbf{H}_{proj} = \begin{bmatrix} \frac{1}{h_3} & 0 & \frac{-h_1}{(h_3)^2} \\ 0 & \frac{1}{h_3} & \frac{-h_2}{(h_3)^2} \end{bmatrix} \quad (166)$$

where  $\mathbf{H}_{proj}(i, j, k, l)$  refers to the block matrix of size  $(k, l)$  with starting index  $(i, j)$ .

In case (ii) where  $k$  refers to the same imaging time but a different camera (such as a stereo partner), we have (see Equation (111)):

$$\mathbf{h} = \frac{C_j}{C_i} \mathbf{R} \begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix} + \rho^{C_j} \mathbf{p}_{C_i} \quad (167)$$

Because in this case the transformation parameters are rigid and known, we need only the derivatives with respect to the unknown feature parameterization:

$$\frac{\partial \mathbf{e}_{fjk}}{\partial \alpha} = \left[ \mathbf{H}_{proj}^{C_j} \mathbf{R} \right] (0, 0, 2, 1) \quad (168)$$

$$\frac{\partial \mathbf{e}_{fjk}}{\partial \beta} = \left[ \mathbf{H}_{proj}^{C_j} \mathbf{R} \right] (0, 1, 2, 1) \quad (169)$$

$$\frac{\partial \mathbf{e}_{fjk}}{\partial \rho} = {}^{C_j} \mathbf{p}_{C_i} \quad (170)$$

In case (iii) where instead the measurement refers to a different time, we can write out the rigid transformation between the anchor and new current camera frame as follows:

$$\begin{aligned} {}^{C_{k,j}} \mathbf{p}_f &= {}^{C_{k,j}} \mathbf{R} {}^{C_{a,i}} \mathbf{p}_f + {}^{C_{k,j}} \mathbf{p}_{C_{a,i}} \\ &= \frac{1}{\rho^I} \mathbf{R}_G^k \mathbf{R}_a^G \mathbf{R}_{C_i}^I \mathbf{R} \begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix} + \\ &\quad {}_G^{C_{k,j}} \mathbf{R} ({}^G \mathbf{p}_{C_{a,i}} - {}^G \mathbf{p}_{C_{k,j}}) \\ &= \frac{1}{\rho^I} \mathbf{R}_G^k \mathbf{R}_a^G \mathbf{R}_{C_i}^I \mathbf{R} \begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix} + \\ &\quad {}_I^C \mathbf{R}_G^k \mathbf{R} ({}^G \mathbf{p}_a + {}_a^G \mathbf{R}^I \mathbf{p}_{C_i} - {}^G \mathbf{p}_k - {}_k^G \mathbf{R}^I \mathbf{p}_{C_j}) \\ &= {}_I^C \mathbf{R}_G^k \mathbf{R}_a^G \mathbf{R}_{C_i}^I \mathbf{R} \left( \frac{1}{\rho} \begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix} - {}^{C_i} \mathbf{p}_I \right) + \\ &\quad {}_I^C \mathbf{R}_G^k \mathbf{R} ({}^G \mathbf{p}_a - {}^G \mathbf{p}_k) + {}^{C_j} \mathbf{p}_I \end{aligned} \quad (171)$$

With this, we have:

$$\begin{aligned} \mathbf{h} &= {}_I^C \mathbf{R}_G^k \mathbf{R}_a^G \mathbf{R}_{C_i}^I \mathbf{R} \left( \begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix} - \rho^{C_i} \mathbf{p}_I \right) + \\ &\quad {}_I^{C_j} \mathbf{R}_G^k \mathbf{R} ({}^G \mathbf{p}_a - {}^G \mathbf{p}_k) + \rho^{C_j} \mathbf{p}_I \end{aligned} \quad (172)$$

We can then take the derivative with respect to each variable:

$$\frac{\partial \mathbf{e}_{fjk}}{\partial \alpha \delta \boldsymbol{\theta}_G} = -\mathbf{H}_{proj} {}_I^{C_j} \mathbf{R}_G^k \mathbf{R}_a^G \mathbf{R} \left[ {}_I^{C_i} \mathbf{R} \left( \begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix} - \rho^{C_i} \mathbf{p}_I \right) \right] \quad (173)$$

$$\frac{\partial \mathbf{e}_{fjk}}{\partial {}^G \mathbf{p}_a} = \mathbf{H}_{proj} {}_I^{C_j} \mathbf{R}_G^k \mathbf{R} \quad (174)$$

$$\begin{aligned} \frac{\partial \mathbf{e}_{fjk}}{\partial {}^k \delta \boldsymbol{\theta}_G} &= \mathbf{H}_{proj} {}_I^{C_j} \mathbf{R} \left[ {}_G^k \mathbf{R}_a^G \mathbf{R}_{C_i}^I \mathbf{R} \left( \begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix} - \rho^{C_i} \mathbf{p}_I \right) + \right. \\ &\quad \left. {}_G^k \mathbf{R} ({}^G \mathbf{p}_a - {}^G \mathbf{p}_k) \right] \end{aligned} \quad (175)$$

$$\frac{\partial \mathbf{e}_{fjk}}{\partial {}^G \mathbf{p}_k} = -\mathbf{H}_{proj} {}_I^{C_j} \mathbf{R}_G^k \mathbf{R} \quad (176)$$

$$\frac{\partial \mathbf{e}_{fjk}}{\partial \alpha} = \left[ \mathbf{H}_{proj} {}_I^{C_j} \mathbf{R}_G^k \mathbf{R}_a^G \mathbf{R}_{C_i}^I \mathbf{R} \right] (0, 0, 2, 1) \quad (177)$$

$$\frac{\partial \mathbf{e}_{fjk}}{\partial \beta} = \left[ \mathbf{H}_{proj} {}_I^{C_j} \mathbf{R}_G^k \mathbf{R}_a^G \mathbf{R}_{C_i}^I \mathbf{R} \right] (0, 1, 2, 1) \quad (178)$$

$$\begin{aligned} \frac{\partial \mathbf{e}_{fjk}}{\partial \rho} &= \mathbf{H}_{proj} (-{}_I^C \mathbf{R}_G^k \mathbf{R}_a^G \mathbf{R}_{C_i}^I \mathbf{R} {}^{C_i} \mathbf{p}_I + \\ &\quad {}_I^C \mathbf{R}_G^k \mathbf{R} ({}^G \mathbf{p}_a - {}^G \mathbf{p}_k) + {}^{C_j} \mathbf{p}_I) \end{aligned} \quad (179)$$

## Appendix D: Relative-Pose Measurement Jacobian

Recall that in Equation (125),  $j$  denotes the query image and  $k$  is the keyframe. We partition the relative-pose residual  $\mathbf{e}_d$  into the relative-orientation residual  $\mathbf{e}_\theta$  and the relative-position residual  $\mathbf{e}_p$ . The Jacobians with respect to the states can be found by perturbation in the same way as before.

$$\begin{aligned} \mathbf{e}_\theta &= 2\text{vec} \left( \begin{bmatrix} \frac{j \delta \boldsymbol{\theta}_G}{2} \\ 1 \end{bmatrix} \otimes {}_G^j \hat{q} \otimes {}_G^k \hat{q}^{-1} \otimes {}_k^j \check{q}^{-1} \right) \\ &= 2\text{vec} \left( \mathcal{R} \left( {}_G^j \hat{q} \otimes {}_G^k \hat{q}^{-1} \otimes {}_k^j \check{q}^{-1} \right) \begin{bmatrix} \frac{j \delta \boldsymbol{\theta}_G}{2} \\ 1 \end{bmatrix} \right) \\ &= (\bar{q}_{r,4} \mathbf{I} + \lfloor \mathbf{q}_r \rfloor)^j \delta \boldsymbol{\theta}_G + \dots \\ \Rightarrow \frac{\partial \mathbf{e}_\theta}{\partial {}^j \delta \boldsymbol{\theta}_G} &= (\bar{q}_{r,4} \mathbf{I} + \lfloor \mathbf{q}_r \rfloor) \end{aligned} \quad (180)$$

Similarly, we perturb the quaternion estimate of the keyframe to compute the corresponding Jacobian as:

$$\begin{aligned} \mathbf{e}_\theta &= 2\text{vec} \left( {}_G^j \hat{q} \otimes {}_G^k \hat{q}^{-1} \otimes \begin{bmatrix} -\frac{k \delta \boldsymbol{\theta}_G}{2} \\ 1 \end{bmatrix} \otimes {}_k^j \check{q}^{-1} \right) \\ &= 2\text{vec} \left( \mathcal{L} \left( {}_j^k \hat{q} \right) \mathcal{R} \left( {}_j^k \check{q} \right)^\top \begin{bmatrix} -\frac{k \delta \boldsymbol{\theta}_G}{2} \\ 1 \end{bmatrix} \right) = \\ &\quad - \left( \left( {}_k^j \hat{q}_4 \mathbf{I} - \lfloor {}_k^j \hat{\mathbf{q}} \rfloor \right) \left( {}_k^j \check{q}_4 \mathbf{I} - \lfloor {}_k^j \check{\mathbf{q}} \rfloor \right) + {}_k^j \hat{\mathbf{q}}_k^j \check{\mathbf{q}}^\top \right)^k \delta \boldsymbol{\theta}_G + \dots \\ \Rightarrow \frac{\partial \mathbf{e}_\theta}{\partial {}^k \delta \boldsymbol{\theta}_G} &= - \left( \left( {}_k^j \hat{q}_4 \mathbf{I} - \lfloor {}_k^j \hat{\mathbf{q}} \rfloor \right) \left( {}_k^j \check{q}_4 \mathbf{I} - \lfloor {}_k^j \check{\mathbf{q}} \rfloor \right) + {}_k^j \hat{\mathbf{q}}_k^j \check{\mathbf{q}}^\top \right) \end{aligned}$$

Again by following a similar procedure, we can find the Jacobians of the relative-position residual with respect to the state as follows:

$$\frac{\partial \mathbf{e}_p}{\partial {}_G^k \delta \mathbf{p}_j} = {}_G^k \hat{\mathbf{R}} \quad (181)$$

$$\frac{\partial \mathbf{e}_p}{\partial {}_G^k \delta \mathbf{p}_k} = -{}_G^k \hat{\mathbf{R}} \quad (182)$$

$$\frac{\partial \mathbf{e}_p}{\partial {}^k \delta \boldsymbol{\theta}_G} = \lfloor {}_G^k \hat{\mathbf{R}} ({}^G \mathbf{p}_j - {}^G \mathbf{p}_k) \rfloor \quad (183)$$