

An Efficient Closed-Form Solution to Full Visual-Inertial State Initialization

Samuel Cerezo , Seong Hun Lee , and Javier Civera , *Member, IEEE*

Abstract—In this letter, we present a closed-form initialization method that recovers the full visual-inertial state without nonlinear optimization. Unlike previous approaches that rely on iterative solvers, our formulation yields analytical, easy-to-implement, and numerically stable solutions for reliable start-up. Our method builds on small-rotation and constant-velocity approximations, which keep the formulation compact while preserving the essential coupling between motion and inertial measurements. We further propose an observability-driven, two-stage initialization scheme that balances accuracy with initialization latency. Extensive experiments on the EuRoC dataset validate our assumptions: our method achieves 10 – 20% lower initialization error than optimization-based approaches, while using $4\times$ shorter initialization windows and reducing computational cost by $5\times$.

Index Terms—Visual-inertial SLAM, state initialization, sensor fusion.

I. INTRODUCTION

VISUAL-INERTIAL SLAM has become a core technology in robotics and mobile devices, enabling real-time localization and mapping by fusing camera and inertial measurements [1], [2], [3]. While camera intrinsics can be calibrated once and remain relatively stable over time, inertial sensors must be re-initialized at each run. This initialization step estimates critical quantities such as the initial velocity, gravity direction, and gyroscope and accelerometer biases, which serve as the seed for subsequent state estimation. Accurate initialization is therefore essential: poor initial estimates may cause optimization or filtering to converge to local minima, leading to scale drift, degraded tracking, or even complete failure of the system. Conversely, proper initialization enhances robustness under challenging conditions and recovers the metric scale that monocular vision alone cannot provide.

A central insight from the SLAM literature is that orientation plays a pivotal role: once rotations are reliably estimated, the recovery of the remaining states becomes more robust. Carlone and Censi [4] showed that accurate orientation estimates prevent spurious minima in pose-graph optimization, and Carlone

Received 24 November 2025; accepted 25 March 2026. Date of publication 9 April 2026; date of current version 20 April 2026. This article was recommended for publication by Associate Editor R. Boutteau and Editor P. Vasseur upon evaluation of the reviewers' comments. This work was supported in part by Spanish Government under Grant PID2021-127685NB-I00 and Grant PID2024-155886NB-I00, in part by AEI/MCIN under Grant PRE2022-103765, and in part by Aragón Government through the Program Odón de Buen POST-DOC_TA_13_25 under Project T45_23R. (Corresponding author: Samuel Cerezo.)

The authors are with the I3A, Universidad de Zaragoza, 50018 Zaragoza, Spain (e-mail: sacerezo@unizar.es; seonghunlee@unizar.es; jcivera@unizar.es).

Digital Object Identifier 10.1109/LRA.2026.3682536

TABLE I
COMPARISON OF EXISTING INITIALIZATION METHODS. SUMMARY OF WHETHER EACH METHOD PROVIDES A CLOSED-FORM SOLUTION AND ESTIMATES GYROSCOPE AND ACCELEROMETER BIASES.

Method	Closed-form	Gyro-bias	Accel-bias
Martinelli [7]	✓	✗	✗
Kaiser et al. [8]	✗	✓	✓
Domínguez et al. [9]	✗	✓	✓
Campos et al. [10], [11]	✗	✓	✓
Concha et al. [13]	✗	✓	✓
Evangelidis and Micušík [16]	✓	✗	✗
Wang et al. [15]	✗	✓	✓
Ours	✓	✓	✓

et al. [5] showed that good rotation initialization significantly improves convergence and accuracy in large-scale SLAM. These results motivate our approach, which relies on simple but accurate decoupled orientation approximations to estimate the rest of the states.

Existing visual-inertial initialization strategies fall into two main categories. Closed-form solutions have been proposed by Martinelli [6], [7], but they assume partial knowledge of the states, specifically of the sensor biases. Later works addressed this limitation using iterative nonlinear optimization [8], [9], [10], [11], which achieves accurate full state initialization but requires careful seeding and incurs significant computational cost. More recent formulations attempt to reduce complexity by decoupling rotation and translation [12], [13], [14], or by combining closed-form solutions with filtering schemes [15]. However, none of these approaches provides a fully analytical solution for the complete visual-inertial state.

In this letter, we present the first closed-form initialization method for all states of a visual-inertial system, including gyroscope and accelerometer biases. Our formulation builds on small-rotation and constant-velocity approximations, which hold for common sensor hardware and motion patterns. We further contribute a two-stage criterion that assesses the observability of the visual-inertial states and enables reliable initialization under diverse conditions. Extensive experiments show that our method achieves higher accuracy than relevant baselines, while significantly reducing the computational cost and initialization delay.

II. RELATED WORK

Early works on closed-form initialization were introduced by Martinelli [6], [7], who derived minimal unique solutions under the assumption of known gyroscope and accelerometer biases. Subsequent approaches shifted towards nonlinear optimization to jointly estimate these biases together with the remaining

states [8], [9], and later refinements such as those by Campos et al. [10], [11] achieved robust results but still relied on iterative procedures requiring careful seeding.

More recently, alternative formulations have been proposed to improve efficiency and robustness. Evangelidis and Micušík [16] revisited the problem with a new closed-form solver for the vi-SfM problem, elegant but better suited for offline reconstruction. Decoupled approaches to relative [17] and multi-view [12] camera motion have also been proposed [13], [14], but they still require nonlinear optimization and in some cases cannot recover accelerometer bias. Hybrid strategies such as the framework of Wang et al. [15] can estimate both biases through an error-state Kalman filter, at the expense of additional algorithmic complexity.

In parallel, approximations have played an important role in related problems: the small-angle approximation in triangulation [18], [19], and the small-rotation approximation in vision-only reconstruction from accidental motion [20] and efficient relative pose solvers [21]. However, to the best of our knowledge, the small-rotation approximation has never been exploited for visual-inertial state initialization.

In summary, existing methods either assume known biases, rely on iterative optimization, or adopt hybrid formulations that increase complexity (see Table I). In contrast, we propose a full-state closed-form solution by only assuming small rotations—a reasonable condition for typical hardware and most use cases. Our approach therefore combines the accuracy of optimization-based methods with the simplicity and efficiency of analytical solvers.

III. NOTATION AND PRELIMINARIES

Vectors are denoted by bold lowercase letters ($\mathbf{x}$), with unit vectors indicated by a hat accent ($\hat{\mathbf{x}}$). Bold upper-case letters ($\mathbf{A}$) represent matrices, and light lower-case letters (a) denote scalars. Rotation matrices $\mathbf{R}_{ab} \in \text{SO}(3)$ refer to the transformation from frame b to frame a , omitting the first subindex when referring to the world frame w , i.e., $\mathbf{R}_a \doteq \mathbf{R}_{wa}$. In axis-angle form, $\mathbf{r} = \theta \hat{\mathbf{r}} \in \mathbb{R}^3$, with the unit rotation axis $\hat{\mathbf{r}} = \frac{\mathbf{r}}{\|\mathbf{r}\|} \in \mathbb{S}^2$ and the rotation angle $\theta = \|\mathbf{r}\| \in \mathbb{R}$. The skew-symmetric matrix of $\mathbf{r}$ is denoted as $\mathbf{r}^\wedge$. The IMU integration between time instants t_i and t_j follows the standard formulation [22]

$$\mathbf{R}_j = \mathbf{R}_i \prod_{k=i}^{j-1} \text{Exp} \left((\tilde{\omega}_k - \mathbf{b}_k^g - \boldsymbol{\eta}_k^{gd}) \Delta t \right), \quad (1)$$

$$\mathbf{v}_j = \mathbf{v}_i + \mathbf{g} \Delta t_{ij} + \sum_{k=i}^{j-1} \mathbf{R}_k (\tilde{\mathbf{a}}_k - \mathbf{b}_k^a - \boldsymbol{\eta}_k^{ad}) \Delta t \quad (2)$$

$$\mathbf{t}_j = \mathbf{t}_i + \sum_{k=i}^{j-1} \left[\mathbf{v}_k \Delta t + \frac{1}{2} \mathbf{g} \Delta t^2 + \frac{1}{2} \mathbf{R}_k (\tilde{\mathbf{a}}_k - \mathbf{b}_k^a - \boldsymbol{\eta}_k^{ad}) \Delta t^2 \right] \quad (3)$$

where $\mathbf{v}_i$ and $\mathbf{v}_j$ denote the linear velocity at times t_i and t_j , $\mathbf{g}$ the gravity vector, and Δt the sampling interval between consecutive measurements. $\tilde{\omega}_k$ and $\tilde{\mathbf{a}}_k$ are the gyroscope and accelerometer readings at time k , with associated biases $\mathbf{b}_k^g$ and $\mathbf{b}_k^a$, respectively. We assume that gyroscope and accelerometer biases remain approximately constant during initialization, i.e., $\mathbf{b}_k^g \approx \mathbf{b}^g$ and $\mathbf{b}_k^a \approx \mathbf{b}^a$. We denote the number of preintegrated IMU measurements by $L = j - i - 1$, and hence $t_j - t_i = L \Delta t$.

IV. ROTATION AND GYRO BIAS INITIALIZATION

For a small ($\theta \rightarrow 0$) rotation vector $\mathbf{r}$, the Rodrigues' formula relates it to a rotation matrix representation as

$$\mathbf{R} = \mathbf{I}_3 + \frac{\sin \theta}{\theta} \mathbf{r}^\wedge + \frac{1 - \cos \theta}{\theta^2} \mathbf{r}^{\wedge^2} \underset{\theta \rightarrow 0}{\approx} \mathbf{I}_3 + \mathbf{r}^\wedge, \quad (4)$$

where $\mathbf{I}_3$ is the 3×3 identity matrix. Notably, under this approximation, rotation matrices commute only up to second-order terms. Approximating $\mathbf{R}_{ab} \approx \mathbf{I} + \mathbf{r}_{ab}^\wedge$ and $\mathbf{R}_{cd} \approx \mathbf{I} + \mathbf{r}_{cd}^\wedge$, the products $\mathbf{R}_{ab} \mathbf{R}_{cd}$ and $\mathbf{R}_{cd} \mathbf{R}_{ab}$ then become

$$\mathbf{R}_{ab} \mathbf{R}_{cd} \approx \mathbf{I} + \mathbf{r}_{ab}^\wedge + \mathbf{r}_{cd}^\wedge + \mathbf{r}_{ab}^\wedge \mathbf{r}_{cd}^\wedge \quad (5)$$

$$\mathbf{R}_{cd} \mathbf{R}_{ab} \approx \mathbf{I} + \mathbf{r}_{cd}^\wedge + \mathbf{r}_{ab}^\wedge + \mathbf{r}_{cd}^\wedge \mathbf{r}_{ab}^\wedge \quad (6)$$

differing only in the neglected second-order cross products of the skew-symmetric matrices $\mathbf{r}_{ab}^\wedge \mathbf{r}_{cd}^\wedge$ and $\mathbf{r}_{cd}^\wedge \mathbf{r}_{ab}^\wedge$, which generally do not commute.¹ However, under the small-rotation assumption, these second-order contributions are negligible relative to the first-order terms, making the two matrix products approximately equal.

Rodrigues' formula also provides the exponential map $\exp : \text{so}(3) \rightarrow \text{SO}(3)$. For small rotation vectors $\mathbf{r}_{ab}$ and $\mathbf{r}_{cd}$, the exponential of their sum approximates the product of exponentials:

$$\begin{aligned} \text{Exp}(\mathbf{r}_{ab} + \mathbf{r}_{cd}) &\underset{\theta \rightarrow 0}{\approx} \mathbf{I}_3 + (\mathbf{r}_{ab} + \mathbf{r}_{cd})^\wedge = \mathbf{I}_3 + \mathbf{r}_{ab}^\wedge + \mathbf{r}_{cd}^\wedge \\ &= (\mathbf{I}_3 + \mathbf{r}_{ab}^\wedge) (\mathbf{I}_3 + \mathbf{r}_{cd}^\wedge) = \text{Exp}(\mathbf{r}_{ab}) \text{Exp}(\mathbf{r}_{cd}) \end{aligned} \quad (7)$$

With this, the exponential maps in (1) can be decomposed into products as shown below

$$\prod_{k=i}^{j-1} \text{Exp}((\tilde{\omega}_k - \mathbf{b}^g) \Delta t) \underset{\theta \rightarrow 0}{\approx} \prod_{k=i}^{j-1} \text{Exp}(\tilde{\omega}_k \Delta t) \text{Exp}(-\mathbf{b}^g \Delta t) \quad (8)$$

By first commuting the products of the right part, and then applying (7), we obtain the following expression

$$\mathbf{R}_{ij} = \mathbf{R}_i^\top \mathbf{R}_j \underset{\theta \rightarrow 0}{\approx} \prod_{k=i}^{j-1} (\text{Exp}(\tilde{\omega}_k \Delta t) \text{Exp}(-\mathbf{b}^g \Delta t)) \quad (9)$$

$$\approx \prod_{k=i}^{j-1} (\text{Exp}(\tilde{\omega}_k \Delta t)) \prod_{k=i}^{j-1} (\text{Exp}(-\mathbf{b}^g \Delta t)) \quad (10)$$

$$= \prod_{k=i}^{j-1} (\text{Exp}(\tilde{\omega}_k \Delta t)) \text{Exp}(-L \mathbf{b}^g \Delta t) \quad (11)$$

from which $\mathbf{b}^g$ can be solved analytically. This yields our first closed-form solution, referred to as **commutative approximation**

$\mathbf{b}^g$ using the commutative approximation
$\mathbf{b}^g \approx -\frac{1}{L \Delta t} \text{Log} \left(\left(\prod_{k=i}^{j-1} \text{Exp}(\tilde{\omega}_k \Delta t) \right)^\top \mathbf{R}_{ij} \right) \quad (12)$

We can derive more efficient closed-form solutions for $\mathbf{b}^g$ by assuming constant rotation angles within each IMU integration. First, we express the left-hand side of (9) as the composition of

¹Skew-symmetric matrices actually commute with their adjoints.

L identical rotations:

$$R_{ij} = \prod_{k=i}^{j-1} \text{Exp}\left(\frac{\text{Log}(R_{ij})}{L}\right) \underset{\theta \rightarrow 0}{\approx} \prod_{k=i}^{j-1} (\text{Exp}(\tilde{\omega}_k \Delta t) \text{Exp}(-\mathbf{b}^g \Delta t)) \quad (13)$$

Now, suppose that $\tilde{\omega}_k$ remains approximately constant within the time interval between two images. Approximating $\tilde{\omega}_k$ in (13) by their average $\bar{\omega}_k$ and dividing both sides of the equation by L identical rotations, we get

$$\text{Exp}\left(\frac{\text{Log}(R_{ij})}{L}\right) \approx \text{Exp}(\bar{\omega}_k \Delta t) \text{Exp}(-\mathbf{b}^g \Delta t), \quad (14)$$

from which we can also solve for $\mathbf{b}^g$ analytically. We can compute $\bar{\omega}_k$ by integrating all IMU measurements or by just averaging them arithmetically, which will give us two closed-form solutions that we will denote as **average approximation** and **arithmetic average approximation** respectively.

$\mathbf{b}^g$ using the average approximation

$$\begin{aligned} \mathbf{b}^g &\approx -\frac{1}{\Delta t} \text{Log}\left(\text{Exp}(\bar{\omega}_k \Delta t)^\top \text{Exp}\left(\frac{\text{Log}(R_{ij})}{L}\right)\right) \\ \text{with } \bar{\omega}_k &\approx \frac{1}{L\Delta t} \text{Log}\left(\prod_{k=i}^{j-1} \text{Exp}(\tilde{\omega}_k \Delta t)\right) \end{aligned} \quad (15)$$

$\mathbf{b}^g$ using the arithmetic average approximation

$$\begin{aligned} \mathbf{b}^g &\approx -\frac{1}{\Delta t} \text{Log}\left(\text{Exp}(\bar{\omega}_k \Delta t)^\top \text{Exp}\left(\frac{\text{Log}(R_{ij})}{L}\right)\right) \\ \text{with } \bar{\omega}_k &= \frac{1}{L} \sum_{k=i}^{j-1} \tilde{\omega}_k \end{aligned} \quad (16)$$

R_{ij} can be estimated from feature correspondences between any pair of frames. However, in practice, it is both feasible and convenient to initialize R_{ij} and $\mathbf{b}^g$ using only the first two frames. First, estimation errors tend to grow over time. Inertial drift accumulates, and our approximations become less accurate as more IMU readings are involved. Second, unlike translation estimates, relative rotation estimates do not necessarily benefit from larger inter-frame motions. In fact, large motions often introduce perspective distortions in feature appearance, reducing the number of reliably tracked points and increasing matching noise [23]. Although R_{ij} is obtained from visual correspondences, our approach does not rely on a full SFM-based pipeline. Rotation is estimated independently of the structure using a certifiably optimal two-view solver, thereby avoiding scale ambiguity and depth-related instabilities. This rotation-only formulation leads to a more robust gyroscope bias estimation than multi-view SFM approaches estimating full six-degrees-of-freedom poses and depths.

V. INITIAL VELOCITY, TRANSLATION, GRAVITY AND ACCELEROMETER BIAS

A. Closed-Form Initialization From Linear Point Constraints

Once R_{ij} and $\mathbf{b}^g$ are initialized, the relation between inertial measurements and image tracks can be expressed linearly. Let the sets $\mathcal{Z}_i = \{\mathbf{z}_i^1, \dots, \mathbf{z}_i^n, \dots, \mathbf{z}_i^N\}$ and $\mathcal{Z}_j = \{\mathbf{z}_j^1, \dots, \mathbf{z}_j^n, \dots, \mathbf{z}_j^N\}$ denote the image coordinates of N visual

features extracted in an image at t_i and tracked up to a later image, not necessarily consecutive, taken at t_j . Assuming a pin-hole camera model, every feature $\mathbf{z}_c^n \in \Omega \subset \mathbb{R}^2$ in both images $c \in \{i, j\}$ can be backprojected to a 3D point $\mathbf{p}_c^n \in \mathbb{R}^3$ in the local camera frame as follows

$$\mathbf{p}_c^n = \lambda_c^n \mathbf{K}^{-1} [\mathbf{z}_c^n^\top \ 1]^\top = \lambda_c^n \boldsymbol{\mu}_c^n, \quad (17)$$

where $\mathbf{K} \in \mathbb{UT}^+(3)$ with $\mathbf{K}[3, 3] = 1$ is the camera calibration matrix, Ω is the image domain, $\boldsymbol{\mu}_j^n \in \mathcal{S}^2$ is the normalized feature bearing and $\lambda_j^n \in \mathbb{R}$ its distance along such bearing. Following [7], a linear constraint can be derived for every feature $\mathbf{z}_c^n$ in every pair of frames taken at t_i and t_j

$$-\frac{(L\Delta t)^2}{2} \mathbf{g} - (L\Delta t) \mathbf{v} + \boldsymbol{\Gamma}_j \mathbf{b}_a + \lambda_i^n \boldsymbol{\mu}_i^n - \lambda_j^n \boldsymbol{\mu}_j^n = \mathbf{s}_j, \quad (18)$$

where the vector $\mathbf{s}_j$ and matrix $\boldsymbol{\Gamma}_j$ are computed by integrating the IMU measurements over the interval $[t_i, t_j]$ as

$$\mathbf{s}_j = \int_{t_i}^{t_j} (t_j - \tau) \mathbf{R}_{1\tau} \tilde{\mathbf{a}}_\tau d\tau, \quad \boldsymbol{\Gamma}_j = \int_{t_i}^{t_j} (t_j - \tau) \mathbf{R}_{i\tau} d\tau. \quad (19)$$

The rotation $\mathbf{R}_{i\tau} \in SO(3)$ comes from the IMU preintegration from time t_i up to time τ using (1). Note that $\mathbf{b}^g$ was already initialized from the first two frames. The vector $\tilde{\mathbf{a}}_\tau$ contains the accelerometer readings. By stacking the constraints from the N tracked features from frame i to frame j , we obtain the overdetermined linear system $\boldsymbol{\Xi} \mathbf{x} = \mathbf{s}$, which yields the following least-squares problem

$\mathbf{g}, \mathbf{v}, \mathbf{b}^a$ initialization

$$\arg \min_{\mathbf{x}} (\boldsymbol{\Xi} \mathbf{x} - \mathbf{s})^2 \rightarrow \hat{\mathbf{x}} = \mathbf{v}_{\min} (\boldsymbol{\Xi}^\top \boldsymbol{\Xi}) \quad (20)$$

which we solve via singular value decomposition (SVD), being the optimal solution $\hat{\mathbf{x}}$ the eigenvector associated to the smallest eigenvalue $\mathbf{v}_{\min}(\boldsymbol{\Xi}^\top \boldsymbol{\Xi})$. Each row of the matrix $\boldsymbol{\Xi}$ corresponds to one instance of (18), leading to $3(M-1)N$ equations and $9 + NM$ unknowns, with M the number of images considered. $\boldsymbol{\Xi}$ denotes the coefficients matrix unnumbered equation shown at the bottom of the next page, with $T_j = -\frac{\Delta t^2}{2} \mathbf{I}_3$. The state to be initialized is defined as

$$\mathbf{x} = [\mathbf{g}^\top, \mathbf{v}^\top, \mathbf{b}^a{}^\top, \lambda_1^1, \dots, \lambda_1^N, \dots, \lambda_j^1, \dots, \lambda_j^N]^\top,$$

and $\mathbf{s}$ contains the measurements obtained by (19)

$$\mathbf{s} = [\mathbf{s}_2^\top, \dots, \mathbf{s}_2^\top, \mathbf{s}_3^\top, \dots, \mathbf{s}_3^\top, \mathbf{s}_j^\top, \dots, \mathbf{s}_j^\top]^\top.$$

This yields an initialization method that initializes R_{ij} and $\mathbf{b}^g$ from just the first two frames, and the rest of the states when they become observable, as illustrated in pseudocode by the Algorithm 1. A detailed analysis of the observability criteria follows in the next subsection. Although feature depths appear in the linear formulation, they are treated as nuisance variables and not estimated recursively. Unlike filter-based methods such as OpenVINS [24], depths are implicitly eliminated and never used beyond initialization. Our observability tests assess numerical conditioning of the reduced system, rather than enforcing depth convergence.

Algorithm 1: Visual-Inertial State Initialization.

Require: Visual-inertial data; observability thresholds $\Delta z_{th}, \Delta \rho_{th}$

Ensure: Visual-inertial states $\mathbf{R}_{ij}, \mathbf{b}^g, \mathbf{t}_j, \mathbf{v}_j, \mathbf{b}_a, \mathbf{g}$

```

1:  $\mathcal{Z}_i \leftarrow \text{get\_first\_image\_and\_extract\_features}()$ 
2:  $\mathcal{Z}_j \leftarrow \text{get\_next\_image\_and\_track\_features}(\mathcal{Z}_i)$ 
3:  $\{\tilde{\omega}_k\}, \{\tilde{\mathbf{a}}_k\} \leftarrow \text{get\_IMU\_readings}()$ 
4:  $\mathbf{R}_{ij} \leftarrow \text{get\_rotation}(\mathcal{Z}_i, \mathcal{Z}_j)$ 
5:  $\mathbf{b}^g \leftarrow \text{get\_bias}(\mathbf{R}_{ij})$   $\triangleright$  Eqs. (12), (15) or (16)
6:  $\text{full\_obs} \leftarrow \text{false}$ 
7: while  $\text{full\_obs} = \text{false}$  do
8:    $\text{t\_obs} \leftarrow \text{translation\_obs\_test}(\mathcal{Z}_i, \mathcal{Z}_j, \mathbf{R}_{ij}, \bar{\Delta z}_{th})$   $\triangleright$  (23)
9:   if  $\text{t\_obs} = \text{true}$  then
10:      $\mathbf{s}_j \leftarrow \text{integrate\_IMU}(\{\tilde{\omega}_k\}, \{\tilde{\mathbf{a}}_k\}, \mathbf{b}^g)$ 
11:      $\text{full\_obs} \leftarrow \text{full\_obs\_test}(\mathcal{Z}_i, \mathcal{Z}_j, \mathbf{s}_j, \Delta \rho_{th})$   $\triangleright$  (25)
12:   if  $\text{full\_obs} = \text{true}$  then
13:      $\mathbf{t}_j, \mathbf{v}_j, \mathbf{b}_a, \mathbf{g} \leftarrow \text{get\_rest\_of\_vi\_states}(\mathbf{s}_j)$   $\triangleright$  (20)
14:      $\text{full\_obs} \leftarrow \text{true}$ 
15:   end if
16: end if
17:  $\mathcal{Z}_j \leftarrow \text{get\_next\_image\_and\_track\_features}(\mathcal{Z}_i)$ 
18:  $\{\tilde{\omega}_k\}, \{\tilde{\mathbf{a}}_k\} \leftarrow \text{get\_IMU\_readings}()$ 
19:  $\mathbf{R}_{ij} \leftarrow \text{get\_rotation}(\mathcal{Z}_i, \mathcal{Z}_j)$ 
20: end while

```

B. Two-Stages Observability Test

While camera rotation and gyroscope bias are observable from any two frames, and in practice we will initialize them from the first two in the sequence, the remaining visual-inertial states are not. Specifically, using the linear system defined in (20), the full state will only be observable and hence we will be able to initialize it when Ξ is full rank. Assessing this criterion at each frame, however, may be expensive, as the matrix Ξ grows quickly with the number of frames. Motivated by this, we will evaluate observability in two stages, that are described next.

Translation Observability Test (Stage 1): This stage assesses the average parallax in the image plane between the *first* frame at t_i and a *single* subsequent frame at t_j , advancing progressively over the image sequence until the criterion is met. Specifically, we first backproject each feature $\mathbf{z}_i^n$ extracted at t_i and project it at t_j assuming rotation-only motion between the two camera frames

$$\hat{\mathbf{z}}_j^n = h \left(\mathbf{K} \mathbf{R}_{ij}^\top \mathbf{K}^{-1} \begin{bmatrix} \mathbf{z}_i^n^\top & 1 \end{bmatrix}^\top \right), \quad (21)$$

where $h(\cdot)$ transforms the image points from homogeneous to Euclidean coordinates. The parallax between t_i and t_j will be reflected in the difference between the tracks predicted with just rotational motion $\hat{\mathbf{z}}_j^n$ and the real tracks $\mathbf{z}_j^n$. We aggregate the parallax of all features using their average:

$$\bar{\Delta z}_j = \frac{1}{N} \sum_{n=1}^N \|\mathbf{z}_j^n - \hat{\mathbf{z}}_j^n\|_2, \quad (22)$$

The condition to be met is expressed as

$$\bar{\Delta z}_j > \bar{\Delta z}_{th}, \quad (\text{Translation Observability Test}) \quad (23)$$

If the condition is not met, we move to the next frame and repeat the test until the threshold is satisfied.

Full Observability Test (Stage 2): This stage is triggered once the motion provides sufficient translational information, i.e., when the condition in (23) holds. Following [25], we assess numerical observability via the Hessian $\mathbf{H}$, as follows

$$\mathbf{H} = \Xi^\top \Sigma^{-1} \Xi, \quad (24)$$

with Σ the residual weights obtained following [26]. Using the full $\mathbf{H}$ is misleading, as nuisance depths λ_j^n , are often ill-conditioned under low parallax, and prone to depressing some singular values. Hence, we marginalize depths via the Schur complement to obtain $\mathbf{H}^*$ over the nine states to be estimated (gravity, velocity, accelerometer bias) and evaluate the observability on this reduced matrix.

We then compute the SVD of $\mathbf{H}^* \in \mathbb{S}_+^9$ and define the singular-value ratio $\rho_j = \sigma_{\max}/\sigma_{\min}$, where $\sigma_{\min}$ and $\sigma_{\max}$ denote the smallest nonzero and largest singular values, respectively. To track stabilization as the data window grows, we monitor the relative change

$$\Delta \rho_j = |(\rho_j - \rho_{j-1})/\rho_{j-1}|, \quad (25)$$

where large values indicate that observability is still changing, whereas consistently small values indicate that the Hessian has stabilized and the state is observable. Once

$$\Delta \rho_j < \Delta \rho_{th}, \quad (\text{Full Observability Test}) \quad (26)$$

with $\Delta \rho_{th}$ a fixed threshold, we deem the system observable and proceed to solve (20). Fig. 1 reports the relative overcost of *Full Observability Test* with respect to *Translation Observability Test*. The overcost of *Full Observability Test* rises steadily with the window size: starting at roughly $\times 3$ – 4 the cost of *Translation Observability Test* for short windows, but exceeding $\times 15$ around ten frames. These results motivate the use of the parallax gate as the primary trigger and reserving the Hessian test as a final verifier.

$$\Xi = \begin{bmatrix} T_2 & -t_2 \mathbf{I}_3 & \Gamma_2 & \mu_1^1 & \mathbf{0}_3 & \mathbf{0}_3 & -\mu_2^1 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ T_2 & -t_2 \mathbf{I}_3 & \Gamma_2 & \mathbf{0}_3 & \mu_1^2 & \mathbf{0}_3 & \mathbf{0}_3 & -\mu_2^2 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ T_2 & -t_2 \mathbf{I}_3 & \Gamma_2 & \mathbf{0}_3 & \mathbf{0}_3 & \mu_1^N & \mathbf{0}_3 & \mathbf{0}_3 & -\mu_2^N & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ T_j & -t_j \mathbf{I}_3 & \Gamma_j & \mu_1^1 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & -\mu_j^1 & \mathbf{0}_3 & \mathbf{0}_3 \\ T_j & -t_j \mathbf{I}_3 & \Gamma_j & \mathbf{0}_3 & \mu_1^2 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & -\mu_j^2 & \mathbf{0}_3 \\ T_j & -t_j \mathbf{I}_3 & \Gamma_j & \mathbf{0}_3 & \mathbf{0}_3 & \mu_1^N & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & -\mu_j^N \end{bmatrix},$$

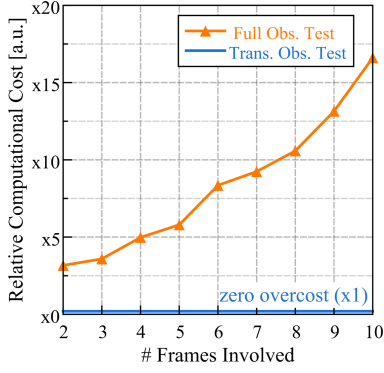


Fig. 1. Relative computational costs of our observability tests. Note the significantly higher cost of the full observability test, which motivates our proposal of a first stage assessing translational observability.

VI. EXPERIMENTS

A. Experimental Setup

All experiments were conducted on a desktop PC equipped with an Intel Core i7-11700 K (3.6 GHz, 64 GB RAM). Our method is implemented in C++, and nonlinear least-squares optimization is performed using Ceres [27]. Unless otherwise specified, all methods are executed on a single thread to ensure fair timing comparisons, and we report wall-clock times. Performance is evaluated in terms of the root mean squared error (RMSE), expressed in the corresponding physical units. In line with previous research, we use the publicly available EuRoC MAV benchmark [28], which provides a diverse set of motion profiles commonly used to benchmark visual-inertial initialization methods, and is therefore adopted as our primary real-world evaluation dataset. EuRoC provides synchronized 752×480 stereo at 20 Hz, inertial data at 200 Hz and accurate ground truth. The dataset spans a range of motion profiles (easy/medium/difficult) and exhibits variations in texture, illumination, and dynamics, making it a suitable testbed for evaluating initialization accuracy, robustness, and runtime under realistic conditions. Following previous works [11], [14], we perform exhaustive initialization every 0.5 s (i.e., every 10 frames) across all sequences, yielding a total of 2248 independent initialization tests.

B. Closed-Form Gyroscope Bias

Synthetic data: We conduct synthetic experiments in which the ground-truth relative rotations R_{ij} from EuRoC are progressively corrupted with zero-mean Gaussian noise, allowing us to isolate the effect of rotational uncertainty on the estimation accuracy of b^g . We report de-aggregated per-sequence metrics to assess generalization over different motion patterns. We compare the three closed-form approximation methods introduced in Section IV with an iterative nonlinear least-squares optimization of b^g based on (1). All methods use the same measurements and relative rotations (estimated by C2P [29]) within identical time windows selected by our observability-driven triggering strategy. As shown in Table II, in an ideal noise-free setup, nonlinear optimization yields the lowest errors, while our closed-form solutions perform only slightly worse. However, as the noise level increases, the errors of our closed-form methods and the

TABLE II
GYROSCOPE BIAS ERRORS [rad/s] IN EUROC, FOR DIFFERENT NOISE SETUPS FOR THE RELATIVE ROTATION R_{ij} . (A) NO NOISE, (B) STANDARD DEVIATION OF 0.006 deg, (C) STANDARD DEVIATION OF 0.06 deg, AND (D) NOISE FROM IMAGE MATCHES. WE COMPARE FOUR METHODS: NONLINEAR OPTIMIZATION (NONLIN. OPT.), COMMUTATIVE APPROXIMATION (COMMUTATIVE.), AVERAGE APPROXIMATION (AVG.) AND ARITHMETIC AVERAGE APPROXIMATION (ARITHM. AVG.). BEST PER ROW IN **BOLD**. EXCEPT FOR THE UNREALISTIC NOISELESS CASE, ALL METHODS BRING SIMILAR PERFORMANCE.

Seq.	Nonlin. Opt.	Closed-form solutions (ours)		
		Commutative	Avg.	Arithm. Avg.
(a) Noise-free relative rotation				
MH01	0.000002	0.000299	0.000042	0.000315
MH02	0.000003	0.000303	0.000043	0.000329
MH03	0.000004	0.000444	0.000055	0.000486
MH04	0.000003	0.000327	0.000045	0.000338
MH05	0.000003	0.000275	0.000042	0.000283
V101	0.000003	0.000446	0.000052	0.000453
V102	0.000004	0.000947	0.000104	0.000965
V103	0.000007	0.001099	0.000123	0.001157
V201	0.000004	0.000439	0.000059	0.000463
V202	0.000004	0.000985	0.000109	0.001049
V203	0.000008	0.001384	0.000166	0.001552
Mean	0.000004	0.000631	0.000076	0.000672
(b) Relative rotation noise of 0.006 deg standard deviation				
MH01	0.004148	0.004166	0.004149	0.004182
MH02	0.004144	0.004144	0.004144	0.004185
MH03	0.004148	0.004169	0.004148	0.004189
MH04	0.004148	0.004147	0.004145	0.004145
MH05	0.004148	0.004161	0.004147	0.004170
V101	0.004148	0.004181	0.004147	0.004183
V102	0.004147	0.004350	0.004158	0.004353
V103	0.004147	0.004422	0.004163	0.004430
V201	0.004148	0.004181	0.004148	0.004180
V202	0.004149	0.004233	0.004144	0.004208
V203	0.003711	0.004116	0.003729	0.004132
Mean	0.004108	0.004206	0.004111	0.004214
(c) Relative rotation noise of 0.06 deg standard deviation				
MH01	0.041481	0.041484	0.041482	0.041498
MH02	0.041481	0.041462	0.041476	0.041502
MH03	0.041481	0.041470	0.041481	0.041471
MH04	0.041481	0.041462	0.041478	0.041458
MH05	0.041481	0.041480	0.041479	0.041486
V101	0.041481	0.041484	0.041480	0.041484
V102	0.041481	0.041571	0.041490	0.041561
V103	0.041481	0.041605	0.041495	0.041581
V201	0.041481	0.041484	0.041481	0.041478
V202	0.041483	0.041432	0.041477	0.041377
V203	0.037116	0.037212	0.037129	0.037109
Mean	0.041084	0.041104	0.041086	0.041091
(d) Relative rotation noise from image matches				
MH01	0.030048	0.030020	0.030046	0.030028
MH02	0.033075	0.033052	0.033072	0.033064
MH03	0.029123	0.029152	0.029126	0.029161
MH04	0.045381	0.045410	0.045384	0.045392
MH05	0.038635	0.038620	0.038633	0.038622
V101	0.044986	0.044941	0.044980	0.044946
V102	0.090967	0.090799	0.090948	0.090825
V103	0.351947	0.351923	0.351939	0.351934
V201	0.048773	0.048756	0.048769	0.048765
V202	0.057977	0.058014	0.057978	0.058041
V203	0.122305	0.122289	0.122303	0.122330
Mean	0.081202	0.081180	0.081198	0.081192

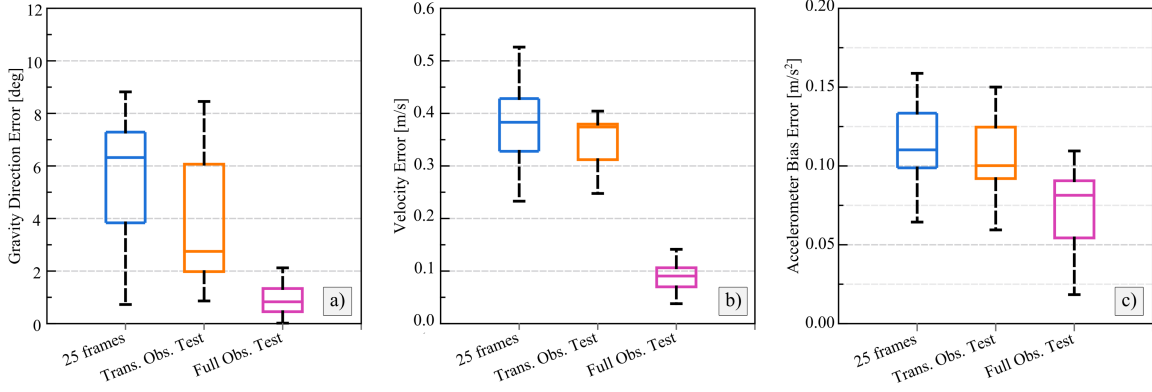


Fig. 2. Initialization accuracy vs. activation policy. We trigger closed-form initialization (i) after a fixed 25-frame window (≈ 1.25 s); (ii) when *Translation Observability Test* is satisfied; and (iii) when *Full Observability Test* is satisfied. Panels report RMSE distributions in (a) gravity direction, (b) velocity, and (c) accelerometer bias. The *Translation Observability Test* already improves initialization accuracy, while the *Full Observability Test* yields the best overall performance.

TABLE III
GYROSCOPE BIAS INITIALIZATION RUNTIME. OUR APPROXIMATIONS ACHIEVE UP TO TWO ORDERS OF MAGNITUDE SPEED-UP COMPARED TO THE NONLINEAR OPTIMIZATION BASELINE.

Method	Mean [μ s]	Std dev. [μ s]
Nonlinear optimization	3687	275
Commutative approximation (ours)	94	6
Avg. approximation (ours)	25	1
Arithm. Avg approximation (ours)	19	1

nonlinear baseline rapidly converge toward each other. Even for unrealistically small rotation errors of standard deviation 0.006 deg, the difference in $\mathbf{b}^g$ error is already below 3 %. For larger perturbations, the differences become negligible. These results indicate that, in practical scenarios with noisy relative rotation estimations, our closed-form solutions achieve a $\mathbf{b}^g$ estimation accuracy comparable to that of nonlinear optimization. Note also that the differences among our approximations are negligible.

Real data: In this experiment, relative rotations $\mathbf{R}_{ij}$ are computed from image correspondences using C2P [29], with outliers removed via RANSAC. This setup emulates realistic conditions in which rotational inputs are derived from real visual data and are therefore affected by noise and matching inaccuracies. Across all initialization experiments, the average outlier ratio was approximately 0.15.

The last rows in Table II summarize the average gyroscope bias estimation errors across all EuRoC sequences, for our closed-form solutions and nonlinear optimization. Table III reports the mean and standard deviation of the execution time for gyroscope bias estimation under the four aforementioned strategies. Relative to the nonlinear baseline, the *Commutative* approximation reduces runtime by a factor 39 $\times$, the *Avg.* approximation by 147 $\times$, and the *Arithm. Avg.* approximation by 194 $\times$. This validates the computational advantage of our approximations. Table III reports isolated gyroscope bias estimation runtimes, assuming that relative rotations $\mathbf{R}_{ij}$ are already available. The computation of $\mathbf{R}_{ij}$ is included in the total initialization time reported in Table V.

As a summary, note how under realistic conditions all our methods achieve comparable accuracy, with noticeable differences only in the unrealistic noise-free case. The errors coming from feature matching are significantly higher than that of our approximations, which are negligible in practice. The accuracy

of our three proposed approximations is similar, while *Arithm. Avg.* is the most efficient, and hence should be the preferred one for implementation.

C. Full State Initialization

Evaluation of Triggering strategies: To validate our observability-driven initialization, we compared three different activation strategies: (i) fixed-time policy after 25 frames (≈ 1.25 s), (ii) activation when the first observability criteria (*Translation Observability Test*) is satisfied, and (iii) activation when the second observability criteria (*Full Observability Test*) is satisfied. Fig. 2 reports the results of this experiment and highlights the advantages of our two-stage strategy over a naive fixed-window policy. Initializing after 25 frames yields noticeably larger and more dispersed errors, particularly in the gravity estimate, as no guarantees exist on the state observability in the chosen interval. By contrast, activating the closed-form full solution at the first *Translation Observability Test* already achieves a clear reduction in error. And even greater improvements are obtained when monitoring the Hessian and postponing activation until the *Full Observability Test* is satisfied, which consistently yields the lowest errors across all estimated state variables. These results confirm that our observability-driven criterion not only enables faster initialization than fixed-time approaches, but also provides more accurate and reliable estimates than a fixed-window strategy. The accelerometer bias is difficult to decouple from gravity under limited excitation. Accordingly, (18) is solved only after the Full Observability Test indicates a well-conditioned Hessian, as premature solving leads to biased estimates (See Fig. 2). Jointly estimating velocity, gravity, and $\mathbf{b}_a$ in a single linear system provides additional numerical robustness without relying on 3D structure.

Comparison against baselines: We compare our initializer against three representative baselines: the tightly-coupled version of DRT (denoted as DRT-tight) [14], the initialization module of VINS-Mono (denoted as VINS-Mono) [1], and the analytical-solution method [30], which extends the ORB-SLAM3 initialization [11] (denoted as AS-MLE). Table IV reports per-sequence initialization errors on the EuRoC dataset. Averaged across all sequences, our method achieves the lowest velocity RMSE (0.09 m/s) and gravity-direction error (0.90 deg), outperforming DRT-tight (0.11 m/s, 1.03 deg), VINS-Mono (0.11 m/s, 1.47 deg), and AS-MLE (0.17 m/s, 2.66 deg). Per-sequence results show that our initializer is best

TABLE IV

PER-SEQUENCE INITIALIZATION ERRORS ON EUROC (ATTEMPTS EVERY 0.5 s) FOR AS-MLE, VINS-MONO, DRT-TIGHT, AND OUR METHOD. BEST PER ROW IS SHOWN IN **BOLD**, SECOND-BEST IS UNDERLINED.

Seq.	Velocity Error [m/s]				Gravity Error [deg]				Accel. Bias Error [m/s ²]				Gyro. Bias Error [rad/s]			
	AS-MLE	VINS-Mono	DRT-tight	Ours	AS-MLE	VINS-Mono	DRT-tight	Ours	AS-MLE	VINS-Mono	DRT-tight	Ours	AS-MLE	VINS-Mono	DRT-tight	Ours
MH01	0.11	0.08	0.11	<u>0.10</u>	1.64	1.17	<u>1.00</u>	0.87	0.093	<u>0.092</u>	–	0.088	0.091	<u>0.070</u>	–	0.030
MH02	0.09	0.08	0.11	0.09	1.36	1.11	<u>0.97</u>	0.65	<u>0.092</u>	0.093	–	0.091	0.090	<u>0.080</u>	–	0.033
MH03	0.30	0.19	<u>0.14</u>	0.09	3.06	<u>1.57</u>	0.96	0.96	<u>0.080</u>	<u>0.081</u>	–	0.080	0.063	<u>0.051</u>	–	0.029
MH04	0.25	<u>0.14</u>	0.19	0.11	1.87	1.23	<u>0.99</u>	0.84	<u>0.087</u>	0.088	–	0.086	0.068	<u>0.060</u>	–	0.045
MH05	0.25	<u>0.17</u>	0.22	0.13	2.61	1.42	<u>0.99</u>	0.83	<u>0.084</u>	0.081	–	0.081	0.075	<u>0.062</u>	–	0.039
V101	0.07	0.06	0.07	0.07	3.37	–	<u>0.99</u>	0.89	<u>0.316</u>	0.319	–	0.128	0.082	–	–	0.045
V102	0.24	0.11	<u>0.06</u>	0.05	5.57	2.60	0.79	<u>1.09</u>	0.082	<u>0.081</u>	–	0.078	0.091	0.077	–	0.091
V103	–	–	0.11	<u>0.12</u>	–	–	<u>1.48</u>	0.87	<u>0.115</u>	0.116	–	0.094	0.352	0.349	–	0.352
V201	0.08	<u>0.07</u>	0.08	0.06	1.79	1.43	<u>1.15</u>	1.06	0.085	<u>0.083</u>	–	0.082	0.067	0.049	–	0.049
V202	0.13	0.08	0.06	<u>0.07</u>	2.68	1.26	0.84	<u>0.94</u>	<u>0.055</u>	<u>0.056</u>	–	0.052	0.068	<u>0.067</u>	–	0.058
V203	–	–	<u>0.09</u>	0.08	–	–	<u>1.17</u>	0.92	0.053	0.053	–	0.062	0.151	0.141	–	0.122
Mean	0.17	0.11	0.11	0.09	2.66	1.47	<u>1.03</u>	0.90	0.104	0.104	–	0.084	0.109	<u>0.101</u>	–	0.081

TABLE V

AVERAGE INITIALIZATION COST [MS] ON EUROC. RUNTIME BREAKDOWN AND COMPARISON AGAINST BASELINES.

Module	AS-MLE	VINS-Mono	DRT-tight	Ours
SfM	30.30	30.35	–	–
$\mathbf{b}^g$ Est.	0.15	0.44	1.95	0.02
Vel. & Grav. Est.	0.08	0.14	2.81	0.52
Point Tri.	0.01	0.01	0.42	0.42
Total	30.54	30.94	5.18	0.96

TABLE VI

INITIALIZATION WINDOW LENGTH AND ADAPTIVITY. WHEREAS BASELINES RELY ON FIXED/LONG WINDOWS, OUR TWO-STAGE CRITERION REDUCES INITIALIZATION DELAYS 4 $\times$.

Method	Initialization time [s]	Adaptive?
Campos et al. [10]	2.16 $\pm$ 0.01	✓
He et al. [14]	2.50	✗
Ours	0.48 $\pm$ 0.27	✓

or second-best in the majority of cases for both velocity and gravity, with only a few isolated sequences where DRT-tight remains competitive (e.g., gravity estimation in V102). In addition to velocity and gravity, Table IV explicitly reports inertial bias errors during initialization. Our method consistently achieves the lowest gyroscope bias error across all sequences, with values in the order of 10^{-2} rad/s, significantly improving over AS-MLE and VINS-Mono. This highlights the benefit of explicitly estimating and compensating the gyroscope bias *prior* to solving the closed-form initialization system, leading to more accurate relative rotations. The improved rotation estimates directly translate into more consistent gravity alignment and lower velocity errors, as observed in Table IV. For the accelerometer bias, our approach also yields the lowest average RMSE (0.084 m/s²), indicating that bias effects are reliably captured during initialization. We note that DRT-tight does not explicitly model inertial biases during its initialization stage, and therefore accelerometer and gyroscope bias errors are not reported for this method. Unlike some initialization methods, we do not enforce a fixed gravity magnitude during initialization. Instead, translation-related states are estimated only after sufficient observability is detected, ensuring a well-conditioned linear system. As shown in Fig. 2 and Table IV, this observability-aware strategy yields lower gravity-direction and velocity errors than methods enforcing gravity magnitude prematurely.

Runtime: We evaluate our overall initialization runtime against representative baselines on the EuRoC dataset. In Table V the full state initialization time is broken down by modules. Our method does not require 3D structure during initialization; therefore, we only include the cost of gyroscope bias estimation and the velocity–gravity solver. For a fair comparison with methods that rely on structure, we additionally report an

optional “Point Tri.” stage executed after initialization. Our total initialization (0.96 ms) is 31.8 $\times$ lower than AS-MLE, 32.2 $\times$ lower than VINS-Mono, and 5.4 $\times$ lower than DRT-tight. At the module level, gyroscope-bias estimation is reduced by a factor 7.5 $\times$ compared to AS-MLE, a factor 22 $\times$ compared to VINS-Mono, and a factor compared to DRT-tight. Our closed-form velocity–gravity solver is 5.4 $\times$ faster than DRT-tight, while being 3.7–6.5 $\times$ slower than AS-MLE/VINS-Mono (which leverage prebuilt structure). Notably, our pipeline avoids the SfM stage ($\approx$ 30 ms in AS-MLE/VINS-Mono), which dominates their cost; if reporting *core-only* initialization (bias + vel. grav.), our time is 0.54 ms, i.e., 56.6 $\times$ lower than AS-MLE’s.

Initialization delay: Table VI compares the initialization windows used by the baselines and our method. Prior approaches require temporal windows of roughly two seconds on average, either fixed or adaptively selected. In contrast, our two-stage criterion achieves reliable initialization with an adaptive window averaging only half a second.

Approximation Robustness Analysis: To quantify the validity range of the three approximations, we plot in Fig. 3 the gyroscope bias estimation error as a function of the accumulated rotation magnitude. For each valid sample, we compare the resulting bias error for: (i) the commutative approximation, (ii) the average approximation, (iii) the arithmetic average approximation and (ii) the nonlinear optimization. Across 2698 valid samples, no sharp degradation is observed up to 1.5–2 deg, where our approximations remains very accurate. Beyond this threshold, the error increases and the gap to nonlinear optimization becomes persistent. In contrast, nonlinear optimization remains stable, with a nearly flat trend and lower error. It should be noted, however, that only 15 % of the consecutive pairs evaluated in the EuRoC dataset have relative rotations greater than 2 deg. This observation supports the practical validity of

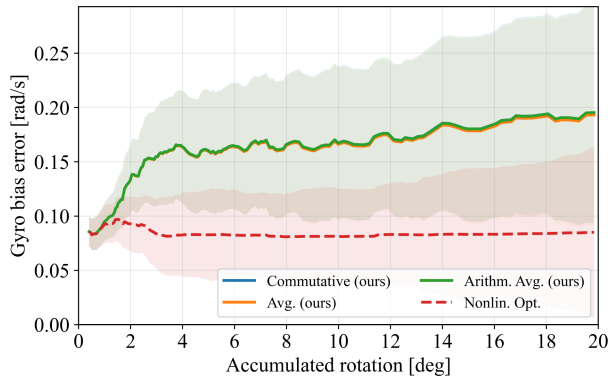


Fig. 3. Gyroscope-bias error vs. accumulated rotation on EuRoC. Our three approximation variants (solid blue/green/orange curves) produce nearly identical results and appear superimposed. In contrast, the nonlinear optimization (dashed red curve) remains comparatively stable up to 20 deg, while approximation-based estimates exhibit increasing error and variability as rotation grows.

our approximations and is consistent with the quantitative results reported in Table II and Table IV.

VII. CONCLUSION

In this work, we introduced a closed-form initialization that recovers the full visual-inertial state without iterative solvers. By leveraging small-rotation and constant-velocity approximations, our method preserves the essential couplings between motion and inertial measurements while remaining analytically compact. A central ingredient is our fast approximation for gyroscope-bias estimation, which replaces costly optimization with a lightweight solver and enables immediate and robust start-up. Across all EuRoC sequences, our method reduces the initialization errors of relevant baselines by 10–20%, reduces initialization windows by $4\times$ and computation times by $5\times$. Importantly, our solutions are also algebraically simpler. Because the proposed approach is designed for short time windows and moderate motion, its underlying approximations may break down under highly dynamic conditions. In such occasional cases, slower nonlinear optimization-based methods may be more suitable. Nevertheless, we believe our closed-form perspective offers a practical alternative for resource-constrained platforms and a strong foundation for future extensions within modern visual-inertial pipelines.

REFERENCES

- [1] T. Qin, P. Li, and S. Shen, “VINS-Mono: A robust and versatile monocular visual-inertial state estimator,” *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, Aug. 2018.
- [2] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel, and J. D. Tardós, “ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM,” *IEEE Trans. Robot.*, vol. 37, no. 6, pp. 1874–1890, Dec. 2021.
- [3] S. Boche, J. Jung, S. B. Laina, and S. Leutenegger, “OKVIS2-X: Open keyframe-based visual-inertial SLAM configurable with dense depth or LiDAR, and GNSS,” *IEEE Trans. Robot.*, vol. 41, pp. 6064–6083, 2025.
- [4] L. Carlone and A. Censi, “From angular manifolds to the integer lattice: Guaranteed orientation estimation with application to pose graph optimization,” *IEEE Trans. Robot.*, vol. 30, no. 2, pp. 475–492, Apr. 2014.
- [5] L. Carlone, R. Tron, K. Daniilidis, and F. Dellaert, “Initialization techniques for 3D SLAM: A survey on rotation estimation and its use in pose graph optimization,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2015, pp. 4597–4604.
- [6] A. Martinelli, “Vision and IMU data fusion: Closed-form solutions for attitude, speed, absolute scale, and bias determination,” *IEEE Trans. Robot.*, vol. 28, no. 1, pp. 44–60, Feb. 2012.
- [7] A. Martinelli, “Closed-form solution of visual-inertial structure from motion,” *Int. J. Comput. Vis.*, vol. 106, no. 2, pp. 138–152, 2014.
- [8] J. Kaiser, A. Martinelli, F. Fontana, and D. Scaramuzza, “Simultaneous state initialization and gyroscope bias calibration in visual inertial aided navigation,” *IEEE Robot. Automat. Lett.*, vol. 2, no. 1, pp. 18–25, Jan. 2017.
- [9] J. Domínguez-Conti, J. Yin, Y. Alami, and J. Civera, “Visual-inertial SLAM initialization: A general linear formulation and a gravity-observing non-linear optimization,” in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2018, pp. 37–45.
- [10] C. Campos, J. M. Montiel, and J. D. Tardós, “Fast and robust initialization for visual-inertial SLAM,” in *Proc. Int. Conf. Robot. Automat.*, 2019, pp. 1288–1294.
- [11] C. Campos, J. M. Montiel, and J. D. Tardós, “Inertial-only optimization for visual-inertial initialization,” in *Proc. Int. Conf. Robot. Automat.*, 2020, pp. 51–57.
- [12] S. H. Lee and J. Civera, “Rotation-only bundle adjustment,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 424–433.
- [13] A. Concha, M. Burri, J. Briales, C. Forster, and L. Oth, “Instant visual odometry initialization for mobile AR,” *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 11, pp. 4226–4235, Nov. 2021.
- [14] Y. He, B. Xu, Z. Ouyang, and H. Li, “A rotation-translation-decoupled solution for robust and efficient visual-inertial initialization,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 739–748.
- [15] R. Wang, X. Qian, X. Gao, and L. Chen, “EDI: An error-state Kalman filter based disjoint initialization framework for visual-inertial SLAM,” *IEEE Robot. Automat. Lett.*, vol. 8, no. 9, pp. 5905–5912, 2023.
- [16] G. D. Evangelidis and B. Micušík, “Revisiting visual-inertial structure-from-motion for odometry and SLAM initialization,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 1415–1422.
- [17] L. Kneip and S. Lynen, “Direct optimization of frame-to-frame rotation,” in *Proc. Int. Conf. Comput. Vis.*, 2013, pp. 2352–2359.
- [18] J. Oliensis, “Exact two-image structure from motion,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 12, pp. 1618–1633, Dec. 2002.
- [19] S. H. Lee and J. Civera, “Closed-form optimal two-view triangulation based on angular errors,” in *Proc. Int. Conf. Comput. Vis.*, 2019, pp. 2681–2689.
- [20] F. Yu and D. Gallup, “3D reconstruction from accidental motion,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3986–3993.
- [21] J. Ventura, C. Arth, and V. Lepetit, “Approximated relative pose solvers for efficient camera motion estimation,” in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 180–193.
- [22] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, “On-manifold preintegration for real-time visual-inertial odometry,” *IEEE Trans. Robot.*, vol. 33, no. 1, pp. 1–21, Feb. 2017.
- [23] A. Fontan, L. Oliva, J. Civera, and R. Triebel, “Model for multi-view residual covariances based on perspective deformation,” *IEEE Robot. Automat. Lett.*, vol. 7, no. 2, pp. 1960–1967, Apr. 2022.
- [24] P. Geneva, K. Eickenhoff, W. Lee, Y. Yang, and G. Huang, “OpenVINS: A research platform for visual-inertial estimation,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 4666–4672.
- [25] S. Cerezo and J. Civera, “GNSS-Inertial state initialization using inter-Epoch baseline residuals,” *IEEE Robot. Automat. Lett.*, vol. 11, no. 2, pp. 1458–1465, Feb. 2026.
- [26] S. Cerezo and J. Civera, “Camera motion estimation from RGB-D-inertial scene flow,” in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshop*, 2024, pp. 841–849.
- [27] S. Agarwal, K. Mierle, and T. C. S. Team, “Ceres solver,” 2023. [Online]. Available: <https://github.com/ceres-solver/ceres-solver>
- [28] M. Burri et al., “The EuRoC micro aerial vehicle datasets,” *Int. J. Robot. Res.*, vol. 35, 2016, pp. 1157–1163.
- [29] J. Tirado-Garní and J. Civera, “From correspondences to pose: Non-minimal certifiably optimal relative pose without disambiguation,” *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 403–412.
- [30] D. Zuniga-Noël, F.-A. Moreno, and J. Gonzalez-Jimenez, “An analytical solution to the IMU initialization problem for visual-inertial systems,” *IEEE Robot. Automat. Lett.*, vol. 6, no. 3, pp. 6116–6122, Jul. 2021.