

An Analytical Solution to the IMU Initialization Problem for Visual-Inertial Systems

David Zuñiga-Noël, Francisco-Angel Moreno, Javier Gonzalez-Jimenez

Abstract—The fusion of visual and inertial measurements is becoming more and more popular in the robotics community since both sources of information complement well each other. However, in order to perform this fusion, the biases of the Inertial Measurement Unit (IMU) as well as the direction of gravity must be initialized first. Additionally, in case of a monocular camera, the metric scale is also needed. The most popular visual-inertial initialization approaches rely on accurate vision-only motion estimates to build a non-linear optimization problem that solves for these parameters in an iterative way. In this paper, we rely on the previous work in [1] and propose an analytical solution to estimate the accelerometer bias, the direction of gravity and the scale factor in a maximum-likelihood framework. This formulation results in a very efficient estimation approach and, due to the non-iterative nature of the solution, avoids the intrinsic issues of previous iterative solutions. We present an extensive validation of the proposed IMU initialization approach and a performance comparison against the state-of-the-art approach described in [2] with real data from the publicly available EuRoC dataset, achieving comparable accuracy at a fraction of its computational cost and without requiring an initial guess for the scale factor. We also provide a C++ open source reference implementation.

I. INTRODUCTION

Visual-Inertial (VI) systems are ubiquitous in robotics nowadays, mainly because visual and inertial sensors are usually inexpensive, have low-power requirements, and provide information complementary to each other. For instance, a monocular camera is an exteroceptive sensor that allows recovering the geometry of the environment up to a scale factor, while an Inertial Measurement Unit (IMU) is a proprioceptive sensor that allows rendering the metric scale observable and provides robust frame-to-frame motion priors. Consequently, VI systems have become minimal set-ups for metric visual SLAM and thus have found applications ranging from AR/VR to autonomous navigation, both aerial and ground [3].

However, before visual and inertial measurements can be fused, some parameters of the IMU need to be initialized, namely: the gravity direction, and the gyroscope and accelerometer biases. Not only that, these biases slowly change over time and the gravity direction depends on the initial pose, so they have to be estimated online. Additionally, knowing the metric scale is also required for VI fusion in case of a monocular system. The initialization of these parameters is critical for the performance of the VI system, and has to be fast and accurate, otherwise the fusion will be delayed and/or convergence issues might appear [2].

Two main approaches to VI initialization can be found in the literature: (i) solving jointly the visual Structure-from-Motion (SfM) problem and the IMU initialization [4], [5] or

(ii) first addressing SfM problem only and then initializing the IMU from the estimate motion [1], [6]. The former assumes that the camera poses can be approximated from inertial measurements and requires the same visual features to be detected across all frames used for initialization. This hinders the practical application of the method, which additionally suffers from numerical stability problems. On the other hand, the latter assumes that the vision-only motion estimation is very accurate so that the IMU parameters can be estimated from them. In case the reader is wondering, this is actually the case for state-of-the-art monocular SLAM systems, such as ORB_SLAM [7], which achieves a localization accuracy in the order of a few centimeters for room-scaled environments. Following this direction, Campos *et al.* [2] recently proposed to solve the involved non-linear optimization problem in a maximum-a-posteriori (MAP) framework, resulting in an accurate and efficient approach. However, due to the iterative nature of the solution, a good initial guess for all parameters is required beforehand. This can be done for the direction of gravity, and the IMU biases can be safely initialized to zero, but estimating the scale factor remains a problem since it cannot be known *a priori* for a monocular vision system.

In this paper, we rely on the previous work of Mur-Artal and Tardós [1] and propose a new solution to the visual-inertial initialization problem within a maximum likelihood estimation (MLE) framework. The bias for the gyroscope is estimated iteratively, as in [1], but our main contribution in this paper lies in an analytic solution for the rest of the parameters (i.e. scale, gravity direction and accelerometer bias). The proposed solution is first validated and then evaluated with real data from the EuRoC [8] dataset and compared to the recent work of Campos *et al.* [2], showing state-of-the-art accuracy while being more efficient (about 7 times faster) and without requiring an initial guess for the scale factor. We also released our implementation for the benefit of the community.

II. RELATED WORK

Martinelli proposed in [4] a closed-form solution to the joint VI-SfM problem. His approach assumes that a set of 3D points can be observed in all frames during initialization and that the IMU measurements can be used to approximate the camera poses. The scale, gravity direction, accelerometer bias, and initial velocity are solved along with the depth of the visual features by building a system of linear equations. Kaiser *et al.* extended this approach in [5] to also take into account the gravity magnitude and estimate the gyroscope bias, resulting in a non-linear optimization problem solved iteratively. However,

the main issue with this approach is that the 3D error of the visual features is minimized instead of the reprojection error, which results in large unpredictable errors as shown in [9].

Modern visual odometry and SLAM systems achieve higher accuracy than IMU integration through Bundle-Adjustment (BA). This inspired Mur-Artal and Tardós [1] to handle the IMU and visual initialization separately. Once the vision-based subsystem is initialized and a first visual BA is performed, they consider the initialization of the IMU parameters. Qin and Shen presented a similar approach in [6] but leaving the estimation of the accelerometer bias to the VI-BA. In both cases, the IMU initialization problem is formulated independently for the gyroscope and the accelerometer as two ordinary least-squares problems. Note that, from a MAP perspective, ignoring the measurement uncertainties yields suboptimal results.

Campos *et al.* addressed these issues in [2] and formulated the IMU initialization problem within a MAP framework. However, the resulting non-linear optimization problem is solved iteratively, and thus requires a proper initial guess of the parameters. In fact, they perform three optimizations with different initial scales in parallel to successfully bootstrap the unknown initial scale factor. In our work, though, we formulate the IMU initialization problem as two independent problems similar to [1], [6] but within a MLE framework. Additionally, we propose an analytical solution for the accelerometer subproblem, thus removing the need of finding iteratively the scale factor and gravity direction, consequently achieving a much faster initialization.

III. PROBLEM FORMULATION

A. IMU measurement model

First, we briefly describe in this section the considered IMU measurement model. For a detailed description and derivations, we refer the interested reader to [10]. An IMU measures both the angular velocity and the specific forces in the body frame with respect to an inertial frame. Ignoring the effects of Earth's rotation, the simplified IMU measurement model can be defined as:

$${}_{\mathbb{B}}\tilde{\boldsymbol{\omega}}(t) = {}_{\mathbb{B}}\boldsymbol{\omega}(t) + \mathbf{b}^g + \boldsymbol{\eta}^g \quad (1)$$

$${}_{\mathbb{B}}\tilde{\mathbf{a}}(t) = \mathbf{R}_{\mathbb{W}\mathbb{B}}^{\top}(t)({}_{\mathbb{W}}\mathbf{a}(t) - {}_{\mathbb{W}}\mathbf{g}) + \mathbf{b}^a + \boldsymbol{\eta}^a, \quad (2)$$

in terms of the instantaneous angular velocity ${}_{\mathbb{B}}\boldsymbol{\omega} \in \mathbb{R}^3$ and the linear acceleration ${}_{\mathbb{W}}\mathbf{a} \in \mathbb{R}^3$ of the sensor. In this notation, we use prefixes to denote in which reference frame the corresponding elements are expressed. For instance, \mathbb{W} and \mathbb{B} refer to the world and body frames, respectively. The biases are represented by $\mathbf{b}^g, \mathbf{b}^a \in \mathbb{R}^3$ for the gyroscope and accelerometer respectively, while ${}_{\mathbb{W}}\mathbf{g} \in \mathbb{R}^3$ stands for the gravity vector. In turn, the pose of the IMU is described by the rigid body transformation $\{\mathbf{R}_{\mathbb{W}\mathbb{B}}, {}_{\mathbb{W}}\mathbf{p}\} \in \text{SE}(3)$, which includes both the rotation and the position, mapping the sensor frame \mathbb{B} coordinates to \mathbb{W} . Finally, the gyroscope and accelerometer measurements are corrupted by zero-mean Gaussian noise: $\boldsymbol{\eta}^g \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}^g)$, $\boldsymbol{\eta}^a \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}^a)$, respectively. Note that the biases are considered fixed in this model, since the IMU

initialization will only last for a short period of time (typically 2-10s).

Applying Euler integration, the pose and instantaneous velocity of the sensor ${}_{\mathbb{W}}\mathbf{v} \in \mathbb{R}^3$ can be expressed in terms of the IMU measurements in discrete time intervals as:

$$\mathbf{R}_{\mathbb{W}\mathbb{B}}(t + \Delta t) = \mathbf{R}_{\mathbb{W}\mathbb{B}}(t) \text{Exp}(({}_{\mathbb{B}}\tilde{\boldsymbol{\omega}}(t) - \mathbf{b}^g - \boldsymbol{\eta}^{gd})\Delta t) \quad (3)$$

$${}_{\mathbb{W}}\mathbf{v}(t + \Delta t) = {}_{\mathbb{W}}\mathbf{v}(t) + {}_{\mathbb{W}}\mathbf{g}\Delta t + \mathbf{R}_{\mathbb{W}\mathbb{B}}(t)({}_{\mathbb{B}}\tilde{\mathbf{a}}(t) - \mathbf{b}^a - \boldsymbol{\eta}^{ad})\Delta t \quad (4)$$

$${}_{\mathbb{W}}\mathbf{p}(t + \Delta t) = {}_{\mathbb{W}}\mathbf{p}(t) + {}_{\mathbb{W}}\mathbf{v}(t)\Delta t + \frac{1}{2}{}_{\mathbb{W}}\mathbf{g}\Delta t^2 + \frac{1}{2}\mathbf{R}_{\mathbb{W}\mathbb{B}}(t)({}_{\mathbb{B}}\tilde{\mathbf{a}}(t) - \mathbf{b}^a - \boldsymbol{\eta}^{ad})\Delta t^2 \quad (5)$$

where Δt represents the integration time and $\text{Exp}(\cdot)$ is the vector-valued exponential map of $\text{SO}(3)$. The covariance of the discrete-time noise for the gyroscope and accelerometer are $\boldsymbol{\Sigma}^{gd} = 1/\Delta t \boldsymbol{\Sigma}^g$ and $\boldsymbol{\Sigma}^{ad} = 1/\Delta t \boldsymbol{\Sigma}^a$, respectively [11].

We adopt the formulation in [3] to preintegrate the IMU measurements from keyframe to keyframe. In this way, the poses and the velocity of the IMU at the i -th and j -th time steps are related by:

$$\begin{aligned} \Delta \mathbf{R}_{ij} &\doteq \mathbf{R}_i^{\top} \mathbf{R}_j \\ &= \prod_{k=i}^{j-1} \text{Exp}((\tilde{\boldsymbol{\omega}}_k - \mathbf{b}^g - \boldsymbol{\eta}^{gd})\Delta t) \end{aligned} \quad (6)$$

$$\begin{aligned} \Delta \mathbf{v}_{ij} &\doteq \mathbf{R}_i^{\top} (\mathbf{v}_j - \mathbf{v}_i - \mathbf{g}\Delta t_{ij}) \\ &= \sum_{k=i}^{j-1} \Delta \mathbf{R}_{ik} (\tilde{\mathbf{a}}_k - \mathbf{b}^a - \boldsymbol{\eta}^{ad})\Delta t \end{aligned} \quad (7)$$

$$\begin{aligned} \Delta \mathbf{p}_{ij} &\doteq \mathbf{R}_i^{\top} (\mathbf{p}_j - \mathbf{p}_i - \mathbf{v}_i\Delta t_{ij} - \frac{1}{2}\mathbf{g}\Delta t_{ij}^2) \\ &= \sum_{k=i}^{j-1} \left(\Delta \mathbf{v}_{ik}\Delta t + \frac{1}{2}\Delta \mathbf{R}_{ik} (\tilde{\mathbf{a}}_k - \mathbf{b}^a - \boldsymbol{\eta}^{ad})\Delta t^2 \right) \end{aligned} \quad (8)$$

where we dropped the reference frame subscripts and used \cdot_t instead of $\cdot(t)$ for simplicity of notation. We refer the interested reader to the original paper for more details and derivations.

On the other hand, in a visual-inertial configuration, the sensor poses at the i -th keyframe, $\{\mathbf{R}_i, \mathbf{p}_i\} \in \text{SE}(3)$, can be independently estimated by the vision system. Thus, we have:

$$\mathbf{R}_i = \bar{\mathbf{R}}_i \mathbf{R}_{\text{cb}} \quad (9)$$

$$\mathbf{p}_i = s \bar{\mathbf{p}}_i + \mathbf{R}_i {}_c\mathbf{t}_{\mathbb{B}} \quad (10)$$

where $\{\bar{\mathbf{R}}_i, \bar{\mathbf{p}}_i\} \in \text{SE}(3)$ represent the keyframe poses as estimated by the vision system, $s \in \mathbb{R}^+$ the scale factor and $\{\mathbf{R}_{\text{cb}}, {}_c\mathbf{t}_{\mathbb{B}}\} \in \text{SE}(3)$ the extrinsic calibration of the camera with respect to the body frame, which is considered fixed.

B. Maximum Likelihood Estimation

The goal of the IMU initialization step is to estimate its initial parameters in order to allow VI fusion. From a MLE perspective, this is:

$$\mathcal{X}^* = \arg \max_{\mathcal{X}} p(\mathcal{X} | \mathcal{Z}) \quad (11)$$

where \mathcal{X} is the set of variables we want to estimate and \mathcal{Z} represents the set of observations. Under the mild assumption of independence between observations, we can write the posterior probability as:

$$p(\mathcal{X} | \mathcal{Z}) \propto p(\mathcal{Z} | \mathcal{X}) = \prod_{k \in \mathcal{K}} p(\mathcal{Z}_k | \mathcal{X}) \quad (12)$$

for each independent observation k . Under the assumption of zero-mean Gaussian noise, the optimization problem in (11) can be written as the sum of the squared residual errors:

$$\mathcal{X}^* = \arg \min_{\mathcal{X}} \sum_{k \in \mathcal{K}} \|\mathbf{r}_k\|_{\Sigma_k}^2, \quad (13)$$

where \mathbf{r}_k are the residual errors and Σ_k their corresponding covariance matrices.

In the following section, we express these MLE-based equations as a function of the parameters to be estimated.

IV. IMU INITIALIZATION

The complete VI-system initialization process, as proposed in [2], can be summarized as follows:

- 1) *Vision-only estimation.* This step usually involves the initialization of the vision system followed by a bundle-adjustment process for a few seconds of data, producing an up-to-scale estimate of the keyframe poses.
- 2) *IMU initialization.* Provides an initial estimate for the IMU-related parameters, as well as for the scale factor, from the previously estimated keyframe poses.
- 3) *Visual-inertial refinement.* Optionally, all previous parameters estimated separately for the vision and inertial systems can be refined in a combined VI-BA.

The first step can be addressed through any visual-based odometry/SLAM system such as ORB-SLAM [7] or PL-SLAM [12], as long as it provides accurate enough keyframe poses estimations, while the third, optional, stage can be accomplished by means of a standard VI bundle-adjustment procedure that considers both visual features and inertial measurements [1].

Is in the second step of the process where our method contributes with an analytic solution for the estimation of the IMU parameters (except the gyroscope bias), as formally stated next. For that, the IMU initialization process is further divided into two subproblems: (i) iteratively estimating the gyroscope bias and (ii) solving for the rest of the parameters. Our approach relies on the mild assumptions that the vision-only poses are very accurate (up to a scale factor), and that the IMU biases can be considered constant during the initialization.

A. Gyroscope bias

First we deal with the estimation of the gyroscope bias, that is, $\mathcal{X}_g \doteq \{\mathbf{b}^g\}$, from the gyroscope measurements and the keyframe orientations, i.e., $\mathcal{Z}_k \doteq \{\Delta \mathbf{R}_{k,k+1}, \bar{\mathbf{R}}_{k:k+1}\}$. For that, we define the residual function from the preintegrated rotation (6) as:

$$\mathbf{r}_k^g(\mathbf{x}) \doteq \text{Log} \left((\Delta \mathbf{R}_{k,k+1} \text{Exp}(\mathbf{J}_{\Delta \mathbf{R}}^g \mathbf{x}))^\top \mathbf{R}_k^\top \mathbf{R}_{k+1} \right), \quad (14)$$

where $\text{Log}(\cdot)$ stands for the inverse of the SO(3) exponential map and $\mathbf{J}_{\Delta \mathbf{R}}^g$ represents the Jacobian of the preintegrated rotation with respect to the bias.

Therefore, the MLE estimation of the gyroscope bias is equivalent to this optimization problem:

$$\mathcal{X}_g^* = \arg \min_{\mathbf{b}^g} \sum_{k \in \mathcal{K}} \|\mathbf{r}_k^g(\mathbf{b}^g)\|_{\Sigma_k^g}^2 \quad (15)$$

which can be solved iteratively using the well-known Levenberg–Marquardt algorithm starting from a zero bias. The analytic expressions for the preintegrated Jacobians and uncertainty propagation can be found in [3].

B. Accelerometer bias, gravity direction and scale factor

For the accelerometer, in turn, we are interested in finding analytically its bias (\mathbf{b}^a), the gravity direction (\mathbf{g}) and the scale (s) of the reconstruction from the measured acceleration and the relative pose. We opt here to eliminate the velocities by considering three consecutive keyframes in order to reduce the complexity of the problem, as proposed in [1]. Therefore, we have:

$$\mathcal{X}_a \doteq \{s, \mathbf{b}^a, \mathbf{g}\} \quad (16)$$

$$\mathcal{Z}_k \doteq \{\Delta \mathbf{v}_{k-1:k+1}, \Delta \mathbf{p}_{k-1:k+1}, \mathbf{b}^g, \bar{\mathbf{R}}_{k-1:k+1}, \bar{\mathbf{p}}_{k-1:k+1}\} \quad (17)$$

Thus, from (7) and (8), we define the residual for the accelerometer as:

$$\mathbf{r}_k^a(\mathbf{x}) = [\boldsymbol{\alpha}_k \quad \mathbf{A}_k \quad \mathbf{B}_k] \mathbf{x} - \boldsymbol{\pi}_k \quad (18)$$

with

$$\mathbf{A}_k = \frac{\mathbf{R}_{k-1} \mathbf{J}_{\Delta \mathbf{p}_{k-1,k}}^a}{\Delta t_{k-1,k}} - \frac{\mathbf{R}_k \mathbf{J}_{\Delta \mathbf{p}_{k,k+1}}^a}{\Delta t_{k,k+1}} - \mathbf{R}_{k-1} \mathbf{J}_{\Delta \mathbf{v}_{k-1,k}}^a \quad (19)$$

$$\mathbf{B}_k = -\frac{1}{2} (\Delta t_{k-1,k} + \Delta t_{k,k+1}) \mathbf{I}_{3 \times 3} \quad (20)$$

$$\boldsymbol{\alpha}_k = \frac{\bar{\mathbf{p}}_{k+1} - \bar{\mathbf{p}}_k}{\Delta t_{k,k+1}} - \frac{\bar{\mathbf{p}}_k - \bar{\mathbf{p}}_{k-1}}{\Delta t_{k-1,k}}, \quad \mathbf{x} = [s \quad \mathbf{b}^a \quad \mathbf{g}]^\top \quad (21)$$

$$\boldsymbol{\pi}_k = \frac{\mathbf{R}_k \Delta \mathbf{p}_{k,k+1}}{\Delta t_{k,k+1}} - \frac{\mathbf{R}_{k-1} \Delta \mathbf{p}_{k-1,k}}{\Delta t_{k-1,k}} + \mathbf{R}_{k-1} \Delta \mathbf{v}_{k-1,k} \quad (22)$$

$$+ \frac{(\bar{\mathbf{R}}_k - \bar{\mathbf{R}}_{k-1})_c \mathbf{t}_b}{\Delta t_{k-1,k}} - \frac{(\bar{\mathbf{R}}_{k+1} - \bar{\mathbf{R}}_k)}{\Delta t_{k,k+1}},$$

where $\mathbf{J}_{\Delta \mathbf{v}_{ij}}^a$ and $\mathbf{J}_{\Delta \mathbf{p}_{ij}}^a$ represent the Jacobian matrices of the preintegrated velocity and position with respect to the accelerometer bias, respectively. As before, the analytic expressions for these Jacobians, as well as their derivations, can be found in [3].

To sum up, the residual function comes from taking the difference $(\mathbf{p}_{k+1} - \mathbf{p}_k) / \Delta t_{k,k+1} - (\mathbf{p}_k - \mathbf{p}_{k-1}) / \Delta t_{k-1,k}$ from the preintegrated position (8) and substituting the expression $\mathbf{v}_{k+1} - \mathbf{v}_k$ involving the velocities with the preintegrated velocity from (7).

Note that, in this case, we are facing a constrained optimization problem, since the magnitude of the gravity vector is known:

$$\mathcal{X}_a^* = \arg \min_{\mathbf{x}} \sum_{k \in \mathcal{K}} \|\mathbf{r}_k^a(\mathbf{x})\|_{\Sigma_k^a}^2 \quad (23)$$

subject to $\|\mathbf{g}\| = G,$

where $G = 9.81$ is the magnitude of the gravity vector.

To solve this problem, first we write the cost function (23) in matrix form (where the superscript \cdot^a in Σ_k^a has been omitted for clarity):

$$\begin{aligned} \mathcal{C}(\mathbf{x}) &= \sum_{k \in \mathcal{K}} \|\mathbf{r}_k^a(\mathbf{x})\|_{\Sigma_k}^2 \\ &= \sum_{k \in \mathcal{K}} (\mathbf{M}_k \mathbf{x} - \boldsymbol{\pi}_k)^\top \Sigma_k^{-1} (\mathbf{M}_k \mathbf{x} - \boldsymbol{\pi}_k) \\ &= \sum_{k \in \mathcal{K}} (\mathbf{x}^\top \mathbf{M}_k^\top \Sigma_k^{-1} \mathbf{M}_k \mathbf{x} - 2\boldsymbol{\pi}_k^\top \Sigma_k^{-1} \mathbf{M}_k \mathbf{x} + \boldsymbol{\pi}_k^\top \Sigma_k^{-1} \boldsymbol{\pi}_k) \\ &= \mathbf{x}^\top \mathbf{M} \mathbf{x} + \mathbf{m}^\top \mathbf{x} + Q, \end{aligned} \quad (24)$$

where

$$\begin{aligned} \mathbf{M} &= \sum_{k \in \mathcal{K}} \mathbf{M}_k^\top \Sigma_k^{-1} \mathbf{M}_k, \quad \mathbf{m}^\top = \sum_{k \in \mathcal{K}} -2\boldsymbol{\pi}_k^\top \Sigma_k^{-1} \mathbf{M}_k, \\ Q &= \sum_{k \in \mathcal{K}} \boldsymbol{\pi}_k^\top \Sigma_k^{-1} \boldsymbol{\pi}_k, \quad \mathbf{M}_k = [\boldsymbol{\alpha}_k \quad \mathbf{A}_k \quad \mathbf{B}_k] \end{aligned} \quad (25)$$

We also write the constraint (23) in quadratic form:

$$\mathbf{x}^\top \mathbf{W} \mathbf{x} = G^2, \quad \mathbf{W} = \begin{bmatrix} \mathbf{0}_{4 \times 4} & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{I}_{3 \times 3} \end{bmatrix}, \quad (26)$$

so that the problem in (23) can be expressed as:

$$\begin{aligned} \mathcal{X}_a^* &= \arg \min_{\mathbf{x}} \mathbf{x}^\top \mathbf{M} \mathbf{x} + \mathbf{m}^\top \mathbf{x} \\ &\text{subject to } \mathbf{x}^\top \mathbf{W} \mathbf{x} = G^2 \end{aligned} \quad (27)$$

Then, we solve this quadratic system with a quadratic constraint using Lagrange's multipliers. For that, let the Lagrangian be:

$$\mathcal{L}(\mathbf{x}) = \mathbf{x}^\top \mathbf{M} \mathbf{x} + \mathbf{m}^\top \mathbf{x} + \lambda (\mathbf{x}^\top \mathbf{W} \mathbf{x} - G^2) \quad (28)$$

Therefore, the necessary condition for optimality is:

$$\frac{\partial \mathcal{L}(\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{x}^\top \mathbf{M} + \mathbf{m}^\top + 2\lambda \mathbf{x}^\top \mathbf{W} = \mathbf{0}^\top, \quad (29)$$

or, equivalently:

$$\mathbf{x} = -(2\mathbf{M} + 2\lambda \mathbf{W})^{-\top} \mathbf{m}. \quad (30)$$

Finally, substituting the relation in (30) back to the constraint in (26), leads to:

$$\mathbf{m}^\top (2\mathbf{M} + 2\lambda \mathbf{W})^{-1} \mathbf{W} (2\mathbf{M} + 2\lambda \mathbf{W})^{-\top} \mathbf{m} = G^2 \quad (31)$$

where λ is now the only unknown. It can be shown that (31) is a sixth-order polynomial in λ (see the appendix for a proof, cf. [13]). Thereby, we reduce the above-defined constrained optimization problem to finding the roots of this polynomial, which can be solved, for instance, by computing the eigenvalues of the corresponding companion matrix [14]. Once solved for λ , the original solution can be recovered using (30).

This analytical procedure does not need any initial guess for the parameters to be estimated and, as proven in our experiments, takes about $7 \times$ less time to get the solution than previous iterative approaches, while achieving similar accuracy.

TABLE I
ERRORS OF THE ESTIMATED PARAMETERS FROM GROUNDTRUTH POSES

Time	Inertial-only Optimization [2]				Proposed IMU Initialization			
	Scale	Gyro Bias	Acc Bias	G. Dir	Scale	Gyro Bias	Acc Bias	G. Dir
1.25s	6.23%	1.16%	99.9%	1.11°	4.61%	1.16%	721%	7.6°
2.5s	3.18%	0.98%	99.9%	1.06°	2.57%	0.94%	299%	3.24°
5s	2.64%	0.83%	99.7%	1.02°	1.60%	0.76%	90.3%	1.18°
12.5s	2.30%	0.72%	99.1%	0.94°	1.21%	0.52%	21.6%	0.42°
18.75s	2.05%	0.47%	98.5%	0.88°	1.11%	0.35%	12.7%	0.29°

V. EXPERIMENTS

We validated and evaluated the proposed IMU initialization approach with real data from the EuRoC dataset [8]. We chose this dataset since it provides accurate groundtruth from a Vicon motion capture system along the whole trajectory, for a total of 11 sequences. In the experiments, an IMU initialization procedure is attempted every 0.5s, with keyframes inserted at a fixed rate of 4Hz. The results are compared with the state-of-the-art IMU initialization approach described in [2] for reference (denoted by *inertial-only* in the following). In turn, the iterative optimization problems are implemented using Ceres Solver [15], while we used Eigen for our proposed analytical solution¹. All experiments were executed in an Intel Core i7-7700HQ CPU with 16 GB of RAM.

A. Validation

In this experiment we executed both, our method and the *inertial-only* optimization, using the keyframe poses from the dataset groundtruth instead of the vision-only estimated ones, in order to serve as a baseline of the expected performance. The estimated IMU parameters, i.e. biases, gravity direction and scale factor, are then compared with the groundtruth parameters provided by the dataset. The results of this experiment are shown in Table I. Note that, for the gravity vector, only the angle is considered for evaluation, while, for the biases, only the magnitude is considered. Besides, attempts of initialization with near constant velocity are unobservable [4], and thus are discarded. Because of this, and in order to get a fair comparison, we use the same heuristic proposed in [2], which discards attempts that exhibit small accelerations (average acceleration within 0.5% of the gravity magnitude).

As shown in the results, our approach performs better in general for the scale factor, even though the true scale factor is used as the initial guess for the *inertial-only* optimization. In turn, the gyroscope bias is solved in both cases using the same iterative algorithm, but we achieve slightly better results since we solve it separately from the rest of the parameters, therefore dealing with a reduced problem. For the accelerometer bias and the direction of the gravity vector, the *inertial-only* approach produces better results for short times of data (about 5s). However our approach achieves higher accuracy when considering longer initialization times, presenting errors with a clear diminishing tendency, in contrast to [2], which decrease at a slower rate.

¹The reference implementation of the proposed IMU initialization approach is publicly available at: https://github.com/dzunigan/imu_initialization

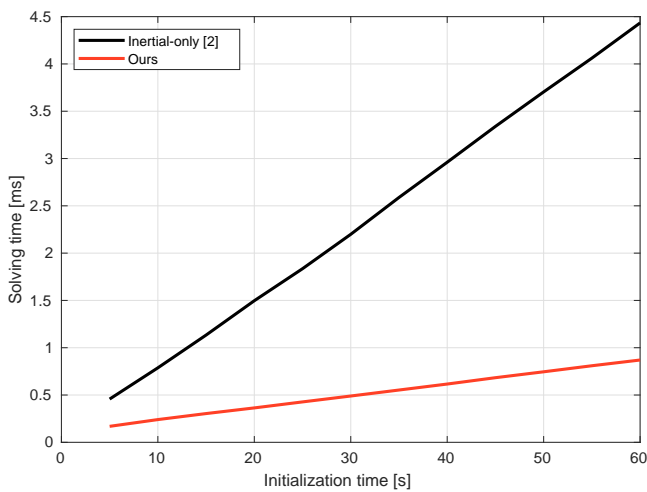


Fig. 1. Required solving time (in average for all dataset sequences) with respect to the length of the considered interval for initialization.

The particular case of the accelerometer bias deserves more attention, though, as it is well-known that such bias is hard to estimate for short initialization times (see [1], for example). To deal with this, Campos *et al.* [2] proposed to introduce a prior residual for this bias in a MAP formulation. For this comparison, we set it to a fixed value of 10^5 (as in the implementation of ORB_SLAM3 [16] system by the same authors), so that the bias remains close to zero for the whole experiment. In contrast, our method does not rely on any prior value, and still is able to estimate the bias more accurately than [2] when using more than 5 s of data.

Finally, in Fig. 1 we show the required time for solving the IMU initialization (average of all the dataset sequences) as a function of the time considered for initialization. Note that, for this test, we set again the initial guess for the metric scale to its true value for the *inertial-only* approach. As can be seen, even in these conditions our solution is faster and scales better compared to the iterative approach in [2].

B. Evaluation

In this experiment, we used the output of a visual odometry system, instead of the groundtruth, in order to perform the IMU initialization. We chose the well-known open source ORB_SLAM2 [17] system for this purpose, processing the images in monocular mode from the left camera only. We also disabled the loop closure and keyframe culling and run ORB_SLAM2 once for each sequence in the dataset. If the tracking was lost, we run the system again from where it failed. As in the previous experiment, we rejected initializations with small accelerations and also used a fixed keyframe insertion rate of 4 Hz. We also tested a *standard* non-linear optimization alternative, that is, the same approach than [2] but without any bias prior. In Table II we show the evaluation results of the IMU initialization process considering 20 keyframes (5 s).

In this case, our approach produced slightly better estimations for the scale factor and the gyroscope bias. However, with only 20 keyframes, the accelerometer bias could not be reliably estimated, consequently affecting the estimation of the gravity

direction. In this experiment, we used the three initial guesses for the scale factor proposed in the original paper [2]: 1, 4, 16, working in parallel. It is worth noting that we only show the required solving time, i.e. without taking into account the IMU preintegration, which is common for all methods (taking about 3.5 ms per keyframe, for a 200 Hz IMU and keyframes captured at 4 Hz). The solution proposed by Campos *et al.* is already very efficient, requiring about 1.3 ms on average to solve the initialization parameters. Comparatively, though, the solution presented in this paper runs about $\times 7$ faster in average.

VI. CONCLUSION

In this work we have addressed the initialization of the IMU parameters in a monocular Visual-Inertial system and proposed an analytical solution for the estimation of the accelerometer bias, the direction of gravity and the scale factor of the vision-based reconstruction.

The proposed formulation is defined within a MLE framework and derives a sixth degree polynomial that can be efficiently solved to obtain optimal values for such parameters. This way, our approach avoids the intrinsic issues of previous iterative solutions, including the need for a proper initial guess of the parameter values, specially the scale factor, and the possibility of divergence of the procedure. Additionally, our implementation incurs in only a fraction of the computational burden reported by state-of-the-art methods, being about 7 times faster.

Our solution has been validated and extensively assessed in a series of experiments with real data from the popular visual-inertial EuRoC dataset and using the vision-based ORB_SLAM2 [17] system, yielding consistent results in all cases. Nevertheless, as the experimental evaluation suggests, the observability of the accelerometer bias should be explicitly addressed in future work. Finally, our implementation of the IMU initialization proposed in this work can be found at: https://github.com/dzunigan/imu_initialization.

REFERENCES

- [1] R. Mur-Artal and J. D. Tardós, “Visual-inertial monocular SLAM with map reuse,” *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, 2017.
- [2] C. Campos, J. M. M. Montiel, and J. D. Tardós, “Inertial-only optimization for visual-inertial initialization,” in *IEEE Int. Conf. Robot. Autom. (ICRA)*, 2020.
- [3] 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, 2017.
- [4] A. Martinelli, “Closed-form solution of visual-inertial structure from motion,” *Int. J. Comput. Vis.*, vol. 106, no. 2, 2014.
- [5] J. Kaiser, A. Martinelli, F. Fontana, and D. Scaramuzza, “Simultaneous state initialization and gyroscope bias calibration in visual inertial aided navigation,” *IEEE Robot. Autom. Lett.*, vol. 2, no. 1, 2017.
- [6] T. Qin and S. Shen, “Robust initialization of monocular visual-inertial estimation on aerial robots,” in *IEEE Int. Conf. Intell. Robot. Syst. (IROS)*, 2017.
- [7] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, “ORB-SLAM: A versatile and accurate monocular SLAM system,” *IEEE Trans. Robot.*, vol. 31, no. 5, 2015.
- [8] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, “The EuRoC micro aerial vehicle datasets,” *Int. J. Robot. Research*, vol. 35, no. 10, 2016.
- [9] C. Campos, J. M. M. Montiel, and J. D. Tardós, “Fast and robust initialization for visual-inertial SLAM,” in *IEEE Int. Conf. Robot. Autom. (ICRA)*, 2019.

TABLE II
INITIALIZATION TIMES AND ERRORS OF THE ESTIMATED PARAMETERS FROM VISUAL POSES

Sequence	Inertial-only Optimization [2]					Standard					Proposed IMU Initialization				
	Solve	Scale	Gyro Bias	Acc Bias	G. Dir	Solve	Scale	Gyro Bias	Acc Bias	G. Dir	Solve	Scale	Gyro Bias	Acc Bias	G. Dir
V1_01	1.34 ms	27.9%	0.72%	99.9%	2.85°	1.09 ms	29.2%	0.86%	56.3%	3.04°	0.18 ms	19.5%	0.84%	58.5%	2.93°
V1_02	1.05 ms	4.50%	0.76%	99.6%	1.01°	1.09 ms	4.15%	1.22%	43.2%	0.88°	0.18 ms	2.69%	1.27%	33.4%	0.81°
V1_03	0.90 ms	13.5%	2.72%	99.6%	2.55°	1.05 ms	15.1%	2.90%	261%	5.12°	0.18 ms	11.3%	2.48%	203%	4.54°
V2_01	1.09 ms	8.66%	0.76%	99.7%	2.16°	1.04 ms	8.78%	0.79%	126%	2.97°	0.19 ms	6.36%	0.76%	96.5%	2.72°
V2_02	1.02 ms	6.54%	0.99%	99.6%	1.49°	1.11 ms	7.18%	0.90%	327%	2.84°	0.18 ms	5.90%	0.75%	125%	1.55°
V2_03	0.93 ms	4.90%	0.77%	99.6%	1.53°	1.04 ms	4.66%	0.60%	118%	1.59°	0.19 ms	2.67%	0.55%	98.2%	1.55°
MH_01	1.32 ms	26.4%	0.83%	99.7%	1.40°	2.97 ms	25.4%	0.83%	475%	7.15°	0.19 ms	15.3%	0.71%	371%	4.68°
MH_02	0.99 ms	12.1%	0.40%	99.7%	1.48°	1.01 ms	12.5%	0.41%	269%	3.73°	0.19 ms	10.0%	0.40%	228%	3.26°
MH_03	1.19 ms	18.2%	0.60%	99.7%	1.39°	1.12 ms	20.4%	0.57%	531%	5.73°	0.19 ms	14.2%	0.50%	299%	3.39°
MH_04	2.23 ms	17.8%	0.95%	99.7%	1.52°	2.15 ms	19.5%	0.93%	736%	7.23°	0.19 ms	15.1%	0.87%	585%	5.85°
MH_05	2.17 ms	19.8%	1.21%	99.7%	6.44°	1.72 ms	20.8%	1.20%	870%	11.4°	0.18 ms	11.7%	1.01%	439%	5.15°
Avg.	1.29 ms	14.6%	0.97%	99.7%	2.17°	1.40 ms	15.2%	1.02%	330%	4.70°	0.19 ms	10.4%	0.92%	230%	3.44°

- [10] M. Kok, J. D. Hol, and T. B. Schön, *Using Inertial Sensors for Position and Orientation Estimation*. Now Foundations and Trends in Signal Processing, 2017.
- [11] N. Trawny and S. I. Roumeliotis, "Indirect Kalman filter for 3D attitude estimation," Multiple Autonomous Robotic Systems Laboratory, Tech. Rep., 2005.
- [12] R. Gomez-Ojeda, F.-A. Moreno, D. Zuñiga-Noël, D. Scaramuzza, and J. Gonzalez-Jimenez, "PL-SLAM: A stereo SLAM system through the combination of points and line segments," *IEEE Trans. Robot.*, vol. 35, no. 3, 2019.
- [13] A. Censi, "An ICP variant using a point-to-line metric," in *IEEE Int. Conf. Robot. Autom. (ICRA)*, 2008.
- [14] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.
- [15] S. Agarwal, K. Mierle, and Others, "Ceres solver." [Online]. Available: <http://ceres-solver.org>
- [16] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "ORB-SLAM3: An accurate open-source library for visual, visual-inertial and multi-map slam," *arXiv preprint arXiv:2007.11898*, 2020.
- [17] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras," *IEEE Trans. Robot.*, vol. 33, no. 5, 2017.

APPENDIX

Proposition: Equation (31) is a sixth-order polynomial in λ .

Proof: Partition the matrix $(2\mathbf{M} + 2\lambda\mathbf{W})$ into four sub-matrices:

$$2\mathbf{M} + 2\lambda\mathbf{W} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{D} + 2\lambda\mathbf{I}_{3 \times 3} \end{bmatrix} \quad (32)$$

Since the \mathbf{W} in the quadratic form (31) is sparse, we only need to compute the last column of $(2\mathbf{M} + 2\lambda\mathbf{W})^{-1}$. Using the matrix inversion lemma we have:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{D} + 2\lambda\mathbf{I}_{3 \times 3} \end{bmatrix}^{-1} = \begin{bmatrix} \star & -\mathbf{A}^{-1}\mathbf{B}\mathbf{Q}^{-1} \\ \star & \mathbf{Q}^{-1} \end{bmatrix} \quad (33)$$

where $\mathbf{Q} = \mathbf{D} - \mathbf{B}^\top\mathbf{A}^{-1}\mathbf{B} + 2\lambda\mathbf{I}_{3 \times 3} \doteq \mathbf{S} + 2\lambda\mathbf{I}_{3 \times 3}$

The quadratic constraint (31) can now be expressed as:

$$\mathbf{m}^\top \mathbf{F} \mathbf{m} = G^2 \quad (34)$$

$$\mathbf{F} = \begin{bmatrix} \mathbf{A}^{-1}\mathbf{B}\mathbf{Q}^{-1}\mathbf{Q}^{-\top}\mathbf{B}^\top\mathbf{A}^{-\top} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{Q}^{-1}\mathbf{Q}^{-\top} \\ \text{symm} & \mathbf{Q}^{-1}\mathbf{Q}^{-\top} \end{bmatrix} \quad (35)$$

where only the matrix \mathbf{Q}^{-1} depends on λ . The analytic expression for this matrix can be written as:

$$\mathbf{Q}^{-1} = (\mathbf{S} + 2\lambda\mathbf{I}_{3 \times 3})^{-1} = \frac{(\mathbf{S} + 2\lambda\mathbf{I}_{3 \times 3})^{\mathbf{A}}}{p(\lambda)} \quad (36)$$

$$= \frac{\mathbf{S}^{\mathbf{A}} + 4\lambda^2\mathbf{I}_{3 \times 3} + 2\lambda(\text{tr}(\mathbf{S})\mathbf{I}_{3 \times 3} - \mathbf{S})}{p(\lambda)} \quad (37)$$

where

$$p(\lambda) \doteq \det(\mathbf{S} + 2\lambda\mathbf{I}_{3 \times 3}), \quad \mathbf{S}^{\mathbf{A}} = \det(\mathbf{S})\mathbf{S}^{-1} \quad (38)$$

Since \mathbf{Q}^{-1} is symmetric, we have:

$$\mathbf{Q}^{-1}\mathbf{Q}^{-\top} = \frac{(\mathbf{S}^{\mathbf{A}} + 4\lambda^2\mathbf{I}_{3 \times 3} + 2\lambda\mathbf{U})^2}{p(\lambda)^2} \quad (39)$$

$$= \frac{16\lambda^4\mathbf{I}_{3 \times 3} + 4\lambda^2(2\mathbf{S}^{\mathbf{A}} + \mathbf{U}^2)}{p(\lambda)^2} \quad (40)$$

$$+ \frac{16\lambda^3\mathbf{U} + 2\lambda(\mathbf{S}^{\mathbf{A}}\mathbf{U} + \mathbf{U}\mathbf{S}^{\mathbf{A}}) + \mathbf{S}^{\mathbf{A}^2}}{p(\lambda)^2}$$

where

$$\mathbf{U} \doteq \text{tr}(\mathbf{S})\mathbf{I}_{3 \times 3} - \mathbf{S} \quad (41)$$

Therefore, (31) can be written as a polynomial expression on λ :

$$16\lambda^4 \mathbf{m}^\top \begin{bmatrix} \mathbf{A}^{-1}\mathbf{B}\mathbf{B}^\top\mathbf{A}^{-\top} & -\mathbf{A}^{-1}\mathbf{B} \\ \text{symm} & \mathbf{I}_{3 \times 3} \end{bmatrix} \mathbf{m} \\ + 16\lambda^3 \mathbf{m}^\top \begin{bmatrix} \mathbf{A}^{-1}\mathbf{B}\mathbf{U}\mathbf{B}^\top\mathbf{A}^{-\top} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{U} \\ \text{symm} & \mathbf{U} \end{bmatrix} \mathbf{m} \\ + 4\lambda^2 \mathbf{m}^\top \begin{bmatrix} \mathbf{A}^{-1}\mathbf{B}\mathbf{X}\mathbf{B}^\top\mathbf{A}^{-\top} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{X} \\ \text{symm} & \mathbf{X} \end{bmatrix} \mathbf{m} \quad (42) \\ + 2\lambda \mathbf{m}^\top \begin{bmatrix} \mathbf{A}^{-1}\mathbf{B}\mathbf{Y}\mathbf{B}^\top\mathbf{A}^{-\top} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{Y} \\ \text{symm} & \mathbf{Y} \end{bmatrix} \mathbf{m} \\ + \mathbf{m}^\top \begin{bmatrix} \mathbf{A}^{-1}\mathbf{B}\mathbf{S}^{\mathbf{A}^2}\mathbf{B}^\top\mathbf{A}^{-\top} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{S}^{\mathbf{A}^2} \\ \text{symm} & \mathbf{S}^{\mathbf{A}^2} \end{bmatrix} \mathbf{m} \\ = G^2 p(\lambda)^2$$

where

$$\mathbf{X} \doteq 2\mathbf{S}^{\mathbf{A}} + \mathbf{U}^2, \quad \mathbf{Y} \doteq \mathbf{S}^{\mathbf{A}}\mathbf{U} + \mathbf{U}\mathbf{S}^{\mathbf{A}} \quad (43)$$

Since $p(\lambda)$ is a cubic function of λ , the expression (42) is a sixth-order polynomial in λ .