Article

Robust Plug-and-Play Joint Axis Estimation Using Inertial Sensors

Fredrik Olsson [1,*] **, Manon Kok** [2] **, Thomas Seel** [3] **and Kjartan Halvorsen** [1,4]

[1] Systems and Control, Department of Information Technology, Uppsala University,
    SE-75105 Uppsala, Sweden; kjartan.halvorsen@it.uu.se
[2] Delft Center for Systems and Control, Delft University of Technology, 2628 CD Delft, The Netherlands;
    m.kok-1@tudelft.nl
[3] Control Systems Group, Technische Universität Berlin, 10623 Berlin, Germany; seel@control.tu-berlin.de
[4] Department of Mechatronics, Campus Estado de Mexico, Tecnologico de Monterrey,
    Monterrey 64849, NL, Mexico
* Correspondence: fredrik.olsson@it.uu.se

check for updates

Abstract: Inertial motion capture relies on accurate sensor-to-segment calibration. When two segments are connected by a hinge joint, for example in human knee or finger joints as well as in many robotic limbs, then the joint axis vector must be identified in the intrinsic sensor coordinate systems. Methods for estimating the joint axis using accelerations and angular rates of arbitrary motion have been proposed, but the user must perform sufficiently informative motion in a predefined initial time window to accomplish complete identifiability. Another drawback of state of the art methods is that the user has no way of knowing if the calibration was successful or not. To achieve plug-and-play calibration, it is therefore important that 1) sufficiently informative data can be extracted even if large portions of the data set consist of non-informative motions, and 2) the user knows when the calibration has reached a sufficient level of accuracy. In the current paper, we propose a novel method that achieves both of these goals. The method combines acceleration- and angular rate information and finds a globally optimal estimate of the joint axis. Methods for sample selection, that overcome the limitation of a dedicated initial calibration time window, are proposed. The sample selection allows estimation to be performed using only a small subset of samples from a larger data set as it deselects non-informative and redundant measurements. Finally, an uncertainty quantification method that assures validity of the estimated joint axis parameters, is proposed. Experimental validation of the method is provided using a mechanical joint performing a large range of motions. Angular errors in the order of $2°$ were achieved using 125–1000 selected samples. The proposed method is the first truly plug-and-play method that overcome the need for a specific calibration phase and, regardless of the user's motions, it provides an accurate estimate of the joint axis as soon as possible.

Keywords: inertial measurement units; gyroscopes and accelerometers; sensor-to-segment calibration; kinematic constraints; joint axis identification; validation on mechanical joint

1. Introduction

Wearable inertial measurement units (IMUs) have become a key technology for a range of applications, from performance assessment and optimization in sports [1], to objective measurements and progress monitoring in health care [2], as well as real-time motion tracking for feedback-controlled robotic or neuroprosthetic systems [3]. In all these application domains, IMUs are used to track or capture the motion of mechatronic or biological joint systems such as robotic or human limbs. In this work we consider such systems where the joint is a hinge joint with one degree of freedom. Examples

of hinge joints include the knee and finger joints, which are essential in applications targeting lower limb [4] and hand [5] kinematics.

In contrast to stationary optical motion tracking systems, miniature IMU networks can be used in ambulatory settings and facilitate motion tracking outside lab environments. While this is an important step towards ubiquitous sensing, one major limitation of the technology is that the IMUs' local coordinate systems must be aligned with the anatomical axes of the joints and body segments to which they are attached. This sensor-to-segment calibration is a crucial step that establishes the connection between the motion of the IMUs and the motion of the joint system to which they are attached.

Several different approaches have been proposed for sensor-to-segment calibration of inertial sensor networks, from trying to align the sensor axes with body axes by precise attachment to predefined calibration poses and motions; see, e.g., [6–9]. However, in all of these cases, the calibration crucially depends on the knowledge and skills of the person who attaches the sensors or the person who performs the calibration procedure. This might be acceptable in supervised settings with trained and able-bodied users, but it represents a major limitation of IMU-based motion tracking and capture in clinical applications and in motion assessment of elderly and children. Finding solutions for these application domains and enabling ubiquitous sensing in daily life requires the development of less restrictive methods for sensor-to-segment calibration.

Ideally, wearable IMU networks should be plug-and-play, and the sensor-to-segment calibration should be performed by the network autonomously, which means without additional effort or requirements on the user's knowledge or on the performed motion. An important step towards this goal was the development of methods that exploit the kinematic constraints of the joints to identify sensor-to-segment calibration parameters from almost arbitrary motions [10,11]. For joints with one degree of freedom (DOF), the feasibility of this approach has been demonstrated [12–15]. Methods have been proposed that require the user to perform a sufficiently informative but otherwise arbitrary motion during an initial calibration time window and determine the functional joint axis in intrinsic coordinates of both IMUs, cf. Figure 1. It was recently shown that almost every motion, including purely sequential motions and simultaneous planar motions, is informative enough to render the joint axis identifiable unless the joint remains stiff throughout the motion [16].
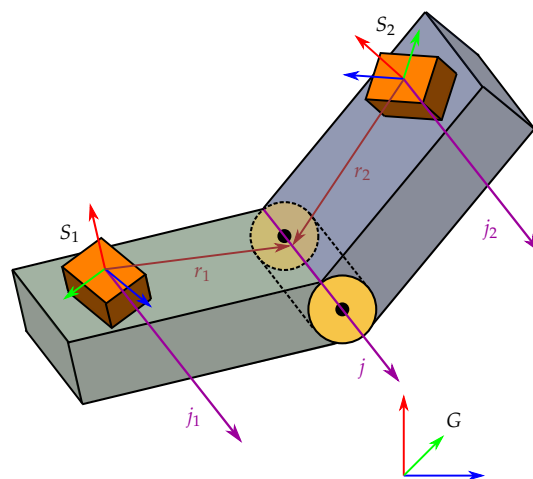

Figure 1. The hinge joint system that we consider. The two segments rotate independently with respect to each other only along the joint axis j. The sensor frames S_i are rigidly fixed to their respective segments and their relative orientation can be described by one joint angle, that corresponds to a rotation about the joint axis. The joint axis expressed in local sensor coordinates is an important sensor-to-segment calibration parameter in joint systems with one degree of freedom (DOF).

Several methods targeting different types of joints or sensor-to-segment calibration parameters have been developed. In [17], a method for identifying the joint axes of a joint with two DOF was proposed. Methods for identifying the position of the joint center relative to sensors attached to adjacent

segments have been proposed in [12,18,19]. A method enabling automatic pairing of sensors to lower limb segments have been proposed in [20].

The published kinematic-constraint-based methods constitute an important step forward but still impose undesirable and unnecessary limitations. If the user does not move during the initial calibration time window or if the motion is not sufficiently informative, the calibration will be wrong and all subsequently derived motion parameters will be subject to unpredictable errors. For a truly plug-and-play system, it is therefore crucial that the IMU network is able to

- Recognize how informative motions are and whether they render the joint axis identifiable;
- Wait for sufficiently informative data to be generated and combine useful data even if it is spread and intermitted by useless data;
- Determine how accurate the current estimate of the joint axis is and provide only sufficiently reliable estimates.

An IMU network with such properties can be used without the aforementioned limitations. Once it is installed, it will autonomously gather all available useful information and provide reliable calibration parameters as soon as possible, which immediately enable calculation of accurate motion parameters from the incoming raw data as well as from already recorded data. To explain the practical value of the proposed concept of plug-and-play calibration, we briefly compare this concept to the aforementioned existing calibration concepts that use predefined motions [6–9] or arbitrary motions [10–15]:

Predefined-Motions: The calibration is based on the assumption that the user performs a sequence of predefined motions and poses with sufficient precision within a predefined initial time interval. The approach fails and provides inaccurate calibration without warning if

(a) The user performs the sequence of predefined motions and poses without sufficient precision;
(b) The user performs the sequence with sufficient precision but not within the predefined initial time interval;
(c) The user performs sufficiently informative but otherwise arbitrary motions;
(d) The user performs no sufficiently informative motion at all, e.g., he/she moves with a stiff joint.

Arbitrary-Motions: The calibration is based on the assumption that the user performs sufficiently informative but otherwise arbitrary motions within a predefined initial time interval. The motion does not need to be precise, and it has been shown that sufficient excitation is provided by almost every motion for which the joint does not remain stiff [16]. However, the approach fails and provides inaccurate calibration without warning if

(a) The user performs a sequence of predefined motions but not within the predefined initial time interval;
(b) The user performs sufficiently informative arbitrary motions but not within the predefined initial time interval;
(c) The user performs no sufficiently informative motion at all, e.g., he/she moves with a stiff joint.

Plug-and-Play: The proposed sensor-to-segment calibration approach. It works well for all mentioned cases and exceptions in the sense that it always provides accurate calibration parameters as soon as the user's motions are sufficiently informative, and it clearly indicates at all times whether the desired calibration accuracy has yet been reached.

It is important to note that the cases without warning are very dangerous, because inaccurate information is provided and claimed as accurate. In many applications, this leads to unacceptable risks. This and the other listed differences between the two existing approaches and the proposed new method have large implications for the way wearable IMU networks can be used in offline and online applications.

Offline Applications include motion capture for ergonomic workplace assessment [21], for monitoring of movement disorders [2] and for sport performance analysis [1]. In state-of-the-art solutions, the user

performs an initial calibration procedure before (or after) recording data from the motions to be analyzed. The user can only hope that the calibration was accurate enough. If the calibration was inaccurate, then all recorded data is corrupted and might lead to false interpretation and conclusions. In contrast, when the calibration is plug-and-play, the user starts recording data from motions that should be analyzed immediately after attaching the sensors. Calibration automatically takes place as soon as sufficiently informative data has been gathered. The system indicates that calibration has been successful, and the user can be sure that all obtained measurements are valid and accurate. The identified calibration parameters are used to evaluate the data that was recorded before and after the moment at which accurate calibration was achieved.

Online Applications include real-time motion tracking for wearable biofeedback systems [22] as well as robotic and neuroprosthetic motion support systems [23]. In state-of-the-art solutions, the user first performs an initial calibration procedure before the sensor system is connected to an assistive device that uses the measurements to provide e.g., biofeedback or motion support. The user can only hope that the calibration was accurate enough. If the calibration was inaccurate, then the provided biofeedback or motion support might be wrong and dangerous. In contrast, when the calibration is plug-and-play, the user instead attaches the sensors and starts moving. As soon as the desired calibration accuracy has been achieved, the sensor system automatically provides measurements to the assistive device. The user can be sure that all provided biofeedback and motion support is based on valid and accurate measurements.

In the present contribution we propose the first joint axis identification method for one-dimensional joints that is plug-and-play in the aforementioned sense. The main contributions of the present work are the following:

1. We leverage recent results on joint axis identifiability [16] to develop a sample selection method that overcomes the limitation of a dedicated initial calibration time window.
2. To assure that the motion needs to fulfill only the minimum required conditions, we combine accelerometer-based and gyroscope-based joint constraints and weight them according to the information contained in both signals.
3. We propose an uncertainty quantification method that assures validity of the estimated joint axis parameters and thereby eradicates the risk of false calibration.
4. We provide an experimental validation in a mechanical joint performing a large range of different motions with different identifiability properties.

In the proposed system, successful calibration no longer depends on performing certain motions in a predefined manner or time window but only on fulfilling the minimum required conditions at some point. Moreover, the system knows when these conditions are fulfilled and provides only reliable calibration parameters.

2. Inertial Measurement Models

Inertial sensors collectively refers to accelerometers and gyroscopes, which are sensors used to measure linear acceleration and angular velocity, respectively. When the sensors have three sensitive axes which are orthogonal to each other, the inertial sensors can measure these quantities in three dimensions. Such sensors are referred to as triaxial. An IMU is a single sensor that contains one triaxial accelerometer and one triaxial gyroscope. The measurements from the IMU are obtained with respect to (w.r.t.) a reference frame, referred to as the sensor frame (S), its axes and origin corresponding to those of the accelerometer triad. The axes of the gyroscope is assumed to be aligned with the axes of the accelerometer. The measured quantities describe the motion of the sensor frame w.r.t. a global frame (G) that is fixed w.r.t. the environment.

The accelerometer measurements at time t_k, where the integer k is used as a sample index, can be modeled as

$$y_a^S(t_k) = R^{SG}(t_k) \left(a^G(t_k) + g^G \right) + b_a^S + e_a^S(t_k), \tag{1}$$

where $a^G \in \mathbb{R}^3$ is the acceleration of the sensor w.r.t. the global frame and $g^G \in \mathbb{R}^3$ is the gravitational acceleration, which is assumed to be constant in the environment. The measurements are corrupted by a constant additive bias b_a^S and noise $e_a^S(t_k) \in \mathbb{R}^3$, which is assumed to be Gaussian $e_a^S(t_k) \sim \mathcal{N}(0, \Sigma_a)$, with zero mean and covariance matrix Σ_a. The superscript S and G are used to denote in which reference frame a quantity is expressed in, and the rotation matrix R^{SG} describes the rotation from the global frame to the sensor frame, i.e., we have that

$$R^{SG}(t_k)\left(a^G(t_k) + g^G\right) = a^S(t_k) + g^S(t_k). \tag{2}$$

The multiplication between a rotation matrix and a vector is equivalent to a change of orthonormal basis.

The gyroscope measurements are modeled as

$$y_\omega^S(t_k) = R^{SG}(t_k)\omega^G(t_k) + b_\omega^S + e_\omega^S(t_k), \tag{3}$$

where $\omega^G \in \mathbb{R}^3$ is the angular velocity of the sensor frame in the global frame. Similar to the accelerometer, the measurements are corrupted by constant additive bias b_ω^S and noise $e_\omega^S(t_k) \in \mathbb{R}^3$, which is assumed to be zero-mean Gaussian $e_\omega^S(t_k) \sim \mathcal{N}(0, \Sigma_\omega)$. Note that the same rotation matrix R^{SG} as in (1) is used to rotate quantities from the global frame to the sensor frame because the accelerometer and the gyroscope are contained in the same IMU and their axes are assumed to be aligned. The gyroscope bias term b_ω^S can be compensated for through pre-calibration of the gyroscopes [24]. In Section 7.5, we will evaluate the effect of uncompensated biases on the proposed method.

Biases and Gaussian measurement noise have been shown to be the dominating error sources, even for low-cost IMUs [25]. However, for longer experiments or for low-quality IMUs, there are other types of errors that may need to be considered. These errors can still be well compensated for by pre-calibration or by online auto-calibration methods. Therefore, we only consider biases in our models, as these are the dominating systematic errors. The bias terms b_a and b_ω are not constant, but drift slowly over time [26]. Sensor manufacturers typically provide a bias stability metric for their sensors, which tells the user the expected rate of the bias drift. Bias instability in inertial sensors is primarily caused by low-frequency flicker noise in the electronics and temperature fluctuations [27]. If the bias drift is significant enough that it needs to be compensated for, there are methods that model the biases as time or temperature dependent, enabling continuous estimation of drifting biases (see, e.g., [28,29]). Such methods can be used in combination with the method proposed in this paper. Low-quality IMUs may be affected by other systematic errors such as non-unit scale factors and misalignments/non-orthogonalities in the sensor axes. If the effect from these types of errors are non-negligible, it is advised to perform a more sophisticated pre-calibration of the sensors to compensate for these errors. Methods for in-field pre-calibration of such errors exist; see, e.g., [30–33].

3. Kinematics

The kinematic model of the hinge joint system has been described in previous works [12,15,16], and is recapitulated here in Sections 3.1 and 3.2 for completeness.

3.1. Kinematic Constraints of Two Segments in a Kinematic Chain

Consider the kinematic chain model where we have two rigid body segments connected by a joint. The joint can have 1, 2 or 3 degrees of freedom (DOF). Furthermore, consider the case where each segment has one IMU rigidly attached to it in an arbitrary position and orientation. We therefore have two sensor frames, denoted by S_1 and S_2, that are fixed in the center of the accelerometer triad of each IMU. The DOF of the joint determines how many angles that are required to describe the orientation of

S_2 w.r.t. S_1 and vice versa. We let subscripts $i \in \{1,2\}$ denote quantities belonging to a specific sensor frame. Rigid body kinematics gives

$$a_i^{S_i}(t) = a_0^{S_i}(t) + \omega_i^{S_i}(t) \times (\omega_i^{S_i}(t) \times r_i^{S_i}) + \dot{\omega}_i^{S_i}(t) \times r_i^{S_i}, \tag{4}$$

where a_i are the accelerations of the sensor frames with $i \in \{1,2\}$, a_0 is the acceleration of the joint center, ω_i and $\dot{\omega}_i$ are the angular velocities and angular accelerations of the sensor frames and t is used to denote time-dependence of the kinematic variables. The positions of the joint center with respect to each sensor frame are denoted by r_i, which we assume to be unknown and constant for each sensor. All quantities in (4) are vectors in $\mathbb{R}^3$ since they describe 3D motion. The acceleration of the joint center expressed in either of the sensor frames has the same magnitude but a different orientation. We have that

$$a_0^G(t) = R^{GS_1}(t)a_0^{S_1}(t) = R^{GS_2}(t)a_0^{S_2}(t) \tag{5}$$

where R^{GS_i} are the rotation matrices that maps a vector expressed in S_i into the global frame.

For convenience we shall for the remainder of this document drop the use of the superscripts except for where it's needed. Hence, the sensor frame of a kinematic variable will be given by subscript $i \in \{1,2\}$. We will also drop the use of t to denote time-dependence unless we want to refer to the kinematic variables at specific time instances. The relationship in (4) is linear in a_0 and r_i and can equivalently be formulated as

$$a_i = a_0^{S_i} + K(\omega_i, \dot{\omega}_i)r_i, \tag{6}$$

where

$$K(\omega, \dot{\omega}) = \begin{bmatrix} -\omega_y^2 - \omega_z^2 & \omega_x\omega_y - \dot{\omega}_z & \omega_x\omega_z + \dot{\omega}_y \\ \omega_x\omega_y + \dot{\omega}_z & -\omega_x^2 - \omega_z^2 & \omega_y\omega_z - \dot{\omega}_x \\ \omega_x\omega_z - \dot{\omega}_y & \omega_y\omega_z + \dot{\omega}_x & -\omega_x^2 - \omega_y^2 \end{bmatrix}, \tag{7}$$

and where subscripts x, y, z denote the elements of the three-dimensional vectors. For convenience of notation we will write $K_i = K(\omega_i, \dot{\omega}_i)$.

3.2. Kinematic Constraints of a Hinge Joint System

For a 1-DOF joint, the two segments can only rotate independently with respect to each other along the joint axis. We let $\| \cdot \|$ denote the Euclidean vector norm, then the joint axis is defined by the unit vector $j \in \mathbb{R}^3, \|j\| = 1$. We refer to such a joint as a hinge joint. We let j_1 and j_2 denote the direction of the joint axis in the respective sensor frames. Since the two IMUs are assumed to be rigidly attached to the segments, j_1 and j_2 are constant. The joint axis j expressed in the global frame must then satisfy

$$j^G(t) = R^{GS_1}(t)j_1^{S_1} = R^{GS_2}(t)j_2^{S_2}, \tag{8}$$

meaning that the vectors j_i expressed in the two sensor frame has the same direction as j in the global frame, see Figure 1, and time-dependence is only caused by the rotations of the sensor frames in the global frame. We can decompose the angular velocities into one component that is parallel to the joint axis and one that is perpendicular to the joint axis

$$\omega_i = \omega_{j_i} + \omega_{j_i^\perp}, \tag{9}$$

$$\omega_{j_i} = j_i^\top \omega_i j_i, \tag{10}$$

$$\omega_{j_i^\perp} = \omega_i - \omega_{j_i} = \omega_i - j_i^\top \omega_i j_i. \tag{11}$$

Since the two segments can only rotate independently along the joint axis, it follows that the perpendicular components must have the same magnitude regardless of reference frame

$$\|\omega_{j_1^\perp}\| = \|\omega_{j_2^\perp}\|. \tag{12}$$

The magnitude of the perpendicular component can also be computed from the cross product between the angular velocity and the joint axis

$$\|\omega_i - j_i^\top \omega_i j_i\| = \|\omega_i \times j_i\|. \tag{13}$$

Combining (12) and (13) we formulate the angular velocity constraint

$$\|\omega_1 \times j_1\| - \|\omega_2 \times j_2\| = 0, \tag{14}$$

which must be satisfied by hinge joint systems.

Looking at the projection of the accelerations onto the joint axis, from (6) we have that

$$j_i^\top a_i j_i = j_i^\top a_0^{S_i} j_i + j_i^\top K_i r_i j_i. \tag{15}$$

Because j_i has the same direction as j in the global frame, it must also be the same for the projection of a_0 onto j_i, it follows from (5) and (8) that

$$j^{G^\top} a_0^G j^G = R^{GS_1} j_1 j_1^\top a_0^{S_1} = R^{GS_2} j_2 j_2^\top a_0^{S_2}$$
$$\Rightarrow j_1^\top a_0^{S_1} = j_2^\top a_0^{S_2}. \tag{16}$$

By projecting the accelerations onto the joint axis and subtracting one from the other we get

$$j_1^\top a_1 - j_2^\top a_2 = j_1^\top a_0^{S_1} - j_2^\top a_0^{S_2} + j_1^\top K_1 r_1 - j_2^\top K_2 r_2$$
$$= j_1^\top K_1 r_1 - j_2^\top K_2 r_2, \tag{17}$$

where we see that only the rotational components of the accelerations remain on the right hand side. The relationship (17) is the exact acceleration constraint of the hinge joint system. The right hand side (r.h.s.) of (17) is zero if and only if either $K_i r_i \perp j_i$ or $K_i = 0$ are satified for all $i \in \{1, 2\}$. It is clear that if the rotational acceleration components along the direction of the joint axis are small ($j_i^\top K_i r_i \approx 0, \forall i$), the r.h.s. will vanish

$$j_1^\top a_1 - j_2^\top a_2 \approx 0, \tag{18}$$

which forms the approximate acceleration constraint for the hinge joint system.

4. Joint Axis Estimation

We assume that we have two IMUs, one attached to each segment of a hinge joint system. Measurements from a completely unspecified motion has been collected. We will use $y_{\omega,i}$ to refer to the gyroscope measurements (3) and $y_{a,i}$ to refer to the accelerometer measurements (1) from Sensor $i \in \{1, 2\}$. We will use the non-indexed y_ω and y_a to refer to measurements from both sensors as

$$y_\omega = \begin{bmatrix} y_{\omega,1}^\top & y_{\omega,2}^\top \end{bmatrix}^\top, \tag{19}$$

and similarly for y_a. We let $\mathcal{D}^N = \{y_\omega^N, y_a^N\}$ denote our data, which consists of N samples of recorded motion. Each sample in the data set is assigned a sample index $k \in \{1, \ldots, N\}$, such that t_k refers to the sampling time of the kth measurement relative to the beginning of the recorded motion.

Given the data $\mathcal{D}^N$ from the two IMUs, the variables we want to estimate are the unit vectors j_i which corresponds to the directions of the joint axis j in the two sensor frames. We let $\widehat{j}_i$ denote the estimate of j_i. Note that the joint axis in one sensor frame can be described by either $\pm j_i$ since a clockwise rotation w.r.t. the positive axis is equivalent to a counter-clockwise rotation w.r.t. the negative axis. However, we require both j_1 and j_2 to have the same sign (direction) to correspond to either $\pm j$ in the global frame, otherwise a clockwise rotation for one sensor might be considered a counter-clockwise rotation for the other sensor and vice versa. That is, the sign pairing of the joint axes in the sensor coordinate frames is important. Consequently, $(\pm j_1, \pm j_2)$ is the correct sign pairing and $(\pm j_1, \mp j_2)$ is the wrong sign pairing.

4.1. Formulating the Optimization Problem

We parametrize j_i using spherical coordinates to enforce the unit vector constraint

$$x = \begin{bmatrix} \theta_1 & \phi_1 & \theta_2 & \phi_2 \end{bmatrix}^\top, \tag{20}$$

$$j_i(x) = \begin{bmatrix} \cos\theta_i \cos\phi_i \\ \cos\theta_i \sin\phi_i \\ \sin\theta_i \end{bmatrix}, \tag{21}$$

which then become the unknown parameters to estimate. The estimation problem for the joint axis is formulated as

$$\widehat{x} = \arg\min_x V(x), \tag{22}$$

$$V(x) = \sum_{k=1}^N [e_\omega(k,x)]^2 + [e_a(k,x)]^2, \tag{23}$$

where $e_\omega(k,x)$ and $e_a(k,x)$ are scalar residual terms, based on the angular velocity constraint (14) and acceleration constraints (18) of the hinge joint system

$$e_\omega(k,x) = w_\omega[\|y_{\omega,1}(t_k) \times j_1(x)\| - \|y_{\omega,2}(t_k) \times j_2(x)\|], \tag{24}$$

$$e_a(k,x) = w_a[j_1^\top(x) y_{a,1}(t_k) - j_2^\top(x) y_{a,1}(t_k)]. \tag{25}$$

Two scalars w_ω and w_a are used to change the relative weighting of the residuals.

4.2. Identifiability and Local Minima

For the gyroscope measurements to contain information about the joint axis, they have to be recorded from motions where the joint angle is excited, i.e., when the two segments rotate independently. These motions should contain either simultaneous planar rotations, where the segments rotate simultaneously in the plane perpendicular to the joint axis, or sequential rotations of the segments. However, stiff joint motions, which can have a significant angular rate but no independent rotation of the segments, do not facilitate identifiability of the joint axis [16]. For the non-informative stiff joint motions, the relative rotation of the two sensors can be described by a time-invariant rotation matrix R and we have that

$$\|\omega_2(t_k) \times j_2\| = \|R(\omega_1(t_k) \times j_1)\| = \|\omega_1(t_k) \times j_1\|, \tag{26}$$

where we see that for any choice of j_1, the vector $j_2 = Rj_1$ will minimize the gyroscope residual (24). Therefore, we want motions where $\|\omega_1(t_k)\| \neq \|\omega_2(t_k)\|$, which implies that the segments are rotating independently and we require motions where $\|\omega_i(t_k)\| > 0$ for at least some time, since $\|\omega_i(t_k)\| = 0 \Rightarrow \|\omega_i(t_k) \times j_i\| = 0, \forall j_i$.

If only acceleration information is considered, we get the following over-determined system of linear equations

$$\underbrace{\begin{bmatrix} a_1^\top(t_1) & -a_2^\top(t_1) \\ \vdots & \vdots \\ a_1^\top(t_M) & -a_2^\top(t_M) \end{bmatrix}}_{=A} \begin{bmatrix} j_1 \\ j_2 \end{bmatrix} = 0, \tag{27}$$

which has a unique solution if $\text{rank}(A) = 5$, in which case $\begin{bmatrix} j_1^\top & j_2^\top \end{bmatrix}^\top$ lies in the null-space of A. This holds when the acceleration constraint holds exactly for all t_k, the accelerations measured are exact and the angular rate and angular accelerations of the sensors are parallel with j [16]. Therefore, for the accelerometer, we want measurements that increase the separation between the column-space and the null-space of A.

The proposed method uses both gyroscope and accelerometer information, and their relative contribution to the cost function is controlled by the weight parameters w_ω and w_a. Figure 2 shows how the weights affect the cost function in the case that $w_a = 1$ and w_ω is allowed to vary. For small w_ω, the local minima corresponds to the correct sign pairing $(\pm j_1, \pm j_2)$, whereas the local maxima corresponds to the wrong sign pairing $(\pm j_1, \mp j_2)$. Note that each local minimum is equally valid for small w_ω because of the periodicity of the spherical coordinates. The acceleration residuals are relatively large whereas the gyroscope residuals are relatively small at the locations corresponding to the wrong sign pairing. Therefore, as w_ω increases the gyroscope residuals will contribute more to the cost function. The peaks associated with the wrong sign pairing are flattened and new local minima will eventually appear at these locations. Therefore, for large w_ω an optimization method (solver) can end up in the wrong local minimum. However, regardless of which sign pairing the solver finds, the opposite sign pairing can always be obtained at $x = \begin{bmatrix} \theta_1 & \phi_1 & -\theta_2 & \phi_2 + \pi \end{bmatrix}^\top$. Therefore, if our solver finds the estimate $\widehat{x}^{(1)}$ we can reinitialize at

$$\begin{bmatrix} \widehat{\theta}_1^{(1)} & \widehat{\phi}_1^{(1)} & -\widehat{\theta}_2^{(1)} & \widehat{\phi}_2^{(1)} + \pi \end{bmatrix}^\top, \tag{28}$$

and obtain a new estimate $\widehat{x}^{(2)}$. Then we select the local minimum with the smallest value of the cost function as our estimate

$$\widehat{x} = \underset{x \in \{\widehat{x}^{(1)}, \widehat{x}^{(2)}\}}{\arg\min} V(x). \tag{29}$$

Therefore, it is possible to find the correct sign pairing as long as $V(\widehat{x}^{(2)})$ is numerically distinguishable from $V(\widehat{x}^{(1)})$. As discussed in this section and shown in Figure 2, the relative weighting of the residuals determines how easy it is to distinguish a correct local minimum from a wrong one. If w_ω is set to be significantly larger than w_a, we expect the acceleration residuals to eventually become so small relative to the gyroscope residuals, that the solver is no longer sensitive enough to detect the difference between correct and wrong local minima.
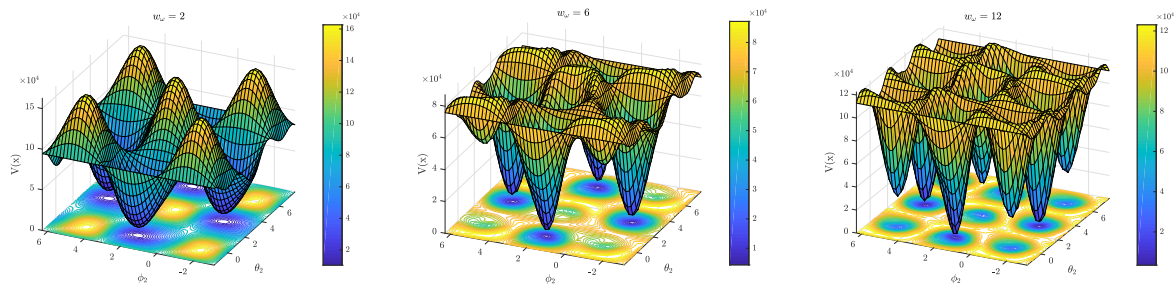
Figure 2. Shape of the cost function $V(x)$ for a motion with simultaneous planar rotations of the segments. The parameters θ_1 and ϕ_1 are fixed near their true values and θ_2 and ϕ_2 are allowed to vary. From left to right, we see how the geometry changes as w_ω increases while $w_a = 1$ is constant. As w_ω increases, new local minima appear near the locations at $\phi_2 + \pi$ from the previously existing local minima. These new local minima correspond to the wrong sign pairing $(\pm j_1, \mp j_2)$.

4.3. Solving the Optimization Problem

The optimization problem (22) is a nonlinear least-squares problem. An efficient solver for such problems is the Gauss–Newton method [34]. Given an initial estimate $\widehat{x}(0)$ the Gauss–Newton method iteratively updates the estimate according to

$$\begin{aligned}
\widehat{x}(k+1) &= \widehat{x}(k) - \alpha \left(J^\top(\widehat{x}(k)) J(\widehat{x}(k)) \right)^{-1} J^\top(\widehat{x}(k)) e(\widehat{x}(k)) \\
&= \widehat{x}(k) - \alpha \Delta_x(k)
\end{aligned} \tag{30}$$

where k is only used here as an integer index denoting the iterations of the method and is not to be confused with the sample index. The method uses the Jacobian matrix $J(x) \in \mathbb{R}^{2N \times 4}$, which contains all first-order partial derivatives of e_ω and e_a

$$J(x) = \begin{bmatrix}
\frac{\partial e_\omega(1,x)}{\partial \theta_1} & \frac{\partial e_\omega(1,x)}{\partial \phi_1} & \frac{\partial e_\omega(1,x)}{\partial \theta_2} & \frac{\partial e_\omega(1,x)}{\partial \phi_1} \\
\vdots & \vdots & \vdots & \vdots \\
\frac{\partial e_\omega(N,x)}{\partial \theta_1} & \frac{\partial e_\omega(N,x)}{\partial \phi_1} & \frac{\partial e_\omega(N,x)}{\partial \theta_2} & \frac{\partial e_\omega(N,x)}{\partial \phi_1} \\
\frac{\partial e_a(1,x)}{\partial \theta_1} & \frac{\partial e_a(1,x)}{\partial \phi_1} & \frac{\partial e_a(1,x)}{\partial \theta_2} & \frac{\partial e_a(1,x)}{\partial \phi_1} \\
\vdots & \vdots & \vdots & \vdots \\
\frac{\partial e_a(N,x)}{\partial \theta_1} & \frac{\partial e_a(N,x)}{\partial \phi_1} & \frac{\partial e_a(N,x)}{\partial \theta_2} & \frac{\partial e_a(N,x)}{\partial \phi_1}
\end{bmatrix}, \tag{31}$$

and $e(x) \in \mathbb{R}^{2N}$ is the residual vector

$$e(x) = \begin{bmatrix} e_\omega(1,x) & \ldots & e_\omega(N,x) & e_a(1,x) & \ldots & e_a(N,x) \end{bmatrix}^\top. \tag{32}$$

The term $\left(J^\top(x) J(x) \right)^{-1}$ is an approximation of the Hessian of $V(x)$, which is given by

$$\begin{aligned}
\frac{d^2 V(x)}{dx^2} &= J^\top(x) J(x) + \sum_{k=1}^{N} e_\omega(k,x) \frac{d^2 e_\omega(k,x)}{dx^2} \\
&\quad + \sum_{k=1}^{N} e_a(k,x) \frac{d^2 e_a(k,x)}{dx^2}
\end{aligned} \tag{33}$$

where the higher-order terms are ignored, yielding

$$\frac{d^2 V(x)}{dx^2} \approx J^\top(x) J(x). \tag{34}$$

The partial derivatives of the residuals (24) and (25) in the Jacobian (31) are computed in the following way using the chain rule

$$\frac{\partial e_\omega(k, x)}{\partial x} = \frac{\partial j}{\partial x} \frac{\partial e_\omega(k, x)}{\partial j} w_\omega(k), \tag{35}$$

$$\frac{\partial e_\omega(k, x)}{\partial j} = \begin{bmatrix} \frac{\partial(\|y_{\omega,1}(t_k) \times j_1(x)\|)}{\partial j_1} \\ -\frac{\partial(\|y_{\omega,2}(t_k) \times j_2(x)\|)}{\partial j_2} \end{bmatrix}, \tag{36}$$

$$\frac{\partial(\|y_{\omega,i}(t_k) \times j_i(x)\|)}{\partial j_i} = \frac{(y_{\omega,i}(t_k) \times j_i) \times y_{\omega,i}(t_k)}{\|y_{\omega,i}(t_k) \times j_i(x)\|}, \tag{37}$$

$$\frac{\partial e_a(k, x)}{\partial x} = \frac{\partial j}{\partial x} \frac{\partial e_a(k, x)}{\partial j}, \tag{38}$$

$$\frac{\partial e_a(k, x)}{\partial j} = \begin{bmatrix} y_{a,1}(t_k) \\ -y_{a,2}(t_k) \end{bmatrix} w_a(k), \tag{39}$$

$$\frac{\partial j}{\partial x} = \begin{bmatrix} \frac{\partial j_1}{\partial x_1} & 0 \\ 0 & \frac{\partial j_2}{\partial x_2} \end{bmatrix}, \tag{40}$$

$$\frac{\partial j_i}{\partial x_i} = \begin{bmatrix} -\sin \theta_i \cos \phi_i & -\cos \theta_i \sin \phi_i \\ -\sin \theta_i \sin \phi_i & \cos \theta_i \cos \phi_i \\ \cos \theta_i & 0 \end{bmatrix}^\top. \tag{41}$$

The term Δ_x in (30) defines the search direction, and $-\Delta_x$ is a descent direction, meaning that moving our estimate in that direction will decrease the value of the cost function. The scalar $0 < \alpha \leq 1$ is known as the step length, which controls how far our estimates move in the descent direction. By using a method known as backtracking line search [35], we find a value for α that is guaranteed to lower the value of the cost function. If no such α is found or the change in the value of $V(x)$ is too small, below a set tolerance level V_{tol}, the Gauss–Newton method terminates and returns the estimate corresponding to the current iteration $\widehat{x} = \widehat{x}(k)$.

The complete joint axis estimation method, including the steps of the Gauss–Newton method and the re-initialization step (28) required to identify the minimum corresponding to the correct sign pairing, is described in Algorithm 1.

Algorithm 1 Joint axis estimation

Require: Data $\mathcal{D}^N = \{y_\omega^N, y_a^N\}$, initial estimate $\widehat{x}(0)$, tolerance V_{tol}, residual weights w_ω and w_a.

1: **for** $i \in \{1, 2\}$ **do**
2:     $k \leftarrow 0$.   ▷ Begin Gauss–Newton.
3:     $\Delta_V \leftarrow V_{\text{tol}}$.
4:     $V(0) \leftarrow V(\widehat{x}(0))$.   ▷ $V(x)$ defined by (23).
5:     **while** $\Delta_V \geq V_{\text{tol}}$ **do**
6:        Compute the Jacobian $J(\widehat{x}(k))$ and the residuals $e(\widehat{x}(k))$ according to (31) and (32).
7:        $\Delta_x(k) \leftarrow \left(J^\top(\widehat{x}(k))J(\widehat{x}(k))\right)^{-1} J^\top(\widehat{x}(k))e(\widehat{x}(k))$.
8:        Obtain step length α using backtracking line search.
9:        $\widehat{x}(k+1) \leftarrow \widehat{x}(k) - \alpha\Delta_x(k)$.
10:       $k \leftarrow k + 1$.
11:       $V(k) \leftarrow V(\widehat{x}(k))$.
12:       $\Delta_V \leftarrow |V(k-1) - V(k)|$.
13:    **end while**
14:    $\widehat{x} \leftarrow \widehat{x}(k)$.   ▷ End Gauss–Newton.
15:    $\widehat{x}^{(i)} = \begin{bmatrix} \widehat{\theta}_1^{(i)} & \widehat{\phi}_1^{(i)} & \widehat{\theta}_2^{(i)} & \widehat{\phi}_2^{(i)} \end{bmatrix}^\top \leftarrow \widehat{x}$.
16:    $\widehat{x}(0) \leftarrow \begin{bmatrix} \widehat{\theta}_1^{(i)} & \widehat{\phi}_1^{(i)} & -\widehat{\theta}_2^{(i)} & \widehat{\phi}_2^{(i)} + \pi \end{bmatrix}^\top$.   ▷ Initialize at $-\widehat{j}_2$.
17: **end for**
18: $\widehat{x} \leftarrow \arg\min_{x \in \{\widehat{x}^{(1)}, \widehat{x}^{(2)}\}} V(x)$.   ▷ Correct sign pairing.
19: **return** $j(\widehat{x})$.

---

5. Sample Selection

A key feature of plug-and-play estimation is that it should not require specific calibration data, recorded from predetermined motions. Rather, such plug-and-play methods should be able to use data recorded from arbitrary motions. Such data sets could be very large, and using all available data for identification is often unnecessary and resource-demanding. It is also possible that very few samples in the data set contain information about the joint axis. In a sense, too much bad information might ruin the good information. To handle this, we propose a method for selecting samples to use for estimation.

In the following sections we assume that we want a maximum of N_{max} gyroscope and accelerometer measurements can be used to identify the joint axis, but that we have $N > N_{\text{max}}$ measurements available to us to choose from.

5.1. Gyroscope

To distinguish between informative and non-informative motions, we use the difference in angular velocity magnitude measured by the two gyroscopes

$$\Delta_\omega(k) = \|y_{\omega,1}(t_k)\| - \|y_{\omega,2}(t_k)\|, \tag{42}$$

which is a sufficient metric for detecting independent rotations of the sensors, and hence the two segments. For stationary segments $\Delta_\omega(k) = 0$. One thing to note is that $\Delta_\omega(k)$ cannot differentiate between informative motions where $\|\omega_1\| \approx \|\omega_2\|$ and non-informative stiff joint rotations. For example, the two segments can undergo simultaneous planar rotations, where the two segments rotate in different directions but with approximately the same magnitude. However, for realistic motions, especially for motions performed by humans, it is unlikely that independent rotations will have the same magnitude, even for short moments.

Each gyroscope measurement is given a score

$$s_\omega(k) = \Delta_\omega(l') \tag{43}$$

$$l' = \arg\min_l |\Delta_\omega(l)|, \quad l \in (k - n, k + n) \tag{44}$$

that is equal to the Δ_ω with smallest magnitude in a window of $2n + 1$ samples. This is to avoid selecting large outliers of Δ_ω. For example, if the system is not completely rigid or the sensors are not rigidly attached, the kinematic constraints are violated, and some samples of stiff joint motion can obtain a large Δ_ω value. However, if the outliers are relatively few, there should be Δ_ω with smaller magnitude among neighboring samples. In some sense, $s_\omega(k)$ assumes a conservative score for each sample.

When the score s_ω has been computed for all measurements, the list of measurements is sorted in descending order such that $s_\omega(k') \geq s_\omega(k' + 1), \forall k' \in (1, N - 1)$, where k' is a new index variable used to denote the sorted order. The first and last $N_{max}/2$ of the sorted gyroscope measurements are selected, or, equivalently, the measurements corresponding to the middle of the list, i.e., with index $k' \in (N_{max}/2 + 1, N - N_{max}/2)$ are removed from the set of measurements. By doing this, the algorithm will make sure that measurements with excitation in both sensors are selected, since $\Delta_\omega > 0$ means that Sensor 1 has larger angular rate than Sensor 2 and vice versa for $\Delta_\omega < 0$. The gyroscope sample selection method is described in Algorithm 2. In essence, the algorithm picks half the required points from either end of the sorted list.

Algorithm 2 Gyroscope sample selection

Require: Gyroscope data y_ω^N, number of allowed measurements, N_{max}, window size n.

1: **if** $N > N_{max}$ **then**

2: 　　Compute $s_\omega(k), \forall k$ according to (43).

3: 　　Obtain the sorted order k' such that $s_\omega(k') \leq s_\omega(k' + 1), \forall k' \in (1, N - 1)$.

4: 　　Remove the $N - N_{max}$ samples $y_\omega(t_k), \forall k' \in (N_{max}/2 + 1, N - N_{max}/2)$ from y_ω.

5: **end if**

6: **return** $y_\omega^{N_{max}}$

5.2. Accelerometer

The acceleration constraint is accurate when the angular rate and angular accelerations are small, since that makes the right hand side of (17) vanish. Note that linear acceleration terms in (17), which are collected in a_0, always cancel out. Therefore, we do not use the energy of the accelerometer measurements to determine if the acceleration constraint is valid. Instead, we give each acceleration measurement a penalty based on the average angular rate energy

$$E_i(k) = \begin{cases} \frac{1}{2n+1} \sum_{l=k-n}^{k+n} \|y_{\omega,i}(t_l)\|^2, & n < k \leq N - n \\ \infty, & \text{otherwise} \end{cases}, \tag{45}$$

where the average is calculated from a window of size $2n + 1$, centered around each sample. This angular rate energy statistic has been shown to be an effective detector of stationarity in foot-mounted inertial navigation [36], so-called zero-velocity detection.

Small $E_i(k)$ indicate that Sensor i is stationary. For the hinge joint system, it is sufficient for one sensor to be stationary since the acceleration components in the plane normal to the joint axis does not change the r.h.s of (17). If one sensor is stationary, then the other sensor can only have accelerations that are induced by independent rotation, which has to be in the plane. For this reason, the penalty given to each pair of acceleration measurements is chosen as

$$s_a(k) = \min\{E_1(t_k), E_2(t_k)\}. \tag{46}$$

As a first step of the accelerometer sample selection, measurements with $s_a(k) > E_{\text{th}}$ are removed, where E_{th} is a scalar threshold parameter, which should be chosen to remove measurements for which it is likely that the motion violates the acceleration constraint.

We also need to consider the conditions for identifiability of the joint axis. That is, we want our measurements to increase the separation between the column-space and the null-space of the matrix A in (27). In practice, A will have full rank regardless of the motion, since the measurements are corrupted by noise and bias and the acceleration constraint does not hold for arbitrary motions. However, if A has one singular value that is relatively small compared to the other singular values, it can be considered to be approximately rank 5. Consider the singular value decomposition (SVD) of A

$$A = U\Sigma W^\top, \tag{47}$$

where the diagonal elements σ_1 to σ_6 of $\Sigma \in \mathbb{R}^{M \times 6}$ are the singular values and the columns of U and W represents orthonormal bases in $\mathbb{R}^N$ and $\mathbb{R}^6$, respectively. The columns of W are known as the right-singular vectors of A, and each is associated with a corresponding singular value, i.e., if

$$\text{diag}(\Sigma) = \begin{bmatrix} \sigma_1 & \sigma_2 & \sigma_3 & \sigma_4 & \sigma_5 & \sigma_6 \end{bmatrix}, \tag{48}$$

$$W = \begin{bmatrix} w_1 & w_2 & w_3 & w_4 & w_5 & w_6 \end{bmatrix}, \tag{49}$$

the right-singular vector w_1 is associated with σ_1. The singular values are ordered $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_6 \geq 0$. We have that w_1 is the direction in $\mathbb{R}^6$ where the rows of A are most coherent, meaning that

$$w_1 = \arg\max_{w, \|w\|=1} |Aw|, \tag{50}$$

which has the interpretation that w_1 is the direction that is most separated from the null-space of A. The information about j that is contained in A is directly linked to the separation between the null-space and the column space of A. The intuition behind this can be seen by comparing the system of linear equations in (27) to the definition of w_1 in (50), where it appears most unlikely that j should be parallel with w_1. In fact, the least-squares estimator for j given by

$$\widehat{j} = \arg\min_{j} \|Aj\|^2, \tag{51}$$

has solutions on the line in $\mathbb{R}^6$, which is spanned by w_6, the right-singular vector associated with the smallest singular value. If we add the constraints $\|j_1\| = \|j_2\| = 1$ the two solutions with correct sign pairing, corresponding to (j_1, j_2) and $(-j_1, -j_2)$ can be obtained through normalization. A problem arises when multiple singular values are close to zero, in which case the value of $\|Aj\|^2$ will be small in more than one direction, and the uncertainty in the estimate increases. If A is only allowed to have N_{max} rows, we should therefore only remove measurements whose rows in A are most coherent with w_1, the direction with most information. This way, we make sure that space is always allocated for measurements with rows that do not align with w_1, which over time should increase the discrepancy between the two smallest singular values and increase the certainty of the least-squares estimator.

The coherence between a row in A and the right-singular vector w_1 is computed as the vector $c \in \mathbb{R}^M$, with the elements

$$c_k = \frac{|A_k w_1|}{\|A_k\| \|w_1\|}, \tag{52}$$

$$A_k = \begin{bmatrix} y_{a,1}^\top(t_k) & -y_{a,2}^\top(t_k) \end{bmatrix}, \tag{53}$$

where A_k is the kth row vector in A, and c_k has a value of 1 if A_k is parallel to w_1 and 0 if they are orthogonal. A $c_k > 0.5$ means that A_k has most of its magnitude in the direction of w_1. Therefore, we choose to remove measurements with the largest $s_a(k)$ where $c_k > 0.5$. This ensures that we also keep good measurements in the w_1 direction, while allocating space for measurements with new information about j. The algorithm for selecting accelerometer samples is described in Algorithm 3.

Algorithm 3 Accelerometer sample selection

Require: Data $\mathcal{D}^N = \{y_\omega^N, y_a^N\}$, number of allowed measurements $N_{\max}$, window size n, threshold

$\quad E_{\text{th}}$.

1: **if** $N > N_{\max}$ **then**
2: $\quad$ Compute $s_a(k), \forall k$ according to (46) using window size n.
3: $\quad$ Remove measurements where $s_a(k) > E_{\text{th}}$ from a.
4: $\quad N \leftarrow |y_a|$.
5: $\quad$ **while** $N > N_{\max}$ **do**
6: $\quad\quad$ Compute the SVD $A = U\Sigma W$, with A given by (27).
7: $\quad\quad$ Compute the coherence c according to (52).
8: $\quad\quad$ Remove the measurement with largest $s_a(k)$ where $c_k > 0.5$ from a.
9: $\quad\quad N \leftarrow |y_a|$. $\qquad\qquad\qquad\qquad\qquad\qquad$ $\triangleright$ A changes in subsequent iterations.
10: $\quad$ **end while**
11: **end if**
12: **return** $y_a^{N_{\max}}$.

5.3. Online Implementation

The two proposed sample selection algorithms can be implemented for an online application. For Algorithm 2, simply save the scores s_ω and re-use them when a new batch of data is available, new s_ω only needs to be computed for the previously unseen measurements. The same principle holds for Algorithm 3 and s_a.

6. Uncertainty Quantification

When identifying an unknown quantity, it is useful for the user of the method to know if they can expect their estimate to be accurate given the data that is available, or if more informative data needs to be collected. Here we propose a method for quantifying both local and global uncertainty of an estimate $\hat{x}$.

The local uncertainty is obtained through estimating the covariance matrix of the estimation errors using the Jacobian of the cost function. Global uncertainty is obtained through solving multiple parallel or sequential optimization problems with different random initializations, then comparing the resulting estimates to see if they correspond to the same joint axis.

The local and global uncertainty metrics are combined into an algorithm that can be used to determine if a current estimate is of acceptable accuracy or if more informative data needs to be collected.

6.1. Local Uncertainty

We approximate the cost function $V(x)$ (23) as a quadratic function near the estimate $\hat{x}$

$$\widehat{V}(x) = V(\hat{x}) + \frac{1}{2}(x - \hat{x})^\top H(\hat{x})(x - \hat{x}), \tag{54}$$

where $H(\hat{x}) \in \mathbb{R}^{4 \times 4}$ is the approximate Hessian of $V(x)$ evaluated at $\hat{x}$ according to (34).

We make the assumption that the uncertainty can be captured by a Gaussian distribution. Given the estimate $\widehat{x}$ and the covariance matrix P_x, the probability that x is the true parameter vector is given by the probability density function (PDF)

$$p(x|\widehat{x}, P_x) = \mathcal{N}(\widehat{x}, P_x) = \frac{1}{\sqrt{(2\pi)^4 |P_x|}} \exp\left(-\frac{1}{2}(x - \widehat{x})^\top P_x^{-1}(x - \widehat{x})\right). \tag{55}$$

This is the same as assuming the estimation errors $x - \widehat{x}$ to be zero-mean Gaussian with covariance P_x. We are interested in finding P_x to quantify the uncertainty of estimates. We now consider the negative log-likelihood of this PDF

$$-\log p(x|\widehat{x}, P_x) = \log\left(\sqrt{(2\pi)^4 |P_x|}\right) + \frac{1}{2}(x - \widehat{x})^\top P_x^{-1}(x - \widehat{x}). \tag{56}$$

Note the similarities to $\widehat{V}(x)$ in (54). If (54) is a good local approximation of the cost function and our estimator is unbiased, the distribution of the estimation errors $x - \widehat{x}$ will be asymptotically ($N \to \infty$) zero-mean Gaussian with covariance matrix [37]

$$P_x \approx \left(J_s(\widehat{x})^\top J_s(\widehat{x})\right)^{-1}, \tag{57}$$

where $J_s(\widehat{x})$ is Jacobian from (31) where the partial derivatives of the gyroscope and acceleration residuals have been scaled by $1/\text{std}(e_\omega(k, \widehat{x}))$ and $1/\text{std}(e_a(k, \widehat{x}))$, respectively. Here, $\text{std}(e(k, x))$ denotes the sample standard deviation of the residuals

$$\text{std}(e(k, x)) = \sqrt{\frac{1}{N-1} \sum_{k=1}^{N} \left(e(k, x) - \frac{1}{N} \sum_{k=1}^{N} e(k, x)\right)^2}. \tag{58}$$

We want to measure the uncertainty in terms of angular deviation

$$\text{AD}(v_1, v_2) = \cos^{-1}\left(\frac{v_1^\top v_2}{\|v_1\| \|v_2\|}\right), \tag{59}$$

where v_1 and v_2 are vectors of the same dimension, $\text{AD}(v_1, v_2)$ returns the positive angle between the two vectors. Let

$$z = h(x) = \begin{bmatrix} \text{AD}(j_1(x), j_1(\widehat{x})) \\ \text{AD}(j_2(x), j_2(\widehat{x})) \end{bmatrix}, \tag{60}$$

$$x_i = \begin{bmatrix} \theta_i & \phi_i \end{bmatrix}^\top, \tag{61}$$

then we want to find the probability distribution of $p(z)$ or its first two moments (mean μ_z and covariance matrix P_z).

We use a Monte Carlo method to estimate the mean μ_z and covariance P_z [38]

$$x_l \sim \mathcal{N}(\mu_x, P_x), \quad l = 1, \dots, L \tag{62}$$

$$z_l = h(x_l) \tag{63}$$

$$\mu_z = \frac{1}{L} \sum_{l=1}^{L} z_l \tag{64}$$

$$P_z = \frac{1}{L-1} \sum_{l=1}^{L} (z_l - \mu_z)(z_l - \mu_z)^\top, \tag{65}$$

where we let $\mu_x = \hat{x}$, P_x is obtained as in (57) and $h(x_l)$ is given by (60). The covariance matrix P_z is estimated by the unbiased sample covariance estimator, hence the division by $L - 1$.

The metric we will be using to determine local uncertainty is the mean plus two standard deviations, $\mu_z + 2\sigma_z$, where $\sigma_z = \sqrt{\text{diag}(P_z)}$.

6.2. Global Uncertainty

The cost function $V(x)$ may have multiple local minima. In the case that the local minima correspond to either the correct or the wrong sign pairing of j_1 and j_2, we can find the correct one by comparing minima located near the opposite sign of either j_1 or j_2. If these minima are not distinctly different in terms of the values of $V(x)$, we expect the estimates to have the correct sign half of the times our method finds a solution given that the initial estimates are uniformly spread over the parameter space. Furthermore, in the case where our data has little information about j, there may be other local minima that corresponds to wrong solutions. Wrong local minima can still have low local uncertainty, meaning that if our estimates are initialized near them, it is likely that wrong solutions are found. Therefore, to be confident that the method has found the global minimum, we need to solve the optimization problem multiple times with different initial estimates and compare the angular deviations of the sequential estimates.

We compute estimates $\hat{j}_i^{(t)}$ for $t = 1, 2, \ldots, T$ as

$$\hat{j}^{(t)} = \begin{cases} \hat{j}, & t = 1 \\ \arg\min_{j \in \{+\hat{j}, -\hat{j}\}} \min_{i \in \{1,2\}} AD(j_i, \hat{j}_i^{(t-1)}), & t > 1 \end{cases}, \tag{66}$$

where $\hat{j}_i^{(t)}$ is chosen as either $\pm\hat{j}$, such that one of the two estimated joint axes $\hat{j}_i$ always has the sign that is most consistent with its previous estimate. Note that this only forces either $\hat{j}_1$ or $\hat{j}_2$ to be consistent with the previous estimate, whereas the other one may still be inconsistent. We then consider the maximum sequential angular deviation as our metric for whether the estimate at time t corresponds to the same minimum as the estimate at time $t - 1$

$$\text{SEQAD}(t) = \begin{cases} 180°, & t \leq 1 \\ \max_{i \in \{1,2\}} AD(\hat{j}_i^{(t)}, \hat{j}_i^{(t-1)}), & t > 1 \end{cases}. \tag{67}$$

The SEQAD(t) metric corresponds to the angular deviation of the joint axis estimate that is most inconsistent with its previous estimate. Consecutive estimates will differ when there is no clear and consistent global minimum. Therefore, if we observe that SEQAD$(t) \to 0$ as t increases, we can be more certain that the local minimum found by our solver corresponds to a global minimum.

6.3. Identifying Estimates with Acceptable Uncertainty

Suppose that we receive data sequentially, i.e., we obtain $\mathcal{D}^{N(t)} = \{y_\omega^{N(t)}, y_a^{N(t)}\}$, $\forall t \in \{0, \ldots, T\}$, the sets of $N(t)$ gyroscope and $N(t)$ accelerometer measurements that have been recorded from time $t = 0$ to t. If we use sample selection according to Algorithms 2–3, then $N(t) \leq N_{\max}$, $\forall t$. For each $\mathcal{D}^{N(t)}$ we obtain an estimate $\hat{j} = j(\hat{x})$ by solving the optimization problem (22). Furthermore, we will select the estimate associated with time t to be $\hat{j}^{(t)}$ as in (66), such that either $\hat{j}_1$ or $\hat{j}_2$ is consistent with the sign of the previous estimate.

We now want to assess if $\hat{j}^{(t)}$ has acceptable uncertainty. Let $E_{\max}$ denote the maximum uncertainty that we accept. We use the following two criteria to determine if the local and global uncertainty is sufficiently small

- We require that $\mu_z + 2\sigma_z < E_{\max}$, where μ_z and σ_z are obtained from (64) and (65) through the procedure described in Section 6.1.

- We require that the sequential angular deviations given by (67) satisfy SEQAD(t) < $E_{\max}$ for a minimum of $n_{\min}$ consecutive estimates, that were randomly initialized uniformly over the parameter space. This is equivalent to

$$\max_{t' \in [t-n_{\min}+1, t]} \left(\text{SEQAD}(t') \right) < E_{\max}. \tag{68}$$

We summarize the method for selecting an estimate $\widehat{j}^{(t)}$ of acceptable uncertainty in Algorithm 4.

Algorithm 4 Identifying an estimate of acceptable uncertainty

---

Require: Data $\mathcal{D}^{N(t)} = \{y_\omega^{N(t)}, y_a^{N(t)}\}$, $\forall t \in \{1, \ldots, T\}$, number of Monte Carlo samples L, maximum acceptable uncertainty $E_{\max}$, threshold for minimum number of sequential estimates with acceptable deviation $n_{\min}$.

1: $n \leftarrow 0$
2: **for** $t \in \{1, \ldots, T\}$ **do**
3:     Obtain an estimate $\widehat{j} = j(\widehat{x})$ by solving the optimization problem (22) using the data $\mathcal{D}^{N(t)}$ and Algorithm 1.
4:     Obtain $\widehat{j}^{(t)}$ from (66).
5:     Compute the covariance matrix P_x according to (57).
6:     Compute μ_z and P_z according to the Monte Carlo method (62)–(65).
7:     Compute SEQAD(t'), $\forall t' \in (t - n_{\min} + 1, t)$ as in (67).
8:     **if** $\mu_z + 2\sigma < E_{\max}$ AND $\max\{\text{SEQAD}(t')\} < E_{\max}$ **then**
9:         **return** $\widehat{j}^{(t)}$.
10:     **end if**
11: **end for**

---

7. Experiment

7.1. Data Acquisition

Data were collected from a 3D printed hinge joint system [39] with one wireless IMU (Xsens MTw) attached to each segment; see Figure 3. The sampling rate was set to 50Hz for both IMUs. The operating ranges of the IMUs were $\pm 160\text{m/s}^2$ for the accelerometers and $\pm 21\text{rad/s}$ for the gyroscopes. The IMUs were attached in sockets such that the joint axis was parallel with the positive y-axes of the sensors and both pointing in the same direction (same sign), that is

$$j_1 = j_2 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^\top. \tag{69}$$

The data consist of 14 recorded motions, listed in order below

1. Stationary system;
2. Free rotation, stiff joint, free joint axis;
3. Sequential rotation, horizontal joint axis;
4. Sequential rotation, tilting joint axis;
5. Simultaneous planar rotation, horizontal joint axis;
6. Simultaneous planar rotation, tilting joint axis;
7. Simultaneous free rotation, free joint axis;
8. [8–14] Same motions as 1–7, respectively, but with faster rotations.

The recorded angular velocity magnitudes for these motions are shown in Figure 4. For the sequential rotation, Segment 1 was always rotated first while Segment 2 was stationary, which was

followed by the converse motion. Horizontal joint axis means that the joint axis was aligned to be approximately orthogonal to the gravitational acceleration vector. For the tilting joint axis case, the angle between the joint axis and the gravitation acceleration vector was maintained at $\approx 45°$ for the duration of the motion. For the free joint axis motions, the joint axis was not constrained to any particular orientation, but rotated freely in space. The hinge joint system was equipped with a screw, which when tightened prevented independent rotation of the two segments. The screw was tightened when the system was stationary and during the stiff joint rotations. Measurements of the transitions from one motion to another were removed from the recorded data, such that only the specified motions of interest could be isolated. The first set of stationary data was used to estimate the gyroscope bias b_ω in (3), which was then subtracted from all subsequent measurements.
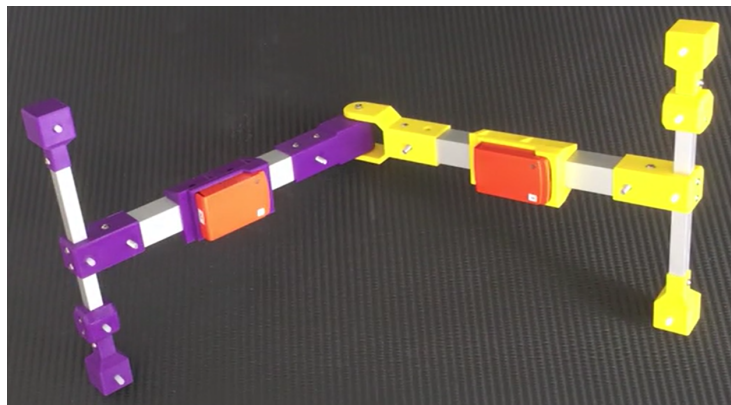
Figure 3. The 3D printed hinge joint system, design by Dustin Lehmann, with the two IMUs (orange boxes, 34×58mm) attached.
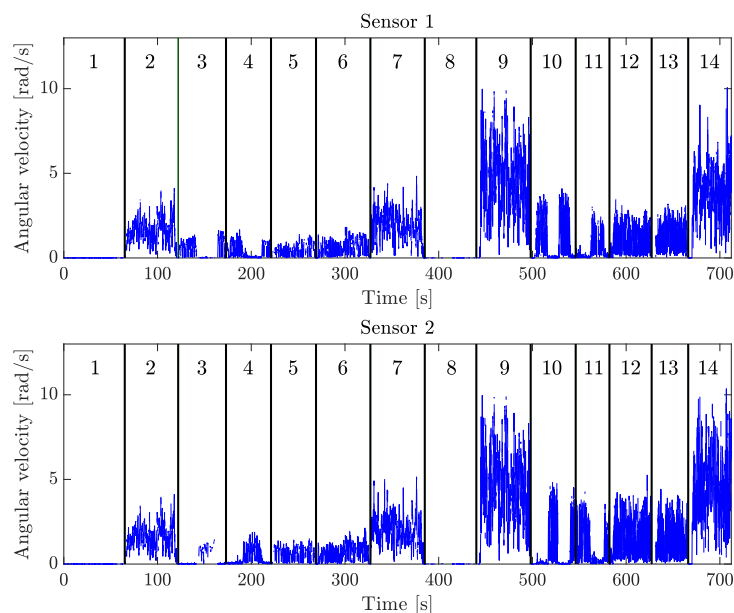
Figure 4. The angular velocity magnitudes for the 14 different motions that were recorded. The vertical lines and numbered sections indicate when the different motions begin and end.

Because the optimization problem (22) is formulated based on the kinematics at each sample time, and does not contain dynamics, we are allowed to shuffle around the measurements in our data set. Using the 14 different motions, four scenarios were designed where the motions appeared in different sequences. The sequences of motion for the different scenarios were

1. $1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14$;
2. $1, 3^*, 10^*, 8, 2, 9, 3, 10, 4, 11, 5, 12, 6, 13, 7, 14$;
   *only 500 samples (10 seconds) from motions 3 and 10, which contain motion in only Segment 1.
3. $6^*, 1, 8, 2, 9$;
   *only 1000 samples from motion 6.
4. $1, 8, 2^\dagger, 9^\dagger, 6^*, 2^\dagger, 9^\dagger$.
   *only 1000 samples (20 seconds) from motion 6.
   †samples divided in half.

Scenario 1 is the original sequence in which the motions were recorded. Scenario 2 starts with the sensors being stationary, then there is motion in only Segment 1, after which the system alternates between the slower and faster motions, starting with non-informative stiff-joint motions. For this scenario we expect to have good estimates of j_1 before we have any excitation in Segment 2. Scenario 3 has early excitation of both segments followed by measurements from a stationary system and non-informative stiff joint motion and Scenario 4 has the converse case where the excitation comes late in the sequence. Scenarios 3 and 4 are also designed to contain more non-informative motions, as the only informative motion is contained in the 1000 samples (20 seconds) from motion 6.

7.2. Evaluating Robustness of the Residual Weighting

To experimentally evaluate the robustness of the proposed method for different weights w_ω and w_a, we estimated the joint axis using data from motions 3 to 7 and 10 to 14. Data from a stationary system and rotations with a stiff joint were not used in these evaluations since the joint axis is not identifiable for these motions. The weights were chosen as

$$w_\omega = \sqrt{w_0}, \quad w_a = \frac{1}{\sqrt{w_0}}, \tag{70}$$

where we let $w_0 = \frac{w_\omega}{w_a}$ determine the relative weighting of the residuals e_ω and e_a. We estimated the joint axis for 100 different values of w_0, which had a logarithmic distribution on the interval $(10^{-3}, 10^{10})$. For all different motions and for each value of w_0 the initial estimates of x were selected deterministically such that all possible sign pairings $(\pm j_1, \pm j_2)$ and $(\pm j_1, \mp j_2)$ were selected equally many times. The initial estimates for j_1 and j_2 were selected from a grid on the unit-sphere of $\mathbb{R}^3$ with 6 grid points the positive and negative axes. With all possible combinations for j_1 and j_2 this resulted in $M = 36$ different initial conditions for the optimization algorithm. Here we use the root-mean-square angular error (RMSAE) for both joint axes as the metric to evaluate performance

$$\text{RMSAE} = \sqrt{\frac{1}{2M} \sum_{k=1}^{M} \text{AD}(j_1, j(\widehat{x}_1^{(k)}))^2 + \text{AD}(j_2, j(\widehat{x}_2^{(k)}))^2}, \tag{71}$$

where AD is the angular deviation metric given by (59), and here we let the superscript k denote the estimates obtained from different initializations. Since we consider $(\pm j_1, \pm j_2)$ to be correct sign pairings of the joint axis, we select the sign of $\widehat{j}_1$ which has the lowest AD. If, as a result of this, $\widehat{j}_1$ changes sign, the sign of $\widehat{j}_2$ is also changed. This way AD for $\widehat{j}_1$ will always be $\leq 90°$ whereas the AD for $\widehat{j}_2$ can be up to $180°$, which corresponds to an error in the sign pairing.

7.3. Evaluating Sample Selection

To evaluate the proposed sample selection in Algorithms 2–3, we use the data according to the four scenarios specified in Section 7.1. Starting with the first second of recorded motion and incrementally

adding subsequent data in small batches of one second at a time. That is, we receive sequential batches of data $\mathcal{D}^{N(t)}, \forall t \in \{1, \ldots, T\}$ where T is the duration of the scenario.

We compute one estimate $\widehat{j}_i^{(t)}$ for each $\mathcal{D}^{N(t)}$. For each new batch, the joint axis is estimated again starting from a new initial estimate (i.e., no warm-start of the optimization method), which is randomly selected from a uniform distribution. The reason for this is that we also want to evaluate if the estimates are consistent over time, regardless of initialization of the optimization method. The relative weighting of the residuals (70) was set to $w_0 = 50$.

We compare the method when the proposed sample selection is used to the case where all available data is used ($N_{\max} = N(t)$) for estimation for all four scenarios. When using the sample selection in Algorithms 2–3, the maximum sample sizes of $N_{\max} \in \{1000, 500, 250, 125\}$ were compared.

Other than N, the other user chosen parameters for the sample selection is related to the angular rate energy penalty (45). The window size of $n = 21$ samples was used, which means the average energy is computed for 0.42s of motion for our sensors. The threshold used to determine if the accelerometer is stationary in Algorithm 3, was set to $E_{\text{th}} = 1\text{rad}^2\text{s}^{-2}$. This is around 10 times higher than the threshold used for the angular rate energy detector suggested for zero-velocity detection in human gait [40]. Measurements that we discard in line 2 of Algorithm 3 are therefore likely to be of significant motion.

7.4. Evaluating Uncertainty Quantification

To evaluate the efficacy of the proposed uncertainty quantification, we will use the same sequential batches of data for all four scenarios $\mathcal{D}^{N(t)}, \forall t \in \{1, \ldots, T\}$ as in Section 7.3, but with a fixed maximum number of samples $N = 1000$ chosen by Algorithms 2–3. Similarly to the procedure used to evaluate the sample selection method one estimate $\widehat{j}_i^{(t)}$ is obtained for each new batch, and initial estimates are independently randomized from a uniform distribution over the parameter space at each t. The relative weighting of the residuals (70) was set to $w_0 = 50$.

Here we use Algorithm 4, which returns an estimate $\widehat{j}$, when the local and global criteria indicate that the uncertainty is acceptable. This requires the user to choose the threshold for acceptable uncertainty, $E_{\max}$, and the minimum number of sequential estimates that should have angular deviations below this threshold, $n_{\min}$. For our evaluation we choose to set $E_{\max} = 3°$, and $n_{\min} \in \{3, 10\}$. Algorithm 4 is then deemed to be successful if $\text{AD}(j_i, \widehat{j}_i) \leq E_{\max}$. This procedure is repeated 100 times for each scenario, with different randomized initial estimates each time.

7.5. Evaluating Robustness to Sensor Bias

We evaluated the robustness of the complete method, which includes Algorithms 1–4, to measurement bias. The measurement bias refers to b_a and b_ω in the measurement models (1)–(3). In the other evaluations presented here, we have compensated for gyroscope bias by estimating b_ω from the initial stationary data (Motion 1) and subtracting this bias from the subsequent measurements. We have not compensated for any accelerometer bias since it cannot be estimated from only one stationary position of the sensor. In this section, we will study the effect of sensor biases by adding artificially generated biases to both the previously bias-compensated gyroscope measurements and to the accelerometer measurements. These artificial biases have fixed magnitudes $\|b_a\| = 1\text{m/s}^2$ and $\|b_\omega\| = 1°\text{/s}$, but their directions are randomized by generating random unit vectors. To evaluate the effect of the added artificial bias, $M = 100$ estimation runs are performed for all four scenarios, with and without the added artificial bias. The artificial biases are first added to the measurements, then the proposed method is applied as described in Section 7.4. Here, we set $N_{\max} = 500$, $E_{\max} = 1°$ and $n_{\min} = 10$, other parameters are the same as in previous sections. We will use the RMSAE metric (71) and the maximum angular error (MAXAE)

$$\text{MAXAE} = \max_{1 \leq k \leq M} \{\text{AD}(j_1, j(\widehat{x}_1^{(k)})), \text{AD}(j_2, j(\widehat{x}_2^{(k)}))\}, \tag{72}$$

to evaluate the performance of the method across all $M = 100$ estimation rounds for all scenarios.

8. Results

8.1. Robustness

The RMSAE for the different motions and weights w_0 are shown in Figure 5. Here, estimation is done separately and using all samples for each motion, i.e., without sample selection. The optimal choice for w_0 appears to be in the interval of $(10^1, 10^5)$, where the errors are small for all motions.
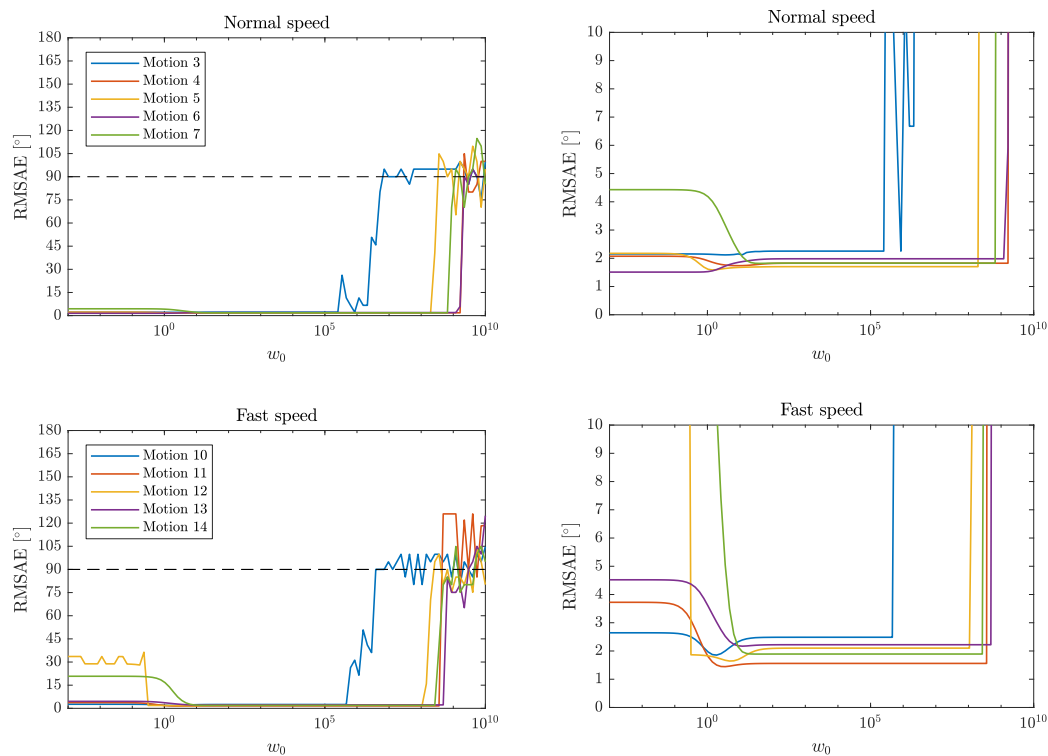

Figure 5. Root-mean-square angular error (RMSAE) (71) for different motions and weights w_0. Normal speed motions are shown in the top plots and faster motions are shown in the bottom plots. The plots on the right show the same results as the plots to their respective lefts, but zoomed in.

8.2. Sample Selection

The angular errors for $\widehat{j}_1$ and $\widehat{j}_2$ obtained from testing the proposed sample selection as described in Section 7.3, are shown in Figure 6. From these results, we can compare the use of Algorithms 2–3 for different sample sizes N_{max}. This includes the case of $N_{max} = N(t)$, where $N(t)$ corresponds to making use of all samples that have been observed up to an integer t number of seconds.

Figure 7 shows which samples were selected for $N_{max} = 1000$, for Scenarios 1 and 2 at the times given by the vertical axes.

8.3. Uncertainty Quantification

With the parameter $n_{min} = 10$, the final errors were below $E_{max} = 3°$ for all 100 estimation rounds for all four scenarios, meaning the estimates obtained from Algorithm 4 were acceptable 100% of the time. With $n_{min} = 3$ and the same E_{max}, the estimates were acceptable 8% of the time for Scenario 1, 81% of the time for Scenario 2, 100% of the time for Scenario 3 and 0% of the time for Scenario 4. Figure 8 compares the local and global uncertainty metrics to the angular errors of a single estimation round for each scenario and shows when Algorithm 4 accepted an estimate $\widehat{j}$ for $n_{min} = 3$ (leftmost vertical dashed lines) and for $n_{min} = 10$ (rightmost vertical dashed lines).
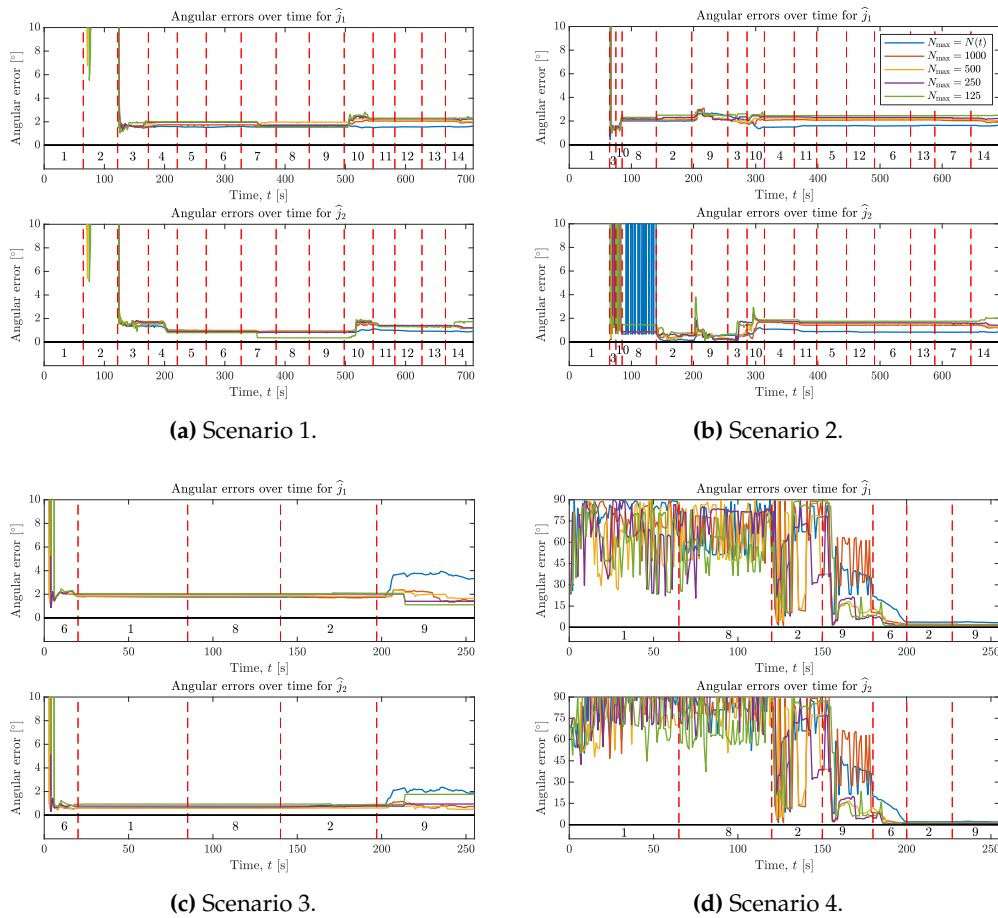
(a) Scenario 1.

(b) Scenario 2.

(c) Scenario 3.

(d) Scenario 4.

Figure 6. Angular errors over time for the four scenarios, see (**a**–**d**). Comparing the case $N_{\max} = N(t)$, where all samples up to time t are used for estimation to $N_{\max} \in \{1000, 500, 250, 125\}$ samples being chosen by Algorithms 2–3 at each integer t seconds. Colored lines define these different cases as given by the legend in the top right. Vertical dashed lines and the numbers 1–14 are used to indicate from which motions (see Section 7.1) the data comes from.

(a) Scenario 1.
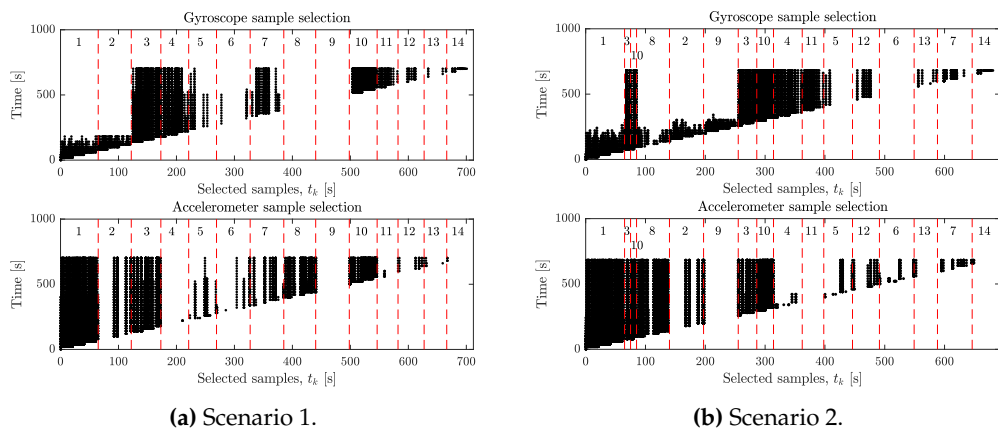
(b) Scenario 2.

Figure 7. The figure shows which samples from Scenario 1 (**a**) and Scenario 2 (**b**) that were selected by Algorithms 2–3 with $N_{\max} = 1000$, at the times given by the vertical axes. Black/white indicates that a sample were selected/not selected respectively. As time increases and more samples become available, we see some previously selected samples being deselected in favor of new samples that are deemed superior by the algorithms. Vertical dashed lines and the numbers 1–14 indicate from which motions (see Section 7.1) the data comes from.
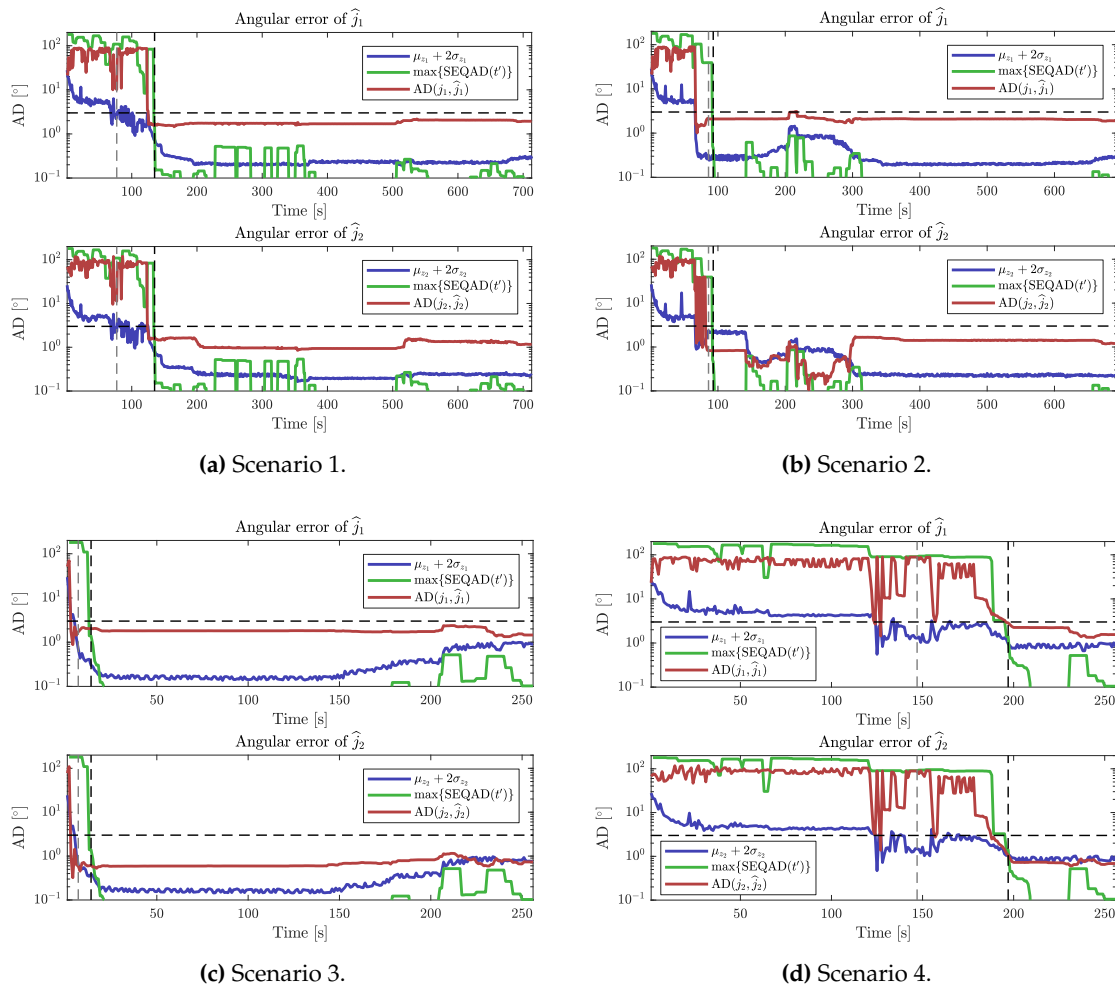
(a) Scenario 1.

(b) Scenario 2.

(c) Scenario 3.

(d) Scenario 4.

Figure 8. The plots shows the local and global uncertainty metrics compared to the angular errors (red) for the four scenarios, see (**a**–**d**). Local uncertainty is quantified by $\mu_z + 2\sigma_z$, where μ_z is the estimated mean AD (64) and σ_z is the standard deviation, computed from the estimated covariance matrix (65). Local uncertainty (blue) is shown for both $\widehat{j}_1$ and $\widehat{j}_2$ for each scenario. Global uncertainty is quantified by (68). The global uncertainty (green) with $n_{\min} = 10$ is shown for each scenario. Horizontal dashed lines show the accuracy threshold $E_{\max} = 3°$. Vertical dashed lines show when estimates $\widehat{j}$ were accepted by Algorithm 4. For each scenario, the leftmost vertical lines show the case of $n_{\min} = 3$ and the rightmost vertical lines show the case of $n_{\min} = 10$, where Algorithm 4 terminates when the estimates have reached the desired accuracy w.r.t. ground truth.

8.4. Robustness to Sensor Bias

The resulting RMSAE and MAXAE for the $M = 100$ estimation rounds with and without added artificial bias are shown for all four scenarios in Table 1. Without the added artificial biases, the errors were at most $2.16°$, and with the added artificial biases of magnitudes $\|b_a\| = 1\mathrm{m/s^2}$ and $\|b_\omega\| = 1°/\mathrm{s}$ the errors were at most $4.84°$.

Table 1. Shows the RMSAE (71) and MAXAE (72) after $M = 100$ runs with and without artificial bias for the four scenarios.

Scenario	$\|b_a\|$ [m/s^2]	$\|b_\omega\|$ [°/s]	RMSAE [°]	MAXAE [°]
1	0	0	1.55	1.67
1	1	1	1.73	4.41
2	0	0	1.58	2.16
2	1	1	1.97	4.84
3	0	0	1.50	2.07
3	1	1	1.58	3.09
4	0	0	1.47	1.98
4	1	1	1.30	2.32

9. Discussion

9.1. The Method Is Not Sensitive to the Relative Weighting w_0

The parameter w_0, which is defined from (70), controls the relative weighting of the residuals e_ω and e_a. As w_0 increases, the relative weighting of the gyroscope residual is increased. As we see in Figure 5, the optimal choice of w_0 for most motions in terms of RMSAE (71), is somewhere in the large range between 10 and 10^5. The errors are also small ($< 3°$) for $w_0 < 10$ for the slower planar motions (3–6), which shows that the acceleration information can be reliable for these motions. However, some larger errors can be observed for small w_0 for the faster planar motion 12 and the errors are also significantly larger for the free axis rotations (motions 7 and 14), which can be explained by the fact that these motions violate the acceleration constraint, meaning that the r.h.s. of (17) is nonzero.

Since we can select any w_0 from within such a large interval and still obtain similar performance, our method is not sensitive to the relative weighting of the residuals. It makes sense that $w_0 > 10$, since the angular velocities, measured in rad/s, have smaller magnitudes than the accelerations, that typically fluctuate around 9.82m/s^2 due to the gravitational acceleration. Furthermore, the angular velocity constraint always holds for a rigid hinge joint system. Hence, we expect the angular velocity information to be more reliable. The method is robust for larger w_0 up to 10^5 where the RMSAE become large for motions 3 and 10. This large increase in RMSAE occurs when the acceleration residual becomes numerically indistinguishable to the tolerance of the optimization algorithm, and it becomes more likely that the method selects an estimate which corresponds to the wrong sign pairing. Therefore, as w_0 increases we see the RMSAE approach $90°$ as the AD for $\widehat{j}_1$ is still small but the probability of selecting $\pm\widehat{j}_2$ is approaching 0.5, meaning that approximately half of the estimates will have the wrong sign pairing. This can also depend on the numerical tolerance and stopping criteria of the optimization method, since a global minimum corresponding to the correct sign pairing might not be significantly different from other local minima that correspond to the wrong sign pairing.

9.2. Sample Selection Offers Substantial Benefits

From the results shown in Figure 6 we see that we can achieve similar, and in some cases even better performance by selecting relatively few measurements to use for estimation out of all $N(t)$ measurements that have been observed up to time t. For Scenario 1, $N_{\max} \in \{1000, 500, 250\}$ have angular errors within $0.5°$ and $N = 125$ have errors within $1°$ from the case with $N_{\max} = N(t)$.

In Scenario 2, the errors for $\widehat{j}_2$ drop below $2°$ at $t = 85$s for the methods using sample selection, but it takes until $t = 135$s for the method where $N_{\max} = N(t)$ to stay consistently below $2°$. However, the method with $N = 125$ again shows a slightly larger deviation from the others, with some momentary spikes in error around $t = 200$s and $t = 300$s. Note that Scenario 2 is designed to have no independent rotation of Segment 2 until $t = 255$s. So the only information about j_2 until then has to come from the accelerometer. This shows that carefully selecting accelerometer samples according to Algorithm 3 is

beneficial, especially if angular velocity information is missing. Comparing the results from Scenarios 1 and 2, the final errors are very similar, indicating that the methods are not sensitive to the sequence of motions.

Scenarios 3 and 4 represent challenging cases where only a small minority of samples contain motion with independent rotation (only 20s of motion 6). In Scenario 3, we note that the final error for $N_{\max} = N(t)$ is significantly larger than the cases with $N_{\max} \in \{1000, 500, 250\}$, and for $N_{\max} = 125$ the final error for $\hat{j}_2$ is at the same level as $N_{\max} = N(t)$. Scenario 4 has a similar performance in terms of final errors. However, Scenario 4 does not have any motions with independent rotations of the segments until around $t = 180$s. The only information about the joint axis until that point comes from the accelerometer, in Motions 1, 8, 2 and 9. The errors start to decrease around $t = 150$s when data from Motion 9 comes in, but do not settle until after Motion 6. The large fluctuations in errors we see for $N_{\max} = 1000$ during Motion 9 indicate that there are still at least two local minima corresponding to the wrong joint axis at this point. Errors for $N_{\max} < 1000$ are smaller during motion 9, but still vary between $5°$ and $20°$.

Using Algorithms 2–3 is therefore beneficial, not only for reducing the computational complexity of the optimization problem, but it can even improve the performance in situations where gyroscope information is limited. However, judging by Scenario 4 in particular, $N_{\max} \geq 1000$ appears to be the best choice in terms of overall performance. Even then, $N_{\max} = 1000$ is only a small fraction of the total number of measurements. With a sample period of 0.02s, we have that $N(t = 700) = 35000$ and $N(t = 250) = 12500$.

Figure 7 shows the samples that were selected over time from Scenarios 1 and 2 with $N_{\max} = 1000$ and which motions these samples come from. For both scenarios, we see that gyroscope samples from non-informative motions 1,2,8,9 are all deselected by the end. Samples from these motions are only kept until enough samples from informative motions have been parsed by the algorithm. For the accelerometer, we see that samples from stationary sensors are preferred since many samples of motions 1 and 8 are kept, which is in line with the penalty we give samples based on the angular rate energy. It is also important that samples from other motions are selected since the criterion for identifiability requires a strong separation between the nullspace and the column space of the matrix A given by (27). Had the selection criterion of the accelerometer only been based on the angular rate energy, we would risk ending up in the situation where all samples are selected from the same stationary position, in which case all rows of A are linearly dependent. Lines 5–10 in Algorithm 3 prevent this by removing the worst samples that are coherent with the right-singular vector of the largest singular value. This can be thought of as allocating space in the A matrix for novel information by removing redundant information.

9.3. Reliability of the Proposed Uncertainty Quantification

We obtained reliable estimates with errors below that of the maximum acceptable error $E_{\max} = 3°$, 100% of the time when the parameter $n_{\min} = 10$. However, estimates were not reliable for $n_{\min} = 3$, where the results were particularly bad for Scenario 1, with 8% of estimates of acceptable error and for Scenario 4 with 0% of estimates of acceptable error. Both of these scenarios contained no informative motions in the beginning, and we found that the estimates that were returned often had not used any batch of informative data for estimation because the criteria for local and global uncertainty were satisfied prematurely by Algorithm 4.

In Figure 8 it can be seen that local uncertainty metric $\mu_z + 2\sigma_z$ can be below $E_{\max}$ (horizontal dashed lines) while the actual angular error fluctuates between values below and above $E_{\max}$. This occurs when there exist multiple other local minima than those corresponding to the true joint axis. Furthermore, Figure 8 shows how the SEQAD remains large as the angular errors fluctuate in the same way as the angular errors. Interestingly, Scenario 2 appears to fluctuate between one correct and one wrong local minimum between $t = 66$s and $t = 82$s. If we assume that the probability of finding the correct local minimum is 0.5, having $n_{\min} = 3$ that means that the probability of ending up in

the wrong local minimum n_{min} times in a row is $0.5^{n_{min}} = 12.5\%$. This matches well with the results obtained for Scenario 2, where 88% of the estimates were acceptable for $n_{min} = 3$.

For Scenarios 1 and 4, where the results were significantly worse for $n_{min} = 3$, it appears that wrong local minima were dominating. These two scenarios have sequences of stiff-joint motion before any informative motions are observed, which can explain why wrong local minima were found more frequently. Scenario 3, which had informative motion in the beginning did not have this issue, and hence 100% of the estimates were acceptable even for $n_{min} = 3$.

We can conclude that setting the parameter n_{min} sufficiently large is important for fully capturing the global uncertainty. Sequential data dominated by non-informative motions in the beginning are more sensitive to the choice of n_{min}. The results showed that Algorithm 4 successfully identified all of the estimates that satisfied the accuracy criteria $E_{max} = 3°$ when $n_{min} = 10$. Here, this corresponds to 10 consecutive estimates (computed once per second), that differed by $3°$ at most.

9.4. The Method Is Robust to Realistic and Uncompensated Sensor Bias

As shown in Table 1, even with added artificial biases of relatively large magnitudes $\|b_a\| = 1 \text{m/s}^2$ and $\|b_\omega\| = 1°/\text{s}$, the errors were at most $4.84°$ across all $M = 100$ estimation runs for all four scenarios. The average errors in terms of RMSAE were less than $2°$ even with the added artificial biases. As a comparison, the IMUs used in our experiments had bias magnitudes in the order of $\|b_a\| = 0.1 \text{ m/s}^2$ and $\|b_\omega\| = 0.5°/\text{s}$, so the artificial biases were significantly larger. This shows that the method is robust to sensor biases of at least these magnitudes. However, we had to lower the threshold E_{max} from $3°$ to $1°$ to achieve this. This means that Algorithm 4 will be more conservative in selecting an estimate. With added artificial bias and $E_{max} = 3°$, the method would sometimes terminate prematurely, when no informative motion had been observed because a global minimum that satisfied this threshold value was found. Therefore, lowering E_{max} was required to achieve robustness to the added artificial biases. It is therefore still highly recommend that pre-calibration of the biases is performed when possible. If bias drift is significant enough to exceed the magnitudes tested here across the duration of the experiment, it is advised to use a method that allows for online compensation of biases alongside the proposed method. Lowering E_{max} is only an optional measure one would take in the unusual case where late bias occurs and is not compensated for.

10. Conclusions

We have proposed a method which facilitates plug-and-play sensor-to-segment calibration for two IMUs attached to the segments of a hinge joint system. The method identifies the direction of the joint axis j in the intrinsic reference frames of each sensor, thus providing the user with information about the sensors' orientation with respect to the joint. Accurate sensor-to-segment calibration is crucial for tracking the motion of the segments.

The method was experimentally validated on data collected from a mechanical joint, which performed a wide range of motions with different identifiability properties. As soon as sufficiently informative data was available, the method achieved a sensor-to-segment calibration accuracy in the order of $2°$, assessed as the angular deviation from the ground truth of the joint axis.

The proposed method includes the following features that were evaluated separately using the experimental data:

- Gyroscope and accelerometer information are weighted and combined, which makes the joint axis identifiable for a wider range of different motions. Experimental evaluation showed that the method is not sensitive to the weighting parameters, and that it performs comparably well for a wide range of different motions across a large interval of weights.
- A method to select a smaller subset of samples to use from a long sequence of recorded motion is proposed. Samples are selected from motions that yield identifiability, and measurements of non-informative motions are automatically discarded. The experimental evaluation showed that using between 125 and 1000 samples can achieve similar and in some cases even better performance

than using all available samples collected from a long sequence of motions. Sample selection was shown to be particularly beneficial when data consisted of more non-informative than informative motions. Furthermore, using less samples for estimation reduces the computational complexity of the estimation.

- A method to quantify local and global uncertainty properties of sequential estimates, which provides the user with an estimate when criteria for acceptable uncertainty are met. The method successfully identified estimates that satisfied the uncertainty criteria ($E_{\max} = 3°$).

The proposed method is the first truly plug-and-play calibration method that directly enables plug-and-play motion tracking in hinge joints. For the first time, the user can simply start using the sensors instead of performing precise or sufficiently informative motion in a predefined initial time window, and the proposed method provides reliable calibration parameters as soon as possible, which immediately enable calculation of accurate motion parameters from the incoming raw data as well as from already recorded data. Regardless of the performed motion, it provides only parameters that are actually accurate, which is not guaranteed by any state of the art method. This enables the kind of truly non-restrictive and reliable motion tracking that is needed in a range of application domains including ubiquitous motion assessment to wearable biofeedback systems.

In future work, the method could be extended to different joint types and be applied to motion tracking in mechatronic and biomechanical systems. For the latter case in particular, it would be of great interest to study the reliability of the method in non-rigid systems, such as human limbs, where motion of soft tissue is significant.

References

1. Camomilla, V.; Bergamini, E.; Fantozzi, S.; Vannozzi, G. Trends supporting the in-field use of wearable inertial sensors for sport performance evaluation: A systematic review. *Sensors* **2018**, *18*, 873. [CrossRef] [PubMed]
2. Jalloul, N. Wearable sensors for the monitoring of movement disorders. *Biomed. J.* **2018**, *41*, 249–253. [CrossRef] [PubMed]
3. Valtin, M.; Seel, T.; Raisch, J.; Schauer, T. Iterative learning control of drop foot stimulation with array electrodes for selective muscle activation. *IFAC Proc. Vol.* **2014**, *47*, 6587–6592. [CrossRef]
4. Picerno, P. 25 years of lower limb joint kinematics by using inertial and magnetic sensors: A review of methodological approaches. *Gait Posture* **2017**, *51*, 239–246. [CrossRef]
5. Kortier, H.G.; Sluiter, V.I.; Roetenberg, D.; Veltink, P.H. Assessment of hand kinematics using inertial and magnetic sensors. *J. Neuroeng. Rehabil.* **2014**, *11*, 70. [CrossRef]
6. Favre, J.; Jolles, B.; Aissaoui, R.; Aminian, K. Ambulatory measurement of 3D knee joint angle. *J. Biomech.* **2008**, *41*, 1029–1035. [CrossRef]
7. O'Donovan, K.J.; Kamnik, R.; O'Keeffe, D.T.; Lyons, G.M. An inertial and magnetic sensor based technique for joint angle measurement. *J. Biomech.* **2007**, *40*, 2604–2611. [CrossRef] [PubMed]
8. Favre, J.; Aissaoui, R.; Jolles, B.M.; De Guise, J.A.; Aminian, K. Functional calibration procedure for 3D knee joint angle description using inertial sensors. *J. Biomech.* **2009**, *42*, 2330–2335. [CrossRef] [PubMed]
9. Cutti, A.G.; Ferrari, A.; Garofalo, P.; Raggi, M.; Cappello, A.; Ferrari, A. 'Outwalk': A protocol for clinical gait analysis based on inertial and magnetic sensors. *Med Biol. Eng. Comput.* **2010**, *48*, 17. [CrossRef]

10. Taetz, B.; Bleser, G.; Miezal, M. Towards self-calibrating inertial body motion capture. In Proceedings of the 19th International Conference on Information Fusion (FUSION), Heidelberg, Germany, 5–8 July 2016; pp. 1751–1759.

11. Nazarahari, M.; Rouhani, H. Semi-automatic sensor-to-body calibration of inertial sensors on lower limb using gait recording. *IEEE Sens. J.* **2019**, *19*, 12465–12474. [CrossRef]

12. Seel, T.; Schauer, T.; Raisch, J. Joint axis and position estimation from inertial measurement data by exploiting kinematic constraints. In Proceedings of the International Conference on Control Applications, Dubrovnik, Croatia, 3–5 October 2012; pp. 45–49.

13. McGrath, T.; Fineman, R.; Stirling, L. An auto-calibrating knee flexion-extension axis estimator using principal component analysis with inertial sensors. *Sensors* **2018**, *18*, 1882. [CrossRef] [PubMed]

14. Küderle, A.; Becker, S.; Disselhorst-Klug, C. Increasing the robustness of the automatic IMU calibration for lower Extremity Motion Analysis. *Curr. Dir. Biomed. Eng.* **2018**, *4*, 439–442. [CrossRef]

15. Olsson, F.; Seel, T.; Lehmann, D.; Halvorsen, K. Joint axis estimation for fast and slow movements using weighted gyroscope and acceleration constraints. In Proceedings of the 22nd International Conference on Information Fusion (Fusion), Ottawa, ON, Canada, 2–5 July 2019; pp. 1–8.

16. Nowka, D.; Kok, M.; Seel, T. On motions that allow for identification of hinge joint axes from kinematic constraints and 6D IMU data. In Proceedings of the 18th European Control Conference (ECC), Naples, Italy, 25–28 June 2019; pp. 4325–4331.

17. Laidig, D.; Müller, P.; Seel, T. Automatic anatomical calibration for IMU-based elbow angle measurement in disturbed magnetic fields. *Curr. Dir. Biomed. Eng.* **2017**, *3*, 167–170. [CrossRef]

18. Salehi, S.; Bleser, G.; Reiss, A.; Stricker, D. Body-IMU autocalibration for inertial hip and knee joint tracking. In Proceedings of the 10th EAI International Conference on Body Area Networks, Sydney, Australia, 28–30 September 2015; pp. 51–57.

19. Olsson, F.; Halvorsen, K. Experimental evaluation of joint position estimation using inertial sensors. In Proceedings of the 20th International Conference on Information Fusion (Fusion), Xi'an, China, 10–13 July 2017; pp. 1–8.

20. Graurock, D.; Schauer, T.; Seel, T. Automatic pairing of inertial sensors to lower limb segments–a plug-and-play approach. *Curr. Dir. Biomed. Eng.* **2016**, *2*, 715–718. [CrossRef]

21. Mark C. Schall, J.; Sesek, R.F.; Cavuoto, L.A. Barriers to the adoption of wearable sensors in the workplace: A survey of occupational safety and health professionals. *Hum. Fact.* **2018**, *60*, 351–362.

22. Passon, A.; Schauer, T.; Seel, T. Hybrid inertial-robotic motion tracking for upper limb rehabilitation with posture biofeedback. In Proceedings of the International Conference on Biomedical Robotics and Biomechatronics (BioRob), Enschede, The Netherlands, 26–29 August 2018; pp. 1163–1168.

23. Salchow-Hömmen, C.; Callies, L.; Laidig, D.; Valtin, M.; Schauer, T.; Seel, T. A tangible solution for hand motion tracking in clinical applications. *Sensors* **2019**, *19*, 208. [CrossRef] [PubMed]

24. Poddar, S.; Kumar, V.; Kumar, A. A comprehensive overview of inertial sensor calibration techniques. *J. Dyn. Syst. Meas. Control* **2017**, *139*. [CrossRef]

25. Kok, M.; Hol, J.D.; Schön, T.B. Using inertial sensors for position and orientation estimation. *Found. Trends®Signal Process.* **2017**, *11*, 1–153. [CrossRef]

26. El-Sheimy, N.; Hou, H.; Niu, X. Analysis and modeling of inertial sensors using allan variance. *Trans. Instrum. Meas.* **2007**, *57*, 140–149. [CrossRef]

27. Woodman, O.J. An introduction to inertial navigation. Technical Report 696; University of Cambridge Computer Laboratory: Cambridge, UK, 2007.

28. Gulmammadov, F. Analysis, modeling and compensation of bias drift in MEMS inertial sensors. In Proceedings of the 4th International Conference on Recent Advances in Space Technologies, Istanbul, Turkey, 11–13 June 2009; pp. 591–596.

29. El Hadri, A.; Benallegue, A. Attitude estimation with gyros-bias compensation using low-cost sensors. In Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with the 28th Chinese Control Conference, Shanghai, China, 15–18 December 2009; pp. 8077–8082.

30. Fong, W.; Ong, S.; Nee, A. Methods for in-field user calibration of an inertial measurement unit without external equipment. *Meas. Sci. Technol.* **2008**, *19*, 085202. [CrossRef]

31. Qureshi, U.; Golnaraghi, F. An algorithm for the in-field calibration of a MEMS IMU. *Sens. J.* **2017**, *17*, 7479–7486. [CrossRef]

32. Olsson, F.; Kok, M.; Halvorsen, K.; Schön, T.B. Accelerometer calibration using sensor fusion with a gyroscope. In Proceedings of the Statistical Signal Processing Workshop (SSP), Palma de Mallorca, Spain, 26–29 June 2016; pp. 1–5.

33. Frosio, I.; Pedersini, F.; Borghese, N.A. Autocalibration of MEMS accelerometers. *Trans. Instrum. Meas.* **2008**, *58*, 2034–2041. [CrossRef]

34. Wright, S.; Nocedal, J. *Numerical Optimization*, 2nd ed.; Springer Series in Operations Research; Springer: New York, NY, USA, 2006.

35. Boyd, S.; Vandenberghe, L. *Convex optimization*; Cambridge University Press: New York, NY, USA, 2004.

36. Skog, I.; Handel, P.; Nilsson, J.O.; Rantakokko, J. Zero-velocity detection—An algorithm evaluation. *Trans. Biomed. Eng.* **2010**, *57*, 2657–2666. [CrossRef] [PubMed]

37. Gustafsson, M.M.F.; Ljung, L.; Milnert, M. *Signal Processing*; Studentlitteratur: Lund, Sweden, 2010.

38. Hendeby, G.; Gustafsson, F. On nonlinear transformations of stochastic variables and its application to nonlinear filtering. In Proceedings of the 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008; pp. 3617–3620.

39. Lehmann, D.; Laidig, D.; Deimel, R.; Seel, T. Magnetometer-Free Inertial Motion Tracking of Arbitrary Joints with Range of Motion Constraints. Available online: https://arxiv.org/abs/2002.00639 (accessed on 22 June 2020).

40. Skog, I.; Nilsson, J.O.; Händel, P. Evaluation of zero-velocity detectors for foot-mounted inertial navigation systems. In Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN), Zurich, Switzerland, 15–17 September 2010; pp. 1–6.