

Multi-kernel Correntropy-based Orientation Estimation of IMUs: Gradient Descent Methods

Shilei Li, Lijing Li, Dawei Shi, Yunjiang Lou, Ling Shi

Abstract—This paper presents two computationally efficient algorithms for the orientation estimation of inertial measurement units (IMUs): the multi-kernel correntropy-based gradient descent (CGD) and the multi-kernel correntropy-based decoupled orientation estimation (CDOE). Traditional methods, such as gradient descent (GD) and decoupled orientation estimation (DOE), rely on the least square (LS) criterion in algorithm derivation, making them vulnerable to external acceleration and magnetic interference. To address this issue, we first demonstrate that the multi-kernel correntropy loss (MKCL) is an optimal objective function under the maximum likelihood estimation (MLE) framework when the noise follows a specific type of heavy-tailed distribution. Then, we provide some important properties of the MKCL as a cost function. By replacing the LS cost with the MKCL, we develop the CGD and CDOE algorithms. We evaluate the effectiveness of our proposed methods by comparing them with existing algorithms in various situations. Experimental results indicate that our proposed methods (CGD and CDOE) outperform their conventional counterparts (GD and DOE), especially when faced with external acceleration and magnetic disturbances. Furthermore, the new algorithms demonstrate significantly lower computational complexity than Kalman filter-based approaches, making them suitable for applications with low-cost microprocessors.

Index Terms—multi-kernel correntropy, gradient descent, orientation estimation, IMUs

I. INTRODUCTION

The use of micro-electro-mechanical-based IMUs has become widespread in various fields, including robotics [1], navigation [2], and localization [3]. Additionally, they can be found in many consumer electronics, such as smartphones, tablet computers, and smartwatches. IMUs can provide information about the carrier's orientation and can give additional localization information when combined with other sensors (e.g., ultra-wideband, camera, ultrasound, and GPS [4], [5]). Compared with optical motion capture systems that suffer from non-line-of-sight conditions, limited capture range, expensiveness, and post-processing, IMUs have the advantages of low cost, small size, portability, and real-time processing. Recently, IMUs have been utilized in health monitoring [6], joint angle estimation [7], and robot control [8]. These applications require IMUs to maintain specific precision in a disturbed

environment and to execute at a faster sampling frequency. Therefore, the development of computationally efficient and robust algorithms is necessary.

IMUs are composed of gyroscopes, accelerometers, and magnetometers. The gyroscope measures the angular rate and the orientation can be obtained by integrating gyroscope readings. However, long-term integration may cause an unbounded orientation error due to gyroscope drift and numerical integration errors. To this end, noisy but drift-free accelerometers and magnetometers are utilized to correct the orientation estimation. The orientation estimation algorithms of IMUs can be broadly categorized into three groups: the complementary filter, the Kalman filter (KF), and the gradient descent (GD). The complementary filter was initially proposed by Mahony et al. [9], [10] for aerospace applications, which was designed on the special orthogonal group (SO_3) and offered robustness to noise. The KF-based methods, which are widely used in commercial IMUs, have higher accuracy than complementary filters but require higher computational complexity and more challenging parameter-tuning procedures. Typical algorithms of this type include the extended KF [11], [12], the indirect KF [13], and the error state KF (ESKF) [14]–[16]. The GDs, proposed by Madgwick et al. [17], and its enhanced version by Seel et al. [18], have fewer parameters to tune and are reported of owing similar performance with the Kalman filter-based methods. Note that some algorithms may fall outside of the aforementioned three categories, such as the sliding mode observer [19] and the Luenberger observer [20]. In this work, we use the GDs [17], [18] and ESKF [15], [16] as the benchmark methods due to their popularity.

The challenges associated with IMU algorithms are their heavy computation cost and sensitivity to external disturbances. The high computation complexity would induce a low execution frequency, which may hinder the real-time control of robots [6]. Another challenge is the accuracy degeneration due to external disturbances, i.e., external acceleration and magnetic interference. Many strategies have been deployed to mitigate the negative impacts of these disturbances. One of the most straightforward approaches may be the normalization of the accelerometer and magnetometer readings, which was employed in [10], [17], [18]. Another way is to model the disturbance as a first-order Markov model and augment it as a new state to attenuate its bad effects [11], [13]–[15]. Some other strategies, e.g., covariance matrices adaption [13], [15], disturbance detection [21], and the combination of them, are also frequently used to increase the robustness of the algorithm. In some indoor applications, magnetic disturbance, caused by surrounding ferromagnetic materials (e.g., iron,

Shilei Li and Ling Shi are with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, China (e-mail: slidk@connect.ust.hk, eesling@ust.hk).

Lijing Li is with the School of Information and Control Engineering, China University of Mining and Technology, China (e-mail: lilijing_29@163.com).

Dawei Shi is with the School of Automation, Beijing Institute of Technology, China (e-mail: daweishi@bit.edu.cn).

Yunjiang Lou is with the State Key Laboratory of Robotics and System, School of Mechanical Engineering and Automation, Harbin Institute of Technology Shenzhen, Shenzhen 518055, China (e-mail: louyj@hit.edu.cn).

steel, and magnets), is a major reason for deteriorating both the heading (yaw) and the inclination (roll and pitch) accuracy. To handle this issue, Seel et al. proposed a decoupled orientation estimation (DOE) scheme that designed an analytical solution for sensor fusion of IMUs so that the magnetometer readings only affected the heading estimation [18]. A similar idea is also deployed in [22] where a decoupled quaternion solution was developed.

Although many algorithms have been developed to counteract the defects caused by disturbance, they directly or indirectly relied on the Gaussian noise assumption and utilized the LS criterion in algorithm derivation [15], [17], [18], [23], [24]. In this work, we demonstrate that great freedom can be obtained by generalizing the LS to the *multi-kernel correntropy loss* (MKCL) where the underlying MLE interpretation is to extend the Gaussian assumption to a type of heavy-tailed distribution. The correntropy is the local similarity measurement of two random variables. It has been successfully utilized in KF [25]–[27], adaptive filtering [28], and machine learning [29]. In [30], we extended the definition of correntropy from random variables to random vectors and defined the terminology *multi-kernel correntropy* (MKC) which greatly alleviates the conservatism of the traditional correntropy (note that the MKC proposed in our previous work is different with the concept in [29]). Based on this idea, in [23], [24], we proposed two MKC-based KFs for orientation estimation of IMUs, which achieved satisfactory performance both with and without disturbance [23], [24].

Although the performances of [23], [24] are satisfactory compared with [15], [17], [18], they have some drawbacks. Firstly, they require accurate covariance matrices of sensors, which increases the difficulty of implementation. Secondly, these algorithms possess heavy computational complexity and are not suitable for low-cost IMUs with limited computational capability. Thirdly, the relationship between the objective function and the underlying noise distribution is not very clear. In this paper, we derive two computationally efficient algorithms, i.e., the CGD and CDOE, which are built upon the GD [17] and DOE [18] and utilize the MKCL as objective functions. Specifically, we first demonstrate that accelerometer and magnetometer noises generally are heavy-tailed the Gaussian assumption is not suitable. Then, we provide some important properties of the MKCL and reveal that the MKCL is linked to a type of heavy-tailed distribution. Finally, we derive two novel algorithms by substituting the LS criterion with the MKCL. It is worth mentioning that the aim of this work is not to develop the “best” orientation estimation algorithms but to illustrate the performance of many existing algorithms can be further improved by matching the objective function with the underlying noise distribution. The contributions of this paper are summarized as follows.

- 1) Based on the maximum likelihood estimation (MLE) framework, we demonstrate that the MKCL is an optimal objective function when the noise follows a type of heavy-tailed distribution. Moreover, some important properties of the MKCL are given.
- 2) We replace the LS with the MKCL and develop two new algorithms, namely the CGD and CDOE. The newly

derived algorithms are robust to external acceleration and magnetic disturbance. Moreover, the CDOE retains the advantage that the magnetometer readings do not affect the inclination estimation.

- 3) We conduct intensive and comprehensive experiments to verify the performance of the proposed algorithms. The results demonstrate that the two proposed methods outperform their traditional counterparts (i.e., GD and DOE in [17], [18]), especially when dealing with external disturbances. Additionally, the two proposed algorithms have comparable accuracy with our previous correntropy-based KF approach [24], but with significantly lower computational cost.

The remainder of this paper is organized as follows. In Section II, we give sensor models of IMUs, provide some properties of the MKCL, and present a general algorithm framework for IMUs. In Section III, we derive two correntropy-based algorithms (i.e., the CGD and CDOE). In Section IV, we validate the performance of the proposed algorithms. In Section V, we draw a conclusion.

II. PRELIMINARIES AND PROBLEM FORMULATION

We first provide the sensor models of IMUs and give the properties of the MKCL. Then, we introduce a general framework for sensor fusion of IMUs and reveal that existing algorithms largely rely on the LS criterion. Finally, we demonstrate that the robustness of many algorithms can be enhanced by substituting the LS with the MKCL.

A. Multi-kernel Correntropy

The correntropy is a local similarity measure of two random variables $X, Y \in \mathbb{R}$ with

$$V(X, Y) = E[\kappa(X, Y)] = \int \kappa(x, y) dF_{XY}(x, y) \quad (1)$$

where $E[\cdot]$ is the expectation operator, $\kappa(x, y)$ is a shift-invariant Mercer kernel, $F_{XY}(x, y)$ is the joint distribution, and x and y are realizations of X and Y . The kernel utilized in this paper is the squared exponential function, i.e., $\kappa(x, y) = G_\sigma(x, y) = \exp(-\frac{e^2}{2\sigma^2})$ where σ is the kernel bandwidth and $e = x - y$ is the realization error. In [24], we extended the correntropy from random variables to random vectors and proposed the *multi-kernel correntropy* (MKC) for random vectors $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^l$:

$$V(\mathcal{X}, \mathcal{Y}) = \sum_{i=1}^l E[\sigma_i^2 \kappa_i(\mathcal{X}_i, \mathcal{Y}_i)] \quad (2)$$

with

$$E[\sigma_i^2 \kappa_i(\mathcal{X}_i, \mathcal{Y}_i)] = \int \sigma_i^2 \kappa_i(x_i, y_i) dF_{\mathcal{X}_i \mathcal{Y}_i}(x_i, y_i)$$

$$\kappa_i(x_i, y_i) = G_{\sigma_i}(x_i, y_i) = \exp(-\frac{e_i^2}{2\sigma_i^2})$$

where σ_i is the bandwidth for random pair \mathcal{X}_i and \mathcal{Y}_i , $e_i = x_i - y_i$ is the realization error for channel i , and $F_{\mathcal{X}_i \mathcal{Y}_i}(x_i, y_i)$ is the joint distribution. In many applications, $F_{\mathcal{X}_i \mathcal{Y}_i}(x_i, y_i)$ is not available whereas only samples $x_{i,k}$ and $y_{i,k}$ can be

obtained. In this case, the simple mean estimator is utilized and MKC can be estimated as

$$\hat{V}(\mathcal{X}, \mathcal{Y}) = \sum_{i=1}^l \sigma_i^2 \hat{V}_i(\mathcal{X}_i, \mathcal{Y}_i) \quad (3)$$

with $\hat{V}_i(\mathcal{X}_i, \mathcal{Y}_i) = \frac{1}{N} \sum_{k=1}^N G_{\sigma_i}(x_{i,k}, y_{i,k})$. Then, the MKCL can be written as

$$J_{CL}(e_k) = \sum_{i=1}^l \sigma_i^2 (1 - \hat{V}_i) = \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^l \sigma_i^2 (1 - G_{\sigma_i}(e_{i,k})). \quad (4)$$

where $e_{i,k} = x_{i,k} - y_{i,k}$ and $e_k = [e_{1,k}, e_{2,k}, \dots, e_{l,k}]^T$. Correspondingly, the LS cost can be expressed as

$$J_{LS}(e_k) = \frac{1}{2N} \sum_{k=1}^N e_k^T e_k = \frac{1}{2N} \sum_{k=1}^N \sum_{i=1}^l e_{i,k}^2. \quad (5)$$

Theorem 1. J_{CL} and J_{LS} in (4) and (5) are identical when setting $\sigma_i \rightarrow \infty$ for $i = 1, 2, \dots, l$.

The proof of this theorem is shown in Appendix VI-A.

Remark 1. It is worth mentioning that the MKC proposed in our previous work [24] is different from the conventional correntropy in two folds: it utilizes different kernel bandwidths for different random pair of variables. Moreover, specified weights are associated with the MKC so that the MKCL is compatible with the LS criterion.

B. Sensor Models and Sensor Fusion Frameworks

In this section, we first introduce the sensor models of IMUs. Then, we provide general frameworks for orientation estimation. Finally, we explain the conventional Gaussian assumption for accelerometers and magnetometers is not suitable.

Definition 1. The orientation from frame A to frame B can be represented by a unit quaternion ${}^B_A q \in \mathbb{R}^4$ with

$${}^B_A q = \begin{bmatrix} \cos(\alpha/2) \\ \sin(\alpha/2)x_{rot} \end{bmatrix}, \|{}^B_A q\|_2 = 1$$

where $\alpha \in \mathbb{R}$ is the rotation angle and $x_{rot} \in \mathbb{R}^3$ is the rotation axis with $\|x_{rot}\|_2 = 1$. In the following section, we restrain the quaternion to be a unit quaternion when representing the orientation without further mentioning. Moreover, the multiplication of two quaternions p and q is denoted as $p \otimes q$.

The gyroscope model is given by

$$y_{G,k} = w_k + b_k + v_{G,k} \quad (6)$$

where $y_{G,k} \in \mathbb{R}^3$ is the sensor reading at time step k , w_k is the angular velocity, b_k is a slow-varying bias, and $v_{G,k}$ is the noise. The accelerometer model has

$$\begin{aligned} y_{A,k} &= -{}^S g_k + {}^S a_k + v_{A,k} \\ {}^S g_k &= R({}^S_E q_k) {}^E g \end{aligned} \quad (7)$$

where $y_{A,k} \in \mathbb{R}^3$ is the sensor reading, ${}^S g_k$ and ${}^E g = [0, 0, 9.81]^T$ is the gravity vector on the sensor frame and earth frame, ${}^S a_k$ is the free acceleration on the sensor frame, ${}^S_E q_k$ is

the orientation (in a quaternion form) from the earth frame to the sensor frame, the operator $R({}^S_E q_k)$ converts the quaternion to a rotation matrix, and $v_{A,k}$ is the noise. The magnetometer model has

$$\begin{aligned} y_{M,k} &= {}^S m_k + {}^S d_k + v_{M,k} \\ {}^S m_k &= R({}^S_E q_k) {}^E m \end{aligned} \quad (8)$$

where $y_{M,k}$ is the sensor reading, ${}^S m_k$ and ${}^E m$ is the earth magnetic vector on the sensor frame and earth frame (note that ${}^E m$ is a constant and hence the subscript k is ignored), ${}^S d_k$ is the magnetic disturbance on the sensor frame, and $v_{M,k}$ is the noise.

As shown in [17], the quaternion ${}^S_E q_k$ at the current step can be obtained by the quaternion at the precious step ${}^S_E q_{k-1}$ and the current angular velocity using

$${}^S_E q_k = {}^S_E q_{k-1} + \frac{1}{2} ({}^S_E q_{k-1} \otimes w_k^q) \Delta t \quad (9)$$

where Δt is the sampling time, $w_k^q = [0, w_k^T]^T$ is the quaternion composed of true angular velocity w_k . In a practical application, we use $y_{G,k}$ to approximate w_k , which gives the following equation

$${}^S_E q_k = {}^S_E q_{k-1} + \frac{1}{2} ({}^S_E q_{k-1} \otimes y_{G,k}^q) \Delta t + v_{G,k} \quad (10)$$

where $y_{G,k}^q = [0, y_{G,k}^T]^T$ and $v_{G,k}$ is the associated noise caused by the approximation of w_k . Denoting $x_k = {}^S_E q_k$, (10) can be generalized as

$$x_k = f(x_{k-1}, y_{G,k}) + v_{G,k} \quad (11)$$

where $f(x_{k-1}, y_{G,k}) = x_{k-1} + ({}^S_E q_{k-1} \otimes y_{G,k}^q) \Delta t$. Similarly, (7) and (8) can be reformulated as

$$y_k = h(x_k) + v_{AM,k}. \quad (12)$$

with

$$y_k = \begin{bmatrix} y_{A,k} \\ y_{M,k} \end{bmatrix}, h(x_k) = \begin{bmatrix} h_A(x_k) \\ h_M(x_k) \end{bmatrix}, v_{AM,k} = \begin{bmatrix} v_{A,k} \\ v_{M,k} \end{bmatrix}$$

where $h_A(x_k) = -R(x_k) {}^E g$, $h_M(x_k) = R(x_k) {}^E m$, $v_{A,k} = {}^S a_k + v_{A,k}$, and $v_{M,k} = {}^S d_k + v_{M,k}$. Equations (11) and (12) provide a general nonlinear state space model for IMUs. We then discuss the existing orientation estimation algorithms under the framework of *maximum a posteriori* (MAP) and *maximum likelihood estimation* (MLE).

Denote measurement set as $\mathbb{M}_{1:N} = \{y_{A,k}, y_{M,k}\}_{k=1}^N$, gyroscope reading set as $\mathbb{G}_{1:N} = \{y_{G,k}\}_{k=1}^N$, and assume that $v_{G,k}$, $v_{AM,k}$ and the noise distribution for the initial guess x_0 are mutually independent. Under the MAP, the posterior distribution of $x_{1:N}$ given $\mathbb{M}_{1:N}$ [note that we regard $\mathbb{G}_{1:N}$ as known inputs of (11)] has

$$\begin{aligned} p(x_{1:N} | \mathbb{M}_{1:N}) &= \frac{p(\mathbb{M}_{1:N} | x_{1:N}) p(x_{1:N})}{p(\mathbb{M}_{1:N})} \\ &= \frac{p(x_0) \prod_{k=1}^N p(y_k | x_k) \prod_{k=1}^N p(x_k | x_{k-1})}{p(\mathbb{M}_{1:N})} \\ &\propto p(x_0) \prod_{k=1}^N p_{v_{AM,k}}(y_k - h(x_k)) \prod_{k=1}^N p_{v_{G,k}}(x_k - f(x_{k-1}, y_{G,k})) \end{aligned} \quad (13)$$

where $p_{v_{AM,k}}(\cdot)$ and $p_{v_{G,k}}(\cdot)$ denotes the corresponding densities of $v_{AM,k}$ and $v_{G,k}$, respectively. By assuming that $v_{AM,k} \sim \mathcal{N}(0, R)$ (we will disclose **this assumption is inappropriate** in the following part), $v_{G,k} \sim \mathcal{N}(0, Q)$, $x_0 \sim \mathcal{N}(\mu, \Pi)$ and ignoring the normalizing constant, the posterior distribution is given by

$$\exp\left(-\frac{1}{2}\|\Pi^{-1/2}(x_0 - \mu)\|_2^2\right) \prod_{k=1}^N \exp\left(-\frac{1}{2}\|R^{-1/2}(y_k - h(x_k))\|_2^2\right) \\ \times \prod_{k=1}^N \exp\left(-\frac{1}{2}\|Q^{-1/2}(x_k - f(x_{k-1}, y_{G,k}))\|_2^2\right). \quad (14)$$

Maximizing (14) is equivalent to minimizing its negative logarithm, which gives

$$x_{1:N} = \arg \min_{x_{1:N}} \frac{1}{2}\|\Pi^{-1/2}(x_0 - \mu)\|_2^2 \\ + \frac{1}{2} \sum_{k=1}^N \|R^{-1/2}(y_k - h(x_k))\|_2^2 \\ + \frac{1}{2} \sum_{k=1}^N \|Q^{-1/2}(x_k - f(x_{k-1}, y_{G,k}))\|_2^2. \quad (15)$$

This optimization problem can be solved by extended Rauch–Tung–Striebel smoother [31] or Gauss-Newton optimization [32].

Under the MLE, by assuming that gyroscope readings $y_{G,k}$ is independent with measurements y_k , the likelihood of observing the data $\{\mathbb{G}_{1:N}, \mathbb{M}_{1:N}\}$ given $x_{1:N}$ has

$$x_{1:N} = \arg \max p(\{\mathbb{G}_{1:N}, \mathbb{M}_{1:N}\} | x_{1:N}) \\ = p(\mathbb{G}_{1:N} | x_{1:N}) p(\mathbb{M}_{1:N} | x_{1:N}) \\ = \prod_{k=1}^N p_{v_{G,k}}(x_k - f(x_{k-1}, y_{G,k})) \prod_{k=1}^N p_{v_{AM,k}}(y_k - h(x_k)). \quad (16)$$

Taking negative logarithm on the right side of (16) gives

$$x_{1:N} = \arg \min_{x_{1:N}} \frac{1}{2} \sum_{k=1}^N \|R^{-1/2}(y_k - h(x_k))\|_2^2 \\ + \frac{1}{2} \sum_{k=1}^N \|Q^{-1/2}(x_k - f(x_{k-1}, y_{G,k}))\|_2^2 \quad (17)$$

which can be seen as a nonlinear regression problem with parameter vector $x_{1:N}$. Solutions for this problem include the Levenberg-Marquardt algorithm [33], Gauss-Newton optimization [32], etc.

In many applications, we are much more interested in the current state x_k based on all previous measurements. Then, under the MAP, the filtering problem becomes

$$x_k = \arg \min_{x_k} \frac{1}{2}\|R^{-1/2}(y_k - h(x_k))\|_2^2 \\ + \frac{1}{2}\|(P_k^-)^{-1/2}(x_k - f(x_{k-1}, y_{G,k}))\|_2^2 \quad (18)$$

where P_k^- is the *a priori* estimate of error covariance. Using (18) as an objective function, one can derive the extended Kalman filter [4], [34]. Under the MLE, the filtering problem

becomes

$$x_k = \arg \min_{x_k} \frac{1}{2}\|R^{-1/2}(y_k - h(x_k))\|_2^2 \\ + \frac{1}{2}\|Q^{-1/2}(x_k - f(x_{k-1}, y_{G,k}))\|_2^2. \quad (19)$$

One solution of (19) is the gradient descent method. Taking partial derivative on the right side of (19) with respect to x_k and setting it to zero gives

$$-\left[\frac{\partial h}{\partial x}\right]_{x_k} R^{-1}(y_k - h(x_k)) + Q^{-1}(x_k - f(x_{k-1}, y_{G,k})) = 0$$

Denoting the *a priori* estimate of the state as $x_k^- = f(x_{k-1}, y_{G,k})$ and approximating $h(x_k)$ as $h(x_k^-)$, it follows that

$$x_k = x_k^- + Q \left[\frac{\partial h}{\partial x}\right]_{x_k^-} R^{-1}(y_k - h(x_k^-)). \quad (20)$$

Assuming that Q and R are diagonal with the same diagonal entities with

$$Q = \text{diag}([q, \dots, q]), R = \text{diag}([r, \dots, r]), \quad (21)$$

it follows that

$$x_k = x_k^- + \lambda \left[\frac{\partial h}{\partial x}\right]_{x_k^-} (y_k - h(x_k^-)) \quad (22)$$

with $\lambda = \frac{q}{r}$ determining how confident we are in believing the measurements compared with the prediction. It is worth mentioning some authors utilize $h(x_{k-1})$ to approximate $h(x_k)$ [17], which gives

$$x_k = x_k^- + \lambda \left[\frac{\partial h}{\partial x}\right]_{x_{k-1}^-} (y_k - h(x_{k-1})). \quad (23)$$

This small difference usually does not have a big impact on the algorithm performance of IMUs since both x_{k-1} and x_k^- provide a good initial guess of x_k .

An alternative is to use the inverse of the nonlinear measurement model rather than the original model in problem formulation, i.e., constructing $x_k = h^{-1}(y_k) + \check{v}_{AM,k}$ and assuming $\check{v}_{AM,k} \sim \mathcal{N}(0, \check{R})$. By this conversion, (19) can be rewritten as

$$x_k = \arg \min_{x_k} \frac{1}{2}\|\check{R}^{-1/2}(x_k - h^{-1}(y_k))\|_2^2 \\ + \frac{1}{2}\|Q^{-1/2}(x_k - f(x_{k-1}, y_{G,k}))\|_2^2. \quad (24)$$

Taking partial derive on the right side of (24) and setting it to zero gives

$$x_k = [Q^{-1} + \check{R}^{-1}]^{-1} Q^{-1} x_k^- + [Q^{-1} + \check{R}^{-1}]^{-1} \check{R}^{-1} h^{-1}(y_k). \quad (25)$$

If Q and \check{R} are diagonal matrices with the same entities on the main diagonal, i.e., $Q = \text{diag}([q, \dots, q])$ and $\check{R} = \text{diag}([\check{r}, \dots, \check{r}])$, the above equation can be reformulated as

$$x_k = \gamma x_k^- + (1 - \gamma) x_{AM,k} \quad (26)$$

where $\gamma = \frac{\check{r}}{q + \check{r}}$ and $x_{AM,k}$ is the state determined by measurement y_k with $x_{AM,k} = h^{-1}(y_k)$. Equation (26) gives a general form of complementary filter for orientation estimation [35]. A remaining question is the obtainment of $x_{AM,k}$. One method

is to solve the following LS-based objective function:

$$x_{AM,k} = \arg \min \frac{1}{2} \|y_k - h(x_{AM,k})\|_2^2. \quad (27)$$

Applying one-step gradient descent update with the initial guess of $x_{AM,k}$ as x_k^- gives

$$x_{AM,k} = x_k^- + \mu \left[\frac{\partial h}{\partial x} \Big|_{x_k^-} \right] (y_k - h(x_k^-)). \quad (28)$$

where μ is the learning rate. Substituting (28) into (26), one obtains

$$\begin{aligned} x_k &= \gamma x_k^- + (1 - \gamma) \left(x_k^- + \mu \left[\frac{\partial h}{\partial x} \Big|_{x_k^-} \right] (y_k - h(x_k^-)) \right) \\ &= x_k^- + \mu(1 - \gamma) \left[\frac{\partial h}{\partial x} \Big|_{x_k^-} \right] (y_k - h(x_k^-)). \end{aligned} \quad (29)$$

One can observe that (29) is identical with (22) by assigning $\mu(1 - \gamma) = \lambda$. This indicates that the optimization problems (19) and (24) are interchangeable under certain situations and the key of (24) is solving (27).

From (18), (19), and (24), one can see that conventional orientation estimation algorithms rely heavily on the Gaussian assumptions [15], [17], [18], [35]. This assumption may be valid for gyroscope noise $v_{G,k}$, but not for accelerometer noise $v_{A,k}$ and magnetometer noise $v_{M,k}$ due to the existence of unknown acceleration $S_{a,k}$ and magnetic disturbance $S_{d,k}$ [4]. A better density representation for $v_{A,k}$ and $v_{M,k}$ should be the heavy-tailed distribution. To illustrate this, we investigate the probability density functions (pdfs) of $v_{A,k}$ and $v_{M,k}$ in two situations: without and with disturbances. We first keep the IMU being static and free of disturbances and the corresponding pdfs of $v_{A,k}$ and $v_{M,k}$ are shown in Figs. 1(a) and 1(c). Then, we manually generate external acceleration and magnetic disturbance, and the corresponding results are shown in Figs. 1(b) and 1(d). The disturbance generation method is described in the caption of Fig. 5(a) (hence is ignored here). Not surprisingly, The densities of $v_{A,k}$ and $v_{M,k}$ are Gaussian-like when without disturbances and are heavy-tailed when involving disturbances. Unfortunately, in many practical applications of IMUs, external acceleration and magnetic disturbance are unavoidable, which degenerates the conventional algorithms significantly.

C. Properties of the MKCL as a Cost Function

In this section, we discuss the properties of the MKCL as an objective function from the perspective of noise distribution, likelihood function, and influence function.

The MLE viewpoint in (16) can be easily extended to the general case of noise distributions, i.e.,

$$\begin{aligned} p(v_{AM}) &\propto \exp(-J_r(R^{-1/2}v_{AM})) \\ p(v_G) &\propto \exp(-J_q(Q^{-1/2}v_G)) \end{aligned}$$

where $J_r(\cdot)$ and $J_q(\cdot)$ are nonlinear functions with respect to its argument. Then, based on (16), we can replace the mean squared errors $\|R^{-1/2}v_{AM}\|_2^2$ and $\|Q^{-1/2}v_G\|_2^2$ by nonlinear functions $J_r(R^{-1/2}v_{AM})$ and $J_q(Q^{-1/2}v_G)$ and construct the

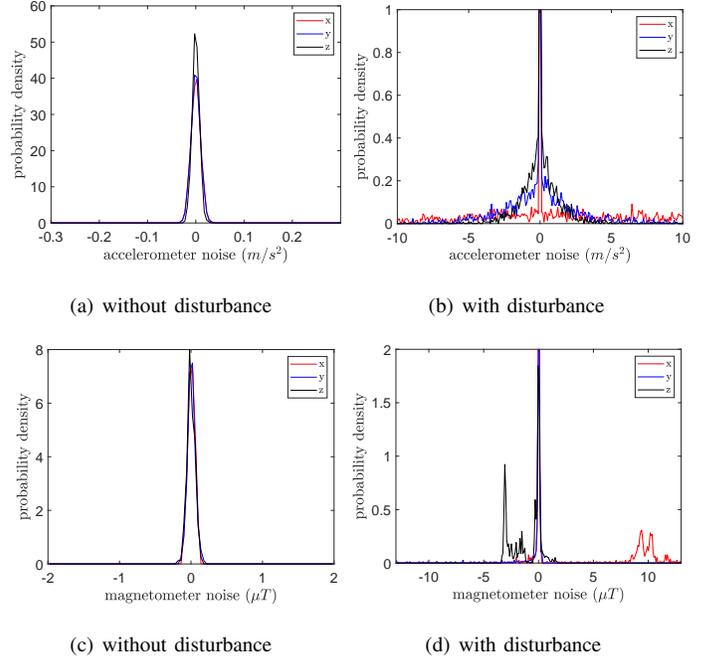


Fig. 1. The pdfs of $v_{A,k}$ and $v_{M,k}$ without and with disturbances. (a) and (c) show the pdfs of $v_{A,k}$ and $v_{M,k}$ when IMU is static and free of disturbances. (b) and (d) show the pdfs of $v_{A,k}$ and $v_{M,k}$ when IMU is with disturbances. The IMU used in this experiment is Xsens MTI-670 and the disturbed data is generated by the translation experiment as shown in Fig. 5(a).

following problem:

$$\begin{aligned} x_{1:N} = \arg \min \sum_{k=1}^N J_r \left(R^{-1/2} (y_k - h(x_k)) \right) \\ + \sum_{k=1}^N J_q \left(Q^{-1/2} (x_k - f(x_{k-1}, y_{G,k})) \right). \end{aligned} \quad (30)$$

The corresponding filtering problem becomes:

$$\begin{aligned} x_k = \arg \min J_r \left(R^{-1/2} (y_k - h(x_k)) \right) \\ + J_q \left(Q^{-1/2} (x_k - f(x_{k-1}, y_{G,k})) \right). \end{aligned} \quad (31)$$

A good candidate for the nonlinear $J_r(\cdot)$ and $J_q(\cdot)$ is the MKCL. To investigate the properties of the MKCL as a cost function, we consider the following nonlinear regression problem:

$$y_k = g(x, u_k) + e_k \quad (32)$$

where $y_k \in \mathbb{R}^l$ is the output, u_k is the input, x is the parameter vector to be estimated, $g(\cdot)$ is the nonlinear function, e_k is the noise, and $k = 1, 2, \dots, N$ is the sample index. We assume that the nominal distribution of e_k has $e_k \sim \mathcal{N}(0, R)$. Then, we compare the following two objective functions:

$$\arg \min_x J_{LS}(\tilde{e}_k) = \frac{1}{2N} \sum_{k=1}^N \sum_{i=1}^l \tilde{e}_{i,k}^2 \quad (33a)$$

$$\arg \min_x J_{CL}(\tilde{e}_k) = \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^l \sigma_i^2 \left(1 - G_{\sigma_i}(\tilde{e}_{i,k}) \right) \quad (33b)$$

where $\tilde{e}_k \triangleq R^{-1/2}(y_k - g(x, u_k))$ and $\tilde{e}_k = [\tilde{e}_{1,k}, \tilde{e}_{2,k}, \dots, \tilde{e}_{l,k}]^T$. Then, we have the following theorem.

Theorem 2. J_{LS} in (33a) is an optimal loss function when $\tilde{e}_k \sim \mathcal{N}(0, I)$ based on MLE. On the contrary, J_{CL} in (33b) is an optimal cost function if $\tilde{e}_{i,k}$ follows

$$p(\tilde{e}_{i,k}) = c_i \exp\left(-\sigma_i^2(1 - \exp(-\frac{\tilde{e}_{i,k}^2}{2\sigma_i^2}))\right) \quad (34)$$

where c_i is a normalization coefficient and $i = 1, 2, \dots, l$.

The proof of this theorem is shown in Appendix VI-B. A comparison of Gaussian distribution $\mathcal{N}(0, 1)$ and $p(\tilde{e}_{i,k})$ in (34) with different kernel bandwidth is shown in Fig. 2. As depicted, when σ_i is large, $p(\tilde{e}_{i,k})$ approaches a Gaussian distribution, which aligns with Theorem 1. When σ_i is relatively small, it represents a type of heavy-tailed distribution. This indicates that the MKCL is a good candidate when the noise is heavy-tailed.

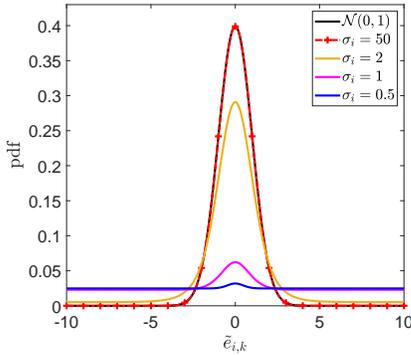


Fig. 2. The pdfs of $\mathcal{N}(0, 1)$ and $p(\tilde{e}_{i,k})$ in (34) with different kernel bandwidths. The error $\tilde{e}_{i,k}$ is assumed to be bounded within $[-20, 20]$ for $p(\tilde{e}_{i,k})$ and the coefficient c_i is obtained by $c_i = 1 / \int_{-20}^{20} \exp\left(-\sigma_i^2(1 - \exp(-\frac{\tilde{e}_{i,k}^2}{2\sigma_i^2}))\right) d\tilde{e}_{i,k}$.

To further explore the properties of the MKCL as a loss function, we use the log-likelihood function as a metric to investigate how well different objective functions explains the data. We consider the following mixture noise distribution with

$$\tilde{e}_k \sim (1 - p)\mathcal{N}(0, 1) + p\mathcal{U}(-20, 20), 0 \leq p < 0.5 \quad (35)$$

where $\mathcal{N}(0, 1)$ is the nominal Gaussian distribution, $\mathcal{U}(-20, 20)$ is a uniform distribution with boundary $[-20, 20]$, and p is a probability that determines \tilde{e}_k generated by which distribution. The log-likelihood for the MKCL can be obtained by $\log \mathcal{L}_{CL} = \frac{1}{N} \sum_{k=1}^N p(\tilde{e}_k)$ where $p(\tilde{e}_k)$ is shown in (34) and the value for the LS has $\log \mathcal{L}_{LS} = \frac{1}{N} \sum_{k=1}^N \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\tilde{e}_k^2}{2}\right)$. We compare $\log \mathcal{L}_{CL}$ and $\log \mathcal{L}_{LS}$ under different p . The results are shown in Fig.3. We observe that $\mathcal{L}_{CL} > \mathcal{L}_{LS}$ always holds when σ is bigger than a certain threshold σ^* and $p > 0$, which indicates that the MKCL is much more preferable when the noise is heavy-tailed. We also observe that the potential benefits of the MKCL over the LS [i.e., maximum values of $(\log \mathcal{L}_{CL} - \log \mathcal{L}_{LS})$] increases with the growth of p , which implies that a higher profit can be obtained by replacing the LS with the MKCL when the

noise has a heavier tail. This fact also reveals that the LS criterion is sensitive to heavy-tailed noises but the MKCL is not, especially when the kernel bandwidth is properly selected.

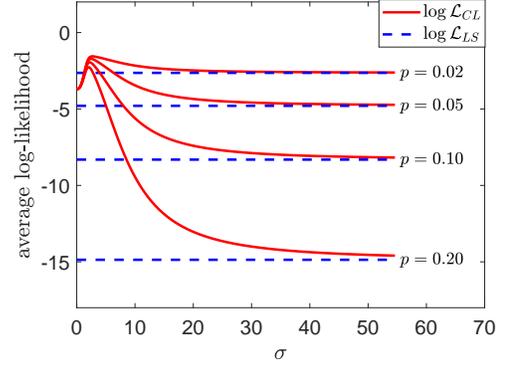


Fig. 3. A comparison of $\log \mathcal{L}_{CL}$ and $\log \mathcal{L}_{LS}$ under differ p .

The influence function measures the derivative of the loss with respect to the residual and quantifies the effect of the size of a residual on the loss [36]. In the case of $l = 1$ and $N = 1$ in (33), one has $J_{LS} = \frac{1}{2}\tilde{e}^2$ and $J_{CL} = \sigma^2(1 - G_\sigma(\tilde{e}))$. Then, we visualize the objective function and influence of J_{LS} and J_{CL} with different kernel bandwidths in Fig. 4. One can see that the LS gives each residual constant influence as shown in Fig. 4(b), which indicates that they are sensitive to outliers. On the contrary, the influence function of the MKCL is close to that of the LS when the residual is small, but goes towards zero with the growth of the residual (when using small kernel bandwidth). This redescending property makes it an attractive option when the underlying noise is heavy-tailed.

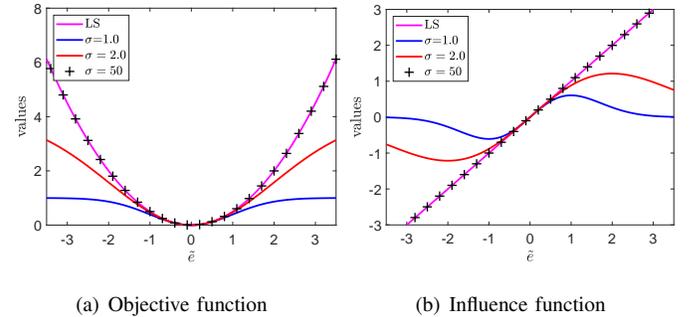


Fig. 4. Objective functions and influence functions for J_{LS} and J_{CL} with different kernel bandwidths.

D. Problem Formulation

As explained in Section II-B, the Gaussian assumption is valid for gyroscopes but not for accelerometers and magnetometers. Taking this prior knowledge into consideration, one can construct the following problem:

$$x_k = \arg \min_{x_k} J_{CL} \left(R^{-1/2}(y_k - h(x_k)) \right) + \|Q^{-1/2}(x_k - f(x_{k-1}, y_{G,k}))\|_2^2. \quad (36)$$

Denoting $\tilde{e}_k = R^{-1/2}(y_k - h(x_k)) \in \mathbb{R}^l$, it follows that $J_{CL}(\tilde{e}_k) = \sum_{i=1}^l \sigma_i^2(1 - G_{\sigma_i}(\tilde{e}_{i,k}))$. By taking partial derivative on the right side of above equation with respect to x_k , one has

$$-\left[\frac{\partial h}{\partial x}\right]_{x_k} \tilde{W} R^{-1}(y_k - h(x_k)) + Q^{-1}(x_k - f(x_{k-1}, y_{G,k})) = 0 \quad (37)$$

where $\tilde{W} = \text{diag}([G_{\sigma_1}(\tilde{e}_{1,k}), G_{\sigma_2}(\tilde{e}_{2,k}), \dots, G_{\sigma_l}(\tilde{e}_{l,k})])$. By analogy with (20), one obtains

$$x_k = x_k^- + Q \left[\frac{\partial h}{\partial x} \right]_{x_k^-} \tilde{W} R^{-1}(y_k - h(x_k^-)). \quad (38)$$

A much more straightforward method is to follow the procedure shown in (24) to (29) and construct the following problem

$$x_{AM,k} = \arg \min J_{CL}(y_k - h(x_{AM,k})). \quad (39)$$

Denoting $e_k = y_k - h(x_k^-) \in \mathbb{R}^l$ and by analogy with (28), it follows that

$$x_{AM,k} = x_k^- + \mu \left[\frac{\partial h}{\partial x} \right]_{x_k^-} W (y_k - h(x_k^-)) \quad (40)$$

where $e_k = [e_{1,k}, e_{2,k}, \dots, e_{l,k}]^T$ and $W = \text{diag}([G_{\sigma_1}(e_{1,k}), G_{\sigma_2}(e_{2,k}), \dots, G_{\sigma_l}(e_{l,k})])$. Substituting (40) into (26) gives

$$x_k = x_k^- + \mu(1 - \gamma) \left[\frac{\partial h}{\partial x} \right]_{x_k^-} W (y_k - h(x_k^-)). \quad (41)$$

One can observe that the only difference between (41) and (29) is the attenuating matrix W . The utility of this matrix can be understood intuitively: when no external disturbances are involved in the measurements, one can expect that the discrepancy $y_k - h(x_k^-)$ is small and hence $W \approx I$. In this case, equation (41) is similar to (29). On the contrary, with the presence of disturbances, the discrepancy $y_k - h(x_k^-)$ is expected to be big, and hence W is close to a zero matrix. By this technique, the MKCL-based methods are much more robust than their LS-based counterparts where the ‘‘robustness level’’ is controlled by the kernel bandwidth.

Remark 2. It is worth mentioning that the purpose of Sections II-B and II-D is to provide a general framework for orientation estimation algorithms. In specific algorithms, different orientation parametrizing strategies, e.g., quaternions, rotation matrices, rotation vectors, and Euler angles, are employed by different authors in different applications. Moreover, some other variables, e.g., the gyroscope bias, the external disturbance, and the magnetic disturbance, may be augmented into the state to enhance the algorithm’s performance. The purpose of this work is not to obtain the best orientation estimation algorithm but to demonstrate that an improvement can be further obtained by replacing the LS cost with the much more generic MKCL. The newly derived algorithms would inherit the advantages of the original method but with enhanced robustness to disturbances.

III. MAIN RESULTS

In this section, we first demonstrate that two popular algorithms, i.e., the GD and DOE, are derived under the LS-based

objective functions. Then we use MKCL to substitute LS and derive two novel algorithms, i.e., the CGD and CDOE. Before proceeding, we denote the normalized accelerometer readings as $y_{A,k} = \frac{y_{A,k}}{\|y_{A,k}\|_2} = [y_{ax}, y_{ay}, y_{az}]^T$ and the normalized magnetometer readings as $y_{M,k} = \frac{y_{M,k}}{\|y_{M,k}\|_2} = [y_{mx}, y_{my}, y_{mz}]^T$.

A. Two Traditional Solutions for IMUs

In this section, we provide a brief overview of the derivation of the GD [17] and DOE [18], emphasizing that they are derived using LS-based objective functions (note that this fact is not pointed out by their original authors).

1) *The GD:* The GD algorithm in [17] is composed of two steps: the *a priori* estimate of orientation by gyroscope readings and the correction by accelerometer and magnetometer readings. Based on (10), the *a priori* estimate of quaternion ${}^S E q_k^-$ has

$${}^S E q_k^- = {}^S E q_{k-1} + \frac{1}{2} \left({}^S E q_{k-1} \otimes y_{G,k}^q \right) \Delta t \quad (42)$$

where ${}^S E q_{k-1}$ is the quaternion at previous time step, and $y_{G,k}^q = [0, y_{G,k}^T]^T$. For accelerometers, the constant gravity vector quaternion is set to be ${}^E g = [0, 0, 0, 1]^T$ and the measured gravity quaternion is ${}^S g = [0, y_{A,k}^T]^T$. The following LS-based objective function is utilized:

$$\min_{{}^S E q_k \in \mathbb{R}^4} f_g({}^E g, {}^S g, {}^S E q_k) = \frac{1}{2} \|E_{g,k}\|^2 \quad (43)$$

where $E_{g,k} = {}^S E q_k^* \otimes {}^E g \otimes {}^S E q_k - {}^S g$ is the error quaternion, ${}^S E q_k = [q_1, q_2, q_3, q_4]^T$ is the quaternion to be determined, and ${}^S E q_k^*$ is the conjugate quaternion of ${}^S E q_k$. For magnetometers, the constant magnetic vector quaternion is set to be ${}^E m = [0, m_x, 0, m_z]^T$ (due to magnetic dip) with $\sqrt{m_x^2 + m_z^2} = 1$ and the measured magnetic vector quaternion is ${}^S m = [0, y_{M,k}^T]^T$. The following LS-based cost function is employed:

$$\min_{{}^S E q_k \in \mathbb{R}^4} f_m({}^E m, {}^S m, {}^S E q_k) = \frac{1}{2} \|E_{m,k}\|^2 \quad (44)$$

where $E_{m,k} = {}^S E q_k^* \otimes {}^E m \otimes {}^S E q_k - {}^S m$. Combining (43) and (44), one has

$$\min_{{}^S E q_k \in \mathbb{R}^4} f_{g,m}({}^E g, {}^E m, {}^S g, {}^S m, {}^S E q_k) = \frac{1}{2} \|E_k\|^2 \quad (45)$$

where $E_k = \begin{bmatrix} E_{g,k} \\ E_{m,k} \end{bmatrix} \in \mathbb{R}^8$ is the aggregated error quaternion. By applying one step gradient descent update with the initial guess ${}^S E q_k^-$, the recursive solution of (45) can be obtained as

$${}^S E q_{AM,k} = {}^S E q_k^- - \mu \nabla f_{g,m} \quad (46)$$

with

$$\nabla f_{g,m} = \frac{\partial E_k}{\partial {}^S E q_k} \frac{\partial f_{g,m}}{\partial E_k} = J_{E_k}({}^S E q_k) E_k \quad (47)$$

where $J_{E_k}({}^S E q_k) = \frac{\partial E_k}{\partial {}^S E q_k} = \left[\frac{\partial E_k}{\partial {}^S E q_k} \Big|_{{}^S E q_k^-} \right]$, ${}^S E q_{AM,k}$ is the quaternion obtained by accelerometer and magnetometer readings and μ is the learning rate. Following the method used in

[17], we modify (46) as

$${}^S_E q_{AM,k} = {}^S_E q_k^- - \mu \frac{\nabla f_{g,m}}{\|\nabla f_{g,m}\|_2}. \quad (48)$$

Remark 3. The benefit of using normalized gradient direction is to increase the robustness with respect to unexpected large E_k , at the price of decreased smoothness when E_k is small (this is the reason why the GD seems much more noisy than the KF-based method when the IMU has slow-varying dynamics).

Combining (42) and (48) in a complementary form gives

$${}^S_E q_k = \gamma {}^S_E q_k^- + (1 - \gamma) {}^S_E q_{AM,k} \quad (49)$$

where γ is a coefficient to balance between the gyroscope estimation and the accelerometer and magnetometer estimation. Substituting (48) into (49), one has

$${}^S_E q_k = {}^S_E q_k^- - \mu(1 - \gamma) \frac{\nabla f_{g,m}}{\|\nabla f_{g,m}\|_2}. \quad (50)$$

By setting $\lambda = (1 - \gamma)\mu$, one obtains

$${}^S_E q_k = {}^S_E q_k^- - \lambda \frac{\nabla f_{g,m}}{\|\nabla f_{g,m}\|_2} \quad (51)$$

where β is the only parameter to be tuned in the GD. The whole algorithm is then summarized as follows:

- Prediction: estimate the orientation using (42).
- Update: correct the orientation using (51).

Remark 4. The detailed derivation procedure of the GD described in the above paragraphs is slightly different from the original work [17], but in fact, they are mathematically equivalent.

2) *The DOE:* The DOE algorithm in [18] shares the same strategy in terms of the *a priori* estimate of orientation using gyroscope readings as shown in (42). The difference is that it employs an analytical solution for the orientation correction. Define

$${}^E g = \begin{bmatrix} 0 \\ E_{r_a} \end{bmatrix}, {}^S g = \begin{bmatrix} 0 \\ S_{a_k} \end{bmatrix}, {}^E m = \begin{bmatrix} 0 \\ E_{r_m} \end{bmatrix}, {}^S m = \begin{bmatrix} 0 \\ S_{m_k} \end{bmatrix}$$

where $E_{r_a} = [0, 0, 1]^T$ and $E_{r_m} = [1, 0, 0]^T$ are the reference z -axis and x -axis in the earth frame, and $S_{a_k} = y_{A,k}$ and $S_{m_k} = y_{M,k}$ are the normalized accelerometer and magnetometer readings. Then, equations (43) and (44) can be equivalently written as

$$\min_{{}^S_E q_k \in \mathbb{R}^4} f_g({}^E g, {}^S g, {}^S_E q_k) = \frac{1}{2} \|E_{g,k}\|^2 \quad (52a)$$

$$\min_{{}^S_E q_k \in \mathbb{R}^4} f_m({}^E m, {}^S m, {}^S_E q_k) = \frac{1}{2} \|E_{m,k}\|^2 \quad (52b)$$

with

$$E_{g,k} = \left\| {}^S_E q_k^* \otimes \begin{bmatrix} 0 \\ E_{r_a} \end{bmatrix} \otimes {}^S_E q_k - \begin{bmatrix} 0 \\ S_{a_k} \end{bmatrix} \right\|^2 \quad (53a)$$

$$E_{m,k} = \left\| {}^S_E q_k^* \otimes \begin{bmatrix} 0 \\ E_{r_m} \end{bmatrix} \otimes {}^S_E q_k - \begin{bmatrix} 0 \\ S_{m_k} \end{bmatrix} \right\|^2. \quad (53b)$$

Note that here $[0, E_{r_m}]^T$ in (52b) is slightly different with E_m in (44) since the magnetic dip is ignored in [18]. The above optimization system is overdetermined since each of the two equations provides two independent constraints on ${}^S_E q_k$, whereas ${}^S_E q_k$ only has three degrees of freedom. Conventionally, we can solve (52) by minimizing both $E_{g,k}$ and $E_{m,k}$ simultaneously. However, it could result in disturbances in the magnetic field affecting the roll and pitch. To solve this problem, an analytical solution of (52) was developed in [18] by constraining the magnetometer readings only affecting the heading. Specifically, for sub-problem (52a), the included angle between S_{r_a} and S_{a_k} has $\alpha_{err,a} = \langle (S_{r_a}, S_{a_k}) \rangle = \arccos\left(\frac{\langle S_{r_a}, S_{a_k} \rangle}{\|S_{r_a}\|_2 \|S_{a_k}\|_2}\right)$ where $\langle \cdot, \cdot \rangle$ is the inner product. Moreover, the rotation axis can be written as $x_a = \frac{S_{r_a} \times S_{a_k}}{\|S_{r_a} \times S_{a_k}\|_2}$. Finally, the analytical solution of (52a) has

$${}^S_E q_{acc,k} = \begin{bmatrix} \cos\left(\frac{1}{2}\alpha_{err,a}\right) \\ \sin\left(\frac{1}{2}\alpha_{err,a}\right)x_a \end{bmatrix} \quad (54)$$

where ${}^S_E q_{acc,k}$ denotes the quaternion obtained by the accelerometer readings. Since the accelerometer measurement is noisy, it is feasible to use only a small portion of $\alpha_{err,a}$ to correct the gyroscope drift, which gives

$${}^S_E q_{ga,k} = {}^S_E q_k^- \otimes q_{accor,k} \quad (55)$$

with

$$q_{accor,k} = \begin{bmatrix} \cos\left(\frac{1}{2}k_a\alpha_{err,a}\right) \\ \sin\left(\frac{1}{2}k_a\alpha_{err,a}\right)x_a \end{bmatrix} \quad (56)$$

where $k_a \in (0, 1)$ is an adjustable weight, ${}^S_E q_k^-$ is the *a priori* estimate of the quaternion as shown in (42), and ${}^S_E q_{ga,k}$ is the obtained quaternion by fusing the gyroscope and accelerometer measurements.

To fulfill the constraint that magnetometer readings only affect the heading, one can project S_{m_k} into the horizontal plane (in the sensor frame), i.e., $S_{\bar{m}_k} = S_{m_k} - (S_{m_k} \cdot S_{r_a}) \cdot S_{r_a}$. Then, the included angle between $S_{\bar{m}_k}$ and S_{r_m} has $\alpha_{err,m} = \langle (S_{\bar{m}_k}, S_{r_m}) \rangle = \arccos\left(\frac{\langle S_{\bar{m}_k}, S_{r_m} \rangle}{\|S_{\bar{m}_k}\|_2 \|S_{r_m}\|_2}\right)$ and the rotation axis has $x_m = \frac{S_{\bar{m}_k} \times S_{r_m}}{\|S_{\bar{m}_k} \times S_{r_m}\|_2}$. Finally, by analogy to (54) and (55), one has

$${}^S_E q_k = {}^S_E q_{ga,k} \otimes q_{mcor,k} \quad (57)$$

with

$$q_{mcor,k} = \begin{bmatrix} \cos\left(\frac{1}{2}k_m\alpha_{err,m}\right) \\ \sin\left(\frac{1}{2}k_m\alpha_{err,m}\right)x_m \end{bmatrix} \quad (58)$$

where $k_m \in (0, 1)$ is an adjustable weight and ${}^S_E q_k$ is fused quaternion at time step k . To further mitigate the gyroscope drift b_k , the following gyroscope bias update equation is utilized

$$b_k = b_{k-1} + k_{b,a}\alpha_{err,a}x_a + k_{b,m}\alpha_{err,m}x_m \quad (59)$$

where $k_{b,a}, k_{b,m} \in (0, 1)$ are two adjustable gains. The bias estimation b_k is then utilized to correct the gyroscope readings $y_{G,k}$ with

$$y_{G,k}^c = y_{G,k} - b_k \quad (60)$$

where $y_{G,k}$ is the raw gyroscope readings and $y_{G,k}^c$ is the corrected one. Correspondingly, $y_{G,k}^q = [0, (y_{G,k}^c)^T]^T$ is

utilized in (42) to replace $y_{G,k}^q = [0, y_{G,k}^T]^T$. The whole algorithm of the DOE is summarized as follows:

- Predict the orientation by gyroscopes using (42) and (60).
- Update the orientation by accelerometers using (55).
- Update the orientation by magnetometers using (57).

B. The CGD

To increase the robustness of the GD with respect to external disturbances, we use the following MKCL-based objective function

$$\min_{\substack{E, q \in \mathbb{R}^4}} f_{cl}(Eg, Em, Sg, Sm, Sq) = \sum_{i=1}^8 \sigma_i^2 \left(1 - G_{\sigma_i}(E_{i,k})\right) \quad (61)$$

where $E_{i,k}$ is i -th element of E_k , σ_i is i -th bandwidth for channel i . Correspondingly, the derivative of f_{cl} with respect to ${}^S E q_k$ can be calculated as

$$\nabla f_{cl} = \left[\frac{\partial E_k}{\partial {}^S E q_k} \Big|_{{}^S E q_k^-} \right] \frac{\partial f_{cl}}{\partial E_k} = J_{E_k}({}^S E q_k) W E_k \quad (62)$$

with

$$W = \text{diag}[G_{\sigma_1}(E_{1,k}), G_{\sigma_2}(E_{2,k}), \dots, G_{\sigma_8}(E_{8,k})]$$

$$E_k = \begin{bmatrix} E_g \\ E_m \end{bmatrix}, J_{E_k}({}^S E q_k) = [J_{E_g}({}^S E q_k), J_{E_m}({}^S E q_k)]$$

$$E_g = \begin{bmatrix} 0 \\ 2(q_2 q_4 - q_1 q_3) - y_{ax} \\ 2(q_1 q_2 + q_3 q_4) - y_{ay} \\ 2(0.5 - q_2^2 - q_3^2) - y_{az} \end{bmatrix}$$

$$E_m = \begin{bmatrix} 0 \\ 2m_x(0.5 - q_3^2 - q_4^2) + 2m_z(q_2 q_4 - q_1 q_3) - y_{mx} \\ 2m_x(q_2 q_3 - q_1 q_4) + 2m_z(q_1 q_2 + q_3 q_4) - y_{my} \\ 2m_x(q_2 q_3 + q_1 q_4) + 2m_z(0.5 - q_2^2 - q_3^2) - y_{mz} \end{bmatrix}$$

$$J_{E_g}({}^S E q_k) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ -2q_3 & 2q_4 & -2q_1 & 2q_2 \\ 2q_2 & 2q_1 & 2q_4 & 2q_3 \\ 0 & -4q_2 & -4q_3 & 0 \end{bmatrix}$$

$$J_{E_m}({}^S E q_k) = \begin{bmatrix} 0 & 0 \\ -2m_z q_3 & 2m_z q_4 \\ -2m_x q_4 + 2m_z q_2 & 2m_x q_3 + 2m_z q_1 \\ 2m_x q_3 & 2m_x q_4 - 4m_z q_2 \\ 0 & 0 \\ -4m_x q_3 - 2m_z q_1 & -4m_x q_4 + 2m_z q_2 \\ 2m_x q_2 + 2m_z q_4 & -2m_x q_1 + 2m_z q_3 \\ 2m_x q_1 - 4m_z q_3 & 2m_x q_2 \end{bmatrix}. \quad (63)$$

where ${}^S E q_k^- = [q_1, q_2, q_3, q_4]^T$ is the quaternion obtained by gyroscope readings as shown in (42). One can observe that the only difference between (47) and (62) is the attenuation matrix W . By analogy to equations (48)–(51), we have

$${}^S E q_k = {}^S E q_k^- - \lambda \frac{\nabla f_{cl}}{\|\nabla f_{g,m}\|_2}. \quad (64)$$

In a practical application, we can use $\sigma_2 = \sigma_3 = \sigma_4 = \sigma_a$ for the accelerometer readings and $\sigma_6 = \sigma_7 = \sigma_8 = \sigma_m$ for the magnetometer readings due to the homogeneous property of sensors at different axes. The selections of σ_1 and σ_5 are

Algorithm 1 The CGD

- 1: Initialize ${}^S E q_0$ by (65)
- 2: **while** $k > 0$ **do**
- 3: Predict the quaternion ${}^S E q_k^-$ using (42)
- 4: Obtain the measurement error and the gradient using (63)
- 5: Update the quaternion using (64)
- 6: $k = k + 1$
- 7: **end while**

redundant since $E_{1,k} = E_{5,k} = 0$ always holds (note that the error quaternion is composed of error vectors). A remaining question is to obtain the initial quaternion ${}^S E q_0$. In this work, we use *ecompass* algorithm [37] to initialize the orientation. Specifically, the initial rotation matrix is calculated by

$$R_0 = [(y_A \times y_M) \times y_A \quad (-y_A) \times y_M \quad (-y_A)] \quad (65)$$

where y_A and y_M are the initial normalized accelerometer and magnetometer readings, and the operator \times denotes the cross product. Then, we convert this rotation matrix to a quaternion and obtain ${}^S E q_0$. The detailed algorithm of the CGD is summarized in Algorithm 1.

C. The CDOE

Based on the algorithm of the DOE described in Section III-A2, we can remodel the correction angles $\alpha_{err,a}$ and $\alpha_{err,m}$ as

$$\begin{aligned} \alpha_{err,a} &= \alpha_{ca} + v_{ca} \\ \alpha_{err,m} &= \alpha_{cm} + v_{cm} \end{aligned} \quad (66)$$

where α_{ca} and α_{cm} is the ground truth correction angle we want to obtain, and v_{ca} and v_{cm} are the noises caused by accelerometer and magnetometer measurement noises. By assuming that v_{ca} and v_{cm} are Gaussian, the obtainment of $k_a \alpha_{err,a}$ and $k_m \alpha_{err,m}$ in (56) and (58) can be given by solving the following LS-based optimization problems:

$$\min_{\alpha_{ca} \in \mathbb{R}} f_a(\alpha_{ca}) = \frac{1}{2} (\alpha_{ca} - \alpha_{err,a})^2, \quad (67a)$$

$$\min_{\alpha_{cm} \in \mathbb{R}} f_m(\alpha_{cm}) = \frac{1}{2} (\alpha_{cm} - \alpha_{err,m})^2. \quad (67b)$$

Solving (67a) and (67b) with the gradient descent strategy, we have

$$\alpha_{ca,t} = \alpha_{ca,t-1} - k_a \nabla f_a \quad (68a)$$

$$\alpha_{cm,t} = \alpha_{cm,t-1} - k_m \nabla f_m \quad (68b)$$

where $\nabla f_a = \frac{\partial f_a}{\partial \alpha_{ca}} \Big|_{\alpha_{ca,t-1}} = \alpha_{ca,t-1} - \alpha_{err,a}$, $\nabla f_m = \frac{\partial f_m}{\partial \alpha_{cm}} \Big|_{\alpha_{cm,t-1}} = \alpha_{cm,t-1} - \alpha_{err,m}$, t is the iteration number and starts from 1, $\alpha_{ca,t-1}$ and $\alpha_{cm,t-1}$ are the correction angles at the previous iteration, and k_a and k_m are step sizes. Note that the initial guess of the correction angle should be zero (since the *a priori* estimate of the correction angle is zero) and it is possible to iterate only one time at each time interval. Thus, we have $\alpha_{ca} = k_a \alpha_{err,a}$ and $\alpha_{cm} = k_m \alpha_{err,m}$ (ignoring the subscript t), which are identical to the corrected angles utilized in quaternion update in (56) and (58).

Algorithm 2 The CDOE

- 1: Initialize ${}^S_E q_0$ using (65) and set $b_0 = 0$
 - 2: **while** $k > 0$ **do**
 - 3: Calculate ${}^S_E q_k^-$ using (42) and (60)
 - 4: Obtain the correction angles $\alpha_{err,a}$ and $\alpha_{err,m}$, and calculate $q_{acor,k}$ and $q_{mcor,k}$ using (70) and (71)
 - 5: Correction by accelerometer readings using (73)
 - 6: Correction by magnetometer readings using (74)
 - 7: Gyroscope bias estimation using (72)
 - 8: $k = k + 1$
 - 9: **end while**
-

Since the Gaussian assumption of v_{ca} and v_{cm} is generally not valid due to the involvement of external acceleration and magnetic disturbance, we construct the following MKCL-based objective functions:

$$\min_{\alpha_{ca} \in \mathbb{R}} f_{cl,a}(\alpha_{ca}) = \sigma_a^2 (1 - G_{\sigma_a}(e_a)), \quad (69a)$$

$$\min_{\alpha_{cm} \in \mathbb{R}} f_{cl,m}(\alpha_{cm}) = \sigma_m^2 (1 - G_{\sigma_m}(e_m)) \quad (69b)$$

where $e_a = \alpha_{ca} - \alpha_{err,a}$, $e_m = \alpha_{cm} - \alpha_{err,m}$, and σ_a and σ_m are two bandwidths. By analogy to the one-step gradient descent update in (68a) and (68b), we have

$$\begin{aligned} \alpha_{ca} &= k_a G_{\sigma_a}(\alpha_{err,a}) \alpha_{err,a} \\ \alpha_{cm} &= k_m G_{\sigma_m}(\alpha_{err,m}) \alpha_{err,m} \end{aligned} \quad (70)$$

and thus

$$q_{acor,k} = \begin{bmatrix} \cos(\frac{1}{2}\alpha_{ca}) \\ \sin(\frac{1}{2}\alpha_{ca})x_a \end{bmatrix}, q_{mcor,k} = \begin{bmatrix} \cos(\frac{1}{2}\alpha_{cm}) \\ \sin(\frac{1}{2}\alpha_{cm})x_m \end{bmatrix}. \quad (71)$$

Similar to (59), the gyroscope drift b_k can be updated as

$$b_k = b_{k-1} + k_{b,a} \alpha_{ca} x_a + k_{b,m} \alpha_{cm} x_m. \quad (72)$$

Then, the quaternion update by accelerometer readings is given as

$${}^S_E q_{ga,k} = {}^S_E q_k^- \otimes q_{acor,k}. \quad (73)$$

Finally, the quaternion update by magnetometer readings is given as

$${}^S_E q_k = {}^S_E q_{ga,k} \otimes q_{mcor,k}. \quad (74)$$

The detailed algorithm of the CDOE is summarized in Algorithm 2.

D. Kernel Bandwidth Tuning

The kernel bandwidths play important roles in the CGD and CDOE. As indicated by Theorem 1, the performance of the CGD and CDOE is almost identical to the GD and DOE when using large kernel sizes. In this case, they perform well under Gaussian noise but are sensitive to external disturbance. On the contrary, a very small kernel bandwidth can reject disturbance effectively at the price of degraded ability in inhibiting gyroscope drift. One method for this trade-off is to minimize the ‘‘distance’’ between the objective function-induced pdf and the practical noise distribution, which falls into the category of MLE. Another method is to treat it as a black-box problem

and solve it with some black-box optimization algorithms, e.g., Bayesian optimization in [24]. Although the above two methods are effective, their performance largely depends on the training set, and obvious generalization errors may occur when the testing situation is different from the training case.

Inspired by the redescending property of the MKCL as shown in Fig. 4(b), one can select kernel bandwidths manually by inspecting the distribution of measurement residuals. Specifically, the influence function of the MKCL in one dimension case has $\nabla J_{GL}(e) = \frac{J_{GL}(e)}{\partial e} = e \exp^{-\frac{e^2}{2\sigma^2}}$, which gives its maximum (and minimum) value at point $e = \sigma$ (and $e = -\sigma$). In the case of $e = \pm 3\sigma$, one has $\nabla J_{GL}(e) \approx \pm 0.011e$, which indicates that the residuals with $|e| \geq 3\sigma$ nearly have no influence on the objective function and hence the residuals polluted by disturbances are expected to locate outside of the region $[-3\sigma, 3\sigma]$. On the contrary, the influence functions of the MKCL and the LS are similar within the region $[-\sigma, \sigma]$, and therefore the residuals under the disturbance-free experiments are expected to be in this region. By these two principles, one can adjust the kernel bandwidth effectively. Empirically, we advise using $\sigma_a = 2d_a$ and $\sigma_m = 2d_m$ where d_a and d_m are the standard deviations of the accelerometer residuals and magnetometer residuals under the disturbance-free experiments.

IV. EXPERIMENTS

In this section, we compare the CGD and CDOE with the GD [17], DOE [18], ESKF [15], and MKMC [24]. We first investigate their performances using a commercial IMU. Then, we implement them on a self-designed low-cost IMU and compare their accuracy under some walking experiments.

A. Performance Validation Using a Commercial IMU

We validate the performance of different algorithms on a commercial sensor, Xsens MTI-670. The sampling frequency is 400 Hz. Two kinds of motions, translation (T) and translation adjoint with rotation (T & R), are considered. The experimental setups are shown in Fig. 5(a) and Fig. 5(b), and the detailed experimental descriptions are summarized in Table. I. It is worth mentioning that the IMU is disturbed either by acceleration disturbance A_d or magnetic disturbance M_d in experiment 1, while it is disturbed by A_d and M_d simultaneously in experiment 5. Moreover, there is almost *no external disturbance* in experiment 2 since the rotation frequency is very low.

TABLE I
EXPERIMENTS DESCRIPTION.

Exp	Movement	Disturbance
1	T	A_d or M_d
2	T & R	$f = 0.1Hz$, almost no disturbance
3	T & R	$f = 0.5Hz$, A_d
4	T & R	$f = 0.1Hz$, M_d
5	T & R	$f = 0.5Hz$, A_d and M_d

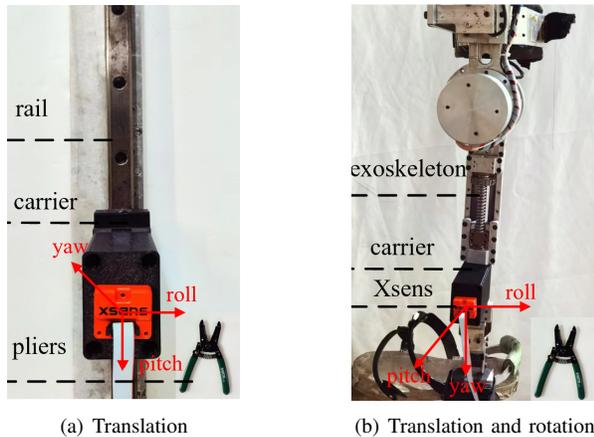


Fig. 5. The experimental setups. In Fig. 5(a), Xsens is attached to a carrier that is connected to a rail. The movement of Xsens is controlled manually. The external acceleration A_d is generated by the movement of Xsens, while the magnetic disturbance M_d is caused by the approaching of pliers. In Fig. 5(b), Xsens is attached to the shank of an exoskeleton, and the exoskeleton is commanded to imitate the walking of a human. The magnitude of A_d depends on the rotation frequency f , while M_d is determined by the distance between the pliers and Xsens. The plastic carrier has a height of 12 cm which is sufficient to isolate the magnetic effects caused by rails or exoskeletons.

The kernel bandwidths of the correntropy-based algorithms are tuned based on the strategies described in Section III-D. As for the other parameters, we use the same values for each comparison pair (i.e., the ESKF and MKMC, the GD and CGD, and the DOE and CDOE). The root-mean-square errors (RMSEs) and maximum errors (MEs) of different algorithms under experiments 1-5 are summarized in Table II. One can see that the performances of the correntropy-based algorithms (i.e., MKMC, CGD, and CDOE) are slightly better than their counterparts (i.e., ESKF, GD, and DOE) in experiment 2, and significantly outperforms the ESKF, GD, and DOE in other experiments. This reveals that the correntropy-based algorithms can increase the robustness of the traditional algorithms against disturbances but almost without sacrificing their performances under Gaussian noises. This feature gives its root in the redescending property of the MKCL's influence function.

We visualize the error performances of different methods in experiments 1 and 5 in Fig. 8. The norm of the sensor readings is shown in Figs. 8(a) and 8(c) while the corresponding errors are shown in Figs. 8(b) and 8(d). One can observe that three correntropy-based algorithms (i.e., the MKMC, CGD, and CDOE) are very robust to external disturbances, especially along the yaw and pitch axes since these two axes are affected by A_d and M_d the most.

To further investigate the effects of kernel bandwidths on the proposed algorithms, we visualize the orientation errors of the CGD and CDOE by applying different bandwidth vectors in experiment 1 and the corresponding results are shown in Figs. 6 and 7. Not surprisingly, the CGD and CDOE are almost identical to the GD and DOE when applying big kernel bandwidths (see the dashed black and yellow lines), and are very robust to disturbances when using relatively small kernel bandwidths. This result is consistent with the log-likelihood

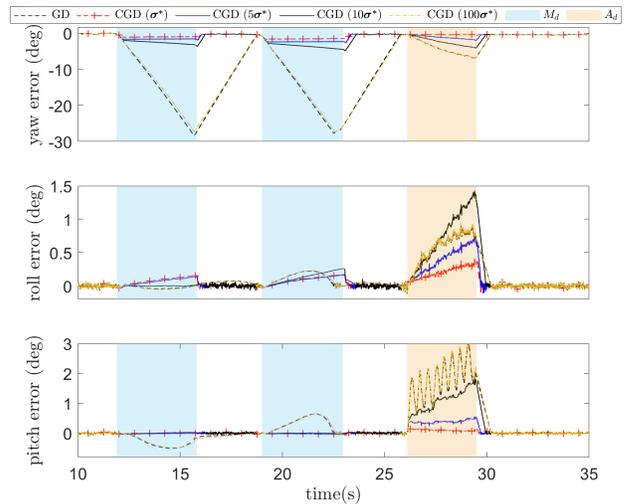


Fig. 6. Performances of the CGD with different kernel bandwidth vectors in experiment 1. The tuned bandwidth vector for the CGD has $\sigma^* = [\sigma_a, \sigma_m]^T = [0.02, 0.01]^T$.

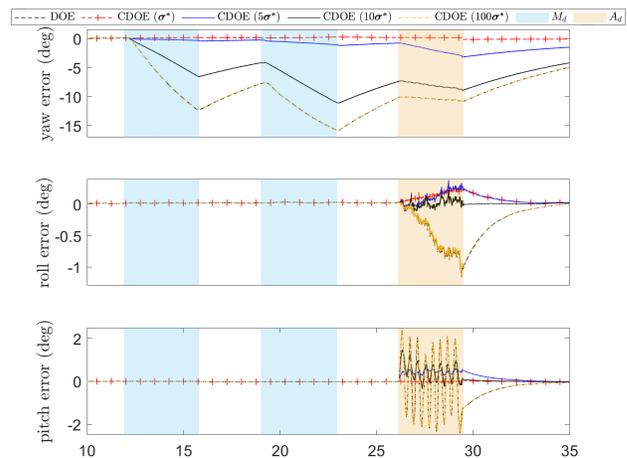


Fig. 7. Performances of the CDOE with different kernel bandwidth vectors in experiment 1. The tuned bandwidth vector for the CDOE has $\sigma^* = [\sigma_a, \sigma_m]^T = [0.05, 0.04]^T$.

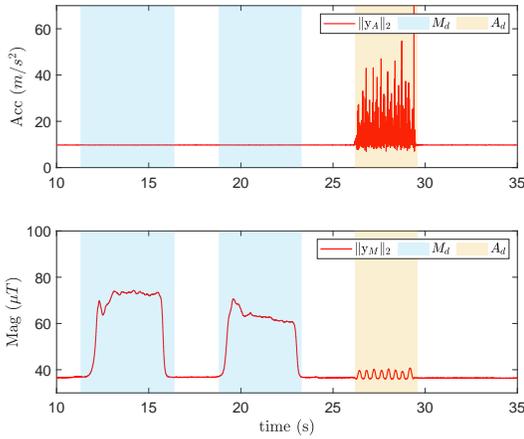
comparison as shown in Fig. 3 which indicates that small kernel bandwidths generally are robust to heavy-tailed noises.

B. Performance Validation on a Low-cost IMU

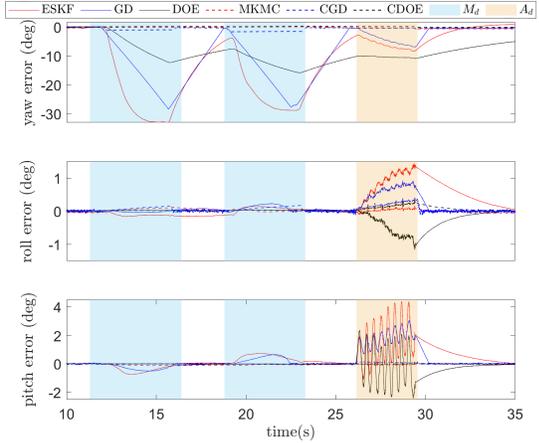
We implement our algorithms to a low-cost IMU (micro-processor: STM32F405RGT6, embedded sensor: ICM20948) which integrates an SD card and an onboard battery. The IMU has been calibrated based on the procedure introduced in [38]. We attach the IMU to a shoe and investigate its performance under some walking tests. The experimental setup is shown in Fig. 9 and a detailed experimental description is shown in Table IV. The ground truth orientation is obtained by the Vicon motion capture system with 8 cameras (Vicon Vero v2.2), which is highly reliable with 0.017 mm (mean) position tracking error [39]. For comparison, the optical data and the IMU output are time-synchronized and aligned. The time synchronization is achieved by correlating the norms of the gyroscope readings and of the angular velocity estimated by

TABLE II
RMSES AND MES OF DIFFERENT ALGORITHMS USING XSSENS.

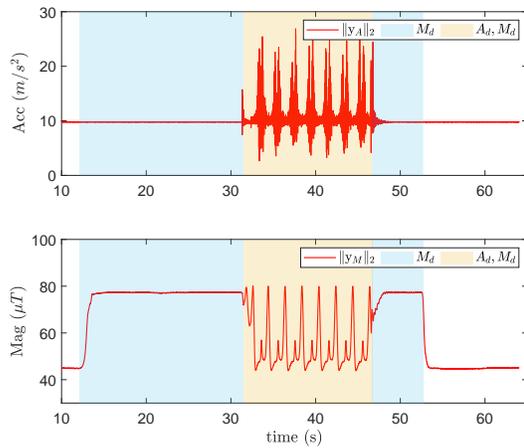
Test	Axis	RMSE (deg)						ME (deg)					
		GD	CGD	DOE	CDOE	ESKF	MKMC	GD	CGD	DOE	CDOE	ESKF	MKCM
1	yaw	10.22	0.69	8.10	0.13	13.54	0.31	28.40	1.68	15.83	0.26	32.95	0.97
	roll	0.21	0.09	0.22	0.06	0.39	0.05	0.93	0.42	1.17	0.33	1.44	0.16
	pitch	0.62	0.04	0.46	0.01	0.77	0.05	3.07	0.16	2.41	0.13	4.40	0.17
2	yaw	1.87	1.84	1.29	0.95	1.34	0.50	5.55	5.41	2.88	2.16	3.82	0.82
	roll	0.60	0.52	0.39	0.21	0.59	0.30	2.02	1.68	1.41	0.67	1.46	0.79
	pitch	0.16	0.27	0.10	0.09	0.51	0.13	0.47	0.88	0.27	0.29	1.38	0.58
3	yaw	0.94	0.44	0.74	0.53	0.67	0.15	2.69	1.29	2.65	1.18	2.01	0.50
	roll	0.79	0.11	1.88	0.23	0.82	0.14	3.26	0.65	7.72	1.11	3.35	0.79
	pitch	0.23	0.16	0.44	0.14	0.33	0.14	0.82	0.77	1.40	0.73	1.47	0.58
4	yaw	17.03	2.38	12.96	2.31	17.77	0.69	31.16	4.74	27.56	3.72	34.25	1.47
	roll	1.67	0.40	0.34	0.17	3.48	0.23	4.94	1.35	1.20	0.58	9.93	0.69
	pitch	1.46	0.17	0.06	0.08	2.56	0.09	4.31	0.79	0.22	0.30	6.32	0.64
5	yaw	16.30	1.29	13.12	1.67	16.60	1.42	26.63	2.64	24.05	2.93	28.81	2.98
	roll	1.07	0.78	1.61	0.20	2.70	0.33	3.79	2.34	7.34	1.05	8.37	1.19
	pitch	0.92	0.16	0.34	0.19	1.82	0.19	3.46	0.66	1.30	0.80	4.87	0.52



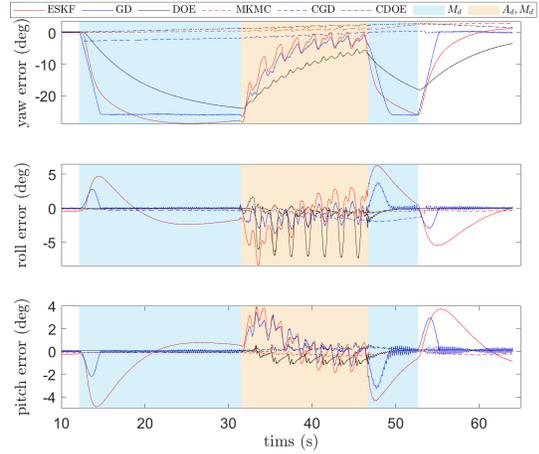
(a) Norm of sensor readings in experiment 1



(b) Orientation errors of different methods in experiment 1



(c) Norm of sensor readings in experiment 5



(d) Orientation errors of different methods in experiment 5

Fig. 8. Performances of different algorithms in experiments 1 and 5. The left two figures show the norm of the accelerometer and magnetometer readings while the right ones show the orientation errors of different algorithms.

the optical system. The alignment is done by an optimization-based method shown in [40]. Four experiments are conducted to cover the situations with or without magnetic disturbance and straight walking or walking with turns.

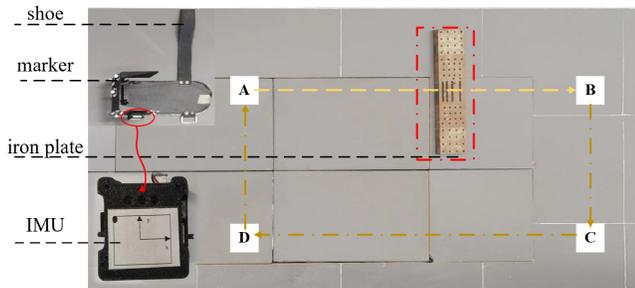


Fig. 9. Experimental setup for the walking test. The low-cost IMU is attached to a shoe. The external acceleration is generated by the moving of the leg while the magnetic disturbance is generated by the iron plate.

We summarize the RMSEs and MEs of different algorithms under different tests in Table III. One can see that the performances of the correntropy-based algorithms again significantly outperform their traditional counterparts, especially in terms of MEs.

We visualize the norm of sensor readings and the error performance of different algorithms under Test 4 in Figs. 10 and 11. The magnetic disturbance area is shown in the blue regions. One can see that obvious yaw errors are caused by the ESKF, GD, and DOE in this region, but are avoided by the MKMC, CGD, and CDOE.

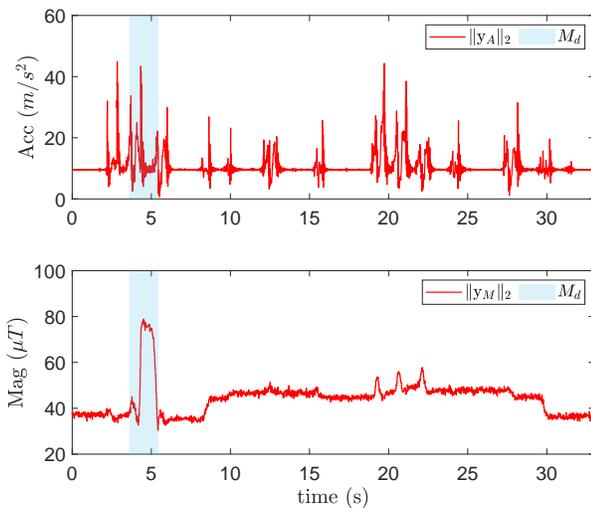


Fig. 10. The norm of sensor readings in Test 4.

C. Algorithm Complexity

Due to the need of matrix inversion in the ESKF and both matrix inversion and Cholesky decomposition in the MKMC, the algorithm complexity of the ESKF and MKMC is heavier than $O(n^3)$ where $n = 12$ (one can refer to [24] for the detailed algorithm of the MKMC). These time-consuming calculations can be avoided by the CGD and CDOE. To explicitly analyze the time consumption of different algorithms

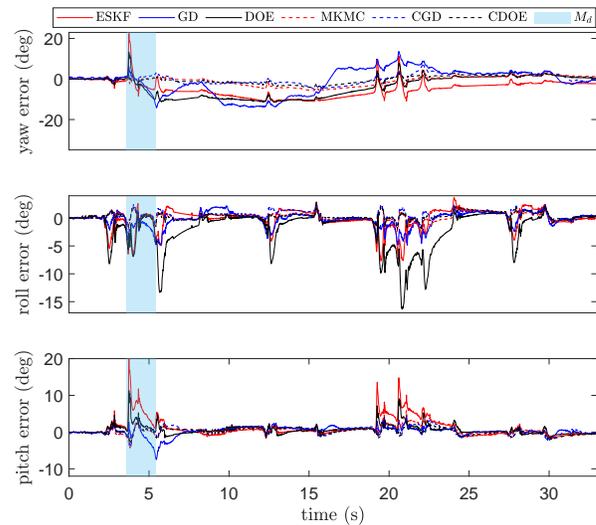


Fig. 11. The orientation errors of different algorithms in Test 4.

in an embedded low-cost microprocessor, we execute them on a low-cost IMU (computation chip: STM32F405RGT6). The average time consumption in one iteration is obtained by averaging the time cost of 10000 execution iterations and the corresponding results are summarized in Table V. One can see that the complexity of the CGD and CDOE are slightly heavier than the GD and DOE, but are significantly lighter than the MKMC and ESKF. The time costs are as large as 2.515 ms and 3.536 ms for the ESKF and MKMC, but merely 0.059 ms and 0.080 ms for the CGD and CDOE.

D. Discussion

From Tables II and III, we observe that the correntropy-based algorithms (i.e., the CGD, CDOE, and MKMC) are significantly better than their counterparts (i.e., the GD, DOE, and ESKF). The effectiveness of the proposed algorithms implies that many existing orientation algorithms can be further improved by re-modeling the objective functions so that the cost-induced distribution much more matches the underlying noise distribution. “Optimizing” the objective function is powerful but is not fully explored in the literature, especially in the field of sensor fusions where the LS criterion is widely used. Our work demonstrates that the MKCL is a suitable candidate when the underlying noise is heavy-tailed. It is worth mentioning that the Gaussian kernel is not the only kernel function for the formulation of the MKCL, but also some other kernels (e.g., generalized Gaussian kernels, Cauchy kernels). Compared with the loss functions utilized in robust statistics, e.g., the Huber loss, least absolute loss, and elastic net, the MKCL is much more flexible (by varying kernel bandwidths) and possesses a redescending influence function. Other types of non-convex penalties should also be useful as long as they can induce suitable heavy-tailed distributions, which gives great freedom to the designers.

We observe that the performance of the CDOE is slightly better than the CGD in our experiments. One reason may be that the modeling of the gyroscope bias is considered in the

TABLE III
RMSES AND MES OF DIFFERENT ALGORITHMS IN THE WALKING TEST.

Test	Axis	RMSE (deg)						ME (deg)					
		GD	CGD	DOE	CDOE	ESKF	MKMC	GD	CGD	DOE	CDOE	ESKF	MKMC
1	yaw	1.00	0.95	1.02	0.46	1.20	0.70	4.06	2.98	6.06	1.80	7.64	2.78
	roll	1.83	0.60	4.34	0.84	2.31	0.95	5.03	1.76	16.95	2.42	7.98	2.54
	pitch	0.75	0.87	0.98	0.74	1.02	0.91	2.59	3.30	5.48	2.56	5.23	3.24
2	yaw	4.78	1.28	7.90	0.69	1.84	0.64	16.08	3.08	13.23	2.10	13.38	2.49
	roll	3.15	0.33	4.24	0.47	2.69	0.34	10.35	1.14	14.34	1.47	10.45	1.23
	pitch	3.12	1.01	1.50	0.73	1.78	0.81	10.54	2.97	8.92	2.42	11.37	2.62
3	yaw	6.35	2.05	2.86	1.30	2.82	1.69	17.78	9.01	8.29	3.81	16.35	5.30
	roll	1.30	0.64	3.67	0.63	2.09	0.74	4.68	2.52	14.94	2.71	10.29	2.47
	pitch	1.42	1.29	1.92	1.07	2.79	1.61	5.43	5.85	9.24	4.57	13.74	7.06
4	yaw	6.52	2.40	6.32	1.88	6.31	2.48	14.18	7.36	13.28	4.16	22.41	5.69
	roll	1.35	0.84	3.59	0.73	1.78	0.78	4.85	2.45	16.36	2.28	7.68	3.11
	pitch	1.26	1.10	1.47	0.91	2.81	1.26	7.45	4.48	11.31	3.98	20.06	6.00

TABLE IV
EXPERIMENTAL DESCRIPTION.

Test	Route	Disturbance
1	A→B	A_d , without iron plate
2	A→B	A_d, M_d , with iron plate
3	A→B→C→D→A	A_d , without iron plate
4	A→B→C→D→A	A_d, M_d , with iron plate

TABLE V
EXECUTION TIME OF DIFFERENT METHODS ON A LOW-COST IMU.

Algorithm	GD	CGD	DOE	CDOE	ESKF	MKMC
Time (ms)	0.053	0.059	0.075	0.080	2.515	3.536

CDOE but ignored in the CGD. Another reason may be that although the proposed methods are very effective in rejecting big disturbances, they are less effective in handling small but lasting disturbances. In many indoor applications, the magnetic field is not strictly homogeneous and may vary with position (one can see the norm of the magnetometer readings in Fig. 10). In this case, the CDOE outperforms the CGD since it constrains that the magnetic readings only affect the heading, but the small lasting magnetic disturbance may deteriorate both the pitch and roll in the CDOE. Conceptually, it is hard to distinguish the small disturbance from the nominal noise since they share similar regions in the density functions. One possible solution may be extending the zero-mean Gaussian kernel to the variable center Gaussian kernel so that the small disturbance can be mitigated by the shifted centers, which may be our future work.

V. CONCLUSION

In this paper, we build a connection between the MKCL and its induced noise distribution and demonstrate that this distribution becomes Gaussian with infinite kernel bandwidth. Some important properties of the MKCL as a cost function are given. Moreover, two MKCL-based algorithms (i.e., the CGD and CDOE) are derived for the orientation estimation of IMUs. The proposed approaches exhibit robustness to external

disturbances, and their performances are verified under extensive experiments. Our proposed two methods bear significantly less complexity compared with the ESKF and MKMC, which should be beneficial when the microprocessor computation resource is limited and can save power consumption in a practical implementation. To avoid the complicated kernel bandwidth procedure, in the future, we will focus on the adaptive kernel bandwidth tuning strategies. Additionally, we plan to extend the Gaussian kernel to the variable center Gaussian kernel to reject the small but lasting disturbances.

VI. APPENDIX

A. Proof of Theorem 1

Proof. Taking Taylor series expansion of $G_{\sigma_i}(e_{i,k})$ gives

$$G_{\sigma_i}(e_{i,k}) = \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n \sigma_i^{2n} n!} e_{i,k}^{2n}$$

By setting $\sigma_i \rightarrow \infty$, it follows

$$\lim_{\sigma_i \rightarrow \infty} \sigma_i^2 (1 - G_{\sigma_i}(e_{i,k})) = e_{i,k}^2 / 2 \quad (75)$$

Substituting this result into (4), one obtains

$$\lim_{\sigma_i \rightarrow \infty} J_{GL} = \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^l e_{i,k}^2 / 2 = J_{LS}. \quad (76)$$

This completes the proof. \blacksquare

B. Proof of Theorem 2

Proof. If $\tilde{e}_k \sim \mathcal{N}(0, I)$, one has $p(\tilde{e}_{i,k}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\tilde{e}_{i,k}^2}{2}\right)$. Then, the likelihood of x given the set $\{y_k, u_k\}_{k=1}^N$ follows

$$\begin{aligned} \mathcal{L}(x; \{y_k, u_k\}_{k=1}^N) &= \prod_{k=1}^N p(\tilde{e}_k) \\ &= \prod_{k=1}^N \prod_{i=1}^l \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\tilde{e}_{i,k}^2}{2}\right). \end{aligned}$$

Based on MLE, it follows that

$$x = \arg \max_x \mathcal{L} \left(x; \{y_k, u_k\}_{k=1}^N \right)$$

which is equivalent to minimizing its negative logarithm function (ignoring the normalization constants) with

$$x = \arg \min_x \frac{1}{2} \sum_{k=1}^N \sum_{i=1}^l \tilde{e}_{i,k}^2.$$

Multiplying a constant $1/N$ on the right side of the above equation, we obtain (33a). This reveals that LS is an optimal metric when the noise is Gaussian under the MLE. On the contrary, if $\tilde{e}_{i,k}$ is heavy-tailed and its density function $p(\tilde{e}_{i,k})$ follows (34), one has

$$\begin{aligned} x &= \arg \max_x \sum_{k=1}^N \sum_{i=1}^l c_i \exp \left(-\sigma_i^2 \left(1 - \exp \left(-\frac{\tilde{e}_{i,k}^2}{2\sigma_i^2} \right) \right) \right) \\ &= \arg \min_x \sum_{k=1}^N \sum_{i=1}^l \sigma_i^2 \left(1 - G_{\sigma_i}(\tilde{e}_{i,k}) \right) \\ &= \arg \min_x \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^l \sigma_i^2 \left(1 - G_{\sigma_i}(\tilde{e}_{i,k}) \right) \\ &= \arg \min_x J_{CL}(\tilde{e}_k). \end{aligned} \quad (77)$$

This indicates that MKCL is an optimal metric when $\tilde{e}_{i,k}$ follows (34) and hence completes the proof. ■

REFERENCES

- [1] T.-H. Huang, S. Zhang, S. Yu, M. K. MacLean, J. Zhu, A. Di Lallo, C. Jiao, T. C. Bulea, M. Zheng, and H. Su, "Modeling and stiffness-based continuous torque control of lightweight quasi-direct-drive knee exoskeletons for versatile walking assistance," *IEEE Transactions on Robotics*, vol. 38, no. 3, pp. 1442–1459, 2022.
- [2] J. Otegui, A. Bahillo, I. Lopetegui, and L. E. Díez, "Simulation framework for testing train navigation algorithms based on 9-DOF-IMU and tachometers," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 7, pp. 5260–5273, 2020.
- [3] Y. Wang, H. Cheng, C. Wang, and M. Q.-H. Meng, "Pose-invariant inertial odometry for pedestrian localization," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021.
- [4] M. Kok, J. D. Hol, and T. B. Schön, "Using inertial sensors for position and orientation estimation," *arXiv preprint arXiv:1704.06053*, 2017.
- [5] J. D. Hol, "Sensor fusion and calibration of inertial sensors, vision, ultra-wideband and gps," Ph.D. dissertation, Linköping University Electronic Press, 2011.
- [6] S. Takano, T. Shimono, K. Masaki, K. Fukunaga, H. Kabata, M. Nishie, T. Ezaki, H. Nakada, J. Hakamata, and A. Hasegawa, "An inhalation device with inertial measurement unit for monitoring inhaler technique," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 4, pp. 2204–2211, 2022.
- [7] C. Yi, B. Wei, Z. Ding, C. Yang, Z. Chen, and F. Jiang, "A self-aligned method of IMU-based 3-DoF lower-limb joint angle estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–10, 2022.
- [8] A. Hooshyar, A. Sayadi, J. Dargahi, and S. Najarian, "Integral-free spatial orientation estimation method and wearable rotation measurement device for robot-assisted catheter intervention," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 2, pp. 766–776, 2022.
- [9] R. Mahony, T. Hamel, and J.-M. Pfimlin, "Complementary filter design on the special orthogonal group SO(3)," in *Proceedings of the 44th IEEE Conference on Decision and Control*, 2005, pp. 1477–1484.
- [10] R. Mahony, T. Hamel, and J.-M. Pfimlin, "Nonlinear complementary filters on the special orthogonal group," *IEEE Transactions on Automatic Control*, vol. 53, no. 5, pp. 1203–1218, 2008.
- [11] A. Sabatini, "Quaternion-based extended Kalman filter for determining orientation by inertial and magnetic sensing," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 7, pp. 1346–1356, 2006.
- [12] B. Candan and H. E. Soken, "Robust attitude estimation using IMU-Only measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021.
- [13] Y. S. Suh, "Orientation estimation using a quaternion-based indirect Kalman filter with adaptive estimation of external acceleration," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 12, pp. 3296–3305, 2010.
- [14] D. Roetenberg, C. T. M. Baten, and P. H. Veltink, "Estimating body segment orientation by applying inertial and magnetic sensing near ferromagnetic materials," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 15, no. 3, pp. 469–471, 2007.
- [15] D. Roetenberg, H. Luinge, C. Baten, and P. Veltink, "Compensation of magnetic disturbances improves inertial and magnetic sensing of human body segment orientation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 3, pp. 395–405, 2005.
- [16] M. S. Mark Pedley and Z. Baranski, "Freescale sensor fusion Kalman filter," <https://github.com/memsindustrygroup/Open-Source-Sensor-Fusion/tree/master/docs>, pp. 24–26, 2014.
- [17] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *Proceedings of IEEE International Conference on Rehabilitation Robotics*, 2011, pp. 1–7.
- [18] T. Seel and S. Ruppel, "Eliminating the effect of magnetic disturbances on the inclination estimates of inertial sensors," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 8798–8803, 2017.
- [19] A. El Hadri and A. Benallegue, "Sliding mode observer to estimate both the attitude and the gyro-bias by using low-cost sensors," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009, pp. 2867–2872.
- [20] D. A. Aligia, B. A. Rocca, C. H. De Angelo, G. A. Magallán, and G. N. González, "An orientation estimation strategy for low cost IMU using a nonlinear luenberger observer," *Measurement*, vol. 173, p. 108664, 2021.
- [21] R. V. Vitali, R. S. McGinnis, and N. C. Perkins, "Robust error-state Kalman filter for estimating imu orientation," *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3561–3569, 2020.
- [22] R. G. Valenti, I. Dryanovski, and J. Xiao, "A linear Kalman filter for marg orientation estimation using the algebraic quaternion algorithm," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 2, pp. 467–481, 2015.
- [23] S. Li, P. Duan, D. Shi, W. Zou, P. Duan, and L. Shi, "Compact maximum correntropy-based error state Kalman filter for exoskeleton orientation estimation," *IEEE Transactions on Control Systems Technology*, pp. 1–8, 2022.
- [24] S. Li, L. Li, D. Shi, W. Zou, P. Duan, and L. Shi, "Multi-kernel maximum correntropy Kalman filter for orientation estimation," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6693–6700, 2022.
- [25] B. Chen, X. Liu, H. Zhao, and J. C. Principe, "Maximum correntropy Kalman filter," *Automatica*, vol. 76, pp. 70–77, 2017.
- [26] M. Kulikova, "Chandrasekhar-based maximum correntropy Kalman filtering with the adaptive kernel size selection," *IEEE Transactions on Automatic Control*, vol. 65, no. 2, pp. 741–748, 2020.
- [27] S. Li, D. Shi, W. Zou, and L. Shi, "Multi-kernel maximum correntropy Kalman filter," *IEEE Control Systems Letters*, vol. 6, pp. 1490–1495, 2022.
- [28] A. Singh and J. C. Principe, "Using correntropy as a cost function in linear adaptive filters," in *Proceedings of International Joint Conference on Neural Networks*, 2009, pp. 2950–2955.
- [29] B. Chen, X. Wang, N. Lu, S. Wang, J. Cao, and J. Qin, "Mixture correntropy for robust learning," *Pattern Recognition*, vol. 79, pp. 318–327, 2018.
- [30] S. Li, D. Shi, W. Zou, and L. Shi, "Multi-kernel maximum correntropy Kalman filter," *IEEE Control Systems Letters*, vol. 6, pp. 1490–1495, 2021.
- [31] S. Särkkä and L. Svensson, *Bayesian filtering and smoothing*. Cambridge university press, 2023, vol. 17.
- [32] M. Kok, J. Dahlin, T. B. Schön, and A. Wills, "Newton-based maximum likelihood estimation in nonlinear state space models," *IFAC-PapersOnLine*, vol. 48, no. 28, pp. 398–403, 2015.
- [33] H. P. Gavin, "The levenberg-marquardt algorithm for nonlinear least squares curve-fitting problems," *Department of civil and environmental engineering, Duke University*, vol. 19, 2019.
- [34] D. Simon, *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, 2006.
- [35] M. Kok and T. B. Schön, "A fast and robust algorithm for orientation estimation using inertial sensors," *IEEE Signal Processing Letters*, vol. 26, no. 11, pp. 1673–1677, 2019.

- [36] A. Aravkin, J. V. Burke, L. Ljung, A. Lozano, and G. Pillonetto, “Generalized Kalman smoothing: Modeling and algorithms,” *Automatica*, vol. 86, pp. 63–86, 2017.
- [37] T. Ozyagcilar, “Implementing a tilt-compensated ecompass using accelerometer and magnetometer sensors,” *Freescale semiconductor AN*, vol. 4248, 2012.
- [38] D. Tedaldi, A. Pretto, and E. Menegatti, “A robust and easy to implement method for imu calibration without external equipments,” in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 3042–3049.
- [39] “Accuracy in motion,” <https://www.vicon.com/wp-content/uploads/2022/07/Vicon-Metrology-Solutions.pdf>, 2022.
- [40] Y. Jin, Y. M. Zhou, C. M. McCann, T. Proietti, C. H. Rycroft, and C. J. Walsh, “A data-based approach to simultaneously align local and global frames between an inertial measurement unit (imu) and an optical motion capture system,” in *2022 9th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechanics (BioRob)*. IEEE, 2022, pp. 1–8.