Multi-kernel Maximum Correntropy Kalman Filter for Orientation Estimation

Shilei Li[1], Lijing Li[2], Dawei Shi[3], Wulin Zou[1], Pu Duan[1], and Ling Shi[4]

***Abstract*— Inertial measurement units (IMUs), composed of gyroscopes, accelerometers, and magnetometers, have been widely used in many fields. However, their performances degenerate remarkably with external acceleration and magnetic disturbance. To handle this issue, we employ a multi-kernel maximum correntropy Kalman filter (MKMCKF) to suppress the adversarial acceleration and magnetic disturbance and use Bayesian optimization (BO) to explore the optimal kernel bandwidths. We validate our algorithm in a set of experiments with different levels of disturbance. Results show that the proposed method is significantly better than the traditional error state Kalman filter (ESKF) and the gradient descent (GD) method, and its root mean square error (RMSE) is less than $0.4629°$on the roll and pitch even under the worst testing case.**

I. INTRODUCTION

Micro-Electro-Mechanical-system based IMUs have attracted continuing research efforts for their advantages of low cost, portability, and being free of restrictions. They are found in most tablet computers, smart-phones and virtual reality (VR) headsets, and have been widely used in the field of motion capture, robotics [1], navigation, localization [2], and rehabilitation.

IMUs are usually composed of gyroscopes, accelerometers, and magnetometers where gyroscopes measure the angular velocity rate while accelerometers and magnetometers detect the gravity vector and the magnetic vector, and are used to correct the long-term gyroscope drift. They are also referred to as attitude and heading reference system (AHRS) or magnetic angular rate and gravity system (MARG), and many algorithms have been designed to solve this orientation estimation issue. Among them, the Kalman filter-based method is a popular one and has been widely used in the motion capture and robotics [3]. Other methods, such as complementary filter [4], gradient descend (GD) [5], and sliding mode observer [6], are also developed with different characteristics and are used for different applications. Challenging problems for IMUs are that the accelerometer is usually contaminated by external accelerations and the magnetometer is easily distorted by ferromagnetic materials, like iron and magnet. Unfortunately, the acceleration disturbance and magnetic disturbance are difficult to be modeled accurately since we do not have *a priori* knowledge about their statistical distributions. Therefore, inevitably, the estimation accuracy would deteriorate remarkably with disturbance. To handle this problem, Li et al. [7] used a Gaussian-Markov model to estimate the external acceleration; Roetenberg et al. [8] employed an adaptive measurement covariance to eliminate the influence of the magnetic disturbance; Madgwick et al. [5] re-normalized the magnetic vector to minimize its influence on inclination portions (i.e., roll and pitch); Seel and Ruppin [9] used a constraint so that the inclination was only affected by accelerometers while the heading was only affected by magnetometers. Those methods, in general, can mitigate the bad effects of external accelerations and magnetic disturbance to some extent. However, their performances are usually unsatisfactory since the estimation accuracy without disturbance is sacrificed.

Maximum correntropy criterion (MCC) provides a potential solution for the aforementioned issue. Correntropy is an information-theoretic measure of the similarity between two arbitrary random variables [10]. It is robust to heavy-tailed noises and has been widely used in regression [11], machine learning [12], Kalman filter (KF) [13] or its variants [14], [15], and adaptive filtering [16]. Aravkin et al. [17] showed that a mixed-norm penalty function (combined ℓ_1 norm and ℓ_2 norm) can effectively suppress the disturbance. Moreover, Liu et al. [10] reported that the correntropy induced metric (CIM) divided space into three regions: when two samples were close, the CIM was similar to an ℓ_2 norm; when they went further apart, the CIM was approaching an ℓ_1 norm; when the distance between the two samples was very large, the CIM was approximated by an ℓ_0 norm. The scale of the norm can be controlled by the kernel bandwidth which gives great flexibility for the algorithm design when considering different types of noises (i.e., the Gaussian noises and the heavy-tailed noises). Using this property, in our previous work, Li et al. [18] designed a multi-kernel maximum correntropy Kalman filter (MKMCKF) which utilized different bandwidths for different channels so that the CIM can be adjusted among an ℓ_2 norm, ℓ_1 norm, and ℓ_0 norm flexibly in accordance with the instantaneous error characteristics. Specifically, when the error is relatively small, the corresponding CIM behaves like an ℓ_2 norm. In this case, the MKMCKF is nearly identical to the KF. On the contrary, when the system involves outliers or non-Gaussian noises,

The work by S. Li and L. Shi is supported by a Hong Kong RGC General Research Fund 16206620.

[1]Shilei Li, Wulin Zou and Pu Duan are with the Control Department, Xeno Dynamics, Co., Ltd, P.R. China (e-mail: lishilei@xeno.com, zouwulin@xeno.com, duanpu@xeno.com).

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

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

[4]Ling Shi is with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, China (e-mail: eesling@ust.hk).

the error is expected to be bigger and the CIM behaves like an ℓ_1 norm or ℓ_0 norm. In this situation, the MKCMKF is very robust to outliers or disturbance. Therefore, this algorithm rejects disturbance effectively with the existence of disturbance and performs well without disturbance.

A remaining issue for the MKMCKF is the kernel bandwidths selection. Generally, we aim to choose a set of kernel bandwidths so that the MKMCKF is similar to the traditional KF with an ℓ_2 norm-based objective function when without disturbance and is similar to an ℓ_0 or ℓ_1 norm-based objective function with disturbance. However, it is difficult to tune the bandwidths manually, especially when the system's dimension is high or with constraints. To handle this issue, in this paper, we employ the Bayesian optimization (BO) for kernel bandwidths selection which is an efficient tool for global minimization where the objective function can be black-box, non-convex, and expensive to be evaluated [19], [20]. We first derive a novel algorithm named multi-kernel maximum correntropy Kalman filter for orientation estimation (MKMCKF-OE), then we explore the optimal bandwidths using BO on the training set. Finally, we validate the performance of the proposed algorithm on the testing set and compare it with the gradient descent (GD) [5], traditional error state Kalman filter (ESKF) [8], [21], improved gradient descent (IGD) [9], and versatile quaternion-based filter (VQF) [22]. The contributions of this paper are summarized as follows: a novel MKMCKF-OE is designed for IMUs, which performs well both with and without disturbance; the BO is utilized to find the optimal kernel bandwidths, which is highly efficient and convenient; the performance of the proposed algorithm is validated on the testing set and is shown to be significantly better than the GD and ESKF, and is superior to the IGD and VQF on the yaw axis with magnetic disturbance.

The remainder of this paper is organized as follows. In Section II, we give some preliminaries. In Section III, we derive a novel algorithm MKMCKF-OE for IMUs and explore the optimal kernel bandwidths by BO. In Section IV, we verify the effectiveness of the proposed methods in experiments. In section V, we draw a conclusion.

Notations: The transpose of a matrix A is denoted by A^T. The symbol $\times$ is the vector product operation and $\mathrm{v}\times$ converts a vector v to a skew symmetric matrix. The Gaussian distribution with mean μ and covariance Σ is denoted by $\mathcal{N}(\mu, \Sigma)$.

II. PRELIMINARIES

In this section, we give a brief introduction about sensor models of IMUs, MKMCKF, and BO.

A. Sensor Models

We give sensor models of gyroscopes, accelerometers, and magnetometers. Firstly, the gyroscope signal $y_{G,k} \in \mathbb{R}^3$ at time step k is composed of the angular velocity w_k, the offset b_k, and the white noise $v_{G,k}$. The offset is modeled as a Markov process with an additional Gaussian noise $v_{b,k}$. Then, we have

$$\begin{aligned} y_{G,k} &= w_k + b_k + v_{G,k} \\ b_k &= b_{k-1} + v_{b,k} \end{aligned} \tag{1}$$

where $v_{G,k} \sim \mathcal{N}(0, Q_{vG})$ and $v_{b,k} \sim \mathcal{N}(0, Q_{vb})$. The accelerometer signal $y_{A,k} \in \mathbb{R}^3$ is composed of the gravity vector in the sensor frame ${}^S g_{A,k}$, the external acceleration ${}^S a_k$ in the sensor frame, and an additional noise $v_{A,k}$. The external acceleration in the global frame ${}^G a_k$ is modeled as a decayed Markov process with coefficient η_a and the noise $v_{a,k}$. Thus, we obtain

$$\begin{aligned} y_{A,k} &= -{}^S g_{A,k} + {}^S a_k + v_{A,k} \\ {}^G a_k &= \eta_a({}^G a_{k-1}) + v_{a,k} \end{aligned} \tag{2}$$

where $0 < \eta_a < 1$, $v_{A,k} \sim \mathcal{N}(0, Q_{vA})$, $v_{a,k} \sim \mathcal{N}(0, Q_{va})$. The free acceleration in the global frame and the sensor frame can be transferred by a rotation matrix ${}^G_S R$ with ${}^G a_k = {}^G_S R({}^S a_k)$.

Finally, the magnetometer signals $y_{M,k} \in \mathbb{R}^3$ is a sum of the magnetic vector ${}^S m_k$, the magnetic disturbance ${}^S d_k$ and the noise $v_{M,k}$. The magnetic disturbance in the global frame ${}^G d_k$ is modeled as a Markov process with a coefficient η_m and an additional noise $v_{d,k}$. Hence, one has

$$\begin{aligned} y_{M,k} &= {}^S m_k + {}^S d_k + v_{M,k} \\ {}^G d_k &= \eta_m({}^G d_{k-1}) + v_{d,k} \end{aligned} \tag{3}$$

where $0 < \eta_m < 1$, $v_{M,k} \sim \mathcal{N}(0, Q_{vM})$, $v_{d,k} \sim \mathcal{N}(0, Q_{vd})$. Similarly, the disturbance in the sensor frame and the global frame can be transferred by ${}^G d_k = {}^G_S R({}^S d_k)$. One can refer to [8], [21] for more information.

B. Multi-kernel Maximum Correntropy Kalman filter

We consider the linear model with

$$\begin{aligned} \mathrm{x}_{k+1} &= \mathrm{A}\mathrm{x}_k + \mathrm{p}_k \\ \mathrm{y}_k &= \mathrm{C}\mathrm{x}_k + \mathrm{q}_k \end{aligned} \tag{4}$$

where $\mathrm{x}_k \in \mathbb{R}^n$, $\mathrm{y}_k \in \mathbb{R}^m$, A and C is state transition matrix and observation matrix, respectively. The noise $\mathrm{p}_k \sim \mathcal{N}(0, \mathrm{Q})$ and $\mathrm{q}_k \sim \mathcal{N}(0, \mathrm{R})$. Equation (4) can be rewritten as

$$\begin{bmatrix} \mathrm{x}_k^- \\ \mathrm{y}_k \end{bmatrix} = \begin{bmatrix} \mathrm{I} \\ \mathrm{C} \end{bmatrix} \mathrm{x}_k + \zeta_k \tag{5}$$

where $\zeta_k = \begin{bmatrix} \mathrm{x}_k^- - \mathrm{x}_k \\ \mathrm{q}_k \end{bmatrix}$ and x_k^- denotes the *a priori* estimate of x_k. The covariance of ζ_k has

$$E(\zeta_k \zeta_k^T) = \begin{bmatrix} \mathrm{P}_k^- & 0 \\ 0 & \mathrm{R} \end{bmatrix} = \begin{bmatrix} \mathrm{T}_p \mathrm{T}_p^T & 0 \\ 0 & \mathrm{T}_r \mathrm{T}_r^T \end{bmatrix} = \mathrm{T}_k \mathrm{T}_k^T$$

where P_k^- is the *a priori* error covariance at time step k, R is the measurement covariance, and T_p, T_r can be obtained by Cholesky decomposition. By left multiplying T_k^{-1} in both sides of (5), we obtain $\mathrm{D}_k = \mathrm{W}_k \mathrm{x}_k + \nu_k$ where

$$\mathrm{D}_k = \mathrm{T}_k^{-1} \begin{bmatrix} \mathrm{x}_k^- \\ \mathrm{y}_k \end{bmatrix}, \mathrm{W}_k = \mathrm{T}_k^{-1} \begin{bmatrix} \mathrm{I} \\ \mathrm{C} \end{bmatrix}, \nu_k = \mathrm{T}_k^{-1} \zeta_k$$

The noise term ν_k is white since $\mathrm{E}(\nu_k \nu_k^T) = \mathrm{E}[(\mathrm{T}_k^{-1}\zeta_k)(\mathrm{T}_k^{-1}\zeta_k)^T] = \mathrm{I}$. The objective of MKMCKF is to solve the following problem:

$$\mathrm{x}_k^+ = \arg\max_{\mathrm{x}_k \in \mathscr{X}} J(\mathrm{x}_k) = \arg\max_{\mathrm{x}_k \in \mathscr{X}} \sum_{i=1}^{n+m} 2\sigma_i^2 G_{\sigma_i}(\mathrm{e}_{i,k}). \tag{6}$$

with $G_{\sigma_i}(\mathrm{e}_{i,k}) = \exp\left(-\frac{\mathrm{e}_{i,k}^2}{2\sigma_i^2}\right)$ where $\mathscr{X}$ is the domain of x_k, $\mathrm{e}_{i,k} = \mathrm{d}_{i,k} - \mathrm{w}_{i,k}\mathrm{x}_k$ denotes the error at time step k in the i-th channel, $\mathrm{d}_{i,k}$ is the i-th element of D_k, and $\mathrm{w}_{i,k}$ is the i-th row of W_k. The details of this algorithm can be found in our previous work [18].

C. Bayesian Optimization

We use the BO to explore the optimal kernel bandwidths for the MKMCKF. For a general function $f : \mathscr{Z} \subset \mathbb{R}^l \to \mathbb{R}$ with $x \in \mathscr{Z}$, we try to find the minima point x^* on a domain $\mathscr{Z}$, i.e.,

$$x^* = \arg\min_{x \in \mathscr{Z}} f(x). \tag{7}$$

In the framework of BO, the distribution f is assumed to be Gaussian with $p(f) = \mathcal{GP}(f; \mu, K)$ where μ is the mean and K is the covariance. With N finite samples $\boldsymbol{x}_{1:N}$, the Gaussian Process (GP) model is a joint Gaussian $\mathbf{f}(\boldsymbol{x}_{1:N}) \sim \mathcal{N}(\mu(\boldsymbol{x}_{1:N}), \mathbf{K})$ where $\mathbf{K}_{i,j} = \kappa(x_i, x_j)$ and $\kappa(\cdot,\cdot)$ is the covariance function. For a new data x_{N+1}, by employing the standard conditioning rules for Gaussian random variables, we have

$$f\left(x_{N+1}\right) \mid \mathcal{D}_N, x_{N+1} \sim \mathcal{N}\left(\hat{\mu}(x_{N+1}), \hat{\Sigma}^2(x_{N+1})\right) \tag{8}$$

with

$$\hat{\mu}(x_{N+1}) = \mu\left(x_{N+1}\right) + \mathbf{k}^T\mathbf{K}^{-1}\left(\mathbf{f}\left(\boldsymbol{x}_{1:N}\right) - \mu\left(\boldsymbol{x}_{1:N}\right)\right)$$
$$\hat{\Sigma}^2(x_{N+1}) = \kappa\left(\mathbf{x}_{N+1}, \mathbf{x}_{N+1}\right) - \mathbf{k}^T\mathbf{K}^{-1}\mathbf{k}.$$

To use a GP, one has to specify the mean function and the covariance function. The mean function is usually set to be zero and the covariance function can be set as Gaussian kernel with $\kappa(x, x') = \exp(\frac{1}{2}(x - x')^T M^{-1}(x - x'))$. The kernel parameters M can be obtained by empirical Bayesian methods [23]; see [24], [25] for more information.

Expected improvement (EI) is a common acquisition function to determine which point should be evaluated at the next iteration. Suppose that f^* is the minimum value of f observed so far. Essentially, EI tries to find a candidate point x that improves upon f^* the most with a utility function $u(x) = \max_{x \in \mathscr{Z}}(0, f^* - f(x))$. The acquisition function can be written as

$$a_{EI}(x) = \hat{\Sigma}(x)\left(\frac{f^* - \hat{\mu}(x)}{\hat{\Sigma}(x)}\Phi(\frac{f^* - \hat{\mu}(x)}{\hat{\Sigma}(x)}) + \phi(\frac{f^* - \hat{\mu}(x)}{\hat{\Sigma}(x)})\right)$$

where Φ is the cumulative distribution function and ϕ is the standard normal probability density function [19], [20]. The next candidate point x_{nc} can be selected as

$$x_{nc} = \arg\max_{x \in \mathscr{Z}} a_{EI}(x). \tag{9}$$

In the applications with the inequity constraint $c(x) \leq \lambda$, we adopt the method in [19] which writes the constrained utility function as $u_c(x) = \Delta(x)\max_{x \in \mathscr{Z}}(0, f^* - f(x)) = \Delta(x)u(x)$ where $\Delta(x) = \{0, 1\}$ is a feasibility indicator and $\Delta(x) = 1$ if $c(x) \leq \lambda$, and $\Delta(x) = 0$ otherwise. Then, the constrained acquisition function can be written as

$$a_{CEI}(x) = Pr(c(x) \leq \lambda) \cdot a_{EI}(x) \tag{10}$$

with $Pr(c(x) \leq \lambda) = \int_{-\infty}^{\lambda} p(c(x) \mid x, \mathcal{D}_N)\mathrm{d}c(x)$. The next candidate point can be obtained by

$$x_{nc} = \arg\max_{\sigma \in \mathscr{Z}} a_{CEI}(x). \tag{11}$$

III. MAIN RESULTS

In this section, we derive a novel MKMCKF-OE algorithm using the maximum correntropy criterion. Then, we tune the hyper-parameters (i.e., kernel bandwidths) with BO to minimize the orientation error.

A. MKMCKF-OE

We use the error state rather than the nominal state to construct the *error model*:

$$x_{\varepsilon,k+1} = \Phi_k x_{\varepsilon,k} + w_k \tag{12}$$

with $x_{\varepsilon,k} = [(\theta_{\varepsilon,k})^T, (b_{\varepsilon,k})^T, ({}^S a_{\varepsilon,k})^T, ({}^S d_{\varepsilon,k})^T]^T$ and $w_k = [w_{\theta,k}^T, w_{b,k}^T, w_{a,k}^T, w_{d,k}^T]$ where the subscript ε denotes the difference of the estimated state and the nominal state, the error state $x_{\varepsilon,k}$ includes the error orientation $\theta_{\varepsilon,k} \in \mathbb{R}^3$ (Euler angle), the error gyroscope offset $b_{\varepsilon,k} \in \mathbb{R}^3$, the error external acceleration in the sensor frame ${}^S a_{\varepsilon,k} \in \mathbb{R}^3$, and the error magnetic disturbance in the sensor frame ${}^S d_{\varepsilon,k} \in \mathbb{R}^3$. The noise w_k is assumed to be Gaussian with covariance Q. Notably, the error state transition matrix $\Phi_k \in \mathbb{R}^{12\times12}$ is a zero matrix since the *a priori* estimate of $x_{\varepsilon,k}^-$ does not depend on the previous *a posteriori* estimate of $x_{\varepsilon,k-1}^+$ (the knowledge about previous errors has been incorporated in the current nominal state.). A more detailed description can be found in [8], [21].

The measurement signals are related with the error state through

$$\begin{aligned} z_{\varepsilon,k} &= \begin{bmatrix} {}^S\hat{g}_{A,k}^- - {}^S\hat{g}_{G,k}^- \\ {}^S\hat{m}_{M,k}^- - {}^S\hat{m}_{G,k}^- \end{bmatrix} + v_k \\ &= H_k x_{\varepsilon,k} + v_k \end{aligned} \tag{13}$$

with

$$\begin{aligned} H_k &= \begin{bmatrix} -{}^S\widehat{g}_{G,k}^-\times, \delta t({}^S\widehat{g}_{G,k}^-\times), I_{3\times3}, 0_{3\times3} \\ -{}^S\widehat{m}_{G,k}^-\times, \delta t({}^S\widehat{m}_{G,k}^-\times), 0_{3\times3}, -I_{3\times3} \end{bmatrix} \\ {}^S g_{A,k}^- &= -y_{A,k} + {}^S\hat{a}_k^-,\ {}^S g_{G,k}^- = \hat{R}_k^-\left({}^G g_k\right) \\ {}^S m_{M,k}^- &= y_{M,k} - {}^S\hat{d}_k^-,\ {}^S m_{G,k}^- = \hat{R}_k^-\left({}^G m_k\right) \end{aligned} \tag{14}$$

where ${}^S\hat{g}_{A,k}^-$ and ${}^S\hat{g}_{G,k}^-$ denotes the *a priori* estimate of gravity vector obtained by the accelerometer model and by the coordinate transformation, ${}^S\hat{m}_{M,k}^-$ and ${}^S m_{G,k}^-$ denotes the *a priori* estimate of geomagnetic vector obtained by the magnetometer model and by the coordinate transformation, ${}^S\hat{a}_k^-$ and ${}^S\hat{d}_k^-$ is the *a priori* estimate of external acceleration and the *a priori* estimate of magnetic disturbance, ${}^G g_k$ and ${}^G m_k$ is the constant gravity vector and the geomagnetic vector, $\hat{R}_k^-$ is the *a priori* estimate of orientation matrix. The measurement noise is v_k with covariance Z_k, δt is the sampling rate, and P_k is the corresponding error covariance matrix. One can refer to [8], [21] for more information.

The nominal state can be written as $x_k = [\theta_k^T, b_k^T, ({}^S a_k)^T, ({}^S d_k)^T]^T$ which includes the orientation (Euler angle), the gyroscope offset, the external acceleration, and the magnetic disturbance. The orientation $\theta_k \in \mathbb{R}^3$ can be written as a rotation matrix $R_k \in \mathbb{R}^{3\times3}$ or a quaternion $q_k \in \mathbb{R}^4$. Obviously, the modeling of the external acceleration and the magnetic disturbance in (2) and (3) is not accurate. The unmodeled part can be seen as

Algorithm 1 MKMCKF-OE

Initialization:
Choose kernel bandwidth vector $\sigma_p \in \mathbb{R}^{12}$, $\sigma_r \in \mathbb{R}^6$.
State Prediction:
$\hat{b}_k^- = \hat{b}_{k-1}^+$, ${}^S\hat{a}_k^- = \eta_a({}^S\hat{a}_{k-1}^+)$, ${}^S\hat{d}_k^- = \eta_m({}^S\hat{d}_{k-1}^+)$
$\hat{\omega}_k^- = y_{G,k} - \hat{b}_k^-$, $\hat{q}_k^- = \hat{q}_{k-1}^+ \Delta q(\hat{\omega}_k^- \delta t)$
Error State Propagation:
$P_k^- = f(P_{k-1}^+, Q)$, details in [8], [21]
Obtain T_p and T_r by $T_p T_p^T = P_k^-$ and $T_r T_r^T = Z_k$
Initialize $x_{\varepsilon,k}^- = x_{\varepsilon,k,0}^+ = 0$
while $\frac{x_{\varepsilon,k,t}^+ - x_{\varepsilon,k,t-1}^+}{x_{\varepsilon,k,t-1}^+} \leq \varepsilon$ or $t = 1$ **do** ▷ t starts from 1
$x_{\varepsilon,k,t}^+ = x_{\varepsilon,k}^- + \tilde{K}_k z_{\varepsilon,k}$ ▷ $z_{\varepsilon,k}$ obtained by (13)
$\tilde{K}_k = \tilde{P}_k^- H_k^T S_k^{-1}$ ▷ H_k is shown in equation (14)
$S_k = H_k \tilde{P}_k^- H_k^T + \tilde{Z}_k$
$\tilde{P}_k^- = T_p \tilde{M}_p^{-1} T_p^T$, $\tilde{Z}_k^- = T_r \tilde{M}_r^{-1} T_r^T$
$\tilde{M}_p = diag(G_{\sigma_p}(e_p))$, $\tilde{M}_r = diag(G_{\sigma_r}(e_r))$
$e_p = T_p^{-1} x_{\varepsilon,k}^- - T_p^{-1} x_{\varepsilon,k,t-1}^+$
$e_r = T_r^{-1} \left[z_{\varepsilon,k} - H_k x_{\varepsilon,k,t-1}^+ \right]$
end while
State Update:
Obtain $\hat{\theta}_{\varepsilon,k}^+$, ${}^S\hat{a}_{\varepsilon,k}^+$ $\hat{b}_{\varepsilon,k}^+$, ${}^S\hat{d}_{\varepsilon,k}^+$ from $x_{\varepsilon,k,t}^+$
$\hat{q}_k^+ = \hat{q}_k^- \Delta q(-\hat{\theta}_{\varepsilon,k}^+)$, $\hat{b}_k^+ = \hat{b}_k^- - \hat{b}_{\varepsilon,k}^+$, ${}^S\hat{a}_k^+ = {}^S\hat{a}_k^- - {}^S\hat{a}_{\varepsilon,k}^+$
${}^S\hat{d}_k^+ = {}^S\hat{d}_k^- - {}^S\hat{d}_{\varepsilon,k}^+$ ▷ if no magnetic jamming
update ${}^G m_k$ based on [21, p.19]
$P_k^+ = (I - \tilde{K}_k H_k) P_k^- (I - \tilde{K}_k H_k)^T + \tilde{K}_k Z_k \tilde{K}_k^T$

non-Gaussian noise which deteriorates the estimate accuracy significantly. To handle this issue, we employ MKMCKF for orientation estimation and construct a novel algorithm MKMCKF-OE. The difference between our algorithm and the ESKF [8], [21] is that we employ Gaussian kernels to increase its robustness to disturbance. The detailed algorithm is summarized in Algorithm 1. Note that the construction of P_k^- and the influence of the magnetic jamming are ignored here, one can refer to [8], [21] for details.

B. Kernel Bandwidths Optimization

In Algorithm 1, we have to choose kernel bandwidth vectors σ_p and σ_r. Based on Theorem 1 in [18], the multi-kernel maximum correntropy Kalman filter is identical to the traditional KF if we set all kernel bandwidths $\sigma_i \to \infty$. Since KF is optimal for a linear system with Gaussian noises, we can use a very large kernel bandwidth, i.e., σ_{inf}, for Gaussian channels. As for non-Gaussian channels, we tune them by BO. In IMUs, the non-Gaussian noises exit in the external acceleration and magnetic disturbance channels due to the first-order dynamic assumption in equations (2) and (3), while noises in other channels can be regarded as Gaussian. Therefore, we have

$$\begin{aligned} \sigma_p &= (\sigma_{\text{inf}} I_{1\times 3}, \sigma_{\text{inf}} I_{1\times 3}, \sigma_1 I_{1\times 3}, \sigma_2 I_{1\times 3})^T \\ \sigma_r &= (\sigma_{\text{inf}} I_{1\times 3}, \sigma_{\text{inf}} I_{1\times 3})^T \end{aligned} \tag{15}$$

where $I_{1\times 3}$ is the identity vector, σ_{inf} is a big constant ($\sigma_{\text{inf}} = 10^8$ in experiments), and σ_1 and σ_2 are two parameters to be optimized. The objective function of BO is to minimize the RMSE on all axes with a constraint on the yaw axis:

$$\begin{aligned} \underset{\sigma_1,\sigma_2 \in \mathscr{Z}}{\arg\min}\, J_{BO} = \underset{\sigma_1,\sigma_2 \in \mathscr{Z}}{\arg\min} \frac{1}{3m} \sum_{i=1}^{3} \sum_{j=1}^{m} \underbrace{\frac{1}{n} \sqrt{\sum_{k=1}^{n} (\hat{\theta}_{i,j,k}(\sigma_1,\sigma_2) - \theta_{i,j,k})^2}}_{RMSE} \\ s.t.\ \frac{1}{m} \sum_{j=1}^{m} \frac{1}{n} \sqrt{\sum_{k=1}^{n} (\hat{\theta}_{3,j,k}(\sigma_1,\sigma_2) - \theta_{3,j,k})^2} \leq \lambda \end{aligned} \tag{16}$$

where m is the number of training sets, $\mathscr{Z}$ is the domain of kernel bandwidths, $\hat{\theta}_{i,j,k}(\sigma_1,\sigma_2)$ represents the estimated angle under i axis (roll, pitch and yaw) at j-th experiment and time step k, $\theta_{i,j,k}$ represents the corresponding ground truth angle (Euler angle). The constant λ gives an upper RMSE bound on the yaw axis ($i = 3$). In total, we conduct m experiments to involve different situations, i.e., with or without acceleration disturbance, with or without magnetic disturbance, and different types of movement (linear motion or rotation).

IV. EXPERIMENTS

In this section, we first obtain the optimal kernel bandwidths through BO. Then, we validate the performance of the MKMCKF-OE in the testing set and compare it with the GD [5], ESKF [8], [21], IGD [9], and VQF [22]. The implementations of the GD, ESKF, and MKMCKF-OE are available at [26].

A. Data Collection

We use Xsens MTI-670 to collect the raw data. The sampling frequency for accelerometers and gyroscopes is $400\ Hz$ while the value for magnetometers is $100\ Hz$. The MKMCKF-OE is executed at $100\ Hz$. The experimental setup is shown in Fig. 1(a) and Fig. 1(b) which corresponds to the linear motion and the rotation, respectively. In Fig. 1(a), the acceleration disturbance A_d is generated by moving the Xsens along the rail manually, while the magnetic disturbance M_d is generated by the approaching of the ferromagnetic pliers. In Fig. 1(b), the Xsens is fixed to the shank of an exoskeleton robot where its leg swings to imitate the gait of a human. The gait frequency is denoted by f and the acceleration disturbance grows with f. The magnetic disturbance M_d is generated by approaching the pliers to Xsens.

The ground truth orientation for the linear motion is set to be the initial mean angle when the Xsens is static without disturbance. The ground truth pitch angle for rotation is obtained by encoders of the exoskeleton robot, while the roll angle and the yaw angle are obtained by the initial mean angles of the corresponding axes when the Xsens is static since there is no rotation at those axes. A total of 12 experiments are conducted and their descriptions are shown in Table I.

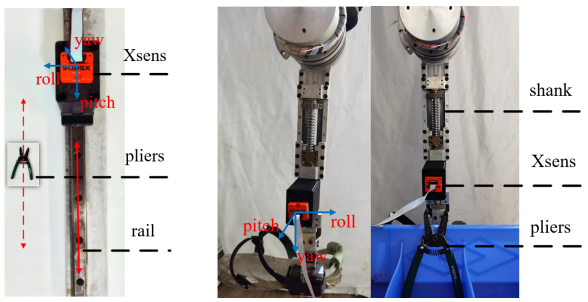
B. Kernel Bandwidths Optimization

The first 5 experiments are utilized as the training set which are shown in Table I. Among them, experiment 5

TABLE II

RMSE OF DIFFERENT ALGORITHMS ON THE ROLL AND PITCH.

Index	Roll (deg)					Pitch (deg)				
	MKMC	GD	ESKF	IGD	VQF	MKMC	GD	ESKF	IGD	VQF
6	0.3683	1.7992	1.4010	0.7596	0.8669	0.4257	0.4549	0.9942	0.2891	0.5410
7	0.1381	0.4985	0.5447	0.1104	0.0996	0.2833	1.3211	0.8961	0.5396	0.4900
8	0.2199	0.4077	0.5463	0.1867	0.2204	0.3811	1.4246	0.9958	0.6671	0.3849
9	0.1827	0.5400	0.4274	0.4394	0.4002	0.4629	1.7052	1.3178	1.4895	0.5611
10	0.1133	1.5521	2.1813	0.0744	0.1039	0.4228	1.9076	3.3010	0.4441	0.2543
11	0.2093	1.2445	1.9321	0.1305	0.1467	0.3022	1.6302	2.9933	0.4670	0.3114
12	0.1759	1.2293	2.0358	0.1809	0.2110	0.2420	1.7925	3.2558	0.7003	0.3308



(a) (b)

Fig. 1. The experimental setup and the IMU coordinate system. The movement can be divided into linear motion in Fig 1(a) and rotation in Fig 1(b). The acceleration disturbance A_d is generated by the movement of Xsens itself and the magnetic disturbance M_d is generated by the pliers.

TABLE I

EXPERIMENTAL DESCRIPTION.

Index	Mode	Description	Usage
1	linear motion	A_d,M_d	training
2	rotation	$f = 0.1Hz$	training
3	rotation	$f = 0.5Hz$	training
4	rotation	$f = 0.1Hz$, M_d	training
5	rotation	$f = 0.5Hz$, M_d	training
6	linear motion	A_d,M_d	testing
7	rotation	$f = 0.2Hz$	testing
8	rotation	$f = 0.3Hz$	testing
9	rotation	$f = 0.4Hz$	testing
10	rotation	$f = 0.2Hz$, M_d	testing
11	rotation	$f = 0.3Hz$, M_d	testing
12	rotation	$f = 0.4Hz$, M_d	testing

has the fastest dynamics with both obvious acceleration disturbance and magnetic disturbance. Since the magnetic disturbance affects the yaw angle the most, our objective is to minimize the overall RMSE with a constraint of the yaw error, i.e., $m = 5$ and $\lambda = 2.0\,\text{deg}$ in (16). The kernel bandwidth domain is set to be $\mathscr{Z} = [0.1, 10]$, which is capable of the majority applications. The initial candidate points are randomly selected and the initial points number is 2. The maximum iteration number for BO is 100. The optimization procedure is conducted on MATLAB 2019b based on equations (7)-(11). The optimal kernel bandwidths are $\sigma_1 = 1.6188$ and $\sigma_2 = 0.4234$.

C. Results

We validate the performance of MKMCKF-OE on experiments 6-12 and compared it with the GD [5], ESKF [8], IGD [9], and VQF [22]. The results are summarized in Tables II and III (we use MKMC to denote the MKMCKF-OE). One can see that the overall performance of the MKMCKF-OE is better than the others, especially on the yaw axis with obvious magnetic disturbance in experiments 6, 10, 11, and 12.

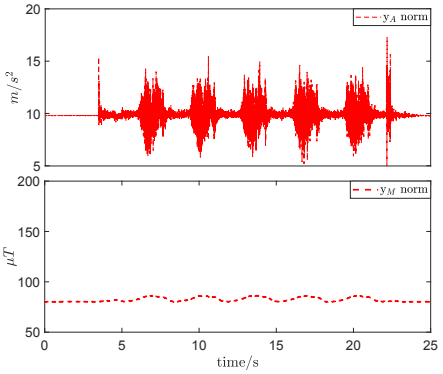


Fig. 2. The norms of sensor readings in experiment 8.

TABLE III

RMSE OF DIFFERENT ALGORITHMS ON THE YAW.

Index	Yaw (deg)				
	MKMC	GD	ESKF	IGD	VQF
6	0.3581	4.4078	7.8615	5.1642	2.9495
7	0.2726	1.9509	1.3227	1.3768	0.8566
8	0.2604	1.6842	1.0041	1.2250	0.5759
9	0.4098	1.7093	1.0888	1.0490	0.3104
10	1.4697	19.7168	20.5097	17.1180	11.7320
11	1.4911	20.8336	21.4068	18.0540	12.9183
12	1.1269	24.4531	24.9301	22.2111	17.2807

We visualize the the norm of accelerometer and magnetometer readings in experiment 8 in Fig. 2 and depict the corresponding orientation errors in Fig. 3. One can see that the overall performance of MKMCKF-OE is superior to the GD, ESKF, IGD, and VQF, especially along the yaw and pitch axes.

We also visualize the orientation error of different approaches in experiment 11 in Fig. 4. One can see that a very big yaw error is caused by the magnetic disturbance in the blue region by the GD, ESKF, IGD, and VQF, while it is avoided by the MKMCKF-OE. In the yellow region where the IMU rotates with both M_d and A_d, the performance of the MKMCKF-OE is again significantly better than the

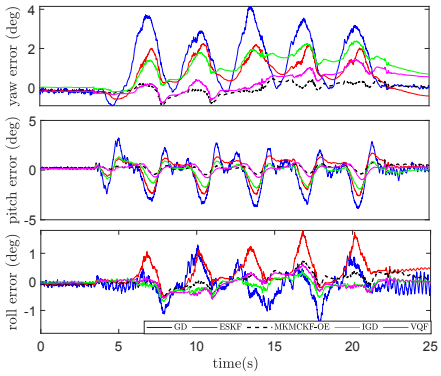


Fig. 3. Estimation errors of different algorithms in experiment 8.

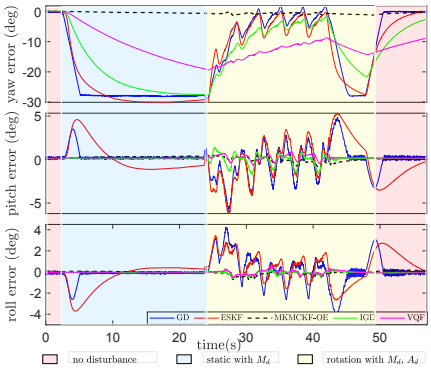


Fig. 4. Estimation errors of the GD, ESKF, MKMCKF-OE, IGD, and VQF in experiment 11.

others, which reveals that our algorithm is very robust to both acceleration disturbance and magnetic disturbance. The IGD and VQF achieve good performance for the inclination estimation (the roll and pitch). However, their performances on the yaw are not satisfactory when there is magnetic disturbance.

V. CONCLUSION

In this paper, we design a new algorithm MKMCKF-OE for orientation estimation which is robust to acceleration disturbance and magnetic disturbance. The kernel bandwidths are obtained by BO which is both efficient and convenient. The effectiveness of the proposed method is validated by a set of experiments and compared with the benchmark methods. Experiments show that our proposed method outperforms the GD and the ESKF significantly, and is superior to the IGD and VQF when there is magnetic disturbance. The RMSE of MKMCKF-OE is less than 1.4911° in the yaw and is smaller than 0.4629° in the roll and pitch even under the worst testing case.

REFERENCES

[1] W. Huo, S. Mohammed, Y. Amirat, and K. Kong, "Fast gait mode detection and assistive torque control of an exoskeletal robotic orthosis for walking assistance," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1035–1052, 2018.

[2] M. Hao, K. Chen, and C. Fu, "Smoother-based 3-D foot trajectory estimation using inertial sensors," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 12, pp. 3534–3542, 2019.

[3] 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.

[4] R. Mahony, T. Hamel, and J.-M. Pflimlin, "Nonlinear complementary filters on the special orthogonal group," *IEEE Transactions on Automatic Control*, vol. 53, no. 5, pp. 1203–1218, 2008.

[5] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *IEEE International Conference on Rehabilitation Robotics*, 2011, pp. 1–7.

[6] A. El Hadri and A. Benallegue, "Sliding mode observer to estimate both the attitude and the gyro-bias by using low-cost sensors," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009, pp. 2867–2872.

[7] L. Xiang and L. Qun, "External acceleration elimination for complementary attitude filter," in *IEEE International Conference on Information and Automation (ICIA)*, 2017, pp. 208–212.

[8] 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.

[9] T. Seel and S. Ruppin, "Eliminating the effect of magnetic disturbances on the inclination estimates of inertial sensors," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 8798–8803, 2017.

[10] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: properties and applications in non-Gaussian signal processing," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286–5298, 2007.

[11] L. Bako, "Robustness analysis of a maximum correntropy framework for linear regression," *Automatica*, vol. 87, pp. 218–225, 2018.

[12] 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.

[13] B. Chen, X. Liu, H. Zhao, and J. C. Principe, "Maximum correntropy Kalman filter," *Automatica*, vol. 76, pp. 70–77, 2017.

[14] X. Liu, Z. Ren, H. Lyu, Z. Jiang, P. Ren, and B. Chen, "Linear and nonlinear regression-based maximum correntropy extended Kalman filtering," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 5, pp. 3093–3102, 2021.

[15] X. Liu, H. Qu, J. Zhao, and P. Yue, "Maximum correntropy square-root cubature Kalman filter with application to SINS/GPS integrated systems," *ISA Transactions*, vol. 80, pp. 195–202, 2018.

[16] A. Singh and J. C. Principe, "Using correntropy as a cost function in linear adaptive filters," in *2009 International Joint Conference on Neural Networks*, 2009, pp. 2950–2955.

[17] 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.

[18] 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.

[19] J. R. Gardner, M. J. Kusner, Z. E. Xu, K. Q. Weinberger, and J. P. Cunningham, "Bayesian optimization with inequality constraints." in *ICML*, vol. 2014, 2014, pp. 937–945.

[20] E. Brochu, V. M. Cora, and N. de Freitas, "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," Tech. Rep. UBC TR-2009-023 and arXiv:1012.2599, 2009.

[21] 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.

[22] D. Laidig and T. Seel, "Vqf: Highly accurate imu orientation estimation with bias estimation and magnetic disturbance rejection," *Information Fusion*, vol. 91, pp. 187–204, 2023.

[23] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT press, 2012.

[24] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2005.

[25] P. I. Frazier, "A tutorial on Bayesian optimization," *arXiv preprint arXiv:1807.02811*, 2018.

[26] S. Li, "Code and data for MKMCKF-OE," 2022. [Online]. Available: https://github.com/lsl-zsj/MKMCKF-OE