






A Nonlinear Maximum Correntropy Information Filter for High-Dimensional Neural Decoding

Xi Liu ^{1,†} , Shuhang Chen ^{2,†} , Xiang Shen ¹ , Xiang Zhang ¹  and Yiwen Wang ^{1,2,*} 

¹ Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, China; lxi1102@163.com (X.L.); xshenai@connect.ust.hk (X.S.); xzhangaz@ust.hk (X.Z.)

² Department of Chemical and Biological Engineering, Hong Kong University of Science and Technology, Hong Kong, China; schenbx@connect.ust.hk

* Correspondence: eewangyw@ust.hk

† Xi Liu is first author and Shuhang Chen serves as co-first author.

Abstract: Neural signal decoding is a critical technology in brain machine interface (BMI) to interpret movement intention from multi-neural activity collected from paralyzed patients. As a commonly-used decoding algorithm, the Kalman filter is often applied to derive the movement states from high-dimensional neural firing observation. However, its performance is limited and less effective for noisy nonlinear neural systems with high-dimensional measurements. In this paper, we propose a nonlinear maximum correntropy information filter, aiming at better state estimation in the filtering process for a noisy high-dimensional measurement system. We reconstruct the measurement model between the high-dimensional measurements and low-dimensional states using the neural network, and derive the state estimation using the correntropy criterion to cope with the non-Gaussian noise and eliminate large initial uncertainty. Moreover, analyses of convergence and robustness are given. The effectiveness of the proposed algorithm is evaluated by applying it on multiple segments of neural spiking data from two rats to interpret the movement states when the subjects perform a two-lever discrimination task. Our results demonstrate better and more robust state estimation performance when compared with other filters.

Keywords: state-observation model; high-dimensional measurements systems; correntropy



Citation: Liu, X.; Chen, S.; Shen, X.; Zhang, X.; Wang, Y. A Nonlinear Maximum Correntropy Information Filter for High-Dimensional Neural Decoding. *Entropy* **2021**, *23*, 743. <https://doi.org/10.3390/e23060743>

Academic Editor: Carlos M. Travieso-González

Received: 30 April 2021
Accepted: 8 June 2021
Published: 12 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Brain machine interface (BMI) establishes a direct communication pathway between the brain and external devices [1–6]. BMI collects the noisy signals from hundreds of neurons in the brain, and estimates a motor intention from these signals [7,8]. This estimated movement intention can then be used to control the robot to assist motor disabled people [9–19]. Signal processing algorithms play a key role in BMI. As a commonly-used state-observation model, the Kalman filter (KF) has been adopted to decode the movement intents as the state from the high-dimensional observations formed by multiple neural firing activity [20–23], in which the movement state evolves over time as described by the linear state model and the observation model reflects how the neuron firing tunes to movement with Gaussian noise. The implementation of the Kalman filter nicely considers the gradual change of the continuous brain state, and thus is especially appropriate for the brain control task where the subject continuously adjusts the brain states to control an external robot.

However, there are some challenges in applying the state-observation model for BMI. Firstly, the nervous systems are nonlinear in general [24,25], and the traditional KF is not fit for nonlinear systems. As the extensions of KF, the extended Kalman filter (EKF) [26], unscented Kalman filter (UKF) [27,28], and cubature Kalman filter (CKF) [29] can deal with a nonlinear system. However, these algorithms can only approximate the

nonlinear function by a finite-order level of its Taylor expansion and lead to an inaccurate state estimation. Secondly, the motor state is controlled by hundreds of neurons, whose activity could be highly connected [30–32]. The components of measurements in traditional KF decoding are assumed to be conditionally independent to update the states, and some connections among these components, e.g., the neural connectivity, may exist in the experiments of BMI that cannot be reflected in the traditional KF. Discriminative Kalman filter (DKF) was applied to decode movement states from high-dimensional neural firings [19]. DKF computes the posterior of the state-given observation by approximating only the mean and variance, which may not be accurate for an arbitrary nonlinear system. The key goal in BMI is to accurately estimate the motor state in real-time through measuring the noisy activities of many neurons that tune nonlinearly and correlatedly to the states. Thirdly, the traditional KF may not be an ideal tool for real-time robust decoding to address the high-dimensional measurement data. This is because the Kalman gain computes the inverse of a correlation matrix with the same dimension of measurements, and the high dimensionality brings a heavy computational burden for online applications. In addition, KF relies strongly on good initial conditions and is not robust to the outliers in the observation. It brings problems to robust online decoding. In particular, the initial movement states in BMI cannot always be known, for example, the limb status of paralyzed patients may not exist. It could be difficult to set the proper initial conditions as well as the uncertainty reflected in the covariance matrix. For online decoding, the performance of the Kalman filter could be significantly affected by outliers in the observation, such as recording noise in neural activity.

We are interested in addressing the nonlinearity and existing connectivity among high-dimensional observations for robust neural decoding. The neural network (NN) has demonstrated an arbitrary nonlinear regression ability in modeling [33], and has been used to decode the motor intention from multiple neuron activities [34–36]. The advantage of using NN is to allow arbitrary dependence among the neurons and provide a good method to reduce the dimensionality from high-dimensional neurons to low-dimensional states. However, a simple multi-layer perception (MLP) network could ignore the time dynamics in brain states. The recurrent neural network (RNN) model considers the possible internal dynamics and has been used to decode the letter shapes for classification [37]. The long short-term memory (LSTM) network has been used in the classification of the motor imagery task based on electroencephalography (EEG) signals [38]. For the brain control in motor BMI, where we adopt invasive signals to decode continuous motor intention, the Kalman filter has been a more popular decoder [21,23,39,40]. The information filter [41] was proposed to address the sensitivity to the initial conditions and reduce the computational complexity, in which the inverse of the covariance matrix defining the uncertainty is propagated instead. However, the current information filter is also only designed for simple linear systems. To address the robustness of the filtering with respect to the outliers, i.e., heavy-tailed non-Gaussian noises [42], the traditional minimum mean square error (MMSE) criterion should be reconsidered. As an alternative, the correntropy theory in information theoretic learning (ITL) is an effective tool, which captures all even-order statistical information rather than the traditional second-order statistics [43–45]. In particular, its optimization criterion, called maximum correntropy criterion (MCC), has been extensively used in robust filters [46–52], mostly for linear systems.

In this paper, we propose a nonlinear maximum correntropy information filter (NM-CIF) to robustly estimate the state from noisy high-dimensional measurements for the nonlinear systems. We utilize the universal nonlinear approximator neural network to preprocess the high-dimensional measurements, which allows dependence among the measurements to be considered. We introduce correntropy as the optimization criterion to improve the robustness of the filter with respect to heavy-tailed non-Gaussian noises. The uncertainty of the state estimation is derived and propagated over time to reduce the computational burden and the sensitivity to the initial conditions. The preferred information matrix and the proof of the robustness of the algorithm are given, and the convergence

speed is analyzed theoretically. The proposed algorithm is applied to real neural recording from two rats while the subjects were manually controlling a two-lever discrimination task. The two-dimensional movement states are estimated from multiple neuronal activities. The performance, convergence time, the sensitivity to the bad initial conditions, and robustness to heavy-tailed noise of the proposed algorithm are compared with those of the traditional Kalman filter and neural network decoder. We also use the mean square error in 2D to evaluate the movement estimation compared with the ground truth.

Our contribution is to propose a new methodology to better estimate the state in the filtering process for the noisy high-dimensional measurement systems. The neural network is utilized to construct the nonlinear measurement model between the high-dimensional measurements and low-dimensional states. The correntropy criterion is considered in deriving the state to cope with the non-Gaussian noise and eliminate uncertainty. The rest of this paper is organized as follows. In Section 2, the NMCIF is introduced. In Section 3, the detailed derivation of NMCIF is given, and the robustness and convergence of the algorithm are analyzed. In Section 4, the real data application on the two-lever discrimination task of rats is used to validate the effectiveness of the proposed algorithm compared with other filters. In Section 5, the discussion and conclusion are given.

2. Nonlinear Maximum Correntropy Information Filter

In general, the state-observation model is written as dynamic systems:

$$\mathbf{x}_k = F(\mathbf{x}_{k-1}) + \mathbf{q}_{k-1}, \quad (1)$$

$$\mathbf{y}_k = H(\mathbf{x}_k) + \zeta_k, \quad (2)$$

where \mathbf{x}_k denotes the n -dimensional state at time k . In our application, it is the 2-dimensional position. F denotes the state transition function. In our application, it is a linear function F estimated from the training data by least square [53]. \mathbf{q}_k denotes the process noise with mean $\bar{\mathbf{q}}_k$ and covariance matrix \mathbf{Q}_k , which is estimated by the residue of the state transition approximation. \mathbf{y}_k denotes the m -dimensional measurements. In our application, it is a high-dimensional vector formed by multiple neural firing activities considering the firing history. H is the nonlinear observation model. In our application, it is estimated from the state vector to each dimension of the neural activity independently. In the linear case, each row of H describes how the neural activity weighted tunes the movement states. ζ_k denotes the measurement noise, which is estimated by the residue of using the observation model. For a nonlinear system with high-dimensional correlated observation, such as the most experiments of BMI, the noisy measurements would be collected from hundreds of neurons, and there exists heavy connectivity among neurons.

The commonly-used second-order statistics method, the MMSE criterion, may not be optimal as the nonlinear system generally does not have Gaussian distributed noise [42]. Correntropy is a key concept in ITL, which measures the similarity between two random variables considering all the even order statistics. Assuming X and Y are two random variables, the correntropy is defined as:

$$V(X, Y) = E[\kappa(X, Y)] = \int \kappa(x, y) dF_{XY}(x, y), \quad (3)$$

where E represents the expectation operator, $F_{XY}(x, y)$ denotes the joint distribution function with respect to the two random variables, and $\kappa(\cdot, \cdot)$ is a shift-invariant Mercer kernel. Gaussian kernel is simple and has been successfully applied in a heavy-tailed noise environment [44,46,47], and we chose the Gaussian kernel to eliminate this noise, i.e.,

$$\kappa(x, y) = G_\sigma(e) = \exp\left(-\frac{e^2}{2\sigma^2}\right), \quad (4)$$

where $e = x - y$, and $\sigma > 0$ is the kernel bandwidth. Taking the Taylor series expansion of the Gaussian kernel, it is noted that the correntropy contains not only the second-order moment but also all higher even order statistical information. In particular, when the kernel bandwidth σ is 20 times larger than the values chosen based on Silverman’s rule [44], the second-order moment plays a key role in correntropy.

In this section, we propose a new method for the state-observation model estimation. Let \mathbf{W}_k be defined as the information matrix of process noise, which is the inverse of the covariance matrix $\mathbf{W}_k = \mathbf{Q}_k^{-1}$. We propose the NMCIF.

The prior mean $\bar{\mathbf{x}}_k$ and corresponding information matrix $\bar{\chi}_k$ are firstly computed as:

$$\bar{\mathbf{x}}_k = \mathbf{F}\hat{\mathbf{x}}_{k-1}, \tag{5}$$

$$\bar{\chi}_k = \mathbf{W}_{k-1} - \mathbf{W}_{k-1}\mathbf{F}(\hat{\chi}_{k-1} + \mathbf{F}^T\mathbf{W}_{k-1}\mathbf{F})^{-1}\mathbf{F}^T\mathbf{W}_{k-1}. \tag{6}$$

and the posterior mean $\hat{\mathbf{x}}_k$ is recursively calculated by the fixed-point iterative methods [54] as:

$$\hat{\mathbf{x}}_k^{(t)} = \bar{\mathbf{x}}_k + \tilde{\mathbf{K}}_k^{(t-1)}(g(\mathbf{y}_k) - \bar{\mathbf{x}}_k), \tag{7}$$

where $g(\cdot)$ denotes the nonlinear preprocess on the observation that deducts the dimensionality of original observation into the new observation with the same dimension as the state, such that:

$$g(\mathbf{y}_k) = \mathbf{x}_k + \mathbf{r}_k. \tag{8}$$

Superscript (t) stands for the fixed-point iteration index and the iterative process stops when:

$$\frac{\|\hat{\mathbf{x}}_k^{(t)} - \hat{\mathbf{x}}_k^{(t-1)}\|}{\|\hat{\mathbf{x}}_k^{(t-1)}\|} \leq \omega, \tag{9}$$

with ω being a small positive value, or the iterative index reaches a preset value. $\tilde{\mathbf{K}}_k^{(t-1)}$ stands for the Kalman gain as:

$$\tilde{\mathbf{K}}_k^{(t-1)} = \left(\hat{\chi}_k^{(t-1)} + \tilde{\mathbf{V}}_k^{(t-1)}\right)^{-1}\tilde{\mathbf{V}}_k^{(t-1)}. \tag{10}$$

Here $\hat{\chi}_k^{(t-1)}$ is the revised prior information matrix denoted as in Equation (12) and $\tilde{\mathbf{V}}_k^{(t-1)}$ is denoted as in Equation (13).

In our application, we use multi-layer perception as a universal approximator to take the high-dimensional neural firing as input and output for the preprocessed states as the observation. Then, the observation model is an identity matrix. \mathbf{r}_k denotes the neural network approximation residue with mean $\bar{\mathbf{r}}_k$ and covariance matrix \mathbf{R}_k . Since the neural network could consider the bias, we then assume the process noise and measurement noise have zero means. \mathbf{V}_k is defined by $\mathbf{V}_k = \mathbf{R}_k^{-1}$.

The corresponding information matrix $\hat{\chi}_k$ of the posterior is updated as:

$$\hat{\chi}_k = \hat{\chi}_k^{(t-1)} + \tilde{\mathbf{V}}_k^{(t-1)}, \tag{11}$$

where for each iteration t ,

$$\hat{\chi}_k^{(t-1)} = \mathbf{S}_{\chi,k}\mathbf{C}_{\chi,k}^{(t-1)}\mathbf{S}_{\chi,k}^T, \tag{12}$$

$$\tilde{\mathbf{V}}_k^{(t-1)} = \mathbf{S}_{V,k}\mathbf{C}_{V,k}^{(t-1)}\mathbf{S}_{V,k}^T, \tag{13}$$

$$\mathbf{C}_{\chi,k}^{(t-1)} = \text{diag}\left(G_\sigma\left(e_{k,1}^{(t-1)}\right), \dots, G_\sigma\left(e_{k,n}^{(t-1)}\right)\right), \tag{14}$$

$$\mathbf{C}_{V,k}^{(t-1)} = \text{diag}\left(G_\sigma\left(e_{k,n+1}^{(t-1)}\right), \dots, G_\sigma\left(e_{k,2n}^{(t-1)}\right)\right), \tag{15}$$

$$e_{k,i}^{(t-1)} = d_{k,i} - \mathbf{m}_{k,i}\hat{\mathbf{x}}_k^{(t-1)}, \tag{16}$$

where $d_{k,i}$ denotes the i -th element of \mathbf{D}_k , $\mathbf{m}_{k,i}$ denotes the i -th row of \mathbf{M}_k , and $\mathbf{D}_k = \mathbf{S}_k \begin{bmatrix} \bar{\mathbf{x}}_k \\ g(\mathbf{y}_k) \end{bmatrix}$, $\mathbf{M}_k = \mathbf{S}_k \begin{bmatrix} \mathbf{I}_n \\ \mathbf{I}_n \end{bmatrix}$, $\mathbf{e}_k = \mathbf{S}_k \delta_k$. Here we take \mathbf{D}_k as the transformed observation, \mathbf{M}_k is the new observation model, and \mathbf{e}_k is the noise term. \mathbf{S}_k can be obtained by making Cholesky decomposition on the inverse of covariance of the noise term $\delta_k = \begin{bmatrix} -(\mathbf{x}_k - \bar{\mathbf{x}}_k) \\ \mathbf{r}_k \end{bmatrix}$, and denoted as $\mathbf{S}_k = \begin{bmatrix} \mathbf{S}_{\chi,k} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{V,k} \end{bmatrix}$.

Note that when the kernel bandwidth becomes increasingly larger, the performance of the algorithm will behave like the corresponding nonlinear information filter (NIF). Especially, if $\sigma \rightarrow \infty$, then $\mathbf{C}_{\chi,k} \rightarrow \mathbf{I}$ and $\mathbf{C}_{V,k} \rightarrow \mathbf{I}$, the proposed algorithm will reduce to the NIF. Thus, the NIF algorithm is a special case of the proposed algorithm, and is summarized as follows:

$$\bar{\mathbf{x}}_k = \mathbf{F}\hat{\mathbf{x}}_{k-1}, \tag{17}$$

$$\bar{\chi}_k = \mathbf{W}_{k-1} - \mathbf{W}_{k-1}\mathbf{F}(\hat{\chi}_{k-1} + \mathbf{F}^T\mathbf{W}_{k-1}\mathbf{F})^{-1}\mathbf{F}^T\mathbf{W}_{k-1}, \tag{18}$$

$$\mathbf{K}_k = (\bar{\chi}_k + \mathbf{V}_k)^{-1}\mathbf{V}_k, \tag{19}$$

$$\hat{\mathbf{x}}_k = \bar{\mathbf{x}}_k + \mathbf{K}_k(g(\mathbf{y}_k) - \bar{\mathbf{x}}_k), \tag{20}$$

$$\hat{\chi}_k = \bar{\chi}_k + \mathbf{V}_k. \tag{21}$$

3. Algorithm Derivation and Analysis

3.1. Derivation of Estimation on the Mean of Posterior

Next, we give a derivation process with respect to the nonlinear maximum correntropy information filter.

The mean of the prior state $\bar{\mathbf{x}}_k$ and the corresponding information matrix are computed the same as Equations (5) and (6). By Equations (1), (8) and (17), we can build the following equation:

$$\begin{bmatrix} \bar{\mathbf{x}}_k \\ g(\mathbf{y}_k) \end{bmatrix} = \begin{bmatrix} \mathbf{I}_n \\ \mathbf{I}_n \end{bmatrix} \mathbf{x}_k + \delta_k, \tag{22}$$

where \mathbf{I}_n is an $n \times n$ identity matrix, and $\delta_k = \begin{bmatrix} -(\mathbf{x}_k - \bar{\mathbf{x}}_k) \\ \mathbf{r}_k \end{bmatrix}$.

We can easily obtain the inverse of covariance of the noise term in Equation (22) as the following equation:

$$\begin{aligned} (\mathbb{E}[\delta_k\delta_k^T])^{-1} &= \begin{bmatrix} \bar{\chi}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_k \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{S}_{\chi,k}\mathbf{S}_{\chi,k}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{V,k}\mathbf{S}_{V,k}^T \end{bmatrix} = \mathbf{S}_k\mathbf{S}_k^T, \end{aligned} \tag{23}$$

where \mathbf{S}_k is the Cholesky decomposition of $(\mathbb{E}[\delta_k\delta_k^T])^{-1}$.

Left multiplying both sides of Equation (22) by \mathbf{S}_k yields:

$$\mathbf{D}_k = \mathbf{M}_k\mathbf{x}_k + \mathbf{e}_k, \tag{24}$$

where $\mathbf{D}_k = \mathbf{S}_k \begin{bmatrix} \bar{\mathbf{x}}_k \\ g(\mathbf{y}_k) \end{bmatrix}$, $\mathbf{M}_k = \mathbf{S}_k \begin{bmatrix} \mathbf{I}_n \\ \mathbf{I}_n \end{bmatrix}$, $\mathbf{e}_k = \mathbf{S}_k\delta_k$.

The correntropy-based cost function is introduced as:

$$J(\mathbf{x}_k) = \frac{1}{2n} \sum_{i=1}^{2n} G_\sigma(d_{k,i} - \mathbf{m}_{k,i}\mathbf{x}_k), \tag{25}$$

where $d_{k,i}$ denotes the i -th component of \mathbf{D}_k , $\mathbf{m}_{k,i}$ denotes the i -th row of \mathbf{M}_k , and n is the dimension of the state.

By the similar derivation in [55], the fixed-point iterative method would be adopted as:

$$\hat{\mathbf{x}}_k^{(t)} = \left(\mathbf{M}_k^T \mathbf{C}_k^{(t-1)} \mathbf{M}_k \right)^{-1} \mathbf{M}_k^T \mathbf{C}_k^{(t-1)} \mathbf{D}_k, \tag{26}$$

where superscript (t) stands for the fixed-point iteration index, and $\mathbf{C}_k^{(t-1)} = \begin{bmatrix} \mathbf{C}_{\chi,k}^{(t-1)} & 0 \\ 0 & \mathbf{C}_{V,k}^{(t-1)} \end{bmatrix}$, and the diagonal elements are defined in Equations (14) and (15). In fact, Equation (26) can be further written as Equation (7) (the detailed derivation can be seen in Appendix A).

3.2. Derivation of Information Matrix

After the iterative process stops, the estimation error information matrix $\hat{\chi}_k$ also needs to update. According to [56], the influence function (IF), defined as an n by 1 matrix, can be used to derive the error covariance matrix of an estimator under an assumed probability distribution, and there is a relationship between the influence function and the asymptotic covariance matrix of the estimation error $\hat{\mathbf{P}}_k$ as follows:

$$\hat{\mathbf{P}}_k = \mathbb{E} \left[\text{IF} \cdot \text{IF}^T \right], \tag{27}$$

where \mathbb{E} is the expectation operator. The detailed derivation can be seen in Appendix B. The IF can be computed as:

$$\text{IF}(e, \mathbf{m}^T, A_0, T) = \frac{\varphi(e)}{\mathbb{E}_{A_0}[\varphi'(e)]} \left[\mathbf{M}_k^T \mathbf{M}_k \right]^{-1} \mathbf{m}_{k,i}^T, i = 1, \dots, 2n, \tag{28}$$

where $\varphi(e) = G_\sigma(e)e$, $\mathbb{E}_{A_0}[\varphi'(e)]$ is the expectation of the first derivative of $\varphi(e)$ at distribution A_0 , and A_0 is the target distribution of e . $\mathbf{m}_{k,i}^T$ is the transposition of the i -th row of \mathbf{M}_k , and \mathbf{M}_k is denoted in Equation (24). The detailed derivation can be seen in Appendix C. Thus the corresponding information matrix is written as:

$$\hat{\chi}_k = \frac{\left\{ \mathbb{E}_{A_0}[\varphi'(e)] \right\}^2}{\mathbb{E}_{A_0}[\varphi^2(e)]} \mathbf{M}_k^T \mathbf{M}_k. \tag{29}$$

Here we prove $\mathbb{E}_{A_0}[\varphi'(e)]$ and $\mathbb{E}_{A_0}[\varphi^2(e)]$ as constants in Equations (A29) and (A30) of Appendix D, then substituting into Equation (29) yields:

$$\hat{\chi}_k = \frac{\sigma^3(\sigma^2 + 2\theta^2)\sqrt{\sigma^2 + 2\theta^2}}{\theta^2(\sigma^2 + \theta^2)^3} \mathbf{M}_k^T \mathbf{M}_k, \tag{30}$$

where σ is the kernel bandwidth, and θ is the variance of the distribution A_0 . Note that the information matrix in Equation (30) can only hold in an asymptotic sense and can be used for the state with a high dimension. However, for the state with a small dimension, especially in the derivation of our algorithm, Equation (30) may not be a good choice. Equation (26) can be referred to as the iterative reweighted least squares (IRLS) method. Thus, assuming the iteration terminates at $t = T$, the final estimate is $\hat{\mathbf{x}}_k^{(T)}$ and the associated covariance $\hat{\mathbf{P}}_k$ can be obtained by:

$$\hat{\mathbf{P}}_k = \left(\mathbf{M}_k^T \mathbf{C}_k^{(t-1)} \mathbf{M}_k \right)^{-1}. \tag{31}$$

Thus, the information matrix is computed as:

$$\begin{aligned}\hat{\chi}_k &= \mathbf{M}_k^T \mathbf{C}_k^{(t-1)} \mathbf{M}_k \\ &= \mathbf{S}_{\chi,k}^T \mathbf{C}_{\chi,k}^{(t-1)} \mathbf{S}_{\chi,k} + \mathbf{S}_{V,k}^T \mathbf{C}_{V,k}^{(t-1)} \mathbf{S}_{V,k}. \\ &= \tilde{\chi}_k^{(t-1)} + \tilde{\mathbf{V}}_k^{(t-1)}\end{aligned}\quad (32)$$

We can see that the information matrix in Equation (30) is equivalent to the least squares (LS) method, i.e., $\mathbf{M}_k^T \mathbf{M}_k$, multiplied by a constant scalar being smaller than one. The information matrix in Equation (32) implies that each row is multiplied by a different scalar depending on the residual error $e_{k,i}$ as in Equation (16). The outlier could be eliminated by multiplying a smaller scalar in calculating the information matrix. Thus compared to Equation (30), using Equation (32) may be relatively better in the small dimension case, especially in the proposed algorithm. Unless otherwise specified, the proposed algorithm uses the information matrix in Equation (32).

3.3. Robustness Analysis

According to [56], the influence function (IF) can also be used to measure quantitatively the robustness of an estimator, i.e., the infinitesimal robustness. If an estimator is infinitesimally robust, the corresponding IF needs to satisfy continuity and boundedness [57]. We show the IF satisfies continuity in Equation (28). Next, we will give the proof with respect to the boundedness of IF. According to Equation (28), $E_{A_0}[\varphi'(e)]$ is a constant (please see in Appendix D) and $(\mathbf{M}_k^T \mathbf{M}_k) \mathbf{m}_{k,i}^T$ is a known matrix, and we can see that the boundedness of IF depends on $\varphi(e)$.

Theorem 1. *The function $\varphi(e)$ is bounded in the interval $(-\infty, +\infty)$.*

Proof of Theorem 1.

$$\varphi(e) = e \times \exp\left(-\frac{e^2}{2\sigma^2}\right) = \frac{e}{\exp\left(\frac{e^2}{2\sigma^2}\right)}, \quad (33)$$

where e is the error and σ is the kernel bandwidth. Note that $\varphi(e)$ is a continuous and odd function according to the boundedness theorem [58], thus it satisfies the boundedness in any closed intervals, i.e., if $\exists \tau > 0$, and $\varphi(e)$ is bounded in the interval $[-\tau, \tau]$. In the interval $[\tau, +\infty]$, according to the L' Hospital's rule [59], we have:

$$\lim_{e \rightarrow +\infty} \varphi(e) = \lim_{e \rightarrow +\infty} \frac{e}{\exp\left(\frac{e^2}{2\sigma^2}\right)} = \lim_{e \rightarrow +\infty} \frac{\sigma^2}{e \exp\left(\frac{e^2}{2\sigma^2}\right)} = 0. \quad (34)$$

Consequently, $\varphi(e)$ is bounded in the interval $[\tau, +\infty)$, and it is bounded in the interval $(-\infty, -\tau]$ due to the odd characteristic of the function $\varphi(e)$. This completes the proof. The IF in Equation (28) satisfies continuity and boundedness, and the proposed algorithm can be regarded as infinitesimally robust. \square

3.4. Convergence Analysis

In Section 3.1, we adopt a fixed-point iterative algorithm to obtain the posterior estimation in Equation (26). In fact, the convergence of the fixed-point iteration is impacted by the kernel bandwidth σ . According to the Banach Fixed-Point Theorem [60], the sufficient condition with respect to the choice of the kernel bandwidth to ensure the convergence of the fixed-point iteration is the same as in NMCIF [55,61].

Here we focus on discussing the convergence rate of the fixed-point iteration. For the sake of this discussion, Equation (26) can be written in an equivalent form as:

$$\mathbf{x}_k = \left(\sum_{i=1}^{2n} G_\sigma(e_{k,i}) \mathbf{m}_{k,i}^T \mathbf{m}_{k,i} \right)^{-1} \sum_{i=1}^{2n} G_\sigma(e_{k,i}) d_{k,i} \mathbf{m}_{k,i}^T, \tag{35}$$

where $e_{k,i}$, $\mathbf{m}_{k,i}$, and $d_{k,i}$ are defined as in Equation (24). Thus Equation (35) forms a fixed point equation as $\mathbf{x}_k = f(\mathbf{x}_k)$. Here we use $\nabla_{\mathbf{x}_k} f(\mathbf{x}_k)$ to denote the $n \times n$ Jacobian matrix of $f(\mathbf{x}_k)$ with respect to \mathbf{x}_k , that is:

$$\nabla_{\mathbf{x}_k} f(\mathbf{x}_k) = \left[\frac{\partial}{\partial x_{k,1}} f(\mathbf{x}_k) \dots \frac{\partial}{\partial x_{k,n}} f(\mathbf{x}_k) \right], \tag{36}$$

with

$$\begin{aligned} \frac{\partial}{\partial x_{k,j}} f(\mathbf{x}_k) = & -\mathbf{N}_{\mathbf{mm}}^{-1} \left(\frac{1}{\sigma^2} \sum_{i=1}^{2n} \left[e_{k,i} m_{k,ij} G_\sigma(e_{k,i}) \mathbf{m}_{k,i}^T \mathbf{m}_{k,i} \right] \right) f(\mathbf{x}_k) \\ & + \mathbf{N}_{\mathbf{mm}}^{-1} \left(\frac{1}{\sigma^2} \sum_{i=1}^{2n} \left[e_{k,i} m_{k,ij} G_\sigma(e_{k,i}) \mathbf{m}_{k,i}^T d_{k,i} \right] \right), \end{aligned} \tag{37}$$

where $\mathbf{N}_{\mathbf{mm}} = \sum_{i=1}^{2n} G_\sigma(e_{k,i}) \mathbf{m}_{k,i}^T \mathbf{m}_{k,i}$ and $m_{k,ij}$ is the j -th element of $\mathbf{m}_{k,i}$.

It is obvious that $G_\sigma(e_{k,i}) \leq 1$, and according to the first sufficient condition $\|f(\mathbf{x}_k)\|_1 \leq \beta$ in [55] and Equation (37), we have:

$$\begin{aligned} \left\| \frac{\partial}{\partial x_{k,j}} f(\mathbf{x}_k) \right\|_1 \leq & \left\| -\mathbf{N}_{\mathbf{mm}}^{-1} \right\|_1 \left\| \frac{1}{\sigma^2} \sum_{i=1}^{2n} \left[e_{k,i} m_{k,ij} G_\sigma(e_{k,i}) \mathbf{m}_{k,i}^T \mathbf{m}_{k,i} \right] \right\|_1 \|f(\mathbf{x}_k)\|_1 \\ & + \left\| \mathbf{N}_{\mathbf{mm}}^{-1} \right\|_1 \left\| \frac{1}{\sigma^2} \sum_{i=1}^{2n} \left[e_{k,i} m_{k,ij} G_\sigma(e_{k,i}) \mathbf{m}_{k,i}^T d_{k,i} \right] \right\|_1 \\ \leq & \frac{1}{\sigma^2} \beta \left\| -\mathbf{N}_{\mathbf{mm}}^{-1} \right\|_1 \sum_{i=1}^{2n} \left\| e_{k,i} m_{k,ij} \mathbf{m}_{k,i}^T \mathbf{m}_{k,i} \right\|_1 \\ & + \frac{1}{\sigma^2} \left\| \mathbf{N}_{\mathbf{mm}}^{-1} \right\|_1 \sum_{i=1}^{2n} \left\| e_{k,i} m_{k,ij} \mathbf{m}_{k,i}^T d_{k,i} \right\|_1, \end{aligned} \tag{38}$$

where $\|\cdot\|_1$ denotes a 1-norm of a vector or an induced norm of a matrix. We can see since β , $\mathbf{N}_{\mathbf{mm}}^{-1}$, $e_{k,i}$, $\mathbf{m}_{k,i}$, and $d_{k,i}$ are bounded, it is evident that if we choose a large enough kernel bandwidth σ , the gradient vector is close to zero with a limited bounded 1-norm. Then the gradient at the optimal point is zero in practice and the algorithm is at least quadratically convergent. In particular, if $\sigma \rightarrow \infty$, we have $G_\sigma(e_{k,i}) \rightarrow 1$, then the fixed-point iteration method changes to the MMSE solution (i.e., the NIF in Section 2), which has a zero gradient vector in the optimal solution and converges in just one step. Consequently, when the kernel bandwidth σ decreases, the gradient of the fixed-point method at the optimal solution will increase, and the order of convergence will decrease from the super-linear to linear order of convergence [62]. Moreover, the robustness of the algorithm increases with a smaller kernel bandwidth.

By the above analysis, we can obtain that the smaller the kernel bandwidth, the more robust the algorithm is and the slower the convergence rate. However the kernel bandwidth has a low bound value to guarantee the convergence of the fixed-point method. On the other hand, when the kernel bandwidth becomes increasingly larger, the convergence rate increases, and the performance of the algorithm will behave like the corresponding NIF. Especially, if $\sigma \rightarrow \infty$, then $\mathbf{C}_k \rightarrow \mathbf{I}$, the proposed algorithm will reduce to the NIF and converge in just one step.

4. Neural Decoding Using Nonlinear Maximum Correntropy Information Filter

In this section, we apply the proposed nonlinear maximum correntropy information filter to decode the two-dimensional movement intention from high-dimensional neural observations, and make comparisons with the Kalman filter and multi-layer perception network.

4.1. Experiment and Data Collection

The BMI experiment paradigm of the rats' two-lever discrimination task was designed and conducted in the Hong Kong University of Science and Technology. All animal handling procedures were approved by the Animal Care Committee of the Hong Kong University of Science and Technology. Two 16-channel microelectrode arrays were implanted in the primary motor cortex (M1) and medial prefrontal cortex (mPFC) respectively on the left hemisphere. The raw data of neural extracellular potential was recorded by a multi-channel acquisition processor (Plexon Inc, Dallas, Texas) with a 40-kHz frequency. The recorded potential was filtered by a 4-pole Butterworth high pass filter at 500 Hz and the action potentials were detected with an artificially set threshold (about 4 times of standard deviation of the filtered potential) in an online recording system. The spike trains were binned by a 100-ms time window without overlap for each channel. The count of spikes in the 100-ms window was assigned as firing rates.

The rats were trained to perform the two-lever discrimination task. In this task, the rats were required to distinguish two audio tones, which were randomly given by the computer at the trial start. The rats needed to use their right paw to press the corresponding lever and hold it for over 500 ms within a 6-second try-out time to get a water reward. With the high pitch (10 kHz), the rats needed to reach the high lever which was 10 cm high. While with the low tone (1.5 kHz), the rats turned to the low lever which was positioned 6 cm lower than the high lever. After successfully holding, the water was provided by a pump with a short feedback tone which had the same frequency as the start cue. The behavioral events and their timing information were recorded by the behavioral chamber (Lafayette Instrument, Lafayette, USA), and synchronized through the Plexon digital input with neural spike activity. There are totally 51 trials of a high lever press and 80 trials of a low lever press in rat A, as well as 72 trials of a high lever press and 128 trials of a low lever press in rat B.

To model the above procedure, we only consider the total successful trials, and the two-dimensional states are denoted as \mathbf{x}_k . To label the behavioral data, all successful trials are segmented from start to 500 ms after the start, and from 500 ms before the press to the press onset. All segments are connected to represent the behavioral process and smoothed by a sigmoid function. After the feedback, the rats spend 200 ms returning rest state and it is smoothed by a sigmoid function as well. To label the behavioral data, the 500 ms before the start cue is set as the rest state with [0,0], and holding a high lever and low lever are set as [1,1] and [1,-1] respectively. We use the current time step of the neural firing rates as the 32-dimensional measurements, which are denoted as \mathbf{y}_k to decode the behavioral state. Figure 1c shows one segment of spike rasters over time in five channels. And the corresponding behavioral states in 2D are shown in Figure 1a,b respectively. Following the above design, the state equations and measurement equations can be written as Equations (1) and (8). The state transition \mathbf{F} in Equation (1) is obtained by the least square approach, the nonlinear relationship $g(\cdot)$ in Equation (8) is approximated by a multi-layer perception network, and the mean and covariance of the process noise \mathbf{q}_k in Equation (1) and measurement noise \mathbf{r}_k in Equation (8) are estimated from the residual errors. Next according to the above model, we apply the proposed algorithm to estimate the two-dimensional states \mathbf{x}_k from the high-dimensional measurements \mathbf{y}_k and compare it with the KF and neural network decoding performance. Note that the neural network takes in the current neural firing rates with the past 400 ms firing rates history and outputs the 2-dimensional movement intention. The number of hidden PE is set as 10. The weights are initialized 20 times and trained by the steepest gradient-descent back propagation

method, which minimizes the mean square error between the output and the ground truth movement. We use 60% of the training data for weight training and 40% for testing. We use the 2D-mean square error between the movement estimation and the ground truth as the criterion to evaluate the decoding performance.

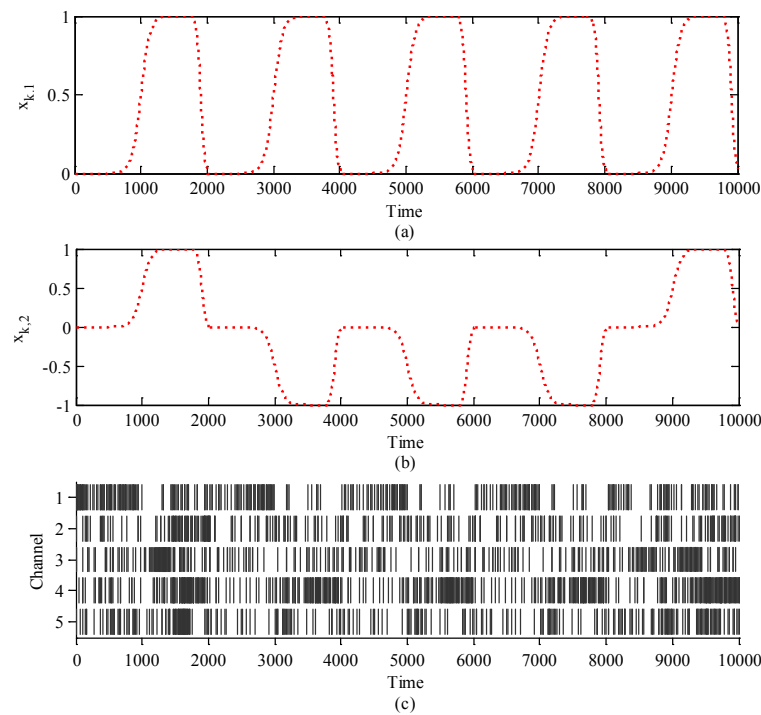


Figure 1. The 2-D behavioral states (a,b) and corresponding spike rasters in five channels of M1 (c).

4.2. Results

Figure 2a,b show a segment of decoded 2-D movements, $x_{k,1}$ and $x_{k,2}$ using KF, with NMCIF and NN presented for comparison. We can see that the curve of NMCIF (the magenta solid curve) is closer to the ground truth (the red dotted curve) than that of KF (the blue solid curve). This is because the proposed method adopts the nonlinear process of observation into the same dimension of the states instead of the linear model, which considers interactions among recorded neurons. Moreover, from the 210th time step to the 220th time step, the 320th time step to the 330th time step, and the 350th time step to the 400th time step in $x_{k,2}$ of Figure 2, we can see that the curve of NMCIF (the magenta solid curve) is smoother than that of NN (the green solid curve) due to the consideration of the time dynamic nature of the state. Table 1 shows the statistical 2D-MSEs of the algorithms between the true values and estimation values across 10 data segments with each rat. Please note that here for the NN, we use a MLP network to approximate the nonlinearity. We can also use RNN, where the results of RNN is 0.3555 ± 0.0959 and 0.3984 ± 0.0735 . In terms of the averaged performance, the 2D-MSE of NMCIF in rat_A decreases by 57.62% and 6.91% compared with those of KF and NN, and the 2D-MSE of NMCIF in rat_B decreases by 25.89% and 5.17%. The results show that the proposed algorithm is superior to the KF and NN.

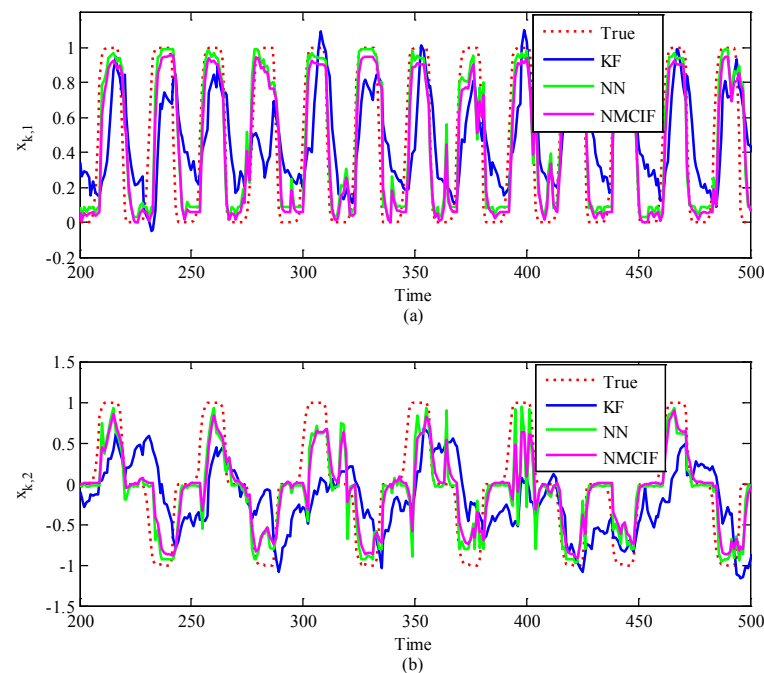


Figure 2. The reconstruction of 2–D movements over time by different models. The X–axis is the time, the Y–axis is the first (a) and the second (b) dimension of the movement. The red dotted curve is the ground truth, the blue solid curve is the Kalman filter, the magenta solid curve is the nonlinear maximum correntropy information filter, and the green solid curve is the neural network.

Table 1. Statistical decoding performances across 10 segments of test data with Different rats.

Method	2D-MSE of x_k in Rat_A	2D-MSE of x_k in Rat_B
KF	0.5783 ± 0.1074	0.5558 ± 0.0653
NN	0.2633 ± 0.0787	0.4119 ± 0.0588
NMCIF	0.2451 ± 0.0684	0.3906 ± 0.0491

We then discuss convergence and sensitivity of the proposed algorithm with the presence of large deviated state initials.

From Section 3.4, we know that the kernel bandwidth plays a key role on the performance of NMCIF. Figure 3 shows the relationships between performance and the kernel bandwidth. We can see that if the kernel bandwidth is too small, the proposed algorithm plays a worse performance, even diverges. If the kernel bandwidth is too large, the performance of NMCIF (the magenta solid curve) is similar to the performance of NIF (the black solid curve). With a proper kernel bandwidth, the NMCIF outperforms the NIF. In particular, when the kernel bandwidth $\sigma = 2$, the NMCIF has its best performance. Figure 4 shows the relationship between the fixed-point iteration number and the kernel bandwidth. As one can see, the larger the kernel bandwidth is, the less the iteration number.

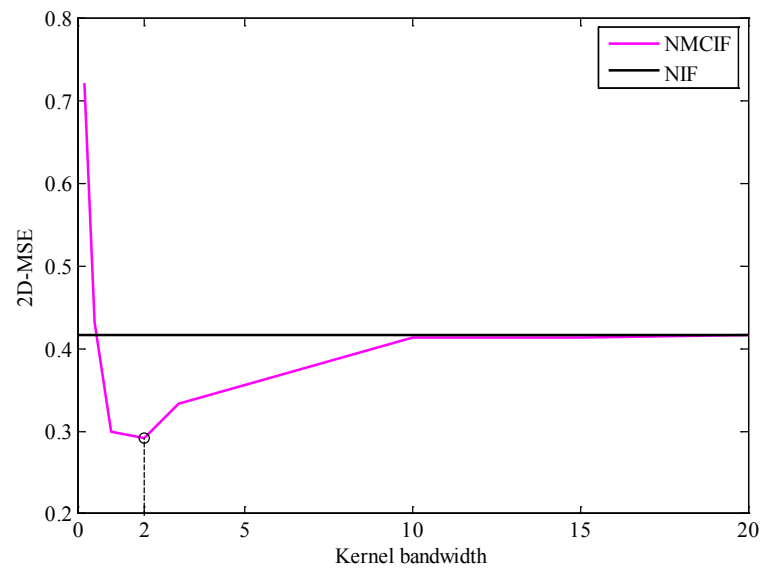


Figure 3. The relationship between the performance and kernel bandwidth.

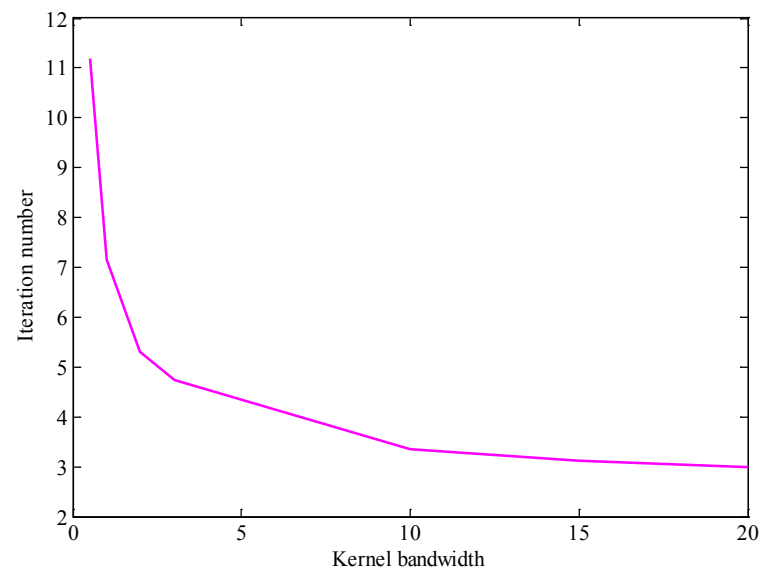


Figure 4. The relationship between the iteration number and kernel bandwidth.

Figure 5 shows the comparison between the KF and proposed NMCIF to exhibit the advantage of the information type filter when the initial condition is largely uncertain. The initial covariance matrix in KF is set to 1, but the initial information matrix in NMCIF is set to a small positive value (10^{-6}). Figure 5a,b show the estimation values of $x_{k,1}$ and $x_{k,2}$ in different algorithms compared with the ground truth on a segment of the data when the initial estimation value deviates greatly from the initial true value. As one can see, the NMCIF (the magenta solid curve) can converge quickly to the neighborhood of the ground truth (the red dotted curve), and KF (the blue solid curve) converges slowly. Table 2 summarizes the statistical decoding error over 10 segments of data with large deviated initial values. In terms of the averaged performance, the 2D-MSE of x_k by NMCIF in rat_A and rat_B decrease by 33.84% and 26.00% respectively compared with that of KF. It is obvious that NMCIF is superior to KF when the initial estimation value deviates greatly from the initial true values.

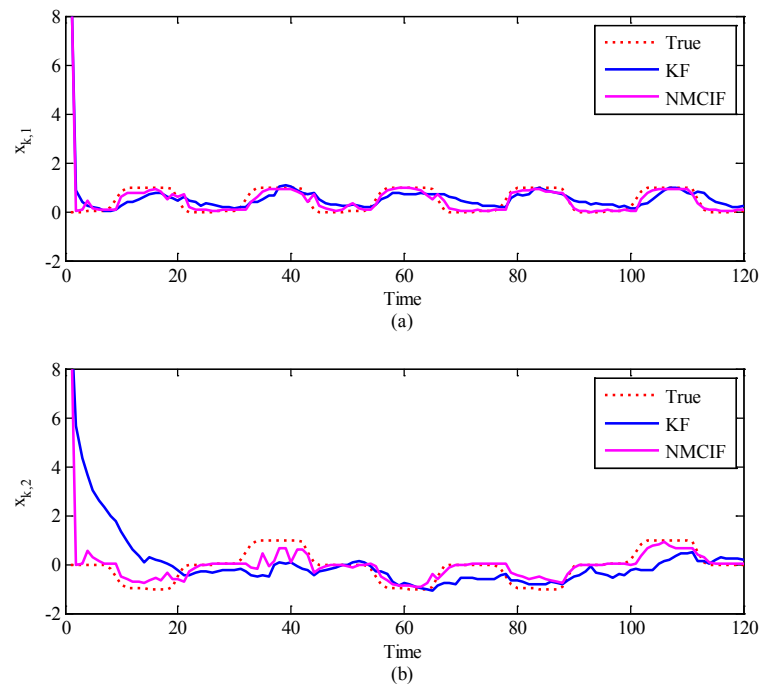


Figure 5. The reconstruction of 2–D movements with the initial value deviates greatly from the truth. The X–axis is the time, the Y–axis is the first (a) and the second (b) dimension of the movement. The red dotted curve is ground truth, the blue solid curve is the Kalman filter, and the magenta solid curve is the nonlinear maximum correntropy information filter.

Table 2. Statistical decoding performances across 10 segments of test data with large deviated initial values.

Method	2D-MSE of x_k in Rat_A	2D-MSE of x_k in Rat_B
KF	2.8686 ± 0.2112	1.8142 ± 0.1218
NMCIF	1.8978 ± 0.0661	1.3425 ± 0.0477

As our algorithm demonstrates the best performance over KF and NN for neural decoding, here we further discuss the robustness of the proposed algorithm with the presence of large noise, which is generally seen in the online recorded neural activity [63]. Here 3.3% of time instances are added randomly with large noise, which corresponds to r_k in Equation (8). Figure 6a,b show one segment of movement decoding using nonlinear information filter and the proposed method. In these figures, the red dotted curve denotes the ground truth and the blue asterisks denote the timing with large noise. The performance of NIF and NMCIF are presented for comparison. From the 410th time step to the 420th time step and the 460th time step to the 470th time step in Figure 6a and the 440th time step to the 450th time step in Figure 6b, we can see that the curve of NMCIF (the magenta solid curve) is smoother than that of NIF (the black solid curve) because of the usage of correntropy against the heavy-tailed non-Gaussian noise. The statistical 2D-MSEs of x_k by NIF and NMCIF across 10 data segments are summarized in Table 3, respectively. Note that NMCIF_A denotes the NMCIF using the information matrix in Equation (30), and NMCIF_B denotes the NMCIF using the information matrix in Equation (32). In terms of the averaged performance, the 2D-MSE of NMCIF_B in rat_A and rat_B decrease by 29.54% and 8.97% respectively compared with that of NIF. Therefore, NMCIF exhibits better performance than NIF in non-Gaussian noise. The 2D-MSE of NMCIF_B is less than that of NMCIF_A, which indicates that the information matrix in Equation (32) may be relatively better in NMCIF.

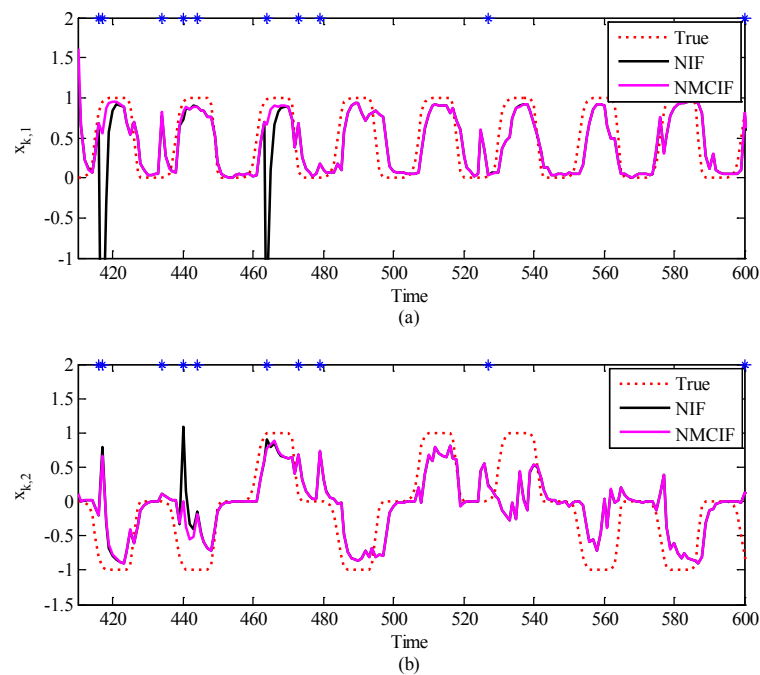


Figure 6. The reconstruction of 2-D movements affected by noisy observation. The X-axis is the time, The Y-axis is the first (a) and the second (b) dimension of the movement. The red dotted curve is ground truth, the black solid curve is the nonlinear information filter, and the magenta solid curve is the nonlinear maximum correntropy information filter.

Table 3. Statistical decoding performances across 10 segments of test data affected by non-Gaussian noise.

Method	2D-MSE of x_k in Rat_A	2D-MSE of x_k in Rat_B
NIF	0.4113 ± 0.1165	0.4962 ± 0.0456
NMCIF_A	0.2933 ± 0.0630	0.4548 ± 0.0453
NMCIF_B	0.2898 ± 0.0637	0.4517 ± 0.0437

5. Discussion and Conclusions

BMI records and decodes the neural signals to predict movement intents with the goal of helping better the lives of motor-impaired patients by restoring motor functions. As a widely-used decoding algorithm, KF is usually adopted to estimate the motor as the states from neural activity observations with the linear assumption and Gaussian noise. However, the nervous systems are nonlinear in general, and the observations are high-dimensional because recordings are collected from hundreds of neurons and could be very noisy, and the initial movement states of disabled patients cannot always be known. All these factors would result in the performance degradation of the KF. In this paper, we propose a nonlinear maximum correntropy information filter to derive a state in the filtering process from the noisy high-dimensional measurements. A nonlinear model is used to preprocess the high-dimensional neural observations into the same dimensionality as the states, which allows the connectivity among neurons to be considered. The correntropy criterion is adopted to address the presence of the non-Gaussian noise. The information matrix of the state is derived with less sensitivity to the state's initial conditions. We provide an analysis of the convergence condition and robustness. The proposed algorithm is applied to decode the movements from real neural data collected from rats performing a two-lever discrimination task. The 2D reconstruction error (2D-MSE) of NMCIF in rat_A decreases by 57.62% and 6.91% compared with those of KF and NN, and the 2D-MSE of NMCIF in rat_B decreases by 25.89% and 5.17%. The results demonstrate that the nonlinear model considering the connection of neural activities contributes to better estimation performance.

When the initial estimation value deviates greatly from the initial true value, the 2D-MSEs of NMCIF in rat_A and rat_B decrease by 33.84% and 26.00% respectively compared with those of KF. These exhibit the superiority of the information matrix type filter. Moreover, the 2D-MSEs of NMCIF in rat_A and rat_B decrease by 29.54% and 8.97% respectively compared with those of NIF in the presence of heavy-tailed non-Gaussian noise showing the proposed algorithm's superiority to the NIF. These results confirm the effectiveness of the proposed algorithm, which potentially improves the decoding performance for BMI. In our study, the NN used as the nonlinear approximator is not limited to MLP but can also be RNN etc., as RNN is usually used in speech recognition and natural language processing. In the area of motor intention estimation, it has been recently used to classify the letter shapes using invasive brain signals [37]. However, the training of MLP is much simpler than the backpropagation through time in RNN. Our method follows the combination of the neural network and simple state-observation model. It inherits a simple explanation on state dynamics and observation model, which is correspondingly the neural tuning property and neural connectivity. On the other hand, the neural tuning property generally changes over time [64,65], the performance of static decoder may decrease if we use the fixed observation model. Thus considering the adaptation in the neural system, the neural network in our method allows the introduction of adaptive nonlinear models [66,67] to replace the stationary observation model. Our study currently uses invasive recordings on rat's single neuron spikes. The rat is an ideal subject for estimating motor intention and is widely used in many papers [68–70], and the Kalman filter has been used in rats [39], monkeys [40], and human subjects [71]. We plan to utilize our algorithm on non-human primates and patients in the future. As more electrodes could be implanted in such subjects, the recorded neural signals would build higher dimensional observations. We would expect the superiority of the proposed algorithm to be clearer on these data.

Author Contributions: Methodology, X.L.; software, X.L. and S.C.; data acquisition and processing, S.C., X.S. and X.Z.; paper writing, X.L.; supervision, Y.W.; project administration, Y.W.; funding acquisition, Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Shenzhen-Hong Kong Innovation Circle (Category D, grant number SGD2019081623021543), the Hong Kong Innovation and Technology Fund (grant number ITS/092/17), the National Natural Science Foundation of China (grant number 61836003), Chau Hoi Shuen Foundation's Donation (grant number R9051), and sponsorship scheme for targeted strategic partnership (grant number FP902).

Institutional Review Board Statement: The experimental paradigm was conducted at the Hong Kong University of Science and Technology. All animal handling procedures were approved by the Animal Ethics Committee at the Hong Kong University of Science and Technology (Protocol code: 2017082, approved at 13 November 2017).

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Derivation Process from Equation (26) to Equation (7)

In this part, we derive Equation (26) to Equation (7). According to Equations (24) and (26), the following equations can be:

$$\mathbf{M}_k = \mathbf{S}_k \begin{bmatrix} \mathbf{I}_n \\ \mathbf{I}_n \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{\chi,k} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{V,k} \end{bmatrix} \begin{bmatrix} \mathbf{I}_n \\ \mathbf{I}_n \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{\chi,k} \\ \mathbf{S}_{V,k} \end{bmatrix}, \quad (\text{A1})$$

$$\mathbf{C}_k^{(t-1)} = \begin{bmatrix} \mathbf{C}_{\chi,k}^{(t-1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{V,k}^{(t-1)} \end{bmatrix}, \quad (\text{A2})$$

$$\mathbf{D}_k = \mathbf{S}_k \begin{bmatrix} \bar{\mathbf{x}}_k \\ g(\mathbf{y}_k) \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{\chi,k} \bar{\mathbf{x}}_k \\ \mathbf{S}_{V,k} g(\mathbf{y}_k) \end{bmatrix}, \quad (\text{A3})$$

$$\left(\mathbf{M}_k^T \mathbf{C}_k^{(t-1)} \mathbf{M}_k \right)^{-1} = \left[\mathbf{S}_{\chi,k}^T \mathbf{C}_{\chi,k}^{(t-1)} \mathbf{S}_{\chi,k} + \mathbf{S}_{V,k}^T \mathbf{C}_{V,k}^{(t-1)} \mathbf{S}_{V,k} \right]^{-1}. \quad (\text{A4})$$

Using the following matrix inversion lemma:

$$\left(\mathbf{A} + \mathbf{B} \mathbf{D}^{-1} \mathbf{C} \right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{D} + \mathbf{C} \mathbf{A}^{-1} \mathbf{B})^{-1} \mathbf{C} \mathbf{A}^{-1}, \quad (\text{A5})$$

$$\begin{aligned} \mathbf{S}_{\chi,k}^T \mathbf{C}_{\chi,k}^{(t-1)} \mathbf{S}_{\chi,k} &\rightarrow \mathbf{A}, \mathbf{I}_n \rightarrow \mathbf{B}, \\ \mathbf{I}_n &\rightarrow \mathbf{C}, \mathbf{S}_{V,k}^T \mathbf{C}_{V,k}^{(t-1)} \mathbf{S}_{V,k} \rightarrow \mathbf{D}^{-1}. \end{aligned} \quad (\text{A6})$$

Equation (A4) can also be written as:

$$\begin{aligned} \left(\mathbf{M}_k^T \mathbf{C}_k^{(t-1)} \mathbf{M}_k \right)^{-1} &= \mathbf{S}_{\chi,k}^{-1} \left(\mathbf{C}_{\chi,k}^{(t-1)} \right)^{-1} \left(\mathbf{S}_{\chi,k}^T \right)^{-1} - \mathbf{S}_{\chi,k}^{-1} \left(\mathbf{C}_{\chi,k}^{(t-1)} \right)^{-1} \left(\mathbf{S}_{\chi,k}^T \right)^{-1} \\ &\quad \times \left(\mathbf{S}_{V,k}^{-1} \left(\mathbf{C}_{V,k}^{(t-1)} \right)^{-1} \left(\mathbf{S}_{V,k}^T \right)^{-1} + \mathbf{S}_{\chi,k}^{-1} \left(\mathbf{C}_{\chi,k}^{(t-1)} \right)^{-1} \left(\mathbf{S}_{\chi,k}^T \right)^{-1} \right)^{-1} \\ &\quad \times \mathbf{S}_{\chi,k}^{-1} \left(\mathbf{C}_{\chi,k}^{(t-1)} \right)^{-1} \left(\mathbf{S}_{\chi,k}^T \right)^{-1}. \end{aligned} \quad (\text{A7})$$

By Equations (A1)~(A3), we have:

$$\mathbf{M}_k^T \mathbf{C}_k^{(t-1)} \mathbf{D}_k = \mathbf{S}_{\chi,k}^T \mathbf{C}_{\chi,k}^{(t-1)} \mathbf{S}_{\chi,k} \bar{\mathbf{x}}_k + \mathbf{S}_{V,k}^T \mathbf{C}_{V,k}^{(t-1)} \mathbf{S}_{V,k} g(\mathbf{y}_k). \quad (\text{A8})$$

By substituting Equations (A7) and (A8) into Equation (26), we have Equation (7).

Appendix B. The Relationship of the Influence Function and Asymptotic Covariance Matrix

Following [57], we derive the relationship of the IF to the asymptotic covariance matrix of the estimation error vector. Using a form of Taylor expansion involving the derivative of the functional to express the estimator $T(A)$ yields:

$$T(A) = T(A_0) + T'(A - A_0) + \text{rem}(A - A_0), \quad (\text{A9})$$

where $\text{rem}(A - A_0)$ is a remainder term. According to the von Mises expansion and assuming the IF exists, Equation (A9) can be written as:

$$T(A) = T(A_0) + \int \text{IF}(e, A_0) d(A - A_0) + \text{rem}(A - A_0). \quad (\text{A10})$$

Since the Fisher consistency at A_0 , obtained by $\int \text{IF}(e, A_0) d(A_0) = 0$, we have:

$$T(A) = T(A_0) + \int \text{IF}(e, A_0) dA + \text{rem}(A - A_0). \quad (\text{A11})$$

For $A = F_m(e)$ with $F_m(e)$ being the empirical distribution function, obtained by $F_m(e) = (1/m) \sum_{i=1}^m u(e - e_i)$ with $u(\cdot)$ being the unit step function, the integral term in Equation (A11) can be written in linear term as:

$$\int \text{IF}(e, A_0) dF_m = \frac{1}{m} \sum_{i=1}^m \text{IF}(e_i, A_0). \quad (\text{A12})$$

Then, we have:

$$\sqrt{m}(T(F_m) - T(A_0)) = \frac{1}{\sqrt{m}} \sum_{i=1}^m \text{IF}_i(e, A_0) + \sqrt{m}rem(F_m - A_0). \tag{A13}$$

If the remainder term converges in probability to zero, then by virtue of the central limit theorem and Slutsky’s lemma, the probability distribution of $\sqrt{m}(T(F_m) - T(A_0))$ tends to $N(0, \Lambda_k)$, with an asymptotic covariance matrix given by:

$$\widehat{\mathbf{P}}_k = E[\text{IF} \cdot \text{IF}^T]. \tag{A14}$$

Appendix C. The Derivation of Influence Function

For the sake of this discuss, we omit the time step k . From Equation (24), we can obtain the residual vector $\mathbf{e} = \mathbf{D} - \mathbf{M}\mathbf{x}$. Consider a ε -contamination model $A = (1 - \varepsilon)A_0 + \varepsilon\Delta e$, where A_0 is the target distribution, Δe is the probability mass at e , and ε is the contaminating ratio, and define the cumulative probability distribution of the residual vector as $A_0(\mathbf{e})$. In general, the state estimate based on correntropy can be obtained by solving Equation (26), which can be written in an equivalent form as:

$$\sum_{i=1}^{2n} \lambda(e_i, \mathbf{m}_i^T, \mathbf{x}) = \sum_{i=1}^{2n} G_\sigma(e_i) e_i \mathbf{m}_i^T = \mathbf{0}. \tag{A15}$$

Let $T(A)$ be an estimator based on correntropy (i.e., state estimate $\widehat{\mathbf{x}}_k$ at the distribution A), which is written in a functional form of A . If each $\lambda(e_i, \mathbf{m}_i^T, \mathbf{x})$ in Equation (A15) is left-multiplied by probability mass $\frac{1}{2n}$, and from the Glivenko–Cantelli theorem [72], Equation (A15) asymptotically tends to:

$$\int \lambda(e, \mathbf{m}^T, T(A)) dA = \mathbf{0}, \tag{A16}$$

where e and \mathbf{m}^T are generalizations of e_i and \mathbf{m}_i^T . According to [56], the asymptotic IF of $T(A)$ is defined as:

$$\text{IF}(e, \mathbf{m}^T, A_0, T) = \left. \frac{\partial T(A)}{\partial \varepsilon} \right|_{\varepsilon=0} = \lim_{\Delta \varepsilon \rightarrow 0} \frac{T((1 - \varepsilon)A_0 + \varepsilon\Delta e) - T(A_0)}{\varepsilon}. \tag{A17}$$

To derive the expression of IF, we substitute the ε -contamination model $A = (1 - \varepsilon)A_0 + \varepsilon\Delta e$ into Equation (A16) to yield:

$$\begin{aligned} & \int \lambda(e, \mathbf{m}^T, T(A)) dA \\ &= \int \lambda(e, \mathbf{m}^T, T(A)) d((1 - \varepsilon)A_0 + \varepsilon\Delta e) \\ &= \int \lambda(e, \mathbf{m}^T, T(A)) dA_0 + \varepsilon \int \lambda(e, \mathbf{m}^T, T(A)) d(\Delta e - A_0) = \mathbf{0}. \end{aligned} \tag{A18}$$

Taking the differentiation of Equation (A18) with respect to ε yields:

$$\begin{aligned} & \frac{\partial}{\partial \varepsilon} \int \lambda(e, \mathbf{m}^T, T(A)) dA_0 + \int \lambda(e, \mathbf{m}^T, T(A)) d(\Delta e - A_0) \\ &+ \varepsilon \frac{d}{d\varepsilon} \int \lambda(e, \mathbf{m}^T, T(A)) d(\Delta e - A_0) = \mathbf{0}. \end{aligned} \tag{A19}$$

Calculating Equation (A19) at $\varepsilon = 0$, and assuming $T(A)$ at the distribution A_0 follows the Fisher consistency and regularity conditions, obtained by $\int \lambda(e, \mathbf{m}^T, T(A_0)) dA_0 = \mathbf{0}$, and utilizing the interchangeability of differentiation and integration theorem, we have:

$$\int \frac{\partial}{\partial \varepsilon} \lambda(e, \mathbf{m}^T, T(A)) dA_0 \Big|_{\varepsilon=0} + \int \lambda(e, \mathbf{m}^T, T(A)) d(\Delta e) \Big|_{\varepsilon=0} = \mathbf{0}. \tag{A20}$$

Utilizing the shifting property of the Δe yields:

$$\int \frac{\partial}{\partial \varepsilon} \lambda(e, \mathbf{m}^T, T(A)) dA_0 \Big|_{\varepsilon=0} + \lambda(e, \mathbf{m}^T, T(A_0)) = \mathbf{0}. \tag{A21}$$

Assuming $\lambda(\cdot)$ is continuous and measurable and $\lambda'(\cdot)$ is measurable, we have:

$$\begin{aligned} & \int \frac{\partial}{\partial \varepsilon} \lambda(e, \mathbf{m}^T, T(A)) dA_0 \Big|_{\varepsilon=0} + \lambda(e, \mathbf{m}^T, T(A_0)) \\ &= \int \frac{\partial \lambda(e, \mathbf{m}^T, T(A))}{\partial T(A)} \Big|_{T(A_0)} \cdot \frac{\partial T(A)}{\partial \varepsilon} \Big|_{\varepsilon=0} dA_0 + \lambda(e, \mathbf{m}^T, T(A_0)) = \mathbf{0}. \end{aligned} \tag{A22}$$

Since $T(A)$ is $\hat{\mathbf{x}}_k$ at the distribution A , we have:

$$\begin{aligned} \text{IF}(e, \mathbf{m}^T, A_0, T) &= \frac{\partial T(A)}{\partial \varepsilon} \Big|_{\varepsilon=0} \\ &= - \left[\int \frac{\partial \lambda(e, \mathbf{m}^T, T(A))}{\partial T(A)} \Big|_{T(A_0)} dA_0 \right]^{-1} \lambda(e, \mathbf{m}^T, T(A_0)) \\ &= - \left[\int \frac{\partial \lambda(e, \mathbf{m}^T, \mathbf{x})}{\partial \mathbf{x}} \Big|_{T(A_0)} dA_0 \right]^{-1} \lambda(e, \mathbf{m}^T, T(A_0)). \end{aligned} \tag{A23}$$

Taking the differentiation of $\lambda(e, \mathbf{m}^T, \mathbf{x})$ in Equation (A23) with respect to \mathbf{x} yields:

$$\frac{\partial \lambda(e, \mathbf{m}^T, \mathbf{x}_k)}{\partial \mathbf{x}_k} = \frac{\partial \varphi(e)}{\partial \mathbf{x}_k} \mathbf{m}^T = \frac{\partial \varphi(e)}{\partial e} \frac{\partial e}{\partial \mathbf{x}_k} \mathbf{m}^T = -\varphi'(e) \mathbf{m}^T \mathbf{m}. \tag{A24}$$

where $\varphi(e) = G_\sigma(e)e$.

By substituting Equations (24), (A15), and (A24) into Equation (A23) yields:

$$\text{IF}(e, \mathbf{m}^T, A_0, T) = \left[\int \varphi'(e) \mathbf{m}^T \mathbf{m} dA_0 \right]_{T(A_0)}^{-1} \varphi(e) \mathbf{m}^T. \tag{A25}$$

The integral in Equation (A25) can be regarded as computing the ensemble average of the function of the term $\varphi'(e)$ for every $(\mathbf{m}_i^T \mathbf{m}_i), i = 1, \dots, 2n$, and it can be factored out as:

$$\int \varphi'(e) \mathbf{m}^T \mathbf{m} dA_0 = E_{A_0}[\varphi'(e)] \mathbf{M}^T \mathbf{M}. \tag{A26}$$

Then, we have:

$$\text{IF}(e, \mathbf{m}^T, A_0, T) = \frac{\varphi(e)}{E_{A_0}[\varphi'(e)]} [\mathbf{M}^T \mathbf{M}]^{-1} \mathbf{m}^T. \quad (\text{A27})$$

Since $E_{A_0}[\varphi'(e)]$ is a constant (please see in Appendix D) and $(\mathbf{M}^T \mathbf{M}) \mathbf{m}^T$ is a known matrix, the IF in Equation (A27) is proportional to $\varphi(e)$. IF can evaluate the influence of each elements of the vector \mathbf{e}_k in Equation (24), i.e.,

$$\text{IF}(e, \mathbf{m}^T, A_0, T) = \frac{\varphi(d_{k,i} - \mathbf{m}_{k,i} \mathbf{x}_k)}{E_{A_0}[\varphi'(e)]} [\mathbf{M}_k^T \mathbf{M}_k]^{-1} \mathbf{m}_{k,i}^T, i = 1, \dots, 2n. \quad (\text{A28})$$

Appendix D. The Solution of $E_{A_0}[\varphi'(e)]$ and $E_{A_0}[\varphi^2(e)]$

Assume the distribution A_0 obeying a Gaussian distribution with zero mean and variance θ , we have:

$$\begin{aligned} E_{A_0}[\varphi'(e)] &= \int_{-\infty}^{+\infty} \varphi'(e) dA_0 \\ &= \int_{-\infty}^{+\infty} \left(\exp\left(-\frac{e^2}{2\sigma^2}\right) - \frac{e^2}{\sigma^2} \exp\left(-\frac{e^2}{2\sigma^2}\right) \right) dA_0 \\ &= \int_{-\infty}^{+\infty} \exp\left(-\frac{e^2}{2\sigma^2}\right) \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{e^2}{2\theta^2}\right) de \\ &\quad - \int_{-\infty}^{+\infty} \frac{e^2}{\sigma^2} \exp\left(-\frac{e^2}{2\sigma^2}\right) \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{e^2}{2\theta^2}\right) de \\ &= \frac{1}{\sqrt{2\pi\theta}} \int_{-\infty}^{+\infty} \exp\left(-\frac{e^2}{2\frac{\sigma^2\theta^2}{\sigma^2+\theta^2}}\right) de - \frac{1}{\sqrt{2\pi\theta\sigma^2}} \int_{-\infty}^{+\infty} e^2 \exp\left(-\frac{e^2}{2\frac{\sigma^2\theta^2}{\sigma^2+\theta^2}}\right) de \\ &= \frac{\sigma}{\sqrt{\sigma^2+\theta^2}} + \frac{1}{\sqrt{2\pi\theta\sigma^2}} \frac{\sigma^2\theta^2}{\sigma^2+\theta^2} \int_{-\infty}^{+\infty} ed \exp\left(-\frac{e^2}{2\frac{\sigma^2\theta^2}{\sigma^2+\theta^2}}\right) \\ &= \frac{\sigma}{\sqrt{\sigma^2+\theta^2}} + \frac{\theta}{\sqrt{2\pi}(\sigma^2+\theta^2)} \left(e \cdot \exp\left(-\frac{e^2}{2\frac{\sigma^2\theta^2}{\sigma^2+\theta^2}}\right) \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} \exp\left(-\frac{e^2}{2\frac{\sigma^2\theta^2}{\sigma^2+\theta^2}}\right) de \right) \\ &= \frac{\sigma}{\sqrt{\sigma^2+\theta^2}} + \frac{\theta}{\sqrt{2\pi}(\sigma^2+\theta^2)} \left(-\frac{\sqrt{2\pi}\sigma\theta}{\sqrt{\sigma^2+\theta^2}} \right) \\ &= \frac{\sigma^3}{(\sigma^2+\theta^2)\sqrt{\sigma^2+\theta^2}} \end{aligned} \quad (\text{A29})$$

$$\begin{aligned}
E_{A_0}[\varphi^2(e)] &= \int_{-\infty}^{+\infty} \varphi^2(e) dA_0 \\
&= \int_{-\infty}^{+\infty} \left(e^2 \exp\left(-\frac{e^2}{\sigma^2}\right) \right) dA_0 \\
&= \int_{-\infty}^{+\infty} e^2 \exp\left(-\frac{e^2}{\sigma^2}\right) \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{e^2}{2\theta^2}\right) de \\
&= \frac{1}{\sqrt{2\pi\theta}} \int_{-\infty}^{+\infty} e^2 \exp\left(-\frac{e^2}{2\frac{\sigma^2\theta^2}{\sigma^2+2\theta^2}}\right) de \\
&= -\frac{1}{\sqrt{2\pi\theta}} \frac{\sigma^2\theta^2}{\sigma^2+2\theta^2} \int_{-\infty}^{+\infty} ed \exp\left(-\frac{e^2}{2\frac{\sigma^2\theta^2}{\sigma^2+2\theta^2}}\right) \\
&= -\frac{\sigma^2\theta}{\sqrt{2\pi}(\sigma^2+2\theta^2)} \left(e \cdot \exp\left(-\frac{e^2}{2\frac{\sigma^2\theta^2}{\sigma^2+2\theta^2}}\right) \right) \Bigg|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} \exp\left(-\frac{e^2}{2\frac{\sigma^2\theta^2}{\sigma^2+2\theta^2}}\right) de \\
&= -\frac{\sigma^2\theta}{\sqrt{2\pi}(\sigma^2+2\theta^2)} \left(-\frac{\sqrt{2\pi}\sigma\theta}{\sqrt{\sigma^2+2\theta^2}} \right) \\
&= \frac{\sigma^3\theta^2}{(\sigma^2+2\theta^2)\sqrt{\sigma^2+2\theta^2}}
\end{aligned} \tag{A30}$$

References

- Lebedev, M.A.; Nicolelis, M.A.L. Brain-machine interfaces: past, present and future. *Trends Neurosci.* **2006**, *29*, 536–546. [[CrossRef](#)] [[PubMed](#)]
- Taylor, D.M.; Tillery, S.I.H.; Schwartz, A.B. Direct Cortical Control of 3D Neuroprosthetic Devices. *Science* **2002**, *296*, 1829–1832. [[CrossRef](#)] [[PubMed](#)]
- O’Doherty, J.E.; Lebedev, M.A.; Hanson, T.L.; Fitzsimmons, N.A.; Nicolelis, M.A.L. A brain-machine interface instructed by direct intracortical microstimulation. *Front. Integr. Neurosci.* **2009**, *3*, 20.
- Orsborn, A.L.; Moorman, H.G.; Overduin, S.A.; Shanechi, M.M.; Dimitrov, D.F.; Carmena, J.M. Closed-Loop Decoder Adaptation Shapes Neural Plasticity for Skillful Neuroprosthetic Control. *Neuron* **2014**, *82*, 1380–1393. [[CrossRef](#)] [[PubMed](#)]
- Vyas, S.; Even-Chen, N.; Stavisky, S.D.; Ryu, S.I.; Nuyujukian, P.; Shenoy, K.V. Neural Population Dynamics Underlying Motor Learning Transfer. *Neuron* **2018**, *97*, 1177–1186. [[CrossRef](#)] [[PubMed](#)]
- Wolpaw, J.R.; Birbaumer, N.; Heetderks, W.J.; McFarland, D.J.; Peckham, P.H.; Schalk, G.; Donchin, E.; Quatrano, L.A.; Robinson, C.J.; Vaughan, T.M. Brain-computer interface technology: a review of the first international meeting. *IEEE Trans. Rehabil. Eng.* **2000**, *8*, 164–173. [[CrossRef](#)] [[PubMed](#)]
- Nicolelis, M.A.L. Brain-machine interfaces to restore motor function and probe neural circuits. *Nat. Rev. Neurosci.* **2003**, *4*, 417–422. [[CrossRef](#)]
- Churchland, M.M.; Cunningham, J.P.; Kaufman, M.T.; Foster, J.D.; Nuyujukian, P.; Ryu, S.I.; Shenoy, K.V. Neural population dynamics during reaching. *Nature* **2012**, *487*, 51–56. [[CrossRef](#)]
- Musallam, S.; Corneil, B.D.; Greger, B.; Scherberger, H.; Andersen, R.A. Cognitive Control Signals for Neural Prosthetics. *Science* **2004**, *305*, 258–262. [[CrossRef](#)] [[PubMed](#)]
- Velliste, M.; Perel, S.; Spalding, M.C.; Whitford, A.S.; Schwartz, A.B. Cortical control of a prosthetic arm for self-feeding. *Nature* **2008**, *453*, 1098–1101. [[CrossRef](#)] [[PubMed](#)]
- Hochberg, L.R.; Serruya, M.D.; Friehs, G.M.; Mukand, J.A.; Saleh, M.; Caplan, A.H.; Branner, A.; Chen, D.; Penn, R.D.; Donoghue, J.P. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* **2006**, *442*, 164–171. [[CrossRef](#)] [[PubMed](#)]
- Hochberg, L.R.; Bacher, D.; Jarosiewicz, B.; Masse, N.Y.; Simeral, J.D.; Vogel, J.; Haddadin, S.; Liu, J.; Cash, S.S.; van der Smagt, P.; et al. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature* **2012**, *485*, 372–375. [[CrossRef](#)] [[PubMed](#)]
- Truccolo, W.; Friehs, G.M.; Donoghue, J.P.; Hochberg, L.R. Primary Motor Cortex Tuning to Intended Movement Kinematics in Humans with Tetraplegia. *J. Neurosci.* **2008**, *28*, 1163–1178. [[CrossRef](#)] [[PubMed](#)]
- Moritz, C.T.; Perlmutter, S.I.; Fetz, E.E. Direct control of paralysed muscles by cortical neurons. *Nature* **2008**, *456*, 639–642. [[CrossRef](#)]
- Collinger, J.L.; Wodlinger, B.; Downey, J.E.; Wang, W.; Tyler-Kabara, E.C.; Weber, D.J.; McMorland, A.J.C.; Velliste, M.; Boninger, L.M.; Schwartz, A.B. High-performance neuroprosthetic control by an individual with tetraplegia. *Lancet* **2013**, *381*, 557–564. [[CrossRef](#)]

16. Gilja, V.; Pandarinath, C.; Blabe, C.H.; Nuyujukian, P.; Simeral, J.D.; Sarma, A.A.; Sorice, B.L.; Perge, J.A.; Jarosiewicz, B.; Hochberg, L.R.; et al. Clinical translation of a high-performance neural prosthesis. *Nat. Med.* **2015**, *21*, 1142–1145. [[CrossRef](#)]
17. Bouton, C.E.; Shaikhouni, A.; Annetta, N.V.; Bockbrader, M.A.; Friedenberg, D.A.; Nielson, D.M.; Sharma, G.; Sederberg, P.B.; Glenn, B.C.; Mysiw, W.J.; et al. Restoring cortical control of functional movement in a human with quadriplegia. *Nat. Med.* **2016**, *533*, 247–250. [[CrossRef](#)]
18. Ajiboye, A.B.; Willett, F.R.; Young, D.R.; Memberg, W.D.; Murphy, B.A.; Miller, J.P.; Walter, B.L.; Sweet, J.A.; Hoyen, H.A.; Keith, M.W.; et al. Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with tetraplegia: a proof-of-concept demonstration. *Lancet* **2017**, *389*, 1821–1830. [[CrossRef](#)]
19. Brandman, D.M.; Hosman, T.; Saab, J.; Burkhart, M.C.; Shanahan, B.E.; Ciancibello, J.G.; Sarma, A.A.; Milstein, D.J.; Vargas-Irwin, C.E.; Franco, B.; et al. Rapid calibration of an intracortical brain–computer interface for people with tetraplegia. *J. Neural Eng.* **2018**, *15*, 1–14. [[CrossRef](#)]
20. Gilja, V.; Nuyujukian, P.; Chestek, C.A.; Cunningham, J.P.; Yu, B.M.; Fan, J.M.; Churchland, M.M.; Kaufman, M.T.; Kao, J.C.; Ryu, S.I.; et al. A high-performance neural prosthesis enabled by control algorithm design. *Nat. Neurosci.* **2012**, *15*, 1752–1757. [[CrossRef](#)]
21. Malik, W.Q.; Truccolo, W.; Brown, E.N.; Hochberg, L.R. Efficient Decoding With Steady-State Kalman Filter in Neural Interface Systems. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2011**, *19*, 25–34. [[CrossRef](#)]
22. Homer, M.L.; Perge, J.A.; Black, M.J.; Harrison, M.T.; Cash, S.S.; Hochberg, L.R. Adaptive Offset Correction for Intracortical Brain-Computer Interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2014**, *22*, 239–248. [[CrossRef](#)]
23. Wu, W.; Gao, Y.; Bienenstock, E.; Donoghue, J.P.; Black, M.J. Bayesian Population Decoding of Motor Cortical Activity Using a Kalman Filter. *Neural Comput.* **2006**, *18*, 80–118. [[CrossRef](#)]
24. Thayer, J.F. On the importance of inhibition: central and peripheral manifestations of nonlinear inhibitory processes in neural systems. *Dose-Response* **2006**, *4*, 2–21. [[CrossRef](#)]
25. Yang, Y.; Dewald, J.P.A.; van der Helm, F.C.T.; Schouten, A.C. Unveiling neural coupling within the sensorimotor system: Directionality and nonlinearity. *Eur. J. Neurosci.* **2018**, *48*, 2407–2415. [[CrossRef](#)]
26. Anderson, B.; Moore, J. *Optimal Filtering*; Prentice-Hall: New York, NY, USA, 1979.
27. Julier, S.; Uhlmann, J.; Durrant-Whyte, H.F. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Trans. Autom. Control.* **2000**, *45*, 477–482. [[CrossRef](#)]
28. Li, Z.; O’Doherty, J.E.; Hanson, T.L.; Lebedev, M.A.; Henriquez, C.S.; Nicolelis, M.A.L. Unscented Kalman Filter for Brain-Machine Interfaces. *PLoS ONE* **2009**, *54*, e6243. [[CrossRef](#)]
29. Arasaratnam, I.; Haykin, S. Cubature Kalman filters. *IEEE Trans. Autom. Control.* **2009**, *54*, 1254–1269. [[CrossRef](#)]
30. Truccolo, W.; Eden, U.T.; Fellows, M.R.; Donoghue, J.P.; Brown, E.N. A Point Process Framework for Relating Neural Spiking Activity to Spiking History, Neural Ensemble, and Extrinsic Covariate Effects. *J. Neurophysiol.* **2005**, *93*, 1074–1089. [[CrossRef](#)]
31. Qian, C.; Sun, X.; Zhang, S.; Xing, D.; Li, H.; Zheng, X.; Pan, G.; Wang, Y. Nonlinear Modeling of Neural Interaction for Spike Prediction Using the Staged Point-Process Model. *Neural Comput.* **2018**, *30*, 3189–3226. [[CrossRef](#)]
32. Qian, C.; Sun, X.; Wang, Y.; Zheng, X.; Wang, Y.; Pan, G. Binless Kernel Machine: Modeling Spike Train Transformation for Cognitive Neural Prostheses. *Neural Comput.* **2018**, *32*, 1863–1900. [[CrossRef](#)] [[PubMed](#)]
33. Ando, T.; Konishi, S. Neural Network Nonlinear Regression Modeling and Information Criteria. In Proceedings of the Advances in Statistics, Combinatorics & Related Areas-selected Papers from the Scra-fim Viii-the Wollongong Conference, Wollongong, Australia, 19–21 December 2002; pp. 11–22.
34. Kim, S.P.; Sanchez, J.C.; Rao, Y.N.; Erdogmus, D.; Carmena, J.M.; Lebedev, M.A.; Nicolelis, M.A.L.; Principe, J.C. A comparison of optimal MIMO linear and nonlinear models for brain–machine interfaces. *J. Neural Eng.* **2006**, *3*, 145–161. [[CrossRef](#)] [[PubMed](#)]
35. Tagliabue, M.; Francis, N.; Hao, Y.; Duret, M.; Brochier, T.; Riehle, A.; Maier, M.A.; Eskiizmirli, S. Estimation of two-digit grip type and grip force level by frequency decoding of motor cortex activity for a BMI application. In Proceedings of the International Conference on Advanced Robotis (ICAR), Istanbul, Turkey, 27–31 July 2015; pp. 308–315.
36. Sadiq, M.T.; Yu, X.; Yuan, Z.; Zeming, F.; Rehman, A.U.; Ullah, I.; Li, G.; Xiao, G. Motor Imagery EEG Signals Decoding by Multivariate Empirical Wavelet Transform-Based Framework for Robust Brain–Computer Interfaces. *IEEE Access* **2019**, *7*, 171431–171451. [[CrossRef](#)]
37. Willett, F.R.; Avansino, D.T.; Hochberg, L.R.; Henderson, J.M.; Shenoy, K.V. High-performance brain-to-text communication via handwriting. *Nature* **2021**, *593*, 249–254. [[CrossRef](#)]
38. Wang, P.; Jiang, A.; Liu, X.; Shang, J.; Zhang, L. LSTM-Based EEG Classification in Motor Imagery Tasks. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 2086–2095. [[CrossRef](#)]
39. Asgharpour, M.; Foodeh, R.; Daliri, M.R. Regularized Kalman filter for brain-computer interfaces using local field potential signals. *J. Neurosci. Methods* **2021**, *350*, 109022. [[CrossRef](#)]
40. Irwin, Z.T.; Schroeder, K.E.; Vu, P.P.; Bullard, A.J.; Tat, D.M.; Nu, C.S.; Vaskov, A.; Nason, S.R.; Thompson, D.E.; Bentley, J.N.; et al. Neural Control of finger movement via intracortical brain-machine interface. *J. Neural Eng.* **2017**, *14*, 066004. [[CrossRef](#)]
41. Simon, D. *Optimal State Estimation: Kalman, H_∞ and Nonlinear Approaches*; Wiley: Hoboken, NJ, USA, 2006.
42. Schick, I.; Mitter, S.K. Robust recursive estimation in the presence of heavy-tailed observation noise. *Ann. Stat.* **1994**, *22*, 1045–1080. [[CrossRef](#)]
43. Principe, J.C. *Information Theoretic Learning: Renyi’s Entropy and Kernel Perspectives*; Springer: New York, NY, USA, 2010.

44. Liu, W.; Pokharel, P.P.; Principe, J.C. Correntropy: Properties and applications in non-Gaussian signal processing. *IEEE Trans. Signal Process.* **2007**, *55*, 5286–5298. [[CrossRef](#)]
45. Santamaria, I.; Pokharel, P.P.; Principe, J.C. Generalized correlation function: definition, properties, and application to blind equalization. *IEEE Trans. Signal Process.* **2006**, *54*, 2187–2197. [[CrossRef](#)]
46. Singh, A.; Principe, J.C. Using correntropy as a cost function in linear adaptive filters. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Atlanta, GA, USA, 14–19 June 2009; pp. 2950–2955.
47. Shi, L.; Lin, Y. Convex combination of adaptive filters under the maximum correntropy criterion in impulsive interference. *IEEE Signal Process. Lett.* **2014**, *21*, 1385–1388. [[CrossRef](#)]
48. Liu, X.; Qu, H.; Zhao, J.; Chen, B. State space maximum correntropy filter. *Signal Process.* **2017**, *130*, 152–158. [[CrossRef](#)]
49. He, R.; Hu, B.G.; Zheng, W.S.; Kong, X.W. Robust principal component analysis based on maximum correntropy criterion. *IEEE Trans. Image Process.* **2011**, *20*, 1485–1494.
50. He, R.; Zheng, W.S.; Hu, B.G. Maximum correntropy criterion for robust face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 1561–1576.
51. Du, B.; Tang, X.; Wang, Z.; Zhang, L.; Tao, D. Robust Graph-Based Semisupervised Learning for Noisy Labeled Data via Maximum Correntropy Criterion. *IEEE Trans. Cybern.* **2019**, *49*, 1440–1453. [[CrossRef](#)]
52. Mandanas, F.D.; Kotropoulos, C.L. Robust Multidimensional Scaling Using a Maximum Correntropy Criterion. *IEEE Trans. Signal Process.* **2017**, *65*, 919–932. [[CrossRef](#)]
53. Bretscher, O. *Linear Algebra with Applications: Fourth Edition*; Pearson Prentice Hall: Hoboken, NJ, USA, 2009.
54. Singh, A.; Principe, J.C. A closed form recursive solution for maximum correntropy training. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Dallas, TX, USA, 15–19 March 2010; pp. 2070–2073.
55. Chen, B.; Liu, X.; Zhao, H.; Principe, J.C. Maximum correntropy Kalman filter. *Automatica* **2017**, *76*, 70–77. [[CrossRef](#)]
56. Hampel, F.R.; Ronchetti, E.M.; Rousseeuw, P.J.; Stahel, W.A. *Robust Statistics: The Approach Based on Influence Functions*; John Wiley & Sons: New York, NY, USA, 1986.
57. Huber, P.J.; Ronchetti, E.M. *Robust Statistics*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2009.
58. Protter, M.H.; Morrey, C.B. *A First Course in Real Analysis*; Springer: New York, NY, USA, 1977.
59. Taylor, A.E. L'Hospital's Rule. *Am. Math. Mon.* **1952**, *59*, 20–24. [[CrossRef](#)]
60. Agarwal, R.P.; Meehan, M.; Regan, D.O. *Fixed Point Theory and Applications*; Cambridge University Press: Cambridge, UK, 2001.
61. Chen, B.; Wang, J.; Zhao, H.; Zheng, N.; Principe, J.C. Convergence of a fixed-point algorithm under maximum correntropy criterion. *IEEE Signal Process. Lett.* **2015**, *22*, 1723–1727. [[CrossRef](#)]
62. Ortega, J.R.; Rheinboldt, W.C. *Iterative Solution of Nonlinear Equations in Several Variables*; Academic Press: London, UK, 1970.
63. Liu, X.; Shen, X.; Chen, S.; Zhang, X.; Huang, Y.; Wang, Y.; Wang, Y. Hierarchical Dynamical Model for Multiple Cortical Neural Decoding. *Neural Comput.* **2021**, *33*, 1372–1401. [[CrossRef](#)]
64. Carmena, J.M.; Lebedev, M.A.; Crist, R.E.; O'Doherty, J.E.; Santucci, D.M.; Dimitrov, D.F.; Patil, P.G.; Henriquez, C.S.; Nicolelis, M.A.L. Learning to Control a Brain–Machine Interface for Reaching and Grasping by Primates. *PLoS Biol.* **2003**, *1*, e42. [[CrossRef](#)]
65. She, X.; Liao, Y.; Li, H.; Zhang, Q.; Wang, Y.; Zheng, X. Clustering and observation on neuron tuning property for brain machine interfaces. In Proceedings of the International Conference on Multisensor Fusion and Information Integration for Intelligent Systems (MFI), Beijing, China, 28–29 September 2014; pp. 1–8.
66. Zhang, Z.; Chen, S.; Yang, Z.; Wang, Y. Tracking the Time Varying Neural Tuning via Adam on Point Process Observations. In Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018; pp. 195–198.
67. Kingma, D.P.; Ba, J.L. Adam: A Method for Stochastic Optimization. In Proceedings of the International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015; pp. 1–15.
68. DiGiovanna, J.; Mahmoudi, B.; Fortes, J.; Principe, J.C.; Sanchez, J.C. Coadaptive Brain-Machine Interface via Reinforcement Learning. *IEEE Trans. Biomed. Eng.* **2009**, *56*, 54–64. [[CrossRef](#)] [[PubMed](#)]
69. Arduin, P.; Fregnac, Y.; Shulz, D.E.; Ego-Stengel, V. “Master” Neurons Induced by Operant Conditioning in Motor Cortex during a Brain-Machine Interface Task. *J. Neurosci.* **2013**, *33*, 8308–8320. [[CrossRef](#)]
70. Boi, F.; Moraitis, T.; Feo, V.D.; Diotalevi, F.; Bartolozzi, C.; Indiveri, G.; Vato, A. A Bidirectional Brain-Machine Interface Featuring a Neuromorphic Hardware Decoder. *Frontiers Neurosci.* **2016**, *10*, 563. [[CrossRef](#)] [[PubMed](#)]
71. Simeral, J.D.; Kim, S.P.; Black, M.J.; Donoghue, J.P.; Hochberg, L.R. Neural control of cursor trajectory and click by a human with tetraplegia 1000 days after implant of an intracortical microelectrode array. *J. Neural Eng.* **2011**, *8*, 025027. [[CrossRef](#)] [[PubMed](#)]
72. Rao, C. *Linear Statistical Inference and its Applications*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 1973.