Nonlinear System Identification Approach under Noisy Input Signals and Impulse Observed Noise by Kernel Adaptive Filtering Algorithm

Ying-Ren Chien* and En-Ting Lin†
* National Taipei University of Technology, Taipei, Taiwan
E-mail: yrchien@ntut.edu.tw
† National Ilan University, Yilan, Taiwan
E-mail: r1241001@niu.edu.tw

***Abstract*—Kernel adaptive filters (KAFs) have garnered significant attention for their efficacy in nonlinear function approximation within the framework of Reproducing Kernel Hilbert Space (RKHS). However, a critical limitation of KAFs lies in their susceptibility to bias induced by input noise and impulse noise at the observation side. To address these issues, we propose a bias-compensated kernel maximum correntropy (BC-KMC) algorithm, which integrates bias compensation mechanisms into KMC to further mitigate noise-induced bias. It is expected that BC-KMC consistently outperforms conventional bias-compensated kernel least mean square (BC-KLMS) in scenarios characterized by low signal-to-noise ratios (SNR) and strong impulsive noise, thereby underscoring its effectiveness in challenging environments.**

***Index Terms*—Nonlinear system identification, kernel adaptive filter algorithm, maximum correntropy criterion, bias compensation, impulse noise robustness.**

I. Introduction

Kernel methods have become indispensable tools for nonlinear system identification, owing to their robust learning capabilities. In the field of signal processing, kernel adaptive filters (KAFs) utilize the reproducing kernel Hilbert space (RKHS) framework [1] to enhance the efficacy of adaptive filtering. Prominent examples include the Kernel Least Mean Square (KLMS) algorithm [2], which facilitates efficient sequential learning.

To enhance robustness, the Kernel Maximum Correntropy (KMC) algorithm [3] integrates KLMS with the Maximum Correntropy Criterion (MCC), mapping input data into RKHS while minimizing discrepancies between the desired output and filter response. This improves nonlinear system modeling accuracy and strengthens resilience against various noise types. Despite their advantages, KAFs remain susceptible to input noise, which can stem from environmental interference, sensor inaccuracies, or system imperfections. Recently, a bias-compensated method for KLMS (BC-KLMS) algorithm has been presented in [4]. However, the impulsive measurement noise may severely degrade the performance of the compensation method. To our knowledge, the system identification problem that accounts for both input noise and impulsive observation noise in KAFs has not been systematically examined.

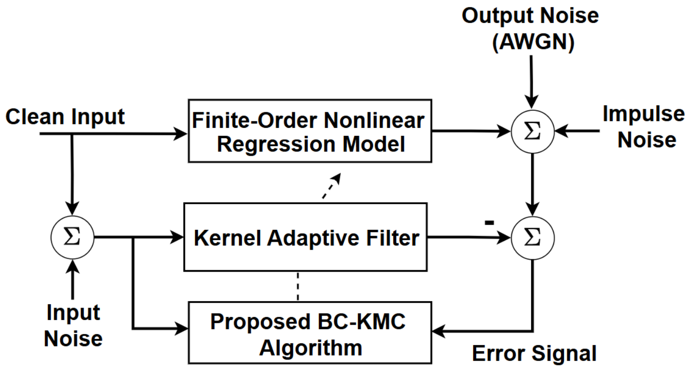

Fig. 1. Block diagram of the nonlinear system identification problem under impulse noise.

II. Key Ideas Proposed: Bias Compensation Kernel Maximum Correntropy Algorithm

A. System Model

Figure 1 illustrates the block diagram for nonlinear system identification. To represent the nonlinear regression function $\psi : \mathbb{R}^O \to \mathbb{R}$, we adopt a finite-order model as follows:

$$y_n = \psi(\mathbf{u}_n; \boldsymbol{\omega}) = \sum_{m=1}^{M} \omega(m)\kappa(\mathbf{u}_n, \mathbf{u}_\omega(m)), \qquad (1)$$

where M is the size of the dictionary $\mathcal{D}$, consisting of stored vectors represented as $\mathcal{D} = \{\mathbf{u}_\omega(1), \ldots, \mathbf{u}_\omega(M)\}$, The input $\mathbf{u}_n$ is an O-dimensionalvector with its o-th element denoted as u_o. Using the Gaussian kernel, the kernelized input for the m-th dictionary element is given by:

$$\kappa\left(\mathbf{u}_n, \mathbf{u}_\omega(m)\right) = \exp\left(-\frac{\|\mathbf{u}_n - \mathbf{u}_\omega(m)\|^2}{2\sigma^2}\right), \qquad (2)$$

where $\sigma > 0$ is the kernel bandwidth. Defining the unknown weight vector as $\boldsymbol{\omega} = \{\omega(1), \ldots, \omega(M)\}$, we rewrite Eq.(1) in a compact matrix form:

$$y_n = \psi(\mathbf{u}_n; \boldsymbol{\omega}) = \boldsymbol{\omega}^\top \kappa_\omega(\mathbf{u}_n), \qquad (3)$$

where $\kappa_\omega(\mathbf{u}_n) \in \mathbb{R}^{M\times 1}$ represents the kernelized input vector, with each element corresponding to a kernel evaluation. The system's observable output is modeled as:

$$d_n = y_n + \nu_n, \tag{4}$$

where ν_n is the observation noise. The output signal-to-noise ratio (SNR) is defined as $\text{SNR}_o = \mathbb{E}\left[y_n^2\right]/\mathbb{E}\left[\nu_n^2\right]$. Similarly, the input noise ζ_n follows an additive white Gaussian noise (AWGN) model with zero mean and variance σ_ζ^2. The input SNR is given by $\text{SNR}_i = \mathbb{E}\left[\overline{u}_n^2\right]/\mathbb{E}\left[\zeta_n^2\right]$.

For the case where noisy inputs are present, i.e., $\overline{u}_n = u_n + \zeta_n$, the KMC filter's output is expressed as:

$$\overline{y}_n = \psi(\overline{\mathbf{u}}_n; \hat{\boldsymbol{\omega}}_n) = \hat{\boldsymbol{\omega}}_n^\top \kappa_\omega(\overline{\mathbf{u}}_n), \tag{5}$$

where $\hat{\boldsymbol{\omega}}_n$ represents the weight vector used by the KMC filter. The standard KMC algorithm updates this weight vector iteratively using the following rule:

$$\hat{\boldsymbol{\omega}}_{n+1} = \hat{\boldsymbol{\omega}}_n + \eta \exp\left(\frac{-\overline{e}_n^2}{2\sigma_{mcc}^2}\right)\overline{e}_n \kappa_\omega(\overline{\mathbf{u}}_n), \tag{6}$$

where η denotes the learning rate and the error signal under noisy conditions is given by $\overline{e}_n = d_n - \overline{y}_n$.

B. Proposed BC-KMC algorithm

In this section, we use the bias introduced by the presence of additive input noise ζ_n

Assuming that the weight vector of the adaptive filter satisfies $\hat{\boldsymbol{\omega}}_n = \boldsymbol{\omega}$, we employ a Taylor series expansion of the nonlinear regression function around ζ_n as follows :

$$\psi(\mathbf{u} + \boldsymbol{\eta}) \approx \psi(\mathbf{u}) + \boldsymbol{\eta}^T \nabla\psi(\mathbf{u}) + \frac{1}{2}\boldsymbol{\eta}^T \mathbf{H}(\mathbf{u})\boldsymbol{\eta}, \tag{7}$$

where $\nabla\psi(\mathbf{u})$ represents the gradient of $\psi(\overline{\mathbf{u}})$ at $\mathbf{u}$ and $\mathbf{H}(\mathbf{u})$ denotes the Hessian matrix of $\psi(\overline{\mathbf{u}})$, given by:

$$\mathbf{H}(\mathbf{u}) = \frac{\partial^2\psi(\mathbf{u})}{\partial u_i \partial u_j}, i, j = 1, \ldots, O. \tag{8}$$

If the additive input noise is assumed to be independent and Gaussian-distributed, the expectation of the bias term can be written as:

$$\mathbb{E}\left[d_n - \overline{y}_n^*\right] = \sum_{o=1}^{O} \frac{-\sigma_{\eta_o}^2}{2} \frac{\partial^2}{\partial u_o^2}\psi(\overline{\mathbf{u}}_n), \tag{9}$$

where $\overline{y}_n^* := \psi(\overline{\mathbf{u}}_n; \boldsymbol{\omega})$. Based on this observation, we propose incorporating the bias term into the conventional MSE cost function, resulting in the following modified expression:

$$J(\hat{\boldsymbol{\omega}}) = \mathbb{E}\left\{\exp\left(-\frac{\overline{e}_n^2 - \gamma\left[\sum_{o=1}^{O} \frac{-\sigma_{\eta_o}^2}{2}\frac{\partial^2}{\partial u_o^2}\psi(\overline{\mathbf{u}}_n)\right]^2}{2\sigma_{mcc}^2}\right)\right\}, \tag{10}$$

Where $\gamma > 0$ serves as a regularization parameter, balancing the trade-off between fitting accuracy on the training data and the algorithm's resilience to input noise. Defining:

$$\beta_n = \exp\left(-\frac{\overline{e}_n^2 - \gamma\left[\sum_{o=1}^{O} \frac{-\sigma_{\eta_o}^2}{2}\frac{\partial^2}{\partial u_o^2}\psi(\overline{\mathbf{u}}_n)\right]^2}{2\sigma_{mcc}^2}\right). \tag{11}$$

Applying the steepest descent method, the gradient of $J(\hat{\boldsymbol{\omega}})$ with respect to $\hat{\boldsymbol{\omega}}$ is derived as:

$$\begin{aligned}\nabla_{\hat{\boldsymbol{\omega}}} J \approx \mathbb{E}\Bigg\{\frac{\beta_n}{\sigma_{mcc}^2}\Bigg(\overline{e}_n\kappa_\omega(\mathbf{u}_n) + \gamma\left[\sum_{o=1}^{O} -\sigma_{\eta_o}^2 \frac{\partial^2}{\partial u_o^2}\psi(\overline{\mathbf{u}}_n)\right] \\ \times \left[\nabla_{\hat{\boldsymbol{\omega}}}\left(\sum_{o=1}^{O}\frac{-\sigma_{\eta_o}^2}{2}\frac{\partial^2}{\partial u_o^2}\psi(\overline{\mathbf{u}}_n)\right)\right]\Bigg)\Bigg\}.\end{aligned} \tag{12}$$

For an input vector of dimension O, the Hessian matrix $\mathbf{H}(\overline{\mathbf{u}}_n)$ can be reduced to a_n as follows :

$$\begin{aligned}a_{o,n} = \sum_{m=1}^{M} \omega(m) \exp\left\{\frac{-\|\overline{\mathbf{u}}_n - \mathbf{u}_\omega(m)\|_2^2}{2\sigma^2}\right\} \\ \times \left[\frac{\left[\overline{u}_{o,n} - u_{\omega,o}(m)\right]^2}{\sigma^4} - \frac{1}{\sigma^2}\right].\end{aligned} \tag{13}$$

Additionally,

$$\nabla_{\hat{\boldsymbol{\omega}}}\left(\frac{-\sigma_{\eta_o}^2}{2}\frac{\partial^2}{\partial u_o^2}\psi(\overline{\mathbf{u}}_n)\right)\Bigg|_{\overline{\mathbf{u}}_n} = \frac{\sigma_{\eta_o}^2}{2}\mathbf{b}_{o,n}, \tag{14}$$

where $\mathbf{b}_{o,n} \in \mathbb{R}^{M\times 1}$ is defined with its m-th element given by:

$$b_{o,n}^m = \exp\left\{\frac{-\|\overline{\mathbf{u}}_n - \mathbf{u}_\omega(m)\|_2^2}{2\sigma^2}\right\}\left[\frac{\left[\overline{u}_{o,n} - u_{\omega,o}(m)\right]^2 - \sigma^2}{\sigma^4}\right]. \tag{15}$$

It is important to note that $\mathbf{u}_\omega(m) = \left[u_{\omega,1}(m), \ldots, u_{\omega,O}(m)\right]^\top$ represents the m-th dictionary vector element. The final expression for the proposed BC-KMC algorithm is given by:

$$\begin{aligned}\boldsymbol{\omega}_{n+1} &= \boldsymbol{\omega}_n \\ &+ \mu\beta_n\left(\overline{e}_n\kappa_\omega(\overline{\mathbf{u}}_n) + \gamma\sum_{o=1}^{O}\frac{\sigma_{\eta_o}^2}{2}a_{o,n}\sum_{o=1}^{O}\frac{\sigma_{\eta_o}^2}{2}\mathbf{b}_{o,n}\right).\end{aligned} \tag{16}$$

REFERENCES

[1] W. Liu, J. C. Principe, and S. Haykin, *Kernel Adaptive Filtering: A Comprehensive Introduction*, 1st ed. Wiley, 2010.

[2] W. Liu, P. P. Pokharel, and J. C. Principe, "The kernel least-mean-square algorithm," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 543–554, 2008.

[3] S. Zhao, B. Chen, and J. C. Príncipe, "Kernel adaptive filtering with maximum correntropy criterion," in *Int. Jt. Conf. Neural Netw. (IJCNN)*, 2011, pp. 2012–2017.

[4] Y.-R. Chien, J.-L. Liu, E.-T. Lin, and G. Qian, "Bias compensation for kernel least-mean-square algorithms," *IEEE Sensors Lett.*, vol. 9, no. 4, pp. 1–4, 2025.