Received: 17 July 2020 | Revised: 15 November 2020 | Accepted: 4 January 2021

DOI: 10.1049/cth2.12097

IET Control Theory & Applications

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH PAPER

The robust multi-innovation estimation algorithm for Hammerstein non-linear systems with non-Gaussian noise

Xuehai Wang[1] | **Feng Ding**[2]

[1] School of Mathematics and Statistics, Xinyang Normal University, PR China

[2] College of Automation and Electronic Engineering, Qingdao University of Science and Technology, PR China

Correspondence
Xuehai Wang, School of Mathematics and Statistics, Xinyang Normal University, Xinyang 464000, PR China.
Email: xuehaiwang735@163.com

Funding information
Science and Technology Department of Henan Province, Grant/Award Number: 202102210297; Key Research Project of Henan Higher Education Institutions, Grant/Award Number: 21A413006; Nanhu Scholars Program for Young Scholars of Xinyang Normal University, Grant/Award Number: 2017B67

Abstract

The characteristic of the external noise has significant influences on system modelling and identification, and the assumption that the noise follows the Gaussian distribution may be invalid due to realistic reasons. This paper discusses the identification issue of Hammerstein non-linear systems with non-Gaussian noise and presents a robust gradient algorithm. The algorithm is derived based on the logarithmic cost function of continuous mixed p-norm of prediction errors, which takes into account each p-norm of errors for $1 \leqslant p \leqslant 2$. The gain at each recursive step adapts to the data quality so that the algorithm has good robustness to non-Gaussian noise. To improve the estimation accuracy, a robust multi-innovation gradient algorithm is proposed by using the multi-innovation identification theory. Two examples are provided to exhibit the validity of the proposed algorithms.

1 | INTRODUCTION

System identification is the theory and methods of establishing the mathematical models of dynamical systems [1–6]. Non-linearity is the essential characteristics of industrial processes [7–10]. Although the dynamic behaviours of many physical plants are always modelled as linear systems in the vicinity of a specific operating point, they can be better represented by non-linear models when they demonstrate strong non-linearities or need to be described in the whole operating range [11, 12]. Because of the diversity and complexity of non-linear phenomena, there is no uniform model structures for describing non-linearities up to now. Various models are exploited for different kinds of non-linearities [13–15]. Typical non-linear models include Volterra models, Hammerstein models and Wiener models [16–18].

The Hammerstein model, which is composed of a static non-linear block followed by a dynamic linear block [19, 20], has been applied to many fields, such as chemical processes, fuel cells, battery and biological processes. Li and Zhang presented a maximum likelihood identification scheme for dual-rate Hammerstein non-linear systems based on the polynomial transformation technique [21]. Wang et al. developed an expectation maximization estimation algorithm for Hammerstein systems by maximizing the expectation of the complete measurements [22]. Rahmani and Farrokhi proposed a frequency domain estimation algorithm for fractional-order Hammerstein systems, in which the input non-linearity is modelled by a radial basis function neural network [23]. These works were accomplished by confining that the external disturbances are the Gaussian noises.

Under the assumption of the Gaussian noise, the ℓ_2-norm minimization based identification algorithms, such as the least-squares-based identification algorithms, can obtain optimal estimation performance [24, 25]. However, the Gaussian assumption is sometimes not realistic due to the appearance of abrupt disturbances, signal interferences and human errors, which induces non-Gaussian noise or outliers [26–28]. In such a scenario, the least-squares-based identification algorithms are sensitive to non-Gaussian noise and their performance may deteriorate seriously since the ℓ_2-norm cost function amplifies the errors such that the outliers are likely to dominate all the observations [29].

To reduce the influence of the non-Gaussian noise, various algorithms have been proposed [30–33]. Stojanovic and Nedic modelled the non-Gaussian noise as an ε-contaminated distribution, and proposed a robust recursive algorithm for linear time-varying output-error systems by taking the expectation of least favorable probability density of prediction errors as the cost function [34]. Li and Zhao presented an M-estimate function-based total least mean algorithm for errors-in-variable systems, where a threshold parameter is designed to control the suppression of the impulsive noise [35]. Liu and Yang applied the expectation-maximization algorithm to the identification of a non-linear state-space model, in which Student's t-distribution is used to describe the non-Gaussian noises with outliers [36]. In these works, the external disturbance is assumed to follow a given distribution in advance.

This paper studies the identification problem of the Hammerstein non-linear system with non-Gaussian noise. The difficulties are that the considered system not only involves the parameters of the linear and non-linear subsystems, but also is corrupted by non-Gaussian noise without prior distribution knowledge. A robust multi-innovation gradient (RMIG) algorithm is presented based on the logarithm continuous mixed p-norm cost function, which takes into consideration each pth moment of errors for $1 \leqslant p \leqslant 2$. Differently from the ℓ_2-norm minimization-based identification algorithms, minimizing the continuous p-norm cost function of the RMIG algorithm can generate an adjustable gain which can make the correction term of the parameter estimation drop to near zero when the non-Gaussian noise is encountered, thus the negative effect of the non-Gaussian noise can be resisted. The main contributions of this paper are as follows.

- Derive a logarithm continuous mixed p-norm cost function to eliminate the detrimental effect of outliers.
- Present a RMIG algorithm for the Hammerstein non-linear system with non-Gaussian noise.
- The RMIG algorithm is found to be robust for non-Gaussian noise processes due to the effect of the varying gain.

The rest of this paper is organised as follows. Section 2 describes the identification problem and gives the identification model of the Hammerstein non-linear systems with non-Gaussian noise. Sections 3 and 4 derive the robust gradient algorithm and the RMIG algorithm, respectively. Section 5 describes the ℓ_1-norm multi-innovation gradient (ℓ_1-MIG) algorithm for comparison. Section 6 gives the simulation examples to illustrate the effectiveness of the proposed algorithms. Section 7 shows some concluding remarks.

2 | SYSTEM DESCRIPTION AND PROBLEM STATEMENT

Consider the following Hammerstein non-linear system,

$$y(t) = A(z)f[u(t)] + B(z)y(t) + v(t), \tag{1}$$

$$f[u(t)] = \mu_1 f_1[u(t)] + \mu_2 f_2[u(t)] + \cdots + \mu_s f_s[u(t)], \tag{2}$$

where $\{u(t)\}$ is the input of the system, $\{y(t)\}$ is the output of the system, $f[u(t)]$ is the non-linear input which can be represented as the pre-specified non-linear basis functions $f_j[u(t)]$'s with unknown coefficients μ_i's, the polynomials $A(z)$ and $B(z)$ are the functions in the unit backward shift operator z^{-1}:

$$A(z) := a_1 z^{-1} + a_2 z^{-2} + \cdots + a_{n_a} z^{-n_a},$$

$$B(z) := b_1 z^{-1} + b_2 z^{-2} + \cdots + b_{n_b} z^{-n_b}.$$

The measurement noise $v(t)$ is a zero-mean non-Gaussian process. Define the parameter vectors $\boldsymbol{\vartheta}$, $\boldsymbol{a}$, $\boldsymbol{\mu}$ and $\boldsymbol{b}$, and the information matrix/vectors $\boldsymbol{F}(t)$, $\boldsymbol{f}(u(t))$ and $\boldsymbol{\varphi}(t)$ as

$$\boldsymbol{\vartheta} := \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{\mu} \\ \boldsymbol{b} \end{bmatrix} \in \mathbb{R}^{n_0},\ n_0 := n_a + n_b + s,$$

$$\boldsymbol{a} := [a_1, a_2, \dots, a_{n_a}]^{\mathrm{T}} \in \mathbb{R}^{n_a},$$

$$\boldsymbol{\mu} := [\mu_1, \mu_2, \dots, \mu_s]^{\mathrm{T}} \in \mathbb{R}^{s},$$

$$\boldsymbol{b} := [b_1, b_2, \dots, b_{n_b}]^{\mathrm{T}} \in \mathbb{R}^{n_b},$$

$$\boldsymbol{F}(t) := [\boldsymbol{f}(u(t-1)), \boldsymbol{f}(u(t-2)), \dots, \boldsymbol{f}(u(t-n_a))]^{\mathrm{T}},$$

$$\boldsymbol{f}(u(t)) := [f_1(u(t)), f_2(u(t)), \dots, f_s(u(t))]^{\mathrm{T}} \in \mathbb{R}^{s},$$

$$\boldsymbol{\varphi}(t) := [y(t-1), y(t-2), \dots, y(t-n_b)]^{\mathrm{T}} \in \mathbb{R}^{n_b}.$$

Inserting (2) into (1) gives

$$\begin{aligned} y(t) &= \sum_{i=1}^{n_a} a_i f(u(t-i)) + \sum_{i=1}^{n_b} b_i y(t-i) + v(t) \\ &= \sum_{i=1}^{n_a} a_i \boldsymbol{f}^{\mathrm{T}}(u(t-i))\boldsymbol{\mu} + \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{b} + v(t) \\ &= \boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\mu} + \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{b} + v(t). \end{aligned} \tag{3}$$

Equation (3) is the identification model of the Hammerstein non-linear system in (1) and (2). Many identification methods are derived based on the identification model in (3) [37–45], which is applied in fields [46–51] such as chemical process control systems.

Assume that the orders n_a, n_b and s are known. When $v(t)$ is a Gaussian noise, the existing approaches such as the over-parameterization algorithm and the least squares algorithm can be applied to (3) [52, 53]. However, these ℓ_2-norm minimization-based identification algorithms are sensitive to outliers and have poor performance under the non-Gaussian noise environment. The objective of this paper is to present efficient identification algorithms with good robustness for estimating the parameters $a_i \in \mathbb{R}$, $b_i \in \mathbb{R}$ and $\mu_i \in \mathbb{R}$ from measurements $\{u(t), y(t):\ t = 1, 2, 3, \dots\}$ with non-Gaussian noise $v(t)$.

Remark 1. If the pair $(\boldsymbol{a}, \boldsymbol{\mu})$ is the solution of (3), then so is $(\boldsymbol{a}/\beta, \boldsymbol{\mu}\beta)$ $(\beta \neq 0)$. This means that the solution of (3) is not unique. To have the parameter identifiability, either $\boldsymbol{a}$ or $\boldsymbol{\mu}$ should be normalized. The following normalization constraint on $\boldsymbol{\mu}$ is adopted.

Assumption 1. $\|\boldsymbol{\mu}\| = 1$ and the first non-zero entry of $\boldsymbol{\mu}$ is positive, that is, the ℓ_2-norm of $\boldsymbol{\mu}$ equals one and the first coefficient of the non-linear input $f[u(t)]$ is positive.

3 | ROBUST GRADIENT ALGORITHM

Note that $v(t)$ is a non-Gaussian impulsive noise process. To suppress the influence of the impulsive noise and to provide robust parameter estimation, the following derives the robust gradient algorithm for Hammerstein non-linear systems.

Define the continuous logarithmic mixed p-norm cost function

$$J_1(t) := \int_1^2 \lambda_t(p)\mathrm{E}[\ln(1 + |v(t)|^p)]\mathrm{d}p,$$

where $\lambda_t(p)$ is the probability density-like weighting function with constraint $\int_1^2 \lambda_t(p)\mathrm{d}p = 1$, $\mathrm{E}(\cdot)$ is the expectation operator. Since the logarithm function is a monotonically increasing function, minimizing $J_1(t)$ is equivalent to minimizing $\int_1^2 \lambda_t(p)\mathrm{E}[|v(t)|^p]\mathrm{d}p$, which can be regarded as an infinite weighted summation of each p-norm $|v(t)|^p$ from $p = 1$ to $p = 2$. When $\mathrm{E}[\ln(1 + |v(t)|^p)]$ is approximated by a point estimate $\ln(1 + |v(t)|^p)$, the cost function $J_1(t)$ can be rewritten as

$$J_2(\boldsymbol{a}, \boldsymbol{\mu}, \boldsymbol{b}) := \int_1^2 \lambda_t(p)[\ln(1 + |v(t)|^p]\mathrm{d}p,$$

where $v(t) = y(t) - \boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\mu} - \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{b}$. Taking the gradient of $J_2(\boldsymbol{a}, \boldsymbol{\mu}, \boldsymbol{b})$ with respect to $\boldsymbol{a}$, $\boldsymbol{\mu}$ and $\boldsymbol{b}$ gives

$$\begin{aligned}
\frac{\partial J_2(\boldsymbol{a}, \boldsymbol{\mu}, \boldsymbol{b})}{\partial \boldsymbol{a}} &= \int_1^2 p\lambda_t(p)\frac{|v(t)|^{p-1}}{1 + |v(t)|^p}\frac{v(t)}{|v(t)|}\frac{\partial v(t)}{\partial \boldsymbol{a}}\mathrm{d}p \\
&= -\xi(t)\boldsymbol{F}(t)\boldsymbol{\mu}[y(t) - \boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\mu} - \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{b}], \\
\frac{\partial J_2(\boldsymbol{a}, \boldsymbol{\mu}, \boldsymbol{b})}{\partial \boldsymbol{\mu}} &= -\xi(t)\boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{a}[y(t) - \boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\mu} - \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{b}], \\
\frac{\partial J_2(\boldsymbol{a}, \boldsymbol{\mu}, \boldsymbol{b})}{\partial \boldsymbol{b}} &= -\xi(t)\boldsymbol{\varphi}(t)[y(t) - \boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\mu} - \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{b}],
\end{aligned}$$

$$\begin{aligned}
\mathrm{grad}[J_2(\boldsymbol{a}, \boldsymbol{\mu}, \boldsymbol{b})] &= \begin{bmatrix} \frac{\partial J_2(\boldsymbol{a},\boldsymbol{\mu},\boldsymbol{b})}{\partial \boldsymbol{a}} \\ \frac{\partial J_2(\boldsymbol{a}, \boldsymbol{\mu}, \boldsymbol{b})}{\partial \boldsymbol{\mu}} \\ \frac{\partial J_2(\boldsymbol{a}, \boldsymbol{\mu}, \boldsymbol{b})}{\partial \boldsymbol{b}} \end{bmatrix} \\
&= -\xi(t)\begin{bmatrix} \boldsymbol{F}(t)\boldsymbol{\mu} \\ \boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{a} \\ \boldsymbol{\varphi}(t) \end{bmatrix}[y(t) - \boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t)\boldsymbol{\mu} - \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{b}],
\end{aligned}$$

where

$$\xi(t) := \int_1^2 p\lambda_t(p)\frac{|v(t)|^{p-2}}{1 + |v(t)|^p}\mathrm{d}p.$$

Let $\hat{\boldsymbol{a}}(t)$ be the estimate of $\boldsymbol{a}$ at instant t. Define the generalised information vector $\boldsymbol{\psi}(t)$ and the innovation $e(t)$ as

$$\boldsymbol{\psi}(t) := \begin{bmatrix} \boldsymbol{F}(t)\hat{\boldsymbol{\mu}}(t-1) \\ \boldsymbol{F}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1) \\ \boldsymbol{\varphi}(t) \end{bmatrix} \in \mathbb{R}^{n_0},$$

$$e(t) := y(t) - \hat{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t)\hat{\boldsymbol{\mu}}(t-1) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t-1) \in \mathbb{R}.$$

Using the negative gradient search and minimizing the continuous logarithmic mixed p-norm cost function $J_2(\boldsymbol{a}, \boldsymbol{\mu}, \boldsymbol{b})$ give

$$\begin{aligned}
\begin{bmatrix} \hat{\boldsymbol{a}}(t) \\ \hat{\boldsymbol{\mu}}(t) \\ \hat{\boldsymbol{b}}(t) \end{bmatrix} &= \begin{bmatrix} \hat{\boldsymbol{a}}(t-1) \\ \hat{\boldsymbol{\mu}}(t-1) \\ \hat{\boldsymbol{b}}(t-1) \end{bmatrix} - \rho(t) \\
&\quad \times \mathrm{grad}[J_2(\hat{\boldsymbol{a}}(t-1), \hat{\boldsymbol{\mu}}(t-1), \hat{\boldsymbol{b}}(t-1))] \\
&= \begin{bmatrix} \hat{\boldsymbol{a}}(t-1) \\ \hat{\boldsymbol{\mu}}(t-1) \\ \hat{\boldsymbol{b}}(t-1) \end{bmatrix} + \rho(t)\bar{\xi}(t)\begin{bmatrix} \boldsymbol{F}(t)\hat{\boldsymbol{\mu}}(t-1) \\ \boldsymbol{F}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1) \\ \boldsymbol{\varphi}(t) \end{bmatrix} \\
&\quad \times [y(t) - \hat{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t)\hat{\boldsymbol{\mu}}(t-1) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t-1)] \\
&= \begin{bmatrix} \hat{\boldsymbol{a}}(t-1) + \rho(t)\bar{\xi}(t)\boldsymbol{F}(t)\hat{\boldsymbol{\mu}}(t-1)e(t) \\ \hat{\boldsymbol{\mu}}(t-1) + \rho(t)\bar{\xi}(t)\boldsymbol{F}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)e(t) \\ \hat{\boldsymbol{b}}(t-1) + \rho(t)\bar{\xi}(t)\boldsymbol{\varphi}(t)e(t) \end{bmatrix}. \qquad (4)
\end{aligned}$$

or

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \rho(t)\bar{\xi}(t)\boldsymbol{\psi}(t)e(t), \qquad (5)$$

where $\rho(t) > 0$ is the step size, and

$$\bar{\xi}(t) := \int_1^2 p\lambda_t(p)\frac{|e(t)|^{p-2}}{1 + |e(t)|^p}\mathrm{d}p.$$

When $\lambda_t(p) = \frac{1}{p\ln 2}$, the constraint $\int_1^2 \lambda_t(p)\mathrm{d}p = 1$ is met and $\bar{\xi}(t)$ can be computed by

$$\begin{aligned}
\bar{\xi}(t) &= \frac{1}{\ln 2}\int_1^2 \frac{|e(t)|^{p-2}}{1 + |e(t)|^p}\mathrm{d}p \\
&= \frac{\ln(1 + |e(t)|^2) - \ln(1 + |e(t)|)}{\ln 2 \cdot (\ln|e(t)|) \cdot |e(t)|^2}. \qquad (6)
\end{aligned}$$

Generally speaking, the estimate $\hat{\boldsymbol{\vartheta}}(t)$ approaches the true value of $\boldsymbol{\vartheta}$ as t increases, and the innovation $e(t)$ may be close to zero. To avoid division by zero, Equation (6) can be modified as

$$\hat{\xi}(t) = \frac{\ln(1+|e(t)|^2) - \ln(1+|e(t)|)}{\ln 2 \cdot (\ln(|\tau(t)|) \cdot (|\tau(t)|)^2}, \tag{7}$$

where $\tau^2(t) := e^2(t) + \tau_0$ and τ_0 is a small positive number. The following derives the computation of the optimal step-size $\rho(t)$ by solving the optimization problem

$$\begin{aligned}\min_{\rho\geqslant 0} h(\rho(t)) &:= J_2(\hat{\boldsymbol{a}}(t), \hat{\boldsymbol{\mu}}(t), \hat{\boldsymbol{b}}(t))\\ &= \int_1^2 \lambda_t(p)[\ln(1+|\varepsilon(t)|^p]\mathrm{d}p\end{aligned}$$

by means of one-dimensional search, where

$$\varepsilon(t) := y(t) - \hat{\boldsymbol{a}}^{\mathrm{T}}(t)\boldsymbol{F}(t)\hat{\boldsymbol{\mu}}(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t). \tag{8}$$

Replacing $\bar{\xi}(t)$ in (4) and (5) by $\hat{\xi}(t)$ and inserting (4) into (8) gives

$$\begin{aligned}\varepsilon(t) &= y(t) - [\hat{\boldsymbol{a}}(t-1) + \rho(t)\hat{\xi}(t)\boldsymbol{F}(t)\hat{\boldsymbol{\mu}}(t-1)e(t)]^{\mathrm{T}}\boldsymbol{F}(t)\\ &\quad \times [\hat{\boldsymbol{\mu}}(t-1) + \rho(t)\hat{\xi}(t)\boldsymbol{F}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)e(t)]\\ &\quad - \boldsymbol{\varphi}^{\mathrm{T}}(t)[\hat{\boldsymbol{b}}(t-1) + \rho(t)\hat{\xi}(t)\boldsymbol{\varphi}(t)e(t)]\\ &= y(t) - \hat{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t)\hat{\boldsymbol{\mu}}(t-1) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t-1)\\ &\quad - \rho(t)\hat{\xi}(t)\hat{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t)\boldsymbol{F}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)e(t)\\ &\quad - \rho(t)\hat{\xi}(t)\hat{\boldsymbol{\mu}}^{\mathrm{T}}(t-1)\boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{F}(t)\hat{\boldsymbol{\mu}}(t-1)e(t)\\ &\quad - \rho^2(t)\hat{\xi}^2(t)\hat{\boldsymbol{\mu}}^{\mathrm{T}}(t-1)\boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{F}(t)\boldsymbol{F}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)e^2(t)\\ &\quad - \rho(t)\hat{\xi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t)e(t)\\ &= \{1 - \rho(t)\hat{\xi}(t)[\|\boldsymbol{F}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)\|^2\\ &\quad + \|\boldsymbol{F}(t)\hat{\boldsymbol{\mu}}(t-1)\|^2 + \|\boldsymbol{\varphi}(t)\|^2]\}e(t)\\ &\quad - \rho^2(t)\hat{\xi}^2(t)\hat{\boldsymbol{\mu}}^{\mathrm{T}}(t-1)\boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{F}(t)\boldsymbol{F}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)e^2(t)\\ &= (1 - \rho(t)\hat{\xi}(t)\|\boldsymbol{\psi}(t)\|^2 - \rho^2(t)\hat{\xi}^2(t)k(t))e(t),\end{aligned} \tag{9}$$

where

$$k(t) := \hat{\boldsymbol{\mu}}^{\mathrm{T}}(t-1)\boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{F}(t)\boldsymbol{F}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)e(t).$$

The optimal step-size $\rho(t)$ can be obtained by letting the gradient of the cost function $h(\rho(t))$ with respect to $\rho(t)$ be zero, that is,

$$\begin{aligned}\frac{\partial h(\rho(t))}{\partial \rho(t)} &= \int_1^2 \lambda_t(p)\frac{\partial[\ln(1+|\varepsilon(t)|^p]}{\partial\rho(t)}\mathrm{d}p\\ &= \int_1^2 p\lambda_t(p)\frac{|\varepsilon(t)|^{p-2}}{1+|\varepsilon(t)|^p}\varepsilon(t)\frac{\partial\varepsilon(t)}{\partial\rho(t)}\mathrm{d}p\\ &= \frac{1}{\ln 2}\int_1^2 \frac{|\varepsilon(t)|^{p-2}}{1+|\varepsilon(t)|^p}\mathrm{d}p \cdot \varepsilon(t)\frac{\partial\varepsilon(t)}{\partial\rho(t)}\\ &= \frac{\ln(1+|\varepsilon(t)|^2) - \ln(1+\varepsilon(t))}{\ln 2 \cdot (\ln|\varepsilon(t)|) \cdot |\varepsilon(t)|^2}\\ &\quad \times [1 - \rho(t)\hat{\xi}(t)\|\boldsymbol{\psi}(t)\|^2 - \rho^2(t)\hat{\xi}^2(t)k(t)]\\ &\quad \times [-\hat{\xi}(t)\|\boldsymbol{\psi}(t)\|^2 - 2\rho(t)\hat{\xi}^2(t)k(t)]e^2(t) = 0.\end{aligned}$$

Note that $\bar{\xi}(t) > 0$ (or $\hat{\xi}(t) > 0$), since the terms $\ln(1+|e(t)|^2) - \ln(1+|e(t)|)$ and $\ln(|e(t)|)$ can keep the same sign when $0 < |e(t)| < 1$ and $|e(t)| > 1$. In the case of $k(t) \neq 0$, the optimal $\rho(t)$ can be given by

$$\begin{aligned}\rho(t) &= \frac{\sqrt{\hat{\xi}^2(t)\|\boldsymbol{\psi}(t)\|^4 + 4\hat{\xi}^2(t)k(t)} - \hat{\xi}(t)\|\boldsymbol{\psi}(t)\|^2}{2\hat{\xi}^2(t)k(t)}\\ &= \frac{2}{\hat{\xi}(t)[\sqrt{\|\boldsymbol{\psi}(t)\|^4 + 4k(t)} + \|\boldsymbol{\psi}(t)\|^2]}.\end{aligned} \tag{10}$$

The other solution

$$\rho(t) = -\frac{\|\boldsymbol{\psi}(t)\|^2}{2\hat{\xi}(t)k(t)}$$

is discarded because the step-size $\rho(t)$ should be non-negative but this solution $\rho(t) < 0$ when $k(t) > 0$. Equation (10) is complicated for computing the step-size $\rho(t)$ and can be modified as

$$\rho(t) := \frac{1}{r(t)}, \tag{11}$$

$$r(t) = r(t-1) + \hat{\xi}(t)\|\boldsymbol{\psi}(t)\|^2,\ r(0) = 1. \tag{12}$$

Thus, we can summarise the following recursive relations:

$$\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{\mu}}(t)\\ \hat{\boldsymbol{b}}(t)\end{bmatrix} = \begin{bmatrix}\hat{\boldsymbol{a}}(t-1)\\ \hat{\boldsymbol{\mu}}(t-1)\\ \hat{\boldsymbol{b}}(t-1)\end{bmatrix} + \frac{1}{r(t)}\hat{\xi}(t)\boldsymbol{\psi}(t)e(t), \tag{13}$$

$$r(t) = r(t-1) + \hat{\xi}(t)\|\boldsymbol{\psi}(t)\|^2, \tag{14}$$

$$\boldsymbol{\psi}(t) = \begin{bmatrix}\boldsymbol{F}(t)\hat{\boldsymbol{\mu}}(t-1)\\ \boldsymbol{F}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1)\\ \boldsymbol{\varphi}(t)\end{bmatrix}, \tag{15}$$

$$\hat{\xi}(t) = \frac{\ln(1 + |e(t)|^2) - \ln(1 + |e(t)|)}{\ln 2 \cdot (\ln|\tau(t)|) \cdot |\tau(t)|^2}, \tag{16}$$

$$\tau^2(t) = e^2(t) + \tau_0, \tag{17}$$

$$e(t) = y(t) - \hat{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t)\hat{\boldsymbol{\mu}}(t-1) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t-1), \tag{18}$$

$$\boldsymbol{F}(t) = [\boldsymbol{f}(u(t-1)), \boldsymbol{f}(u(t-2)), \dots, \boldsymbol{f}(u(t-n_a))]^{\mathrm{T}}, \tag{19}$$

$$\boldsymbol{f}(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_s(u(t))]^{\mathrm{T}}, \tag{20}$$

$$\boldsymbol{\varphi}(t) = [y(t-1), y(t-2), \dots, y(t-n_b)]^{\mathrm{T}}. \tag{21}$$

To guarantee the parameter identifiability, the following normalization constraint of $\hat{\boldsymbol{\mu}}(t)$ should be imposed,

$$\bar{\boldsymbol{\mu}}(t) := \mathrm{sgn}[\hat{\mu}_1(t)]\frac{\hat{\boldsymbol{\mu}}(t)}{\|\hat{\boldsymbol{\mu}}(t)\|}, \tag{22}$$

where $\mathrm{sgn}[\hat{\mu}_1(t)]$ represents the sign of the first non-zero entry of the estimate $\hat{\boldsymbol{\mu}}(t)$, and we let $\hat{\boldsymbol{\mu}}(t) := \bar{\boldsymbol{\mu}}(t)$. Equations (13)–(22) construct the robust gradient (RG) algorithm for the Hammerstein non-linear system in (1)–(2).

Remark 2. It can be seen from (16) that the third term $|\tau(t)|^2$ in the denominator of $\hat{\xi}(t)$ plays a dominant role. When the Hammerstein non-linear system encounters outliers, the term $|\tau(t)|^2$ in (16) abruptly increases and the gain $\hat{\xi}(t)$ sharply decreases such that the parameter estimate $\hat{\boldsymbol{\vartheta}}(t)$ in (13) has small changes. It means that the robust gradient algorithm can automatically adjust the gain to resist the influence of non-Gaussian noise.

Remark 3. From (6) and (7), we have

$$\frac{\hat{\xi}(t)}{\bar{\xi}(t)} = \frac{(\ln|e(t)|) \cdot |e(t)|^2}{(\ln(|\tau(t)|) \cdot (|\tau(t)|)^2}. \tag{23}$$

Let $g(x) := x^2 \ln x$. The derivative of $g(x)$ with respect to x is

$$g'(x) = x(2\ln x + 1).$$

Note that $|e(t)| < |\tau(t)|$. When $x = |e(t)| > \frac{1}{\sqrt{\mathrm{e}}}$, $g'(x) > 0$ and $g(x)$ is monotonically increasing and $0 < \hat{\xi}(t) < \bar{\xi}(t)$. When $0 < |e(t)| < \frac{1}{\sqrt{\mathrm{e}}}$, $g'(x) < 0$ and $g(x)$ is monotonically decreasing and $0 < \bar{\xi}(t) < \hat{\xi}(t)$. It indicates from (13) that when the system encounters the non-Gaussian noise and $|e(t)| > \frac{1}{\sqrt{\mathrm{e}}}$, using $\hat{\xi}(t)$ in place of $\bar{\xi}(t)$ can reduce the influence of non-Gaussian noise to parameter estimate $\hat{\boldsymbol{\vartheta}}(t)$. When the system is corrupted by white noise and $0 < |e(t)| < \frac{1}{\sqrt{\mathrm{e}}}$, using $\hat{\xi}(t)$ in place of $\bar{\xi}(t)$ may slightly increase the parameter estimate error. However, if τ_0 is taken as a very small positive number, then $|\tau(t)| \approx |e(t)|$, $\hat{\xi}(t) \approx \bar{\xi}(t)$ and the influence of this approximation on the results is trivial.

Remark 4. The robust gradient algorithm is based on the continuous logarithmic mixed p-norm cost function $J_2(\boldsymbol{a}, \boldsymbol{\mu}, \boldsymbol{b}) = \int_1^2 \lambda_t(p)[\ln(1 + |v(t)|^p]\mathrm{d}p$, which takes into consideration each p-norm of errors for $1 \leqslant p \leqslant 2$ and keeps the merit of the various error p-norms. The continuous changes of the parameter p adapt noisy environments without resorting to a priori knowledge of noise.

Remark 5. From (10) and (11), we have

$$\begin{aligned} r(t) &= \frac{1}{2}\hat{\xi}(t)\left(\sqrt{\|\boldsymbol{\psi}(t)\|^4 + 4k(t)} + \|\boldsymbol{\psi}(t)\|^2\right) \\ &= \hat{\xi}(t)\|\boldsymbol{\psi}(t)\|^2 + \tilde{r}(t), \end{aligned} \tag{24}$$

where $\tilde{r}(t) := \frac{1}{2}\hat{\xi}(t)(\sqrt{\|\boldsymbol{\psi}(t)\|^4 + 4k(t)} - \|\boldsymbol{\psi}(t)\|^2)$. When the parameter estimate $\hat{\boldsymbol{\vartheta}}(t)$ approaches its true parameter $\boldsymbol{\vartheta}$ with t increasing and $e(t) \to 0$, we have $k(t) \to 0$ and $\tilde{r}(t) \to 0$. Thus $r(t) = \hat{\xi}(t)\|\boldsymbol{\psi}(t)\|^2$ and the gain vector $L(t) := \frac{1}{r(t)}\hat{\xi}(t)\boldsymbol{\psi}(t)$ of the correction term in (13) becomes $\frac{\boldsymbol{\psi}(t)}{\|\boldsymbol{\psi}(t)\|^2}$, which does not vanish as t increases and will make $\hat{\boldsymbol{\vartheta}}(t)$ deviate from $\boldsymbol{\vartheta}$. After approximating (24) by (12), replacing t in (12) with $t - j$ $(j = 1, 2, \dots, t-1)$ and successive substitutions give

$$\begin{aligned} r(t) &= \hat{\xi}(t)\|\boldsymbol{\psi}(t)\|^2 + \hat{\xi}(t-1)\|\boldsymbol{\psi}(t-1)\|^2 + r(t-2) \\ &= \sum_{j=0}^{t-1}\hat{\xi}(t-j)\|\boldsymbol{\psi}(t-j)\|^2 + r(0) \\ &= \hat{\xi}(t)\|\boldsymbol{\psi}(t)\|^2 + \bar{r}(t), \end{aligned} \tag{25}$$

where $\bar{r}(t) := \sum_{j=1}^{t-1}\hat{\xi}(t-j)\|\boldsymbol{\psi}(t-j)\|^2 + 1$. Compared with $\tilde{r}(t)$ in (24), the modified $\bar{r}(t)$ in (25) satisfies $\bar{r}(t) > \bar{r}(t-1) \geqslant 1$ and is monotonically increasing. As $t \to \infty$, $\bar{r}(t) \to \infty$ and the gain vector $L(t) \to 0$. Thus the parameter estimation error $\tilde{\boldsymbol{\vartheta}}(t) := \hat{\boldsymbol{\vartheta}}(t) - \boldsymbol{\vartheta}$ in (13) is close to zero and the performance of the algorithm can be guaranteed.

4 | ROBUST MULTI-INNOVATION GRADIENT ALGORITHM

The robust gradient algorithm updates the parameter estimates by using the measurement $\{u(t), y(t)\}$ and the innovation $e(t)$ at current instant. To improve the estimation accuracy by making full use of data information, the following derives the RMIG algorithm.

Consider the measurements from $t-l+1$ to t and define the cost function

$$J_3(\boldsymbol{a},\boldsymbol{\mu},\boldsymbol{b}) := \sum_{j=0}^{l-1}\int_1^2 \lambda_t(p)[\ln(1+|v(t-j)|^p]\mathrm{d}p,$$

where the integer l is the innovation length, and $v(t-j)=y(t-j)-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t-j)\boldsymbol{\mu}-\boldsymbol{\varphi}^{\mathrm{T}}(t-j)\boldsymbol{b}$. Taking the gradient of $J_3(\boldsymbol{a},\boldsymbol{\mu},\boldsymbol{b})$ with respect to $\boldsymbol{a}$, $\boldsymbol{\mu}$ and $\boldsymbol{b}$ gives

$$\begin{aligned}
\frac{\partial J_3(\boldsymbol{a},\boldsymbol{\mu},\boldsymbol{b})}{\partial \boldsymbol{a}} &= \sum_{j=0}^{l-1}\int_1^2 p\lambda_t(p)\frac{|v(t-j)|^{p-1}}{1+|v(t-j)|^p}\\
&\quad\times\frac{v(t-j)}{|v(t-j)|}\frac{\partial v(t-j)}{\partial \boldsymbol{a}}\mathrm{d}p\\
&= -\sum_{j=0}^{l-1}\xi(t-j)\boldsymbol{F}(t-j)\boldsymbol{\mu}\\
&\quad\times[y(t-j)-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t-j)\boldsymbol{\mu}-\boldsymbol{\varphi}^{\mathrm{T}}(t-j)\boldsymbol{b}],\\
\frac{\partial J_3(\boldsymbol{a},\boldsymbol{\mu},\boldsymbol{b})}{\partial \boldsymbol{\mu}} &= -\sum_{j=0}^{l-1}\xi(t-j)\boldsymbol{F}^{\mathrm{T}}(t-j)\boldsymbol{a}\\
&\quad\times[y(t-j)-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t-j)\boldsymbol{\mu}-\boldsymbol{\varphi}^{\mathrm{T}}(t-j)\boldsymbol{b}],\\
\frac{\partial J_3(\boldsymbol{a},\boldsymbol{\mu},\boldsymbol{b})}{\partial \boldsymbol{b}} &= -\sum_{j=0}^{l-1}\xi(t-j)\boldsymbol{\varphi}(t-j)\\
&\quad\times[y(t-j)-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t-j)\boldsymbol{\mu}-\boldsymbol{\varphi}^{\mathrm{T}}(t-j)\boldsymbol{b}],\\
\mathrm{grad}[J_3(\boldsymbol{a},\boldsymbol{\mu},\boldsymbol{b})] &= \begin{bmatrix}\dfrac{\partial J_3(\boldsymbol{a},\boldsymbol{\mu},\boldsymbol{b})}{\partial \boldsymbol{a}}\\ \dfrac{\partial J_3(\boldsymbol{a},\boldsymbol{\mu},\boldsymbol{b})}{\partial \boldsymbol{\mu}}\\ \dfrac{\partial J_3(\boldsymbol{a},\boldsymbol{\mu},\boldsymbol{b})}{\partial \boldsymbol{b}}\end{bmatrix}\\
&= -\sum_{j=0}^{l-1}\xi(t-j)\begin{bmatrix}\boldsymbol{F}(t-j)\boldsymbol{\mu}\\ \boldsymbol{F}^{\mathrm{T}}(t-j)\boldsymbol{a}\\ \boldsymbol{\varphi}(t-j)\end{bmatrix}\\
&\quad\times[y(t-j)-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t-j)\boldsymbol{\mu}-\boldsymbol{\varphi}^{\mathrm{T}}(t-j)\boldsymbol{b}]\\
&= -\sum_{j=0}^{l-1}\xi(t-j)\boldsymbol{\psi}(t-j)[y(t-j)\\
&\quad-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t-j)\boldsymbol{\mu}-\boldsymbol{\varphi}^{\mathrm{T}}(t-j)\boldsymbol{b}],
\end{aligned}$$

where

$$\xi(t-j) := \int_1^2 p\lambda_t(p)\frac{|v(t-j)|^{p-2}}{1+|v(t-j)|^p}\mathrm{d}p.$$

To facilitate the representation of the RMIG algorithm, define the stacked vectors/matrices

$$\boldsymbol{Y}(l,t) := [y(t),y(t-1),\ldots,y(t-l+1)]^{\mathrm{T}}\in\mathbb{R}^l, \tag{26}$$

$$\boldsymbol{\Psi}(l,t) := [\boldsymbol{\psi}(t),\boldsymbol{\psi}(t-1),\ldots,\boldsymbol{\psi}(t-l+1)]\in\mathbb{R}^{n_0\times l}, \tag{27}$$

$$\begin{aligned}\boldsymbol{\Omega}(l,t) := [&\boldsymbol{F}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1),\boldsymbol{F}^{\mathrm{T}}(t-1)\hat{\boldsymbol{a}}(t-1),\ldots,\\ &\boldsymbol{F}^{\mathrm{T}}(t-l+1)\hat{\boldsymbol{a}}(t-1)]\in\mathbb{R}^{s\times l},\end{aligned} \tag{28}$$

$$\boldsymbol{\Phi}(l,t) := [\boldsymbol{\varphi}(t),\boldsymbol{\varphi}(t-1),\ldots,\boldsymbol{\varphi}(t-l+1)]\in\mathbb{R}^{n_b\times l}, \tag{29}$$

$$\begin{aligned}\boldsymbol{E}(l,t) :=& \begin{bmatrix} y(t)-\hat{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t)\hat{\boldsymbol{\mu}}(t-1)\\ y(t-1)-\hat{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t-1)\hat{\boldsymbol{\mu}}(t-1)\\ \vdots\\ y(t-l+1)-\hat{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t-l+1)\hat{\boldsymbol{\mu}}(t-1)\end{bmatrix}\\
&-\begin{bmatrix}\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t-1)\\ \boldsymbol{\varphi}^{\mathrm{T}}(t-1)\hat{\boldsymbol{b}}(t-1)\\ \vdots\\ \boldsymbol{\varphi}^{\mathrm{T}}(t-l+1)\hat{\boldsymbol{b}}(t-1)\end{bmatrix}\\
=&\ \boldsymbol{Y}(l,t)-\boldsymbol{\Omega}^{\mathrm{T}}(l,t)\hat{\boldsymbol{\mu}}(t-1)-\boldsymbol{\Phi}^{\mathrm{T}}(l,t)\hat{\boldsymbol{b}}(t-1)\in\mathbb{R}^l,\end{aligned} \tag{30}$$

$$\hat{\boldsymbol{\Xi}}(l,t) := \mathrm{diag}\{\hat{\xi}(t),\hat{\xi}(t-1),\ldots,\hat{\xi}(t-l+1)\}\in\mathbb{R}^{l\times l}. \tag{31}$$

Similar to the derivation of the robust gradient algorithm, using the negative search and minimizing $J_3(\boldsymbol{a},\boldsymbol{\mu},\boldsymbol{b})$ yield

$$\begin{aligned}\begin{bmatrix}\hat{\boldsymbol{a}}(t)\\ \hat{\boldsymbol{\mu}}(t)\\ \hat{\boldsymbol{b}}(t)\end{bmatrix} &= \begin{bmatrix}\hat{\boldsymbol{a}}(t-1)\\ \hat{\boldsymbol{\mu}}(t-1)\\ \hat{\boldsymbol{b}}(t-1)\end{bmatrix}-\frac{1}{r(t)}\\
&\quad\times\mathrm{grad}[J_3(\hat{\boldsymbol{a}}(t-1),\hat{\boldsymbol{\mu}}(t-1),\hat{\boldsymbol{b}}(t-1)]\\
&= \begin{bmatrix}\hat{\boldsymbol{a}}(t-1)\\ \hat{\boldsymbol{\mu}}(t-1)\\ \hat{\boldsymbol{b}}(t-1)\end{bmatrix}+\frac{1}{r(t)}\sum_{j=0}^{l-1}\xi(t-j)\begin{bmatrix}\boldsymbol{F}(t-j)\boldsymbol{\mu}\\ \boldsymbol{F}^{\mathrm{T}}(t-j)\boldsymbol{a}\\ \boldsymbol{\varphi}(t-j)\end{bmatrix}\\
&\quad\times[y(t-j)-\hat{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t-j)\hat{\boldsymbol{\mu}}(t-1)\\
&\quad-\boldsymbol{\varphi}^{\mathrm{T}}(t-j)\hat{\boldsymbol{b}}(t-1)]\\
&= \begin{bmatrix}\hat{\boldsymbol{a}}(t-1)\\ \hat{\boldsymbol{\mu}}(t-1)\\ \hat{\boldsymbol{b}}(t-1)\end{bmatrix}+\frac{1}{r(t)}\boldsymbol{\Psi}(l,t)\hat{\boldsymbol{\Xi}}(l,t)\boldsymbol{E}(l,t),\end{aligned} \tag{32}$$

$$r(t) = r(t-1)+\hat{\xi}(t)\|\boldsymbol{\psi}(t)\|^2, \tag{33}$$

$$\boldsymbol{\psi}(t) = \begin{bmatrix} \boldsymbol{F}(t)\hat{\boldsymbol{\mu}}(t-1) \\ \boldsymbol{F}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1) \\ \boldsymbol{\varphi}(t) \end{bmatrix}, \tag{34}$$

$$\hat{\xi}(t) = \frac{\ln(1+|e(t)|^2) - \ln(1+|e(t)|)}{\ln 2 \cdot (\ln|\tau(t)|) \cdot |\tau(t)|^2}, \tag{35}$$

$$\tau^2(t) = e^2(t) + \tau_0, \tag{36}$$

$$e(t) = y(t) - \hat{\boldsymbol{a}}^{\mathrm{T}}(t-1)\boldsymbol{F}(t)\hat{\boldsymbol{\mu}}(t-1) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{b}}(t-1), \tag{37}$$

$$\boldsymbol{F}(t) = [\boldsymbol{f}(u(t-1)), \boldsymbol{f}(u(t-2)), \dots, \boldsymbol{f}(u(t-n_a))]^{\mathrm{T}}, \tag{38}$$

$$\boldsymbol{f}(u(t)) = [f_1(u(t)), f_2(u(t)), \dots, f_s(u(t))]^{\mathrm{T}}, \tag{39}$$

$$\boldsymbol{\varphi}(t) = [y(t-1), y(t-2), \dots, y(t-n_b)]^{\mathrm{T}}. \tag{40}$$

Equations (22) and (26)–(40) form the RMIG algorithm for the Hammerstein non-linear system in (1)–(2).

The steps of computing $\hat{\boldsymbol{a}}(t)$, $\hat{\boldsymbol{\mu}}(t)$ and $\hat{\boldsymbol{b}}(t)$ in the RMIG algorithm (22) and (26)–(40) are listed in the following.

1. Let $t=1$, set $\hat{\boldsymbol{a}}(0) = \mathbf{1}_{n_a}/p_0$, $\hat{\boldsymbol{\mu}}(0) = \mathbf{1}_s/p_0$, $\hat{\boldsymbol{b}}(0) = \mathbf{1}_{n_b}/p_0$, $r(0)=1$, $\tau_0 = 1/p_0$, where $p_0 = 10^6$.
2. Collect the input–output data $u(t)$ and $y(t)$.
3. Construct $\boldsymbol{\varphi}(t)$, $\boldsymbol{f}(u(t))$, $\boldsymbol{F}(t)$ and $\boldsymbol{\psi}(t)$ using (40), (39), (38) and (34). Compute $e(t)$, $\tau(t)$ and $\hat{\xi}(t)$ using (37), (36) and (35).
4. Form $\boldsymbol{Y}(l,t)$, $\boldsymbol{\Psi}(l,t)$, $\boldsymbol{\Omega}(l,t)$, $\boldsymbol{\Phi}(l,t)$ and $\hat{\boldsymbol{\Xi}}(l,t)$ using (26)–(31). Compute $\boldsymbol{E}(l,t)$ using (30).
5. Compute $r(t)$ using (33). Update the estimates $\hat{\boldsymbol{a}}(t)$, $\hat{\boldsymbol{\mu}}(t)$ and $\hat{\boldsymbol{b}}(t)$ using (32). Normalize $\hat{\boldsymbol{\mu}}(t)$ using (22).
6. Increase t by 1 and go to Step 2.

Remark 6. Compared with the robust gradient algorithm, the scalar gain $\hat{\xi}(t)$ in (13) is expanded into the gain matrix $\boldsymbol{\Xi}(l,t)$ in (32) of the RMIG algorithm. Thus, the RMIG algorithm can be illustrated as a weighted MIG estimation algorithm. In addition, the RMIG algorithm utilises not only the current data $\{u(t), y(t)\}$, but also the past data $\{u(t-j), y(t-j),\ j = 1, 2, \dots, l-1\}$ at each recursive step, which improves the estimation accuracy by using the observations repeatedly. When $l=1$, the RMIG algorithm reduces to the robust gradient algorithm.

5 | ℓ_1-NORM-BASED MULTI-INNOVATION GRADIENT ALGORITHM

To show the advantage of the RMIG algorithm for the Hammerstein non-linear system, the following simply describes the ℓ_1-norm-based MIG (ℓ_1-MIG) algorithm for comparison.

Define the cost function

$$J_4(\boldsymbol{a}, \boldsymbol{\mu}, \boldsymbol{b}) := \sum_{j=0}^{l-1} \sqrt{(y(t-j) - \boldsymbol{a}^{\mathrm{T}}\boldsymbol{F}(t-j)\boldsymbol{\mu} - \boldsymbol{\varphi}^{\mathrm{T}}(t-j)\boldsymbol{b})^2},$$

where $J_4(\boldsymbol{a}, \boldsymbol{\mu}, \boldsymbol{b})$ represents the ℓ_1-norm of the error. Using the negative search and minimizing $J_4(\boldsymbol{a}, \boldsymbol{\mu}, \boldsymbol{b})$ yield

$$\begin{aligned}\begin{bmatrix} \hat{\boldsymbol{a}}(t) \\ \hat{\boldsymbol{\mu}}(t) \\ \hat{\boldsymbol{b}}(t) \end{bmatrix} &= \begin{bmatrix} \hat{\boldsymbol{a}}(t-1) \\ \hat{\boldsymbol{\mu}}(t-1) \\ \hat{\boldsymbol{b}}(t-1) \end{bmatrix} - \frac{1}{r(t)} \\ &\quad \times \operatorname{grad}[J_4(\hat{\boldsymbol{a}}(t-1), \hat{\boldsymbol{\mu}}(t-1), \hat{\boldsymbol{b}}(t-1)] \\ &= \begin{bmatrix} \hat{\boldsymbol{a}}(t-1) \\ \hat{\boldsymbol{\mu}}(t-1) \\ \hat{\boldsymbol{b}}(t-1) \end{bmatrix} + \frac{1}{r(t)}\boldsymbol{\Psi}(l,t)\hat{\boldsymbol{\Lambda}}(l,t)\boldsymbol{E}(l,t), \end{aligned} \tag{41}$$

$$r(t) = r(t-1) + \|\boldsymbol{\psi}(t)\|^2, \tag{42}$$

$$\boldsymbol{\psi}(t) = \begin{bmatrix} \boldsymbol{F}(t)\hat{\boldsymbol{\mu}}(t-1) \\ \boldsymbol{F}^{\mathrm{T}}(t)\hat{\boldsymbol{a}}(t-1) \\ \boldsymbol{\varphi}(t) \end{bmatrix}, \tag{43}$$

$$\hat{\boldsymbol{\Lambda}}(l,t) = \operatorname{diag}\{\hat{\zeta}(t), \hat{\zeta}(t-1), \dots, \hat{\zeta}(t-l+1)\}, \tag{44}$$

$$\hat{\zeta}(t) = \frac{1}{1+|e(t)|}. \tag{45}$$

Equations (22), (26)–(30) and (37)–(45) construct the ℓ_1-norm-based MIG algorithm for Hammerstein non-linear system. The proposed robust multi-innovation estimation algorithm for Hammerstein non-linear systems with non-Gaussian noise in this paper can combine some mathematical tools [54–56] to study the parameter identification of some linear and non-linear systems with coloured noises and can be applied to other fields [57–60] such as the information processing and transportation communication systems [61–67] and so on.

6 | EXAMPLES

Example 1. Consider the following Hammerstein non-linear system,

$$y(t) = A(z)f[u(t)] + B(z)y(t) + v(t),$$

$$f[u(t)] = \mu_1 u(t) + \mu_2 u^2(t) = 0.60u(t) + 0.80u^2(t),$$

$$A(z) = a_1 z^{-1} + a_2 z^{-2} = 0.24z^{-1} + 0.20z^{-2},$$

$$B(z) = b_1 z^{-1} + b_2 z^{-2} = 0.35z^{-1} - 0.45z^{-2}.$$

The parameters μ_1 and μ_2 of the non-linear block meet Assumption 1, and the input non-linearity $f[u(t)]$ is a quadratic

TABLE 1 The RMIG estimates and errors under different l for Example 1 ($\alpha = 1.60$)

l	t	a_1	a_2	μ_1	μ_2	b_1	b_2	δ (%)
1	100	0.04774	0.06281	0.74261	0.66972	0.23743	−0.31790	29.43038
	200	0.05657	0.06912	0.74212	0.67027	0.24530	−0.33157	28.15767
	500	0.06575	0.07643	0.74177	0.67066	0.25188	−0.35510	26.65769
	1000	0.06910	0.07840	0.74151	0.67094	0.26009	−0.35978	26.09881
	1500	0.07072	0.07943	0.74138	0.67109	0.26315	−0.36166	25.86165
	2000	0.07105	0.07977	0.74134	0.67113	0.26401	−0.36329	25.77354
3	100	0.10393	0.13114	0.67269	0.73992	0.39193	−0.54477	17.36385
	200	0.12640	0.14684	0.66910	0.74317	0.37722	−0.48785	13.49518
	500	0.14484	0.15722	0.66967	0.74266	0.36264	−0.49476	12.20478
	1000	0.15150	0.16079	0.66891	0.74334	0.36798	−0.48011	11.42197
	1500	0.15445	0.16137	0.66846	0.74374	0.36438	−0.46657	10.97777
	2000	0.15445	0.16148	0.66831	0.74388	0.36229	−0.46607	10.94064
7	100	0.20146	0.18817	0.60392	0.79704	0.46369	−0.55253	13.27942
	200	0.22616	0.22041	0.59430	0.80424	0.38923	−0.46081	4.03328
	500	0.23812	0.21295	0.59992	0.80006	0.33836	−0.48870	3.56125
	1000	0.23483	0.21189	0.60056	0.79958	0.34246	−0.46880	2.01726
	1500	0.23758	0.21192	0.60031	0.79977	0.34974	−0.46224	1.44732
	2000	0.23580	0.21017	0.60029	0.79978	0.34903	−0.46260	1.40531
True values		0.24000	0.20000	0.60000	0.80000	0.35000	−0.45000	

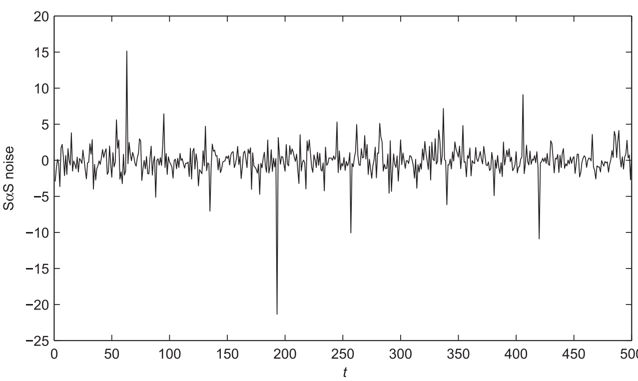

FIGURE 1 The impulsive noise versus t for Example 1

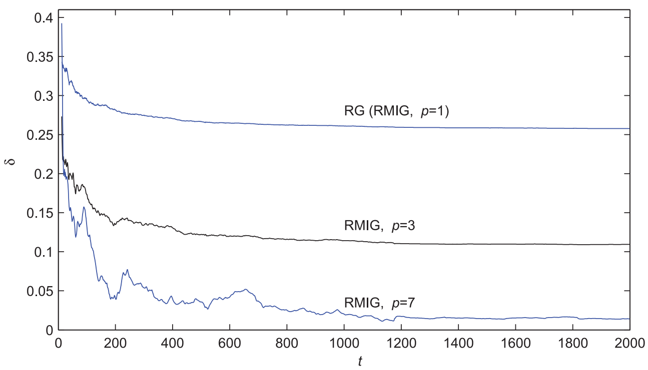

FIGURE 2 The RMIG estimation errors δ versus t under different l for Example 1

polynomial function. The parameter vector to be estimated is

$$\begin{aligned}\boldsymbol{\vartheta} &:= [a_1, a_2, \mu_1, \mu_2, b_1, b_2]^{\mathrm{T}}\\ &= [0.24, 0.20, 0.60, 0.80, 0.35, -0.45]^{\mathrm{T}}.\end{aligned}$$

In simulation, the input $\{u(t)\}$ is taken as a persistent excitation signal sequence, $\{v(t)\}$ is a non-Gaussian noise which is modelled by the symmetric α-stable ($S\alpha S$) distribution, and can be described as the characteristic function:

$$g_\alpha(t) = \exp\{-\gamma|t|^\alpha\},$$

where $0 < \gamma \leqslant 1$ is a constant, and α ($1 < \alpha \leqslant 2$) is a shape parameter. As α decreases, the outliers of the $S\alpha S$ distribution have higher amplitudes and are more likely more impulsive noise [68, 69]. Figure 1 depicts the $S\alpha S$ impulsive noise process with $\alpha = 1.2$.

Take the data length $L = 2000$. Under the shape parameter $\alpha = 1.6$ of the noise process, the RMIG algorithm is applied to identify the system, and the RMIG estimates and errors are shown in Table 1 and Figure 2. The RMIG estimates with $l = 7$ versus t are illustrated in Figures 3 and 4.

To test the effect of the non-Gaussian noise $v(t)$ to the RMIG algorithm, Table 2 and Figure 5 compare the RMIG estimates and errors under the different shape parameters α.

Example 2. Consider the following Hammerstein non-linear system,

$$y(t) = A(z)f[u(t)] + B(z)y(t) + v(t),$$

$$f[u(t)] = \mu_1 \cos(u(t)) + \mu_2 \cos^2(u(t))$$

TABLE 2 The RMIG estimates and errors under different α for Example 1 (l = 7)

α	t	a_1	a_2	μ_1	μ_2	b_1	b_2	δ (%)
1.2	100	0.21698	0.20490	0.59771	0.80171	0.45852	−0.55871	13.03064
	200	0.24055	0.23056	0.58653	0.80992	0.39017	−0.46881	4.72874
	500	0.24910	0.22101	0.59399	0.80447	0.33799	−0.49620	4.48247
	1000	0.24097	0.21729	0.59507	0.80367	0.34361	−0.47839	2.88591
	1500	0.24328	0.21645	0.59464	0.80399	0.34956	−0.46675	2.06572
	2000	0.24027	0.21351	0.59468	0.80396	0.34887	−0.46776	1.95428
2.0	100	0.19347	0.18244	0.60640	0.79516	0.46011	-0.55252	13.30230
	200	0.21869	0.21694	0.59691	0.80231	0.38351	-0.45739	3.68683
	500	0.23156	0.21039	0.60206	0.79845	0.33562	-0.48484	3.36034
	1000	0.23138	0.20498	0.60264	0.79801	0.33838	-0.46166	1.63638
	1500	0.23656	0.20591	0.60257	0.79806	0.34833	-0.45715	0.88391
	2000	0.23441	0.20364	0.60257	0.79807	0.34926	-0.45713	0.86406
True values		0.24000	0.20000	0.60000	0.80397	0.35000	−0.45000	

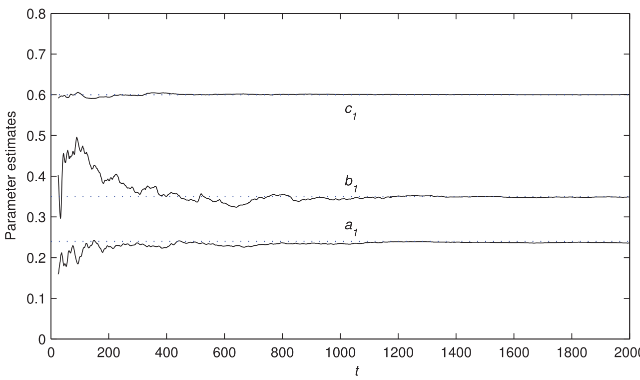

FIGURE 3 The RMIG estimates $\hat{a}_1(t)$, $\hat{b}_1(t)$ and $\hat{c}_1(t)$ versus t for Example 1

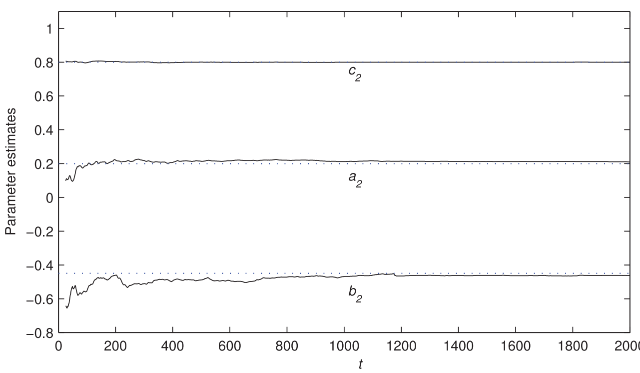

FIGURE 4 The RMIG estimates $\hat{a}_2(t)$, $\hat{b}_2(t)$ and $\hat{c}_2(t)$ versus t for Example 1

$$= 0.72\cos(u(t)) + 0.69397\cos^2(u(t)),$$

$$A(z) = a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3}$$

$$= 0.40 z^{-1} + 0.30 z^{-2} + 0.25 z^{-3},$$

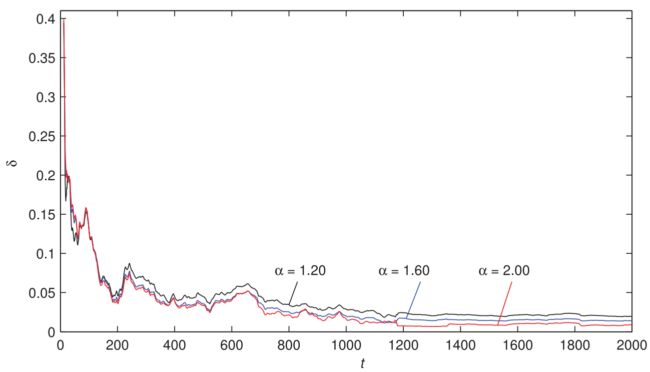

FIGURE 5 The RMIG estimation errors δ versus t under different shape parameters α for Example 1

$$B(z) = b_1 z^{-1} + b_2 z^{-2} + b_3 z^{-3}$$

$$= 0.60 z^{-1} - 0.15 z^{-2} - 0.50 z^{-3}.$$

The parameters μ_1 and μ_2 of the non-linear block meet Assumption 1, the input non-linearity $f[u(t)]$ is a trigonometric function, and $v(t)$ follows the ε-contaminated distribution which can be approximated by the following mixed Gaussian distribution [70]:

$$\mathcal{P}_\varepsilon = \{\mathcal{P} : \mathcal{P} = (1-\varepsilon)\mathcal{N}(0,\sigma_1^2) + \varepsilon\mathcal{N}(0,\sigma_2^2)\},$$

$\mathcal{N}(0,\sigma_1^2)$ and $\mathcal{N}(0,\sigma_2^2)$ represent the normal distribution with zero mean and variances σ_1^2 and σ_2^2 ($\sigma_2^2 \gg \sigma_1^2$), respectively, and ε $(0<\varepsilon<1)$ is the contamination degree. The normal distribution with larger variance σ_2^2 produces outliers. The parameter vector to be estimated is

$$\boldsymbol{\vartheta} := [a_1, a_2, a_3, \mu_1, \mu_2, b_1, b_2, b_3]^{\mathrm{T}}$$

$$= [0.40, 0.30, 0.25, 0.72, 0.69397, 0.60, -0.15, -0.50]^{\mathrm{T}}.$$

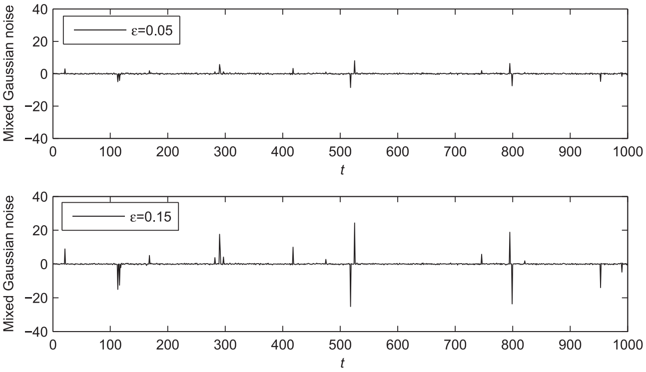

FIGURE 6 The mixed Gaussian noise with different contaminations for Example 2

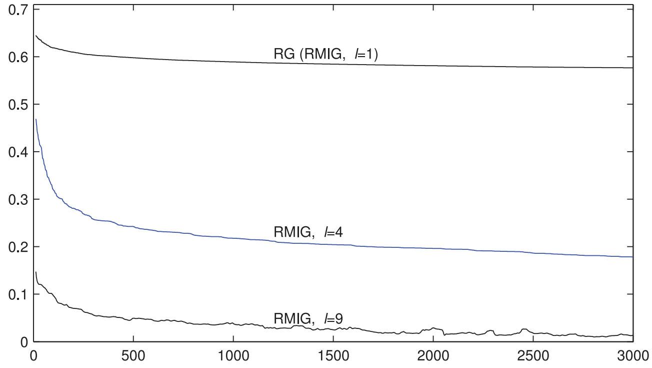

FIGURE 7 The RMIG estimation errors δ versus t under different l for Example 2

To show the influences of contamination degree ε to the noise distribution, under the variances $\sigma_1^2 = 0.50^2$ and $\sigma_2^2 = 10.00^2$, Figure 6 depicts the mixed Gaussian noise process with $\varepsilon = 0.05$ and $\varepsilon = 0.15$, respectively. It can be seen from Figure 6 that a larger contamination degree ε corresponds to higher amplitudes of the outliers.

Take the data length $L = 3000$. Under the noise variances $\sigma_1^2 = 0.50^2$ and $\sigma_2^2 = 10.00^2$ and the contamination degree $\varepsilon = 0.05$, apply the RMIG algorithm to identify the system, and the parameter estimates and errors are shown in Table 3 and Figure 7.

To test the effect of the non-Gaussian noise $v(t)$ to the RMIG algorithm, under the same noise variance $\sigma_1^2 = 0.50^2$ and the same contamination degree $\varepsilon = 0.05$, Table 4 and Figure 8 compare the RMIG estimates and errors under different noise variance σ_2^2. Under the same noise variances $\sigma_1^2 = 0.50^2$ and $\sigma_2^2 = 10.00^2$, Table 5 and Figure 9 compare the RMIG estimates and errors under different contamination degree ε.

To show the advantage of the RMIG algorithm, Table 6 and Figure 10 compare the estimation errors of the RMIG algorithm and the ℓ_1-MIG algorithm under the variances $\sigma_1^2 = 0.50^2$ and $\sigma_2^2 = 10.00^2$ and the contamination degree $\varepsilon = 0.05$.

From Tables 1–6 and Figures 1–10, the following conclusions can be drawn.

- As t increases, the RMIG estimation errors decay, and a larger innovation length l results in the higher estimation

TABLE 3 The RMIG estimates and errors for Example 2

l	t	a_1	a_2	a_3	μ_1	μ_2	b_1	b_2	b_3	δ (%)
1	100	0.03438	0.03353	0.03344	0.71954	0.69445	0.07862	−0.02992	−0.04590	61.85030
	200	0.04124	0.03908	0.03881	0.71959	0.69440	0.08850	−0.02527	−0.04783	60.97056
	500	0.05122	0.04934	0.04880	0.71968	0.69431	0.10120	−0.01999	−0.04793	59.77824
	1000	0.05764	0.05646	0.05544	0.71973	0.69426	0.11178	−0.01459	−0.04837	58.91180
	2000	0.06344	0.06267	0.06168	0.71979	0.69420	0.12121	−0.01055	−0.04947	58.10741
	3000	0.06597	0.06539	0.06456	0.71981	0.69417	0.12611	−0.00899	−0.05136	57.66379
4	100	0.20897	0.19708	0.17903	0.70365	0.71055	0.25647	−0.15333	−0.33353	31.93819
	200	0.23382	0.22003	0.19929	0.70475	0.70946	0.30545	−0.12202	−0.33045	28.07873
	500	0.25843	0.24404	0.22210	0.70590	0.70831	0.35138	−0.10427	−0.33904	24.22781
	1000	0.27085	0.25723	0.23526	0.70653	0.70768	0.38255	−0.09261	−0.35004	21.76806
	2000	0.28171	0.26766	0.24596	0.70705	0.70717	0.40851	−0.08625	−0.36253	19.64055
	3000	0.28577	0.27224	0.25036	0.70723	0.70698	0.42796	−0.08694	−0.38060	17.85144
9	100	0.29003	0.29175	0.28405	0.71387	0.70028	0.55910	−0.10563	−0.45485	9.90403
	200	0.31150	0.29556	0.27446	0.71294	0.70122	0.58185	−0.10992	−0.48332	7.43986
	500	0.33276	0.29871	0.27450	0.71384	0.70031	0.58287	−0.13373	−0.50789	5.47061
	1000	0.34654	0.29810	0.27171	0.71479	0.69934	0.58951	−0.13583	−0.50729	4.39230
	2000	0.38491	0.30959	0.27679	0.71550	0.69862	0.61418	−0.13185	−0.50642	2.91299
	3000	0.38343	0.30434	0.26182	0.71556	0.69855	0.60494	−0.14879	−0.51035	1.76635
True values		0.40000	0.30000	0.25000	0.72000	0.69397	0.60000	−0.15000	−0.50000	

TABLE 4 The RMIG estimates and errors under σ_2^2 for Example 2 ($\sigma_1^2 = 0.50^2, \varepsilon = 0.05$)

σ_2^2	t	a_1	a_2	a_3	μ_1	μ_2	b_1	b_2	b_3	δ (%)
20.00^2	100	0.26014	0.27370	0.28518	0.71292	0.70125	0.58774	−0.12380	−0.48234	10.81290
	200	0.27220	0.28119	0.28552	0.71298	0.70118	0.59737	−0.13696	−0.49283	9.69389
	500	0.29049	0.29451	0.28581	0.71350	0.70066	0.60172	−0.15015	−0.50523	8.30756
	1000	0.31181	0.29728	0.28482	0.71437	0.69977	0.59785	−0.14486	−0.50085	6.83796
	2000	0.36869	0.31553	0.28804	0.71622	0.69787	0.61840	−0.14219	−0.50343	3.99844
	3000	0.36913	0.31487	0.28130	0.71616	0.69794	0.60250	−0.15129	−0.50272	3.36270
15.00^2	100	0.28000	0.28386	0.28495	0.71354	0.70061	0.58540	−0.12084	−0.47672	9.50550
	200	0.29524	0.28959	0.28013	0.71318	0.70098	0.59706	−0.12714	−0.49058	8.08258
	500	0.31057	0.29793	0.27701	0.71357	0.70058	0.59623	−0.14450	−0.50521	6.75980
	1000	0.32864	0.30051	0.27757	0.71449	0.69964	0.59642	−0.14430	−0.50238	5.53834
	2000	0.37460	0.31366	0.28291	0.71576	0.69834	0.62067	−0.13765	−0.50518	3.62677
	3000	0.37263	0.31079	0.27165	0.71569	0.69842	0.60382	−0.15110	−0.50450	2.69113
True values		0.40000	0.30000	0.25000	0.72000	0.69397	0.60000	−0.15000	−0.50000	

TABLE 5 The RMIG estimates and errors under ε for Example 2 ($\sigma_1^2 = 0.50^2, \sigma_2^2 = 10.00^2$)

ε	t	a_1	a_2	a_3	μ_1	μ_2	b_1	b_2	b_3	δ (%)
0.30	100	0.24525	0.26614	0.28584	0.70808	0.70613	0.56847	−0.10194	−0.48427	12.46475
	200	0.25520	0.27287	0.28796	0.70833	0.70588	0.58507	−0.12966	−0.49681	11.12454
	500	0.27975	0.29157	0.29286	0.70919	0.70502	0.59803	−0.14184	−0.50918	9.28297
	1000	0.30505	0.29627	0.28866	0.71005	0.70415	0.59724	−0.14385	−0.50126	7.44251
	2000	0.36422	0.31422	0.28578	0.71197	0.70221	0.60990	−0.14623	−0.50124	3.93263
	3000	0.36729	0.31545	0.28228	0.71201	0.70217	0.60105	−0.15089	−0.50136	3.57481
0.15	100	0.26986	0.27921	0.28622	0.71120	0.70299	0.59012	−0.12002	−0.48275	10.17400
	200	0.28352	0.28706	0.28496	0.71115	0.70304	0.59890	−0.13319	−0.49385	8.90975
	500	0.30160	0.29822	0.28288	0.71162	0.70256	0.59979	−0.14787	−0.50551	7.50162
	1000	0.32323	0.30003	0.28127	0.71255	0.70162	0.59726	−0.14431	−0.50133	6.01094
	2000	0.37516	0.31385	0.28237	0.71411	0.70003	0.61831	−0.14054	−0.50447	3.49270
	3000	0.37401	0.31229	0.27442	0.71403	0.70012	0.60263	−0.15141	−0.50330	2.79181
True values		0.40000	0.30000	0.25000	0.72000	0.69397	0.60000	−0.15000	−0.50000	

TABLE 6 The ℓ_1-MIG estimates and errors for Example 2

t	a_1	a_2	a_3	μ_1	μ_2	b_1	b_2	b_3	δ (%)
100	0.44656	0.17891	0.21748	0.72417	0.68962	0.67284	−0.25040	−0.39727	15.01719
200	0.43645	0.14388	0.15430	0.71844	0.69559	0.68205	−0.20640	−0.34914	18.63443
500	0.43372	0.19034	0.19665	0.72172	0.69218	0.70810	−0.25647	−0.41470	15.43090
1000	0.42028	0.20910	0.21380	0.72234	0.69154	0.67766	−0.22615	−0.44474	11.31038
2000	0.42529	0.28865	0.23021	0.72320	0.69064	0.61087	−0.19391	−0.43570	6.15207
3000	0.36463	0.28386	0.26647	0.72142	0.69250	0.60193	−0.19409	−0.48427	4.52457
True values	0.40000	0.30000	0.25000	0.72000	0.69397	0.60000	−0.15000	−0.50000	

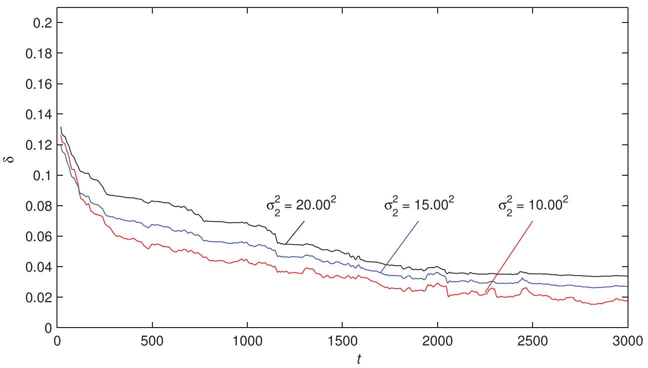

FIGURE 8 The RMIG estimation errors δ versus t under different noise variances σ_2^2 for Example 2

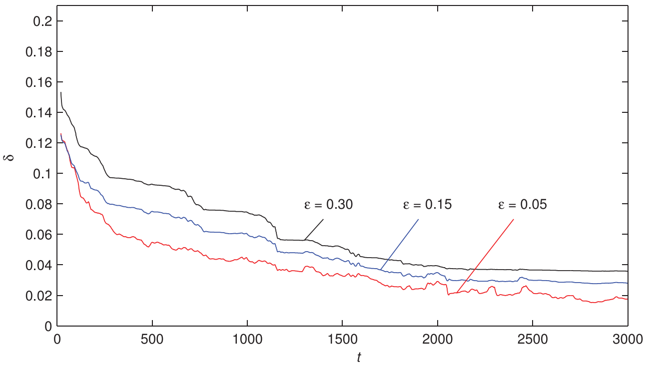

FIGURE 9 The RMIG estimation errors δ versus t under different contamination degree ε for Example 2

accuracy—see Tables 1 and 3, and Figures 2 and 7. It shows that the RMIG algorithm is effective for the Hammerstein non-linear system with non-Gaussian noise.

- The RMIG estimates can rapidly reach the vicinity of the true values with increasing t—see Figures 3 and 4.
- With the shape parameter α declining, the RMIG estimation errors have a small increase and two estimation error curves drop to below 0.02 at instant $t = 2000$—see Table 2 and Figure 5. This suggests that the RMIG algorithm is not sensitive to the variation of SαS noise to some extent.

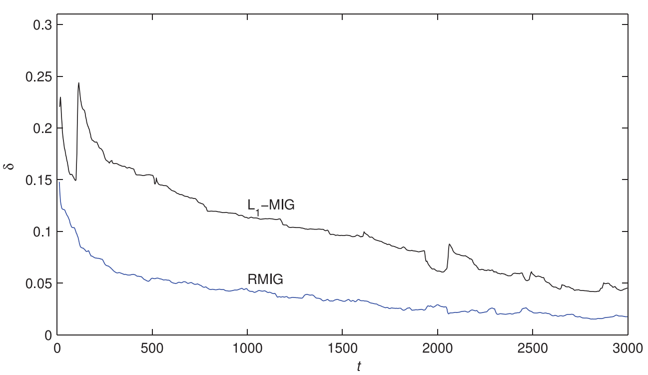

FIGURE 10 The estimation errors of the RMIG algorithm and the ℓ_1-MIG algorithm for Example 2

- Increasing the noise variance σ_2^2 and decreasing the contamination degree ε lead to stronger noise interference and lower RMIG estimation accuracy, but the reduced accuracy is small—see Figures 8 and 9. It indicates that the RMIG algorithm is robust to the ε-contaminated noise.
- Under the same data length and the same noise environment, the RMIG algorithm has higher estimation accuracy and better robustness than the ℓ_1-MIG algorithm—see Figure 10.

7 | CONCLUSIONS

A robust gradient algorithm and a RMIG algorithm are developed to identify the Hammerstein non-linear system corrupted by non-Gaussian noise. The algorithms are based on the approximation of the expectation of the logarithmic p-norm of prediction errors. The continuous combination of error norms generates an adjustable gain in the recursive algorithms and yields good robustness to non-Gaussian noise, which is verified by the results of two simulation examples. In future, there are still some interesting topics which can be discussed, for example, how to devise proper weighting functions to enhance the performance of the RMIG algorithm. Additionally, how to design the identification algorithms with higher precision by means of some acceleration techniques such as the Aitken method. These topics will remain as open issues.

ACKNOWLEDGEMENTS

This work was supported by the Science and Technology Department of Henan Province (China) (No. 202102210297), the Key Research Project of Henan Higher Education Institutions (China) (No. 21A413006) and Nanhu Scholars Program for Young Scholars of XYNU (China) (No. 2017B67).

ORCID

Feng Ding https://orcid.org/0000-0002-2721-2025

REFERENCES

1. Ding, F.: Modern Control Theory. Tsinghua University Press, Beijing (2018)
2. Ding, F.: System Identification - New Theory and Methods. Science Press, Beijing (2013)
3. Ding, F.: System Identification - Performances Analysis for Identification Methods. Science Press, Beijing (2014)
4. Ding, F.: System Identification - Auxiliary Model Identification Idea and Methods. Science Press, Beijing (2017)
5. Ding, F.: System Identification - Iterative Search Principle and Identification Methods. Science Press, Beijing (2018)
6. Ding, F.: System Identification - Multi-Innovation Identification Theory and Methods. Science Press, Beijing (2016)
7. Na, J., et al.: Finite-time convergence adaptive neural network control for nonlinear servo systems. IEEE Trans. Cybern. 50(6), 2568–2579 (2020)
8. Wang, L.J., et al.: Decomposition-based multiinnovation gradient identification algorithms for a special bilinear system based on its input-output representation. Int. J. Robust Nonlinear Control 30(9), 3607–3623 (2020)
9. Wang, L.J., et al.: Hierarchical recursive generalized extended least squares estimation algorithms for a class of nonlinear stochastic systems with colored noise. J. Frankl. Inst. 356(16), 10102–10122 (2019)

10. Li, M.H., Liu, X.M.: Maximum likelihood least squares based iterative estimation for a class of bilinear systems using the data filtering technique. Int. J. Control Autom. Syst. 18(6), 1581–1592 (2020)
11. Kapetina, M.N., Rapaic, M.R., Pisano, A.: Adaptive parameter estimation in LTI systems. IEEE Trans. Autom. Control 64(10), 4188–4195 (2019)
12. Esmaeilani, L.L., Ghaisari, J., Bagherzadeh, M.A.: Bayesian approach to identify Hammerstein-Wiener non-linear model in presence of noise and disturbance. IET Control Theory Appl. 13(3), 367–376 (2019)
13. Chen, G.Y., et al.: Modified Gram-Schmidt method-based variable projection algorithm for separable nonlinear models. IEEE Trans. Neural Netw. Learn. Syst. 30(8), 2410–2418 (2019)
14. Zhu, Q.M., Yu, D.L., Zhao, D.Y.: An Enhanced Linear Kalman Filter (EnLKF) algorithm for parameter estimation of nonlinear rational models. Int. J. Syst. Sci. 48(3), 451–461 (2017)
15. Fan, Y.M., Liu, X.M.: Two-stage auxiliary model gradient-based iterative algorithm for the input nonlinear controlled autoregressive system with variable-gain nonlinearity. Int. J. Robust Nonlinear Control 30(14), 5492–5509 (2020)
16. Ji, Y., et al.: Parameter estimation for block-oriented nonlinear systems using the key term separation. Int. J. Robust Nonlinear Control 30(9), 3727–3752 (2020)
17. Li, J.H., et al.: Parameter estimation of Wiener systems based on the particle swarm iteration and gradient search principle. Circuits Syst. Signal Process. 39(7), 3470–3495 (2020)
18. Ji, Y., Jiang, X.K., Wan, L.J.: Hierarchical least squares parameter estimation algorithm for two-input Hammerstein finite impulse response systems. J. Frankl. Inst. 357(8), 5019–5032 (2020)
19. Wang, X.H., et al.: Combined state and multi-innovation parameter estimation for an input nonlinear state space system using the key term separation. IET Control Theory Appl. 10(13), 1503–1512 (2016)
20. Hou, J., et al.: Fixed point iteration-based subspace identification of Hammerstein state-space models. IET Control Theory Appl. 13(8), 1173–1181 (2019)
21. Li, J.H., Zhang, J.L.: Maximum likelihood identification of dual-rate Hammerstein output-error moving average system. IET Control Theory Appl. 14(8), 1078–1090 (2020)
22. Wang, D.Q., et al.: A novel EM identification method for Hammerstein systems with missing output data. IEEE Trans. Ind. Inform. 16(4), 2500–2508 (2020)
23. Rahmani, M.R., Farrokhi, M.: Fractional-order Hammerstein state-space modeling of nonlinear dynamic systems from input-output measurements. ISA Trans. 96, 177–184 (2020)
24. Liu, Y., Bai, E.W.: Iterative identification of Hammerstein systems. Automatica 43(2), 346–354 (2007)
25. Kazemi, M., Arefi, M.M.: A fast iterative recursive least squares algorithm for Wiener model identification of highly nonlinear systems. ISA Trans. 67, 382–388 (2017)
26. Isaza-Hurtado, J.A., Botero-Castro, H.A., Alvarez, H.: Robust estimation for LPV systems in the presence of non-uniform measurements. Automatica 115, 108901 (2020)
27. Liu, Q.Y., et al.: Quadratic estimation for discrete time-varying non-Gaussian systems with multiplicative noises and quantization effects. Automatica 113, 108714 (2020)
28. Dong, S.J., et al.: Robust extended recursive least squares identification algorithm for Hammerstein systems with dynamic disturbances. Digit. Signal Process. 101, 102716 (2020)
29. He, Y.S., et al.: Robust matrix completion via maximum correntropy criterion and half-quadratic optimization. IEEE Trans. Signal Process. 68, 181–195 (2020)
30. Alessandri, A., Awawdeh, M.: Moving-horizon estimation with guaranteed robustness for discrete-time linear systems and measurements subject to outliers. Automatica 67, 85–93 (2016)
31. Silva, M.E., Pereira, I., McCabe, B.: Bayesian outlier detection in non-Gaussian autoregressive time series. J. Time Ser. Anal. 40(5), 631–648 (2019)
32. Li, J.H., Deng, F., Chen, J.: A fast distributed variational Bayesian filtering for multisensor LTV system with non-Gaussian noise. IEEE Trans. Cybern. 49(7), 2431–2443 (2019)
33. Zayyani, H.: Continuous mixed-norm adaptive algorithm for system identification. IEEE Signal Process. Lett. 21(9), 1108–1110 (2014)
34. Stojanovic, V., Nedic, N.: Identification of time-varying OE models in presence of non-Gaussian noise: Application to pneumatic servo drives. Int. J. Robust Nonlinear Control 26(18), 3974–3995 (2016)
35. Li, L., Zhao, H.Q.: A robust total least mean M-estimate adaptive algorithm for impulsive noise suppression. IEEE Trans. Circuits Syst. II, Exp. Briefs 67(4), 800–804 (2020)
36. Liu, X., Yang, X.Q.: Robust identification approach for nonlinear state-space models. Neurocomputing 333, 329–338 (2019)
37. Gu, Y., Zhu, Q., Nouri, H.: Bias compensation-based parameter and state estimation for a class of time-delay nonlinear state-space models. IET Control Theory Appl. 14(15), 2176–2185 (2020)
38. Ding, J., et al.: Weighted parameter estimation for Hammerstein nonlinear ARX systems. Circuits Syst. Signal Process. 39(4), 2178–2192 (2020)
39. Xu, L., Chen, L., Xiong, W.L.: Parameter estimation and controller design for dynamic systems from the step responses based on the Newton iteration. Nonlinear Dyn. 79(3), 2155–2163 (2015)
40. Xu, L.: The damping iterative parameter identification method for dynamical systems based on the sine signal measurement. Signal Process. 120, 660–667 (2016)
41. Xu, L., et al.: Hierarchical parameter estimation for the frequency response based on the dynamical window data. Int. J. Control Autom. Syst. 16(4), 1756–1764 (2018)
42. Xu, L., Song, G.L.: A recursive parameter estimation algorithm for modeling signals with multi-frequencies. Circuits Syst. Signal Process. (2020) 39(8), 4198–4224
43. Bu, N., Pang, J.X., Deng, M.: Robust fault tolerant tracking control for the multi-joint manipulator based on operator theory. J. Frankl. Inst. 357(5), 2696–2714 (2020)
44. Chen, J., Zhu, Q.M., Liu, Y.J.: Modified Kalman filtering based multi-step-length gradient iterative algorithm for ARX models with random missing outputs. Automatica 118, 109034 (2020)
45. Chen, J., et al.: Stochastic average gradient algorithm for multirate FIR models with varying time delays using self-organizing maps. Int. J. Adapt. Control Signal Process. 34(7), 955–970 (2020)
46. Ji, F., et al.: Self-reconfiguration batteries with stable voltage during the full cycle without the DC-DC converter. J. Energy Storage 28, 101213 (2020)
47. Wan, X.K., et al.: Heartbeat classification algorithm based on one-dimensional convolution neural network. J. Mech. Med. Biol. 20(7), 2050046 (2020)
48. Wan, X.K., et al.: Ventricular repolarization instability quantified by instantaneous frequency of ECG ST intervals. Technol. Health Care 29(1), 73–83 (2021). https://doi.org/10.3233/THC-202377
49. Zhang, Y., Yan, Z., Zhou, C.C.: Capacity allocation of HESS in micro-grid based on ABC algorithm. Int. J. Low-Carbon Technol. 15(4), 496–505 (2020)
50. Zhao, Z.Y., et al.: A health performance evaluation method of multirotors under wind turbulence. Nonlinear Dyn. 102(3), 1701–1715 (2020)
51. Jin, X., et al.: Deep-learning prediction model with serial two-level decomposition based on Bayesian optimization. Complexity 2020, 4346803 (2020)
52. Bai, E.W.: An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems. Automatica 34(3), 333–338 (1998)
53. Wang, X.H., Zhu, F., Ding, F.: Bias correction-based recursive estimation for dual-rate output-error systems with sampling noise. Circuits Syst. Signal Process. 39(9), 4297–4319 (2020)
54. Jiang, C.M., et al.: Synchronization of bidirectional N-coupled fractional-order chaotic systems with ring connection based on antisymmetric structure. Adv. Differ. Equ. 2019(1), 456 (2019)
55. Li, X.Y., Wu, B.Y.: A new kernel functions based approach for solving 1-D interface problems. Appl. Math. Comput. 380, 125276 (2020)
56. Geng, F.Z.: Piecewise reproducing kernel-based symmetric collocation approach for linear stationary singularly perturbed problems. AIMS Math. 5(6), 6020–6029 (2020)

57. Yang, G.C., et al.: Rapid relocation method for mobile robot based on improved ORB-SLAM2 algorithm. Remote Sens. 11(2), 149 (2019). https://doi.org/10.3390/rs11020149
58. Lin, J., Li, Y., Yang, G.C.: FPGAN: Face de-identification method with generative adversarial networks for social robots. Neural Networks 133, 132–147 (2021)
59. Tian, S.S., et al.: Application of $C_6F_{12}O/CO_2$ mixture in 10 kV medium-voltage switchgear. IET Sci. Meas. Technol. 13(9), 1225–1230 (2019)
60. Zhang, G.Z., et al.: Ladder-wise calculation method for z-coordinate of transformer PD source based on planar layout UHF antenna sensors. IEEJ Trans. Electr. Electr. Eng. 15(3), 340–345 (2020)
61. Su, S., et al.: An energy-efficient train operation approach by integrating the metro timetabling and eco-driving. IEEE Trans. Intell. Transp. Syst. 21(10), 4252–4268 (2020)
62. Su, S., et al.: Design of running grades for energy-efficient train regulation: A case study for Beijing Yizhuang line. IEEE Intell. Transp. Syst. Mag. (2021). https://doi.org/10.1109/MITS.2019.2907681
63. Cao, Y., et al.: Bio-inspired speed curve optimization and sliding mode tracking control for subway trains. IEEE Trans. Veh. Technol. 68(7), 6331–6342 (2019)
64. Cao, Y., et al.: Fault diagnosis of train plug door based on a hybrid criterion for IMFs selection and fractional wavelet package energy entropy. IEEE Trans. Veh. Technol. 68(8), 7544–7551 (2019)
65. Cao, Y., Wen, J.K., Ma, L.C.: Tracking and collision avoidance of virtual coupling train control system. Alexandria Engineering Journal 60(2), 2115–2125 (2021)
66. Zhang, L., Tang, S.Y., Lv, L.L.: An finite iterative algorithm for sloving periodic Sylvester bimatrix equations. J. Frankl. Inst. 357(15), 10757–10772 (2020)
67. Cao, Y., Li, P., Zhang, Y.: Parallel processing algorithm for railway signal fault diagnosis data based on cloud computing. Future Gener. Comput. Syst. 88, 279–283 (2018)
68. Zhao, H.Q., et al.: Adaptive recursive algorithm with logarithmic transformation fornonlinear system identification in alpha-stable noise. Digit. Signal Process. 46, 120–132 (2015)
69. Lu, L., Zhao, H.Q., Chen, B.D.: Robust adaptive algorithm for smart antenna system with alpha-stable noise. IEEE Trans. Circuits Syst. II, Exp. Briefs 65(11), 1783–1787 (2018)
70. Stojanovic, V., Prsic, D.: Robust identification for fault detection in the presence of non-Gaussian noises: Application to hydraulic servo drives. Nonlinear Dyn. 100(3), 2299–2313 (2020)

How to cite this article: Wang X, Ding F. The robust multi-innovation estimation algorithm for Hammerstein non-linear systems with non-Gaussian noise. *IET Control Theory Appl.* 2021;15:989–1002. https://doi.org/10.1049/cth2.12097