2024 European Control Conference (ECC)
June 25-28, 2024. Stockholm, Sweden

Strategies to Alleviate the Impact of Noise in Data-Driven Model Order Reduction by Moment Matching

Zichen Zhao, Junyu Mao and Giordano Scarciotti

***Abstract*— In this paper we propose strategies to alleviate the impact of noise on data-driven model-order reduction by moment matching. We classify the noise affecting the data-driven methods as *interconnection noise* and *measurement noise*. We then consider two statistical models of the noise, namely Gaussian (white noise) and Student's t, to represent noise in a variety of applications. We propose and study the use of Wavelet denoising for dealing with white noise and the use of Huber regression for the Student's t-distribution. We demonstrate by means of extensive simulations how these strategies improve the accuracy and robustness of the data-driven algorithms.**

I. INTRODUCTION

To address the growing complexity of dynamical systems, the field of model order reduction proposes methods to reduce the dimension of the state of the system. A review of linear model order reduction methods is available in [1]. More recently, there has been growing interest in data-driven model order reduction methods. For this class of methods, the full-order model is potentially unknown and a reduced-order model is constructed directly from measurements. This paper focuses on the family of data-driven model order reduction methods firstly introduced in [2]. Therein a method to construct one-sided reduced-order models by moment matching from time-domain measurements of the signals of certain systems' interconnections has been proposed. The method has then been extended in various directions. In particular the data-driven swapped interconnection problem and the two-sided interconnection problem for linear systems have been solved in [3] and [4], respectively, and applied to power systems with classical [5] and renewable [6] generation. However, while these papers have focused on theoretical development, they have seldom considered the effect that noise may have on the estimated reduced-order models. In this paper we fill this gap by proposing strategies to alleviate the impact that noise may have on the data-driven moment matching methods introduced in those papers. We consider all possible sources of noise (from interconnection of modules and from measurement) and study two different noise distributions. We propose two methods to improve the construction of reduced-order models from data and we illustrate the results by means of extensive simulations.

The rest of the paper is structured as follows: Section II reviews the data-driven model order reduction methods. Section III discusses how noise impacts data-driven moment matching and the statistical models used in the paper. Section IV explores the white noise scenarios and studies the effect of Wavelet denoising. Section V explores the Student's t-distribution scenarios and studies the effect of Huber regression. Section VI contains our concluding remarks.

All the simulations in the paper are performed on the Los Angeles University Hospital building model from the SLICOT benchmark library [7].

Notation. In this paper, we employ standard notation. $\mathbb{R}_{\geq 0}$ and $\mathbb{R}_{>0}$ denote the sets of non-negative and positive real numbers, respectively. The set of complex numbers with a negative real part is represented as $\mathbb{C}_{<0}$, while $\mathbb{C}_0$ refers to the set of complex numbers with zero real part. The identity matrix is indicated by I, and $\sigma(A)$ represents the spectrum of a matrix $A \in \mathbb{R}^{n\times n}$. The Kronecker product is denoted by $\otimes$. The induced Euclidean or supremum matrix norms are indicated by $\|A\|_2$ and $\|A\|_\infty$, respectively. The vectorization of a matrix $A \in \mathbb{R}^{n\times m}$ is indicated by $\text{vec}(A)$, which is the $nm \times 1$ vector formed by stacking the matrix columns, *i.e.* $\text{vec}(A) = [a_1^\top, a_2^\top, \dots, a_m^\top]^\top$, where $a_i \in \mathbb{R}^n$ is the i^{th} column of A and $\top$ denotes transposition.

II. PRELIMINARIES

In this section, we recall model-based model order reduction by moment matching and its data-driven version. The exposition summarizes the relevant results of the survey [8].

A. Model-Based Moment Matching

Consider a linear, single-input, single-output, continuous-time, minimal system described by the equations

$$\dot{x} = Ax + Bu, \qquad y = Cx, \tag{1}$$

with state $x(t) \in \mathbb{R}^n$, input $u(t) \in \mathbb{R}$, output $y(t) \in \mathbb{R}$, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times 1}$, and $C \in \mathbb{R}^{1\times n}$. The moments of system (1) are defined as follows.

Definition 1. Let $s_i \in \mathbb{C}\backslash\sigma(A)$. The 0-*moment of system (1) at* s_i is the complex number $\eta_0(s_i) = C(s_iI - A)^{-1}B$. The k-*moment of system (1) at* s_i is the complex number $\eta_k(s_i) = \frac{(-1)^k}{k!}\left[\frac{d^k}{ds^k}(C(sI-A)^{-1}B)\right]_{s=s_i}$, with $k \geq 1$ integer.

Consider two sets of interpolation points $\mathcal{I}_1 = \{s_1, s_2, \dots, s_\nu\} \subset \mathbb{C}\backslash\sigma(A)$ and $\mathcal{I}_2 = \{s_{\nu+1}, s_{\nu+2}, \dots, s_{2\nu}\} \subset \mathbb{C} \setminus \sigma(A)$, with $\mathcal{I}_1 \cap \mathcal{I}_2 = \emptyset$, and consider two matrices $S \in \mathbb{R}^{\nu\times\nu}$ and $Q \in \mathbb{R}^{\nu\times\nu}$ such that $\sigma(S) = \mathcal{I}_1$ and $\sigma(Q) = \mathcal{I}_2$, and two vectors $L \in \mathbb{R}^{1\times\nu}$ and $R \in \mathbb{R}^{\nu\times 1}$ such that the pair (S, L) is observable and the pair (Q, R) is reachable. There exist invertible matrices T and $\bar{T}$ such that $C\Pi = \begin{bmatrix} \eta_0(s_1) & \dots & \eta_0(s_\nu) \end{bmatrix} T$ and $\Upsilon B =$

Zichen Zhao, Junyu Mao and Giordano Scarciotti are with the Department of Electrical and Electronic Engineering, Imperial College London, London, SW7 2AZ, UK. {zichen.zhao22, junyu.mao18, g.scarciotti}@imperial.ac.uk.

$\widetilde{T}\begin{bmatrix}\eta_0(s_{\nu+1}) & \dots & \eta_0(s_{2\nu})\end{bmatrix}^\top$, where $\Pi \in \mathbb{R}^{n\times\nu}$ and $\Upsilon \in \mathbb{R}^{\nu\times n}$ are the unique solutions of the Sylvester equations

$$\Pi S = A\Pi + BL, \tag{2a}$$
$$Q\Upsilon = \Upsilon A + RC. \tag{2b}$$

In *e.g.* [8] it has been shown that the matrices $C\Pi$, ΥB and $\Upsilon\Pi$ are all that is needed to construct reduced-order models by moment matching. In this section we recall how to obtain data-driven estimates of these three matrices. Consider the signal generator

$$\dot\omega = S\omega, \qquad \theta = L\omega, \tag{3}$$

with $\omega(t) \in \mathbb{R}^\nu$ and $\theta(t) \in \mathbb{R}$, and the interconnection (with $u = \theta$) between this generator and system (1), namely

$$\dot\omega = S\omega, \qquad \dot x = Ax + BL\omega, \qquad y = Cx. \tag{4}$$

This interconnection is called *direct interconnection*. Consider the filter

$$\dot\varpi = Q\varpi + R\eta, \tag{5}$$

with $\varpi(t) \in \mathbb{R}^\nu$ and $\eta(t) \in \mathbb{R}$, and the interconnection (with $\eta = y$) between this filter and system (1), namely

$$\dot x = Ax + Bu, \quad \dot\varpi = Q\varpi + RCx, \tag{6}$$

with $u = \delta_0$, where δ_0 indicates the Dirac-delta generalised function. This interconnection is called *swapped interconnection*. Finally, consider the signal generator (3), the filter (5), and the interconnection between these and system (1), yielding the system

$$\dot\omega = S\omega, \quad \dot x = Ax + BL\omega, \quad \dot\varpi = Q\varpi + RCx. \tag{7}$$

This interconnection is called *two-sided interconnection*.

In [2], it has been shown that under certain assumptions, the quantity

$$\operatorname{vec}(\widetilde{C\Pi}_k) := (\widetilde{W}_k^\top \widetilde{W}_k)^{-1}\widetilde{W}_k^\top \widetilde{Y}_k, \tag{8}$$

where, with $h \geq \nu$,

$$\begin{aligned}\widetilde{W}_k &= \begin{bmatrix}\omega(t_{k-h+1}) & \dots & \omega(t_{k-1}) & \omega(t_k)\end{bmatrix}^\top,\\ \widetilde{Y}_k &= \begin{bmatrix} y(t_{k-h+1}) & \dots & y(t_{k-1}) & y(t_k)\end{bmatrix}^\top,\end{aligned} \tag{9}$$

is an online estimate of $C\Pi$, namely there exists a sequence $\{t_k\}$ such that $\widetilde{C\Pi}_k$ is well-defined for all k and $\lim_{k\to\infty}\widetilde{C\Pi}_k = C\Pi$.

In [3], it has been shown that under certain assumptions, the quantity

$$\operatorname{vec}(\widetilde{\Upsilon B}_k) = (\widetilde{E}_k^\top \widetilde{E}_k)^{-1}\widetilde{E}_k^\top \widetilde{P}_k, \tag{10}$$

where, with $\widetilde{q} \geq 1$,

$$\begin{aligned}\widetilde{P}_k &= \begin{bmatrix}\varpi(t_{k-\widetilde{q}+1})^\top & \cdots & \varpi(t_{k-1})^\top & \varpi(t_k)^\top\end{bmatrix}^\top,\\ \widetilde{E}_k &= \begin{bmatrix} e^{Q^\top t_{k-\widetilde{q}+1}} & \dots & e^{Q^\top t_{k-1}} & e^{Q^\top t_k}\end{bmatrix}^\top,\end{aligned} \tag{11}$$

is an online estimate of ΥB.

Finally, in [4] it has been shown that under certain assumptions there exists $\gamma > 0$ such that as k approaches infinity,

$$||\widetilde{\Upsilon\Pi}_k - \Upsilon\Pi||_\infty \leq \gamma, \tag{12}$$

where

$$\operatorname{vec}(\widetilde{\Upsilon\Pi}_k) := (\widetilde{O}_k^\top \widetilde{O}_k)^{-1}\widetilde{O}_k^\top \widetilde{D}_k, \tag{13}$$

where, with $p \geq \nu$,

$$\widetilde{O}_k = \begin{bmatrix}\omega(t_{k-p+1})^\top \otimes I_\nu\\ \vdots\\ \omega(t_{k-1})^\top \otimes I_\nu\\ \omega(t_k)^\top \otimes I_\nu\end{bmatrix}, \ \widetilde{D}_k = \begin{bmatrix}\hat d(t_{k-p+1}) - \varpi(t_{k-p+1})\\ \vdots\\ \hat d(t_{k-1}) - \varpi(t_{k-1})\\ \hat d(t_k) - \varpi(t_k)\end{bmatrix}, \tag{14}$$

where $\hat d$ is the solution of

$$\dot{\hat d} = Q\hat d + \widetilde{\Upsilon B}_i L\omega, \qquad \hat d(0) = 0. \tag{15}$$

Alternatively, $\widetilde{\Upsilon\Pi}_k$ can be obtained by solving the Sylvester equation

$$Q\widetilde{\Upsilon\Pi}_k - \widetilde{\Upsilon\Pi}_k S = R\widetilde{C\Pi}_k - \widetilde{\Upsilon B}_k L, \tag{16}$$

where $\widetilde{C\Pi}_k$ is an estimate of $C\Pi$ and $\widetilde{\Upsilon B}_k$ is an estimate of ΥB.

III. Interconnection Noise, Measurement Noise and Statistical Modelling

In this section we describe the ways in which noise can affect the data-driven moment matching methods recalled in the previous section. In particular, we consider "interconnection noise" and "measurement noise". Then we describe the statistical models considered in the paper to describe the noise (Gaussian and Student's t-distribution).

A. Points of Introduction of Noise

Noise can affect the data-driven moment matching methods in two ways.

Interconnection noise: In data-driven model order reduction by moment matching, noise can be introduced during the interconnection of independent modules, such as the signal generator and the system in (4). Whether connecting electronic systems via wires, or mechanical systems via gears and rods, these interconnections may introduce noise. For example, imperfect electrical contact points can introduce resistance-based noise, while mechanical vibrations can result in noise from small physical displacements. We call noise introduced due to imperfect connections *interconnection noise*. Interconnection noise may affect data-driven model order reduction methods when connecting modules (system, generator, filter). In the direct interconnection (4), noise may be introduced when connecting the signal generator with the system. In the swapped interconnection (6), noise may be introduced when connecting the system and the filter. In the two-sided interconnection (7), noise may be introduced between generator and system and between system and filter.

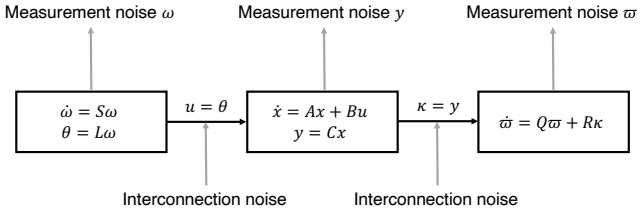

Fig. 1. Diagrammatic illustration of all sources of noise for the two-sided interconnection.

Measurement noise: Data-driven moment matching is based on the collection of time-domain samples to construct the data-matrices in (9), (11) and (14). When using instruments to measure these data, systematic errors may be introduced due to the inaccuracy of the instruments. These systematic errors are introduced in the form of noise which is added to the signal. We call this type of noise *measurement noise*. Measurement noise is introduced at the output of modules that need to be measured.

Fig. 1 provides an illustration of the various types of noise that can affect the two-sided interconnection. The figure shows that this experimental setup can be affected by measurement noise on three different quantities and interconnection noise on two points. Note also that the first half of this figure illustrates the direct interconnection, while the second half illustrates the swapped interconnection.

B. Statistical Modeling of Noise

We consider two noise distributions: Gaussian (*i.e.* white noise) and Student's t.

Gaussian distribution: Gaussian distribution is one of the most common ways of modelling noise in science and it represents white noise. White noise has constant power spectral density and statistical independence between different samples. These are properties which are used to represent a variety of different random perturbations. The mathematical model of white noise is a Gaussian distribution with probability density function

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},$$

where μ is the mean and σ is the standard deviation. In the simulations in this paper we have set $\mu = 0$ and changed σ depending on the signal-to-noise ratio (SNR) which we needed to produce.

Student's t-distribution: Considering noise with more abrupt or extreme fluctuations, we turn to the Student's t-distribution. This distribution can model the presence of outliers more effectively due to its long-tailed characteristic. The probability density function of the Student's t-distribution is given by

$$\varphi(x) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{\pi v}\Gamma(\frac{v}{2})} \left[1 + \frac{x^2}{v}\right]^{-\frac{v+1}{2}}, \tag{17}$$

where v is the number of degrees of freedom which determines whether the long-tailed property is significant or not and Γ is the gamma function. In the simulations in this paper we have set $v = 1$.

IV. Reducing the Impact of White Noise via Wavelet Denoising

In this section we first briefly recall the Wavelet denoising technique and explain the rationale of using it. Then we show the effect of the technique, first on the estimation of the parameters (*e.g.* $C\Pi$) and then on constructing the reduced-order model.

A. Wavelet Denoising and its Applicability

Since the Ordinary Least Squares (OLS) method is the best linear unbiased estimator when dealing with signals that include white noise, and formulas (8), (10) and (13) are already OLS, we turn to signal preprocessing prior to parameter estimation for this kind of noise. We have evaluated several denoising techniques such as time-domain filtering, Fourier Transform denoising, Short-Time Fourier Transform denoising, and Wavelet denoising. Overall, the best results were achieved with Wavelet denoising, which we now describe. Wavelet denoising is based on the Wavelet transform, defined by

$$W_{f(a,b)} = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t)\psi\left(\frac{t-b}{a}\right) dt, \tag{18}$$

where f is the function (signal), ψ is the wavelet, a is the scale factor, related to frequency, and b is the translation factor, related to time. Wavelet denoising adapts its window size based on signal characteristics, thereby fine-tuning time and frequency resolution without spectral leakage. Its localized approach is especially effective for handling white noise-induced changes. In particular, Wavelet denoising methods prove to be beneficial if the signal and noise are effectively separated at different intensities and scales. This is the case in data-driven moment matching methods based on the direct and on the two-sided interconnections as the quantities in these two interconnections are driven by a signal generator of known frequencies. On the other hand, the data-driven method based on the swapped interconnection utilises the impulse response as input instead of signals with narrow frequency bands. As such, we do not expect Wavelet denoising to be as beneficial for the estimation of ΥB.

B. Effect of Wavelet Denoising on Parameter Estimation

To assess the performance of Wavelet denoising in the estimation of $C\Pi$, ΥB and $\Upsilon\Pi$, we employ the ℓ_2-error. Let q be a quantity of interest and $\hat{q}$ its estimated value. The ℓ_2-error is defined as $\|\frac{\hat{q}-q}{q}\|_2$. The ℓ_2-error captures the cumulative error across all dimensions, representing the energy distribution of the error in the entire vector space.

In the simulation experiment we consider the direct interconnection (4) and the swapped interconnection (6) affected by interconnection noise and measurement noise. In particular, for the direct interconnection we have

$$\dot{\omega} = S\omega, \qquad \dot{x} = Ax + B(L\omega + w_1^i), \tag{19}$$

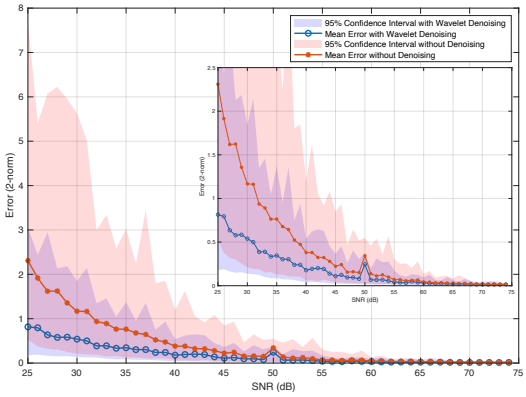

Fig. 2. ℓ_2 error of $C\Pi$, with (blue) and without (red) Wavelet denoising, for white noise. Insert: detail.

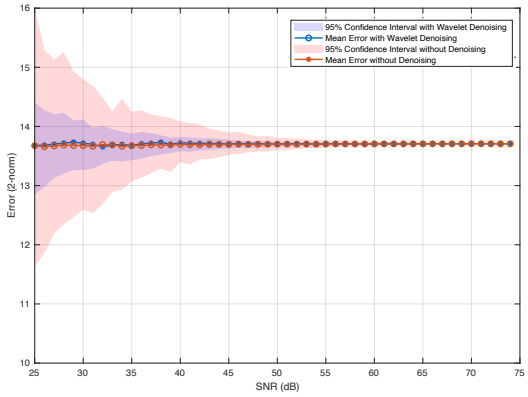

Fig. 3. ℓ_2 error of ΥB, with (blue) and without (red) Wavelet denoising, for white noise.

with

$$\omega_{\text{meas}} = \omega + w_1^m, \quad y_{\text{meas}} = y + w_2^m, \tag{20}$$

while for the swapped interconnection we have

$$\varpi_{\text{meas}} = \varpi + w_3^m, \tag{21}$$

where w_1^i is a white interconnection noise and w_1^m, w_2^m and w_3^m represent white measurement noise. We have omitted interconnection noise in the swapped interconnection because all our results are illustrated on the basis of SNR; since in the swapped interconnection the interconnection variable goes to zero exponentially, the SNR would change and this would make a comparison with the other results difficult.

In the experiment, we consider a SNR ranging from 25dB to 75dB. The data-driven model order reduction process was executed 500 times at each SNR level, using the non-denoised data and the Wavelet denoised data[1]. In the direct interconnection, the quantity $C\Pi$ was estimated using the OLS formula (8) with $h = 10\nu$. In the swapped interconnection, the quantity ΥB was estimated using the OLS formula (10) with $\widetilde{q} = 10$. The quantity $\Upsilon\Pi$ was estimated using (16). The experimental results for $C\Pi$ and ΥB are shown in Fig. 2 and 3, respectively (the behaviour of $\Upsilon\Pi$ is analogous to that of $C\Pi$ and thus the figure is omitted).

The blue curves in the figures show the mean ℓ_2-errors when using Wavelet denoised data across varying SNR levels, while the red curves show the mean errors when non-denoised data is used. The shaded region represents the 95% confidence interval in the two cases. Consistent lower blue curves and narrower blue regions confirm that the data-driven algorithms for the estimation of $C\Pi$ (and $\Upsilon\Pi$) benefit from pre-processing the data with Wavelet denoising when the data are affected by white interconnection and measurement noise. As expected, Wavelet denoising does not improve as much the results for ΥB when the swapped interconnection is affected by measurement noise. However, even though the mean error does not improve, the 95% confident region is still narrower when using Wavelet denoising. This suggests that using Wavelet denoising in this case still has the benefit of increasing the consistency of results for low values of SNR (large noise).

[1]This has been implemented in MATLAB with the following settings: Wavelet: Daubechies; number: 10; denoising method: Bayes; level of Wavelet decomposition: 10; thresholding rule: soft; method of estimating variance of noise: level-independent.

Remark 1. If the objective is to improve the estimation of ΥB, one can increase the parameter $\widetilde{q}$ in (10), as demonstrated in [6]. In fact, for all parameters ($C\Pi$, ΥB and $\Upsilon\Pi$) better results can be obtained by increasing the number of data points in the corresponding ordinary least square algorithms. However, the point of this paper is to study specifically the effect of data preprocessing in addition to the use of least squares.

C. Effect of Wavelet Denoising on Reduced-Order Models

The above experiment shows that Wavelet denoising provides more accurate estimates of $C\Pi$ and $\Upsilon\Pi$ (and more consistent estimates of ΥB) under the influence of white noise. We now look at the effect that these improvements have on the reduced-order model.

We measure the error using the $\mathcal{H}_\infty$-norm, comparing the difference between the transfer function of the reduced-order model and the transfer function of the original system. The experimental setup is identical to that of the previous section and thus it is not repeated here. The only difference is that in this section we repeat the experiment 1000 times for each SNR level to increase the accuracy of the results. Since the obtained reduced-order models may be unstable under the influence of noise, a stability check is performed before calculating the error. We compute the $\mathcal{H}_\infty$-error for the stable systems and we discard (but count the number of) the unstable systems. Fig. 4 shows the $\mathcal{H}_\infty$-errors of the Wavelet denoised (blue curve) and non-denoised (red curve) reduced-order models. The shaded areas (histogram) represent the 95% confidence interval of the error (with the same color coding) with an additional twist: the opacity of each bar in the histogram represents the proportion of stable systems in each of the 1000 iterations at each SNR level. The higher the opacity, the greater the proportion of stable systems. It can be observed that the blue curve is lower than the red curve and that the shaded blue region is narrower

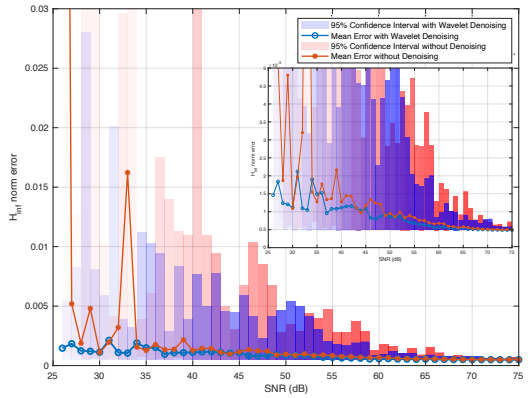

Fig. 4. $\mathcal{H}_\infty$-error of the reduced-order models, with (blue) and without (red) Wavelet denoising, for white noise. The opacity of the bars in the histogram represents the proportion of stable reduced-order models at each SNR level. Insert: detail.

than the red region at most SNR levels. This is particularly noticeable at smaller SNR (larger noise). Moreover, one can notice that the blue region is darker (more opaque) than the red region at each SNR level, indicating that Wavelet denoising increases the chance of obtaining a stable reduced-order model.

V. Reducing the Impact of Student's t Noise via Huber Regression

In this section we briefly recall the Huber regression method and explain the rationale of using it. Then we show the effect of the technique, first on the estimation of the parameters (*e.g.* $C\Pi$) and then on constructing the reduced-order model.

A. Huber Regression

We consider now noise drawn from the Student's t-distribution. Since this is not white noise, OLS is not necessarily the best approximating method. This distribution has long-tail and symmetric distribution characteristics and we choose Huber regression as an improvement. Huber regression combines ℓ_1 and ℓ_2 loss functions and exhibits excellent robustness to outliers, while being computationally efficient. Compared with some nonlinear methods or complex statistical models, the parameters of Huber regression are easier to interpret and select. We also note that Huber regression can handle many types of noise distributions (*i.e.* different from Student's t) without requiring specific modifications for each noise distribution.

The Huber loss function for a single data point (x, y) is defined as

$$L_K(y, f(x)) = \begin{cases} K|y - f(x)| - \frac{K^2}{2}, & |y - f(x)| > K, \\ \frac{(y-f(x))^2}{2}, & |y - f(x)| \leq K, \end{cases} \tag{22}$$

where $f(x)$ is the estimator and $K > 0$ is the robustness parameter that balances bias and robustness. In this work we select $K = 1.345$ (the default value in MATLAB).

B. Effect of Huber Regression on Parameter Estimation

In this section, we use the swapped and the two-sided interconnection, and consider only measurement noise. Thus, for the swapped interconnection we have

$$\varpi_{\text{meas}} = \varpi + w_4^m, \tag{23}$$

and for the two-sided interconnection we have

$$\omega_{\text{meas}} = \omega + w_5^m, \quad y_{\text{meas}} = y + w_6^m, \quad \varpi_{\text{meas}} = \varpi + w_7^m, \tag{24}$$

where w_4^m, w_5^m, w_6^m and w_7^m are drawn from from the Student's t-distribution. No interconnection noise is introduced. In the experiment, we consider an SNR ranging from 25dB to 75dB. The data-driven model order reduction process was executed 250 times at each SNR level, using OLS and Huber regression. In the swapped interconnection, the quantity ΥB was estimated using the OLS formula (10) with $\widetilde{q} = 10$ and Huber regression. In the two-sided interconnection the quantity $C\Pi$ was estimated using the OLS formula (8) with $h = 10\nu$ and Huber regression, whereas the quantity $\Upsilon\Pi$ was estimated using the OLS formula (13) with $p = 10\nu$ and Huber regression. Fig. 5 and 6 show the ℓ_2-error of $C\Pi$ and ΥB, respectively (the behaviour of $\Upsilon\Pi$ is analogous to that of $C\Pi$ and thus the figure is omitted). The color coding is the same as that of the figures in Section IV-B. From the figures it is evident that the blue curves are always much lower than the red curves at all SNR levels. This shows that when Huber regression is used for parameter estimation, the error obtained is significantly lower than that obtained using ordinary least square. Similarly, the blue shaded areas are significantly narrower than the red shaded areas at all SNR levels. This indicates that the estimates exhibit less sensitivity to noise when Huber regression is used, significantly mitigating the adverse effects of the randomness of noise on the accuracy of the estimates.

C. Effect of Huber Regression on Reduced-Order Models

The above experiment shows that Huber regression provides more accurate estimates of $C\Pi$, ΥB and $\Upsilon\Pi$ under the influence of Student's t noise. We now look at the effect that Huber regression has on the reduced-order model.

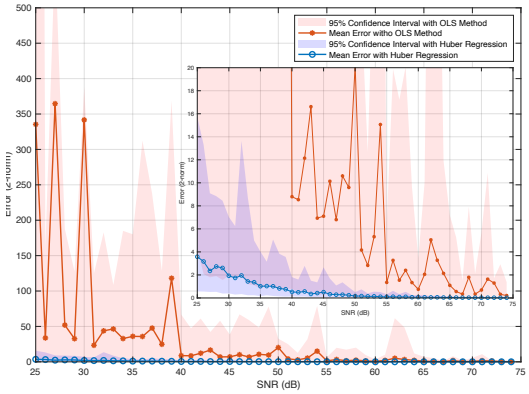

Fig. 5. ℓ_2 error of $C\Pi$, with Huber regression (blue) and OLS (red), for Student's t-noise. Insert: detail.

Fig. 7 shows the $\mathcal{H}_\infty$-errors of the reduced-order models obtained with the Huber regression (blue curve) and with OLS (red curve). The shaded areas (histogram) represent the 95% confidence interval of the error (with the same color coding). The opacity of each bar in the histogram represents the proportion of stable systems in each of the 250 iterations at each SNR level. The higher the opacity, the greater the proportion of stable systems. The opacity of the blue shaded bars is much higher than that of the red shaded bars for all SNR levels, suggesting that models reduced using Huber regression are much more likely to be stable. It is worth noting that at lower SNR levels, OLS was unable to produce a stable system in any of the 250 experimental repetitions. In contrast, the method using Huber regression was able to generate a small number of stable systems under these challenging conditions.

Finally, Fig. 8 shows a comparison of the Bode plots for an SNR of 65dB. The green curve represents the original system. The blue and red shaded areas represent the 95% confidence interval obtained by constructing reduced-order models using Huber regression and OLS, respectively. The figure clearly demonstrates that the reduced-order models obtained using Huber regression are closer to the Bode plot of the original system and are less sensitive to noise.

Together, these observations highlight that the method utilizing Huber regression not only produce better reduced-order models but it also increases the chance of obtaining a stable reduced-order model.

VI. Conclusions

In this paper we have presented strategies to reduce the impact of noise in the context of data-driven model order reduction by moment matching. By means of extensive simulations, we have shown that Wavelet denoising is beneficial in the presence of white noise in the estimation of $C\Pi$ and $\Upsilon\Pi$ and that overall it reduces the sensitivity of the reduced-order model to randomness. We have also shown that Huber regression outperformed OLS substantially in the presence of Student's t-distribution noise: all three parameters ($C\Pi$, ΥB and $\Upsilon\Pi$) greatly benefited from Huber regression and the chance of obtaining a stable reduced-order model was significantly increased.

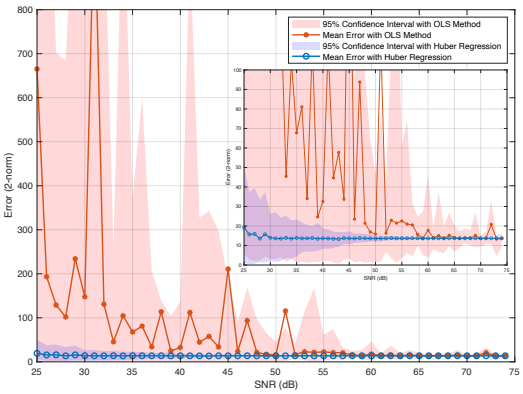

Fig. 6. ℓ_2 error of ΥB, with Huber regression (blue) and OLS (red), for Student's t-noise. Insert: detail.

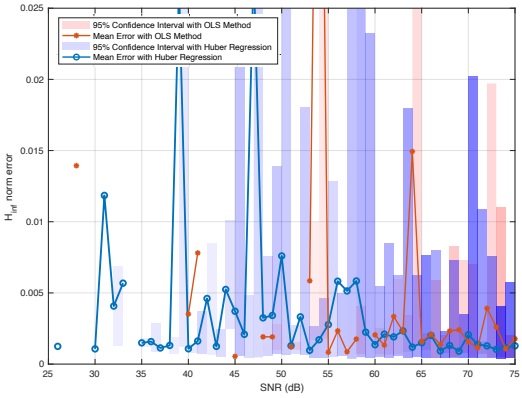

Fig. 7. $\mathcal{H}_\infty$-error of the reduced-order models, with Huber regression (blue) and OLS (red), for Student's t noise. The opacity of the bars in the histogram represents the proportion of stable reduced-order models at each SNR level.

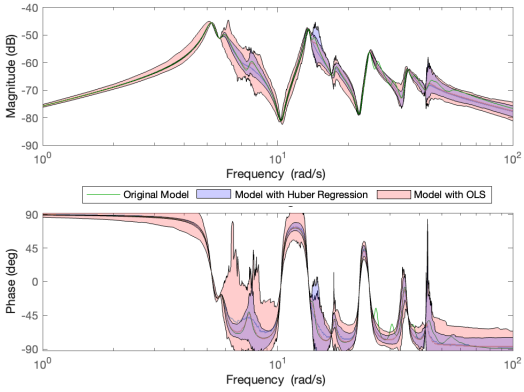

Fig. 8. Bode plot with Student's t-noise at an SNR of 65dB: original (green), Huber (blue) and OLS (red).

References

[1] A. C. Antoulas, *Approximation of Large-Scale Dynamical Systems*. Philadelphia, PA: SIAM Advances in Design and Control, 2005.

[2] G. Scarciotti and A. Astolfi, "Data-driven model reduction by moment matching for linear and nonlinear systems," *Automatica*, vol. 79, pp. 340–351, 2017.

[3] J. Mao and G. Scarciotti, "Data-driven model reduction by moment matching for linear systems through a swapped interconnection," in *Proceedings of the 20th European Control Conference*, 2022, pp. 1690–1695.

[4] ——, "Data-driven model reduction by two-sided moment matching," *arXiv:2212.08589*, 2022.

[5] G. Scarciotti, "Low computational complexity model reduction of power systems with preservation of physical characteristics," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 743–752, 2017.

[6] Z. Gong, J. Mao, A. Junyent-Ferre, and G. Scarciotti, "Model order reduction of large-scale wind farms: A data-driven approach," *Submitted to IEEE Transactions on Sustainable Energy*, 2024.

[7] SLICOT, "Benchmark examples for model reduction," http://slicot.org/20-site/126-benchmark-examples-for-model-reduction, accessed: 2017-02-13.

[8] G. Scarciotti and A. Astolfi, "Interconnection-based model order reduction - A survey," *European Journal of Control*, vol. 75, p. 100929, 2024.