Research Article

Mixed L_p Estimators Variety for Model Order Reduction in Control Oriented System Identification

Christophe Corbier^{1,2,3} and Jean-Claude Carmona⁴

¹Université de Lyon, 42023 Saint-Étienne, France

²Université de Saint-Étienne (Jean Monnet), 42000 Saint-Étienne, France

³LASPI, IUT Roanne, 42334 Roanne, France

⁴Laboratoire des Sciences de l'Information et des Systèmes, UMR CNRS, ENSAM, 13617 Aix-en-Provence, France

Correspondence should be addressed to Christophe Corbier; christophe.corbier@univ-st-etienne.fr

Received 5 May 2014; Accepted 2 July 2014

Academic Editor: Guido Maione

Copyright © 2015 C. Corbier and J.-C. Carmona. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A new family of MLE type L_p estimators for model order reduction in dynamical systems identification is presented in this paper. A family of L_p distributions proposed in this work combines L_{p_2} ($1 < p_2 < 2$) and L_{p_1} ($0 < p_1 < 1$) distributions which are quantified by four parameters. The main purpose is to show that these parameters add degrees of freedom (DOF) in the estimation criterion and reduce the estimated model complexity. Convergence consistency properties of the estimator are analysed and the model order reduction is established. Experimental results are presented and discussed on a real vibration complex dynamical system and pseudo-linear models are considered.

1. Introduction

The choice of the norm is a fundamental problem in system identification. For some special cases, the least squares (LS) ($p = 2$), the least absolute deviation (LAD) ($p = 1$), and the Chebyshev estimator ($p = \infty$) methods have been proposed by many authors [1, 2]. It is well known that the L_1 norm estimators are efficient when the noise follows the Laplace distribution, the L_2 norm estimators are efficient when the noise follows the Gaussian distribution, and the L_∞ norm estimators are efficient when the noise follows the uniform distribution. Furthermore, it has been shown that the L_p norm estimator is the MLE if noise is a generalized p -Gaussian [3]. Thus, we can always find a suitable exponent p for the L_p norm method. In the reference paper [4], the authors discuss smooth and sensitive L_p norms for prediction error system identification when disturbances are magnitude-bounded. They show that a necessary condition for L_p norm to be statistically robust with respect to the family $\tilde{\mathcal{P}}_p(\bar{\omega})$ of the $\bar{\omega}$ -contaminated distribution model with support $[-\beta, \beta]$ for some arbitrary $\beta > 0$ is that its second

derivative does not vanish on the support and must be strictly positive with respect to $\tilde{\mathcal{P}}_p(\bar{\omega})$. This latter condition is fundamental in our framework by ensuring the parameter estimate variance convergence rate (see Proposition 4.2 and Corollary 4.3 [4]). However, many drawbacks come from the use of the L_p norm. Firstly, p is the unique DOF of the norm. Secondly, the support is a restricted condition to deal with the outliers [5]. Thirdly, the choice of p may prevent robustness to the large data. Finally, most of the time, L_∞ norm is used to the model reduction for robust control [6] with the restricted case where the model parameters number is limited. The L_p norms are widely used in many domains [7, 8]. Even though the formal framework presents some difficulties, previous works show the interest to use these estimators. In system identification, Chen et al. in [9] propose an ARMA robust system identification using a generalized L_p norm estimation algorithm. The authors use this norm to identify a system with a non-Gaussian noise, since the classical LSE is related to the situation when the noise distribution is white-Gaussian. Likewise, these authors in [10] propose the parameter estimation of linear systems

with input-output noisy data, supposed to be corrupted by measurement noise unknown distributions.

Here, we deal with the problem of the L_p estimation in system identification in the presence of atypical data (outliers) in the vibration dynamical system output signal, in order to reduce the model order for the controller design. In such a vibration dynamical system, *natural outliers* occur in the output signal, increasing considerably the model order in the estimation procedure.

To avoid this problem, a new variety of MLE type L_p estimators are considered. We define a mixed L_p distribution with multiple DOF containing $1 < p_2 < 2$ and $0 < p_1 < 1$ [11], parametrized by a threshold named *scaling factor*, denoted by $\tilde{\eta}$. To ensure both the parameter estimate variance convergence and the model order reduction and apply conditions in [4] related to the second derivative of the L_p norm on the support with respect to $\tilde{\mathcal{P}}_p(\bar{\omega})$, we show that there exists a tradeoff between p_2 , p_1 , and $\tilde{\eta}$. This reduction is efficient only with $\tilde{\eta} < 1$ and $p_2 < 2$, involving $p_1 < 1$.

Model reduction has been subject to considerable interest in many domains. In [12], a method of the model reduction of the nonlinear complex Ginzburg-Landau equation is presented. In [13] the authors propose dimension reduction and dynamics of a spiking neural network model for decision making under neuromodulation. The model order reduction is a fundamental step, especially in control design. To estimate a low order model of a system, several possibilities exist. The most obvious one is to directly estimate a lower order model. As known from, for example, Ljung [14], the prediction/output error estimate automatically gives models that are L_2 approximations of the *true system* in frequency-weight norm, determined by the input spectrum and noise model. See [14, chapter 8] for more details. A second possibility is to estimate a high order model which is then subjected to model reduction to the desired order. Indeed, some contributions that take into account that the high order model is obtained through an identification experiment when performing model reduction are Porat and Friedlander [15], Porat [16], Söderström et al. [17], Stoica and Söderström [18], Zhu and Backx [19, chapter 7] and Tjärnström and Ljung [20], and Tjärnström [21]. Porat and Friedlander study an ARMA parameter estimation via covariance estimates. Recently, in time domain, [22] presents a relevant paper on a new model order reduction algorithm based on general orthogonal polynomials. The following contributors deal with models having input signals. Söderström et al. [17] focus on model structures that can be embedded in larger structures which are easy to estimate, such as ARX structures. After estimating the high order structure, they reduce the estimate to the low order structure in a weighted nonlinear least-squares sense. The method is called *indirect prediction error method*. In [20], the authors focus on L_2 model reduction. They study FIR and OE models. The former is estimated directly from data and the latter is computed by reducing a high order model, by L_2 model reduction. For OE models, they show that the reduced model has the same variance as the directly estimated one, if the reduced model class used contains the true system. Recently, in [23], the authors illustrate procedures to identify

a state-space representation of a lossless or dissipative system from a given noise-free trajectory. The idea is to perform model reduction by obtaining a balanced realization directly from data and truncating it to obtain a reduced order model. However, these different model order reduction methods are performed after estimation phases with high model orders. On the other hand, this reduction is made in favorable conditions, where no outliers are present in the dataset and the prediction errors. In practice, we all know that real conditions lead to model complexity strongly increased. The lack of robustness of these methods damages the estimation and the model order. Different works have been performed, using robust estimators. A first alternative is the use of the L_1 norm [24–26], as a possibility to reduce the model complexity to identify a complex dynamical system, for example, an acoustic duct for active noise control (ANC). The authors propose relevant OE (output error) L_1 models with complexity $d_M^{L_1}$ equal to 21. The second one is to use a symmetric convex Huber's function [27]. Recently, in [28, 29], the identification of the same plant, using $L_2 - L_1$ estimators with low values of the scaling factor in Huber's function, proposes OE- $L_2 - L_1$ models with complexity d_M^H equal to 17.

Here, we will tackle the $L_{p_2} - L_{p_1}$ distribution approach and, from this, we will propose a L_p parameterized robust estimation criterion (PREC) using a symmetric convex function, offering to the user more *flexibility* in terms of an extra DOF, in order to improve the estimation balance robustness/performances. We will show that the number of DOF in this function facilitates the model order reduction. We will see that the reduced order models are provided by the $L_{p_2} - L_{p_1}$ estimation itself, where classical approaches separate the phases, estimation and then reduction. Indeed, we will demonstrate that the L_p model order thus obtained remains lower than the corresponding $L_2 - L_1$ estimate. Finally, experimental results on the acoustic duct will confirm these results.

We start in Section 2 by introducing the mathematical background related to definition and properties of the L_p estimator. We develop the consistency and convergence properties of this estimator. The application to the system identification is presented in Section 3. Section 4 proposes the main results concerning the model order reduction. In Section 5, the improvement of the model order reduction with respect to the number of DOF is presented. Experimental results based on a vibration dynamical system are presented and discussed in Section 6. Conclusions and perspectives are drawn in Section 7.

2. Definition and Properties of the L_p Estimators

The new family of L_p -distributions proposed in this paper combines L_{p_2} and L_{p_1} distributions, quantified by four parameters, p_2 ($1 < p_2 < 2$), p_1 ($0 < p_1 < 1$), ϕ ($\phi > 0$), and $\tilde{\eta}$ ($0 < \tilde{\eta} < 1$). Hence, it can be viewed as a generalization of the *redescending* mixed distributions, which corresponds to the M-estimates $p_2 = 2$ and $p_1 = 1$ (see [27, chapter 4, page 84–85]). A formal definition of a probability density function (pdf) of a random variable is as follows.

Definition 1. Consider a function $f_p : \mathbb{R} \rightarrow \mathbb{R}_+^*$, $X \mapsto f_p(X)$ such that

$$f_p(X) = \begin{cases} f_{p_2}(X) = Ce^{-|X|^{p_2}/2\phi^{p_2}} & \text{for } |X| \leq \tilde{\eta} \\ f_{p_1}(X) = Ce^{-\tilde{\eta}|X|^{p_1}/\phi^{p_2} + \tilde{\eta}^2/2\phi^{p_2}} & \text{for } |X| > \tilde{\eta}, \end{cases} \quad (1)$$

where $C = 1/2(K_1 + K_2)$ with

$$\begin{aligned} K_1 &= e^{\tilde{\eta}^2/2\phi^{p_2}} \frac{\phi^{p_2/p_1}}{p_1 \tilde{\eta}^{1/p_1}} \Gamma\left(\frac{1}{p_1}, \frac{\tilde{\eta}^{p_1+1}}{\phi^{p_2}}\right) \quad \text{for } |X| > \tilde{\eta}, \\ K_2 &= \frac{\phi^{2^{1/p_2}}}{p_2} \left[\Gamma\left(\frac{1}{p_2}\right) - \Gamma\left(\frac{1}{p_2}, \frac{\tilde{\eta}^{p_2}}{2\phi^{p_2}}\right) \right] \quad \text{for } |X| \leq \tilde{\eta}, \end{aligned} \quad (2)$$

where $\Gamma(a)$ and $\Gamma(a, X)$ are, respectively, the complete and incomplete Euler's gamma functions. We can easily verify that, for all $X \in \mathbb{R}$, $f_p(X) \geq 0$ and $\int_{\mathbb{R}} f_p(X) dX = 1$, which ensure that f_p is a probability density. The parameter $\tilde{\eta}$ is defined as a threshold named *scaling factor*. Its value depends on the degree of the contamination of the random variables X . Moreover, a low value will ensure a good robustness and a high value a good efficiency of the estimation.

The pdf $f_p(X)$ belongs to the maximum likelihood estimators (MLE) class. Then, there exists a symmetric function variety $\rho_{\tilde{\eta}}^p(X)$ such that $\rho_{\tilde{\eta}}^p(X) \sim -\log f_p(X)$.

Definition 2. Let one define a new estimator variety as a L_p estimator, denoted by $\hat{\theta}_N^p$. Let $(S, \mathcal{S}, P)$ be a probability space and let $\{X_t\}_{t=1}^N$ be a sequence of *i.i.d.r.v.s* with values in S . Let Θ be a Borel subset in $\mathbb{R}^d$. Let $\rho_{\tilde{\eta}}^p : S \times \Theta \rightarrow \mathbb{R}$ be a symmetric function such that $\rho_{\tilde{\eta}}^p(\bullet(\theta)) : S \rightarrow \mathbb{R}$ is measurable for each $\theta \in \Theta$. The L_p estimator $\hat{\theta}_N^p$ is defined by a minimum of the form

$$N^{-1} \sum_{t=1}^N \rho_{\tilde{\eta}}^p(X_t(\hat{\theta}_N^p)) = \inf_{\theta \in \Theta} N^{-1} \sum_{t=1}^N \rho_{\tilde{\eta}}^p(X_t(\theta)), \quad (3)$$

where

$$\rho_{\tilde{\eta}}^p(X) = \begin{cases} \frac{|X|^{p_2}}{2} & \text{for } |X| \leq \tilde{\eta} \\ \tilde{\eta}|X|^{p_1} - \frac{\tilde{\eta}^2}{2} & \text{for } |X| > \tilde{\eta} \end{cases} \quad (4)$$

with $1 < p_2 < 2$ and $0 < p_1 < 1$ for all $\tilde{\eta} > 0$.

In robust statistics, continuity conditions are necessary to ensure good estimations [27, chapter 2, page 24-25]. Therefore, these conditions must be applied for both $\rho_{\tilde{\eta}}^p$ and its first derivative. This latter must be a bounded continuous function. From these considerations, we define the dimensional continuity conditions (DCC) as follows:

$$\text{DCC} : \begin{cases} \rho_{\tilde{\eta}}^{p_2}(\pm\tilde{\eta}) = \rho_{\tilde{\eta}}^{p_1}(\pm\tilde{\eta}) \\ \left(\frac{\partial}{\partial X} \rho_{\tilde{\eta}}^{p_2}(X) \right)_{\pm\tilde{\eta}} = \left(\frac{\partial}{\partial X} \rho_{\tilde{\eta}}^{p_1}(X) \right)_{\pm\tilde{\eta}}, \end{cases} \quad (5)$$

where $\rho_{\tilde{\eta}}^{p_2}(X) = |X|^{p_2}/2$ and $\rho_{\tilde{\eta}}^{p_1}(X) = \tilde{\eta}|X|^{p_1} - \tilde{\eta}^2/2$.

The dimensional p_1 is given from DCC and, after straightforward calculations, we obtain

$$p_1 = \frac{p_2}{1 + \tilde{\eta}^{2-p_2}}. \quad (6)$$

Equation (6) allows to show that, for the conditions $\tilde{\eta} < 1$ and $p_2 < 2$, we get $p_1 < 1$. We now define the gross error model (GEM) in the mixed L_p framework as a generalization of the GEM given in [27, chapter 1, page 12].

Definition 3. Let F_{p_2} be a fixed probability distribution having a twice differentiable density f_{p_2} , such that $-\log f_{p_2}$ is convex on the convex support of F_{p_2} . Let $\bar{\omega} > 0$ be the level of contamination and let $\tilde{\mathcal{P}}_p(\bar{\omega})$ be the set of all probability distributions arising from F_{p_2} through the $\bar{\omega}$ -contamination model

$$\tilde{\mathcal{P}}_p(\bar{\omega}) = \{F_p \mid F_p = (1 - \bar{\omega})F_{p_2} + \bar{\omega}H, H \in \mathcal{M}\}. \quad (7)$$

Here, $\mathcal{M}$ is the set of all probability measures on the real line. The corresponding pdf of F_p , contained in $\tilde{\mathcal{P}}_p$, is given by (1), and it puts all contamination outside $[-\tilde{\eta}, \tilde{\eta}]$. A suitable value of $\tilde{\eta}$ should efficiently reduce the propagation of outlier in the prediction errors. This is done by $\rho_{\tilde{\eta}}^{p_1}$. Accordingly, the L_{p_1} treatment of the estimated residuals is focused on the outlier occurrence, so the following errors rapidly decrease and they are advantageously treated by $\rho_{\tilde{\eta}}^{p_2}$.

Now, our goal is to provide the relation between the level of contamination $\bar{\omega}$ and the scaling factor $\tilde{\eta}$. In practice, the robustness is defined by a suitable value of $\tilde{\eta}$ that fixes the $\bar{\omega}$ -contamination model.

Theorem 4. If one adjusts the scaling factor $\tilde{\eta}$ in $\rho_{\tilde{\eta}}^p$, such that $\bar{C} = (1 - \bar{\omega})C$, from the condition $\int_{-\infty}^{\infty} \bar{C} e^{-\rho_{\tilde{\eta}}^p(X)/\sigma^{p_2}} dX = 1$, $\bar{\omega}$ and $\tilde{\eta}$ are linked by

$$\mathcal{A} + \mathcal{B} = \frac{1}{1 - \bar{\omega}} \quad (8)$$

with

$$\begin{aligned} \mathcal{A} &= \frac{\phi^{2^{1/p_2}}}{p_2(K_1 + K_2)} \left[\Gamma\left(\frac{1}{p_2}\right) - \Gamma\left(\frac{1}{p_2}, \frac{\tilde{\eta}^{p_2}}{2\phi^{p_2}}\right) \right] \\ \mathcal{B} &= \frac{e^{\tilde{\eta}^2/2\phi^{p_2}} \phi^{p_2/p_1}}{(K_1 + K_2) p_1 \tilde{\eta}^{1/p_1}} \Gamma\left(\frac{1}{p_1}, \frac{\tilde{\eta}^{p_1+1}}{\phi^{p_2}}\right), \end{aligned} \quad (9)$$

where $\Gamma(a, X) = \int_X^{\infty} Z^{a-1} e^{-Z} dZ$, $a > 0$, is incomplete Euler's gamma function.

See Appendix A for the proof. Notice that in the limit case, where $(p_2, p_1) = (2, 1)$ (M-estimates), we obtain, for $\eta = k\sigma$ ($\phi = \sigma$), $\mathcal{A} = 1 - 2\Phi(-k)$ and $\mathcal{B} = 2\phi(k)/k$. See [27, chapter 4, page 84].

Figure 1 shows an example of the level of contamination $\bar{\omega}$ as a function of $\tilde{\eta}$, p_1 , parameterized by p_2 , with $\phi = \sigma = 1$. A large value of the scaling factor means that the probability

FIGURE 1: Level of contamination $\bar{w}$ is plotted as a function of $\tilde{\eta}$, p_1 , parameterized by p_2 and $\phi = \sigma = 1$.

distribution of the prediction errors is weakly disturbed, corresponding to $\bar{w} < 0.1$ (see Figure 1). Conversely, a small value corresponds to a pdf more disturbed. For $p_2 \leq 1.7$, the curves have a minimum, meaning that a particular value of the scaling factor renders the corresponding pdf weakly contaminated.

Definition 5. Let $\rho_{\tilde{\eta}}^p(X)$ be a symmetric function given by

$$\rho_{\tilde{\eta}}^p(X) = \frac{|X|^{p_2}}{2} \mathbb{I}_{\{|X| \leq \tilde{\eta}\}} + \left(\tilde{\eta} |X|^{p_1} - \frac{\tilde{\eta}^2}{2} \right) \mathbb{I}_{\{|X| > \tilde{\eta}\}}, \quad (10)$$

where $\mathbb{I}_{\{Y\}}$ is the unit function with the condition Y . Let us introduce two index subsets ν_{p_2} and ν_{p_1} , respectively, related to the L_{p_2} and L_{p_1} contributions, defined by $\nu_{p_2}(\theta) = \{t : |\varepsilon_t(\theta)| \leq \tilde{\eta}\}$ and $\nu_{p_1}(\theta) = \{t : |\varepsilon_t(\theta)| > \tilde{\eta}\}$. From this, the L_p -PREC to be minimized is given by

$$g_N^p(\theta) = \frac{1}{N} \sum_{t \in \nu_{p_2}(\theta)} \frac{|\varepsilon_t(\theta)|^{p_2}}{2} + \frac{\tilde{\eta}}{N} \sum_{t \in \nu_{p_1}(\theta)} \left(|\varepsilon_t(\theta)|^{p_1} - \frac{\tilde{\eta} s_t^2(\theta)}{2} \right), \quad (11)$$

where $\varepsilon_t(\theta)$ is the estimated modeling error and $s_t(\theta)$ is the sign function such that $s_t(\theta) = 0$ for $|\varepsilon_t(\theta)| \leq \tilde{\eta}$, $s_t(\theta) = -1$ for $\varepsilon_t(\theta) < -\tilde{\eta}$, and $s_t(\theta) = 1$ for $\varepsilon_t(\theta) > \tilde{\eta}$.

In order to prove the consistency and convergence of the robust estimator $\hat{\theta}_N^p$, we need the following lemma related to the moment of order r of the random variable U .

Lemma 6. For the considered pdf f_p , the moments of order r ($r \in \mathbb{N}$) of a random variable U are

$$E(|U|^r) = \mathcal{J}_1^r + \mathcal{J}_2^r \quad (12)$$

with

$$\begin{aligned} \mathcal{J}_1^r &= \left(\frac{\phi^{r+1} 2^{(r+1)/p_2}}{p_2} \right. \\ &\quad \times \left[\Gamma\left(\frac{r+1}{p_2}\right) - \Gamma\left(\frac{r+1}{p_2}, \frac{\tilde{\eta}^{p_2}}{2\phi^{p_2}}\right) \right] \\ &\quad \times (K_1 + K_2)^{-1}, \\ \mathcal{J}_2^r &= \left(e^{\tilde{\eta}^2/2\phi^{p_2}} \frac{\phi^{(p_2/p_1)(r+1)}}{p_1 \tilde{\eta}^{(r+1)/p_1}} \right. \\ &\quad \times \Gamma\left(\frac{r+1}{p_1}, \frac{\tilde{\eta}^{p_1+1}}{\phi^{p_2}}\right) \\ &\quad \times (K_1 + K_2)^{-1}. \end{aligned} \quad (13)$$

See Appendix B for the proof.

Theorem 7. Let Θ be a compact subset of $\mathbb{R}^d$ and let $\hat{\theta}_N^p$ be the estimator obtained as a solution of the problem

$$\inf_{\theta \in \Theta} N^{-1} \sum_{t=1}^N (-\ln f_p(\varepsilon_t(\theta))). \quad (14)$$

Let θ^* be the global minimum obtained as a solution of the problem

$$\inf_{\theta \in \Theta} N^{-1} \sum_{t=1}^N (-E \ln f_p(\varepsilon_t(\theta))), \quad \text{as } N \rightarrow \infty, \quad (15)$$

and then $\hat{\theta}_N^p$ is a consistent estimate of θ^* ; that is,

$$\hat{\theta}_N^p \rightarrow \theta^*, \quad \text{w.p.1.} \quad (16)$$

In order to do so, we use the (MLE) consistency result by Newey and McFadden [30] and show that all the assumptions of their theorem hold.

Theorem 8. Suppose that $z_i (i = 1, 2, \dots)$ are iid rv's with pdf $f(z_i | \theta^*)$ and

- (i) $\theta \neq \theta^* \Rightarrow f(z_i | \theta) \neq f(z_i | \theta^*)$;
- (ii) $\theta^* \in \Theta$, which is compact;
- (iii) $\ln f(z_i | \theta)$ is continuous at each $\theta \in \Theta$ w.p.1;
- (iv) $E[\sup_{\theta \in \Theta} |\ln f(z_i | \theta)|] < \infty$.

Then $\hat{\theta}_N \xrightarrow{\text{prob}} \theta^*$.

See Appendix C for the proof.

Remark 9. Before presenting the theorem related to the asymptotic distribution of the L_p estimator, let us notice that the standard asymptotic normality results for MLE requires that the PREC be twice continuously differentiable, which is not the case here. There exist, however, asymptotic normality results for nonsmooth functions and we will hereafter use the one proposed by Newey and McFadden [30] and Andrews [31]. The basic insight of their approaches is that the smoothness condition of the PREC, $g_N^p(\theta)$, can be

replaced by a smoothness of its limit, which in the standard maximum likelihood case corresponds to the expectation $-E \ln f_p(\varepsilon_t(\theta)) = g^p(\theta)$, with the requirement that certain remainder terms are small. Hence, the standard differentiability assumption is replaced by a stochastic differentiability condition, which can then be used to show that the MLE $\hat{\theta}_N^p$ is consistent and asymptotically normal. Let us consider the Ψ -function as the derivative with respect to θ of ρ_{η}^p denoted by $\Psi_t(\varepsilon; \theta)$. If this function is differentiable in θ , one can establish the asymptotic normality of $\hat{\theta}_N^p$ by expanding $\sqrt{N}(\hat{\theta}_N^p - \theta^*)$ about θ^* using element by element mean value expansions. This is the standard way of establishing asymptotic normality of the estimator. In a variety of applications, however, $\Psi_t(\varepsilon; \theta)$ is not differentiable in θ , or not even continuous, due to the appearance of a sign function. In such a case, one can still establish asymptotic normality of the estimator provided $E\Psi_t(\varepsilon; \theta)$ is differentiable in θ . Since the expectation operator is a smoothing operator, $E\Psi_t(\varepsilon; \theta)$ is often differentiable in θ even though $\Psi_t(\varepsilon; \theta)$ is not.

Let us consider the following theorem establishing the asymptotic distribution of the estimator.

Theorem 10. Let $\varepsilon_1(\hat{\theta}_N^p), \dots, \varepsilon_N(\hat{\theta}_N^p)$ be iid rv's from the pdf f_p with an unknown parameter θ^* , $\theta^* \in B^\circ$ with B a compactness and B° interior of B . Then the MLE $\hat{\theta}_N^p$ of θ^* is consistent and asymptotically normal:

$$\sqrt{N}(\hat{\theta}_N^p - \theta^*) \xrightarrow{d} \mathcal{N}(0, \mathcal{P}(\theta^*)), \quad (17)$$

where $\mathcal{P}(\theta^*)$ is the asymptotic covariance matrix.

In order to do so, we use Theorem 7.1 given by Newey and McFadden [30] and show that all the assumptions of their theorem hold.

Theorem 11. Suppose $g_N^p(\hat{\theta}_N^p) \geq \sup_{\theta \in D_M} g_N^p(\theta) - o_p(N^{-1})$, $\hat{\theta}_N^p \xrightarrow{prob} \theta^*$, and

- (i) $g^p(\theta)$ is maximized on D_M at θ^* ;
- (ii) θ^* is an interior point of D_M ;
- (iii) $g^p(\theta)$ is twice differentiable at θ^* with nonsingular second derivative $\bar{g}''^p(\theta)$;
- (iv) $\sqrt{N}(E(\partial g_N^p(\theta)/\partial \theta))_{\hat{\theta}_N^p} \xrightarrow{d} \mathcal{N}(0, Q(\theta^*))$;
- (v) for any $\delta_N \rightarrow 0$, $\sup_{\|\hat{\theta}_N^p - \theta^*\| \leq \delta_N} |\hat{R}_N(\hat{\theta}_N^p)/(1 + \sqrt{N}\|\hat{\theta}_N^p - \theta^*\|)| \xrightarrow{prob} 0$ with the remainder

$$\begin{aligned} \hat{R}_N(\theta) &= \sqrt{N} \frac{g_N^p(\theta) - g_N^p(\theta^*) - (\partial g_N^p(\theta)/\partial \theta)_{\theta^*}(\theta - \theta^*)}{\|\theta - \theta^*\|} \\ &\quad + \frac{-g^p(\theta) + g^p(\theta^*)}{\|\theta - \theta^*\|}, \end{aligned} \quad (18)$$

and then $\sqrt{N}(\hat{\theta}_N^p - \theta^*) \xrightarrow{d} \mathcal{N}(0, \mathcal{P}(\theta^*))$.

The proof is given in Appendix D.

3. Application to the System Identification

Having defined the mixed L_p framework, our purpose is now to apply this approach to the system identification by introducing the L_p -PREC and its first and second derivatives, to express the asymptotic L_p covariance matrix of the L_p estimator. Before this, let us recall the prediction error context.

3.1. Prediction Error Approach. Throughout the paper, we denote the input signal by u_t , the output signal by y_t , and the disturbance signal by e_t as a random variable sequence with zero mean and variances λ , and N is the total number of measured data. We assume that y_t is generated according to

$$y_t = G_0(q) u_t + H_0(q) e_t, \quad (19)$$

where $G_0(q)$ is a linear time-invariant system, usually referred to as the *true system* and $H_0(q)$ is the noise filter. In the case of OE models, the model set is parameterized by a d -dimensional real-valued parameter vector θ ; that is,

$$y_t = G(q, \theta) u_t + H(q, \theta) e_t = \frac{\mathcal{B}(q, \theta)}{\mathcal{F}(q, \theta)} u_t + e_t, \quad (20)$$

with $\mathcal{F}(q, \theta) = 1 + \sum_{i=1}^{n_F} f_i q^{-i}$, $\mathcal{B}(q, \theta) = \sum_{i=1}^{n_B} b_i q^{-i}$, and $\theta = [b_1 \dots b_{n_B} f_1 \dots f_{n_F}]^T$. Moreover, q^{-1} is the lag operator such that $q^{-l} y_t = y_{t-l}$, $l \in \mathbb{N}$. The prediction error (also named estimated residuals) is $\varepsilon_t(\theta) = y_t - \hat{y}_t(\theta)$ with $\hat{y}_t(\theta) = \varphi_t^T(\theta) \theta$ the prediction model, $\varphi_t(\theta) = [u_{t-1} \dots u_{t-n_B} - \hat{y}_{t-1}(\theta) \dots - \hat{y}_{t-n_F}(\theta)]^T$ the regressor, and $\psi_t(\theta) = (\partial/\partial \theta) \hat{y}_t(\theta)$ the gradient of $\hat{y}_t(\theta)$ [14].

3.2. Limit Hessian and Q_{ω}^p -Matrix. The main purpose of this section is to express the asymptotic covariance matrix of the estimator $\hat{\theta}_N^p$ for prediction error estimates in case, essentially, a true description of the system is available in the model set. The condition then becomes $\mathcal{S} \in \mathcal{M}$, where $\mathcal{M}$ is a model structure, a mapping from a parameter space to a set of models. Let us consider the convergence domain defined by $D_c = \arg \min_{\theta \in D_M} g^p(\theta) = \{\theta \mid \theta \in D_M, g^p(\theta) = \min_{\tilde{\theta} \in D_M} g^p(\tilde{\theta})\}$. We put ourselves in the case where the global minimum is equal to the true parameter vector θ_0 ; in this case, $D_c = \{\theta^*\} = \{\theta_0\}$. To establish this asymptotic covariance matrix, we need the limit Hessian and the Q_{ω}^p -matrix. Using Theorem 10, the covariance matrix is given by

$$\text{cov} \hat{\theta}_N^{L_p} \sim \frac{\bar{g}''_{\omega}(\theta_0)^{-1} Q_{\omega}^p(\theta_0) \bar{g}''_{\omega}(\theta_0)^{-1}}{N} = \frac{\mathcal{P}(\theta_0)}{N} \quad (21)$$

with the limit Hessian and the Q_{ω}^p -matrix yielded, respectively, by

$$\begin{aligned} \bar{g}''_{\omega}(\theta_0) &= \frac{p_2(p_2 - 1)}{2} m_{\mathcal{H},2}^0 \bar{E} \Phi_{\nu_{p_2},t}^0 \\ &\quad + \tilde{\eta} p_1(p_1 - 1) m_{\mathcal{H},1}^0 \bar{E} \Phi_{\nu_{p_1},t}^0, \end{aligned} \quad (22)$$

$$Q_{\omega}^p(\theta_0) = \frac{p_2^2}{4} m_{\mathcal{H},3}^0 \bar{E} \Phi_{\nu_{p_2},t}^0 + \tilde{\eta}^2 p_1^2 m_{\mathcal{H},4}^0 \bar{E} \Phi_{\nu_{p_1},t}^0,$$

where $\bar{E}\Phi_{\nu_{p_1},t}^0 = \bar{E}\psi_{\nu_{p_1},t}^0 \psi_{\nu_{p_1},t}^{0T}$, $i = 1, 2$. See Appendix E for the proof.

Remark 12. It is easy to show that the second derivative of $g_N^p(\theta)$ is given by $g_N^{p''}(\theta) = (1/N) \sum_{t=1}^N (\partial^2 / \partial \theta \partial \theta^T) \rho_{\tilde{\eta}}^p(\varepsilon_t(\theta))$ with

$$\begin{aligned} \frac{\partial^2}{\partial \theta \partial \theta^T} \rho_{\tilde{\eta}}^p(\varepsilon_t(\theta)) &= \frac{p_2(p_2 - 1)}{2} m_{\mathcal{H},2} \psi_{\nu_{p_2},t}(\theta) \psi_{\nu_{p_2},t}^T(\theta) \\ &\quad + \tilde{\eta} p_1(p_1 - 1) m_{\mathcal{H},1} \psi_{\nu_{p_1},t}(\theta) \psi_{\nu_{p_1},t}^T(\theta). \end{aligned} \quad (23)$$

Since $p_1 < 1$, the second term on the right-hand side of (23) is negative. Accordingly, there exists a limit value of p_1 denoted by p_1^L such that (23) is positive and then verifies Proposition 4.2 and Corollary 4.3 in [4].

3.3. L_p Robust Final Prediction Error. Similar to the RFPE in Huber's framework [29], we now establish a quality measure for a given model. Let $\hat{m}_N^p = \hat{\mathbf{M}}(\hat{\theta}_N^p)$ be the robust model to be evaluated. It is estimated within the model structure $\tilde{\mathbf{M}}$. Let us denote $\bar{\mathcal{F}}_{\hat{\theta}_N^p}(\tilde{\mathbf{M}}; \bar{\omega})$, this quality measure that depends on the actual data record Z^N . We classically have

$$\bar{\mathcal{F}}_{\hat{\theta}_N^p}(\tilde{\mathbf{M}}; \bar{\omega}) = E \bar{g}_{\bar{\omega}}^p(\hat{\theta}_N^p), \quad (24)$$

where $\bar{g}_{\bar{\omega}}^p(\theta)$ is the limit criterion. Let θ_0 be the argument of the minimum of this limit criterion and let us expand $\bar{g}_{\bar{\omega}}^p(\theta)$ around θ_0 , and we obtain

$$\bar{g}_{\bar{\omega}}^p(\hat{\theta}_N^p) \approx \bar{g}_{\bar{\omega}}^p(\theta_0) + \frac{1}{2} (\hat{\theta}_N^p - \theta_0)^T \bar{g}_{\bar{\omega}}^{p''}(\tilde{\zeta}_N^p) (\hat{\theta}_N^p - \theta_0). \quad (25)$$

Similarly, since $g_N^{p'}(\hat{\theta}_N^p) = 0$,

$$g_N^p(\hat{\theta}_N^p) \approx g_N^{p'}(\theta_0) - \frac{1}{2} (\hat{\theta}_N^p - \theta_0)^T g_{N,\bar{\omega}}^{p''}(\tilde{\zeta}_N^p) (\hat{\theta}_N^p - \theta_0). \quad (26)$$

From the Hessian uniform convergence theorem and the expectations of (25) and (26), we obtain for the L_p -RFPE

$$\bar{\mathcal{F}}_{\hat{\theta}_N^p}(\tilde{\mathbf{M}}; \bar{\omega}) \approx E g_N^p(\hat{\theta}_N^p) + \frac{1}{N} \text{tr} \left(Q_{\bar{\omega}}^p(\theta_0) \bar{g}_{\bar{\omega}}^{p''}(\theta_0)^{-1} \right). \quad (27)$$

If a single run is taken into account, then we can classically replace the expectation of $g_N^p(\hat{\theta}_N^p)$ with only one observation available. Therefore,

$$\bar{\mathcal{F}}_{\hat{\theta}_N^p}(\tilde{\mathbf{M}}; \bar{\omega}) \approx g_N^p(\hat{\theta}_N^p) + \frac{1}{N} \text{tr} \left(Q_{\bar{\omega}}^p(\theta_0) \bar{g}_{\bar{\omega}}^{p''}(\theta_0)^{-1} \right). \quad (28)$$

Let us denote $\mathcal{S}_{\bar{\omega}}^p(\theta_0) = Q_{\bar{\omega}}^p(\theta_0) \bar{g}_{\bar{\omega}}^{p''}(\theta_0)^{-1}$ and $\mathcal{T}_{\nu_{p_1}, \nu_{p_2}}^p(\theta_0) = \bar{E}\Phi_{\nu_{p_1},t}^0 (\bar{E}\Phi_{\nu_{p_2},t}^0)^{-1}$. We obtain

$$\mathcal{S}_{\bar{\omega}}^p(\theta_0) = \mathcal{Z}^0 I - \mathcal{X}_{\bar{\omega}}^0 \mathcal{T}_{\nu_{p_1}, \nu_{p_2}}^p(\theta_0) - \mathcal{Y}_{\bar{\omega}}^0 \left(\mathcal{T}_{\nu_{p_1}, \nu_{p_2}}^p(\theta_0) \right)^2, \quad (29)$$

where I is the identity matrix. Now, in order to completely define the expression of the L_p -RFPE, we need the following result due to the Schur complement.

Lemma 13. Let M be a symmetric matrix of the form $M = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$, where A , B , and C are real and positive definite matrices. If C is invertible, then the following property holds:

$$(i) \quad M > 0 \quad \text{iff} \quad C > 0, \quad A - BC^{-1}B^T > 0 \implies C > A^{-1}. \quad (30)$$

In the mixed L_p context, the estimation procedure is mainly a L_{p_2} estimation since the impulsional effect of the outlier is rapidly treated by the L_{p_1} . Accordingly, $\text{card}\{\nu_{p_2}\} > \text{card}\{\nu_{p_1}\}$. Moreover, $\bar{E}\Phi_{\nu_{p_1},t}^0 = \bar{E}\psi_{\nu_{p_1},t}^0 \psi_{\nu_{p_1},t}^{0T}$, $i = 1, 2$, are positive definite matrices. Using Lemma 13 with $B = I^d$, $A = (\bar{E}\Phi_{\nu_{p_1},t}^0)^{-1}$, and $C = \bar{E}\Phi_{\nu_{p_2},t}^0$, we have

$$\left(\bar{E}\Phi_{\nu_{p_2},t}^0 \right) > \left(\bar{E}\Phi_{\nu_{p_1},t}^0 \right) \implies \mathcal{T}_{\nu_{p_1}, \nu_{p_2}}^p(\theta_0) = \bar{\alpha} I^d, \quad (31)$$

where d is the model complexity and $0 < \bar{\alpha} < 1$. Therefore, $(\mathcal{T}_{\nu_{p_1}, \nu_{p_2}}^p(\theta_0))^2 = \bar{\alpha}^2 I^d$ and the trace of $\mathcal{S}_{\bar{\omega}}^p(\theta_0)$ becomes

$$d_M^p(\mathcal{Z}^0 - \bar{\alpha} \mathcal{X}_{\bar{\omega}}^0 - \bar{\alpha}^2 \mathcal{Y}_{\bar{\omega}}^0) \quad (32)$$

with d_M^p being the model complexity. The L_p -RFPE becomes

$$\bar{\mathcal{F}}_{\hat{\theta}_N^p}(\tilde{\mathbf{M}}; \bar{\omega}) \approx g_N^p(\hat{\theta}_N^p) + \frac{d_M^p}{N} (\mathcal{Z}^0 - \bar{\alpha} \mathcal{X}_{\bar{\omega}}^0 - \bar{\alpha}^2 \mathcal{Y}_{\bar{\omega}}^0). \quad (33)$$

4. Model Order Reduction

Let two model classes $\mathbf{M}$ and $\tilde{\mathbf{M}}$ be given, respectively, related to Huber's and L_p estimators. To avoid some drawbacks with the notation of the parameter vectors, model classes $\mathbf{M}$ and $\tilde{\mathbf{M}}$ are parameterized by θ and $\tilde{\theta}$, respectively. Recall that the PREC in Huber's approach is defined as follows:

$$W_N(\theta) = \frac{1}{N} \sum_{t \in \gamma_2(\theta)} \frac{\varepsilon_t^2(\theta)}{2} + \frac{\eta}{N} \sum_{t \in \gamma_1(\theta)} \left(|\varepsilon_t(\theta)| - \frac{\eta s_t^2(\theta)}{2} \right). \quad (34)$$

See [29, chapter 4, page 50]. Let us consider the following conditions C1:

- (1) $\eta < 1$ and $\omega < 0.4$;
- (2) $\tilde{\eta} < 1$ and $\bar{\omega} < 0.4$;
- (3) $1.8 \leq p_2 < 2$ with $p_1 < 1$;
- (4) $\phi < 2$.

Theorem 14. Assume that the conditions C1 are verified. Let $|\Omega_i| = \sup_{(t, \theta, \tilde{\theta})} (|\varepsilon_t(\theta)|, |\varepsilon_t(\tilde{\theta})|)$ be the innovation outlier; then

$$\sup_{\tilde{\theta}} |g_N^p(\tilde{\theta})| < \sup_{\theta} |W_N(\theta)|. \quad (35)$$

TABLE 1: DOF in the symmetric function.

Estimator	Function	D_f	DOF
LSE	$\frac{X^2}{2}$	0	0
LSAD	$ X $	0	0
Huber	$\begin{cases} \frac{X^2}{2} & \text{for } X \leq \eta \\ \eta X - \frac{\eta^2}{2} & \text{for } X > \eta \end{cases}$	1	η
Mixed- L_p	$\begin{cases} \frac{ X ^{p_2}}{2} & \text{for } X \leq \tilde{\eta} \\ \tilde{\eta} X ^{p_1} - \frac{\tilde{\eta}^2}{2} & \text{for } X > \tilde{\eta} \end{cases}$	2	$\tilde{\eta}, p_2$

See the proof in Appendix F. Let $\hat{\theta}_N^H$ be Huber's estimator. Using $\bar{\mathcal{F}}_{\hat{\theta}_N^H}(\mathbf{M}; \omega)$ in [29, chapter 7, page 125] and Lemma 13, since $\alpha < 1$ and $\omega < 0.4$, then

$$\sup_{\hat{\theta}_N^H} \left| \bar{\mathcal{F}}_{\hat{\theta}_N^H}(\mathbf{M}; \omega) \right| = 1 + |\Omega_i| + \frac{2\hat{\lambda}(\omega) d_M^H}{N}. \quad (36)$$

From (33) and conditions C1, since $|\mathcal{X}_{\bar{\omega}}^0| < 1/2$ and $|\mathcal{Y}_{\bar{\omega}}^0| < 1/2$, we obtain

$$\sup_{\hat{\theta}_N^{L_p}} \left| \bar{\mathcal{F}}_{\hat{\theta}_N^{L_p}}(\bar{\mathbf{M}}; \bar{\omega}) \right| = 1 + |\Omega_i|^{p_1} + \frac{d_M^p}{N} (1 + |\mathcal{Y}^0|). \quad (37)$$

From the curves of $\mathcal{Y}^0$ and using the conditions C1, then

$$\mathcal{Y}^0 \approx \frac{2p_2^2}{5p_2(p_2 - 1)} (2\tilde{\eta} + 1). \quad (38)$$

Let us denote $\mathcal{R}_d = 2\hat{\lambda}(\omega)/(1 + |\mathcal{Y}^0|)$, where $\omega = g(\eta)$ and $(\tilde{\eta}, p_2)$ are the DOF, respectively, related to Huber's and L_p symmetric functions.

Theorem 15. If $\sup_{\hat{\theta}_N^p} |\bar{\mathcal{F}}_{\hat{\theta}_N^p}(\bar{\mathbf{M}}; \bar{\omega})| < \sup_{\hat{\theta}_N^H} |\bar{\mathcal{F}}_{\hat{\theta}_N^H}(\mathbf{M}; \omega)|$ and $\mathcal{R}_d < 1$, then

$$d_M^p < \frac{2\hat{\lambda}(\omega)}{1 + |\mathcal{Y}^0|} d_M^H \implies d_M^p < d_M^H. \quad (39)$$

The proof is given in Appendix G. This tuning offers to the user the necessary DOF to estimate and validate the reduced order models, as described in the following section.

5. Effect of DOF in the Symmetric Function to the Model Order Reduction

Here, we focus on a new vision on the link between the model reduction objective and its link to the estimation chosen, in particular the number of DOF of this symmetric function. Let D_f be the number of DOF. We present in Table 1 different symmetric function related to the MLE context and the corresponding D_f with their associated tuning parameters.

TABLE 2: Effect of the DOF in the model order reduction.

Dimensional	D_f	Condition	DOF
$d_M^{L_1} < d_M^{L_2}$	0	$\hat{\lambda}_N < 1$	0
$d_M^H < d_M^{L_1}$	1	$\frac{1}{2\hat{\lambda}(\omega)} < 1, \omega = g(\eta)$	η
$d_M^p < d_M^H$	3	$\mathcal{R}_d < 1$	$\eta, \tilde{\eta}, p_2$

First, the reason for which the LSE and LSAD estimators do not allow the model order reduction seems to be due to the number of DOF; here, $D_f = 0$. From the comparison between Akaike's FPE and the L_1 -FPE [24], we obtain $d_M^{L_1} < \hat{\lambda}_N d_M^{L_2}$. The condition $d_M^{L_1} < d_M^{L_2}$ is ensured only by the value of the variance $\hat{\lambda}_N$ of the L_2 estimation. Sometimes, this one is obtained with difficulty, especially in the presence of some outliers in the prediction errors. Since $D_f = 0$, the model complexity reduction remains difficult. From our point of view, the introduction of DOF in the estimation symmetric function facilitates the model complexity reduction. For example, the DOF in Huber's symmetric function may ensure both the robustness and the model order reduction. Therefore, comparing (36) and the L_1 -FPE, we have $d_M^H < (1/2\hat{\lambda}(\omega)) d_M^{L_1}$. Consequently, the model order reduction may be obtained by tuning η in Huber's function to have the condition $1/2\hat{\lambda}(\omega) < 1$. By extension, with three DOF, the condition $\mathcal{R}_d < 1$ in (G.2) may be obtained by tuning ω (through η), $\tilde{\eta}$, and p_2 . Thus, the *flexibility* of the L_p symmetric function ensures the model order reduction. Table 2 summarizes these main results related to the effect of the norm's DOF on the model order reduction.

6. Experimental Results

6.1. Presentation. This section presents different results related to the identification of a real system, namely, an acoustic duct plant, in order to estimate and validate reduced order models, using L_p estimators. We compare the L_p estimation results with those given by Huber's [28, 29], LSAD, LSE [24], and L_∞ estimators ($\hat{\theta}_N^{\infty} = \min_{\theta} \max_t |\varepsilon_t(\theta)|$). As shown in

TABLE 3: Main results from L_∞ , LSE, LSAD, M , and L_p estimators for model order reduction.

Estimator	Model	Fit (%)	k	η	ω	$\hat{\lambda}(\omega)$	$\tilde{\eta}$	$\bar{\omega}$	$\mathcal{R}_d$	d_M
L_∞	(31, 11)	12	—	—	—	—	—	—	—	42
LSE	(2, 0)-(14, 9)	48	—	—	—	—	—	—	—	23
LSAD	(0, 1)-(14, 7)	72	—	—	—	—	—	—	—	21
M	(2, 1)-(9, 8)	69.45	0.75	0.6159	0.22	0.2819	—	—	—	17
L_p	(1.9, 0.9924)-(8, 7)	67.95	—	—	—	—	0.71	0.265	0.185	15
	(1.8, 0.939)-(7, 7)	65.52	—	—	—	—	0.64	0.351	0.182	14

TABLE 4: Parameters from the Huber estimation (2, 1)-(9, 8) (b_n^H, f_n^H) and from the L_p estimation (1.8, 0.939)-(7, 7) (b_n^P, f_n^P).

n	1	2	3	4	5	6	7	8	9
b_n^H	-0.0856	-0.1642	0.1062	0.2241	0.1460	0.0440	-0.1929	-0.0707	—
f_n^H	0.4241	-0.5992	-0.4307	-0.0207	0.2362	0.3232	-0.1615	0.0462	0.0744
b_n^P	-0.0927	-0.1415	0.1829	0.1553	0.0447	0.0412	-0.1857	—	—
f_n^P	-0.0699	-0.7319	0.0578	0.0387	-0.0021	0.3249	-0.2332	—	—

Figure 3, we have to find discrete-time dynamical models with reduced orders of the plant including the secondary loudspeaker, its amplifier, the acoustic secondary path, the measurement microphone, and its amplifier/antialiasing filter. For more details, the reader could refer to [24]. The identification has been realized from a PRBS (pseudorandom binary sequence) as excitation input signal, with a length $\mathcal{L} = 2^{10} - 1$ and level $\pm 3V$, approximating a white noise upon the frequency range $[0; 1000 \text{ Hz}]$. The sampling period is $T_s = 500 \mu\text{s}$ and the number of data is $N = 1024$. The pure delay d of the duct is given as an integer of the sampling period. Here, $d = 7$. Figure 2 shows the output signal given by the microphone during the recording phase of the data set Z^N of the process. An estimations campaign has been conducted to get the value of p_1^L ensuring Proposition 4.2 and Corollary 4.3 in [4]. p_1 was varied over the range $[0.5; 1[$ with an incremental step of 0.01. The convergence is ensured for $p_1 \geq 0.88$. We deduce that $p_1^L = 0.88$. In the sequel, for the L_p estimation context, $1.8 \leq p_2 \leq 2$, $\tilde{\eta} < 1$, and $0.88 \leq p_1 < 1$, and we will denote all the models by $(\alpha, \beta) - (n_F, n_B)$.

6.2. Reduced Order Models. Table 3 presents the main results from L_∞ , LSE, LSAD, M , and L_p estimators. The model order reduction is clearly shown and we can notice that the complexity reduction between LSE and the low dimension of the new mixed L_p estimator is 40%. In Table 4, the parameters of the (2, 1)-(9, 8) and (1.8, 0.939)-(7, 7) models are given. Figures 4 and 5 show the frequency responses of the (2, 1)-(9, 8) and (1.8, 0.939)-(7, 7) models, respectively, versus the spectral estimation of the process. Even though the L_p model presents a lower complexity than Huber's model, its frequency response especially in low frequencies particularly interesting in controller design is good. Remark that, for L_∞ estimator, the fit is low and the model order increases considerably with $d_M^\infty = 42$.

FIGURE 2: Output signal given by the microphone during the recording phase of the data set Z^N . It clearly shows some outliers with distinguished levels.

7. Conclusion

In this paper, a new robust estimator based on a mixed L_p distribution has been proposed in order to reduce the estimated model complexity. The main consistency and convergence properties of the robust estimator have been established and it has been shown that the tuning of DOF in the symmetric function improves the model order reduction. In effect, the additional DOF in the symmetric function of the estimation criterion present the advantage to tune its parameters. This flexibility offers to the user the choice to estimate and reduce the estimated model complexity in the same procedure. We presented experimental results from a vibration real process and we showed the effective reductions. Many aspects of these

FIGURE 3: Experimental setup.

FIGURE 4: Frequency response of the (2, 1)-(9, 8) model (solid line) versus spectral estimation of the process (dashed line).

FIGURE 5: Frequency response of the (1.8, 0.939)-(7, 7) model (solid line) versus spectral estimation of the process (dashed line).

studies are opened to further researches. First, from a formal point of view, we should study the generalization of multiple DOF in a mixed symmetric function by investigating the link between the number of DOF and the model reduction. Then, it would be relevant to apply our L_p framework to a process with a level of contamination of the data more significant.

Appendices

A. Proof of Theorem 4

For all $\bar{\omega} > 0$, the pdf $(1 - \bar{\omega}) Ce^{-\rho_{\bar{\eta}}^p(X)/\phi^{p_2}}$ has a total mass 1. Hence,

$$\int_{-\infty}^{\infty} Ce^{-\rho_{\bar{\eta}}^p(X)/\phi^{p_2}} dX = \frac{1}{1 - \bar{\omega}}. \quad (\text{A.1})$$

Since $\rho_{\bar{\eta}}^p$ is a symmetric function, then (A.1) becomes

$$\mathcal{A} + \mathcal{B} = \frac{1}{1 - \bar{\omega}} \quad (\text{A.2})$$

with

$$\begin{aligned} \mathcal{A} &= 2C \int_0^{\bar{\eta}} e^{-X^{p_2}/2\phi^{p_2}} dX; \\ \mathcal{B} &= 2C \int_{\bar{\eta}}^{\infty} e^{-\bar{\eta}X^{p_1}/\phi^{p_2} + \bar{\eta}^2/2\phi^{p_2}} dX. \end{aligned} \quad (\text{A.3})$$

Case 1 (for $\mathcal{A}$). With $Z = X^{p_2}/2\phi^{p_2}$, we get $dZ = (p_2/2\phi^{p_2})X^{p_2-1}dX$. Moreover, $X^{p_2-1} = \phi^{p_2-1}(2Z)^{(1-1/p_2)}$ and $dX = (\phi^{1/p_2}/p_2)Z^{1/p_2-1}dZ$. We obtain

$$\begin{aligned}\mathcal{A} &= \frac{2\phi^{1/p_2} \int_0^{\tilde{\eta}^{p_2}/2\phi^{p_2}} Z^{1/p_2-1} e^{-Z} dZ}{p_2} \\ &= \frac{\phi^{1/p_2}}{p_2(K_1 + K_2)} \int_0^{\tilde{\eta}^{p_2}/2\phi^{p_2}} Z^{1/p_2-1} e^{-Z} dZ.\end{aligned}\quad (\text{A.4})$$

The integral in (A.4) is equal to $\Gamma(1/p_2) - \Gamma(1/p_2, \tilde{\eta}^{p_2}/2\phi^{p_2})$, which proves $\mathcal{A}$.

Case 2 (for $\mathcal{B}$). From straightforward calculations, expression $\mathcal{B}$ only necessitates the result of the integral $I = \int_{\tilde{\eta}}^{\infty} e^{-\tilde{\eta}X^{p_1}/\phi^{p_2}} dX$. With $Z = \tilde{\eta}X^{p_1}/\phi^{p_2}$, we obtain $dZ = (\tilde{\eta}p_1/\phi^{p_2})X^{p_1-1}dX$. Moreover, $X^{p_1-1} = \phi^{p_2/p_1-p_2}Z^{1/p_1-1}/\tilde{\eta}^{1/p_1-1}$ and we deduce $dX = \phi^{p_2/p_1}Z^{1/p_1-1}/p_1\tilde{\eta}^{1/p_1}dZ$. Therefore,

$$\begin{aligned}I &= \frac{\phi^{p_2/p_1}}{p_1\tilde{\eta}^{1/p_1}} \int_{\tilde{\eta}^{p_1+1}/\phi^{p_2}}^{\infty} Z^{1/p_1-1} e^{-Z} dZ \\ &= \frac{\phi^{p_2/p_1}}{p_1\tilde{\eta}^{1/p_1}} \Gamma\left(\frac{1}{p_1}, \frac{\tilde{\eta}^{p_1+1}}{\phi^{p_2}}\right),\end{aligned}\quad (\text{A.5})$$

$$\mathcal{B} = \frac{e^{\tilde{\eta}^2/2\phi^{p_2}} \phi^{p_2/p_1}}{(K_1 + K_2) p_1 \tilde{\eta}^{1/p_1}} \Gamma\left(\frac{1}{p_1}, \frac{\tilde{\eta}^{p_1+1}}{\phi^{p_2}}\right),$$

which proves $\mathcal{B}$.

B. Proof of Lemma 6

Since f_p is a symmetric function, it is easy to show that

$$\begin{aligned}E(|U|^r) &= 2C \int_0^{\tilde{\eta}} u^r e^{-u^{p_2}/2\phi^{p_2}} du \\ &\quad + 2C \int_{\tilde{\eta}}^{\infty} u^r e^{-\tilde{\eta}u^{p_1}/\phi^{p_2} + \tilde{\eta}^2/2\phi^{p_2}} du \\ &= \mathcal{J}_1^r + \mathcal{J}_2^r.\end{aligned}\quad (\text{B.1})$$

For $\mathcal{J}_1^r$, let us denote $\mathcal{J}_1^r = \int_0^{\tilde{\eta}} u^r e^{-u^{p_2}/2\phi^{p_2}} du$. With $Z = u^{p_2}/2\phi^{p_2}$, we get $dZ = (p_2/2\phi^{p_2})u^{p_2-1}du$. Moreover, $u^{p_2-1} = \phi^{p_2-1}(2Z)^{(1-1/p_2)}$ and $du = (\phi^{1/p_2}/p_2)Z^{1/p_2-1}dZ$. We obtain

$$\begin{aligned}\mathcal{J}_1^r &= \int_0^{\tilde{\eta}^{p_2}/2\phi^{p_2}} \phi^r 2^{r/p_2} Z^{r/p_2} \frac{\phi}{p_2} Z^{1/p_2-1} e^{-Z} dZ \\ &= \frac{\phi^{r+1} 2^{(r+1)/p_2}}{p_2} \int_0^{\tilde{\eta}^{p_2}/2\phi^{p_2}} Z^{(r+1)/p_2-1} e^{-Z} dZ,\end{aligned}\quad (\text{B.2})$$

where the last integral is equal to $\Gamma((r+1)/p_2) - \Gamma((r+1)/p_2, \tilde{\eta}^{p_2}/2\phi^{p_2})$, which proves $\mathcal{J}_1^r$.

For $\mathcal{J}_2^r$, let us denote $\mathcal{J}_2^r = \int_{\tilde{\eta}}^{\infty} u^r e^{-\tilde{\eta}u^{p_1}/\phi^{p_2}} du$. With $Z = \tilde{\eta}u^{p_1}/\phi^{p_2}$, we obtain $dZ = (\tilde{\eta}p_1/\phi^{p_2})u^{p_1-1}du$. Moreover, $u^{p_1-1} = \phi^{p_2/p_1-p_2}Z^{1/p_1-1}/\tilde{\eta}^{1/p_1-1}$ and we deduce

$du = (\phi^{p_2/p_1} Z^{1/p_1-1}/p_1\tilde{\eta}^{1/p_1})dZ$. After straightforward calculations,

$$\begin{aligned}\mathcal{J}_2^r &= \frac{\phi^{(p_2/p_1)(r+1)}}{p_1\tilde{\eta}^{(r+1)/p_1}} \int_{\tilde{\eta}^{p_1+1}/\phi^{p_2}}^{\infty} Z^{(r+1)/p_1-1} e^{-Z} dZ \\ &= \frac{\phi^{(p_2/p_1)(r+1)}}{p_1\tilde{\eta}^{(r+1)/p_1}} \Gamma\left(\frac{r+1}{p_1}, \frac{\tilde{\eta}^{p_1+1}}{\phi^{p_2}}\right),\end{aligned}\quad (\text{B.3})$$

which proves $\mathcal{J}_2^r$ and finally the lemma.

C. Proof of Theorem 7

To prove Theorem 7, we need to describe system $\mathcal{S}$ with output y_t by

$$y_t = \varphi_t^T(\theta) \theta + e_t, \quad (\text{C.1})$$

where $\varphi_t^T(\theta)$ is the regressor of the model and e_t is the disturbance signal with zero mean and variance λ . Let $\varepsilon_t(\theta)$ be the estimation error equal to $y_t - \varphi_t^T(\theta)\theta$. At θ^* , we have $\varepsilon_t(\theta^*) = e_t^*$.

(i) We can say $\theta = \theta^* \Rightarrow f_p(\varepsilon_t(\theta)) = f_p(e_t^*)$; that is,

$$f_p(y_t - \varphi_t^T(\theta)\theta) = f_p(y_t - \varphi_t^T(\theta^*)\theta^*). \quad (\text{C.2})$$

We obtain

$$f_p(\varphi_t^T(\theta^*)\theta^* - \varphi_t^T(\theta)\theta + e_t^*) = f_p(e_t^*). \quad (\text{C.3})$$

Let us consider $V_t = \varphi_t^T(\theta^*)\theta^* - \varphi_t^T(\theta)\theta + e_t^*$. We then get

$$f_{p,V}(v) = f_p(V_t - \varphi_t^T(\theta^*)\theta^* + \varphi_t^T(\theta)\theta). \quad (\text{C.4})$$

Using (C.3), we get that $f_{p,V}(v) = f_p(v)$, which combined with (C.4) yields $\varphi_t^T(\theta^*)\theta^* = \varphi_t^T(\theta)\theta \Rightarrow \theta^* = \theta$.

(ii) The compactness condition (ii) is ensured by considering that $\theta^* \in D_M \subset \Theta$, where D_M is a compact subset of $\mathbb{R}^d$.

(iii) The continuity condition is trivially verified since $\ln f_p(\varepsilon_t(\theta))$ is continuous at each $\theta \in D_M$ w.p.1.

(iv) The boundedness condition requires that, for all $\theta \in D_M$, $E[\sup_{\theta \in \Theta} \ln f_p(\varepsilon_t(\theta))] < K < \infty$. From (1), we then have

$$\ln f_p(\varepsilon_t(\theta)) = \ln C - \frac{1}{\phi^{p_2}} \rho_{\tilde{\eta}}^p(\varepsilon_t(\theta)). \quad (\text{C.5})$$

With $\rho_{\tilde{\eta}}^p$ defined in (10), we obtain for $\theta \in D_M$

$$\begin{aligned}E|\ln f_p(\varepsilon_t(\theta))| &\leq |\ln C| + \frac{1}{2\phi^{p_2}} E[|\varepsilon_t(\theta)|^{p_2}] \\ &\quad + \frac{1}{2\phi^{p_2}} \tilde{\eta}^2 + \frac{\tilde{\eta}}{\phi^{p_2}} E[|\varepsilon_t(\theta)|^{p_1}].\end{aligned}\quad (\text{C.6})$$

From Lemma 6, we deduce $E[|\varepsilon_t(\theta)|^{p_2}]$ and $E[|\varepsilon_t(\theta)|^{p_1}]$. Then, for

$$\begin{aligned}K &= \sup_{\theta \in D_M} \left[|\ln C| + \frac{1}{2\phi^{p_2}} \tilde{\eta}^2 + \frac{1}{2\phi^{p_2}} (\mathcal{J}_1^{p_2} + \mathcal{J}_2^{p_2}) \right. \\ &\quad \left. + \frac{\tilde{\eta}}{\phi^{p_2}} (\mathcal{J}_1^{p_1} + \mathcal{J}_2^{p_1}) \right]\end{aligned}\quad (\text{C.7})$$

and by compactness of D_M , we have $K < \infty$, so that, for each $\theta \in D_M$, we get $E[\sup_{\theta \in D_M} |\ln f_p(\varepsilon_t(\theta))|] \leq K < \infty$, which proves $\hat{\theta}_N^p \rightarrow \theta^*$ w.p.1.

D. Proof of Theorem 11

(i) From $E \ln f_p(\varepsilon_t(\theta))$, we can deduce that

$$\theta^* = \arg \max_{\theta \in D_M} \left(\frac{-1}{N} \sum_{t=1}^N \rho_{\tilde{\eta}}^p(\varepsilon_t(\theta)) \right), \quad \text{as } N \rightarrow \infty, \quad (\text{D.1})$$

which is equivalent to

$$\arg \min_{\theta \in D_M} \left(\frac{1}{N} \sum_{t=1}^N \rho_{\tilde{\eta}}^p(\varepsilon_t(\theta)) \right), \quad \text{as } N \rightarrow \infty. \quad (\text{D.2})$$

Since $g^p(\theta) = E(\rho_{\tilde{\eta}}^p(\varepsilon_t(\theta)))$, then $g^p(\theta)$ is maximized on D_M at θ^* .

(ii) The interior condition of Theorem 11 is equivalent to the assumption $\theta \in D_M^\circ$, where D_M° is the interior of D_M .

(iii) From (11), the gradient and the Hessian are, respectively,

$$g_N^{p'}(\theta) = \frac{-p_2}{2N} \sum_{t \in \nu_{p_2}(\theta)} \psi_t(\theta) |\varepsilon_t(\theta)|^{p_2-1} s_t(\theta) - \frac{\tilde{\eta} p_1}{N} \sum_{t \in \nu_{p_1}(\theta)} \psi_t(\theta) |\varepsilon_t(\theta)|^{p_1-1} s_t(\theta), \quad (\text{D.3})$$

$$\begin{aligned} g_N^{p''}(\theta) &= \frac{p_2(p_2-1)}{2N} \sum_{t \in \nu_{p_2}(\theta)} \psi_t(\theta) \psi_t^T(\theta) |\varepsilon_t(\theta)|^{p_2-2} \\ &\quad - \frac{p_2}{2N} \sum_{t \in \nu_{p_2}(\theta)} |\varepsilon_t(\theta)|^{p_2-1} s_t(\theta) \frac{\partial}{\partial \theta} \psi_t^T(\theta) \\ &\quad + \frac{\tilde{\eta} p_1(p_1-1)}{N} \sum_{t \in \nu_{p_1}(\theta)} \psi_t(\theta) \psi_t^T(\theta) |\varepsilon_t(\theta)|^{p_1-2} \\ &\quad - \frac{\tilde{\eta} p_1}{N} \sum_{t \in \nu_{p_1}(\theta)} |\varepsilon_t(\theta)|^{p_1-1} s_t(\theta) \frac{\partial}{\partial \theta} \psi_t^T(\theta). \end{aligned} \quad (\text{D.4})$$

From (D.3), the Ψ -function is

$$\begin{aligned} \Psi_{t,\tilde{\eta}}^p(\varepsilon; \theta) &= \frac{-p_2}{2} \psi_{\nu_{p_2},t}(\theta) |\varepsilon_{\nu_{p_2},t}(\theta)|^{p_2-1} s_{\nu_{p_2},t}(\theta) \\ &\quad - \tilde{\eta} p_1 \psi_{\nu_{p_1},t}(\theta) |\varepsilon_{\nu_{p_1},t}(\theta)|^{p_1-1} s_{\nu_{p_1},t}(\theta). \end{aligned} \quad (\text{D.5})$$

In the sequel, we adopt Ljung's notations: $\bar{E}f_t = \lim_{N \rightarrow \infty} (1/N) \sum_{t=1}^N E f_t$ (see [14]), where E is the expectation with respect to the considered probability distribution. Moreover, write for short in the general case $h_t(\theta^*) = h_t^*$. In order to render differentiable the limit Hessian, we use the stochastic differentiable condition; therefore, at the global

parameter vector θ^* , since $Es_{\nu_{p_i},t}^* \xrightarrow{\text{AS}} 0$ ($i = 1, 2$), the limit Hessian is

$$\begin{aligned} \overline{g_{\omega}^{p''}}(\theta^*) &= \frac{p_2(p_2-1)}{2} m_{\mathcal{H},2}^* \bar{E}\Phi_{\nu_{p_2},t}^* \\ &\quad + \tilde{\eta} p_1(p_1-1) m_{\mathcal{H},1}^* \bar{E}\Phi_{\nu_{p_1},t}^*, \end{aligned} \quad (\text{D.6})$$

with $\bar{E}\Phi_{\nu_{p_i},t}^* = \bar{E}\psi_{\nu_{p_i},t}^* \Psi_{\nu_{p_i},t}^{*T}$, $i = 1, 2$. $m_{\mathcal{H},2}^*$ and $m_{\mathcal{H},1}^*$ are the moments, respectively, given by $E|\varepsilon_{\nu_{p_2},t}^*|^{p_2-2}$ and $E|\varepsilon_{\nu_{p_1},t}^*|^{p_1-2}$ established by Lemma 6. Expression (D.6) satisfies the stochastic differentiable condition and, for a fixed scaling factor, the matrices $\bar{E}\Phi_{\nu_{p_i},t}^*$ are nonsingular. Therefore, $\overline{g_{\omega}^{p''}}(\theta^*)$ is invertible.

(iv) Using the mean value theorem, we get

$$\left(E \frac{\partial^2 g_N^p(\xi)}{\partial \xi \partial \xi^T} \right)_{\tilde{\theta}_N} (\hat{\theta}_N^p - \theta^*) = \left(E \frac{\partial g_N^p(\theta)}{\partial \theta} \right)_{\hat{\theta}_N^p} - \left(E \frac{\partial g_N^p(\theta)}{\partial \theta} \right)_{\theta^*} \quad (\text{D.7})$$

with $\hat{\theta}_N^p \leq \tilde{\theta}_N \leq \theta^*$. For $N \rightarrow \infty$, $\tilde{\theta}_N \rightarrow \theta^*$, $(E(\partial g_N^p(\theta)/\partial \theta))_{\theta^*} = \bar{E}\Psi_{t,\tilde{\eta}}^p(\varepsilon; \theta^*) = 0$ and $(E(\partial^2 g_N^p(\xi)/\partial \xi \partial \xi^T))_{\tilde{\theta}_N} \rightarrow \overline{g_{\omega}^{p''}}(\theta^*)$. One has

$$\sqrt{N}(\hat{\theta}_N^p - \theta^*) = (\overline{g_{\omega}^{p''}}(\theta^*))^{-1} \sqrt{N} \left(E \frac{\partial g_N^p(\theta)}{\partial \theta} \right)_{\hat{\theta}_N^p}. \quad (\text{D.8})$$

The asymptotic distribution of $\sqrt{N}(\hat{\theta}_N^p - \theta^*)$ only depends on the asymptotic distribution of $\sqrt{N}(E(\partial g_N^p(\theta)/\partial \theta))_{\hat{\theta}_N^p}$. Therefore,

$$\begin{aligned} & - \sqrt{N} \left(E \frac{\partial g_N^p(\theta)}{\partial \theta} \right)_{\hat{\theta}_N^p} \\ &= \sqrt{N} \left\{ \frac{1}{N} \sum_{t=1}^N [\Psi_{t,\tilde{\eta}}^p(\varepsilon; \hat{\theta}_N^p) - E\Psi_{t,\tilde{\eta}}^p(\varepsilon; \hat{\theta}_N^p)] \right\} \\ &\quad - \sqrt{N} \frac{1}{N} \sum_{t=1}^N \Psi_{t,\tilde{\eta}}^p(\varepsilon; \hat{\theta}_N^p) \\ &= \frac{1}{\sqrt{N}} \sum_{t=1}^N (\Psi_{t,\tilde{\eta}}^p(\varepsilon; \hat{\theta}_N^p) - E\Psi_{t,\tilde{\eta}}^p(\varepsilon; \hat{\theta}_N^p)) \\ &\quad - \frac{1}{\sqrt{N}} \sum_{t=1}^N \Psi_{t,\tilde{\eta}}^p(\varepsilon; \hat{\theta}_N^p). \end{aligned} \quad (\text{D.9})$$

Let us denote $S_N(\theta) = (1/\sqrt{N}) \sum_{t=1}^N (\Psi_{t,\tilde{\eta}}^p(\varepsilon; \theta) - E\Psi_{t,\tilde{\eta}}^p(\varepsilon; \theta))$; then

$$\begin{aligned} & - \sqrt{N} \left(E \frac{\partial g_N^p(\theta)}{\partial \theta} \right)_{\hat{\theta}_N^p} = (S_N(\hat{\theta}_N^p) - S_N(\theta^*)) + S_N(\theta^*) \\ &\quad - \frac{1}{\sqrt{N}} \sum_{t=1}^N \Psi_{t,\tilde{\eta}}^p(\varepsilon; \hat{\theta}_N^p). \end{aligned} \quad (\text{D.10})$$

Since $(1/N) \sum_{t=1}^N \Psi_{t,\eta}^p(\varepsilon; \hat{\theta}_N^p) = 0$, the third term on the right-hand side of (D.10) is $o_p(1)$. Its first term is $o_p(1)$ provided $\{S_N(\bullet), N \geq 1\}$ is stochastically equicontinuous and $\hat{\theta}_N^p \xrightarrow{\text{prob}} \theta^*$ (see Theorem 7). This follows because, given any $\alpha > 0$ and $\beta > 0$, there exists a $\delta > 0$ such that

$$\begin{aligned} & \overline{\lim}_{N \rightarrow \infty} P\left(|S_N(\hat{\theta}_N^p) - S_N(\theta^*)| > \alpha\right) \\ & \leq \overline{\lim}_{N \rightarrow \infty} P\left(|S_N(\hat{\theta}_N^p) - S_N(\theta^*)|, \right. \\ & \quad \left. \left\|\rho_{\eta}^p(\varepsilon_t(\hat{\theta}_N^p)) - \rho_{\eta}^p(\varepsilon_t(\theta^*))\right\| \leq \delta\right) \\ & \quad + \overline{\lim}_{N \rightarrow \infty} P\left(\left\|\rho_{\eta}^p(\varepsilon_t(\hat{\theta}_N^p)) - \rho_{\eta}^p(\varepsilon_t(\theta^*))\right\| > \delta\right) \\ & \leq \overline{\lim}_{N \rightarrow \infty} P\left(\sup_{\theta \in D_M} |S_N(\theta) - S_N(\theta^*)| > \alpha\right) < \beta, \end{aligned} \quad (\text{D.11})$$

where the second inequality uses $\hat{\theta}_N^p \xrightarrow{\text{prob}} \theta^*$ and the third uses the stochastic equicontinuity. Accordingly, this shows that, for N tends to infinity, we have in law $\mathcal{L}$

$$\mathcal{L}\left(\sqrt{N}(\hat{\theta}_N^p - \theta^*)\right) \underset{N \rightarrow \infty}{\sim} \mathcal{L}(S_N(\theta^*)) \quad (\text{D.12})$$

with

$$S_N(\theta^*) = \frac{1}{\sqrt{N}} \sum_{t=1}^N \left(\frac{d}{d\theta} \rho_{\eta}^p(\varepsilon_t(\theta)) - E \frac{d}{d\theta} \rho_{\eta}^p(\varepsilon_t(\theta)) \right)_{\theta^*}. \quad (\text{D.13})$$

The main purpose is to prove that $S_N(\theta^*)$ in law is a normal asymptotic distribution. For this, we must show that the terms of $S_N(\theta^*)$ are independent. However, if we show that the dependence between distant terms decreases, that is, if we show that they are m -dependent (i.e., independent beyond lags of length m), then we show the asymptotic normal behavior of $S_N(\theta^*)$. For our proof, we will use the following technical assumptions.

(H1). The dataset Z^∞ is such that, for some filters $\{\alpha_{t,k}(\theta), \beta_{t,k}(\theta), \gamma_{t,k}(\theta), \delta_{t,k}(\theta)\} = f_{t,k}(\theta)$, the estimated error is

$$\begin{aligned} \varepsilon_t(\theta) &= \underbrace{\sum_{k \geq 1} \alpha_{t,k}(\theta) u_{t-k} + \sum_{k \geq 0} \beta_{t,k}(\theta) e_{t-k}}_{t \in \mathcal{V}_{p_2}} \\ &\quad + \underbrace{\sum_j \Omega_j(\theta) \delta_{t,j}}_{t \in \mathcal{V}_{p_1}}. \end{aligned} \quad (\text{D.14})$$

- (1) $\{u_t\}$ is a bounded, deterministic, exogenous, and input sequence with $\sup_t |u_t| = C_u$.
- (2) $\{e_t\}$ is a sequence of independent rv's with zero mean values and bounded moments of order $4 + \delta$, for $\delta > 0$.

- (3) The family of filters $f_{t,k}(\theta)$, $t = 1, 2, \dots$, is uniformly stable for all t, θ , with $f_{t,k}(\theta) < \mu_k$ and $\sum_{k \geq 0} \mu_k < \infty$.
- (4) $\{u_t\}$ and $\{y_t\}$ are jointly quasistationary.
- (5) Innovation outliers $\Omega_j(\theta)$ and $\phi_j(\theta)$ are bounded for all θ and j , $\sup_{j,\theta} |\Omega_j(\theta)| = \widehat{\Omega}$ and $\sup_{j,\theta} |\phi_j(\theta)| = \widehat{\phi}$.

Let us consider the following short expressions: $\varepsilon_{i,t}^* = \varepsilon_{i,t}^*(\theta^*) = f_{t,k}^*$. We then have

$$\begin{aligned} \varepsilon_{2,t}^* &= \varepsilon_{2,t}^{*,m} + \tilde{\varepsilon}_{2,t}^{*,m} \\ &= \sum_{k \geq 1} \alpha_{t,k}^* u_{t-k} + \sum_{k=0}^m \beta_{t,k}^* e_{t-k} + \sum_{k=m+1}^{\infty} \beta_{t,k}^* e_{t-k}, \end{aligned} \quad (\text{D.15})$$

where m is an integer. Moreover,

$$\varepsilon_{1,t}^* = \sum_j \Omega_j^* \delta_{t,j}^K, \quad (\text{D.16})$$

with $\Omega_j^* = \Omega_j(\theta^*)$ and $\delta_{t,j}^K$ is Kronecker's function. Analogously, since $\psi_t(\theta) = -(\partial/\partial\theta)\varepsilon_t(\theta)$, we have

$$\begin{aligned} \psi_{2,t}^* &= \psi_{2,t}^{*,m} + \tilde{\psi}_{2,t}^{*,m} \\ &= \sum_{k \geq 1} \gamma_{t,k}^* u_{t-k} + \sum_{k=0}^m \delta_{t,k}^* e_{t-k} + \sum_{k=m+1}^{\infty} \delta_{t,k}^* e_{t-k}, \\ \psi_{1,t}^* &= \sum_j \phi_j^* \delta_{t,j}^K, \end{aligned} \quad (\text{D.17})$$

with $\phi_j^* = \phi_j(\theta^*)$. Let us consider the following lemmas.

Lemma D.1. Consider the sum of doubly indexed rv's $\{x_{t,N}\}$ such that $S_N = \sum_{t=1}^N x_{t,N}$, where $Ex_{t,N} = 0$ and where $\{x_{1,N}, \dots, x_{s,N}\}$ and $\{x_{t,N}, x_{t+1,N}, \dots, x_{n,N}\}$ are independent of $t - s > m$ as follows:

$$\lim_{N \rightarrow \infty} \sup_{t=1}^N Ex_{t,N}^2 < \infty,$$

$$\lim_{N \rightarrow \infty} \sum_{t=1}^N E|x_{t,N}|^{2+\delta} = 0, \quad \delta > 0, \quad \text{Lyapunov's condition.} \quad (\text{D.18})$$

Then S_N is asymptotically normally distributed with zero mean and covariance matrix $Q = \lim_{N \rightarrow \infty} ES_N S_N^T$. See [32, 33].

Lemma D.2. Let $S_N = Z_{m,N} + X_{m,N}$, $m, N = 1, 2, \dots$, such that

- (i) $EX_{m,N}^2 \leq C_m$, $\lim_{m \rightarrow \infty} C_m = 0$;
- (ii) $P(Z_{m,N} \leq z) = F_{m,N}(z)$.

Then $\lim_{m \rightarrow \infty} \lim_{N \rightarrow \infty} P(Z_{m,N} \leq z) = F(z)$. See [34, 35].

$S_N(\theta^*)$ can be written as $S_N(\theta^*) = Z_{m,N}(\theta^*) + X_{m,N}(\theta^*)$, with

$$Z_{m,N}(\theta^*) = \frac{1}{\sqrt{N}} \sum_{t=1}^N \left(\frac{d}{d\theta} \rho_{\tilde{\eta}}^p(\varepsilon_t^m(\theta)) - E \frac{d}{d\theta} \rho_{\tilde{\eta}}^p(\varepsilon_t^m(\theta)) \right)_{\theta^*}, \quad (\text{D.19})$$

$$X_{m,N}(\theta^*) = \frac{1}{\sqrt{N}} \sum_{t=1}^N \frac{d}{d\theta} \left[\rho_{\tilde{\eta}}^p(\varepsilon_t(\theta)) - \rho_{\tilde{\eta}}^p(\varepsilon_t^m(\theta)) \right] - E \frac{d}{d\theta} \left[\rho_{\tilde{\eta}}^p(\varepsilon_t(\theta)) - \rho_{\tilde{\eta}}^p(\varepsilon_t^m(\theta)) \right]_{\theta^*}. \quad (\text{D.20})$$

Part 1. From (D.19) and using Lyapunov's condition, we obtain

$$\begin{aligned} & E \left| \frac{d}{d\theta} \rho_{\tilde{\eta}}^p(\varepsilon_t^m(\theta)) - E \frac{d}{d\theta} \rho_{\tilde{\eta}}^p(\varepsilon_t^m(\theta)) \right|^{\delta+2} \\ & \leq 2^{\delta+1} E \left(\left| \frac{d}{d\theta} \rho_{\tilde{\eta}}^p(\varepsilon_t^m(\theta)) \right|^{\delta+2} + E \left| \frac{d}{d\theta} \rho_{\tilde{\eta}}^p(\varepsilon_t^m(\theta)) \right|^{\delta+2} \right) \\ & \leq 2^{\delta+2} E \left| \frac{d}{d\theta} \rho_{\tilde{\eta}}^p(\varepsilon_t^m(\theta)) \right|^{\delta+2}. \end{aligned} \quad (\text{D.21})$$

On the other hand,

$$\begin{aligned} & \left| \frac{d}{d\theta} \rho_{\tilde{\eta}}^p(\varepsilon_t^m(\theta)) \right|^{\delta+2} \\ & \leq \left(\frac{p_2}{2} |\psi_{2,t}^{*,m}| |\varepsilon_{2,t}^{*,m}|^{p_2-1} + \tilde{\eta} p_1 |\psi_{1,t}^{*,m}| |\varepsilon_{1,t}^{*,m}|^{p_1-1} \right)^{\delta+2} \\ & \leq 2^{\delta+1} \left(\left(\frac{p_2}{2} \right)^{\delta+2} |\psi_{2,t}^{*,m}|^{\delta+2} |\varepsilon_{2,t}^{*,m}|^{(p_2-1)(\delta+2)} \right. \\ & \quad \left. + (\tilde{\eta} p_1)^{\delta+2} |\psi_{1,t}^{*,m}|^{\delta+2} |\varepsilon_{1,t}^{*,m}|^{(p_1-1)(\delta+2)} \right). \end{aligned} \quad (\text{D.22})$$

Lyapunov's condition becomes

$$\begin{aligned} & 2^{\delta+2} E \left| \frac{d}{d\theta} \rho_{\tilde{\eta}}^p(\varepsilon_t^m(\theta)) \right|^{\delta+2} \\ & \leq 2^{\delta+1} p_2^{\delta+2} E |\psi_{2,t}^{*,m}|^{\delta+2} |\varepsilon_{2,t}^{*,m}|^{(p_2-1)(\delta+2)} \\ & \quad + 2^{2\delta+3} (\tilde{\eta} p_1)^{\delta+2} E |\psi_{1,t}^{*,m}|^{\delta+2} |\varepsilon_{1,t}^{*,m}|^{(p_1-1)(\delta+2)}. \end{aligned} \quad (\text{D.23})$$

Using Schwarz's inequality, we then have

$$\begin{aligned} & 2^{\delta+2} E \left| \frac{d}{d\theta} \rho_{\tilde{\eta}}^p(\varepsilon_t^m(\theta)) \right|^{\delta+2} \\ & \leq 2^{\delta+1} p_2^{\delta+2} \left(E |\psi_{2,t}^{*,m}|^{2\delta+4} E |\varepsilon_{2,t}^{*,m}|^{(p_2-1)(2\delta+4)} \right)^{1/2} \\ & \quad + 2^{2\delta+3} (\tilde{\eta} p_1)^{\delta+2} \left(E |\psi_{1,t}^{*,m}|^{2\delta+4} E |\varepsilon_{1,t}^{*,m}|^{(p_1-1)(2\delta+4)} \right)^{1/2}. \end{aligned} \quad (\text{D.24})$$

The first and second terms on the right-hand side of (D.24) are, respectively, denoted by A^* and B^* .

(i) For A^* , in ν_{p_2} , for all t and θ^* , $|\varepsilon_{2,t}^{*,m}| \leq \tilde{\eta}$. Therefore,

$$\begin{aligned} E |\psi_{2,t}^{*,m}|^{2\delta+4} & \leq 2^{2\delta+3} \left[C_u^{2\delta+4} \left(\sum_{k \geq 1} \mu_k \right)^{2\delta+4} \right. \\ & \quad \left. + E |e_{t-k}|^{2\delta+4} \left(\sum_{k=0}^m \mu_k \right)^{2\delta+4} \right]. \end{aligned} \quad (\text{D.25})$$

From (H1), we have $E |\psi_{2,t}^{*,m}|^{2\delta+4} \leq C^*$ and after straightforward calculations $A^* \leq C$.

(ii) For B^* , from (H1) we obtain $\sup_{t,\theta} |\varepsilon_{1,t}^{*,m}| = \widehat{\Omega}$. We deduce that $E |\varepsilon_{1,t}^{*,m}|^{(p_1-1)(2\delta+4)}$ and $E |\psi_{1,t}^{*,m}|^{2\delta+4}$ are bounded. Accordingly, $B^* \leq C$.

By inserting $(1/\sqrt{N})^{\delta+2}$, we finally obtain

$$E \left| \frac{1}{\sqrt{N}} \left[\frac{d}{d\theta} \rho_{\tilde{\eta}}^p(\varepsilon_t^m(\theta)) - E \frac{d}{d\theta} \rho_{\tilde{\eta}}^p(\varepsilon_t^m(\theta)) \right] \right|^{\delta+2}_{\theta^*} \leq \frac{C}{N^{1+\delta/2}}. \quad (\text{D.26})$$

Then,

$$\begin{aligned} & \lim_{N \rightarrow \infty} \sum_{t=1}^N E \left| \frac{1}{\sqrt{N}} \left[\frac{d}{d\theta} \rho_{\tilde{\eta}}^p(\varepsilon_t^m(\theta)) - E \frac{d}{d\theta} \rho_{\tilde{\eta}}^p(\varepsilon_t^m(\theta)) \right] \right|^{\delta+2}_{\theta^*} \\ & \leq \lim_{N \rightarrow \infty} \frac{C}{N^{\delta/2}} \rightarrow 0. \end{aligned} \quad (\text{D.27})$$

Expression (D.27) proves (D.18) in Lemma D.1 with $Q_m(\theta^*) = \lim_{N \rightarrow \infty} E Z_{m,N}(\theta^*) Z_{m,N}^T(\theta^*)$.

Part 2. In $X_{m,N}(\theta^*)$, we can write

$$\begin{aligned} & \frac{d\rho_{\tilde{\eta}}^p(\varepsilon_t(\theta))}{d\theta} - \frac{d\rho_{\tilde{\eta}}^p(\varepsilon_t^m(\theta))}{d\theta} \\ & = \frac{\partial \rho_{\tilde{\eta}}^p(\varepsilon_t(\theta))}{\partial \varepsilon_t(\theta)} \frac{\partial \varepsilon_t(\theta)}{\partial \theta} - \frac{\partial \rho_{\tilde{\eta}}^p(\varepsilon_t(\theta))}{\partial \varepsilon_t(\theta)} \frac{\partial \varepsilon_t^m(\theta)}{\partial \theta} \\ & \quad + \frac{\partial \rho_{\tilde{\eta}}^p(\varepsilon_t(\theta))}{\partial \varepsilon_t(\theta)} \frac{\partial \varepsilon_t^m(\theta)}{\partial \theta} - \frac{\partial \rho_{\tilde{\eta}}^p(\varepsilon_t^m(\theta))}{\partial \varepsilon_t^m(\theta)} \frac{\partial \varepsilon_t^m(\theta)}{\partial \theta}. \end{aligned} \quad (\text{D.28})$$

That is,

$$\begin{aligned} & \frac{d\rho_{\tilde{\eta}}^p(\varepsilon_t(\theta))}{d\theta} - \frac{d\rho_{\tilde{\eta}}^p(\varepsilon_t^m(\theta))}{d\theta} \\ & = \frac{\partial \rho_{\tilde{\eta}}^p(\varepsilon_t(\theta))}{\partial \varepsilon_t(\theta)} \left(\frac{\partial \varepsilon_t(\theta)}{\partial \theta} - \frac{\partial \varepsilon_t^m(\theta)}{\partial \theta} \right) \\ & \quad + \left(\frac{\partial \rho_{\tilde{\eta}}^p(\varepsilon_t(\theta))}{\partial \varepsilon_t(\theta)} - \frac{\partial \rho_{\tilde{\eta}}^p(\varepsilon_t^m(\theta))}{\partial \varepsilon_t^m(\theta)} \right) \frac{\partial \varepsilon_t^m(\theta)}{\partial \theta}. \end{aligned} \quad (\text{D.29})$$

Using mean value theorem, we then get

$$\frac{\partial \rho_{\eta}^p(\varepsilon_t(\theta))}{\partial \varepsilon_t(\theta)} - \frac{\partial \rho_{\eta}^p(\varepsilon_t^m(\theta))}{\partial \varepsilon_t^m(\theta)} = \frac{\partial^2 \rho_{\eta}^p(\xi_t(\theta))}{\partial \xi_t(\theta)^2} (\varepsilon_t(\theta) - \varepsilon_t^m(\theta)). \quad (\text{D.30})$$

Hence,

$$\begin{aligned} & \left| \frac{d\rho_{\eta}^p(\varepsilon_t(\theta))}{d\theta} - \frac{d\rho_{\eta}^p(\varepsilon_t^m(\theta))}{d\theta} \right| \\ & \leq \left| \frac{\partial \rho_{\eta}^p(\varepsilon_t(\theta))}{\partial \varepsilon_t(\theta)} \right| \left| \frac{\partial \varepsilon_t(\theta)}{\partial \theta} - \frac{\partial \varepsilon_t^m(\theta)}{\partial \theta} \right| \\ & \quad + \left| \frac{\partial^2 \rho_{\eta}^p(\xi_t(\theta))}{\partial \xi_t(\theta)^2} \right| |\varepsilon_t(\theta) - \varepsilon_t^m(\theta)| \left| \frac{\partial \varepsilon_t^m(\theta)}{\partial \theta} \right|. \end{aligned} \quad (\text{D.31})$$

From the regularity conditions C1 in [36] given by

- (i) $\|\partial \rho(\varepsilon)/\partial \varepsilon\| \leq C|\varepsilon|, \theta \in D_M, \text{ all } t,$
- (ii) $\|\partial \rho(\varepsilon)/\partial \theta\| \leq C|\varepsilon|^2, \theta \in D_M, \text{ all } t,$
- (iii) $\|\partial^2 \rho(\varepsilon)/\partial \varepsilon^2\| \leq C,$

we then have

$$\begin{aligned} & \left| \frac{d\rho_{\eta}^p(\varepsilon_t(\theta))}{d\theta} - \frac{d\rho_{\eta}^p(\varepsilon_t^m(\theta))}{d\theta} \right| \\ & \leq C|\varepsilon_t(\theta)| \left| \frac{\partial \varepsilon_t(\theta)}{\partial \theta} - \frac{\partial \varepsilon_t^m(\theta)}{\partial \theta} \right| \\ & \quad + C|\varepsilon_t(\theta) - \varepsilon_t^m(\theta)| \left| \frac{\partial \varepsilon_t^m(\theta)}{\partial \theta} \right|. \end{aligned} \quad (\text{D.32})$$

In detail,

$$\begin{aligned} & \left| \frac{\partial \varepsilon_t(\theta)}{\partial \theta} - \frac{\partial \varepsilon_t^m(\theta)}{\partial \theta} \right| = |\tilde{\psi}_t^m(\theta)| \\ & \leq \sum_{k=m+1}^{\infty} |\delta_{t,k}| |e_{t-k}| \\ & \leq \sum_{k=m+1}^{\infty} \mu_k |e_{t-k}|, \\ & |\varepsilon_t(\theta) - \varepsilon_t^m(\theta)| = |\tilde{\varepsilon}_t^m(\theta)| \\ & \leq \sum_{k=m+1}^{\infty} |\beta_{t,k}| |e_{t-k}| \\ & \leq \sum_{k=m+1}^{\infty} \mu_k |e_{t-k}|. \end{aligned} \quad (\text{D.33})$$

Expression (D.32) becomes

$$\begin{aligned} & \left| \frac{d\rho_{\eta}^p(\varepsilon_t(\theta))}{d\theta} - \frac{d\rho_{\eta}^p(\varepsilon_t^m(\theta))}{d\theta} \right| \\ & \leq C \left(\sum_{k=m+1}^{\infty} \mu_k |e_{t-k}| \right) (|\varepsilon_t(\theta)| + |\psi_t^m(\theta)|). \end{aligned} \quad (\text{D.34})$$

Moreover,

$$\begin{aligned} |\varepsilon_t(\theta)| + |\psi_t^m(\theta)| & \leq 2C_u \sum_{k \geq 1} \mu_k + 2 \sum_{k=0}^m \mu_k |e_{t-k}| \\ & \quad + \sum_{k=m+1}^{\infty} \mu_k |e_{t-k}|. \end{aligned} \quad (\text{D.35})$$

Accordingly,

$$\left| \frac{d\rho_{\eta}^p(\varepsilon_t(\theta))}{d\theta} - \frac{d\rho_{\eta}^p(\varepsilon_t^m(\theta))}{d\theta} \right| \leq \alpha_t + \beta_t + \gamma_t, \quad (\text{D.36})$$

with

$$\begin{aligned} \alpha_t & = 2CC_u \left(\sum_{k \geq 1} \mu_k \right) \left(\sum_{k=m+1}^{\infty} \mu_k |e_{t-k}| \right), \\ \beta_t & = 2C \left(\sum_{k=0}^m \mu_k \right) \left(\sum_{k=m+1}^{\infty} \mu_k |e_{t-k}| \right), \\ \gamma_t & = C \left(\sum_{k=m+1}^{\infty} \mu_k |e_{t-k}| \right)^2. \end{aligned} \quad (\text{D.37})$$

Therefore,

$$\begin{aligned} X_{m,N}(\theta^*) & \leq \underbrace{\frac{1}{\sqrt{N}} \sum_{t=1}^N (\alpha_t - E\alpha_t)}_{X_{m,N}^{\alpha}(\theta^*)} + \underbrace{\frac{1}{\sqrt{N}} \sum_{t=1}^N (\beta_t - E\beta_t)}_{X_{m,N}^{\beta}(\theta^*)} \\ & \quad + \underbrace{\frac{1}{\sqrt{N}} \sum_{t=1}^N (\gamma_t - E\gamma_t)}_{X_{m,N}^{\gamma}(\theta^*)}. \end{aligned} \quad (\text{D.38})$$

Each term on the right-hand side of (D.38) verifies the corollary to Lemma 2B.1 in [14, page 57]. Hence, as $m \rightarrow \infty$,

$$\begin{aligned} E(X_{m,N}^{\alpha}(\theta^*))^2 & \leq K \left(\sum_{k \geq 1} \mu_k \right) \left(\sum_{k=m+1}^{\infty} \mu_k \right) \rightarrow 0, \\ E(X_{m,N}^{\beta}(\theta^*))^2 & \leq K \left(\sum_{k=0}^m \mu_k \right) \left(\sum_{k=m+1}^{\infty} \mu_k \right) \rightarrow 0, \\ E(X_{m,N}^{\gamma}(\theta^*))^2 & \leq K \left(\sum_{k=m+1}^{\infty} \mu_k \right)^2 \rightarrow 0. \end{aligned} \quad (\text{D.39})$$

Hence, $Z_{m,N}(\theta^*) \in \mathcal{AS}\mathcal{N}(0, Q_m(\theta^*))$ and $S_N(\theta^*) \in \mathcal{AS}\mathcal{N}(0, Q(\theta^*))$ with $Q(\theta^*) = \lim_{m \rightarrow \infty} Q_m(\theta^*)$, which proves the point (iv) of Theorem 11.

(v) Expanding $g^p(\theta)$ into Taylor series around θ^* , we then get

$$\begin{aligned} g^p(\hat{\theta}_N^p) &= g^p(\theta^*) + \frac{1}{2}(\hat{\theta}_N^p - \theta^*)^T g''^p(\theta^*)(\hat{\theta}_N^p - \theta^*) \\ &\quad + o(\|\hat{\theta}_N^p - \theta^*\|^2). \end{aligned} \quad (\text{D.40})$$

On the other hand, the Hessian $g''^p(\theta^*)$ is positive definite and nonsingular, so there is $C > 0$ and a neighborhood of θ^* with

$$\begin{aligned} \frac{1}{2}(\hat{\theta}_N^p - \theta^*)^T g''^p(\theta^*)(\hat{\theta}_N^p - \theta^*) \\ + o(\|\hat{\theta}_N^p - \theta^*\|^2) \leq C\|\hat{\theta}_N^p - \theta^*\|^2, \end{aligned} \quad (\text{D.41})$$

and we obtain $g^p(\hat{\theta}_N^p) \leq g^p(\theta^*) + C\|\hat{\theta}_N^p - \theta^*\|^2$. Moreover,

$$\begin{aligned} g_N^p(\hat{\theta}_N^p) - g_N^p(\theta^*) + o_p\left(\frac{1}{N}\right) \\ = g^p(\hat{\theta}_N^p) - g^p(\theta^*) + \left(\frac{\partial g_N^p(\xi)}{\partial \xi}\right)_{\theta^*}^T (\hat{\theta}_N^p - \theta^*) \\ + \|\hat{\theta}_N^p - \theta^*\| \hat{R}_N(\hat{\theta}_N^p) + o_p\left(\frac{1}{N}\right). \end{aligned} \quad (\text{D.42})$$

Therefore,

$$\begin{aligned} g_N^p(\hat{\theta}_N^p) - g_N^p(\theta^*) \\ \leq C\|\hat{\theta}_N^p - \theta^*\|^2 \\ + \left\| \frac{\partial g_N^p(\xi)}{\partial \xi} \right\|_{\theta^*} \|\hat{\theta}_N^p - \theta^*\| \\ + \|\hat{\theta}_N^p - \theta^*\| (1 + \sqrt{N}\|\hat{\theta}_N^p - \theta^*\|) o_p\left(\frac{1}{\sqrt{N}}\right) \\ + o_p\left(\frac{1}{N}\right). \end{aligned} \quad (\text{D.43})$$

Since $\|\partial g_N^p(\xi)/\partial \xi\|_{\theta^*} \rightarrow 0$, as $N \rightarrow \infty$, then

$$\begin{aligned} g_N^p(\hat{\theta}_N^p) - g_N^p(\theta^*) \leq (C + o_p(1))\|\hat{\theta}_N^p - \theta^*\|^2 \\ + o_p\left(\frac{1}{\sqrt{N}}\right)\|\hat{\theta}_N^p - \theta^*\| + o_p\left(\frac{1}{N}\right). \end{aligned} \quad (\text{D.44})$$

The remainder $\hat{R}_N(\hat{\theta}_N^p)$ can be written as

$$\hat{R}_N(\hat{\theta}_N^p) \leq \sqrt{N}\|\hat{\theta}_N^p - \theta^*\|(K + o_p(1)) \quad (\text{D.45})$$

and then

$$\frac{\hat{R}_N(\hat{\theta}_N^p)}{1 + \sqrt{N}\|\hat{\theta}_N^p - \theta^*\|} \leq \frac{\sqrt{N}\|\hat{\theta}_N^p - \theta^*\|(K + o_p(1))}{1 + \sqrt{N}\|\hat{\theta}_N^p - \theta^*\|}. \quad (\text{D.46})$$

Since $\sqrt{N}\|\hat{\theta}_N^p - \theta^*\| \xrightarrow{\text{prob}} 0$, then

$$\sup_{\|\hat{\theta}_N^p - \theta^*\| \leq \delta_N} \left| \frac{\hat{R}_N(\hat{\theta}_N^p)}{1 + \sqrt{N}\|\hat{\theta}_N^p - \theta^*\|} \right| \xrightarrow{\text{prob}} 0, \quad (\text{D.47})$$

which proves the point (v) and finally Theorem 11.

E. Proof of the Limit Hessian and the Q_{ω}^p -Matrix

The moments $m_{\mathcal{H},2}^0$ and $m_{\mathcal{H},1}^0$ are, respectively, given by $E|\varepsilon_{\gamma_{p_2},t}^0|^{p_2-2}$ and $E|\varepsilon_{\gamma_{p_1},t}^0|^{p_1-2}$. In detail, for $m_{\mathcal{H},2}^0$, we have

$$m_{\mathcal{H},2}^0 = 2(1 - \bar{\omega})C\hat{A} \quad (\text{E.1})$$

with $\hat{A} = \int_0^{\tilde{\eta}} X^{p_2-2} e^{-X^{p_2}/2\phi^{p_2}} dX$ and $C = 1/2(K_1 + K_2)$. After straightforward calculations, $\hat{A}$ is given by

$$\frac{2^{1-1/p_2}\phi^{p_2-1}}{p_2} \left[\Gamma\left(1 - \frac{1}{p_2}\right) - \Gamma\left(1 - \frac{1}{p_2}, \frac{\tilde{\eta}^{p_2}}{2\phi^{p_2}}\right) \right]. \quad (\text{E.2})$$

We can deduce $m_{\mathcal{H},2}^0$. For $m_{\mathcal{H},1}^0$, we have

$$m_{\mathcal{H},1}^0 = 2(1 - \bar{\omega})Ce^{\tilde{\eta}^2/2\phi^{p_2}}\hat{B} \quad (\text{E.3})$$

with $\hat{B} = \int_{\tilde{\eta}}^{\infty} X^{p_1-2} e^{-\tilde{\eta}X^{p_1}/\phi^{p_2}} dX$. With straightforward calculations,

$$\hat{B} = \frac{\phi^{p_2-p_2/p_1}}{p_1\tilde{\eta}^{1-1/p_1}} \int_{\tilde{\eta}^{p_1+1}/\phi^{p_2}}^{\infty} Z^{-1/p_1} e^{-Z} dZ. \quad (\text{E.4})$$

In (E.4), $\int_{\tilde{\eta}^{p_1+1}/\phi^{p_2}}^{\infty} Z^{-1/p_1} e^{-Z} dZ$ is convergent. Indeed, since $e^{-Z}/Z^{1/p_1} < 1/Z^{1/p_1}$, then, for $X > 0$, $\int_X^{\infty} e^{-Z}/Z^{1/p_1} dZ < \int_X^{\infty} 1/Z^{1/p_1} dZ$. Moreover, since $p_1 < 1$, then $1/p_1 > 1$ and $1/Z^{1/p_1} \sim 1/Z^{\alpha}$ ($\alpha > 1$), that is, the general term of Riemann's integral. We deduce

$$m_{\mathcal{H},1}^0 = 2(1 - \bar{\omega})e^{\tilde{\eta}^2/2\phi^{p_2}} C \frac{\phi^{p_2-p_2/p_1}}{p_1\tilde{\eta}^{1-1/p_1}} \hat{I} \quad (\text{E.5})$$

with $\hat{I} = \int_{\tilde{\eta}^{p_1+1}/\phi^{p_2}}^{\infty} Z^{-1/p_1} e^{-Z} dZ$.

Now, to express $Q_{\bar{\omega}}^p$ -matrix, we need the expectation of $g_N^{p'}(\theta_0)g_N^{p'}(\theta_0)^T$ as $N \rightarrow \infty$. More precisely, $Q_{\bar{\omega}}^p(\theta_0) = \lim_{N \rightarrow \infty} NEg_N^{p'}(\theta_0)g_N^{p'}(\theta_0)^T$. In detail, $NEg_N^{p'}(\theta_0)g_N^{p'}(\theta_0)^T$ contains three terms given by

$$\frac{p_2^2}{4N} \sum_{t \in \nu_{p_2}} \sum_{u \in \nu_{p_2}} E|\varepsilon_t^0|^{p_2-1} |\varepsilon_u^0|^{p_2-1} Es_t^0 s_u^0 E\psi_t^0 \psi_u^{0T}, \quad (\text{E.6})$$

$$\frac{\tilde{\eta} p_1}{N} \sum_{t \in \nu_{p_2}} \sum_{u \in \nu_{p_1}} E|\varepsilon_t^0|^{p_2-1} |\varepsilon_u^0|^{p_1-1} Es_t^0 s_u^0 E\psi_t^0 \psi_u^{0T}, \quad (\text{E.7})$$

$$\frac{\tilde{\eta}^2 p_1^2}{N} \sum_{t \in \nu_{p_1}} \sum_{u \in \nu_{p_1}} E|\varepsilon_t^0|^{p_1-1} |\varepsilon_u^0|^{p_1-1} Es_t^0 s_u^0 E\psi_t^0 \psi_u^{0T}. \quad (\text{E.8})$$

Since $Es_t^0 s_u^0 = 0$, if $(t, u) \in \nu_{p_i} \cap \nu_{p_j}$, with $i \neq j$, then (E.7) is zero and (E.6), (E.8) become respectively

$$\frac{p_2^2}{4N} \sum_{t \in \nu_{p_2}} E|\varepsilon_t^0|^{2p_2-2} E\Phi_t^0, \quad (\text{E.9})$$

$$\frac{\tilde{\eta}^2 p_1^2}{N} \sum_{t \in \nu_{p_1}} E|\varepsilon_t^0|^{2p_1-2} E\Phi_t^0.$$

Accordingly, the $Q_{\bar{\omega}}^p$ -matrix is

$$Q_{\bar{\omega}}^p(\theta_0) = \frac{p_2^2}{4} m_{\mathcal{H},3}^0 \bar{E}\Phi_{\nu_{p_2},t}^0 + \tilde{\eta}^2 p_1^2 m_{\mathcal{H},4}^0 \bar{E}\Phi_{\nu_{p_1},t}^0, \quad (\text{E.10})$$

with the moments $m_{\mathcal{H},3}^0 = E|\varepsilon_{\nu_{p_2},t}^0|^{2p_2-2}$ and $m_{\mathcal{H},4}^0 = E|\varepsilon_{\nu_{p_1},t}^0|^{2p_1-2}$. These moments are

$$m_{\mathcal{H},3}^0 = 2(1-\bar{\omega}) C \bar{A}^* \quad (\text{E.11})$$

with

$$\bar{A}^* = \frac{2^{2-1/p_2} \phi^{2p_2-1}}{p_2} \left[\Gamma\left(2 - \frac{1}{p_2}\right) - \Gamma\left(2 - \frac{1}{p_2}, \frac{\tilde{\eta} p_2}{2\phi^{p_2}}\right) \right],$$

$$m_{\mathcal{H},4}^0 = 2(1-\bar{\omega}) C e^{\tilde{\eta}^2/2\phi^{p_2}} \bar{B}^* \quad (\text{E.12})$$

with

$$\bar{B}^* = \frac{\phi^{(p_2/p_1)(p_2+p_1-1)}}{p_1 \tilde{\eta}^{(p_2+p_1-1)/p_1}} \Gamma\left(\frac{p_2+p_1-1}{p_1}, \frac{\tilde{\eta}^{p_1+1}}{\phi^{p_2}}\right). \quad (\text{E.13})$$

Now, let us denote

$$\alpha_{\bar{\omega}}^0 = \frac{p_2(p_2-1)}{2} m_{\mathcal{H},2}^0; \quad \beta_{\bar{\omega}}^0 = \tilde{\eta} p_1(p_1-1) m_{\mathcal{H},1}^0,$$

$$\gamma_{\bar{\omega}}^0 = \frac{p_2^2}{4} m_{\mathcal{H},3}^0; \quad \delta_{\bar{\omega}}^0 = \tilde{\eta}^2 p_1^2 m_{\mathcal{H},4}^0,$$

$$\mathcal{R}_{\bar{\omega}}^0 = \frac{\beta_{\bar{\omega}}^0}{\alpha_{\bar{\omega}}^0}; \quad \mathcal{V}^0 = \frac{\gamma_{\bar{\omega}}^0}{\alpha_{\bar{\omega}}^0},$$

$$\mathcal{X}_{\bar{\omega}}^0 = \frac{\gamma_{\bar{\omega}}^0 \mathcal{R}_{\bar{\omega}}^0 - \delta_{\bar{\omega}}^0}{\alpha_{\bar{\omega}}^0}; \quad \mathcal{Y}_{\bar{\omega}}^0 = \frac{\delta_{\bar{\omega}}^0 \mathcal{R}_{\bar{\omega}}^0}{\alpha_{\bar{\omega}}^0}. \quad (\text{E.14})$$

In the restricted case, where $1.8 \leq p_2 \leq 2$ with $\tilde{\eta} < 1$, $|\mathcal{R}_{\bar{\omega}}^0| \ll 1$. Then, the limit Hessian matrix can be written as

$$\overline{g_{\bar{\omega}}^{p''}}(\theta_0)^{-1} = \frac{[\bar{E}\Phi_{\nu_{p_2},t}^0 + \mathcal{R}_{\bar{\omega}}^0 \bar{E}\Phi_{\nu_{p_1},t}^0]^{-1}}{\alpha_{\bar{\omega}}^0}. \quad (\text{E.15})$$

Since $|\mathcal{R}_{\bar{\omega}}^0| \ll 1$, an approximation of $\overline{g_{\bar{\omega}}^{p''}}(\theta_0)^{-1}$ is considered; therefore,

$$[\bar{E}\Phi_{\nu_{p_2},t}^0 + \mathcal{R}_{\bar{\omega}}^0 \bar{E}\Phi_{\nu_{p_1},t}^0]^{-1} \approx (\bar{E}\Phi_{\nu_{p_2},t}^0)^{-1} - \mathcal{R}_{\bar{\omega}}^0 (\bar{E}\Phi_{\nu_{p_2},t}^0)^{-1} \bar{E}\Phi_{\nu_{p_1},t}^0 (\bar{E}\Phi_{\nu_{p_2},t}^0)^{-1} \quad (\text{E.16})$$

and we obtain

$$\overline{g_{\bar{\omega}}^{p''}}(\theta_0)^{-1} \approx \frac{1}{\alpha_{\bar{\omega}}^0} \left[(\bar{E}\Phi_{\nu_{p_2},t}^0)^{-1} - \mathcal{R}_{\bar{\omega}}^0 (\bar{E}\Phi_{\nu_{p_2},t}^0)^{-1} \bar{E}\Phi_{\nu_{p_1},t}^0 (\bar{E}\Phi_{\nu_{p_2},t}^0)^{-1} \right]. \quad (\text{E.17})$$

On the other hand,

$$Q_{\bar{\omega}}^p(\theta_0) = \gamma_{\bar{\omega}}^0 \bar{E}\Phi_{\nu_{p_2},t}^0 + \delta_{\bar{\omega}}^0 \bar{E}\Phi_{\nu_{p_1},t}^0. \quad (\text{E.18})$$

F. Proof of Theorem 14

From (34), we have

$$|W_N(\theta)| \leq \frac{1}{2N} \eta^2 (\text{card}\{\nu_2(\theta)\} + \text{card}\{\nu_1(\theta)\}) + \frac{\eta}{N} |\Omega_i| \text{card}\{\nu_1(\theta)\}. \quad (\text{F.1})$$

Since $\text{card}(\nu_2(\theta)) + \text{card}(\nu_1(\theta)) = N$ with $\text{card}(\nu_1(\theta)) < N$, then for all and $\eta < 1$

$$\sup_{\theta} |W_N(\theta)| = 1 + |\Omega_i|. \quad (\text{F.2})$$

From (11) and DCC (5), with $\text{card}(\nu_{p_1}(\tilde{\theta})) < N$, then for all $\tilde{\theta}$

$$\sup_{\tilde{\theta}} |g_N^p(\tilde{\theta})| = 1 + |\Omega_i|^{p_1}. \quad (\text{F.3})$$

Since $p_1 < 1$, then $|\Omega_i|^{p_1} < |\Omega_i|$, which proves the theorem.

G. Proof of Theorem 15

Let us denote $\Delta \bar{\mathcal{F}} = \sup_{\bar{\theta}_N^H} |\bar{\mathcal{F}}_{\bar{\theta}_N^H}(\bar{\mathbf{M}}; \bar{\omega})| - \sup_{\bar{\theta}_N^H} |\bar{\mathcal{F}}_{\bar{\theta}_N^H}(\mathbf{M}; \omega)|$. From (36) and (37), we have

$$\Delta \bar{\mathcal{F}} = |\Omega_i|^{p_1} - |\Omega_i| + \frac{d_M^p}{N} (1 + |\mathcal{V}^0|) - \frac{2\hat{\lambda}(\omega) d_M^H}{N}. \quad (\text{G.1})$$

Since $p_1 < 1$, then $|\Omega_i|^{p_1} < |\Omega_i|$. $\Delta \bar{\mathcal{F}} < 0$ if and only if $(d_M^p/N)(1 + |\mathcal{V}^0|) - (2\hat{\lambda}(\omega) d_M^H/N) < 0$. We then obtain

$$d_M^p < \frac{2\hat{\lambda}(\omega)}{1 + |\mathcal{V}^0|} d_M^H. \quad (\text{G.2})$$

The model order reduction condition $\mathcal{R}_d < 1$ is tuned by the main parameters ω , $\tilde{\eta}$, and p_2 by means of $\mathcal{V}^0$, which proves the theorem.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] G. Appa and C. Smith, "On L_1 and Chebyshev estimation," *Mathematical Programming*, vol. 5, pp. 73–87, 1973.
- [2] R. W. Farebrother, "Unbiased L_1 and L_∞ estimation," *Communications in Statistics A Theory and Methods*, vol. 14, no. 8, pp. 1941–1962, 1985.
- [3] T. T. Pham and R. J. P. de Figueiredo, "Maximum likelihood estimation of a class of non-Gaussian densities with application to l_p ," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 1, pp. 73–82, 1989.
- [4] H. Akcay, H. Hjalmarsson, and L. Ljung, "On the choice of norms in system identification," *IEEE Transactions on Automatic Control*, vol. 41, no. 9, pp. 1367–1372, 1996.
- [5] A. J. Fox, "Outliers in time series," *Journal of the Royal Statistical Society*, vol. 34, no. 3, pp. 350–363, 1972.
- [6] K. C. Tan and Y. Li, "Evolutionary L_∞ identification and model reduction for robust control," *Proceedings of the Institution of Mechanical Engineers I*, vol. 214, no. 3, pp. 231–237, 2000.
- [7] D. W. McMichael, "Robust recursive L_p estimation," *IEE Proceedings D: Control Theory and Applications*, vol. 137, no. 2, pp. 67–76, 1990.
- [8] O. V. Ivanov and I. V. Orlovskii, "Asymptotic normality of L_p -estimators in nonlinear regression models with weak dependence," *Theory of Probability and Mathematical Statistics*, no. 79, pp. 57–72, 2009.
- [9] B. S. Chen, J. M. Chen, and S. C. Shern, "ARMA robust system identification using a generalized l_p norm estimation algorithm," *IEEE Transactions on Signal Processing*, vol. 42, no. 5, pp. 1063–1073, 1994.
- [10] J. M. Chen, B. S. Chen, and W. S. Chang, "Parameter estimation of linear systems with input-output noisy data: a generalized l_p norm approach," *Signal Processing*, vol. 37, no. 3, pp. 345–356, 1994.
- [11] M. Schreiber, "Some remarks on the characterisations of L_p spaces with $0 \leq p < 1$," *Annales de l'Institut Henri Poincaré B*, vol. 8, no. 1, pp. 83–92, 1972.
- [12] M. Ilak, S. Bagheri, L. a. Brandt, and D. S. Henningson, "Model reduction of the nonlinear complex Ginzburg-Landau equation," *SIAM Journal on Applied Dynamical Systems*, vol. 9, no. 4, pp. 1284–1302, 2010.
- [13] P. Eckhoff, K. Wong-Lin, and P. Holmes, "Dimension reduction and dynamics of a spiking neural network model for decision making under neuromodulation," *SIAM Journal on Applied Dynamical Systems*, vol. 10, no. 1, pp. 148–188, 2011.
- [14] L. Ljung, *System Identification: Theory for the User*, Prentice Hall, New York, NY, USA, 1999.
- [15] B. Porat and B. Friedlander, "Asymptotic accuracy of ARMA parameter estimation methods based on sample covariances," in *Proceedings of the 7th IFAC Symposium on Identification and System Parameter Estimation*, pp. 963–968, York, UK, 1985.
- [16] B. Porat, "On the estimation of the parameters of vector Gaussian processes from sample covariances," in *Proceedings of the 25th Conference on Decision and Control*, pp. 2002–2005, Athens, Greece, December 1986.
- [17] T. Söderström, P. Stoica, and B. Friedlander, "An indirect prediction error method for system identification," *Automatica*, vol. 27, no. 1, pp. 183–188, 1991.
- [18] P. Stoica and T. Söderström, "On reparametrization of loss functions used in estimation and the invariance principle," *Signal Processing*, vol. 17, no. 4, pp. 383–387, 1989.
- [19] Y. Zhu and T. Backx, *Identification of Multivariable Industrial Processes for Simulation, Diagnosis and Control*, Springer, Berlin, Germany, 1993.
- [20] F. Tjårnström and L. Ljung, " L_2 model reduction and variance reduction," *Automatica*, vol. 38, no. 9, pp. 1517–1530, 2002.
- [21] F. Tjårnström, "Variance analysis of L_2 model reduction when undermodeling—the output error case," *Automatica*, vol. 39, no. 10, pp. 1809–1815, 2003.
- [22] Y. Jiang and H. Chen, "Time domain model order reduction of general orthogonal polynomials for linear input-output systems," *IEEE Transactions on Automatic Control*, vol. 57, no. 2, pp. 330–343, 2012.
- [23] P. Rapisarda and H. L. Trentelman, "Trentelman Identification and data-driven model reduction of state-space representations of lossless and dissipative systems from noise-free data," *Automatica*, vol. 22, no. 3, pp. 295–308, 2011.
- [24] J. C. Carmona and V. M. Alvarado, "Active noise control of a duct using robust control theory," *IEEE Transactions on Control Systems Technology*, vol. 8, no. 6, pp. 930–938, 2000.
- [25] J. C. Carmona and V. Alvarado, "A new approach model validation: application to time delay systems," in *Proceedings of the 2nd IFAC Workshop on Linear Time Delay Systems*, Ancona, Italy, September 2000.
- [26] J. C. Carmona and V. M. Alvarado, " L_1 prediction error approach in system identification," in *Proceedings of the American Control Conference*, vol. 4, pp. 3219–3223, Anchorage, Alaska, USA, May 2002.
- [27] P. J. Huber and E. M. Ronchetti, *Robust Statistics*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2009.
- [28] C. Corbier, J.-C. Carmona, and V. Alvarado, " $L_1 - L_2$ robust estimation in prediction error system identification," in *Proceedings of the CINVESTAV IEEE Congress*, Toluca, Mexico, November 2009.
- [29] C. Corbier, *Contribution to the robust estimation of dynamic models: application to the command of complex dynamic systems [Thesis]*, Arts et Métiers ParisTech, 2012.
- [30] W. K. Newey and D. L. McFadden, "Large sample estimation and hypothesis testing," in *Handbook of Econometric*, vol. 4, pp. 2113–2247, Elsevier Science, 1994.
- [31] D. W. K. Andrews, "Large sample estimation and hypothesis testing," in *Handbook of Econometric*, vol. 4, pp. 2248–2292, Elsevier Science, 1994.
- [32] S. Orey, "A central limit theorem for m -dependent random variables," *Duke Mathematical Journal*, vol. 25, pp. 543–546, 1958.
- [33] B. Rosén, "On the central limit theorem for sums of dependent random variables," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 7, no. 1, pp. 48–82, 1967.
- [34] P. H. Diananda, "Some probability limit theorems with statistical applications," *Proceedings of Cambridge Philosophical Society*, vol. 49, pp. 239–246, 1953.

- [35] T. W. Anderson, "On asymptotic distributions of estimates of parameters of stochastic difference equations," *Annals of Mathematical Statistics*, vol. 30, pp. 676–687, 1959.
- [36] L. Ljung, "Convergence analysis of parametric identification methods," *IEEE Transactions on Automatic Control*, vol. 23, no. 5, pp. 770–783, 1978.