Robust estimate of excitations in mechanical systems using M-estimators – Theoretical background and numerical applications

Paolo Pennacchi
Dipartimento di Meccanica, Politecnico di Milano, Via La Masa 34, I-20156 Milano, Italy
paolo.pennacchi@polimi.it

Abstract

Model based methods are often used along with least squares to estimate (or to identify in equivalent but in more engineering terms) dynamic forces, parameters and malfunctions in mechanical systems, starting from experimental vibrations.

The effectiveness of these methods, broadly proven and documented by several cases of study, can be reduced if the model of the system is not accurate or if the experimental data are corrupted by noise, especially if the mean value of the noise is not null or if biases are present.

A possible solution is the use of robust estimation techniques instead of traditional least squares in the ambit of model based identification. The author proposes the application of the M-estimators and discusses the problems related to their application to excitation identification in mechanical systems.

In this paper the necessary theory is presented in detail, introducing several concepts of Statistics, in order to properly introduce the concept of robust estimation and the required algorithms (based on iterative re-weighted least squares) are described. Then the different types of M-estimators proposed in literature are introduced. Their performances with regard to mechanical applications are evaluated by means of a theoretical analysis and a couple of simple numerical examples: a single input – single output and a multiple inputs – multiple outputs systems. Moreover the problem of the scale parameter, which is not discussed in literature for complex numbers, as the vibrations are, is analyzed and a solution is proposed using a concept related to the data depth.

Keywords: *Robust estimation; identification; parameter estimation; inverse problems; M-estimators; least squares; iterated re-weighted least squares; numerical results.*

1. Introduction

The least squares estimate is widely used in mechanical systems to solve identification problems. These can be for instance the estimate of the excitations starting from the dynamical response and from the knowledge of the system parameters. Conversely also the system parameters can be evaluated by identifying the experimental frequency response function (FRF). A comprehensive overview is given for instance in [1], by limiting to the mechanical field only, where also some remarks are given about the sensitivity of least squares to bias. The sensitivity of least squares to data corruption is the main drawback of this method and several improvements have been proposed in literature. One of the simplest is the introduction of weighted least squares and

some successful applications in model based identification of faults in rotor-dynamics are presented for instance in [2][3]. In these applications the faults are represented by means of equivalent excitations.

Another recent proposal [4] to improve the method of least squares method is focused on the application of suitable regularization filters, namely the *truncated singular value decomposition filter* and the *Tikhonov's filter*, along with total least squares. Another approach proposed in [5] uses non-linear least squares to estimate parameters of a multi-frequency signal from discrete-time observations corrupted by additive noise.

Some of the proposed techniques in literature have the general task to obtain the maximum efficiency in the estimate. In [6] a maximum likelihood estimator is used to generalize total least squares and it is shown that this estimate has the maximum efficiency. Anyhow, under a statistical point of view, the maximum efficiency is antithetic to the robustness of the estimate, as it will be shown later on in the paper.

In any case practically everyone agrees about the fact that an estimator should be robust, but also the concept of robustness is somewhat vague. In some cases the robustness is intended as something that makes the estimator less sensitive to noise, both on system input and output. Some algorithms that have the task of reducing this sensitivity, both introducing the noise model in the parametric identification (the *generalized total least squares* [7]) and maintaining the nearly maximum likelihood property (the *bootstrapped total least squares* [8]), have been proposed and compared in [9] when they are applied to modal analysis. In other cases the introduction of a rigorous concept of robustness is avoided and the term *immunity*, similar to that of living bodies, is used, along with the developing of a modified least squares algorithm [10]. Anyhow the problem of the robust estimation is not new in Statistics and useful references can be found in [11][12][13][14].

The aim of the author is to give a precise definition of the robustness and to apply robust estimate to mechanical problems, trying to preserve a high level of efficiency. A suitable trade-off between robustness and efficiency is represented by the class of M-estimators. Their application to excitation identification of the in mechanical system is discussed in the paper. M-estimate has never been applied to mechanical systems, to the author's knowledge. In Statistics the data sets are generally composed of real numbers. In Mechanics vibrations are conveniently represented by complex numbers and the M-estimate application to complex data sets has never been presented before. Therefore its implementation is fully discussed, by first reviewing the related theory and then introducing the different types of M-estimators proposed in statistical literature. Since it is not possible to define *a priori* which one of the M-estimators is more suitable to mechanical applications, their performances are evaluated by means of a theoretical analysis and two simple numerical examples of a single input – single output (SISO) and a multiple inputs – multiple outputs (MIMO) systems, in which the knowledge of the model is perfect and noise affects only the output(s).

The use of M-estimate implies the evaluation of a scale parameter for the data sample. This problem is not discussed in statistical literature because normally data are real numbers. Since

vibrations are conveniently represented by complex numbers, data samples are complex and the evaluation of the scale parameter is not trivial. Also this problem is analyzed and a solution is proposed by using a concept related to the data depth.

2. Estimation, least squares, robustness and influence function

Not only in the mechanical field, there are many cases in which two or more variables are related by themselves and the relationship is made explicit by means of a model. The estimate of model parameters is often made using the least squares method (described in appendix A).

Anyhow, the estimators in the least squares sense rely on some fundamental hypotheses on the distribution of errors between the variable of interest and the data. In fact the noise that corrupts the data is assumed to have null mean value and this implies the estimate of a not distorted parameter. If the variance of the noise is known, a minimal variance estimate of the parameters, can be obtained by using suitable weights for the data.

Several studies have shown that least squares estimators are vulnerable to the violation of these hypotheses. For example the distribution of the errors could be asymmetric or prone to extreme outliers. Sometimes even a "bad" observation only can completely perturb the least squares estimate. Therefore many robust techniques have been proposed, but the robustness is a vague property of statistical procedures and is well outlined by Bickel [12]: *"A robust procedure, like a robust individual, performs well not only under ideal conditions, the model assumptions that have been postulated, but also under departures from the ideal. The notion is vague insofar as the type of departure and the meaning of "good performance" need to be specified"*. In the following, the robustness concept that the author intends is focused.

2.1. Evaluation of the influence function, of the robustness and of the efficiency

One of the causes of the inaccuracy of the probabilistic model assumed for the data is the presence of errors due to *outliers*, i.e. anomalous values, defined as observations far away from the majority of the data.

The analytical tool that allows the evaluation of the robustness of an estimator in presence of outliers and gross errors is the *influence function*, introduced by Hampel [15][16]. In order to allow a more intuitive interpretation of its definition (as suggested in [17] and [18]), it is introduced as the limit of the *sensitivity curve*, proposed by Tukey [19] to evaluate the stability of an estimator. Let $(X_1, X_2, \ldots, X_n)$ be a random sample extracted from $X \sim F(x;\theta)$ and the parameter θ is going to be estimated by means of the estimator $T_n = T(X_1, X_2, \ldots, X_n)$. Estimator T_n can be defined using the empirical distribution function:

$$F_n(u) = \frac{1}{n}\sum I\left(X_i \le u\right), \tag{1}$$

which attributes to every observation equal probability $1/n$. The operator $I(\cdot)$ indicates the number of i that satisfy the condition $\cdot$ inside the parentheses. Obviously, the knowledge of the random

sample or of the empirical distribution function is equivalent, because the random sample is immediately inferred from $F_n(x)$ and, vice versa, the sample is immediately inferred from the empirical distribution function because $F_n(x)$ introduces some jumps in correspondence of the sample values. This justifies denoting the estimator as a functional of the empirical distribution function: $T_n = T[F_n]$.

Now, if an arbitrary observation x is added to the random sample $(X_1, X_2, \ldots, X_n)$, a new sample $(X_1, X_2, \ldots, X_n, x)$ is obtained, the empirical distribution function of which is $F_{n+1}(u;x)$. It is easy to show that:

$$F_{n+1}(u;x) = \frac{1}{n+1}\sum_{i=1}^{n} I\left(X_i \le u\right) + \frac{1}{n+1} I\left(x \le u\right) = \frac{n}{n+1} F_n(u) + \frac{1}{n+1} I\left(x \le u\right). \tag{2}$$

If $\delta_x(u)$ indicates the distribution function of a random anomalous variable in x and $\varepsilon = \frac{1}{n+1}$, then:

$$F_{n+1}(u;x) = (1-\varepsilon)F_n(u) + \varepsilon\,\delta_x(u) = (1-\varepsilon)F_n + \varepsilon\,\delta_x. \tag{3}$$

Similarly, when the estimator $T_n = T[F_n]$, defined on $(X_1, X_2, \ldots, X_n)$, shifts on the extended sample $(X_1, X_2, \ldots, X_n, x)$, the functional becomes:

$$T_{n+1}(x) = T\left[F_{n+1}(u;x)\right] = T\left[(1-\varepsilon)F_n + \varepsilon\,\delta_x\right]. \tag{4}$$

The condition that makes T_n an acceptable estimate for θ is that the addition of an observation x to the random sample does not strongly modify its value. The most straightforward way to measure the effect of this addition is to consider the difference between $T_{n+1}(x)$ and T_n and to compare it to the weight of the added observation (measured by ε), that obviously is inversely proportional to the sample number. This defines the *sensitivity curve SC* of the estimator T_n [19]:

$$SC_n(x,T) = \frac{T_{n+1}(x) - T_n}{\varepsilon} = \frac{T\left[(1-\varepsilon)F_n + \varepsilon\,\delta_x\right] - T[F_n]}{\varepsilon}. \tag{5}$$

The sensitivity curve can be studied as a function of x for an estimator T_n. If the observation x is far away from the majority of the data, the curve of sensitivity shows what happens to the estimator when this outlier is present in the sample. Therefore it is necessary that $SC_n(x,T)$ is limited, so that the effect of an outlier on the estimator is always restricted within defined limits.

If $n \to \infty$, the Glivenko-Cantelli's theorem states that $F_n(x)$ converges uniformly and in distribution to $F(x)$. Moreover for Fisher's consistent estimators $T[F_n] \to T[F] = \theta$ and $T[F_n]$ can be replaced asymptotically by $T[F]$. The *influence function IF* of the estimator T_n with respect to $F(x;\theta)$ is:

$$IF(x,T,F) = \lim_{\varepsilon \to 0} \frac{T\left[(1-\varepsilon)F + \varepsilon\,\delta_x\right] - T[F]}{\varepsilon}. \tag{6}$$

The influence function shows the asymptotic variation of $T[F]$ due to an infinitesimal contamination of the distribution $F(x)$, related to the contamination entity. The influence function depends on x, as the sensitivity curve, but also on $F(x)$, that is on the parametric model supposed

for the data. Therefore it represents a parametric tool that can be used to verify the behaviour of an estimator in the cases in which the "actual" distribution is similar to the hypothesized distribution $F(x)$. The influence function will be used in paragraph 3.1.2 to discuss the robustness of the least squares. It allows also a robustness measure to be introduced and the robustness concept to be finally pointed out. The *gross error sensitivity* γ [12][16] is defined as:

$$\gamma(T,F) = \sup_x \left| IF(x,T,F) \right|. \tag{7}$$

If $\gamma(T,F) < +\infty$, the estimator is *robust* with respect to outliers, i.e. to anomalous values.

The robustness requires instead that the influence function is superiorly limited, but this requirement is in contrast to the estimator efficiency, which is often the target of the identification methods in mechanical systems [6]. In fact a theorem presented in [16] states: let $X \sim F(x,\theta)$ and $T_n = T(X_1, X_2, \ldots, X_n)$ be an estimator for θ and let $V_n'(\theta)$ be the related score function; if T_n is Fisher's consistent, under regularity conditions valid for Cramér-Rao's inequality [20], then T_n is efficient for θ if $IF(x,T,F) \propto V_n'(\theta)$. The *score function* $V_n'(\theta)$ is defined as:

$$V'(x;\theta) = \frac{\partial}{\partial\theta} \log f(x;\theta) = \frac{f'(x;\theta)}{f(x;\theta)}, \tag{8}$$

where f is the probability density function. The score function is normally unbounded, so that the influence function has to be both limited and unlimited (proportional to the score function) to achieve respectively estimator robustness and efficiency. To solve this paradox, the M-estimator class has been introduced [21] [22] [23].

2.2. Definition of M-estimate

Instead of estimate parameter $\hat{\theta}$ by means of the minimization of the quadratic error:

$$\sum_{i=1}^{n} (x_i - \theta)^2 = \sum_{i=1}^{n} r_i^2, \tag{9}$$

as it is done in the standard least squares, the following quantity is minimized:

$$\sum_{i=1}^{n} \rho(x_i - \theta) = \sum_{i=1}^{n} \rho(r_i), \tag{10}$$

where the type of function ρ will be discussed in section 3. To obtain the minimum, eq. (10) is derived with respect to r_i and put equal to zero. Let $\underline{\theta} = [\theta_1, \ldots, \theta_m]$ be the vector of the parameters to be estimated.

The *M-estimator* of $\underline{\theta}$ based on function $\rho(r_i)$ is the vector $\underline{\theta}$ solution of the m equations:

$$\sum_{i=1}^{n} \psi(r_i) \frac{\partial r_i}{\partial \theta_j} = 0 \quad \text{for} \quad j = 1, \ldots, m, \tag{11}$$

where the derivate:

$$\psi(r_i) = \frac{\mathrm{d}\,\rho(r_i)}{\mathrm{d}\,r_i}, \tag{12}$$

is proportional to the influence function of ρ. This is shown in appendix B.

If the estimator is robust, the influence of a single observation is insufficient to cause a significant error. A weight function is then defined as:

$$w(r_i) = \frac{\psi(r_i)}{r_i}, \tag{13}$$

so that eq. (11) is rewritten as:

$$\sum_{i=1}^{n} w(r_i) r_i \frac{\partial r_i}{\partial \theta_j} = 0 \quad \text{for} \quad j = 1, ..., m. \tag{14}$$

The equation system (14) is corresponding to that obtained when the iterated re-weighted least squares (IRLS) problem [24][25] is solved:

$$\min \sum_{i=1}^{n} w\left(r_i^{(t-1)}\right) r_i^2, \tag{15}$$

where t is the number of iteration index and the weights $w\left(r_i^{(t-1)}\right)$ have to be calculated per each iteration. The detailed description of the IRLS algorithm is presented in paragraph 4.1.

2.3. Conditions on the ρ functions

In order to be robust and to have good computational characteristics, an M-estimator should comply with some conditions that are reflected on its ρ function:

1) The influence function eq. (12) has to be bounded, as shown in paragraph 2.1 [18].
2) ρ function must have these properties [24]:
   - $\rho(r) \geq 0$;
   - $\rho(0) = 0$;
   - $\rho(r) = \rho(-r)$;
   - $\rho(r_i) \geq \rho(r_{i'}), \quad |r_i| > |r_{i'}|$;
   - ρ is less increasing than a quadratic function.

   In practice it is required that the ρ function has the same "good" properties of the least squares with the additionally obvious condition that ρ is less increasing than a quadratic function in order to limit the influence of the outliers.
3) The robust estimator should be unique, i.e. the objective function eq. (10) should have a unique minimum. This requires that the individual ρ function is convex in the variable $\underline{\theta}$, which is equivalent to impose that $\partial^2 \rho / \partial \underline{\theta}^2$ is non-negative definite [26].
4) A practical requirement is that whenever $\partial^2 \rho / \partial \underline{\theta}^2$ is singular, the objective function should have a gradient, that is $\partial \rho / \partial \underline{\theta} \neq 0$. This avoids having to search for a minimum through the complete parameter space.

Not all the M-estimators proposed in literature actually comply with all of these conditions. In the following section 3, the types proposed are analyzed in detail and some performance characteristics can be forecasted, while the applications in sections 4 and 5 make explicit the advantages and the drawbacks in mechanical applications.

3. Scale estimate and types of M-estimators

Before considering the possible types of M-estimators, it is necessary to introduce a discussion about the scale of the sample. In paragraph 2.2 it has been implicitly assumed a unitary scale, while in general the scale of the sample is not unitary and it should also be estimated. The problem of eq. (11) should be stated as:

$$\sum_{i=1}^{n} \psi\left(\frac{x_i - \theta_j}{\hat{\sigma}}\right)\frac{\partial r_i}{\partial \theta_j} = 0\,. \tag{16}$$

Obviously the same scale estimate $\hat{\sigma}$ has to be robust. Since in the cases commonly analyzed in Statistics, the observed data are real values, $x \in \mathbb{R}$, the estimates of scale are real values too and normally the *MAD* (median absolute deviation) is used [21][24][26].

Anyhow, parameter estimate in mechanical systems uses vibration data that are complex numbers, $x \in \mathbb{C}$. No studies exist regarding the application of M-estimate to complex quantities to the author's knowledge. Nevertheless a rich literature exists about the non-parametric estimate of robust location parameters in multidimensional distributions [27][28][29][30][31] and many of them are based on the concept of *data depth* [32]. These studies can be extended to the present case, in which the residues are complex numbers, $r \in \mathbb{C}$, and a scale estimate for bivariate data should be employed. A proposal, which will be adopted in following by the author, is based on the extension of the *MAD* to the complex field using the Tukey's median instead of the conventional median. Let $\mathbf{X} = \{\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_n\}$, $\overline{x}_i \in \mathbb{C}$, be the bivariate data vector composed by the complex vibration measures and $T^*(\cdot)$ the Tukey's median operator. The extension of the *MAD* is here defined as *TMAD* (Tukey's median absolute deviation):

$$TMAD(\mathbf{X}) = \mathrm{Med}\left(\left|\mathbf{X} - T^*(\mathbf{X})\right|\right). \tag{17}$$

Details on Tukey's median calculation are presented in [33][34].

3.1. Least Powers

This is a wide class of ρ functions that does not depend on the pre-emptive knowledge of a measure of scale. The general definition is:

$$\rho\left(x-\theta\right) = \left|x-\theta\right|^{\nu} \quad \leftrightarrow \quad \rho\left(r\right) = \left|r\right|^{\nu}. \tag{18}$$

Depending on the value of the exponent ν, several estimators are defined. Note that the conditions expressed in point 2 of paragraph 2.3 are satisfied also if $r \in \mathbb{C}$.

3.1.1 L_1 - Least Absolute or Absolute Value

In this case $\nu = 1$ and:

$$\rho(r) = |r|; \quad \psi(r) = \text{sgn}(r); \quad w(r) = \frac{1}{|r|}. \tag{19}$$

This estimator has a bounded influence function $\psi(r)$, thus reduces the influence of large errors, but has the disadvantage of possible numerical instability, because the ρ function $|r|$ is not strictly convex in r, since second derivative is unbounded in 0 and an indeterminate solution may result.

3.1.2 L_2 - Least Squares

The classical least squares are obtained for $\nu = 2$ and results:

$$\rho(r) = \frac{r^2}{2}; \quad \psi(r) = r; \quad w(r) = 1. \tag{20}$$

Although this estimator is convex, the influence function $\psi(x)$ is a straight line in r, insofar $\psi(x)$ is not bounded and this estimator is not robust, as well-known.

3.1.3 $L_1 - L_2$

This estimator does not derive directly from eq. (18), but is intended to maintain the advantages of both L_1 to reduce the influence of large errors and L_2 to be convex. It behaves like L_2 for small residues r and like L_1 otherwise.

$$\rho(r) = 2\left(\sqrt{1+\frac{r^2}{2}} - 1\right); \quad \psi(r) = \frac{r}{\sqrt{1+\frac{r^2}{2}}}; \quad w(r) = \frac{1}{\sqrt{1+\frac{r^2}{2}}}. \tag{21}$$

3.1.4 L_p - Least power

From the consideration of L_1 and L_2, it appears that as smaller the exponent ν of eq. (18) is, as smaller the incidence of great residues is on the estimate of θ; that is ν has to be enough small to give robust estimators or, in other words, to give an estimator poorly perturbed by the outliers. The functions are:

$$\rho(r) = \frac{|r|^\nu}{\nu}; \quad \psi(r) = \text{sgn}(r)|r|^{\nu-1}; \quad w(r) = |r|^{\nu-2}. \tag{22}$$

The investigations about the selection of an optimal ν have indicated that 1.2 is a suitable value [13][18][35].

3.2. Huber's function

The aim of this estimator is to find the simplest imaginable function that is consistent with the conditions of robustness. Huber introduced it [13][12][21] to give the min-max solution to eq. (16) for normal distributions of data affected by noise, under the hypothesis of known scale parameter $\hat{\sigma}$ and then extended its use to general distributions. He started from the maximum likelihood estimator (MLE) and reduced its sensitivity to the outliers. For a data sample with distribution density f, the MLE maximizes:

$$\sum \log f(x-\theta). \tag{23}$$

It can be defined by considering that, if the sample distribution is unknown, the most reasonable assumption is to suppose the symmetry:

$$f(x-\theta) \approx f(\theta-x). \tag{24}$$

The Taylor's expansion in proximity of the centre of the symmetric distribution is:

$$f(x) \approx f(\theta) - \left(x-\theta\right)^2 / 2. \tag{25}$$

In the proximity of the centre can be written that:

$$\log f(x-\theta) \approx \frac{(x-\theta)^2}{2}, \tag{26}$$

plus a constant. Therefore the parabola:

$$\rho(x-\theta) = \frac{(x-\theta)^2}{2}, \tag{27}$$

is the optimal choice in the proximities of the centre. This notwithstanding eq. (27) coincides to the least squares, which are not robust. A possibility is to limit the influence function if the residue r exceeds a certain value c, called *tuning parameter* [25]. The resulting functions are:

$$\rho(r) = \begin{cases} \dfrac{r^2}{2} & \text{if } |r| \le c \\ c\left(|r| - \dfrac{c}{2}\right) & \text{if } |r| > c \end{cases}; \quad \psi(r) = \begin{cases} r & \text{if } |r| \le c \\ c\,\mathrm{sgn}(r) & \text{if } |r| > c \end{cases}; \quad w(r) = \begin{cases} 1 & \text{if } |r| \le c \\ \dfrac{c}{|r|} & \text{if } |r| > c \end{cases}. \tag{28}$$

This ρ function is so satisfactory that has been recommended for any situation. However depending on the value of the tuning parameter, different estimators can be obtained. Since the estimator has been developed originally for normal distributions, the optimal tuning parameter c is calculated in order to have the 95% of asymptotic efficiency with respect to a normal distribution and results $c = 1.3450\,\hat{\sigma}$, where $\hat{\sigma}$ is the scale parameter. This value and the calculation of the tuning parameters of the other functions can be found by starting from the score function of eq. (8) as shown in [23][26].

3.3. Modified Huber's function

Even if Huber's estimator has outstanding performances, it could cause calculation problems related to lack of stability of the gradient values (as noted in [18]) due to the discontinuity of the second derivative of the ρ function. In fact:

$$\frac{d^2\rho(r)}{dr^2}=\begin{cases}1 \text{ if } |r|\le c\\ 0 \text{ if } |r|>c\end{cases}. \tag{29}$$

The modification proposed in [18] is:

$$\rho(r)=\begin{cases}c^2\left(1-\cos\left(\frac{r}{c}\right)\right) \text{if } \frac{|r|}{c}\le\frac{\pi}{2}\\ c|r|+c^2\left(1-\frac{\pi}{2}\right) \text{if } \frac{|r|}{c}>\frac{\pi}{2}\end{cases};\quad \psi(r)=\begin{cases}c\sin\left(\frac{r}{c}\right) \text{if } \frac{|r|}{c}\le\frac{\pi}{2}\\ c\,\text{sgn}(r) \text{ if } \frac{|r|}{c}>\frac{\pi}{2}\end{cases};\quad w(r)=\begin{cases}\frac{c}{r}\sin\left(\frac{r}{c}\right) \text{if } \frac{|r|}{c}\le\frac{\pi}{2}\\ \frac{c}{|r|} \quad \text{if } \frac{|r|}{c}>\frac{\pi}{2}\end{cases}. \tag{30}$$

The 95% of asymptotic efficiency with respect to a normal distribution is obtained in this case results $c=1.2107\hat{\sigma}$, where $\hat{\sigma}$ is the scale parameter.

Anyhow the modified Huber's function in this form is not suitable in case of complex data and residuals, since the trigonometric functions in eq. (30) have complex arguments and this determines the weights $w(r)$ to be complex and $\rho(r)$ is not compliant with the conditions of point 2 of paragraph 2.3. For the mechanical systems presented in the paper we use a "corrected" modified Huber's function that avoids the problem of complex weights:

$$\rho(r)=\begin{cases}c^2\left(1-\cos\left(\frac{|r|}{c}\right)\right) \text{if } \frac{|r|}{c}\le\frac{\pi}{2}\\ c|r|+c^2\left(1-\frac{\pi}{2}\right) \text{if } \frac{|r|}{c}>\frac{\pi}{2}\end{cases};\quad \psi(r)=\begin{cases}c\,\text{sgn}(r)\sin\left(\frac{|r|}{c}\right) \text{if } \frac{|r|}{c}\le\frac{\pi}{2}\\ c\,\text{sgn}(r) \quad \text{if } \frac{|r|}{c}>\frac{\pi}{2}\end{cases};$$
$$w(r)=\begin{cases}\frac{c}{|r|}\sin\left(\frac{|r|}{c}\right) \text{if } \frac{|r|}{c}\le\frac{\pi}{2}\\ \frac{c}{|r|} \quad \text{if } \frac{|r|}{c}>\frac{\pi}{2}\end{cases}. \tag{31}$$

3.4. Fair function

It has been already pointed out that also the scale parameter $\hat{\sigma}$ should be estimated and that in many cases, like in Huber's estimator definition, an "average" scale factor is used. The Fair function has been defined in [18] with the aim to have low sensitivity to the scale factor, so that the estimator has low sensitivity to the tuning parameter. It is defined as:

$$\rho(r)=c^2\left(\frac{|r|}{c}-\log\left(1+\frac{|r|}{c}\right)\right);\quad \psi(r)=\frac{r}{1+\frac{|r|}{c}};\quad w(r)=\frac{1}{1+\frac{|r|}{c}}. \tag{32}$$

This ρ function is robust as well as yields to a unique solution (since it has everywhere defined continuous derivatives up to the third order). The 95% asymptotic efficiency with respect to a normal distribution is obtained with a tuning parameter $c = 1.3998\,\hat{\sigma}$.

3.5. Cauchy's function

The name of this function derives from the fact that it is optimal for data having a Cauchy's distribution. Anyhow it does not guarantee a unique minimum. The decreasing first order derivative can yield to erroneous solution that cannot be observed and the influence of large errors only decreases linearly with their size. The definition is:

$$\rho(r) = \frac{c^2}{2}\log\left(1+\left(\frac{|r|}{c}\right)^2\right); \quad \psi(r) = \frac{r}{1+\left(\frac{|r|}{c}\right)^2}; \quad w(r) = \frac{1}{1+\left(\frac{|r|}{c}\right)^2}. \tag{33}$$

The 95% of asymptotic efficiency on a normal distribution is obtained for $c = 2.3849\,\hat{\sigma}$, where $\hat{\sigma}$ is the scale parameter. By studying the relationships between tuning constant and efficiency, this function appears to be the best among the functions proposed in literature that does not comply with the satisfaction of the property 3 of paragraph 2.3 (the best among the worse ones).

3.6. Welsch's function

This function has been introduced in order to further reduce the effect of large errors, but has not a unique minimum. It is defined as:

$$\rho(r) = \frac{c^2}{2}\left(1-\exp\left(-\left(\frac{|r|}{c}\right)^2\right)\right); \quad \psi(r) = r\exp\left(-\left(\frac{|r|}{c}\right)^2\right); \quad w(r) = \exp\left(-\left(\frac{|r|}{c}\right)^2\right). \tag{34}$$

The tuning parameter for 95% asymptotic efficiency with respect to a normal distribution is $c = 2.9846\,\hat{\sigma}$.

3.7. Tukey's function

This function has been proposed by Tukey and is also called *biweight function* [36]. Its aim is to suppress the outliers as shown by the definition of $w(r)$:

$$\rho(r)=\begin{cases}\dfrac{c^2}{6}\left(1-\left(1-\left(\dfrac{r}{c}\right)^2\right)^3\right) & \text{if } |r|\le c\\ \dfrac{c^2}{6} & \text{if } |r|>c\end{cases};\quad \psi(r)=\begin{cases} r\left(1-\left(\dfrac{r}{c}\right)^2\right)^2 & \text{if } |r|\le c\\ 0 & \text{if } |r|>c\end{cases};$$
$$w(r)=\begin{cases}\left(1-\left(\dfrac{r}{c}\right)^2\right)^2 & \text{if } |r|\le c\\ 0 & \text{if } |r|>c\end{cases}. \tag{35}$$

It is misleading due to the lack of a unique minimum and the 95% asymptotic efficiency with respect to a normal distribution is obtained with a tuning parameter $c=4.6851\hat{\sigma}$.

3.8. German-McClure's function

The aim of this function is reduce the effect of large errors without introducing a tuning parameter, but it has not a unique minimum. It is defined as:

$$\rho(r)=\frac{r^2/2}{1+r^2};\quad \psi(r)=\frac{r}{\left(1+r^2\right)^2};\quad w(r)=\frac{1}{\left(1+r^2\right)^2}. \tag{36}$$

Due to the not compliance with condition 3) of paragraph 2.3, the performances of this function will not be good *a priori*. It is included only for completeness, since it has been used in some applications related to image recognition [37][38].

4. Numerical application to a SISO system

The simplest mechanical dynamical system is represented by a single degree of freedom (d.o.f.) system shown in figure 1. The parameters of the system are the mass m, the damping c and the stiffness k. The system d.o.f. is described by means of the system displacement $x(t)$. An external harmonic excitation F causes the forced vibration of the system.

The equation of motion of the considered system is simply:

$$m\,\ddot{x}+c\,\dot{x}+k\,x=F(t). \tag{37}$$

By considering that the forcing system is harmonic:

$$F(t)=\overline{F}_0\,\mathrm{e}^{\mathrm{i}\Omega t}=F_0\,\mathrm{e}^{\mathrm{i}\varphi}\mathrm{e}^{\mathrm{i}\Omega t}, \tag{38}$$

also the solution of eq. (38) has to be harmonic:

$$x(t)=\overline{X}\,\mathrm{e}^{\mathrm{i}\Omega t}=X\,\mathrm{e}^{\mathrm{i}\phi}\mathrm{e}^{\mathrm{i}\Omega t}. \tag{39}$$

Replacing eq. (39) in eq. (37), the well-known steady-state solution can be obtained as:

$$\left(-\Omega^2 m+\mathrm{i}\Omega c+k\right)\overline{X}=F_0\,\mathrm{e}^{\mathrm{i}\varphi} \tag{40}$$

and finally:

$$\bar{X} = \frac{F_0 \, \mathrm{e}^{\mathrm{i}\varphi}}{-\Omega^2 m + \mathrm{i}\Omega c + k} = H(\Omega) F_0 \, \mathrm{e}^{\mathrm{i}\varphi}. \tag{41}$$

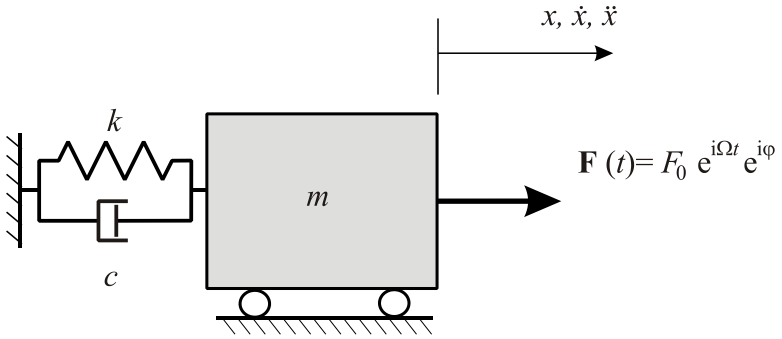

Figure 1. Single degree of freedom system.

Note that the solutions given by eq. (41) as function of the frequency Ω of the forcing system are complex. Since eq. (41) gives the response of the SISO system to the external force, in a deterministic context, in which the mass, the damping and the stiffness of the system are exactly know, the knowledge of the system displacement amplitude and phase at a single frequency only allows to determine the amplitude and the phase of the forcing system.

Now, let us consider a stochastic environment, in which the displacements of the mass are measured for a given set of frequencies, once the system has reached the steady-state. Measures could be corrupted by noise, biases, systematic errors and so on.

For each measure $\bar{x}_i$ of the displacement (in amplitude and phase) at the frequency Ω_i, it is possible to write the equation:

$$\bar{x}_i = h_i(\Omega_i) F_0 \, \mathrm{e}^{\mathrm{i}\varphi} \quad i = 1, ..., n. \tag{42}$$

The system of all the n equations (42) has the only unknown represented by the force (in amplitude and phase) and therefore it is over-determined. Normally this system is solved by means of least squares. If the measures are perturbed by white-noise, usually quite accurate results are obtained in any case. Conversely if a systematic error affect the measures, L_2 estimate does not produce any more accurate results.

This fact can be shown simply by means of a simulated case. Let us consider a system in which m = 10 kg, k = 1000 N/m and c = 60 Ns/m (the damping is equal to 30% of the critical damping). If the forcing system has the amplitude F_0 equal to 100 N and the phase φ of 45°, the nominal, i.e. non-corrupted, system response calculated in the frequency range from 0 to 30 rad/s with a step of 0.1 rad/s is shown as Bode plot in figure 2. Now a systematic error is applied on the system response.

Systematic errors can be chosen in infinite ways. To test the robustness, the criterion selected is to use a fixed step in the frequency. This reproduces the presence of electromagnetic disturbance on the experimental signals, at a certain frequency and its multiples. A constant magnification of amplitude and phase shift is used to increase the systematic characteristic of the error:

- Each value, in the ordered vector of the measures, starting from 0 rad/s with a step of 1 rad/s, has the amplitude increased of 500% and the phase rotated of +45°. This corresponds to the $\bar{x}_k$ in which $k = 1 + 10j,\ j \in \mathbb{N}_0$.
- Each value starting from 0.4 rad/s with a step of 1 rad/s has the amplitude reduced of 50% and the phase rotated of −20°. This corresponds to the $\bar{x}_k$ in which $k = 5 + 10j,\ j \in \mathbb{N}_0$.

The resulting corrupted system response is shown in figure 3.

If L_2 estimate is used, the amplitude and the phase of the external force results:

$$F_0 = 125.2641[\mathrm{N}], \quad \varphi = 60.8731°. \tag{43}$$

The error is considerable on both amplitude and phase.

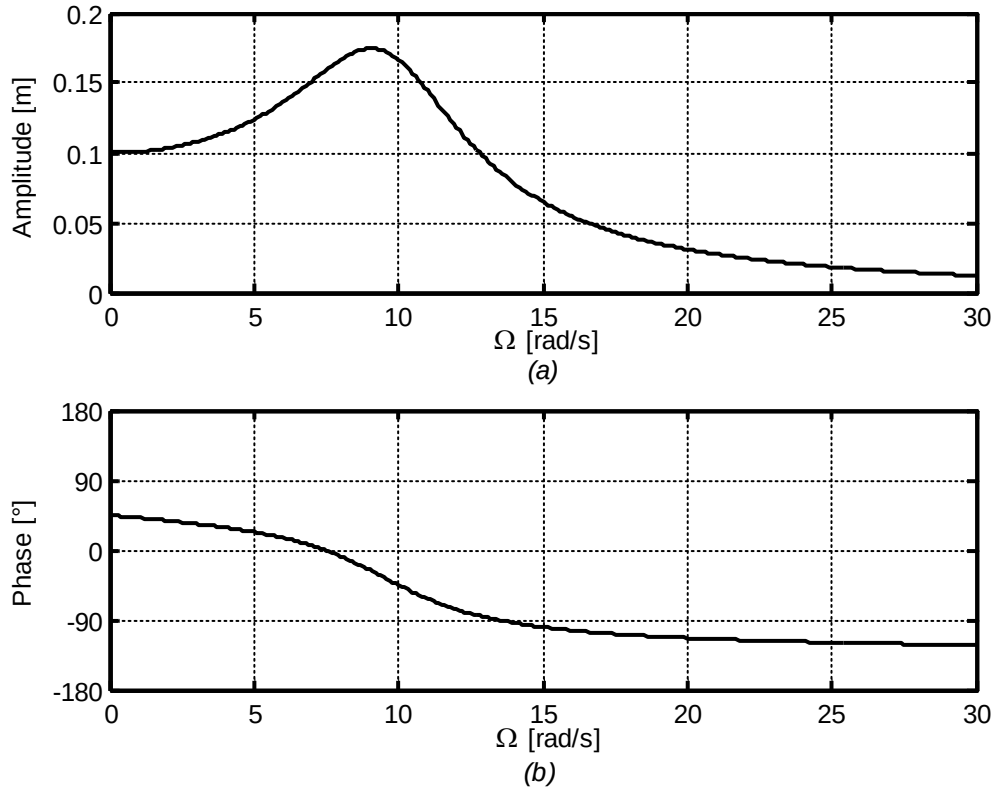

Figure 2. Nominal response of the system: (a) amplitude, (b) phase.

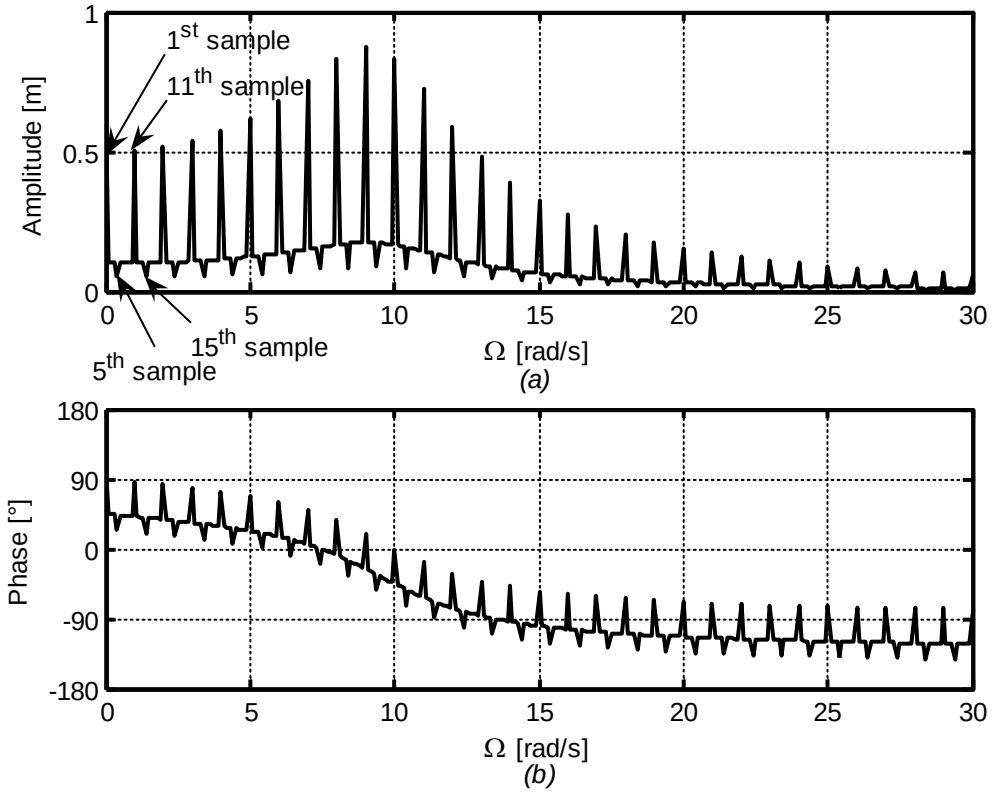

Figure 3. Corrupted response of the system: (a) amplitude, (b) phase.

4.1. Calculation of the M-estimate by the implementation of the ILRS algorithm

The excitation, the external force, is now estimated using M-estimate. The objective function to be minimized is:

$$\sum_{i=1}^{n} \rho(r_i) = \sum_{i=1}^{n} \rho\left(\frac{\overline{x}_i - h_i(\Omega_i) F_0 \, \mathrm{e}^{\mathrm{i}\varphi}}{\hat{\sigma}} \right), \tag{44}$$

where $\hat{\sigma}$ is the scale estimate calculated using eq. (17).

The solution of the minimization of eq. (44) follows the theoretical arguments presented in paragraph 2.2. Let ψ be the first derivative of ρ with respect to the unknown force:

$$\psi(r) = \frac{\partial \rho(r)}{\partial F_0 \, \mathrm{e}^{\mathrm{i}\varphi}} \ . \tag{45}$$

The minimization of eq. (44) is obtained by differentiating the objective function with respect to the unknown force and setting the partial derivatives to 0:

$$\sum_{i=1}^{n} \psi\left(\frac{\overline{x}_i - h_i(\Omega_i) F_0 \, \mathrm{e}^{\mathrm{i}\varphi}}{\hat{\sigma}} \right) h_i(\Omega_i) = 0 \, . \tag{46}$$

If the weight function $w(r)$ is defined as:

$$w(r) = \frac{\psi(r)}{r} \ , \tag{47}$$

then the equations in (46) can be written as:

$$\sum_{i=1}^{n} w\left(\frac{\overline{x}_i - h_i(\Omega_i)F_0\,\mathrm{e}^{\mathrm{i}\varphi}}{\hat{\sigma}}\right)=0\,, \tag{48}$$

which is equivalent to minimize the least squares problem:

$$\sum_{i=1}^{n} w_i^2 r_i^2\,. \tag{49}$$

Anyhow, the weights $w(r)$ depend upon the residuals r, the residuals depend upon the estimated excitation $\overline{F}_0$ and the estimated excitation depends upon the weights $w(r)$. To solve this loop, an iterative solution, called IRLS, is used. The algorithm is the following:

1. The initial estimate of the force amplitude and phase is selected using the results of least squares calculation, thus is that of eq. (43).
2. At each iteration t, the residuals $r_i^{(t)}$ and the associated weights $w_i^{(t)}$ are calculated from the previous iteration.
3. The new weighted least squares estimate is:

$$\left(F_0\,\mathrm{e}^{\mathrm{i}\varphi}\right)^{(t+1)} = \left[\mathbf{H}^{\mathrm{T}}\mathbf{W}^{(t)}\mathbf{H}\right]^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{W}^{(t)}\mathbf{X}\,, \tag{50}$$

where:

$$\mathbf{W}^{(t)} = \mathrm{diag}\left[w_i^{(t)}\right]. \tag{51}$$

Steps 2 and 3 are repeated until the estimated coefficients converge.

The main advantage of the ILRS algorithm is its simplicity, practically standard numerical methods to calculate weighted least squares are iteratively used.

The first drawback is that, once a ρ function is chosen, the weight function w descends automatically. Therefore ρ function should be "good" (see paragraph 2.3). The second is the iteration stop condition. In this case, being the estimated coefficient complex, the algorithm is repeated until the maximum normalized difference between the real and the imaginary parts of the coefficient value in the present iteration and those of the previous iteration is less than 1e–4, i.e:

$$\max\left(\begin{array}{l}\left|\left(\mathrm{Re}(\overline{F}_0^{(t-1)}) - \mathrm{Re}(\overline{F}_0^{(t)})\right)\Big/\mathrm{Re}(\overline{F}_0^{(t-1)})\right| \\ \left|\left(\mathrm{Im}(\overline{F}_0^{(t-1)}) - \mathrm{Im}(\overline{F}_0^{(t)})\right)\Big/\mathrm{Im}(\overline{F}_0^{(t-1)})\right|\end{array}\right) < 1\mathrm{e}-4\,. \tag{52}$$

If this convergence is not reached, the algorithm stops after 100 iterations. Similar stopping rule is also used in [39].

The results with different types of M-estimators are reported in table 1. As predicted by the theoretical analysis (section 3), some of the M-estimators (Cauchy, German-McClure, Welsch and Tukey) give bad results. In all of these four cases, the algorithm has stopped having reached the maximum number of iteration without satisfying the convergence condition. Actually the trends of the estimated amplitude and phase of the force presented an oscillating behaviour as function of the

iteration number, with average values close to the correct ones. This fact will be discussed in detail in the next section.

The algorithm, using the other M-estimators, does not presented numerical oscillations, stops in few iterations and gives remarkable good results. In two cases the results are definitely exact even with the corrupted data.

Table 1. Results of the excitation estimate.

Estimator	F_0 [N]	φ
L_1 (absolute value)	100	45°
L_1-L_2 (absolute value and least squares)	100	45.0001°
L_p, ν=1.2	100	45.0002°
Fair	99.9999	45.0002°
Huber	100	45°
Cauchy	123.4596	58.548°
German-McClure	135.5781	62.5372°
Welsch	140.8528	64.1209°
Tukey	142.6836	64.642°

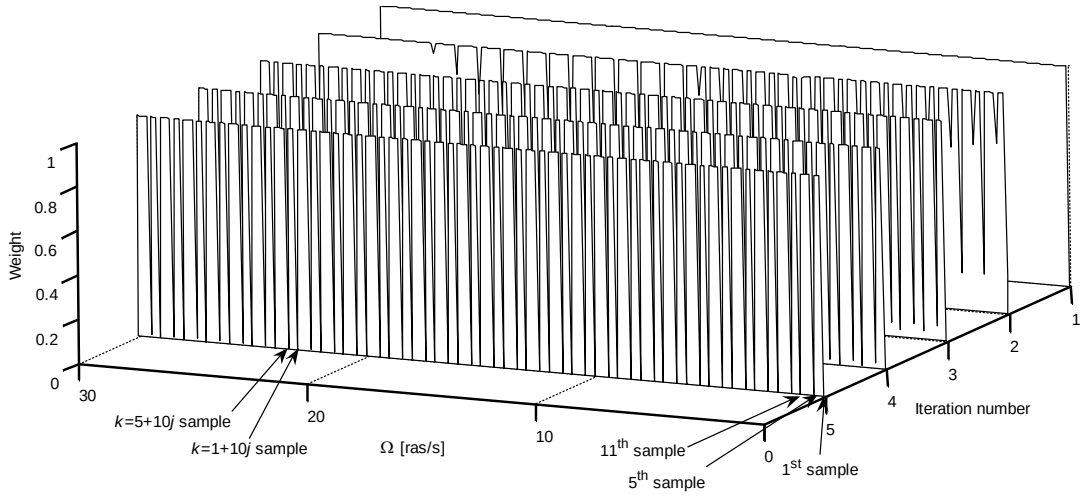

Figure 4. Weights attributed per iteration for the Huber's estimator.

The explanation of these surprisingly results can be given by considering the weights w_i that the algorithm attributes to each measure $\overline{x}_i$. For example, figure 4 shows the weights as a function of the measure order number and of the iteration for the Huber's estimator: after few iterations the weights given to the corrupted measures become near 0 and the algorithm practically discard them. In this case, the algorithm stops in 5 steps.

5. Numerical application to a MIMO system

In this example, not only the performances of the different types of M-estimators are now evaluated for a MIMO system, but also the convergence to a stable solution for the IRLS is analyzed depending on the type of the M-estimator. The considered MIMO system is a simple linear mechanical system with 2 d.o.f.s, shown in figure 5. The physical parameters of the system are known, i.e. the model of the system is known and reliable, and are grouped in the mass $\mathbf{M}$, in the damping $\mathbf{C}$ and in the stiffness $\mathbf{K}$ matrices. The system d.o.f.s are described by means of the vector of the mass displacements $\mathbf{x}$. The unknown external force system $\mathbf{F}$, acting on the masses, causes the forced vibration of the system.

The equations of motion of the considered linear system are simply:

$$\mathbf{M}\ddot{\mathbf{x}}+\mathbf{C}\dot{\mathbf{x}}+\mathbf{K}\mathbf{x}=\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}\begin{Bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{Bmatrix}+\begin{bmatrix} c_1+c_2 & -c_2 \\ -c_2 & c_2+c_3 \end{bmatrix}\begin{Bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{Bmatrix}+\begin{bmatrix} k_1+k_2 & -k_2 \\ -k_2 & k_2+k_3 \end{bmatrix}\begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix}= \\ =\begin{Bmatrix} F_1(t) \\ F_2(t) \end{Bmatrix}=\mathbf{F}(t). \tag{53}$$

By considering a harmonic forcing system:

$$\mathbf{F}(t)=\begin{Bmatrix} \overline{F}_{01} \\ \overline{F}_{02} \end{Bmatrix}\mathrm{e}^{\mathrm{i}\Omega t}=\begin{Bmatrix} F_{01}\,\mathrm{e}^{\mathrm{i}\varphi_1} \\ F_{02}\,\mathrm{e}^{\mathrm{i}\varphi_2} \end{Bmatrix}\mathrm{e}^{\mathrm{i}\Omega t}, \tag{54}$$

also the steady-solution of (53) has to be harmonic:

$$\mathbf{x}(t)=\mathbf{X}\,\mathrm{e}^{\mathrm{i}\Omega t}=\begin{Bmatrix} \overline{X}_1 \\ \overline{X}_2 \end{Bmatrix}\mathrm{e}^{\mathrm{i}\Omega t}=\begin{Bmatrix} X_1\,\mathrm{e}^{\mathrm{i}\phi_1} \\ X_2\,\mathrm{e}^{\mathrm{i}\phi_2} \end{Bmatrix}\mathrm{e}^{\mathrm{i}\Omega t}. \tag{55}$$

Replacing eq. (55) in eq. (53), the steady-state solution can be obtained as:

$$\left(-\Omega^2\mathbf{M}+\mathrm{i}\,\Omega\mathbf{C}+\mathbf{K}\right)\mathbf{X}=\mathbf{F}_0\ , \tag{56}$$

and finally:

$$\mathbf{X}=\left(-\Omega^2\mathbf{M}+\mathrm{i}\,\Omega\mathbf{C}+\mathbf{K}\right)^{-1}\mathbf{F}_0=\mathbf{H}(\Omega)\,\mathbf{F}_0\,. \tag{57}$$

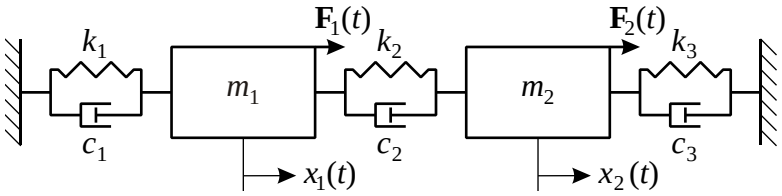

Figure 5. System with 2 degrees of freedom.

Note that the solution given by eq. (57) as function of the frequency Ω of the forcing system is complex, i.e. $\overline{X}_{1,2}\in\mathbb{C}$. Similarly to previous section 4, a stochastic environment is considered. The displacements of the masses are measured for a given set of excitations/frequencies, once the system has reached the steady-state and we want to estimate (identify) the forcing system. The measures, repeated for *n* different excitations/frequencies, could be corrupted by noise, biases, systematic errors and so on.

For each measure $\mathbf{x}_i = \left\{ \overline{x}_{1i}\ \overline{x}_{2i} \right\}^{\mathrm{T}}$ of the displacements (in amplitude and phase) at the frequency Ω_i, it is possible to write the equation:

$$\mathbf{x}_i = \mathbf{h}_i(\Omega_i)\mathbf{F}_0 \ , \tag{58}$$

If a systematic error affects the measures, L_2 estimate does not produce any more accurate results, as analytically proven in the previous sections and as exemplified by means of this simulated case.

Let consider a system like that of figure 5 and eq. (53) in which, for simplicity, $m_1 = m_2 = 10\text{kg}$, $k_1 = k_2 = k_3 = 1000\text{N/m}$ and $c_1 = c_2 = c_3 = 60\text{Ns/m}$. If the forcing system has amplitudes equal to $F_{01} = 100\text{N}$ and $F_{02} = 50\text{N}$ and phases $\varphi_1 = 45°$ and $\varphi_2 = -60°$, the nominal, i.e. non-corrupted, system response calculated in the frequency range from 0 to 30 rad/s with a step of 0.1 rad/s is shown as Bode plot in figure 6. Now a systematic error is applied to the system response: every value of the system response in the range starting from 0 rad/s with a step of 0.5 rad/s has the amplitude increased of 500% and the phase rotated of +45°, i.e. for $\overline{x}_{1,2k} : k = 1+10\,j,\ j \in \mathbb{N}_0$.

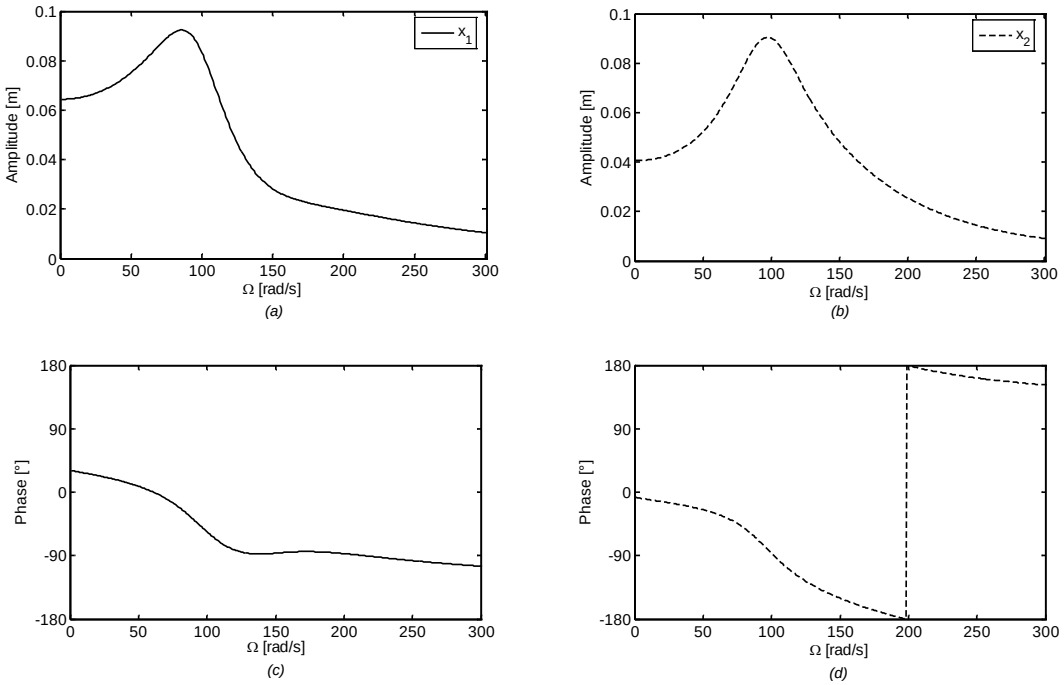

Figure 6. Nominal response of the system: (a) amplitude of x_1, (b) amplitude of x_2, (c) phase of x_1, (d) phase of x_2.

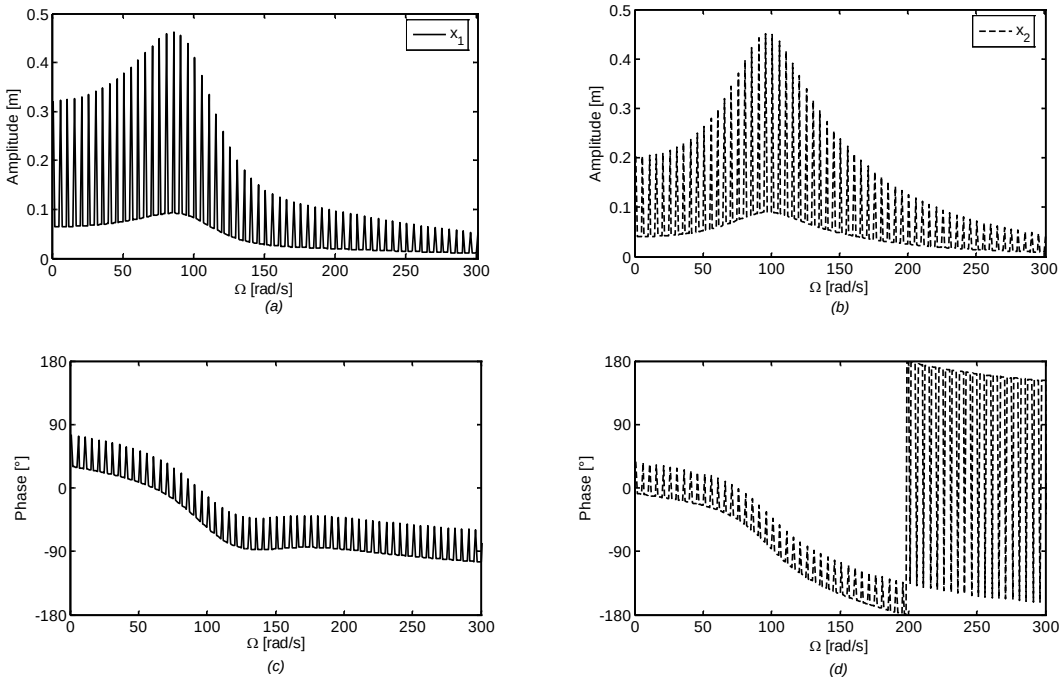

Figure 7. Corrupted response of the system: (a) amplitude of x_1, (b) amplitude of x_2, (c) phase of x_1, (d) phase of x_2.

The resulting corrupted system response is shown in Figure 7.

If L_2 estimate is used to identify the amplitude and the phase of the external forces using the corrupted measures, the objective function is:

$$\sum_{i=1}^{n}\left(r_i\right)^2=\sum_{i=1}^{n}\left(\sum_{1,2}\left(\mathbf{x}_i-\mathbf{h}_i(\Omega_i)\mathbf{F}_0\right)\right)^2 \tag{59}$$

and it results:

$$\begin{aligned} F_{01}&=167.1089\text{N},\ \varphi_1=70.2677^\circ \\ F_{02}&=70.2677\text{N},\ \varphi_2=-34.7705^\circ \end{aligned}. \tag{60}$$

The errors are considerable on both the amplitude and the phase. If M-estimate is used, r_i is the residue of the *i*-th observation corresponding to the frequency Ω_i and the objective ρ function to be minimized is:

$$\sum_{i=1}^{n}\rho\left(r_i\right)=\sum_{i=1}^{n}\rho\left(\sum_{1,2}\left(\frac{\mathbf{x}_i-\mathbf{h}_i(\Omega_i)\mathbf{F}_0}{\hat{\sigma}}\right)\right). \tag{61}$$

The IRLS algorithm is practically the same, with the necessary adjusts to take into consideration the d.o.f.s of the system. Therefore eq. (50) becomes:

$$\hat{\mathbf{F}}_0^{(t+1)}=\left[\mathbf{A}^{\mathrm{T}}\mathbf{A}\right]^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}^{(t)}\mathbf{Y}. \tag{62}$$

where:

$$\mathbf{A}=\mathbf{W}^{(t)}\underline{\mathbf{H}}\,\mathbf{v}, \tag{63}$$

$$\mathbf{W}^{(t)} = \mathrm{diag}\left[w_i^{(t)} \right], \tag{64}$$

$$\underline{\mathbf{H}} = \mathrm{diag}\left[\mathbf{H}\left(\Omega_i\right) \right], \tag{65}$$

$$\mathbf{v} = \left[\cdots \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdots \right]^{\mathrm{T}}, \tag{66}$$

$$\mathbf{Y} = \left\{ \cdots \overline{x}_{1i}\ \overline{x}_{2i} \cdots \right\}^{\mathrm{T}}. \tag{67}$$

Matrix **W** is a diagonal matrix, the elements of which are the weights, calculated per each iteration, of the measures of the d.o.f.s, $\underline{\mathbf{H}}$ is a band matrix with the elements along the main and secondary diagonals represented by the transfer function **H** corresponding to the frequency at which the measures are acquired and vector **v** is a "localization vector" that indicate the d.o.f.s on which the excitations are acting. The stop condition is similar to that of previous paragraph 4.1:

$$\max \left(\begin{array}{l} \max\limits_{1,2} \left| \left(\mathrm{Re}(\overline{F}_{0i}^{(t-1)}) - \mathrm{Re}(\overline{F}_{0i}^{(t)}) \right) \Big/ \mathrm{Re}(\overline{F}_{0i}^{(t-1)}) \right| \\ \max\limits_{1,2} \left| \left(\mathrm{Im}(\overline{F}_{0i}^{(t-1)}) - \mathrm{Im}(\overline{F}_{0i}^{(t)}) \right) \Big/ \mathrm{Im}(\overline{F}_{0i}^{(t-1)}) \right| \end{array} \right) < 1\mathrm{e}-4. \tag{68}$$

The results obtained by the different types of M-estimators, described in section 3, are reported in table 2, while the calculated values during the iterations are shown in figure 8 to figure 17.

Table 2. Identification results

M-estimator Type	F_{01} [N]	φ_1 [°]	F_{02} [N]	φ_2 [°]	Iterations
L_1 (least absolute)	100.0001	45.0001	50	–59.9999	6
$L_1 - L_2$	100.0007	45.0007	50.0006	–59.9996	8
$L_p, \nu = 1.2$ (least powers)	100.0047	45.0038	50.0024	–59.9962	10
Fair	100.0005	45.0004	50.0003	–59.9997	11
Huber	100.0002	45.0002	50.0002	–59.9999	8
Modified Huber	100.0001	45.0001	50.0001	–60	8
Cauchy	136.8991	62.5361	65.8223	–37.9421	100
German-McClure	167.8775	69.749	83.7188	–40.6262	100
Welsch	174.7285	72.1534	87.9188	–26.2993	100
Tukey	169.3348	69.54	80.0968	–32.0012	100

The results of table 2 confirm what stated about the M-estimator functions that do not have unique solution. Cauchy's, German-McClure's, Welsch's and Tukey's estimators can be badly deemed, and in all the four cases the algorithm has stopped having reached the maximum number of iteration without satisfying the convergence condition. As anticipated Cauchy's estimator behaves better than German-McClure's, Welsch's and Tukey's one.

Actually the trends of the estimated amplitude and phase of the force presented an oscillating behaviour as function of the iteration number, with average values closer to the correct ones (dash-dot lines in figure 14 to figure 17) than those of L_2. Therefore they could be dealt less severely with a different stop condition of the IRLS algorithm, based for instance on the convergence of the average value.

The other M-estimators do not present numerical oscillations (figure 8 to figure 13), stop in few iterations and give good results, being very robust with respect to data corruption. As expected Huber's and modified Huber's have excellent results while the good performance of L_1 has to be carefully considered due to the drawback underlined in paragraph 3.1.1.

Similarly to the previous SISO example, the explanation of these surprisingly results, which are in any case based on the perfect knowledge of the system model, can be given by considering the weights w_i that the algorithm attributes to each measure $\overline{x}_i$ in each iteration. After few iterations, the weights given to the corrupted measures become 0 and the algorithm practically and

automatically discards them. This is evident if the weights are plotted as a function of the measure order number and of the iteration (see figure 18 for the Huber's estimator).

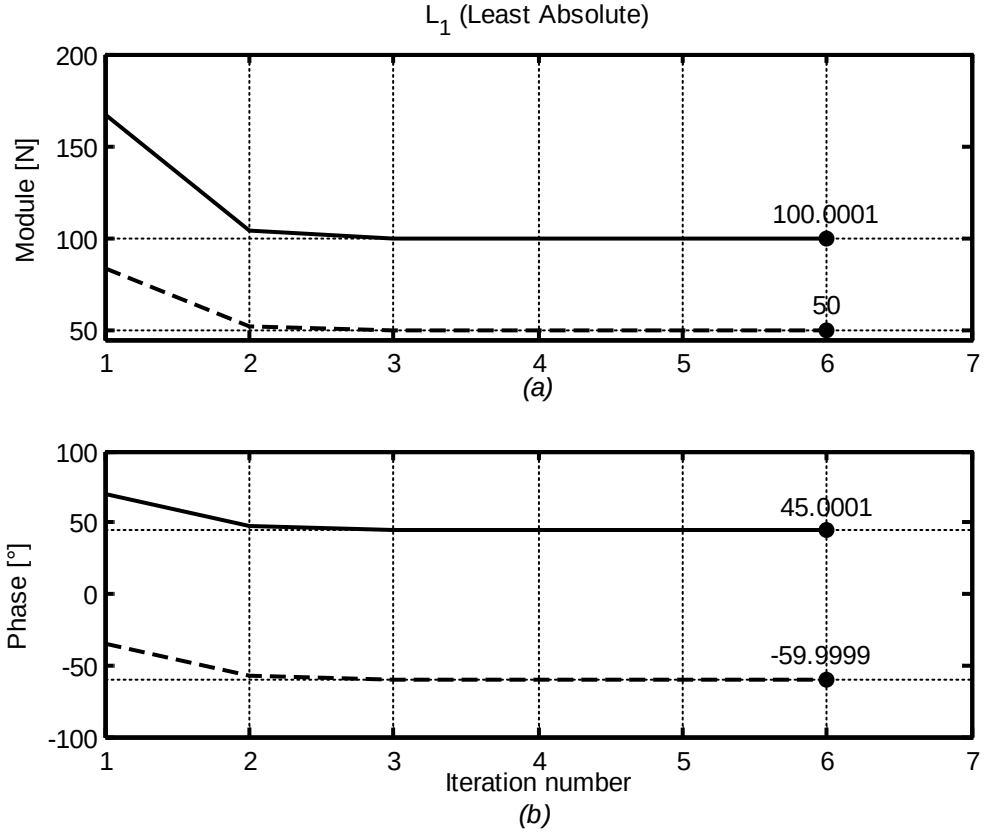

Figure 8. Estimated values vs. iteration for the L_1 estimate: (a) force amplitudes, (b) force phases.

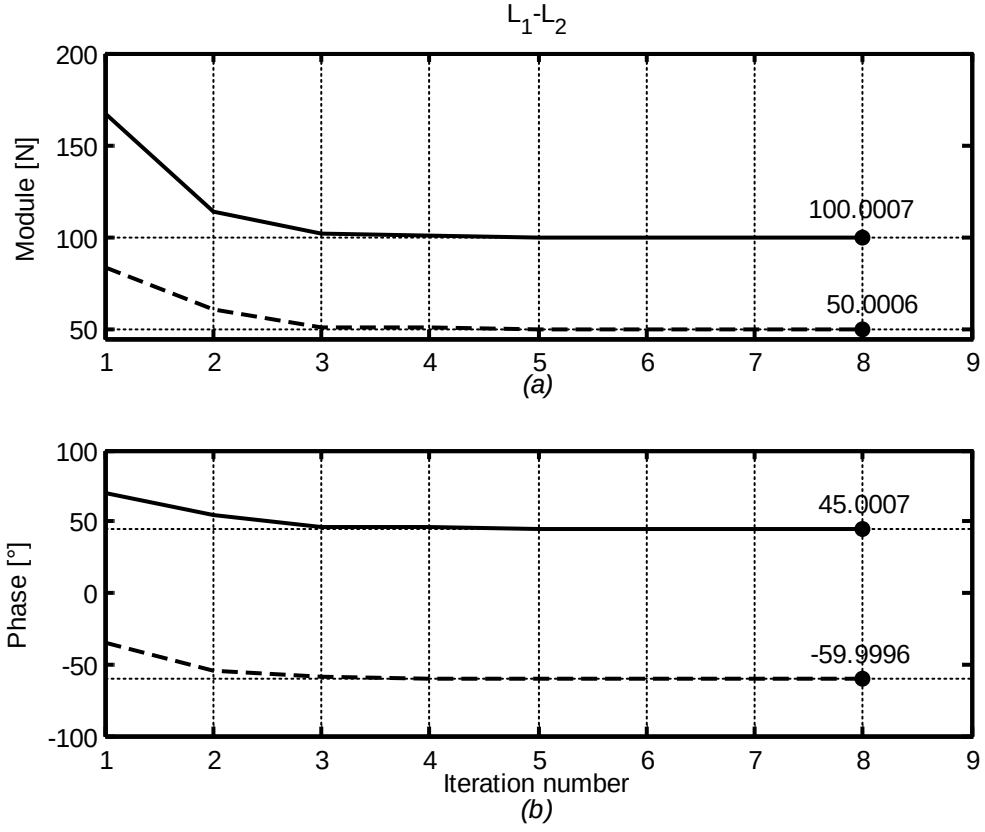

Figure 9. Estimated values vs. iteration for the L_1-L_2 estimate: (a) force amplitudes, (b) force phases.

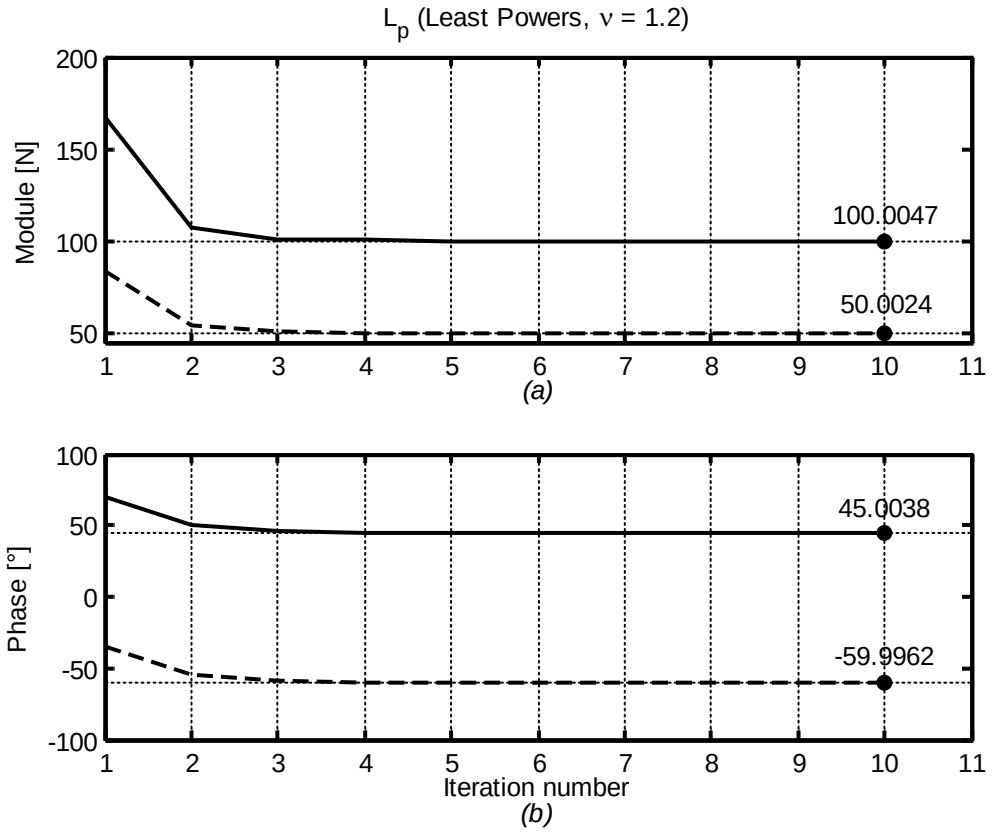

Figure 10. Estimated values vs. iteration for the L_p estimate: (a) force amplitudes, (b) force phases.

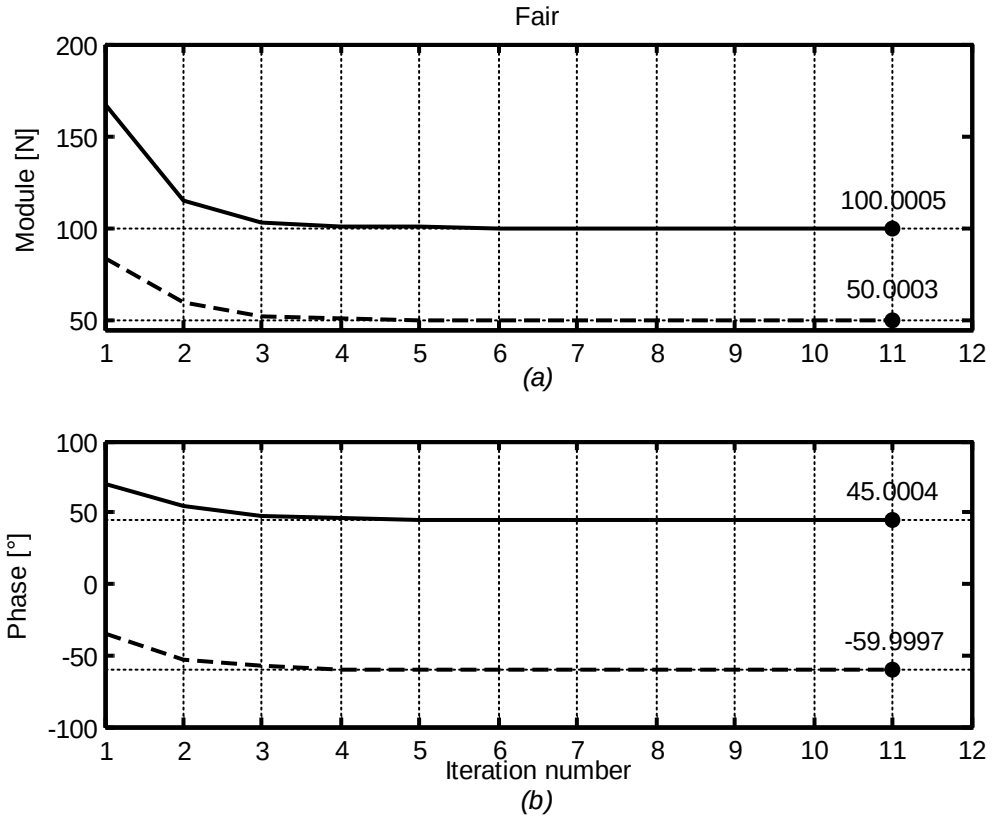

Figure 11. Estimated values vs. iteration for the Fair estimate: (a) force amplitudes, (b) force phases.

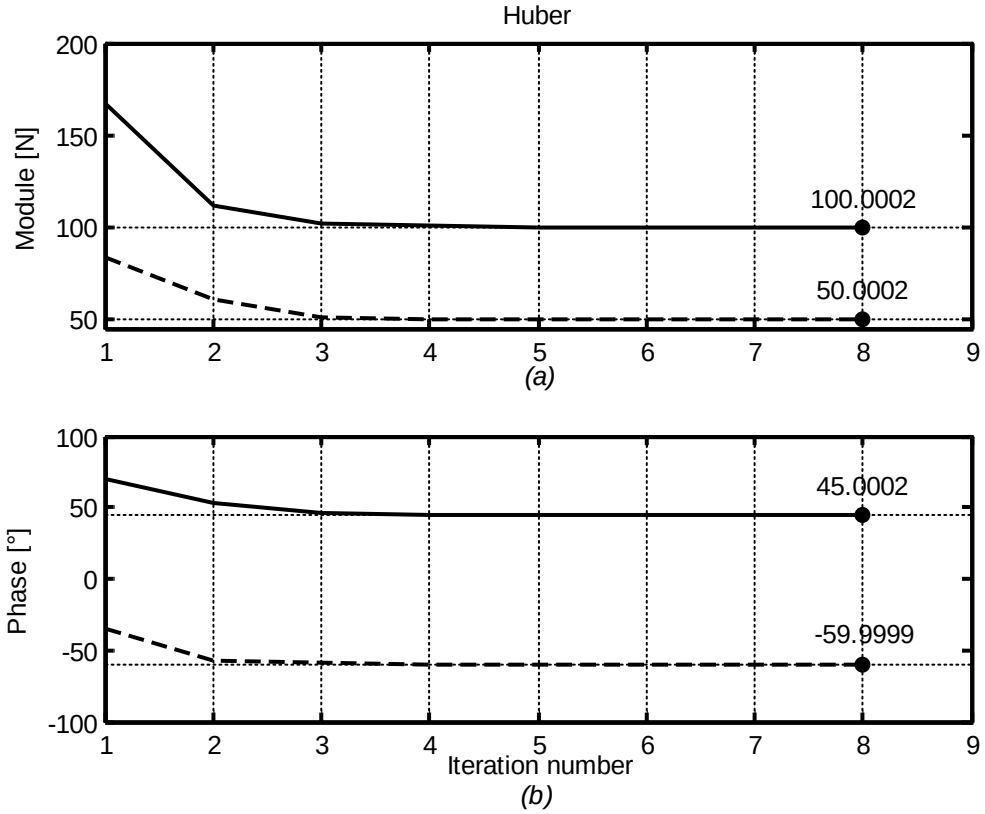

Figure 12. Estimated values vs. iteration for the Huber's estimate: (a) force amplitudes, (b) force phases.

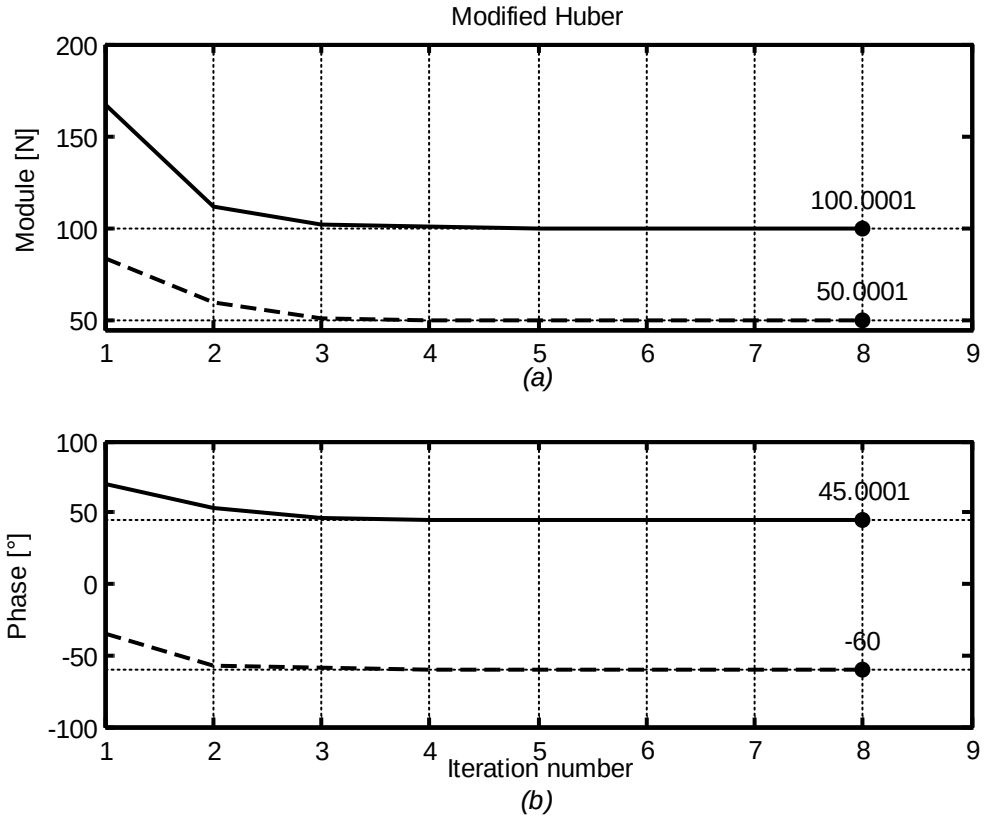

Figure 13. Estimated values vs. iteration for the Modified Huber's M-estimate: (a) force amplitudes, (b) force phases.

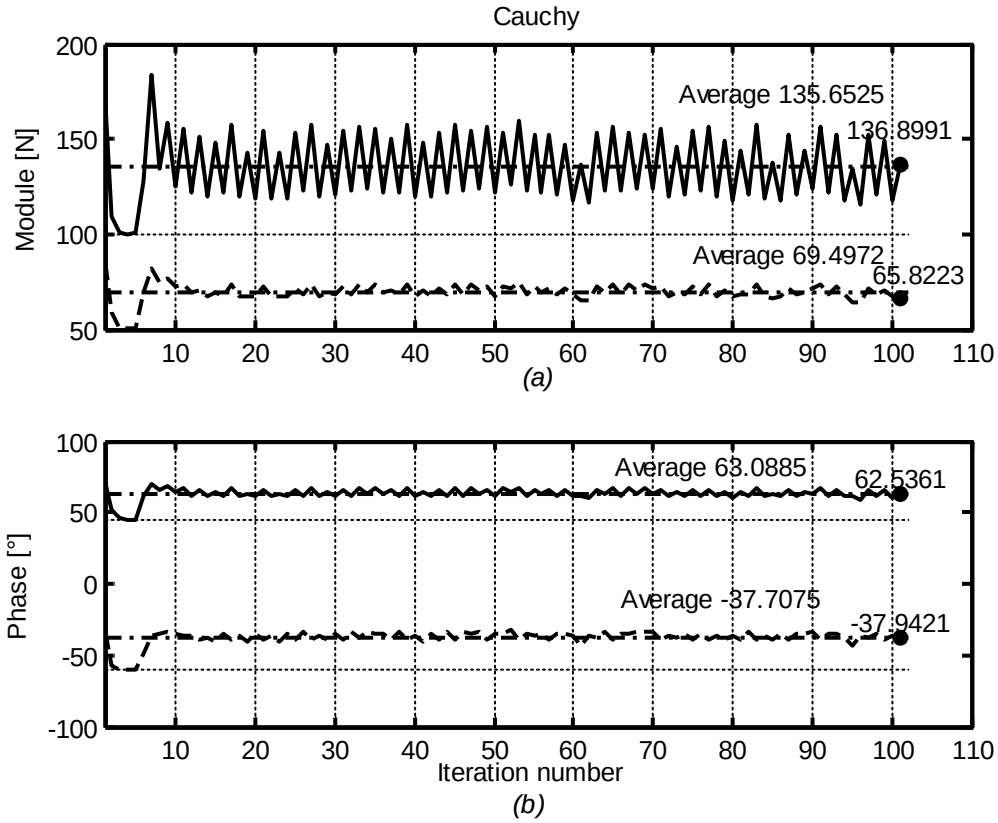

Figure 14. Estimated values vs. iteration for the Cauchy's estimate: (a) force amplitudes, (b) force phases.

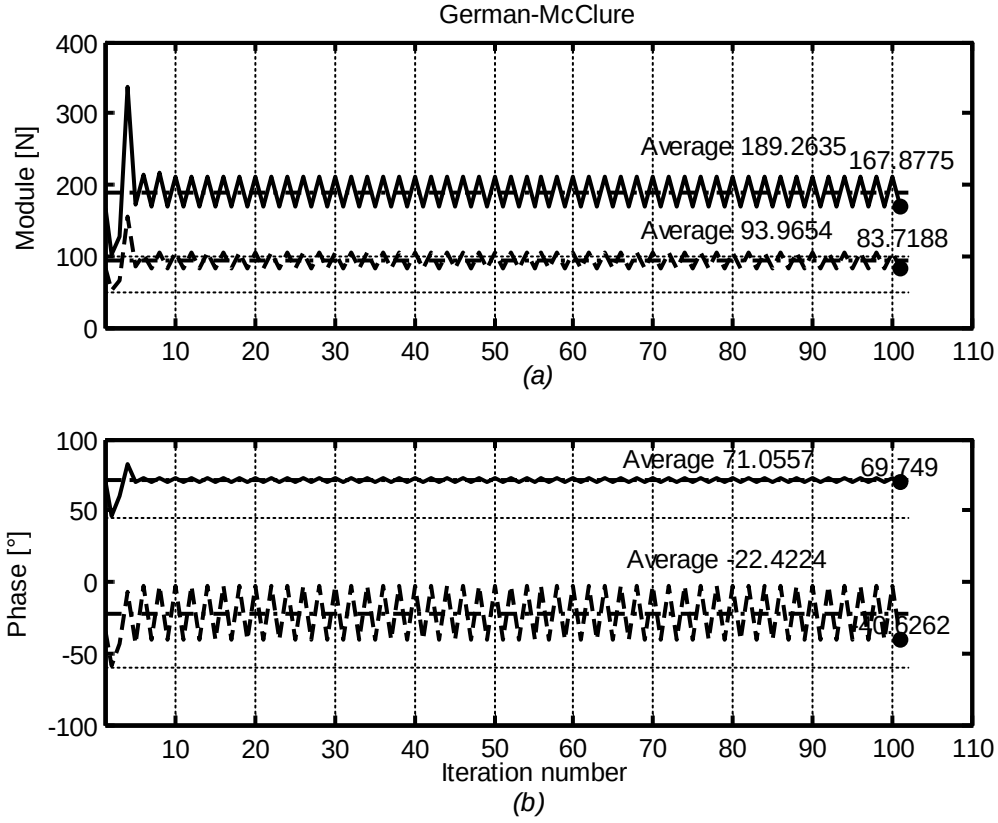

Figure 15. Estimated values vs. iteration for the German-McClure's estimate: (a) force amplitudes, (b) force phases.

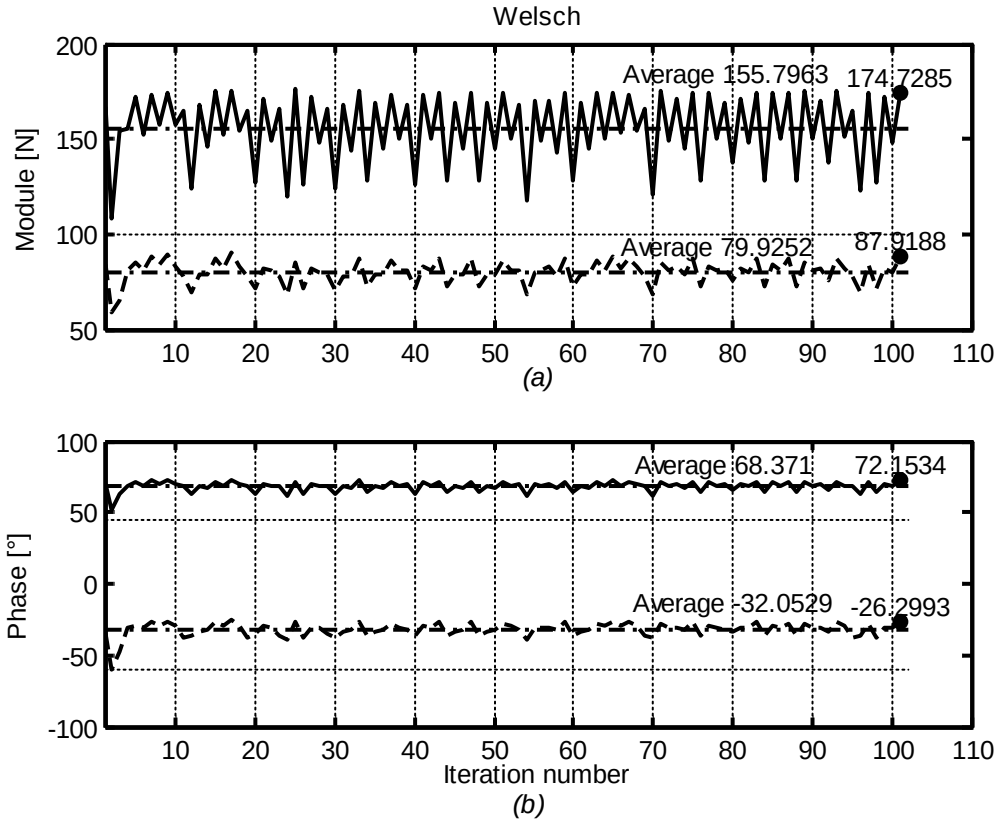

Figure 16. Estimated values vs. iteration for the Welsch's estimate: (a) force amplitudes, (b) force phases.

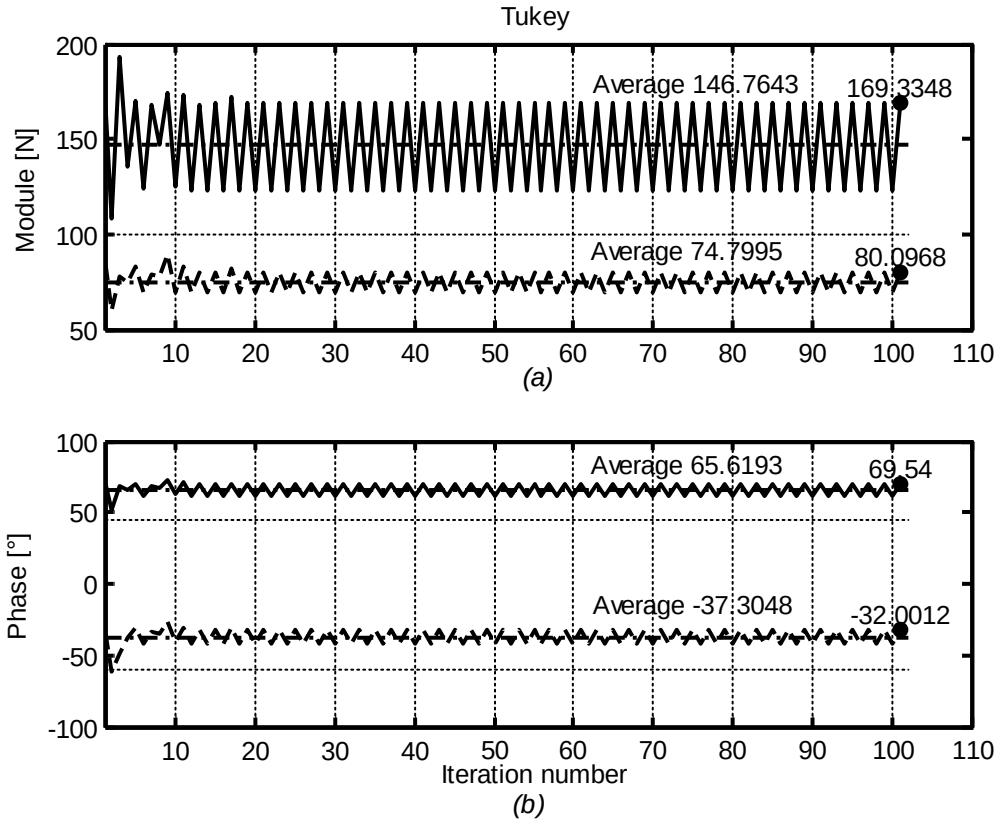

Figure 17. Estimated values vs. iteration for the Tukey's estimate: (a) force amplitudes, (b) force phases.

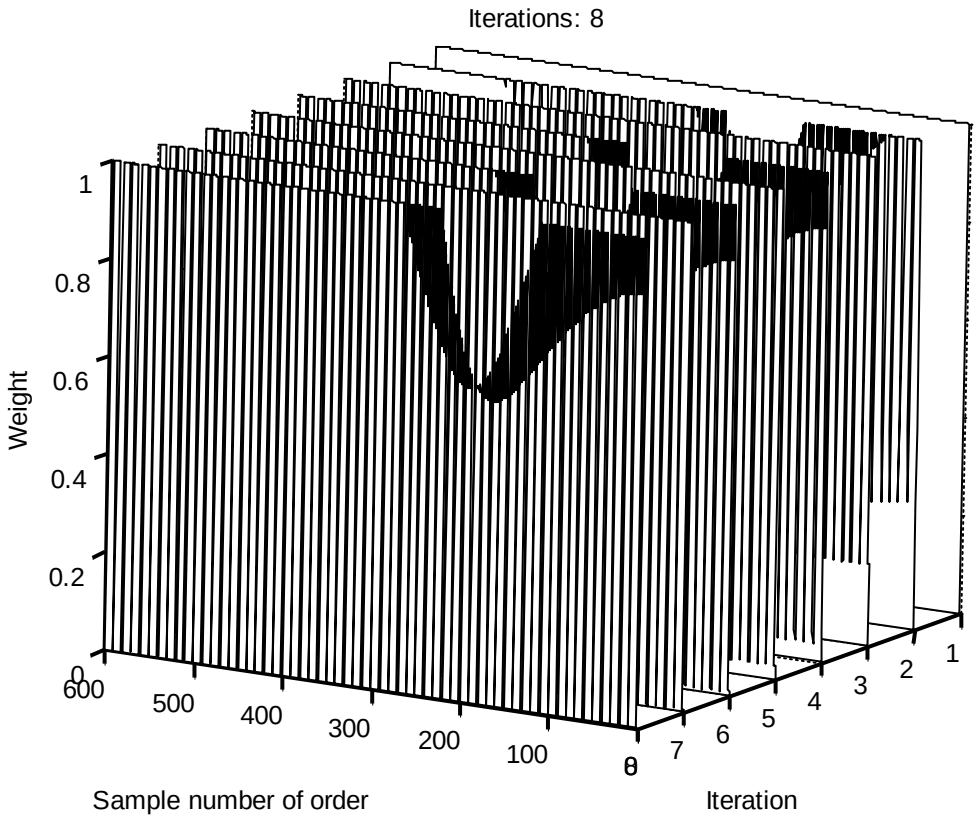

Figure 18. Weighs attributed per iteration for the Huber's estimate.

6. Conclusions

The paper presents in detail the theory of the M-estimation and discusses its application to the identification of excitations in mechanical systems, instead of the more tradition least squares. The rigorous definition of estimator robustness is introduced and it is analytically shown that least squares are definitely not robust. The different types of M-estimators proposed in literature are discussed and the modifications required to be applied to mechanical systems are presented. Their performances are theoretically forecast and verified by means of two applications of excitation identification in a SISO and a MIMO mechanical systems. In particular the IRLS algorithm necessary for the calculations is presented. It is very simple in its implementation, since it represent an iterated application of the standard algorithm used for weighted least squares.

The results obtained in the numerical applications show that few of the proposed M-estimators are suitable for identification problems in mechanical system and that practically only those are characterized by unique solution obtain remarkable results, being very robust not only under an analytical point of view.

References

[1] Friswell M.I. and Mottershead J.E. 1995, *Finite Element Model Updating in Structural Dynamics*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

[2] Bachschmid N., Pennacchi P. and Vania A., 2002, "Identification of Multiple Faults in Rotor Systems", *Journal of Sound and Vibration*, **254**(2), 327-366.

[3] Pennacchi P., Bachschmid N., Vania A., Zanetta G.A. and Gregori L., 2006, "Use of Modal Representation for the Supporting Structure in Model Based Fault Identification of Large Rotating Machinery: Part 1 – Theoretical Remarks", *Mechanical Systems and Signal Processing*, **20**(3), 662-681.
[4] Liu Y. and Shepard S. Jr., 2005, "Dynamic Force Identification Based on Enhanced Least Squares and Total Least-squares Schemes in the Frequency Domain", *Journal of Sound and Vibration*, **282**(1-2), 37-60.
[5] Zhu L.-M., Li H.-X. and Ding H., 2005, "Estimation of multi-frequency signal parameters by frequency domain non-linear least squares", *Mechanical Systems and Signal Processing*, **19**(5), 955-973.
[6] White P.R., Tan M.H. and Hammond J.K., 2006, "Analysis of the maximum likelihood, total least squares and principal component approaches for frequency response function estimation", *Journal of Sound and Vibration*, **290**(3-5), 676-689.
[7] Verboven P., Guillaume P., Cauberghe B., Parloo E. and Vanlanduit S., 2004, "Frequency-domain generalized total least-squares identification for modal analysis", *Journal of Sound and Vibration*, **278**(1-2), 21-38.
[8] Verboven P., Cauberghe B., Guillaume P., Vanlanduit S. and Parloo E., 2004, "Modal parameter estimation and monitoring for on-line flight flutter analysis", *Mechanical Systems and Signal Processing*, **18**(3), 587-610.
[9] Verboven P., Cauberghe B. and Guillaume P., 2005, "Improved total least squares estimators for modal analysis", *Computers and Structures*, **83**(25-26), 2077-2085.
[10] Luh G.-C. and Cheng, W.-C., 2004, "Identification of immune models for fault detection", *Proceedings of the Institution of Mechanical Engineers. Part I: Journal of Systems and Control Engineering*, **218**(5), 353-367.
[11] Krasker W.S., 1983, "Robust Regression", in *Encyclopedia of Statistical Sciences*, Eds. S. Kotz, N.L. Johnson, C.B. Read, N. Balakrishnan and B. Vidakovic, Vol. 8, John Wiley & Sons, Ltd, 166-169.
[12] Bickel P., 1983, "Robust Estimation", in *Encyclopedia of Statistical Sciences*, Eds. S. Kotz, N.L. Johnson, C.B. Read, N. Balakrishnan and B. Vidakovic, Vol. 8, John Wiley & Sons, Ltd, 157-163.
[13] Huber P.J., 1981, *Robust Statistics*, John Wiley, New York.
[14] Bai Z.D. and Wu Y., 1997, "General M-estimation", *Journal of Multivariate Analysis*, **63**(1), 119-135.
[15] Hampel F.R., 1968, *Contributions on the Theory of Robust Estimation*, Ph.D. dissertation, University of California, Berkeley.
[16] Hampel F.R., Ronchetti E.M., Rousseeuw P.J.,and Stahel W.A., 1986, *Robust Statistics: The Approach Based on Influence Functions*, John Wiley, New York.
[17] Piccolo D., 2000, *Statistica*, Il Mulino, Milano, 798-805.
[18] Rey W.J.J., 1983, *Introduction to Robust and Quasi-Robust Statistical Methods*, Springer, Berlin.
[19] Mosteller F. and Tukey J.W., 1977, *Data Analysis and Regression*, Addison-Wesley, Reading, MA.
[20] Pettitt A.N., 1982, "Cramér-Rao lower bound", in *Encyclopedia of Statistical Sciences*, Eds. S. Kotz, N.L. Johnson, C.B. Read, N. Balakrishnan and B. Vidakovic, Vol. 2, John Wiley & Sons, Ltd, 219.
[21] Ruppert D., 1983, "M-Estimators", in *Encyclopedia of Statistical Sciences*, Eds. S. Kotz, N.L. Johnson, C.B. Read, N. Balakrishnan and B. Vidakovic, Vol. 5, John Wiley & Sons, Ltd, 443-449.
[22] Chen H. and Meer P., 2003, "Robust Regression with Projection Based M-estimators", *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV'03)*, 0-7695-1950-4, 1-8.
[23] Rousseeuw P.J. and Leroy A.M., 1987, *Robust Regression and Outlier Detection*, John Wiley, New York, 202.
[24] Fox J., "Robust Regression", *An R and S-PLUS Companion to Applied Regression*, http://cran.r-project.org/.
[25] Coleman D., Holland P., Kaden N., Klema V. and Peters S.C., 1980, "A System of Subroutines for Iteratively Reweighted Least Squares Computations", *ACM Trans. Math. Softw.*, **6**, 327-336.
[26] Zhang Z., 1997, "Parameter estimation techniques: a tutorial with application to conic fitting", *Image and Vision Computing*, **15**, 59-76.
[27] Collins J.R., 2000, "Robustness Comparisons of Some Classes of Location Parameter Estimators", *Ann. Inst. Statist. Math.*, **52**(2), 351-366.
[28] Liu R.Y., Parelius J.M. and Singh K., 1999, "Multivariate Analysis by Data Depth: Descriptive Statistics, graphics and Inference", *The Annals of Statistics*, **27**(3), 783-858.
[29] Hwang J., Jorn H. and Kim J., 2004, "On the Performance of Bivariate Robust Location Estimators under Contamination", *Computational Statistics & Data Analysis*, **44**, 587-601.
[30] Ma Y. and Genton M.G., 2001, "Highly Robust Estimation of Dispersion Matrices", *Journal of Multivariate Analysis*, **78**(1), 11-36.
[31] Donoho D.L. and Gasko M., 1992, "Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness", *Ann. Statist.*, **20**(4), 1803-1827.
[32] Barnett V., 1976, "The Ordering of Multivariate Data", *Journal of Royal Statistical Society - Series A*, **139**, 319-354.
[33] Tukey J.W., 1975, "Mathematics and Picturing of Data", *Proc. of International Congress of Mathematicians*, Vancouver, 523-531.
[34] Pennacchi P., Vania A. and Bachschmid N., 2005, "Bivariate analysis of complex vibration data: An application to condition monitoring of rotating machinery", *Mechanical Systems and Signal Processing*, doi:10.1016/j.ymssp.2005.05.008.
[35] Andrews D.F., Bickel P., Hampel F.R., Huber P.J., Rogers W.H. and Tukey J.W., 1972, *Robust Estimates of Location: Survey and Advances*, Princeton Univ. Press, Princeton.

[36] Gross A.M., 1981, "Confidence Intervals for Bisquare Regression Estimates", *J. Amer. Statist. Ass.*, 72, 341-354.
[37] Singh R. and Saha M., 2003, "Identifying Structural Motifs in Proteins", *Proc. of 8th Pacific Symposium on Biocomputing*, Lihue, Hawaii, USA, January 3-7, 2003, 1-13.
[38] De la Torre F., Vitria J., Radeva, P. and Melenchon J., 2000, "Eigenfiltering for flexible eigentracking (EFE)", *Proc. of IEEE 15th International Conference on Pattern Recognition*, Barcelona, Spain, Volume 3, September 3-7, 2000, 1106-1109.
[39] Lee M.-S., Katkovnik V., Kim K. and Kim Y.H., 2004, "Robust M-Estimation for Multipath CDMA Communication Systems with Phased Array Receiver", *Proc. of IEEE 5th Workshop on Signal Processing Advances in Wireless Communications*, Lisbon, Portugal, July 11-14, 2004, 160-164.

Appendix A

In mathematical terms, considering only the simplest case with two variables (a SISO system), the model that relates a variable of interest x to the observed data z is called *empirical* model. It is reasonable to assume that the mean value of the random variable x is related to z by:

$$E[x \mid z] = \theta_0 + \theta_1 z \,. \tag{69}$$

Although the mean value of x is a linear function of z, the actual values observed of x usually do not follow exactly a straight line. The appropriate way to generalize this situation is to assume that the expected value of x is a linear function of z, but also that the expected value of x is given by the linear model with the addiction of an error e, for a fixed value of z. This corresponds to a simple linear regression with a single predictor z and a response variable x:

$$x = \theta_0 + \theta_1 z + e \,. \tag{70}$$

The error e is assumed to have normal distribution $N(0,\sigma)$, null mean value and variance equal to σ^2. Once z is known, the mean value and the variance of x are insofar:

$$\begin{aligned} E[x \mid z] &= \theta_0 + \theta_1 z \\ V[x \mid z] &= \sigma^2 \end{aligned} \,. \tag{71}$$

The estimate of θ_0 and θ_1 is made by means of the best approximation of the data. The most known and more used method was introduced by Gauss, which proposed the estimate of the parameters θ_0 and θ_1 by means of the minimization of the sum of the vertical distances between the observations (the data) and the line of linear interpolation. This is the *method of the least squares*. The n observations can be expressed as:

$$x_i = \theta_0 + \theta_1 z_i + e_i, \quad i = 1,\ldots,n \,. \tag{72}$$

The sum of the squares of the distances of the observations from the actual regression line is:

$$L = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(x_i - \theta_0 - \theta_1 z_i \right)^2 \,. \tag{73}$$

The estimators of θ_0 and θ_1 in the least squares sense are named $\hat{\theta}_0$ and $\hat{\theta}_1$ and have to satisfy the so called *normal equations of the least squares*:

$$\left.\frac{\partial L}{\partial \theta_0}\right|_{\hat{\theta}_0,\hat{\theta}_1} = -2\sum_{i=1}^{n}\left(x_i - \theta_0 - \theta_1 z_i\right) = 0$$
$$\left.\frac{\partial L}{\partial \theta_1}\right|_{\hat{\theta}_0,\hat{\theta}_1} = -2\sum_{i=1}^{n}\left(x_i - \theta_0 - \theta_1 z_i\right) z_i = 0 \quad . \tag{74}$$

The solution of the normal equations (74) gives the least squares estimate of $\hat{\theta}_0$ and $\hat{\theta}_1$. The estimated regression line is therefore:

$$\hat{x} = \hat{\theta}_0 + \hat{\theta}_1 z \,. \tag{75}$$

Note that whatever couple of observations satisfies the equation:

$$x_i = \theta_0 + \theta_1 z_i + r_i \,, \tag{76}$$

where the quantity $r_i = x_i - \hat{x}_i$ is called *residue*: it evaluates the error in the model approximation of the *i*-th observation x_i.

Appendix B

In order to show that $\psi(r_i)$ is proportional to the influence function of ρ, let's start from eq. (11) that can be written in general terms as:

$$\sum_{i=1}^{n}\psi(X_i;T) = 0 \tag{77}$$

or:

$$\int \psi\left(x, T[F]\right) \mathrm{d}F = 0 \,. \tag{78}$$

The effect of the added arbitrary observation (see eq. (3)) is:

$$\frac{\partial}{\partial \varepsilon}\int \psi\left(x, T\left[(1-\varepsilon)F + \varepsilon\,\delta_x\right]\right) \mathrm{d}\left[(1-\varepsilon)F + \varepsilon\,\delta_x\right]\Big|_{\varepsilon=0} = 0 \,, \tag{79}$$

which becomes, by changing the order of the integration and differentiation:

$$\begin{aligned} &\int \psi\left(x, T\left[(1-\varepsilon)F + \varepsilon\,\delta_x\right]\right) \mathrm{d}\left[\,\delta_x - F\right]\Big|_{\varepsilon=0} + \\ &+\int \frac{\partial}{\partial T[F]}\psi\left(x, T\left[(1-\varepsilon)F + \varepsilon\,\delta_x\right]\right) \times \frac{\partial}{\partial \varepsilon} T[F]\Big|_{\varepsilon=0} \mathrm{d}F\Big|_{\varepsilon=0} = 0 \end{aligned} \,, \tag{80}$$

and simplifying:

$$\int \psi\left(x, T[F]\right) \mathrm{d}\left[\,\delta_x - F\right] + \frac{\partial}{\partial \varepsilon} T[F]\Big|_{\varepsilon=0} \times \int \frac{\partial}{\partial T[F]}\psi\left(x, T[F]\right) \mathrm{d}F = 0 \,. \tag{81}$$

Therefore, $IF(x, T, F)$ is also equal to (see eq. (6)):

$$IF(x,T,F)=\frac{\partial}{\partial\varepsilon}T[F]\Big|_{\varepsilon=0}=\frac{\psi\left(x,T[F]\right)}{-\int\frac{\partial}{\partial T[F]}\psi\left(x,T[F]\right)\mathrm{d}F}, \tag{82}$$

providing that the denominator is non-zero. Therefore the influence function is proportional to $\psi\left(x,T[F]\right)$:

$$IF(x,T,F)\propto\psi\left(x,T[F]\right). \tag{83}$$