CSMA 2019
14ème Colloque National en Calcul des Structures
13-17 Mai 2019, Presqu'île de Giens (Var)

# Material parameter identification using set-valued inverse problem and detection of outliers in the noisy measurements

K. Shinde[1], P. Feissel[1], S. Destercke[2]

[1] *Roberval, Université de Technologie de Compiègne, {krushna.shinde,pfeissel}@utc.fr*
[2] *Heudiasyc, Université de Technologie de Compiègne, sebastien.destercke@hds.utc.fr*

**Résumé** — The present work describes a new parameter identification strategy using a set-valued inverse method. We applied this strategy to identify the elastic parameters of an isotropic material, and it allows to take into account both prior information about the parameters as well as measurement uncertainty in the form of sets (interval or boxes) during the inversion process. The main advantage of this strategy is not only it helps to obtain a feasible set of the parameters but is also able to detect the outliers in the noisy measurements.
**Key words** — inverse problem, set uncertainty, outlier detection.

## 1 Introduction

Material characterization is necessary to design optimal structures. Characterization of material behavior requires (under heterogeneous conditions, e.g. complex loading, geometry or material) a method to identify the material parameter. The deterministic identification problem is generally ill-posed, and one way to resolve this issue is to add uncertainties. These uncertainties pervade multiple aspects of the problem : in the model, in the measurements, in expert knowledge. Modeling all these uncertainties by probabilities is questionable, especially when information is partial or missing [7].

Many methods are proposed in the literature such as deterministic least squares and non-deterministic Bayesian inverse method for identification [2] [4]. The first does not give error bounds on the identified parameter, and it is sensitive to the presence of outliers in the data used to fit a model [8]. The latter adopts subjective probabilistic modeling [1]. In the literature, a lot of researchers argued that probabilistic methods such as Bayesian inverse methods are not well suited for representing and propagating uncertainty when information is missing [3] [9] [5]. This method is sensitive to outliers in the data and to mis-specification of priors [10], and requires a prior assumption on the uncertainties, e.g. measurement error usually considered as Gaussian. Hence, it is crucial to have an identification framework which should consider all kind of information, and should be robust to outliers.

This work addresses these issues, by proposing a set-valued inverse method which is not only able to identify a feasible set of the parameters but is also able to detect outliers in the measurements. The work is structured as follows. Firstly, the identification strategy for set-valued inverse problem is described which includes the theoretical approach and its implementation. Secondly, this strategy is applied to identify the elastic parameters of an isotropic material from full-field displacement measurements and detect outliers in the measurements.

## 2 Identification strategy

Our strategy aims at identifying the set of the model parameters by making use of available information such as measurement data as well as prior information about the parameters using sets to model the information.

### 2.1 Set-valued inverse problem

Let us consider an inverse problem where we want to identify some parameters of a model from measurements. A direct problem yields the relationship between the model parameters and the measurements as shown in Equation (1).

$$y = f(x) \tag{1}$$

where $x$ denotes the parameters of the model such that $x \in \mathbb{R}^{N_x}$, $N_x$ is the number of parameters, and $y$ denotes the measurements such that $y \in \mathbb{R}^{N_y}$ , $N_y$ is the number of measurements. A typical example introduced in Section 3 is the one where $x$ corresponds to elastic Lamé parameters and $y$ is full-field displacement data. The measured quantity corresponding to $y$ is denoted as $\tilde{y}$. The discrepancy between $y(x)$ and $\tilde{y}$ is mainly due to model and measurement errors. In the proposed approach, sets describe the uncertainty in both the measurements and the identified parameters. Hence, the solution of the inverse problem can be obtained thanks to a set inversion process. The uncertainty in the measurements is described through the set $\mathbb{S}_y \subset \mathbb{R}^{N_y}$ as shown in Equation (2) with each measurement having a lower bound $\underline{\tilde{y_k}}$ and an upper bound $\overline{\tilde{y_k}}$.

$$\mathbb{S}_y = \prod_{k=1}^{N_y} [\underline{\tilde{y_k}}, \overline{\tilde{y_k}}] \tag{2}$$

Given a set $\mathbb{S}_y \subset \mathbb{R}^{N_y}$ describing the uncertainty on $\tilde{y}$ , the set $\mathbb{S}_x \subset \mathbb{R}^{N_x}$ describing the solution of the inverse problem is defined as shown in Equation (3) where $\mathbb{S}_{ox} \subset \mathbb{R}^{N_x}$ is a prior parameter set.

$$\mathbb{S}_x = \{x \in \mathbb{S}_{ox} \mid f(x) \in \mathbb{S}_y\} \tag{3}$$

In the current work, it is possible to obtain a solution set for each individual measurement, denoted as $\mathbb{S}_x^k$ shown in Equation (4) and $\mathbb{S}_x$ can be obtained as the intersection of the $\mathbb{S}_x^k$ by using Equation (5).

$$\mathbb{S}_x^k = \{x \in \mathbb{S}_{ox} \mid y_k(x) \in [\underline{\tilde{y_k}}, \overline{\tilde{y_k}}]\} \tag{4}$$

$$\mathbb{S}_x = \bigcap_{k=1}^{N_y} \mathbb{S}_x^k \tag{5}$$

In case of inconsistent measurements, the set-valued inverse method gives an empty solution set $\mathbb{S}_x = \emptyset$ corresponding to $\bigcap_{k=1}^{N_y} \mathbb{S}_x^k = \emptyset$. There may be several reasons for the inconsistency of the measurements with respect to the model : presence of measurement outliers, model error.

## 2.2 Outlier detection

In case of inconsistency, a way to restore consistency is to remove incompatible measurements, i.e. outliers. To do this, we must evaluate the degree of consistency of each measurement. We will now propose such a measure.

For any two solution sets $\mathbb{S}_x^k$ and $\mathbb{S}_x^{k'}$ corresponding to $\tilde{y_k}$ and $\tilde{y_{k'}}$ measurement respectively, $(k,k') \in \{1,...,N_y\}^2$, we define the following indicators of the degree of inclusion with one another as shown in Equation (6) where $\mathcal{A}(\mathbb{S}_x^k)$ corresponds to the area of the set $\mathbb{S}_x^k$.

$$C_{k'k} = \frac{\mathcal{A}(\mathbb{S}_x^k \cap \mathbb{S}_x^{k'})}{\mathcal{A}(\mathbb{S}_x^k)} \quad \text{and} \quad C_{kk'} = \frac{\mathcal{A}(\mathbb{S}_x^k \cap \mathbb{S}_x^{k'})}{\mathcal{A}(\mathbb{S}_x^{k'})} \tag{6}$$

These indicators follow the properties shown in Equation (7).

$$C_{k'k} = \begin{cases} 1 & \text{iff } \mathbb{S}_x^k \subseteq \mathbb{S}_x^{k'} \\ 0 & \text{iff } \mathbb{S}_x^k \cap \mathbb{S}_x^{k'} = \emptyset \end{cases} \quad \text{and} \quad C_{kk'} = \begin{cases} 1 & \text{iff } \mathbb{S}_x^{k'} \subseteq \mathbb{S}_x^k \\ 0 & \text{iff } \mathbb{S}_x^{k'} \cap \mathbb{S}_x^k = \emptyset \end{cases} \tag{7}$$

Furthermore, the value of $C_{k'k}$ or $C_{kk'}$ will always be between 0 and 1 when $\mathcal{A}(\mathbb{S}_x^k)$ or $\mathcal{A}(\mathbb{S}_x^{k'})$ is non-zero. The larger these indicators are, the larger is the overlapping of $\mathbb{S}_x^k$ and $\mathbb{S}_x^{k'}$, hence the higher the degree of inclusion between the corresponding measurements.

By using the pairwise degree of inclusion of the measurements, we define the global degree of consistency (GDOC) of any $k^{th}$ measurement as shown in Equation (8).

$$GDOC(k) = \frac{\sum_{k'=1}^{N_y} \frac{\mathcal{A}(\mathbb{S}_x^k \cap \mathbb{S}_x^{k'})}{\mathcal{A}(\mathbb{S}_x^k)} + \sum_{k'=1}^{N_y} \frac{\mathcal{A}(\mathbb{S}_x^k \cap \mathbb{S}_x^{k'})}{\mathcal{A}(\mathbb{S}_x^{k'})}}{2N_y} \tag{8}$$

The GDOC follows some properties as shown in Equation (9).

$$GDOC(k) = \begin{cases} 1 & \text{iff } \mathbb{S}_x^1 = \mathbb{S}_x^2, ..., = \mathbb{S}_x^k \\ 0 & \text{iff } \mathbb{S}_x^k \cap \mathbb{S}_x^{k'} = \emptyset, \ \forall\, k' \in \{1, ..., N_y\} \end{cases} \tag{9}$$

The value of $GDOC(k)$ will always be between 0 and 1, and also one thing to notice that the condition that allows GDOC =1 is very strong to satisfy. If $GDOC(k) = 0$ then the $k^{th}$ measurement is fully inconsistent with all other measurements. A high value of GDOC for the $k^{th}$ measurement indicates that it has a high consistency with other measurements. Hence, $GDOC$ can detect less consistent measurements from the set of measurements that can be considered as outliers. The idea is to remove measurements that have 0 or very low value of GDOC from the set of measurements to obtain non-empty solution set.

### 2.3 Implementation

In order to compute the set inversion, we have to choose a discrete description of the sets. This could be done through a subpaving of boxes strategy, as it is used in the SIVIA algorithm [6]. Here, we choose to use the same description of the sets as the one used in [5] through a grid of points, $x_i$, $i \in \{1, ..., N_g\}$ where $N_g$ is the number of grid points. Such a description is convenient when comparing or intersecting the sets since the grid of points is the same for any set. Any set $\mathbb{S}_x \subset \mathbb{S}_{ox}$ is then characterized through its discrete characteristic function, defined at any point $x_i \in \mathbb{S}_{ox}$ of the grid as shown in Equation (10).

$$\chi_{\mathbb{S}_x}(x_i) = \begin{cases} 1 & \text{if } x_i \in \mathbb{S}_x \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

In the current application, a uniform grid is chosen to describe prior parameter set $\mathbb{S}_{ox}$, but it is not mandatory. In our method, each $\mathbb{S}_x^k$ is therefore described by its discrete characteristic function, defined at any point of the grid as shown in Equation (11) where $y_k(x_i)$ represents the model data at any grid point.

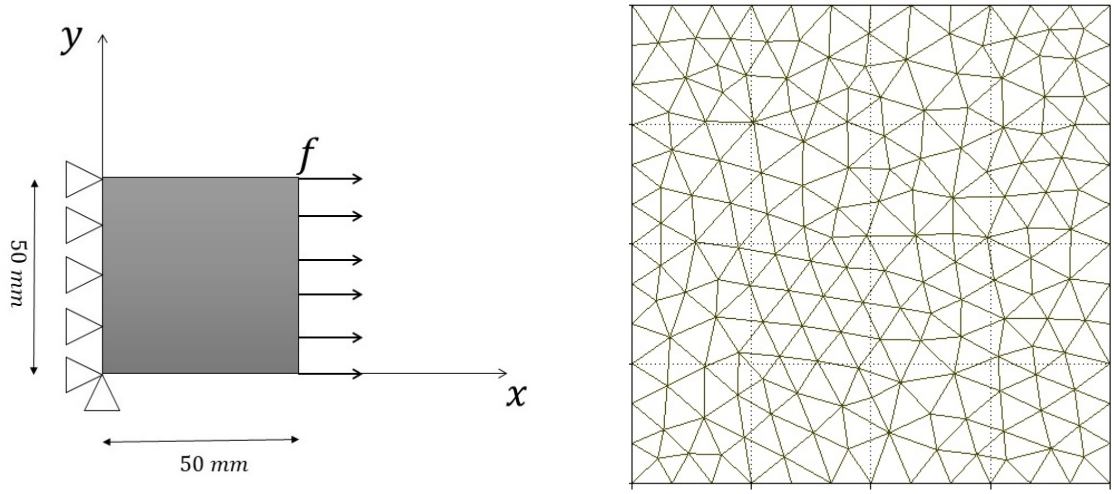
$$\chi_{\mathbb{S}_x^k}(x_i) = \begin{cases} 1 & \text{if } \underline{\tilde{y}_k} \le y_k(x_i) \le \overline{\tilde{y}_k} \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

These discrete characteristic function can be collected in a $N_g \times N_y$ matrix $X$ as columns of boolean values. By taking advantage of the matrix $X$, a $N_y \times N_y$ matrix $T = X^T X$ can be obtained which is symmetric. The diagonal element $T_{kk}$ of $T$ represents the number of grid points from a prior parameter set $\mathbb{S}_{ox}$ for which the $k^{th}$ measurement is consistent and it is proportional to $\mathcal{A}(\mathbb{S}_x^k)$. The non-diagonal element $T_{kk'}$ of $T$ represents the number of grid points from a prior parameter set $\mathbb{S}_{ox}$ for which both $k^{th}$ and $k'^{th}$ measurements are consistent and it is proportional to $\mathcal{A}(\mathbb{S}_x^k \cap \mathbb{S}_x^{k'})$. Hence, GDOC can be computed from matrix $T$ for any $k^{th}$ measurement as shown in Equation (12).

$$GDOC(k) = \frac{\sum_{k'=1}^{N_y} \frac{T_{k'k}}{T_{kk}} + \sum_{k'=1}^{N_y} \frac{T_{kk'}}{T_{k'k'}}}{2N_y} \tag{12}$$

## 3 Applications

In this Section, we applied the set-valued inverse method to identify elastic properties (Lamé parameters : $\lambda$ and $\mu$) of a homogeneous 2D plate as shown in Figure 1(a). The plate is clamped on the left side and loaded on the right side by a uniform traction $f = 1000\ N$. To generate displacement measurement data $\tilde{y}$ (386 measurements), exact displacement data $y^{Ref}$ is simulated by FE model (193 nodes, 336 elements) as shown in Figure 1(b) with taking reference value : $\lambda_0 = 1.15 \times 10^5\ MPa$ and $\mu_0 = 7.69 \times 10^4\ MPa$. Then, the measurement $\tilde{y}$ is created from $y^{Ref}$, by adding a Gaussian white noise with standard deviation $\sigma$. In the current work, $\sigma$ was taken as 5% of the average of all the exact displacement values and it can be assumed that $\sigma$ can be deduced from the measurement technique.




(a) 2D homogeneous plate (b) FEM mesh

FIGURE 1 – A homogeneous plate and its model

For the set-valued inverse method, the uncertainty on the measurements is described in interval form. Therefore, each measurement is considered in interval form with lower and upper bounds ($[\ \tilde{y}_k - 2\sigma,\ \tilde{y}_k + 2\sigma\ ]$). The bounds on the measurement should be greater than the standard deviation $\sigma$ to capture the solution set of the parameter. Hence in the current work, it was chosen as $2\sigma$. Prior information about the parameters ($\mathbb{S}_{ox}$) is considered as a uniform grid $\lambda \times \mu$ with $\lambda = [0.72 \times 10^5,\ 1.90 \times 10^5]$ and $\mu = [\ 7.2 \times 10^4,\ 8.15 \times 10^4]$. The method is studied with different natures of the uncertainty in the data such as when there is no noise in the data, when there is random noise in the data, when data is corrupted because of the local inclusion in the material corresponding to a non-homogeneous material.

## 3.1 Identification from exact data

In this section, we applied the set-valued inverse method to identify the set of elastic parameters when there is no noise in the data. The measurement data was chosen to be exact such that $\tilde{y} = y^{Ref}$ and the information on the measurement $\tilde{y}$ was described in an interval form : $[\ \tilde{y} - 2\sigma,\ \ \tilde{y} + 2\sigma\ ]$.

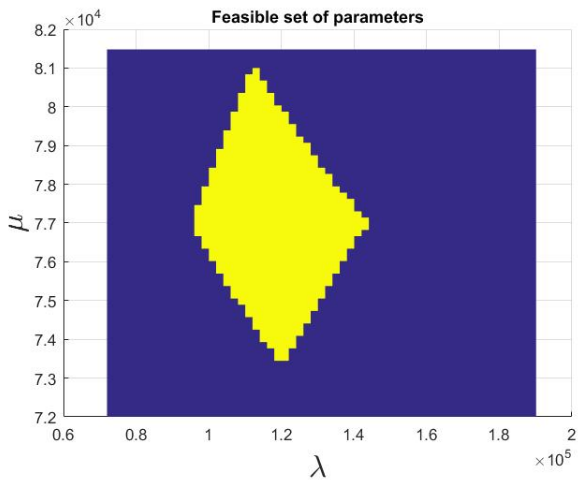
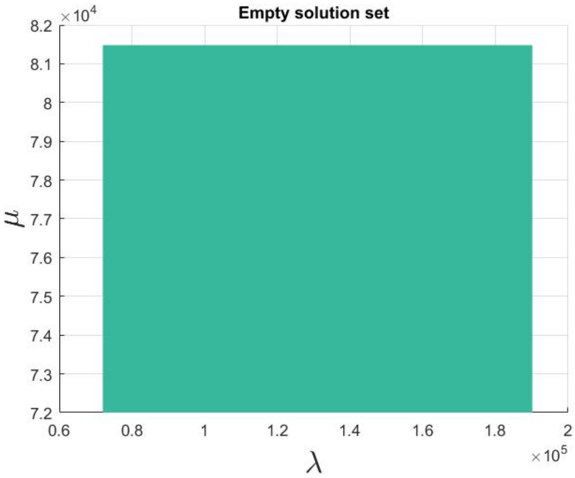



FIGURE 2 – Feasible set of parameters

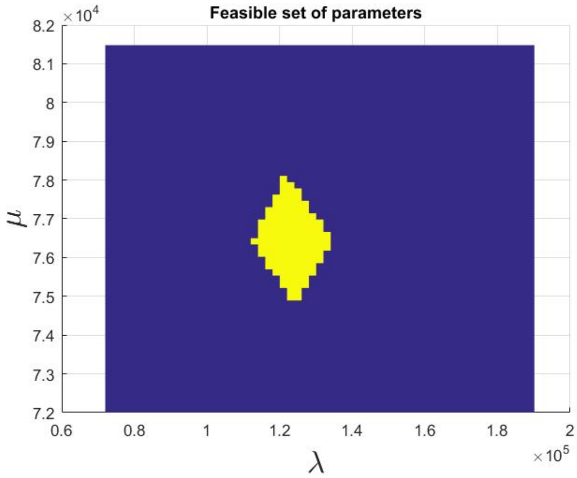
Figure 2 shows the feasible set (yellow color) of the identified parameter which is consistent with all 386 measurements using the set-valued inverse method. The size of the feasible solution set depends on the magnitude of the upper and lower bounds of the measurement interval.

### 3.2 Identification from data with random noise

In this section, we applied the set-valued inverse method to identify the set of elastic parameters when there is random noise in the data. The measurement $\tilde{y}$ is created form $y^{Ref}$, by adding a Gaussian white noise with standard deviation $\sigma$ and the information on the measurement $\tilde{y}$ was described in an interval form : $[\,\tilde{y}-2\sigma, \quad \tilde{y}+2\sigma\,]$.




(a) Empty solution set




(b) Solution set after detecting outlier

FIGURE 3 – Outlier detection

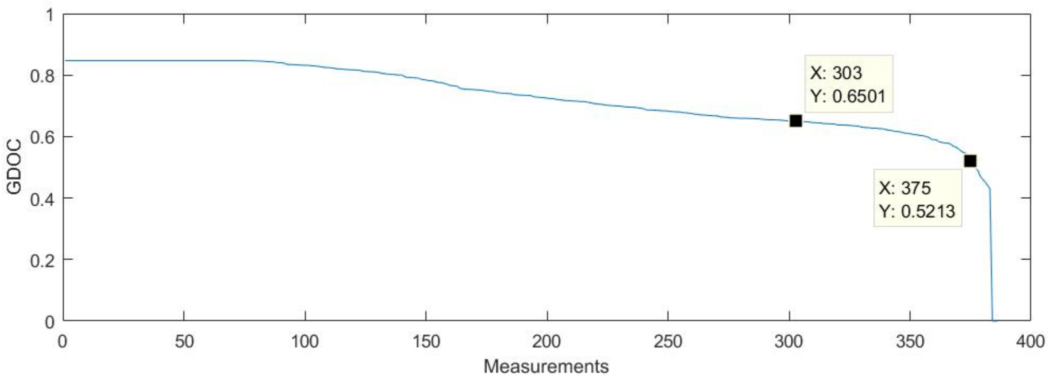



FIGURE 4 – GDOC

Figure 3(a) shows that the identified set (green color) is empty due to inconsistency within the measurements. Hence, to obtain a non-empty solution set, we need to detect the outliers. A way to detect outliers is to know GDOC of each measurement. Figure 4 shows GDOC of all the measurements in decreasing order, and it can be observed that the value of GDOC starts to decrease abruptly on this example where GDOC is between 0.65 and 0.52. A possible criterion for outlier detection could be to detect this abrupt decrease of the estimator of consistency of the data. This detection should be performed automatically, yet in this first example, it was done manually by choosing a threshold of 0.64. The corresponding identified set (yellow color) is presented in Figure 3(b), where 75 measurements were removed.

### 3.3 Identification from data corrupted by a local inclusion in the material

In this section, we applied the set-valued inverse method to identify the set of elastic parameters when data is corrupted because of a less stiff local inclusion in the material which results in a non-homogeneous material. The measurement $\tilde{y}$ (corrupted data) is created by FE model as shown in Figure 5 with taking reference value of $\lambda_1 = \lambda_0 = 1.15 \times 10^5\ MPa$, $\mu_1 = \mu_0 = 7.69 \times 10^4\ MPa$ for the region of the mesh outside the red boundary and $\lambda_2 = 2.77 \times 10^4\ MPa$, $\mu_2 = 4.16 \times 10^4\ MPa$ for the region of the mesh inside the red boundary. The information on the measurement $\tilde{y}$ was described in an interval form : $[\,\tilde{y}-2\sigma,\;\; \tilde{y}+2\sigma\,]$.

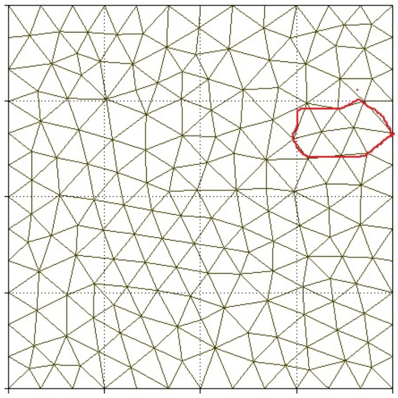
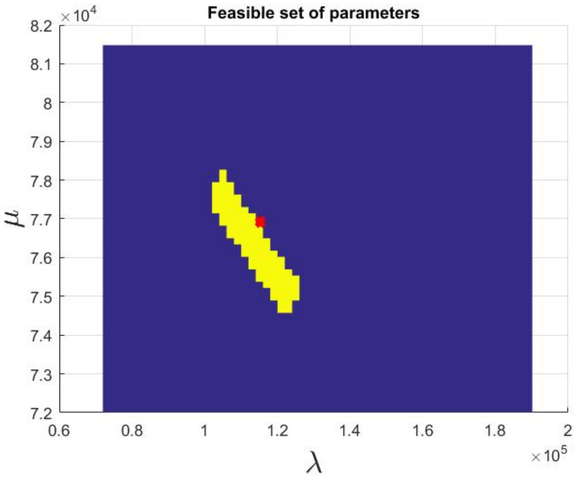
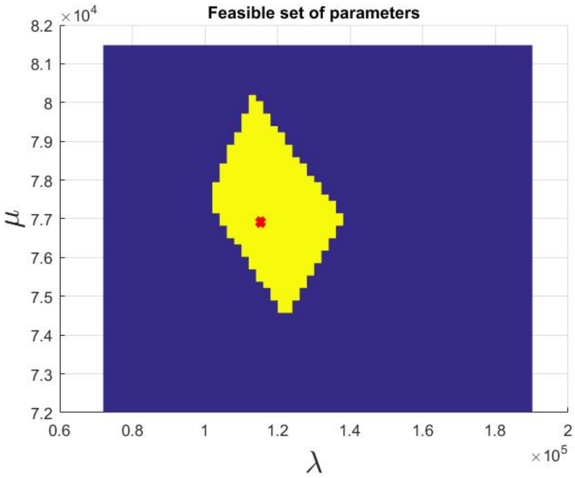


FIGURE 5 – FEM mesh




(a) Solution set before detcting outliers




(b) Solution set after detecting outliers

FIGURE 6 – Outlier detection

Figure 6 (a) shows the solution set of the identified parameters which is consistent with all 386 measurements and it can be observed that the identified set does not include the reference value of Lamé parameters ($\lambda_0$, $\mu_0$) shown by red mark. Hence, in this case, we want to reject the measurements which are corrupted. The same procedure as the one proposed in section 3.2 is then applied to detect outliers based on a manual detection of the brutal drop-off of the GDOC curve presented in Figure 7. In this particular case, the threshold is chosen as 0.68 and the identified set is presented in Figure 6(b) with 20 measurements removed. The removed measurements are close to a less stiff inclusion, and such a removal allows to identify a set which includes the reference value of Lamé parameters ($\lambda_0$, $\mu_0$) of the bulk material shown by a red mark.

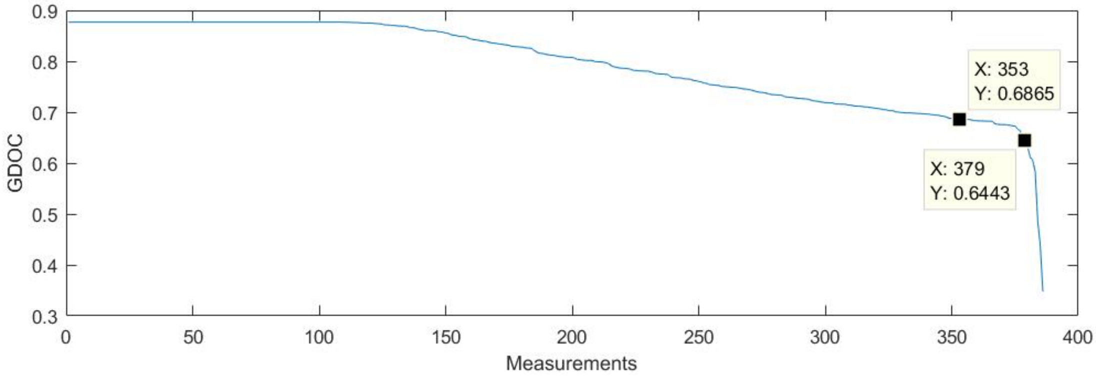



FIGURE 7 – GDOC

# 4 Conclusions

In this work, we presented a new parameter identification strategy relying on the set theory. This strategy is robust to outliers. We applied this strategy to identify the elastic properties of an isotropic material with three different cases of measurement data as described in section 3.1, 3.2 and 3.3 respectively. The results showed that the identification strategy is not only helpful to obtain a feasible set of the parameters but is also able to detect the outliers in the noisy measurements.

The detection of outliers in the measurements is based on a global degree of consistency (GDOC) of each measurement. The main challenge is then to decide the threshold value of GDOC to remove outliers. In the current work, the threshold was detected manually based on the drop-off of the GDOC curve. The current work focuses on the automatic detection of the drop-off and the other outlier detection criterions. These are under investigation and shall be compared for the presentation.

Then the next step in this work is to compare this strategy with the least squares method or Bayesian inference. Then, we intend to apply this strategy to structural health monitoring problems, identification of non-linear material model.

## Acknowledgements

The research reported in this paper has been supported by the project Labex MS2T, financed by the French Government through the program Investments for the future managed by the National Agency for Research (Reference ANR-11-IDEX-0004-02).

### Références

[1] A. Gelman. *Objections to Bayesian statistics.* Bayesian Anal., 3 :445-449, 2008.

[2] C. Gogu, R. Haftka, R.L. Riche, J.Molimard and A. Vautrin. *Introduction to the bayesian approach applied to elastic constants identification.* AIAA journal, 8(5) :893–903, 2010.

[3] I. Elishakoff. *Essay on uncertanties in elastic and viscoelastic structures : from a. m. freudenthal's criticisms to modern convex modeling.* Computers and Structures, Vol. 56. No. 6. pp. 871-895, 1995.

[4] J. Beck and S. Au. *Bayesian Updating of Structural Models and Reliability using Markov Chain Monte Carlo Simulation.* Journal of Engineering Mechanics, 4 :380–391, 2002.

[5] L. Sui, L. P. Feissel and T.Denœux. *Identification of Elastic Properties in the Belief Function Framework.* International Journal of Approximate Reasoning, Vol. 101, Pages 69-87, 2018.

[6] L. Jaulin and E. Walter. *Set Inversion via Interval Analysis for Nonlinear Bounded-error Estimation.* Auloma-tica, Vol. 29, No. 4, pp. 1053-10164, 1993.

[7] P. Walley. *Statistical Reasoning with Imprecise Probabilities.* Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Vol. 42, 1991.

[8] R. Blais. *Least Squares for Practitioners.* Mathematical Problems in Engineering, 10.1155/2010/508092., 2013.

[9] S. Ferson, L. R. Ginzburg. *Different methods are needed to propagate ignorance and variability.* Reliability Engineering & System Safety, 54(2) :133–144, 1996.

[10] Y. Chen, P. A. Breen and N.L. Andrew. *Impacts of outliers and mis-specification of priors on Bayesian fisheries-stock assessment.* Canadian Journal of Fisheries and Aquatic Sciences, 57(11) :2293-2305, 2000.