Case study of non-linear inverse problems: mammography and non-destructive evaluation

O. Kosheleva and S. Cabrera[a];
R. Osegueda, C. Ferregut, S. Nazarian, D. L. George, M. J. George, V. Kreinovich[b];
and K. Worden[c]

[a]Department of Electrical and Computer Engineering and [b]FAST Center,
University of Texas at El Paso, El Paso, TX 79968, USA

[c]Department of Mechanical Engineering,
University of Sheffield, Sheffield S1 3JD, United Kingdom

ABSTRACT

The inverse problem is usually difficult because the signal (image) that we want to reconstruct is weak. Since it is weak, we can usually neglect quadratic and higher order terms, and consider the problem to be linear. Since the problem is linear, methods of solving this problem are also, mainly, linear (with the notable exception of the necessity to take into consideration, e.g., that the actual image is non-negative).

In most real-life problems, this linear description works pretty well. However, at some point, when we start looking for a better accuracy, we must take into consideration non-linear terms. This may be a minor improvement for normal image processing, but these non-linear terms may lead to a major improvement and a great enhancement if we are interested in outliers such as faults in non-destructive evaluation or bumps in mammography. Non-linear terms (quadratic or cubic) give a great relative push to large outliers, and thus, in these non-linear terms, the effect of irregularities dominate. The presence of the non-linear terms can serve, therefore, as a good indication of the presence of irregularities.

We describe the result of the experiments in which these non-linear terms are really helpful.

Keywords: Non-linear inverse problem, non-linear data compression, non-destructive evaluation, mammography

1. INTRODUCTION: AN INVERSE PROBLEM CAN BE LINEAR, OR IT CAN HAVE A SMALL NON-LINEARITY

Active vs. passive inverse problems. In many applied problems, it is difficult (or even impossible) to directly measure the desired characteristics $c_1, \ldots, c_n$ of the object in which we are interested. So, to find the values of these characteristics, we measure some other characteristics $y_1, \ldots, y_m$ which are related to c_j, and then reconstruct the values of the desired parameters c_j from the results $\widetilde{y}_1, \ldots, \widetilde{y}_m$ of measuring these characteristics. The problem of reconstructing the values c_j from the measurement results $\widetilde{y}_1, \ldots, \widetilde{y}_m$ is called the *inverse problem*.

Examples of such problem range from engineering (e.g., detecting and measuring cracks and other faults in aerospace structures), to science (e.g., reconstructing brightness of a distant astronomical source), to medicine (e.g., detecting and measuring parameters of possible tumors), etc.

Depending on the scope of measurement possibilities, we can divide inverse problems into two classes:

- First, we have inverse problems in which only a relatively small amount of data is available; examples of such problems include reconstructing an images of a distant quasar or a distant galaxy, reconstructing a deep-site geophysical structure from the available earthquake data, etc. In such problems, often, even if we use all available information, we can still have at best rather crude estimates of the important parameters. Therefore, the only possible strategy here is to use all available information when solving the inverse problem. It is natural to call such problems *passive* because we have no simple way to actively solicit any more information; the only thing we can do is simply observe what is there and make conclusions based on these observations. For passive

First author's e-mail olga@ece.utep.edu

control problems, the main problem is: *Given the measurement results, how can we reconstruct the values of the desired characteristics?* Usually, we know how the measured characteristics $y_1, \ldots, y_m$ depend on the desired ones c_j, i.e., we know that $y_i = f_i(c_1, \ldots, c_n)$ for some known functions $f_1, \ldots, f_m$. So, in mathematical terms, the problem is to solve the system of m (approximate) equations $\widetilde{y}_i \approx f_i(c_1, \ldots, c_n)$, where $\widetilde{y}_i$ are known (measured) values, and $c_1, \ldots, c_n$ are unknowns.

- In many other situations, however, the inverse problem is *active* in the sense that we can send different signals $x_1, \ldots, x_k$ to the object of our interest, and get different measurement results based on this choice of signals. For example, in medical imaging, we can send ultrasound or X-ray or other signals and measure the result of these signals passing through (or reflected by) the human body; in geophysics, we can send the signals to the Earth crust and measure the results, etc.

In this paper, we will analyze active inverse problems.

For the active inverse problem, we still have the same problem as we had for the passive ones: *Given the measurement results, how can we reconstruct the values of the desired characteristics?* For active problems, the measured characteristics y_i depend not only on the desired quantities c_j, but also on the signals, i.e., $y_i = f_i(c_1, \ldots, c_n, x_1, \ldots, x_k)$ for some known functions $f_1, \ldots, f_m$. So, in mathematical terms, this problem looks slightly more complicated than in the passive case: we must solve a system of m (approximate) equations $\widetilde{y}_i \approx f_i(c_1, \ldots, c_n, x_1, \ldots, x_k)$, where $\widetilde{y}_i$ are known (measured) values, x_j are known values of applied signals, and $c_1, \ldots, c_n$ are unknowns.

However, in addition to the original problem, now that we can control what exactly we measure, we also need to solve a different problem: *What exactly should we measure?* In other words, which values $x_1, \ldots, x_k$ of the signal should we choose?

There may be different criteria for selecting these signals: e.g., we may want to be able to reconstruct the values of the desired characteristics with the best possible accuracy, or we may want to minimize the total cost consisting of the detection cost and losses caused by possible un-detected faults, etc.

The dependence of y_i on x_j is usually linear. If the signal-to-noise ratio is high, then the measurement errors can be neglected, i.e., we can safely assume that the measured values $\widetilde{y}_i$ coincide with the actual values y_i of the measured quantities. In such situations, to reconstruct the desired values c_j, we have m exact equations with n unknowns: $\widetilde{y}_i = f_i(c_1, \ldots, c_n, x_1, \ldots, x_k)$. In general, if we have at least as many equations as unknowns, this system has a unique solution; there exist standard numerical methods for getting such solutions. In this case, the inverse problem is easily solvable.

The real problem starts when the signals x_i are relatively weak, and as a result, the signal-to-noise ratio is not very high. It may happen because we are simply unable to send a strong signal: e.g., when we analyze the geophysical structures several kilometers below, even the strongest test explosions or vibrations lead to very small and hard-to-detect reflected signals. Sometimes, it is potentially possible to send a stronger signal, but this stronger signal may damage or even destroy the very object that we are interested in: e.g., a strong test vibration may damage the tested plane, powerful X-rays may hurt or even kill a patient, etc. In such situations, we are forced to restrict ourselves to weak signals only.

The weakness of the signals is both a curse and a blessing. On one hand, from the viewpoint of the *result*, it is clearly a curse because when the signals are weak, the signal-to-noise ratio goes down, and the quality of reconstructing the desired values c_j deteriorates. On the other hand, from the viewpoint of the computational complexity, weakness is a blessing. Indeed, when the dependencies f_i are smooth (and they are normally smooth), we can expand each function f_i in Taylor series in x_j:

$$f_i(x_1, \ldots, x_n) = f_i(0, \ldots, 0) + \sum_j \frac{\partial f_i}{\partial x_j} \cdot x_j + \frac{1}{2!} \sum_{j,k} \frac{\partial^2 f_i}{\partial x_j \partial x_k} \cdot x_j \cdot x_k + \ldots \tag{1}$$

Since the values x_j are small, we can neglect terms which are quadratic and of higher order in x_j and safely assume that the dependence of y_i on x_j is linear. Thus, we have a system of equations in which the functions f_i are linear in x_j. This linearity makes computations simpler than in the general non-linear case.

Sometimes, the dependence of y_i on x_j is slightly non-linear. The quadratic terms in the equation (1) can be neglected if the signal x_j is weak, and if the second derivatives are reasonably small (i.e., if the dependencies f_i

are sufficiently smooth). If one of these two conditions is not satisfied, then the quadratic terms can no longer be neglected, and we have a non-linear dependence. So, if a signal is weak but the function f_i is not very smooth, we must take non-linearity into consideration when solving an active inverse problem.

Traditional viewpoint: this minor non-linearity is of minor importance. Traditionally, the above-described non-linearity is viewed (correctly) as a second order term. It is definitely extremely important to take non-linear terms into consideration if we want a superaccurate reconstruction. However, in most practical problems, when this super accuracy is not needed, this non-linearity is a minor term that is rather a nuisance than a serious problem.

What we are planning to do. In this papers, we will show, by analyzing the cases of non-destructive evaluation of aerospace structures and mammograms, that this small non-linearity can actually be of great practical importance.

2. NON-LINEARITY CAN BE OF GREAT PRACTICAL IMPORTANCE

In many practical problems, it is very important to test smoothness. In many practical problems, we must check whether a given object is smooth or whether it has non-smooth areas. For example, when we test an aerospace structure, we must check whether it is still safe to fly, or it has cracks, holes, or other faults. Similarly, the main problem of mammography is to detect small non-smoothnesses in the mammal (small clots, cracks, etc.), which may indicate a tumor. In both cases, we must detect possible non-smoothness.

Non-smoothness leads to non-linearity. If a tested structure has no faults, then the surface is usually smooth. As a result, the dependencies f_i are also smooth. Since we are sending relatively weak signals x_i (strong signals can damage the plane), we can neglect quadratic (and higher order) terms in Taylor series and only consider linear terms in these series; thus, the dependency will be *linear*.

A fault (e.g., a crack) is, usually, a violation of smoothness. Thus, if there is a fault, the structure stops being smooth; hence, the function f_j stops being smooth, and therefore, linear terms are no longer sufficient. Thus, *in the absence of fault, the dependence is linear, but with the faults, the dependence is non-linear.*

So, we can detect the fault by checking whether the dependency between y_j and x_i is linear.

Comment. The idea that non-linear terms can be helpful has been suggested some time ago; see, e.g., Ref. 2.

Non-linear terms simplify the detection of non-smoothness. When a smooth object acquires a fault, two changes occur in the dependence of y_i on x_j: first, linear terms change; second, non-linear terms appear. Thus, there are two possible methods of detecting non-smoothness: we can compare the linear response with the response of an ideal (smooth) object, or we can try to detect non-smoothness by finding non-linearity.

The first method works well if we do know the ideal response function f_i. However, in some cases, e.g., in mammography, we *do not know* the linear response of the fault-less object. Or, alternatively, we may know the original response, but we also know that this response can change not only because of faults, but also because of stress, material wariness, and other factors that do not necessarily mean that there is a dangerous fault inside. In such cases, we *cannot* detect the fault by comparing the current *linear* response with the ideal one; however, if we detect *non-linear* terms, it is a clear indication that there are some faults inside.

The resulting proposal: main idea. As a result of the above analysis, we propose the following way of detecting faults:

- We apply different signals x_j to the object, and measure the response y_i.
- If the measurement results are consistent with the linear dependence of y_i on x_j, this means that there are no faults, and no further testing is needed.
- If the measurement results are inconsistent with the linear model, this means that there is a fault, and so further thorough tests are needed.

Checking linearity is easy. As a result, for non-destructive evaluation of aerospace structures, we get a simple test that enables us to save time and resources (necessary for the detailed solution of the inverse problem) by limiting this detalization only to the cases when the presence of the faults was revealed by non-linearity.

Let us confirm that non-smoothness leads to non-linearity. To show the above non-linearity is indeed practically detectable and thus, practically useful, we will present mechanical analysis and experimental results.

3. MECHANICAL ANALYSIS OF NON-LINEARITY

In this section, we present a simplified mechanical explanation of non-linearity. This explanation is too oversimplified to explain the *quantitative* experimental results, but it explains, on the *qualitative* level, why non-linearities do occur.

In order to understand how non-linear effects can occur, let us first describe how the signal travels through a fault-less plate. In this case, at the location of the transmitter, we send, at any given moment of time t, the signal $x(t) = A \cdot \cos(\omega \cdot t)$. This signal travels to the receiver (measuring device) with a velocity equal to the speed of sound. For simplicity, we can assume that the plate is homogeneous, so at any point, we have the same speed of sound v. Thus, while traveling from the transmitter to the receiver, the signal gets delayed by the amount of time $\Delta t = d/v$, where d is the distance between the transmitter and the receiver. As a result, at any moment of time t, the values of the observed signal $y(t)$ is proportional to value $x(t - \Delta t)$ that the input signal had Δt seconds ago: $y(t) = k \cdot x(t - \Delta t)$, where the coefficient k describes the loss of amplitude.

Thus, for a fault-less plate, we indeed have a linear dependence between the transmitted signal $x(t)$ and the measured signal $y(t)$.

Let us now consider the case when a fault lies between the transmitter and the receiver. This fault may be a crack or a hole. In this case, we can also use the formula $y(t) = k \cdot x(t - \Delta t)$, where Δt is the delay. However, this delay can no longer be computed simply as d/v, because, in addition to going straight through the material, the signal has to go either through or around the crack. In both cases, the presence of the crack changes the travel time:

- If the ultrasound has to travel through air, then it is delayed because the speed of sound in the air is smaller than the speed of sound in the solid body.
- If the ultrasound has to go around the crack, then the speed of sound stays the same, but the length of the path increases, and so the signal is also delayed.

In both cases, the delay Δt between the transmitter and the receiver can be computed as $\Delta t = d/v + k_f \cdot d_0$, where d_0 is the linear size of the fault, i.e., the distance between the front and the rear borders ("walls") of the fault area (front and rear with respect to the transmitter), and the coefficient k_f describes how fast the signal passes the fault area. As a result, the measured signal is equal to $y(t) = k \cdot x(t - \Delta t) = k \cdot A \cdot \cos(\omega \cdot t - \omega \cdot \Delta t)$. Since we are interested in detecting small faults, the value d_0 is small, so we can expand the expression for $y(t)$ in terms of d_0 and keep only the first few terms. As a result, we get the following formula

$$y(t) = A \cdot \cos(\omega \cdot t - \omega \cdot d_f/v) + k_f \cdot d_0 \cdot A \cdot \sin(\omega \cdot t\omega \cdot d_f/v) + o(d_0). \tag{2}$$

Before we send the signal, the plate is immobile, and the distance d_0 stays constant: $d_0(t) = d_0^{(0)}$. However, as we transmit the signal $x(t)$, the plate starts vibrating, and this vibration changes the position of both borders and therefore, changes the distance d_0: $d_0 = d_0(t)$. In order to describe this change, let us denote the distance between the transmitter and the fault's front border by d_f. By the time the signal reaches this left border, it is delayed by the time d_f/v, i.e., takes the form $x_{\rm front}(t) = k_{\rm front} \cdot A \cdot \cos(\omega \cdot t - \omega \cdot d_f/v)$. This vibration causes the corresponding change in the location of this front border: instead of being equal exactly to d_f, this location oscillates around x_f. At any given moment of time, the change in location is proportional to the amplitude $x_{\rm front}(t)$ of oscillating signal:

$$d_{\rm front}(t) = d_f + k_{\rm mov} \cdot x_{\rm front}(t) = d_f + k_{\rm mov} \cdot k_{\rm front} \cdot A \cdot \cos(\omega \cdot t - \omega \cdot (d_f/v)),$$

for some coefficient $k_{\rm mov}$.

Similarly, the signal that passes to the rear border gets delayed by $\approx d_f/v + k_f \cdot d_0^{(0)}$. Thus, the location location of the rare border also changes, as

$$d_{\rm rear}(t) = d_f + k_{\rm mov} \cdot x_{\rm rear}(t) = d_f + k_{\rm mov} \cdot k_{\rm front} \cdot A \cdot \cos(\omega \cdot t - \omega \cdot (d_f/v) - \omega \cdot k_f \cdot d_0^{(0)}).$$

As a result of these slightly different oscillations, the size $d_0(t) = d_{\rm real}(t) - d_{\rm front}(t)$ also changes with time. We have already mentioned that the size d_0 is small, so we can expand the expression for $d_0(t)$ in terms of $d_0^{(0)}$ and keep only the first few terms. As a result, we get the following formula:

$$d_0(t) = d_0^{(0)} + k_{\rm mov} \cdot k_{\rm front} \cdot A \cdot \omega \cdot k_f \cdot d_0^{(0)} \cdot \sin(\omega \cdot t - \omega \cdot (d_f/v)) + o(d_0^{(0)}). \tag{3}$$

Substituting (3) into (2), we get, in $y(t)$, in addition to terms proportional to $\cos(\omega t)$, also quadratic terms $\sin^2(\omega t)$ which lead to double frequency terms in the Fourier transform of $y(t)$. These terms are proportional to A^2.

Similarly, we get cubic terms, etc.

4. EXPERIMENTAL CONFIRMATION OF NON-LINEARITY

First experiments: pseudo-random signals. The first experimental confirmation that for an ultrasonic scan, faults do cause non-linear terms, was presented in Ref 9. Namely, it was known that for a fault-less plate, the dependence between the transmitted signal $x(t)$ and the measured signal $y(t)$ is linear, i.e., $y(t) = \int A(t-s) \cdot x(s)\,ds$ for some function $A(t)$. It turned out that for a plate with a fault, this dependence is non-linear: namely, cubic terms must be taken into consideration. To detect this non-linearity, the authors of Ref. 9 used *pseudo-random signals* that combine components of several different frequencies with pseudo-random amplitudes and pseudo-random phases.

The data from[9] shows that the amplitude of the cubic term is roughly proportional to the cube of the linear fault size. Thus, not only the non-linear terms indicate the *presence* of the fault, but also the value of the cubic term can be used to determine the *size* of the fault.

Pseudo-random signals are difficult to generate, so, it is preferable to use simpler test signals. In practice, it is difficult to generate pseudo-random signals. It is therefore desirable to confirm that non-linearity can be also observed for simpler signals, e.g., for sinusoid signals.

Experiment with sinusoid signals: a hardware part. In our experiment, as a signal x_j, we sent an ultrasound wave. To generate this wave, a sinusoid electric signal $x(t) = A \cdot \cos(\omega \cdot t)$ was sent to the transducer, which then generated an ultrasonic wave in the tested object. The transducer was set at an angle of incidence of 31° with the plate, so that a wave would go along the surface of the plate (such waves are called *Lamb waves*; see, e.g., Refs. 2–5,8.

If the transducer was ideally linear, then we would get an ultrasonic wave of the exact same frequency and of the same sinusoid shape as the original electric signal. In this case, to detect the non-linearity of the plate, it would be sufficient to place a single sensor on the plate and check whether the signal $y(t)$ measured by this sensor depends linearly on $x(t)$.

In reality, however, the transducer is somewhat non-linear; as a result, the ultrasonic signal sent to the plate contained components at frequencies different from the original frequency ω: it has components which are slightly different from ω, and it also has higher harmonics, i.e., frequencies close to 2ω, 3ω, etc. We chose $\omega = 500KHz$; for this frequency, the ultrasonic signal is mainly located in the frequency area from 350 to 650 KHz.

To separate the non-linearity of the transducer from the non-linearity of the plate itself, we placed *two* sensors on the plate: the first sensor is located near the transducer, and it measures the ultrasonic wave $x_1(t)$ that the transducer generates; the second sensor is located at a distance from the transducer, and it measure the wave $x_2(t)$ changed after passing through the plate. Then, we check whether $x_2(t)$ linearly depends on $x_1(t)$.

How to check non-linearity: general discussion. The detection of non-linearity is based on the fact that the general linear time-invariant dependency has the form $x_2(t) = \int A(t-s) \cdot x_1(s)\,ds$ for some function $A(t)$. In terms of Fourier components, this dependency takes a simple form $\hat{x}_2(\omega) = \hat{A}(\omega) \cdot \hat{x}_1(\omega)$. Thus, to check whether the dependence is linear, it is sufficient to check whether, for each ω, the Fourier component $\hat{x}_2(\omega)$ is a linear function of the Fourier component $\hat{x}_1(\omega)$.

How to check non-linearity: ideal case. If the signal $x_1(t)$ is purely harmonic $x_1(t) = A \cdot \cos(\omega \cdot t)$, then it has only one Fourier component, and all we have to do to check non-linearity is to take different amplitudes A, and to plot the absolute value of the corresponding Fourier component $|\hat{x}_2(\omega)|$ of the signal measured by the second sensor as a function of $A = |\hat{x}_1(\omega)|$. Instead of the absolute values of the Fourier components, we could take their energies $E_2 = |\hat{x}_2(\omega)|^2$ and $E_1 = |\hat{x}_1(\omega)|^2$; if the dependency of $x_2(t)$ on $x_1(t)$ is linear, then E_2 is a linear function of E_1: $E_2 = k \cdot E_1$.

Due to the inevitable noise, the measured energy at the second sensor also contains a noise component, i.e., $E_2 = k \cdot E_1 + n$, where n is the energy of the noise. Crudely speaking, if the dependence between E_1 and E_2 is linear, this means that $x_2(t)$ linearly depends on $x_1(t)$, otherwise, the dependence of $x_2(t)$ on $x_1(t)$ is non-linear.

How to check non-linearity: main idea of the practical method. In our case, the original electric signal has only one Fourier component with the frequency $\omega = 500$ KHz, but, as we have mentioned, due to the non-linearity of the transducer, the resulting ultrasound wave has components in a certain vicinity of this original frequency. As a result, the signal's energy is distributed over the resulting range of frequencies. Thus, instead of the value $|\hat{x}_2(\omega)|$ corresponding to a single frequency, we took the total energy $E_2 = \int |\hat{x}_2(\omega)|^2\,d\omega$ of the signal in the frequency range, where the integral is taken over the entire range of frequencies $[350\text{ KHz}, 650\text{Khz}]$. Similarly, for the first sensor, we take an integral $E_1 = \int |\hat{x}_1(\omega)|^2\,d\omega$. We then check whether E_2 is a linear function of E_1.

The choice of an object. In our experiments, as a sample object, we took an aluminum 6065 plate; its size is 36×18 in, its thickness is 1/16 in. Initially, we performed the measurements on the undamaged plate. Then, we simulated a crack by sawing across the 18 in width of the plate with a fine tooth hand saw. The crack is at the middle of the plate. The two sensors were placed at an equal distance from the crack (or, for the un-damaged plate, at an equal distance from the center line where we later cut in a crack).

The choice of a signal. We wanted to make sure that the first sensor really measures the original ultrasound wave. Therefore, we restricted our signals only to the first moments of time after the beginning of the experiments, before the wave reflected from the plate's borders gets back to the location of the first sensor. To be able to separate the original signal from its later reflections, we generated only five cycles of the 500 KHz wave.

To check for non-linearity, we repeated this experiment at several different voltage levels of the original electric signal: 0V (pure noise), 6V, 7V, 8V, and 9V.

The choice of sampling frequency. We used a sampling frequency of 10 million samples per second (MSPS), i.e., 20 samples per cycle (we first tried 5 MSPS, but the noise was too high to make any conclusions, so we had to double the sampling frequency).

Filtering out reflections and the original noise. Based on the geometry of the plate and on the known speed of sound waves, we estimated (and later experimentally confirmed) that the reflection starts in at least 250 points after the original signal, and that the entire signal (before reflection) occurs in the first 2500 data points, so we only measured the first 2500 data points.

The entire 2500-point data starts as noise (no signal), then contains the signal, and then has the signal mixed with the reflections. To separate the signal from the original noise and from the following reflections, we selected 256 points out of the 2500 available. As a criterion for selecting the front edge of the data, it is natural to chose the first instance when the measured signal exceeds a certain portion of the maximum amplitude. Based on our observations, we have chosen 1/6 as this portion. So, for each sensor k ($k = 1, 2$), we computed the largest value $A_{\max}$ of all 2500 amplitudes, found the first point t_i at which the measured value $x_k(t_i)$ was larger than or equal to $A_{\max}/6$, and counted a total of 256 points $x_k(t_i), x_k(t_{i+1}), \ldots, x_k(t_{i+255})$. Then, we applied FFT to the selected data, and used this FFT to compute the total energy E_k of the signal in the frequency interval [350 KHz, 650Khz].

We further decreased noise by repeating the measurements. To decrease the noise, we repeated each five-wave burst 200 times, and averaged the signals before processing them. To estimate the measurement accuracy, we repeated the same 200-burst experiment ten time. Then, as a result of measuring energy, we took an interval $[E_k^-, E_k^+]$ between the smallest and the largest of the resulting ten values.

How to check non-linearity: formulation of the problem. As a result of the measurements, we got several intervals $[E_1^-(V), E_1^+(V)]$ and $[E_2^-(V), E_2^+(V)]$ corresponding to different voltages V. We know that for each voltage, the actual (unknown) values of the energy $E_1(V)$ and $E_2(V)$ lie within the corresponding intervals. The question is: is this data consistent with the assumption that $E_2(V)$ is a linear function of $E_1(V)$? Or, in other words, it is possible to find real numbers $k > 0$, n, and values $E_1(V) \in [E_1^-(V), E_1^+(V)]$ and $E_2(V) \in [E_2^-(V), E_2^+(V)]$ for which $E_2(V) = k \cdot E_1(V) + n$?

How to check non-linearity: derivation of an algorithm. For each V, we want to have a value of $E_2(V)$ that satisfies the following two properties: first, it belongs to the interval $[E_2^-(V), E_2^+(V)]$, and second, it can be represented as $k \cdot E_1(V) + n$ for some $E_1(V) \in [E_1^-(V), E_1^+(V)]$.

Let us first assume that the values $k > 0$ and n are given. Since $k > 0$, the function $k \cdot E_1(V) + n$ is increasing, and so, for each V, when $E_1(V)$ takes values from the interval $[E_1^-(V), E_1^+(V)]$, the expression $k \cdot E_1(V) + n$ takes values from the interval $[k \cdot E_1^-(V) + n, k \cdot E_1^+(V) + n]$. Thus, the above two conditions on $E_2(V)$ mean that $E_2(V)$ must belong to two different intervals: $[E_2^-(V), E_2^+(V)]$ and $[k \cdot E_1^-(V) + n, k \cdot E_1^+(V) + n]$. This is possible if and only if these two intervals have a non-empty intersection, i.e., if $E_2^-(V) \le k \cdot E_1^+(V) + n$ and $k \cdot E_1^-(V) + n \le E_2^+(V)$.

Now, the question is: when is it possible to find $k > 0$ and n for which these inequalities hold for all V? Let us first assume that k is given. Then, by moving n into one side of each inequality, we can reformulate the above inequalities in the following way: $E_2^-(V) - k \cdot E_1^+(V) \le n$ and $n \le E_2^+(V) - k \cdot E_1^-(V)$. Such a value n exists if and only if all the lower bounds for n are smaller than or equal to all the upper bounds for n, in other words, if $E_2^-(V) - k \cdot E_1^+(V) \le E_2^+(V') - k \cdot E_1^-(V')$ for all possible values of V and V'.

So, the original question can be reformulated as follows: does there exist a value k for which this inequality is true for all V and V'? We can somewhat simplify this inequality by moving all terms which contain k to one side and all other terms to another side. As a result, we get the inequality $k\cdot(E_1^-(V')-E_1^+(V))\le E_2^+(V')-E_2^-(V)$. In our case, the energy of the wave monotonically increases with the voltage V, so that if $V<V'$, then $E_1^+(V)<E_1^-(V')$. Hence, when $V<V'$, the above inequality is equivalent to

$$k\le\frac{E_2^+(V')-E_2^-(V)}{E_1^-(V')-E_1^+(V)},$$

and when $V>V'$, the above inequality turns into

$$\frac{E_2^-(V)-E_2^+(V')}{E_1^+(V)-E_1^-(V')}\le k.$$

Such a value k exists if and only if all lower bounds for k are smaller than or equal to all the upper bounds for k, i.e., when

$$\max_{V'<V}\frac{E_2^-(V)-E_2^+(V')}{E_1^+(V)-E_1^-(V')}\le\max_{V<V'}\frac{E_2^+(V')-E_2^-(V)}{E_1^-(V')-E_1^+(V)}. \tag{4}$$

How to check non-linearity: the resulting algorithm. To check non-linearity, we must check the inequality (4).

Experimental results.

V	$[E_1^-(V),E_1^+(V)]$ undamaged, 10^6	$[E_2^-(V),E_2^+(V)]$ undamaged, 10^6	$[E_1^-(V),E_1^+(V)]$ damaged,10^5	$[E_2^-(V),E_2^+(V)]$ damaged, 10^4
0V	$[0.00,0.01]$	$[0.00,0.01]$	$[0.02,0.03]$	$[0.06,0.11]$
6V	$[2.65,2.66]$	$[1.59,1.61]$	$[0.69,0.70]$	$[0.23,0.28]$
7V	$[3.12,3.14]$	$[1.86,1.88]$	$[0.87,0.92]$	$[0.14,0.23]$
8V	$[3.62,3.64]$	$[2.16,2.18]$	$[1.05,1.08]$	$[4.75,4.84]$
9V	$[4.59,4.69]$	$[2.70,2.80]$	$[1.28,1.32]$	$[5.57,5.80]$

In the undamaged case, we clearly have a linear dependency ($E_2(V)\approx 0.6\cdot E_1(V)$), while in the damaged case, the dependence is clearly non-linear.

Why did nobody notice this non-linearity before? Our experiments do not require very complicated and accurate equipment, they use standard sensors and transducers. So why did not anybody make these experiments before? The main reason is that before, people used just one signal level (e.g., the highest possible), to detect the faults. The use of only one signal level is justified if the response is linear: then, another input level will not lead to any new information. To detect non-linear terms, however, we must use at least two different input levels.

5. PRACTICAL RECOMMENDATION: BRIEF SUMMARY

Main recommendation. To detect the faults, we must use at least *two* different signal levels. If the increase in the signal level x_j leads to a proportional increase in the measured values y_i, then most probably the object is smooth. If the dependence of y_i on x_j is non-linear, then, most probably, there is a fault, so further analysis is needed.

Auxiliary recommendations. Since it is important to detect non-linearity, it is important not to smoothen and linearize the signal (or the image) if this signal (image) is to be compressed. To overcome the smoothing aspects of lossy compression and following decompression, it is therefore important to enhance the image before compression. For example, one of the successful enhancement methods consists of replacing the brightness value x at each pixel by a so-called *selective median* of its neighbors, i.e., by a median m if $|x-m|\le\varepsilon$ for some fixed ε, and by the original value x otherwise.

A general comment about checking non-linearity. After K measurements, we have K sets of data $\tilde{x}_1^{(k)},\ldots,\tilde{x}_n^{(k)},\tilde{y}_1^{(k)},\ldots,\tilde{y}_m^{(k)}$, $1\le k\le K$. Often, we do not know the probabilities of different measurement errors, we only know the upper bounds for these errors. So, we know the *intervals* $X_1^{(k)},\ldots,X_n^{(k)},Y_1^{(k)},\ldots,Y_m^{(k)}$ of possible values of the measured quantities. We want to check whether this dependence can be linear, i.e., whether these exist coefficients c_{ij} for which, for every k and j, $\sum c_{ij}\cdot x_j^{(k)}\in Y_j^{(k)}$ for some $x_i^{(k)}\in X_i^{(k)}$. This is a known problem of interval computations: check whether the given system of interval linear equations is solvable (here, the unknowns are c_{ij}, interval coefficients are $X_i^{(k)}$ and $Y_j^{(k)}$).

Our main concern is not to miss the fault, so we need *guaranteed* methods. Thus, we need to use interval (guaranteed) methods for solving linear interval systems (see, e.g., Refs. 1,6,7,10,11).

ACKNOWLEDGMENTS

This work was supported in part by NASA under cooperative agreement NCC5-209, by NSF grants No. DUE-9750858 and CDA-9522207, by the United Space Alliance, grant No. NAS 9-20000 (P.O. 297A001153), and by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number F49620-95-1-0518.

REFERENCES

1. G. Alefeld and J. Herzberger, *Introduction to Interval Computations*, Academic Press, N. Y. 1983.
2. D. E. Chimenti, "Guided waves in plates and their use in materials characterization", *Appl. Mech. Rev.*, **50**, No. 5, pp. 247–287, 1997.
3. R. W. Clough and J. Penzien, *Dynamics of structures*, McGraw Hill, N.Y., 1986.
4. I. Grabec and W. Sachse, *Synergetics of measurement, prediction, and control*, Springer Verlag, Berlin-Heidelberg, 1997.
5. A. K. Mal and S. J. Singh, *Deformation of elastic bodies*, Prentice Hall, Englewood Cliffs, NJ, 1991.
6. A. Neumaier, *Interval Methods for Systems of Equations*, Cambridge University Press, Cambridge, 1990.
7. S. M. Rump, "Solving Algebraic Problems with High Accuracy", in: U. Kulisch and W. Miranker, eds., *A New Approach to Scientific Computation*, Academic Press, N.Y., 1983, pp. 51–120.
8. I. A. Viktorov, *Rayleigh and Lamb waves: Physical theory and applications*, Plenum Press, N.Y., 1967.
9. C. Wolters, S. A. Billings, S. G. Pierce, K. Worden, and G. R. Tomlinson, "Model-based fault detection using Lamb waves", *Proceedings of the International Workshop on Intelligent NDE Sciences for Aging and Futuristic Aircraft*, El Paso, TX, September 30–October 2, pp. 229–244, 1997.
10. K. Worden, R. Osegueda, C. Ferregut, S. Nazarian, D. L. George, M. J. George, V. Kreinovich, O. Kosheleva, and S. Cabrera, "Interval Methods in Non-Destructive Testing of Aerospace Structures and in Mammography", *International Conference on Interval Methods and their Application in Global Optimization (INTERVAL'98), April 20–23, Nanjing, China, Abstracts*, pp. 152–154, 1998.
11. K. Worden, R. Osegueda, C. Ferregut, S. Nazarian, E. Rodriguez, D. L. George, M. J. George, V. Kreinovich, O. Kosheleva, and S. Cabrera, "Interval Approach to Non-Destructive Testing of Aerospace Structures and to Mammography", In: Goetz Alefeld and Raul A. Trejo (eds.), *Interval Computations and its Applications to Reasoning Under Uncertainty, Knowledge Representation, and Control Theory. Proceedings of MEXICON'98, Workshop on Interval Computations, 4th World Congress on Expert Systems*, México City, México, 1998.