Interval Approach to Phase Measurements Can Lead to Arbitrarily Complex Sets – A Theorem and Ways Around It

Bharat C. Mulupuru, Vladik Kreinovich, and Roberto Osegueda
FAST Center for Structural Integrity of Aerospace Structures
The University of Texas at El Paso
El Paso, Texas 79968, USA
contact email vladik@cs.utep.edu

Abstract

We are often interested in phases of complex quantities; e.g., in non-destructive testing of aerospace structures, important information comes from phases of Pulse Echo and magnetic resonance.

For each measurement, we have an upper bound Δ on the measurement error $\Delta x = \widetilde{x} - x$, so when the measurement result is $\widetilde{x}$, we know that the actual value x is in $[\widetilde{x} - \Delta, \widetilde{x} + \Delta]$. Often, we have no information about probabilities of different values, so this interval is our only information about x. When the accuracy is not sufficient, we perform several repeated measurements, and conclude that x belongs to the intersection of the corresponding intervals.

For real-valued measurements, the intersection of intervals is always an interval. For phase measurements, we prove that an arbitrary closed subset of a circle can be represented as an intersection of intervals.

Handling such complex sets is difficult. It turns out that if we have some statistical information, then the problem often becomes tractable. As a case study, we describe an algorithm that uses both real-valued and phase measurement results to determine the shape of a fault. This is important: e.g., smooth-shaped faults gather less stress and are, thus, less dangerous than irregularly shaped ones.

Keywords: Aerospace Structures, Phase Intervals, Fault Shapes

AMS Subject Classification: 65G20, 65G40, 65G30, 68U10, 74Rxx

1 Interval Uncertainty for Real-Valued Measurements

In most measurements, the measured quantity is a real number; see, e.g., [14].

Measurements are never 100% accurate; as a result, the measurement result $\widetilde{x}$ is usually different from the actual (unknown) value x of the measured quantity. For each measuring instrument, the manufacturer provides an upper bound Δ on the (absolute value of the) measurement error $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$: $|\Delta x| \le \Delta$. (If no such bound is provided, this means that an arbitrarily large and/or arbitrarily small value of x is possible, so $\widetilde{x}$ is rather an estimate and not a measurement.)

Often, in addition to this upper bound, we know the probabilities of different values of Δx. However, in many practical situations, we have no information about these probabilities. In such

cases, after we performed the measurement and found the measurement result $\widetilde{x}$, the only information about the (unknown) actual value x is that x cannot differ from $\widetilde{x}$ by more than Δ – i.e., in other words, that x belongs to the interval $\mathbf{x} = [\widetilde{x} - \Delta, \widetilde{x} + \Delta]$.

Often, measurement results serve as inputs to complex data processing algorithms, algorithms that use the measurement results $\widetilde{x}_1, \ldots, \widetilde{x}_n$ to estimate the values of the quantity y that are difficult (or even impossible) to measure directly. There exist techniques – known as *interval computations* (see, e.g., [5, 7, 8, 11]) – that analyze how the interval uncertainty $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in the inputs x_i propagates to the uncertainty $\mathbf{y}$ of the result y of data processing.

Sometimes, the measurement error is too large, so the accuracy resulting from a single measurement is not sufficient. In this case, a natural idea is to perform repeated measurements of the same quantity. After each measurement, we get an interval $\mathbf{x}^{(j)}$ that contains the actual value x of the measured quantity. After N measurements, we know that the value x belongs to all n intervals $\mathbf{x}^{(j)}$; therefore, the actual value x belongs to the *intersection* $\bigcap \mathbf{x}^{(j)}$ of these intervals. This intersection is always an interval, so using this intersection instead of the original (wider) interval does not increase the complexity of the corresponding data processing.

This method is not only in accordance with common sense: it can actually be proven (see, e.g., [17, 18]) that under reasonable conditions, for large N, the intersection is indeed much narrower than each of the original intervals – in other words, repeated measurements do drastically improve the measurement accuracy.

2 Phase Measurements: Necessity

In most measurements, the measured quantity is a real number; however, in many cases, the measured quantity is determined by the delay between the two waves. In such situations, it is often impossible to determine the actual delay, because if the delay coincides with the full period ($2 \cdot \pi$ radians), then the two waves – original and delayed one – are practically indistinguishable. In such situations, we cannot measure the actual delay, we can only measure the relative *phase* φ of the two waves, the phase that takes values from 0 to 2π in such a way that 0 and $2 \cdot \pi$ are indistinguishable. In geometric terms, we can describe the phase φ by a point on the unit circle whose radius forms an angle φ with the OX axis. Let us give two examples of phase measurements.

In Very Large Baseline Interferometry (VLBI; see, e.g., [15]), we use two (or more) distant antennas (separated by several thousand miles) to record the signal from the same extra-galactic radio source. Each antenna site is equipped with a super-precise clock, so we are able to exactly reference each observation to time and thus, to compare the times that it takes for the signal to reach the two antennas. Unfortunately, to be able to effectively amplify the signal, we must restrict ourselves to a narrow frequency band. Within this narrow band, the signal is so close to being periodic that we cannot effectively measure the actual delay between the two recorded signals – only the phase shift between these signals.

Another case when phase measurements are very important is ultrasonic testing of structural integrity; see, e.g., [2, 3]. In this testing, a transmitter emits an ultrasonic wave; part of this wave goes directly to the sensor, part is first reflected by the fault. The delay between the two detected signals indicates how far away the fault is. Similarly to the VLBI case, often, by comparing the two waves, we cannot determine the delay exactly, we can only determine the phase shift between the two waves.

In both examples – VLBI and non-destructive testing – there exist efficient methods for handling the phases. In addition to the above-cited sources, we can mention [1, 4, 9] for radioastronomical data processing, and [19] for data processing in non-destructive testing.

3 Phase Measurements: Interval Uncertainty

Similarly to the case of real-valued measurements, phase measurements are never 100% accurate. The measurement error of a phase measurement can be described by a *distance* $d(x, \widetilde{x})$ between the actual (unknown) value of the phase x and the measured value $\widetilde{x}$. On a unit circle, this distance can be defined as the length of the shortest of the two arcs that connect the corresponding points. In analytical terms, the distance between the two values from 0 to $2 \cdot \pi$ can be defined as

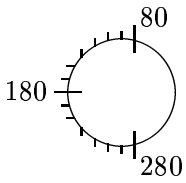
$$d(x, \widetilde{x}) = \min(|x - \widetilde{x}|, |x - \widetilde{x} - 2 \cdot \pi|, |x - \widetilde{x} + 2 \cdot \pi|).$$

For example, the distance between the values 0 and 6 is equal to the smallest of 6 and $2 \cdot \pi - 6 \approx 0.28$, i.e., to ≈ 0.28.

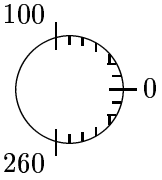
Similar to the real-valued measurements, in many real-life situations, the only information we have about the measurement error of the phase measurement is the upper bound Δ on the distance between x and $\widetilde{x}$. In this case, once we have performed the measurement and measured the value $\widetilde{x}$, the only information that we have about the actual value x of the phase is that the distance between x and $\widetilde{x}$ cannot exceed Δ, i.e., $d(x, \widetilde{x}) \le \Delta$. Once can easily see that this is equivalent to the condition that x belongs to the interval $[\widetilde{x} - \Delta, \widetilde{x} + \Delta]$; see, e.g., [9].

For simplicity, let us illustrate these intervals in terms of degrees (not radians); in terms of degrees, the full circle is 360°.

- If we measured the phase as $\widetilde{x} = 180^\circ$, and the upper bound on the measurement accuracy is $\Delta = 100^\circ$, then the actual value of the phase can be anywhere between $180 - 100 = 80^\circ$ and $180 + 100 = 280^\circ$.



- If we measured the phase as $\widetilde{x} = 0^\circ$, and the upper bound on the measurement accuracy is $\Delta = 100^\circ$, then the actual value of the phase can be anywhere between $0 - 100 = -100^\circ = 260^\circ$ and $0 + 100 = 100^\circ$. In terms of angles from 0 to 360, this interval goes from 260 to 360 (which is the same as 0) and then from 0 to 100.



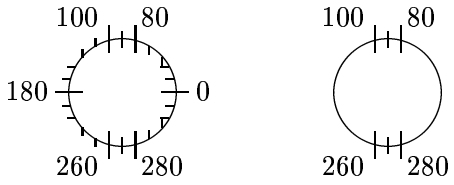
Similarly to the real-valued measurements, if we are not satisfied with the accuracy of a single measurement, a natural idea is to perform repeated measurements of the same quantity and then take the intersection of the corresponding intervals. For real-valued measurement, this intersection is always an interval. In contrast, for phase measurements, the intersection of two intervals may no longer be an interval. Indeed, for the above two measurements, the intersection of the intervals

$$[180 - 100, 180 + 100] = [80, 280]$$

and

$$[0 - 100, 0 + 100] = [-100, 100] = [260, 360] \cup [0, 100]$$

is not a single interval, but a union of two disjoint intervals: $[80, 100]$ and $[260, 280]$.



How complex can such an intersection be?

4 How Complex Can Such an Intersection Be?

We have already seen that the intersection of two intervals can consist of two disjoint intervals. If we add the third interval $[270 - 175, 270 + 175]$ to the above intersection, we conclude that the triple intersection consists of three disjoint intervals: $[260, 280]$, $[80, 85]$, and $[95, 100]$. Our main result is that this intersection can be arbitrarily complex:

Theorem. *An arbitrary closed subset of a circle can be represented as an intersection of intervals.*

Restriction to closed sets is necessary because each interval is a closed set, and the intersection of closed sets is always closed. So, this theorem says, in effect, that the interval approach to phase measurements can lead to arbitrarily complex sets.

Comment. This theorem says, in effect, that a simple problem of measuring the angle (or, to be more precise, measuring the value of an angular physical quantity) becomes much more complicated if we take interval uncertainty into consideration:

- If we do not take the interval uncertainty into consideration, i.e., if we assume that the measurements are absolutely accurate, then, as a result of these measurements, we get a single value – the actual value of the measured angle. We can repeat this measurement several times, and, within our assumption, we get the exact same value every time.
- On the other hand, if we take the interval uncertainty into consideration, then after each measurement, instead of a single value of the angle, we get an interval that contains the (unknown) actual value x. If we repeat this measurement several times, we get several intervals each of which contains the actual angle. So, after these measurements, the information that we have about the actual value x is that x belongs to the *intersection* of the intervals corresponding to individual measurements. According to the Theorem, this intersection can be a arbitrary closed set – and thus, it can be much more complex than a single number corresponding to the case when there is no interval uncertainty.

The fact that taking interval uncertainty into consideration leads to an increase in complexity is in line with other similar situations; for example (see, e.g., [10]):

- If the values $x_1, \ldots, x_n$ are known exactly, then computing the value of a given polynomial $f(x_1, \ldots, x_n)$ for these values x_i is a straightforward and easy problem, solvable by known polynomial-time algorithms. On the other hand, if we only know the inputs x_i with interval uncertainty, i.e., if we only know the intervals $\mathbf{x}_i$ of possible values of x_i, then the natural problem is to compute the range $\{f(x_1, \ldots, x_n \,|\, x_1 \in \mathbf{x}_1, \ldots, \mathbf{x}_n\}$ of the given polynomial f on these intervals. The problem of computing such a range is NP-hard even for quadratic polynomials f.

- If we know the exact values of the coefficients a_{ij} and of the right-hand sides b_i, then the problem of solving an $n \times n$ system of linear equations $\sum_j a_{ij} \cdot x_j = b_i$ can be solved by known polynomial-time algorithms. On the other hand, if we only know the intervals $\mathbf{a}_{ij}$ and $\mathbf{b}_i$ of possible values of these coefficients, then, depending on which values $a_{ij} \in \mathbf{a}_{ij}$ and $b_i \in \mathbf{b}_i$ we choose, we get different values of x_j. The problem of computing, for a given j, the range of possible values of x_j is also NP-hard.

Proof of the Theorem. For clarity, we prove this result for the unit circle. One can easily see that this result is true for an arbitrary circle.

Let S be a closed subset of the unit circle C. Then, its complement $-S$ is an open set (see, e.g., [6]). By definition, an open set contains, together with each of its points α, an open ball I_α. On a circle with the above-defined metric d, an open ball is an open interval, so for every point $\alpha \in -S$, there exists an open interval $I_\alpha \subseteq -S$ for which $\alpha \in I_\alpha$.

- Since each of the open intervals I_α is contained in the set $-S$, the union $\bigcup I_\alpha$ of these open intervals is also a subset of $-S$: $\bigcup I_\alpha \subseteq -S$.
- Since every element α of the set $-S$ belongs to the corresponding open interval I_α and therefore, belongs to their union, we can also conclude that $-S$ is a subset of the union $\bigcup I_\alpha$: $-S \subseteq \bigcup I_\alpha$.

Thus, the complement $-S$ is equal to the union $\bigcup I_\alpha$ of the open intervals I_α.

Due to de Morgan's laws, we can now conclude that

$$S = -(-S) = -\left(\bigcup_\alpha I_\alpha\right) = \bigcap_\alpha (-I_\alpha).$$

On a circle, a complement $-I_\alpha$ to an open interval is a closed interval. Therefore, the set S can indeed be represented as an intersection of closed intervals. The theorem is proven.

Comment 1. In this proof, we represent a closed set as an intersection of continuum many open intervals. It is, however, possible to get a similar representation with no more than countably many intervals. Indeed, the open set $-S$ can be represented as a union of its connected components J_α. Each connected component is an open interval. Each component contains a rational number. There are countably many rational numbers, and different components cannot contain the same number; thus, there are at most countably many components. Therefore, $-S$ is a union of at most countably many open intervals J_α. Hence, S is an intersection of at most countably many closed intervals $-J_\alpha$. (The authors are thankful to the anonymous referees for this comment.)

Comment 2. It is worth mentioning that the statement of the theorem is not true if we replace the circle with a real line. The arguments that we used in the proof do not apply to intervals on the real number line because on this line, in contrast to the circle, a complement to an open interval is *not* a closed interval.

5 What Can We Do: Case Study

In the previous section, we have proved that an arbitrary (in particular, arbitrarily complex) closed subset of a circle can be represented as an intersection of intervals.

Handling such complex sets is difficult. It turns out that if we have some statistical information, then the problem often becomes tractable.

5.1 Shape Detection and Why It Is Important

As a case study, we describe an algorithm that uses both real-valued and phase measurement results to determine the shape of a fault; see [12] for details. This shape detection is important: e.g., smooth-shaped faults gather less stress and are, thus, less dangerous than irregularly shaped ones.

Faults are usually detected as outliers, i.e., as points in which the value of some physical quantity are drastically different from the usual values of this quantity. Detecting shapes of regions formed by outlier points is useful in other applications as well; for example:

- in military applications, we want to be able to distinguish between a tank and a heap of rubbish;
- in medical imaging, we must be able to detect the shapes of skin formations: regularly shaped formations are mostly harmless, but the irregularly shaped ones could mean cancer.

As a test case, we used a benchmark B-52 plate provided by Boeing which contains 16 artificially induced smooth-shaped (circular) and angular-shaped (square) faults of 4 different sizes both inside and on the edge: 4 inside squares of sizes $1/2'' \times 1/2''$, $3/8'' \times 3/8''$, $1/4'' \times 1/4''$, and $1/8'' \times 1/8''$, 4 edge squares of the same sizes, 4 inside circles with diameters $1/2''$, $3/8''$, $1/4''$, and $1/8''$, and 4 edge circles of the same diameters. Seven different measurements were done on this plate: two Pulse Echo measurements at different frequencies, four measurements of magnetic resonance, and one measurement of Eddy current. Six of these 7 measurements measure phase (Eddy current is the only exception).

None of these 7 measurements detects all the faults; e.g., Eddy current only detects circular faults, etc. We therefore need to combine ("fuse") the results of these measurements.

5.2 Possible Interval Approach to Data Fusion and Fault Detection: Brief Explanation and Related Difficulties

As we have mentioned, 6 of 7 measurements measure phases. For each point A and for each such measurement $x_i(A)$, we know the upper bound Δ_i on the measurement error, i.e., on the distance $d(x_i(A), \widetilde{x}_i(A))$ between the actual (unknown) value of this phase $x_i(A)$ and the measurement result $\widetilde{x}_i(A)$.

If this upper bound is the only information that we have about the measurement error, and we have no information about the probabilities of different possible values of measurement error, then the only information that we have about the actual value of $x_i(A)$ is that $x_i(A)$ belongs to the interval $\{x \,|\, d(x, \widetilde{x}_i(A)) \le \Delta_i\}$, the interval that we, in the previous section, denoted by $[\widetilde{x}_i(A) - \Delta_i, \widetilde{x}_i(A) + \Delta_i]$.

In many practical cases, the measurement error is close to π, so the resulting interval is close to the entire circle. In such cases, before the measurement, we know that the phase is somewhere on this circle; after the measurement, all we added to this original knowledge is that we excluded values from a small portion of this circle as impossible. This exclusion does not add much knowledge, so no wonder that very little can be deduced from the results of such measurements.

To bring in more information, we can perform several measurements of the same phase-valued quantities. Since every measurement adds a little bit of information, we can expect that after performing sufficiently many independent measurements, we will gather enough information to make meaningful conclusions about the faults.

Theoretically, this conclusion sounds reasonable, but in practice, when we tried this approach, we encountered the problem that we described in Section 4. Namely, as a result of each measurement, we get an interval that covers almost the while circle. After two measurements, we get two such intervals, so we can conclude that the actual value of $x_i(A)$ belongs to their intersection. As we have mentioned, this intersection often consists of two disconnected intervals – the union of which still covers almost the entire circle. After the third measurement, we get one more almost circular interval, and the intersection often further increases the number of disconnected components that form the set $X_i(A)$ of possible values of $x_i(A)$.

In short, the more measurements we undertake, the more accuracy we want, the more complex the resulting set becomes. In principle, it is possible to *describe* such a set, but it is extremely difficult to use the information that $x_i(A) \in X_i(A)$ in any data processing algorithm. Due to the huge number of components, this information simply means that $x_i(A)$ belongs to one of the many components.

Traditional data processing techniques are ill-equipped for constraints that contain the word "or" between inequalities; in most cases, the best we can do is consider each of these cases separately. We could do that for each individual point A, but the multiple-component phenomenon occurs for numerous points A. So, we have to consider all possible combinations of such components – and this make this approach practically non-feasible.

What can we do? As we have mentioned, the above difficulty occurs if the only information that we have about the measurement errors is the upper bound on the measurement error. In practice, often, in addition to this upper bound, we also have some information about the probabilities of different values of this error – actually, in many cases, the observation data are consistent with the assumption that this measurement error is normally distributed. We will show that this additional statistical information indeed helps – it provides a way out of the complexity caused by the interval approach to angle measurements.

5.3 Existing Methods of Fault Detection

Several statistics-based methods have been proposed that fuse the results of different measurements and thus, detect the faults [3]; the best of the known fusion methods is the following one (see [13] for more detail).

This method is based on the fact that faults can be detected by unusual values of different measured quantities; in statistical terms, we can say that faults can be detected as *outliers*. For each plate, and for each measurement type x_i, the probability distribution of measurement result for regular (non-fault) points is close to Gaussian. As a result, it is natural to declare a point A an outlier if the corresponding value $x_i(A)$ is outside the corresponding 2 sigma interval (or 3 sigma), i.e., for which $|x_i(A) - a_i| > 2 \cdot \sigma_i$, where a_i and σ_i are the mean and standard deviation of the corresponding Gaussian distribution[1].

How can we compute the values a_i and σ_i? If we had a plate with no faults, then we could simply compute a_i as the average of all the values $x_i(a)$, and, accordingly, σ_i as $\sqrt{(1/N) \cdot \sum_A (x_i(A) - a_i)^2}$, where N is the total number of pixels. However, the whole point is that there are faults, and if we take the average of all the values $x_i(a)$, including the fault points A, we get a biased estimate for a_i. To avoid this bias, we can apply an *iterative* method in which we sequentially re-calculate the values a_i and σ_i and also mark points as possible outliers. At first, none of the points are marked. At each step, we:

[1] In this paper, capital letters A, B will denote points on a plate.

- compute the new value of a_i as the average of $x_i(A)$ over all un-marked points, and then compute σ_i as $\sqrt{(1/N) \cdot \sum_A (x_i(A) - a_i)^2}$, where N is the total number of un-marked points;
- then, we additionally mark points for which $|x_i(A) - a_i| > 2 \cdot \sigma$ as outliers.

We stop when no new points are marked.

Based on the resulting estimates of a_i and σ_i, we compute, for every measurement i and for every pixel A, the normalized value $z_i(A) \stackrel{\text{def}}{=} (x_i(A) - a_i)/\sigma_i$. According to normal distribution, the probability that a non-fault point A has value $z_i(A)$ is proportional to $\exp(-(z_i(A))^2)$.

It is reasonable to assume that the measurement results x_i and x_j are statistically independent (if x_j was strongly dependent on x_i, then measuring x_j would not make much sense after we have already measured x_i). In this case, the normalized values are also independent, and so the probability for a non-fault point to have values $(z_1(A), \ldots, z_n(A))$ is proportional to

$$p = \prod_{i=1}^{n} \exp(-(z_i(A))^2) = \exp\left(-\sum_{i=1}^{n} (z_i(A))^2\right).$$

If this probability is very small (smaller than a certain threshold p_0), then this point cannot be a regular point and is, therefore, an outlier. Turning to logarithms, we can transform the criterion $p \le p_0$ into an equivalent form $\sum (z_i(A))^2 \ge t_0$ for some new threshold t_0.

How can we determine this value t_0? For outliers, we have already selected a 2 sigma criterion. According to this criterion, even if we start with a population that is perfectly normally distributed, we will classify 5% of this population as not belonging to this distribution. In other words, even in the absence of any faults, with a normally distributed population of regular points, 5% of these perfectly normal points will be (mis)classified as outliers. It is reasonable to accept a similar 5% criterion for selecting the value t_0. This leads to a $t_0 = n \cdot (1 + 2 \cdot \sqrt{2/n})$. So, we mark a point A a fault if the sum $\sum (z_i(A))^2$ exceeds this threshold t_0.

To make this algorithm work better, we need to make two minor modifications:

- First, we have to process edges separately and the interior of the plate separately. Reason: crudely speaking, faults are points where the plate is thinner; near the edges, it is also drastically thinner, so if we combine the edge pixels with the interior ones, then the entire edge will show as one big fault. We said "crudely speaking" because, in reality, naturally occurring small variations in plate thickness does not cause any trouble; however, a drastic change in thickness – e.g., the change near the edges – does affect the results.
- Second, some of the outlier values $x_i(A)$ are caused by a malfunctioning of the measuring instrument. To avoid marking such points as outliers, we mark a point as fault only if, in addition to the condition $|z_i(A)| > 2$ (corresponding to the two sigma deviation), at least one more mode $j \ne i$ detects an outlier either at this very point A or at some point B in the nearest neighborhood of this point A (i.e., for which $\rho(A, B) \le c$ for some small $c > 0$). If for all other measurements $j \ne i$ in this small neighborhood, we have $|z_j(B)| \le 2$, we dismiss the value $z_i(A)$ as a probable malfunctioning of the measuring instrument. For this pixel, we thus combine only 6 remaining values $z_j(A)$ $(j \ne i)$ instead of the usual seven.

On a test plate and on several other plates with known fault locations, the resulting method detects the faults reasonably well, in the sense that it has a smaller number of false positives (regular points erroneously marked as faults) and false negatives (fault points erroneously marked as regular) than the previously known methods.

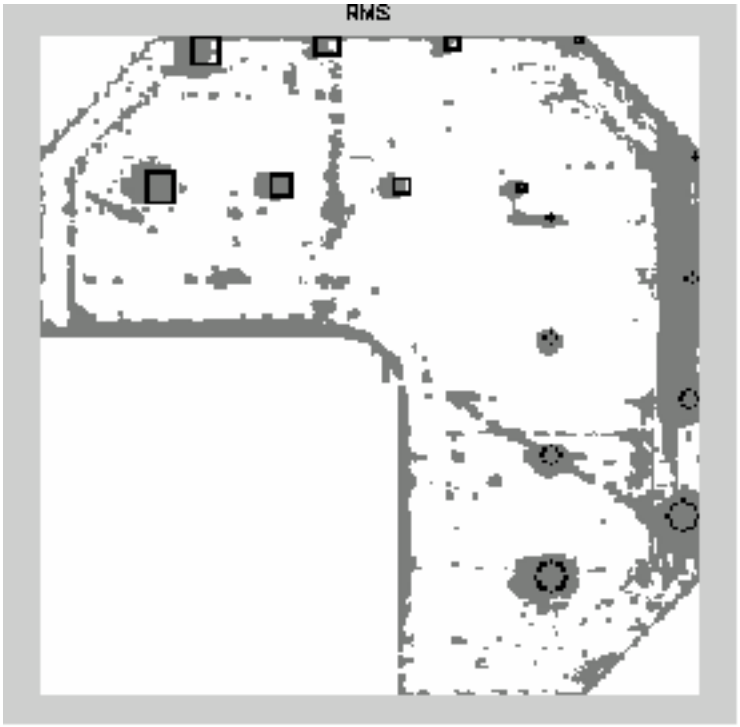


Figure 1: Existing Method

The results of applying this method to the test plate are described in Fig. 1. In this figure, actual faults are outlined in black: by black squares and (broken) circular contours. The gray points are the ones that the algorithm detected as faults.

5.4 Problems with the Existing Methods of Fault Detection

No material is flawless, so we are bound to find some faults. The important question is: how to distinguish really dangerous faults that require repairs (or even a replacement of this part) from the minor ones that do not present an immediate danger (but that may indicate the need for further monitoring).

The fault's degree of danger depends on its size, location, and shape. The dependence on the size is straightforward: larger faults are more dangerous, miniscule faults can be safely ignored. The dependence on the fault's location is also pretty straightforward: faults at the edges are usually dangerous (irrespective of their shape), because when they grow, they can easily get to the edge and thus, provide a serious damage. In contrast, faults inside a plate are sometimes reasonably harmless.

To what extent a fault inside a plant constitutes a danger depends not only on its size but also on its shape: smooth-shaped faults gather less stress and are, thus, less dangerous than angular-shaped ones.

Of these three criteria – location, size, and shape – the existing methods of fault detection detect the location and size reasonably well. However, the *shape* of the reconstructed set of fault

points is not reproduced well: some square faults look like circles and vice versa.

It is therefore desirable to supplement the existing method – which is reasonably good at detecting, locating, and gauging the size of the faults – with an additional algorithm that would better determine the shape of the discovered faults.

In other words, once we applied the original method to detected the faults, we should then use the new method to get a better idea of the shapes of these faults, and thus, to make a decision on whether the plate needs repair or replacement.

5.5 A Supplemental Method for Shape Detection

The main reason why the existing method is not very good in detecting shape in that in the existing method, our main objective was not to miss any faults – because faults are dangerous. Therefore, when there was good evidence to support both hypotheses: that the pixel A is a fault and that the pixel A is not a fault – we tended to declare it a fault. As a result, we "padded" the set of fault points with extra points – thus distorting the shape of the set of all the fault points.

To get the shape better, it is therefore reasonable to treat the two hypotheses equally. Specifically, we consider two hypotheses: H_0 that a point is not a fault and H_1 that the point is a fault, and we use the standard techniques of hypothesis testing (see, e.g., [16]) to decide which of these hypotheses is more probable: we choose H_1 if the probability P_1 of the hypothesis H_1 is larger than the probability P_0 of the hypothesis H_0.

According to Bayes Theorem, the ratio P_1/P_0 is equal to $(p_1 \cdot P_1^{\rm prior})/(p_0 \cdot P_0^{\rm prior})$, where $P_0^{\rm prior}$ and $P_0^{\rm prior}$ are prior probabilities of these hypotheses, p_0 is the probability (density) of the observed data under the hypothesis H_0, and p_1 is the probability (density) of the observed data under the hypothesis H_1. Thus, the criterion $P_1 > P_0$ for choosing H_1 can be reformulated as $P_1/P_0 > 1$, or, equivalently, as $p_1/p_0 > t$, where $t \stackrel{\rm def}{=} P_0^{\rm prior}/P_1^{\rm prior}$.

We already know how to describe the probability p_0 corresponding to H_0. It turns out that the distribution of $z_i(A)$ for fault points is also approximately Gaussian – of course, with different values a_i^f and σ_i^f. (We can estimate the values a_i^f and σ_i^f by processing the points that are marked as outliers after the iterative algorithm for computing a_i and σ_i.) As a result, after taking logarithms of both sides, we can transform the criterion $p_1/p_0 \geq t$ to an equivalent form

$$\sum_{i=1}^{n} \left((z_i(A))^2 - \left(\frac{z_i(A) - a_i^f}{\sigma_i^f} \right)^2 \right) \geq t_0.$$

The same arguments as before lead us to choose the same value for t_0.

Comment. As we have mentioned, from the *practical* viewpoint, it is mostly important to detect the shapes of the faults that are located inside the plate – because all the faults near the edge are dangerous, irrespective of their shape. From the *theoretical* viewpoint, however, it would be nice to be able to detect the shape of edge faults as well. We cannot do that with the above algorithm because to apply this algorithm, we need to have enough fault points to be able to conclusively estimate a_i^f and σ_i^f; there are enough such points on the edge. Thus, if a need would appear to detect shapes of edge faults, new methods must be invented.

5.6 Comparing the Quality of Shape Detection under the Two Methods

In order to find out whether the new method indeed improves the quality of shape detection, we must be able to gauge how well the shape is reconstructed. In other words, we need a numerical

characteristic that describes how well a method captures a shape. In the vicinity of each fault, we have two sets: the true fault F_0 (whose shape, for the test plate, we know), and the set F of all the points that are marked as faults in this neighborhood. How can we compare the true fault F_0 with the reconstructed fault F?

As we have mentioned, our main objective is to reconstruct the *shape* of the fault. Specifically, we want to be able to distinguish between angular and smooth (circular-type) faults. From this viewpoint, if the reconstruction method preserves the fault set F_0 intact but shifts it a little bit as a whole – i.e., if F can be obtained from F_0 by shift – we would say that the shape was preserved perfectly. Similarly, if the reconstructed set F can be obtained from F_0 by scaling, it is natural to say that the shape was preserved.

If we cared not only about the shape but also about the exact location and size of the fault, then it would be natural to gauge the difference between the actual fault F_0 and the reconstructed fault F by counting the total number of false positives and false negatives, i.e., in mathematical terms, the total number of pixels $|F_0 \Delta F|$ in the symmetric difference $F_0 \Delta F$ between the sets F_0 and F.

From our viewpoint, however, this measure of difference is not fuzzy adequate because if F has exactly the same shape as F_0 but differs by a shift, the above distance measure can be large. So, since we do not mind if F and F_0 differ by shift and by scaling, it is natural to define a different measure of distance: instead of taking $|F_0 \Delta F|$, we take the minimum of the values $|T(F_0) \Delta F|$ for all possible combinations T of shifts and scalings. In this case, if F indeed has the same shape as F_0 but differs from it only by a shift and/or a scaling, the resulting distance will be 0. Vice versa, if the resulting metric is 0, it means that the sets F and $T(F_0)$ are identical, i.e., that the reconstructed shape F can indeed be obtained from the actual shape F_0 by some combination T of a shift and a scaling.

On the test plate, we have square faults (whose axes are parallel to coordinate axes) and circular faults. For a square fault F_0, the shifted and scaling also result in a square (with the same direction of axes); moreover, any axes-parallel square can be obtained from F_0 by an appropriate combination of shift and scaling. Therefore, for such faults, sets $T(F_0)$ corresponding to all possible T are exactly all possible squares S whose axes are parallel to the coordinate axes. Thus, for such faults, the above-defined distance is equal to the minimum of the value $|S \Delta F|$ over all possible axes-parallel squares S. Hence, for such faults, to gauge how well the reconstructed fault F reproduced the shape of the original fault, we must find the axes-parallel square S that is the closest to F (in the sense that the total number of pixels in the symmetric set difference is the smallest), and then estimate the number of false positives and false negatives by comparing F and S.

Similarly, for circular faults, any circular disc can be obtained from F_0 by an appropriate shift and scaling. Thus, to gauge the quality of reproducing a circular shape, it is sufficient to compare the set F with the closest circular disc C.

To gauge the overall quality of shape detection, we added the numbers of false positives over all 8 inside faults, and we also added the numbers of false negatives over these faults. Here is the result of our comparison:

- When we applied this procedure to the original method from [13], we got 2,443 false positives and 19 false negatives inside the plate.
- For the new method, we got 1,895 false positives and 11 false negatives inside the plate.

The results of applying the new method to the test plate are given in Fig. 2. In this figure, for each of the 8 faults, in addition to the black contour of the actual fault F_0, we also marked, in black, the contour of the set $S = T(F_0)$ that is the closest to the detected shape F.

Figure 2: New Method

Conclusion: if, after using the original method to detect the faults, we run the new method, we indeed get a better understanding of the shape of the faults.

Acknowledgments

This work was supported in part by NASA under cooperative agreement NCC5-209 and grant NCC2-1232, 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 grants numbers F49620-95-1-0518 and F49620-00-1-0365, by NSF grants CDA-9522207, EAR-0112968, EAR-0225670, and 9710940 Mexico/Conacyt, and by IEEE/ACM SC2001 and SC2002 Minority Serving Institutions Participation Grants.

The authors are thankful to all the participants of SCAN'2002, especially to Bill Walster, for valuable discussions, and to the anonymous referees for important suggestions.

References

[1] A. F. Dravskikh, A. M. Finkelstein, and V. Kreinovich. "Astrometric and geodetic applications of VLBI arc method," *Modern Astrometry, Proceedings of the IAU Colloquium No. 48*, Vienna, 1978, pp. 143–153.

[2] C. Ferregut, R. A. Osegueda, and A. Nuñez (eds.), *Proceedings of the International Workshop on Intelligent NDE Sciences for Aging and Futuristic Aircraft*, El Paso, TX, September 30–October 2, 1997.

[3] X. E. Gros, *NDT Data Fusion*, J. Wiley, London, 1997.

[4] V. S. Gubanov, A. M. Finkelstein, and P. A. Fridman, *Introduction to Radioastrometry*, Leningrad, Nauka, 1983 (in Russian).

[5] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control and Robotics*, Springer-Verlag, London, 2001.

[6] I. Kaplansky, *Set Theory and Metric Spaces*, Chelsea Publ., New York, 1972.

[7] R. B. Kearfott, *Rigorous Global Search: Continuous Problems*, Kluwer, Dordrecht, 1996.

[8] R. B. Kearfott and V. Kreinovich (eds.), *Applications of Interval Computations*, Kluwer, Dordrecht, 1996.

[9] V. Kreinovich, A. Bernat, O. Kosheleva, and A. Finkelstein, "Interval estimates for closure phase and closure amplitude imaging in radio astronomy", *Interval Computations*, 1992, No. 2(4), pp. 51–71.

[10] V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational complexity and feasibility of data processing and interval computations*, Kluwer, Dordrecht, 1997.

[11] R. E. Moore, *Methods and Applications of Interval Analysis*, SIAM, Philadelphia, 1979.

[12] B. C. Mulupuru, *Differentiating between angular and smooth shapes in noisy computer images, on the example of non-destructive testing of aerospace structures*, M.Sc. Thesis, Department of Computer Science, University of Texas at El Paso, 2002.

[13] R. A. Osegueda, S. Seelam, A. Holguin, V. Kreinovich, and C.-W. Tao, "Statistical and Dempster-Shafer Techniques in Testing Structural Integrity of Aerospace Structures", *Int'l J. of Uncertainty, Fuzziness, Knowledge-Based Systems (IJUFKS)*, 2001, Vol. 9, No. 6, pp. 749–758.

[14] S. Rabinovich, *Measurement Errors: Theory and Practice*, American Institute of Physics, New York, 1993.

[15] A. R. Thompson, J. M. Moran, and G. W. Swenson, Jr., *Interferometry and synthesis in radio astronomy*, Wiley, N.Y., 2001.

[16] H. M. Wadsworth, Jr. (eds.), *Handbook of statistical methods for engineers and scientists*, McGraw-Hill Publishing Co., New York, 1990.

[17] G. W. Walster, "Philosophy and practicalities of interval arithmetic", In: *Reliability in Computing*, Academic Press, N.Y., 1988, pp. 309–323.

[18] G. W. Walster and V. Kreinovich, "For unknown-but-bounded errors, interval estimates are often better than averaging", *ACM SIGNUM Newsletter*, 1996, Vol. 31, No. 2, pp. 6–19.

[19] 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 Material Structures", *Reliable Computing*, 2001, Vol. 7, No. 4, pp. 341–352.