1-2000

Aircraft Integrity and Reliability

Carlos M. Ferregut
The University of Texas at El Paso, ferregut@utep.edu

Roberto A. Osegueda
The University of Texas at El Paso

Yohanz Mendoza

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Timothy J. Ross

Comments:
UTEP-CS-00-03.
Published in: Jane Booker, Jerry Parkinson, and Timothy R. Ross (eds.), *Combined Fuzzy Logic and Probability Applications*, SIAM, Philadelphia, 2000, pp. 219-242.

CHAPTER 10
Aircraft Integrity and Reliability

Carlos Ferregut[1], Roberto A. Osegueda[1], Yohans Mendoza[1],
Vladik Kreinovich[1], and Timothy J. Ross[2]

[1]FAST Center for Structural Integrity of
Aerospace Systems
University of Texas at El Paso
El Paso, TX 79968, USA
contact email: vladik@cs.utep.edu
[2]Department of Civil Engineering
University of New Mexico
Albuquerque NM 87131-1351, USA
email ross@unm.edu

Abstract. In his recent paper "Probability theory needs an infusion of fuzzy logic to enhance its ability to deal with real-world problems", L. A. Zadeh explains that probability theory needs an infusion of fuzzy logic to enhance its ability to deal with real-world problems. In this chapter, we give an example of a real-world problem for which such an infusion is indeed successful: the problems of aircraft integrity and reliability.

1 Case Study: Aircraft Structural Integrity. Formulation of the Problem

1.1 Aerospace Testing: Why

One of the most important characteristics of the plane is its weight: every pound shaved off the plane means a pound added to the carrying ability of this plane. As a result, planes are made as light as possible, with their "skin" as thin as possible. However, the thinner the layer, the more vulnerable is the resulting

structure to stresses and faults, and a flight is a very stressful experience. Therefore, even minor faults in the plane's structure, if undetected, can be disastrous. To avoid possible catastrophic consequences, before the flight, we must thoroughly check the structural integrity of the plane.

1.2 Aerospace Testing: How

Some faults, like cracks, holes, etc., are external, and can, therefore, be detected during the visual inspection. However, to detect internal cracks, internal holes, and other internal faults, we must somehow scan the inside of the thin plate that forms the skin of the plane. This skin is not transparent to light or to other electromagnetic radiation; very energetic radiation, e.g., X-rays or gamma-rays, can go through the metal, but it is difficult to use on such a huge object as a modern plane.

The one thing that easily penetrates the skin is vibration. Therefore, we can use sound, ultrasound, etc., to detect the faults. Usually, a wave easily glosses over obstacles whose size is smaller than its wavelength. Therefore, since we want to detect the smallest possible faults, we must choose the sound waves with the smallest possible wavelength, i.e., the largest possible frequency. This frequency is usually higher than the frequencies that we hear, so it corresponds to ultrasound.

Ultrasonic scans are indeed one of the main non-destructive NDE tools; see, e.g, [Chimenti, 1997], [Clough et al., 1986], [Grabec et al., 1997], [Mal et al., 1991], [Viktorov, 1967].

1.3 Aerospace Integrity Testing is Very Time-Consuming and Expensive

One possibility is to have a point-by-point ultrasound testing, the so called S-scan. This testing detects the exact locations and shapes of all the faults. Its main drawback, however, is that since we need to cover every point, we get a very time-consuming (and therefore, very expensive) testing process.

A faster idea is to send waves through the material so that with each measurement, we will be able to test not just a single point, but the entire line between the transmitter and the receiver. To make this procedure work, we need signals at special frequencies called Lamb waves.

There are other testing techniques. All these techniques aim at determining whether there is a fault, and if there are faults, what is the location and the size of each fault. All these methods require a lot of computation time. How can we

speed up the corresponding data processing?

2 Our Main Idea of Solving the Problem

2.1 Main Idea

The amount of data coming from the ultrasonic test is huge, and processing this data takes a lot of time. It is therefore desirable to uncover some structure in the data and to use this structure to speed up the processing of this data.

The first natural idea is to divide the tested structure into pieces and consider these pieces as different clusters. However, physically, the tested piece is a solid body, so the observed vibrations of different points are highly correlated and cannot be easily divides into clusters.

Instead of division in the original space, we propose to make a division, crudely speaking, in frequency domain, i.e., a division into separate different vibration modes.

For each vibration mode, we can estimate the energy density at each point; if this measured energy density is higher than in the original (undisturbed) state, this is a good indication that a fault may be located at this point. The larger the increase in energy density, the larger the probability of a fault. After we get the probabilities related to different modes, we must combine them into an overall probability of having a fault at this particular point.

2.2 Steps Necessary for Implementing the Main Idea

So, our idea leads to following steps:

- first, we must be able to transform the information about the excess energy of each mode at different point into the probabilities of having a fault at the corresponding point;
- second, for each point, we must combine the probabilities coming from different modes into a single probability of a fault.

To make the first step more accurate, we should take into consideration not only the value of the excess energy, but also the value of the original mode-related energy. For example, if the original energy at some point is 0, this means that the vibrations corresponding to this mode have zero amplitude at this point; in other words, this mode does not affect our point at all and therefore, cannot give us any

information about the faults at this point. Similarly, points with small mode energy can give little information about the presence of the fault. So, when computing the probability of a fault in different points based on different modes, we must take into consideration not only the excess energy, but also the original mode energy at this point.

To apply our idea, we must know:

- the function which transforms the value of the excess energy (and of the original mode energy) into the probability of a fault, and
- the combination function which transforms probabilities coming from different modes into a single probability.

2.3 We Do Not Have Sufficient Statistical Data, So We Must Use Expert Estimates

Ideally, we should get all these probability functions from the experiments; however, in real-life, we do not have enough statistics to get reliable estimates for probabilities; we have to complement the statistics with expert estimates. In other words, we must use intelligent methods for non-destructive testing as described, e.g., in [Ferregut et al., 1997].

2.4 Soft Computing

One of the most natural formalisms for describing expert estimates is fuzzy theory, which together with neural networks, genetic algorithms, simulated annealing, etc., form a combined intelligent methodology called soft computing. Therefore, in our chapter, we will be using methods of soft computing (including fuzzy).

2.5 The Choice of Transformation and Combination Functions is Very Important

The quality of fault detection essentially depend on the choice of these methods:

- for some choices of a transformation function and of a combination method, we get a very good fault detection,
- while for others, the quality of detection is much worse.

It is therefore desirable to find the optimal transformation and combination functions.

2.6 How Can We Solve the Corresponding Optimization Problem

This optimization problem is very hard, for two reasons:

- first, due to the presence of expert uncertainty, it is difficult to formulate this problem as a precise mathematical optimization problem;
- second, even when we succeed in formalizing this problem, it is usually a complicated non-linear optimization problem which is extremely difficult to solve by using traditional optimization techniques.

In our previous work (see, e.g., [Nguyen et al., 1997]), we have developed a general methodology for finding the optimal uncertainty representation. In this paper, we show how this general methodology can be used to find the optimal uncertainty representations for this particular problem. Namely:

- the problem of assigning probability to excess energy is solved similarly to problem of finding the best *simulated annealing* technique, and
- the problem of finding the best approximation to the probability of detection (POD) curve (describing the dependence of probability of detection on the mode energy, see [Barbier et al., 1993], [Gros 1997], [Hovey et al., 1988]) is solved similarly to the problem of finding the best activation function in *neural networks*.
- we also have a solution to the probability combination problem; it turns out that under certain reasonable conditions, probability combination methods can be described by so-called Frank's t-norms; this problem is solved using *fuzzy* techniques.

The results of this chapter have been partly published in conference proceedings [Kosheleva et al., 1999], [Krishna et al., 1999], [Osegueda, Mendoza, et al., 1999], [Ross et al., 1999], [Yam et al., 1999].

3 How to Determine Probabilities From Observed Values of Excess Energy: Optimal Way (Use of Simulated Annealing)

3.1 An Expression For Probabilities

For every point x, we estimate the value of the excess energy $J(x)$ at this point. We want to transform these values into the probabilities $p(x)$ that different points contain the fault.

The larger the value of $J(x)$, the more probable it is that the point x contains a fault, i.e., the larger the value $p(x)$. So, a natural first guess would be to take $p(x) = f(J(x))$ for some increasing function $f(z)$.

However, this simple first guess does not work: In many applications, there is usually only one fault. As a result, the total probability $\sum_x p(x)$ that a fault is located somewhere should be equal to 1. If we simply take $p(x) = f(J(x))$, then this condition is not satisfied. In order to satisfy this condition, we must *normalize* the values $f(J(x))$, i.e., consider the probabilities

$$p(x) = \frac{f(J(x))}{\sum_y f(J(y))}. \tag{1}$$

Which function $f(z)$ should we choose? This question is very important, because numerical experiments show that different choices lead to drastically different efficiency of the resulting method; so, to increase detection rate, we would like to choose the best possible function $f(z)$.

3.2 Best In What Sense?

What do we mean by "the best"? It is not so difficult to come up with different criteria for choosing a function $f(z)$:

- We may want to choose the function $f(z)$ for which the resulting fault location error is, on average, the smallest possible $P(f) \to \min$ (i.e., for which the *quality of the answer* is, on average, the best).
- We may also want to choose the function $f(z)$ for which the *average computation time* $C(f)$ is the smallest (average in the same of some reasonable probability distribution on the set of all problems).

At first glance, the situation seems hopeless: we cannot estimate these numerical criteria even for a single function $f(z)$, so it may look like we therefore cannot undertake an even more ambitious task of finding the *optimal* function $f(z)$. Hopefully, the situation is not as hopeless as it may seem, because there is a symmetry-based formalism (actively used in the foundations of fuzzy, neural, genetic computations, see, e.g., [Nguyen et al., 1997]) which will enable us to find the optimal function $f(z)$ for our situation too. (Our application will be mathematically similar to the optimal choice of a non-linear scaling function in genetic algorithms [Kreinovich et al., 1993], [Nguyen et al., 1997].)

Before we make a formal definition, let us make two comments.

The first comment is that our goal is to find probabilities. Probabilities are always non-negative numbers, so the function $f(z)$ must also take only non-negative values.

The second comment is that all we want from the function $f(z)$ is the probabilities. These probabilities are computed according to the formula (1). From this expression (1), one can easily see that if we multiply all the values of this function $f(z)$ by an arbitrary constant C, i.e., if we consider a new function $\widetilde{f}(z) = C \cdot f(z)$, then this new function will lead (after the normalization involved in (1)), to exactly the same values of the probabilities. Thus, whether we choose $f(z)$ or $\widetilde{f}(z) = C \cdot f(z)$, does not matter. So, what we are really choosing is not a *single* function $f(z)$, but a *family* of functions $\{C \cdot f(z)\}$ (characterized by a parameter $C > 0$).

In the following text, we will denote families of functions by capital letters, such as F, F', G, etc.

3.3 An Optimality Criterion Can Be Non-Numeric

Traditionally, optimality criteria are *numerical*, i.e., to every family F, we assign some value $J(F)$ expressing its quality, and choose a family for which this value is minimal (i.e., when $J(F) \leq J(G)$ for every other alternative G). However, it is not necessary to restrict ourselves to such numeric criteria only.

For example, if we have several different families F that have the same average location error $P(F)$, we can choose between them the one that has the minimal computational time $C(F)$. In this case, the actual criterion that we use to compare two families is not numeric, but more complicated: A family F_1 is better than the family F_2 if and only if either $P(F_1) < P(F_2)$, or $P(F_1) = P(F_2)$ and $C(F_1) < C(F_2)$.

The only thing that a criterion *must* do is to allow us, for every pair of families (F_1, F_2), to make one of the following conclusions:

- the first family is better with respect to this criterion (we'll denote it by $F_1 \succ F_2$, or $F_2 \prec F_1$);
- with respect to the given criterion, the second family is better ($F_2 \succ F_1$);
- with respect to this criterion, the two families have the same quality (we'll denote it by $F_1 \sim F_2$);
- this criterion does not allow us to compare the two families.

Of course, it is necessary to demand that these choices be consistent. For example, if $F_1 \succ F_2$ and $F_2 \succ F_3$ then $F_1 \succ F_3$.

3.4 Optimality Criterion Must Be Final

A natural demand is that this criterion must choose a *unique* optimal family (i.e., a family that is better with respect to this criterion than any other family). The reason for this demand is very simple.

If a criterion *does not choose* any family at all, then it is of no use.

If *several* different families are the best according to this criterion, then we still have the problem of choosing the best among them. Therefore we need some additional criterion for that choice, like in the above example: If several families $F_1, F_2, \ldots$ turn out to have the same average location error ($P(F_1) = P(F_2) = \ldots$), we can choose among them a family with minimal computation time ($C(F_i) \to \min$).

So what we actually do in this case is abandon that criterion for which there were several "best" families, and consider a new "composite" criterion instead: F_1 is better than F_2 according to this new criterion if either it was better according to the old criterion, or they had the same quality according to the old criterion and F_1 is better than F_2 according to the additional criterion.

In other words, if a criterion does not allow us to choose a unique best family, it means that this criterion is not final, we'll have to modify it until we come to a final criterion that will have that property.

3.5 The Criterion Must Not Change If We Change the Measuring Unit for Energy

The exact mathematical form of a function $f(z)$ depends on the exact choice of units for measuring the excess energy $z = J(x)$. If we replace this unit by a new unit that is λ times larger, then the same physical value that was previously described by a numerical value $J(x)$ will now be described, in the new units, by a new numerical value $\widetilde{J}(x) = J(x)/\lambda$.

How will the expression for $f(z)$ change if we use the new units? In terms of $\widetilde{J}(x)$, we have $J(x) = \lambda \cdot \widetilde{J}(x)$. Thus, if we change the measuring unit for energy, the same probabilities $p(x) \sim f(J(x))$ that was originally represented by a function $f(z)$, will be described, in the new units, as $p(x) \sim f(\lambda \cdot \widetilde{J}(x))$, i.e., as $p(x) \sim \widetilde{f}(\widetilde{J}(x))$, where $\widetilde{f}(z) = f(\lambda \cdot z)$.

There is no reason why one choice of unit should be preferable to the other. Therefore, it is reasonable to assume that the relative quality of different families should not change if we simply change the units, i.e., if the family F is better than a family G, then the transformed family $\widetilde{F}$ should also be better than the family $\widetilde{G}$.

We are now ready for the formal definitions.

3.6 Definitions and the Main Result

Definition 1. *Let $f(z)$ be a differentiable strictly increasing function from real numbers to non-negative real numbers. By a family that corresponds to this function $f(z)$, we mean a family of all functions of the type $\widetilde{f}(z) = C \cdot f(z)$, where $C > 0$ is an arbitrary positive real number. (Two families are considered equal if they coincide, i.e., consist of the same functions.)*

In the following text, we will denote the set of all possible families by Φ.

Definition 2. *By an optimality criterion, we mean a consistent pair $(\prec, \sim)$ of relations on the set Φ of all alternatives which satisfies the following conditions, for every $F, G, H \in \Phi$:*

(1) *if $F \prec G$ and $G \prec H$ then $F \prec H$;*
(2) *$F \sim F$;*
(3) *if $F \sim G$ then $G \sim F$;*
(4) *if $F \sim G$ and $G \sim H$ then $F \sim H$;*
(5) *if $F \prec G$ and $G \sim H$ then $F \prec H$;*
(6) *if $F \sim G$ and $G \prec H$ then $F \prec H$;*
(7) *if $F \prec G$ then $G \not\prec F$ and $F \not\sim G$.*

Comment. The intended meaning of these relations is as follows:

- $F \prec G$ means that with respect to a given criterion, G is better than F;
- $F \sim G$ means that with respect to a given criterion, F and G are of the same quality.

Under this interpretation, conditions (1)–(7) have simple intuitive meaning; e.g., (1) means that if G is better than F, and H is better than G, then H is better than F.

Definition 3.

- *We say that an alternative F is optimal (or best) with respect to a criterion $(\prec, \sim)$ if for every other alternative G either $F \succ G$ or $F \sim G$.*
- *We say that a criterion is final if there exists an optimal alternative, and this optimal alternative is unique.*

Definition 4. *Let $\lambda > 0$ be a positive real number.*

- *By a λ-rescaling of a function $f(x)$ we mean a function $\widetilde{f}(x) = f(\lambda \cdot x)$.*
- *By a λ-rescaling $R_\lambda(F)$ of a family of functions F we mean the family consisting of λ-rescalings of all functions from F.*

Definition 5. *We say that an optimality criterion on Φ is unit-invariant if for every two families F and G and for every number $\lambda > 0$, the following two conditions are true:*

i) if F is better than G in the sense of this criterion (i.e., $F \succ G$), then $R_\lambda(F) \succ R_\lambda(G)$;

ii) if F is equivalent to G in the sense of this criterion (i.e., $F \sim G$), then $R_\lambda(F) \sim R_\lambda(G)$.

Theorem 1. *If a family F is optimal in the sense of some optimality criterion that is final and unit-invariant, then every function $f(z)$ from this family F has the form $C \cdot z^\alpha$ for some real numbers C and α.*

Comment. Experiments show that for non-destructive testing, the best choice is $\alpha \approx 1$.

4 How to Determine the Probability of Detection (POD): Optimal Way (Use of Neural Networks)

In the previous section, we assumed that the probability of having a fault at a certain point x depends only on the value of the excess energy $J(x)$ at this point x. This assumption, however, is not always true. For example, if at some point x, the energy $E(x)$ of the original (no-fault) vibration mode is equal to 0, this means that this point does not participate in this vibration mode at all; therefore, from observing this vibration mode, we cannot deduce whether there is a fault at this point or not. Similarly, if the strain energy $E(x)$ is small, this means that this point x is barely moving and hardly participating in the vibration. Therefore, this vibration mode is barely affected by the presence or absence of the fault at this point.

With this in mind, we can say that the probability $p(x)$ which we described in the previous section is not the "absolute" probabilities of a fault at a point x based on this mode, but rather a *conditional* probability that there is fault $P(fault|C)$ – under the condition C that the analysis of this mode can detect a fault at a point x. The actual probability $P(fault)$ of detecting a fault from the measurements related to this mode can be therefore computed as a product $P(fault) = P(fault|C) \cdot P(C)$.

We already know how to compute $P(fault|C)$. Therefore, to compute the desired probability $P(fault)$, we must find the probability $P(C)$ that the analysis of this mode can detect a fault at a given point x. We have already mentioned that this detection probability depends on the energy $E(x)$ of the original vibration mode at the point x: $P(fault) = p(E(x))$: if $E = 0$, then $p(E) = 0$; if E is small, then $p(E)$ is small; if E is large enough, then $p(E)$ is close to 1. Let us describe this Probability of Detection (POD) dependence $p(E)$.

Our application will be mathematically similar to the optimal choice of an activation function in neural networks [Nguyen et al., 1997].)

4.1 The POD Function Must Be Smooth and Monotonic

If we change the energy E slightly, the probability $p(E)$ of detecting the fault should not change drastically. Thus, we expect the dependency $p(E)$ to be *smooth* (differentiable).

For a POD function, the probability of detection should be equal to 0 when the point is not affected by the vibration ($E = 0$) and should be equal to 1 when the point is highly affected ($E \to \infty$). The larger the energy E, the more probable it is that we will be able to find the fault; thus, the dependence $p(E)$ should be

monotonic.

4.2 We Must Choose a Family of Functions, Not a Single Function

For practical applications, we need the function $p(E)$ which would determine the probability that if a point with an energy E is presented to a certain NDE technique, then the corresponding fault will be detected. In order to determine this function empirically, we must have a statistics of samples which were presented to this techniques and for which, later on, the fault was discovered; from this statistics, we can determine the desired probability.

This probability, however, depends on how we select the samples presented to the NDE techniques. For example, most structures are inspected visually before using a more complicated NDE technology. Some aerospace structures are easier to inspect visually, so we can detect more faults visually, and only harder-than-usual faults are presented to the NDE technique; as a result of this pre-selection, for such structures, the success probability $p(E)$ is lower than in other cases. Other structures are more difficult to inspect visually; for these structures, all the faults (including easy-to-detect ones) are presented to the NDE techniques, and the success probabilities $p(E)$ will be higher.

In view of this pre-selection, for one and the same NDE technique we may have different POD functions depending on which structures we apply it to. So, instead of looking for a *single* function $p(E)$, we should look for a *family* of POD functions which correspond to different pre-selections.

How are different functions from this family related to each other? Pre-selection means, in effect, that we are moving from the original unconditional detection probability to the conditional probability, under the condition that this particular sample has been pre-selected. In statistics, the transformation from an unconditional probability $P_0(H_i)$ of a certain hypothesis H_i to its conditional probability $P(H_i|S)$ (under the condition S that a sample was pre-selected) is described by the Bayes formula

$$P(H_i|S) = \frac{P(S|H_i) \cdot P_0(H_i)}{\sum_j P(S|H_j) \cdot P_0(H_j)}.$$

In mathematical terms, the transformation from $p(E) = P_0(H_i)$ to $\widetilde{p}(E) = P(H_i|S)$ is *fractionally linear*, i.e., has the form $p(E) \to \widetilde{p}(E) = \varphi(p(E))$, where

$$\varphi(y) = \frac{k \cdot y + l}{m \cdot y + n}$$

for some real numbers k, l, m, and n. So, instead of looking for a single function $p(E)$, we should look for a family of functions $\{\varphi(p(E))\}$, where $p(E)$ is a fixed

function and $\varphi(y)$ are different fractionally linear transformations. In the following text, when we say "a family of functions", we will mean a family of this very type.

Similarly to the previous section, we are looking for a family which is optimal with respect to some final optimality criterion, and it is reasonable to require that this criterion should not change if we change the measuring unit for energy. Thus, we arrive at the following definitions:

4.3 Definition and the Main Result

Definition 6.

- *By a probability function, we mean a smooth monotonic function $p(E)$ defined for all $E \geq 0$ for which $p(0) = 0$ and $p(E) \to 1$ as $E \to \infty$.*
- *By a family of functions we mean the set of functions that is obtained from a probability function $p(E)$ by applying fractionally linear transformations.*

Theorem 2. *If a family F is optimal in the sense of some optimality criterion that is final and unit-invariant, then every function p from the family F is equal to*

$$p(E) = \frac{A \cdot E^\beta}{A \cdot E^\beta + 1} \tag{2}$$

for some A and $\beta > 0$.

5 How To Combine Probabilities (Use Of Fuzzy Techniques)

Based on the formulas given in the previous two sections, we can determine, for each point x and for each mode i, the probability $p_i(x)$ that, based on this mode, there is a fault at this point x. Since there are several modes, we must therefore *combine*, for each point x, these probabilities $p_i(x)$ into a single probability $p(x)$ that there is a fault structure at this point. In this section, we show how different techniques can help to find the best combination.

5.1 Traditional Probabilistic Approach: Maximum Entropy

For each point x, after we get the probabilities $p_i(x)$ ($1 \leq i \leq n$) that this point has a fault in it, we must combine these n probabilities into a single probability $p(x)$ that there is a fault. A fault is there if it is detected by one of the modes, i.e., if it is detected by the 1-st mode, *or* detected by the 2-nd mode, etc. In other words, if $p_i(x)$ is a probability of the event D_i "a fault at point x is detected by i-th mode", then $p(x)$ is the probability of the disjunction $D_1 \vee \ldots \vee D_n$.

In general, if we only know the probabilities of events D_i, then it is not possible to uniquely determine the probability of the disjunction; to select a unique probability, we use a *maximum entropy* approach. The idea of this approach is as follows: to find the probabilities of all possible logical combinations of the events D_i, it is sufficient to determine the probabilities of all 2^n events W of the type $D_1^{\varepsilon_1} \& \ldots \& D_n^{\varepsilon_n}$, where $\varepsilon_i \in \{-, +\}$, D_i^+ means D_i, and D_i^- means its negation $\neg D_i$. So, we must determine the probabilities $p(W)$ so that $\sum p(W) = 1$, and for every i, the sum of $p(W)$ for all events for which D_i is true is equal to $p_i(x)$. There are many distribution of this type; we select the one for which the entropy $-\sum p(W) \cdot \log(p(W))$ takes the largest possible value.

For this distribution, the desired probability of the disjunction is equal to $p(x) = 1 - \prod_{i=1}^{n}(1 - p_i(x))$ (see, e.g., [Kreinovich et al., 1996]).

5.2 Traditional Approach Is Not Always Sufficient

From the statistical viewpoint, the MaxEnt formula corresponds to the case in which all modes are statistically independent. In reality, the detection errors in different modes can be caused by the same cause and therefore, these are not necessarily independent.

If we knew the correlation between these errors, then we could use traditional statistical methods to combine the probabilities $p_i(x)$. In reality, however, we typically do not have sufficient information about the correlation between the components. So, we need to find a new method of fusing probabilities $p_1(x), \ldots, p_n(x)$ corresponding to different modes.

5.3 Main Idea: Describe General Combination Operations

We must choose a method for combining probabilities. In mathematical terms, we must describe, for every n, a function, which, takes n numbers

$p_1, \ldots, p_n \in [0,1]$ as inputs and returns the "fused" (combined) probability $f_n(p_1, \ldots, p_n)$.

This description if further complicated by the fact that the division into modes is rather subjective; e.g., for three close modes, we can

- either divide the vibration into these three modes 1, 2, and 3;
- or divide the vibration into two mode 3 and a "macromode" $\{1, 2\}$ combining modes 1 and 2,
- or divide it into 1 and $\{2, 3\}$.

Depending on the division, we get different expressions for the resulting probability:

- If we divide the vibration into three modes, the resulting probability is $p = f_3(p_1, p_2, p_3)$.
- If we divide the vibration into two macromodes $\{1, 2\}$ and 3, then for the first macromode, we get $p_{12} = f_2(p_1, p_2)$, and thus, $p = f_2(p_{12}, p_3) = f_2(f_2(p_1, p_2), p_3)$.
- Similarly, if we divide the vibration into two macromodes 1 and $\{2, 3\}$, then for the second macromode, we get $p_{23} = f_2(p_2, p_3)$, and thus, $p = f_2(p_1, p_{23}) = f_2(p_1, f_2(p_2, p_3))$.

The resulting probability should not depend on the (subjective) subdivision into modes. As a result, we should get $f_3(p_1, p_2, p_3) = f_2(f_2(p_1, p_2), p_3) = f_2(p_1, f_2(p_2, p_3))$. In other words, we can make two conclusions:

- First, the combination function for arbitrary $n > 2$ can be expressed in terms of a combination function corresponding to $n = 2$, as $f_n(p_1, \ldots, p_n) = f_2(p_1, f_2(p_2, \ldots, f_2(p_{n-1}, p_n) \ldots))$.
- Second, the function f_2 which describes the combination of two probabilities should be *associative* (i.e., $f_2(a, f_2(b, c)) = f_2(f_2(a, b), c)$ for all $a, b, c \in [0, 1]$).

Associativity is a reasonably strong property, but associativity alone is not sufficient to determine the operation $f_2(a, b)$, because there are many different associate combination operations. Hopefully, there is another property which we can use. To describe this property, let us recall that for fault detection, the function $f_2(p_1, p_2)$ has the following meaning:

- p_1 is the probability $P(D_1)$ of the event D_1 defined as "the first mode detected a fault";
- p_2 is the probability $P(D_2)$ of the event D_2 defined as "the second mode detected a fault"; and

- $f_2(p_1, p_2)$ is the probability $P(D_1 \vee D_2)$ that one of the two modes detected a fault.

So far, our main goal was to detect a fault; this problem is difficult because faults are small and therefore, they only show up on one of the modes. Different faults present different danger to the aerospace structure: small fault can be potentially dangerous, but if they are small enough they may not require grounding the plane. On the other hand, large faults are definitely dangerous. Therefore, in addition to detecting all the faults, we would like to know which of them are large (if any). A large fault is probably causing strong changes in all vibration modes. Thus, we can expect a large fault if we have detected a fault in all the modes. In other words, in this problem, we are interested in the value $P(D_1 \& D_2)$. From the probability theory, we know that $P(D_1 \& D_2) = P(D_1) + P(D_2) - P(D_1 \vee D_2)$, i.e., $P(D_1 \& D_2) = p_1 + p_2 - f_2(p_1, p_2)$. We can describe this expression as $P(D_1 \vee D_2) = g_2(p_1, p_2)$, where $g_2(a, b) = a + b - f_2(a, b)$.

Similar to the case of "or"-combination of different components, we can describe the probability $P(D_1 \& D_2 \& D_3)$ (that all three modes detect a fault) in two different ways:

- either as $P((D_1 \& D_2) \& D_3) = g_2(g_2(p_1, p_2), p_3)$,
- or as $P(D_1 \& (D_2 \& D_3)) = g_2(g_2(p_1, p_2), p_3)$.

The expression for $P(D_1 \& D_2 \& D_3)$ should not depend on how we compute it, and therefore, we should have $g_2(g_2(p_1, p_2), p_3) = g_2(p_1, g_2(p_2, p_3))$. In other words, not only the function $f_2(a, b)$ should be associative, but also the function $g_2(a, b) = a + b - f_2(a, b)$ should be associative.

5.4 The Notions of T-Norms and T-Conorms

In the above text, we described everything in terms of combining probabilities. However, from the mathematical viewpoint, the resulting requirements were exactly the requirements traditionally used for combining membership values in fuzzy approach (see, e.g., [Klir Yuan, 1995], [Nguyen et al., 1999]):

- The function $f_2(p_1, p_2)$ which describes the degree of truth of the statement $D_1 \vee D_2$, provided that we know the degrees of truth for statement D_1 and D_2, is a *t-conorm*.
- The function $g_2(p_1, p_2)$ which describes the degree of truth of the statement $D_1 \& D_2$, provided that we know the degrees of truth for statement D_1 and D_2, is a *t-norm*.

5.5 Main Result: Frank's T-Norms

Functions $f(a,b)$ for which both the function itself and the expression $g(a,b) = a + b - f(a,b)$ are associative have been classified in [Frank, 1979]. These functions (called *Frank's t-conorms*) are described by a formula

$$f_2(a,b) = 1 - \log_s \left[1 + \frac{(s^{1-a}-1)\cdot(s^{1-b}-1)}{s-1}\right],$$

the corresponding t-norms are:

$$g_2(a,b) = \log_s \left[1 + \frac{(s^{a}-1)\cdot(s^{b}-1)}{s-1}\right],$$

for some constant s. As a particular case of this general formula, we get the expression $1-(1-a)\cdot(1-b)$ (for $s \to 1$).

Thus, to combine the probabilities $p_i(x)$ coming from different modes, we should use Frank's t-conorms.

Comment. Fuzzy techniques have been successfully used in nondestructive testing in particular (for a latest survey, see, e.g., [Ferregut et al. 1997]) and in damage assessment in general (see, e.g., [Terano et al., 1987], [Ulieru et al., 1993]).

The fact that for probabilistic data, we get similar formulas, makes us hope that this algebraic approach will be able to combine probabilistic and fuzzy data.

6 Preliminary Results

As a case study, we applied the new method to the problem of non-destructive evaluation of structural integrity of Space Shuttle's vertical stabilizer. To prove the applicability of our method, we applied this techniques to measurement results for pieces with known fault locations.

The value s was determined experimentally so as to achieve the best performance; it turned out that the best value is $s \approx 1$ (corresponding to the independent modes). For this value s, our method detected all the faults in $\approx 70\%$ of the cases, much larger proportion than with any previously known techniques (for details, see [Andre, 1999], [Osegueda, Revilla, 1999], [Pereyra et al., 1999], [Stubbs et al., 1998]).

For other values of s, we got an even better detection, but at the expense of false alarms.

We are currently trying different data fusion techniques (as described, e.g., in [Gros, 1997]) to further improve the method's performance.

7 Alternative Approach to Fusing Probabilities: Fuzzy Rules

7.1 Main Problems with the Above Approach

There are two main problems with the above approach:

- first, due to the fact that we used several different (and reasonably complicated) formalisms, the resulting computational models are rather time-consuming and not very intuitive;
- second, although we got better fault detection that all previously known methods, but there is a still quite some room for improvement.

7.2 The Use of Fuzzy Rules

The main problem we face is the problem of complexity of the computational models we use. Complex models are justified in such areas as fundamental physics, when simpler first approximation models have been tried and turned out not exactly adequate. However, in our case, the computational models are chosen not because simpler models have been tried, but because these complex models were the only ones which we could find which fit our data and are consistent with the expert knowledge.

The very fact that a large part of our knowledge comes from expert estimates, which have a high level of uncertainty, makes us believe that within this uncertainty, we can find simpler computational models which will work equally well. How can we find such models?

A similar situation, when unnecessarily complex models were produced by the existing techniques, started the field of fuzzy logic. Namely, L. Zadeh proposed to use, instead of traditional analytical models, new simplified models based on the direct formalization of expert's knowledge.

In view of the success of fuzzy techniques, it is reasonable to use a similar approach in fault detection as well. Let us first describe the corresponding rules.

7.3 Expert Rules for Fault Detection

For each location, as a result of the measurements, we get five different values of the excess energy $E_1, \ldots, E_5$ which correspond to 5 different modes. An expert can look at these values and tell whether we have a definite fault here, or a fault with a certain degree of certainty, or definitely no fault at all.

Before we formulate the expert rules, we should note that for each node, the *absolute* values of excess energy are not that characteristic because, e.g., a slight increase or decrease in the original activation can increase or decrease all the values of the excess energy, while the fault locations remain the same. Therefore, it is more reasonable to look at *relative* values of the excess energy. Namely, for each mode i, we compute the mean square average σ_i of all the values, and then divide all values of the excess energy by this means square value to get the corresponding relative value of the excess energy $x_i = J_i/\sigma_i$.

In accordance with the standard fuzzy logic methodology, we would like to describe some of these values as "small positive" (SP), some as "large positive" (LP), etc. To formalize these notions, we must describe the corresponding membership functions $\mu_{SP}(x)$ and $\mu_{LP}(x)$.

Some intuition about the values x_i comes from the simplified situation in which the values of excess energy J_i are random, following a normal distribution with 0 average. In this simplified situation, the mean square value σ_i is (practically) equal to the standard deviation of this distribution. For normal distributions, deviations which exceed $2\sigma_i$ are rare and are therefore usually considered to be definitely large; on the hand, deviations which are smaller than the average σ_i are, naturally, definitely small. Deviations $J_i \geq 2\sigma_i$ correspond to the values $x_i = J_i/\sigma_i \geq 2$, and deviations $J_i \leq \sigma_i$ correspond to $x_i = J_i/\sigma_i \leq 1$. Therefore, can conclude that values $x_i \geq 2$ are definitely large, and positive values $x_i \leq 1$ are definitely small.

So, for the fuzzy notion "small", we know that:

- values from 0 to 1 are definitely small, i.e., $\mu_{SP}(x_i) = 1$ for these values, and
- values 2 and larger are definitely not small, i.e., $\mu_{SP}(x_i) = 0$ for these values.

These formulas determine the value of the membership function for all positive values of x_i, except for the values from 1 to 2. In accordance with the standard fuzzy techniques, we use the simplest – linear – interpolation to define $\mu_{SP}(x_i)$ for values from this interval, i.e., we take $\mu_{SP}(x_i) = 2 - x_i$ for $x_i \in [1, 2]$.

Similarly, we define the membership function for "large" as follows: $\mu_{LP}(x_i) = 0$ for $x_i \in [0, 1]$; $\mu_{LP}(x_i) = x_i - 1$ for $x_i \in [1, 2]$; and $\mu_{LP}(x_i) = 1$

for $x_i \geq 2$.

Similarly, we describe the membership functions corresponding to "small negative" (SN) and "large negative" (LN): in precise terms, for $x_i < 0$, we set $\mu_{SN}(x_i) = \mu_{SP}(|x_i|)$ and $\mu_{LN}(x_i) = \mu_{LP}(|x_i|)$.

This takes care of fuzzy terms used in the condition of expert rules. To describe the conclusion, we determined that experts use 5 different levels of certainty, from level 1 to level 5 (absolute certainty). We can identify these levels with numbers from 0.2 to 1.

Now, we are ready to describe the rules.

1. *If the "total" excess energy $x_1 + \ldots + x_5$ attains its largest possible value, or is close to the largest possible value (by ≤ 0.06), then we definitely have a fault at this location (this conclusion corresponds to level 5).*

2. *If all 5 modes show increase, then we have a level 4 certainty that there is a fault at this location.*

3. *If 4 modes show increase, and one mode shows small or large decrease, then level 4.*

4. *If 3 modes show increase and 2 show small decreases then level 4.*

5. *If 3 modes show increase, and we have either 1 small and 1 large decrease, or 2 large decreases, then level 3.*

6. *If 2 modes show large increase and 3 modes show small decrease, then level 3.*

7. *If 2 modes show large increase, 1 or 2 modes show large decrease, and the rest show decrease, then level 2.*

8. *If 1 mode shows large increase, 1 mode shows small increase, and 3 modes show small decrease, then level 2.*

9. *In all other cases, level 1.*

7.4 The Problem With This Rule Base and How We Solve It

The technique of fuzzy modeling and fuzzy control enables us to translate rule bases (like the one above) into an algorithm which transforms the inputs $x_1, \ldots, x_n$ into a (defuzzified) value of the output y. In principle, we can apply this technique to our rule base, but the problem is that we will need too many rules. Indeed, standard rules are based on the conditions like "if x_1 is A_1, ..., and x_n is A_n, then y is B". In our case, we have 5 input variables, each of which can take 4 different fuzzy values (LN, SN, SP, and LP). So, to describe all possible combinations of inputs, we must use $4^5 = 1,024$ rules. It is doable, but it is definitely not the simplification for which we were looking.

To decrease the number of the resulting rules, we can use the fact that all the rules do not distinguish between different modes. Therefore, if we permute

the values x_i (e.g., swap the values x_1 and x_2), the expert's conclusion will not change. Hence, instead of considering all possible combinations of x_i, we can first apply some permutation to decrease the number of possible combinations. One such permutation is *sorting* the values of x_i, i.e., re-ordering these values in the decreasing order. Let us show that if we apply the rules to thus re-ordered values, then we can indeed drastically decrease the number of resulting fuzzy rules.

Let $y_1 \geq y_2 \ldots \geq y_5$ denote the values $x_1, \ldots, x_5$ re-ordered in decreasing order. Let us show how, e.g., Rules 2, 3, and 4 from the above rule base can be reformulated in terms of these new values y_i:

Rule 2. To say that all five values x_i are positive is the same as to say that the smallest of these values is positive, so the condition of Rule 2 can be reformulated as $y_5 > 0$.

Rule 3. When 4 modes are positive and the fifth is negative, it means that $y_4 > 0$ and $y_5 < 0$.

We can notice that since Rules 2 and 3 have the same conclusion, they can be combined into a single rule with a new (even simpler) condition $y_4 > 0$. (Indeed, we either have $y_5 > 0$ and $y_5 \leq 0$; if $y_4 > 0$ and $y_5 > 0$, then the conclusion is true because of Rule 2; if $y_4 > 0$ and $y_5 < 0$, then the conclusion is true because of Rule 3.)

Rule 4. Similarly, its condition can be reformulated as $y_3 > 0$, y_4 is SN, and y_5 is SN.

As a result, we get the following new (simplified) rule base:

1. *If the "total" excess energy $y_1 + \ldots + y_5$ attains its largest possible value, or is close to the largest possible value (by ≤ 0.06), then level* 5.
2. *If $y_4 > 0$, then level* 4.
3. *If $y_3 > 0$, y_4 is SN, and y_5 is SN, then level* 4.
4. *If $y_3 > 0$, $y_4 < 0$, and y_5 is LN, then level* 3.
5. *If y_2 is LP, $y_3 < 0$, and y_5 is SN, then level* 3.
6. *If y_2 is LP, $y_3 < 0$, and y_5 is LN, then level* 2.
7. *If y_1 is LP, y_2 is SP, $y_3 < 0$, y_4 is SN, and y_5 is LN, then level* 2.
8. *In all other cases, level* 1.

To transform these fuzzy rules into a precise algorithm, we must select a fuzzy "and"-operation (t-norm) and a fuzzy "or"-operation (t-conorm), e.g., $\min(a, b)$ and $\max(a, b)$, and a defuzzification; in this paper, we use centroid defuzzification.

For each rule (except for the last one), we can compute the degree of satisfaction for each of the conditions. The rule is applicable if its first condition holds, *and* the second condition holds, etc. So, to find the degree with which the

rule is applicable, we apply the chosen "and"-operation to the degrees with which different conditions of this rule hold.

For each level > 1, we have two rules leading to this level. The corresponding degree of certainty is achieved if either the first *or* the second of these rules is applicable. Therefore, to find a degree to which this level is justified, we must apply the chosen "or"-operation to the degrees to which these two rules are applicable.

As a result, we get the degrees $d(l)$ with which we can justify levels $l = 2 \div 5$. Since the last rule (about level 1) says that this rule is applicable when no other rule applies, we can compute $d(1)$ as $1 - d(2) - \ldots - d(5)$. Now, centroid defuzzification leads to the resulting certainty $1 \cdot d(1) + 2 \cdot d(2) + \ldots + 5 \cdot d(5)$. This is the value that the system outputs as the degree of certainty (on a 1 to 5 scale) that there is a fault at a given location.

7.5 Experimental Results

We have applied the resulting fuzzy model to the beams with known fault locations. The results are as follows:

When there is *only one fault*, this fault can be determined as the location where the degree of certainty attains its largest value 5. This criterion leads to a *perfect fault localization*, with no false positives and no false negatives.

When there are *several faults*, all the faults correspond to locations with degree 4 or larger. This criterion is not perfect; it *avoids the most dangerous errors of false negatives* (i.e., all the faults are detected), but it has false positives, i.e., sometimes faults are wrongly indicated in the areas where there are none.

To make the fuzzy algorithm better, we take into consideration that the vibration corresponding to each mode has points in which the amplitude of this vibration is 0. The corresponding locations are not affected by this mode and therefore, the corresponding excess energy values cannot tell anything about the presence or absence of a fault. Therefore, it makes sense to only consider those values x_i for which the corresponding mode energy is at least, say, 10% of its maximum. If we thus restrict the values x_i, then the number of false positives decreases.

Similar results fold for 2D cases.

We tried different t-norms and t-conorms. So far, we have not found a statistically significant difference between the results obtained by using different t-norms and t-conorms; therefore, we recommend to use the simplest possible operations: $\min(a, b)$ and $\max(a, b)$.

8 Applications to Aircraft Reliability

A similar approach works not only for fault detection, but also for another important problem – reliability.

8.1 Reliability: General Problem

A typical system (e.g., an airplane) consists of several heterogeneous components. For a system to function normally, it is important that all these components function well.

> For example, for an airplane to function normally, it is important that its structural integrity is intact, that its engines are functioning normally, that its communication system is functioning OK, and that the controlling software is functioning well.

The reliability of each component is normally analyzed by different engineering disciplines which use slightly different techniques. As a result of this analysis, we get the probabilities f_i of each component's failure (or, equivalently, the probability $p_i = 1 - f_i$ that i-th component functions correctly). To estimates the reliability of the entire system, we must combine these probabilities that each component functions correctly into a single probability p that the whole systems functions correctly.

8.2 Traditional Approach to Reliability

The simplest case typically covered by statistical textbooks is when all components are independent. In this case, for reliability, the probability p of the system's correct functioning is equal to the product of the correctness probabilities for components:

$$p = p_1 \cdot \ldots \cdot p_n.$$

Another case with known answer is when all failures are caused by one and the same cause (case of full correlation). In this case, the failure probability is determined by its weakest link, so $f = \max(f_1, \ldots, f_n)$ and $p = \min(p_1, \ldots, p_n)$.

8.3 Traditional Approach is Not Always Sufficient: a Problem

Most real-world situations lie in between these two extremes (see, e.g., [Petroski, 1994], [Ross, 1998]):

- components are not completely independent (e.g., a structural fault can also damage sensors and thus, computational ability suffers), but
- components are not fully correlated.

Typically, we do not have sufficient information about the correlation between the components.

8.4 Proposed Approach to Fusing Probabilities: Main Idea

We must choose a method for combining probabilities. In mathematical terms, we must describe, for every n, a function, which, takes n numbers $p_1, \ldots, p_n \in [0,1]$ as inputs and returns the "fused" (combined) probability $f_n(p_1, \ldots, p_n)$.

Similarly to modes, the division into components is rather subjective; e.g., for three components, we can

- either divide the system into these three subsystems 1, 2, and 3;
- or divide the system into two "macrocomponents" $\{1,2\}$ and 3,
- or divide it into 1 and $\{2,3\}$.

Depending on the division, we get different expressions for the resulting probability:

- If we divide the system into three components, the resulting probability is $p = f_3(p_1, p_2, p_3)$.
- If we divide the system into two macrocomponents $\{1,2\}$ and 3, then for the first macrocomponent, we get $p_{12} = f_2(p_1, p_2)$, and thus, for a system as a whole, $p = f_2(p_{12}, p_3) = f_2(f_2(p_1, p_2), p_3)$.
- Similarly, if we divide the system into two macrocomponents 1 and $\{2,3\}$, then for the second macrocomponent, we get $p_{23} = f_2(p_2, p_3)$, and thus, for a system as a whole, $p = f_2(p_1, p_{23}) = f_2(p_1, f_2(p_2, p_3))$.

The resulting probability should not depend on the (subjective) subdivision into components. As a result, we should get $f_3(p_1, p_2, p_3) = f_2(f_2(p_1, p_2), p_3) = f_2(p_1, f_2(p_2, p_3))$. In other words, we can make two conclusions:

- First, the combination function for arbitrary $n > 2$ can be expressed in terms of a combination function corresponding to $n = 2$, as $f_n(p_1, \ldots, p_n) = f_2(p_1, f_2(p_2, \ldots, f_2(p_{n-1}, p_n) \ldots))$.
- Second, the function f_2 which describes the combination of two probabilities should be *associative* (i.e., $f_2(a, f_2(b, c)) = f_2(f_2(a, b), c)$ for all $a, b, c \in [0, 1]$).

For reliability, the function $g_2(p_1, p_2)$ has the following meaning:

- p_1 is the probability $P(C_1)$ of the event C_1 defined as "the first component is functioning correctly";
- p_2 is the probability $P(C_2)$ of the event C_1 defined as "the first component is functioning correctly"; and
- $f_2(p_1, p_2)$ is the probability $P(C_1 \& C_2)$ that both components are functioning correctly.

In some reliability problems, several components serve as back-ups for one another; in such situations, the system as a whole functions correctly if at least one of the components functions correctly. In other words, in such problems, we are interested in the value $P(C_1 \vee C_2)$. From the probability theory, we know that $P(C_1 \vee C_2) = P(C_1) + P(C_2) - P(C_1 \& C_2)$, i.e., $P(C_1 \vee C_2) = p_1 + p_2 - f_2(p_1, p_2)$. We can describe this expression as $P(C_1 \vee C_2) = g_2(p_1, p_2)$, where $g_2(a, b) = a + b - f_2(a, b)$.

Similar to the case of "and"-combination of different components, we can describe the probability $P(C_1 \vee C_2 \vee C_3)$ (that are least one of three components functions correctly) in two different ways:

- either as $P((C_1 \vee C_2) \vee C_3) = g_2(g_2(p_1, p_2), p_3)$,
- or as $P(C_1 \vee (C_2 \vee C_3)) = g_2(g_2(p_1, p_2), p_3)$.

The expression for $P(C_1 \vee C_2 \vee C_3)$ should not depend on how we compute it, and therefore, we should have $g_2(g_2(p_1, p_2), p_3) = g_2(p_1, g_2(p_2, p_3))$. In other words, not only the function $f_2(a, b)$ should be associative, but also the function $g_2(a, b) = a + b - f_2(a, b)$ should be associative.

8.5 Resulting Solution

We already know that functions $f(a, b)$ for which both the function itself and the expression $a + b - f(a, b)$ are associative have been classified and are known as Frank's t-norms

$$f_2(a, b) = \log_s \left[1 + \frac{(s^a - 1) \cdot (s^b - 1)}{s - 1}\right]$$

for some constant s. As a particular case of this general formula, we get the above two expressions $a \cdot b$ (for $s \to 1$) and $\min(a, b)$ (for $s \to 0$).

9 Potential Applications to Mammography

The main problem of mammography is to detect small faults in the mammory (small clots, cracks, etc.), which may indicate a tumor. When formulated in these terms, the problem sounds very similar to the problem of aerospace testing: in both cases, we must detect possible faults. Thus, we can use the above ideas in mammography as well.

Acknowledgments

This work was supported in part by NASA under cooperative agreement NCC5-209, by NSF grants No. DUE-9750858 and CDA-9522207, by United Space Alliance, grant No. NAS 9-20000 (PWO C0C67713A6), 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, and by the National Security Agency under Grant No. MDA904-98-1-0561.

References

Aczel, J. (1966) *Lectures on functional equations and their applications*, Academic Press, NY-London.

Andre, G. (1999) *Comparison of Vibrational Damage Detection Methods in an Aerospace Vertical Stabilizer Structure*, Master Thesis, The University of Texas at El Paso, Civil Engineering Department.

Barbier, P., and P. Blondet (1993) *Using NDT techniques in the maintenance of aeronautical products*, Aerospatiale France Report No. 93-11587/1/GAL.

Chimenti, D. E. (1997) "Guided waves in plates and their use in materials characterization", *Appl. Mech. Rev.* **50** (5), 247–287.

Clough, R. W., and J. Penzien (1986) *Dynamics of Structures*, McGraw Hill, N.Y.

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

Frank, M. J. (1979) "On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$", *Aequationes Mathematicae* **19** (2–3), 194–226.

Grabec, I., and W. Sachse (1997) *Synergetics of Measurement, Prediction, and Control*, Springer Verlag, Berlin-Heidelberg.

Gros, X. E. (1997) *NDT Data Fusion*, J. Wiley, London.

Hovey, P. W., and A. P. Berens (1988) "Statistical evaluation of NDE reliability in the aerospace industry", *Review of Progress in QNDE*, Plenum, N.Y., **7B**, 1761–1768.

Klir, G. and B. Yuan (1995) *Fuzzy sets and fuzzy logic: theory and applications*, Prentice Hall, Upper Saddle River, NJ.

Kosheleva, O., L. Longpré, and R. A. Osegueda (1999) "Detecting Known Non-Smooth Structures in Images: Fuzzy and Probabilistic Methods, with Applications to Medical Imaging, Non-Destructive Testing, and Detecting Text on Web Pages", *Proceedings of The Eighth International Fuzzy Systems Association World Congress IFSA'99*, Taipei, Taiwan, August 17-20, 269–273.

Kreinovich, V., H. T. Nguyen, and E. A. Walker (1996) "Maximum entropy (MaxEnt) method in expert systems and intelligent control: new possibilities and limitations", In: K. M. Hanson and R. N. Silver, Eds., *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht, 93–100.

Kreinovich, V., C. Quintana, and O. Fuentes (1993) "Genetic algorithms: what fitness scaling is optimal?" *Cybernetics and Systems: an International Journal*, **24** (1), 9–26.

Krishna, M., V. Kreinovich, and R. A. Osegueda (1999) "Fuzzy Logic in Non-Destructive Testing of Aerospace Structures", *Proceedings of the 1999 IEEE Midwest Symposium on Circuits and Systems*, Las Cruces, New Mexico, August 8–11 (to appear).

Mal, A. K. and S. J. Singh (1991) *Deformation of Elastic Bodies*, Prentice Hall, Englewood Cliffs, NJ.

Nguyen, H. T., and V. Kreinovich (1997) *Applications of Continuous Mathematics to Computer Science*, Kluwer, Dordrecht.

Nguyen, H. T., and E. A. Walker (1999) *A first course in fuzzy logic*, CRC Press, Boca Raton, Florida.

Osegueda, R. A., Y. Mendoza, O. Kosheleva, and V. Kreinovich (1999) "Multi-Resolution Methods in Non-Destructive Testing of Aerospace Structures and in Medicine", *Proceedings of the 14th IEEE International Symposium on Intelligent Control/Intelligent Systems and Semiotics ISIC/ISAS'99*, Cambridge, Massachusetts, September 15–17, 208–212.

Osegueda, R. A., A. Revilla, L. Pereyra, and O. Moguel (1999) "Fusion of modal strain energy differences for localization of damage", In: A. K. Mal (ed.), *Nondestructive Evaluation of Aging Aircraft, Airports, and Aerospace Hardware III*, Proceedings of SPIE, **3586**, Paper 3586–28.

Pereyra, L. R., R. A. Osegueda, C. Carrasco, and C. Ferregut (1999) "Damage detection in a stiffened plate using modal strain energy differences", Ibid, Paper 3586–29.

Petroski, H. (1994) *Design paradigms: case histories of error and judgment in engineering*, Cambridge University Press, Cambridge, MA.

Ross, T. J. (1998) “Case studies in civil engineering: fuzzy logic applications”, In: B. Papadopoulos and A. Syropoulos (eds.), *Current Trends and Developments in Fuzzy Logic, Proceedings of the First International Workshop*, Thessaloniki, Greece, October 16–20, 211–236.

Ross, T. J., C. Ferregut, R. A. Osegueda, and V. Kreinovich (1999) “System Reliability: A Case When Fuzzy Logic Enhances Probability Theory’s Ability to Deal With Real-World Problems”, *Proceedings of the 18th International Conference of the North American Fuzzy Information Society NAFIPS’99*, New York City, June 10–12, 81–84.

Stubbs, N. S., T. Broom, and R. A. Osegueda (1998) “Non-destructive construction error detection in large space structures”, *AIAA ADM Issues of the International Space Station*, AIAA, Williamsburg, Virginia, April, 47–55.

Terano, T., K. Asai, and M. Sugeno (1987) *Fuzzy Systems Theory and its Applications*, Academic Press, San Diego, CA.

Ulieru, M. and R. Isermann (1993) “Design of a fuzzy-logic based diagnostic model for technical processes”, *Fuzzy Sets and Systems*, **52** (3), 249–272.

Viktorov, I. A. (1967) *Rayleigh and Lamb Waves: Physical Theory and Applications*, Plenum Press, N.Y.

Yam, Y., R. A. Osegueda and V. Kreinovich (1999) “Towards Faster, Smoother, and More Compact Fuzzy Approximation, with an Application to Non-Destructive Evaluation of Space Shuttle’s Structural Integrity”, *Proceedings of the 18th International Conference of the North American Fuzzy Information Society NAFIPS’99*, New York City, June 10–12, 243–247.

Appendix: Proofs

Proof of Theorem 1

This proof is based on the following lemma:

Lemma. *If an optimality criterion is final and unit-invariant, then the optimal family F_{opt} is also unit-invariant, i.e., $R_\lambda(F_{opt}) = F_{opt}$ for every number λ.*

Proof of the Lemma. Since the optimality criterion is final, there exists a unique family F_{opt} that is optimal with respect to this criterion, i.e., for every other F, either $F_{opt} \succ F$, or $F_{opt} \sim F$.

To prove that $F_{opt} = R_\lambda(F_{opt})$, we will first show that the re-scaled family $R_\lambda(F_{opt})$ is also optimal, i.e., that for every family F: either $R_\lambda(F_{opt}) \succ F$, or $R_\lambda(F_{opt}) \sim F$.

If we prove this optimality, then the desired equality will follow from the fact that our optimality criterion is final and therefore, there is only one optimal family (so, since the families F_{opt} and $R_\lambda(F_{opt})$ are both optimal, they must be the same family).

Let us show that $R_\lambda(F_{opt})$ is indeed optimal. How can we, e.g., prove that $R_\lambda(F_{opt}) \succ F$? Since the optimality criterion is unit-invariant, the desired relation is equivalent to $F_{opt} \succ R_{\lambda^{-1}}(F)$. Similarly, the relation $R_\lambda(F_{opt}) \sim F$ is equivalent to $F_{opt} \sim R_{\lambda^{-1}}(F)$.

These two equivalences allow us to complete the proof of the lemma. Indeed, since F_{opt} is optimal, we have one of the two possibilities: either $F_{opt} \succ R_{\lambda^{-1}}(F)$, or $F_{opt} \sim R_{\lambda^{-1}}(F)$. In the first case, we have $R_\lambda(F_{opt}) \succ F$; in the second case, we have $R_\lambda(F_{opt}) \sim F$.

Thus, whatever family F we take, we always have either $R_\lambda(F_{opt}) \succ F$, or $R_\lambda(F_{opt}) \sim F$. Hence, $R_\lambda(F_{opt})$ is indeed optimal and thence, $R_\lambda(F_{opt}) = F_{opt}$. The lemma is proven.

Let us now prove the theorem. Since the criterion is final, there exists an optimal family $F_{opt} = \{C \cdot f(z)\}$. Due to the lemma, the optimal family is unit-invariant.

From unit-invariance, it follows that for every λ, there exists a real number $A(\lambda)$ for which $f(\lambda \cdot z) = A(\lambda) \cdot f(z)$. Since the function $f(z)$ is differentiable, we can conclude that the ratio $A(\lambda) = f(\lambda \cdot z)/f(z)$ is differentiable as well. Thus, we can differentiate both sides of the above equation with respect to λ, and substitute $\lambda = 1$. As a result, we get the following differential equation for the unknown function $f(z)$:

$$z \cdot \frac{df}{dz} = \alpha \cdot f,$$

where by α, we denoted the value of the derivative $dA/d\lambda$ taken at $\lambda = 1$. Moving terms dz and z to the right-hand side and all the term containing f to the left-hand side, we conclude that

$$\frac{df}{f} = \alpha \cdot \frac{dz}{z}.$$

Integrating both sides of this equation, we conclude that $\ln(f) = \alpha \cdot \ln(z) + C$ for some constant C, and therefore, that $f(z) = \text{const} \cdot z^\alpha$. The theorem is proven.

Proof of Theorem 2

Since the optimality criterion is final, there exists an optimal family F_{opt}. Similarly to the proof of Theorem 1, we prove that this optimal family is unit-invariant, i.e., $R_\lambda(F_{\text{opt}}) = F_{\text{opt}}$ for all real numbers $\lambda > 0$.

So, if a function $p(E)$ belongs to the optimal family $F_{\rm opt}$, then, for every $\lambda > 0$, the re-scaled function $p(\lambda \cdot E)$ of multiplying E to this function f belongs to $F_{\rm opt}$, i.e., due to definition of a family, there exist values $k(\lambda)$, etc., for which

$$p(\lambda \cdot E) = \frac{k(\lambda) \cdot p(E) + l(\lambda)}{m(\lambda) \cdot p(E) + n(\lambda)}. \tag{3}$$

The solution to this functional equation is, in essence, described in [Aczel, 1966]. For completeness, let us describe the proof in detail.

For $\lambda = 1$, we have $k = n = 1$ and $l = m = 0$, so, since p is smooth (hence continuous), for $\lambda \approx 1$, we have $n(\lambda) \neq 0$; hence, we can divide both the numerator and the denominator of (3) by $n(\lambda)$ and thus, get a similar formula with $n(\lambda) = 1$. If we multiply both sides of the resulting equation by the denominator, we get the following formula:

$$m(\lambda) \cdot p(E) \cdot p(\lambda \cdot E) + p(E) = k(\lambda) \cdot p(E) + l(\lambda).$$

If we fix λ and take three different values of E, we get three linear equations for determining three unknowns $k(\lambda)$, $l(\lambda)$, and $m(\lambda)$, from which we can determine these unknowns using Cramer's rule. Cramer's rule expresses every unknown as a fraction of two determinants, and these determinants polynomially depend on the coefficients. The coefficients either do not depend on λ at all (like $p(E)$) or depend smoothly ($p(\lambda \cdot E)$ smoothly depends on λ because $p(E)$ is a smooth function). Therefore, these polynomials are also smooth functions of λ, and so are their ratios $k(\lambda)$, $l(\lambda)$, and $m(\lambda)$.

Now that we know that all the functions in the equations (3) are differentiable, we can differentiate both sides with respect to λ and set $\lambda = 1$. As a result, we get the following differential equation:

$$E \cdot \frac{dp}{dE} = C_0 + C_1 \cdot p + C_2 \cdot p^2$$

for some constants C_i. To solve this equation, we can separate the variables, i.e., move all the terms related to E to one side and all the terms related to p to the other side, and get the differential equation

$$\frac{dp}{C_0 + C_1 \cdot p + C_2 \cdot p^2} = \frac{dE}{E}. \tag{4}$$

Let us first show that $C_2 \neq 0$. Indeed, if $C_2 = 0$ and $C_1 = 0$, then $p/C_0 = \ln(E) + \text{const}$, which contradicts to our assumption that $p(0) = 0$. If $C_2 = 0$ and $C_1 \neq 0$, then we get $C_1^{-1} \cdot \ln(C_1 \cdot p + C_0) = \ln(E) + \text{const}$ hence $C_1 \cdot p + C_0 = A \cdot E^\alpha$, which for $\alpha < 0$, contradicts to the assumption that $p(0) = 0$, and for $\alpha > 0$, contradicts to the assumption that $p(E) \to 1$ as $E \to \infty$. Thus, the case $C_2 = 0$ is impossible, and $C_2 \neq 0$. For $C_2 \neq 0$, in general, the left-hand side of the equation (4) can be represented as a linear combination of elementary

fractions $(p+z_1)^{-1}$ and $(p+z_2)^{-1}$ (where z_i are – possibly complex – roots of a quadratic polynomial $C_1 + C_1 \cdot p + C_2 \cdot p^2$):

$$\frac{1}{C_0 + C_1 \cdot p + C_2 \cdot p^2} = c \cdot \left(\frac{1}{p+z_1} - \frac{1}{p+z_2} \right).$$

(the case of a double root can be handled in a similar manner.) Thus, integrating the equation (4), we conclude that

$$c \ln \left(\frac{p+z_1}{p+z_2} \right) = \ln(E) + \text{const},$$

and

$$\frac{p+z_1}{p+z_2} = P \cdot E^{\beta}$$

for some A and β. So, the expression $A \cdot E^{\beta}$ can be obtained from $p(E)$ by a fractional linear transformation; hence, by applying the inverse transformation (and it is known that the inverse to a fractionally linear transformation is also fractionally linear) we conclude that

$$p(E) = \frac{A \cdot E^{\beta} + B}{C \cdot E^{\beta} + D}$$

for some numbers A, B, C, and D. One can easily check that only for real values $A - D$ and β, we get a monotonic everywhere defined function $p(E)$.

If $\beta < 0$, then we can multiply both numerator and denominator by $E^{-\beta}$ and get a similar expression with $\beta > 0$. Thus, without losing generality, we can assume that $\beta > 0$. Now, the condition that $p(0) = 0$ leads to $B/D = 0$ and hence, to $B = 0$. The condition leads to $A = C$, i.e., to

$$p(E) = \frac{A \cdot E^{\beta}}{A \cdot E^{\beta} + D}.$$

Since $p(E)$ is not identically equal to 1, we have $D \neq 0$. Therefore, we can divide both the numerator and the denominator of this fraction by D, and get the desired expression (1). The theorem is proven.