

Multi-Resolution Methods in Non-Destructive Testing of Aerospace Structures and in Medicine

Roberto Osegueda¹, Yohans Mendoza^{2,1}, Olga Kosheleva³, and Vladik Kreinovich^{1,2}

¹FAST Center for Structural Integrity of Aerospace Systems

²Department of Computer Science

³Department of Electrical and Computer Engineering

University of Texas at El Paso

El Paso, TX 79968, USA

emails osegueda@utep.edu, yohans@sol.mmtlc.utep.edu, olga@ece.utep.edu, vladik@cs.utep.edu

Abstract

A fault in an aerospace structure can lead to catastrophic consequences; therefore, it is extremely important to test these structures regularly. Thorough testing of a huge aerospace structures results in a large amount of data, and processing this data takes a lot of time. To decrease the processing time, we use a “multi-resolution” technique, in which we first separate the data into data corresponding to different vibration modes, and then combine these data together. There are many possible ways to transform each mode’s data into the probability of a fault, and many possible way of combining these mode-based probabilities; different approaches lead to different results. In this paper, we show how a general methodology for choosing the optimal uncertainty representation can be used to find the optimal uncertainty representations for this particular problem. Namely, we we show that the problem of finding the best approximation to the probability of detection (POD) curve (describing the dependence of probability $p(a)$ of detection on the size a of the fault, see [2, 6, 7]) can be solved similarly to the problem of finding the best activation function in neural networks. A similar approach can be used in detecting faults in medical images (e.g., in mammography).

1 Introduction

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 its “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 (cracks, holes, etc.), are *external*, and can, therefore, be detected during the visual inspection. However, to detect *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 it 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. [3, 5, 9, 11].

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 special signals 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.

1.4 How Can We Save Time and Money?

In spite of many time-saving ideas, for each of these methods, we must still scan a huge area for potential small faults. As a result, testing requires lots of time, and is very expensive. How can we save the time and cost of testing? Our main idea is this:

The existing testing procedures are very expensive and time-consuming because they try not only to check whether there is a fault, but also to find its location and size. If our only goal is to detect the fault, and we are not interested in its exact location, then the problem becomes much simpler and hopefully, easier to solve. Therefore, we suggest to make a two-step testing:

- First, we apply a simpler test to check whether there is a fault.
- Only when the first test detects the presence of a fault, we run more expensive tests to locate and size this fault.

This two-step procedure is very similar to medical testing: In medical testing, first, the basic parameters are tested such as body temperature, blood pressure, pulse, etc. If everything is OK, then the person is considered healthy. Only if something is not OK, then the whole battery of often expensive and time-consuming tests is used to detect what exactly is wrong with the patient.

So the question is: How can we detect the presence of a fault?

1.5 Our 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 hierarchical structure and to apply multi-resolution techniques 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 divided into clusters.

Instead of division in the original space, we propose to make a division, crudely speaking, in frequency domain, i.e., 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 combine them into an overall probability of having a fault at this particular point.

Ideally, we should get all these probabilities 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 [4].

The intelligent methods, however, come with a large uncertainty: we can use different t-norms and t-conorms to combine degrees of belief in fuzzy-logic approach; we can use different activation functions in a neural network approach, etc. The resulting probabilities essentially depend on these choices: for some choices, we get a very good fault detection, while for others, the quality of detection is much worse. It is desirable to find the optimal uncertainty representations.

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., [10]), we have developed a general methodology for finding the optimal uncertainty representation. In [8], we have shown how this general methodology can be used to find the optimal uncertainty representations for this particular problem: namely, the problem of assigning probability $p(a)$ to excess energy $E(a)$ is solved similarly to problem of finding the best simulated annealing technique. The

resulting (optimal) function is

$$p(a) = \frac{(E(a))^\alpha}{\sum_b (E(b))^\alpha}$$

for some real number α .

In this paper, we show that the problem of finding the best approximation to the probability of detection (POD) curve (describing the dependence of probability $p(a)$ of detection on the size a of the fault, see [2, 6, 7]) can be solved similarly to the problem of finding the best activation function in neural networks.

2 Theoretical Explanation for the Empirical Probability of Detection (POD) Curve: A Neural Network-Motivated Approach

Recently, an empirical formula has been found which described this dependence [2, 6, 7]:

$$p(a) = \frac{A \cdot a^\beta}{1 + A \cdot a^\beta}. \quad (1)$$

Since important decisions are based on this formula, it is desirable to find out how reliable it is, i.e., whether it is a crude empirical approximation or a precise formula which has deep theoretical justifications.

In this paper, we show that this formula (1) can indeed be theoretically justified. Our justification for this formula will use methods motivated by the neural network approach (see, e.g., [10]).

2.1 The POD Function Must Be Smooth and Monotonic

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

For a POD function, the probability of detection should be equal to 0 when the fault is too small ($a = 0$) and should be equal to 1 when the fault is very large ($a \rightarrow \infty$).

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

For practical applications, we need the function $p(a)$ which would determine the probability that if a sample with a fault size a is presented to a certain NDE technique, then this 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 $p(a)$ 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 $p(a)$ 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 $p(a)$, 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

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

In mathematical terms, the transformation from $p(a) = P_0(H_i)$ to $\tilde{p}(a) = P(H_i|S)$ is *fractionally linear*, i.e., has the form $p(a) \rightarrow \tilde{p}(a) = \varphi(p(a))$, 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 $p(a)$, we should look for a family of functions $\{\varphi(p(a))\}$, where $p(a)$ 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.

2.3 What family is the best?

Among all such families, we want to choose the best one. In formalizing what “the best” means we follow the general idea outlined in [10]. The criteria to choose may be *approximation accuracy* (i.e., accuracy with which these functions approximate the empirical

data about the dependence of the POD of the fault size), *computational simplicity*, or something else. In mathematical optimization problems, numeric criteria are most frequently used, when to every family we assign some value expressing its performance, and choose a family for which this value is maximal. However, it is not necessary to restrict ourselves to such numeric criteria only. For example, if we have several different families that have the same approximation accuracy A , we can choose between them the one that has the smallest computational complexity C . 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 $A(F_1) > A(F_2)$ or $A(F_1) = A(F_2)$ and $C(F_1) < C(F_2)$* . A criterion can be even more complicated. What a criterion *must* do is to allow us for every pair of families to tell whether the first family is better with respect to this criterion (we'll denote it by $F_1 > F_2$), or the second is better ($F_1 < F_2$) or these families have the same quality in the sense of this criterion (we'll denote it by $F_1 \sim F_2$). Of course, it is necessary to demand that these choices be consistent, e.g., if $F_1 > F_2$ and $F_2 > F_3$ then $F_1 > F_3$.

Another 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 a problem to choose among those “best”. Therefore, we need some additional criterion for that choice. For example, if several families turn out to have the same approximation accuracy, we can choose among them a family with minimal computational complexity. 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 according to the old criterion they had the same quality 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 ultimate; we have to modify it until we come to a final criterion that will have that property.

The next natural condition that the criterion must satisfy is that the relative quality of the two families should not depend on the choice of the units in which we measure the size of the fault. Suppose that instead of the original unit of length, we consider a new unit of length which is λ times larger than the original one. How will

the POD curve change, i.e., what will be the new function $\tilde{p}(\tilde{a})$ describing the dependence of the probability of detection on the size \tilde{a} in the new units? One new unit is equal to λ old units, therefore, the length \tilde{a} in the new units means the length $a = \lambda \cdot \tilde{a}$ in the old units. So, the probability $\tilde{p}(\tilde{a})$ is equal to $\tilde{p}(\tilde{a}) = p(\lambda \cdot a)$. The optimality criterion should be invariant with respect to these rescaling transformations.

We arrive at the following definitions:

2.4 Definitions

Definition 1.

- By a *probability function*, we mean a smooth monotonic function $p(a)$ defined for all $a \geq 0$ for which $p(0) = 0$ and $p(a) \rightarrow 1$ as $a \rightarrow \infty$.
- By a *family of functions* we mean the set of functions that is obtained from a probability function $p(a)$ by applying fractionally linear transformations.
- A pair of relations $(<, \sim)$ is called *consistent* if it satisfies the following conditions: (1) if $F < G$ and $G < H$ then $F < H$; (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 $F < G$ and $G \sim H$ then $F < H$; (6) if $F \sim G$ and $G < H$ then $F < H$; (7) if $F < G$ then $G < F$ or $G \sim F$ are impossible.
- Assume a set \mathcal{F} is given. Its elements will be called *alternatives*. By an *optimality criterion* we mean a consistent pair $(<, \sim)$ of relations on the set \mathcal{F} of all alternatives. If $F > G$, we say that F is *better* than G ; if $F \sim G$, we say that the alternatives F and G are *equivalent* with respect to this criterion. We say that an alternative F is *optimal* (or *best*) with respect to a criterion $(<, \sim)$ if for every other alternative G , either $F > G$ or $F \sim G$.
- We say that a criterion is *final* if there exists an optimal alternative, and this optimal alternative is *unique*.

In the present section we consider optimality criteria on the set \mathcal{F} of all families.

Definition 2. Let $\lambda > 0$. By the λ -rescaling $S_\lambda(p)$ of a function $p(a)$, we mean a function $\tilde{p}(a) = p(\lambda \cdot a)$. By the λ -rescaling $S_\lambda(F)$ of the family F , we mean the family of the functions that are obtained from $p \in F$ by λ -rescaling.

Definition 3. We say that an optimality criterion on \mathcal{F} is scale-invariant if for every two families F and G and for every number $\lambda > 0$, the following two conditions are true:

- if F is better than G in the sense of this criterion (i.e., $F > G$), then $S_\lambda(F) > S_\lambda(G)$;
- if F is equivalent to G in the sense of this criterion (i.e., $F \sim G$), then $S_\lambda(F) \sim S_\lambda(G)$.

2.5 Main Result

Theorem. If a family F is optimal in the sense of some optimality criterion that is final and scale-invariant, then every function p from the family F is equal to $p(a) = (A \cdot a^\beta)/(A \cdot a^\beta + 1)$ for some A and $\beta > 0$.

So, the empirical formula (1) is justified.

2.6 Proof

The idea of this proof is as follows: first we prove that the optimal family is scale-invariant (in part 1), and from that, in part 2, we conclude that any function f from F satisfies some functional equations, whose solutions are known.

1. Let us first prove that the optimal family F_{opt} exists and is *scale-invariant* in the sense that $S_\lambda(F_{\text{opt}}) = F_{\text{opt}}$ for all real numbers $\lambda > 0$.

Indeed, we assumed that the optimality criterion is final; therefore, there exists a unique optimal family F_{opt} . Let's now prove that this optimal family is scale-invariant. The fact that F_{opt} is optimal means that for every other F , either $F_{\text{opt}} > F$ or $F_{\text{opt}} \sim F$. In particular, for every family F and for every $\lambda > 0$, either $F_{\text{opt}} > S_{\lambda^{-1}}(F)$ or $F_{\text{opt}} \sim S_{\lambda^{-1}}(F)$. Since the optimality criterion is scale-invariant, we conclude that for every family F , either $S_\lambda(F_{\text{opt}}) > S_\lambda(S_{\lambda^{-1}}(F)) = F$, or $S_\lambda(F_{\text{opt}}) \sim F$. Thus, $S_\lambda(F_{\text{opt}})$ is also an optimal family. Since the optimality criterion is final, there is only one optimal family hence, $S_\lambda(F_{\text{opt}}) = F_{\text{opt}}$ for all real numbers $\lambda > 0$. Scale-invariance is proven.

2. Let us now deduce the actual form of the functions p from the optimal family.

Due to part 1 of this proof, if the function $p(a)$ belongs to the optimal family F_{opt} , then, for every $\lambda > 0$, the re-scaled function $p(\lambda \cdot a)$ of multiplying a to this function f belongs to F_{opt} , i.e., due to definition of a family, there exist values $k(\lambda)$, etc., for which

$$p(\lambda \cdot a) = \frac{k(\lambda) \cdot p(a) + l(\lambda)}{m(\lambda) \cdot p(a) + n(\lambda)}. \quad (2)$$

The solution to this functional equation is, in essence, described in [1]. 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 (2) 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:

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

If we fix λ and take three different values of a , 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 $p(a)$) or depend smoothly ($p(\lambda \cdot a)$ smoothly depends on λ because $p(a)$ 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 (2) are differentiable, we can differentiate both sides with respect to λ and set $\lambda = 1$. As a result, we get the following differential equation:

$$a \cdot \frac{dp}{da} = C_0 + C_1 \cdot p + C_2 \cdot p^2$$

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

$$\frac{dp}{C_0 + C_1 \cdot p + C_2 \cdot p^2} = \frac{da}{a}. \quad (3)$$

Let us first show that $C_2 \neq 0$. Indeed, if $C_2 = 0$ and $C_1 = 0$, then $p/C_0 = \ln(a) + \text{const}$, which contradicts to our assumption that $p(0) = 0$. If $C_2 = 0$ and $C_1 \neq 0$, then we get $C_1^{-1} \cdot \ln(C_1 \cdot p + C_0) = \ln(a) + \text{const}$ hence $C_1 \cdot p + C_0 = A \cdot a^\alpha$, which for $\alpha < 0$, contradicts to the assumption that $p(0) = 0$, and for $\alpha > 0$, contradicts to the assumption that $p(a) \rightarrow 1$ as $a \rightarrow \infty$. 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 (3) 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 (3), we conclude that

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

and

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

for some A and β . So, the expression $A \cdot a^\beta$ can be obtained from $p(a)$ 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

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

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 $p(a)$.

If $\beta < 0$, then we can multiply both numerator and denominator by $a^{-\beta}$ 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

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

Since $p(a)$ 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.

3 Preliminary Results

We have applied our techniques to testing pieces of Space Shuttle. This technique indeed worked better than previously known methods. We are currently trying different data fusion techniques (as described, e.g., in [6]) to further improve the method's performance.

4 Potential Applications to Mammography

The main problem of mammography is to detect small faults in the mammal (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 idea 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 the 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, by the National Security Agency under Grant No. MDA904-98-1-0564, and by Hong Kong RGC grant 4138/97E.

References

- [1] J. Aczel, "Lectures on functional equations and their applications", Academic Press, NY-London, 1966.
- [2] P. Barbier and P. Blondet, "Using NDT techniques in the maintenance of aeronautical products", Aerospatiale France Report No. 93-11587/1/GAL, 1993.
- [3] D. E. Chimenti, "Guided waves in plates and their use in materials characterization", Appl. Mech. Rev., 1997, Vol. 50, No. 5, pp. 247-287.
- [4] C. Ferregut, R. Osegueda, and A. Nunez (eds.), Proceedings of the International Workshop on Intelligent NDE Sciences for Aging and Futuristic Aircraft, El Paso, TX, September 30-October 2, 1997.
- [5] I. Grabec and W. Sachse, "Synergetics of Measurement, Prediction, and Control", Springer Verlag, Berlin-Heidelberg, 1997.
- [6] X. E. Gros, "NDT Data Fusion", J. Wiley, London, 1997.
- [7] P. W. Hovey and A. P. Berens, "Statistical evaluation of NDE reliability in the aerospace industry", Review of Progress in QNDE, Plenum, N.Y., 1988, Vol. 7B, pp. 1761-1768.
- [8] O. Kosheleva, L. Longpré, and R. Osegueda, "Detecting Known Non-Smooth Structures in Images", Proceedings of The Eighth International Fuzzy Systems Association World Congress IFSA'99, Taipei, Taiwan, August 17-20, 1999 (to appear).
- [9] A. K. Mal and S. J. Singh, "Deformation of Elastic Bodies", Prentice Hall, Englewood Cliffs, NJ, 1991.
- [10] H. T. Nguyen and V. Kreinovich, "Applications of Continuous Mathematics to Computer Science", Kluwer, Dordrecht, 1997.
- [11] I. A. Viktorov, "Rayleigh and Lamb Waves: Physical Theory and Applications", Plenum Press, N.Y., 1967.