POLISH ACADEMY OF SCIENCES - MATERIALS SCIENCE COMMITTEE
SILESIAN UNIVERSITY OF TECHNOLOGY OF GLIWICE
INSTITUTE OF ENGINEERING MATERIALS AND BIOMATERIALS
ASSOCIATION OF ALUMNI OF SILESIAN UNIVERSITY OF TECHNOLOGY

AMME'2002

Conference Proceedings

11th INTERNATIONAL SCIENTIFIC CONFERENCE
ACHIEVEMENTS IN MECHANICAL & MATERIALS ENGINEERING

Application of non probabilistic information

M. Pappalardo, A. Pellegrino

DIMEC Department of Mechanical Engineering, University of Salerno
Via P. don Melillo, 84084 Fisciano (SA), Italy

In this paper we analyse the applications of *MaxInf Principle* in a real problem of optimisation of a typical industrial manufacturing. After we reminded the theory of information we examine an example where is possible to determine a polynomial equation that approximates a set of value with a constant deviation that represents the same error for every point or the same information.

1. INTRODUCTION

The design of a complex system involves analysis, organisation and calculation of various elements under external constraints. The evaluation of criteria adopted may differ in importance because for the low level of information the weigh of all parameters is not well known. In the correct classification of a plan *D*, characterised by the set $\{D_i\}$of criteria, in each phase, needs to know and appraise the contribution of new elements introduced in the planning. For each new introduced criteria D_n it needs to determine the effect, since the condition $D_n \supset D_{n-1}$ is not generally valid. The new effect could be evaluated with the calculus of probabilities [1]. In presence of complex problems, it needs to adopt some algorithms that allow simplifying and keeping the process under control. If we are dealing in a space of probability distributions *M* represents the *distance* between two probability distributions. If *q* is *a priori* distribution and we need to select the distribution *p* closeness to *q***.** for satisfying the constraints of probability distribution we can use the measure of *cross-entropy* developed by Kullbach and Leibler $D(\mathrm{p:q}) = \sum_i p_i \, ln \, p_i / q_i$. *D(p:q)* is the *distance* of the *a priori* distribution *q* from the distribution *p*. In order to optimisation a distribution we can use the Minimum Cross Entropy Principle: *From all probability distribution satisfying given constraint we must choose the distribution p that minimise the measure D(p:q)*. This is the well know *MinEnt* principle. If the *a priori* distribution is of maximum uncertainty then *q* $= \underbrace{1/n, ..., 1/n}_{n}$. Therefore, when the *a priori* distribution *q* has the maximum uncertainty, for minimising *D(p:q)* we should choose values of *p* maximising *entropy S(p)*. This is the *Jaynes' maximum principle (MaxEnt).* Jaynes, the principal proponent of *MaxEnt Principle* in axiomatic way, suggests that in all probability distribution, when we have *only* the constraint that $\{p_i \geq 0\}$ and $\sum_i p_i = 1$, we should choose that one have maximum entropy. The use of probability distribution with less then maximum entropy implies the use of additional information [2].

2. AXIOMS OF NEW THEORY OF INFORMATION (EXTENDED THEORY)

The idea of information, in the theories of Fisher and Wiener-Shannon [3][4], is a measure only on probabilistic and repetitiveness events. The idea of information is larger than the probability and the axioms of Wiener–Shannon can be extended to the non-probabilistic and repetitiveness events. Let Ω to be the field of all events ω, probabilistic or non-probabilistic, and $\Im$ a class of parts of Ω, $\Im \subset \wp_{arts}(\Omega)$. With $A \subset \Im$ we can assume the next two axioms:

AXIOM I: The value of information *J(A)* is always non-negative*:*

$$J(A) \ : \ \Im \rightarrow \mathrm{R}^{+} \tag{1}$$

AXIOM II: The value of information *J(A)* is monotonous in regard to inclusion:

$$\forall A, B \in \Im \quad , \quad B \subset A \quad , \quad J(B) \geq J(A) \tag{2}$$

Now it is possible the construction of new algorithms in terms of information, founded only on the first and second axioms [5]. For independent events it is opportune to assume a third axiom:

AXIOM III: If the events $A, B \in \Im$ are independent for all the values of information we have:

$$\forall A, B \in \Im \qquad \mathrm{J}(B \cap A) = J(B) + J(A) \tag{3}$$

The third axiom shows that when we are in presence of independent events it is possible to add up information. If Ω is a certain event and ϕ the impossible event than, for an universal validity of $J(A)$ and $J(\phi)$, for all $\Omega, \Im$ and J must be:

$$J(\Omega) = 0 \ \ , \ J(\phi) = +\infty \tag{4}$$

The expression $J(\Omega) = 0$ means that Ω is a certain event without needs of information. The expression $J(\phi) = +\infty$ means that if ϕ is an impossible event with the needs of infinite information. In a metric space Ω, if ω is an event in $\Im \subset \wp_{arts}(\Omega)$, its measure will be always incorrect. The knowledge of ω is not given by its coordinates in Ω, but it is possible only to assert that ω is limited in a subset $A_i \in \Im$. If $d(A_i)$ is the diameter of set A_i, than, more is the precision of measures, less is the measure of diameter of event A_i. If we assume that is a set $\{P_{x,y}\}$ of ideal data in a continuous closed bounded subset $\Omega \in [D]$, given any $\varepsilon >$ 0, there is a set $\{M_{x,y}\}$of values of measures with sufficiently high precision such that

$$\left|P_{x,y} - M_{x,y}\right| < \varepsilon \ \ \text{for } (x, y) \in \Omega \tag{5}$$

But the probability *p* of an exact measure is in inverse proportion to the precision, so the ideal measure of point's coordinates of has null probability to be obtained: it is an impossible event. The impossible event ϕ and the certain event Ω are always independent from J and *A*: they are universal values. All three axioms have correspondent axioms in Wiener-Shannon theory.

3. THE EXTENDED PRINCIPLE OF MAX-INFORMATION (MaxInf)

It is possible to define a new principle on basis of the New Theory of Information. On the analogy of *MaxEnt principle*, the name is *Max Information Principle* (*MaxInf*) [6]. In the New

Theory of Information, instead of probability it is possible the utilisation of a finite number of appropriate proportion subject to a set of constraints that add up to one. In observance of the Axioms, let $d_1, d_2, ..., d_n$ be *n* non-negative real numbers, let [7] [8]

$$\sum_{i=1}^{n} d_i \neq 0 \quad \rho_i = \frac{d_i}{d_1 + d_2 + ... + d_n} \quad \sum_{i=1}^{n} \rho_i = 1 \ \ (\rho_i \geq 0 \ \ \forall i) \tag{6}$$

We can use the measure of information the relation

$$J(\rho) = J(\rho_1, \quad \rho_2, ..., \quad \rho_n) = -\sum_{i=1}^{n} \rho_i \ln \rho_i \tag{7}$$

So that

J(ρ) is maximum when $\rho_1 = \ \rho_2 = ... = \ \rho_n$

J(ρ) is minimum when: $\forall i$ only one number is $\neq$ zero

In metric space, using Euclidean's distances the information maybe

$$J = -\sum_{ij} \left| d\left(x_i, x_j\right) \right| \Big/ \sum_{ij} \left| d\left(x_i, x_j\right) \right| \log \left(\left| d\left(x_i, x_j\right) \right| \Big/ \sum_{ij} \left| d\left(x_i, x_j\right) \right| \right) \tag{8}$$

The value of information *J(ρ)* is a measure of equality of numbers among themselves. Applying the same formalism of MaxEnt Principle it is easy to define the MaxInf Principle on the basis of so called *Laplace's Principle* of insufficient reason.

MaxInf Principle: *Out of all knowledge, choose the solution closet to the uniform distribution of information.*

In the situations in which we have no reasons for to prefer a solution, it is better choose the solution with uniform distribution, or the closet to the uniform distribution of information.

4. APPLICATION IN METRIC SPACE

The MaxInf principle can be is useful in problems of approximation as criteria to find polynomials for to represent a given set $E = \{(x_i, y_i), ..\}$ of empirical points [9]. Ideally, this process should take in account the reliability of the observations, so the more reliable points will have grater weight on approximating function. In absence of knowledge, on basis of MaxInf principle, we must use a polynomial, which in representing points, the deviation from them choose the solution closet the uniform distribution of information. In metric space, let be $y = f(x_i)$ the approximating function from which we obtain, from the points (x_i, y_i), the *n* deviation $d_i = (f(x_i) - y_i)$. The estimator vector is

$$\overline{d} = (d_1, d_2, .., d_n)^T \tag{9}$$

As function for to measure the information we can use the function

$$J = \sum_i \frac{1}{\left|(f(x_i) - y_i)\right|} \tag{10}$$

From MaxInf we have the max value for *J* when

$$J_1 = J_2 = ... = J_n \qquad \frac{1}{\left|(f(x_1) - y_1)\right|} = \frac{1}{\left|(f(x_2) - y_2)\right|} = ... = \frac{1}{\left|(f(x_n) - y_n)\right|} \qquad \forall i \tag{11}$$

The max of information is obtained when the approximating function $y = f(x_i)$ *has the same error* from all the *n* points (x_i, y_i).

$$\left|(f(x_1)-y_1)\right| = \left|(f(x_2)-y_2)\right| = ... = \left|(f(x_n)-y_n)\right| = h \tag{12}$$

The estimator vector has the distributions

$$\overline{d} = \Big(\underbrace{h,\ h,..,h}_{n}\Big)^T \tag{13}$$

If the approximation function is a polynomial

$$f(x) = a + bx + cx^2 + ... + dx^n + ... \tag{14}$$

The deviations from the points $\{(x_i, y_i),..\}$ of the function $f(x_i)$ evaluated at certain abscissa and the given ordinate corresponding to the same abscissa:

$$\begin{aligned} f(x_1) - y_1 &= a + bx_1 + cx_1^{\ 2} + ... + dx_1^{\ k} + ... - y_1 = (-1)^1 h \\ &\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ f(x_n) - y_n &= a + bx_n + cx_n^{\ 2} + ... + dx_n^{\ k} + ... - y_n = (-1)^n h \end{aligned} \tag{15}$$

From the solution of the linear system $A\overline{x} = \overline{b}$ with $\overline{b} = (y_1, y_2,..., y_n)^T$ and $\overline{x} = (a,b,c,...,h)^T$ we have the solution of polynomials and the value of h from which can be evaluated the approximation with the max information.

5. APPLICATION

Let us illustrate the method with a very simple numerical application, using as polynomial equation a parabolic curve ($p(x) = ax^2 + bx + c$), that fits the points *A*, *B*, *C* and *D*. In manufacturing industries is important to verify the economic set of production to optimise production. It is possible to evaluate *a priori* the economic set with the usual analysis. For mechanical parts, unitary cost function is given by the sum of a cost to equip the apparatus of production, a cost to realise a single product and a cost to store the products.

$$C_t = C_1 + C_2 + C_3 \tag{16}$$

Comparing the three values to the number of produced pieces is possible to determine a hyperbole, a horizontal straight line and one tilted.

$$C_1 = \frac{K_e}{N} \qquad C_2 = K \qquad C_3 = K_i N \tag{17}$$

The optimisation (economic set of production) isn't in a very narrow range, and in this range it's assimilable to a parable.

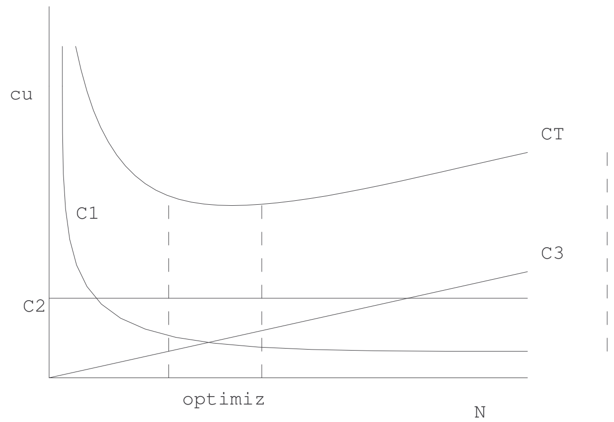

Figure 1. Diagram: unitary cost–number of elements

It's possible, after imposing points given historically or experimentally that are approximately in the field of optimisation of the set, to determine the real curve, hypothesising, in according to (12) that the input data are far from it of a minimum constant value 6h. We assume that in diagram (Number of elements – Cost) there are 4 known points in the range of optimised production, obtained experimentally, and we want to evaluate the value 6h of the approximation with the MaxInf principle.
Input data are:

$$A(1800-138)\quad B(2100-133)\quad C(2700-130)\quad D(3600-132) \tag{18}$$

with x = number of mechanical elements and y = unitary cost.
Imposing that

$$p(x)-y=\pm h \tag{19}$$

curve of constant error, or

$$ax^2+bx+c\pm h=y \tag{20}$$

we have, alternatively

$$\begin{array}{lllll} 3240000a & +1800b & +c & +h & =138 \\ 4410000a & +2100b & +c & -h & =133 \\ 7290000a & +2700b & +c & +h & =130 \\ 12960000a & +3600b & +c & -h & =132 \end{array} \tag{21}$$

and in matricial term

$$\begin{vmatrix} 3240000 & 1800 & 1 & 1 \\ 4410000 & 2100 & 1 & -1 \\ 7290000 & 2700 & 1 & 1 \\ 12960000 & 3600 & 1 & -1 \end{vmatrix} \begin{vmatrix} a \\ b \\ c \\ h \end{vmatrix} = \begin{vmatrix} 138 \\ 133 \\ 130 \\ 132 \end{vmatrix} \tag{22}$$

we have as solution the parabolic curve

$$p(x)=6.851*10^{-6}x^2+0.0397x+186.75 \tag{23}$$

where it's easy to verify that the value obtained with *MaxInf principle* is h=60.55
So it's possible to calculate the curve of optimisation, that approximates with the max information.

6. CONCLUSION

The result obtained isn't absolute. It is possible to recalculate the curve with a new value in input that modifies the value of h. The *MaxInf Principle for non probabilistic events* can be utilised when we need to evaluate an approximate function by a cloud of points, preserving a constant value of information *h*. The applied method has the same numerical result of the theory of the approximation of Cebysev.

REFERENCES

1. N.Siddal, *Probabilistic Engineer Design,* M.Dekker JHG N.York 1983.
2. Jaynes E.T. *Probability Theory: The logic of Science*. http: // omega.albany.edu: 8008/ Jaynes Book.html
3. N. Wiener, *Cybernetics,* Paris Herman (Act. Sc. 1053) 1948.
4. C. E. Shannon, *The Bell System Technical Journal* –Vol. 27, July, October, 1948.
5. J. Kampe de Feriet, *Note di Teoria dell'Informazione,* Anno Accademico 1971-72-Annali dell'Istituto di Matematica Applicata dell'Università di Roma. 1972.
6. M. Pappalardo, *Information in Design Process,* 2nd International Conference on Planned Maintenance, Reliability and Quality - Oxford , England 2nd -3rd April 1998.
7. Nam P. Suh: *The Principles of Design.* Oxford University Press New York 1990.
8. N. Kapur, K. Kesavan, *Entropy Optimization Principles*, Academic Press Inc. Boston 1992.
9. M. Pappalardo, *Information on non-probabilistic events*, DIMEC UNISA Reports n. 10, May 2002