A novel approach to the solution of indirect measurement problems with minimal error propagation

A novel approach to the solution of indirect measurement problems with minimal error propagation

M. K. Smit
Delft University of Technology, Dept of Electrical Engineering, Microwave Laboratory, Delft, The Netherlands

A general solution method for indirect-measurement problems is developed, and implemented in a FORTRAN-IV program. A comparison with the conventional approach shows that the accuracy improvements gained can be very substantial. Use of the method for testing the validity of measurement set-ups, and for improving them is discussed. Attention is also paid to the theoretical implications of the approach. An analysis of conventional solution techniques is given which illustrates the basic weakness associated with these methods. A comparison is made with the theory of stochastic estimation and with the formal theory of measurement, and a fundamental problem in the formulation of the latter is presented.

List of symbols

$b(y)$	binary function describing the Y-space uncertainty region
$\hat{F}$	operator transforming x into y values
$\hat{F}_i x$	ith component of $\hat{F}x$
$\hat{G}$	analytical operator transforming y-values into x-values
$\hat{G}_i y$	ith component of $\hat{G}y$
L	linear approximation of $N\hat{F}$
m	number of X-space dimensions
$\hat{M}$	measurement operation
n	number of Y-space dimensions
N	Y-space normalisation operator
N_m	number of measured parameters
N_u	number of unknown parameters
$\hat{O}_N$	measurement transformation
$\langle p \rangle$	set of primary qualities
R_E	empirical relation
R_N	numerical relation
$\langle s \rangle$	set of secondary qualities
V	$\int_Y b(y)\,dy$
x	point in the X-space
X	value of x corresponding to Y
ΔX	set of maximum errors in the components of X
$(X,\Delta X)$	X-space uncertainty region
Y	set of meter readings
ΔY	set of maximum errors in the components of Y
$(Y,\Delta Y)$	Y-space uncertainty region
σ_i	standard deviation of Y_i
Σ_y	covariance matrix of Y

Symbols used in the examples

A	area of the triangle B_1–B_2–P
B_1	radar beacon at $(-c,0)$
B_2	radar beacon at $(c,0)$
c	half distance between B_1 and B_2
h	height of the radar beacons
m	mass
P	ship position
r_1	distance P–B_1
r_2	distance P–B_2
V	volume
x_1	position co-ordinate
x_2	position co-ordinate
θ	angle B_1–P–B_2
ρ	mass-density

1 Introduction

General approaches to the solution of indirect measurement problems, including the propagation of measurement errors, are not yet available. Only solutions of particular cases are known.

Basic works in the field of measurement theory, such as Krantz *et al* (1971) and Ellis (1966), are concerned mainly with direct measurement. An extension of the formal theory, as it is developed by these authors, to the case of indirect measurement, is given by Finkelstein (1975, 1976) and Leaning and Finkelstein (1979).

A practical problem in applying the formal theory to indirect measurement problems is that the formula expressing the unknown parameters in terms of the measured ones is taken for granted. In measurement practice, such a formula is not always at hand, and it will be shown that finding a good formula may be very troublesome.

A second problem arises from the treatment of measurement errors. These errors can basically upset the fundamental relations on which the formal theory is based. The problem has been recognised by many authors (Krantz *et al*, 1971; Finkelstein, 1975; Leaning and Finkelstein, 1979; Destouches, 1975; Gonella, 1975, 1979). A probabilistic extension of the formal theory to the treatment of measurement errors has been given by Leaning and Finkelstein (1979).

In this paper we will follow an approach which avoids the above problems but differs fundamentally from the formal theory.

The principles of the new approach are outlined in Section 2. The section concludes with the formulation of a general optimal solution to the problem of computing the unknown parameters from the measured ones, i e, a solution with minimal propagation of the measurement errors. In Section 4 this solution is discussed in more detail. Section 3 comprises an analysis of the conventional solution methods, in terms of the theory of Section 2, which shows that these methods are principally sub-optimal. Section 4 reverts to the analysis of Section 3.

Sections 2–4 contain the information necessary for applying the approach in computing the unknown parameters from the measured ones. The new method may also be applied for testing measurement set-ups or improving them. These applications are discussed briefly in Section 4.4.

Sections 5 and 6 can be omitted by readers who are interested only in the practical use of the method. In Section 5 it is shown that the present approach, although not statistical, is closely related to the theory of stochastic estimation. In Section 6 a comparison is made with the formal theory of measurement, which turns out to be inconsistent with the present approach. These sections require some prior knowledge of the theories discussed.

The present theory has developed from problems in the field of dielectric measurement, where it has been shown to yield considerable accuracy improvement. In this article we will use a more transparent problem to illustrate the terms and concepts which are to be introduced.

Let P be the position (x_1, x_2) of a ship, and B_1 and B_2 radar beacons, located at $(-c, 0)$ and $(c, 0)$ respectively. We choose the co-ordinate system as shown in Fig 1. On the ship we measure the distances r_1 and r_2 to B_1 and B_2 respectively, and the angle θ under which both beacons are seen. Using these measurements we want to determine the co-ordinates x_1 and x_2 of the ship.

An obvious approach to solve such a problem is to derive two equations which express x_1 and x_2 in terms of r_1, r_2 and θ. The following pair offers such a solution: *

$$x_1 = \frac{r_1^2 - r_2^2}{4c} \qquad \ldots(1)$$

$$x_2 = \frac{r_1 r_2 \sin\theta}{2c} \qquad \ldots(2)$$

In the following sections it will be shown that such an approach is not optimal. Initially we consider the principles of indirect measurement.

2 Theory of indirect measurement

2.1 Concepts and terminology

Measurement will be conceived of as assigning numerical entities to qualities of the real world. These qualities will be denoted as empirical qualities ('empirical' is used here as the counterpart of 'numerical'). The theory will be developed for measuring a set of qualities, measuring only one quality being a special case of the general approach.

Most measurements are direct measurements, that is, they are based upon a measurement operation which directly assigns numerical entities to the empirical qualities

* The first formula is found by calculating the x_1 co-ordinate of P using Pythagoras: $x_2^2 = r_1^2 - (x_1 + c)^2 = r_2^2 - (x_1 - c)^2$. The latter identity yields Eqn 1. Eqn 2 is found by considering the area of the triangle: $A = x^2 \cdot c = \frac{1}{2} r_1 r_2 \sin\theta$.

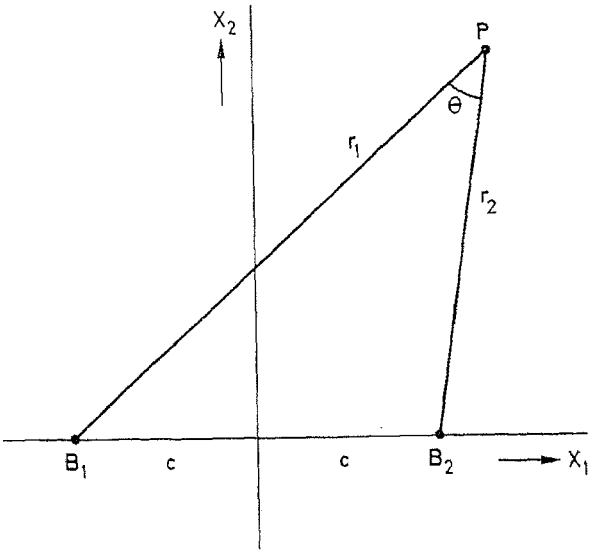

Fig 1 Geometry of the measurement configuration

which are to be measured. When there is no suitable direct measurement operation, we have to apply another procedure. This can be done when the measurement object has, in addition to the qualities in which we are interested, some other qualities which are uniquely related to them, and which can be measured directly.

Fig 2 gives a diagrammatic representation of indirect measurement. The left-hand part of the picture indicates the domain of empirical qualities. The qualities of interest are denoted as $\langle p \rangle$, where p stands for primary. In our example these are the co-ordinates of the ship. The circle in the picture denotes the domain of all possible manifestations of $\langle p \rangle$. The qualities that we use to obtain information about $\langle p \rangle$ are denoted as $\langle s \rangle$, where s stands for secondary. For the position determination the distances r_1 and r_2 and the angle θ are the secondary empirical qualities $\langle s \rangle$. They are geometrically related to $\langle p \rangle$. We will denote this relation as the empirical relation R_E (i e, a relation in the empirical domain).

Now we assume that for the primary, as well as for the secondary entities, there exists a scale. Because $\langle p \rangle$ and $\langle s \rangle$ may represent several qualities, this scale may be multi-dimensional, each dimension representing a scale for one

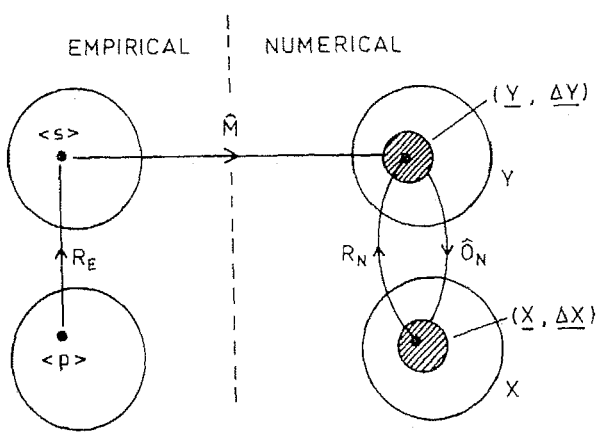

Fig 2 Diagrammatic representation of indirect measurement

single quality. The scale for the primary qualities will be called the X-space, and for the secondary ones the Y-space. These are indicated in the right-hand part of the figure. Now we will confine ourselves to real-valued spaces. This means that qualities which are usually represented by one complex number will now have to be represented by two real ones (real and imaginary part, or modulus and argument).

2.2 The measurement operation

The measurement operation (in Fig 2: the operation $\hat{M}$) will be conceived of as an operation which assigns a region of numerical values to a manifestation of $\langle s \rangle$, so as to account for the measurement inaccuracy. In the figure this region is symbolically denoted as $(Y, \Delta Y)$, which indicates that it describes the meter reading Y as well as the maximal measurement error ΔY. It will be called the Y-space uncertainty region, and defined as follows:

Definition 1: A Y-space uncertainty region is a region in the Y-space containing the values y of *all* manifestations of $\langle s \rangle$ which could have caused the actual meter reading Y, given the measurement inaccuracy ΔY.

To describe this region numerically, we introduce the following function:

$$b(y) = 1 \quad \text{if } |Y_i - y_i| \leqslant \Delta Y_i \quad \text{for all } i = 1, 2, \ldots, n$$
$$= 0 \quad \text{otherwise} \qquad \ldots(3)$$

in which Y_i, y_i and ΔY_i are the ith components of Y, y, and ΔY, respectively, and n is the number of dimensions of the Y-space. The uncertainty region can now be described as the unit-value region of the function $b(y)$. Fig 3 gives a picture of such a region in a three-dimensional Y-space.

It is noted that this definition deviates from the formal definition of measurement as assigning numbers to qualities of the real world (Finkelstein, 1975). It is more in line with suggestions made by Destouches (1975) and Gonella (1975, 1979) to conceive of measurement as assigning intervals instead of numbers.

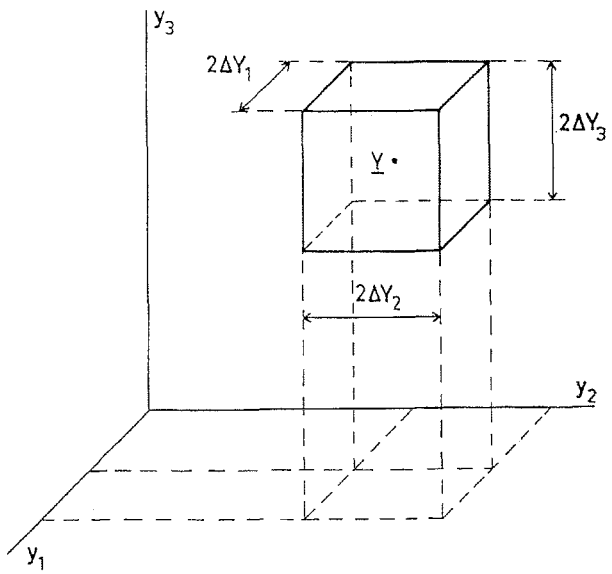

Fig 3 Uncertainty region in a three-dimensional Y*-space*

2.3 The measurement transformation

When the measurement result is obtained, the next problem is how to derive information about the possible values of x, based on the measured Y-space uncertainty region. To solve this problem we define an X-space uncertainty region $(X, \Delta X)$ similar to the Y-space uncertainty region:

Definition 2: An X-space uncertainty region is a region in the X-space containing the values x of *all* manifestations of $\langle p \rangle$ which could have caused the actual meter reading Y, given the measurement inaccuracy ΔY.

Our problem can now be formulated as how to transform a measured Y-space uncertainty region into an X-space uncertainty region. A transformation of this type will be called a *measurement transformation*. It can be written as follows:

$$(X, \Delta X) = \hat{O}_N(Y, \Delta Y) \qquad \ldots(4)$$

in which $\hat{O}_N$ is a numerical operator.

In the example of Section 1, Eqns (1) and (2) can be used to construct such an operator. They must be completed with an operator transforming the maximum measurement error ΔY into a maximum error ΔX. Such an operator can be easily inferred from relations like Eqns (1) and (2), as will be shown in Section 3.

2.4 The optimal measurement transformation

In the following it will be shown that there exist a number of possible transformations $\hat{O}_N$, each with its own error-propagation properties. This means that one transformation will produce a greater uncertainty region than another. This leads to the question: What is the best transformation, ie, which transformation yields the smallest X-space uncertainty region?

To answer this question we will have to look more closely to Definition 2. This definition excludes the possibility that there are manifestations of $\langle p \rangle$ which could have been responsible for the actual meter reading, for which the values x are not contained within $(X, \Delta X)$. But it includes the possibility that $(X, \Delta X)$ contains values x of $\langle p \rangle$, which could never have caused the actual meter reading. If this is the case, our uncertainty region is greater than is strictly necessary. When we exclude this possibility we have the smallest uncertainty region which is possible with the given accuracy ΔY. We will call it the optimal uncertainty region. It can be defined as follows:

Definition 3: An optimal X-space uncertainty region is a region in the X-space containing the values x of *all and only those* manifestations of $\langle p \rangle$ which could have caused the actual meter reading Y, given the measurement inaccuracy ΔY.

Using this definition, we introduce the optimal transformation as the transformation which yields the optimal X-space uncertainty region. Because this uncertainty region is unique, all optimal transformations are equivalent with respect to their error-propagation properties. This is not true for sub-optimal transformations.

In order to find a mathematical formulation for the optimal measurement transformation, we have to pay some attention to the empirical relation R_E of Fig 2. (In our example it is the relation between the measured angle and

distances as the position co-ordinates.) It has a numerical counterpart R_N which can be written as follows:

$$y = \hat{F}x \qquad \ldots(5)$$

in which $\hat{F}$ is a numerical operator. The measurement transformation $\hat{O}_N$ can be conceived of as a particular inversion of $\hat{F}$.

We emphasise that the relation $y = \hat{F}x$ is more fundamental than inversions of $\hat{F}$, such as in Eqns (1) and (2). In many measurement configurations, the dependence of the secondary on the primary empirical qualities will have the form of a relatively simple set of equations. Their inversion, however, may be quite complicated, e g, when we measure field patterns in order to determine some underlying parameters, etc. Furthermore, the relation $y = \hat{F}x$ is unique, which does not hold for its inversions, as will be shown in Section 3.

In our example it is easily verified that R_N has the following form:

$$y_1 = \sqrt{(x_1 + c)^2 + y^2} \qquad \ldots(6)$$

$$y_2 = \sqrt{(x_1 - c)^2 + y^2} \qquad \ldots(7)$$

$$y_3 = \arctan\frac{c - x_1}{x_2} + \arctan\frac{c + x_1}{x_2} \qquad \ldots(8)$$

in which y_1, y_2 and y_3 stand for r_1, r_2 and θ, respectively.

With the use of Eqn (5), Definition 3 can be formulated mathematically as:

$$(X, \Delta X)_{\text{opt}} = \{x \in R^m / \hat{F}x \in (Y, \Delta Y)\} \qquad \ldots(9)$$

or, in other words, the optimal X-space uncertainty region is the region (of an m-dimensional space of real numbers) containing all points x whose transformations $y = \hat{F}x$ fall within the Y-space uncertainty region.

With the use of Eqn (3), Eqn (9) can be written as:

$$(X, \Delta X)_{\text{opt}} = \{x \in R^m / b(\hat{F}x) = 1\} \qquad \ldots(10)$$

Thus the optimal X-space uncertainty region coincides with the unit-value region of the function $b(\hat{F}x)$.

This is a very important result. It gives us a criterion with which to determine whether a point x belongs to the optimal uncertainty region or not. When we have such a criterion the problem has been solved, in principle. The solution can now be traced with conventional multi-dimensional scanning algorithms. A discussion of some algorithms is given in Smit and van Vliet (1983).

Before proceeding to a more detailed discussion of the properties of the optimal measurement transformation, we will analyse the properties of the classical solution of indirect measurement problems.

3 The classical inversion approach

3.1 Mathematical formulation

The most common way to determine x, once y is measured, is by inverting the formula $y = \hat{F}x$. The inversion formula can generally be written as:

$$x = \hat{G}y \qquad \ldots(11)$$

in which $\hat{G}$ is the inversion operator. It is clear that Eqns (1) and (2) offer a solution of this type. In general, all solutions which express x as an analytical function of y are of this type, which we have here termed the classical inversion formula.

Formula (11) does not constitute a valid measurement transformation, because it does not transform a region but a point. This problem can be solved in theory by transforming all points within the Y-space uncertainty region. The X-space region is then found as the set of all transformed points. When $\hat{G}$ behaves reasonably linearly within $(Y, \Delta Y)$, the size of this region can be approximated by:

$$\Delta X_i = \sum_{j=1}^{n} \left| \frac{\partial \hat{G}_i y}{\partial y_j} \right| \Delta Y_j, \qquad i = 1, 2, \ldots, m \qquad \ldots(12)$$

in which $\hat{G}_i y$ is the ith component of $\hat{G}y$ and m the number of dimensions of the X-space. If no explicit expressions for the x_i can be found, $\hat{G}_i y$ has to be replaced by x_i. This form is known in the literature as the total error differential.

Other approaches are also possible. When combined with such an error evaluation, Eqns (1) and (2) constitute a valid measurement transformation. It must be stressed that Eqn (12) or an equivalent is an essential part of the solution procedure. Measurement values without accuracy indication have only meaning if the user knows something about the accuracy in advance.

3.2 Error propagation properties

Now we will prove that Eqns (1) and (2), together with Eqn (12), form a sub-optimal measurement transformation. We assume that the ship of our example is at a position with co-ordinates $x_1 = 1000$ and $x_2 = 1000$. In our notation, $x = (1000, 1000)^T$, in which T denotes transposition. With $c = 100$, we should measure for r_1, r_2, and θ 1487, 1345 and 6 respectively.* Therefore, in our notation, $y = (1487, 1345, 6)^T$. We will assume that the distances can be measured with an accuracy of ± 10 m, and the angle with an accuracy of ± 2 degrees, i e, $\Delta Y\ (10, 10, 2)^T$. With these data, Eqn (12) gives us:

$$\Delta x_1 = \pm 140 \text{ m}$$

and

$$\Delta x_2 = \pm 360 \text{ m}$$

Substitution of the above y-value in Eqns (1) and (2) yields $x_1 = 1005$, $x_2 = 1045$. The corresponding X-space uncertainty region is indicated in Fig 4 by the dashed line.

It is easily shown that this X-space uncertainty region is far from optimal. When we take, for example, a point from the upper right-hand corner, such as (1140, 1400) and compute the distances and angle corresponding to these values using Eqns (6) to (8), we find the following value for y: $(1870, 1744, 5)^T$. As can be easily verified, this point falls far outside the uncertainty region, so that this position of the ship could never have caused the actual measurement result. This means that the X-space uncertainty region is sub-optimal (see Def 3). Consequently the same holds for the classical inversion method and, considering the large discrepancy between the y computed above and the uncertainty region, we question the efficiency of the method.

The cause of this trouble can be seen as follows. We start from the basic relation $y = \hat{F}x$. Once we have measured

* r_1 and r_2 are expressed in metres, θ in degrees. All entities are rounded off to whole numbers.

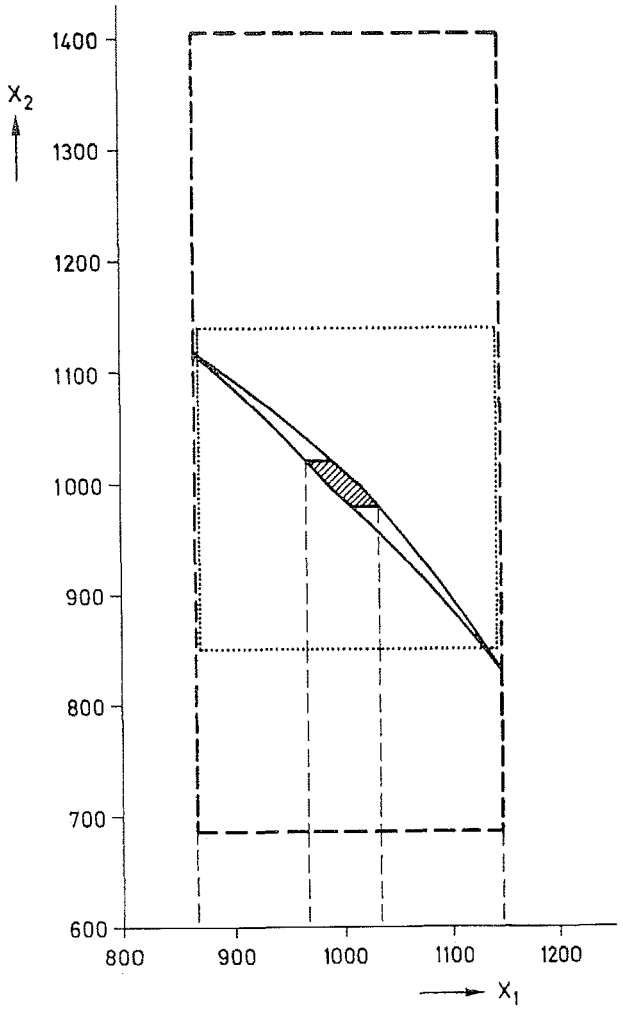

Fig 4 X-space uncertainty regions corresponding to the Y-space uncertainty region with:

$Y^T = (1487, 1345, 6)$, *and* $\Delta Y^T = (10, 10, 2)$:

– – – *based on Eqns (1) and (2)*

.... *based on Eqns (1) and (13)*

——— *optimal X-space uncertainty region without a priori information about* x_2

▨ *optimal X-space uncertainty region with* $X_2 = 1000$ *and* $\Delta X_2 = 20$

the three components of Y we can substitute their values in this relation. This gives us three equations with two unknowns (x_1 and x_2). Consequently the set is overdetermined, and there are two possibilities: either the set is dependent or it is inconsistent.

In the absence of measurement errors, the set will be dependent, but when there are measurement errors, it will most probably be inconsistent. This means that there is no exact solution. There are several ways to find an approximate solution. One way is to select that value of x for which the Euclidian distance between its transform and the measured value (i e, $|\hat{F}x - Y|$) is minimal. This is known as the least-squares approximation (see Section 5).

Another way to circumvent the inconsistency is to remove one of the three equations, and to solve x_1 and x_2 from the other two. It is easily seen that this yields $\binom{3}{2}$ possibilities (assuming that each sub-set of two equations is independent). By manipulating the original set of relations,

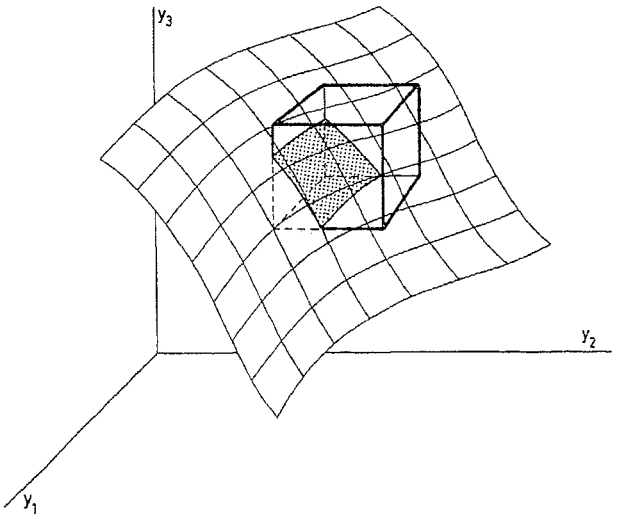

Fig 5 Cross-section of a two-dimensional curved subspace with a three-dimensional uncertainty region (the cross-section has been shaded)

this number can be increased. This is the way in which we arrive at the formulae (1) and (2). Obviously, they are merely one possibility out of many.

The consequences of the latter procedure for the error-propagation are far from negligible. For the subject problem, in which the Y-space is three-dimensional and the X-space two-dimensional, the relation $y = \hat{F}x$ describes a curved plane in the three-dimensional Y-space. Fig 5 shows what such a plane might look like. All points within the plane are 'possible' points: they can really occur. For these points $Y = \hat{F}x$, conceived of as a set of equations with x as unknown, is consistent. For all points outside the plane it is inconsistent. These points are geometrically 'impossible'; without measurement errors they can never be measured.

With this knowledge we can restrict our Y-space uncertainty region; by rejecting all 'impossible' values within this region, we can reduce it to its cross-section with the plane of physical values (the shaded region in Fig 5). To find the X-space uncertainty region, we only have to transform this part of the Y-space to the X-space. The reason for the sub-optimality of the classical approach is that it determines the X-space uncertainty region over the transform $x = Gy$ of both 'possible' and 'impossible' points. It is the transformation of just these 'impossible' points which leads to the sub-optimality of the classical method.

3.3 Potential improvements

An important question is whether the deficiency described above can be avoided. To eliminate it fully we will have to discriminate between 'possible' and 'impossible' values of y. This possibility we have lost, however, in the classical approach, through the removal of one of the three equations. Taking this equation into account so as to increase the accuracy will also drastically change the classical inversion method that it will no longer be an inversion method. The reasons for this will be discussed in more depth in Section 6.

This does not mean, however, that nothing can be done within the framework of the classical approach. We mentioned already that the applied formula is only one of a multitude of possible formulae. For 'possible' points all

these formulae will yield equal results, because of the consistency of the set $y = \hat{F}x$ for these points. For 'impossible' points, however, they will differ because of the inconsistency of the complete set. An investigation into the properties of the applied formula as well as some of the other possible formulae is thus justified.

In our example we may substitute for Eqn (2) the following equation:*

$$x_2 = [\{(r_1 + r_2)^2 - 4c^2\}\ \{4c^2 - (r_1 - r_2)^2\}]^{1/2}/4c \quad \ldots(13)$$

Together with Eqns (1) and (12), this yields the dotted region in Fig 4 as X-space uncertainty region corresponding to the Y-space uncertainty region described in the previous paragraph. This set turns out to be better than the previous one. Analysis of the error-propagation properties in the vicinity of the origin will point out, however, that in this region the previous set is better. There is no single set of equations with optimal properties throughout the whole region.

An alternative way to improve the accuracy can be applied when $\langle s \rangle$ depends on, in addition to $\langle p \rangle$, some other parameters, which can be freely chosen. In our example, the height h of the radar beacons is such a parameter. It can easily be accounted for in Eqns (6) to (8). This yields three equations with three unknowns, from which x_1, x_2 and h can be solved analytically. Another possibility is to measure h and substitute its value in the Eqn $y = \hat{F}x$ (substituting $h = 0$ will yield Eqns (6) to (8). We can now analyse the error propagation properties of the formulae for x_1 and x_2 for different values of h, and choose h such that these properties are optimal. In our case this dependence is weak as long as $h \ll r_1$, r_2, the optimum lying at $h = 0$.

Obviously, our example is less suitable for demonstrating the importance of parameters like h in optimising measurement accuracy. It is important, however, to realise that parameters like these, whose influence is obscured because they do not occur in the inversion formulae for the relevant parameters, may greatly affect the accuracy with which these parameters can be determined.

4 The optimal measurement transformation

4.1 General character

In Section 2.4 we concluded that the problem of optimally solving an indirect measurement problem can be reduced to the problem of tracing a region in an m-dimensional space with the function $b(\hat{F}x)$ as discriminant.

Solution method. In the absence of an analytical solution (see Section 6), we have been reduced to dependence on numerical solution procedures. An advantage of the present approach is that the solution can be formulated most generally. Scanning algorithms can be programmed for a variable number of dimensions of the X-space, as well as the function b, so that the only problem-specific entity remaining is the operator $\hat{F}$. If $\hat{F}$ is given in the form of a user-supplied sub-routine computing y on x as input, the solution can be programmed to a high level of generality. Such a program has been developed in the FORTRAN-IV language and is described in Smit and van Vliet (1983).

* This formula is found by solving x_2 from Eqns (6) and (7).

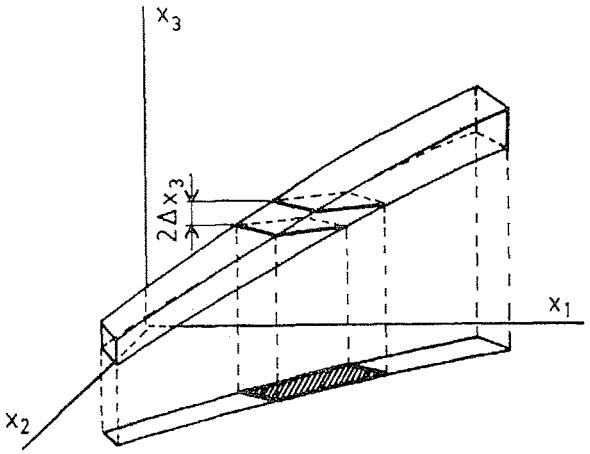

Fig 6 A three-dimensional X-space uncertainty region, showing the effect of a priori *knowledge about* x_3 *on the size of the projection on the* X_1–X_2 *co-ordinate plane*

Representation of the solution. A problem which is independent of the way in which the uncertainty region is detected is its representation. For one-dimensional regions, this region is fully described by its minimum and maximum values (assuming the region is singly connected). For multi-dimensional regions, we can follow the same approach and describe it by the extreme values of each parameter within the uncertainty region.

A disadvantage of this approach may be that we obtain no information about the shape of the uncertainty region. In a two-dimensional X-space this disadvantage can be met by making a plot of all regions. For more-than-two-dimensional regions we can use projections on the different co-ordinate planes. Such a projection represents the region of possible values of the two parameters of the plane independent of the others.

Utilisation of additional information. An important feature of the optimal measurement transformation is that it has the potential to increase the accuracy for a parameter by utilising additional information about other parameters.

This potential is illustrated in Fig. 4. The solid line indicates the optimal X-space uncertainty region based on the same data that were used to compute the broken line and the dotted line region.† Suppose now that there is a dike parallel to the x_1-axis, that we know its position, and that we can measure the distance to this dike with an accuracy of ±20 m, so that we know our x_2 co-ordinate within ±20 m. This means that we can reduce the above uncertainty region to the hatched region in Fig. 4. In fact, we cut out a slice of the total region. The effect of this on the accuracy with which the other parameter can be determined may be considerable, as is demonstrated in the figure. Fig 6 shows the same effect in a three-dimensional X-space. If we choose h as third dimension, the uncertainty region will be a near-vertical cylinder because of the weak dependence of r_1, r_2 and θ on h, so that the projection of the slice comes close to the projection of the total region. The larger the dependence, however, the larger the effect will be.

†It follows from Definition 3 that each valid uncertainty region should fully contain the optimal region. Obviously this is not the case for the dotted region in Fig 4. This is due to the non-linearity of Eqn (12), which causes the linear approximation of Eqn (12) to be slightly erroneous.

Mathematical formulation of the problem. The distinction between X and Y parameters, which is prerequisite to the formulation of the operator $\hat{F}$, is not always evident.

Suppose, for example, that we measure the mass m and the volume V of an object and that we want to determine the mass density ρ. Superficial analysis yields two y-parameters (m and V) and one x-parameter (ρ). But then $\hat{F}$ should have two components, and we have only one equation, namely $m = \rho V$ (or $V = m/\rho$).

This problem disappears if we conceive of n (the dimension of the Y-space) not as the number of measured parameters, but as the number of independent equations, relating the measured and unknown parameters to each other.

If we have N_m measured parameters, N_u unknown parameters and n equations interrelating them, we have the following possibilities:

- If $N_m < n$, the measurement problem cannot be solved.
- If $N_m = n$, the x- and y-parameters fall together with the unknown and the measured parameters.
- If $N_m > n$, we select n of the N_m parameters as y-parameters and treat the resulting $N_m - n$ parameters as x-parameters, thus making the X-space $(N_u + N_m - n)$ dimensional. The measured x-parameters are treated as indicated in the paragraph on 'utilisation of additional information'. Which of the measured parameters are selected as x- or y-parameters is arbitrary from a mathematical point of view.

The latter case is illustrated in Fig 7 for the mass-density measurement. The mass m is conceived of as the y-parameter, while ρ and V are taken as x-parameters. R_N then becomes $m = \rho V$. The two hyperbolas represent the curves $\rho V = m_{\min}$, and $\rho V = m_{\max}$, $m_{\min}$ and $m_{\max}$ being the extrema of the (one-dimensional) Y-space uncertainty region. The area between these curves is the X-space uncertainty region, unbounded because $m < n$. It becomes bounded when we restrict the possible values of V based on a measurement of V, as illustrated in the figure. The results, however, are equivalent to those obtained with the classical approach.

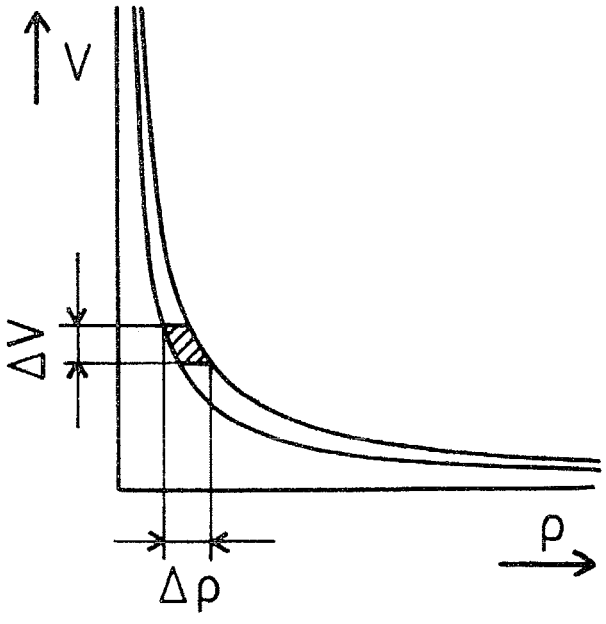

Fig 7 Uncertainty region with n = *1 and* m = *2*

Applicability. Based on the foregoing and on previous experience with the present approach in the field of dielectric measurement (Smit and van Vliet, 1983), the following conclusions seem valid:

- If the number of unknown parameters N_u equals the number of equations n, the classical inversion method is near-optimal. In Fig 4 this can be seen by comparing the size (not the area) of the optimal-uncertainty region with the dotted-line region ($N_u = 2, n = 2$).
- If the number of unknown parameters is less than the number of equations n, the classical inversion method is sub-optimal. The sub-optimality grows with increasing difference between the number of unknown parameters and the number of equations. This can be seen by comparing the solid-line region with the broken-line region ($N_u = 2, n = 3$) and the hatched region with the broken-line region ($N_u = 1, n = 3$).*

4.2 Possible causes of inaccuracy

Although the outlined approach inherently accounts for measurement inaccuracy, there are two types of inaccuracy for which the measurement transformation does not account.

The first type results if the formula $y = \hat{F}x$ contains constants which are known with finite accuracy. If we account in Eqns (6) to (8) for the height of the radar beacons, this height is such a parameter. In physical measurements there are often several entities which affect the measurement result, and which are supposed to be constant. Their values appear as constants in $\hat{F}$, but these constants usually have only finite accuracy.

This type of inaccuracy can be removed, if necessary, by conceiving of these constants as measured x-parameters, and treating them as indicated in Section 4.1.

The second type of inaccuracy occurs when the idealised description $y = \hat{F}x$ differs structurally from the empirical relation R_E. In our example the omission of h causes such a structural deficiency, although it will be a small one, which can, moreover, be eliminated by incorporating h in Eqns (6) to (8). In physical measurements a manageable formula $y = \hat{F}x$ can often be obtained only by making some idealising assumptions.

Inaccuracies of the latter type are very difficult to overcome, because the quantitative description of their origins is very complicated, if not impossible. This fact also makes it very difficult to estimate the size of the errors which they introduce. Yet the latter type of inaccuracy may ultimately become the bottle-neck when the potential of the optimal measurement transformation to increase measurement accuracy by increasing the difference $(n_m - N_u)$ is fully utilised, as indicated in the following section. We note that underestimation of the measurement error ΔY may affect the size of $(X, \Delta X)$ considerably out of proportion. The reasons for this are discussed in Section 4.3.1.

4.3 Other applications

In addition to solving indirect measurement problems, the optimal transformation may be applied for some other purposes. Two of them will be discussed.

* The latter comparison seems to be unfair because the inversion method does not make use of the information in x_2. There is no means, however, to make this information effective within the frame of the applied formula (Eqn 1), although better inversion formulae are certainly feasible.

4.3.1 Testing indirect measurement set-ups

In those cases in which $m < n$, the optimal measurement transformation offers a possibility to test whether the operator $\hat{F}$ exhibits structural inadequacy, as described in Section 4.2. In Section 3.2 it was shown that the equation $y = \hat{F}x$ defines a sub-space in the Y-space (see Fig 5). Because a structural difference between the actual relation R_E and its idealised numerical description $y = \hat{F}x$ will cause the actual sub-space to differ from the idealised one, it will become possible for the Y-space uncertainty region to have no intersection with the sub-space, thus yielding an empty X-space uncertainty region. The chance of finding such an empty sub-space increases with increasing departure from the ideal, and with increasing difference between n and m.*

A drawback of the above property is that it cannot prove adequacy, but only inadequacy. Application of the optimal measurement transformation on a number of measurements on the same object may give us an indication of the validity of the description $y = \hat{F}x$, however. Experience in the field of dielectric measurement indicates that structural inadequacy may ultimately become a bottle-neck when an optimal transformation is used.

4.3.2 Improving indirect measurement set-ups

The optimal transformation makes it possible to increase the measurement accuracy by increasing the number of measurement probes. Although this will not be possible in each case, it applies to all measurements, in which a certain pattern is measured to determine the underlying parameters (e g, a diffraction pattern or a standing-wave pattern). When an inversion method is used, such an approach will most probably increase our measurement inaccuracy. The reason for this is that, when we find an inversion formula containing the new measurement variable(s), the number of terms in Eqn (12) will increase, whereas there is no reason to believe that the partial differentials will become smaller.

It would thus appear that a good approach for optimising measurement accuracy, when using an inversion method, is to minimise the number of measured parameters in the inverse equations and to search for the optimal equations as indicated in Section 3.3.

When an optimal method is used, each new measurement variable places new restrictions on the region of possible outcomes, so that the final result can only become more accurate. This is in accordance with our intuitive expectation.

5 The relation to the theory of stochastic estimation

In the previous sections we have used a binary description for the uncertainty region (Eqn (3)), i e, a description which discriminates between member and non-member points, but assigns no probabilities to the member points.

One may alternatively consider the X-space uncertainty region in a statistical way as in the following.

What we are interested in is the distribution of x given the measurement result Y, i e, the pdf (probability density function) $p(x/Y)$. This pdf can be considered to be the statistical analogue of the X-space uncertainty region.

According to Bayes' formula, this pdf can be calculated from:

$$p(x/Y) = \frac{p(Y/x)\,p(x)}{\int_X p(Y/x)\,p(x)dx} \qquad \ldots(14)$$

in which dx is an infinitesimal volume element, and the integration is over the whole X-space. In this formula $p(Y/x)$ can be obtained from $p(Y/y)$,† which is the statistical analogue of the Y-space uncertainty region, by substituting $y = \hat{F}x$. The pdf $p(x)$ embodies *a priori* knowledge about x.

When there is no *a priori* information, it is usual to assume $p(x)$ to be uniformly distributed throughout the X-space, which will cause $p(x)$ to vanish from Eqn (14).

For the multi-dimensional problems we have a representation problem, similar to the one we encountered for uncertainty regions exceeding two dimensions. This problem can be solved by computing the one-dimensional marginal distribution function:

$$p(x_i/Y) = \frac{p(Y/x_i)\,p(x_i)}{\int_{X_i} p(Y/\chi_i)\,p(\chi_i)\,d\chi_i} \qquad \ldots(15)$$

in which $p(Y/\chi_i)\,p(\chi_i)$ can be found by integrating $p(Y/x)$ $p(x)$ over the sub-space $x_i = \chi_i$. This distribution function is the statistical analogue of the one-dimensional projection.

The computational labour involved in applying the statistical approach can be reduced by applying the technique of stochastic estimation. Stochastic estimation is conceived of as the process of extracting information about a parameter or a signal function from noise-corrupted observations (Nahi, 1969). These observations are not required to be direct observations of the parameter itself, so that a link between stochastic estimation and indirect measurement solution procedures is evident.

Using this technique, we can derive an optimal estimate‡ for the value of x which has to be assigned to Y. Although several approaches are possible, we will confine ourselves to the LSE (Least Squares Estimation) approach. It can be proved (Nahi, 1969) that the expected value $\hat{x}$ of x given Y:

$$\hat{x} = \int_X x\,p(x/Y)\,dx \qquad \ldots(16)$$

minimises the expected value of the square Euclidian distance $|x - x_{est}|^2$, making it a very attractive estimator for our problem.

---

* Finding an empty sub-space may also be due to a too optimistic estimation of ΔY. We will assume that the latter possibility is precluded.

† If we assign equal probability to all points within the Y-space uncertainty region, we obtain the uniform distribution, using Eqn (3):

$$p(Y/y) = V^{-1}\,b(Y)$$

in which V is the integral of $b(y)$ over the whole Y-space. If we choose a normal error description we obtain:

$$p(Y/y) = \frac{1}{(2\pi)^{n/2}\,|\Sigma_y|^{1/2}} \exp\left[-\tfrac{1}{2}(Y-y)^T \Sigma_y^{-1} (Y-y)\right]$$

in which Σ_y is the covariance matrix of the measurement error $(Y_{meas} - Y_{real})$.

‡ An optimal (or Bayes) estimator is defined as an estimator which minimises the expected value $\hat{C}$ of a cost function which is defined on the difference $x - x_{est}$, in which x_{est} is the estimated value of x:

$$\hat{C} = \int_X C(x - x_{est}) \cdot p(x)\,dx.$$

The integration required in Eqn (16) can be eliminated if $p(Y/y)$ can be described by a normal distribution, and if $\hat{F}$ can be linearised in the region of non-neglectable values of $p(Y/y)$. If these conditions are fulfilled, $\hat{x}$ coincides with the minimum of the exponent in the distribution function:

$$(Y-\hat{F}x)^T \Sigma_y^{-1} (Y-\hat{F}x).$$

By normalising the Y-space such that $N^T \Sigma_y^{-1} = I$ (N being the normalisation operator, and I the identity matrix), the exponent will take the form:

$$(Y'-Lx)^T (Y'-Lx)$$

in which $Y' = NY$, and L is the linear approximation of $N\hat{F}$. This form is the square Euclidean distance between Y and $\hat{F}x$ in the normalised Y-space.

Furthermore it can be proved (Gelb *et al*, 1974) that the minimum x_{est} of $(Y'-Lx)^T (Y'-Lx)$ can be computed from:

$$x_{est} = (L^T L)^{-1} L^T Y' \qquad \ldots(17)$$

which is a solution of the equation $Y' = Lx$ in the least-squares sense. When we denote the matrix $(L^T L)^{-1} L^T$ (which is called the pseudo inverse of L) as $L^{\#}$, the covariance matrix of x_{est} follows from:

$$\Sigma_x = L^{\#} (L^{\#})^T \qquad \ldots(18)$$

Eqns (17) and (18) are the statistical analogue of an optimal transformation. From Eqn (14) it can be seen that the X-space uncertainty region coincides with the region of non-neglectable values of $p(x/Y)$. Therefore, Eqns (17) and (18), being the centre and the covariance matrix of this distribution, will be comparable to an optimal transformation with respect to their error-propagation properties.

We may alternatively compute x_{est} by applying a conventional optimalisation procedure to find the value of x for which $|\hat{N}(Y-\hat{F}x)|^2$ is minimal. If the measurement errors in the different components of Y are uncorrelated, the normalised distance reduces to:

$$|\hat{N}(Y-\hat{F}x)|^2 = \sum_{i=1}^{n} \left(\frac{Y_i - \hat{F}_i x}{\sigma_i} \right)^2 \qquad \ldots(19)$$

in which σ_i is the spread in Y_i due to the measurement error. We note that searching for the minimum distance in the non-normalised Y-space will yield sub-optimal results.

The optimisation procedure has the advantage that it can easily be adapted to account for *a priori* knowledge about some components of x by restricting the region to be searched to the region of possible values. Using Eqns (17) and (18), we must account for a non-constant pdf $p(x)$, which complicates the evaluation of these expressions.

Note: An attractive way to obtain information about $p(x/Y)$, at variance with the foregoing one, would appear to be the optimal transformation of some Y-space uncertainty regions, with different confidence levels, e g, the 80, 90, 95 and 98% confidence regions. A problem with this approach, however, is that for $n \neq m$ the optimal transformation fails to obey the law of 'conservation of probability', i e, the confidence levels of a Y-space region and its corresponding X-space region will not, in general, be equal. The reasons underlying this lack of conservation are discussed in Section 6.

6 The relation to the formal theory of measurement

The theory of indirect measurement, as it is developed in the preceding sections, is not without consequence for the formal concept of indirect measurement. To show this we will apply the formalism developed by Finkelstein (1975, 1976), and Leaning and Finkelstein (1979), to our concept (Section 2.1).*

In the formal approach it is assumed that the secondary empirical qualities $\langle s \rangle$, are logically independent, and that to each element $\langle s \rangle$ there corresponds one $\langle p \rangle$ (Finkelstein, 1976, Section 2.5). Next it is assumed that for $\langle s \rangle$ there exists a scale Y with elements $\mathbf{Y}$, that there exists a mapping $\hat{G}$ of Y in X, and that the set of all manifestations $X = \hat{G}(Y)$ constitutes a scale for $\langle p \rangle$.

In this formalism, the mapping $\hat{G}$ is essentially an inversion operator; it is equivalent to Eqn (11). As a consequence, each solution procedure which obeys this formalism is a sub-optimal procedure (presupposed that $m < n$).

The crucial links in this formalism are the assumptions that the secondary qualities are logically independent, and that to each manifestation of them there corresponds one manifestation of $\langle p \rangle$. If $N_u < n$, the set of possible values is a sub-set of the Y-space, as explained in Section 3. As a consequence, the components of $\langle s \rangle$ cannot be logically independent.

To demonstrate the consequences of the second assumption, consider the case of Fig 8, which shows the mapping $y = \hat{F}x$ of a one-dimensional X-space in a two-dimensional Y-space. Let us consider two slightly different Y-space uncertainty regions as indicated in the figure. The optimal X-space uncertainty regions corresponding with them are the transforms of the cross-sections of these Y-space regions with the sub-set $y = \hat{F}x$ into the X-space. As can be seen, these transforms will be disjunct. Nevertheless, the Y-space regions have a part of the Y-space in common (the hatched region). Consequently, when we want to conceive of the optimal transformation as a mapping, this part has to be mapped into one part of the X-space one time, and in another part the other time, which makes clear that an optimal transformation differs fundamentally from a mapping. This means that, if $N_u < n$, an optimal measurement transformation is principally non-analytical; or otherwise, that each analytical solution (Eqn (11)) is principally sub-optimal. It is this character of the optimal transformation which explains the lack of 'conservation of probability' that we mentioned in Section 5; this conservation is restricted to analytical transformations.

The foregoing discussion illustrated that, in fact, regions and not numbers are transformed. As a consequence, the formal theory of measurement, which is based on the concept of measurement as assigning numbers to empirical qualities, is incapable of accounting for the properties of

* We extend the formalism to the case of a multi-dimensional scale for the primary qualities. The terminologies are related as follows:

Present paper	Finkelstein (1976)
$\langle p \rangle$	q_0
$\langle s \rangle$	$q = \langle q_1, \ldots, q_n \rangle$
Y-space	$\prod_{i=1}^{n} Z_i$
X-space	z_0
Y	$M(q)$
X	z_0
$\hat{G}$	Φ

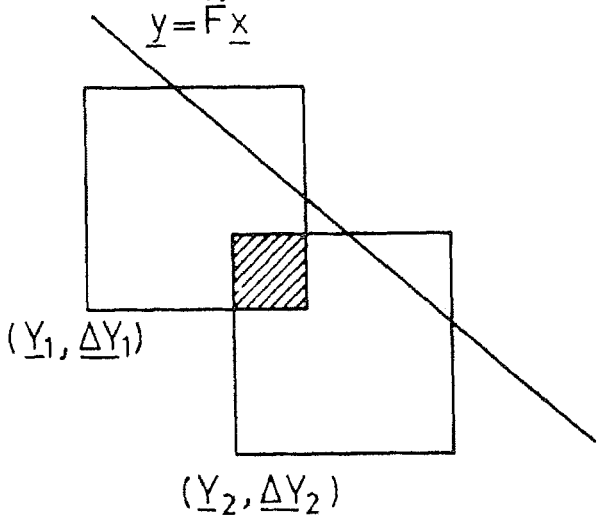

Fig 8 Two overlapping Y-space uncertainty regions with disjunct corresponding X-space uncertainty regions

optimal transformations, including optimal estimators as, for example, the LSE.

In order to avoid this deficiency, the concept of measurement as assigning numbers has to be replaced by that of assigning intervals or regions.

The above deficiency is restricted to cases for which $N_u < n$, and consequently, will not occur in direct measurement. We wonder, however, if the formal concept of measurement should not also be abandoned in the direct measurement field in favour of the above concept, two different concepts covering the entire field of measurement being most unsatisfactory.

7 Conclusions

An analysis of the conventional methods for solving indirect measurement problems has pointed out that these methods make insufficient use of the available information.

A new method, making optimal use of this information, has been developed. It can be formulated at a high level of generality, which makes it suitable for implementation in a computer program that is applicable to a broad class of indirect measurement problems. The method also has a potential for testing the validity of measurement set-ups and improving them. The error-propagation properties of the new method were shown to be comparable with those of optimal estimators.

Comparison with the formal theory of measurement has pointed out that the formal concept of measurement is not adequate to account for the properties of the present approach, and those of optimal estimators. This deficiency may also have consequences in the field of direct measurement. To avoid it, the concept of measurement as assigning numbers will have to be abandoned in favour of the concept of measurement as assigning regions or intervals.

Acknowledgements

The author wishes to thank Prof ir L. Krul and Prof ir B. P. Th. Veltman for their encouragement and valuable comments. Furthermore he is much indebted to Mr Ch. van Vliet, who was closely involved in the elaboration of the theory in the dielectric field, and to Ms J. P. van Soldt for typing the manuscript.

References

Destouches, J. L. 1975. 'On basic problems about measurement and errors', *Proc IMEKO Symp on Measurement Theory–Measurement Error Analysis*, Enschede, The Netherlands.

Ellis, B. 1966. *Basic concepts of measurement*, Cambridge University Press, Cambridge.

Finkelstein, L. 1973. 'Fundamental concepts of measurement., *Proc 6th Congress of International Measurement Confederation*, Dresden, **1**, 11–27.

Finkelstein, L. 1975. 'Fundamental concepts of measurement: definition and scales', *Measurement and Control*, **8**, 105–111.

Finkelstein, L. 1976. 'The relation between the formal theory of measurement and pattern recognition', *Proc 7th Congr of International Measurement Confederation*, London, **1**, 97–105.

Gelb, A. *et al.* 1974. *Applied optimal estimation*, MIT Press, Massachusetts.

Gonella, L. 1975. 'A proposed reformulation of the basic terminology and the theoretical approach concerning measurements', *Proc IMEKO Symp on Measurement Theory–Measurement Error Analysis*, Enschede, The Netherlands.

Gonella, L. 1979. 'Problems in theory of measurement today', *Proc 8th Congr of International Measurement Confederation*, Moscow, **1**, 103–110.

Krantz, D. H., Luce, R. D., Suppes, P. and Tversky, A. 1971. *Foundations of measurement*, Academic Press, New York.

Leaning, M. S. and Finkelstein, L. 1979. 'A probabilistic treatment of measurement uncertainty in the formal theory of measurement', *Proc 8th Congr of International Measurement Confederation*, Moscow, **1**, 73–81.

Nahi, N. E. 1969. *Estimation theory and applications*, John Wiley & Sons, New York.

Smit, M. K. and van Vliet, Ch. 1983. 'OMTP: a program for optimally solving indirect measurement problems', *Measurement*, **1**(4), 209–211.