USCIPI REPORT #122

Heuristically Contrained Estimation for Intelligent Signal Processing

by

Robert F. Popoli and Jerry M. Mendel

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Room 404
Los Angeles, CA 90089-2564 U.S.A.

February 1988

USCSIPI Report No. 122

Heuristically Constrained Estimation for Intelligent Signal Processing

by

Robert F. Popoli and Jerry M. Mendel

Accepted for publication as a chapter in Artificial Intelligence and Expert Systems in Petroleum Exploration, F. Aminzadeh and M. Simaan, co-editors, JAI Press.

HEURISTICALLY CONSTRAINED ESTIMATION FOR INTELLIGENT SIGNAL PROCESSING

Robert F. Popoli and **Jerry M. Mendel**

Signal and Image Processing Institute

Department of Electrical Engineering - Systems

University of Southern California

Los Angeles, CA

ABSTRACT

The solution of many estimation problems can be greatly enhanced by the incorporation of inexact knowledge or vague human reasoning. For such estimation problems, two distinct forms of problem knowledge can be identified: statistical (objective) knowledge and heuristic (subjective) knowledge. This chapter discusses a systematic way of expressing and integrating these two forms of knowledge into the estimation process. This work can be interpreted as a fuzzification of standard constrained optimization. Fuzzy set theory is used to form a fuzzy constraint which represents the domain-specific knowledge of human experts. This work may also be interpreted as a systematization of the use of subjective priors by Bayesians. Although our work is of general applicability, we demonstrate the use of heuristically constrained estimation to the particular problem of seismic deconvolution. These results show that the incorporation of heuristic knowledge (albeit vague) yields better results than if such knowledge is ignored.

INTRODUCTION

Several authors have noted or predicted that advances in seismic signal processessing will focus on the blurring of the distinction between seismic processing and seismic interpretation [15][16]. They note that methodologies must be devised to allow the seismic interpreter to play a more direct role in the processing routines. Such applications demand novel approaches that provide for the systematic integration of inexact knowledge or vague human reasoning into the processing algorithms. The development of estimation techniques for such problems is the focus of our work.

We demonstrate an approach for the integration of human reasoning directly into the seismic deconvolution process. Seismic deconvolution is performed to estimate the earth's layer structure by processing echoes from an energy source at the earth's surface. This problem can be viewed as one of detecting overlapping wavelets in noise. The standard maximum likelihood procedure for this difficult problem [7] leads to an objective function that can rank drastically different earth layer structures as being equally likely. Most of these structures, however, do not represent any heuristically feasible true earth structure; that is, a human data interpreter would easily reject most of these solutions as impossible. One might attempt to use standard search space constraints to eliminate these heuristically infeasible structures; however, human criteria for such judgements involve vague concepts which cannot be coerced into crisply defined constraints.

Our approach shows how human heuristic rules of thumb, expressed with vague linguistic concepts, can be systematically used to heuristically constrain the search. This procedure selects the likely solutions which are more consistent with the interpreter's common sense.

Our methodology, which can be employed in a wide class of signal processing applications, is developed by drawing upon the basic ideas of fuzzy optimization from decision theory, fuzzy logic from fuzzy set theory, and the basic concepts of rule based systems from AI.

Techniques That Are Of Age

The work of many individuals in the diverse fields of mathematics, signal processing, systems engineering, and computer science have set the stage for the advent of "Intelligent" or "Expert Signal Processing". That the time is ripe for such beginnings is evidenced by notable trends in current research.

During the past decade, a good deal of attention in mathematical research has been focused on studying fuzzy set theory as a mathematically sound formulation for human knowledge representation [1][2][3]. This work provides a systematic methodology for the expression of vague human concepts. Another trend, spawned by the introduction of fuzzy set theory, is a general tendency of researchers to fuzzify previously developed crisp methods [8][9]. In particular, the fuzzification of the notion of constrained optimization in decision theory has received at least limited attention [9][10]. During this same period, computer science research has established structures and algorithms for machine intelligence (e.g., [4]). We shall demonstrate how theoretical results from these fields combine to define heuristically constrained estimation (*H.C.E.*), a technique for intelligent signal processing.

Recent literature cites novel signal processing techniques which have begun to tie the available theoretical foundations to today's demanding signal processing applications [5][6][10]. Although not all these authors emphasize the artificially intelligent nature of their novel approaches, the existence of such seedlings is undeniable.

Our Approach

Our approach is to organize the heuristic knowledge of the data interpreter in a simple "If-Then" rule base (as might be done for an expert system). Conceptual meanings of the linguistic elements of these rules are encoded by fuzzy set membership functions. We then employ an inexact reasoning scheme based on fuzzy set theory. This reasoning scheme operates on the rule base to arrive at a fuzzy constraint set. This fuzzy set is then used to heuristically constrain the search space.

Our work is closest technically to the somewhat limited [10] work done in fuzzy optimization. Although there are some differences in formulation between our work and that done in fuzzy optimization (a discussion of the similarities and differences follows later), the major differences are more a matter of the spirit of application. Even though estimation theory can formally be thought of as a branch of decision theory, the emphasis in fuzzy optimization has had a decidedly operations research flavor. Applying the general notion of a fuzzy constraint to estimation applications provides interesting interpretations of classical constrained estimation techniques and of the use of subjective priors by Bayesians.

Expert Signal Processing: An AI Interpretation

We wish to stress the AI nature of heuristically constrained estimation. We now set out to describe what we mean by intelligent signal processing and how the concept is achieved by *H.C.E.*.

To help establish a definition of intelligent signal processing , we quickly review the definitions of general Artificial Intelligence. Elaine Rich in her book entitled *Artificial Intelligence* [4] offers an appropriate definition:

> *Artificial Intelligence is the study of how to make computers do things at which, at the moment, people are better.*

Equally appropriate is Patrick Winston's definition in his book, also entitled *Artificial Intelligence* [12]:

> *Artificial Intelligence is the study of ideas which enable computers to do the things that make people seem intelligent.*

In keeping with such definitions, we define intelligent signal processing as *the study of techniques which extend conventional signal processing toward applications which are normally thought to require some human intervention.* More specifically, intelligent signal processing is the study of ways to merge conventional signal processing goals and techniques with incomplete human *a priori* knowledge and vague human reasoning.

Conventional estimation theory is the study of methods which employ statistical knowledge of a problem to the development of estimators. Our work investigates how this same objective statistical knowledge can be coordinated with the use of domain-specific subjective knowledge to achieve estimators capable of results that normally require human intervention.

Human reasoning is incorporated by virtue of the fuzzy constraints. This amounts to the codification of part of the design process of the estimator. By providing a means with which to codify the reasoning behind the selection of optimization constraints, such constraints may be expressed in the form of a rule base. This brings the heurisitic part of the design process to the surface in a form which is understandable and modifiable by the end user of the estimation algorithm. The end result is an algorithm with less "hard-wired" behavior; one that is more adaptable to the user's subjective viewpoint.

In addition, the use of fuzzy logic to represent conceptual human reasoning allows heuristic constraints to remain vague in nature. This allows for use of subjective domain-specific knowledge which would otherwise be intractable. Further, the automation of such reasoning allows the interaction of heuristic human reasoning to permeate the optimization procedure in a way not otherwise possible. Partial solutions may be checked against human criteria as the optimization proceeds. Such interaction can act to cut down the effort of optimization over large search spaces.

The end result is that *H.C.E.* is a way of coordinating probabilistic knowledge with heuristic domain-specific knowledge for solving hard estimation problems. By developing a methodology for the coordination of these two distinct forms of problem knowledge, we can address the following problematic issues:

- Some problem knowledge is not easily coerced into a statistical model.
- Some problem knowledge can not be tractably coerced into standard (i.e., crisp) optimization constraints.
- Some problem knowledge is too variable to be of general use for all data sets, thus leading to the tendency to off-handedly reject such special purpose information.

The rest of this chapter is organized as follows. First, we outline the nature of the seismic estimation problem. This will provide a frame of reference for the rest of the document. In doing so, we will examine the available problem knowledge. Next, we show how this knowledge can be mathematically encoded into an objective function and fuzzy constraints. We then demonstrate comparative results using and excluding the heuristic knowledge. Finally, we return to our rule set to show how it can be expanded to begin to include some of the practical considerations with which interpreters deal.

A PARTICULAR APPLICATION

Although *H.C.E.* is generally applicable to a wide range of estimation problems with ill defined search space constraints, we are particularly interested here with its application to seismic deconvolution. The central theme of *H.C.E.* is the coordination of objective and subjective problem knowledge. Although it is common practice to include a full discussion of the statistical (objective) modeling assumptions, it is rare to include or attach any significance to heuristic (subjective) problem knowledge. This sort of knowledge is usually considered either useless or untractable and is typically abstracted out of the problem statement. Our application description, on the other hand, will include a discussion of what is known both statistically and heuristically.

The purpose of seismic signal processing is to map out the gross layer structure of the earth in order to detect geological structures which could serve as traps for oil reserves. To accomplish this goal, a seismic experiment along the lines of that shown in Figure 1 is performed.

An explosion serves as a source of acoustic energy which propagates downward into the earth. Images of the acoustic source wavelet $h(k)$ are reflected upward by the interfaces of layers of different acoustic impedances. The reflected images are recorded, in the presence of measurement noise, by a sensor placed at the surface. The measurements are modeled by the convolution of the wavelet h(k) with a spike train u(k), where the time between spikes is related to the distance between layers,

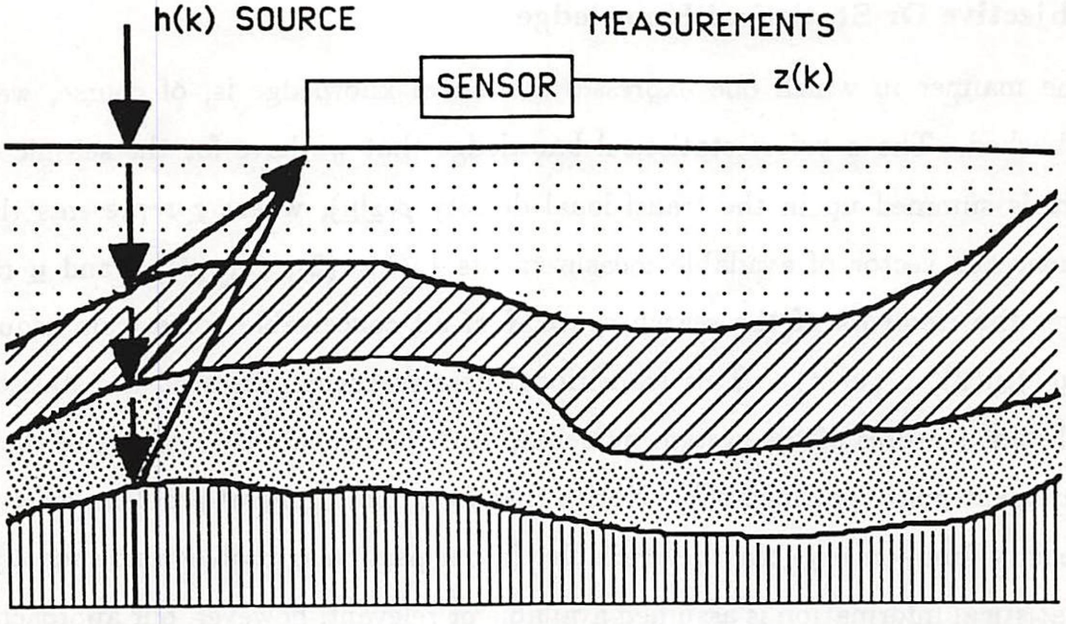

Figure 1: *The Essence Of The Seismic Experiment*

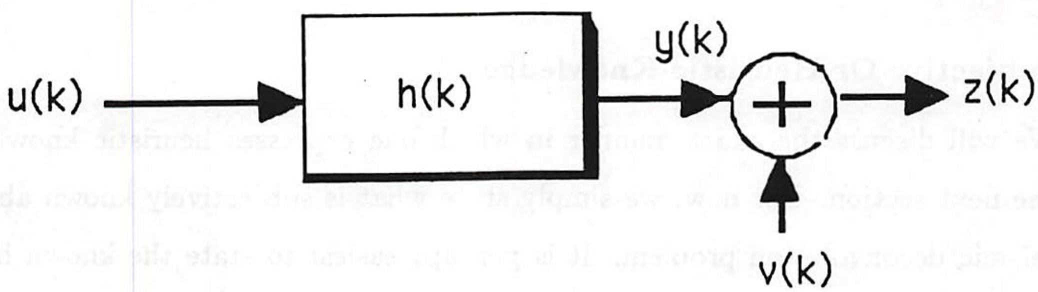

Figure 2: *The Deconvolution Problem*

and the magnitude of each spike is related to changes in acoustic impedance between adjacent layers. Since convolution is commutative, the noise-free measurements $y(k)$ can be viewed as the result of subjecting a linear time-invariant system, whose impulse response is $h(k)$, to a spike train input $u(k)$, as depicted in Figure 2 [14].

The goal in deconvolution is to process the noise corrupted measurements $z(k)$ so as to determine a $u(k)$ which best explains the observed data; i.e., the $u(k)$ which is most consistent with the observed data and the *a priori* knowledge we have about the problem.

Objective Or Statistical Knowledge

The manner in which one expresses statistical knowledge is, of course, well established. The *a priori* statistical knowledge that we have for the seismic problem is summed up in the transitional density $p(\underline{z}|\underline{u})$, where $\underline{z}$ represents the sequence or vector of available measurements $\{z(1), z(2), \ldots, z(N)\}$, and $\underline{u}$ ranges over the elements of the search space S of all conceivable spike train sequences $\{u(1), u(2), \ldots, u(N)\}$. This transitional density encompasses all the statistically relevant features of the system modeling, and the probabilistic nature of the measurement noise sequence $v(k)$. Notice that for a more general problem, an *a priori* density function $p(\underline{u})$ may be available. In our particular problem, no such *a priori* statistical information is assumed available or relevant; however, our approach does not restrict the use of such information. Hence, we do not need to employ the widely used random reflector model which is controversial from a physical point of view [14].

Subjective Or Heuristic Knowledge

We will discuss the exact manner in which one expresses heuristic knowledge in the next section. For now, we simply state what is subjectively known about the seismic deconvolution problem. It is perhaps easiest to state the known heuristic information in the form of rules of thumb. Consider the following:

1) It is best to explain the received measurements by a $u(k)$ which consists of a relatively small number of significant events.
2) In any subsequence of length L, there should not be too many events.
3) In any subsequence of length L, there should not be significant events close together.

The heuristic rationale for these rules is as follows. If our layered model of the earth is correct, rule 1 simply says that we are interested in explaining our measured data by an earth with a small number of distinct layers. Rules 2 and 3 deal with the local nature of the layering. Rule 2 simply says that many layers should not be packed in any small depth region L. Finally, rule 3 says that layers of greatly different impedances should not be very thin.

We will show how the above heuristic rules of thumb can be expressed in a simple mathematical form which is interpreted as a fuzzy subset of the search space S. The general idea is to use these rules of thumb to form a fuzzy constraint set. In some sense, this heuristic knowledge is not as tidy as the statistical information. There may always be more rules to consider, some of which may be conditional on other heuristic inferences made by either the user of the estimator or by other rule based systems. For example, the seismologist may expect a "bright spot" at some approximate depth. The interpreter could easily add a rule which codified this expectation. We will return to the issue of expanding the rule base later in this chapter. Ultimately, an expert system of the form suggested in [2],[3] or [13] could be used to generate the fuzzy constraint set. Such a system could be set up to interface with the seismic interpreter directly, allowing for on-line tailoring of the constraint set.

KNOWLEDGE REPRESENTATION

We have identified two forms of knowledge (objective and subjective) that can be applied to the seismic deconvolution problem. We now turn to the problem of codifying this knowledge. We are primarily interested in examining a methodology for the representation of the subjective problem knowledge; since, probabilistic representations for objective knowledge are so standard.

Before demonstrating how the heuristic rules of thumb listed earlier can be expressed by fuzzy set notions, we first briefly summarize the elements of fuzzy set theory which are important to us.

Fuzzy Set Primer

Consider a normal "crisp" subset A of the universe of discourse X. Denote by x an element of X. Then, the set $A \subset X$ may be defined by identifying the elements $x \in A$. One way to do this is to specify a condition by which $x \in A$. Thus, A can be defined as

$$A = \{x|\ x \text{ meets some condition}\} \tag{1}$$

For example, let X be the universe of discourse composed of all people. Let the set A, for example, be the set of all females. Thus,

$$A = \{x|\ x \text{ is a female}\}. \tag{2}$$

An equivalent way of specifying A is through its membership function.

$$A \rightleftharpoons \mu_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \tag{3}$$

We say that A is mathematically equivalent to its membership function $\mu_A(x)$ in the sense that knowing $\mu_A(x)$ is the same as knowing A itself. One should also realize that knowing the set A is mathematically equivalent to knowing the concept of female (at least in the sense of who is and who is not female).

Next, consider the "fuzzy" set $B \subset X$. Again, B can be defined by a condition any x must meet in order to be in B; except, now the condition can be vague.

$$B = \{x|\ x \text{ meets some vague condition}\} \tag{4}$$

For example, let B be the set of all *tall* people. The problem is that tall is not a crisp notion. Certainly, if person x has height 5'-0", he is not *tall;* whereas, if person x has height 6'-5", he is a member of the set B. On the other hand, what if he has height 5'-7"? The idea behind a fuzzy set is to assign to each $x \in X$ a number in the range [0,1] which represents the degree of agreement between the nature of x and the defining concept of the set. Thus we say that B is mathematically equivalent to its membership function

$$B \rightleftharpoons \mu_B : X \mapsto [0,1] \tag{5}$$

Note that a normal crisp set is indeed a special case of a fuzzy set. The exact definition of this membership function represents the subjective notion of the concept

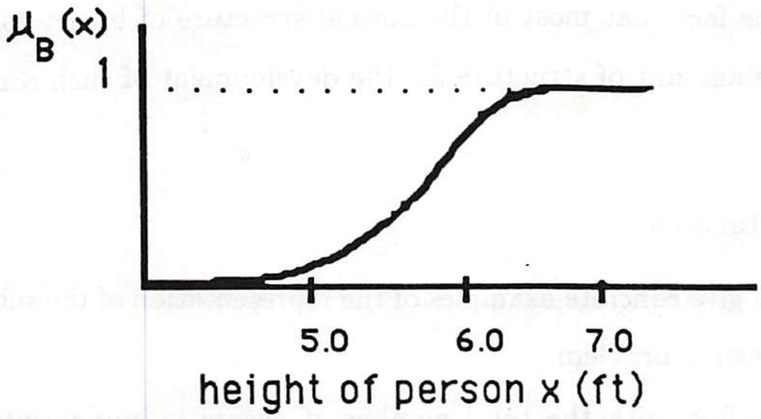

Figure 3: *Membership function expressing the notion of a tall person*

tall. For example, a suitable membership function for the concept of *tall people* represented by the set B may be defined in terms of person x's height as shown in Figure 3.

An important observation is that the notion of *tall* has been encoded by the subset B. Thus, we can think of the subset B as being the subjective concept *tall.*

Finally, we can define fuzzy set logical connectives, **and, or,** and **not** as

$$
\begin{array}{lll}
\cap: & A \cap B \rightleftharpoons & \mu_{A\cap B}(x) = min[\mu_A(x), \mu_B(x)] \\
\cup: & A \cup B \rightleftharpoons & \mu_{A\cup B}(x) = max[\mu_A(x), \mu_B(x)] \\
\overline{\ }: & \overline{B} \rightleftharpoons & \mu_{\overline{B}}(x) = 1 - \mu_B(x)
\end{array} \tag{6}
$$

Note that these definitions correspond to the normal notions of intersection, union and complement when applied to crisp sets (i.e., sets with membership functions that take on only values of 1 and 0). Note, also, that the following properties usually associated with binary logic hold: commutativity, associativity, idempotence, distributivity, De Morgan's law, and absorption. Note, however, that the normal notion of excluded middle does not hold, in that

$$\overline{B} \cup B \neq X \quad \text{and} \quad \overline{B} \cap B \neq \emptyset \tag{7}$$

These logical connectives can be used to define concepts which are composites of already defined concepts. For example, the concept of a *tall female* can be defined by $A \cap B$. The fact that most of the normal structure of binary logic still holds provides a large amount of structure for the development of such composite concepts.

Example Representations

In this sub-section, we give concrete examples of the representation of the subjective information for the seismic problem.

The first rule dealing with the total number of events is implemented as a normal crisp constraint. That is, the optimization procedure is designed not to consider any elements from the search space that have more than a pre-specified number of spikes. For the simulation results presented later this number was set equal to 20. Note however, that in principle, there is no reason why such a constraint could not also be made fuzzy.

The other two rules of thumb can be expressed by the following simple "IF-THEN" rule base, which can be used to infer the subjective acceptability of a member x of the search space S based on examining all subsequences s of length L, where L was subjectively set equal to 30.

r1: **IF:** x has any subsequence s (i.e., **"or"** over all subsequences) with *too-many* events; **THEN:** x is a bad sequence under *r1*.

r2: **IF:** event i is a ***significant-event*** **"and"** event j is a *significant-event*; **THEN:** the pair i, j is a ***significant-pair***.

r3: **IF:** events i and j are close; **THEN:** the pair i, j is a *close-pair*.

r4: **IF:** subsequence s has any pair of events i, j (i.e., **"or"** over all pairs) that is a ***significant-pair*** **"and"** is a *close-pair*; **THEN:** subsequence s has *close-significant-events*.

r5: **IF:** x has any subsequence s (i.e., **"or"** over all subsequences) with *close-significant-events* **THEN:** x is a bad sequence under *r5*.

r6: **IF:** x is a bad sequence under rule *r1* **"or"** *r5*; **THEN:** x is a bad sequence.

r7: **IF:** x is not a bad sequence; **THEN:** x is a good sequence.

This rule base starts with the concepts *close*, *significant-event*, and *too-many* and through the use of the logical connectives generates the concept of the good members of the search space S. Thus this rule base can be used to delineate a fuzzy subset of the search space (in a way, a fuzzy answer) which can serve as the optimization constraint space.

What has been left out of the above discussion is the subjective definitions used to define the root concepts of *close*, *significant-event*, and *too-many*. These concepts are defined in terms of fuzzy subsets of the following domains of discourse: S, the search space consisting of all conceivable spike trains x; B, the space of all subsequences s of any particular sequence x; E, the space of all events within a subsequence s where an event is simply the amplitude and location of a sample point of x; and, D, the space of all interspike distances $|i-j|$ where i and j are event indices in a sequence x.

For our subjective membership functions, we have chosen:

$$\mu_{close}(i-j) = max\left[\frac{15-|i-j|}{15}, 0\right], \tag{8}$$

a subset of D which codifies the concept of all inter-spike distances considered to be close;

$$\mu_{significant\text{-}event}(\text{event} \in s) = \frac{|\text{amplitude of event} \in s|}{|\text{amplitude of max event} \in s|}, \tag{9}$$

a subset of E which codifies the concept of all significant events within the subsequence s of length L of sequence x; and,

$$\mu_{too\text{-}many}(s) = \frac{1}{L}\sum_{i=1}^{L} \mu_{significant\text{-}event}(\text{event i}), \tag{10}$$

a subset of B which codifies the concept of all subsequences s which have *too-many* events.

By using the logical connectives and their properties, the rules which define the composite concepts *close-pair*, *significant-pair*, and *close-significant-events* can be translated as follows:

$$\mu_{close\text{-}pair}(i,j) = \mu_{close}(i-j) \qquad \forall i,j \in E; \tag{11}$$

$$\mu_{significant\text{-}pair}(i,j) = min[\mu_{significant\text{-}event}(i), \mu_{significant\text{-}event}(j)] \qquad \forall i,j \in E; \tag{12}$$

and,

$$\mu_{close\text{-}significant\text{-}events}(s) = \max_{\text{event pairs i j} \in s}[min(\mu_{significant\text{-}pair}(i,j), \mu_{close\text{-}pair}(i,j))] \qquad \forall s \in B \tag{13}$$

Finally, the constraint space is defined, using De Morgan's Law, by the fuzzy subset $C \rightleftharpoons \mu_c(x)$ as

$$\mu_c(x) = \min_{s \in B}[1 - max[\mu_{too\text{-}many}(s), \mu_{close\text{-}significant\text{-}events}(s)]] \qquad \forall x \in S \tag{14}$$

Thus the approach to subjective knowledge representation is to establish the fuzzy sets described by the root functions and then to construct higher order sets from these using fuzzy intersection, union and not. The process chains until finally the constraint set can be established. This process can be diagramatically represented by a tree of fuzzy sets shown in Figure 4. Each rectangle in the tree represents a fuzzy set (note that not all sets need be explicitly named). At the bottom of the tree are the fuzzy sets defined by the root membership functions. Fuzzy sets which are constructed from these lower sets are connected by lines. Lines without an arc represent the **"or"** operation while those with the arc represent the **"and"** operation. The **"not"** operation is represented by a tilde. The tree shows exactly how the constraint set is ultimately constructed from the root functions.

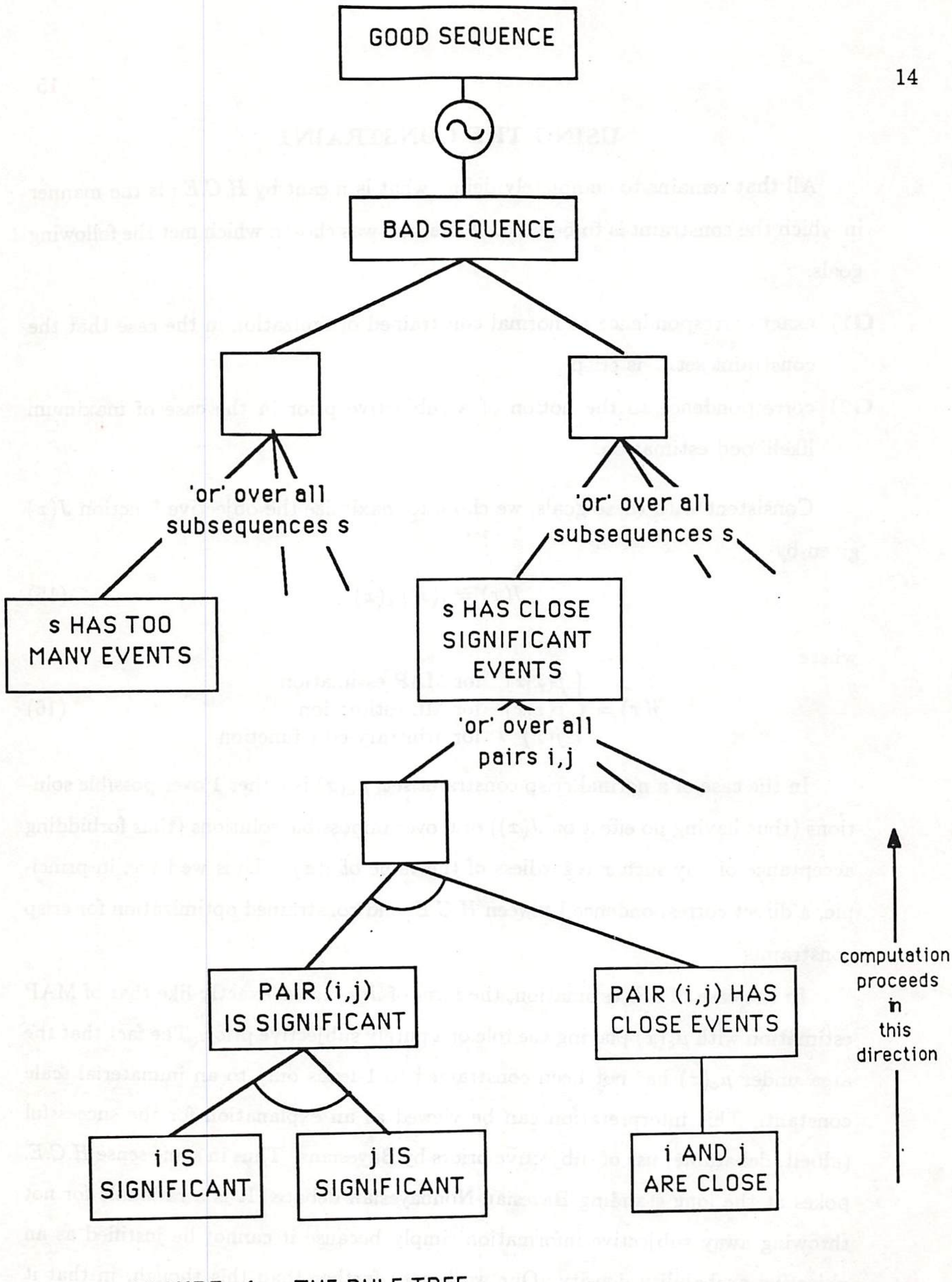

FIGURE 4 THE RULE TREE

USING THE CONSTRAINT

All that remains to completely define what is meant by *H.C.E.*, is the manner in which the constraint is to be used. A method was chosen which met the following goals:

G1) exact correspondence to normal constrained optimization in the case that the constraint set C is crisp

G2) correspondence to the notion of a subjective prior in the case of maximum likelihood estimation.

Consistent with these goals, we chose to maximize the objective function $J(x)$ given by

$$J(x) = j(x)\mu_c(x) \tag{15}$$

where

$$j(x) = \begin{cases} p(x|z) & \text{for MAP estimation} \\ p(z|x) & \text{for ML estimation} \\ j(x) & \text{for arbitrary cost function} \end{cases} \tag{16}$$

In the case of a normal crisp constraint set, $\mu_c(x)$ is either 1 over possible solutions (thus having no effect on $J(x)$) or 0 over impossible solutions (thus forbidding acceptance of any such x regardless of the value of $j(x)$). Thus we have, in principle, a direct correspondence between *H.C.E.* and constrained optimization for crisp constraints.

In the case of ML estimation, the form of $J(x)$ looks exactly like that of MAP estimation with $\mu_c(x)$ playing the role of a purely subjective prior. The fact that the area under $\mu_c(x)$ has not been constrained to 1 leads only to an immaterial scale constant. This interpretation can be viewed as an explanation for the successful (albeit, debatable) use of subjective priors by Bayesians. Thus in some sense *H.C.E.* pokes at the long standing Bayesian-Nonbayesian debate. It is a rationale for not throwing away subjective information simply because it cannot be justified as an objective probability density. Our work goes further than this though, in that it provides a methodology for integrating information that even a subjectivist would not attempt to coerce into a prior density function. Fuzzy logic is in effect a way of systematizing the use, construction, and justification of such subjective densities.

The form of Eq. (15) suggests that one should form a new objective function $J(x)$ from the old one, $j(x)$, and then maximize this new objective function through some gradient search scheme. Although this can certainly be done, there are practical advantages to casting the optimization problem as an equivalent tree search and using the $\mu_c(x)$ to guide the search, which is, in fact, how the simulation results described below were obtained. This approach also motivated writing the rule base to examine subsequences of a potential x. By monitoring subsequences of the solution as the search proceeds, bad branches of the tree are detected and the search proceeds down more profitable branches. The details of the tree search formulation are too involved to be included here, but will be described in a forthcoming publication.

Relation To Fuzzy Optimization

This subsection discusses the similarities and differences between our work and that done in the field of fuzzy optimization as described by Negoita [10]. As defined by Negoita

> *An optimization problem in a fuzzy environment is a system of fuzzy constraints f_i together with an objective function $P: X \mapsto R$ (where R is the real line).*

Under this definition, the solution to the fuzzy optimization problem with m fuzzy constraints corresponds to the points that maximize

$$D(x) = min(P(x), \quad f_1(x), f_2(x), \ldots, f_m(x)) \tag{17}$$

In light of our previous discussion, this may be thought of as

$$D(x) = min(P(x), \quad [f_1 \cap f_2 \cap, \ldots, \cap f_m](x)) \tag{18}$$

Thus *H.C.E.* apparently differs from fuzzy optimization in two ways. First, we allow more complicated forms of connectives in the generation of the overall constraint.

This is actually an insignificant difference in that no restriction was ever placed on the individual constructions of the f_i; thus, they themselves would allow for arbitrary complexity. The second difference is a little more significant. We suggest the product of $P(x)$ and $\mu_c(x)$ rather than the min. We do this so as to achieve the goals **G1** and **G2** set out earlier. Although the min operator could achieve goal 1 with proper normalization of $J(x)$, such a normalization is not always possible, since the range of $J(x)$ is not usually pre-determinable. Furthermore, min will not achieve goal 2.

This difference can be reconciled, though, by first interpreting $P(x)$ itself as some sort of membership function for the set of probable x and then $min[P(x), \mu_c(x)]$ may be interpreted as an intersection. If this is what is intended, then there are other fuzzy set operations [9] (the so called stochastic operators) which allow the interpretation of intersection as membership product. The contortions of the above argument are not as farfetched as one might initially think; since, both probability measures and possibility measures are fuzzy measures [8]. Thus, we see that *H.C.E.*, although different in implementation, shares some of the motivational intent of fuzzy optimization.

RESULTS

This section presents results of applying *H.C.E.* to a maximum likelihood estimator for the seismic problem. Figure 5 shows the overall measurement modelling. The figure shows: an input spike train $u(k)$ which models the layered structure of the earth; the impulse response $h(k)$ of the linear time invariant system which represents the source wavelet; the noise-free convolution $y(k)$ of the spike train and the system's impulse response (note the severe overlapping nature of the wavelet copies); and the available noise corrupted measurements $z(k)$.

Presented in Figures 6 and 7 are simulation results showing the benefit of the integration of heuristic information into the seismic problem solution. The spikes shown were estimated; while, the circles mark the true locations and amplitudes of the actual spikes. Both represent constrained ML estimation of the spike train.

The "no heuristic" case was constrained only by the crisp constraint of "no more that 20 spikes." The "using heuristics" case in addition used the rule base described earlier. The influence of the heuristic is obvious. It is clear that the beginning of the "no heuristic" case is a subjectively unlikely earth layer formation. There are many significant spikes within the early subsequences. The large overlapping wavelets are the cause for the non-heuristically constrained ML estimator's poor performance. The heuristic also managed to note the low possibility of the doublet near time 240.

EXPANDING THE RULE SET

The primary goal of this chapter has been to present a novel approach to signal processing which directly incorporates geophysical reasoning into the estimation process. Our goal was not to demonstrate this geophysical reasoning *per se*. To this end we had set up only a bare minimum of rules.

Collecting the appropriate geophysical knowledge for the specification of a complete knowledge base is not a simple matter. How does one recognize pertinent subjective knowledge and collect it for codification and ultimate use? This problem has its parallel in the development of expert systems. The problematic issue of how one decides what subjective knowledge is useful and how to extract it from the experts of the particular domain of application has been well recognized by the *AI* community. This general topic (called knowledge engineering) has plagued expert system development [18]. It is well recognized by the *AI* community that much of an expert's knowledge has subconscious elements to it. The intricacies in which data is interpreted are not always immediately expressible in words.

The geophysical signal processing engineer must take a far greater responsibility to understand the intricacies of the specific estimation problem at hand. Simple abstracted models are no longer sufficient. Expert sources for this interpretive geophysical knowledge must be sought out and interviewed in an attempt to codify their knowledge.

The goal of this section is to demonstrate the process of knowledge extraction by way of an example. Several good texts have been published which may serve as

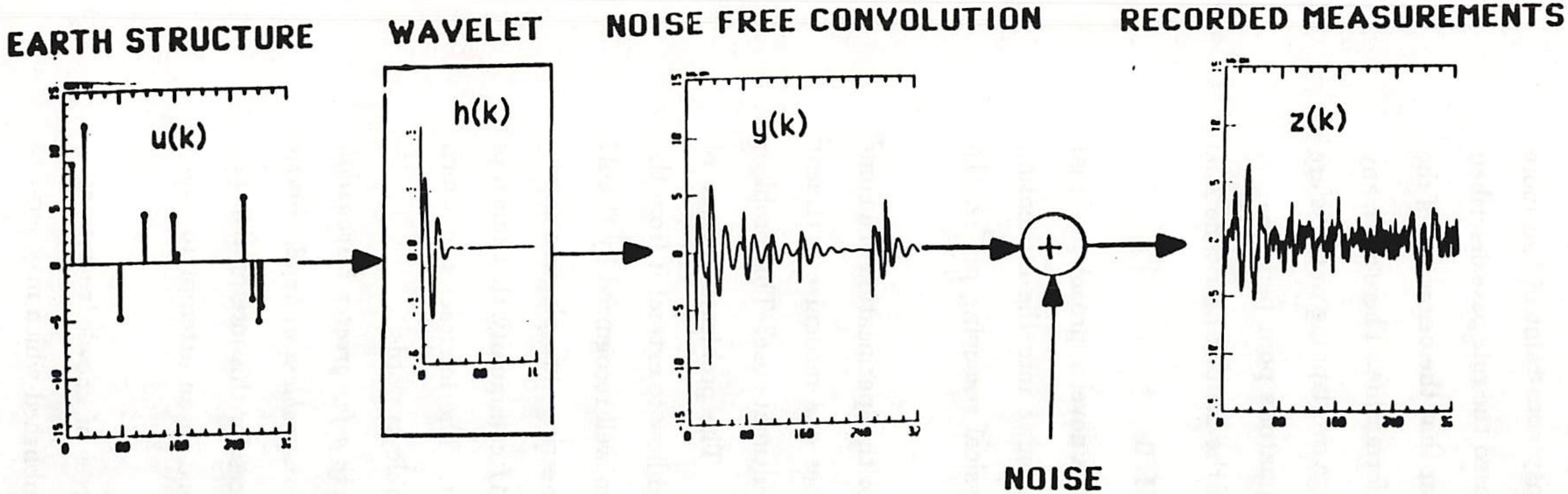

Figure 4: *Measurement Modelling*

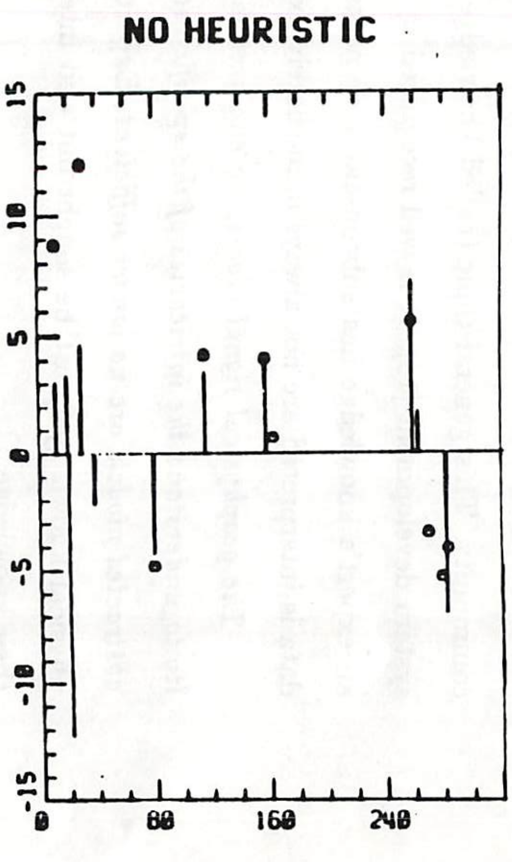

Figure 5: *Standard Deconvolution Results*

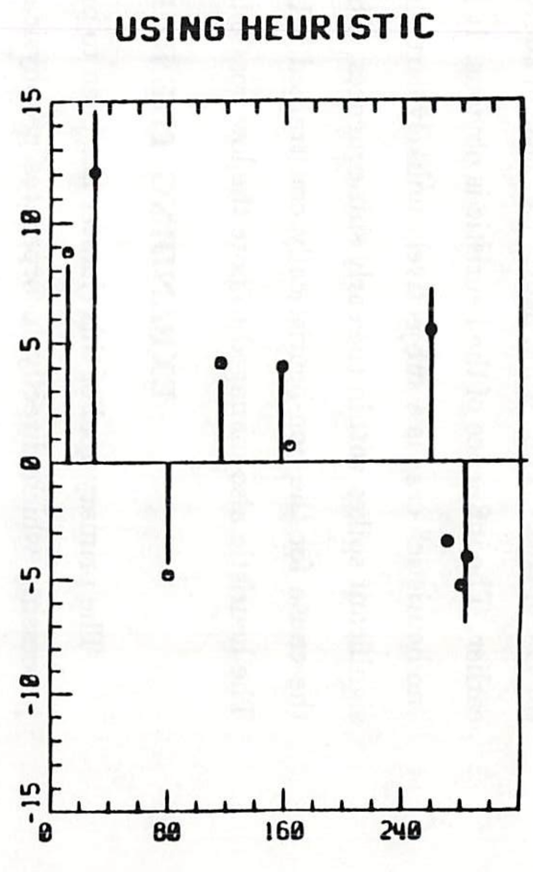

Figure 6: *Deconvolution Results Using Heuristics*

sources of expert geophysical interpretation knowledge. Anstey [17] is a particularly good source and Denham [15] points out several others, as well as providing a general overview of seismic interpretation himself. We will refer to Anstey [17] and Lindseth [16] to show how knowledge about "bright-spots" can be integrated into the small rule base which we have already presented.

Accounting For Bright Spots

Trends in modern seismic signal processing include care to recover and avoid distortion of seismic trace attributes. It has been found that trace attributes such as amplitude can provide valuable intrepretive information. In particular, so called direct hydrocarbon detection often relies on sudden contrast changes in trace amplitude as an indication of the presence of a hydrocarbon accumulation. It has been noted that relatively flat regions of large amplitude returns correlate well with the presence of hydrocarbons, particularly in sandstone source reservoir rock encased in shale. A graphical representation of this sort of trap is shown in Figure 8.

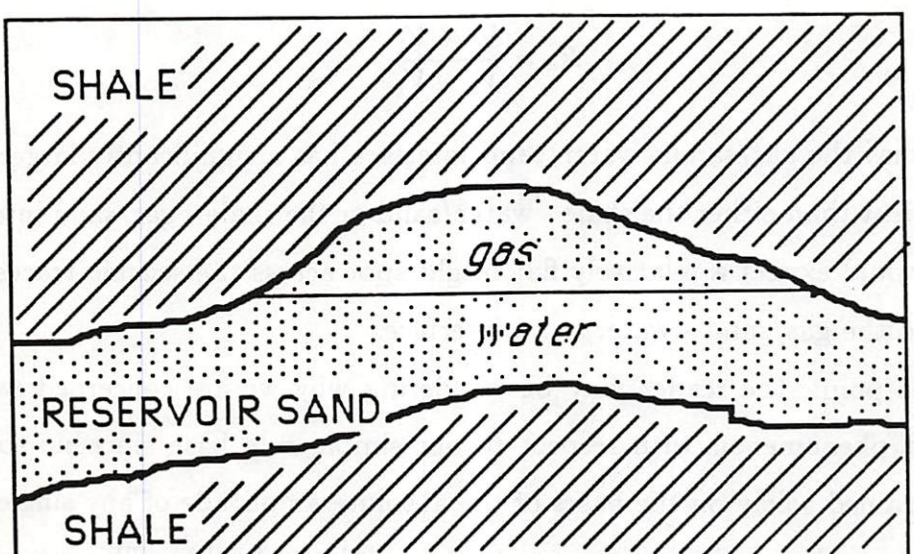

Figure 8: *Physical Model Leading To A Bright Spot*

The geophysical reasons for the bright spot in the recorded data are well understood. The velocity through the porous material of the reservoir rock is a weighted average (weighted by a porosity factor) of the velocity of the dry rock matrix and the velocity through the fluid which fills this matrix. Roughly the composite average velocity V_a in a fluid filled matrix is:

$$\frac{1}{V_a} = \frac{\theta}{V_f} + \frac{1-\theta}{V_m} \tag{19}$$

where

V_a: The average velocity

V_m: Rock matrix velocity

V_f: Fluid velocity

θ: Porosity, expressed as a decimal fraction

Given the velocity in shale, dry sand, gas, and water and by using Eq. (19), the velocities for the various regions shown in Figure 8 will be known, namely: V_{shale}, $V_{gas/sand}$, and $V_{water/sand}$. The amplitudes recorded on the seismic trace correspond to the reflection coefficient R_c at each of the interfaces. Given that R_c at the interface of any regions 1 and 2 is roughly given by

$$R_c = \frac{V_2 - V_1}{V_2 + V_1}, \tag{20}$$

it is clear that the gas/sand - water/sand interface has a significantly larger reflection coefficient than either the shale - water/sand or the shale - gas/sand interfaces. Thus, we would expect a relatively flat bright spot across the seismic traces corresponding to the gas/sand - water/sand interface.

At this point, the reader may be wondering why we are concerned with the bright spot phenomenon with respect to our simple rule base since no solutions were constrained solely on the bases of the absolute amplitude of any single event. Notice however that such sand reservoirs are usually fairly thin. This would imply that in case of the geological structure of Figure 8, it is reasonable to have three reflectors relatively close together. Such a solution would indeed be partially constrained by the heuristic data base as presented. The question arises as to how one might maintain the good effects of our closeness constraint while not de-emphasizing legitimate solutions as depicted in Figure 8.

Consider the following additional pieces of expert geophysical knowledge (direct quotes from expert source [17]):

1) The clearest bright spots exist at times less than about 1.75 - 2 seconds.

2) Fluid contacts are observed, though only when the pay zone is thick.

3) In the identification of fluid contacts (and reservoir boundaries in general), we are often concerned with fairly subtle indicators of interference.

4) The gas-liquid contact must necessarily be a positive reflector.

5) Polarity inversions are observed occasionally. If (and only if) the reservoir material in its water-saturated state has an acoustic impedance greater than that of the overlying material, and if (and only if) the local replacement of water by gas depresses the acoustic impedance of the gas-saturated zone to be less than that of the over lying material, then (and only then) the reflection from the top of the reservoir shows a polarity inversion over the gas.

Fact 1 indicates that our previous rules need only be modified for times less than approximately 2 seconds. Fact 2 affirms our initial worry about the need to note closely spaced events when the reservoir rock is thick relative to the sampling resolution of the trace (but still thin with respect to our close spike concept). Fact 3 says that no special inference can be made with respect to the polarity or large size of any but the gas - water contact. Fact 4 clearly indicates that a bright spot has known polarity.

A suitable extension to our small rule set would be:

r5: **IF:** x has any subsequence s (i.e., "or" over all subsequences) with ***close-significant-events*** which do not constitute a brightspot **THEN:** x is a bad sequence under *r5*.

r8: **IF:** subsequence s is in the ***typical-bright-region*** and is bright **THEN:** subsequence s constitutes a brightspot.

r9: **IF:** subsequence s has a double bright spot **"or"** triple bright spot **THEN:** subsequence is bright.

r10: **IF:** subsequence has two significant events **"and"** one is *locally-significant* **"and"** *positive* **THEN:** subsequence has a double bright spot.

r11: **IF:** subsequence has three significant events **"and"** only the center event is *locally-significant* **"and"** *positive* **THEN:** the subsequence has a triple bright spot.

The goal of these rules is to assure that the geophysically reasonable closely spaced spikes which can occur at a bright spot will not be incorrectly heuristically constrained. Bright spots which do not contain closely spaced spikes require no special attention since they were not heuristically constrained by the original rule set. The required root membership functions necessary to implement these rules (i.e., those for the concepts *typical-bright-region*, *locally-significant*, and *positive*) pose no special problems and can be constructed as we had done with our previous root membership functions. Figure 9 shows the appropriate modifications to the rule tree required by the bright-spot reasoning.

Going Even Further

Notice that all the rules given so far dealt with the codification of inferences made along a single trace. To be sure, many more such inferences could be sought out and codified. However, the scope of our inference structure would be greatly increased by additionally codifying inferences across seismic traces. For example in the case of bright spots: bright spots due to a gas - water contact are expected to be horizontally flat across traces. Another example for an across trace inference is the codification of Fact 5 given above which requires the deduction of a gas - water interface based

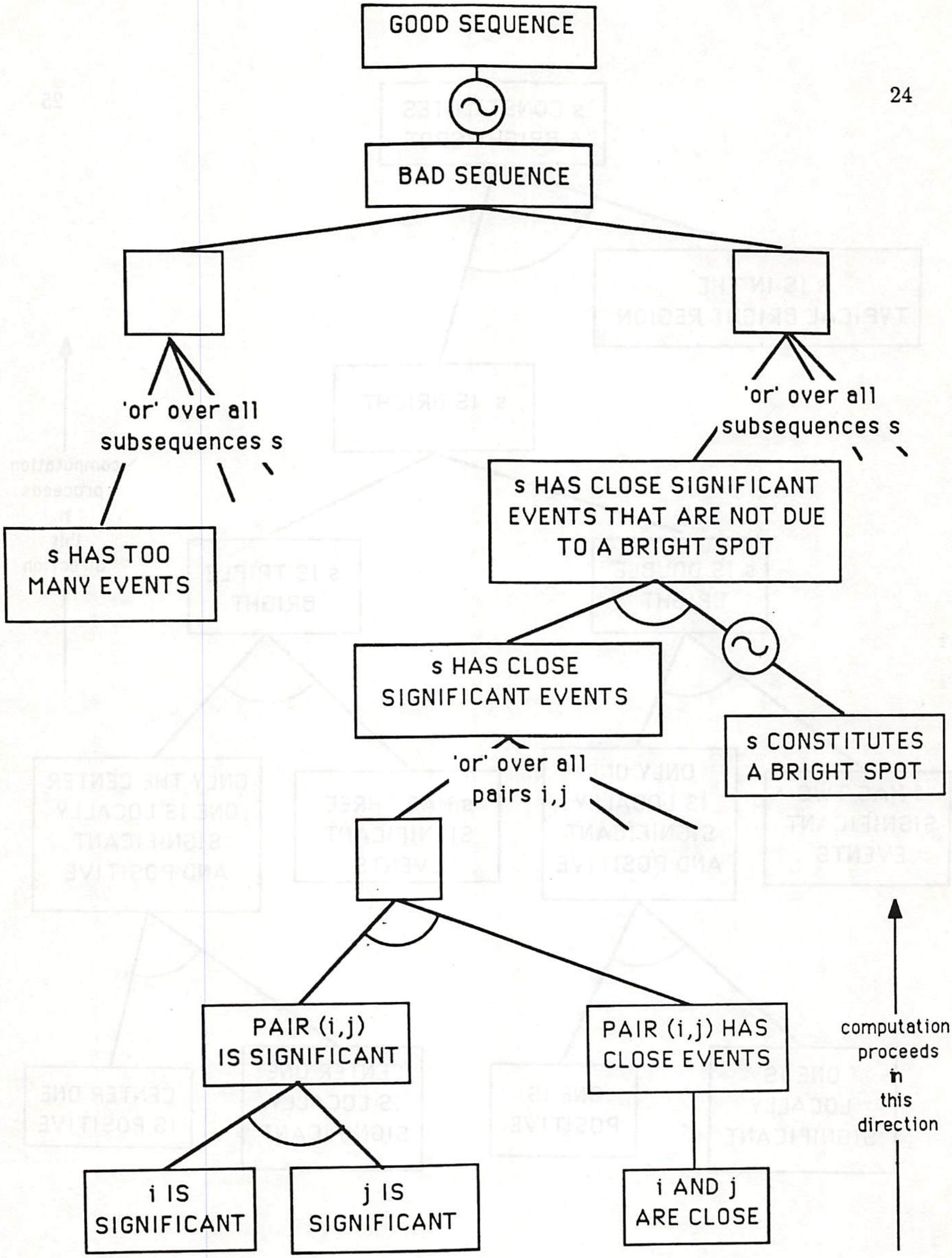

FIGURE 9a EXPANDED TREE

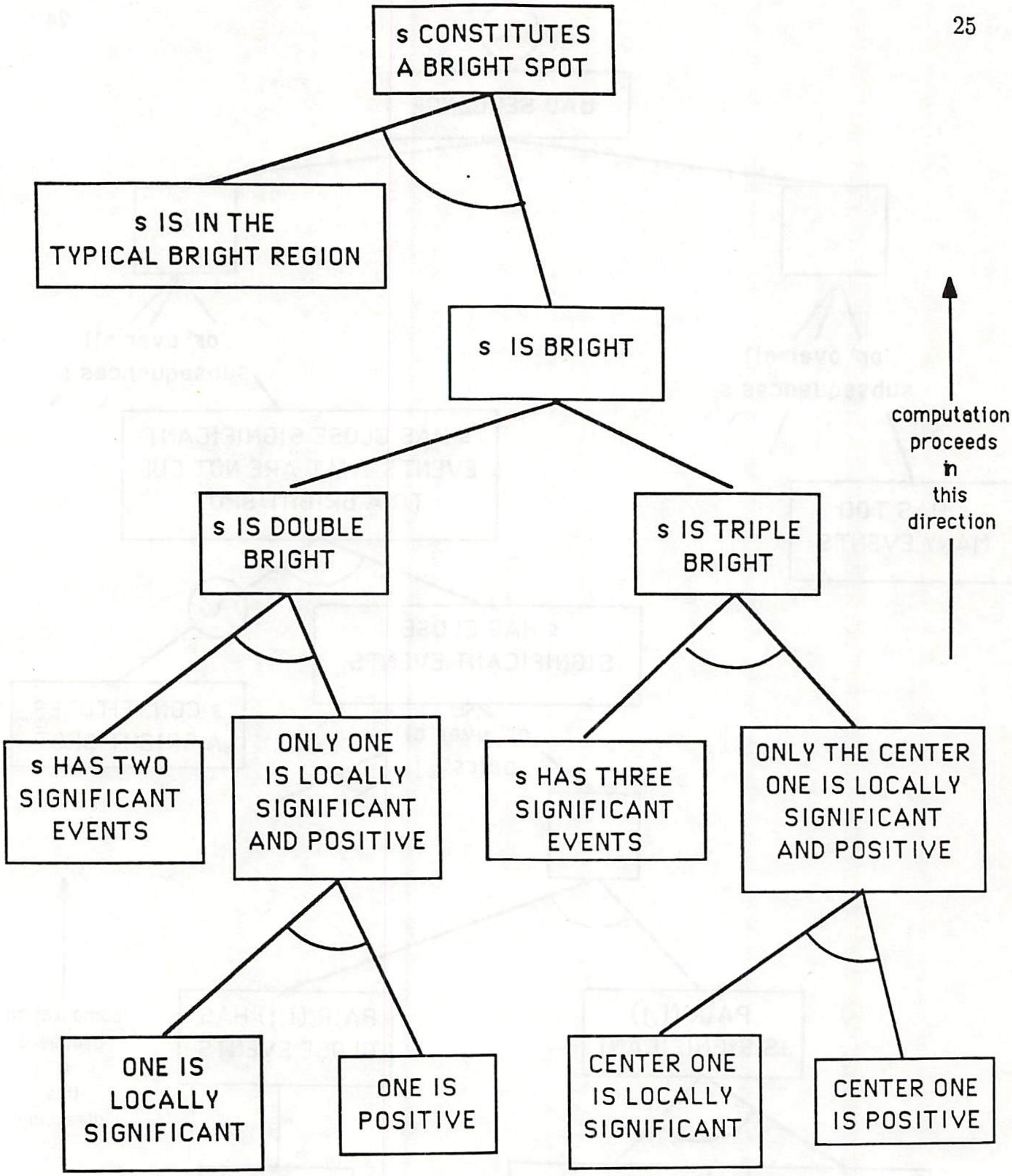

FIGURE 9b EXPANDED TREE

on a polarity reversal for traces which intersect this interface.

We wish to stress that our work is not important because of the (admittedly sparse) rule set which we have presented; rather, our work is important because we have offered a general methodology for the incorporation of expert knowledge and have demonstrated the ease and effectiveness of doing so.

CONCLUSIONS

Our work focuses on the substantiation of three primary ideas:

- For estimation, there are two distinct forms of problem knowledge: objective or statistical, and subjective or heuristic.
- Subjective knowledge can be mathematically codified in a simple tractable rule base form.
- Subjective information can be coordinated with objective information in a manner which is an extension of classical constrained estimation.

We have shown that for the seismic deconvolution problem there exists subjective information which can lead to an enhanced estimator. This information is usually abstracted out of the problem description because it is normally thought to be intractable. We have shown that this subjective knowledge is, in fact, readily codifiable using ideas from fuzzy set theory. Once codified, this heuristic knowledge can be incorporated as a fuzzy constraint on the search space. The overall methodology of *H.C.E.* can be viewed as a fuzzification of classical constrained estimation.

The presented codification of heuristics readily lends itself to a rule base implementation. Thus, there is a natural connection between rule based or expert systems and the *H.C.E.* procedure. Heuristically constrained estimation can be viewed as a methodology to perform *intelligent* or *expert signal processing.*

We have presented simulation results which demonstrate the advantages of incorporating both objective and subjective knowledge. In short, we have shown that mathematical techniques are now available to systematically include a type of knowledge which is essentially ignored by classical approaches. Taking advantage of this additional heuristic knowledge leads to estimators which are more versatile, achieve subjectively better performance, and can automate some human data interpretation functions.

REFERENCES

[1] J. A. Goguen, Jr., "Concept representation in natural and artificial languages: axioms, extensions and applications for fuzzy sets," in *Fuzzy Reasoning And Its Applications,* E. H. Mamdani, B. R. Gaines, Ed., New York, NY: Academic Press, pp. 67-115, 1981.

[2] L. A. Zadeh, "The role of fuzzy logic in the management of uncertainty in expert systems," *Fuzzy Sets And Systems,* Vol. 11, pp. 199-227, 1983.

[3] C. V. Negoita, *Expert Systems and Fuzzy Systems,* Menlo Park, CA: Benjamin/Cummings, 1985.

[4] E. Rich, *Artificial Intelligence,* New York, NY: McGraw-Hill, 1983.

[5] M. R. Civanlar, H. J. Trussell, "Digital signal restoration using fuzzy sets," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-34, pp. 919-936, Aug. 1986.

[6] I. R. Goodman, "PACT: possibilistic approach to correlation and tracking," *Naval Ocean Systems Center Report,* San Diego, CA, 1982.

[7] J. M. Mendel, *Lessons In Digital Estimation Theory,* Englewood Cliffs, NJ: Prentice Hall, 1987.

[8] D. Dubois, H. Prade, *Fuzzy Sets And Systems: Theory and Application,* New York, NY: Academic Press, 1980.

[9] A. Kandel, *Fuzzy Mathematical Techniques with Applications,* Reading, MA: Addison-Wesley, 1986.

[10] C. V. Negoita "The current interest in fuzzy optimization," *Fuzzy Sets and Systems,* Vol. 6, No. 3, pp. 261-270, 1981.

[11] L. A. Zadeh, "Making computers think like people," *IEEE Spectrum,* vol. 21, pp. 26-32, Aug., 1984.

[12] P. H. Winston, *Artificial Intelligence,* Reading MA: Addison-Wesley, 1977.

[13] R. F. Popoli, S. S. Blackman, "Expert system allocation of the electronically scanned antenna radar," *Proc. American Control Conference,* Minneapolis, MN, June 10-12, 1987.

[14] J. M. Mendel, *Optimal Seismic Deconvolution: An Estimation-Based Approach,* New York, NY: Academic Press, 1983.

[15] L. R. Denham, "Seismic Interpretation," *Proceedings Of The IEEE,* Vol. 72, pp.1255-1265, Oct., 1984.

[16] R. O. Lindseth, *Digital Processing Of Geophysical Data: A Review,* Teknica Resource Development LTD., 600,633 Sixth Avenue SW, Calgary, Alberta, Canada, Sept., 1982.

[17] N. A. Anstey, *Seismic Interpretation: The Physical Aspects,* Boston, MA: International Human Resources Development Corporation, 1977.

[18] H.L. Dreyfus, S.E. Dreyfus, "Why Expert Systems Do Not Exhibit Expertise," *IEEE EXPERT,* Vol. 1, Number 2, Summer, 1986.