



Adaptive Distributed Multi-Sensor Detection Using Information Elasticity Approach

RAM NARAYANAN
PENNSYLVANIA STATE UNIVERSITY

01/22/2020
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

DISTRIBUTION A: Distribution approved for public release.

REPORT DOCUMENTATION PAGE					<i>Form Approved</i> OMB No. 0704-0188	
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.						
1. REPORT DATE (DD-MM-YYYY) 20-01-2020		2. REPORT TYPE Final Report			3. DATES COVERED (From - To) 01 Jan. 2017 - 31 Dec. 2019	
4. TITLE AND SUBTITLE Adaptive Distributed Multi-Sensor Detection Using Information Elasticity Approach				5a. CONTRACT NUMBER		
				5b. GRANT NUMBER FA9550-17-1-0032		
				5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) Narayanan, Ram, M Liu, Andrew, Z Singerman, Paul, G				5d. PROJECT NUMBER		
				5e. TASK NUMBER		
				5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The Pennsylvania State University Office of Sponsored Programs 110 Technology Center University Park, PA 16802					8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research 875 N. Randolph, Ste.325, Rm. 3112 Arlington VA 22203-1768					10. SPONSOR/MONITOR'S ACRONYM(S) AFOSR	
					11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT						
13. SUPPLEMENTARY NOTES						
14. ABSTRACT It is believed that better and more effective decisions can be made in radar sensor processing if more 'information' is available. However, this is not always the case since due to the problem of information overload. The concept of information elasticity is useful in determining the relationship and role of information, qualitatively and quantitatively, to its interactions with the surroundings. While Shannon's theory describes information in quantitative terms by assigning information value based on event probabilities, in many instances, information obtained can only be described in fuzzy and qualitative terms. In this project, we apply the concept of information elasticity to radar system performance for various scenarios.						
15. SUBJECT TERMS Information elasticity; information overload; radar cross section; clutter; constant false alarm rate (CFAR); pseudorandom binary sequences; phase code modulation; constraint-based modeling; signal-to-clutter ratio (SCR); uncertainty immunity; robust optimization; cost function design; constraining statements; multi-objective optimization; language modeling						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON	
a. REPORT	b. ABSTRACT	c. THIS PAGE			Ram M Narayanan	
U	U	U	SAR	79	19b. TELEPHONE NUMBER (Include area code) 814-863-2602	

TABLE OF CONTENTS

Background and Motivation.....	3
Information Elasticity Formulation.....	11
References Cited.....	14
List of Publications.....	15
Reprints of Publications.....	16
Information Elasticity in Radar Systems.....	17
Information Elasticity in Radar Systems – Author Interview.....	19
Information Elasticity in Pseudorandom Code Pulse Compression.....	20
Information Elasticity in Ultra-Wideband Target Detection amongst Distributed Clutter...	29
Robust Decision Making Method for Adaptive Ordered-Statistics CFAR Technique using Information Elasticity.....	41
Cost Function Design for Modeling Information Overload in Radar Systems.....	50
Multi-Objective Fully-Adaptive Radar Utilizing Language-Based Cost Functions.....	61
An Information Elasticity Framework for the Adaptive Matched Filter.....	67
Students Supported.....	79
Collaborators.....	79

Background and Motivation

Information Overload and its Effects

One major concern of researchers interested in how individuals exploit available information is the determination of the maximum amount of information that a consumer (i.e., user) can effectively process before experiencing decreased accuracy in assessing the value (utility) of available choice alternatives. Currently, advocates for providing substantial amounts of information often act as if this maximum is a large number. Although it is difficult to dispute the concept of the consumers' or the users' "right to know," some researchers have warned that policies aimed at increasing information can lead to poorer decisions, citing numerous laboratory experiment findings that show humans to be cognitively limited (Keller and Staelin, 1987). We believe that similar is the case for sensors.

In ordinary language, the term "information overload" is often used to convey the simple notion of receiving too much information. Within the research community, this everyday use of the term has led to various constructs, such as cognitive overload, sensory overload, communication overload, knowledge overload, and information fatigue (Eppler and Mengis, 2004). The main focus of these studies is how the performance (in terms of adequate decision making) of an individual varies with the amount of information he or she is exposed to. Researchers across various disciplines have found that the performance (i.e., the quality of decisions or reasoning in general) of an individual correlates positively with the amount of information he or she receives—up to a certain point. If further information is provided beyond this point, the performance of the individual will rapidly decline. The information provided beyond this point will no longer be integrated into the decision-making process and information overload will be the result. The burden of a heavy information load will confuse the individual, affect his or her ability to set priorities, and make prior information harder to recall. Figure 1 provides a schematic version of this discovery. It is generally referred to as the inverted U-curve, which is also expected to hold for sensors (as will be shown later).

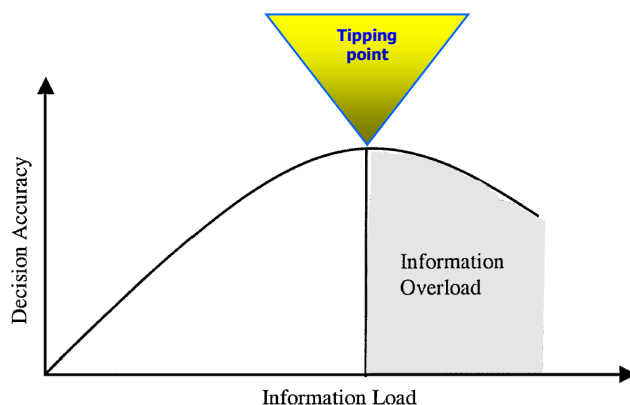


Fig. 1: Information overload as the inverted U-curve.

In the information load paradigm, the dependent variable of interest is the ability of the consumer to make "a correct choice" under different information load conditions. This can be modelled in terms of the probability of making a correct choice (Malhotra et al, 1982). The probability of correct choice can be estimated in several ways. However, in view of the binary nature of the

dependent variable and the fact that the probability of making a correct choice under each experimental condition (cell) varies between 0 and 1, the linear logistic model becomes a particularly suited technique for analysis. The probability of making a correct choice in each experimental cell could be calculated as a function of the number of brands and the number of attributes per brand on which information is provided. Where information overload does occur, the probability of making correct choices should decrease significantly. The linear logistic framework offers an attractive approach for analyzing the information load paradigm and does not require the researcher to adopt a particular definition of “total information.” The methodology is flexible enough to enable the researcher to examine the effects of the number of brands, the number of attributes or other variables of interest on the occurrence of information overload. Moreover, such effects can be examined individually or jointly. Yet the linear logistic model permits the researcher to model various definitions of “total information” and to examine complex interaction effects.

Information overload can be caused by both intrinsic and extraneous factors. Intrinsic factors are fundamental elements of the information overload problem and directly influence information overload. Extraneous factors indirectly contribute to information overload and have a direct effect on intrinsic factors. These are summarized in Table I (Jackson and Farzaneh, 2012). Many of these factors are clearly directly applicable to sensor processing.

TABLE I: FACTORS ASSOCIATED WITH INFORMATION OVERLOAD

Intrinsic Factors	Extraneous Factors	Intrinsic & Extraneous Factors
<i>Quantity of Information</i> • Amount of Information available • Information processing <i>Available Time</i>	<i>Characteristics of Information</i> • Complexity of information • Ambiguity of information • Uncertainty of information <i>Quality of Information</i> • Information usefulness ➢ Relevance ➢ Validity <i>Task and Process Parameters</i> • Task complexity • Task novelty • Task interruption • Multi-tasking <i>Personal Factors</i> • Level of prior experiences • Personal skills • Cognitive style • Individual motivation • Personal situation	<i>Sources of Information</i> • Push versus Pull

Tipping Point is defined as the point at which information overload occurs, i.e. the information load at which the decision accuracy is maximized. It is the critical point of the presented model by which the model is lead from one state to another, overloaded or not-overloaded. Each unit has its

own individual contribution to information overload, which means an increase in scope of one unit can increase or decrease the likelihood of information overload happening. As a result, units can be divided into two groupings: 'For' factors, which increase the likelihood of information overload occurring, and 'Against' factors, which when increased, lessen the possibility of information overload. In addition the classification of the units, shown by Fig. 2, was also determined by how they could counterbalance each other. The success of the model depends on the units being placed on the correct side of the model providing appropriate counterbalance.

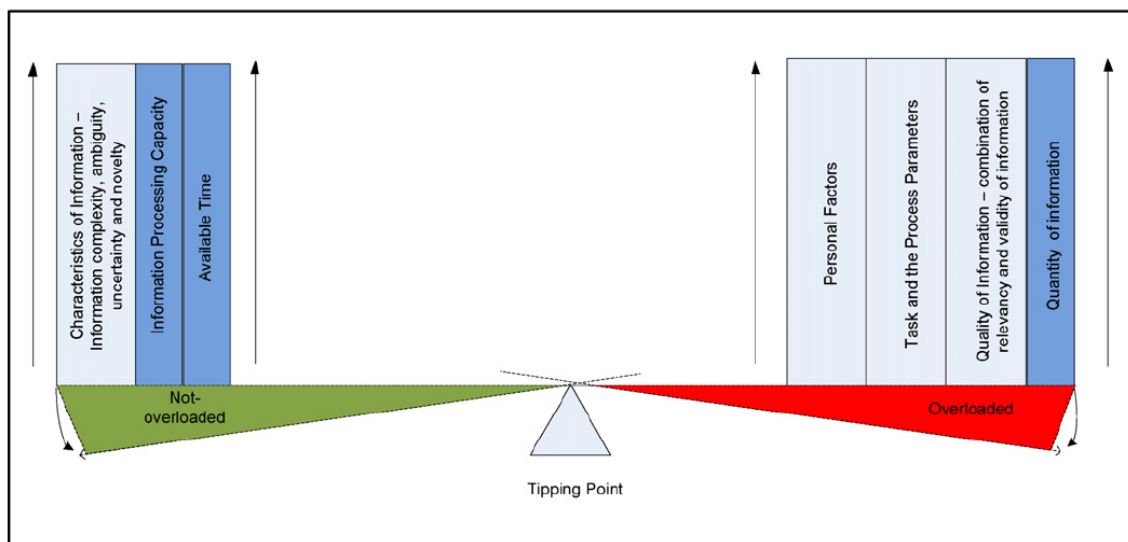


Fig. 2: The conceptual information overload model.

A major problem in decision making is understanding the methods by which users combine information to make evaluative decisions (Einhorn, 1971). The combining process has typically been conceptualized in terms of a linear model in which components are added together to form an overall evaluation. Consequently, linear regression techniques have commonly been used to approximate the cognitive processes underlying decision making. The use of this model implies that the decision maker somehow adds together component cues in order to come to some overall evaluation or decision. One reason for this good fit of the linear model in many different areas of decision making may be that the linear model can still provide a very good approximation even when there are significant nonlinear relationships in the data. While researchers have dealt with the linear model, there are other data-combination models, called nonlinear, noncompensatory models, for combining information. A decision rule is noncompensatory if a decision, determined by some attributes of an object, cannot be reversed by other attributes of the object. These models, viz., the conjunctive, disjunctive, and lexicographic are primarily conceptual. Using a theoretical model based on utility assumptions, precise mathematical functions have been developed to approximate both the conjunctive and disjunctive models.

Elrod et al (2004) proposed and tested a model of decision making that integrated variations of a compensatory and two noncompensatory (i.e., conjunctive and disjunctive) decision strategies. It was capable of providing probabilistic predictions for objects anywhere on the closed interval [0,1]. The model offers a number of important advantages for decision researchers, such as (a) it allows for the identification of compensatory and noncompensatory decision strategies from experimental or nonexperimental choices, eliminating the need for protocol data or self-reports;

(b) it remains true to the conceptualization of noncompensatory heuristics as being cognitively less demanding than compensatory decision making; (c) it allows for the possibility that decision makers may use a compensatory strategy for some attributes, conjunctive or disjunctive for others, and a combination of compensatory and noncompensatory for still other attributes, all within a single decision; and (d) it tests for a pervasive influence of cutoffs, where objects that do not quite invoke a conjunctive or disjunctive rule receive evaluations that are nonetheless influenced by that rule. That is, objects that nearly fail a conjunctive criterion may be devalued, and those that nearly satisfy a disjunctive criterion may be valued more highly, relative to objects that do not nearly invoke the rule.

Boyd and Sworder (1999) developed a mathematical model to simulate information overload. A decision maker's task is to evaluate units of information arriving at random times. The decision maker (user) forms an opinion, based on the (possibly inconsistent) indications of the considered information. This information may or may not be relevant to his or her task, and any unit might or might not accurately indicate the true situation. Determination that a unit of information is irrelevant takes a small amount of time; evaluation of relevant information takes longer. The elapsed time to a prescribed level of confidence provides a metric of decision maker efficiency. The subject's effectiveness depends on how frequently relevant information is presented and to what extent he or she is distracted by the accompanying irrelevant data. The model permits prediction of the optimal rate of presentation of relevant reports as well as analysis of the effects of distracting data on decision maker effectiveness.

The elasticity of information is the term mainly concerned with the usability properties of its qualitative nature and its interactions with the surroundings (Gospodarek, 2008). It is somewhat related to its quantitative aspects as derived from Shannon's theory. Elasticity of information is also connected with the rationality and accuracy of decision making. This concept is based on the definition that information is the processing and organization of data or signals, which causes an increasing of knowledge level of the user or which induces a defined activity. There are several instances where information obtained can only be described in fuzzy and qualitative terms. These include statistical, syntactic, semantic, pragmatic, and apobetic properties of information, each of which has a bearing on the "information value" and its effect on decision making.

Environmental policy has intensively focused on information-based instruments that seek to change agents' behavior through information provision. This information provision is generally considered as likely to ultimately improve environmental quality. (Bougherara et al, 2007) suggest a new and complementary way to consider information-based instruments. They argue that information provision differs from information impact by introducing the concept of informational elasticity. According to their research, beyond an optimum level, an additional information load, regardless of the information quality, could do more harm than good. Indeed, some perverse effects could occur, resulting in a worse overall impact. A similar case can be made for information acquired and processed by sensor systems. A conceptual view of the general mechanism sustaining those policies is shown in Fig. 3.

The information source is the party that holds private information to be delivered to the information provider. The latter gathers information, standardizes it and releases it to the decider. The decider is the target of the policy whose decision is to be influenced. The decider usually takes decisions

in a context more or less loaded with information. Moreover, the decider's decisions may vary in quality. Quality of decisions may be defined as the accurateness of the decider's behavior aimed at leading to a given impact depending on the quantity of information. The decision quality can be viewed as the concordance between the ideal choice according to the agent's preferences among a set of available alternatives and the real choice achieved after information releases.

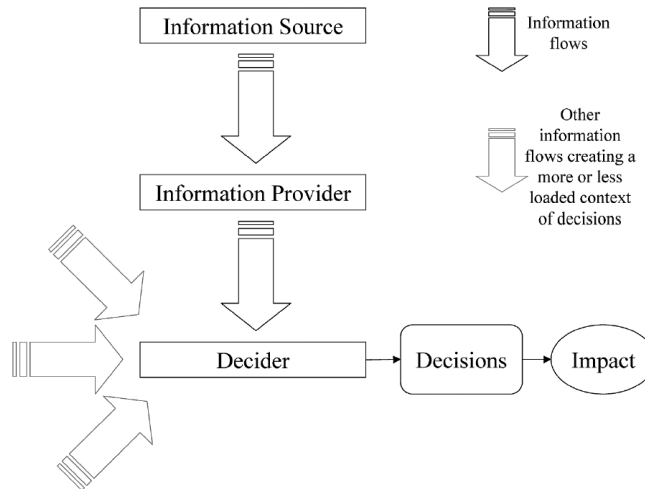


Fig. 3: General mechanism for information-based policies.

(Narayanan, 2012) derived an expression for the ultrawideband (UWB) normalized radar cross section (NRCS) of distributed clutter as a function of the NRCS value at the center frequency and the fractional bandwidth. Higher the fractional bandwidth β at a given frequency, smaller is the value of the range resolution cell, and better is the range resolution. He showed the application of these results in deducing the UWB signal-to-clutter ratio (SCR) for representative targets. Note that the UWB SCR increases with increasing β initially (up to about 35%), and levels off at higher values of β , and then starts to decrease beyond the tipping point, as depicted in Fig. 4.

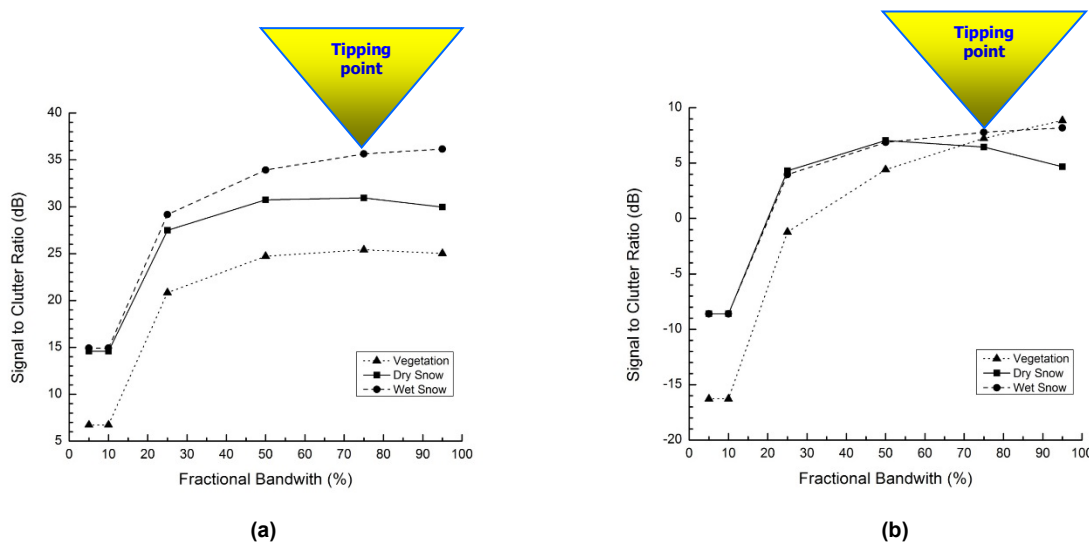


Fig. 4: SCR as a function of fractional bandwidth for (a) corner reflector at a center frequency of 3 GHz, and (b) sphere at a center frequency of 10 GHz.

The implications of these results are quite significant. With the available electromagnetic (EM) spectrum becoming increasingly scarce, a crucial requirement is one of multimodal sensor operation with fully adaptive waveform capability. Operating a radar at the highest bandwidth (and the best resolution) at all times has the disadvantage of requiring more processing and not leaving any bandwidth for other applications, such as communications or telemetry. If SCR enhancement is the goal, then it appears that not much advantage is gained beyond a fractional bandwidth of approximately 35%, after which it appears that the information becomes inelastic.

Information Elasticity Models

Let u_i^* represent the ideal utility for an alternative i based on the available information, where the cost of processing this information is not included in $\hat{u}_i$. Let $\hat{u}_i$ be defined as the user's utility assessment formed under non-ideal environmental and/or processing conditions for alternative i , where $\varepsilon_i = u_i^* - \hat{u}_i$. Small absolute values of ε_i indicate that the user is effective in assessing the ideal or true utility value of the alternative while large values are associated with ineffective assessments. The decision effectiveness y_i^* is then defined as

$$y_i^* = \frac{u_i^*}{u_i^* + |\varepsilon_i|}.$$

Thus, larger values of y_i^* imply greater decision effectiveness. Also, y_i^* is independent of the metric chosen for u_i^* and $\hat{u}_i$.

Both quality as well as quantity of information affect decision effectiveness. It is postulated that (1) increasing the quantity of attribute information while fixing the quality of information decreases the decision effectiveness, and (2) increasing the quality of information while fixing the quantity of information increases the decision effectiveness. Mathematically,

$$y_i^* = -g_1(N) + g_2(TQ) + \varepsilon_{2i}$$

where N is the quantity of information, TQ is the total quality of information, ε_{2i} is the error term, and $g_i(\cdot)$ are the functions reflecting the effects of quantity and quality on decision effectiveness, with derivatives $g'_i(\cdot) > 0$ and $g''_i(\cdot) < 0$. In a radar detection context for example, N could be the fractional bandwidth while TQ could be the SCR.

Although the constructs for the quantity and quality of information are conceptually distinct, they are indeed related. If we define AQ to be the average quality level of the available attributes, we can state, to a first order, that $TQ = AQ \cdot N$. Then,

$$y_i^* = -g_1(N) + g_2(AQ \cdot N) + \varepsilon_{2i}$$

i.e., the decision effectiveness is a function of the quantity and the average quality of the available information.

The information elasticity E is defined as

$$E = \frac{\frac{\Delta TQ}{TQ}}{\frac{\Delta N}{N}}.$$

From Fig. 4, we note that E is proportional to the slope of the SCR (i.e., TQ) plot versus the fractional bandwidth (N), i.e., $E \propto d(TQ)/dN$. The tipping point occurs when the slope is zero, i.e., when $E = 0$.

The operator is usually interested in the overall impact or effectiveness I caused by a change in the release of information. I is a function of both TQ and C , where C is the *context* in which the decisions are made. As stated earlier, $TQ = AQ \cdot N$, i.e., TQ is itself a function of N . Thus, we can write

$$I = I(TQ(N), C(N)).$$

The total differential of I is then

$$dI = \frac{\partial I}{\partial(TQ)} \left(\frac{\partial(TQ)}{\partial N} dN + \frac{\partial(TQ)}{\partial C} \cdot \frac{\partial C}{\partial N} dN \right).$$

The first term in the equation above, namely, $\frac{\partial I}{\partial(TQ)} \cdot \frac{\partial(TQ)}{\partial N} dN$, corresponds to the effect on I caused by a variation in the quality of decisions due to an increase in the quality of information. The overall sign of the ‘quality of decisions effect’ depends upon the sign of $d(TQ)/dN$, which is indeterminate (it can be positive, zero, or negative). On the other hand, the second term, $dI = \frac{\partial I}{\partial(TQ)} \cdot \frac{\partial(TQ)}{\partial C} \cdot \frac{\partial C}{\partial N} dN$, is the indirect effect on I caused by an increase in the informational load of the context. Since the more loaded the context, the harder it is to process the information, the ‘context effect’ is negative. Thus, an adverse effect of the overall impact may occur when the negative effects exceed the positive effects.

In a general case, the quantity of information N could be a vector $\mathbf{N}$ consisting of several parameters.

Application to Adaptive Distributed Multi-Sensor Suite in a Contextual Paradigm

In a typical theater of operation, one can envisage a large number of adaptive distributed networked sensors of different types and modalities (IR, optical, radar, etc.) acquiring and processing different types of data from the target(s) of interest. The varied data are processed within the fusion space to make decisions about the target characteristics to aid in classification. Having *a priori* calibrated knowledge of the scene and the target characteristics, through modeling, simulation, or experiments, will significantly aid in the parametrization and optimal selection of data inputs from each sensor by ensuring that the inputs lie within the region of positive information elasticity.

The multi-agent paradigm may be used in many different ways for structuring the process of fusing data in large networks of information sources. The general goal of such networks is the maximization of the quality of the result to be obtained at minimum cost, e.g. within a minimum time interval using many agents or before a given deadline using as few agents as possible. To this end, agents compete for scarce resources or they work in parallel. Both the general granularity of the agent society and the competence assigned to each individual agent determine the information flow in the network. The multitude of parameters involved makes it difficult for the designer to optimally adapt the structure of the network to a given class of sensing tasks. We propose to develop appropriate network structures and present an approach for analyzing a number of important parameters characterizing the network at a relatively abstract level. The abstraction enables a comparison of different network structures. The methods for the analysis may, however, be readily refined to evaluate a specific problem. A model of cognitive control in sensor networks is proposed as a result of this comparison. This model is based on the concept of collaborative cooperation between pairs of autonomous sensor agents. Typically, distributed sensor agents communicate with each other, thus enabling them to fuse the information they acquired. This implementation is facilitated through network self-organization, i.e., the automatic adaptation to changing environmental conditions. The cooperation is akin to human social behavior in the case of a task being addressed sequentially by team members with potentially different capabilities and reasonings. Suitable cooperation protocols may be developed to achieve the desired coordination. Our qualitative reasoning will be enhanced by theoretical and simulation results to investigate and assess lateral over pure vertical coordination, specifically under confounding environmental conditions, such as sensor degradation or sensor failure.

Specific attention must be paid to managing constraints. It is important to capture and represent a certain type of contextual knowledge in the system as a list of contextual constraints. Constraints are what keep the sensors from reaching the overall goal. In the information processing context, these include size, weight, and power (SWaP) constraints, which will limit the processing power of the hardware in order to meet the SWaP requirements. Another constraint is the maximum available clock speed, which will put a cap on the throughput.

Information Elasticity Formulation

We define elasticity as

$$\varepsilon = \frac{dE/E}{dQ/Q}$$

where E is the decision effectiveness and Q is the information quantity.

The decision effectiveness is a function of the decision metric D and the constraint function C , i.e.

$$E(Q) = f(D(Q), C(Q)).$$

Generally, $D(Q)$ initially increases with Q , levels off, and may eventually decrease at high values of Q , under no constraints. $C(Q)$ always decreases as Q increases. Thus, we have

$$dD/dQ > 0$$

and

$$dC/dQ < 0.$$

This allows us to express

$$dC/dQ = -|dC/dQ|.$$

The function $C(Q)$ could be expressed in many ways to capture its negative trend with Q , such as:

$C(Q) = Q^{-p}$ (assuming that the processing time T is positively correlated with Q , with no limitations on T)

or as:

$C(Q) = Q_{\max} - Q$ (assuming that there is a maximum processing time allowed, $T_{\max}$, which occurs at some $Q_{\max}$).

A simple model for E is a exponentially weighted product of D and C , i.e.

$$E = D^m C^n.$$

Thus,

$$\ln E = m \ln D + n \ln C.$$

Therefore, differentiating with respect to Q , we get

$$\frac{1}{E}(dE/dQ) = \frac{m}{D}(dD/dQ) + \frac{n}{C}(dC/dQ).$$

Multiplying throughout by Q , we can write

$$\frac{Q}{E}(dE/dQ) = \frac{mQ}{D}(dD/dQ) + \frac{nQ}{C}(dC/dQ),$$

which can be simplified as

$$\frac{dE/E}{dQ/Q} = m \frac{dD/D}{dQ/Q} + n \frac{dC/C}{dQ/Q}.$$

We define the following partial elasticities with respect to D and C ,

$$\begin{aligned}\varepsilon_D &= \frac{dD/D}{dQ/Q} \\ \varepsilon_C &= \frac{dC/C}{dQ/Q}\end{aligned}$$

wherein ε_D is positive and ε_C is negative.

Thus,

$$\varepsilon = m\varepsilon_D + n\varepsilon_C = m\varepsilon_D - n|\varepsilon_C|.$$

At the tipping point, $\varepsilon = 0$, and thus, the tipping point occurs at a Q value when

$$m\varepsilon_D = n|\varepsilon_C|.$$

Clearly, the tipping point depends upon the functional dependence models $D(Q)$ and $C(Q)$, and also on the selected weights m and n . These are within the control of the analyst.

If several decision metrics D_k ($k = 1, \dots, K$) and constraint functions C_l ($l = 1, \dots, L$) are used following the same multiplicative product model, we can infer that the tipping point will occur when (see below):

$$\sum_{k=1}^K m_k \varepsilon_{D_k} = \sum_{l=1}^L n_l |\varepsilon_{C_l}|.$$

General model (more than one decision metric and/or one constraint function)

This model assumes there are several decision quality metrics $D_k(Q)$, $k = 1, \dots, K$ and several constraint functions $C_l(Q)$, $l = 1, \dots, L$.

Then,

$$E(Q) = f(D_1(Q), \dots, D_K(Q), C_1(Q), \dots, C_L(Q)).$$

Using a similar multiplicative model, we have

$$E = \prod_{k=1}^K D_k^{m_k} \prod_{l=1}^L C_l^{n_l}.$$

We define partial elasticities as follows:

$$\begin{aligned} \varepsilon_{D_k} &= \frac{dD_k/D_k}{dQ/Q} \\ \varepsilon_{C_l} &= \frac{dC_l/C_l}{dQ/Q} \end{aligned}.$$

Then, we can show that

$$\varepsilon = \sum_{k=1}^K m_k \varepsilon_{D_k} + \sum_{l=1}^L n_l \varepsilon_{C_l} = \sum_{k=1}^K m_k \varepsilon_{D_k} - \sum_{l=1}^L n_l |\varepsilon_{C_l}|.$$

At the tipping point, $\varepsilon = 0$, and thus, the tipping point occurs when

$$\sum_{k=1}^K m_k \varepsilon_{D_k} = \sum_{l=1}^L n_l |\varepsilon_{C_l}|.$$

For $k = l = 1$, we can write $\varepsilon_{D_1} = \varepsilon_D$ and $\varepsilon_{C_1} = \varepsilon_C$. Also assuming $m_1 = n_1 = 1$, we have, at the tipping point,

$$\varepsilon_D = |\varepsilon_C|.$$

References Cited

- Boyd, J.E., and Sworder, D.D. 1999. Modeling information overload. In *Proceedings of the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*, Orlando, FL, November–December 1999, pp. 1-9.
- Bougherara, D., Grolleau, G., and Mzoughi, N. 2007. Is more information always better? An analysis applied to information-based policies for environmental protection. *International Journal of Sustainable Development* **10**(3): 197-213.
- Einhorn, H.J. 1971. Use of nonlinear, noncompensatory models as a function of task and amount of information. *Organizational Behavior and Human Performance* **6**(1): 1-27.
- Elrod, T., Johnson, R.D., and White, J. 2004. A new integrated model of noncompensatory and compensatory decision strategies. *Organizational Behavior and Human Decision Processes* **95**(1): 1-19.
- Eppler, M.J., and Mengis, J. 2004. The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The Information Society* **20**(5): 325-344.
- Gospodarek, T. 2008. Elasticity of information. In *Proceedings of 14th International Congress of Cybernetics and Systems of WOSC*, Wrocław, Poland, September 2008, pp. 511-520.
- Jackson, T.W., and Farzaneh, P. 2012. Theory-based model of factors affecting information overload. *International Journal of Information Management* **32**(6): 523-532.
- Keller, K.L., and Staelin, R. 1987. Effects of quality and quantity of information on decision effectiveness. *Journal of Consumer Research* **14**(2): 200-213.
- Malhotra, N.K., Jain, A.K., and Lagakos, S.W. 1982. The information overload controversy: An alternative viewpoint. *Journal of Marketing* **46**(2): 27-37.
- Narayanan, R.M. 2012. Ultrawideband normalized radar cross sections of distributed clutter. In *Proceedings of IEEE Radar Conference (RadarCon 2012)*, Atlanta, GA, May 2012, pp. 796-800.

List of Publications

R.M. Narayanan, A.Z. Liu, P.G. Singerman, and M. Rangaswamy, “Information Elasticity in Radar Systems,” *Electronics Letters*, Vol. 54, No. 17, pp. 1049–1051, 23 August 2018.

Author Interview: *Electronics Letters*, Vol. 54, No. 17, p. 1014, 23 August 2018.

A.Z. Liu, R.M. Narayanan, and M. Rangaswamy, “Information Elasticity in Pseudorandom Code Pulse Compression,” *Proc. SPIE Conference on Radar Sensor Technology XXII*, Orlando, FL, Vol. 10633, pp. 106330F-1–106330F-9, doi: 10.1117/12.2305853, April 2018.

P.G. Singerman, R.M. Narayanan, and M. Rangaswamy, “Information Elasticity in Ultra-Wideband Target Detection amongst Distributed Clutter,” *Proc. SPIE Conference on Radar Sensor Technology XXII*, Orlando, FL, Vol. 10633, pp. 106330G-1–106330G-12, doi: 10.1117/12.2305854, April 2018.

A.Z. Liu, R.M. Narayanan, and M. Rangaswamy, “Robust Decision Making Method for Adaptive Ordered-Statistics CFAR Technique using Information Elasticity,” *Proc. SPIE Conference on Radar Sensor Technology XXIII*, Baltimore, MD, Vol. 11003, pp. 1100308-1–1100308-9, doi: 10.1117/12.2519677, April 2019.

P.G. Singerman, R.M. Narayanan, and M. Rangaswamy, “Cost Function Design for Modeling Information Overload in Radar Systems,” *Proc. SPIE Conference on Radar Sensor Technology XXIII*, Baltimore, MD, Vol. 11003, pp. 110030A-1–110030A-11, doi: 10.1117/12.2519726, April 2019.

P.G. Singerman, S.M. O’Rourke, R.M. Narayanan, and M. Rangaswamy, “Multi-Objective Fully-Adaptive Radar Utilizing Language-Based Cost Functions,” *Proc. IEEE International Radar Conference*, Washington, DC, April 2020 (accepted for publication).

R.M. Narayanan, A.Z. Liu, and M. Rangaswamy, “An Information Elasticity Framework for the Adaptive Matched Filter,” *IEEE Transactions on Aerospace and Electronic Systems* (under preparation).

Reprints of Publications

The following pages include reprints of all papers, both published and under preparation, as listed in the previous section.

Information elasticity in radar systems

R.M. Narayanan[✉], A.Z. Liu, P.G. Singerman and M. Rangaswamy

It is generally assumed that better and more effective decisions can be made in radar sensor processing if more ‘information’ is available. However, this is not always the case since more information may not necessarily translate to better sensor performance due to information overload. The concept of information elasticity is useful in determining the relationship and role of information, in both qualitative and quantitative terms, to its interactions with the surroundings. While Shannon’s well-known theory describes information in quantitative terms by assigning information value based on event probabilities, there are several instances where information obtained can only be described in fuzzy and qualitative terms. In this letter, the application of information elasticity to radar system performance is introduced.

Introduction: A major concern in information exploitation is the determination of the optimum information required to execute a specific task. Decision effectiveness is related to information quality and information quantity. Generally, decision effectiveness first increases to a peak value and then decreases as the amount of available information increases, for a fixed average quality level of the informational environment [1]. This result is a consequence of information overload [2]. The burden of a heavy information load will confuse the decision maker, affect the ability to set priorities, and make prior information harder to recall. Fig. 1 shows the relationship between decision effectiveness and the quantity of information, which is referred to as the inverted U-curve.

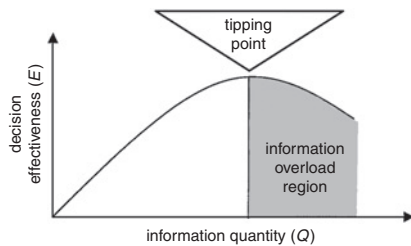


Fig. 1 Typical inverted U-curve depicting decision effectiveness as a function of information load or quantity

Information elasticity: Recently, the concept of ‘information elasticity’ has been proposed which explores the usability properties of information and its interactions with the surroundings [3]. The elasticity of any differentiable function is defined as the ratio of incremental change of the function with respect to the incremental change of its argument. While Shannon’s formulation presents a quantitative descriptor of information, there are several cases where it can only be described in fuzzy and qualitative terms [3].

The information elasticity, ε , is defined as [4]

$$\varepsilon = \frac{dE/E}{dQ/Q}, \quad (1)$$

where E is the decision effectiveness and Q is the information quantity. The decision effectiveness is a function of, in general, several decision quality metrics $D_k(Q)$, $k = 1, \dots, K$ as well as contextual constraint functions $C_l(Q)$, $l = 1, \dots, L$, both types depending on Q . $D_k(Q)$ always increases with Q . On the other hand, $C_l(Q)$ always decreases with Q . Thus, we have $dD_k/dQ > 0$ and $dC_l/dQ < 0$. We can express $dC_l/dQ = -|dC_l/dQ|$. The variation of $E(Q)$ with Q follows the dependence shown in Fig. 1.

A simple model for E is an exponent-weighted product of all D_k and C_l values, i.e. $E(Q) = f(D_1(Q), \dots, D_K(Q), C_1(Q), \dots, C_L(Q))$ given as

$$E = \prod_{k=1}^K D_k^{m_k} \prod_{l=1}^L C_l^{n_l}, \quad (2)$$

where m_k and n_l are the respective exponents. Defining the partial elasticities as $\varepsilon_{D_k} = ((dD_k/D_k)/(dQ/Q))$ and $\varepsilon_{C_l} = ((dC_l/C_l)/(dQ/Q))$,

we can show, via differentiation and rearrangement, that

$$\varepsilon = \sum_{k=1}^K m_k \varepsilon_{D_k} + \sum_{l=1}^L n_l \varepsilon_{C_l} = \sum_{k=1}^K m_k \varepsilon_{D_k} - \sum_{l=1}^L n_l |\varepsilon_{C_l}|. \quad (3)$$

The optimum value of Q for the specific application under the specific constraints occurs at the tipping point where $dE/dQ = 0$, i.e. where $\varepsilon = 0$. This occurs at a value of Q where

$$\sum_{k=1}^K m_k \varepsilon_{D_k} = \sum_{l=1}^L n_l |\varepsilon_{C_l}|. \quad (4)$$

In typical radar applications, examples of the decision quality metrics D_k include probability of detection and signal-to-clutter ratio (SCR), while examples of the constraint functions C_l include sensor processing time and available power. The decision effectiveness may also be construed as a utility metric characterising overall radar performance.

In the examples considered below, we used a single decision quality metric $D_1 = D$ and a single constraint function $C_1 = C$. We also assumed $m_1 = m = 1$ and $n_1 = n = 1$ for a simplified analysis.

Phase-coded modulation: Phase-coded waveforms are frequently used in radar to achieve good range resolution and to reduce interference between proximate radars. A long pulse is subdivided into sub-pulses of equal duration, and the phase of each sub-pulse is selected in accordance with a given code sequence [5]. Maximal-length sequences (MLS), generated using linear-feedback shift-registers, have a structure similar to random sequences (thus also known as pseudorandom binary sequences) and possess desirable autocorrelation properties. MLS waveforms exist only in discrete sequence lengths of $N = 2^n - 1$ bits, where n is the number of stages in the shift-register generator.

One of the problems in high-resolution radars is the occurrence of range sidelobes, sometimes referred to as time sidelobes. These sidelobes may sometimes pick out a stronger target which may vie for attention with weaker targets in the main lobe. Therefore, a lower peak-to-sidelobe ratio (PSLR) is desirable. In our analysis, the PSLR is used as the decision quality metric. Since the pulse compression varies for each cyclic permutation of a sequence, a cross-correlation must be taken for each cyclic permutation. The maximum sidelobe level is found for each permutation, and the permutation with the lowest maximum sidelobe is selected.

The PSLR is taken from the best sequence permutation at each discrete sequence length. As the sequence length increases, the PSLR increases, effectively improving both pulse compression and decision quality. The increase in PSLR slows down at higher sequence lengths. For large N , it has been shown that the PSLR increases linearly as $\sqrt{N}$ [6].

However, longer sequences require longer correlations; thus, the processing time for the cross-correlation operation increases with sequence length. The actual processing time for cross-correlation, measured at each discrete sequence length, shows a near linear relationship between processing time and sequence length.

We define the decision effectiveness as the ratio of the PSLR to the processing time required to produce that pulse compression. The unit for this is s^{-1} , which essentially describes the PSLR achieved per second of processing time. This variable is used to exhibit trade-off qualities between two parameters as the sequence length increases.

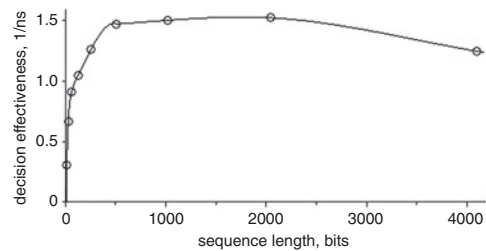


Fig. 2 Detection effectiveness versus sequence length

Fig. 2 shows the plot between decision effectiveness (in ns^{-1}) and information quantity, which is measured by the sequence length. This plot shows that our decision effectiveness increases with sequence length up until length 2047. Beyond this point, the trade-off of

processing time begins to take over, lowering the decision effectiveness. At 2047 bits, the plot displays the maximum value at the tipping point. Beyond the tipping point, information overload occurs, which is represented by the decreasing decision effectiveness in this specific case.

Target detection in clutter: Ultrawideband (UWB) radars are used to enhance spatial resolution in the range direction since resolution is inversely proportional to the bandwidth. Owing to the competing needs for the scarce electromagnetic (EM) spectrum in multifunctional systems, its preservation is critical in many circumstances. Operating a radar at maximum bandwidth has the major drawback of needlessly using up a larger piece of the EM spectrum which can limit the use of other devices.

An important decision quality metric in target detection is the SCR which depends on the bandwidth used. Continuous, uniformly distributed land clutter, such as snow, sand, or dirt, has a radar cross-section (RCS) that increases with frequency, at the rate of f^n where $n \approx (1/2)$, typically. Radar targets show a similar frequency dependence with $n \in [-4, +2]$ depending on target type. It has been shown for several combinations of target and clutter types that the UWB SCR increases at the rate of ~ 0.75 dB for a 1% increase in fractional bandwidth (FB) in the range 10–30%, levelling off at about 30–35% FB, and then decreasing slowly [7].

Data processing time increases with FB to perform required operations such as digitisation, matched filtering, and correlation, and thus acts as a constraint. If the size of the frequency bins is kept constant, a higher FB will have more frequency bins causing an increase in processing time. Combining the processing time along with the SCR, an appropriate model for data effectiveness can be developed. In the range of FB where SCR increases, the corresponding increase in the processing time can result in the onset of zero and negative information elasticity.

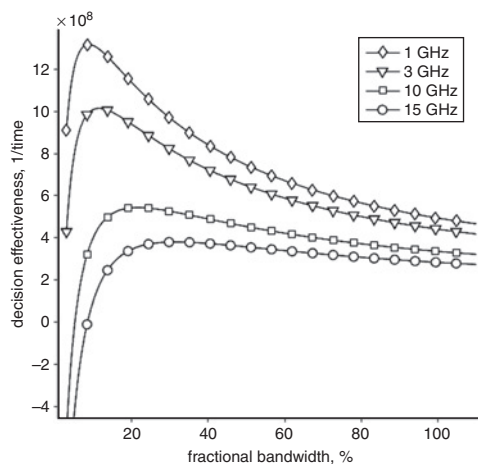


Fig. 3 Decision effectiveness versus FB for different centre frequencies for detecting a metal sphere in wet snow clutter

Fig. 3 shows the data effectiveness for detecting a 1 m radius metal sphere in wet snow clutter as a function of FB percentage. Since the metal sphere circumference is over a magnitude larger than the wavelength at the lowest frequency, the monostatic RCS is constant with respect to frequency and is equal to the projected disk area. Data on

the wet snow clutter were obtained from measurements reported in [8]. At a higher centre frequency, the data effectiveness is observed to be lower at all FB values. This is due to increased RCS response at higher frequencies from the clutter while the metal sphere's RCS response stays constant. Furthermore, at higher frequencies, the total usable bandwidth (in Hz) is greater for the same FB, which in turn increases the number of frequency bins causing a rise in processing time. The tipping point in the data effectiveness ranges from an FB value of 7.5% at a centre frequency of 1 GHz and to an FB value of 30% at a centre frequency of 15 GHz. This indicates that increasing the FB and therefore the actual bandwidth beyond the tipping point may not be beneficial for detecting the metal sphere in wet snow clutter under the processing time constraint.

Conclusions: The information elasticity formulation permits the selection of the proper quantity of information for optimum decision effectiveness in typical radar applications. The tipping point for the decision effectiveness will depend on the assigned weights for each decision quality metric and constraint function.

Acknowledgment: This work was supported by the US Air Force Office of Scientific Research Contract Number FA9550-17-1-0032 (POC: R.D. Riecken).

© The Institution of Engineering and Technology 2018

Submitted: 30 January 2018 E-first: 11 July 2018

doi: 10.1049/el.2018.0295

One or more of the Figures in this Letter are available in colour online.

R.M. Narayanan, A.Z. Liu and P.G. Singerman (Department of Electrical Engineering, The Pennsylvania State University, University Park, PA, USA)

✉ E-mail: ram@engr.psu.edu

M. Rangaswamy (US Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, USA)

References

- 1 Keller, K.L., and Staelin, R.: 'Effects of quality and quantity of information on decision effectiveness', *J. Consum. Res.*, 1987, **14**, (2), pp. 200–213
- 2 Eppler, M.J., and Mengis, J.: 'The concept of information overload: a review of literature from organization science, accounting, marketing, MIS, and related disciplines', *Inf. Soc.*, 2004, **20**, (5), pp. 325–344
- 3 Gospodarek, T.: 'Elasticity of information'. 14th Int. Congress of Cybernetics and Systems of WOSC, Wroclaw, Poland, September 2008, pp. 511–520
- 4 Bougherara, D., Grolleau, G., and Mzoughi, N.: 'Is more information always better? An analysis applied to information-based policies for environmental protection', *Int. J. Sust. Dev.*, 2007, **10**, (3), pp. 197–213
- 5 Farnett, E.C., and Stevens, G.H.: 'Pulse compression radar', in Skolnik, M.I. (Ed.): 'Radar handbook' (McGraw-Hill, New York, 1990, 2nd edn.), pp. 10.1–10.39
- 6 Dmitriev, D., and Jedwab, J.: 'Bounds on the growth rate of the peak sidelobe level of binary sequences', *Adv. Math. Commun.*, 2007, **1**, (4), pp. 461–475
- 7 Narayanan, R.M.: 'Ultrawideband normalized radar cross sections of distributed clutter'. 2012 IEEE Radar Conf., Atlanta, GA, May 2012, pp. 796–800
- 8 Ulaby, F.T., and Dobson, M.C.: 'Handbook of radar scattering statistics for terrain' (Artech House, Norwood, MA, 1989)

interview

Ram Narayanan

“Our work aims to determine the right amount of information required to maximise the radar utility metric for a given application”

Ram Narayanan from Pennsylvania State University in the US, talks to *Electronics Letters* about the paper ‘Information elasticity in radar systems’, page 1049.



utility metric for a given application.

What is the advance that you have reported?

Our paper develops a mathematical model which combines information quality metrics and asso-

Tell us about your field of research

My field of research encompasses radar system development, radar signal processing, radar information exploitation, and radar networking. After graduating from the Indian Institute of Technology, I started working in the area of microwave communications and radar systems. After working for seven years, I came to the United States and completed my Ph.D. from the University of Massachusetts. My area of research was the development of a millimetre-wave radar system that I used to study the scattering properties of vegetation and fallen snow at these wavelengths. I continued to work in the general field of radar as a faculty member, first at the University of Nebraska and now at the Pennsylvania State University. The most interesting and rewarding aspects of my field are that I work on all parts of the system, starting from circuit design, subsystem and integration, data acquisition, and signal processing. Currently, I am most interested in optimising radar system performance through intelligent waveform design and signal processing, in order to fully exploit the potential of radar to provide significant performance improvements through finesse rather than brute force.

Can you describe the background to the work that is presented in your *Electronics Letters* paper?

It is generally assumed that better and more effective decisions can be made in radar sensor processing if more “information” is available. However, this is not always the case since more information may not necessarily translate to better radar detection performance. In ordinary language, the term “information overload” is often used to convey the simple notion of receiving too much information. The concept of information elasticity is useful in determining the relationship and role of information, in both qualitative and quantitative terms, with respect to its interactions with the surroundings. The performance (i.e., the quality of decisions or reasoning in general) of a sensor correlates positively with the amount of information it receives... up to a certain point. If further information is provided beyond this point, the performance of the sensor will rapidly decline. Our work aims to determine the right amount of information required to maximise the radar

utility metric for a given application. This is done using the information elasticity construct, which is the ratio of the relative change in decision effectiveness to the relative change in information quantity. This permits the selection of the optimum quantity of information for the most effective decisions in typical radar applications.

In the era of big data, the conventional viewpoint is that more data or information translates to more effective decisions. However, beyond a certain quantity of information, the performance of the sensor will rapidly decline since it will continue to expend resources processing the information without making a decision. Additional information can no longer be effectively integrated into the decision-making process owing to information overload. The burden of a heavy information load will confuse the sensor, affect its ability to set priorities, and make prior information harder to recall. The challenge is to optimise the trade-offs between sensor metrics and constraints with respect to available information to achieve the best decision with minimal wastage of computational resources.

What do you think the impact of this work will be?

In the short term, we expect that our approach can be exploited to optimise the performance of single sensor systems with several metrics and constraints, as discussed in our paper. Subsequently, the approach can be extended to multiple sensors imposing additional constraints and to include factors such as data quality.

How do you think the field will develop over the next ten years?

The field has changed over the past few years since advances in data acquisition and storage have provided us with massive amounts of data. This has begun to place strains on the signal processor which is unable to cope with the data deluge. Since future sensors have to conform to size, weight, and power (SWaP) requirements dictated by the platforms that carry them, we believe that over the next ten years, researchers will be exploring ways and means by which sensors can perform their intended tasks effectively in the required time without wasting computational resources. We believe that our approach provides such a road map.

DISTRIBUTION A: Distribution approved for public release.

Information elasticity in pseudorandom code pulse compression

Andrew Z. Liu^a, Ram M. Narayanan^a, and Muralidhar Rangaswamy^b

^aThe Pennsylvania State University
University Park, PA 16802, USA

^bAir Force Research Laboratory, AFRL/RYP,
Wright-Patterson AFB, OH 45433, USA

ABSTRACT

Information elasticity is a new concept which characterizes the role of information in making effective decisions in sensor processing. Information elasticity is defined as the ratio of the fractional increase in decision effectiveness to the fractional increase in information. Increasing the quantity of information used in radar processing has the ability to decrease the performance of the radar in certain contexts, depending on what constraints and objectives exist. Because of this phenomenon (known as information overload), it is advantageous to find the optimal amount of information tailored to the specific context the radar is used in. This paper analyzes the process of finding the point of information overload using the information elasticity model. This model is used in observation of different contexts in pseudorandom code pulse compression. In this model, the length of the pseudorandom code represents the amount of information. Increasing this quantity affects both the quality of pulse compression and constraints of the system. We observe this relationship between the constraints and information quantity by developing constraint functions. In this paper, two decision metrics are created for pseudorandom pulse compression, the first based on the peak to side-lobe ratio and the second based on the detection region of the radar.

Keywords: Information elasticity, information overload, pseudorandom-binary sequences, phase code modulation, constraint-based modeling

1. INTRODUCTION

Information elasticity is a new concept proposed to assess the role of information in making effective decisions in sensor processing.¹ In a decision making context, a decision maker is often required to make decisions with the goal of improving performance. Typically, as the amount of information available to the decision maker is increased, the effectiveness of the decisions are also increased. However, in many situations, the amount of information can be increased to a point beyond which the effectiveness begins to decrease.² This point of maximum decision effectiveness is known as the information overload point. This phenomenon is displayed in Figure 1, and is otherwise known as the inverted U-curve.

In order to find this overload point, an information elasticity model is developed. First the decision maker must clearly define the objective of the decisions in question as well as be able to measure it. This parameter is known as the decision metric. Furthermore, the decision maker should be able to define the different parts of the system that constrain the decision effectiveness. The information elasticity model shows the relationship that exists between the decision effectiveness, the decision metric and the system constraints. Information elasticity is a term that describes the usability and qualitative properties of information.³ In this way, the decision metric and system constraints are dependent on the context and what information the decision maker considers to be useful.

This elasticity model is applied in a radar information processing context. In radar, a decision maker has many different parameters to choose that affect the radar's performance. Thus, these different parameters

Further author information: (Send correspondence to Andrew Z. Liu or Ram M. Narayanan)
Andrew Z. Liu: E-mail: azl5361@psu.edu, Telephone: 1 716 969 2105
Ram M. Narayanan: E-mail: rnarayanan@engr.psu.edu, Telephone: 1 814 863 2602

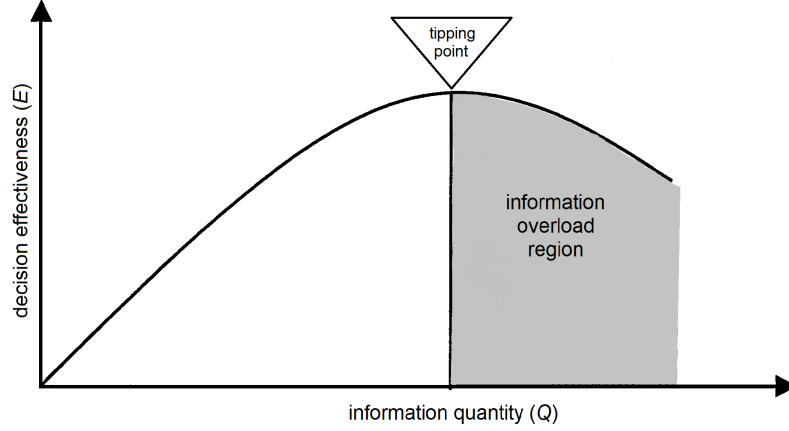


Figure 1. Information overload demonstrated by the inverted U Curve.

represent the x -axis (information quantity) seen in Figure 1, and the radar's performance represents the y -axis (decision effectiveness). Assuming that the decision maker's objective is to improve the performance, the goal of this elasticity model is to determine the parameter values that will maximize performance. This characteristic of performance is entirely defined by the decision maker and is also context dependent.

This paper separately focuses on two different measures of radar performance for phase-coded modulation. These performance measures are maximized when an appropriate modulation code is selected. Thus, we choose the parameter of modulation code length to represent our information quantity. The performance measures are also based on different decision metrics that pertain to the decision maker's context. The first of these is the peak-to-sidelobe ratio (PSLR) and the second is the detection region. Both of these quantities are shown to increase monotonically with increasing modulation code length, producing a concave curve, but not an inverted U-Curve as in Figure 1. In order for an information overload region to exist, the matched filter processing time is introduced as a constraining parameter. This is due to the fact that this processing time increases with code length, and is often considered detrimental to radar performance.

2. MATHEMATICAL BACKGROUND

2.1 Information Elasticity

Information elasticity has been defined as the ratio between the incremental change of decision effectiveness and the incremental change of information quantity:⁴

$$\varepsilon = \frac{dE/E}{dQ/Q} \quad (1)$$

where ε is the information elasticity, E is the decision effectiveness, and Q is the information quantity.

We define E to be a function of the decision metric D and system constraint C . In this paper, we define E to have the following form:

$$E(Q) = \frac{D(Q)}{C(Q)} \quad (2)$$

However, it is important to note that E can take any form, as long as it is a function of D and C . Also, in this paper, all decision metrics and constraint functions increase with increasing Q . Thus:

$$dD/dQ > 0, \quad \text{and} \quad dC/dQ > 0$$

From (2), we write:

$$\ln(E(Q)) = \ln(D(Q)) - \ln(C(Q))$$

Differentiating with respect to Q we obtain:

$$\frac{1}{E} \frac{dE}{dQ} = \frac{1}{D} \frac{dD}{dQ} - \frac{1}{C} \frac{dC}{dQ}$$

Multiplying through by Q , we obtain:

$$\frac{dE/E}{dQ/Q} = \frac{dD/D}{dQ/Q} - \frac{dC/C}{dQ/Q}$$

We define partial elasticities of both D and C as:

$$\varepsilon_D = \frac{dD/D}{dQ/Q} \quad (3)$$

$$\varepsilon_C = \frac{dC/C}{dQ/Q} \quad (4)$$

$$\varepsilon = \varepsilon_D - \varepsilon_C \quad (5)$$

At the overload point, $\varepsilon = 0$. Thus, in order for information overload to exist in a system, there must be a value of Q that exists such that $\varepsilon_D = \varepsilon_C$.

2.2 Maximal Length Sequences

Range side lobes are a common problem in high resolution radars. The existence of these side lobes have the potential to mislead a radar system into picking out a strong target in the side lobes over weaker targets in the main lobe. Thus, a lower PSLR is a desirable trait in many applications and is chosen to be a decision metric in this paper. This decision metric is observed in a phase coded modulation radar context. Furthermore, maximal length sequences (MLS) are selected as the modulation code. This is due to their ease in generation, good autocorrelation properties, and resulting low side lobes.

MLS are a type of pseudorandom binary sequence. These sequences are produced using a linear feedback shift register and only exist in finite lengths of $2^n - 1$, where n is the number of stages in the shift register.⁵ Additionally, the different cyclic permutations of a given MLS produce different autocorrelations, and thus produce different PSLRs when the sequence is matched filtered.⁶ Thus, all of the sequences that are selected are chosen such that their cyclic permutation produces a higher PSLR than any other cyclic permutation.

3. INFORMATION ELASTICITY

3.1 PSLR Decision Metric

In the first analysis on information elasticity, PSLR acts as the decision quality metric and MLS length acts as the information quantity. Figure 2 shows the relationship between the decision metric and information quantity. Note that only the points at which maximal length sequences exist are plotted, and a spline of best fit is added. In this unconstrained system, the decision metric increases with increasing Q but begins to taper off at high values of Q . Furthermore, there is no point at which information overload occurs, meaning that the most effective decision is to choose a sequence of arbitrarily large length.

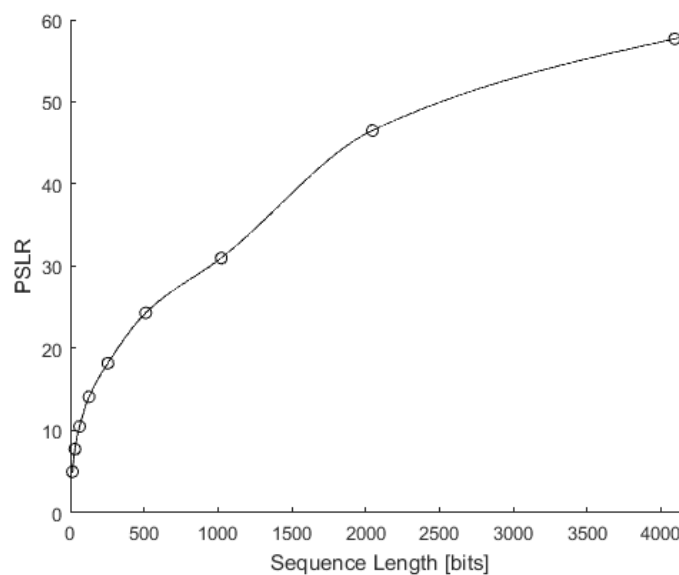


Figure 2. PSLR (decision metric) versus sequence length (information quantity).

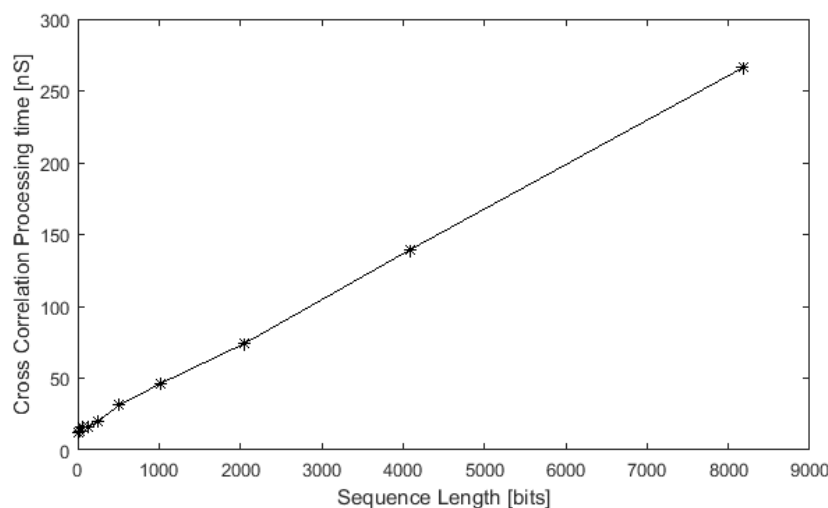


Figure 3. Processing time required to perform cross correlation (constraint) versus sequence length (information quantity).

In this specific observed context, the constraint function is chosen to be the processing time required for MATLAB to perform the cross correlation. The trend for this constraint function is shown in Figure 3. In this specific application, the cross correlation processing time has a linear relationship with the number of elements being cross correlated. Note that this does not necessarily have to be the case. Different processing time trends (i.e. faster, non-linear, etc.) in other applications would also be acceptable, however an overload point may exist in a different location, or may not exist at all.

Equation (2) is applied, where $D(Q)$ is chosen to be the function represented in Figure 2 and $C(Q)$ is chosen to be the function represented in Figure 3. The resulting function (which represents decision effectiveness), is a ratio of PSLR to processing time. This ratio is named the utility factor, since it represents what counts as

usefulness of data in this application. For example, a high PSLR and low processing time would produce a high utility factor and vice versa. Furthermore, its units are represented as s^{-1} , which is meant to describe the PSLR achieved per second of processing time.

The relationship between utility factor and information quantity can be seen in Figure 4. As the units suggest, the utility factor is also used to represent a trade off nature between the decision metric and constraint function. At sequence lengths from 15-511, the utility factor increases with sequence length at a high rate. However, as the trend in PSLR vs. sequence length begins to taper off, the utility factor also begins to taper off (sequence lengths 511-2047). Figure 4 shows that the utility factor begins to decrease at a sequence length of 2047. Beyond this point, information overload occurs. Thus, a sequence length of 2047 provides optimal decision effectiveness for this application.

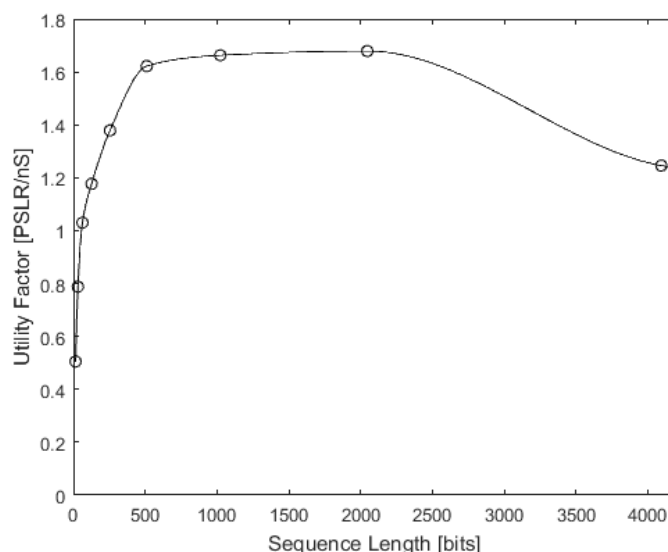


Figure 4. Utility Factor (decision effectiveness) versus sequence length (information quantity).

3.2 Detection Region Decision Metric

A second information elasticity model is developed using a parameter called detection region as a decision metric. For this analysis, a radar simulation observes two simulated targets at set specified locations. The transmitted waveform is pulsed with a PRF of 5 kHz, and is again modulated with a MLS sequence. Pulse integration is used to improve the radar return. The information quantity is still represented by the length of the sequence modulating the waveform, and the constraint function is still represented by the processing time required to cross correlate the chosen MLS sequences. The goal of this analysis is to find the modulating sequence (with the chip length being kept fixed) that produces the best decision effectiveness.

In the radar simulation, the radar return is sent through a matched filter in order to find the range of both simulated targets. For this specific setup, the main lobe of one of the targets is much stronger than that of the second target. The main lobe of the second target is therefore much closer to the noise floor of the matched filter. In this paper, the detection region is defined as the ratio between the peak voltage of the second target's main lobe in the matched filter and the peak voltage of the matched filter's noise floor. This ratio of the voltages is then expressed in decibels. This parameter is used for the decision metric for this experiment. Since the noise floor of the matched filter is different for every radar return, the detection region is averaged from 1000 simulations.

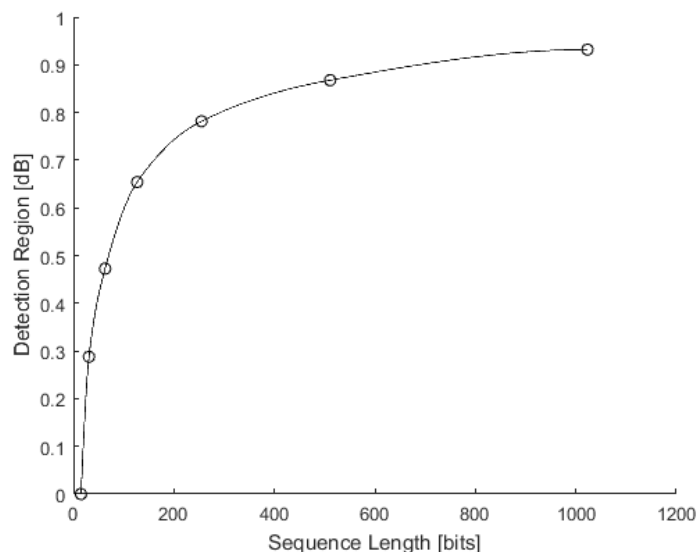


Figure 5. Detection region (decision metric) versus sequence length (information quantity).

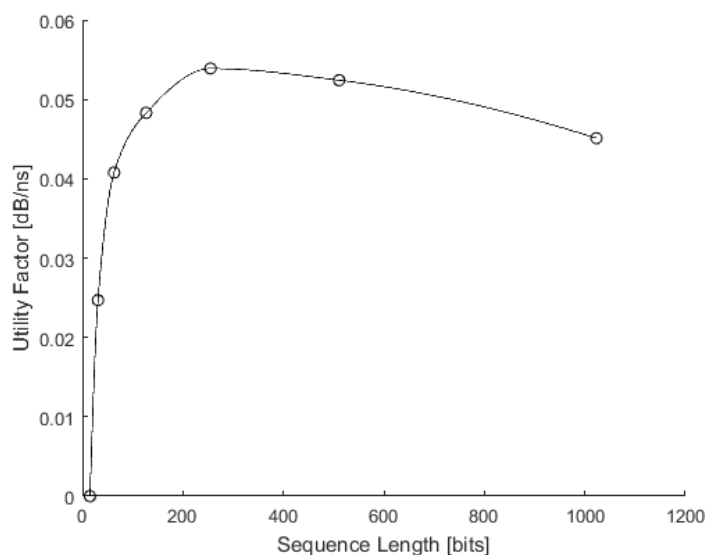


Figure 6. Utility Factor (decision effectiveness) versus sequence length (information quantity).

Figure 5 shows the relationship between the averaged detection region and the sequence length. At higher sequence lengths, the main lobe became much higher than the noise floor. Note that at a sequence length of 15, the detection region is chosen to be 0 dB, meaning that the target became indistinguishable from the noise floor at this sequence length. Once again, the decision metric begins to taper off at higher sequence lengths. Note that although this tapering off characteristic is not required for the selection of a decision metric, it makes the existence of information overload in a system more likely.

Equation (2) is applied, where $D(Q)$ is chosen to be the function represented in Figure 5 and $C(Q)$ is chosen to be the function represented in Figure 3. This produces another type of utility factor, this time with units of

$\frac{dB}{ds}$, representing trade-off between detection region decibels and processing time seconds. As shown in Figure 6, the utility factor increases with sequence length until a length of 255, after which information overload occurs. Thus the MLS with length 255 produces the maximum decision effectiveness in this system.

3.3 Elasticities

When considering the two utility factors discussed, both experiments use the same constraint function (cross correlation processing time). Thus, Figure 7 shows the partial elasticities of both decision metrics and the partial elasticity of their shared constraint function. As shown, the partial elasticity of the detection region metric approaches and crosses the partial elasticity of the constraint first (at its overload point of 255). Similarly, the partial elasticity of the PSLR metric reaches equality with the constraint partial elasticity at its overload point of 2047. Furthermore, Figure 8 shows the full elasticities of both utility factors. Note that the overload point occurs at the zero crossings of each elasticity.

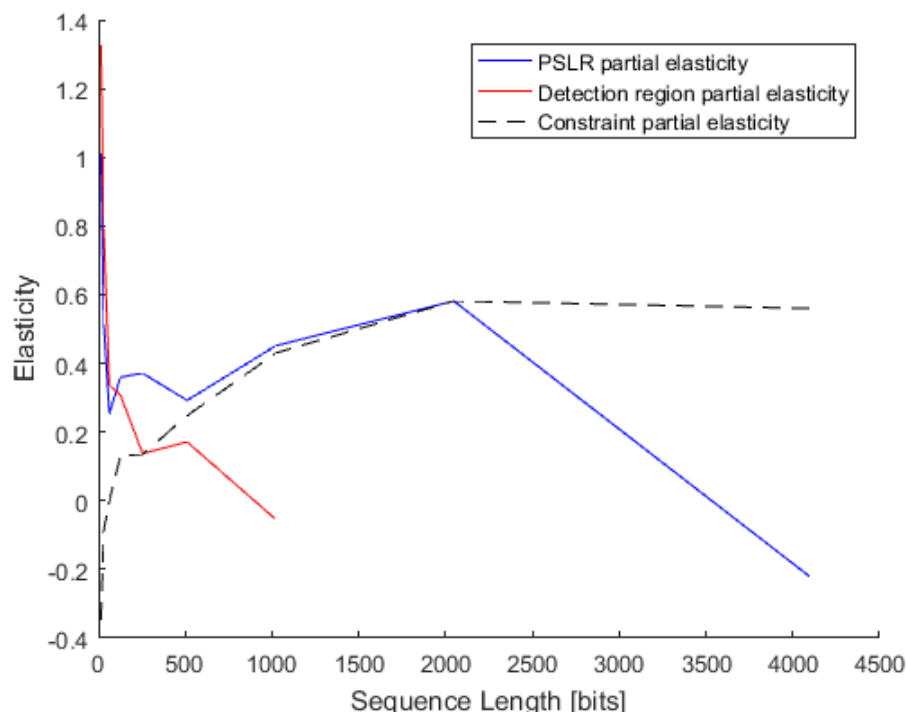


Figure 7. Partial elasticities of both decision metrics plotted with the partial elasticity of the constraint function. The point where the partial elasticity of the decision metric crosses the partial elasticity of the constraint represents the overload point.

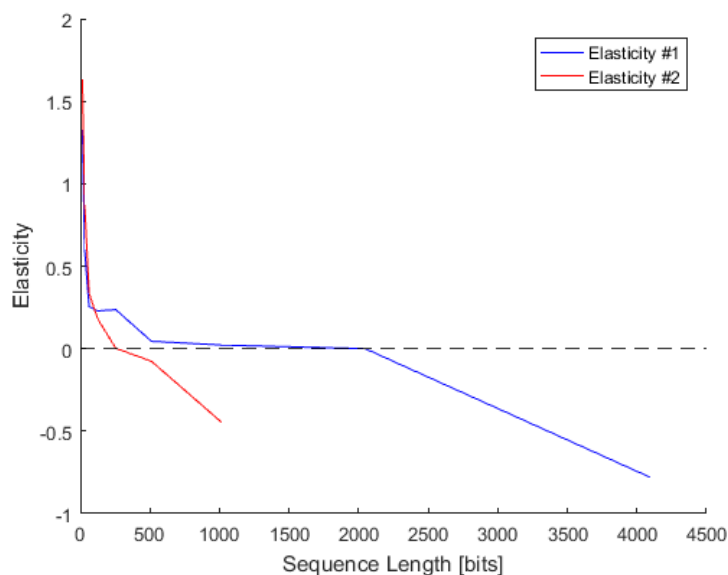


Figure 8. Elasticities of both utility factors. Elasticity #1 represents the PSLR/s utility factor. Elasticity #2 represents the Detection Region/s utility factor.

Note that when using Equations (1), (3) and (4) to calculate elasticities, all differential components were estimated using the differential of a polynomial of best fit. This is due to the fact that the functions of E , C , and D in this context are all discrete functions. The piecewise cubic Hermite interpolation method is used, due to its splines having monotonicity between intervals and a continuous first derivative.⁷ When interpolating the decision effectiveness E , the piecewise spline before the maximum should be monotonically increasing, and the interval after the maximum should be monotonically decreasing. Furthermore, since the first derivative must be continuous, it follows that $\frac{dE}{dQ}$ will always be equal to 0 exactly at the maximum when using this type of interpolation. This produces the desired outcome that the point of information overload occurs at the maximum.

Interpolating the decision metric D to estimate the partial elasticity is slightly different. Let D_i be the interpolation of D and let C_i and E_i be cubic Hermite interpolations of E and C respectively. D_i is thus chosen to be:

$$D_i = E_i \cdot C_i$$

This is due to the fact that the partial elasticities of D_i and C_i are used to create the elasticity of E_i , which has already been carefully defined to be equal to zero at the overload point. Using a cubic Hermite interpolation for both D and C will produce a spline for E that is different than the one previously defined.

4. CONCLUSION

Two different application contexts of phase coded modulation are analyzed. By defining decision effectiveness to be a function of a decision metric and a constraint, a point at which maximum decision effectiveness occurs is found. Owing to the different trade-offs in the system, choosing a different point would be detrimental. Since the measure of decision effectiveness is highly dependent on the context that decisions are made in, the information elasticity model is better suited for systems where the relationship between information and context is well defined. Thus, the development of way to characterize information quality in different contexts is a future goal for the information elasticity model.

ACKNOWLEDGMENTS

This work was supported by the US Air Force Office of Scientific Research Grant Number FA9550-17-1-0032 (POC: R.D. Riecken).

REFERENCES

- [1] Narayanan, R. M., Liu, A. Z., Singerman, P. G., and Rangaswamy, M., “Information elasticity in radar systems,” *submitted for publication* (Jan. 2018).
- [2] Keller, K. L. and Staelin, R., “Effects of quality and quantity of information on decision effectiveness,” *Journal of Consumer Research* **14**, 200–213 (1987).
- [3] Gospodarek, T., “Elasticity of information,” *Proc. 14th International Congress of Cybernetics and Systems of WOSC*, Wrocaw, Poland, 511–520 (Sept. 2008).
- [4] Bougherara, D., Grolleau, G., and Mzoughi, N., “Is more information always better? an analysis applied to information-based policies for environmental protection,” *International Journal of Sustainable Development* **10**, 197–213 (Feb. 2007).
- [5] Golomb, S. W., [*Shift Register Sequences*], Aegean Park Press, Laguna Hills, CA (1981).
- [6] Boehmer, A., “Binary pulse compression codes,” *IEEE Transactions on Information Theory* **13**, 156–167 (April 1967).
- [7] Fritsch, F. N. and Carlson, R. E., “Monotone piecewise cubic interpolation,” *SIAM Journal on Numerical Analysis* **17**, 238–246 (1980).

Information Elasticity in Ultra-Wideband Target Detection Amongst Distributed Clutter

Paul G. Singerman^a, Ram M. Narayanan^{a*}, Muralidhar Rangaswamy^b,

^aThe Pennsylvania State University, University Park, PA 16802, USA

^bAir Force Research Laboratory, AFRL/RYP, Wright-Patterson AFB, OH 45433, USA

ABSTRACT

Information elasticity is a new concept proposed by us to assess the role of information in making effective decisions in sensor processing. Information elasticity is defined as the ratio of the fractional increase in decision effectiveness to the fractional increase in information. One such application is the role of bandwidth when applied to target detection in ground clutter. Continuous, uniformly distributed, land clutter, such as snow, sand, or dirt, has a normalized radar cross section (NRCS) that increases with frequency in the microwave region of the electromagnetic (EM) spectrum. If a target's radar cross section (RCS) increases slower than the clutter that surrounds it as the signal bandwidth increases, this can limit the amount of useful fractional bandwidth that can be used for achieving high range resolution. Using equations for the RCS of a typical sphere target and the overall NRCS of land clutter, it is possible to determine the signal-to-clutter ratio (SCR) at differing fractional bandwidth percentages. If the size of the frequency bins measured is kept constant, a higher fractional bandwidth will have more frequency bins causing an increase in processing time and therefore a decrease in the elasticity. Several different target shapes and types of distributed clutter were simulated, and it is shown that there can be a decrease in elasticity when the fractional bandwidth exceeds a certain value. Furthermore, factoring in the sensor processing time shows that there exists a tipping point beyond which information overload occurs resulting in negative elasticity.

Keywords: information overload, information elasticity, signal-to-clutter ratio, radar cross section, constraint-based modeling

1. INTRODUCTION

Information elasticity is a new concept proposed by us to assess the role of information in making effective decisions in sensor processing.¹ Due to the electromagnetic (EM) spectrum having only a finite amount of usable bandwidth, the preservation of it is critical in many circumstances. Operating a radar at maximum bandwidth has the major drawback of using up a large piece of the EM spectrum which can limit the use of other devices.² If a desired signal-to-clutter ratio (SCR) can be reached with a smaller fractional bandwidth this would allow saving of the EM spectrum for other crucial tasks.

Continuous, uniformly distributed, land clutter, such as snow, sand, or dirt, has a NRCS that increases with frequency in the microwave region of the EM spectrum.³ A metal sphere, on the other hand, has an average RCS that stays reasonably fixed as frequency increases as long as the radius to wavelength ratio is large.⁴ Therefore, a large metal sphere in land clutter will give back a similar average RCS response independent of the frequency, while the land clutter returns a greater RCS as frequency increases; this can limit the useful fractional bandwidth in some cases.

When a target's RCS increases with increasing frequency at a rate less than the land clutter that surrounds it, a decrease in the elasticity (rate of increase) of the SCR can result; eventually at higher frequencies, an overload (decrease) in the SCR can, in some cases, occur. By combining processing time along with the SCR, a better model for the total utility of the data can be obtained. The processing time will increase with extra data at a rate determined by the complexity of the target detection algorithm. The processing time should be inverse to

Further author information: (Send correspondence to R.M.N.)

R.M.N.: E-mail: ram@engr.psu.edu, Telephone: 1 814 863 2602, Fax: 1 814 865 7065

the total utility of the data while the SCR should be directly proportional. This emulates a real sensor system as extra SCR is beneficial while extra processing time is disadvantageous.

2. INFORMATION ELASTICITY IN TARGET DETECTION

2.1 Information Elasticity Formulation

The elasticity of any differentiable function is defined as the ratio of incremental change of the function with respect to the incremental change of its argument. The information elasticity, ϵ , is defined as⁵

$$\epsilon = \frac{dE/E}{dQ/Q} \quad (1)$$

where E is the decision effectiveness and Q is the information quantity. The decision effectiveness is a function of a number of decision quality metrics $D_k(Q)$ for $k = 1, \dots, K$, along with various contextual constraint functions $C_l(Q)$ for $l = 1, \dots, L$. Here, both the decision quality and constraints depend on Q , with $D_k(Q)$ being always increasing, although often leveling off, and $C_l(Q)$ always decreasing. Thus, we have $dD_k/dQ \geq 0$ and $dC_l/dQ \leq 0$. This leads to the variation of $E(Q)$ with Q following the dependence shown in Figure 1.

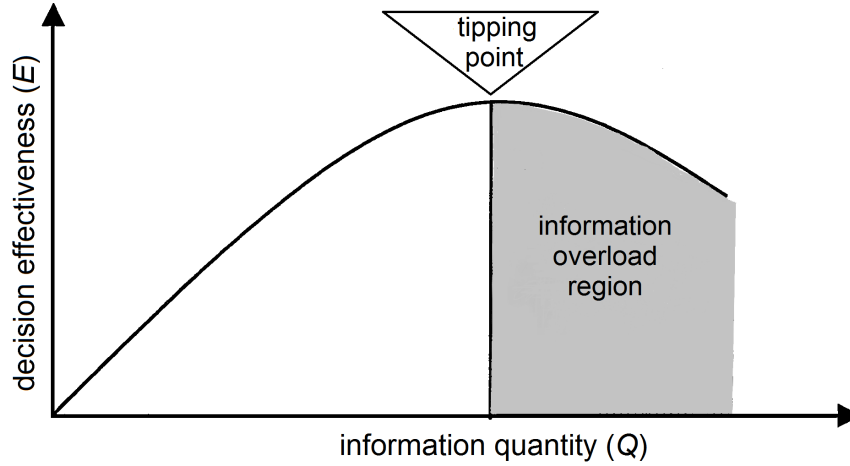


Figure 1: Typical inverted U-curve depicting decision effectiveness, $E(Q)$, as a function of information load or quantity, Q .

2.2 Classification of Clutter

As stated earlier, it has been found that continuous, uniformly distributed, land clutter, such as snow, sand, or dirt, has an NRCS that generally increases with frequency in the microwave region of the EM spectrum.³ It would be valuable to be able to model different types of clutter and even more valuable to be able to classify this clutter based on measurements taken from a radar in real time during some type of target detection mission. Frequency response of the NRCS of continuous, uniformly distributed, land clutter is typically given by⁶

$$\sigma_0 = \sigma_0^0 \left(\frac{f}{f_0} \right)^n \quad (2)$$

By taking measurements of NRCS at two frequencies f_1 and f_2 , we can find n which will give us a function representing the specific clutter we are measuring.

$$n = \frac{\log \frac{\sigma_0(f_1)}{\sigma_0(f_2)}}{\log \frac{f_1}{f_2}} \quad (3)$$

2.3 Simulation of Clutter RCS Response

We can show that it is possible to classify clutter just by have two variables used to represent it. It is necessary to first create an appropriate waveform that would be similar to what we would expect to get from continuous, uniform clutter. A good choice is a linear frequency modulated (LFM) waveform with an appropriate power spectral density (PSD). Therefore, we simulated a LFM waveform with 1 GHz bandwidth centered at 1 GHz frequency (100% fractional bandwidth) as parameters. This was followed by adding white Gaussian noise at a signal-to-noise ratio (SNR) of 20 dB. The waveform's amplitude increased linearly with frequency to generally mimic real continuous, uniformly distributed, land clutter. Figure 2 shows a plot of one pulse of the LFM waveform. We then found the approximate power at a frequencies of 1.00 and 1.25 GHz by taking the average

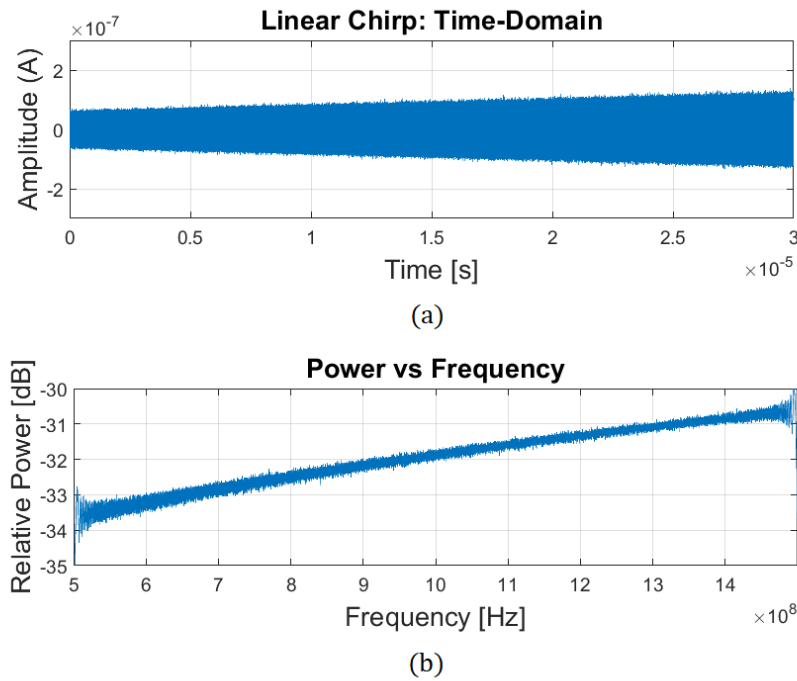


Figure 2: (a) One pulse of the simulated linear chirp with center frequency 1 GHz and bandwidth 1 GHz (100% fractional bandwidth) and increasing power with frequency in the time domain. (b) The same LFM waveform in the frequency domain.

of about 300 data points surrounding them (this was in order to minimize the effect of the noise). Here 1 GHz was the center frequency and 1.25 GHz was at the 75th percentile of the total bandwidth. Assuming that power and RCS are interchangeable for finding the SCR, we can use equation (3) and solve for n using these measured values.

We found that in this case $\sigma_0(f_1) = -31.89$, $\sigma_0(f_2) = -31.22$, $f_1 = 1.00$ GHz, and $f_2 = 1.25$ GHz. This gives us an n value of 0.6985. With other studies showing n being between 0 and 1 for land clutter this is support for the waveform being a suitable recreation of the desired clutter type.⁷ Now, using equation (2) again and substituting the measured values for $\sigma_0(f_1)$ converted from dB, f_1 , and n and converting to dB scale, we get

$$\sigma_0 = 10 \log_{10} \left(10^{-31.89/10} \left(\frac{f}{10^9} \right)^{0.6985} \right) \quad (4)$$

Figure 3 shows this function plotted over different frequency values from 0.5 GHz to 1.5 GHz. Now we can

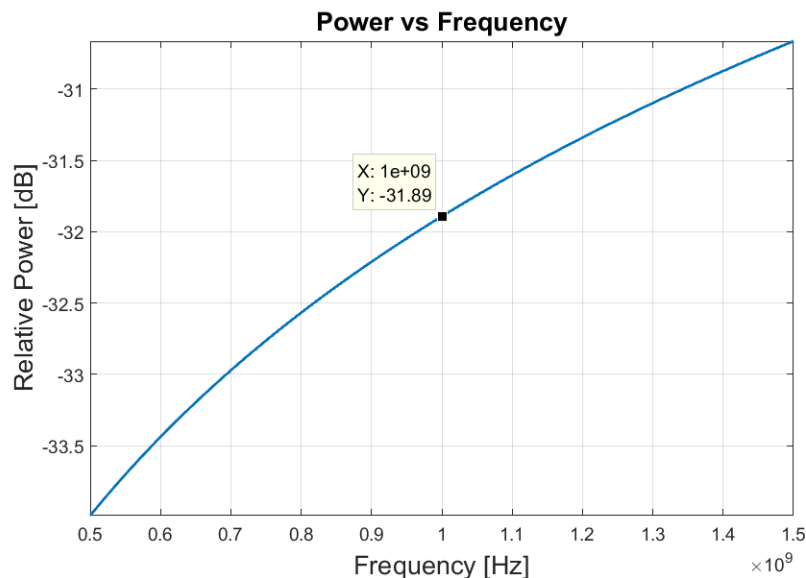


Figure 3: The predicted power vs frequency curve for the n value of 0.6985 found from a simulated LFM waveform from 0.5 GHz to 1.5 GHz frequency with increasing amplitude.

plot both the simulated waveform with a SNR of 20 dB and our calculated RCS response on the same graph to compare and confirm if we achieved a good approximation using the measured n value of 0.6985. This is shown in Figure 4 below.

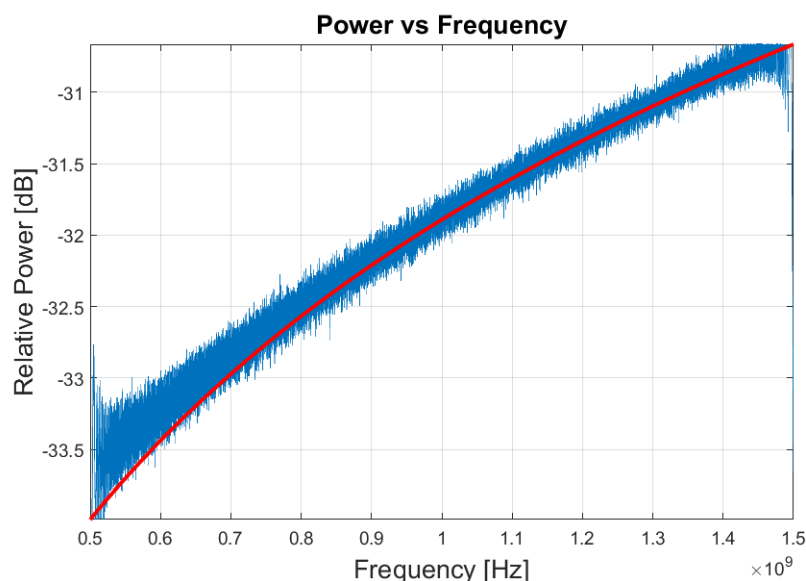


Figure 4: The predicted power vs frequency curve generated from equation (2) for the n value of 0.6985 found from a simulated LFM waveform (red) and the simulated waveform used, with an SNR of 20 dB (blue).

Figure 4 confirms that by calculating the scaling factor n via different RCS and measurements at the center frequency and the 75th percentile frequency, we were able to get a very good approximation of the simulated clutter's RCS response at different frequencies. This also means we can classify different continuous, uniformly

distributed, land clutter by just measuring it and solving for its scaling constant. We do not need to use subjective language, such as describing something as “snow”. Snow has many different types (wet, dry, dense, etc..) with different clutter properties and is therefor not the most effective descriptor for radar. We can instead describe the clutter mathematically to a much more precise degree, at least in the area of target detection. Figure 5 shows the predicted trend as frequency gets beyond the range we used to classify the clutter. This is what we

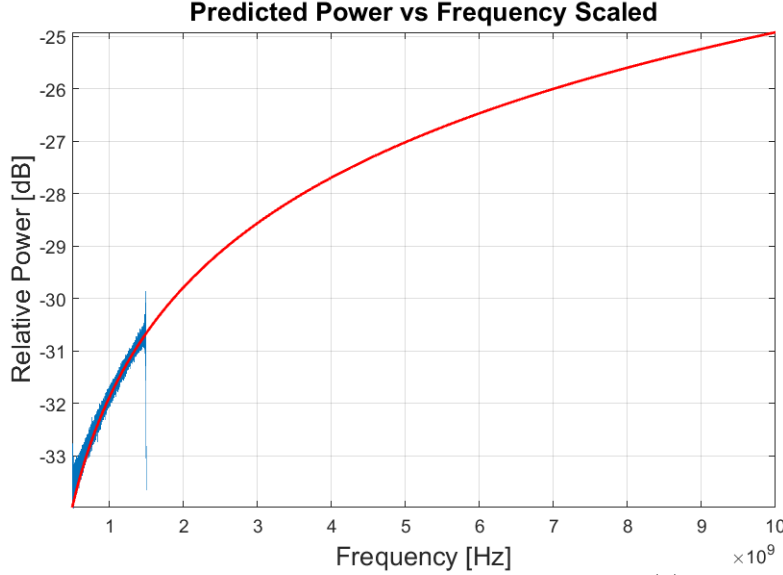


Figure 5: The predicted power vs frequency curve generated from equation (2) for the n value of 0.6985 found from a simulated LFM waveform scaled to 10 GHz (red) and the simulated waveform used, with an SNR of 20 dB (blue).

employ to predict the relative RCS power as the frequency approaches 10 GHz. At 10 GHz, we would expect the relative power to be about -25 dB for continuous, uniformly distributed, land clutter with an n value of 0.6985.

2.4 Fractional Bandwidth Optimization for Target Detection in Uniform Land Clutter

It is possible to use this power vs. frequency curve that was derived and supported in the last section in order to find specific fractional bandwidths for achieving optimal signal-to-clutter ratios (SCRs) for target detection in continuous, uniformly distributed, land clutter. Consider a metal sphere as our target of interest with a radius of 3 meters. In the optical region ($r \gg \lambda$), the sphere's σ_0 (RCS) is given as⁶

$$\sigma = \pi r^2 \quad (5)$$

where r is the radius of the sphere. Now, it is necessary to find the ultrawideband (UWB) RCS of the sphere for a frequency independent antenna case, which is given by⁸

$$\sigma_{\text{UWB}(f)} = \sigma \left(\frac{4}{4 - \beta^2} \right) \quad (6)$$

where β is the fractional bandwidth. Therefore, the UWB RCS of a sphere can be expressed as

$$\sigma_{\text{UWB}(f)}^{(t)} = \pi r^2 \left(\frac{4}{4 - \beta^2} \right) \quad (7)$$

Here $\sigma_{\text{UWB}(f)}^{(t)}$ is the UWB RCS value of the sphere, which represents the target. The SCR is the ratio of the target's RCS to the RCS of the clutter patch that is illuminated. The equation of uniform clutter's UWB RCS is represented as⁶

$$\sigma_{\text{UWB}(f)}^{(c)} = \frac{A\sigma_0^0\beta^{n-2}}{n-1} \quad (8)$$

Here $\sigma_{\text{UWB}(f)}^{(c)}$ is the UWB RCS value of the clutter and A is the area of the illuminated clutter patch. The equation for the SCR of a sphere then becomes

$$\text{SCR}_{\text{sphere}} = \frac{\sigma_{\text{UWB}(f)}^{(t)}}{\sigma_{\text{UWB}(f)}^{(c)}} = \frac{\pi r^2 \left(\frac{4}{4-\beta^2} \right)}{\frac{A\sigma_0^0\beta^{n-2}}{n-1}} \quad (9)$$

which can then be reduced to

$$\text{SCR}_{\text{sphere}} = \frac{4\pi r^2(n-1)}{A\sigma_0^0(4\beta^{n-2} - \beta^n)} \quad (10)$$

where n is the scaling constant calculated using known values for f_1 and f_2 and measured values for $\sigma_0(f_1)$ and $\sigma_0(f_2)$, utilizing equation (3).

Now using the simulated waveform we made in the last section and the corresponding calculated n value along with the measured value of σ_0^0 , we obtain the SCR vs. fractional bandwidth percentage (FBP) plot shown in Figure 6.

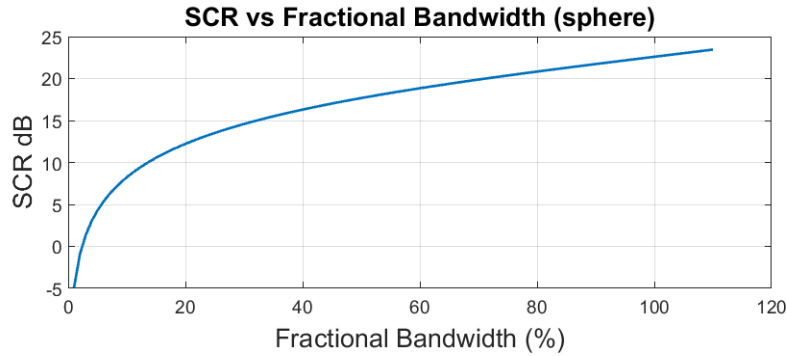


Figure 6: Predicted SCR for different fractional bandwidth percents for a 3 radius meter metal sphere in clutter with n value 0.6985 and center frequency at 1 GHz.

This approach was then repeated for the apex and the corner reflector targets. We also included a 3 GHz center frequency signal with 3 GHz bandwidth shown in Figure 7.

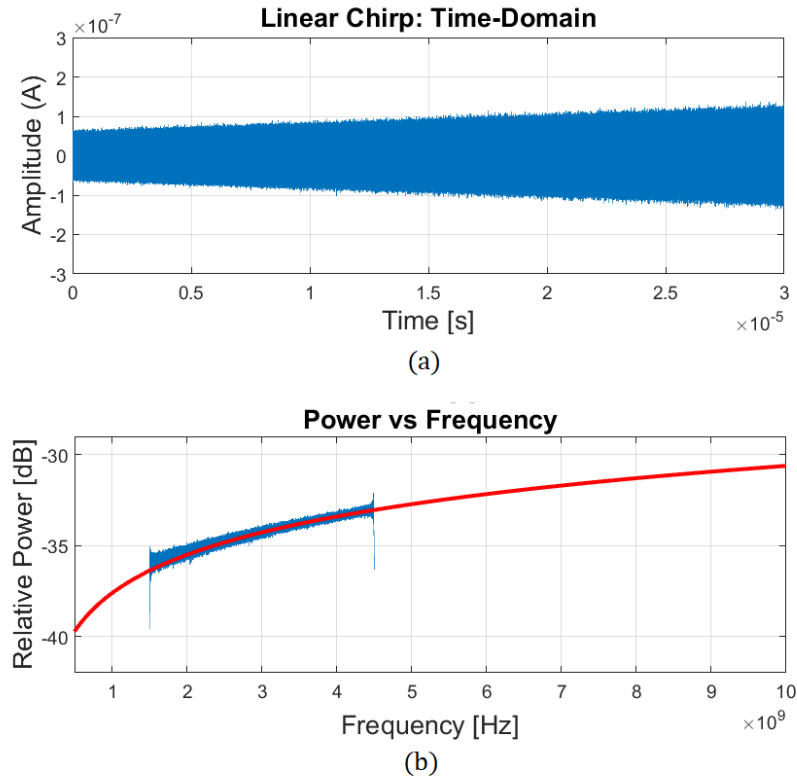


Figure 7: (a) One pulse of the simulated linear chirp with center frequency 3 GHz and bandwidth 3 GHz and increasing power with frequency in the time domain. (b) The same LFM waveform in the frequency domain with the measured value of the center frequency RCS and the calculated n value.

We can then use the UWB RCS equations for an apex and a corner reflector (CR)^{6,8}

$$\text{Apex : } \sigma = \frac{\lambda^2}{16\pi}; \quad \sigma_{\text{UWB}(f)}^{(t)} = \sigma \left(\frac{64}{(4 - \beta^2)^3} \left(1 + \frac{\beta^2}{12} \right) \right) \quad (11)$$

$$\text{Corner Reflector : } \sigma = \frac{\pi l^4}{3\lambda^2}; \quad \sigma_{\text{UWB}(f)}^{(t)} = \sigma \quad (12)$$

Then, by performing the same steps we find the SCRs for 1, 3, and 10 GHz center frequencies for all three target shapes. This is shown in Figure 8 below.

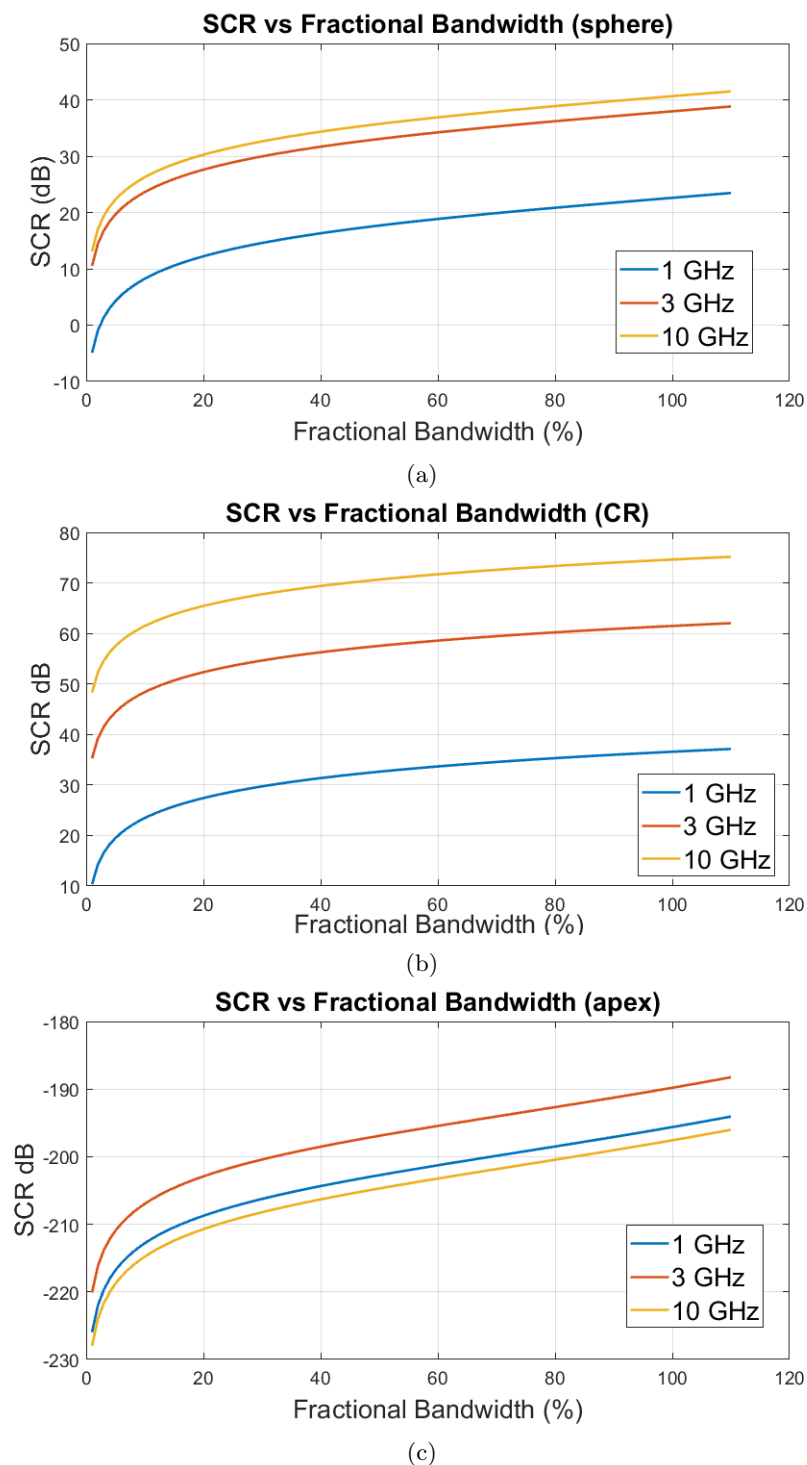


Figure 8: SCR vs. FBP at 1 and 3 GHz center frequencies in simulated clutter for (a) sphere ($r = 3$ m), (b) corner reflector ($l = 3$ m), and (c) apex ($l = 3$ m).

Using this information, it would be simple to find what fractional bandwidth percentage would be needed at every center frequency in order to achieve a desired SCR when looking for a target in any distributed uniform clutter, while still leaving as much EM spectrum available for other uses.

2.5 Scaling Constant of Wet Snow

Although it is generally assumed that $n = 0.5$ for land clutter, it is critical to verify if the n value used was correct, specifically for wet snow clutter. This way it was possible to calculate n by knowing four values: the center frequency (f_0), some other frequency (f) (preferably relatively distant from the center frequency), the RCS response at the starting frequency (σ_0^0), and the RCS response at the other frequency chosen ($\sigma_0(f)$). All relevant measured values were obtained from Ref. [3] and are shown in Table 1. By using equation (3) and any arbitrary frequency value as the center frequency and any arbitrary frequency as the final frequency, the n values could be calculated. The goal was to find a more or less constant RCS curve no matter what data points were used from these measurements. Using equation (2), we can plot the predicted RCS vs. frequency curves by

Table 1: Measured values of f_0 , f , σ_0^0 , and $\sigma_0(f)$ from³ and the corresponding calculated n value.

f_0 (GHz)	f (GHz)	$\sigma_0^0(f)(10^{-3})$	$\sigma_0(f)(10^{-3})$	n
1	3	2.5	5	1.585
1	10	2.5	20	1.107
1	15	2.5	39.8	0.9785
3	10	5	20	0.8685
3	15	5	39.8	0.7758
10	15	20	39.8	0.589

using the values for σ_0^0 and f_0 . This is shown in Figure 9.

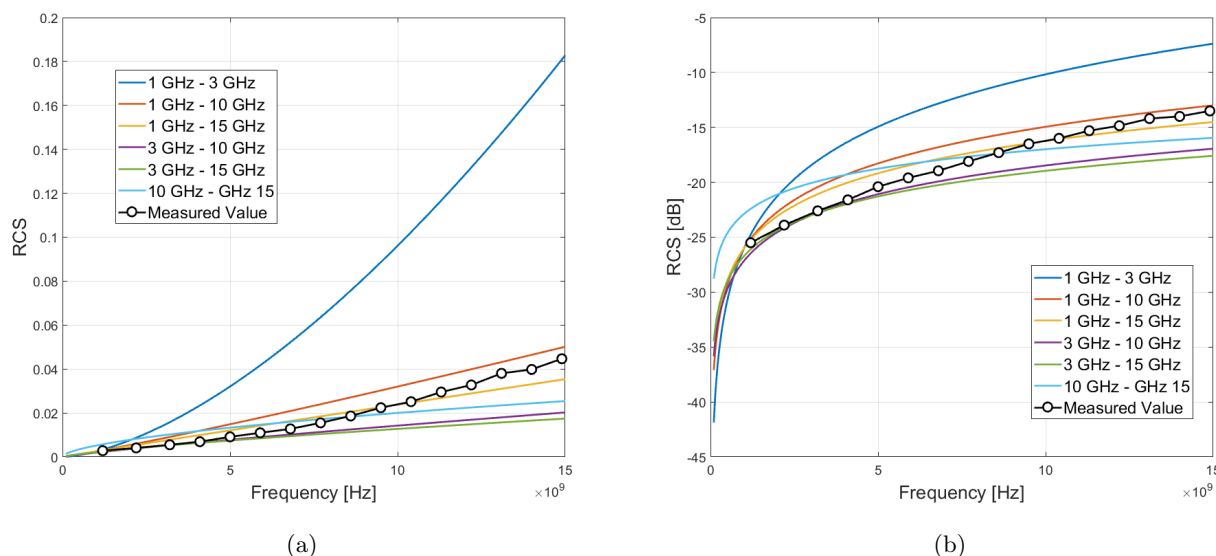


Figure 9: (a) Predicted RCS vs. frequency for different frequencies used to predict n and the actual measured frequency. (b) Same plot in dB scale.

All plots give a very similar shape with the 1-3 GHz measurement being a bit of an anomaly compared to the other five measurements. With the n value for the 1-3 GHz measurements being 1.585 and the others being between 1.1 and 0.589, the variation in the other high frequency measurements between the last 5 are similar; thus, $n = 0.6$ was used for the RCS curve with increasing frequency for wet snow. Now that an approximation of the n value was found for wet snow, it is now possible to use equation (10), derived earlier, to get the SCR vs. fractional bandwidth curves for different center frequencies. Relevant plots for different center frequencies of 1, 3, 10, and 15 GHz are shown in Figure 10.

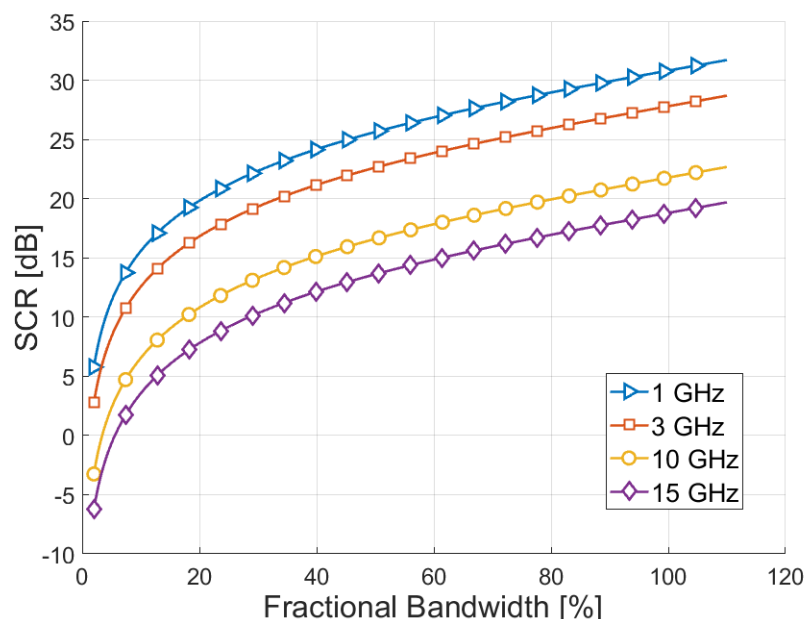


Figure 10: Signal to clutter ratios compared to fractional bandwidths for different center frequencies for a 3 meter radius metal sphere in wet snow clutter.

Figure 9 was generated with n values of 0.6 which represented wet snow. This led to the question of what shape curves would be generated with different types of land clutter that had different n values ranging from 0 to 0.9 (e.g. sand, dry snow, etc.). Thus, using the same sphere and setting the center frequency to 15 GHz and incrementing n from 0 to 0.9 by 0.1 increments, a plot for several values of n was generated. The resulting plot is shown in Figure 11.

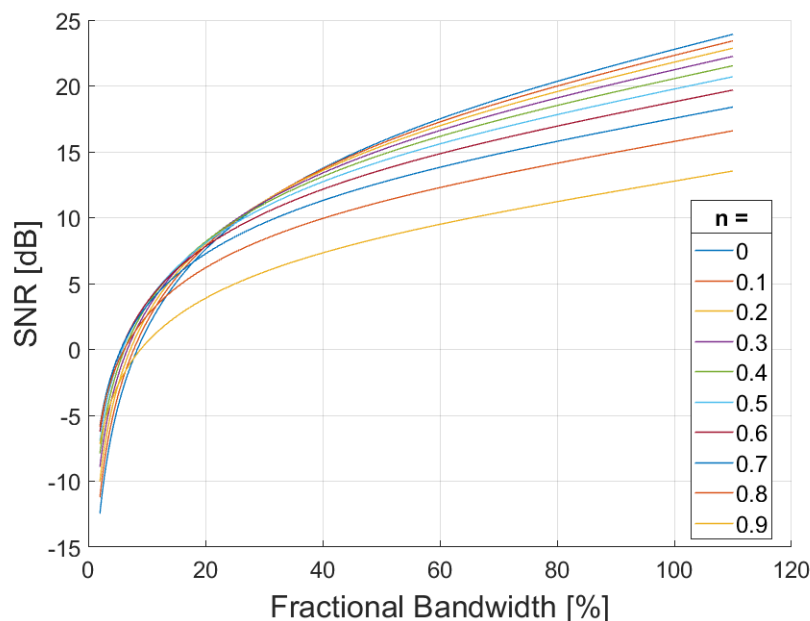


Figure 11: Signal to clutter ratios compared to fractional bandwidths at 15 GHz center frequency for a metal sphere in different types of uniform land clutter represented by different n values.

The plot shows that as n increases from 0 (the blue, highest line) to 0.9 (the yellow lowest line) the SNR has a similarly shaped curve with sharp increase in SNR from 0 to 35% fractional bandwidth and a more moderate increase as FBP goes higher.

2.6 Utility of Fractional Bandwidth

Here, we define information elasticity as the overall gain in the metrics deemed to be important that a system obtains from extra information. We also define information overload at the point in which extra information stops giving the expected benefit and causes a loss in the metrics deemed important.⁹ We use processing time and SCR as the two metrics deemed important. Processing can be related to different fractional bandwidths by including the time spent processing, since using higher FBP would cause an increase in the amount of data to be processed. Processing time is crucial for information elasticity as more processing time is a trade-off that should be taken into account when trying to optimize a waveform to have a certain fractional bandwidth.

Up until now we have been analyzing the RCS of targets and clutter without processing time considerations. However, in a real operational scenario, processing time is critical for making timely decisions. This processing will undoubtedly be using frequency bins, and if the frequency bins are to stay constant, then additional frequency bins incur extra processing. By finding how the processing time increases with extra data input, namely the big O of the algorithm doing the processing, then it is possible to relate the extra fractional bandwidth to extra total frequency which is then divided up into frequency bins. The sensor utility factor is defined as:

$$\text{Utility Factor} = \frac{\text{SCR}}{\text{Processing Time}} \quad (13)$$

It is important to stress how important different algorithm complexities and differing data amounts change the final processing time. Radar data can get extensive in size.¹⁰ Therefore, this should also drastically change how much utility the data is providing. By plotting different utility factors vs. fractional bandwidths, we can visualize how utility trends with an increase in FBP. We choose to use an $O(n)$ complexity algorithm with 100 data sets per frequency bin, for several different center frequencies. By using equation (13) and assuming a metal sphere in wet snow clutter, we plotted the Utility Factor vs. FBP, which is shown in Figure 12.

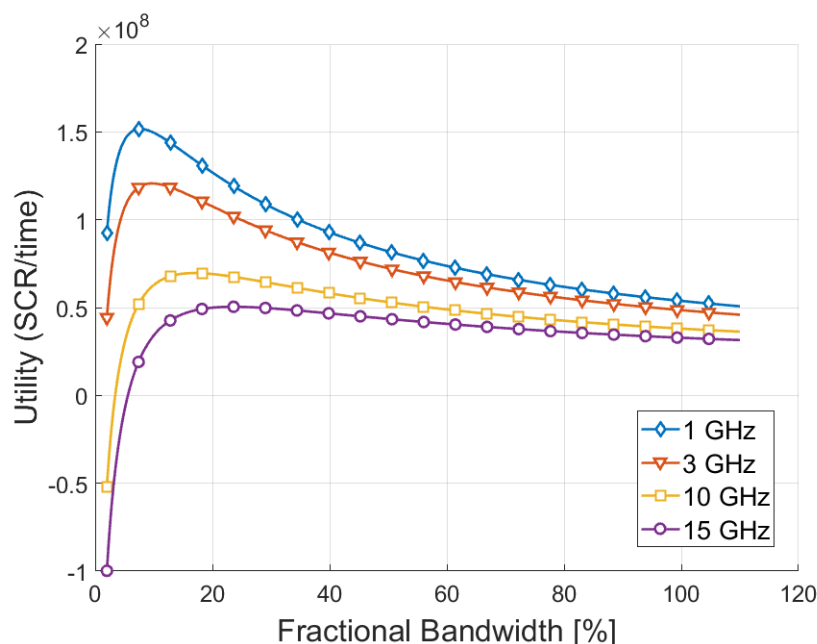


Figure 12: Utility compared to fractional bandwidths for different center frequencies for a metal sphere in wet snow clutter with $O(n)$ complexity and 100 data sets per frequency bin.

These curves have clear maximums varying around 10-25% fractional bandwidth. This is the tipping or the overload point. Using more fractional bandwidth in this case is not beneficial and is wasting resources such as the EM spectrum and processing time, both of which can be used for other tasks. It is important to note that here we are using arbitrary weights for how valuable the EM spectrum and extra processing time is. It is possible to add different weights according to the importance given to these metrics. This will change where the overload point will occur.

3. CONCLUSION

In conclusion, we simulated waveforms that mimicked continuous, uniform, land clutter. We then used this to show that we could generate a frequency dependence exponent (n), which would be used to predict the clutter's RCS response at higher or lower frequencies. We then showed that the n value prediction is close to previously measured results. Using this frequency response prediction, we were able to find the relationship between the signal-to-clutter ratio and the operating fractional bandwidth. We simulated three different simple geometric shapes for these results. By using processing time as a metric, we were able to characterize the effect of fractional bandwidth on the overall utility factor.

ACKNOWLEDGMENTS

This work was supported by the US Air Force Office of Scientific Research Grant Number FA9550-17-1-0032 (POC: R.D. Riecken).

REFERENCES

1. R. M. Narayanan, A. Z. Liu, P. G. Singerman, and M. Rangaswamy, "Information elasticity in radar systems," *submitted for publication*, Jan. 2018.
2. H. T. Hayvaci and B. Tavli, "Spectrum sharing in radar and wireless communication systems: A review," in *2014 International Conference on Electromagnetics in Advanced Applications (ICEAA)*, pp. 810–813, (Palm Beach, Aruba), Aug. 2014.
3. F. T. Ulaby and M. G. Dobson, *Handbook of Radar Scattering Statistics for Terrain*, Artech House, Norwood, MA, 1989.
4. R. Wang, Y. Qin, B. Deng, and H. Wang, "Radar cross section of the metal sphere from microwave to the optical frequency," in *2014 39th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, Sep 2014.
5. D. Bougherara, G. Grolleau, and N. Mzoughi, "Is more information always better? an analysis applied to information-based policies for environmental protection," *International Journal of Sustainable Development* **10**(3), pp. 197–213, 2007.
6. R. M. Narayanan, "Ultrawideband normalized radar cross sections of distributed clutter," in *IEEE Radar Conference*, pp. 796–800, (Atlanta, GA), May 2012.
7. J. Reilly, R. McDonald, G. Dockery, and J. Stapleton, "RF-environment models for the ADSAM program," Final Report A1A97U-070, JHU/APL, 1997.
8. R. M. Narayanan, "Ultrawideband radar cross sections of radar calibration targets," in *Proc. NATO Specialist Meeting on Ultra-Wideband Radar Systems*, pp. 8–1–8–8, (Toulouse, France), Oct 2008.
9. J. Boyd and D. Sworder, "Modeling information overload," in *Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)*, pp. 1–9, (Orlando, FL), Nov.-Dec. 1999.
10. A. Jayapraksh and G. Reddy, "Small target detection within sea clutter based on fractal analysis," *Procedia Technology* **24**, pp. 988–995, 2016.

Robust decision making method for adaptive ordered-statistics CFAR technique using information elasticity

Andrew Z. Liu^a, Ram M. Narayanan^a, and Muralidhar Rangaswamy^b

^aThe Pennsylvania State University
University Park, PA 16802, USA

^bAir Force Research Laboratory, AFRL/RYP,
Wright-Patterson AFB, OH 45433, USA

ABSTRACT

Some performance characteristics of ordered-statistics CFAR (OS-CFAR), such as probability of false alarm (P_{FA}) and probability of detection (P_D), are controlled by many parameters. Some of these parameters can be considered decision parameters, since they are user defined to achieve a fixed P_{FA} , and an optimized P_D . However, other parameters that control these probabilities have the tendency to fluctuate based on the radar environment and operating conditions. These environmental variables can sometimes be difficult to predict and may affect performance. In this paper, a robust decision making method is presented, which selects decision parameters that provide robust performance even in the presence of these variations. The relevant environmental variables investigated in this paper are the number of interfering targets within the detection window and the signal-to-noise ratio (SNR). The Forward Automatic Order Selection Ordered Statistics Detector (FAOSOSD) is used to provide an estimate for the number of interfering targets, and the accuracy of this estimate is observed as a function of SNR. The proposed method defines a performance metric and observes its mean and variance over the uncertain parameter SNR. A trade-off behavior is shown between this mean and variance, and using information elasticity analysis, a decision is selected.

Keywords: OS-CFAR, uncertainty immunity, robust optimization, information overload

1. INTRODUCTION

Constant false alarm rate (CFAR) is a technique used to dynamically set a threshold of detection using an estimate of the local background interference. This interference estimate is often based on the voltages of range cells surrounding a cell under test (CUT). It is assumed that some or most of these cells contain background interference voltages. However, in many cases, the cells surrounding the CUT contain interfering targets which can affect the interference estimate and cause a loss in detection performance.

Ordered statistics CFAR (OS-CFAR) has been proposed as a solution to this detection loss.¹ This technique operates by using the k^{th} largest cell (from a window of N total cells) as an interference estimate. A multiplier α is applied to this term to obtain the detection threshold, where α is selected to achieve a constant probability of false alarm (P_{FA}) set by the user. The parameter k may also be user selected, as certain values of k offer higher probability of detection (P_D).² Thus, both k and α can be thought of as decision parameters that alter important performance metrics P_{FA} and P_D .

It has been shown that P_{FA} and P_D are also dependent on the number of interfering targets present.^{1,3} Thus, in order to properly select k and α to provide desired operating conditions, it is important to know the number of interfering targets. In ref. 2, an adaptive version of OS-CFAR is proposed, which selects the value of k based on some inference of the background situation. One such algorithm that accomplishes this is the

Further author information: (Send correspondence to Andrew Z. Liu or Ram M. Narayanan)
Andrew Z. Liu: E-mail: azl5361@psu.edu, Telephone: 1 716 969 2105
Ram M. Narayanan: E-mail: rnarayanan@engr.psu.edu, Telephone: 1 814 863 2602

Forward Automatic Order Selection Ordered Statistics Detector (FAOSOSD).⁴ This technique uses information theoretic criteria to estimate the number of interfering targets present.

Ideally, this estimate would be used to select the decision (α, k) that fixes P_{FA} and maximizes P_D . However, the accuracy of this estimate is dependent on the SNR between the interfering targets and the background interference cells, which is unknown. As a result, a given decision may be suitable at one value of SNR, but unsuitable at another. Thus, the objective of this paper is to employ a decision making method which selects an (α, k) that is robust to uncertainties in the SNR. A robust decision should offer high performance as well as low sensitivity to environmental variables (i.e. SNR). These characteristics can be achieved by maximizing the mean performance while minimizing the variance with respect to the environmental variables.⁵

A function ψ is created to represent the performance. Since the goal is to maximize P_D and fix P_{FA} , ψ is dependent on both P_D and P_{FA} . The performance ψ is calculated at a collection of different 2-tuple decision points $\tilde{\mathbf{x}} = (\alpha, k)$ over a range of SNR values. From this, the mean and variance of ψ over SNR is calculated for each decision. If the mean of ψ represents average performance and the variance of ψ represents robustness, a direct trade-off nature can be shown in the Pareto frontier of these traits.

This paper employs the technique of information elasticity analysis, which creates a measure of decision effectiveness based on the robustness and average performance. In many scenarios, it can be shown using this technique that as the robustness increases, the decision effectiveness will increase up to a point, beyond which the decision effectiveness will decrease. This phenomenon is known as information overload⁶ and the point of maximum decision effectiveness is known as the overload point. The decision making method in this paper seeks to select the decision that occurs at the overload point.

2. PERFORMANCE OF OS-CFAR

2.1 Formulation of P_D and P_{FA}

Ref. 3 derives a formula for P_D and P_{FA} as a function of the number of interfering targets present. However, this formula uses an approximation that does not hold for lower values of SNR. Since this paper aims to select robust decisions, even at low values of SNR, it is important to know how P_D and P_{FA} behave in this region. In this section, an equation for P_D and P_{FA} that holds at all values of SNR is derived.

Let y be the power within the CUT and z be the power within the background cells. Consider that hypothesis 0 describes when a cell contains only interference, and hypothesis 1 describes when a cell contains a target and interference. Thus, $f_0(\cdot)$ and $F_0(\cdot)$ respectively represent the PDF and CDF of a cell under hypothesis 0, and $f_1(\cdot)$ and $F_1(\cdot)$ respectively represent the PDF and CDF of a cell under hypothesis 1. Furthermore, let the PDF of the k^{th} highest cell be $g_k(z)$. Therefore, P_{FA} and P_D are given by:²

$$\begin{aligned} P_{FA} &= E_Z\{P[Y > \alpha Z | H_0]\} \\ &= \int_{-\infty}^{\infty} \left\{ g_k(z) \int_{\alpha z}^{\infty} f_0(y) dy \right\} dz \end{aligned} \quad (1)$$

$$P_D = \int_{-\infty}^{\infty} \left\{ g_k(z) \int_{\alpha z}^{\infty} f_1(y) dy \right\} dz \quad (2)$$

Using ordered statistics, an expression can be found for $g_k(z)$. Ref. 7 describes that the k^{th} statistic of N independent but not necessarily identical random variables is given by:

$$g_k(z) = \frac{1}{(k-1)!(N-k)!} \sum_p F_{i_1}(z) \dots F_{i_{k-1}}(z) \cdot f_{i_k}(z) \cdot \{1 - F_{i_{k+1}}(z)\} \dots \{1 - F_{i_N}(z)\}$$

where $F_{i_1}, \dots, F_{i_N}$ and $f_{i_1}, \dots, f_{i_N}$ represent the distributions of the N random variables, and $\sum_p$ represents the summation over all $N!$ permutations $(i_1, i_2, \dots, i_N)$ of $(1, 2, \dots, N)$. This large number of summed terms can be significantly reduced if we consider that all interference within cells are identically distributed and all targets have the same SNR.¹ Both of these assumptions are used in this paper. Let J represent the number of interfering

targets. Using combinatorics, we obtain:
When $J = 0$:

$$g_k(z) = \frac{n!}{(k-1)!(n-k)!} \{ (F_0^{k-1} f_0 (1-F_0)^{n-k} \} \quad (3)$$

When $J = N$:

$$g_k(z) = \frac{n!}{(k-1)!(n-k)!} \{ (F_1^{k-1} f_1 (1-F_1)^{n-k} \} \quad (4)$$

Otherwise:

$$g_k(z) = \frac{J!(N-J)!}{(k-1)!(N-k)!} \left\{ \sum_{\ell_2=\max(0, J-N+k)}^{\min(k-1, J)} F_1^{\ell_2} F_0^{k-1-\ell_2} \cdot f_0 \cdot \{1-F_1\}^{J-\ell_2} \{1-F_0\}^{N-k-J+\ell_2} \binom{k-1}{\ell_2} \binom{N-k}{J-\ell_2} \right. \\ \left. + \sum_{\ell_1=\max(0, J-1-N+k)}^{\min(k-1, J-1)} F_1^{\ell_1} F_0^{k-1-\ell_1} \cdot f_1 \cdot \{1-F_1\}^{J-1-\ell_1} \{1-F_0\}^{N-k-J+1+\ell_1} \binom{k-1}{\ell_1} \binom{N-k}{J-1-\ell_1} \right\} \quad (5)$$

For this paper, f_0 , f_1 , F_0 , and F_1 are chosen to follow exponential distributions. This is true if we assume a Swerling 1 modelled target with Rayleigh distributed voltage, Rayleigh distributed interference, and usage of a square law detector.^{8,9} However, other distributions may also be used. We have, for our assumptions,

$$f_0(z) = \frac{1}{\mu_i} e^{-z/\mu_i}$$

$$f_1(z) = \frac{1}{\mu_i} (1 + \text{SNR}) e^{-z \frac{1}{\mu_i} (1 + \text{SNR})}$$

where μ_i is the mean interference power. Using these distributions, equations (1) and (2) become:

$$P_{FA} = \int_0^\infty g_k(z) \exp\left(-\frac{\alpha z}{\mu_i}\right) dz \quad (6)$$

$$P_D = \int_0^\infty g_k(z) \exp\left(-\frac{\alpha z}{\mu_i(1 + \text{SNR})}\right) dz \quad (7)$$

Table 1: Selected values of α for $P_{FA} = 1 \cdot 10^{-4}$, $N = 20$, as a function of k , J , and SNR

SNR = 100		$k = 7$	$k = 8$	$k = 9$	$k = 10$	$k = 11$	$k = 12$	$k = 13$
	$J = 1$	45.9585	35.2190	28.0072	22.8583	19.0065	16.0160	13.6222
	$J = 2$	43.2375	33.0535	26.2126	21.3256	17.6662	14.8210	12.5385
	$J = 3$	40.5140	30.8846	24.4135	19.7873	16.3188	13.6167	11.4423
	$J = 4$	37.7870	28.7114	22.6091	18.2420	14.9621	12.4001	10.3295
	$J = 5$	35.0561	26.5330	20.7978	16.6876	13.5934	11.1669	9.1929
	$J = 6$	32.3201	24.3480	18.9777	15.1214	12.2083	9.9103	8.0210
	$J = 7$	29.5777	22.1543	17.1459	13.5388	10.7998	8.6184	6.7908
	$J = 8$	26.8269	19.9491	15.2979	11.9331	9.3562	7.2686	5.4487
SNR = 90		$k = 7$	$k = 8$	$k = 9$	$k = 10$	$k = 11$	$k = 12$	$k = 13$
	$J = 1$	45.9585	35.2190	28.0072	22.8583	19.0065	16.0160	13.6222
	$J = 2$	43.2411	33.0565	26.2151	21.3279	17.6683	14.8230	12.5404
	$J = 3$	40.5212	30.8906	24.4188	19.7920	16.3231	13.6208	11.4464
	$J = 4$	37.7981	28.7207	22.6172	18.2493	14.9690	12.4067	10.3360
	$J = 5$	35.0712	26.5457	20.8090	16.6978	13.6031	11.1763	9.2025
	$J = 6$	32.3394	24.3642	18.9921	15.1347	12.2211	9.9231	8.0346
	$J = 7$	29.6014	22.1745	17.1640	13.5559	10.8166	8.6358	6.8105
	$J = 8$	26.8554	19.9737	15.3203	11.9547	9.3782	7.2930	5.4803

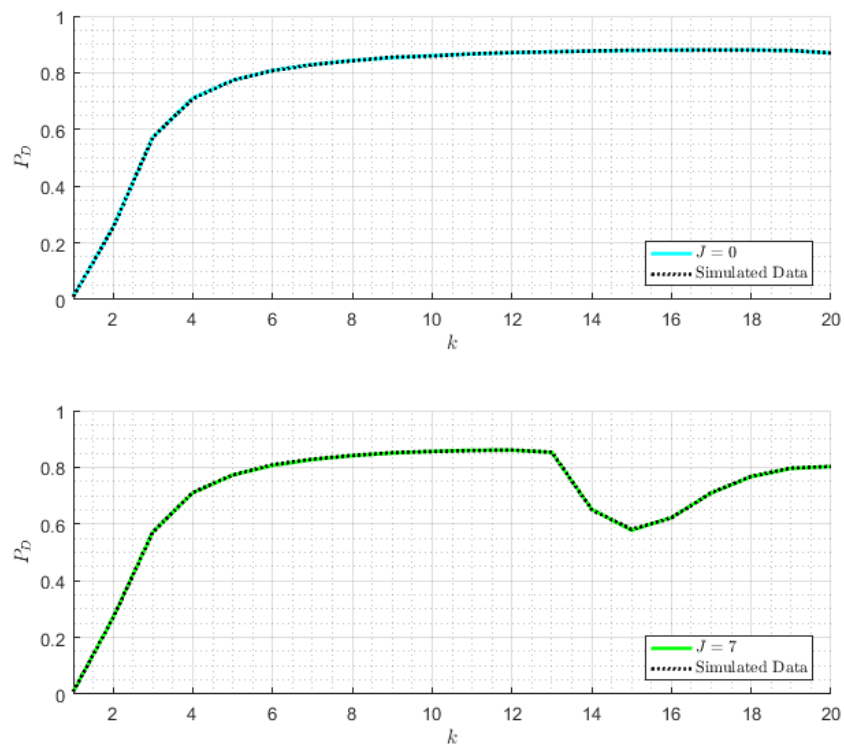


Figure 1: P_D shown as a function of k at $\text{SNR} = 100$. The α values are chosen such that P_{FA} is constant at each value of k and J , as in table 1 ($P_{FA} = 1 \cdot 10^{-4}$, $N = 20$).

Since $g_k(z)$ is quite cumbersome, the integrals in (6) and (7) are calculated numerically. Note that both P_{FA} and P_D are functions of α , k , J , N and SNR . For a fixed P_{FA} and N , we numerically solve (6) for α at all values of k and J , and select values of SNR . Table 1 shows a small collection of these values for $P_{FA} = 1 \cdot 10^{-4}$ and $N = 20$.

Figure 1 displays the P_D versus k curves at two different values of J . Values on the solid colored curves are calculated numerically from equation (7). The simulated P_D (shown on the black dotted curves) is simply the relative number of detections from $1 \cdot 10^5$ Monte Carlo simulated CFAR windows. These curves are plotted together to show very close agreement between theoretical data and simulated data. Note that all simulated CFAR windows in this paper use exponentially distributed interference and interference + target.

2.2 Implementation of FAOSOSD

The FAOSOSD algorithm is used to provide an estimate $\hat{J}$ for the number of interfering targets present. This paper chooses the minimum description length as the information theoretic criteria used to obtain this estimate. Monte Carlo simulations are used to approximate the general performance of this algorithm. $1 \cdot 10^6$ CFAR windows are simulated with a fixed value of J . The FAOSOSD is applied on each of these simulated windows, resulting in $1 \cdot 10^6$ estimates $\hat{J}$. Using the relative frequency of each observation of $\hat{J}$, a PMF $P(\hat{J}|J)$ is obtained. This process is repeated using each value of J .

This approximated PMF can be used to find an a posteriori distribution $P(J|\hat{J})$. This distribution represents the distribution of J given the estimate from the FAOSOSD. Intuitively, the FAOSOSD can be considered accurate when $P(J|\hat{J})$ is high for $J = \hat{J}$ (estimate is equal to the ground truth) and low for all other values of

J . Using Bayes' rule:

$$P(J|\hat{J}) = \frac{P(J)P(\hat{J}|J)}{\sum_{J=0}^N P(J)P(\hat{J}|J)}$$

Assuming a uniformly distributed prior distribution $P(J)$, this can be simplified as:

$$P(J|\hat{J}) = \frac{P(\hat{J}|J)}{\sum_{J=0}^N P(\hat{J}|J)}$$

A non-uniform prior distribution may also be used if at hand. Otherwise ignorance is assumed, and each possible value of J is considered equally likely to occur.

Note that in ref. 4, the FAOSOSD algorithm is also used to set $k = N - (\hat{J} + 1)$. In this paper, this step is omitted, as our application only requires $P(J|\hat{J})$. OS-CFAR has been shown to break down in performance if $k \geq N - J$.³ This will occur if the FAOSOSD over-estimates the number of targets present, i.e. $\hat{J} > J$. Thus, $P(J|\hat{J})$ is obtained to show the likelihood that the FAOSOSD will overestimate/underestimate J at different SNR values. Ultimately, it is desired that a decision (α, k) is selected, whose performance is robust even when the FAOSOSD is less accurate.

Figure 2 shows $P(J|\hat{J} = 5)$ at different SNR values. The FAOSOSD is shown to be fairly accurate at a high SNR of 20 dB, since the PMF is maximized when $J = \hat{J} = 5$, and has fairly low probabilities everywhere else. As the SNR decreases to 17 dB, the performance is shown to be slightly worse, but is still able to correctly estimate J . When the SNR reaches 7 dB, the accuracy is shown to degrade significantly. It should also be noted that this degradation in accuracy increases for higher values of $\hat{J}$.

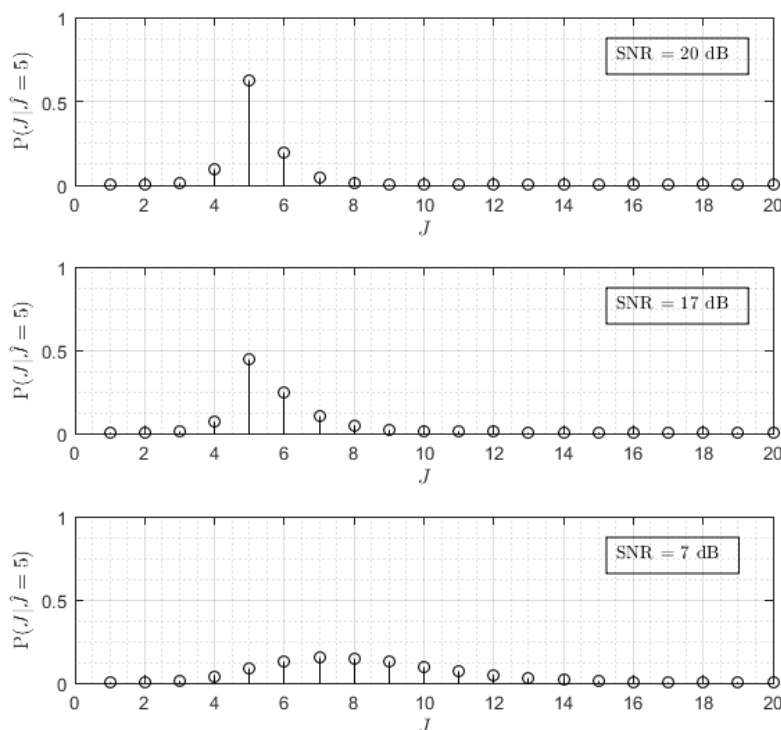


Figure 2: $P(J|\hat{J} = 5)$ shown at varying values of SNR.

3. FORMULATION OF PERFORMANCE FUNCTION

In this section, a function is created to represent the performance of a decision. Since performance may be subjective between decision makers, this function may vary from application to application. For the application in this paper, it is desired for the performance function to be large when P_D is large and P_{FA} is close to its specified value P_{FA}^* . These traits can be obtained using the following function ψ_0 :

$$\psi_0 = P_D^2 \exp \left(- \left| 1 - \frac{P_{FA}}{P_{FA}^*} \right| \right) \quad (8)$$

This function is plotted in Figure 3. For a fixed N , the P_D and P_{FA} are both dependent on α , k , J and SNR, as shown by (5), (6), and (7). Thus, these probabilities can be written as $P_D(\tilde{\mathbf{x}}, J, \text{SNR})$ and $P_{FA}(\tilde{\mathbf{x}}, J, \text{SNR})$, and ψ_0 can be written as $\psi_0(\tilde{\mathbf{x}}, J, \text{SNR})$.

Thus, each decision $\tilde{\mathbf{x}}$ can be thought to produce a different ψ_0 value under unknown conditions J and SNR. However, by obtaining $P(J|\hat{J})$, we are aware that certain values of J are more probable given $\hat{J}$. Thus, we can take the conditional mean of ψ_0 with respect to J :

$$\psi_{\hat{J}}(\tilde{\mathbf{x}}, \text{SNR}) = \sum_{J=0}^N \psi_0(\tilde{\mathbf{x}}, J, \text{SNR}) P(J|\hat{J}) \quad (9)$$

This can be thought of as a weighted sum of ψ_0 values, where $\psi_0(\tilde{\mathbf{x}}, J, \text{SNR})$ is weighted more if J is more probable.

The PMFs in Figure 2 can be used for the following example. First, $\hat{J} = 5$ is obtained from FAOSOSD. Say that the performance of decision $\tilde{\mathbf{y}}$ at SNR = 20 dB (i.e. $\psi_5(\tilde{\mathbf{y}}, 20 \text{ dB})$) is desired. From Figure 2, it is shown that only probabilities of $J = \{4, 5, 6, 7\}$ are relevant at this SNR, and all other probabilities appear to be 0. Thus, even if decision $\tilde{\mathbf{y}}$ performs poorly at $J = 20$, this J is so unlikely that it can be thought of as irrelevant, and $\psi_5(\tilde{\mathbf{y}}, 100)$ is not affected. Only $\tilde{\mathbf{y}}$'s performance at $J = \{4, 5, 6, 7\}$ will impact $\psi_5(\tilde{\mathbf{y}}, 100)$.

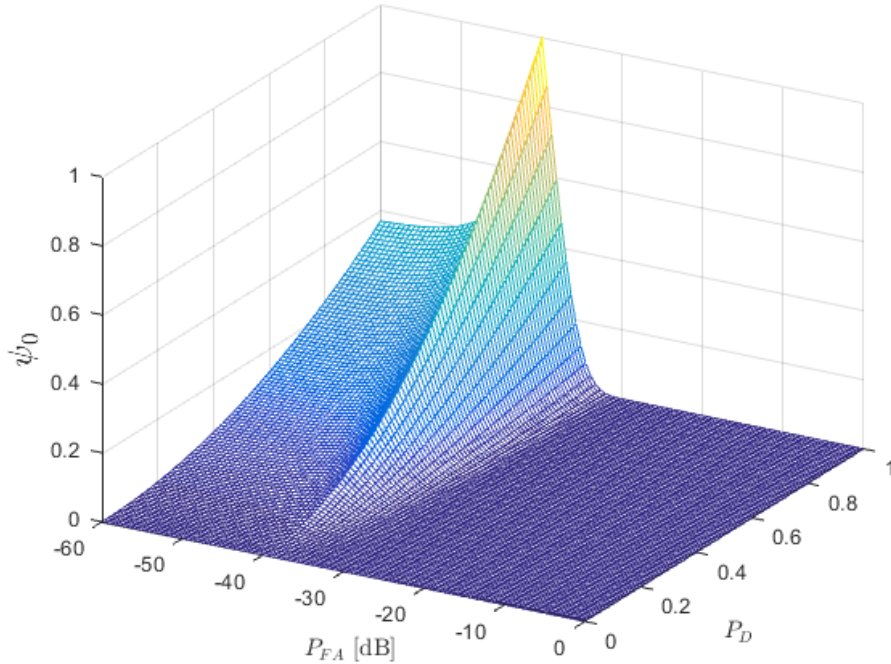


Figure 3: ψ_0 for $P_{FA}^* = 1 \cdot 10^{-4}$.

4. ROBUST DECISION MAKING METHOD

Recall that this robust decision making method seeks to maximize the absolute performance of a decision $\tilde{\mathbf{x}}$ while also minimizing its sensitivity to SNR. To accomplish this, two measures are defined, as in Ref. 5. The absolute performance is represented by the mean of $\psi_j(\tilde{\mathbf{x}}, \text{SNR})$ in a neighborhood around a specified SNR value. Similarly, the variance of ψ_j in this same neighborhood is chosen to represent the sensitivity to SNR. The specified SNR around which the neighborhood is defined is represented by $\hat{\text{SNR}}$.

These measures are estimated by finding the sample mean and variance of several points within the specified neighborhood. Let M represent the number of points evaluated in this neighborhood. The estimates are formulated as follows:

$$\hat{\mu}(\tilde{\mathbf{x}}, \hat{\text{SNR}}) = \frac{1}{M} \sum_{i=1}^M \psi_j(\tilde{\mathbf{x}}, \hat{\text{SNR}} + \Delta \mathbf{z}_i) \quad (10)$$

$$\hat{\text{Var}}(\tilde{\mathbf{x}}, \hat{\text{SNR}}) = \frac{1}{M-1} \sum_{i=1}^M \left| \psi_j(\tilde{\mathbf{x}}, \hat{\text{SNR}} + \Delta \mathbf{z}_i) - \hat{\mu}(\tilde{\mathbf{x}}, \hat{\text{SNR}}) \right|^2 \quad (11)$$

where $\Delta \mathbf{z}_i$ is a vector of small numbers such that $\hat{\text{SNR}} + \Delta \mathbf{z}_i$ are points within the neighborhood. For the purposes of this paper, this vector is generated deterministically, however, stochastic methods may also be used.⁵ The specified $\hat{\text{SNR}}$ and its surrounding neighborhood are selected by the user to define the region the decision maker would like the system to be robust within.

Figure 4 displays these measures for the case when the specified neighborhood is $\text{SNR} : [5, 95]$. Thus, $\hat{\text{SNR}} = 50$, $M = 20$, and $\Delta \mathbf{z} = \{-45, -40, -35, \dots, 45\}$. The $\hat{\mu}$ and $\hat{\text{Var}}$ values are plotted for 4400 decision points $\tilde{\mathbf{x}} = (\alpha, k)$. Each decision $k = 1, 2, \dots, N$ is considered, and only values of α that satisfy equation (1) are considered. These α values must produce $P_{FA} = 1 \cdot 10^{-4}$ for at least some SNR and some J . The intuition for

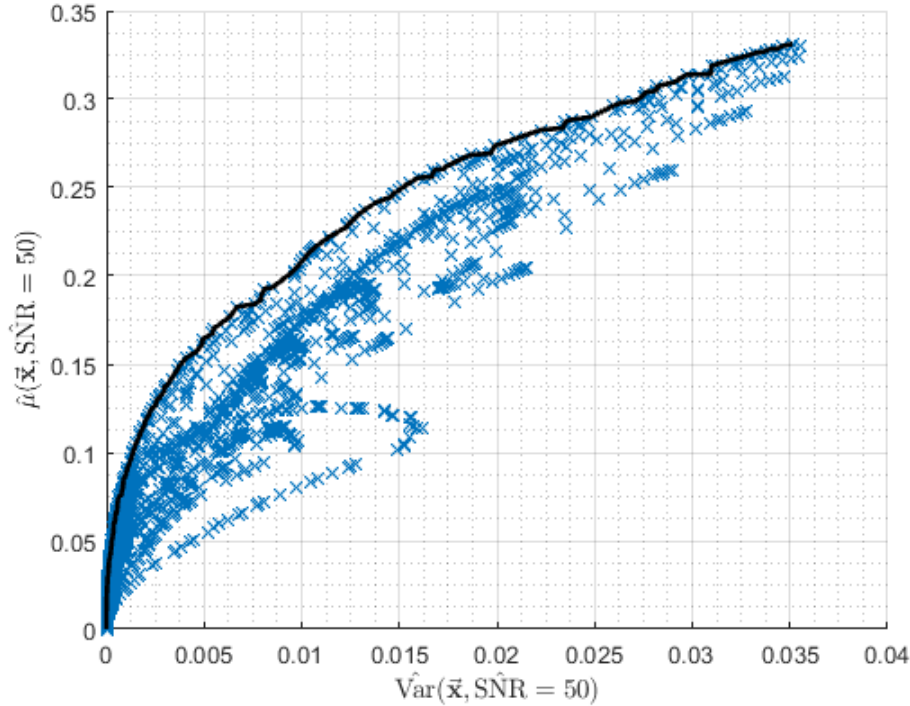


Figure 4: $\hat{\mu}$ and $\hat{\text{Var}}$ of many decision points for when $\hat{\text{SNR}} = 50$ and $\hat{J} = 5$ ($N = 20$, $P_{FA} = 1 \cdot 10^{-4}$). The Pareto frontier is shown by the black curve.

this is that if α is such that $P_{FA} = 1 \cdot 10^{-4}$ is not satisfied anywhere, its values for equation (8) and (9) would be so low, that this decision would be rejected anyway.

Recall that the goal of this decision making method is to maximize $\hat{\mu}$, and minimize $\hat{\text{Var}}$. The Pareto frontier is found with this in mind, and is shown in Figure 4 by the black curve. The points on this curve represent all of the Pareto efficient decisions. A decision $\tilde{\mathbf{x}}$ is considered to be Pareto efficient, if there is no decision $\tilde{\mathbf{x}}^*$ such that $\hat{\mu}(\tilde{\mathbf{x}}^*, \hat{\text{SNR}}) > \hat{\mu}(\tilde{\mathbf{x}}, \hat{\text{SNR}})$ and $\hat{\text{Var}}(\tilde{\mathbf{x}}^*, \hat{\text{SNR}}) < \hat{\text{Var}}(\tilde{\mathbf{x}}, \hat{\text{SNR}})$ simultaneously. From this curve, a direct trade-off between the absolute performance and robustness can be seen.

5. INFORMATION ELASTICITY ANALYSIS

To select a decision along this Pareto frontier, information elasticity analysis is employed. This analysis seeks to define a metric of decision effectiveness, E , and observe its behavior to change in some other parameter. E is typically a function of a decision quality metric D , which should be a parameter where large values are desirable, and a constraint function C , which should be a parameter where large values are undesirable. In this application, $D = \hat{\mu}(\tilde{\mathbf{x}}, \hat{\text{SNR}})$ since a high mean is desired, and $C = \hat{\text{Var}}(\tilde{\mathbf{x}}, \hat{\text{SNR}})$ since a low variance is desired.

Ref. 10 proposes a model of decision effectiveness using the ratio $E = D/C$. This simple model is inadequate for our application, since there are many decisions along the pareto curve where $C \approx 0$. Instead, a new model for decision effectiveness that takes the L_2 distance from a model point to a decision point is used. A model point is a user defined $(\hat{\mu}, \hat{\text{Var}})$ pair that the decision maker considers to be exemplary. Since $\hat{\text{Var}} = 0$ is ideal, this point will always have the form $(\mu_0, 0)$, where μ_0 is a user defined value between 0 and 1. E is thus defined as:

$$E = \sqrt{(\mu_0 - \hat{\mu})^2 + (\hat{\text{Var}}/\mu_0)^2} \quad (12)$$

In equation (12), the $\hat{\text{Var}}$ is normalized by μ_0 so as to represent the variability of the decision around the specified μ_0 value. Since this quantity represents distance from an exemplar point, the decision is considered

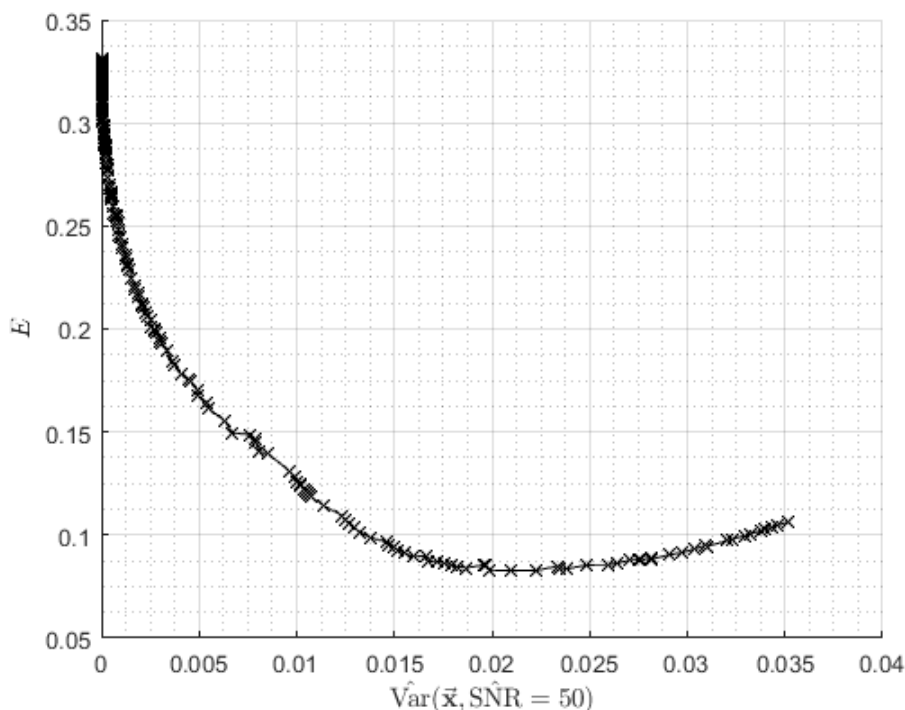


Figure 5: E as a function of $\hat{\text{Var}}$. Note that the overload point occurs at $\hat{\text{Var}} = 0.02221$.

to be more effective the smaller this distance E is. Figure 5 shows an example of this approach, using the Pareto frontier shown in Figure 4. Here, μ_0 is selected to be 0.3308, since this is the maximum $\hat{\mu}$ value that can be obtained for the specified neighborhood of SNR values (can be seen in Figure 4). We note that as the variance begins to decrease, the decisions appear to get more effective. Eventually, an overload point is reached at $\hat{\text{Var}} = 0.02221$, and the decisions begin to get less effective. The decision pair that occurs at this overload point is thereby selected. As for the example in Figure 5, the decision that offers the minimum value of E is $\tilde{\mathbf{x}} = (\alpha = 17.9850, k = 9)$.

6. CONCLUSIONS

The detection and false alarm probabilities for OS-CFAR are calculated as a function of decision variables α and k , and environmental variables J and SNR. ψ_0 is defined to be a function of these probabilities, and represents performance. The FAOSOSD algorithm is used to obtain $\hat{J}$, an estimate for J . Using this estimate, we obtain $\psi_{\hat{J}}(\tilde{\mathbf{x}}, \text{SNR})$: the conditional mean of ψ_0 with respect to J and given $\hat{J}$. The sample mean and variance of this quantity are calculated over a user-defined range of SNRs.

When observing this performance mean and variance at each decision point $\tilde{\mathbf{x}}$, a direct trade-off between these quantities is shown. The decision effectiveness is defined to be the L_2 distance from a user-defined model point to a decision point, and the decision with the smallest distance is selected. It is shown that under certain conditions, an overload point may exist when comparing decision effectiveness and performance variance.

ACKNOWLEDGMENTS

This work was supported by the US Air Force Office of Scientific Research Grant Number FA9550-17-1-0032 (POC: R.D. Riecken).

REFERENCES

- [1] Blake, S., "Os-cfar theory for multiple targets and nonuniform clutter," *IEEE Transactions on Aerospace and Electronic Systems* **24**, 785–790 (Nov 1988).
- [2] Gandhi, P. P. and Kassam, S. A., "Analysis of cfar processors in nonhomogeneous background," *IEEE Transactions on Aerospace and Electronic Systems* **24**, 427–445 (July 1988).
- [3] Levanon, N., "Detection loss due to interfering targets in ordered statistics cfar," *IEEE Transactions on Aerospace and Electronic Systems* **24**, 678–681 (Nov 1988).
- [4] Magaz, B., Belouchrani, A., and Hamadouche, M., "Automatic threshold selection in os-cfar radar detection using information theoretic criteria," *Progress In Electromagnetics Research B* **30**, 157–175 (01 2011).
- [5] Jin, Y. and Sendhoff, B., "Trade-off between performance and robustness: An evolutionary multiobjective approach," in [*Evolutionary Multi-Criterion Optimization*], Fonseca, C. M., Fleming, P. J., Zitzler, E., Thiele, L., and Deb, K., eds., 237–251, Springer Berlin Heidelberg, Berlin, Heidelberg (2003).
- [6] Narayanan, R. M., Liu, A. Z., Singerman, P. G., and Rangaswamy, M., "Information elasticity in radar systems," *Electronics Letters* **54**(17), 1049–1051 (2018).
- [7] David, H. and Nagaraja, H., [*Order Statistics*], Wiley Series in Probability and Statistics, Wiley (2004).
- [8] Richards, M., Holm, W., and Scheer, J., [*Principles of Modern Radar: Basic Principles*], Electromagnetics and Radar, Institution of Engineering and Technology (2010).
- [9] Rohling, H., "Radar cfar thresholding in clutter and multiple target situations," *IEEE Transactions on Aerospace and Electronic Systems* **AES-19**, 608–621 (July 1983).
- [10] Liu, A. Z., Narayanan, R. M., and Rangaswamy, M., "Information elasticity in pseudorandom code pulse compression," in [*Radar Sensor Technology XXII*], *Proc. SPIE* **10633** (2018).

Cost Function Design for Modeling Information Overload in Radar Systems

Paul G. Singerman^a, Ram M. Narayanan^{a*}, Muralidhar Rangaswamy^b,

^aThe Pennsylvania State University, University Park, PA 16802, USA

^bAir Force Research Laboratory, AFRL/RYP, Wright-Patterson AFB, OH 45433, USA

ABSTRACT

It is a common assumption that more effective decisions can be made in radar sensor processing if more “information” is available. However, this is not always the case since more information may not necessarily translate to better sensor performance due to a phenomenon known as information overload. What determines the effectiveness of a decision is rarely only dependent on a single performance metric. Instead, there are often multiple performance metrics involved. This necessitates the use of cost functions for modeling a loss in decision effectiveness due to deviations from the optimal values of the performance metrics. These optimal values are also often determined in a fuzzy/subjective manner by an individual using language. Two different types of statements are looked at and modeled, namely, soft and hard constraining statements. These models, called language-based cost functions (LBCFs), are derived and their properties are explored. These can then be easily combined, via a summation, to obtain a smooth objective function. The objective function is then used to find the utility of the radar system. By altering the systems parameter values the amount of information introduced into the system can be varied and the point at which the utility of the radar system decreases due to excess information can be found. This is then regarded as the information overload point or the point at which more information causes a decrease in the utility of the system. An example of a radar system with imposed constraints is provided and its resulting information overload point is found by utilizing LBCFs.

Keywords: information overload; performance metrics; radar sensor processing; radar systems; cost function design, constraining statements, multi-objective optimization, language modeling

1. INTRODUCTION

Information overload has been defined as the point at which more information starts to negatively affect the total utility of a system.¹ That which determines the utility of a system is rarely only dependent on a single performance metric. Instead, there are often multiple performance metrics involved. This requires cost functions to map from the partially ordered set of multiple performance metrics to a totally ordered set, $\mathbb{R}$. In a previous paper, the cost function was arbitrarily chosen to be a fraction with the performance metrics that were positively related to the utility in the numerator and the performance metrics that were negatively related to the utility in the denominator.² This included very little context on the operating goals and requirements on the radar system. A more complete theory of information overload in radar systems should take this context into account.

When designing a radar system, the specific performance objectives given to it must be kept in mind and accounted for. Often, there are both performance requirements and performance goals that must be met. Requirements are fundamental to the system design and cannot be violated. Goals, on the other hand, can be violated but the radar system would have far more utility if they were met. This often leads to goals being more difficult to achieve than requirements.

Habitually, goals and requirements are explained in terms of performance metrics or sensor constraints for the radar system. By altering sensor parameters (e.g. operating bandwidth, pulse width) performance metrics can obtain different values (e.g. maximum detection range, time-to-detect, etc...). The question then becomes, “From a list of goals and performance metrics for a radar system, what is the best combination of sensor parameters?”

Further author information: (Send correspondence to R.M.N.)

R.M.N.: E-mail: ram@engr.psu.edu, Telephone: 1 814 863 2602, Fax: 1 814 865 7065

Radar Sensor Technology XXIII, edited by Kenneth I. Ranney, Armin Doerry, Proc. of SPIE Vol. 11003, 110030A · © 2019 SPIE · CCC code: 0277-786X/19/\$18 · doi: 10.1117/12.2519726

DISTRIBUTION A: Distribution approved for public release.
Proc. of SPIE Vol. 11003 110030A-1

An issue that arises is that a list of goals and requirements does not provide enough context to determine what exactly is “best”. In this case, the “best” system could be the system that achieves all requirements and fulfills the most goals. Unfortunately, by this definition, this will often end in a tie as there could be uncountably many different combinations of sensor parameters that could be considered “best”. While this lowers the pool of options to some extent, a decision as to which combination should be selected still needs to be made.

If more context could be given or ascertained from goals and requirements, then a single “best” solution could be found instead of a Pareto optimal surface with many “best” solutions. This context could contain information such as: which goals are more important and by how much, which value (even if unrealistic) is optimal for a signal metric or parameter, and/or what is the improvement gained from parameter values sitting close to goals compared to barely satisfying the requirements. This added context can be modeled by a novel method utilizing language-based cost functions (LBCFs). Different types of statements correspond to different types of LBCFs and depending on the context given in the statements, intuitively appealing cost functions are output. The summation of all LBCFs is then treated as an objective function for which existing optimization techniques can be used.

Using context rich LBCFs, the information overload point can be calculated by holding fixed the sensor parameters that do not impact how much information is input into the system and finding the utility over different values of the other parameters. This will give a utility versus information curve. By taking the derivative of this curve, the information elasticity at different points can be calculated.

New technology has opened the door for real-time adaptation of radar sensor parameters. Using this proposed framework the radar sensor parameters could be chosen such that they give a result that is optimal based on the constraints given. This is primarily the job of the controller block in a cognitive radar.^{3,4} This framework could easily be extended into the domain of machine cognition.

2. LANGUAGE-BASED COST FUNCTIONS FOR SOFT CONSTRAINING STATEMENTS

To aid in designing a radar system there are often goals and requirements placed on the sensor parameters and/or the performance metrics. Requirements must be met while goals are not crucial to the design of the radar, but are still desirable. Goals are typically more difficult to achieve than requirements for this reason.

The idea of utility is to quantify how good or bad a system performs, with high utility signifying good performance and vice-versa. This is a simple problem if there is only one sensor parameter and the goal is to just maximize/minimize a performance metric dependent on this parameter. This becomes more difficult when there are several parameters and the goal is to maximize/minimize a performance metric that depends on all of these sensor parameters. In real world systems, not only are there several sensor parameters but also there are several performance metrics. This means in order to maximize a system composed of these performance metrics we need a way of distinguishing how “important” these performance metrics are compared to each other. This has proven to be the most difficult part of trying to characterize information overload as it is often impossible to find an objective way to combine the different performance metrics into a utility function.

A possible solution is to take the subjective opinion of domain experts. Even though this method of combination is subjective, it is still valuable, much like a prior distribution in Bayesian statistics is subjective but still widely accepted to be valuable. The prior has several interpretations: encoding previously observed related observations (pseudo data), biasing the parameter estimation towards more realistic or probable values, or imposing constraints on the values a parameter can take, and estimating parameters with regularization.⁵ Cost functions can also be used to represent the validity of a decision. With lower cost representing a good decision and vice-versa. This is opposite to how utility is treated with high utility indicating a good decision. A natural way to relate the utility and loss (total cost) is to make the utility the inverse of the loss

$$U(\theta) = \frac{1}{L(\theta)}. \quad (1)$$

Here, U is the total utility of the system, L is the total cost or loss of the system, and $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ represents the set of sensor parameters. This, of course, means that as the loss goes to zero, the utility goes to infinity and vice-versa.

Cost functions which, like the prior, take pseudo information and encode it along with imposing constraints, would be extremely valuable in determining the loss and, therefore by equation (1), utility of a radar system. These cost functions should also be easy to design and require little to no knowledge about the underlying mathematics. They should allow a system designer to turn statements about a radar system into a full model of its utility. A novel way of modeling this subjective information has been developed and is called the language-based cost function (LBCF).

Since there are both goals and requirements to a radar system this leads to the idea of two different types of constraining statements. One type of constraining statement is called a soft constraining statement. A soft constraining statement is what models a goal of a radar system. It is “soft” because the constraint can be broken but with an abrupt penalty. Shown below are three examples of soft constraining statements:

- (1) The total detection time **should** be below 5 milliseconds.
- (2) The bandwidth **should** be below 10 MHz.
- (3) The total power **should** be above 5 watts.

In all three examples the word “should” is the key for the softness of the constraint as it suggests a desire or goal for the radar system rather than an absolute. The three soft constraining statements shown above each could correspond to a goal given to a system designer that, if met, would improve the utility of the radar system. Just as the utility should sharply increase when a goal is met the cost should sharply increase when the goal is not met. This task could be easily solved with piece-wise functions. Unfortunately, piece-wise functions are usually not differentiable everywhere which would make optimization more difficult later on, so another solution should be found.⁶ Therefore, a continuous function with a sharp increase at a specific value (the constraint) which then levels off again is desired.

2.1 Exponential-Sigmoid Mixture Functions

In the first two examples, the constraining statements use the preposition “below”. This indicates that the optimal parameter will be less than the constraint. This necessitates the need for a function that has a low value at values less than the constraint and also has the properties described earlier. A function that accomplishes this task is the sigmoid function which is used many fields including: machine learning, network utility maximization, and political economy.⁷⁻⁹ A general equation for the sigmoid function is

$$C(\theta) = \frac{T}{1 + \exp(-r(\theta - c))}. \quad (2)$$

This general sigmoid has a domain $(-\infty, \infty)$ and a range $(0, T)$ with a sharp increase around c . For this use case, c can be thought of as representing the soft constraint, T represents the amplitude of the increase, and r represents the “sharpness” of the increase when approaching the constraint. Since this function should scale properly with the constraint, the parameter r should be chosen such that it is a function of the constraint. An arbitrary but sufficient choice for r is $r = \frac{10^4}{c}$. As can be seen in Figure 1, this choice for r produces a function that scales properly and has a jump around the constraint that is always equally sharp no matter how big or small the constraint.

This function encodes the model with the context of the soft constraint. However, there are still two more implied contexts given in the statement that should be accounted for in the model. In the first two example statements, the preposition below indicated that lower parameter values are better. This means that there should be some penalty due to the parameter value increasing regardless of where the constraint is located (directional context). Furthermore, if the directional context is extended to all positive values then the optimal value, in this case, should be 0 (even though 0 is often not realistic it is still optimal). To account for this added context in the model, two more terms are added to the sigmoid function

$$C(\theta) = \frac{T}{1 + \exp\left(-\frac{10^4}{c}(\theta - c)\right)} + \exp\left(\frac{\theta}{c}\right) + \beta. \quad (3)$$

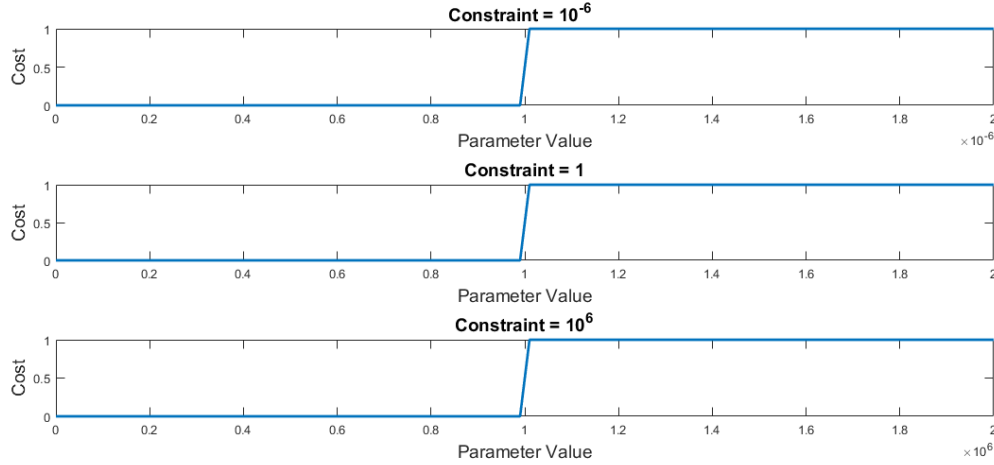


Figure 1: Three sigmoid functions modeled using equation (2) with $T = 1$ and $r = \frac{10000}{c}$ for different constraints of (top) $c = 10^{-6}$, (middle) $c = 1$, and (bottom) $c = 10^6$.

Here, $\exp\left(\frac{\theta}{c}\right)$ represents the context that lower values of the parameter θ give the system higher utility and the further the parameter is from the optimal value the more the cost per unit distance. The parameter, θ , is divided by the constraint c to assure that the cost function is, once again, scale invariant. The term β is a constant that makes $f(0) = 0$ which represents minimum cost at a parameter value of zero. Solving for β by letting $f(0) = 0$ gives $\beta = -1$.

Now, the function is of the form

$$C(\theta) = \frac{T}{1 + \exp\left(-\frac{10^4}{c}(\theta - c)\right)} + \exp\left(\frac{\theta}{c}\right) - 1. \quad (4)$$

The parameter T still remains to be explained, which represents the amplitude of the sigmoid jump. For the purpose of preserving the scalability of the function no matter the value of the constraint, T should be chosen to represent the percent jump in cost that results when the soft constraint is broken. Based on this requirement, the numerator of the sigmoid term ends up being $T = A(\exp(1) - 1)$, where A is the ratio representing the percent increase ($A = 1 = 100\%$ increase). This gives the solution

$$C(\theta) = \frac{A(\exp(1) - 1)}{1 + \exp\left(-\frac{10^4}{c}(\theta - c)\right)} + \exp\left(\frac{\theta}{c}\right) - 1. \quad (5)$$

It is easy to see how the value of A determines the size of the jump at the constraint, meaning A represents how much cost is incurred from breaking the soft constraint. The parameter A gives context between the direction cost and the cost incurred by the constraint.

If there are multiple parameters, there also needs to be context between each parameter. This context will let the model account for which parameters are more or less important. To account for this, the whole function is multiplied by an importance term I . This gives the final solution for the model of a soft constraint/goal,

$$C(\theta) = I \left(\frac{A(\exp(1) - 1)}{1 + \exp\left(-\frac{10^4}{c}(\theta - c)\right)} + \exp\left(\frac{\theta}{c}\right) - 1 \right). \quad (6)$$

Figure 2 shows how the amplitude and importance terms affect the resulting models. It can be seen that A only affects the constraint penalty while I affects both the direction cost and the constraint penalty. This function, termed an Exponential-Sigmoid Mixture Function (ESMF), is used to represent LBCFs for parameters when the parameter's optimal value is 0 and the constraint is soft.

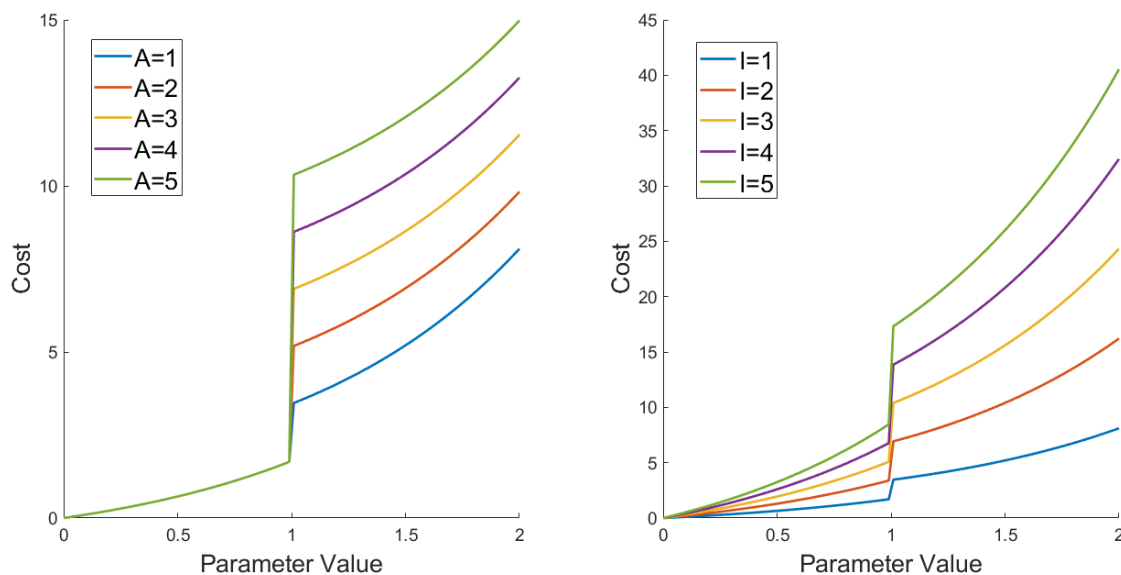


Figure 2: (Left) Five sigmoid functions are modeled using equation (5) with $c = 1$ for different constraint amplitude values A ranging from 1 to 5 and (right) five sigmoid functions are modeled using equation (7) with $c = 1$ and $A = 1$ for different importance values I ranging from 1 to 5.

2.2 Inverse Exponential-Sigmoid Mixture Functions

The ESMF is a cost function used to model a soft constraint that has its optimal value at zero, which is true for the first two example soft constraining statements given earlier. The third example, instead, uses the preposition “above”. By way of the same logic used before, it can be reasoned that the optimal parameter value is infinity. This behooves the need for a function with all the same properties as the ESMF except that its cost approaches zero as the parameter value approaches infinity and vice-versa.

The function that represents a soft constraining statement of this kind can be found by following the same steps as earlier, but instead constraining the function to approach zero as it approaches infinity. This yields

$$C(\theta) = I \left(\frac{A (\exp(1) - 1)}{1 + \exp(-10^4 (\frac{c}{\theta} - 1))} + \exp\left(\frac{c}{\theta}\right) - 1 \right). \quad (7)$$

This function is called an inverse exponential-sigmoid mixture function (IESMF). Figure 3 illustrates an example of the cost function this function generates along with a comparison to an ESMF. It is apparent that the jumps at the constraint are identical but the IESMF grows larger faster than the ESMF by design.

3. LANGUAGE-BASED COST FUNCTIONS FOR HARD CONSTRAINING STATEMENTS

While the ESMF and the IESMF together take care of soft constraining statements corresponding to system goals, there are still hard constraining statements corresponding to system requirements left to address. The cost functions modeling the hard constraining statements should still be continuous and not represented by a piece-wise function. They should also be infinitely differentiable to aid in optimization later on.

3.1 Exponential-Logit Mixture Functions

Requirements, by definition, are not to be violated. This means they are integral to the design of the system. A cost function representing a requirement would be similar to what was obtained for a goal except that once the constraint is violated the cost should jump to an arbitrarily large number and stay there as the parameter value increases. This arbitrarily large value should be chosen such that a digital system could not distinguish it from

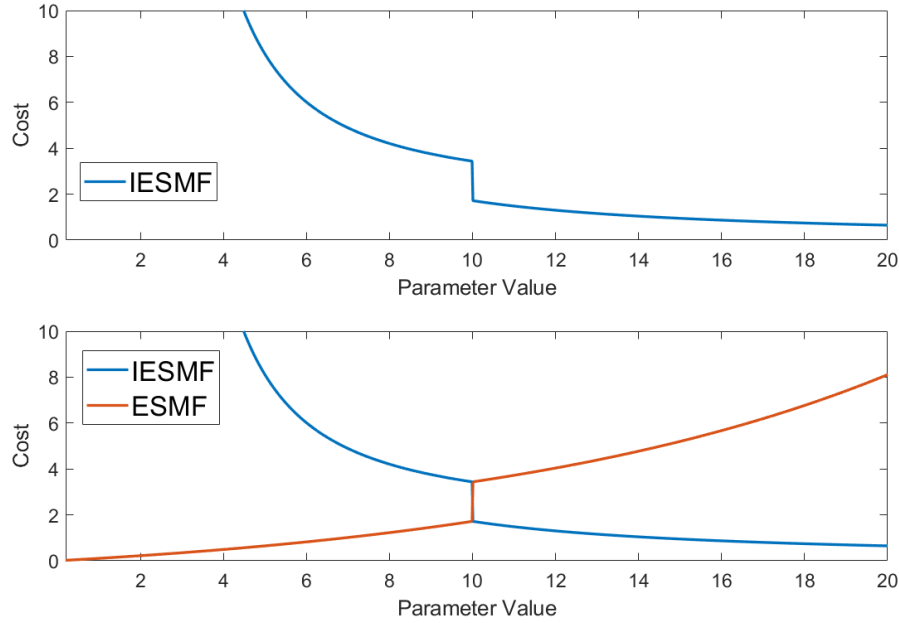


Figure 3: (Top) an IESMF representing a sensor parameter with an optimal value of infinity and a soft constraint at 10 and (bottom) the same IESMF is compared to an identical ESMF.

infinity. A function that accomplishes this task for a parameter which has an optimal value at 0 is logit-like and is defined by

$$C(\theta) = I \left(\ln \left(\frac{1}{1 - \frac{1}{1 + \exp\left(-\frac{10^4}{c}(\theta - c)\right)}} \right) + \exp\left(\frac{c}{\theta}\right) - 1 \right). \quad (8)$$

Here the parameter of a regular logit function is replaced with a sigmoid with amplitude one. This is called an exponential-logit mixture function (ELMF).

When the optimal value is infinity the LBCF is

$$C(\theta) = I \left(\ln \left(\frac{1}{1 - \frac{1}{1 + \exp\left(-10^4\left(\frac{c}{\theta} - 1\right)\right)}} \right) + \exp\left(\frac{c}{\theta}\right) - 1 \right). \quad (9)$$

This function is called an inverse exponential-logit mixture function (IELMF). Figure 4 provides an example of both an ELMF and an IELMF. When a parameter has both a soft and hard constraint the logit-like term and sigmoid term should be added along with the directional context. A parameter with both a hard and soft constraint with an optimal value at zero would have an LCBF of the form

$$C(\theta) = I \left(\ln \left(\frac{1}{1 - \frac{1}{1 + \exp\left(-10^4\left(\frac{c_h}{\theta} - 1\right)\right)}} \right) + \frac{A(\exp(1) - 1)}{1 + \exp\left(-\frac{10^4}{c_s}(\theta - c_s)\right)} + \exp\left(\frac{c_s}{\theta}\right) - 1 \right). \quad (10)$$

Here, c_s is the soft constraint and c_h is the hard constraint. Even though, two constraints are being modeled for the same parameter, there is still only one directional context term, $\exp\left(\frac{c_s}{\theta}\right) - 1$, and the term representing directional context is scaled by the soft constraint.

Utilizing these four different LBCFs, a vast number of language based constraints can be modeled.

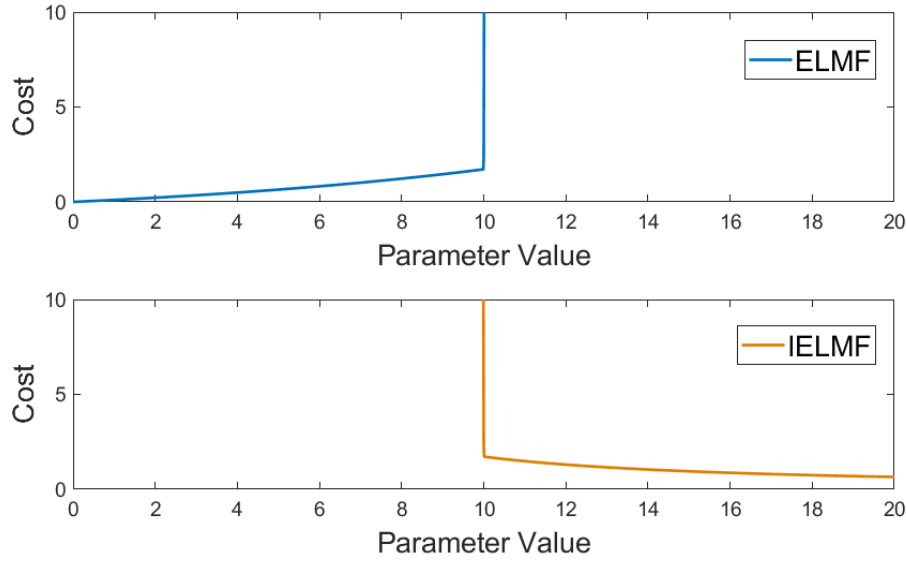


Figure 4: (Top) an ELMF representing a sensor parameter with an optimal value of 0 and a hard constraint at 10 and (bottom) an IELMF representing a sensor parameter with an optimal value of infinity and a hard constraint at 10.

4. MULTI-OBJECTIVE OPTIMIZATION

In the previous section, cost functions for four different types of constraining statements were modeled. These cost functions should allow the modeling of a broad amount of constraining statements imposed on both performance metrics and radar sensor parameters. Individual cost functions must still be combined into one loss function which can then be minimized. The parameter values that minimize this function are the optimal set of parameter values according to the constraints.

The importance term I in the last section makes the process of finding the loss function exceedingly simple. The importance term acts as a weight to give the parameters context between one another, so the loss function is just an unweighted summation of the individual LBCFs. Considering both sensor and performance metrics can be given LBCFs, we can model the objective function as

$$L(\theta) = \sum_{i=1}^k (C_i^s(\theta_i)) + \sum_{i=1}^m (C_i^p(f_i(\theta))). \quad (11)$$

Here $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ is the set of tunable sensor parameters, k is the number of constrained, tunable sensor parameters, m is the number of constrained performance metrics, $C_i^s(\cdot)$ is the LBCF for the i th parameter θ_i , $C_i^p(\cdot)$ is the LBCF for the i th performance metric, and $f_i(\cdot)$ is the function between all the sensor parameters and the i th performance metric.

The goal is now to find θ which minimizes $L(\theta)$, that is

$$\hat{\theta} = \arg \min_{\theta} L(\theta). \quad (12)$$

This optimization problem is beyond the scope of this paper so the reader is referred to literature where this is addressed.^{10, 11}

5. INFORMATION OVERLOAD

The information overload point of a radar system is the point at which extra information starts to negatively affect the utility of the system. As discussed previously, the utility of a system is a subjective quantity and the

system with high utility will change based on the user and the operational context. By this same argument, the appropriate amount of information also changes based on the context. Using LBCFs, which were constructed earlier, the context for specific tasks can be modeled. By varying one parameter that has an impact on total information and using equations (1) and (11) we can find the utility of the radar system at any distinct information value. This allows the information versus utility curve to be visualized and the corresponding overload point to be found.

5.1 A Radar System Example

A simple example system with constraints made with language will be explored to provide a better understanding of information overload and its causes. The system will be a basic pulsed radar with the task of target detection. Three simple constraining statements on the radar system are made that represent statements that would likely be issued by someone in a directorial position, such as a military officer or CEO. These statements are:

- (1) Maximum detection range **must** be greater than 30 km.
- (2) Maximum detection range **should** be greater than 50 km.
- (3) Total detection time **should** be less than 100 milliseconds.

In this case, there are two constraining statements for the maximum detection range and one for the detection time. It is straightforward to see that statements (2) and (3) are soft constraining statements and statement (1) is a hard constraining statement. The LBCF generated for the maximum detection range by statements (1) and (2) is

$$C_1^p(R_{\max}) = \frac{\exp(1) - 1}{1 + \exp\left(-10^4 \left(\frac{3 \times 10^4}{R_{\max}} - 1\right)\right)} + \ln \left(\frac{1}{1 - \frac{1}{1 + \exp\left(-10^4 \left(\frac{5 \times 10^4}{R_{\max}} - 1\right)\right)}} \right) + \exp\left(\frac{3 \times 10^4}{R_{\max}}\right) - 1. \quad (13)$$

Here, the leftmost term is generated by statement (2) (soft constraining statement), the second term is generated by statement (1) (hard constraining statement), the third term gives the model context that a higher maximum range is an improvement over a lower maximum range, and $A = I = 1$ because the statements do not make any distinction between which parameters or goals are more important. Therefore, there is no context to be modeled.

The LBCF generated by statement (3) for the total detection time is

$$C_2^p(t_D) = \frac{\exp(1) - 1}{1 + \exp(-10^5 (t_D - 0.1))} + \exp(10t_D) - 1. \quad (14)$$

These cost functions are plotted in Figure 5. Total detection time and maximum detection range are both performance metrics. To simplify this example, the pulse width τ is the only sensor parameter that is tunable. The pulse width affects the bandwidth, which in turn affects the amount of information the system generates and has to deal with.

Here, the constraints are placed on performance metrics rather than directly on the radar sensor parameters. This means that the functional relationships between the sensor parameters and the performance metrics must be modeled. These functions will correspond to the functions $f_i(\cdot)$ in equation (11).

The pulse width and maximum detection range can be related directly by the radar range equation

$$f_1(\tau) = R_{\max} = \left(\frac{P_t G^2 \lambda^2 \sigma \tau}{(4\pi)^3 k T_e F L (\text{SNR})} \right)^{\frac{1}{4}}. \quad (15)$$

This equation makes the maximum detection range a function of the pulse width.

Unlike the maximum detection range, a few separate components impact the total detection time. The first is the time it takes for the signal to go through its entire generation, or τ seconds. The second is the travel time

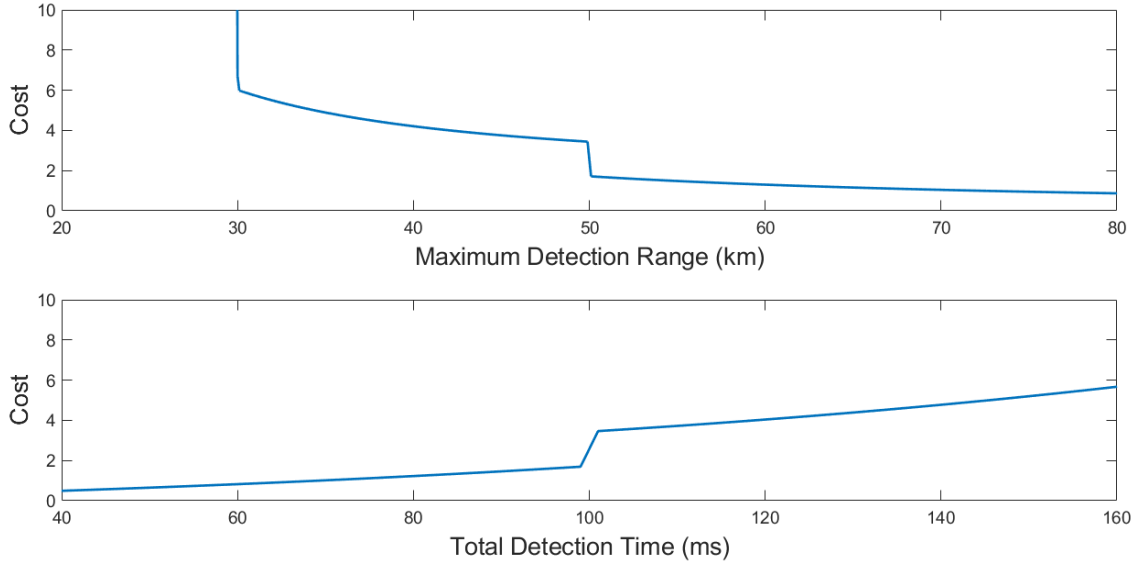


Figure 5: (Top) The LBCF corresponding to the hard and soft constraints given for the maximum detection range and (bottom) the LBCF corresponding to the soft constraint given for the detection time.

the signal takes to reach its target and return to the receiver. While this may be negligible in most situations, if there is a strict constraint on time, this may cause it to become critical in assessing the total utility of the radar system. The third is the amount of pulses used. While extra pulses can increase the detection range, they also add to the detection time. Lastly, the total time the detection algorithm takes to perform the target detection will be critical in determining the values of τ that optimize utility.

The time needed for the signal to travel to maximum range and back is

$$t(\tau) = \frac{2R_{max}}{c} = \left(\frac{16P_t G^2 \lambda^2 \sigma \tau}{(4\pi)^3 k T_e F L (\text{SNR}) c^4} \right)^{\frac{1}{4}}. \quad (16)$$

The time needed for extra pulses will be a function of the pulse repetition time, T . Therefore, the total detection time for a radar utilizing multiple pulses will be

$$t(\tau) = (N - 1)T + \left(\frac{16P_t G^2 \lambda^2 \sigma \tau}{(4\pi)^3 k T_e F L (\text{SNR}) c^4} \right)^{\frac{1}{4}} + \tau, \quad (17)$$

where N is the number of pulses. The $(N - 1)$ term comes from the fact that on the last pulse the full duration of the pulse repetition time should not be included. The total processing time will be based on the sampling rate, complexity of the detection algorithm used, processor speed, and pulse width. This can be modeled as

$$t(\tau) \approx \frac{(s * \tau)^n}{\text{CPU}_f}, \quad (18)$$

where s is the sampling rate in samples per second, $n \in \mathbb{N}$ is the algorithmic complexity, and CPU_f is the number of operations per second of the CPU used to process the radar data. We can now obtain the total detection time by adding equations (17) and (18) to get

$$f_2(\tau) = t_D = (N - 1)T + \left(\frac{16P_t G^2 \lambda^2 \sigma \tau}{(4\pi)^3 k T_e F L (\text{SNR}) c^4} \right)^{\frac{1}{4}} + \tau + \frac{(s * \tau)^n}{\text{CPU}_f}. \quad (19)$$

This gives a functional relationship between the pulse width and the total detection time, t_D . With both the functions relating the sensor parameter to the performance metrics and the corresponding LBCFs, the loss function can be defined as

$$L(\tau) = C_1^p(f_1(\tau)) + C_2^p(f_2(\tau)). \quad (20)$$

The parameter values used for this example are: peak power $P_t = 100$ kW, operating frequency $f_0 = 5.6$ GHz, antenna gain $G = 20$ dB, radar losses $L = 6$ dB, noise figure $F = 3$ dB, target RCS $\sigma = 1$ m², SNR = 15 dB, CPU frequency $CPU_f = 3$ GHz, sampling rate $s = 10$ MHz, number of pulses $N = 1$, and an algorithm complexity of $O(n)$. The loss vs. pulse width with these parameter values is shown in Figure 6. Using equation (1) we can find the utility of the system as a function of pulse width. Figure 7 shows the utility as a function of τ . Figure 7 is an information overload curve due to τ 's linear relationship with information. This curve has minimum loss/maximum utility at a pulse width of 24.4 ms. At this pulse width, the system should have a maximum detection range of 76.32 km and a total detection time of 25 ms which satisfies the requirement and both goals.

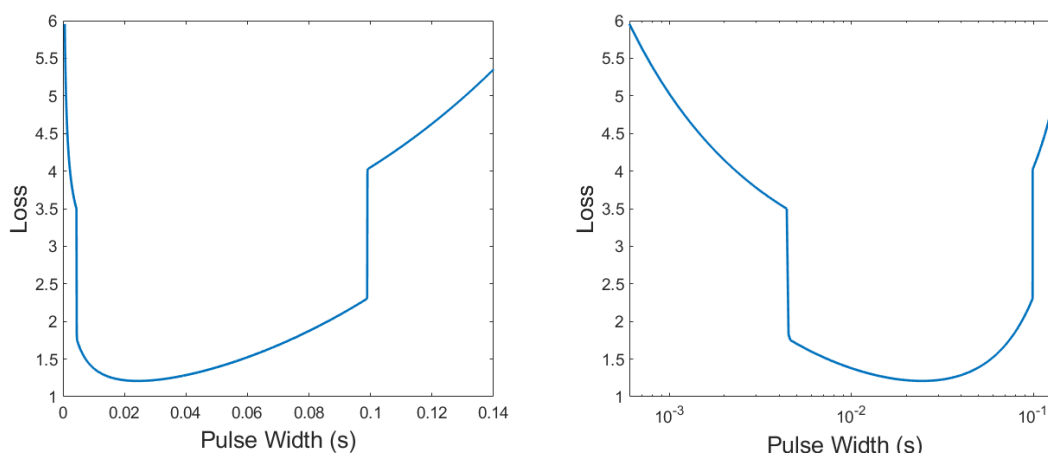


Figure 6: (Left) the pulse width vs the corresponding loss due to the constraints given to the performance metrics and (right) the same plot with the pulse width in log scale.

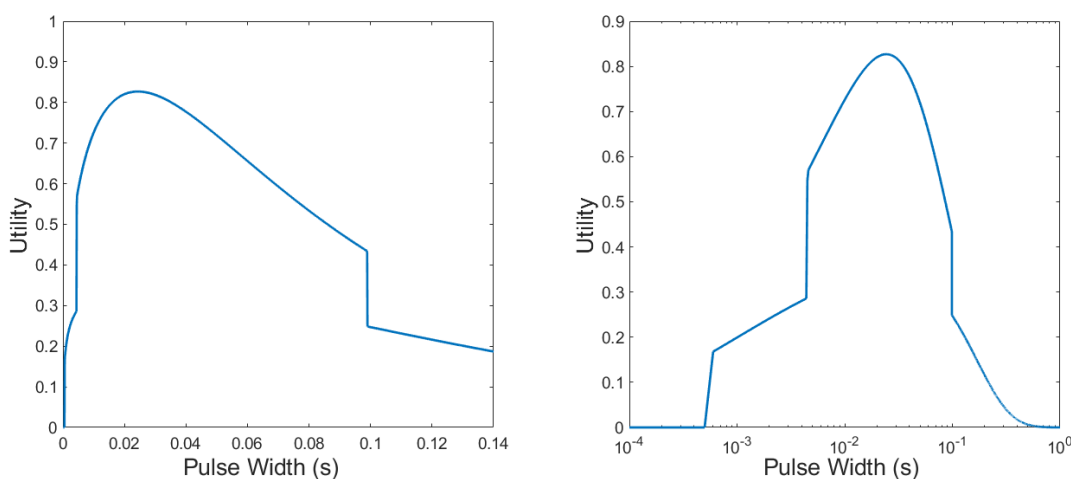


Figure 7: (Left) the pulse width vs the corresponding utility due to the constraints given to the performance metrics and (right) the same plot with the pulse width in log scale.

If instead, the parameter selection strategy was to select a parameter value that fulfilled the most goals, the solution would be a range of pulse widths from 2.4 ms to 98.9 ms as all values of the pulse width that fall in this range fulfill both goals. By making use of extra implied context, the optimal solution could be narrowed down to just one value. Beyond this critical/optimal value, the system will experience information overload where added information will reduce the overall utility of the system.

6. CONCLUSION

In conclusion, a framework called LBCFs for modeling hard and soft constraining statements was introduced. The soft constraining statements used sigmoidal functions to represent cost jumps for violating constraints. The hard constraints utilized logit-like functions that would jump to an arbitrarily high value when the hard constraint was violated. The functions also modeled the directional context with the functions' output representing whether an increase in the parameter value was beneficial or detrimental. A framework for relating parameters to one another via an importance term was proposed which allowed the final loss function to be a simple unweighted summation of the individual LBCFs. The LBCF framework was then used on an example radar system to illustrate information overload. In the example, constraints were placed on maximum detection range and total detection time, both performance metrics. LBCFs representing the constraints were generated. The only variable parameter was the pulse width which was noted to have a linear relationship with the amount of information in the system. The equations relating the pulse width to the performance metrics were derived and used in the objective function. The optimal pulse width was found and this point was noted to be the information overload point. This point represents when extra added information starts to negatively impact the utility of the radar system.

ACKNOWLEDGMENTS

This work was supported by the US Air Force Office of Scientific Research Grant Number FA9550-17-1-0032 (POC: R.D. Riecken).

REFERENCES

1. R. Narayanan, A. Liu, P. Singerman, and M. Rangaswamy, "Information elasticity in radar systems," *Electronics Letters* **54**(17), pp. 1049–1051, 2018.
2. P. G. Singerman, R. M. Narayanan, and M. Rangaswamy, "Information elasticity in ultra-wideband target detection amongst distributed clutter," in *Radar Sensor Technology XXII*, **10633**, p. 106330G, International Society for Optics and Photonics, 2018.
3. S. Haykin, "Cognitive radar: a way of the future," *IEEE Signal Processing Magazine* **23**(1), pp. 30–40, 2006.
4. K. L. Bell, C. J. Baker, G. E. Smith, J. T. Johnson, and M. Rangaswamy, "Cognitive radar framework for target detection and tracking," *IEEE Journal of Selected Topics in Signal Processing* **9**(8), pp. 1427–1439, 2015.
5. C. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, Springer Science & Business Media, 2007.
6. K. Deb, "Multi-objective optimization," in *Search Methodologies*, pp. 403–449, Springer, 2014.
7. C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
8. M. Fazel and M. Chiang, "Network utility maximization with nonconcave utilities using sum-of-squares method," in *Proceedings of the 44th IEEE Conference on Decision and Control*, pp. 1867–1874, IEEE, 2005.
9. M. Friedman and L. J. Savage, "The utility analysis of choices involving risk," *Journal of Political Economy* **56**(4), pp. 279–304, 1948.
10. J. Nocedal and S. Wright, *Numerical Optimization*, Springer Science & Business Media, 2006.
11. M. Udell and S. Boyd, "Maximizing a sum of sigmoids." http://stanford.edu/~boyd/papers/max_sum_sigmoids.html, 2013.

Multi-Objective Fully-Adaptive Radar Utilizing Language-Based Cost Functions

Paul G. Singerman*, Sean M. O'Rourke[†], Ram M. Narayanan* and Muralidhar Rangaswamy[†]

*School of Electrical Engineering and Computer Science

The Pennsylvania State University, University Park, PA 16802

Email: pgs16@psu.edu and rmn12@psu.edu

[†]Air Force Research Laboratory

Sensors Directorate, WPAFB, OH 45433

Email: sean.orourke.3@us.af.mil and Muralidhar.Rangaswamy@wpafb.af.mil

Abstract—There is a push to create radar systems that optimize their own performance in real time. In order to accomplish this, the radar system's operating context must be known. Often, this context is given in natural language statements about the goals and requirements of the radar system by individuals in charge of the “big picture.” In this paper, we propose a method to optimize the performance of a fully adaptive radar (FAR) target tracker when there are multiple performance metrics and multiple tunable sensor parameters. We use a previously developed method of encoding context known as language-based cost functions (LBCFs). Unfortunately, in adaptive target tracking, processing time is severely constrained and the final objective function is often non-convex. Therefore, we develop a catch-all iterative algorithm utilizing a clever initialization scheme, followed by a local search to find relaxed, efficient solutions. The algorithm is then tested on a basic simulation of a human running. Results show that the combination of LBCFs and the optimization algorithm allow the radar to far outperform a classical non-adaptive radar. Results also show that the specific initialization scheme used vastly changes the radar's ability to find satisfactory solutions which in turn can dramatically alter the tracker's performance.

Index Terms—Multi-objective optimization; non-convex optimization; fully adaptive radar; target tracking

I. INTRODUCTION

Imagine an individual is operating a radar system that is performing some task (e.g. target detection) and the radar's parameters, such as power, bandwidth, etc., were calibrated for the specific case of a target at 4-km range under optimal weather conditions. If the scenario is as specified, the radar should perform optimally according to theory. Often, these exact specifications are not true. The target may be closer or farther than 4 km which means more or less power could be used to get the same performance. Environmental conditions are also likely to change. This means the radar's parameter set could become inadequate if originally calibrated for clear conditions or might become excessive and resource wasteful if originally calibrated for worst-case conditions. Unfortunately, it is not very feasible to have a radar engineer on hand to manually change the operating parameters of the radar system every time the target moves or the weather changes. Therefore, to solve this problem, we need a radar that is able to sense the target and environment and change its parameters itself.

In the past, a radar system would be made with a fixed set of parameters, which were custom selected based on existing theory to perform optimally in a specific operating context [1]. With recent increases in computer processing power and more advanced hardware, radar systems with a fixed parameter set are on their way out. Instead, as pioneered by Simon Haykin [2], radar systems are starting to be able to intelligently adapt their parameters in response to both operational context and the external environment. With this increased flexibility, the performance of the radar can be ensured while at the same time minimizing the resources used [3], [4], [5].

A radar's operational context fully specifies how a radar is expected to perform. This context is often given in the form of goals and requirements that the system is expected to meet. For instance, parameters that perform well for target radial velocity estimation likely will be inadequate when given the task of target range estimation. Therefore, to determine what parameters are optimal, the goals and requirements or operational context for a radar system must be specified.

To accomplish this, we require a method of encoding this operational context for the radar. This can be easy in the case of a single performance metric as one may only be interested in minimizing or maximizing the metric. The far more difficult scenario comes when there are multiple competing performance metrics. It is often not possible to put each performance metric at its individually optimal value which means some trade-offs must take place.

In this paper, we explore fully adaptive radar (FAR) target tracking with multiple objectives. To deal with the multiple, often competing objectives, we use a previously developed method known as language-based cost functions (LBCFs) [6] to encode the goals and requirements. These are developed to allow the objective function to be a simple unweighted summation of the individual LBCFs. Once obtained, the objective function can be minimized where the solution is the best set of radar parameters (for the given operational context) for the next measurement.

In Section II, the radar target tracking model is introduced. In Section III, LBCFs are explained. In Section IV, LBCFs are shown to be a solution to an optimization issue arising from hard constraints on functions of parameters, and an op-

timization algorithm for handling non-convex functions while under severe time constraints is proposed. In Section V, the functional relationships, parameters, and performance metrics are found. In Section VI, the simulation is presented, and in Section VII, the results are shown and discussed.

II. FAR TARGET TRACKING MODEL

The state space model being employed is:

$$\mathbf{x}_k = \mathbf{A}(\boldsymbol{\theta}_k)\mathbf{x}_{k-1} + \mathbf{w}_k(\boldsymbol{\theta}_k) \quad (1)$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{v}_k(\boldsymbol{\theta}_k, \mathbf{x}). \quad (2)$$

The state vector, $\mathbf{x}_k = [r, \dot{r}]^\top$, is the target's state at time k where r is the radial distance to the target and $\dot{r}$ is the range rate of the target. The measurement vector, $\mathbf{y}_k = [r, f_D]^\top$ consists of the natural parameters measured by the radar, where f_D is the Doppler-frequency of the received signal. The tunable sensor parameters $\boldsymbol{\theta}_k = [f_{\text{PRF}}, N, B]^\top$ are the parameters values that can change in real time where f_{PRF} is the pulse repetition frequency, N is the number of pulses coherently integrated for one measurement, and B is the signal bandwidth. The state transition matrix,

$$\mathbf{A}(\boldsymbol{\theta}_k) = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \quad (3)$$

propagates the states forward in time where $\Delta t = N/f_{\text{PRF}}$ is the time between state estimations. The measurement, or output, matrix

$$\mathbf{C} = \begin{bmatrix} 1 & 0 \\ 0 & -\frac{f_c}{c} \end{bmatrix} \quad (4)$$

relates the states to the measurements where f_c is the center frequency of the radar waveform used for measurements, and c is the speed of light in air. Both the noise terms are zero-mean Gaussian random variables, with the process noise $\mathbf{w}_k(\boldsymbol{\theta}) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{ww}(\boldsymbol{\theta}))$ and the measurement noise $\mathbf{v}_k(\boldsymbol{\theta}, \mathbf{x}) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{vv}(\boldsymbol{\theta}, \mathbf{x}))$.

The process noise was modeled as nearly constant velocity (NCV) with discrete white noise acceleration (DWNA) [1]. This gives a process covariance of

$$\mathbf{R}_{ww} = \beta \begin{bmatrix} \Delta t^4/4 & \Delta t^3/2 \\ \Delta t^3/2 & \Delta t^2 \end{bmatrix}. \quad (5)$$

The design parameter for this process noise model is β . This is typically set such that

$$\frac{1}{2}a_m < \beta < a_m \quad (6)$$

where a_m is the maximum acceleration of the target.

The measurement noise was modeled by finding the individual CRLBs for the range [7] and the range rate [1] separately and assuming they are independent of each other. That is

$$\mathbf{R}_{vv} = \begin{bmatrix} \alpha \frac{R^4}{BN} & 0 \\ 0 & \gamma \frac{f_{\text{PRF}}^2 R^4 B}{N^3} \end{bmatrix}. \quad (7)$$

where

$$\alpha = \frac{24\pi f_c^2 k T_0 F}{P_t G^2 \sigma} \quad (8)$$

and

$$\gamma = \frac{192\pi f_c^4 k T_0 F}{c^4 P_t G^2 \sigma} \quad (9)$$

Here, R is the target's range, k is Boltzmann's constant, T_0 is the noise temperature, F is the noise figure of the radar system, P_t is the transmitted power, G is the gain of the antenna, and σ is the radar cross section (RCS) of the target.

III. LANGUAGE-BASED COST FUNCTIONS

To aid in designing a radar system there are often *goals* and *requirements* placed on both the performance metrics and sensor parameters. *Requirements* must be met and are crucial to the performance of the radar system. On the other hand, meeting *goals* is not absolutely essential but a system that does so will have better performance compared to one that does not, given that all requirements are met by both systems. Goals are typically more difficult to achieve than requirements for this reason.

When designing a system, there is often an implied directional context on each parameter and performance metric. For example, a system that could achieve the same performance goals as another system while using less power would be a better system. This implies that, given all goals and requirements are met, the lower the power the better. This means power has a directional context where zero would be the optimal value. Of course, zero will likely never be a solution as it would not give a system that meets all the other performance goals.

In [6], we developed quasi-convex LBCFs that can take, as input, the goals, requirements, and directional context of a system and output a single objective function. Our LBCF concept is a more rigorous formalization of field-of-view penalization functions for sensor adaptation in [8], [9]. This act of reducing a multi-objective optimization (MOO) problem to a single-objective optimization problem is known as scalarization [10]. By doing this, we output a single value for the performance of the system at different sensor parameter values. This was done using sigmoid functions to represent the goals and barrier functions to represent the requirements, while also including normalized exponential to represent the directional context. An importance term was also introduced that allowed the goals to have different levels of "importance." More important goals, when not achieved, would give more cost compared to less important goals. Thus, the system that gives the lowest cost gives the best performance while the system with the highest cost would give the worst performance.

The objective function obtained from the LBCFs is given as

$$L(\boldsymbol{\theta}) = \sum_{i=1}^k (C_i^s(\theta_i)) + \sum_{i=1}^m (C_i^p(f_i(\boldsymbol{\theta}))) \quad (10)$$

where $\boldsymbol{\theta} = \{\theta_1, \theta_2, \dots, \theta_n\}$ is the set of tunable sensor parameters, k is the number of constrained, tunable sensor parameters,

m is the number of constrained performance metrics, $C_i^s(\cdot)$ is the LBCF for the i th parameter θ_i , $C_i^p(\cdot)$ is the LBCF for the i th performance metric, and $f_i(\cdot)$ is the function representing the relationship between the sensor parameters and the i th performance metric. The optimal parameters are then found via an unconstrained optimization problem over the sensor parameters, namely,

$$\hat{\theta} = \arg \min_{\theta} L(\theta). \quad (11)$$

Once the objective function is acquired an optimization algorithm can be used to find a solution that minimizes it. This solution will be a vector of values for the tunable sensor parameters that gives the radar system the lowest cost and therefore the highest performance.

IV. OPTIMIZATION ALGORITHM

A. Solution to a Hard Constraint Problem via LBCFs

In general, the solutions to any optimization problem must be feasible; that is, they must meet all hard constraints exactly. Consequently, as a parameter approaches infeasibility (that is, its related hard constraint), the objective function will sharply grow towards infinity, ensuring that infeasible parameters will not be selected. When hard constraints are placed directly on the sensor parameters, this is generally not an issue because the set of allowable sensor parameters Θ_s is straightforward.

Since the performance metrics are functionally related to the sensor parameters θ , hard constraints on these metrics induce their own set of allowable sensor parameters Θ_p in the parameter space. This then produces a total set of allowable parameters Θ , which is the intersection of the direct parameter constraints and the indirect constraints via the performance metrics. That is,

$$\Theta = \Theta_s \cap \Theta_p \quad (12)$$

which is illustrated in Figure 1.

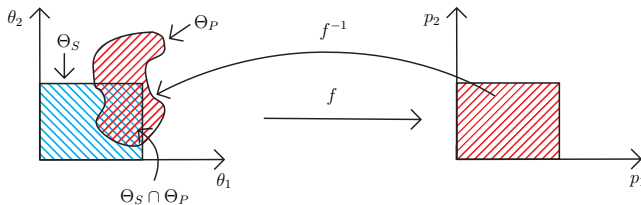


Fig. 1. Illustration of direct hard constraints on the space of performance metrics (right/red) and direct hard constraints on the space of tunable sensor parameters (left/blue) where the hard constraints on the performance metrics cause indirect hard constraints on the space of tunable sensor parameters (left/red).

However, an issue arises when there are hard constraints on “complex functions” of the parameters. The issue comes from trying to find exactly what the set Θ_p contains. This would require we know the inverse mapping f^{-1} which relates the points in the performance metric space to points in the parameter space. That is, if we have $p = \{p_1, \dots, p_m\}$ as a

point in the m dimensional performance metric space then the inverse mapping of this would be

$$f^{-1}(p) = \{\theta : f(\theta) = p\} \quad (13)$$

which is the set of values θ such that $f(\theta) = p$.

If there are many sensor parameters, many performance metrics, or the functional relationships between the sensor parameters and the performance metrics are too complex, Θ_p and, by extension, Θ may be impractical to find. Furthermore if, the optimization routine is initialized by a parameter vector $\theta_k^{\text{int}} \notin \Theta$ then the algorithm, which relies on information about the function around the point, will be unable to gather any information. This means the algorithm either gets stuck at the initial point or just takes ignorant steps relying on luck to find a point inside the feasible set.

Just a single point within Θ is needed to start the optimization algorithm. Once this is obtained, the optimization algorithm can gather information about the function and make an informed step towards, at least, a local minimum.

One simple, but problematic, solution is to perform a grid search over Θ_s in search of a point which has a finite cost. However, if θ is high-dimensional, the grid search would suffer from the curse of dimensionality [11], a phenomenon caused by the exponential increase in the volume of the parameter space as its dimension is increased. Thus, a grid search quickly becomes impractical if time is in any way a factor.

LBCFs provide a ready-made solution for this problem, particularly by providing a mechanism for soft constraints with finite cost. The issue with having an infinite cost for constraints is that the optimization algorithm must be initialized in Θ . If the constraints on the performance metrics used a large but not infinite cost, then the algorithm would be able to start anywhere within Θ_s and could proceed to find at least a local minimum. Since LBCFs also incorporate directional context, this would hopefully help the optimization to “roll downhill” as the directional context will give information about the direction of the minimum. Therefore, the algorithm will be able to approach an optimal solution $\hat{\theta}$ that puts the performance metrics within their constraints.

B. Normal Random Perturbation Initialization for Non-convex Optimization within Time Constraints

When performing FAR target tracking, the controller attempts to choose optimal parameter values in between each measurement [4]. Therefore, the optimization algorithm cannot take longer than the time it has between measurements. At the same time, we do not want to be limited only to performance metrics and/or parameters that have convex relationships between them. Therefore, we want an algorithm that is robust enough to tune parameters even when the accompanying functional relationships are non-convex. Pseudo-global optimization algorithms do exist and can be used for validating theory [10] but they are plagued by their unpredictable convergence speed, which is unacceptable for a radar system that is updating parameter values between measurements.

This problem then simplifies to that of quickly finding optimal or nearly-optimal solutions while leveraging LBCFs to represent hard constraints as “large” soft constraints. A possible solution would be to cleverly choose the initialization value θ_k^{int} in such a way that the algorithm, while it may not find a satisfactory solution in one time step, is expected to approach an acceptable solution in several time steps.

Now, assuming we only have time between measurements to perform one local search, a way to choose an initialization value would be to introduce a distribution over the space of tunable sensor parameters in which the value could be sampled. The important step here, is how we parameterize the distribution. The goal is to choose the distribution’s parameters in such a way as to make the distribution encode information about the last local search such that it can use this information to improve the next local search. This will hopefully allow the optimization algorithm to escape any inadequate local minimums it gets stuck in.

We present a novel method that accomplishes this and denote it as the normal random perturbation (NRP) scheme. This is where the initialization value for time step k is selected from a normal distribution with mean $\hat{\theta}_{k-1}$ and variance $f(L(\hat{\theta}_{k-1}))\hat{\theta}_{k-1}$. That is

$$\theta_k^{\text{int}} \sim \mathcal{N}(\hat{\theta}_{k-1}, f(L(\hat{\theta}_{k-1}))\text{diag}(\hat{\theta}_{k-1})) \quad (14)$$

where $f(\cdot)$ is a function that changes the variance based on the cost of the previous locally optimal value $\hat{\theta}_{k-1}$. Simply put, a higher cost means that the optimization algorithm should search more aggressively for a better local minimum. once a solution with a low cost is found, the search can focus on fine tuning the parameters within the low cost parameter region. This can allow for a wider and more exploratory search when constraints are not met in the previous time step.

The $\text{diag}(\hat{\theta}_{k-1})$ term in the variance is included, instead of an identity matrix, because some of the sensor parameters may have many orders of magnitude between their lower bound and upper bound. Therefore, if the variance was not dependent on the scale of the parameters, the algorithm could get stuck in either the higher or lower regions with very low probability of transitioning.

V. FUNCTIONAL RELATIONSHIPS

The posterior conditional Cramér–Rao lower bound (PC-CRLB) is used as the functional relationship between the tunable sensor parameters and the variance of the range and range rate estimates. This is the inverse of the Bayesian information matrix (BIM) [12]. The BIM is given as

$$B = \left[A(\theta) \tilde{P} A(\theta)^\top + R_{ww}(\theta) \right]^{-1} + R_{vv}^{-1}(\theta, \hat{x}), \quad (15)$$

where $\tilde{P}$ is the filtered covariance at time step $k-1$ and $\hat{x}$ is the expected value of x at time k . This is simply the sum of the Fisher information from the prior distribution and expected Fisher information of the next measurement.

The variances of the range and range rate estimates are the diagonal entries of the PC-CRLB.

TABLE I
FAR TARGET TRACKING ALGORITHM

<u>Initialization</u>
1) Set $\hat{x}(k-1 k-1) = \hat{x}_0$.
2) Set $\tilde{P}(k-1 k-1) = \tilde{P}_0$.
<u>Parameter Optimization</u>
3) $B(k k-1) = \left[A(\theta) \tilde{P}(k-1 k-1) A(\theta)^\top + R_{ww}(\theta) \right]^{-1}$ $+ C R_{vv}^{-1}(\theta, \hat{x}) C^\top$
4) $\hat{\theta}_k = \arg \min_{\theta} L(\theta, B_{1,1}^{-1}(k k-1), B_{2,2}^{-1}(k k-1))$
<u>Prediction</u>
5) $\hat{x}(k k-1) = A(\hat{\theta}_k) \hat{x}(k-1 k-1)$
6) $\tilde{P}(k k-1) = A(\hat{\theta}_k) \tilde{P}(k-1 k-1) A^\top(\hat{\theta}_k)$
<u>Measurement</u>
7) Using the sensor parameters $\hat{\theta}_k$, measure y_k .
<u>State Estimation</u>
8) $K = \left(\tilde{P}(k k-1) C^\top \right) \left(C \tilde{P}(k k-1) C^\top + R_{vv}(\theta, \hat{x}) \right)^{-1}$
9) $\hat{x}(k k) = \hat{x}(k k-1) + K [y_k - C \hat{x}(k k-1)]$
10) $\tilde{P}(k k) = (I - KC) \tilde{P}(k k-1)$
<u>Recursion</u>
11) $\hat{x}(k k) \rightarrow \hat{x}(k-1 k-1)$
12) $\tilde{P}(k k) \rightarrow \tilde{P}(k-1 k-1)$
13) Continue to step 3.

The objective function L generated from applying LBCFs to the diagonal entries of the PC-CRLB can be non-convex and riddled with local minimum depending on the tunable sensor parameters. This is the reason we developed the more general optimization algorithm in Section V.

VI. SIMULATION

A first-order nonlinear system was used to simulate a moving target. This is expressed as

$$r = 2.5 \cos(t) + 3.5 \quad (16)$$

$$\dot{r} = 2.5 \sin(t); \quad (17)$$

where the model was taken from [5] and vaguely mimics a human running.

Three statements on three different performance metrics were given:

- 1) The standard deviation on the range estimate must be less than 2×10^{-3} meters.
- 2) The standard deviation on the range rate estimate must be less than 2×10^{-3} meters per second.
- 3) The time between estimations must be less than 60 milliseconds.

Here, the verb “must” in all three statements indicates hard constraints. As noted in Section IV, we instead use soft constraints with importance terms, $I = 1000$, in lieu of hard constraints. This indicates that the requirements are 1000 times more important than the least important goal. Hard constraints on the parameters were also introduced to simulate a real radar system that would not, for example, have access to 200% fractional bandwidth or be capable of an infinite pulse

repetition frequency. These constraints were the same as in [3], except the minimum bandwidth was lowered to 1000 Hz. With the constraints on the performance metrics and sensor parameters now specified, the operational context of the radar system is fully expressed. LBCFs are then generated from the statements.

A standard Kalman filter was then implemented in Matlab for target tracking, the state-space model for which was shown in Section II. To initialize the Kalman filter, priors for the state estimates and the covariance matrix were chosen. These were

$$\hat{x}_0 = \begin{bmatrix} 3.5 \\ 0 \end{bmatrix} \quad \text{and} \quad \tilde{P}_0 = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} \quad (18)$$

where $\hat{x}$ is the state estimate and $\tilde{P}_0$ is the covariance. We also initialized the recursion by choosing a set of parameter values. The starting parameter values were PRF = 5 kHz, $N = 500$ pulses, and $B = 500$ MHz. The FAR algorithm was carried out as shown in Table 1.

VII. RESULTS

The importance of the initialization distribution's variance being a function of the previous cost is shown in Figures 2 and 3. Figure 2 shows that with the perturbation being independent of the cost, the optimization routine can get stuck in a local minimum which is outside a constraint (the standard deviation of the range estimate). Figure 3 shows the full NRP scheme where variance is a function of the last $L(\hat{\theta}_{k-1})$. This shows that a solution exists that has the performance metrics within the constraints and that the perturbation variance being dependent on $L(\hat{\theta}_{k-1})$ finds this solution.

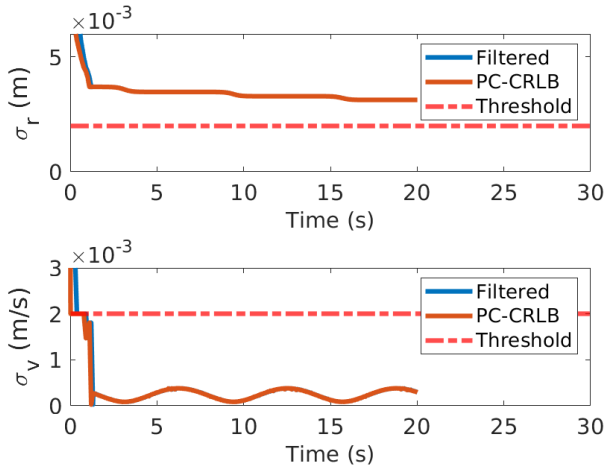


Fig. 2. The predicted and filtered range and range rate standard deviations over time for the NRP scheme without letting the perturbation variance depend on the previous cost. The red dashed line is the hard constraint.

Figure 4 shows the parameters changing with time for the NRP scheme. In this case, both the PRF and bandwidth are maximized and minimized, respectively, while the number of coherently integrated pulses varies from 263 to 64 after convergence. This shows that the radar is adapting. Since the

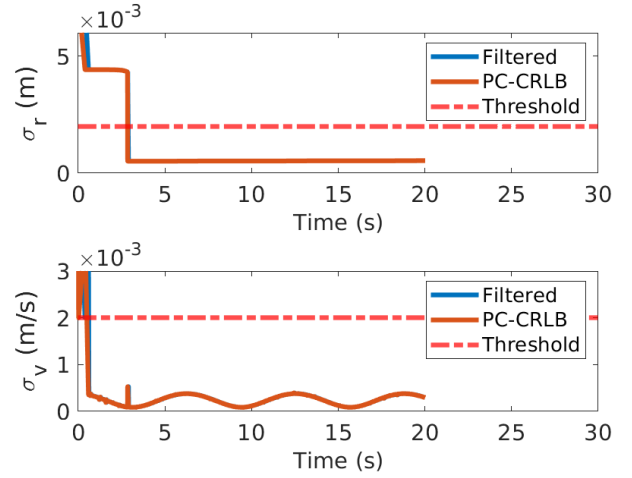


Fig. 3. The predicted and filtered range and range rate standard deviations over time for the NRP scheme letting the perturbation variance depend linearly on the previous cost. The red dashed line is the hard constraint.

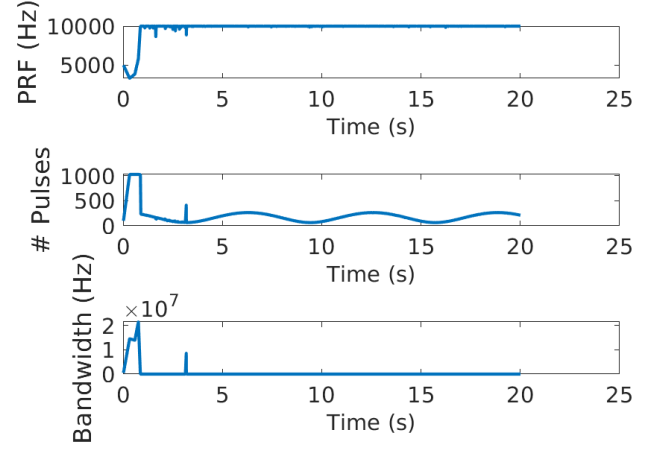


Fig. 4. The sensor parameter values over time when the NRP scheme is used.

radar is only changing the number of pulses it coherently integrates before making a decision the amount of data that the radar processes will change with time. This is related to our previous work in information elasticity where the optimal amount of information a system uses is shown to be finite [13]. As the target moves away from the radar, the SNR decreases which in turn increases the measurement variance. The controller adapts to this and increases the number of integrated pulses to get more accurate measurements. The controller deemed the SNR increase from the extra pulses was worth the trade off of more process noise due to a greater amount of time between measurements.

We can compare the tracks of each optimization algorithm to the non adaptive case by examining the mean squared error (MSE) of the filters estimate compared to the ground truth. Figures 5 and 6 compare the NRP initialization scheme to three other initialization schemes as well as the static parameter

case.

The three other initialization schemes were:

- 1) Previously optimal (PO) – the next initialization point is set equal to the previous optimal point $\hat{\theta}_{k-1}$,
- 2) Uniform random (UR) – the initial point is sampled from a uniform distribution over the set of allowed sensor parameter values Θ_s ,
- 3) Uniform Random Perturbation (URP) – the initial point is sampled from a uniform distribution centered at $\hat{\theta}_{k-1}$ which is scaled by the cost but still inside Θ_s .

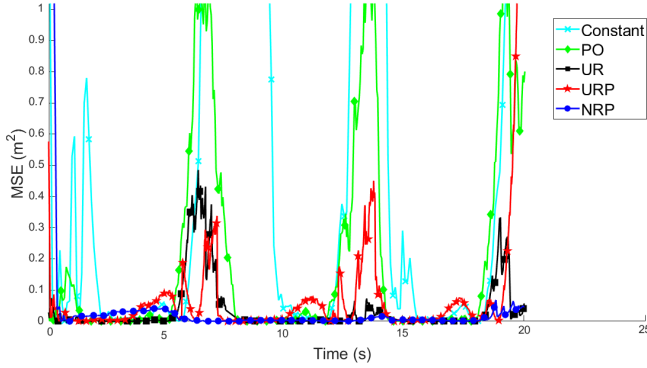


Fig. 5. The MSE of the range estimations from the ground truth over time for the constant parameter case and the four different adaptive algorithms.

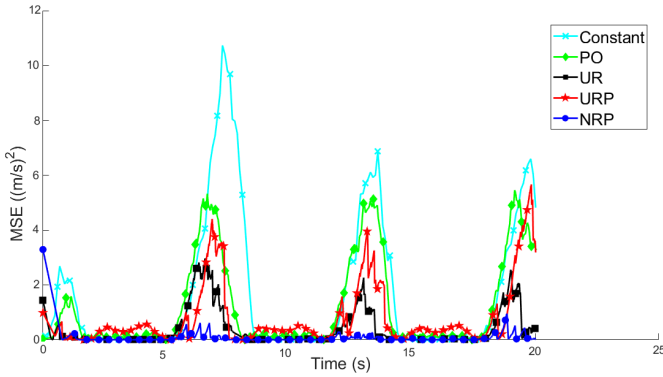


Fig. 6. The MSE of the range rate estimations from the ground truth over time for the constant parameter case and the four different adaptive algorithms.

As we can see from Figures 5 and 6, all four different initialization schemes performed, for the most part, uniformly better than the non-adaptive case. This indicates that even if an adaptation algorithm is not do enough exploration to always find satisfactory solutions it still outperforms the non-adaptive case. The NRP scheme achieves a much lower MSE on both the range and range rate tracks than any of the other initialization methods. This is because it does not suffer from some of the downfalls the other initialization schemes do. For example, the POI schemes gets stuck in a local minimum because it does not explore the parameter space looking for better solutions. Meanwhile, the URI and URWP schemes both jump rapidly in the parameter space because they only do large

searches and do not convert to finer searches once a sufficient solution has been found. This leads us to conclude that the POI does not explore the parameter space enough, while the URI and URWP schemes explore *too much*. The NRP scheme takes the best of both and focuses on a volume of the parameter space where satisfactory solutions are found.

VIII. CONCLUSION

In this paper, we showed that LBCFs are a simple and practical tool to encode operational context for FAR target tracker when there are multiple competing objectives. A novel iterative initialization method was developed that allows for relaxed solutions to non-convex optimization problems under severe time constraints. A radially moving target was simulated and four different initialization methods were tested as well as the classical static parameter case for a radar performing target tracking. All four initialization methods for parameter adaptation were shown to outperform the static parameter case, indicating that parameter adaptation using LBCFs is beneficial even when the optimization algorithm is stuck in bad local minimums. The NRP initialization scheme was shown to be superior due to its ability to focus in on the parameter volume with satisfactory solutions.

ACKNOWLEDGMENT

This work was supported by the US Air Force Office of Scientific Research Grant Number FA9550-17-1-0032.

REFERENCES

- [1] M. A. Richards, J. Scheer, W. A. Holm, and W. L. Melvin, *Principles of modern radar*. SciTech Publishing, 2010.
- [2] S. Haykin, "Cognitive radar: a way of the future," *IEEE Signal Processing Magazine*, vol. 23, no. 1, pp. 30–40, 2006.
- [3] A. E. Mitchell, G. E. Smith, K. L. Bell, A. J. Duly, and M. Rangaswamy, "Cost function design for the fully adaptive radar framework," *IET Radar, Sonar & Navigation*, vol. 12, no. 12, pp. 1380–1389, 2018.
- [4] K. L. Bell, C. J. Baker, G. E. Smith, J. T. Johnson, and M. Rangaswamy, "Cognitive radar framework for target detection and tracking," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 8, pp. 1427–1439, 2015.
- [5] K. L. Bell, J. T. Johnson, G. E. Smith, C. J. Baker, and M. Rangaswamy, "Cognitive radar for target tracking using a software defined radar system," in *2015 IEEE Radar Conference (RadarCon)*, pp. 1394–1399, IEEE, 2015.
- [6] P. G. Singerman, R. M. Narayanan, and M. Rangaswamy, "Cost function design for modeling information overload in radar systems," in *Radar Sensor Technology XXIII*, vol. 11003, p. 110030A, International Society for Optics and Photonics, 2019.
- [7] S. M. Kay, *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.
- [8] S. M. O'Rourke and A. L. Swindlehurst, "Closed-loop tracking using multimodal rf/eo sensors," in *2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers*, pp. 1662–1666, IEEE, 2010.
- [9] S. M. O'Rourke and A. L. Swindlehurst, "Limited field-of-view multimodal sensor adaptation for data association," in *2012 IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 241–244, IEEE, 2012.
- [10] P. M. Pardalos, A. Žilinskas, and J. Žilinskas, *Non-convex multi-objective optimization*. Springer, 2017.
- [11] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.
- [12] H. L. Van Trees and K. L. Bell, *Bayesian bounds for parameter estimation and nonlinear filtering/tracking*. John Wiley, 2007.
- [13] R. Narayanan, A. Liu, P. Singerman, and M. Rangaswamy, "Information elasticity in radar systems," *Electronics Letters*, vol. 54, no. 17, pp. 1049–1051, 2018.

An Information Elasticity Framework for the Adaptive Matched Filter

Ram Narayanan, *Fellow, IEEE*, Andrew Liu, and Muralidhar Rangaswamy, *Fellow, IEEE*

Abstract—The adaptive matched filter (AMF) uses a number of training samples observed by the radar to estimate the unknown disturbance covariance matrix of a cell under test. In general, as the number of homogeneous training samples increases, the detection performance of the AMF improves up to an theoretical limit (defined by the performance of a matched filter detector where the disturbance covariance is known). However, radar data are non-homogeneous in practice. Consequently, a high number of training samples is typically undesirable, since non-homogeneous training data cause detection performance to suffer. Thus, a decision maker (DM) must consider these trade-offs when selecting this number of training samples, along with other decision parameters for the AMF. Using the concept of information elasticity, this trade-off behavior is characterized for decisions that are relevant to a decision maker. A simple user defined constraint function is proposed, characterizing the relative cost of selecting different decisions. Using a multi-objective optimization (MOO) technique known as compromise programming, information overload is observed, in that increasing the cost of decisions improves performance up to a point, beyond which increasing the cost no longer provides meaningful benefit. Using this framework, a cost-efficient solution is selected.

I. INTRODUCTION

THE detection of targets in the presence of clutter, noise and other disturbances is a critical signal processing problem for many radar systems. The technique known as space-time adaptive processing (STAP) seeks to address this detection problem for radar systems using multiple spatio-temporal elements. STAP has been well studied [1], and a plethora of different approaches to this problem have been proposed. One such receiver, proposed by Robey et al. in [2], uses a test statistic which is derived using a generalized likelihood ratio test of Gaussian distributed data with a known disturbance covariance matrix. This “true” covariance matrix is then replaced with an estimated covariance matrix, yielding the test statistic for the adaptive matched filter (AMF).

In order to obtain the test statistic for the AMF, the estimated covariance matrix being used must be invertible. Thus, the covariance matrix estimation method must be capable of producing a non-singular estimate. When sample covariance estimation is used, the estimate is non-singular as long as $K > N$ [2], where K is the number of training/secondary samples and N is the spatio-temporal product/dimensionality of the detector. Note that the AMF test statistic is derived

using the assumption that these K secondary samples are independent and identically distributed as disturbance only [2], and are thus homogeneous. However, in practice, the training data are often non-homogeneous [3], since some of the training samples may have different disturbance statistics or may include target information. When non-homogeneous training samples are used, the detection performance is degraded, often by a considerable amount [3].

In general, the detection performance increases as the number of homogeneous training samples increases [2]. Although only $K = N$ samples are required to obtain a test statistic for the AMF, the detection performance has been shown to be very poor when K is close to N . In fact, the work of Reed, Mallet, and Brennan analyzes the random loss in SNR when using the filter weights obtained through sample matrix inversion (SMI), also known as RMB loss. In order for the expectation of the RMB loss to be less than 3 dB, at least approximately $K = 2N$ training samples are required [4]. Furthermore, the AMF suffers an additional amount of SNR loss [2] due to the constant false alarm rate (CFAR) loss [5].

In certain cases, such as when the dimensionality N is considerably large, it is unrealistic that $K = N$ homogeneous training samples are available in an environment, let alone $K = 2N$ homogeneous samples. Thus, approaches to covariance estimation other than SMI have been proposed to obtain better performance in these data starved cases. Many of these approaches, such as the fast maximum likelihood (FML) estimator [6] or the rank constrained maximum likelihood (RCML) estimator [7], assume a certain rank structure on the true disturbance covariance matrix. The RCML estimation method assumes that this matrix can be decomposed into a covariance matrix of the noise, which is full rank, and a covariance of the clutter, which is rank deficient but has a known rank. The rank of this clutter covariance matrix can be found under certain conditions (land clutter from an airborne radar) using Brennan’s rule [8].

Whether RCML or SMI is used to obtain the AMF test statistic, the interference environment, disturbance statistics, and other system factors determine the amount of homogeneous training samples that is available for a detector [7]. In order to avoid non-homogeneous data among the training samples, the choice of K should take these factors into account. Given these constraints on K , a decision maker (DM) may consider that a larger K is more costly, since increasing K increases the likelihood that the secondary samples contain a non-homogeneity [9] [10].

In [11], a framework known as information elasticity is proposed, which characterizes both the benefit and cost of

This work was supported by the US Air Force Office of Scientific Research Contract Number FA9550-17-1-0032 (POC: R.D. Riecken).

R. M. Narayanan, A. Z. Liu, Department of Electrical Engineering, The Pennsylvania State University, State College, PA, USA, E-mail: (ram@engr.psu.edu). M. Rangaswamy, US Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, USA

using data/information that is relevant to a DM. This concept has been applied to different problems in radar signal processing [12], [13], [14], demonstrating that given the cost and benefit of relevant information, increasing the information quantity often results in a point of maximum benefit (also known as overload or tipping point). Implicitly, increasing the information quantity beyond this overload point degrades the system performance. In this paper, this framework is applied to the AMF (under either SMI or RCML) to determine the quantity of training samples that causes this overload point to occur. In order to do this, a measure of decision effectiveness for the detector is defined, based on defined measures of cost and benefit for K .

The organization of this paper is as follows. Section II provides background on the AMF and analyzes a measure of performance based on the additional SNR required for the AMF to achieve the same detection probability as the clairvoyant detector. This section provides a closed form approximation for this performance metric when using SMI, as well as a description of the empirical methods used to obtain this metric for RCML. Section III provides an overview of the information elasticity framework. In this section, a user-defined constraint function describing the cost of decisions is presented, and the trade-off behavior between this function and the performance metric given in Section II is analyzed. Section IV presents a discussion of experimental results obtained using the decision making method on SMI and RCML matrices.

II. PERFORMANCE METRIC FOR THE ADAPTIVE MATCHED FILTER

A. AMF Background

As discussed in Section I, the AMF considers N -dimensional data [2]. The “primary” data vector $\mathbf{x}$ is considered to follow one of the two following hypotheses:

$$H_0 : \mathbf{x} = \mathbf{d} \quad (1)$$

$$H_1 : \mathbf{x} = b \mathbf{s} + \mathbf{d} \quad (2)$$

where $\mathbf{d}$ represents the random disturbance vector, $\mathbf{s}$ represents the steering vector, and b represents a complex scalar attributed to the target’s reflectivity and the channel’s propagation effects [15]. Let $\mathcal{CN}(\mu, \Sigma)$ denote the complex Gaussian distribution with mean vector μ and covariance matrix Σ . For $\mathbf{d} \sim \mathcal{CN}(0, \Sigma)$, the likelihood ratio test using the maximum likelihood estimate for b yields the following [2], [15]:

$$\frac{|\mathbf{s}^H \Sigma^{-1} \mathbf{x}|^2}{\mathbf{s}^H \Sigma^{-1} \mathbf{s}} \underset{H_0}{\overset{H_1}{\gtrless}} \alpha_{CV} \quad (3)$$

where $(\cdot)^H$ denotes the Hermitian transpose, and α represents the user defined threshold used to set the probability of false alarm (P_{FA}).

Equation (3) represents the test statistic for a detector where the disturbance covariance matrix Σ is known. This detector is also known as the clairvoyant detector. As discussed in Section I, the disturbance covariance matrix is unknown in practice and must be estimated from training samples either by sample covariance estimation or RCML estimation. Let $\hat{\Sigma}$

represent this estimated disturbance covariance matrix. This estimate is substituted into (3) in place of Σ , yielding the AMF test statistic:

$$\frac{|\mathbf{s}^H \hat{\Sigma}^{-1} \mathbf{x}|^2}{\mathbf{s}^H \hat{\Sigma}^{-1} \mathbf{s}} \underset{H_0}{\overset{H_1}{\gtrless}} \alpha \quad (4)$$

B. Performance of AMF Under Sample Covariance Estimation

Let $\mathbf{x}_k$ represent the k^{th} training data vector. The sample covariance estimate is given as follows:

$$\hat{\Sigma} = \frac{1}{K} \sum_{k=1}^K \mathbf{x}_k \mathbf{x}_k^H \quad (5)$$

It is well known that the sample covariance matrix is the maximum likelihood estimate for Σ . This is true as long as the training samples are independent and identically distributed as disturbance only (H_0). As the maximum likelihood estimate, the sample covariance matrix is also consistent [16] in that it converges to Σ in probability as the sample size K grows to infinity. Thus, as K increases, the test statistic for the AMF (4) generally approaches the test statistic for the clairvoyant detector (3).

This convergence with regard to sample size also occurs in the P_D for these detectors. The exact formulations for P_D for these detectors are well known [2] and can be used to show that the detection probability (P_D) of the AMF asymptotically approaches that of the clairvoyant detector. In this paper, the performance of the AMF is characterized by how close it is operating to its asymptotic limit (clairvoyant performance). One way to characterize this is by considering the additional SNR required for the AMF to achieve the same P_D as the clairvoyant detector, as shown in Figure 1. This additional

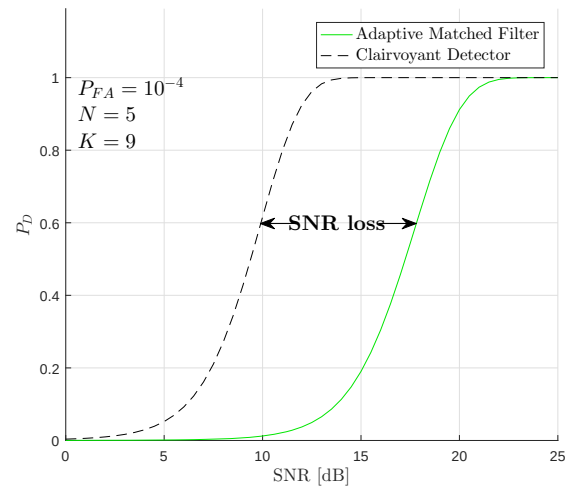


Fig. 1: P_D vs SNR curves shown for the clairvoyant detector and AMF for $P_{FA} = 10^{-4}$, $N = 5$, and $K = 9$. The additional SNR required for the AMF to achieve the same P_D as the clairvoyant detector is denoted by the arrows labeled “SNR loss”.

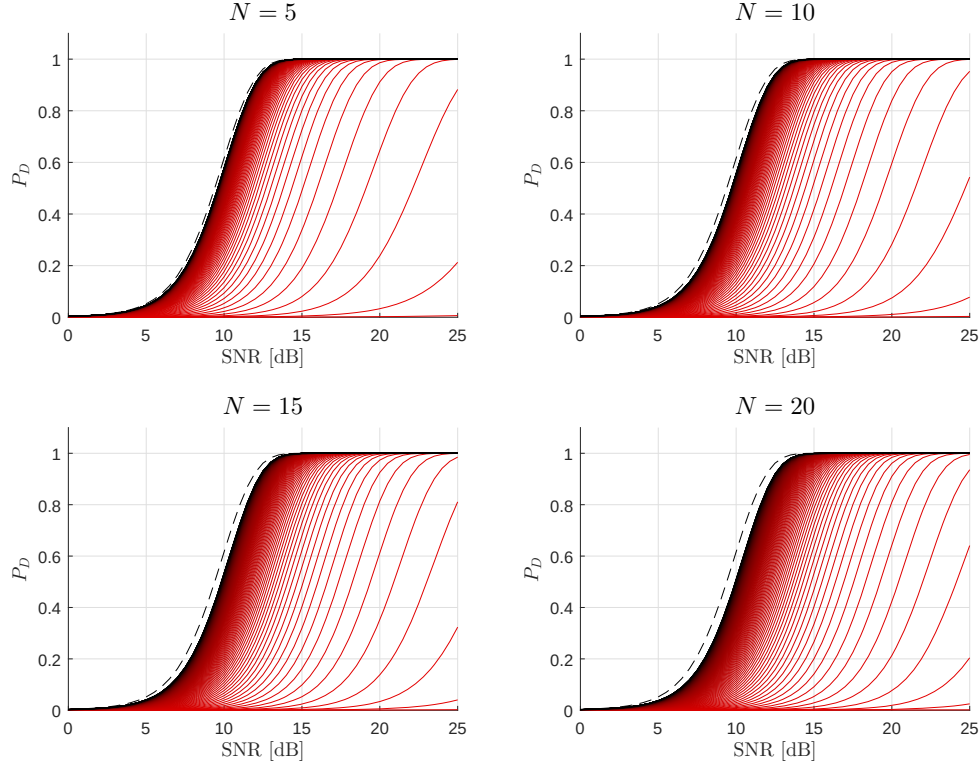


Fig. 2: P_D vs SNR at different N values for $P_{FA} = 10^{-4}$, and $K = \{N, \dots, 200\}$. The shade of the solid curves portrays the value of K used to generate it. Darker shades represent higher K values. The dashed curve represents the performance of the clairvoyant detector.

SNR required metric can also be thought of as a loss in SNR [5].

This SNR loss term can be found using the known equations for P_D and P_{FA} . The P_{FA} for the clairvoyant detector (denoted by $P_{FA,CV}$) is dependent on the threshold term α_{CV} given in (3) only. On the other hand, the P_{FA} for the AMF is dependent on α (given in (4)), N , and K . Their exact forms are given below [2]:

$$P_{FA,CV} = e^{-\alpha_{CV}} \quad (6)$$

$$P_{FA} = \frac{K!}{L!(N-2)!} \int_0^1 \frac{\rho^L (1-\rho)^{N-2}}{(1+\alpha\rho)^L} d\rho \quad (7)$$

where $L = K - N + 1$ and ρ is known as the random loss factor, which is statistically identical to the RMB loss. It should be noted that while ρ represents a random loss of the effective SNR, it is different than the SNR loss shown in Figure 1 [5], which shows the additional SNR required to reach clairvoyant detection probability.

Equations (6) and (7) are used to find the threshold value α required to achieve a prescribed P_{FA} value. For the clairvoyant case, this is relatively simple, since (6) can be rewritten as:

$$\alpha_{CV} = -\ln P_{FA,CV}$$

However for the AMF detector, (7) does not easily permit a closed form for α . Thus, α is evaluated using numerical methods. In this paper, numerical integration is used to solve

for P_{FA} for a given α , and line search methods are used to find the value of α that produces the prescribed P_{FA} .

The P_D for the clairvoyant detector (represented by $P_{D,CV}$) is dependent on the signal-to-noise ratio (SNR) and the α , while P_D for the AMF is dependent on the SNR, α , K , and N . Their exact forms are given below [2]:

$$P_{D,CV} = e^{-\text{SNR}} \sum_{m=0}^{\infty} \frac{\text{SNR}^m}{m!} \Gamma(\alpha_{CV}, m+1) \quad (8)$$

$$P_D = 1 - \left[\int_0^1 \frac{1}{(1+\alpha\rho)^L} \sum_{m=1}^L \binom{L}{m} (\alpha\rho)^m \Gamma\left(\frac{\text{SNR} \cdot \rho}{1+\alpha\rho}, m\right) \frac{K! \rho^L (1-\rho)^{N-2}}{L!(N-2)!} d\rho \right] \quad (9)$$

where $\Gamma(\cdot, \cdot)$ is the normalized upper incomplete Gamma function. Note that these detection probabilities are dependent on P_{FA} , by merit of α . Figure 2 compares $P_{D,CV}$ and P_D at different values of N and K . In these figures, the P_D vs SNR curves appear to shift left as K increases, converging toward the $P_{D,CV}$ vs SNR curve. As discussed, this convergence is characterized by the SNR loss, which approaches 0 as K increases.

This SNR loss can be found as follows. Equations (8) and (9) can be used to find the SNR values required to achieve a given P_D . Unfortunately, (8) and (9) do not easily permit

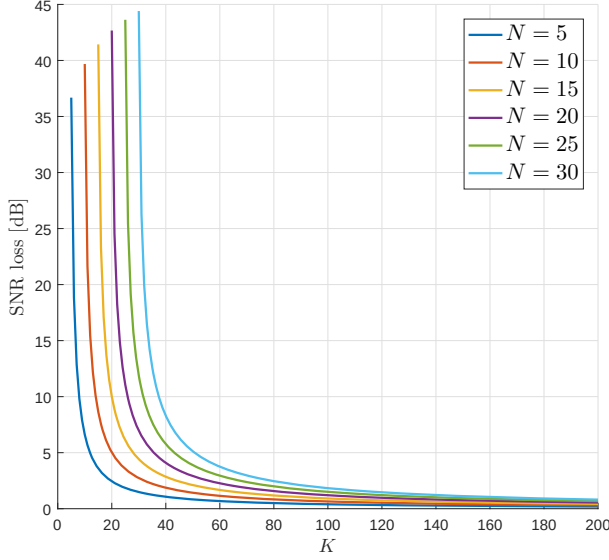


Fig. 3: Additional SNR required to achieve the same P_D and P_{FA} as the clairvoyant detector, as a function of K . $P_{FA} = 10^{-4}$ and $P_D = 0.6$.

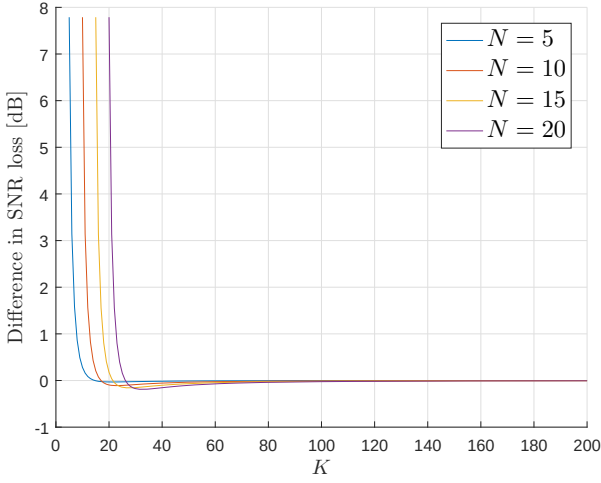


Fig. 4: Difference between SNR losses at $P_D = 0.9$ and $P_D = 0.1$ for $P_{FA} = 10^{-4}$.

a closed form for deriving the SNR. Furthermore, the P_D in (9) is typically obtained using numerical integration [2]. Thus, numerical integration and line search methods are used to find these SNR values. Once these values are obtained, the SNR loss is simply the ratio between these SNR values (the two are subtracted instead if decibels are used).

The SNR loss calculated using these numerical methods are shown in Figure 3 as a function of K . The curves all appear to converge towards 0. It is also shown that the rate of convergence differs depending on the dimensionality N . As a general trend, the SNR loss appears to converge more quickly for smaller N values. It should also be noted that, while these results are for $P_D = 0.6$, the SNR loss does

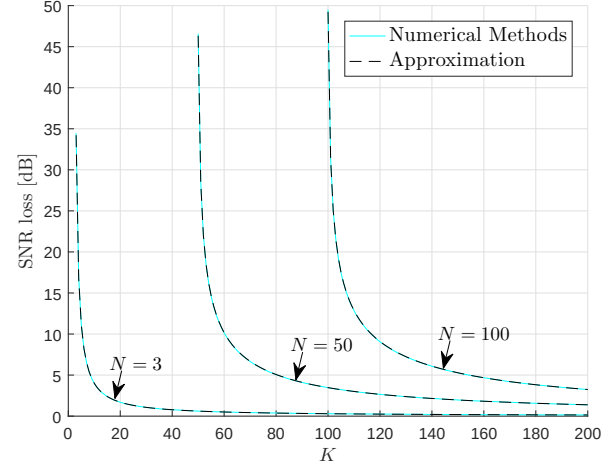


Fig. 5: SNR loss as a function of K . Calculations from numerical methods as well as approximation are displayed together, showing close agreement.

not change very much if it is recalculated for a different P_D value. This is because the P_D vs SNR curves of the AMF are nearly parallel with that of the clairvoyant detector. In fact, these curves become more parallel to the clairvoyant curve as K increases. This is demonstrated in Figure 4, where the difference in SNR loss at $P_D = 0.9$ and $P_D = 0.1$ is shown. Note that this difference is shown to be quite large for $K = N$. However, it is assumed that this degenerate case is not used in practice, since its associated SNR loss is so high.

While the SNR loss is able to characterize the performance of the AMF, the numerical methods involved in the process of finding the SNR loss are quite slow and cumbersome. Thus, this paper provides a closed form approximation for the SNR loss. This approximation is found using the equation for P_D when the Swerling I fluctuation model is being considered. However, the approximation is also shown to apply for the non-fluctuating case as well. The approximation is as follows:

$$\text{SNR loss} \approx \left(\alpha \frac{P_D^{1/L}}{1 - P_D^{1/L}} - \frac{K}{L} \right) \frac{\log(P_D)}{\log(\frac{P_{FA}}{P_D})} \quad (10)$$

Note that α is still calculated via (9) using line search methods, however, using this approximation reduces the computational costs substantially. The derivation for this approximation is provided in Appendix A.

C. Performance of AMF Under RCML

The work in [17] has shown that the AMF using RCML greatly outperforms the AMF using SMI, especially when the number of training samples is near N . Furthermore, [17] has also shown that increasing K also improves the detection performance of the RCML AMF detector. Thus, this paper observes the SNR loss between the AMF using RCML and the clairvoyant detector, and uses this performance metric to compare different decisions within the information elasticity framework.

Unlike sample covariance estimation, RCML estimation assumes that the true covariance matrix Σ follows a specific structure, given as follows:

$$\Sigma = \sigma^2 I + \Sigma_c \quad (11)$$

where σ^2 is the noise power, I is the $N \times N$ identity matrix, and Σ_c is the clutter matrix. As discussed in Section I, this clutter matrix must be positive semi-definite and rank deficient, in that it has rank $r < N$ (rank is obtained under certain conditions via the Brennan rule).

The method for obtaining the RCML estimate is given in [18], [7], and is briefly described here. Using K training samples, the sample covariance matrix is obtained, as in (5). Denote this sample covariance matrix as $\mathbf{S} = \frac{1}{K} \sum_{k=1}^K \mathbf{x}_k \mathbf{x}_k^H$, and denote its scaled form as $\mathbf{S}' = \frac{1}{\sigma^2} \mathbf{S}$. Clearly, $\mathbf{S}'$ is obtained if the noise power σ^2 is known. This term can typically be estimated by finding the thermal noise (also known as “ kTB ” noise) and applying the relevant noise factor terms, or by measuring receiver data while the radar is in receive only mode [19], [9]. Note also that [7] also provides an RCML estimator for the case when σ^2 is unknown, but a lower bound on the noise power can be found.

The RCML estimator is derived from the log likelihood of observing K secondary/training vectors $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K\}$ given the true disturbance covariance matrix Σ . Again, these secondary vectors are assumed to be homogeneous, identically distributed as $\mathbf{x}_k \sim \mathcal{CN}(0, \Sigma)$ for $1 \leq k \leq K$. This log likelihood is taken to be the cost function of an optimization problem, which is able to be simplified to the following convex function:

$$\mathbf{d}^T \lambda - \mathbf{1}^T \log \lambda \quad (12)$$

where $\mathbf{d}$ is an N length vector containing eigenvalues of $\mathbf{S}'$, and λ is an N length vector containing eigenvalues of $\sigma^2 \mathbf{R}^{-1}$. Note that minimizing this function maximizes the log likelihood function. Furthermore, the λ that minimizes (12) (denoted by λ^*) is used to find $\hat{\Sigma}$ as follows:

$$\hat{\Sigma} = \sigma^2 \mathbf{V} \Lambda^{*-1} \mathbf{V}^H \quad (13)$$

where $\mathbf{V}$ comes from the eigendecomposition $\mathbf{S} = \mathbf{V} \mathbf{D} \mathbf{V}^H$, and Λ^* is a diagonal matrix containing the elements of λ^* . The work in [7] presents a closed form for λ^* , which is used for the results in this paper.

When this estimator is used in the AMF (given in (4)), P_D becomes a function of the SNR and $\hat{\Sigma}$ [7]. Unlike the P_D for the SMI case, P_D does not easily allow a closed form when RCML estimation is used. Thus, the P_D vs SNR curves are generated empirically. In this paper, true covariance matrices (which follow the required structure for RCML, given in (11)) of different sizes are generated. These matrices are used to generate training samples for RCML estimation, which are then used to find the α value required to achieve the desired P_{FA} . Once α is found, Monte Carlo simulations are used to find P_D at different SNR values. Note that this method of obtaining P_D vs SNR is identical to the method used to find these curves in [17].

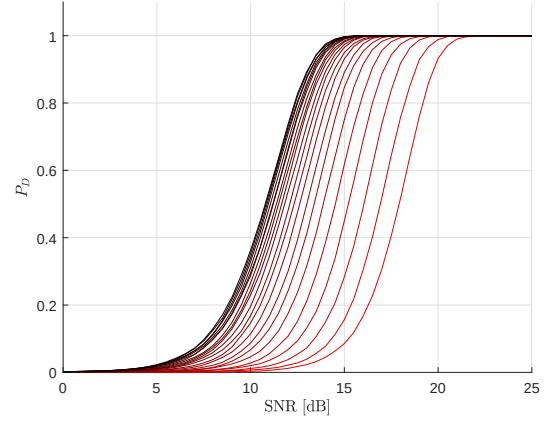


Fig. 6: P_D vs SNR for $N = 16$, $K = \{3, 4, \dots, 24\}$, $r = 7$, and $P_{FA} = 10^{-4}$.

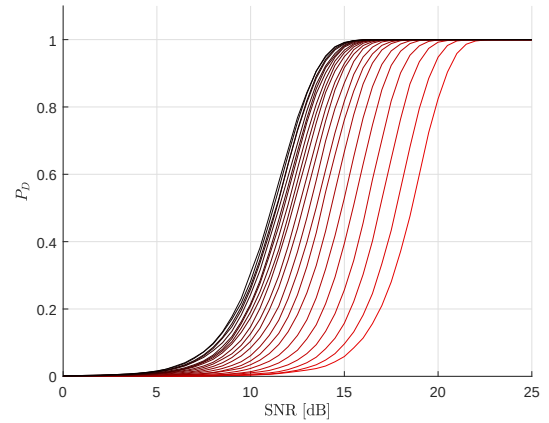


Fig. 7: P_D vs SNR for $N = 24$, $K = \{3, 4, \dots, 24\}$, $r = 9$, and $P_{FA} = 10^{-4}$.

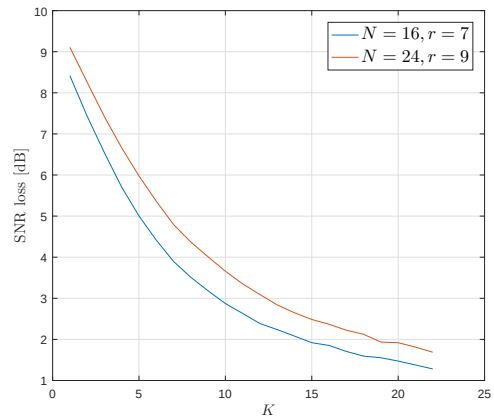


Fig. 8: SNR loss at $P_D = 0.6$ vs K for the data shown in Figures 6 and 7.

Several P_D vs SNR curves generated in this empirical manner are shown in Figures 6 and 7. The SNR loss at $P_D = 0.6$ is found using the data from these curves. The

SNR loss is taken at $P_D = 0.6$, since the P_D vs SNR curves generally exhibit linear behavior in this region. This allows for linear interpolation to be used to approximate the SNR required to produce $P_D = 0.6$ for the RCML AMF. The SNR at which $P_D = 0.6$ for the clairvoyant detector is obtained via the same method in Section II-B (using line search methods on equation (6)). Again, the ratio of these two SNR values is taken, yielding the SNR loss. The SNR losses for the examples in Figures 6 and 7 are shown as a function of K in Figure 8. As deduced from Figure 8, this relationship between SNR loss and K is clearly dependent on the dimensionality N and the rank of the clutter matrix r .

III. INFORMATION ELASTICITY FRAMEWORK

The concept of information elasticity considers a measure of decision effectiveness and a quantitative decision unit, represented by E and Q respectively. These parameters are typically user defined to fit the context they are being used in. The Q parameter should typically be a quantitative unit that the decision maker has the ability to control. In the context of the AMF and other such detectors, typically a DM must select the desired P_{FA} as well as the number of secondary vectors used to estimate the disturbance covariance matrix. Thus, let Q be represented as a vector containing both K and P_{FA} : $\mathbf{Q} = \{P_{FA}, K\}$. Note that while the P_{FA} represents a probability rather than a quantitative value, the P_{FA} is directly analogous to the quantity of false alarms per unit time (false alarm rate) [9].

Several factors are typically taken into account when selecting these parameters. Typically, less false alarms, and thus lower values of P_{FA} are more desirable for a DM. However, the receiver operating characteristic (ROC) of the detector must also be taken into account, and the DM must ensure that P_D levels are satisfactory at the desired P_{FA} . On the other hand, a DM typically selects K based on number of homogeneous samples that are available in a given environment. Thus, the selection of K is heavily dependent on the type of clutter and interference models being considered [10]. Additionally, the selection of K must also, in general, provide satisfactory levels of P_D .

Taking these factors into account, an appropriate selection of K and P_{FA} must consider the trade-offs presented. Thus, a constraint function $C(Q)$ is defined which characterizes the factors influencing a DM to reduce the number of false alarms and training samples. Furthermore, a decision quality metric $D(Q)$ is defined, which characterizes the factors influencing a DM to increase these decision parameters. The effectiveness of a decision, E , is thus a function of $C(Q)$ and $D(Q)$.

In this framework, the DM specifies a domain for the decision variables consisting only of decisions that are relevant to the DM. The interval of allowable decision values is defined as follows:

$$a \leq P_{FA} \leq b \quad (14)$$

$$c \leq K \leq d \quad (15)$$

For example, a user may define a if the ROC curve shows that P_D is unsatisfactory for P_{FA} values below a . The b parameter

would be defined as the maximum P_{FA} that he or she would tolerate in their system. A user might define c if the amount of SNR loss was too high for values of K below c . Finally, a user might define d if there were typically only d homogeneous training samples available in a given environment.

The decision metric, $D(\mathbf{Q})$, is defined to be the SNR loss, since it generally describes how well the detector is performing (how close the performance is to clairvoyant performance). Clearly, a lower SNR loss, and thus a lower $D(\mathbf{Q})$ is more desirable for a DM. Furthermore, increasing either K or P_{FA} will cause the SNR loss to decrease. Thus, the DM must increase P_{FA} and K to minimize $D(\mathbf{Q})$.

The constraint function, $C(\mathbf{Q})$, should describe the level of detriment associated with selecting a given (K, P_{FA}) pair. Since this generally varies between different decision makers and contexts, this paper seeks to define a generalized cost function that is user-defined and is able to be fine-tuned by the decision maker. This function describes the relative “cost” of using a given (K, P_{FA}) pair. These cost functions monotonically increase as either K or P_{FA} increase. This is intuitively understood for P_{FA} , since a larger number of false alarms is typically more costly for a decision maker. On the other hand, a larger K value can be thought to increase the likelihood that one of the training samples includes returns from an interfering target, clutter edge, or some other non-homogeneity [9][10]. Thus, in contrast to $D(\mathbf{Q})$, the DM must decrease P_{FA} and K to minimize $C(\mathbf{Q})$.

A. Formulation of user-defined constraint function

Consider the following functions:

$$C_{P_{FA}} = \left(\frac{P_{FA} - a}{b - a} \right)^n \quad \text{for } a \leq P_{FA} \leq b \quad (16)$$

$$C_K = \left(\frac{K - c}{d - c} \right)^m \quad \text{for } c \leq K \leq d \quad (17)$$

Equation (16) represents the component of the cost associated with P_{FA} . Similarly, (17) represents the component of the cost associated with K . Since these functions describe “relative cost,” it is important to define a scale of cost over which decisions can be compared. Thus, 1 is defined to be the maximum cost and 0 is defined to be the minimum cost. With this in mind, both $C_{P_{FA}}$ and C_K increase monotonically from 0 to 1. Furthermore, the n and m parameters allow the DM to change the rate at which these functions increase.

The cost function of the decision pair (P_{FA}, K) is given to be a linear combination of the costs defined in (17) and (16), as shown below:

$$C(\mathbf{Q}) = C(P_{FA}, K) = \lambda_1 C_{P_{FA}} + \lambda_2 C_K \quad (18)$$

$$\text{for } a \leq P_{FA} \leq b \quad \text{and} \quad c \leq K \leq d$$

$$\text{s.t. } \lambda_1 + \lambda_2 = 1$$

The λ_1 and λ_2 parameters represent the weight of each cost component. The DM defines the ratio $\frac{\lambda_1}{\lambda_2}$ to represent the relative level of importance over $C_{P_{FA}}$ over C_K . Note also that these parameters must sum to 1 in order for the maximum $C(P_{FA}, K)$ cost to be equal to 1.

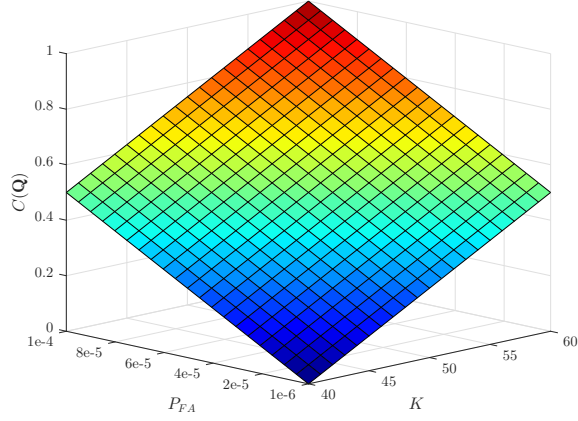


Fig. 9: Constraint function $C_1(\mathbf{Q})$ for $\lambda_1 = \lambda_2 = 0.5$, $n = m = 1$, $a = 10^{-4}$, $b = 10^{-6}$, $c = 40$, and $d = 60$.

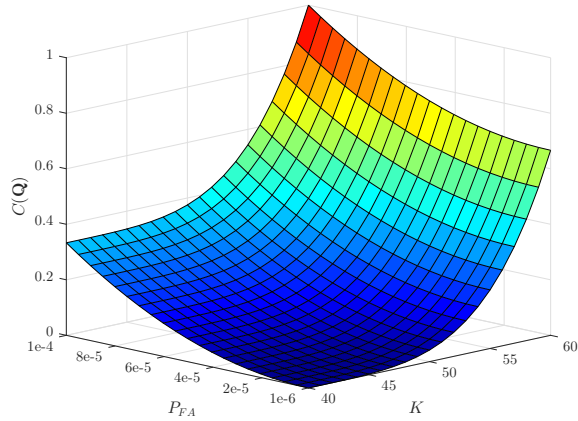


Fig. 10: Constraint function $C_2(\mathbf{Q})$ for $\lambda_1 = 1/3$, $\lambda_2 = 2/3$, $n = 2$, $m = 4$, $a = 10^{-4}$, $b = 10^{-6}$, $c = 40$, and $d = 60$.

It is also possible that a decision maker would want to define a cost function for either P_{FA} or K , but not both. This might be the case if it is required that a DM selects a specific P_{FA} value but is given freedom over the choice in K , or vice versa. In this case, the DM would use either (17) or (16) as the constraint function, which is equivalent to using (18) and setting either λ_1 or λ_2 to be 0.

Figures 9 and 10 show examples of two possible constraint functions, labeled $C_1(\mathbf{Q})$ and $C_2(\mathbf{Q})$ respectively. Both functions share the same values for a , b , c , and d . For these examples, 60 is the maximum allowable value for K , perhaps because the decision maker is dealing with a scenario where there are typically only 60 homogeneous samples available for use. These examples also consider a minimum K value of 40. This may be the case for a decision maker who is using a radar system where $N = 20$. As explained in Section I, at least $K = 2N$ training samples are required to ensure that the random SNR loss ρ does not have an expectation below 3 dB [4]. Thus, if $N = 20$, the decision maker may choose the minimum value of K to be 40 so as to not let ρ reach

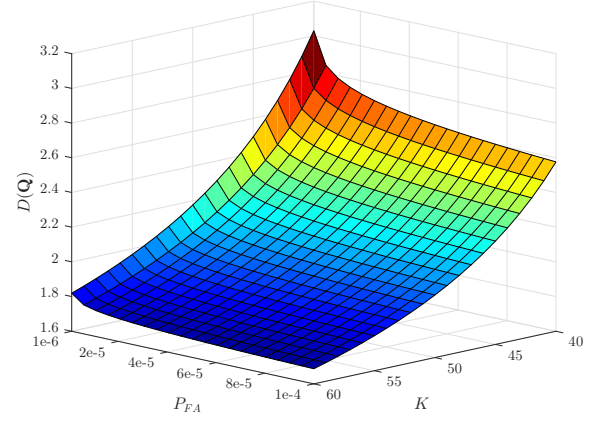


Fig. 11: Decision metric $D(\mathbf{Q})$ for the AMF using sample covariance estimation and $N = 20$. Domain parameters are $a = 10^{-4}$, $b = 10^{-6}$, $c = 40$, and $d = 60$. Note that the P_{FA} and K axes are inverted versions of the axes in Figures 9 and 10.

unsuitable levels.

For $C_1(\mathbf{Q})$, λ_1 and λ_2 are selected such that K and P_{FA} equally contribute to the constraint function. $C_2(\mathbf{Q})$, on the other hand, weights its component associated with K to contribute twice as much as its component associated with P_{FA} . Clearly, $C_2(\mathbf{Q})$ would be more suitable for a decision maker who considers a large K to be generally more costly than a large P_{FA} value.

Note also that $C_2(\mathbf{Q})$ uses $n = 2$ and $m = 4$, while $C_1(\mathbf{Q})$ uses $n = m = 1$. For $C_1(\mathbf{Q})$, increasing K or P_{FA} causes a linear increase in the cost. However, this is not the case for $C_2(\mathbf{Q})$, since the individual components of $C_2(\mathbf{Q})$ are exponentiated. As deduced from Figure 10, the gradient of $C_2(\mathbf{Q})$ becomes quite large when the decision variables are near their upper extremities. Similarly the gradient decreases when K or P_{FA} are near their lower extremities. From Figures 9 and 10, it is clear that due to the effects of m and n , $C_2(\mathbf{Q})$ sees a higher concentration of low cost decisions than $C_1(\mathbf{Q})$.

While $C_1(\mathbf{Q})$ and $C_2(\mathbf{Q})$ are nuanced differently based on DM preference, Both functions share two specific points. Clearly from the figures, $C_1(\mathbf{Q}_0) = C_2(\mathbf{Q}_0) = 0$ where $\mathbf{Q}_0$ is the point of minimum cost: $\mathbf{Q}_0 = \{a, c\}$. Similarly, $C_1(\mathbf{Q}_1) = C_2(\mathbf{Q}_1) = 0$ where $\mathbf{Q}_1$ is the point of maximum cost: $\mathbf{Q}_1 = \{b, d\}$.

B. Formulation of decision effectiveness using compromise programming

Recall that a lower $D(\mathbf{Q})$ value is desirable for a DM since it represents the SNR loss parameter described in Section II. Figure 11 shows the decision metric for an AMF detector using SMI with dimensionality $N = 20$. Note that this function is maximized at the point of minimum cost (in this case, $P_{FA} = 1e-6$ and $K = 40$) and is minimized at the point of maximum cost ($P_{FA} = 1e-4$ and $K = 60$).

From the way $C(\mathbf{Q})$ and $D(\mathbf{Q})$ have been defined, an increase in either P_{FA} or K will produce a decrease in the

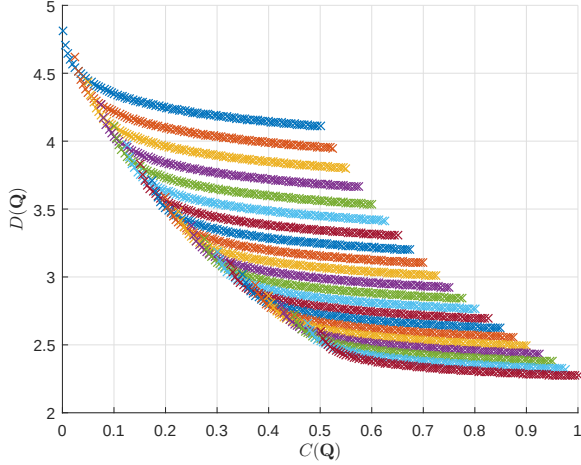


Fig. 12: $C_1(\mathbf{Q})$ and $D(\mathbf{Q})$ for different decision points given $N = 20$. Note that points of a shared color represent data of a shared K value.

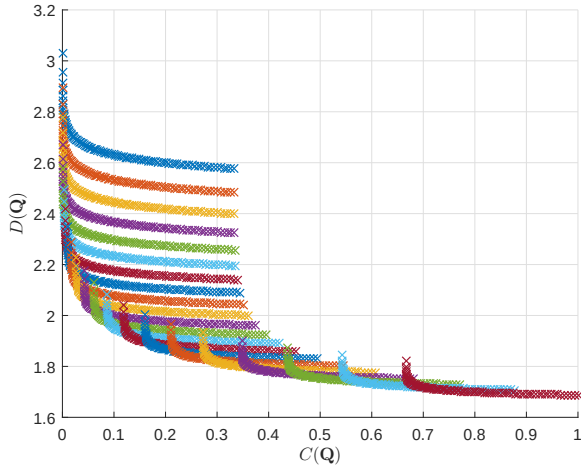


Fig. 13: $C_2(\mathbf{Q})$ and $D(\mathbf{Q})$ for different decision points given $N = 20$. Again, points of a shared color represent data of a shared K value.

decision metric and an increase in the constraint function. This behavior is demonstrated by the examples in Figures 9, 10, and 11. To analyze this trade-off behavior between the two functions, $C(\mathbf{Q})$ and $D(\mathbf{Q})$ of different decision pairs $\mathbf{Q} = (P_{FA}, K)$ are graphed on the same plot. This is shown in Figure 12, which displays $C_1(\mathbf{Q})$ and $D(\mathbf{Q})$ for different decision points, and Figure 13 which shows $C_2(\mathbf{Q})$ and $D(\mathbf{Q})$ for different decision points.

These two figures show where each decision point within the domain specified in (15) and (14) exist in the criterion space (where $C(\mathbf{Q})$ and $D(\mathbf{Q})$ are the only two criteria). Note that only 100 evenly spaced P_{FA} values are represented in this figure for each K value. However, more P_{FA} values can be used depending on how finely a decision maker wants to resolve P_{FA} . Within this figure, solutions with smaller $C(\mathbf{Q})$ and $D(\mathbf{Q})$ values are more desirable. Thus, there is a

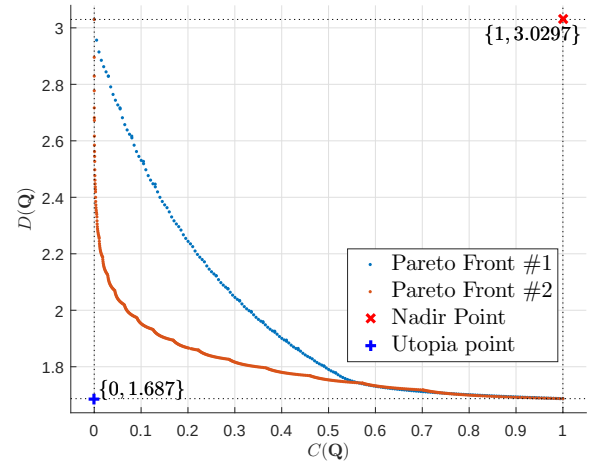


Fig. 14: Pareto front #1 shows non-dominated solutions for $C_1(\mathbf{Q})$ and $D(\mathbf{Q})$, while Pareto front #2 shows non-dominated solutions for $C_2(\mathbf{Q})$ and $D(\mathbf{Q})$.

clearly defined Pareto frontier where non-dominated or Pareto efficient solutions exist. More formally, the decision point (P_{FA}^*, K^*) is considered Pareto efficient as long as there is no other decision (P_{FA}', K') such that $C(P_{FA}', K') < C(P_{FA}^*, K^*)$ and $D(P_{FA}', K') < D(P_{FA}^*, K^*)$.

The Pareto frontiers of both examples are shown in Figure 14. One commonly used method used to evaluate decisions along this frontier is known as compromise programming [20]. This method considers two specific points within the criterion space. The first point is known as the nadir point, and it is the point at which each objective function is maximized, i.e. $F^1 = \left\{ \max_{\mathbf{Q}} C(\mathbf{Q}), \max_{\mathbf{Q}} D(\mathbf{Q}) \right\}$. From the way that $C(\mathbf{Q})$ and $D(\mathbf{Q})$ have been defined, it is clear that $C(\mathbf{Q})$ has a maximum value of 1 and $D(\mathbf{Q})$ is maximized at the point of minimum cost, i.e. $\mathbf{Q}_0$. Thus, the nadir point is given as:

$$F_1 = \{1, D(\mathbf{Q}_0)\} \quad (19)$$

Compromise programming is primarily focused on the utopia point, which is the point at which each objective function is minimized, i.e. $F_0 = \left\{ \min_{\mathbf{Q}} C(\mathbf{Q}), \min_{\mathbf{Q}} D(\mathbf{Q}) \right\}$. It is clear that $C(\mathbf{Q})$ has a minimum value of 0 and $D(\mathbf{Q})$ is minimized at the point of maximum cost, i.e. $\mathbf{Q}_1$. Thus the utopia point is given as:

$$F_0 = \{0, D(\mathbf{Q}_1)\} \quad (20)$$

Note that the utopia point exists only in the criterion space [21], in that there is no decision that can achieve the performance at F_0 . The utopia point serves only as an idealized baseline where the optimum value of each criterion is represented. The “compromise” solution is a solution that minimizes some measure of distance from the utopia point.

This measure of distance is specified such that each criterion is normalized and weighted. This normalization is necessary, since otherwise the distance measure shows bias towards objectives with smaller values [21]. The weightings are used

to allow the decision maker to specify how important each objective is to him/her. Let w_1 represent the weight on $C(\mathbf{Q})$ and w_2 represent the weight on $D(\mathbf{Q})$. The distance measure/decision effectiveness is given as:

$$E = \left(w_1 [C(\mathbf{Q})]^2 + w_2 \left[\frac{D(\mathbf{Q}) - D(\mathbf{Q}_1)}{D(\mathbf{Q}_0) - D(\mathbf{Q}_1)} \right]^2 \right)^{1/2} \quad (21)$$

This formulation considers the weighted and normalized L_2 distance from the utopia point to points in the performance space. The weighting and normalization are defined such that $E = 0$ at the utopia point and $E = 1$ at the Nadir point. Thus, values of E will always lie between 0 and 1.

IV. RESULTS AND DISCUSSION

For sake of comparison, let DM 1 be a decision maker using $C_1(\mathbf{Q})$ as their constraint function, and let DM 2 be a decision maker using $C_2(\mathbf{Q})$ as their constraint function. Figures 15 and 16 show the decision effectiveness of Pareto efficient solutions as a function of their constraint values for DM 1 and DM 2 respectively. Each figure shows two different decision effectiveness curves, one where the constraint objective is weighted higher ($w_1 > w_2$) and another where the decision metric is weighted higher ($w_1 < w_2$). While all four of these curves are different, they each exhibit information overload, in the context of [11]. Essentially, as the Pareto efficient solutions become more costly (and thus larger quantities of the $\mathbf{Q}$ parameters are being used), the decision effectiveness improves up to minimum point known as the overload point. Increasing the cost beyond this point only results in a decision with a worse decision effectiveness. The compromise solution [20], at which E is minimized, clearly occurs at this overload point.

The decision at the overload point along with its associated SNR loss and cost are provided in Tables I and II. It is clear in both of these tables that using a larger w_1 yields a solution with a lower cost, and that using a larger w_2 yields a solution with a lower SNR loss. Thus, these weights operate as expected, allowing a decision maker to emphasize the importance of different criteria. Note also that the P_{FA} values for DM 2 are much larger than the P_{FA} values for DM 1. This is likely due to the fact that $C_2(\mathbf{Q})$ uses a smaller λ_1 value than $C_1(\mathbf{Q})$. In general, this causes larger P_{FA} values to have a smaller relative cost for DM 2 than for DM 1. This is clearly shown in Figures 9 and 10. Thus, higher P_{FA} values are more cost efficient for DM 2.

TABLE II: Decisions at which E is minimized using constraint function $C_2(\mathbf{Q})$.

	SNR loss	Cost	K	P_{FA}	α
$w_1 < w_2$	2.73 dB	0.1831	54	$2.7 \cdot 10^{-5}$	0.5633
$w_1 > w_2$	2.91 dB	0.0987	52	$2 \cdot 10^{-5}$	0.6400

TABLE I: Decisions at which E is minimized using constraint function $C_1(\mathbf{Q})$.

	SNR loss	Cost	K	P_{FA}	α
$w_1 < w_2$	2.80 dB	0.3952	55	$5 \cdot 10^{-6}$	0.6458
$w_1 > w_2$	3.32 dB	0.2452	49	$5 \cdot 10^{-6}$	0.8717

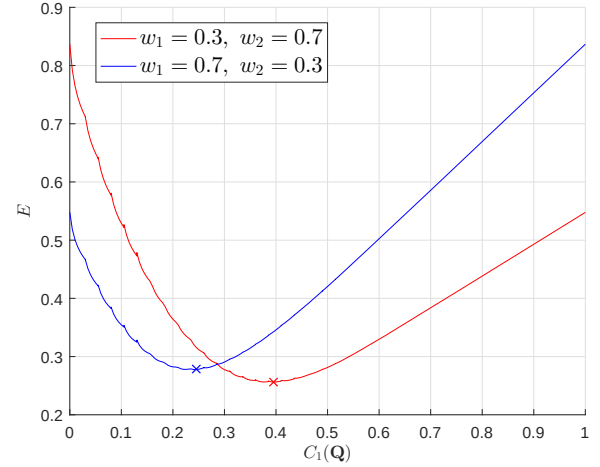


Fig. 15: Decision effectiveness E of Pareto efficient decisions shown as a function of their cost $C_1(\mathbf{Q})$. Overload points are marked on each curve.

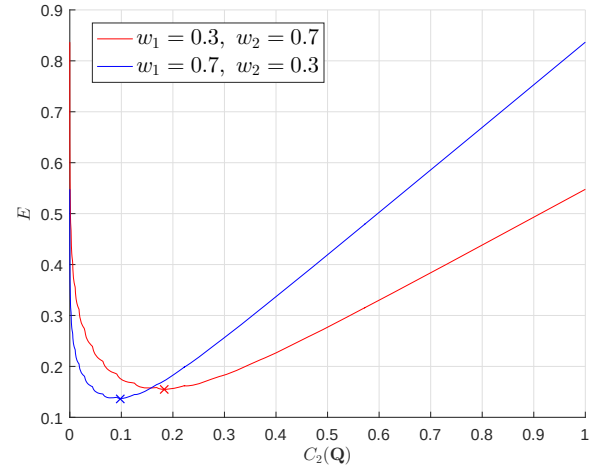


Fig. 16: Decision effectiveness E of Pareto efficient decisions shown as a function of their cost $C_2(\mathbf{Q})$. Overload points are marked on each curve.

Note also that DM 2 has selected decisions that are associated with SNR losses that are lower than that of DM 1. From the criterion spaces of these decisions (which can be seen in Figures 12 and 13), it is clear that the decisions associated with lower SNR losses are generally lower in cost for DM 2 than for DM 1. This can be analyzed further by observing the decision spaces (in Figures 9, 10, and 11), where the $C_2(\mathbf{Q})$ is shown to have a larger gradient than $C_1(\mathbf{Q})$ within the region of the decision space where the SNR loss is low. This is clearly due to the fact that $C_1(\mathbf{Q})$ considers a linear increase in cost while $C_2(\mathbf{Q})$ uses $n = 2$ and $m = 4$. It should also be noted that while DM 2 sees decisions that have lower costs than DM 1, their costs should not directly be compared, since each decision maker considers a different cost function.

Table III provides parameters for more constraint functions, and Table IV shows specifications for more decision metrics.

TABLE III: Constraint function parameters

	λ_1	λ_2	n	m
$C_a(\mathbf{Q})$	1/4	3/4	4	2
$C_b(\mathbf{Q})$	3/4	1/4	4	2
$C_c(\mathbf{Q})$	1/4	3/4	5	10
$C_d(\mathbf{Q})$	3/4	1/4	5	10
$C_e(\mathbf{Q})$	1/4	3/4	15	15
$C_f(\mathbf{Q})$	3/4	1/4	15	15

TABLE IV: Specification for decision metrics

	Estimation Type	N	r	a	b	c	d
$D_a(\mathbf{Q})$	SMI	16	N/A	10^{-4}	10^{-6}	30	40
$D_b(\mathbf{Q})$	RCML	16	7	10^{-4}	10^{-6}	10	24
$D_c(\mathbf{Q})$	SMI	24	N/A	10^{-4}	10^{-6}	50	60
$D_d(\mathbf{Q})$	RCML	24	9	10^{-4}	10^{-6}	15	24

TABLE V: Decisions at which E is minimized for different constraint functions and decision metrics. Note that $w_1 = w_2$ for each decision.

Decision Metric	Constraint	SNR loss	Cost	K	P_{FA}
$D_a(\mathbf{Q})$	$C_a(\mathbf{Q})$	3.89 dB	0.2013	35	4.9e-5
	$C_b(\mathbf{Q})$	3.61 dB	0.1329	37	3.5e-5
	$C_c(\mathbf{Q})$	3.66 dB	0.0687	36	7.3e-5
	$C_d(\mathbf{Q})$	3.40 dB	0.0855	38	5.9e-5
	$C_e(\mathbf{Q})$	3.36 dB	0.0322	38	7.8e-5
	$C_f(\mathbf{Q})$	3.26 dB	0.0532	39	6.7e-5
$D_b(\mathbf{Q})$	$C_a(\mathbf{Q})$	3.03 dB	0.1929	17	3.9e-5
	$C_b(\mathbf{Q})$	2.30 dB	0.1294	20	2.3e-5
	$C_c(\mathbf{Q})$	2.06 dB	0.0432	20	5.9e-5
	$C_d(\mathbf{Q})$	1.82 dB	0.0545	22	2.7e-5
	$C_e(\mathbf{Q})$	1.72 dB	0.0286	21	8e-5
	$C_f(\mathbf{Q})$	1.63 dB	0.0343	22	7.5e-5
$D_c(\mathbf{Q})$	$C_a(\mathbf{Q})$	3.25 dB	0.2080	55	5.4e-5
	$C_b(\mathbf{Q})$	3.04 dB	0.1704	58	3.5e-5
	$C_c(\mathbf{Q})$	3.06 dB	0.1328	57	7e-5
	$C_d(\mathbf{Q})$	2.99 dB	0.0889	58	6.3e-5
	$C_e(\mathbf{Q})$	2.97 dB	0.0366	58	8.1e-5
	$C_f(\mathbf{Q})$	2.91 dB	0.0548	59	7e-5
$D_d(\mathbf{Q})$	$C_a(\mathbf{Q})$	3.50 dB	0.1819	19	6.1e-5
	$C_b(\mathbf{Q})$	2.96 dB	0.1404	21	4.5e-5
	$C_c(\mathbf{Q})$	2.79 dB	0.0728	22	5.5e-5
	$C_d(\mathbf{Q})$	2.58 dB	0.0841	23	4e-5
	$C_e(\mathbf{Q})$	2.71 dB	0.0386	22	8.5e-5
	$C_f(\mathbf{Q})$	2.48 dB	0.0438	23	6.5e-5

Note that some of these decision metrics are defined for AMF detectors using SMI while other are defined for AMF detectors using RCML. Using different combinations of these metrics, the decision at which E is minimized (overload solution) is shown in Table V along with its associated SNR loss and cost. Just as discussed above, increasing m and n causes the relative cost of the overload solution to decrease for the decisions made in Table V.

Note that the decision metrics associated with the RCML AMF detectors consider much smaller domains for the K variable. Specifically, $D_b(\mathbf{Q})$ and $D_d(\mathbf{Q})$ only consider values of

K up to 24, while $D_a(\mathbf{Q})$ considers K values up to 40 and $D_c(\mathbf{Q})$ considers K values up to 60 (the latter two decision metrics are for SMI AMF detectors). This is done to highlight the fact that RCML estimation is able to perform well in cases where the amount of usable training data is limited. For example, consider that the detectors associated with $D_a(\mathbf{Q})$ and $D_b(\mathbf{Q})$ both have a dimensionality of 16. Given the same decision constraint function $C_f(\mathbf{Q})$, the overload solution yields $K = 22$ for the RCML detector and $K = 39$ for the SMI detector. Even though the overload solution SMI detector selects a much higher training size, the overload solution for the SMI detector outperforms it in SNR loss by a wide margin (1.63 dB SNR loss compared to 3.26 dB).

The only case where the overload solution for RCML does not outperform the overload solution for SMI is when $C_a(\mathbf{Q})$ is used for the detectors associated with $D_c(\mathbf{Q})$ and $D_d(\mathbf{Q})$. However, note that the overload solution for the RCML detector selects $K = 19$ and the overload solution for the SMI detector selects $K = 55$. This is because the $D_d(\mathbf{Q})$ is defined to represent a scenario where the homogeneous training data are much more scarce, while $D_c(\mathbf{Q})$ is defined to be much more forgiving. From the way that $C_a(\mathbf{Q})$ is defined, there is not very many low cost solutions near the decision region where SNR is low, causing the overload solution to select values of K that produce a large SNR loss. However, it is shown that the RCML detector given in $D_d(\mathbf{Q})$ greatly outperforms the SMI detector given in $D_c(\mathbf{Q})$ when all other constraint functions are used, even though the decisions are selected in the data starved scenario.

V. CONCLUSION

A decision making scheme for selecting parameters for the AMF detector by Robey et. al [2] based on a DM's context and application is proposed. This scheme does not claim to select the most optimal or best decision. Rather, an information elasticity framework is used given a domain of usable decisions and parameters defining the cost of decisions, both defined by a DM. The cost and benefit of each decision is compared, and a solution that balances the trade-offs between them is selected.

The cost of decisions is represented using a simple constraint function with tunable parameters. The benefit of decisions is represented using a decision metric function, which is defined to be the additional SNR required for the AMF to reach the detection probability of the clairvoyant detector. This trait is analyzed for AMF detectors using sample covariance estimation (for which an approximation is derived) and RCML estimation (for which data are collected empirically). The constraint function and decision quality metric are defined as criteria that the DM wishes to minimize, and a Pareto front of non-dominated solutions is generated in this criterion space. Compromise programming is used to find a specified measure of distance in criterion space from the utopia point to the Pareto efficient decisions.

APPENDIX
PROOF OF THE CONVERGENCE OF THE AMF

The approximation shown in (10) is found using the P_D equation for the Swerling I fluctuation model. This model is used when the radar cross section of targets in question are constant over a given pulse but fluctuate in between scans [22]. This model is used here because its P_D equation is much simpler than P_D for the non-fluctuating case (given in (9)). This equation is as follows [2]:

$$P_D = \mathbb{E}_\rho \left[\left(\frac{1 + \text{SNR}\rho}{1 + (\alpha + \text{SNR})\rho} \right)^L \right] \quad (22)$$

$$= \int_0^1 \left(\frac{1 + \text{SNR}\rho}{1 + (\alpha + \text{SNR})\rho} \right)^L \frac{K! \rho^L (1 - \rho)^{N-2}}{L!(N-2)!} d\rho$$

where $\mathbb{E}(\cdot)$ is the expectation operator, and ρ is the SNR loss factor. The clairvoyant detection probability under Swerling I is given as [2]:

$$P_{D,CV} = \exp \left(\frac{-\alpha_{CV}}{1 + \text{SNR}} \right) \quad (23)$$

Note that both α and α_{CV} are obtained the same way as in Section II-B. Consider the P_D conditioned on a given loss factor ρ :

$$P_{D|\rho} = \left(\frac{1 + \text{SNR}\rho}{1 + (\alpha + \text{SNR})\rho} \right)^L$$

$$= \left(\frac{\frac{1}{\rho} + \text{SNR}}{\frac{1}{\rho} + \alpha + \text{SNR}} \right)^L \quad (24)$$

The leading idea behind this approximation is that the variance of $P_{D|\rho}$ is generally extremely small. Thus, this variance is assumed to be negligible, and the expectation $\mathbb{E}(1/\rho)$ is used in place of $1/\rho$ in (24).

Note that the loss factor ρ is beta distributed with parameters $L+1$ and $N-1$. Thus, the variance of $\frac{1}{\rho}$ is found as follows:

$$\text{Var}(1/\rho) = \mathbb{E}(1/\rho^2) - \mathbb{E}(1/\rho)^2$$

$$\mathbb{E}(1/\rho^2) = \int_0^1 \frac{1}{\rho^2} \frac{K! \rho^L (1 - \rho)^{N-2}}{L!(N-2)!} d\rho$$

$$= \frac{K(K-1)}{L(L-1)} \underbrace{\int_0^1 \frac{(K-2)! \rho^{L-2} (1 - \rho)^{N-2}}{(L-2)!(N-2)!} d\rho}_{=1}$$

$$= \frac{K(K-1)}{L(L-1)}$$

$$\mathbb{E}(1/\rho) = \int_0^1 \frac{1}{\rho} \frac{K! \rho^L (1 - \rho)^{N-2}}{L!(N-2)!} d\rho$$

$$= \frac{K}{L} \underbrace{\int_0^1 \frac{(K-1)! \rho^{L-1} (1 - \rho)^{N-2}}{(L-1)!(N-2)!} d\rho}_{=1}$$

$$= \frac{K}{L}$$

$$\text{Var}(1/\rho) = \frac{K(N-1)}{L^2(L-1)}$$

$$= \frac{m(N-1)}{(m-1)(N(m-1)+1)^2} \quad (25)$$

where m is the ratio $m = K/N$. Recall that $K > N$ in order for the covariance matrix estimate to be invertible. Thus, m must be greater than 1. Consider $P_{D|\rho}$ when the SNR is low. In this case, $P_{D|\rho}$ is generally very close to 0. The fact that low SNRs are generally associated with low detection probabilities is well known, and is also shown in Figure 2. This behavior is regardless of the choice of m , which is explained as follows:

1) m is close to 1: In this case, we can consider that K is close to but not less than N . This choice of parameters causes the PDF of ρ to have large likelihoods at small ρ values. Due to the nature of (7), α becomes very large when this is the case, to account for the fact that ρ is more likely to be small. In this case, $\alpha \gg (1/\rho + \text{SNR})$ (for low SNR) and thus $P_{D|\rho} \approx 0$.

However, as the SNR increases, this is no longer true. In other words, the SNR will eventually increase to be large enough such that $\alpha \gg (1/\rho + \text{SNR})$ is not true, and thus $P_{D|\rho} \not\approx 0$. However, when this is the case, the SNR is also large enough such that $\text{SNR} \gg \text{Var}(1/\rho)$. Consider, $\text{Var}(1/\rho)$ is fairly large when m is close to 1, as shown by equation (25). However, since α is so large, the SNR at which $P_{D|\rho} \not\approx 0$ is generally much larger than $\text{Var}(1/\rho)$ when m is close to 1.

2) $m \gg 1$: In this case, we can consider that K is much larger than N . Furthermore, since $L = K - N + 1$, we can consider that $L \gg 1$. Since α is positive, $\frac{\frac{1}{\rho} + \text{SNR}}{\frac{1}{\rho} + \text{SNR} + \alpha}$ is always less than 1. Consider, this term decreases as the SNR decreases. When the SNR is very low, this term becomes small enough that $\left(\frac{\frac{1}{\rho} + \text{SNR}}{\frac{1}{\rho} + \text{SNR} + \alpha} \right)^L \approx 0$. Thus, $P_{D|\rho} \approx 0$ in this case.

However, as the SNR increases, it will eventually become large enough that $P_{D|\rho} \not\approx 0$. When this is the case, it is also generally true that $\text{SNR} \gg \text{Var}(1/\rho)$. This is because $\text{Var}(1/\rho)$ is generally extremely small when $m \gg 1$, as shown by equation (25).

In summary, either the SNR is low enough such that $P_{D|\rho} \approx 0$, or the SNR is large enough such that $\text{SNR} \gg \text{Var}(1/\rho)$, regardless of what m value is selected. When $\text{SNR} \gg \text{Var}(1/\rho)$ is true, the variance of $1/\rho$ can be thought of as negligible, and we can replace $1/\rho$ with its expectation $\mathbb{E}(1/\rho)$, yielding the following approximation:

$$\mathbb{E}_\rho(P_{D|\rho}) \approx \left(\frac{\mathbb{E} \left[\frac{1}{\rho} \right] + \text{SNR}}{\mathbb{E} \left[\frac{1}{\rho} \right] + \alpha + \text{SNR}} \right)^L$$

$$P_D \approx \left(\frac{\frac{K}{L} + \text{SNR}}{\frac{K}{L} + \alpha + \text{SNR}} \right)^L \quad (26)$$

Note that all of the statements made above are verified using Monte Carlo simulations for many different values of m , SNR, and P_{FA} . Equations (23) and (26) easily allow a closed form

for the clairvoyant SNR and the AMF SNR. Using these SNR values, the SNR loss approximation is obtained:

$$\text{SNR loss} \approx \left(\alpha \frac{P_D^{1/L}}{1 - P_D^{1/L}} - \frac{K}{L} \right) \frac{\log(P_D)}{\log\left(\frac{P_{FA}}{P_D}\right)}$$

Note that while this approximation was derived for detectors using the Swerling I fluctuation model, this approximation is also applicable for the non-fluctuating case. This is because the SNR loss vs K is generally the same between the two cases. Figure 5 shows that this is indeed the case.

REFERENCES

- [1] W. L. Melvin, "A STAP overview," *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 19–35, Jan 2004.
- [2] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 1, pp. 208–216, Jan 1992.
- [3] B. Himed and W. L. Melvin, "Analyzing space-time adaptive processors using measured data," in *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers*, vol. 1, Pacific Grove, CA, Nov 1997, pp. 930–935.
- [4] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-10, no. 6, pp. 853–863, Nov 1974.
- [5] E. J. Kelly, "Adaptive detection in non-stationary interference. Part I and Part II," MIT Lincoln Laboratory, Lexington, MA, Tech. Rep. 724, May 1985.
- [6] M. Steiner and K. Gerlach, "Fast converging adaptive processor or a structured covariance matrix," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 4, pp. 1115–1126, Oct 2000.
- [7] B. Kang, V. Monga, and M. Rangaswamy, "Rank-constrained maximum likelihood estimation of structured covariance matrices," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 1, pp. 501–515, January 2014.
- [8] J. Ward, "Space-time adaptive processing for airborne radar," MIT Lincoln Laboratory, Tech. Rep. 105, Dec 1994.
- [9] M. Richards, W. Holm, and J. Scheer, *Principles of Modern Radar: Basic Principles*. Raleigh, NC: SciTech Publishing, 2010.
- [10] M. Weiss, "Analysis of some modified cell-averaging cfar processors in multiple-target situations," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-18, no. 1, pp. 102–114, Jan 1982.
- [11] R. M. Narayanan, A. Z. Liu, P. G. Singerman, and M. Rangaswamy, "Information elasticity in radar systems," *Electronics Letters*, vol. 54, no. 17, pp. 1049–1051, 2018.
- [12] A. Z. Liu, R. M. Narayanan, and M. Rangaswamy, "Information elasticity in pseudorandom code pulse compression," in *SPIE Conference on Radar Sensor Technology XXII*, Orlando, FL, Apr 2018. [Online]. Available: <https://doi.org/10.1117/12.2305853>
- [13] P. G. Singerman, R. M. Narayanan, and M. Rangaswamy, "Information elasticity in ultra-wideband target detection amongst distributed clutter," in *SPIE Conference on Radar Sensor Technology XXII*, Apr 2018. [Online]. Available: <https://doi.org/10.1117/12.2305854>
- [14] A. Z. Liu, R. M. Narayanan, and M. Rangaswamy, "Robust decision making method for adaptive ordered-statistics CFAR technique using information elasticity," in *SPIE Conference on Radar Sensor Technology XXIII*, Baltimore, MD, Apr 2019. [Online]. Available: <https://doi.org/10.1117/12.2519677>
- [15] E. J. Kelly, "An adaptive detection algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-22, no. 2, pp. 115–127, March 1986.
- [16] E. Lehmann, *Elements of Large-Sample Theory*. New York, NY: Springer-Verlag, 1999.
- [17] B. Kang, V. Monga, and M. Rangaswamy, "On the practical merits of rank constrained ML estimator of structured covariance matrices," in *2013 IEEE Radar Conference (RadarCon13)*, Ottawa, ON, Apr 2013. [Online]. Available: <https://doi.org/10.1109/RADAR.2013.6586015>
- [18] V. Monga and M. Rangaswamy, "Rank constrained ML estimation of structured covariance matrices with applications in radar target detection," in *2012 IEEE Radar Conference*, Atlanta, GA, May 2012, pp. 0475–0480.
- [19] M. I. Skolnik, *Radar Handbook, Third Edition*. New York, NY: McGraw-Hill, 2008.
- [20] G. O. Odu and O. E. Charles-Owaba, "Review of multi-criteria optimization methods - theory and applications," *IOSR Journal of Engineering*, vol. 3, no. 10, pp. 1–14, 2013.
- [21] K. Chang, *e-Design: Computer-Aided Engineering Design*. London, UK: Academic Press, 2016.
- [22] M. I. Skolnik, *Introduction to Radar Systems*. New York, NY: McGraw-Hill, 1962.

Students Supported

The following graduate students were supported on this research grant:

1. Andrew Z. Liu
2. Paul G. Singerman

Collaborators

The following served as AFRL collaborators for this research grant:

1. Dr. Muralidhar Rangaswamy
2. Dr. Sean M. O'Rourke