Formulation for min-max clairvoyant fusion based on monotonic recalibration of statistics

James Theiler

Los Alamos National Laboratory, Los Alamos, NM, USA

ABSTRACT

A formulation for min-max clairvoyant fusion (also known as continuum fusion) is developed that exploits the invariance of hypothesis testing statistics to monotonic transformations. In addition to generalizing an earlier formulation based on manipulated thresholds, the new formulation leads to efficient algorithms for two of the most widely advocated fusion "flavors:" one based on combining detectors with the same false alarm rate, and one based on constant detection rate. These algorithms are used to investigate and compare the performance of different detectors for a class of problems that arises from detecting small or weak targets in hyperspectral imagery. The experiments are performed on simulated data from well-defined distributions so as to isolate the effect of different flavors of fusion from the effects of model mismatch.

Keywords: composite hypothesis testing, statistical data analysis, continuum fusion, clairvoyant fusion, min-max detector, target detection, multispectral, hyperspectral imagery

1. INTRODUCTION

Hyperspectral imagery – with upwards of hundreds of spectral channels, and sometimes millions of pixels – provides a target detection tool with potentially exquisite discriminative ability. But such detection requires good models both of the desired target and of the background clutter, and these models are inevitably plagued by uncertainty, complexity, and ambiguity.

Statistical hypothesis testing provides a mathematical framework that addresses these issues in a way that is both principled and practical. In this framework, the two hypotheses (target present and target absent) are distinguished by likelihood functions of a measurement (typically the radiance or reflectance spectrum at a pixel, possibly augmented by neighborhood pixel information). The problem becomes more challenging when one or both of the competing hypotheses are *composite*, which is to say that they involve unknown nuisance parameters. The generalized likelihood ratio (GLR) has been the workhorse solution to this problem for decades, but a new mathematical approach – originally called continuum fusion[1–9] (CF), and more recently min-max fusion[10–12] or clairvoyant fusion[8,13] (also CF) – has recently been suggested as a way to address this composite hypothesis testing problem.

This paper advances the development of CF by presenting a generalized formulation of the CF process, and by providing algorithms that implement CF for two of the "flavors" of CF that have been proposed: one based on constant false alarm rate (CF-cfar) and one based on constant probability of detection (CF-cpd). This in turn will enable numerical comparison of CF and GLR detectors.

Section 2 sets up the background for this paper, describing statistical hypothesis testing in general, the distinction between decision rules and statistics (the latter are more general), and a formulation for CF based on manipulating the threshold. In Section 3, the new formulation of CF is presented, based on monotonic recalibration, and Section 4 exploits this new formulation to provide algorithms that can efficiently apply different flavors of CF to data drawn from arbitrary distributions. Section 5 uses these algorithms to investigate the CF-cfar and CF-cpd flavors of CF for detecting additive targets against non-Gaussian backgrounds. Conclusions are presented in Section 6.

2. STATISTICAL HYPOTHESIS TESTING

In the statistical hypothesis testing framework, one begins with two competing hypotheses. The "null" hypothesis $\mathcal{H}_o$ is that there is no target, while the "alternative" $\mathcal{H}_1$ says that a target is present. From a measurement $\mathbf{x}$, the task is to choose between $\mathcal{H}_o$ and $\mathcal{H}_1$. A *decision rule* is a binary function $\mathcal{B}(\mathbf{x})$ whose value (0 or 1) specifies the choice of hypothesis for each $\mathbf{x}$. There are two kinds of errors that can be made: if there is no target but $\mathcal{B}(\mathbf{x}) = 1$, that is a false alarm (also known as a Type I error); if a target is present but $\mathcal{B}(\mathbf{x}) = 0$, that is a missed detection (or Type II error).

2.1 Decision rule vs. Statistic

Many detectors are implemented in terms of a *statistic*: a real-valued function $\mathcal{T}(\mathbf{x})$ which characterizes the "target-ness" corresponding the measurement $\mathbf{x}$. A larger $\mathcal{T}(\mathbf{x})$ generally indicates more confidence that there is a target where $\mathbf{x}$ was measured. By comparing a statistic to a threshold, one obtains a decision rule. This can be written $\mathcal{B}(\mathbf{x}) = \{\mathcal{T}(\mathbf{x}) \lesseqgtr \lambda\}$, where $\mathcal{T}(\mathbf{x})$ is the statistic and λ is the threshold. The '$\lesseqgtr$' symbol corresponds to the notion that when $\mathcal{T}(\mathbf{x}) < \lambda$, then $\mathbf{x}$ is declared a background element; and when $\mathcal{T}(\mathbf{x}) > \lambda$, then $\mathbf{x}$ is declared a target.[†] Different values of λ correspond to nested surfaces in $\mathbf{x}$ space.

2.2 Performance measures and ROC curves

Let $p_b(\mathbf{x})$ be the distribution of a measurement $\mathbf{x}$ when it is drawn from the background; and let $p_t(\mathbf{x})$ be the target distribution. The false alarm rate P_{fa} corresponds to the fraction of background pixels which are declared targets: $P_{\text{fa}} = \int \mathcal{B}(\mathbf{x}) p_b(\mathbf{x})\, d\mathbf{x}$. The detection rate is the fraction of actual target pixels which are declared targets: $P_{\text{d}} = \int \mathcal{B}(\mathbf{x}) p_t(\mathbf{x})\, d\mathbf{x}$. In terms of detectors that are implemented as statistics, we can write

$$P_{\text{fa}}(\lambda) = \int_{\mathcal{T}(\mathbf{x})>\lambda} p_b(\mathbf{x})\, d\mathbf{x} \tag{1}$$

$$P_{\text{d}}(\lambda) = \int_{\mathcal{T}(\mathbf{x})>\lambda} p_t(\mathbf{x})\, d\mathbf{x} \tag{2}$$

[†]The case when $\mathcal{T}(\mathbf{x}) = \lambda$ can be dealt with in an elegant and consistent way,[14] but it requires further formalism that does not address the issues of concern in this paper. For the examples considered here, $\mathcal{T}(\mathbf{x}) = \lambda$ corresponds to a measure-zero set of $\mathbf{x}$ values anyway.

where the integration is over the points $\mathbf{x}$ for which $\mathcal{T}(\mathbf{x}) > \lambda$. Now, P_{fa} and P_{d} are functions of λ. By sweeping over a range of values of λ, one can trace out the ROC (receiver operator characteristic) curve of P_{d} as a function of P_{fa}.

For the simple hypothesis testing problem, the optimal detector is given by the likelihood ratio (LR):

$$\mathcal{T}_{\text{LR}}(\mathbf{x}) = \frac{p_t(\mathbf{x})}{p_b(\mathbf{x})} \lessgtr \lambda. \tag{3}$$

That the likelihood ratio is optimal is known as the Neyman-Pearson theorem.[14,15] Note that the log likelihood ratio (LLR)

$$\mathcal{T}_{\text{LLR}}(\mathbf{x}) = \log p_t(\mathbf{x}) - \log p_b(\mathbf{x}) \lessgtr \lambda' = \log(\lambda) \tag{4}$$

is equivalent to Eq. (3), with $\lambda' = \log(\lambda)$ playing the role of a "sweep variable." This logarithmic formulation is often more convenient.

In general, for any monotonic† function h, and any statistic $\mathcal{T}(\mathbf{x})$, the decision rule

$$\mathcal{T}'(\mathbf{x}) = h(\mathcal{T}(\mathbf{x})) \lessgtr h(\lambda) = \lambda' \tag{5}$$

is equivalent to the rule $\mathcal{B}(\mathbf{x}) = \{\mathcal{T}(\mathbf{x}) \lessgtr \lambda\}$. Equivalent statistics exhibit identical ROC curve performance. This is the monotonic invariance property that will be exploited in the reformulation of min-max clairvoyant fusion.

2.3 Composite hypothesis testing

In the composite hypothesis testing problem, we do not have a single target distribution and a single background distribution. Instead, we have families of distributions for the target and/or background. We parameterize these families with $\theta_t \in \Theta_t$ and $\theta_b \in \Theta_b$ respectively. That is, $p_t(\mathbf{x}; \theta_t)$ is the distribution on $\mathbf{x}$ when the target is present, and $p_b(\mathbf{x}; \theta_b)$ is the distribution when the target is absent. The problem is that the parameters, θ_t and/or θ_b, upon which the target and/or background distributions depend, are themselves unknown.

This ambiguity in $\theta = (\theta_t, \theta_b) \in \Theta$ also complicates the measuring of detector performance. The definitions of P_{fa} and P_{d} in Eqs. (1,2) require that p_b and p_t be known. Each value of $\theta \in \Theta$ produces a different ROC curve. In place of a single performance measure, or even a single ROC curve, we have a different performance for each θ.

2.4 Clairvoyant statistic

In the unrealistic case where θ_t and θ_b are both known, one can write the optimal statistic as the simple likelihood ratio:

$$\mathcal{T}(\mathbf{x}; \theta_t, \theta_b) = \frac{p_t(\mathbf{x}; \theta_t)}{p_b(\mathbf{x}; \theta_b)}. \tag{6}$$

This special statistic, which depends on *a priori* knowledge of the parameter values, is sometimes called the "clairvoyant" statistic.[15]

†A scalar function h is monotonic if $z_1 > z_2$ implies $h(z_1) > h(z_2)$.

For composite hypothesis testing, the parameters are by definition *not* known. Nonetheless, the clairvoyant statistic is a useful concept for building detectors that can be used in a composite hypothesis testing problem. Also, some clairvoyant detectors, even though they are optimal only for for a specific θ, may nonetheless exhibit reasonable performance over the whole range of $\theta \in \Theta$.

2.5 Generalized likelihood ratio (GLR) test

The most common approach is known as the generalized likelihood ratio (GLR).[15]

$$\mathcal{T}_{\mathrm{GLR}}(\mathbf{x}) = \frac{\max\limits_{\theta_t \in \Theta_t} p_t(\mathbf{x}; \theta_t)}{\max\limits_{\theta_b \in \Theta_b} p_b(\mathbf{x}; \theta_b)}. \tag{7}$$

One interpretation is that θ_t and θ_b are effectively re-estimated from each data element $\mathbf{x}$ using maximum likelihood:

$$\hat{\theta}_t(\mathbf{x}) = \mathrm{argmax}_{\theta_t \in \Theta_t} p_t(\mathbf{x}; \theta_t) \tag{8}$$

$$\hat{\theta}_b(\mathbf{x}) = \mathrm{argmax}_{\theta_b \in \Theta_b} p_b(\mathbf{x}; \theta_b) \tag{9}$$

and then

$$\mathcal{T}_{\mathrm{GLR}}(\mathbf{x}) = \mathcal{T}(\mathbf{x}; \hat{\theta}_t(\mathbf{x}), \hat{\theta}_b(\mathbf{x})) = \frac{p_t(\mathbf{x}; \hat{\theta}_t(\mathbf{x}))}{p_b(\mathbf{x}; \hat{\theta}_b(\mathbf{x}))}. \tag{10}$$

Thus the GLR is like the clairvoyant statistic except that instead of *knowing* the parameter values θ_t and θ_b, one *estimates* them from the measurement $\mathbf{x}$, using maximum likelihood.

2.6 Min-max clairvoyant fusion

Schaum[1] proposed CF as a generalization of the GLR. To motivate it, let us (following later treatments by Schaum[3–5] and Bajorski[10–12]) write the GLR as a min-max operation over the clairvoyant statistics:

$$\mathcal{T}_{\mathrm{GLR}}(\mathbf{x}) = \frac{\max\limits_{\theta_t \in \Theta_t} p_t(\mathbf{x}; \theta_t)}{\max\limits_{\theta_b \in \Theta_b} p_b(\mathbf{x}; \theta_b)} \tag{11}$$

$$= \min_{\theta_b \in \Theta_b} \max_{\theta_t \in \Theta_t} \frac{p_t(\mathbf{x}; \theta_t)}{p_b(\mathbf{x}; \theta_b)} = \min_{\theta_b \in \Theta_b} \max_{\theta_t \in \Theta_t} \mathcal{T}(\mathbf{x}; \theta_t, \theta_b). \tag{12}$$

The GLR *decision rule* compares this statistic to a threshold: $\mathcal{B}_{\mathrm{GLR}}(\mathbf{x}) = \{\mathcal{T}_{\mathrm{GLR}}(\mathbf{x}) \lessgtr \lambda\}$. The CF idea is to make λ depend on the parameters θ_t and θ_b, but this requires that the threshold be brought "inside" the min-max operator. Thus, instead of

$$\mathcal{B}_{\mathrm{GLR}}(\mathbf{x}) = \left\{ \min_{\theta_b \in \Theta_b} \max_{\theta_t \in \Theta_t} \mathcal{T}(\mathbf{x}; \theta_t, \theta_b) \lessgtr \lambda \right\}, \tag{13}$$

we write

$$\mathcal{B}_{\mathrm{CF}}(\mathbf{x}) = \left\{ \min_{\theta_b \in \Theta_b} \max_{\theta_t \in \Theta_t} \left[\mathcal{T}(\mathbf{x}; \theta_t, \theta_b) - \lambda(\theta_t, \theta_b) \right] \lessgtr 0 \right\}, \tag{14}$$

or, equivalently,

$$\mathcal{B}_{\mathrm{CF}}(\mathbf{x}) = \left\{ \min_{\theta_b \in \Theta_b} \max_{\theta_t \in \Theta_t} \left[\mathcal{T}(\mathbf{x}; \theta_t, \theta_b) / \lambda(\theta_t, \theta_b) \right] \lessgtr 1 \right\}. \tag{15}$$

It is important to observe that this manipulated-threshold formulation of CF provides a CF *decision rule*, not a CF *statistic*. To compute ROC curves for instance, one must produce many decision rules.

The "art" of CF lies in the choice of $\lambda(\theta_t, \theta_b)$; different strategies for making this choice are called *flavors*.[1] Choosing λ that is constant with respect to θ_t and θ_b recovers the GLR, but the following two other flavors have also been advocated:

2.6.1 CF-cfar flavor

One of the continuum fusion flavors that is worth calling out specifically is CFAR (constant false alarm rate) fusion.[†] Here $\lambda(\theta_t, \theta_b)$ is chosen so that the associated clairvoyant statistic has a constant false alarm rate. Specifically,

$$\lambda(\theta_t, \theta_b, \alpha) = \min_{\lambda} \lambda \ \text{ s.t.} \int_{\mathcal{T}(\mathbf{x}; \theta_t, \theta_b) > \lambda} p_b(\mathbf{x}, \theta_b) \, d\mathbf{x} \leq \alpha. \tag{16}$$

Given this expression for λ, the binary decision rule takes the form

$$\mathcal{B}_{\mathrm{CF}}(\mathbf{x}, \alpha) = \left\{ \min_{\theta_b \in \Theta_b} \max_{\theta_t \in \Theta_t} \mathcal{T}(\mathbf{x}, \theta_t, \theta_b) / \lambda(\theta_t, \theta_b, \alpha) \lessgtr 1 \right\} \tag{17}$$

For many distributions, Eq. (16) is a difficult implicit integral criterion, though Schaum has shown how this integral condition can be re-expressed in terms of partial differential equations.[1] And in some cases, simple analytical expressions can be found.[1]

It bears remarking that this formulation of CF does provide a family of decision rules (parameterized by α), but this family of decision rules cannot be described in terms of a single statistic and a variable threshold. Further, although the individual decision rules have been calibrated to have individual false alarm rates of α, the false alarm rate for the fused decision rule will be larger than that. To produce a CF decision rule with a specified false alarm rate, it is generally a matter of trial and error to find the α that leads to this specified rate.

2.6.2 CF-cpd flavor

The natural counterpoint to the CF-cfar flavor is one based on P_{d} instead of P_{fa}. In this case,

$$\lambda(\theta_t, \theta_b, \beta) = \max_{\lambda} \lambda \ \text{ s.t.} \int_{\mathcal{T}(\mathbf{x}; \theta_t, \theta_b) > \lambda} p_t(\mathbf{x}, \theta_t) \, d\mathbf{x} \geq \beta \tag{18}$$

ensures that the individual decision rules (the fusees) all have a detection rate of β.

[†]Schaum writes:[4] "It appears that one flavor of CF fills a particular theoretical niche. In CH [composite hypothesis] problems that admit an optimal (UMP [uniformly most powerful]) solution, one particular flavor (CFAR) appears to always generate it. This is a coveted property of any method of solving CH problems, a property lacking in the GLRT."

3. MONOTONIC RECALIBRATION OF CLAIRVOYANT STATISTICS

In this section, a more general formulation for CF is presented. To motivate this new formulation, consider Eq. (15), and observe that just as $\mathcal{T}(\mathbf{x};\theta_t,\theta_b)$ is a clairvoyant statistic, so is $\mathcal{T}(\mathbf{x};\theta_t,\theta_b)/\lambda(\theta_t,\theta_b)$ for any $\lambda(\theta_t,\theta_b)>0$. Indeed, one can think of this as a θ-dependent recalibration of the clairvoyant statistic. But for any scalar function $h(z,\theta_t,\theta_b)$ that is monotonic in its first argument, it is clear that $\mathcal{T}^*(\mathbf{x};\theta_t,\theta_b)=h(\mathcal{T}(\mathbf{x};\theta_t,\theta_b),\theta_t,\theta_b)$ is also a clairvoyant statistic. One can then fuse this family of monotonically recalibrated clairvoyant statistics to obtain:

$$\mathcal{T}_{\mathrm{CF}}(\mathbf{x}) = \min_{\theta_b\in\Theta_b}\max_{\theta_t\in\Theta_t}\mathcal{T}^*(\mathbf{x};\theta_t,\theta_b) \tag{19}$$

$$= \min_{\theta_b\in\Theta_b}\max_{\theta_t\in\Theta_t} h\left(\mathcal{T}(\mathbf{x};\theta_t,\theta_b),\theta_t,\theta_b\right). \tag{20}$$

What this formulation of CF provides is a single statistic, the $\mathcal{T}_{\mathrm{CF}}(\mathbf{x})$ defined in Eq. (20), that can be used to detect targets based on measured pixel value $\mathbf{x}$. It bears remarking, however, that although the new formulation can produce new *statistics*, it does not produce any new *decision rules*. In fact, from the point of view of individual decision rules, the two formulations are equivalent. This will be clarified with two formal statements: the first asserts that any decision rule obtained from the original formulation in Eq. (15) can be produced by the new formalism, and the second asserts the converse.

Statement 1: Given any positive function $\lambda(\theta_t,\theta_b)$, there exists a monotonic function $h(z,\theta_t,\theta_b)$ and a threshold h_o such that the decision rule $\mathcal{B}_{\mathrm{CF}}(\mathbf{x})$ defined in Eq. (15) can be obtained in terms of the statistic defined in Eq. (20) as $\{\mathcal{T}_{\mathrm{CF}}(\mathbf{x})\lesseqgtr h_o\}$.

To see this, take $h(z,\theta_t,\theta_b)=z/\lambda(\theta_t,\theta_b)$ and $h_o=1$. Then:

$$\{\mathcal{T}_{\mathrm{CF}}(\mathbf{x})\lesseqgtr h_o\} = \left\{\min_{\theta_b\in\Theta_b}\max_{\theta_t\in\Theta_t}\left[h(\mathcal{T}(\mathbf{x};\theta_t,\theta_b),\theta_t,\theta_b)\right]\lesseqgtr h_o\right\} \tag{21}$$

$$= \left\{\min_{\theta_b\in\Theta_b}\max_{\theta_t\in\Theta_t}\left[\mathcal{T}(\mathbf{x};\theta_t,\theta_b)/\lambda(\theta_t,\theta_b)\right]\lesseqgtr 1\right\} \tag{22}$$

$$= \mathcal{B}_{\mathrm{CF}}(\mathbf{x}). \tag{23}$$

Here, Eq. (21) follows from the definition of $\mathcal{T}_{\mathrm{CF}}(\mathbf{x})$ in Eq. (20); Eq. (22) follows from Eq. (21) by substitution of the assignments made to $h(z,\theta_t,\theta_b)$ and h_o; and Eq. (23) follows from Eq. (22) and the definition of $\mathcal{B}_{\mathrm{CF}}(\mathbf{x})$ in Eq. (15).

Statement 2: For any choice of monotonic function $h(z,\theta_t,\theta_b)$ and scalar threshold h_o, there exists a function $\lambda(\theta_t,\theta_b)$ such that the decision rule given by $\{\mathcal{T}_{\mathrm{CF}}(\mathbf{x})\lesseqgtr h_o\}$ is the same as $\mathcal{B}_{\mathrm{CF}}(\mathbf{x})$.

To see this, first observe that since h is a monotonic function, we can define $h^{-1}(y,\theta_t,\theta_b)$ to be the value z for which $h(z,\theta_t,\theta_b)=y$. Then, observe that for any values of θ_t, θ_b, and h_o, the following three expressions are equivalent:

$$h\left(\mathcal{T}(\mathbf{x};\theta_t,\theta_b),\theta_t,\theta_b\right)\lesseqgtr h_o, \tag{24}$$

$$\mathcal{T}(\mathbf{x};\theta_t,\theta_b)\lesseqgtr h^{-1}(h_o,\theta_t,\theta_b), \tag{25}$$

$$\mathcal{T}(\mathbf{x};\theta_t,\theta_b)/h^{-1}(h_o,\theta_t,\theta_b)\lesseqgtr 1. \tag{26}$$

It follows that

$$\min_{\theta_b \in \Theta_b} \max_{\theta_t \in \Theta_t} h\left(\mathcal{T}(\mathbf{x}; \theta_t, \theta_b), \theta_t, \theta_b\right) \lessgtr h_o \tag{27}$$

is equivalent to

$$\min_{\theta_b \in \Theta_b} \max_{\theta_t \in \Theta_t} \mathcal{T}(\mathbf{x}; \theta_t, \theta_b) / h^{-1}(h_o, \theta_t, \theta_b) \lessgtr 1, \tag{28}$$

which tells us that by using $\lambda(\theta_t, \theta_b) = h^{-1}(h_o, \theta_t, \theta_b)$ as the adjustable threshold function in Eq. (15), we obtain the decision rule $\{\mathcal{T}_{\mathrm{CF}}(\mathbf{x}) \lessgtr h_o\}$.

4. FAST ALGORITHMS FOR TWO FUSION FLAVORS

What the formulation in Section 3 provides is a way to think about detector fusion in terms of a min-max over clairvoyant statistics that have been recalibrated using the monotonic function h. For the CF-cfar flavor, we recalibrate so that $\mathcal{T}^*(\mathbf{x}; \theta_t, \theta_b)$ indicates the false alarm rate; for CF-cpd $\mathcal{T}^*(\mathbf{x}; \theta_t, \theta_b)$ corresponds to detection rate. This formulation provides two advantages.

One advantage is that the decision rules will be of the form $\mathcal{T}(\mathbf{x}) \lessgtr \lambda$. In some cases, the original manipulated-threshold formulation of CF can lead to detectors of this form, but that is not generally the case for CF-cfar or CF-cpd. For CF-cfar, as seen in Eq. (17), the original formulation gives a decision rule of the form $\mathcal{T}(\mathbf{x}, \alpha) \lessgtr 1$. Similarly, for CF-cpd, we would obtain a decision rule of the form $\mathcal{T}(\mathbf{x}, \beta) \lessgtr 1$.

A second advantage of this formulation is that it leads to efficient implementations for the CF-cfar and CF-cpd flavors. These fast algorithms apply to the case in which the target parameters are unknown, but the background parameters are *known*. This is usually the case in hyperspectral imagery, since there are typically many pixels from which to estimate the background (and with which to fit parameters that might otherwise be unknown), but the target estimation is done separately for each pixel.

4.1 CF-cfar

For CF-cfar, we want to recalibrate all of the clairvoyant statistics so that, for a given threshold, they all have the same false alarm rate. A simple way to achieve this is to individually recalibrate each statistic so that $\mathcal{T}^*(\mathbf{x}, \theta) = h(\mathcal{T}(\mathbf{x}, \theta), \theta)$ is equal to one minus the false alarm rate. That is: the decision rule $\{\mathcal{T}^*(\mathbf{x}, \theta) \lessgtr \lambda\}$ has a false alarm rate equal to $1 - \lambda$. This approach is described in detail in Algorithm 1. In this implementation, w_{ji} corresponds the jth recalibrated clairvoyant statistic evaluated at $\mathbf{x}_i$; that is, $w_{ji} = \mathcal{T}^*(\mathbf{x}_i, \theta_j)$. In terms of the uncalibrated clairvoyant statistic $\mathcal{T}(\mathbf{x}, \theta)$, this false alarm rate is given by the fraction of $\mathbf{x}_i$ values for which $\mathcal{T}(\mathbf{x}_i, \theta) > \mathcal{T}(\mathbf{x}, \theta)$. Thus, $w_{ji} = \#\{i' \,:\, v_{ji'} > v_{ji}\}/n$, where $v_{ji} = \mathcal{T}(\mathbf{x}_i, \theta_j)$ and n is the number of samples in the base data set. This can be rapidly computed by first sorting the values of v_{ji} (with j fixed, and the i index sorted over), and then assigning w_{ji} to the rank of v_i in that list. Since this is done for each of k values of θ, the total computation complexity is $O(kn \log n)$.

As written, Algorithm 1 can be used to search for rare targets. The input is all the pixels in the image, and the output is a score u_i associated with each pixel $\mathbf{x}_i$. Those pixels with the largest values are the most likely to be targets.

Algorithm 1 CF-cfar flavor of min-max clairvoyant fusion; for the case when background is known and target parameter θ_t is unknown

Require: Clairvoyant statistic $\mathcal{T}(\mathbf{x}, \theta_t)$
Require: Finite representative set $\{\theta_1, \theta_2, \ldots, \theta_k\} \subset \Theta_t$
Require: Base data set $\mathbf{x}_1, \ldots, \mathbf{x}_n$ known (or assumed) to be non-target pixels
  for each θ_j in $\{\theta_1, \ldots, \theta_k\}$ **do**
    Let $v_{ji} = \mathcal{T}(\mathbf{x}_i, \theta_j)$ for $i = 1 \ldots n$
    Let $w_{ji} = \mathrm{Rank}(v_{ji})/n$ for $i = 1 \ldots n$
  end for
  Let $u_i = \max_j w_{ji}$ for $i = 1 \ldots n$
Output: $u_1, \ldots, u_n$ corresponding to the CF-cfar statistic evaluated at $\mathbf{x}_1, \ldots, \mathbf{x}_n$.

It is also possible to apply this algorithm "offline" to new data points $\mathbf{x}$ that were not in the original base dataset. The details are described in Algorithm 2. The w_j in this algorithm corresponds to the recalibration of the jth clairvoyant statistic $\mathcal{T}^*(\mathbf{x}, \theta_j)$. The maximum of the w_j values corresponds to the maximum in Eq. (20).

Algorithm 2 Offline application of CF-cfar

Require: Input value $\mathbf{x}$
Require: Clairvoyant statistic $\mathcal{T}(\mathbf{x}; \theta_t)$
Require: Finite representative set $\{\theta_1, \theta_2, \ldots, \theta_k\} \subset \Theta_t$
Require: Values v_{ji} and w_{ji} for $j = 1 \ldots k$ and $i = 1 \ldots n$; these will have been computed in Algorithm 1.
  for each θ_j in $\{\theta_1, \ldots, \theta_k\}$ **do**
    Let $v = \mathcal{T}(\mathbf{x}, \theta_j)$
    Let $i_1 = \mathrm{argmax}_i\ v_{ji}$ s.t. $v_{ji} \leq v$
    Let $i_2 = \mathrm{argmin}_i\ v_{ji}$ s.t. $v \leq v_{ji}$
    Let $p = (v - v_{ji_1})/(v_{ji_2} - v_{ji_1})$
    {Note that i_1, i_2, p satisfy: $v = (1-p)v_{ji_1} + pv_{ji_2}$.}
    Let $w_j = (1-p)w_{ji_1} + pw_{ji_2}$
  end for
  Let $u = \max_j w_j$.
Output: u is the CF-cfar statistic for $\mathbf{x}$

4.2 CF-cpd

For the CF-cpd flavor, the situation is complicated by the fact that the target distribution $p_t(\mathbf{x})$ depends on the target strength θ_t, so the fusion step is not as straightforward. In particular, besides the base data set $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ that was needed for the CF-cfar flavor, a target data set is also needed. Specifically, $\mathbf{y}_{ij}$ is the measurement (*e.g.,* radiance spectrum) that would be observed at the ith pixel if the target strength in that pixel were given by θ_j. For many problems of interest, $\mathbf{y}_{ij}$ is a simple function of the target-free pixel value $\mathbf{x}_i$ and the target strength θ_j;

for instance, in the additive target model described in the next section, $\mathbf{y}_{ij} = \mathbf{x}_i + \theta_j \mathbf{s}$, where $\mathbf{s}$ is the target signature.

For any given θ_t, we want to recalibrate the clairvoyant statistic so that it corresponds to the detection rate P_d; specifically, we will construct h so that the decision rule $\{\mathcal{T}^*(\mathbf{x}, \theta) \lesseqgtr \lambda\}$ will have a detection rate $P_\mathrm{d} = 1 - \lambda$. That is, the recalibrated $\mathcal{T}^*(\mathbf{x}, \theta_j)$ is assigned to the fraction of points for which $\mathcal{T}(\mathbf{y}_{ij}, \theta_j) < \mathcal{T}(\mathbf{x}, \theta_j)$. As with the CF-cfar case, we use w_{ji} to correspond to the recalibrated statistic: $\mathcal{T}^*(\mathbf{x}_i, \theta_j)$. The details are given in Algorithm 3.

Algorithm 3 CF-cpd flavor of min-max clairvoyant fusion

Require: Clairvoyant statistic $\mathcal{T}(\mathbf{x}; \theta_t)$
Require: Finite representative set $\{\theta_1, \theta_2, \ldots, \theta_k\} \subset \Theta_t$
Require: Base data set $\mathbf{x}_1, \ldots, \mathbf{x}_n$
Require: Target data set $\mathbf{y}_{ij}$ for $i = 1, \ldots, n$ and $j = 1, \ldots, k$
  for each θ_j in $\{\theta_1, \ldots, \theta_k\}$ **do**
    Let $v_i = \mathcal{T}(\mathbf{x}_i, \theta_j)$ for $i = 1 \ldots n$
    Let $v_i^+ = \mathcal{T}(\mathbf{y}_{ij}, \theta_j)$ for $i = 1 \ldots n$
    Let $w_{ji} = \#\{i' \ : \ v_{i'}^+ < v_i\}/n$ for $i = 1 \ldots n$
  end for
  Let $u_i = \max_j w_{ji}$ for $i = 1 \ldots n$
Output: $u_1, \ldots, u_n$ corresponding to the CF-cpd statistic evaluated at $\mathbf{x}_1, \ldots, \mathbf{x}_n$.

By sorting the v_i and v_i^+ arrays, which takes $O(n \log n)$ time, one can perform the assignment to w_{ji} for all values of i in $O(n)$ time. Thus, the full algorithm can be run in $O(kn \log n)$ time.

5. ADDITIVE TARGET SIGNAL

Although CF has been demonstrated on a variety of problems,[1–13] that variety has been restricted to those problems for which a CF solution (in particular, CF-cfar and in some cases CF-cpd) could be derived analytically. Armed with the algorithms in the previous section, that list of problems will be extended to include a class with additive targets on various background distributions.

In the additive target model, the background is given by a known distribution $p_b(\mathbf{x})$, and to that background, a target signal $t\mathbf{s}$ has been added. Here the target "signature" $\mathbf{s}$ is known, but the target strength is given by the parameter $\theta_t = t$, and is unknown. The null hypothesis is that $t = 0$, and the (one-sided) alternative is that $t > 0$.

$$\mathcal{H}_o : \mathbf{x}_b \sim p_b(\mathbf{x}) \tag{29}$$

$$\mathcal{H}_1 : \mathbf{x}_t = \mathbf{x}_b + t\mathbf{s} \sim p_t(\mathbf{x}) = p_b(\mathbf{x} - t\mathbf{s}) \tag{30}$$

The additive target model has proved useful in particular for plume detection in hyperspectral imagery.[16–19]

5.1 Gaussian background

An instructive special case is the Gaussian:

$$p_b(\mathbf{x}) = (2\pi)^{-d/2}|R|^{-1/2} \exp\left[-(\mathbf{x}-\boldsymbol{\mu})^T R^{-1}(\mathbf{x}-\boldsymbol{\mu})/2\right] \tag{31}$$

where $\boldsymbol{\mu}$ is the mean, R is the covariance, and d is the number of spectral channels. We can invoke a linear whitening transform of the data,

$$\mathbf{x} \leftarrow R^{-1/2}(\mathbf{x}-\boldsymbol{\mu}), \tag{32}$$

$$t \leftarrow t\sqrt{\mathbf{s}^T R^{-1}\mathbf{s}}, \tag{33}$$

$$\mathbf{s} \leftarrow R^{-1/2}\mathbf{s}/\sqrt{\mathbf{s}^T R^{-1}\mathbf{s}}. \tag{34}$$

In these coordinates, we have $\langle \mathbf{x} \rangle = 0$ and $\left\langle \mathbf{x}\mathbf{x}^T \right\rangle = I$. We also have a normalized signature $\mathbf{s}^T\mathbf{s} = 1$, and t measures signal strength as a dimensionless quantity. And in these coordinates, we can rewrite Eq. (31)

$$p_b(\mathbf{x}) = (2\pi)^{-d/2} \exp\left[-\mathbf{x}^T\mathbf{x}/2\right]. \tag{35}$$

The clairvoyant detector is given by the the likelihood ratio

$$\mathcal{T}(\mathbf{x},t) = \frac{p_t(\mathbf{x},t)}{p_b(\mathbf{x})} = \exp\left[-(\mathbf{x}-t\mathbf{s})^T(\mathbf{x}-t\mathbf{s})/2 + \mathbf{x}^T\mathbf{x}/2\right] = \exp\left[t\mathbf{s}^T\mathbf{x} - t^2/2\right] \lessgtr \lambda \tag{36}$$

which is equivalent to the expression

$$\mathcal{T}^*(\mathbf{x},t) = \mathbf{s}^T\mathbf{x} \lessgtr \frac{1}{t}\log\lambda + \frac{t}{2} = \lambda^*. \tag{37}$$

Here, the test statistic itself ($\mathbf{s}^T\mathbf{x}$) does not depend on the signal strength t. Further, as Kay notes in reference to Equation (6.3) in his book,[15] the dependence of the threshold on t is "illusory."[†] The threshold λ^* that achieves a given false alarm rate, for instance, does not depend on t. This statistic is called the *adaptive* matched filter (AMF)[20,21] since in the original (unwhitened) coordinates, it is given by $\mathcal{T}(\mathbf{x}) = \mathbf{s}^T R^{-1}(\mathbf{x}-\boldsymbol{\mu})$, which is adapted to the mean $\boldsymbol{\mu}$ and covariance R of the original data.

This problem – additive target with Gaussian background – is one of the few composite hypothesis testing problems that admits a uniformly most powerful (UMP) solution; and the AMF is the UMP solution. For $P_{\text{fa}} > 0.5$, however, the GLR fails to reproduce the AMF.[‡] However, the CF-cfar flavor works for all values of P_{fa} and for this reason is favored.[4,8]

A number of departures from this ideal of additive target with Gaussian background are discussed in Ref. [22], but one in particular is to consider distributions $p_b(\mathbf{x})$ that are not

[†]See Chapter 3.4 in Lehmann and Romano[14] for a more formal treatment, which speaks of families of distributions $p(\mathbf{x},t)$ having a "monotone likelihood ratio" when $p(\mathbf{x},t')/p(\mathbf{x},t)$ depends monotonically on a function $\mathcal{T}(\mathbf{x})$ whenever $t' > t$. In this case, $\mathcal{T}(\mathbf{x}) = \mathbf{s}^T\mathbf{x}$.

[‡]Many authors get around this by taking the sign of the square root to be the sign of $\mathbf{s}^T\mathbf{x}$, but this is not justified by the GLR formulation.

Gaussian. A natural extension that does not require the introduction of a lot of new parameters is the elliptically-contoured distribution.[23–28] We still perform the whitening transformation in Eq. (32), but we end up with two new distributions: multivariate-t and Laplacian. Both distributions exhibit fatter-than-Gaussian tails, which is often observed in multispectral data. The multivariate-t has the flexibility of an adjustable parameter ν, and is asymptotically fatter than the Laplacian. On the other hand, the Laplacian is (sometimes) algebraically simpler than the multivariate-t and does not require the estimation of a parameter ν.

5.2 Multivariate-t distribution

The multivariate-t distribution[25–27] is given by

$$p(\mathbf{x}, \nu) = c \left[1 + \frac{\mathbf{x}^T \mathbf{x}}{\nu - 2} \right]^{-(d+\nu)/2}, \tag{38}$$

where c is the normalizing constant, d is the number of spectral channels, and ν is a scalar parameter that characterizes the distribution. Smaller ν corresponds to fatter tails; and as $\nu \to \infty$, the distribution becomes Gaussian. (Also, recall that $\mathbf{x}$ has been whitened and has unit covariance.)

The clairvoyant statistic for the multivariate-t distribution is given by:

$$\mathcal{T}(\mathbf{x}, t) = \frac{2t\mathbf{s}^T\mathbf{x} - t^2}{1 + \mathbf{x}^T\mathbf{x}/(\nu - 2)}. \tag{39}$$

In the matched-filter-residual (MFR) projection,[22,29] we write $\mathbf{x}$ in terms of its matched filter component, and a residual component that corresponds to the magnitude of the signal with the matched filter projected out:

$$\mathrm{MF} = \mathbf{s}^T\mathbf{x} \tag{40}$$

$$\mathrm{R} = \sqrt{\mathbf{x}^T\mathbf{x} - (\mathbf{s}^T\mathbf{x})^2} \tag{41}$$

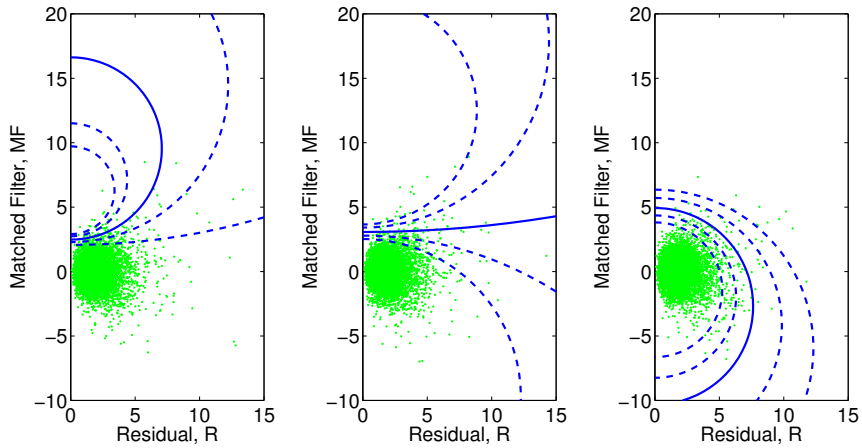
Note that MF and R have the same "units" as $\mathbf{x}$, and that $(\mathrm{MF})^2 + (\mathrm{R})^2 = \mathbf{x}^T\mathbf{x}$. We remark that this is a kind of "folded space" discussed in Ref. [4]. In Fig. 1, the clairvoyant decision rule of Eq. (39) is seen to be a semicircle in MFR space.

The GLR solution for the multivariate-t distribution is given by[31]

$$\mathcal{T}(\mathbf{x}) = \frac{\mathbf{s}^T\mathbf{x}}{\sqrt{1 + \mathbf{x}^T\mathbf{x}/(\nu - 2)}} \tag{42}$$

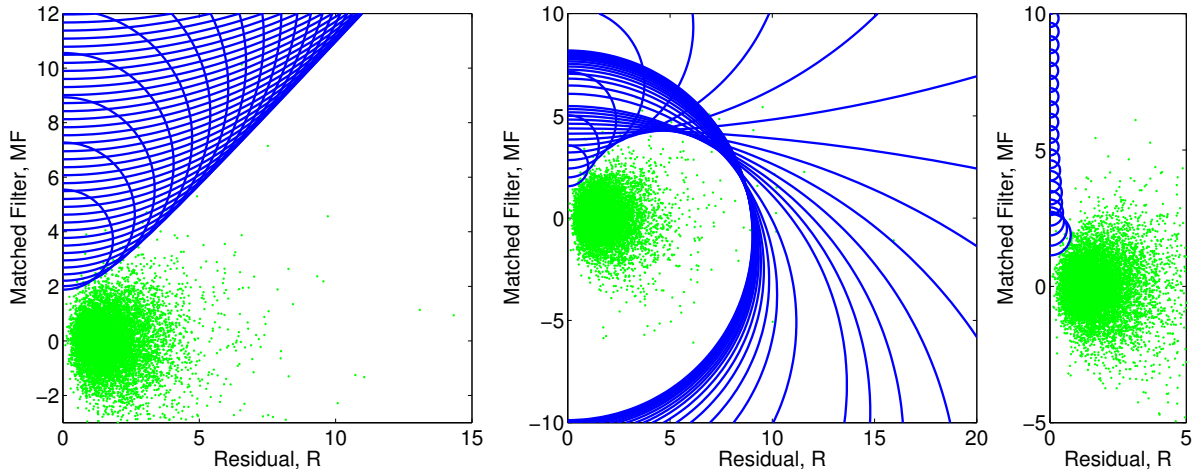
which lies "between" the adaptive matched filter (AMF)[20,21] and the adaptive coherence estimator (ACE).[32,33] In particular, the $\nu \to \infty$ limit, which corresponds to a Gaussian distribution, leads to the AMF detector; and the $\nu \to 2$ limit, which corresponds to a very heavy-tailed distribution, leads to the ACE detector.[†]

[†]By an odd coincidence, this detector has the same form as the Kelly detector,[34] but with $(\nu - 2)$ replacing the number n of data samples. The two are qualitatively different: in addition to being derived under different assumptions (multivariate-t versus Gaussian), the size of n is generally much larger than $(\nu - 2)$.



(a) $t = 4$ (b) $t = 6$ (c) $t = 20$

Figure 1. Clairvoyant detectors plotted in MFR coordinates for the additive target model, with the background given by the multivariate-t distribution with $d = 5$ spectral channels and parameter value $\nu = 5$. Each panel shows a single clairvoyant detector, optimized for a specific signal strength t, with the threshold adjusted so that the false alarm rates are: 0.02 (dashed), 0.01 (dashed), 0.005 (solid), 0.002 (dashed), and 0.001 (dashed). In all cases, the detector is a family of semi-circles, though the the semicircles may curve upward or downward (or, at the transition between those two, they may be horizontal lines, same as the adaptive matched filter). For the upward curving semicircles (*e.g.,* in panel (a)), the detections are the points inside the circles; for the downward curving semicircles (*e.g.,* in panel (c)) the detections are outside the circles. The circles are the optimal boundaries between the non-target pixels (indicated as dots on these plots), and the target pixels (which are not shown, but would have a similar distribution as the plotted dots, except translated upward by an amount t). Looking specifically at panel (a), although this detector is optimal for $t = 4$, it has the unusual property that it would fail to detect a pixel with very high signal strength; *e.g.,* a pixel of strength 20 would be outside the $P_{\text{fa}} = 0.005$ circle that was optimized for detecting pixels with $t = 4$. In panel (c), we see that for large signal $t = 20$, the detector will continue to detect even stronger detections; but this detector exhibits a different unusual property: pixels with large *negative* matched filter values will be treated as detections. The $t \to \infty$ limit of this clairvoyant detector approaches the classic RX detector.[30] Note the computation of threshold required to observe specified false alarm rates was based on a random sample of 10^6 points, but for clarity the figure only shows 10^4 points.



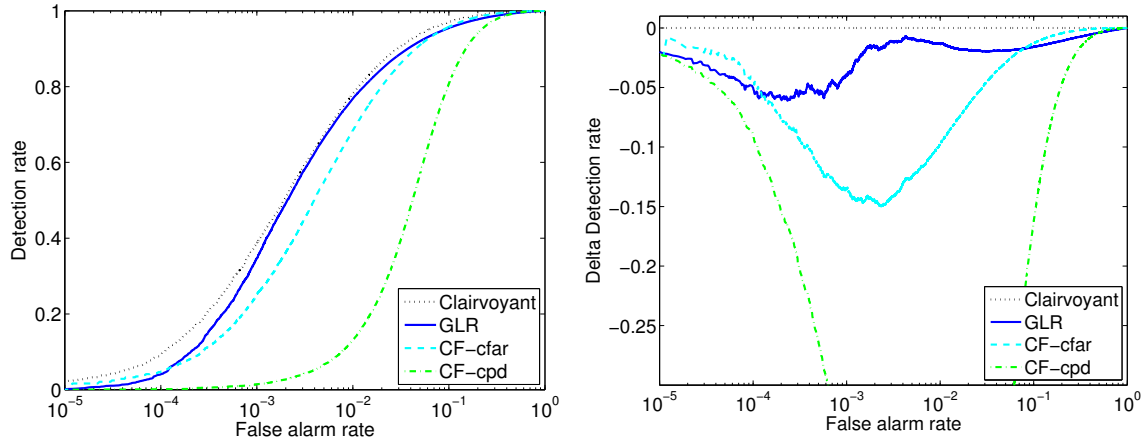
(a) GLR (b) CF-cfar (c) CF-cpd

Figure 2. The clairvoyant detectors in Fig. 1 are fused to produce GLR, CF-cfar, and CF-cpd detectors; in all three cases, parameters are adjusted so that $P_{\text{fa}} = 0.005$. (a) For the GLR, the clairvoyant detectors (plotted here with $t = 0.5, 1.0, 1.5, \ldots, 20$) have thresholds chosen so that they all have the same likelihood ratio; in the continuum limit, this fusion of semicircles leads to a hyperbolic curve. (b) For CF-cfar, the clairvoyant detectors (plotted here with $t = 1, 2, 3, \ldots, 15$ and $t = 20, 25, 30, \ldots, 100$) all have the same individual false alarm rate; in this case $\alpha = 0.0013$. (c) For CF-cpd, the clairvoyant detectors (plotted here with $t = 0, 1, 2, \ldots, 10$) all have the same detection rate; in this case $\beta = 0.0011$.

Fig. 2 illustrates decision rules obtained as a fusion of clairvoyant decision rules using GLR, CF-cfar, and CF-cpd fusion. The performance of these three detectors are shown for a particular signal strength ($t = 3$) in Fig. 3, and compared to the clairvoyant detector, which of course outperforms them all. Here the ROC curves show that CF-cfar is somewhat better in the very low false alarm rate regime and the very high detection rate regime, whereas the GLR is better over an intermediate range of false alarm rates. For this problem, CF-cpd does not exhibit much utility.

Although we do not have the luxury of an analytic closed-form solution for the statistic $\mathcal{T}(\mathbf{x})$ that corresponds to CF-cfar (or to CF-cpd) for this problem, we can efficiently compute $\mathcal{T}(\mathbf{x})$ for every point in the dataset using Algorithm 1 (or Algorithm 3), and this enables us to efficiently compute the ROC curves shown in Fig. 3.

We can also compute performance as a function of signal strength t. The plots in Fig. 4 and Fig. 5 compare the "power" (*i.e.*, detection rate at $P_{\text{fa}} = 0.005$) of the different fusion detectors. The CF-cpd actually does best at very low signal strength, the CF-cfar does best at very high signal strength, and GLR shows the best performance over intermediate ranges. One can further observe that the GLR always exhibits *increasing* detection rate with *increasing* target strength, even though the individual clairvoyant statistics do not. CF-cfar also exhibits this property, whereas CF-cpd does not.



(a) (b)

Figure 3. (a) ROC curves for three different flavors of min-max clairvoyant fusion, where the target strength is given by $t = 3$, and the background distribution is multivariate-t with $d = 5$ and $\nu = 5$. (b) To accentuate the differences, this plot shows the change in detection rate compared to the clairvoyant detector. Both experiments are based on a simulated dataset with $n = 10^6$ points.

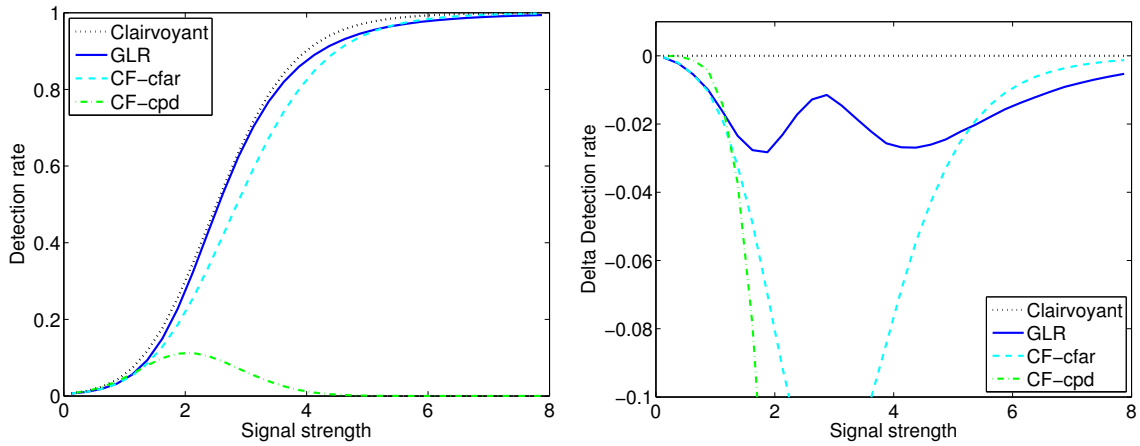


Figure 4. Power plot for three different flavors of min-max clairvoyant fusion detectors. In this case, the distribution is multivariate-t, with $d = 5$ and $\nu = 5$. The left panel is detection rate versus dimensionless signal strength, and the right panel displays the same data, but with the optimal clairvoyant performance subtracted. The detection rate is computed at a threshold value that gives $P_{\text{fa}} = 0.005$ for all detectors, and the runs are based on simulated imagery with $n = 10^6$ pixels. The plots show that the GLR detector is a useful general-purpose detector with good performance over the whole range of signal strengths. The CF-cfar detector exhibits poorer performance at small signal strengths, but as the signal strength increases, CF-cfar outperforms GLR. The CF-cpd detector exhibits generally poor performance overall, but at very small signal strength, it is the best of the three.

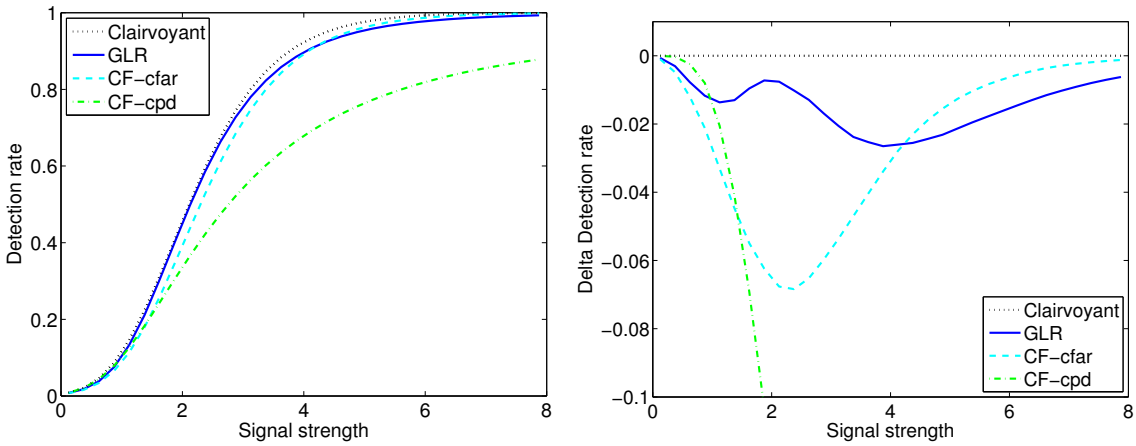


Figure 5. Same as Fig. 4, but with $d = 128$ instead of $d = 5$, which is in the hyperspectral regime. As seen in the $d = 5$ case, GLR behaved well overall and was best for intermediate signal strength, but CF-cfar performed somewhat better for strong signals and CF-cpd did somewhat better for very weak signals.

5.3 Laplacian distribution

Another elliptically-contoured distribution is the Laplacian:[28]

$$p(\mathbf{x}) = c \exp\left[-\left(\mathbf{x}^T\mathbf{x}\right)^{1/2}\right]. \tag{43}$$

For the Laplacian distribution, the clairvoyant statistic is given by[28]

$$\mathcal{T}(\mathbf{x}, t) = \sqrt{\mathbf{x}^T\mathbf{x}} - \sqrt{(\mathbf{x} - t\mathbf{s})^T(\mathbf{x} - t\mathbf{s})}. \tag{44}$$

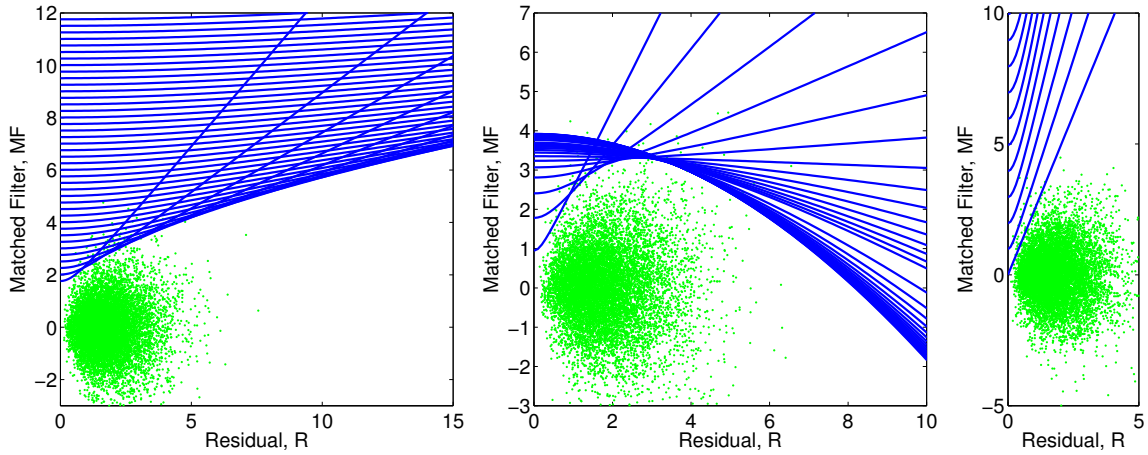
Whereas the clairvoyant detectors for the multivariate-t are semicircles in MFR space, they are hyperbolic for the Laplacian, as seen in Fig. 6. They do not, for instance, exhibit the counter-intuitive property of *decreasing* detection rate with *increasing* target strength that was seen in the multivariate-t case.

The GLR solution for the Laplacian is given by[7]

$$\mathcal{T}(\mathbf{x}) = \mathrm{sign}(\mathbf{s}^T\mathbf{x})\left[\sqrt{\mathbf{x}^T\mathbf{x}} - \sqrt{\mathbf{x}^T\mathbf{x} - (\mathbf{s}^T\mathbf{x})^2}\right]. \tag{45}$$

Fig. 6(a) illustrates the GLR detector as a fusion of clairvoyant detectors.

Fig. 6(b,c) illustrates the CF-cfar and CF-cpd fusion detectors. As with the multivariate-t, we do not have simple closed-form solutions for these detectors, but we can obtain them numerically using the algorithms in Section 4. The performance of these detectors is compared to the GLR and to the ideal (and unattainable) clairvoyant detector in Fig. 7. Similar to what was seen for the multivariate-t in Fig. 4 and Fig. 5, the detector performance depends on signal strength: CF-cpd is best for very small signals, CF-cfar is best for large signals, and GLR did well over the full range of signal strengths.



(a) GLR (b) CF-cfar (c) CF-cpd

Figure 6. Similar to Fig. 2, but using the Laplacian instead of the multivariate-t as the non-Gaussian background distribution. Here the clairvoyant detectors, defined by Eq. (44), are hyperbolas. (a) For the GLR, the detectors have a common likelihood ratio; this fusion of hyperbolas leads to a parabolic curve. (b) For the CF-cfar detector, the clairvoyant detectors have thresholds chosen so that the individual false alarm rates are the same; in this case $\alpha = 0.0021$. (c) The CF-cpd detector is based on the fusion of detectors with the same detection rate; the result is equivalent to the $t = 0$ clairvoyant detector, which is equivalent to the ACE detector. In all three cases, the final false alarm rate is $P_{\text{fa}} = 0.005$.

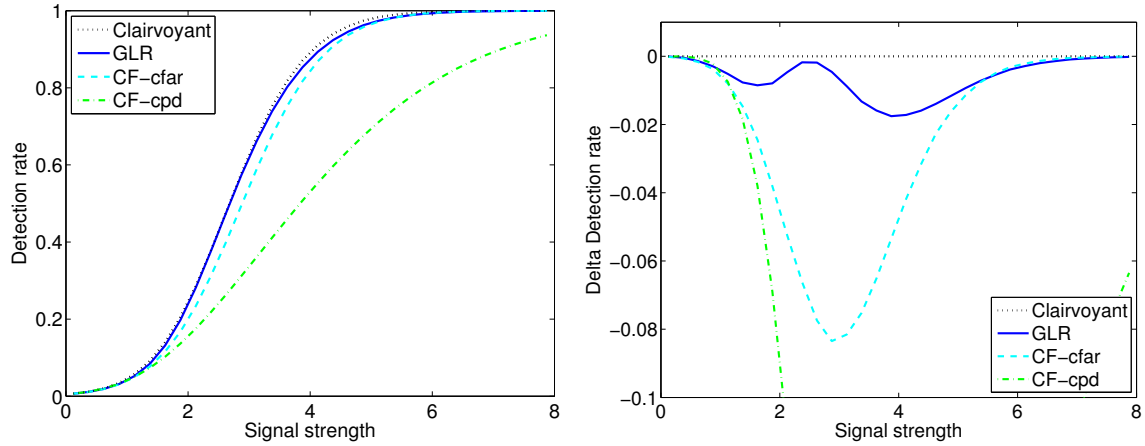


Figure 7. Same as Fig. 4, but using the Laplacian instead of the multivariate-t as the non-Gaussian background distribution.

6. CONCLUSIONS

Min-max clairvoyant fusion (CF) provides an intriguing alternative to the generalized likelihood ratio (GLR) that includes GLR as a special case; in that sense, it is an "even more generalized" likelihood ratio. The manipulated-threshold formulation of CF can be further generalized by using monotonic recalibration of clairvoyant statistics before the min-max fusion step. This new formulation allows us to produce new statistics, but it is important to acknowledge that it does not produce new decision rules. It is also important to point out that not every one-parameter family of CF decision rules can be converted into a statistic. And even for those that can, this paper provides an explicit procedure for constructing that statistic only for the CF-cfar and CF-cpd cases.

Although CF offers seemingly open-ended flexibility (in terms of the function $\lambda(\theta_t, \theta_b)$ in Eq. (15), or the function $h(\mathbf{x}, \theta_t, \theta_b)$ in Eq. (20)), three of the flavors – GLR, CF-cfar, and CF-cpd – are unambiguously specified and do not require a user to exercise domain expertise or arbitrary judgment. In the original formulation of CF, however, CF-cfar and CF-cpd are difficult to implement except in special cases. The monotonic recalibration approach leads to algorithms for CF-cfar and CF-cpd that work when the background distribution is specified in terms of data (instead of a formula), and these algorithms permit more extensive numerical testing and comparison of CF detectors.

The numerical results presented here suggest that there are regimes where CF-cfar is better (high signal strength and/or low false alarm rate), and regimes where CF-cpd is better (very low signal strength), but that over a wide range of intermediate regimes, GLR does very well. Further studies in more operational circumstances may be called for, and it is hoped that this formulation will facilitate those studies.

REFERENCES

1. A. Schaum, "Continuum fusion: a theory of inference, with applications to hyperspectral detection," *Optics Express* **18**, pp. 8171–8181, 2010.
2. A. Schaum, "Continuum fusion: a new methodology for creating hyperspectral detection algorithms," *Proc. SPIE* **7695**, p. 769502, 2010.
3. A. Schaum, "Algorithms with attitude," *Proc. 39th IEEE Applied Imagery Pattern Recognition (AIPR) Workshop* , 2010.
4. A. Schaum, "Continuum fusion: a new methodology for overcoming imperfect sensors and the limitations of statistical detection models," *Proc. MSS (Military Sensing Symposia)* , 2011.
5. A. Schaum, "Design methods for continuum fusion detectors," *Proc. SPIE* **8048**, p. 804803, 2011.
6. A. Schaum, "The continuum fusion theory of signal detection applied to a bi-modal fusion problem," *Proc. SPIE* **8064**, p. 806403, 2011.
7. B. Daniel and A. Schaum, "Linear log-likelihood ratio (L^3R) for spectral target detection," *Proc. SPIE* **8048**, p. 804804, 2011.
8. A. Schaum, "CFAR fusion: a replacement for the generalized likelihood ratio test for Neyman-Pearson problems," *Proc. 40th IEEE Applied Imagery Pattern Recognition (AIPR) Workshop* , 2011.
9. A. Schaum, "Progress in the theory of continuum fusion," *Proc. SPIE* **8390**, p. 839002, 2012.

10. P. Bajorski, "Min-max detection fusion for hyperspectral images," *Proc. 3rd IEEE Worskhop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)* , 2011.
11. P. Bajorski, "Generalized fusion: a new framework for hyperspectral detection," *Proc. SPIE* **8048**, p. 804802, 2011.
12. P. Bajorski, "Generalized detection fusion for hyperspectral images," *IEEE Trans. Geoscience and Remote Sensing* **50**, pp. 1199–1205, 2012.
13. J. Theiler, "Confusion and clairvoyance: some remarks on the composite hypothesis testing problem," *Proc. SPIE* **8390**, p. 839003, 2012.
14. E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, Springer, New York, 2005.
15. S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, vol. II, Prentice Hall, New Jersey, 1998.
16. A. Hayden, E. Niple, and B. Boyce, "Determination of trace-gas amounts in plumes by the use of orthogonal digital filtering of thermal-emission spectra," *Applied Optics* **35**, pp. 2802–2809, 1996.
17. B. R. Foy, R. R. Petrin, C. R. Quick, T. Shimada, and J. J. Tiee, "Comparisons between hyperspectral passive and multispectral active sensor measurements.," *Proc. SPIE* **4722**, pp. 98–109, 2002.
18. S. J. Young, "Detection and quantification of gases in industrial-stack plumes using thermal-infrared hyperspectral imaging," Tech. Rep. ATR-2002(8407)-1, The Aerospace Corporation, 2002.
19. J. Theiler, B. R. Foy, and A. M. Fraser, "Characterizing non-Gaussian clutter and detecting weak gaseous plumes in hyperspectral imagery," *Proc. SPIE* **5806**, pp. 182–193, 2005.
20. I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerospace and Electronic Systems* **10**, pp. 853–863, 1974.
21. F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerospace and Electronic Systems* **28**, pp. 208–216, 1992.
22. J. Theiler, B. R. Foy, and A. M. Fraser, "Beyond the adaptive matched filter: nonlinear detectors for weak signals in high-dimensional clutter," *Proc. SPIE* **6565**, p. 656503, 2007.
23. S. Cambanis, S. Huang, and G. Simons, "On the theory of elliptically contoured distributions," *J. Multivariate Analysis* **11**, pp. 368–385, 1981.
24. M. Rangaswamy, D. Weiner, and A. Ozturk, "Non-Gaussian random vector identification using spherically invariant random processes," *IEEE Trans. Aerospance and Electronic Systems* **29**, pp. 111–124, 1993.
25. E. Conte and M. Longo, "Characterisation of radar clutter as a spherically invariant random process," *IEE Proc. F (Communications, Radar and Signal Processing)* **134**, pp. 191–197, 1987.
26. D. Manolakis, D. Marden, J. Kerekes, and G. Shaw, "On the statistics of hyperspectral imaging data," *Proc. SPIE* **4381**, pp. 308–316, 2001.
27. D. B. Marden and D. Manolakis, "Using elliptically contoured distributions to model hyperspectral imaging data and generate statistically similar synthetic data," *Proc. SPIE* **5425**, pp. 558–572, 2004.
28. A. Schaum and E. Allman, "Detection in hyperspectral imagery based on non-Gaussian elliptically contoured distributions," *Proc. MSS (Military Sensing Symposium) on Camouflage, Concealment, and Deception* , 2006.
29. B. R. Foy, J. Theiler, and A. M. Fraser, "Decision boundaries in two dimensions for target detection in hyperspectral imagery," *Optics Express* **17**, pp. 17391–17411, 2009.

30. I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoustics, Speech, and Signal Processing* **38**, pp. 1760–1770, 1990.
31. J. Theiler and B. R. Foy, "EC-GLRT: Detecting weak plumes in non-Gaussian hyperspectral clutter using an elliptically-contoured generalized likelihood ratio test," *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* , p. I:221, 2008.
32. L. L. Scharf and L. T. McWhorter, "Adaptive matched subspace detectors and adaptive coherence estimators," *Proc. 30th Asilomar Conference on Signals, Systems, and Computers* , pp. 1114–1117, 1996.
33. S. Kraut, L. L. Scharf, and R. W. Butler, "The Adaptive Coherence Estimator: a uniformly most-powerful-invariant adaptive detection statistic," *IEEE Trans. Signal Processing* **53**, pp. 427–438, 2005.
34. E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerospace and Electronic Systems* **22**, pp. 115–127, 1986.