

# Beyond the Adaptive Matched Filter: Nonlinear Detectors for Weak Signals in High-Dimensional Clutter

James Theiler<sup>a</sup>, Bernard R. Foy<sup>b</sup>, and Andrew M. Fraser<sup>a</sup>

<sup>a</sup>Space and Remote Sensing Sciences Group, and

<sup>b</sup>Physical Chemistry and Applied Spectroscopy Group;  
Los Alamos National Laboratory, Los Alamos, NM, USA.

## ABSTRACT

For known signals that are linearly superimposed on gaussian backgrounds, the linear adaptive matched filter (AMF) is well-known to be the optimal detector. The AMF has furthermore proved to be remarkably effective in a broad range of circumstances where it is not optimal, and for which the optimal detector is not linear. In these cases, nonlinear detectors are theoretically superior, but direct estimation of nonlinear detectors in high-dimensional spaces often leads to flagrant overfitting and poor out-of-sample performance. Despite this difficulty in the general case, we will describe several situations in which nonlinearity can be effectively combined with the AMF to detect weak signals. This allows improvement over AMF performance while avoiding the full force of dimensionality's curse.

**Keywords:** matched filter, hyperspectral imagery, adaptive signal detection

## 1. INTRODUCTION

Hyperspectral imagery is big. You may think space is big, but space is only three dimensional. Hyperspectral images live in spaces with hundreds of dimensions. This is both a blessing, and – as Bellman<sup>1</sup> has famously noted – a curse. For finding weak signals in this high-dimensional space, the adaptive matched filter (AMF), often attributed to Reed *et al.*,<sup>2</sup> is an effective and time-honored tool. The AMF is a linear function that maps a high-dimensional  $\mathbb{R}^d$  to a scalar  $\mathbb{R}$ . This scalar is monotonically related to the confidence with which a pixel can be said to contain the signal of interest. For known signals linearly superimposed on a gaussian background with a known covariance matrix, the AMF is the optimal detector.

But in practice, hyperspectral data distributions are not gaussian. Alternative models include convex simplexes (*i.e.*, positive linear combinations of a few “endmembers”),<sup>3–5</sup> mixtures of gaussians,<sup>6</sup> elliptically-contoured (EC) distributions,<sup>7–10</sup> mixtures of EC distributions,<sup>11</sup> and more recently, lower-dimensional nonlinear manifolds.<sup>12</sup> The EC distributions are particularly appealing because they maintain the covariance structure of the gaussian distribution, but account for the fatter tails that are seen in real data. This simplifies the analysis, though it also predicts that tails are equally fat in all directions, and that is not observed in real data.<sup>13</sup>

Furthermore, the effect of the target on the measured signal is not always well-modeled as a straightforward addition: for a gaseous plume, there is both emission and absorption; for an opaque target, the target signal replaces (rather than adds to) the background signal.

Finally, the target signature may not be precisely known: the strength of the target signature is often unknown; there may be small uncertainties or field variations in the target spectrum; the target might be an unknown mixture of several known signatures; or, in the case of anomaly detection, the target may be completely unknown.

For all of these deviations from the ideal of a known signal added to a known gaussian background, the optimal detector need not be, and usually is not, the AMF. Even restricted to *linear* filters, the AMF need not be optimal, though in practice, the AMF is often found to be very nearly the best linear filter.<sup>14</sup> (For a class of elliptically-contoured distributions, one can show that the AMF *is* the optimal linear filter.)

---

Email: {jt,bfoy,afraser}@lanl.gov

To ask for the optimal nonlinear detector without further restrictions in a high-dimensional space invites overfitting. In this paper we will survey a number of nonlinear detectors that have been proposed for various departures from the linear gaussian ideal. By controlling the degrees of freedom or the dimensionality of the solutions, these methods can successfully exploit nonlinearity in the solutions. In many (though not all) of these cases, these nonlinear detectors can be readily visualized in a two-dimensional space. In this space, one axis is a matched-filter projection of the data and the other axis is the magnitude of the residual after matched-filter subtraction.

Our aim in this paper is not so much to provide a review that compares established detection algorithms (such as provided recently by Manolakis<sup>15</sup>), but to consider how various nonlinearities arise in hyperspectral target detectors as the restrictive assumptions that lead to the standard AMF are systematically lifted. We start, in Section 2, by deriving the AMF in two different ways, to illustrate two ways of thinking about AMF: either as optimizing signal-to-clutter for any distribution, or for producing a Bayes-optimal solution when the distribution is gaussian. We introduce the MFR plot in Section 3; many of the detectors that we describe are functions in the two-dimensional MFR space. The next three sections cover deviations from the ideal of a gaussian background to which a known signal has been directly added: Section 4 surveys detectors that arise when optimizing for non-gaussian backgrounds; Section 5 explains effects that arise when the signature is not known; and Section 6 describes detectors that are optimized for cases where the signal is not strictly additive.

## 2. DERIVATION OF THE ADAPTIVE MATCHED FILTER

The aim of the matched filter is to discriminate between pixels that are background clutter, and pixels that include signal. If  $\mathbf{x} \in \mathbb{R}^d$  is the observed spectrum of a background pixel, we write  $\boldsymbol{\xi}(\mathbf{x})$  for the spectrum of that pixel if signal has been added. For the linear adaptive matched filter (AMF), the effect of the signal is modeled as an additive perturbation.

$$\boldsymbol{\xi}(\mathbf{x}) = \mathbf{x} + \epsilon \mathbf{b}, \quad (1)$$

where  $\epsilon$  is a measure of the strength of the signal, and  $\mathbf{b}$  is the signal's spectrum. For a gaseous plume, this would correspond to the chemical spectrum of the gas. Other models for the effect of the signal on the the background are different functions  $\boldsymbol{\xi}(\mathbf{x})$ . We will discuss two other models later: a multiplicative absorption model in Section 6.1 and a linear combination in Section 6.2.

The matched filter is a vector  $\mathbf{q} \in \mathbb{R}^d$  and the matched filter output for a given pixel is the scalar value  $\mathbf{q}^T \mathbf{x}$ . The goal is to find a filter  $\mathbf{q}$  such that  $\mathbf{q}^T \boldsymbol{\xi}(\mathbf{x})$  is distinguishable from  $\mathbf{q}^T \mathbf{x}$ .

### 2.1. Maximizing Signal-to-Clutter

One can derive the matched filter without actually *assuming* that the data is gaussian. We use the second moments of the data to express the average squared signal and the average squared clutter. We remark, however, that the mapping of this optimization criterion to other optimality properties – such as highest detection or lowest false alarm rates – *does* invoke a gaussian assumption.

If  $\mathbf{b}$  is the signal of interest, then  $\mathbf{q}^T \mathbf{b}$  is the effect of the matched filter on the signal, and one wants this to be a large value. If  $\mathbf{x} \in \mathbb{R}^d$  is a hyperspectral pixel value without signal, then  $\mathbf{q}^T \mathbf{x}$  is the effect of the matched filter on the clutter. The average signal-to-clutter ratio (SCR) is given by

$$\text{SCR} = \frac{|\mathbf{q}^T \mathbf{b}|^2}{\text{Var}(\mathbf{q}^T \mathbf{x})} \quad (2)$$

where  $\text{Var}$  indicates the variance. It is given by

$$\text{Var}(\mathbf{q}^T \mathbf{x}) = \langle (\mathbf{q}^T \mathbf{x} - \mathbf{q}^T \boldsymbol{\mu})^2 \rangle = \mathbf{q}^T \langle (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \rangle \mathbf{q} = \mathbf{q}^T K \mathbf{q} \quad (3)$$

with  $\boldsymbol{\mu} = \langle \mathbf{x} \rangle$  the mean and  $K = \langle (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \rangle$  the covariance.

Since the magnitude of  $\mathbf{q}$  does not affect the SCR in Eq. (2), we are free to impose a constraint on the magnitude of  $\mathbf{q}$ . Thus, optimizing the SCR in Eq. (2) is equivalent to maximizing  $\mathbf{q}^T \mathbf{b}$  subject to constant  $\mathbf{q}^T K \mathbf{q}$ . We take the convention that  $\mathbf{q}^T K \mathbf{q} = 1$ , which leads to the Lagrangian formulation:

$$\ell(\mathbf{q}; \lambda, \mathbf{b}, K) = -\mathbf{q}^T \mathbf{b} + \lambda(\mathbf{q}^T K \mathbf{q} - 1) \quad (4)$$

Minimizing this loss function  $\ell$  is equivalent to maximizing the signal to clutter ratio. In particular, if we let

$$\mathbf{u} = 2\lambda K^{1/2} \mathbf{q} - K^{-1/2} \mathbf{b}, \quad (5)$$

then we can write the scalar value

$$\begin{aligned} \mathbf{u}^T \mathbf{u} &= (2\lambda K^{1/2} \mathbf{q} - K^{-1/2} \mathbf{b})^T (2\lambda K^{1/2} \mathbf{q} - K^{-1/2} \mathbf{b}) \\ &= 4\lambda^2 \mathbf{q}^T K \mathbf{q} - 4\lambda \mathbf{q}^T \mathbf{b} + \mathbf{b}^T K^{-1} \mathbf{b}, \end{aligned} \quad (6)$$

so Eq. (4) becomes

$$\ell(\mathbf{q}; \lambda, K, \mathbf{b}) = \frac{1}{4\lambda} \mathbf{u}^T \mathbf{u} - \lambda - \frac{1}{4\lambda} \mathbf{b}^T K^{-1} \mathbf{b}, \quad (7)$$

and it is clear that this is minimized when  $\mathbf{u} = \mathbf{0}$ , from which Eq. (5) implies  $\mathbf{q} = \frac{1}{2\lambda} K^{-1} \mathbf{b}$ . Since  $\mathbf{q}^T K \mathbf{q} = 1$ , we have  $\lambda = \frac{1}{2} \sqrt{\mathbf{b}^T K^{-1} \mathbf{b}}$ , or

$$\mathbf{q} = \frac{K^{-1} \mathbf{b}}{\sqrt{\mathbf{b}^T K^{-1} \mathbf{b}}}. \quad (8)$$

## 2.2. Likelihood ratio

The likelihood ratio test provides a way to derive the AMF in a way that optimizes what we really care about – detection and false alarm rates – subject to an explicit assumption that the distribution is gaussian. Let  $K$  be the covariance matrix of the clutter, and  $\boldsymbol{\mu}$  be its mean. Then the probability density for a point  $\mathbf{x}$  in the background clutter is given by

$$P_{\text{bkg}}(\mathbf{x}) = (2\pi)^{-d/2} |K|^{-1/2} \exp \left[ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T K^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]; \quad (9)$$

if  $\mathbf{x}$  includes plume, then that is equivalent to saying that  $\mathbf{x} - \epsilon \mathbf{b}$  is background. Thus,  $P_{\text{signal}}(\mathbf{x}) = P_{\text{bkg}}(\mathbf{x} - \epsilon \mathbf{b})$ . The likelihood ratio is then given by:

$$\mathcal{L}(\mathbf{x}) = \frac{P_{\text{signal}}(\mathbf{x})}{P_{\text{bkg}}(\mathbf{x})} = \frac{P_{\text{bkg}}(\mathbf{x} - \epsilon \mathbf{b})}{P_{\text{bkg}}(\mathbf{x})} = \frac{\exp \left[ -\frac{1}{2} (\mathbf{x} - \epsilon \mathbf{b} - \boldsymbol{\mu})^T K^{-1} (\mathbf{x} - \epsilon \mathbf{b} - \boldsymbol{\mu}) \right]}{\exp \left[ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T K^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]}. \quad (10)$$

So the negative log likelihood ratio is given by

$$\begin{aligned} -\log \mathcal{L}(\mathbf{x}) &= \frac{1}{2} (\mathbf{x} - \epsilon \mathbf{b} - \boldsymbol{\mu})^T K^{-1} (\mathbf{x} - \epsilon \mathbf{b} - \boldsymbol{\mu}) - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T K^{-1} (\mathbf{x} - \boldsymbol{\mu}) \\ &= -\frac{1}{2} \epsilon (\mathbf{x} - \boldsymbol{\mu})^T K^{-1} \mathbf{b} - \frac{1}{2} \epsilon \mathbf{b}^T K^{-1} (\mathbf{x} - \boldsymbol{\mu}) + \frac{1}{2} \epsilon^2 \mathbf{b}^T K^{-1} \mathbf{b} \\ &= -\epsilon \mathbf{b}^T K^{-1} (\mathbf{x} - \boldsymbol{\mu}) + \frac{1}{2} \epsilon^2 \mathbf{b}^T K^{-1} \mathbf{b} \\ &= -\epsilon (K^{-1} \mathbf{b})^T \mathbf{x} + \text{const} \end{aligned} \quad (11)$$

And in particular, contours of constant  $\mathcal{L}(\mathbf{x})$  will be given by  $\mathbf{q}^T \mathbf{x} = \text{const}$ , where the filter  $\mathbf{q} \propto K^{-1} \mathbf{b}$  is the AMF. Again, we take the convention that leads to the normalization in Eq. (8).

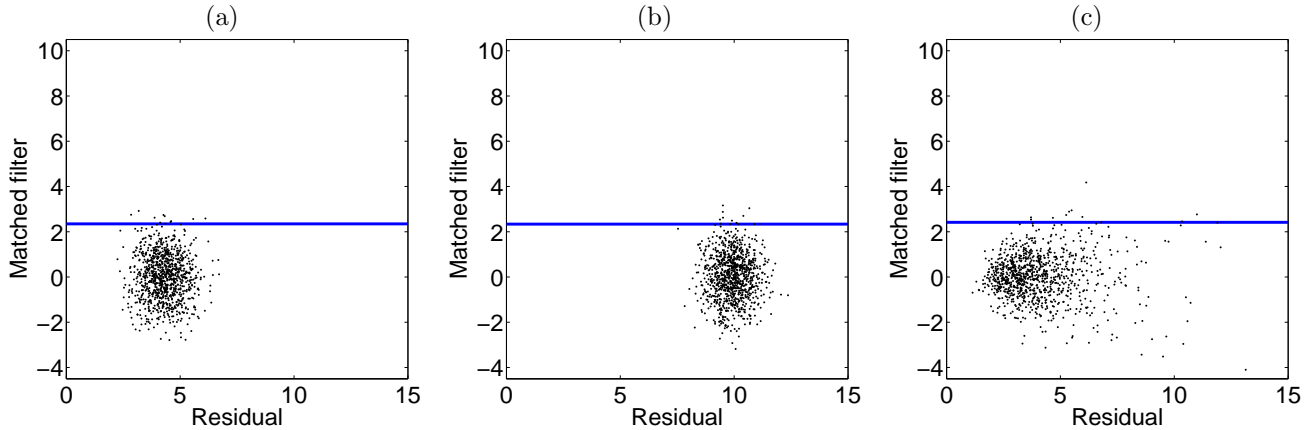
## 3. MFR PLOT

In this section, we introduce the matched-filter-residual (MFR) plot, which provides for a number of detectors a concise two-dimensional visualization of behavior that is nominally high-dimensional. In this plot, the vertical axis corresponds to the matched-filter direction, and the horizontal axis is the magnitude of the residual after the matched-filter has been projected out. The data set is first whitened, with the transform

$$\mathbf{w} = K^{-1/2} (\mathbf{x} - \boldsymbol{\mu}). \quad (12)$$

Note that  $\langle \mathbf{w} \mathbf{w}^T \rangle = \langle K^{-1/2} (\mathbf{x} - \boldsymbol{\mu}) (\mathbf{x} - \boldsymbol{\mu})^T K^{-1/2} \rangle = K^{-1/2} \langle (\mathbf{x} - \boldsymbol{\mu}) (\mathbf{x} - \boldsymbol{\mu})^T \rangle K^{-1/2} = K^{-1/2} K K^{-1/2} = I$ , so that the transformed data has unit covariance. The matched-filter amount is given by  $s = \mathbf{q}^T (\mathbf{x} - \boldsymbol{\mu})$ . This can be expressed

$$s = \tilde{\mathbf{q}}^T \mathbf{w}, \quad (13)$$



**Figure 1.** MFR plot showing the AMF as a solid horizontal line. Points above the line in this space will be identified as on-plume with a false alarm rate of 1%. Also shown are whitened data points ( $n = 1000$ ) drawn from three distributions: **(a)** multivariate gaussian distribution ( $d = 20$ ); **(b)** multivariate gaussian distribution ( $d = 100$ ); and **(c)** elliptically-contoured (EC) multivariate- $t$  distribution ( $\nu = 5, d = 20$ ). That the EC distribution has fatter tails than the gaussian distribution is most evident on the residual (horizontal) axis of this plot.

where  $\tilde{\mathbf{q}} = K^{1/2}\mathbf{q}$  is the matched filter in the whitened space. Note that  $\tilde{\mathbf{q}} = \tilde{\mathbf{b}}/|\tilde{\mathbf{b}}|$ , where  $\tilde{\mathbf{b}} = K^{-1/2}\mathbf{b}$  is the whitened signature. The residual magnitude is given by

$$t = \sqrt{\mathbf{w}^T \mathbf{w} - s^2}. \quad (14)$$

Fig. 1 shows a scatterplot of  $s$  versus  $t$  for simulated data drawn from different distributions. For the gaussian distributions, the AMF is the optimal detector.

## 4. NON-GAUSSIAN BACKGROUND

### 4.1. Parameterized

If the background is known (or at any rate modeled) to be an explicitly parameterized distribution and the signal is precisely known and additively combined, then we can use the likelihood ratio to derive optimal detectors. If the signal strength  $\epsilon$  is known, the likelihood test gives

$$\mathcal{L}_\epsilon(\mathbf{x}) = \frac{P_{\text{bkg}}(\mathbf{x} - \epsilon\mathbf{b})}{P_{\text{bkg}}(\mathbf{x})}. \quad (15)$$

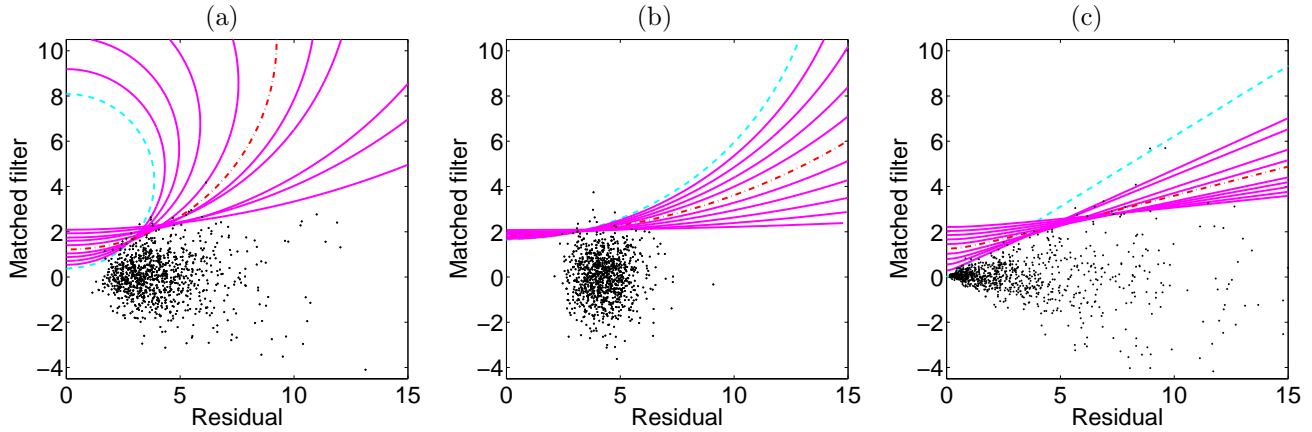
The elliptically-contoured (EC) multivariate  $t$ -distributions that have been advocated by Manolakis *et al.*<sup>8,9</sup> provide a handy example. As well as being analytically tractable, they also do a good job of describing the leptokurtic tails that are observed with real data.\* For a given  $\nu > 2$ , we can write

$$P_{\text{bkg}}(\mathbf{x}) \propto |K|^{-1/2} \left[ 1 + \frac{1}{\nu - 2} (\mathbf{x} - \boldsymbol{\mu})^T K^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]^{-(d+\nu)/2} \quad (16)$$

with the proportionality constant depending only on  $\nu$  and  $d$ . We can use this in Eq. (15) to define optimal detection contours, and these contours are shown in Fig. 2(a,b) for two different  $t$ -distributions.

Schaum and Allman<sup>16</sup> describe an additive signal model in which the background distribution is EC with an exponential tail, and the signal strength is known. Despite the restrictive assumptions, the model was found to

\*We have observed that the tails of real data are often more gaussian in the “thin” directions (*i.e.*, directions  $\mathbf{z}$  where  $\mathbf{z}K\mathbf{z}$  is small) and more heavy-tailed in directions with larger variance.<sup>13</sup> This deviates from what would be observed if the data were truly drawn from an EC distribution.



**Figure 2.** MFR plot showing various detectors, against a backdrop of whitened EC-distributed data, for two different EC distributions. The optimal detector, based on a likelihood ratio test, for non-gaussian distributions depends on on the strength  $\epsilon$  of the signal. Shown in these plots are a range of detectors, each optimized for a different signal strength. The dashed curve is the  $\epsilon \rightarrow 0$  limit, given by the derivative-based method in Eq. (20). The dashed-dotted line corresponds to  $\epsilon$  set to two sigmas, as described in Section 4.2.2. Shown are (a) the multivariate  $t$  distribution ( $\nu = 5, d = 20$ ); (b) a more nearly-gaussian  $t$  distribution ( $\nu = 50, d = 20$ ); and (c) the exponential-tailed EC distribution. Note that if the distribution were gaussian, then all of these detectors would be identical to each other.

be effective on some real hyperspectral target detection data. What is particularly appealing about this model is that the detector takes on a very simple hyperbolic form:

$$\mathcal{L}(\mathbf{w}) = \|\mathbf{w}\| - \|\mathbf{w} - \epsilon \tilde{\mathbf{b}}\|, \quad (17)$$

where  $\mathbf{w}$  and  $\tilde{\mathbf{b}}$  are the whitened coordinates for the measured spectrum and for the signal of interest, respectively. Fig. 2(c) shows the hyperbolic decision boundaries that this model produces for a range of  $\epsilon$  values.

The Johnson family of EC distributions has also been recommended for hyperspectral data.<sup>10</sup>

## 4.2. Unknown signal strength

When the background is gaussian and the signal is additive, then the optimal detector, the AMF, does not depend on the strength of the signal. This is a valuable invariance because in practice, the signal strength is rarely known. But if the background is not gaussian, then this invariance is broken, and the optimal detector *does* depend on signal strength. This means that when the signal strength is not known, some kind of non-optimal detector must be employed. Several approaches have been suggested for deriving a detector when the signal strength is not known.

### 4.2.1. Derivative method

The derivative method is based on the premise that the signal of interest is weak, and so the limit of  $\epsilon \rightarrow 0$  is taken. Informally, we would write  $\mathcal{L}(\mathbf{x}) = \lim_{\epsilon \rightarrow 0} \mathcal{L}_\epsilon(\mathbf{x})$ , but, as written, that limit is simply 1. Now, optimizing  $\mathcal{L}_\epsilon(\mathbf{x})$  for any given  $\epsilon$  is equivalent to optimizing the expression

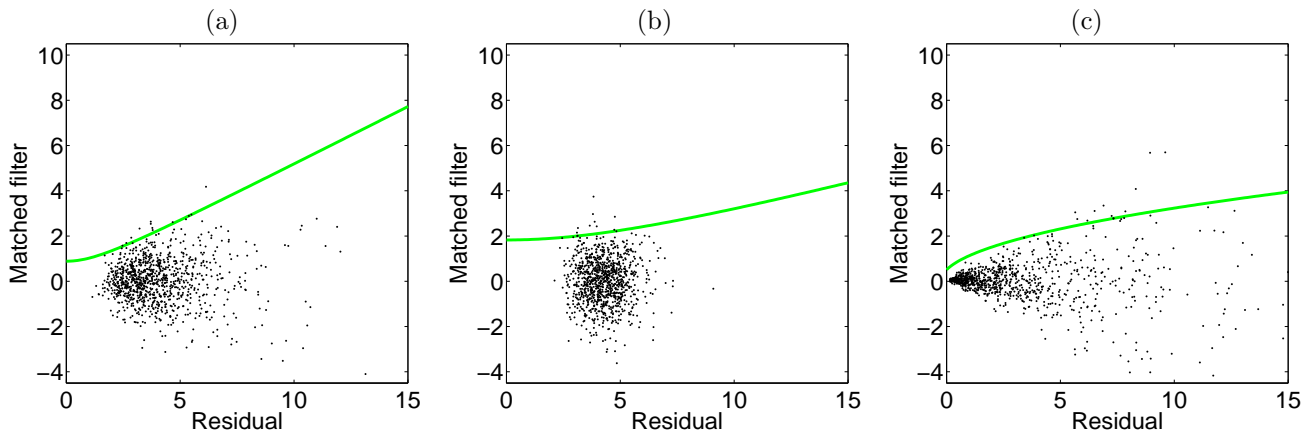
$$\mathcal{L}_\epsilon^*(\mathbf{x}) = (\mathcal{L}_\epsilon(\mathbf{x}) - 1)/\epsilon \quad (18)$$

and for small  $\epsilon$ , this looks like a derivative:

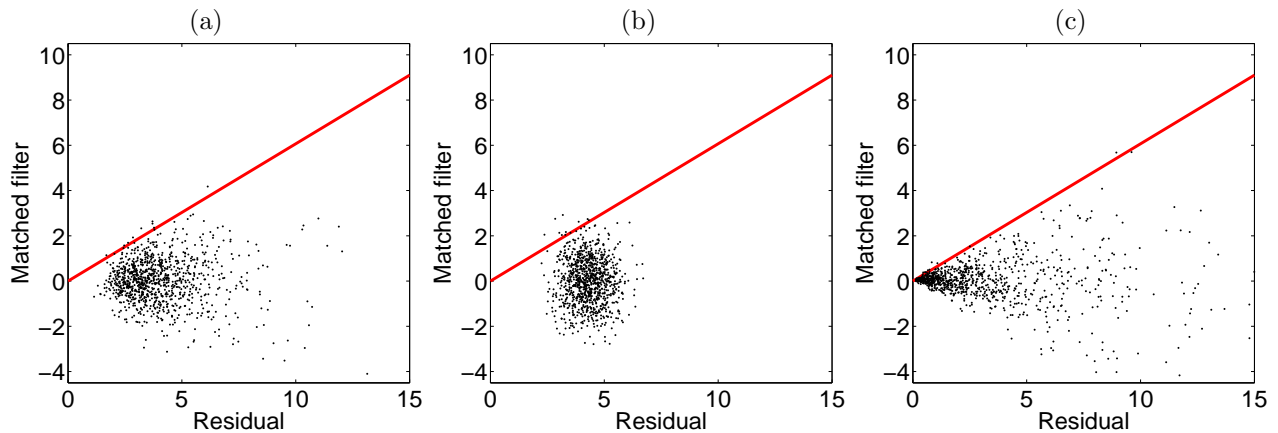
$$\mathcal{L}_\epsilon^*(\mathbf{x}) = \frac{\mathcal{L}_\epsilon(\mathbf{x}) - 1}{\epsilon} = \frac{\frac{d}{d\epsilon} P_{\text{bkg}}(\mathbf{x} - \epsilon \mathbf{b})}{P_{\text{bkg}}(\mathbf{x})} + O(\epsilon) \quad (19)$$

and now the limit  $\mathcal{L}^*(\mathbf{x}) = \lim_{\epsilon \rightarrow 0} \mathcal{L}_\epsilon^*(\mathbf{x})$  provides a nontrivial detector that does not depend on  $\epsilon$

$$\mathcal{L}^*(\mathbf{x}) = \frac{\frac{d}{d\epsilon} P_{\text{bkg}}(\mathbf{x} - \epsilon \mathbf{b})}{P_{\text{bkg}}(\mathbf{x})} = \frac{-\mathbf{b} \cdot \nabla P_{\text{bkg}}(\mathbf{x})}{P_{\text{bkg}}(\mathbf{x})} = -\mathbf{b} \cdot \nabla [\log P_{\text{bkg}}(\mathbf{x})]. \quad (20)$$



**Figure 3.** MFR plot showing GLRT detectors, against a backdrop of whitened EC-distributed data, for the same three distributions shown in Fig. 2.



**Figure 4.** The Adaptive Coherence Estimator (ACE) produces a contour in the MFR plot that is a straight line that passes through the origin. Its performance is independent of how fat the tails of the distribution are, and that leads to particularly favorable performance for particularly fat-tailed distributions. The same detector is shown with three distributions: (a) multivariate  $t$  distribution with  $\nu = 5.0$ , (b) gaussian; and (c) an EC distribution with an exponentially distributed tail.

When  $P_{\text{bkg}}(\mathbf{x})$  is the EC-distribution given in Eq. (16), then the derivative-based detector boundary is seen to be circular on the MFR plot, as seen in Fig. 2(a). This detector exhibits the counter-intuitive property that while moderate values of matched-filter output correspond to evidence for the presence of signal, very large values of matched-filter output argue for the absence of signal. For the exponentially-tailed distribution used by Schaum and Allman,<sup>16</sup> the detectors do not exhibit this property.

#### 4.2.2. Two sigma method

The idea is to achieve optimality for signals whose strength is about at the level of detectability. Stronger signals won't require an optimal detector, and weaker signals won't be seen in any case. Why two and not six sigmas or half a sigma? The choice depends on the ultimate false alarm rate that is desired. If an ultra-low false alarm rate is needed, then a larger number of sigmas will be better. For an  $n$ -sigma signal, we need  $\epsilon = n/\sqrt{\mathbf{b}^T K^{-1} \mathbf{b}}$ . The likelihood ratio looks like

$$\mathcal{L}(\mathbf{x}) = \frac{P_{\text{bkg}}(\mathbf{x} - \left[ \frac{n}{\sqrt{\mathbf{b}^T K^{-1} \mathbf{b}}} \right] \mathbf{b})}{P_{\text{bkg}}(\mathbf{x})} \quad (21)$$

#### 4.2.3. Generalized Likelihood Ratio Test (GLRT)

When signal strength *is* known, a likelihood ratio identifies the optimal detector for a given background distribution. Rather than treat the signal strength as very small ( $\epsilon \rightarrow 0$  limit) or as some other specific value (*e.g.*,  $\epsilon$  is two sigmas), the "generalized" likelihood ratio test (GLRT) uses the maximum likelihood estimate of  $\epsilon$  as if it were the known strength. Since the detector is already a likelihood ratio, this means an explicit estimate of the maximum likelihood  $\epsilon$  is not actually required.

$$\mathcal{L}_{\text{GLRT}}(\mathbf{x}) = \frac{\max_{\epsilon} P_{\text{bkg}}(\mathbf{x} - \epsilon \mathbf{b})}{P_{\text{bkg}}(\mathbf{x})} \quad (22)$$

Although the GLRT (sometimes known in the statistics literature simply as a likelihood ratio test<sup>17</sup>) has been used at least since the 1960's, its use in statistical processing, particularly for radar, was developed and advocated by Kelly.<sup>18</sup> It is not known to be optimal, except in special cases.<sup>19</sup> On the other hand, as seen in Fig. 3, and unlike what is seen in Fig. 2(a), and to a lesser extent in Fig. 2(b), the boundaries do not curve back on themselves. The larger the matched-filter output, the greater the evidence for the detection of signal.

#### 4.2.4. Bayesian Likelihood Ratio Test (BLRT) – Bayes Factor

The Bayesian likelihood ratio test (BLRT) provides an alternative to GLRT that employs Bayes rule.<sup>20</sup> This requires a prior, a distribution of (non-zero) signal strengths  $P_o(\epsilon)$ , possibly based on auxiliary information and/or domain knowledge, that encapsulates a best guess of the signal strength, given that there *is* a signal. This is the guess that would be made prior to seeing the data. Note that the method in Section 4.2.2 is a special case of the BLRT in which  $P_o(\epsilon)$  is a delta function at the two-sigma value of  $\epsilon$ . The BLRT is given by the ratio

$$\mathcal{L}_{\text{BLRT}}(\mathbf{x}) = \frac{\int P_{\text{bkg}}(\mathbf{x} - \epsilon \mathbf{b}) P_o(\epsilon) d\epsilon}{P_{\text{bkg}}(\mathbf{x})}. \quad (23)$$

Although it is an improper prior, a uniform  $P_o(\epsilon) = 1$  produces a result that leads to contours that roughly, though not exactly, mimic the GLRT.

For gaussian distributions with an additive signal model, all of the methods in this section give identical results. Schaum<sup>20</sup> describes some other cases in which GLRT can be treated as a special case of BLRT. One theoretical advantage of the BLRT over the GLRT is that, so long as  $P_o(\epsilon)$  truly represent the prior signal strength probability distribution, the BLRT is the provably optimal detector. A second advantage, described below in Section 4.3.2, is that this detector lends itself to being fit by machine learning methods.

### 4.3. Nonparametric

#### 4.3.1. Adaptive Coherence Estimator

The invariance approach of Scharf<sup>21</sup> leads to detectors that are not optimized for specific distributions, but instead are optimal over their invariance classes. In particular, the Adaptive Coherence Estimator (ACE)<sup>22–24</sup> was developed for a situation (apparently common in radar) in which the structure of the covariance matrix was fairly well estimated, but the overall magnitude of the clutter variance was nonstationary. The ACE is invariant to magnitude of clutter variance, and in fact it is invariant over all EC distributions with a given covariance matrix. This invariance, it turns out, produces a detector that performs well against fat-tailed distributions.

One way to think about the ACE detector is that is like the AMF, but instead of being applied directly to the spectral data  $\mathbf{x}$ , it is applied to the normalized value  $\mathbf{x}/|\mathbf{x}|$ . In  $\mathbb{R}^d$ , the ACE detector is a cone (versus the AMF detector which is a plane), and the MFR plot, shown in Fig. 4, indicates a detector that is a straight line that goes through the origin. Further, for a given false-alarm rate, the ACE detector is the same for any EC distribution.

#### 4.3.2. Machine learning in a two-dimensions

The appeal of machine learning methods, such as the support vector machine, for hyperspectral detection<sup>25</sup> and classification<sup>26, 27</sup> problems, is that it draws detection contours in a manner that is driven more by the data than the model. This is particularly attractive when the clutter is known to be structured, but that structure is not known *a priori*. In fitting this clutter to a parametric model such as a multivariate gaussian, the deviations of real data from the model will not be used in drawing the detection contours. On the other hand, attempting to learn a generic model in  $\mathbb{R}^d$  when  $d$  is large is to invite overfitting – and furthermore, this generic model doesn't incorporate what should be a valuable hint: that the AMF is a good place to start.

In Foy *et al.*,<sup>28</sup> we describe a machine learning approach that draws boundaries in the two-dimensional matched-filter-residual space. This compromise avoids direct parametric modelling of the background distribution, and permits a data-driven solution, but does so in a restricted space that diminishes the potential for overfitting, and that takes advantage of the AMF by using that as one of the two coordinate axes.

## 5. UNKNOWN SIGNATURE

There are a number of variations in which the signal of interest is not precisely known beforehand. Previously, in Section 4.2, we discussed the issues arising from the most common uncertainty – that of the signal strength. In this section, we will describe the effects of other kinds of uncertainties in the target signature.

### 5.1. Gaussian uncertainty in signature

Here,  $\boldsymbol{\xi}(\mathbf{x}) = \mathbf{x} + \epsilon\mathbf{b}_*$ , where  $\mathbf{b}_*$  is a random variable drawn from a distribution centered at  $\mathbf{b}$ . If the distribution is modeled to be gaussian – let's say  $P_{\text{gauss}}(\mathbf{b}_*; \mathbf{b}, K_b)$  – then we can write the likelihood function

$$\mathcal{L}(\mathbf{x}) = \frac{P_{\text{bkg}}(\mathbf{x} - \epsilon\mathbf{b}_*)}{P_{\text{bkg}}(\mathbf{x})} = \frac{\int P_{\text{bkg}}(\mathbf{x} - \epsilon\mathbf{b}_*)P_{\text{gauss}}(\mathbf{b}_*; \mathbf{b}, K_b) d\mathbf{b}_*}{P_{\text{bkg}}(\mathbf{x})} \quad (24)$$

and the negative log likelihood is (disregarding constants that do not depend on  $\mathbf{x}$ ) given by

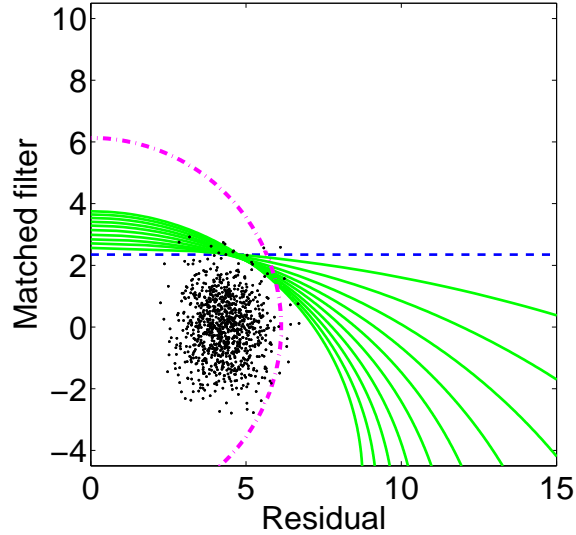
$$-\log \mathcal{L}(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu} - \epsilon\mathbf{b})^T (K + K_b)^{-1} (\mathbf{x} - \boldsymbol{\mu} - \epsilon\mathbf{b}) - (\mathbf{x} - \boldsymbol{\mu})^T K^{-1} (\mathbf{x} - \boldsymbol{\mu}) \quad (25)$$

As  $K_b \rightarrow 0$ , the error in the target vanishes, and Eq. (25) reverts to the AMF. For  $K_b \ll K$ , we can write

$$(K + K_b)^{-1} \approx K^{-1} - K^{-1}K_bK^{-1} \quad (26)$$

and we remark that that leaves Eq. (25) with a quadratic term whose coefficient is  $K^{-1}K_bK^{-1}$ . In the special case that  $K_b$  is proportional to  $K$ , this simplifies Eq. (25), and the solution can be expressed in the two-dimensional space of the MFR plot, as seen in Fig. 5.





**Figure 5.** Uncertainty in the signal alters the optimal detector; in the limit of zero uncertainty (shown as a dashed line), the AMF is obtained. But as the uncertainty increases, the optimal detector “curves down” away from the signal. This is because the variance of the signal+background is larger than the variance of the background alone. Shown are curves corresponding to the case  $K_b = \theta K$ , for increasing values of  $\theta$ . In the limit of a completely unknown signal, the problem devolves to that of anomaly detection, which is shown here with the the dot-dashed curve.

## 5.2. Completely unknown signature

For the anomaly detection problem, the signature is entirely unknown, and the anomaly detector is chosen as a level curve for  $P_{\text{bkg}}(\mathbf{x})$ . One way to think about this, discussed for instance in Refs. [29,30], is as a likelihood ratio where the “signal” distribution is uniform over a region larger than the extent of the data. For instance, taking  $K_b$  much larger (*i.e.*, broader) than  $K$  in Eq. (25), we obtain that the detection contour is an ellipse that is characterized by  $\mathbf{x}^T K \mathbf{x}$ , corresponding to the elliptical contours of the data.

## 5.3. Multiple targets

When there are multiple target signatures (such as the chemicals that might make up a gaseous plume) whose effect is a linear combination of the individual signals, then the unknown signal strength  $\epsilon$  becomes a vector of signal strengths associated with the different components. This situation also arises when uncertainty in the target signature is best modeled as a (possibly bounded) subspace instead of an explicit probability distribution. This is the subspace detection problem, and it has been widely studied in the literature<sup>19, 21, 31-33</sup> (see Schaum<sup>34</sup> for a readable review). Even for gaussian backgrounds, the detector surfaces are nonlinear,<sup>33</sup> but we remark that in contrast to many of the previous cases we discussed, the subspace solutions do not, unfortunately, fit neatly in an MFR plot.

# 6. NON-ADDITIVE SIGNAL

## 6.1. Absorptive plume detection

For a strictly absorptive plume, Beer’s Law<sup>35</sup> says that for any wavelength  $\lambda$  we have  $\xi_\lambda(x_\lambda) = x_\lambda \exp(-\epsilon b_\lambda)$ , where again  $\epsilon$  is the plume strength. Let us define a “signature matrix”  $B$ , derived from the chemical signature  $b_\lambda$ , which is a diagonal matrix whose diagonal elements are given by:

$$B_{\lambda\lambda} = \frac{1 - \exp(-\epsilon b_\lambda)}{\epsilon} \quad (27)$$

Note that the  $\epsilon \rightarrow 0$  limit of this equation has  $B_{\lambda\lambda} = b_\lambda$ . Now, we can write  $\xi(\mathbf{x})$  in honest matrix-vector notation:

$$\xi(\mathbf{x}) = \mathbf{x} - \epsilon B \mathbf{x} \quad (28)$$

This looks similar to the expression for  $\xi(\mathbf{x})$  in the additive case, but there is an extra factor of  $\mathbf{x}$  on the additive component. One simple way to address the variability of the  $B\mathbf{x}$  term is to take the mean: That is, let  $\xi(\mathbf{x}) \approx \mathbf{x} - \epsilon B \langle \mathbf{x} \rangle = \mathbf{x} - \epsilon B \boldsymbol{\mu}$ , and now the expression  $B\boldsymbol{\mu}$  plays the same role as  $\mathbf{b}$  in the AMF, and that leads to a matched filter  $\mathbf{q}$  given by  $\mathbf{q} = K^{-1}B\boldsymbol{\mu}$ . (This approach was used, for example, in Ref. [36].)

From Eq. (28), we can write likelihoods for plume and for background.

$$\mathcal{L}(\mathbf{x}) = \frac{P_{\text{signal}}(\mathbf{x})}{P_{\text{bkg}}(\mathbf{x})} = \frac{P_{\text{bkg}}(\mathbf{x} + \epsilon B\mathbf{x})}{P_{\text{bkg}}(\mathbf{x})} = \frac{\exp\left[-\frac{1}{2}(\mathbf{x} + \epsilon B\mathbf{x} - \boldsymbol{\mu})^T K^{-1}(\mathbf{x} + \epsilon B\mathbf{x} - \boldsymbol{\mu})\right]}{\exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T K^{-1}(\mathbf{x} - \boldsymbol{\mu})\right]}. \quad (29)$$

And the negative log likelihood is:

$$\begin{aligned} -\log \mathcal{L}(\mathbf{x}) &= \frac{1}{2}(\mathbf{x} + \epsilon B\mathbf{x} - \boldsymbol{\mu})^T K^{-1}(\mathbf{x} + \epsilon B\mathbf{x} - \boldsymbol{\mu}) - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T K^{-1}(\mathbf{x} - \boldsymbol{\mu}) \\ &= \frac{1}{2}\epsilon \mathbf{x}^T B^T K^{-1}(\mathbf{x} - \boldsymbol{\mu}) + \frac{1}{2}\epsilon(\mathbf{x} - \boldsymbol{\mu})^T K^{-1}B\mathbf{x} + \epsilon^2 \mathbf{x}^T B^T K^{-1}B\mathbf{x}, \end{aligned} \quad (30)$$

which, to leading order in  $\epsilon$ , is

$$-\log \mathcal{L}(\mathbf{x}) = \frac{1}{2}\epsilon(\mathbf{x} - \boldsymbol{\mu})^T (K^{-1}B + B^T K^{-1})(\mathbf{x} - \boldsymbol{\mu}) + \epsilon \boldsymbol{\mu}^T B^T K^{-1}(\mathbf{x} - \boldsymbol{\mu}) + O(\epsilon^2) \quad (31)$$

Thus, the modified quadratic matched filter is given by

$$\mathbf{q}^T(\mathbf{x} - \boldsymbol{\mu}) + \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T Q(\mathbf{x} - \boldsymbol{\mu}) = \text{const} \quad (32)$$

where  $\mathbf{q} = K^{-1}B\boldsymbol{\mu}$  is the linear matched filter, and the quadratic covariance coefficient is given by:

$$Q = K^{-1}B + B^T K^{-1}. \quad (33)$$

As a rough order of magnitude, the quadratic term is a factor of  $|\mathbf{x} - \boldsymbol{\mu}|/|\boldsymbol{\mu}|$  the size of the linear term.

## 6.2. Finite Target Matched Filter

Introduced by Schaum and Stocker,<sup>37</sup> the finite-target matched-filter (FTMF) is a nonlinear detector that arises in subpixel target detection. When the target  $\mathbf{b}$  is precisely known, then the signal observed at a pixel in which a fraction  $f$  of that pixels contains target is given by

$$\xi(\mathbf{x}) = (1 - f)\mathbf{x} + f\mathbf{b}. \quad (34)$$

If the target strength  $f$  is known beforehand, then we can use a likelihood ratio test

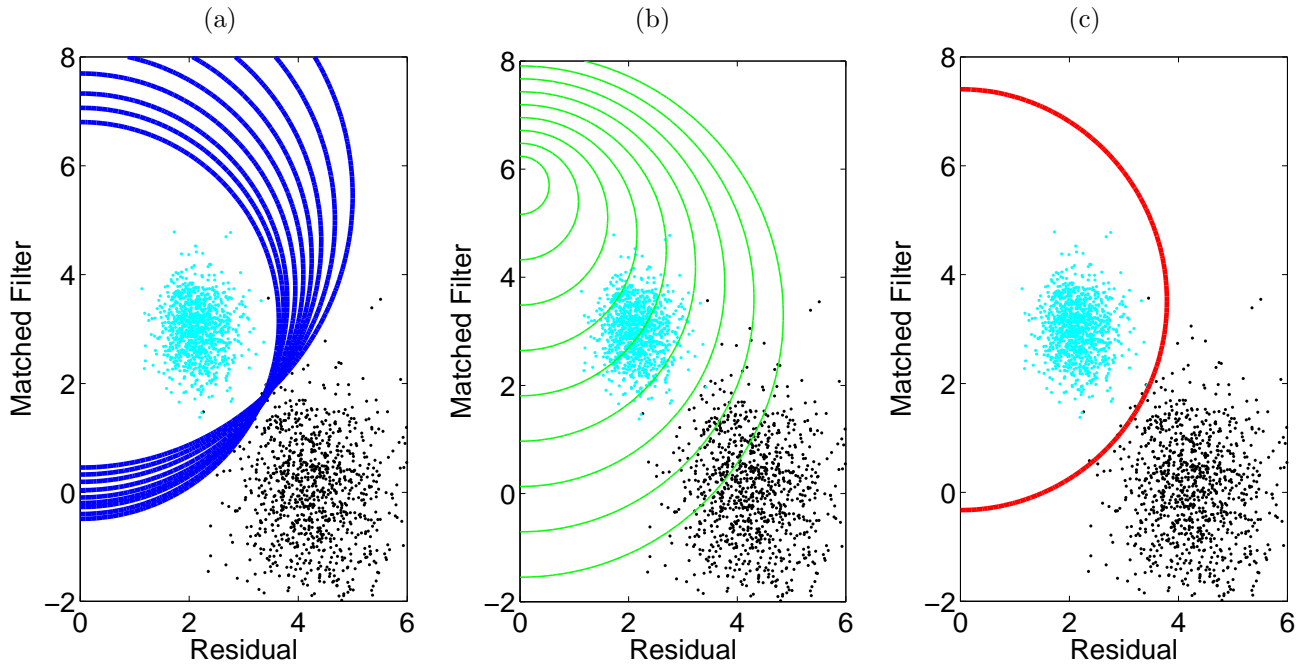
$$\mathcal{L}(\mathbf{x}) = \frac{P_{\text{signal}}(\mathbf{x})}{P_{\text{bkg}}(\mathbf{x})} = \frac{P_{\text{gauss}}(\mathbf{x}; (1 - f)\boldsymbol{\mu} + f\mathbf{b}, (1 - f)^2 K)}{P_{\text{gauss}}(\mathbf{x}; \boldsymbol{\mu}, K)} \quad (35)$$

where the expression on the right assumes a gaussian background distribution. The log likelihood will be a quadratic function of  $\mathbf{x}$  and the boundaries will be hyperellipsoids, and they will appear as ellipses in an MFR plot, as shown in Fig. 6(a) for a range of values of  $f$ .

If the target strength  $f$  is unknown, then we can employ maximum likelihood to estimate it:

$$\hat{f}(\mathbf{x}) = \underset{f \geq 0}{\operatorname{argmax}} P_{\text{gauss}}(\mathbf{x}; (1 - f)\boldsymbol{\mu} + f\mathbf{b}, (1 - f)^2 K) \quad (36)$$

Note that this leads to a different  $\hat{f}$  at each position  $\mathbf{x}$ ; the value can be expressed in closed form (see eq (12) of Ref. [37]); the contours of  $\hat{f}$ , as they appear in MFR space, are shown in Fig. 6(b). Using the expression for  $\hat{f}$  in the likelihood ratio, we can obtain the GLRT boundary shown in Fig. 6(c); unlike the LRT curves in Fig. 6(a), this is no longer a quadratic function of  $\mathbf{x}$  (sometimes called ‘‘hyperquadratic’’ in these high-dimensional spaces).



**Figure 6.** Finite-target matched filter (FTMF) solution for gaussian data in  $d = 20$  with a target that is six sigmas from the mean of the data. Shown are two data clouds: the lower larger cloud is the data; the smaller cloud to the upper left is the data that would be observed using Eq. (34) with  $f = 0.5$ . (a) Likelihood ratio boundaries are shown for a range of fixed values of  $f$ , in each case with the threshold chosen so that the false alarm rate is 1%. (b) Contours of the maximum likelihood estimate of  $f$  is shown for  $f = 0.1$  to  $f = 0.9$ . (c) GLRT contour.

## REFERENCES

1. R. E. Bellman, *Adaptive Control Processes*, Princeton University Press, Princeton, NJ, 1961.
2. 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.
3. J. W. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Summaries of the Fourth Annual JPL Airborne Geoscience Workshop*, R. O. Green, ed., pp. 11–14, 1994.
4. J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of AVIRIS data," in *Summaries of Fifth Annual JPL Airborne Earth Science Workshop*, R. O. Green, ed., pp. 23–26, 1995.
5. M. E. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proc. SPIE* **3753**, pp. 266–275, 1999.
6. C. C. Funk, J. Theiler, D. A. Roberts, and C. C. Borel, "Clustering to improve matched filter detection of weak gas plumes in hyperspectral imagery," *IEEE Trans. Geoscience and Remote Sensing* **39**, pp. 1410–1420, 2001.
7. M. Rangaswamy, D. Weiner, and A. Ozturk, "Non-Gaussian random vector identification using spherically invariant random processes," *IEEE Trans. Aerospace and Electronic Systems* **29**, pp. 111–124, 1993.
8. D. Manolakis, D. Marden, J. Kerekes, and G. Shaw, "On the statistics of hyperspectral imaging data," *Proc. SPIE* **4381**, pp. 308–316, 2001.
9. 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.
10. E. C. Meidunas and S. C. Gustafson, "Johnson distribution models of hyperspectral image data clusters," *Proc. SPIE* **6233**, 2006.
11. D. B. Marden and D. Manolakis, "Modeling hyperspectral imaging data," *Proc. SPIE* **5093**, pp. 253–262, 2003.

12. C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Exploiting manifold geometry in hyperspectral imagery," *IEEE Trans. Geoscience and Remote Sensing* **43**, pp. 441–454, 2005.
13. 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.
14. B. R. Foy, J. Theiler, and A. M. Fraser, "Unreasonable effectiveness of the adaptive matched filter," *Proc. MSS (Military Sensing Symposia) Passive Sensors Conference*, 2006.
15. D. Manolakis, "Taxonomy of detection algorithms for hyperspectral imaging applications," *Optical Engineering* **44**, pp. 66403–1–11, 2005.
16. 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.
17. T. S. Ferguson, *Mathematical Statistics: a decision theoretic approach*, Academic Press, New York, 1967.
18. E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerospace and Electronic Systems* **22**, pp. 115–127, 1986.
19. L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Processing* **42**, pp. 2146–2156, 1994.
20. A. Schaum, "Hyperspectral target detection using a Bayesian likelihood ratio test," *Proc. IEEE Aerospace Conference* **3**, pp. 1537–1540, 2002.
21. L. L. Scharf, *Statistical Signal Processing*, Addison-Wesley, Reading, MA, 1990.
22. E. Conte, M. Lops, and G. Ricci, "Asymptotically optimum radar detection in compound-gaussian clutter," *IEEE Trans. Aerospace and Electronic Systems* **31**, pp. 617–625, 1995.
23. E. Conte, M. Lops, and G. Ricci, "Adaptive matched filter detection in spherically invariant noise," *IEEE Signal Processing Letters* **3**, pp. 248–250, 1996.
24. L. L. Scharf and L. T. McWhorter, "Adaptive matched subspace detectors and adaptive coherence estimators," *Proc. Asilomar Conference on Signals, Systems, and Computers*, 1996.
25. T. Tarumi, G. W. Small, R. J. Combs, and R. T. Kroutil, "Remote detection of heated ethanol plumes by airborne passive fourier transform infrared spectrometry," *Applied Spectroscopy* **57**, pp. 1432 – 1441, 2003.
26. J. A. Gualtieri and R. F. Crompt, "Support vector machines for hyperspectral remote sensing classification," *Proc. SPIE* **3584**, pp. 221–232, 1999.
27. J. A. Gualtieri and S. Chettri, "Support vector machines for classification of hyperspectral data," *Proc. IGARSS 2000* **2**, pp. 813–815, 2000.
28. B. R. Foy, J. Theiler, and A. M. Fraser, "Decision boundaries in two dimensions for target detection in hyperspectral imagery," Tech. Rep. LA-UR-07-0037, Los Alamos National Laboratory, 2007.
29. T. Hastie, R. Tibshirani, and J. Friedman, *Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer-Verlag, New York, 2001. This anomaly detection approach is developed in Chapter 14.2.4, and neatly illustrated in Fig 14.3.
30. J. Theiler and D. M. Cai, "Resampling approach for anomaly detection in multispectral images," *Proc. SPIE* **5093**, pp. 230–240, 2003.
31. S. Kraut, L. L. Scharf, and L. T. McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Processing* **49**, pp. 1–16, 2001.
32. A. Schaum and A. Stocker, "Joint subspace detection in hyperspectral sensing," *Proc. MSS (Military Sensing Symposium) on Camouflage, Concealment, and Deception*, 2003.
33. A. Schaum and E. Allman, "Nonlinear signature-based hyperspectral target detection: Beyond the matched filter," *Proc. MSS (Military Sensing Symposium) on Camouflage, Concealment, and Deception*, 2004.
34. A. Schaum, "Spectral subspace matched filtering," *Proc. SPIE* **4381**, pp. 1–17, 2001.
35. A. Beer, "Bestimmung der absorption des rothen lichts in farbigen flussigketiten," *Ann. Physik* **86**, pp. 78–88, 1852.
36. J. Theiler and B. R. Foy, "Effect of signal contamination in matched-filter detection of the signal on a cluttered background," *IEEE Geoscience and Remote Sensing Letters* **3**, pp. 98–102, 2006.
37. A. Schaum and A. Stocker, "Spectrally selective target detection," *Proc. International Symposium on Spectral Sensing Research*, 1997.