BAYESIAN VS GENERALIZED LIKELIHOOD RATIO DETECTION OF SOLID SUB-PIXEL TARGETS

James Theiler

Space Remote Sensing and Data Science Group,
Los Alamos National Laboratory, Los Alamos, NM 87545 USA
Email: `jt@lanl.gov`

ABSTRACT

Numerical experiments compare Bayesian and non-Bayesian Generalized Likelihood Ratio Test (GLRT) detection algorithms for opaque sub-pixel hyperspectral targets of unknown abundance. A simplified problem is identified, which allows the full range of Bayesian priors to be explored. When one seeks to minimize the false alarm rate at a fixed detection rate, one finds that GLRT detection outperforms Bayesian detection for *any* choice of prior. By contrast, when the criterion is detection rate at fixed false alarm rate, Bayesian detection is better. The results hold over a wide range of parameters, and appear to contradict known optimality results for Bayesian detectors. The apparent discrepancy is explained, and a case is made for the practical use of GLRT-based detection statistics.

***Index Terms*—** Algorithm, Spectral imagery, Target detection, Likelihood ratio, Composite hypothesis testing, Detection statistic, Bayes, GLRT, Multivariate t distribution

1. BACKGROUND: TARGET DETECTION

For an opaque sub-pixel target with a d-channel reflectance spectrum given by the vector $\mathbf{t}$, sub-pixel abundance given by scalar a (the fraction of the pixel covered by the target), and atop a background spectrum $\mathbf{z}$, the observed mixed spectrum is given by the replacement target model [1, §11.1.4]:

$$\mathbf{x} = (1-a)\mathbf{z} + a\mathbf{t}. \tag{1}$$

Under the null hypothesis, the target is absent, so $a = 0$ and $\mathbf{x} = \mathbf{z}$. Under the alternative *composite* hypothesis, the target is present with a nonzero abundance a. If the target abundance a were known, the alternative hypothesis would be *simple*, and the optimal detector would be given by the likelihood ratio:

$$L(a, \mathbf{x}) = \frac{P_{\text{target}}(\mathbf{x})}{P_{\text{bkg}}(\mathbf{x})} = \frac{(1-a)^{-d} P_{\text{bkg}}\left(\dfrac{\mathbf{x} - a\mathbf{t}}{1-a}\right)}{P_{\text{bkg}}(\mathbf{x})}, \tag{2}$$

where $P_{\text{bkg}}(\mathbf{z})$ is the probability distribution function that models the variability of the backgound pixels $\mathbf{z}$.

1.1. Detectors vs Detection Statistics

Formally speaking, the likelihood ratio in Eq. (2) is not by itself a detector; rather, it is a *detection statistic*; it is effectively a family of detectors, with each member of the family an individual detector specified by a threshold η. Thus, if $\mathcal{D}(\mathbf{x})$ is a detection statistic, then the detector is given by:

$$\begin{aligned} &\mathcal{D}(\mathbf{x}) < \eta \ \text{ indicates no target in pixel } \mathbf{x}; \\ &\mathcal{D}(\mathbf{x}) \geq \eta \ \text{ indicates target present in pixel } \mathbf{x}. \end{aligned} \tag{3}$$

Each *detector* is a binary function of $\mathbf{x}$, and in general any binary function can be a detector, but detectors built form detection statistics are especially useful. An adjustable threshold enables measures of confidence to be associated with individual detections, and furthermore permits the construction of receiver operating characteristic (ROC) curves.

1.2. ROC Curve-Based Measures of Quality

What ultimately characterizes the quality of a detector is the number of true detections (target pixels correctly identified as target pixels) and false alarms (background pixels incorrectly identified as target pixels).

For a family of detectors characterized by a detection statistic and an adjustable threshold, a ROC curve can be produced by plotting detection rate against false alarm rate for different choices of the threshold.

Specific scalar-valued criteria are useful for comparing detection statistics, and these typically take one of two forms. Write DR@FAR=x (with x typically 0.05 or smaller) to indicate the detection rate at the threshold for which the false alarm rate is x; and write FAR@DR=x (with x typically a half or larger) to indicate the false alarm rate at the threshold for which the detection rate is x. For applications where targets are relatively rare (*e.g.*, for sub-pixel targets in imagery), the false alarm rates of interest are very low, and also vary over a large dynamic range (*e.g.*, as the target strength a varies). For these reasons, the FAR@DR=x criteria are often more attractive as a research tool in evaluating target detection algorithms.

2. DETECTION STATISTICS

2.1. Clairvoyant Detection

If the target abundance *is* known to be $a = a_o$ (put another way: if the target abundance is restricted to be either $a = 0$ [target absent] or $a = a_o$ [target present]), then the likelihood ratio provides the optimal detection statistic:

$$\mathcal{D}_{\text{Clairvoyant}}(\mathbf{x}) = L(a_o, \mathbf{x}) \tag{4}$$

In most practical applications of interest, the target abundance is *unknown*. In that case, we have a composite hypothesis testing problem, and there is no single optimum solution. Several approaches have been taken to develop detectors in this composite case.

2.2. GLRT Detection

The traditional approach is to employ the generalized likelihood ratio test (GLRT). That is: find $\widehat{a}(\mathbf{x})$ that maximizes the likelihood in Eq. (2), and use $L(\widehat{a}(\mathbf{x}), \mathbf{x})$ as the detector. That is:

$$\mathcal{D}_{\text{GLRT}}(\mathbf{x}) = \max_a L(a, \mathbf{x}) = L(\widehat{a}(\mathbf{x}), \mathbf{x}). \tag{5}$$

For the replacement-model likelihood function in Eq. (2), using the elliptically-contoured multivariate t-distributed backgroun in Eq. (9), it turns out that it is possible to solve for $\widehat{a}(\mathbf{x})$ analytically and to thus obtain a closed-form solution for the detector in Eq. (5) [2, 3].

2.2.1. *Restricted GLRT Detection*

In the traditional version of the sub-pixel target detection problem, the abundance a is only known to be in the range $0 \leq a \leq 1$. For the experiments conducted here, a simplified version of the problem is considered, in which the target strength (when a target is present) is known to be an element of a discrete set. In particular, we will take $a \in \mathbf{a} = \{a_1, \ldots, a_n\}$. In this case, we modify Eq. (5) to consider the maximum over this restricted set; thus

$$\mathcal{D}_{\text{RGLRT}}(\mathbf{x}) = \max_{a \in \mathbf{a}} L(a, \mathbf{x}) \tag{6}$$

is the Restricted GLRT (RGLRT). When the targets really are restricted to this discrete set, then the RGLRT is generally expected to be a better detector than the GLRT.

Note that the RGLRT is straightforward (and relatively inexpensive) to implement. A closed-form solution for Eq. (5) requires taking a derivative, setting it to zero, solving the resulting equation for $\widehat{a}(\mathbf{x})$, and plugging that expression into $L(\widehat{a}(\mathbf{x}), \mathbf{x})$. By contrast, Eq. (6) is already in closed form.

2.3. Bayesian Detection

The Bayesian approach to target detection (*e.g.*, see [4]) first posits a prior $q(a)$. Then, instead of taking the peak value of the likelihood function, we take the weighted average over the range $0 \leq a \leq 1$:

$$\mathcal{D}_{\text{Bayes}}(\mathbf{x}) = \int_0^1 L(a, \mathbf{x}) q(a) da. \tag{7}$$

A natural choice, given that we do not have *a priori* information about target abundance, is to use the uniform prior: $p(a) = 1$. This choice, unfortunately, does not readily lead to a closed-form solution for $\mathcal{D}_{\text{Bayes}}(\mathbf{x})$.

2.3.1. *Discrete/Restricted Bayesian Detection*

In parallel with §2.2.1, the Bayesian detector also simplifies when the set of available target strengths is discrete. If $a \in \mathbf{a} = \{a_1, \ldots, a_n\}$, then with corresponding weights $w_1, \ldots, w_n$, we can write $q(a) = \sum_{i=1}^n w_i \delta(a - a_i)$, so that Eq. (7) simplifies to

$$\mathcal{D}_{\text{DiscreteBayes}}(\mathbf{x}) = \sum_{i=1}^n w_i L(a_i, \mathbf{x}), \tag{8}$$

which, unlike Eq. (7), is in closed form.[1] Note that Eq. (8) can also be used as a closed-form approximation to Eq. (7) by judicious choice of a_i's and w_i's [6].

2.3.2. *Sculpting priors*

For many Bayesian data analysis problems, the prior corresponds to a subjective probabilistic model of what target abundances we might expect to see, *a priori*; that is, prior to seeing the data. But for target detection, as noted in Lehmann and Romano [7, §1.6], this interpretation of the prior is often not useful; instead, they suggest, it "expresses the importance that the experimenter attaches to the various values of" the unknown parameter. This opens up the possibility of "sculpting" the prior, of crafting the function $q(a)$ to optimize a (possibly specialized, problem-specific) performance criterion. One particularly simple numerical scheme is to design the prior with a single appropriately chosen delta function [8].

2.4. Power and Admissibility

The *power* of a detector is its detection rate, and it depends on the parameter a. The false alarm rate, meanwhile, depends only on the background distribution; it does not depend on a. For two detectors with the same false alarm rate, one is said to be more powerful than the other if its detection rate is higher. If the detection rate is as high or higher than the other detector for every value of a, then it is *uniformly* more powerful. If a detector is uniformly more powerful than *every* other detector, then it is a *uniformly most powerful* (UMP) detector [9, §6.3]. Unfortunately, many problems of interest, including the one

[1] We can also deal with $a_1 \to 0$, which would enable us to optimize for arbitrarily weak targets, but this requires special handling [5].

studied here, do not admit a UMP solution. In that case, our next best bet is to find *admissible* detectors. These are detectors for which no other detector is uniformly more powerful.

One of the motivations[2,3] for investigating the Bayesian detector is that Bayesian detectors are known to be *admissible* [7, §1.8]. By contrast, GLRT detectors (while often quite useful in practice) do not have this guarantee; Ref. [13] shows an example for which the GLRT detector is uniformly worse than the associated Bayesian detector.

Finally, it is important to emphasize that the concepts of (and theorems about) power and admissibility formally refer to individual detectors, not to detection statistics.

3. NUMERICAL EXPERIMENTS AND RESULTS

The experiments described here use the replacement target model in Eq. (1), where the background is given by the multivariate t distribution with mean $\boldsymbol{\mu}$ and covariance R:

$$P_{\text{bkg}}(\mathbf{z}) = c\left[(\nu-2) + (\mathbf{z}-\boldsymbol{\mu})' R^{-1} (\mathbf{z}-\boldsymbol{\mu})\right]^{-(d+\nu)/2}, \tag{9}$$

Here: d is the dimension of the vector $\mathbf{x}$ (*i.e.*, number of channels in the spectral image); $\boldsymbol{\mu}$ and R are the mean (vector) and covariance (matrix) of the distribution; ν characterizes fatness of the tail; and c is a constant pre-factor.

In practice, $\boldsymbol{\mu}$ and R are estimated from the data, but unlike the estimate of the target strength $\widehat{a}$ (which is done separately for each pixel), these estimates are based on the full dataset and for the purposes of the experiments here, are assumed to be accurately estimated. In fact, we will be working with whitened data, with zero mean and unit covariance.

Note that this is a "two step" process; with μ and R estimated in the first step; then fixed as $\widehat{a}$ is estimated for each pixel in the second step. If the total number of background pixels were limited, a "one step" procedure (in which μ, R, and a are simultaneously estimated) might be preferred [14].

Although we have performed experiments over a range of parameters, the results shown in Figs. 1,2 are based on $d = 9$, $\nu = 5$, and target strength $|\mathbf{t}| = 3$. In these experiments, we consider a discrete set $\mathbf{a} = \{0.3, 0.5\}$ of target strengths. In this case the Bayesian detector is based on weights w_1, w_2 with $q(a) = w_1\delta(a-0.3) + w_2\delta(a-0.5)$. Since we constrain $w_1 + w_2 = 1$, we only have one free parameter, w_1, which ranges from 0 to 1. Numerical results were based on $N = 10^8$ background pixels, and an equal number of target-implanted pixels for each value of a.

[2] Another advantage of the Bayesian approach is that it works well with non-parametric models of $P_{\text{bkg}}(\mathbf{x})$, which can be described as a sum over simple (typically kernel-like) terms. This is because the integral operator can be interchanged with the summation operator, leading to a sum of integrals of the individual terms [10].

[3] Yet another advantage of the Bayesian approach arises when the background distribution does not admit an analytical model. In that case, matched-pair machine learning [11,12] can be used to implement the detector by first implanting targets of abundances drawn from the prior and then by using binary classification to distinguish background from background+target.

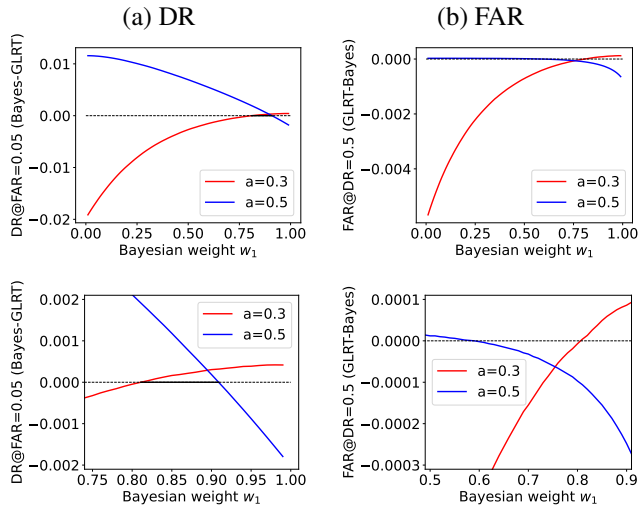

Fig. 1: Comparison of GLRT and Bayesian detectors over a range of values for the first weight in the Bayesian prior. The difference is plotted, with the direction of the subtraction set so that values above zero correspond to the Bayesian detector outperforming the GLRT detector, and for values below zero, the GLRT detector is better. Observe that the $a = 0.3$ (red) curves increase with increasing w_1, while the $a = 0.5$ (blue) curves decrease. The larger w_1, the more weight on the clairvoyant $a = 0.3$ detector, and the better the detector performs at $a = 0.3$. For the DR@FAR=0.05 criterion, we see that there is a range ($0.811 < w_1 < 0.909$) over which the Bayesian detector is better for all values of the unknown target strength (which, in this simplified problem, corresponds to just the two values $a = 0.3$ and $a = 0.5$). By contrast, for the FAR@DR=0.5 criterion, there is no value of w_1 for which the Bayesian detector is better at both target strengths. The bottom panels show the same data as the top panels, but are zoomed in to more clearly illustrate how the intersection of the red and blue lines are above the zero line for DR and below the zero line for FAR.

Fig. 1 compares the standard GLRT with this discrete Bayesian detector, and finds that for DR@FAR=0.05, there is a range of weights for which the Bayesian detector outperforms the GLRT detector at all (*i.e.*, both) target strengths. These Bayesian detectors are thus uniformly more powerful than the GLRT detector. By contrast, for FAR@DR=0.5, we find that there is no choice of prior for which the Bayesian detector uniformly outperforms (or even equals) the GLRT detector. These results are also seen in Fig. 2, which compares to the restricted RGLRT.

These experiments have been repeated with different parameter values, and while the same basic results were seen, some further observations were also made:

1. Using DR@FAR=x for $x = 0.005$ instead of $x = 0.05$, it was still possible to observe a range of w_1 values over which

the Bayesian outperformed the GLRT, but it was a different range than observed with $x = 0.05$. Thus, there is no single choice of Bayesian prior that leads to superior results at *all* false alarm rates.

2. A few cases were observed in which the Bayesian outperformed the GLRT using the FAR@DR=0.5 criterion, but no cases were found for which it outperformed the RGLRT detector at both target strengths.

3. With Gaussian background instead of multivariate t, the range of weights over which Bayesian beat RGLRT, using the DR@FAR=0.5 measure, was found to be *much* narrower; indeed, to within numerical precision, the best Bayesian and the RGLRT detectors exhibited equal performance.

4. DISCUSSION AND CONCLUSIONS

To be perfectly clear: what this paper describes is numerical experiments on simulated data; no theorems are proved, no actual data are investigated. What is demonstrated, however, is that statements about the theoretical superiority of Bayesian target detection (including remarks made by me [13]) do not apply as broadly as may have been assumed.

An example is provided for which *no* choice of prior enables the Bayesian detection algorithm to uniformly outperform the GLRT, based on the criterion of low false alarm rate at fixed detection rate. This was initially a surprise, but it can be understood by more carefully distinguishing individual detectors from detection statistics. The FAR@DR=0.5 criterion evaluates the performance, not of an individual detector, but of a full detection statistic; that's because the DR=0.5 constraint leads to a threshold that depends on the value of a.

This explains why GLRT *can* outperform Bayes on the FAR@DR=0.5 criterion, but does not explain why GLRT *does* outperform Bayes in this situation, and there may be situations where it does not. Given the practical utility of the FAR@DR=x performance criteria, this also justifies the use of the GLRT for target detection applications, and provides impetus for further research into potentially non-admissible detection schemes, such as clairvoyant fusion [15].

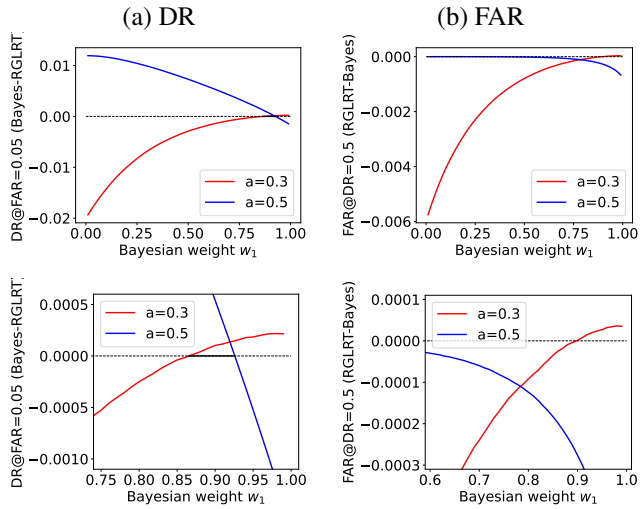

Fig. 2: Like Fig. 1 but with Restricted GLRT in place of the standard GLRT. The same qualitative behavior is observed.

5. REFERENCES

[1] D. G. Manolakis, R. B. Lockwood, and T. W. Cooley, *Hyperspectral Imaging Remote Sensing: Physics, Sensors, and Algorithms*, Cambridge University Press, Cambridge, 2016.

[2] A. Schaum and A. Stocker, "Spectrally selective target detection," *Proc. ISSSR (International Symposium on Spectral Sensing Research)*, p. 23, 1997.

[3] J. Theiler, B. Zimmer, and A. Ziemann, "Closed-form detector for solid sub-pixel targets in multivariate t-distributed background clutter," in *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2018, pp. 2773–2776.

[4] A. Schaum, "Hyperspectral target detection using a Bayesian likelihood ratio test," in *Proc. IEEE Aerospace Conference*, 2002, vol. 3, pp. 1537–1540.

[5] J. Theiler, "Homeopathic priors?," *arXiv*:2212.02725, 2022.

[6] J. Theiler, S. Matteoli, and A. Ziemann, "Bayesian detection of solid subpixel targets," in *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2021, pp. 3213–3216.

[7] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, Springer, New York, 2005.

[8] J. Theiler, "Veritas: an admissible detector for targets of unknown strength," *Proc. SPIE*, vol. 11727, pp. 117270B, 2021.

[9] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, vol. II, Prentice Hall, New Jersey, 1998.

[10] S. Matteoli, M. Diani, and G. Corsini, "Bayesian nonparametric detector based on the replacement model," in *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2022, pp. 871–874.

[11] J. Theiler, "Matched-pair machine learning," *Technometrics*, vol. 55, pp. 536–547, 2013.

[12] A. Ziemann, M. Kucer, and J. Theiler, "A machine learning approach to hyperspectral detection of solid targets," *Proc. SPIE*, vol. 10644, pp. 1064404, 2018.

[13] J. Theiler, "Confusion and clairvoyance: some remarks on the composite hypothesis testing problem," *Proc. SPIE*, vol. 8390, pp. 839003, 2012.

[14] F. Vincent and O. Besson, "One-step generalized likelihood ratio test for subpixel target detection in hyperspectral imaging," *IEEE Trans. Geoscience and Remote Sensing*, vol. 58, pp. 4479–4489, 2020.

[15] A. Schaum, "Continuum fusion: a theory of inference, with applications to hyperspectral detection," *Optics Express*, vol. 18, pp. 8171–8181, 2010.