ELLIPSOID-SIMPLEX HYBRID FOR HYPERSPECTRAL ANOMALY DETECTION

James Theiler

Space and Remote Sensing Sciences, Los Alamos National Laboratory; Los Alamos, NM 87544, USA

ABSTRACT

The problem of anomaly detection in hyperspectral imagery is expressed in terms of a minimal volume set in a high-dimensional space that encloses the bulk of the data samples. The venerable RX algorithm employs an ellipsoid for this volume, but endmember methods can be used to create a simplex volume. This paper considers a hybrid ellipsoid-simplex volume and characterizes its performance on hyperspectral imagery by computing a plot of volume versus false alarm rate. This plot provides a generic measure of quality (smaller volumes are better) without requiring the identification of specific anomalies in the data.

***Index Terms*—** hyperspectral imagery, anomaly detection, ellipsoid, simplex, endmember, volume, false alarm rate

1. INTRODUCTION

Anomaly detection is a kind of target detection problem in which the target is unknown. It's like looking for needles in a haystack, without knowing what a needle is. Despite the apparent philosophical conundrum inherent in the anomaly detection problem [1], there is nonetheless a practical utility in identifying those data elements that are unusual with respect to the others. This practical need has led to a variety of algorithms (see Refs. [2, 3] for recent surveys), including mathematically rigorous machine learning approaches [4, 5].

Let $\mathbf{x} \in \mathbb{R}^d$ represent a data element (such as a pixel in a d-channel hyperspectral image). The anomaly detector is a scalar function $\mathcal{A}(\mathbf{x})$, which corresponds to the "anomalousness" of pixel $\mathbf{x}$. When $\mathcal{A}(\mathbf{x}) > \eta$ for some threshold η, then $\mathbf{x}$ is declared an anomalous pixel. Given a labeled data set of anomalous and non-anomalous pixels, we can characterize the performance of $\mathcal{A}(\mathbf{x})$ with a receiver-operator characteristic (ROC) curve that compares false alarm rate (fraction of non-anomalous[1] pixels for which $\mathcal{A}(\mathbf{x}) > \eta$) to detection rate (fraction of anomalous pixels with $\mathcal{A}(\mathbf{x}) > \eta$), over a range of values of η. This is classic target detection methodology, but anomalies are unusual targets.

A representative and statistically meaningful set of anomalous pixels is not generally available, so detection rate is not readily measureable. A solution is to posit a distribution for anomalous pixels. To reflect the fact that we don't really know what anomalies are, that distribution should be broad and uninformative; a natural choice is a uniform distribution over a volume extends well beyond the non-anomalous data.[2] The detection rate is then the fraction of that volume for which $\mathcal{A}(\mathbf{x}) > \eta$. More measurable is the volume for which $\mathcal{A}(\mathbf{x}) < \eta$. Without making a fuss over what the constant multiplier is, it is clear that this volume is proportional to the *missed* detection rate. This motivates the plotting of volume versus false alarm rate to characterize the performance of an anomaly detector. The best anomaly detectors will be minimal volume sets that enclose most of the data.

2. MODELS FOR THE DISTRIBUTION OF THE (NON-ANOMALOUS) BACKGROUND

2.1. Ellipsoidal contours

One of the simplest and most useful models for the distribution of hyperspectral pixels is a multivariate Gaussian. Here,

$$P(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d |W|}} \exp\left[-\frac{1}{2} r^2(\mathbf{x})\right] \tag{1}$$

where

$$r(\mathbf{x}) = \left[(\mathbf{x}-\mu)^T W^{-1} (\mathbf{x}-\mu)\right]^{1/2} \tag{2}$$

is the Mahalanobis distance from the centroid. Traditionally, one obtains the centroid vector μ and the covariance matrix W from the first and second moments of the data, but variants have been suggested to accomodate estimation error arising either from the limited number of data samples or from deviations from the assumption of Gaussianity.

The contours of constant $P(\mathbf{x})$ will be ellipsoids centered at μ, and an anomaly detector that is fit to a Gaussian (or any other elliptically contoured distribution [8]) will have the form

$$\mathcal{A}(\mathbf{x}) = r(\mathbf{x}) = \left[(\mathbf{x}-\mu)^T W^{-1} (\mathbf{x}-\mu)\right]^{1/2}. \tag{3}$$

The volume enclosed by the d-dimensional ellipsoid is a function of the radius r:

$$V(r) = \frac{\pi^{d/2} |W|^{1/2}}{\Gamma(1+d/2)}\, r^d. \tag{4}$$

This work was supported by the Laboratory Directed Research and Development (LDRD) program at Los Alamos National Laboratory.

[1] Another word for this is "normal," but that word often connotes Gaussian distributions, and the aim here is to be more general than that.

[2] It is possible to extend the uniform distribution over all of $\mathbb{R}^d$; in this limit, it is not strictly a *distribution*, but is a *measure* [5].

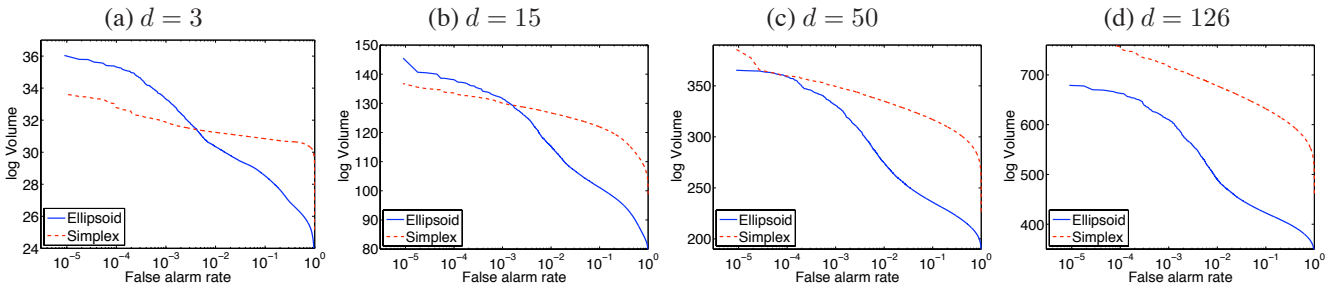

Fig. 1. Coverage plot shows the volume of a simplex or ellipsoid necessary to cover a fraction $1 - \alpha$ of the data, where α corresponds to the false alarm rate (*i.e.*, the fraction of pixels that are outside the volume). Models are fit to half of the data, and the volume-vs-α curves are computed on the other half. A smaller volume at a given α indicates a better model. The curves shown are for the 126-channel visible and near-infrared "blind test" radiance data [6], with dimension reduced using principal components analysis. It is observed for lower dimensions d that the simplex method (in this case, the Max-D [7] algorithm was used) produces better results at low false alarm rates (which is the regime of most interest in an anomaly detection scenario), but that advantage begins to fail at larger dimensions.

Having provided a way, in Eq. (2), to compute radius r for each point $\mathbf{x}$ in a dataset; and having an expression, in Eq. (4), for volume as a function of r, we can produce the coverage plots that were suggeted in Ref. [9]. For a given threshold radius, the false alarm rate is the fraction of the data for which r is larger than the threshold. As we sweep over threshold values, we can plot volume versus false alarm rate. The smaller the volume (corresponding to a smaller missed detection rate), the better the performance. These coverage curves are shown in Fig. 1; the solid lines are for ellipsoids.

2.2. Simplicial contours

A popular and often effective model for hyperspectral data is a k-dimensional simplex, defined by a set of $k + 1$ vertices (called "endmembers") in the d-dimensional space [10]. The model is motivated by the notion that pixels in the image are convex linear combinations of a relatively small number of distinct materials whose spectra are realized in the endmembers.

Let $E \in \mathbb{R}^{d\times(k+1)}$ be the matrix whose columns are the $k + 1$ endmembers

$$E = [\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_k] \tag{5}$$

where $\mathbf{e}_i \in \mathbb{R}^d$ is the ith endmember. We presume that endmembers span the k dimensional space (none of the endmembers can be written as linear combinations of the others).

A point $\mathbf{x}$ is in the *subspace* of the simplex if $E\mathbf{a} = \mathbf{x}$ for a vector $\mathbf{a}$ (of "abundances") which satisfies the sum-to-one rule; *i.e.*, $\mathbf{1}^T\mathbf{a} = 1$, where $\mathbf{1} \in \mathbb{R}^{k+1}$ is a vector of all 1's. That is:

$$\begin{bmatrix} \mathbf{1}^T \\ E \end{bmatrix} \mathbf{a} = \begin{bmatrix} 1 \\ \mathbf{x} \end{bmatrix}. \tag{6}$$

If the abundances are non-negative ($\mathbf{a} \succeq 0$), then $\mathbf{x}$ is in the *interior* of the simplex.

Using this simplex as a starting point, we will generate a family of nested simplices, and use them for anomaly detection in the same way that the nested ellipsoids were used for anomaly detection in the previous subsection. Define

$$\mathbf{e}_i(r) = (1 - r)\overline{\mathbf{e}} + r\mathbf{e}_i, \tag{7}$$

where $\overline{\mathbf{e}}$ is the centroid of the endmembers. Then for any r, we have $k + 1$ vertices $\mathbf{e}_i(r)$. Here r plays the role of a scalar-valued radius; when $r = 0$, then the simplex contracts to a point, and when $r = 1$, the original simplex is recovered. In general, as r increases, a larger simplex is obtained. (See Fig. 2.)

Further, in terms of these new endmembers, $\mathbf{e}_i(r)$, we can write[3] the abundances $\mathbf{a}(r)$ associated with the point $\mathbf{x}$ as

$$\mathbf{a}(r) = \frac{1}{r}\left[\mathbf{a} + \left(\frac{r-1}{k+1}\right)\mathbf{1}\right] \tag{8}$$

where $\mathbf{a}$ is the solution when $r = 1$. The point $\mathbf{x}$ is inside a simplex of radius r if $\mathbf{a}(r) \succeq 0$, or equivalently, if $\mathbf{a} \succeq (1-r)/(k+1)$. The radius of a point $\mathbf{x}$ is therefore given by

$$r(\mathbf{x}) = 1 - (k + 1)\min(\mathbf{a}), \tag{9}$$

where $\mathbf{a}$ solves Eq. (6). Just as Mahalanobis radius provides a measure of anomalousness for the ellipsoidal contours, we can use the simplex radius as a measure of anomalousness for the simplex contours: $\mathcal{A}(\mathbf{x}) = r(\mathbf{x})$.

To compute volume, choose one of the endmembers, say $\mathbf{e}_0$ (it will not matter which endmember is chosen), as a reference point. Write the reduced $d \times k$ matrix as differences with respect to this reference:

$$\widehat{E} = [\mathbf{e}_1 - \mathbf{e}_0, \mathbf{e}_2 - \mathbf{e}_0, \ldots, \mathbf{e}_k - \mathbf{e}_0]. \tag{10}$$

[3] This is a straightforward, though nontrivial, derivation.

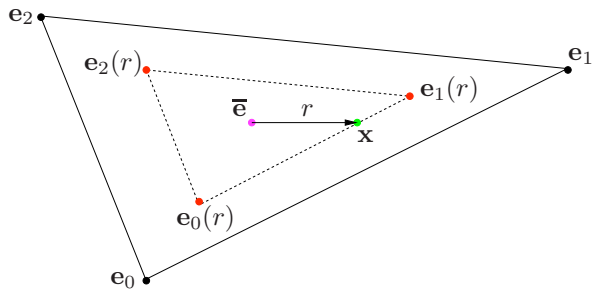

Fig. 2. Simplex delineated by three endmembers: $\mathbf{e}_0$, $\mathbf{e}_1$, and $\mathbf{e}_2$. The centroid of the simplex is $\overline{\mathbf{e}}$. The "radius" associated with a point $\mathbf{x}$ in the subspace of the simplex is the smallest r such that $\mathbf{x}$ is still inside the simplex given by $\mathbf{e}_0(r)$, $\mathbf{e}_1(r)$, and $\mathbf{e}_2(r)$, where $\mathbf{e}_i(r) = (1-r)\overline{\mathbf{e}} + r\mathbf{e}_i$.

When $d = k$, then $\widehat{E}$ is square, and its determinant is related to the volume of the simplex. But for $d \geq k$, we can create a full rank square matrix $\widehat{E}^T\widehat{E} \in \mathbb{R}^{k\times k}$, and use the square root of its determinant. In particular, we have

$$V_{\text{simplex}} = |\widehat{E}^T\widehat{E}|^{1/2}/k! \tag{11}$$

When using $\mathbf{e}_i(r)$ as endmembers, then it is straightforward to show that

$$V(r) = \frac{|\widehat{E}^T\widehat{E}|^{1/2}}{k!}\, r^k. \tag{12}$$

Although simplicial contours can exhibit improved anomaly detection performance (ie, more coverage with less volume), these improvements often fade at higher dimension. See Fig. 1; the dashed lines are for simplex-based anomaly detectors. For a hyperspectral image with, say, 126 channels, one would need 127 endmembers to describe the data. Thus, in addition to being computationally expensive, the simplex models appear ineffective at such high dimensions.

Although the simplex was motivated in terms of endmembers, any choice of simplex will work. In fact, this analysis provides an alternative measure of goodness for fitting a simplex to data: rather than qualitatively assess how accurately the $\mathbf{e}_i$ match "real" spectra, one can quantitatively compute simplex volume versus false alarm rate.

2.3. Ellipsoid-simplex hybrid

While hyperspectral data is not typically Gaussian, there is a sense (*e.g.*, see Refs. [11, 12, 13]) that it can be Gaussian in "some directions," particularly the directions with lower variance. This intuition suggests a hybrid model for characterizing hyperspectral data: simplex-like in the large-variance directions and ellipsoid-like in the low-variance directions.

We begin with a k-dimensional simplex, defined by $k+1$ endmembers. This simplex spans a k-dimensional subspace of the d dimensional hyperspectral data, and we will use that to express the d dimensional pixels $\mathbf{x}$ as the sum of a k-dimensional component and a $d-k$ dimensional component.

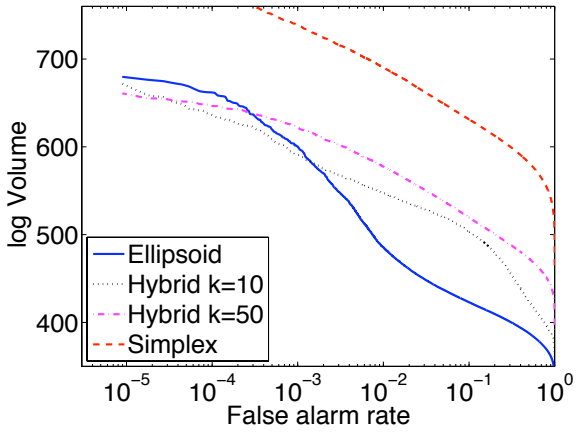

Fig. 3. Coverage plot for the ellipsoid-simplex hybrid, where k of the components are from the simplex and the remaining $d-k$ are the ellipsoid. While the pure ellipsoid does a better job of modeling the "core" of the data distribution, the periphery is better fit by the hybrid. At this high dimension, the pure simplex requires a huge volume to cover most of the data.

That is,

$$\mathbf{x} = \mathbf{x}_S + \mathbf{x}_E \tag{13}$$

where $\mathbf{x}_S$ is in the affine subspace of the simplex, and $\mathbf{x}_E$ is the residual. Here, $\mathbf{x}_S$ is modeled by a k-dimensional simplex, and $\mathbf{x}_E$ is modeled by a $d-k$ dimensional Gaussian.

Specifically, in terms of $\widehat{E} \in \mathbb{R}^{d\times k}$ defined in Eq. (10), we can write $\mathbf{x}_S = \mathbf{e}_0 + \widehat{E}\widehat{E}^{\#}(\mathbf{x}-\mathbf{e}_0)$, where $\widehat{E}^{\#} \in \mathbb{R}^{k\times d}$ is the pseudo-inverse of $\widehat{E}$. Then, $\mathbf{x}_E = \mathbf{x} - \mathbf{x}_S = (I - \widehat{E}\widehat{E}^{\#})(\mathbf{x}-\mathbf{e}_0)$.

For a given $\mathbf{x}$, we then compute two distinct radii: $r_E(\mathbf{x})$ is given by Eq. (2) applied to $\mathbf{x}_E$; and $r_S(\mathbf{x})$ is given by Eq. (9) applied to $\mathbf{x}_S$. Informally speaking, r_E and r_S indicate two qualitatively different senses of anomalousness: a small r_E and large r_S indicates an "in-plane" anomaly that is in the main subspace of the data but on the far outskirts of it; a small r_S and a large r_E is an "out-of-plane" anomaly that is near the centroid of the data, but off of the main subspace. To get a single measure, we take $r = \max(r_E, \beta r_S)$, where β is chosen so that the different radii have the same scale. In particular, we choose β so that the median of r_E is equal to the median of βr_S. In terms of r, then we have the volume of this hybrid as the product of the two components:

$$V_{\text{hybrid}} = \frac{\pi^{(d-k)/2}|W|^{1/2}}{\Gamma(1+(d-k)/2)}\,\frac{|\widehat{E}^T\widehat{E}|^{1/2}}{k!}\,\beta^{-k}r^d \tag{14}$$

See Fig. 3 and Fig. 4.

3. DISCUSSION

This paper presents two innovations. One is the use of an "elastic" simplex for anomaly detection. It is observed that

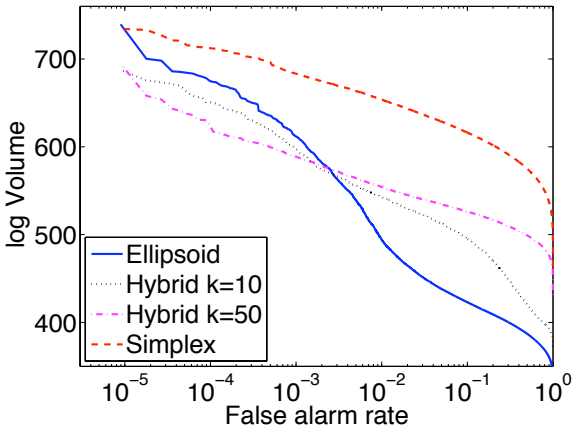

Fig. 4. Same as Fig. 3, except that a different simplex algorithm is used (NFINDR [14] instead of Max-D [7]), which leads to smaller volume simplices and ellipsoid-simplex hybrids.

for low-dimensional projections of hyperspectral data, *e.g.*, as seen in Fig. 1(a,b), the simplex provides a more effective shape than an ellipsoid for characterizing the tail of the data distribution.

Since anomalies might be lost in the projection to lower dimension, it is valuable to do anomaly detection in the full dimensional space. But for higher-dimensional data, the ellipsoid outperforms the simplex, as seen in Fig. 1(c,d). This leads to the second innovation, the hybrid ellipsoid-simplex model. As seen in Fig. 3, and even more so in Fig. 4, this hybrid produces the smallest-volume coverage of the data at the lowest false alarm rates.

There is ample room for improvement, here. In addition to optimizing the free parameters in this model (k and β), there are potentially better ways to choose the ellipsoid [9] and simplex – *e.g.*, see Fig. 4 – so as to optimize the volume versus false alarm rate criterion.

Stepping back from these particular examples, a key notion here is that the anomaly detection problem can be expressed in terms of minimum volume shapes that enclose most the data; and that these shapes can be adapted to the data. In addition to providing better anomaly detection performance, knowing the properties of shapes that are effective will provide insight into the nature of hyperspectral data distributions.

4. REFERENCES

[1] D. Rumsfeld, "As we know, there are known knowns. There are things we know we know. We also know there are known unknowns. That is to say we know there are some things we do not know. But there are also unknown unknowns, the ones we don't know we don't know," 2002, Department of Defense news briefing.

[2] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Processing Magazine*, vol. 19, pp. 58–69, Jan 2002.

[3] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE A&E Systems Magazine*, vol. 25, pp. 5–27, 2010.

[4] S. Ben-David and M. Lindenbaum, "Learning distributions by their density levels: A paradigm for learning without a teacher," *J. Computer and System Sciences*, vol. 55, pp. 171–182, 1997.

[5] I. Steinwart, D. Hush, and C. Scovel, "A classification framework for anomaly detection," *J. Machine Learning Research*, vol. 6, pp. 211–232, 2005.

[6] D. Snyder, J. Kerekes, I. Fairweather, R. Crabtree, J. Shive, and S. Hager, "Development of a web-based application to evaluate target finding algorithms," *Proc. IGARSS*, vol. 2, pp. 915–918, 2008.

[7] P. Bajorski, E. J. Ientilucci, and J. R. Schott, "Comparison of basis-vector selection methods for target and background subspaces as applied to subpixel target detection," *Proc. SPIE*, vol. 5425, pp. 97–108, 2004.

[8] D. Manolakis, D. Marden, J. Kerekes, and G. Shaw, "On the statistics of hyperspectral imaging data," *Proc. SPIE*, vol. 4381, pp. 308–316, 2001.

[9] J. Theiler and D. Hush, "Statistics for characterizing data on the periphery," *Proc. IGARSS*, pp. 4764–4767, 2010.

[10] J. W. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Summaries of the Fourth Annual JPL Airborne Geoscience Workshop*, Robert O. Green, Ed., 1994, pp. 11–14.

[11] S. M. Adler-Golden, "Improved hyperspectral anomaly detection in heavy-tailed backgrounds," *Proc. WHISPERS*, 2009.

[12] P. Bajorski, "Maximum Gaussianity models for hyperspectral images," *Proc. SPIE*, vol. 6966, pp. 69661M, 2008.

[13] J. Theiler, B. R. Foy, and A. M. Fraser, "Characterizing non-Gaussian clutter and detecting weak gaseous plumes in hyperspectral imagery," *Proc. SPIE*, vol. 5806, pp. 182–193, 2005.

[14] M. E. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proc. SPIE*, vol. 3753, pp. 266–275, 1999.