# RANDOMIZED RX FOR TARGET DETECTION

*Fatih Nar[1], Adrián Pérez-Suay[2], José Antonio Padrón[2], Gustau Camps-Valls[2]*

[1]Konya Food and Agriculture University, Konya, Turkey
[2]Image Processing Lab (IPL), Universitat de València, València, Spain

## ABSTRACT

This work tackles the target detection problem through the well-known global RX method. The RX method models the clutter as a multivariate Gaussian distribution, and has been extended to nonlinear distributions using kernel methods. While the kernel RX can cope with complex clutters, it requires a considerable amount of computational resources as the number of clutter pixels gets larger. Here we propose random Fourier features to approximate the Gaussian kernel in kernel RX and consequently our development keep the accuracy of the nonlinearity while reducing the computational cost which is now controlled by an hyperparameter. Results over both synthetic and real-world image target detection problems show space and time efficiency of the proposed method while providing high detection performance.

***Index Terms—*** Anomaly and target detection, Reed-Xiaoli (RX), kernel methods, random Fourier features

## 1. INTRODUCTION

Anomaly Detection (AD) in remote sensing data analysis has been (and will continue to be) one of the main important research topics for various applications [1, 2]. Anomaly decision is made by analyzing the difference of the pixel under test (PUT) and the background. Various anomaly detectors have been proposed in the literature under a number of assumptions: 1) the choice of the background region and the distribution, 2) the way of calculating the distance between the PUT and the background statistics, and 3) the definition of the decision threshold or test statistic [3].

Among the large variety of AD methods, the Reed-Xiaoli (RX) is widely used for its simplicity, good practical results and efficiency in operational settings [1, 2]. The RX algorithm, which probably is the most known AD method for hyperspectral images, it is based on assuming a multivariate Gaussian distributed background and to deal with, it uses

the Mahalanobis distance between the PUT and the background [4]. With regard to the background region, there are two main different approaches. In the global approach the whole image is defined as the background while in the local approach pixels around the PUT are defined as the background by means of a sliding window or a segmentation based approach [5, 6]. Also, possible anomalies can be censored to prevent degradation on the background statistics [7].

Even though RX is computationally efficient and works well for simple backgrounds (mostly linear), it can not cope with more complex, nonlinear and/or non-Gaussian backgrounds. To deal with its limitations, more powerful techniques like kernel methods can be used to cope with non-Gaussian distributed data [8–10]. However, kernel RX is a computationally demanding method because it involves matrix inversion in the size of the sample count, e.g. for more than thousand pixels. On the contrary, covariance matrix in RX has fixed size which is proportional to spectral dimension of the image, e.g. around ten for multispectral images and around hundred for hyperspectral images. Since AD is generally the first step in further recognition tasks, the execution time is as (or even more) important as the detection accuracy [11]. Besides, processor capacity (CPU or GPU), memory storage, and execution budget can be limited in operational applications [5]. In [12], a fast recursive kernel RX method was recently introduced, which processes the data in a causal manner. Although this method is fast, threshold estimation for target decision is a problematic task since the detection result is dependent on the location of the PUT in the image.

In this study, we propose an alternative way to speed up the kernel RX method while maintaining its accuracy: we propose to approximate the kernel function with random Fourier features [13] in order to project data into a nonlinear feature space where a standard, cheap and efficient linear RX method can be readily applied. Previous use of the randomized approach in remote sensing has considered classification and regression problems [14, 15]. The proposed randomized RX approach is able to significantly lower the execution time and the memory storage compared to its kernel RX counterpart, while preserving almost the same detection accuracy. We give empirical evidence in both synthetic and real-world remote sensing target detection problems.

The remainder of the paper is organized as follows. First, section 2 briefly reviews the proposed method used in this study. Then, section 3 presents the performance of the randomized RX method. Finally, we conclude in section 4 with some remarks and prospective future work.

## 2. RANDOMIZED RX

### 2.1. The RX algorithm

Let us define a hyperspectral image in matrix form $\mathbf{X} \in \mathbb{R}^{n \times d}$, where $n$ is the number of pixels and $d$ is the dimensionality of each pixel, i.e. number of spectral channels. A generic pixel element in $I$ is denoted as the (column) feature vector $\mathbf{x}_i \in \mathbb{R}^d$.

Among the various AD proposed in the literature, one of the most frequently used anomaly detector is the (spectral only version of the) Reed-Xiaoli (RX) detector [4] that is often used as a benchmark to which other methods are compared. The RX detector characterizes the background by its spectral mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}_{RX} = \frac{1}{d}\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}}$, where $\tilde{\mathbf{X}}$ is the centered $\mathbf{X}$ matrix. The actual detector calculates the Mahalanobis distance between the pixel under test, $\mathbf{x}_*$, and the background as follows

$$D_{RX}(\mathbf{x}_*) = (\mathbf{x}_* - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}_{RX}^{-1}(\mathbf{x}_* - \boldsymbol{\mu}). \qquad (1)$$

Note that we only need to invert the covariance matrix once. However, the global RX detector characterizes the background of the complete scene by a single multivariate normal probability density function (pdf). In many scenes, this model is not adequate. For this reason, several variations of the global RX detector have been proposed using kernel functions, such as the kernel RX (KRX) [9].

### 2.2. The kernel RX algorithm

Notationally, let us map all pixels to a higher dimensional Hilbert feature spaces $\mathcal{H}$ by means of the feature map $\phi : \mathbf{x} \to \phi(\mathbf{x})$. The mapped training data matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$ is now denoted as $\boldsymbol{\Phi} \in \mathbb{R}^{n \times d_{\mathcal{H}}}$. Let us define a kernel function $K$ that, by virtue of the Riesz theorem, can evaluate (reproduce) the dot product between samples in $\mathcal{H}$, i.e. $K(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle \in \mathbb{R}$. Now, in order to estimate the anomalousness (distance) for a test example $\mathbf{x}_*$, we first map it $\phi(\mathbf{x}_*)$ and then apply the RX formula in (1):

$$D_{KRX}(\mathbf{x}_*) = \phi(\mathbf{x}_*)^\top (\boldsymbol{\Phi}^\top \boldsymbol{\Phi})^{-1} \phi(\mathbf{x}_*),$$

which, after some linear algebra, can be expressed in terms of kernel matrices [8, 16]:

$$D_{KRX}(\mathbf{x}_*) = \mathbf{k}_*^\top (\mathbf{K}\mathbf{K})^{-1}\mathbf{k}_*, \qquad (2)$$

where $\mathbf{k}_* = [K(\mathbf{x}_*, \mathbf{x}_1), \dots, K(\mathbf{x}_*, \mathbf{x}_n)]^\top \in \mathbb{R}^n$ contains the similarities between $\mathbf{x}_*$ and all training points in $\mathbf{X}$ using $K$, and $\mathbf{K} \in \mathbb{R}^{n \times n}$ stands for the kernel matrix containing all training data similarities. Note that, if $n$ is very

large constructing and inverting kernel matrix is not feasible hence generally $N$ samples are randomly sampled to obtain computational tractability. Recently, the kernel RX has been also extended to non-Gaussian settings by defining a complete family of anomalous change detectors with kernels [10].

### 2.3. The randomized RX algorithm

An outstanding result in the recent kernel methods literature makes use of a classical definition in harmonic analysis to improve approximation and scalability [13]. The Bochner's theorem states that a continuous shift-invariant kernel $K(\mathbf{x}, \mathbf{x}') = K(\mathbf{x} - \mathbf{x}')$ on $\mathbb{R}^d$ is positive definite (p.d.) if and only if $K$ is the Fourier transform of a non-negative measure. If a shift-invariant kernel $K$ is properly scaled, its Fourier transform $p(\mathbf{w})$ is a proper probability distribution. This property is used to approximate kernel functions with linear projections on a number of $D$ random features, as follows:

$$K(\mathbf{x}, \mathbf{x}') \approx \frac{1}{D} \sum_{j=1}^{D} \exp(-\mathrm{i}\mathbf{w}_j^\top \mathbf{x}) \exp(\mathrm{i}\mathbf{w}_j^\top \mathbf{x}'),$$

where $\mathrm{i} = \sqrt{-1}$, and $\mathbf{w}_i \in \mathbb{R}^d$ is randomly sampled from a data-independent distribution $p(\mathbf{w})$ [17]. Note that we can define a $D$-dimensional *randomized* feature map $\mathbf{z} : \mathbb{R}^d \to \mathbb{C}^D$, which can be *explicitly* constructed as $\mathbf{z}(\mathbf{x}) := [\exp(\mathrm{i}\mathbf{w}_1^\top \mathbf{x}), \dots, \exp(\mathrm{i}\mathbf{w}_D^\top \mathbf{x})]^\top$.

Therefore, given $n$ data points (pixels), the kernel matrix $\mathbf{K} \in \mathbb{R}^{n \times n}$ can be approximated with the explicitly mapped data, $\mathbf{Z} = [\mathbf{z}_1 \cdots \mathbf{z}_n]^\top \in \mathbb{R}^{n \times D}$, and will be denoted as $\hat{\mathbf{K}} \approx \mathbf{Z}\mathbf{Z}^\top$. However, we do not use such approximation in Eq. (2) which would lead to a mere approximation with extra computational cost. Instead, we run the linear RX in Eq. (1) with explicitly mapped points onto random Fourier features, which reduces to

$$D_{RRX} = \Re\left(\mathbf{z}_*^\top (\mathbf{Z}^\top \mathbf{Z})^{-1}\mathbf{z}_*\right), \qquad (3)$$

where $\mathbf{z}_* = \mathbf{z}(\mathbf{x}_*)$, and $\Re$ is the real part function $\Re(a + \mathrm{i}b) = a$. This leads to a nonlinear randomized RX (RRX) that approximates the KRX in expectation. Essentially, we map the original data $\mathbf{x}_i$ into a nonlinear space through the explicit mapping $\mathbf{z}(\mathbf{x}_i)$ to a $D$-dimensional space (instead of the potentially infinite feature space with $\phi(\mathbf{x}_i)$), and then use the linear RX formula. This allows to control the space and time complexity explicitly through $D$, as one has to store matrices of $n \times D$ and invert matrices of size $D \times D$ only (see Table 1). Typically, parameter $D$ satisfies $D \ll n$ in practical applications, which turns to be a beneficial regularization effect in the solution.

## 3. EXPERIMENTS

This section reports some empirical results comparing RX, the KRX and the proposed RRX, in both synthetic and real experiments which are run on Intel i7 4.0 GHz PC.

**Table 1**: Space and time complexity for all methods.

| Method | Space | | Time | | | |
|---|---|---|---|---|---|---|
| | $T$ | $C^{-1}$ | $T$ | $C$ | $C^{-1}$ | $AD$ |
| RX | $-$ | $d^2$ | $-$ | $nd^2$ | $d^3$ | $nd^2$ |
| KRX | $nN$ | $N^2$ | $ndN$ | $N^3$ | $N^3$ | $nN^2$ |
| RRX | $nD$ | $D^2$ | $ndD$ | $nD^2$ | $D^3$ | $nD^2$ |

$T$ is transformation of image into a nonlinear space.
$C$ is covariance matrix and $C^{-1}$ is its inverse.
$AD$ is anomaly detection.

### 3.1. Synthetic dataset

To compare the detectors in a controlled manner, we created a synthetic test image of size $100 \times 100$ pixels with two spectral channels. Fig. 1 shows data distribution in a density colormap, here dark points are background and bright points are anomalies. Besides, Fig.2.a shows image as false colored (the first and the second dimensions assigned to the green and blue channels, respectively). In this data, $2.72\%$ of the pixels are cast as anomalies (Fig. 2.d).
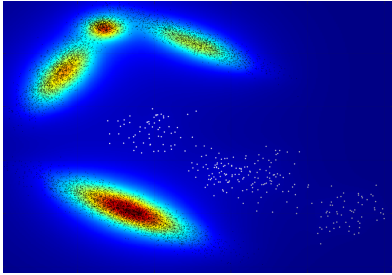


**Fig. 1**: Distribution of synthetic data.

We added a small regularization value $\lambda = 10^{-2}$ to better condition matrix inversions, and fixed the kernel lengthscale $\sigma$ as the median of all pairwi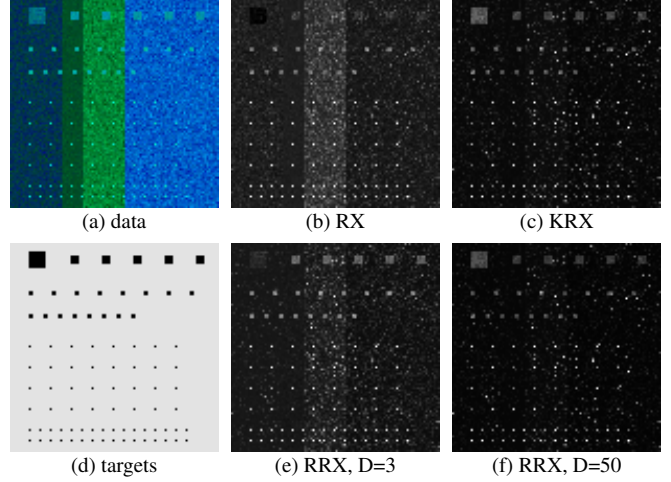se distances in both KRX and RRX. Detection responses are given in Fig. 2 using whole image in training for RX and RRX, and randomly selected subset ($N = 3000$) in training for KRX. Note that background in synthetic data (cf. Fig. 1) is complex and non-Gaussian distributed. For this data with complex background, RRX $D = 3$ fails since 3 Fourier basis is not sufficient to model the background and RX also fails since background is non-Gaussian distributed. However, both RRX with only 50 basis ($D = 50$) and KRX ($N = 3000$) are able to model background properly.

In Fig. 3 appears the ROC curves for all methods and they are given in log scale (abcissa) where area under the curve (AUC) values and execution times are given in legends. Since background is not Gaussian distributed, RX has the lowest performance. KRX and the proposed RRX with $D = 50$ basis show a similar superior performance, which is an expected result. However, computational load of KRX increases rapidly as the number of training samples increase. In this paper, we used $N = 3000$ randomly sampled pixels for KRX to achieve satisfactory detection results. The RRX detector with just $D = 50$ basis has similar detection performance yet two orders of magnitude faster.



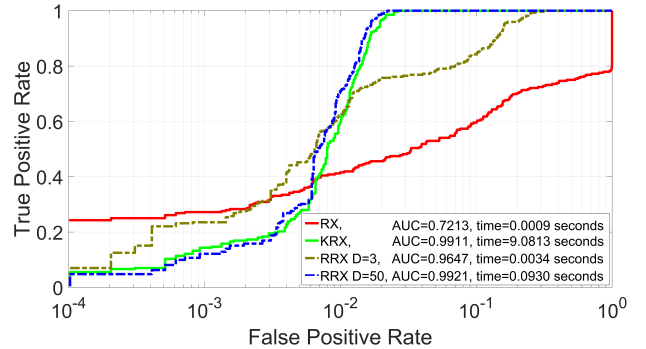**Fig. 2**: Detection responses for synthetic data



**Fig. 3**: ROC curve for synthetic data.

### 3.2. Real-world dataset

For the real-world test, a multispectral image of California with 12 spectral channels and size of $964 \times 332$ is used (see RGB composite in Fig. 4). California image is acquired from Sentinel-2 satellite at 8 August 2017 and downloaded from Google Earth Engine (GEE).



**Fig. 4**: California dataset (3 channels are used for false color)

The image was divided into 16 non-overlapping patches, each having size of $100 \times 100$ pixels and anomalous pixels were synthetically fused into these patches using the mask given in Fig.2.d. Half of the patches are used as a training set and the remaining half as the validation set. The regularization parameter $\lambda$ was varied between $10^{-5}$ and $10^0$, and the

lengthscale of the kernel was varied by taking values between $[0.05, 5]$ and considering this particular values multiplied with $m$, where $m$ is the median of all pairwise distances. For each $\lambda$ and $c$ pairs, an average AUC value is calculated over the validation set with $8$ patches. The obtained optimal hyperparameters were $\lambda = 10^{-4}$ and $c = 0.5$.

Test results are evaluated for the red square with size of $100 \times 100$ which is shown in Fig. 4. Test results are given in Fig. 5 in log scale where AUC values and execution times are given in legends. Proposed RRX can obtain similar performance with KRX using $D = 100$ basis and hence it is two orders of magnitude faster for this data set. Thus, better accuracies, robustness and computational efficiency is attained with the proposed RRX.
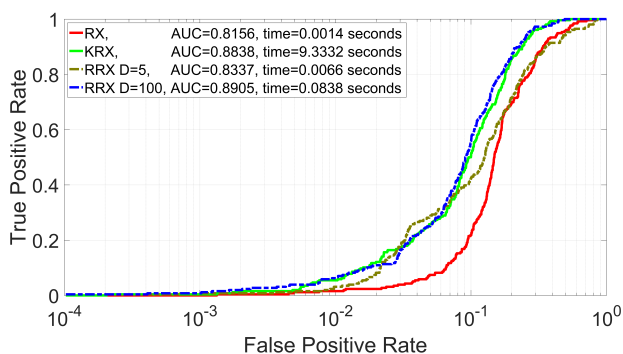


**Fig. 5**: ROC curve for California data.

## 4. CONCLUSIONS

We introduced the use of random Fourier features in the context of anomaly detection. In particular, we focused on approximating the KRX method to cope with complex backgrounds in an computationally efficient manner. The proposed RRX provides space and time efficient solutions for anomaly detection, while providing high detection accuracy in complex nonlinear and non-Gaussian background distributions. In our experiments, we observed that proposed RRX method can provide two orders of magnitude speedup compared to KRX method with no loss in accuracy. Other kernelized detectors can benefit from this strategy, which will be the object of our future research.

## 5. REFERENCES

[1] 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, no. 1, pp. 58–69, 2002.

[2] N. Keshava, "Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries," *IEEE Trans. Geosc. Rem. Sens.*, vol. 42, no. 7, pp. 1552–1565, 2004.

[3] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aerospace and Electronic Systems Magazine*, vol. 25, no. 7, pp. 5–28, 2010.

[4] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1760–1770, 1990.

[5] J. M. Molero, E. M. Garzn, I. Garca, and A. Plaza, "Analysis and optimizations of global and local versions of the RX algorithm for anomaly detection in hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 2, pp. 801–814, 2013.

[6] Chunhui Zhao, Yulei Wang, Bin Qi, and Jia Wang, "Global and local real-time anomaly detectors for hyperspectral remote sensing imagery," *Remote Sensing*, vol. 7, no. 4, pp. 3966–3985, 2015.

[7] F. Nar, E. Okman, A. Ozgur, and M. Cetin, "Fast target detection in radar images using Rayleigh mixtures and summed area tables," *Digital Signal Processing*, vol. 77, pp. 86–101, 2018.

[8] G. Camps-Valls and L. Bruzzone, Eds., *Kernel methods for remote sensing data analysis*, Wiley & Sons, UK, Dec 2009.

[9] H. Kwon and N. Nasrabadi, "A comparative analysis of kernel subspace target detectors for hyperspectral imagery," *EURASIP Journal of of Advances in Signal Proc.*, vol. 2007, no. 29250, 2007.

[10] Nathan Longbotham and G. Camps-Valls, "A family of kernel anomaly change detectors," in *IEEE Workshop on Hyperspectral Image and Signal Processing, Whispers*, June 2014.

[11] S. Y. Chen, Y. Wang, C. C. Wu, C. Liu, and C. I. Chang, "Real-time causal processing of anomaly detection for hyperspectral imagery," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 2, pp. 1511–1534, 2014.

[12] Chunhui Zhao, Xifeng Yao, and Bormin Huang, "Real-time anomaly detection based on a fast recursive kernel RX algorithm," *Remote Sensing*, vol. 8, no. 12, 2016.

[13] Ali Rahimi and Benjamin Recht, "Random features for large-scale kernel machines," in *Advances in Neural Information Processing Systems 20*, pp. 1177–1184. 2008.

[14] V. Laparra, D.M. Gonzalez, D. Tuia, and G. Camps-Valls, "Large-scale random features for kernel regression," in *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*, 2015, pp. 17–20.

[15] A. Perez-Suay, J. Amoros, L. Gomez-Chova, V. Laparra, Mu noz Marí, and G. Camps-Valls, "Randomized kernels for large scale earth observation applications," *Remote Sensing of Environment*, vol. 1, no. 1, pp. 1, 2017.

[16] John Shawe-Taylor and Nello Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.

[17] Ali Rahimi and Benjamin Recht, "Weighted sums of random kitchen sinks: Replacing minimization with randomization in learning," in *Advances in Neural Information Processing Systems 21*, pp. 1313–1320. Curran Associates, Inc., 2009.