Matched shrunken subspace detectors for hyperspectral target detection

Ziyu Wang[a,b], Jing-Hao Xue[b,*]

[a] *Department of Security and Crime Science, University College London, London WC1E 6BT, UK*
[b] *Department of Statistical Science, University College London, London WC1E 6BT, UK*

Abstract

In this paper we propose a new approach, called the matched shrunken subspace detector (MSSD), to target detection from hyperspectral images. The MSSD is developed by shrinking the abundance vectors of the target and background subspaces in the hypothesis models of the matched subspace detector (MSD), a popular subspace-based approach to target detection. The shrinkage is achieved by introducing simple l_2-norm regularisation (also known as ridge regression or Tikhonov regularisation). We develop two types of MSSD, one with isotropic shrinkage and termed MSSD-i and the other with anisotropic shrinkage and termed MSSD-a. For these two new methods, we provide both the frequentist and Bayesian derivations. Experiments on a real hyperspectral imaging dataset called Hymap demonstrate that the proposed MSSD methods can outperform the original MSD for hyperspectral target detection.

Keywords: Matched subspace detector (MSD), matched shrunken subspace detector (MSSD), shrinkage estimation, target detection, hyperspectral image (HSI)

*Corresponding author. Tel.: +44-20-7679-1863; Fax: +44-20-3108-3105
Email addresses: `ziyu.wang.12@ucl.ac.uk` (Ziyu Wang), `jinghao.xue@ucl.ac.uk` (Jing-Hao Xue)

1. Introduction

Target detection or anomaly detection is an important task of hyperspectral image (HSI) analysis [1, 2, 3, 4, 5, 6]. To target detection, the matched subspace detector (MSD) [7, 8] is one of the most widely-used subspace-based approaches, underlying which is the idea of the linear mixing model (LMM) [9].

The LMM [9] is a typical approach to unmixing a mixed pixel. Suppose there are p spectral bands and thus a mixed pixel $\mathbf{x}$ is represented by a p-dimensional vector/spectrum. Let us assume there are K types of materials potentially constituting a pixel; these component materials are often referred to as endmembers, the spectra of which can be represented by $\mathbf{m}_1, \ldots, \mathbf{m}_K$, where each $\mathbf{m}_k$ is a p-dimensional vector. Then the LMM of pixel $\mathbf{x}$ models the spectral signature of $\mathbf{x}$ as a linear combination of endmembers $\mathbf{m}_1, \ldots, \mathbf{m}_K$ with corresponding abundance fractions $a_1, \ldots, a_K$. More specifically, $\mathbf{x} = [x_1, \ldots, x_p]^T$ can be expressed as an additive mixture of K endmembers $\mathbf{m}_k$ plus noise:

$$\mathbf{x} = \Sigma_{k=1}^{K} a_k \mathbf{m}_k + \mathbf{n} = \mathbf{M}\mathbf{a} + \mathbf{n}, \tag{1}$$

where $\mathbf{M}$ is a $p \times K$ matrix whose columns are the K endmember spectra $\mathbf{m}_k = [m_{k,1}, \ldots, m_{k,p}]^T$ for $k = 1, \ldots, K$, respectively; $\mathbf{a} = [a_1, \ldots, a_K]$ denotes the abundance vector; and $\mathbf{n} = [n_1, \ldots, n_p]^T$ represents the additive Gaussian white noise, i.e. $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, where $\mathbf{I}$ is a $p \times p$ identity matrix. In classical unmixing problems, the abundances $a_1, \ldots, a_K$ need to satisfy two conditions, which are the non-negative constraint and the sum-to one constraint, i.e. $a_k \geq 0$ and $\Sigma_{k=1}^{K} a_k = 1$, respectively. However, in target detection problems, as explained in [9], both constraints will complicate the solution; as usually is the case, we can relax both constraints in target detection.

To achieve an HSI target detection, the MSD determines whether a test pixel can be represented by a linear combination of target spectral signatures and background spectral signatures. To this end, two subspaces are constructed: the target subspace and the background subspace. In each subspace, the MSD assumes that each basis vector represents an endmember, which is in line with the assumption of the LMM for HSI analysis.

To construct the two subspaces, the MSD usually acquires their basis vectors from the eigen-decomposition of covariance matrices of the training samples [1, 10]. The eigenvectors with dominant eigenvalues, termed leading eigenvectors, are selected as bases to span the subspaces, while those with small eigenvalues are discarded. This is essentially a scheme of basis selection, or say 0/1 weighting, which extracts a subspace out of the full eigenspace.

In fact, the 0/1 weighting scheme of the MSD implicitly imposes a sparseness constraint or say an l_0-norm regularisation while building its LMM. However, it is well known that such a "hard" selection may exhibit high variance on the selected leading eigenvectors. Alternatively, explicit sparse representation (SR)-based techniques have also been developed in hyperspectral target detection [11, 12, 13], with selection of a small number of atoms from a large dictionary. That is, these SR methods model a test HSI pixel as a linear combination of only few atoms from an over-complete dictionary; atoms in the dictionary are usually also samples, hence these SR methods can be viewed as being developed in the original sample space. Regarding the construction of the dictionary, [11] propose to construct a background spectra dictionary and a target spectra dictionary separately; on the other hand, [12, 13] propose to construct an over-complete dictionary including both background spectra and target spectra.

To avoid the problem of high variance from such a "hard" selection, shrinkage methods [14] have been developed in statistical learning, mainly due to such a problem in regression analysis. Among the shrinkage methods, the most popular one is called ridge regression, also known as Tikhonov regularisation [15] in other disciplines; it shrinks the regression coefficients through imposing an l_2-norm constraint. In this way, the estimates of the coefficients become more stable and therefore can improve the performance of regression.

The l_2-norm regularisation has been investigated for analysing hyperspectral imagery [16, 17, 4, 18, 19, 20]. For the HSI classification, [16] and [17] assume that a test pixel can be collaboratively represented by raw spectral signatures. It is shown that l_2-norm constraints can actually improve the classification, instead of the "competitive" nature imposed by sparseness constraints (as l_1-

norm or l_0-norm regularisation). For the HSI target detection, [4, 18, 19, 20] add a scaled identity matrix to the background clutter covariance matrix before inverting it, in order to avoid an ill-conditioned problem. It is worth noting that these l_2-norm regularisation methods are developed in the original sample space, rather than in the eigenspace as this work.

In this paper, focusing on the popular MSD, we propose a new approach, called the matched shrunken subspace detector (MSSD), to target detection from hyperspectral images. Our MSSD is developed by shrinking the abundance vectors of the target and background subspaces in the hypothesis models of the MSD. The shrinkage is simply achieved by introducing l_2-norm regularisation into the MSD. We develop two types of the MSSD, one with isotropic shrinkage (and termed MSSD-i) and the other with anisotropic shrinkage (and termed MSSD-a). For these two new methods, we provide both the frequentist and Bayesian derivations. Experiments on a real hyperspectral imaging dataset called Hymap demonstrate that the proposed MSSD-i and MSSD-a can outperform the original MSD for hyperspectral target detection.

The main contributions of this paper are two-fold. 1) Through introducing the l_2-norm regularisation terms into the MSD, we shrink the abundance vectors so that the variance in each basis direction of the subspaces is also reduced, leading to a more stable estimation. 2) We derive the proposed MSSD-i and MSSD-a from both the frequentist and Bayesian perspectives, with the latter showing how the proposed methods preserve Gaussian prior distributions of the abundance vectors, instead of the uniform prior distribution which is implicitly imposed by the original MSD.

The rest of this paper is organised as follows. Section 2 reviews the original MSD. In section 3.1 and section 3.2, detailed formulation of the two proposed method, MSSD-i and MSSD-a, are introduced. Then the two proposed methods are derived from the Bayesian perspective and shown in section 4. The links of MSD, MSSD-i and MSSD-a are discussed in section 5. Section 6 presents the experimental results, with the whole work concluded in section 7.

2. Matched subspace detector (MSD)

2.1. Overview of the binary hypothesis testing model

From a statistical perspective, target detection is typically derived from a binary hypothesis testing problem [3]. It is based on the likelihood ratio of the conditional probability density functions (pdfs) of two competing hypotheses, given that the spectral signature of an HSI pixel $\mathbf{x}$ is treated a continuous random vector:

$$\begin{aligned} &H_0 : \mathbf{x} \text{ is a background pixel,} \\ &H_1 : \mathbf{x} \text{ is a target pixel,} \\ &\Rightarrow D(\mathbf{x}) = \frac{f_{\mathbf{x}|H_1}(\mathbf{x})}{f_{\mathbf{x}|H_0}(\mathbf{x})} \underset{H_0}{\overset{H_1}{\gtrless}} \nu, \end{aligned} \tag{2}$$

where $f_{\mathbf{x}|H_0}(\mathbf{x})$ and $f_{\mathbf{x}|H_1}(\mathbf{x})$ are two conditional pdfs of $\mathbf{x}$ under the null hypothesis H_0 and the alternative hypothesis H_1, respectively; ν is the detection threshold; and $D(\mathbf{x})$ is an output detector. In reality, the conditional pdfs are usually not available and are expressed parametrically. Hence, the generalised likelihood ratio test (GLRT) [21] is commonly used to replace the unknown parameters by their maximum likelihood estimates (MLEs):

$$\begin{aligned} D_{GLRT}(\mathbf{x}) &= \frac{f_{\mathbf{x}|H_1}(\mathbf{x};\hat{\omega}_1)}{f_{\mathbf{x}|H_0}(\mathbf{x};\hat{\omega}_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \nu \\ &= \frac{\max_{\omega_1}\{f_{\mathbf{x}|H_1}(\mathbf{x};\omega_1)\}}{\max_{\omega_0}\{f_{\mathbf{x}|H_0}(\mathbf{x};\omega_0)\}} \underset{H_0}{\overset{H_1}{\gtrless}} \nu, \end{aligned} \tag{3}$$

where ω_0 and ω_1 are unknown parameters of pdf $f_{\mathbf{x}|H_0}(\mathbf{x};\omega_0)$ and pdf $f_{\mathbf{x}|H_1}(\mathbf{x};\omega_1)$, respectively; and $\hat{\omega}_0$ and $\hat{\omega}_1$ are their MLEs. In this paper, "ˆ" denotes the estimates of unknown parameters.

2.2. Formulation of the matched subspace detector (MSD)

Following the idea of LMM (1) [9], the MSD models a test pixel by a linear combination of target spectral endmembers and background spectral endmembers, and these endmembers are represented by the basis vectors of the target subspace and the background subspace, respectively.

That is, derived from the binary hypothesis model (2), the MSD model [7] is constructed as

$$
\begin{aligned}
&H_0 : \mathbf{x} = \mathbf{B}\boldsymbol{\beta} + \mathbf{n}_0,\ \mathbf{x} \text{ is a background pixel,}\\
&H_1 : \mathbf{x} = \mathbf{T}\boldsymbol{\gamma} + \mathbf{B}\boldsymbol{\beta} + \mathbf{n}_1,\ \mathbf{x} \text{ is a target pixel,}
\end{aligned} \tag{4}
$$

where $\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_{r_t}]$ is a $p \times r_t$ matrix representing the target subspace, and $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_{r_b}]$ is a $p \times r_b$ matrix representing the background subspace; $\mathbf{T}$ is derived from a training target matrix $\mathbf{M}_T \in \mathbb{R}^{p\times N_t}$ whose columns are the N_t target spectra, and $\mathbf{B}$ is derived from a training background matrix $\mathbf{M}_B \in \mathbb{R}^{p\times N_b}$ whose columns are the N_b background spectra; $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are the corresponding abundance vectors of the subspaces $\mathbf{T}$ and $\mathbf{B}$, respectively; and $\mathbf{n}_0$ and $\mathbf{n}_1$ are p-dimensional vectors of Gaussian white noise: $\mathbf{n}_0 \sim \mathcal{N}(\mathbf{0}, \sigma_0^2\mathbf{I})$ and $\mathbf{n}_1 \sim \mathcal{N}(\mathbf{0}, \sigma_1^2\mathbf{I})$.

In general, a set of orthogonal basis vectors that spans the corresponding subspace are used as the column vectors of $\mathbf{T}$ or $\mathbf{B}$. In common practice, the leading eigenvectors of the target covariance matrix $\mathbf{C}_T$ and those of the background covariance matrix $\mathbf{C}_B$ are used as the columns of $\mathbf{T}$ and $\mathbf{B}$, respectively, as with [10][1]. In other words, when the test pixel $\mathbf{x}$ is a target pixel, it is decomposed into two components by linear combinations of the bases of $\mathbf{B}$ and $\mathbf{T}$, denoted by model H_1. When $\mathbf{x}$ is a background pixel, it is adequately described by model H_0, which is a reduced order model.

Let $\mathbf{V}$ be the concatenated matrix of $\mathbf{T}$ and $\mathbf{B}$, i.e. $\mathbf{V} = [\mathbf{T}\ \mathbf{B}] = [\mathbf{t}_1, \ldots, \mathbf{t}_{r_t}, \mathbf{b}_1, \ldots, \mathbf{b}_{r_b}]$, then the abundance vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ of model H_1 can be concatenated into a single vector, denoted as $\boldsymbol{\alpha}$, i.e. $\boldsymbol{\alpha} = \begin{bmatrix}\boldsymbol{\gamma}\\ \boldsymbol{\beta}\end{bmatrix} = [\gamma_1, \ldots, \gamma_{r_t}, \beta_1, \ldots, \beta_{r_b}]^T$. Hence model H_1 can be written as

$$
\begin{aligned}
H_1 : \mathbf{x} &= \mathbf{T}\boldsymbol{\gamma} + \mathbf{B}\boldsymbol{\beta} + \mathbf{n}_1\\
&= \begin{bmatrix}\mathbf{T} & \mathbf{B}\end{bmatrix}\begin{bmatrix}\boldsymbol{\gamma}\\ \boldsymbol{\beta}\end{bmatrix} + \mathbf{n}_1\\
&= \mathbf{V}\boldsymbol{\alpha} + \mathbf{n}_1,
\end{aligned} \tag{5}
$$

and thus the MSD model (4) becomes

$$\begin{aligned} &H_0 : \mathbf{x} = \mathbf{B}\boldsymbol{\beta} + \mathbf{n}_0,\ \mathbf{x} \text{ is a background pixel,} \\ &H_1 : \mathbf{x} = \mathbf{V}\boldsymbol{\alpha} + \mathbf{n}_1,\ \mathbf{x} \text{ is a target pixel,} \end{aligned} \tag{6}$$

where now the unknown parameters are $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and those of $\mathbf{n}_0$ and $\mathbf{n}_1$.

The corresponding estimate of the likelihood ratio is the generalised likelihood ratio (GLR) of the MSD, formulated as

$$\begin{aligned} \hat{l}(\mathbf{x}) &= \frac{l(\hat{\boldsymbol{\alpha}}, \hat{\sigma}_1^2; \mathbf{x})}{l(\hat{\boldsymbol{\beta}}, \hat{\sigma}_0^2; \mathbf{x})} \\ &= \left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_0^2}\right)^{-p/2} \exp\left\{-\frac{1}{2\hat{\sigma}_1^2}\left\|\hat{\mathbf{n}}_1\right\|_2^2 + \frac{1}{2\hat{\sigma}_0^2}\left\|\hat{\mathbf{n}}_0\right\|_2^2\right\}. \end{aligned} \tag{7}$$

The MLEs $\hat{\sigma}_0^2$ and $\hat{\sigma}_1^2$ are equal to $\frac{1}{p}\left\|\hat{\mathbf{n}}_0\right\|_2^2$ and $\frac{1}{p}\left\|\hat{\mathbf{n}}_1\right\|_2^2$, respectively. Taking the $2/p$ power of (7), we have the following GLR of the MSD:

$$\begin{aligned} L_{MSD}(\mathbf{x}) &= (\hat{l}(\mathbf{x}))^{2/p} \\ &= \left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_0^2}\right)^{-1} = \frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2} \\ &= \frac{\left\|\hat{\mathbf{n}}_0\right\|_2^2}{\left\|\hat{\mathbf{n}}_1\right\|_2^2} = \frac{\left\|\mathbf{x} - \mathbf{B}\hat{\boldsymbol{\beta}}\right\|_2^2}{\left\|\mathbf{x} - \mathbf{V}\hat{\boldsymbol{\alpha}}\right\|_2^2}. \end{aligned} \tag{8}$$

The MLEs of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ in (8) are given by

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}}\left\{f_{\mathbf{x}|H_0}(\mathbf{x}; \boldsymbol{\beta}, \sigma_0^2)\right\} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\left\{\frac{1}{2\sigma_0^2}\left\|\mathbf{x} - \mathbf{B}\boldsymbol{\beta}\right\|_2^2\right\} \tag{9}$$

and

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmax}}\left\{f_{\mathbf{x}|H_1}(\mathbf{x}; \boldsymbol{\alpha}, \sigma_1^2)\right\} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}}\left\{\frac{1}{2\sigma_1^2}\left\|\mathbf{x} - \mathbf{V}\boldsymbol{\alpha}\right\|_2^2\right\}, \tag{10}$$

and thus

$$\hat{\boldsymbol{\beta}} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{x} = \mathbf{B}^T\mathbf{x} \tag{11}$$

and

$$\hat{\boldsymbol{\alpha}} = (\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T\mathbf{x}. \tag{12}$$

It is to be noted that the bases $[\mathbf{b}_1, \ldots, \mathbf{b}_{r_b}]$ of $\mathbf{B}$ are orthogonal, therefore $(\mathbf{B}^T\mathbf{B})^{-1}$ is an identity matrix and $\hat{\boldsymbol{\beta}}$ can be simplified to $\mathbf{B}^T\mathbf{x}$, but the bases $[\mathbf{t}_1, \ldots, \mathbf{t}_{r_t}, \mathbf{b}_1, \ldots, \mathbf{b}_{r_b}]$ of $\mathbf{V}$ are not orthogonal to each other.

Based on (11) and (12), the residual sums of squares (RSS) e_0 and e_1 given model H_0 and model H_1 are computed as

$$H_0 : e_0 = \|\hat{\mathbf{n}}_0\|_2^2 = \left\|\mathbf{x} - \mathbf{B}\hat{\boldsymbol{\beta}}\right\|_2^2 = \mathbf{x}^T(\mathbf{I} - \mathbf{B}\mathbf{B}^T)\mathbf{x}, \tag{13}$$

and

$$H_1 : e_1 = \|\hat{\mathbf{n}}_0\|_2^2 = \|\mathbf{x} - \mathbf{V}\hat{\boldsymbol{\alpha}}\|_2^2 = \mathbf{x}^T(\mathbf{I} - \mathbf{V}(\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T)\mathbf{x}, \tag{14}$$

where $\mathbf{I}$ is a $p \times p$ identity matrix. The final GLRT detector of the MSD model is then given by

$$D_{MSD}(\mathbf{x}) = \frac{e_0}{e_1} = \frac{\mathbf{x}^T(\mathbf{I} - \mathbf{B}\mathbf{B}^T)\mathbf{x}}{\mathbf{x}^T(\mathbf{I} - \mathbf{V}(\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T)\mathbf{x}} \underset{H_0}{\overset{H_1}{\gtrless}} \nu. \tag{15}$$

The value of D_{MSD} is compared to a threshold ν to make the final decision of which hypothesis should be rejected for the test pixel $\mathbf{x}$. Two tuning parameters should be determined for the MSD, which are the numbers of leading eigenvectors to be preserved in the subspace $\mathbf{B}$ and $\mathbf{T}$, i.e. r_b and r_t, respectively.

3. Matched shrunken subspace detector (MSSD)

In the MSD, the eigenvectors spanning the eigenspace are either preserved or discarded to build the subspaces. Rather than applying this selection scheme, it is desirable to adopt shrinkage schemes to reduce the variance induced by selection [14], in order to develop a more stable statistical method like the MSD, in particular for high-dimensional data like hyperspectral pixels. In the l_2-norm regularised shrinkage methods, all the available features/eigenvectors are preserved and their coefficients are shrunk. In other words, r_b and r_t are fixed to the maximal numbers of available features/eigenvectors. We propose to introduce l_2-norm regularisation into the MSD, to shrink the abundance vectors of the target and background subspaces in the hypothesis models of the MSD. We call this approach the matched shrunken subspace detector (MSSD).

It is worth noting that, in the hyperspectral target detection practice, we often have only one target spectrum as a priori information for training, and this single target spectrum usually comes from the spectrum library. If this is the case, the target training sample $\mathbf{M}_T$ is a single vector, not a matrix, and thus the typical eigen-decomposition cannot be applied on $\mathbf{M}_T$ to get $\mathbf{T}$. To this end and as usually is the case, we use the normalised mean-corrected target spectrum as the only basis vector of the target subspace $\mathbf{T}$. As a result, we have $r_t = 1$ and $\mathbf{T} \in \mathbb{R}^{p\times 1}$, and the MSD does not discard this basis vector. Similarly, we do not shrink the abundance $\boldsymbol{\gamma}$ for the target subspace $\mathbf{T}$ when there is only one target spectrum available in practice, as also discussed in section 6.

In the following sections, we shall develop two types of the MSSD, MSSD-i with isotropic shrinkage and MSSD-a with anisotropic shrinkage, and provide both the frequentist and Bayesian derivations of them.

3.1. MSSD with isotropic shrinkage (MSSD-i)

While preserving all available eigenvectors, we introduce l_2-norm regularisation terms $\theta_0 \|\boldsymbol{\beta}\|_2^2$ and $\theta_1 \|\boldsymbol{\alpha}\|_2^2$ as constraints to the hypothesis models H_0 and H_1 of the MSD, respectively. The shrunken estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ now become

$$\hat{\boldsymbol{\beta}}_{iso} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\{\|\mathbf{x} - \mathbf{B}\boldsymbol{\beta}\|_2^2 + \theta_0 \|\boldsymbol{\beta}\|_2^2\} \tag{16}$$

and

$$\hat{\boldsymbol{\alpha}}_{iso} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}}\{\|\mathbf{x} - \mathbf{V}\boldsymbol{\alpha}\|_2^2 + \theta_1 \|\boldsymbol{\alpha}\|_2^2\}, \tag{17}$$

where θ_0 and θ_1 are the parameters that control the degree of shrinkage imposed on the size of abundance vectors $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, respectively. In this sense, the same shrinkage degree is applied to all eigenvectors, as done in (16) and (17), and we call this new method the MSSD with isotropic shrinkage, shortened as MSSD-i.

The test likelihood ratio of the MSSD-i is thus given by

$$L_{MSSD_{iso}}(\mathbf{x}) = \frac{\min_{\boldsymbol{\beta}}\{\|\mathbf{x} - \mathbf{B}\boldsymbol{\beta}\|_2^2 + \theta_0 \|\boldsymbol{\beta}\|_2^2\}}{\min_{\boldsymbol{\alpha}}\{\|\mathbf{x} - \mathbf{V}\boldsymbol{\alpha}\|_2^2 + \theta_1 \|\boldsymbol{\alpha}\|_2^2\}} \underset{H_0}{\overset{H_1}{\gtrless}} \nu, \tag{18}$$

and the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ in the MSSD-i are readily given as

$$\hat{\boldsymbol{\beta}}_{iso} = ((1 + \theta_0)\mathbf{I}_0)^{-1}\,\mathbf{B}^T\mathbf{x} \tag{19}$$

and

$$\hat{\boldsymbol{\alpha}}_{iso} = (\mathbf{V}^T\mathbf{V} + \theta_1\mathbf{I}_1)^{-1}\mathbf{V}^T\mathbf{x}, \tag{20}$$

where $\mathbf{I}_0$ is a $r_b \times r_b$ identity matrix and $\mathbf{I}_1$ is $(r_t+r_b)\times(r_t+r_b)$ identity matrix. Hence the RSS e_0 and e_1 given models H_0 and H_1 are computed as

$$H_0 : e_0^{iso} = \left\|\mathbf{x} - \mathbf{B}\hat{\boldsymbol{\beta}}_{iso}\right\|_2^2 = \mathbf{x}^T(\mathbf{I} - \mathbf{B}\left((1+\theta_0)\mathbf{I}_0)^{-1}\mathbf{B}^T\right)\mathbf{x}, \tag{21}$$

and

$$H_1 : e_1^{iso} = \|\mathbf{x} - \mathbf{V}\hat{\boldsymbol{\alpha}}_{iso}\|_2^2 = \mathbf{x}^T(\mathbf{I} - \mathbf{V}(\mathbf{V}^T\mathbf{V} + \theta_1\mathbf{I}_1)^{-1}\mathbf{V}^T)\mathbf{x}. \tag{22}$$

As with (15), the detector of the MSSD-i model is finally given by

$$D_{MSSD_{iso}}(\mathbf{x}) = \frac{e_0^{iso}}{e_1^{iso}} = \frac{\mathbf{x}^T(\mathbf{I} - \mathbf{B}\left((1+\theta_0)\mathbf{I}_0)^{-1}\mathbf{B}^T\right)\mathbf{x}}{\mathbf{x}^T(\mathbf{I} - \mathbf{V}(\mathbf{V}^T\mathbf{V} + \theta_1\mathbf{I}_1)^{-1}\mathbf{V}^T)\mathbf{x}} \underset{H_0}{\overset{H_1}{\gtrless}} \nu, \tag{23}$$

To be noticed, the MSSD-i also has two tuning parameters, but not the r_b and r_t of the MSD: this time the tuning parameters are the shrinkage parameters θ_0 and θ_1.

3.2. MSSD with anisotropic shrinkage (MSSD-a)

Besides the directions represented by eigenvectors, the values of eigenvalues also reflect the information about distributions, in particular variances, of the data in the background and target subspaces. Therefore in addition to the MSSD-i, we propose another new method which preserves not just the useful information from all the available eigenvectors, but also the information of all the eigenvalues, while constructing the l_2-norm regularisation terms for the MSD.

Let $\mathbf{\Lambda_B}$ denote the background eigenvalue matrix with the eigenvalues of the background eigenvectors $\lambda_1^b, \ldots, \lambda_{r_b}^b$ on the diagonal, i.e. $\mathbf{\Lambda_B} = \text{diag}([\lambda_1^b, \ldots, \lambda_{r_b}^b]^T)$; and let $\mathbf{\Lambda_T}$ denote the target eigenvalue matrix with the eigenvalues of the target eigenvectors $\lambda_1^t, \ldots, \lambda_{r_t}^t$ on the diagonal, i.e. $\mathbf{\Lambda_T} = \text{diag}([\lambda_1^t, \ldots, \lambda_{r_t}^t]^T)$.

It is known that small eigenvalues correspond to the eigenvectors having small variances, therefore we aim to shrink these directions the most. To this end, we can add the inverse of the eigenvalue matrix, $\mathbf{\Lambda_B^{-1}}$, to the regularisation term $\boldsymbol{\beta}^T\boldsymbol{\beta}$, for example. The shrunken estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ now become

$$\hat{\boldsymbol{\beta}}_{aniso} = \underset{\boldsymbol{\beta}}{\text{argmin}}\left\{(\mathbf{x} - \mathbf{B}\boldsymbol{\beta})^T(\mathbf{x} - \mathbf{B}\boldsymbol{\beta}) + \theta_0\boldsymbol{\beta}^T\mathbf{\Lambda_B^{-1}}\boldsymbol{\beta}\right\} \tag{24}$$

and

$$\hat{\boldsymbol{\alpha}}_{aniso} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \left\{ (\mathbf{x} - \mathbf{V}\boldsymbol{\alpha})^T (\mathbf{x} - \mathbf{V}\boldsymbol{\alpha}) + \theta_1 \boldsymbol{\alpha}^T \boldsymbol{\Lambda}_{\mathbf{V}}^{-1} \boldsymbol{\alpha} \right\}, \tag{25}$$

where θ_0 and θ_1 are again the parameters for the shrinkage degrees, and $\boldsymbol{\Lambda}_{\mathbf{V}}$ is a concatenated matrix formed as

$$\boldsymbol{\Lambda}_{\mathbf{V}} = \begin{bmatrix} \boldsymbol{\Lambda}_{\mathbf{T}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{\mathbf{B}} \end{bmatrix}. \tag{26}$$

Compared with (16) and (17) which shrink isotropically over features in MSSD-i, both (24) and (25) shrink anisotropically over features. Hence we call this new method the MSSD with anisotropic shrinkage, shortened as MSSD-a.

As with (18), the test likelihood ratio of the MSSD-a is given by

$$L_{MSSD_{aniso}}(\mathbf{x}) = \frac{\min_{\boldsymbol{\beta}} \{ \|\mathbf{x} - \mathbf{B}\boldsymbol{\beta}\|_2^2 + \theta_0 \boldsymbol{\beta}^T \boldsymbol{\Lambda}_{\mathbf{B}}^{-1} \boldsymbol{\beta} \}}{\min_{\boldsymbol{\alpha}} \{ \|\mathbf{x} - \mathbf{V}\boldsymbol{\alpha}\|_2^2 + \theta_1 \boldsymbol{\alpha}^T \boldsymbol{\Lambda}_{\mathbf{V}}^{-1} \boldsymbol{\alpha} \}} \underset{H_0}{\overset{H_1}{\gtrless}} \nu, \tag{27}$$

and the estimates of $\boldsymbol{\beta}_{aniso}$ and $\boldsymbol{\alpha}_{aniso}$ are

$$\hat{\boldsymbol{\beta}}_{aniso} = (\mathbf{I}_0 + \theta_0 \boldsymbol{\Lambda}_{\mathbf{B}}^{-1})^{-1} \mathbf{B}^T \mathbf{x} \tag{28}$$

and

$$\hat{\boldsymbol{\alpha}}_{aniso} = (\mathbf{V}^T \mathbf{V} + \theta_1 \boldsymbol{\Lambda}_{\mathbf{V}}^{-1})^{-1} \mathbf{V}^T \mathbf{x}. \tag{29}$$

The RSS e_0^{aniso} and e_1^{aniso} given models H_0 and H_1 are then computed as

$$\begin{aligned} H_0 : e_0^{aniso} &= \left\| \mathbf{x} - \mathbf{B}\hat{\boldsymbol{\beta}}_{aniso} \right\|_2^2 \\ &= \mathbf{x}^T (\mathbf{I} - \mathbf{B}(\mathbf{I}_0 + \theta_0 \boldsymbol{\Lambda}_{\mathbf{B}}^{-1})^{-1} \mathbf{B}^T) \mathbf{x} \end{aligned} \tag{30}$$

and

$$\begin{aligned} H_1 : e_1^{aniso} &= \| \mathbf{x} - \mathbf{V}\hat{\boldsymbol{\alpha}}_{aniso} \|_2^2 \\ &= \mathbf{x}^T (\mathbf{I} - \mathbf{V}(\mathbf{V}^T \mathbf{V} + \theta_1 \boldsymbol{\Lambda}_{\mathbf{V}}^{-1})^{-1} \mathbf{V}^T) \mathbf{x}. \end{aligned} \tag{31}$$

As with (15) and (23), the detector of the MSSD-a model can be written as

$$\begin{aligned} D_{MSSD_{aniso}}(\mathbf{x}) &= \frac{e_0^{aniso}}{e_1^{aniso}} \\ &= \frac{\mathbf{x}^T (\mathbf{I} - \mathbf{B}(\mathbf{I}_0 + \theta_0 \boldsymbol{\Lambda}_{\mathbf{B}}^{-1})^{-1} \mathbf{B}^T) \mathbf{x}}{\mathbf{x}^T (\mathbf{I} - \mathbf{V}(\mathbf{V}^T \mathbf{V} + \theta_1 \boldsymbol{\Lambda}_{\mathbf{V}}^{-1})^{-1} \mathbf{V}^T) \mathbf{x}} \underset{H_0}{\overset{H_1}{\gtrless}} \nu, \end{aligned} \tag{32}$$

Similar to MSSD-i, only two tuning parameters are need to be determined in the proposed MSSD-a: the shrinkage parameters θ_0 and θ_1.

4. Bayesian derivations of MSSD-i and MSSD-a

From the Bayesian perspective, the estimation of parameters $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ in the MSSD-i and the MSSD-a can be translated as the maximisation of a posteriori probability (MAP). Taking $\boldsymbol{\beta}$ for example, Bayes' theorem [14] says

$$f(\boldsymbol{\beta}|\mathbf{x}) = \frac{f(\mathbf{x}|\boldsymbol{\beta})f(\boldsymbol{\beta})}{f(\mathbf{x})}, \tag{33}$$

where $f(\mathbf{x}|\boldsymbol{\beta})$ is a likelihood function of $\mathbf{x}$ and $f(\boldsymbol{\beta})$ is a prior distribution of $\boldsymbol{\beta}$. Therefore the MAP estimate of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}}\, f(\boldsymbol{\beta}|\mathbf{x}) = \underset{\boldsymbol{\beta}}{\operatorname{argmax}}\, f(\mathbf{x}|\boldsymbol{\beta})f(\boldsymbol{\beta}). \tag{34}$$

As the noise term $\mathbf{n}_0$ is assumed to be a multivariate Gaussian distribution $\mathbf{n}_0 \sim \mathcal{N}(\mathbf{0}, \sigma_0^2\mathbf{I})$ in the LMM [9] and the MSD [7], the likelihood function $f(\mathbf{x}|\boldsymbol{\beta})$ can be formulated as

$$f(\mathbf{x}|\boldsymbol{\beta}) \propto \exp\left\{-\frac{1}{2\sigma_0^2}\|\mathbf{x} - \mathbf{B}\boldsymbol{\beta}\|_2^2\right\}. \tag{35}$$

In the conventional MSD, an improper uniform (non-informative) prior distribution is actually assumed for parameter $\boldsymbol{\beta}$ of the selected leading eigenvectors. In the proposed MSSD-i and MSSD-a, adding l_2-norm regularisation in fact imposes Gaussian prior distributions on $\boldsymbol{\beta}$.

4.1. Prior distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ in MSSD-i

For the MSSD-i, the prior distribution of $\boldsymbol{\beta}$ is in fact assumed to be

$$\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \sigma_B^2\mathbf{I}_0), \tag{36}$$

with equal variance σ_B^2 in each element β_i of $\boldsymbol{\beta}$ for $i = 1, \ldots, r_b$. Thus $f(\boldsymbol{\beta})$ is given by

$$f(\boldsymbol{\beta}) \propto \exp\left\{-\frac{1}{2\sigma_B^2}\|\boldsymbol{\beta}\|_2^2\right\}. \tag{37}$$

Placing (35) and (37) into (34) and taking logarithm, we have

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{iso} &= \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \log\{f(\boldsymbol{\beta}|\mathbf{x})\} \\ &= \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \log\{f(\mathbf{x}|\boldsymbol{\beta})f(\boldsymbol{\beta})\} \\ &= \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \left\{-\frac{1}{2\sigma_0^2}\|\mathbf{x}-\mathbf{B}\boldsymbol{\beta}\|_2^2 - \frac{1}{2\sigma_B^2}\left\|\boldsymbol{\beta}\right\|_2^2\right\} \\ &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{\|\mathbf{x}-\mathbf{B}\boldsymbol{\beta}\|_2^2 + \theta_0\left\|\boldsymbol{\beta}\right\|_2^2\right\},\end{aligned} \tag{38}$$

where $\theta_0 = \sigma_0^2/\sigma_B^2$. The estimate of $\boldsymbol{\beta}$ in (38) is exactly the same as the MSSD-i estimate in (16). In this fashion, parameter θ_0 effectively controls the degree of shrinkage through the ratio of two variances σ_0^2 and σ_B^2.

Similarly, the prior distribution of $\boldsymbol{\gamma}$ is in fact assumed to be

$$\boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \sigma_T^2\mathbf{I}_t), \tag{39}$$

where $\mathbf{I}_t$ is a $r_t \times r_t$ identity matrix and therefore it results in a zero mean distribution of $\boldsymbol{\alpha}$ with an $(r_t + r_b) \times (r_t + r_b)$ diagonal covariance matrix

$$\begin{bmatrix} \sigma_T^2\mathbf{I}_t & \mathbf{0} \\ \mathbf{0} & \sigma_B^2\mathbf{I}_0 \end{bmatrix}. \tag{40}$$

Then $f(\boldsymbol{\alpha})$ is given by

$$f(\boldsymbol{\alpha}) = \prod_{i=1}^{r_t+r_b} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{1}{2\sigma_i^2}\alpha_i^2\right\}, \tag{41}$$

where $\sigma_i = \sigma_T$ for $i = 1, \ldots, r_t$ and $\sigma_i = \sigma_B$ for $i = r_t + 1, \ldots, r_t + r_b$. When $\sigma_B = \sigma_T$ and we let both of them to be σ_α, (41) can be simplified to

$$\begin{aligned} f(\boldsymbol{\alpha}) &= \frac{1}{(2\pi\sigma_\alpha^2)^{(r_t+r_b)/2}} \exp\left\{-\frac{1}{2\sigma_\alpha^2}\left\|\boldsymbol{\alpha}\right\|_2^2\right\} \\ &\propto \exp\left\{-\frac{1}{2\sigma_\alpha^2}\left\|\boldsymbol{\alpha}\right\|_2^2\right\}. \end{aligned} \tag{42}$$

Then placing the likelihood function and the prior distribution (42) into the MAP estimate of $\boldsymbol{\alpha}$ we have

$$\hat{\boldsymbol{\alpha}}_{iso} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \left\{\|\mathbf{x}-\mathbf{V}\boldsymbol{\alpha}\|_2^2 + \theta_1\left\|\boldsymbol{\alpha}\right\|_2^2\right\}, \tag{43}$$

where $\theta_1 = \sigma_1^2/\sigma_\alpha^2$ is the shrinkage parameter. This is also in the same form of the MSSD-i estimate of $\boldsymbol{\alpha}$ in (17), in particular if we assume $\sigma_T = \sigma_B$.

We can further generalise (43) to a slightly-adaptive shrinkage model:

$$\hat{\boldsymbol{\alpha}}_{iso} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \left\{ \|\mathbf{x} - \mathbf{V}\boldsymbol{\alpha}\|_2^2 + \sum_{i=1}^{r_t+r_b} \theta_{1i}\alpha_i^2 \right\}. \tag{44}$$

In (44), when $i = 1, \dots, r_t$, we have $\theta_{1i} = \sigma_1^2/\sigma_T^2$, and when $i = r_t+1, \dots, r_t+r_b$, we have $\theta_{1i} = \sigma_1^2/\sigma_B^2$.

4.2. Prior distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ in MSSD-a

For MSSD-a, the prior distribution of $\boldsymbol{\beta}$ is in fact assumed to be

$$\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \theta_B \boldsymbol{\Lambda}_\mathbf{B}), \tag{45}$$

where $\boldsymbol{\Lambda}_\mathbf{B}$ is a $r_b \times r_b$ diagonal matrix with eigenvalues $\lambda_1^b, \dots \lambda_{r_b}^b$ on the diagonal, and θ_B is a parameter scaling the eigenvalue matrix $\boldsymbol{\Lambda}_\mathbf{B}$. It means that each β_i, for $i = 1, \dots, r_b$, is assumed to have its own variance instead of an equal variance assumed in the MSSD-i. Then $f(\boldsymbol{\beta})$ in MSSD-a is given by

$$f(\boldsymbol{\beta}) \propto \exp\left\{ -\frac{1}{2} \boldsymbol{\beta}^T (\theta_B \boldsymbol{\Lambda}_\mathbf{B})^{-1} \boldsymbol{\beta} \right\}. \tag{46}$$

Placing (35) and (46) into (34) and taking logarithm, we have the MAP estimator of $\boldsymbol{\beta}$ in MSSD-a:

$$\hat{\boldsymbol{\beta}}_{aniso} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ (\mathbf{x} - \mathbf{B}\boldsymbol{\beta})^T (\mathbf{x} - \mathbf{B}\boldsymbol{\beta}) + \theta_0 \boldsymbol{\beta}^T \boldsymbol{\Lambda}_\mathbf{B}^{-1} \boldsymbol{\beta} \right\}, \tag{47}$$

where $\theta_0 = \sigma_0^2/\theta_B$. This is the same as the MSSD-a estimate of $\boldsymbol{\beta}$ in (24).

The prior distribution of $\boldsymbol{\gamma}$ is assumed to be

$$\boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \theta_T \boldsymbol{\Lambda}_\mathbf{T}), \tag{48}$$

where $\boldsymbol{\Lambda}_\mathbf{T}$ is a $r_t \times r_t$ diagonal matrix with different eigenvalues $\lambda_1^t, \dots \lambda_{r_t}^t$ on the diagonal, and θ_T is a parameter scaling the eigenvalue matrix $\boldsymbol{\Lambda}_\mathbf{T}$. Therefore the distribution of $\boldsymbol{\alpha}$ is a zero mean distribution with a $(r_t + r_b) \times (r_t + r_b)$ diagonal covariance matrix

$$\begin{bmatrix} \theta_T \boldsymbol{\Lambda}_\mathbf{T} & \mathbf{0} \\ \mathbf{0} & \theta_B \boldsymbol{\Lambda}_\mathbf{B} \end{bmatrix}. \tag{49}$$

If we let θ_T and θ_B both be equal to θ_v, then the prior distribution of $\boldsymbol{\alpha}$ will be $\boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0}, \theta_v \boldsymbol{\Lambda}_\mathbf{V})$, where $\boldsymbol{\Lambda}_\mathbf{V} = \begin{bmatrix} \boldsymbol{\Lambda}_\mathbf{T} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_\mathbf{B} \end{bmatrix}$.

Similar to (41), $f(\boldsymbol{\alpha})$ is given by

$$f(\boldsymbol{\alpha}) \propto \exp\left\{-\frac{1}{2}\boldsymbol{\alpha}^T(\theta_v \boldsymbol{\Lambda}_\mathbf{V})^{-1}\boldsymbol{\alpha}\right\}. \tag{50}$$

Then the MAP estimate of $\boldsymbol{\alpha}$ becomes

$$\hat{\boldsymbol{\alpha}}_{aniso} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \left\{(\mathbf{x} - \mathbf{V}\boldsymbol{\alpha})^T(\mathbf{x} - \mathbf{V}\boldsymbol{\alpha}) + \theta_1 \boldsymbol{\alpha}^T \boldsymbol{\Lambda}_\mathbf{V}^{-1}\boldsymbol{\alpha}\right\}, \tag{51}$$

where $\theta_1 = \sigma_1^2/\theta_v$. This is also exactly the same as the MSSD-a estimate of $\boldsymbol{\alpha}$ in (25).

Again, we can generlise (51) to a slightly-adaptive shrinkage model:

$$\hat{\boldsymbol{\alpha}}_{aniso} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \left\{\|\mathbf{x} - \mathbf{V}\boldsymbol{\alpha})\|_2^2 + \sum_{i=1}^{r_b+r_t} \frac{\theta_{1i}}{\lambda_i}\alpha_i^2\right\}. \tag{52}$$

In (52), when $i = 1, \ldots, r_t$, we have $\theta_{1i} = \sigma_1^2/\theta_T$ and $\lambda_i = \lambda_i^t$, and when $i = r_t + 1, \ldots, r_t + r_b$, we have $\theta_{1i} = \sigma_1^2/\theta_B$ and $\lambda_i = \lambda_{i-r_t}^b$.

To sum up, in contrast to the improper uniform distributions assumed in the MSD, two different prior distributions are assumed by the proposed MSSD-i and MSSD-a for the abundance vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ for the background and target subspaces. In the MSSD-i, a common variance is assumed on each coefficient in the form of a scaled identity matrix (see (37) and (39)). In the MSSD-a, unequal variances are assumed for individual coefficients in the form of a scaled eigenvalue matrix (see (46) and (48)).

5. Underlying links among MSD, MSSD-i and MSSD-a

The conventional MSD preserves the leading eigenvectors to form the subspaces $\mathbf{B}$ and $\mathbf{T}$, which is essentially a basis selection process. Specifically, it drops eigenvectors of small eigenvalues, effectively forcing these eigenvalues to be 0. At the same time, eigenvalues of the preserved eigenvectors are effectively forced to be equal to each other. The proposed MSSD-i and MSSD-a on the other hand, preserve all available eigenvectors and control the degrees of shrinkage of abundance by imposing l_2-norm regularisation. Specifically, the MSSD-i imposes an isotropic shrinkage over the full eigenspace, while the MSSD-a is anisotropic using eigenvalues to adapt the shrinkage for different directions.

From the Bayesian perspective, the conventional MSD implies a non-informative uniform distribution for the coefficient vectors over infinite interval. Different from the MSD, the proposed MSSD-i and MSSD-a imply Gaussian prior distributions for the coefficient vectors: the MSSD-i assumes an equal variance for each coefficient, while the MSSD-a assumes different variances for different coefficients which are based on eigenvalues.

Nevertheless, it is readily seen that the MSSD-i is equivalent to a ridge regression on the eigenspace. Also, as a kind of dual representation, the proposed MSSD-a can also be derived as a ridge regression on the original sample space. Specifically regarding this derivation of MSSD-a, if we apply the LMM in the original N_b-dimensional sample space of the $p \times N_b$ training sample matrix $\mathbf{M}_B$ under model H_0 with mean-corrected measurement. That is, supposing $\mathbf{M}_B$ is a mean-corrected matrix and pixel $\mathbf{x}$ is represented as a linear mixture of N_b samples, we have

$$\mathbf{x} = \mathbf{M}_B\mathbf{a} + \mathbf{n}, \tag{53}$$

where $\mathbf{a}$ is an $N_b \times 1$ coefficient vector, and the ridge regression problem becomes

$$\hat{\mathbf{a}}_{iso} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}}\{\|\mathbf{x} - \mathbf{M}_B\mathbf{a}\|_2^2 + \theta_M \|\mathbf{a}\|_2^2\}, \tag{54}$$

where $\hat{\mathbf{a}}_{iso}$ is the shrunken estimator of $\mathbf{a}$ and θ_M is the parameter controlling the shrinkage. The solution of $\hat{\mathbf{a}}_{iso}$ is

$$\hat{\mathbf{a}}_{iso} = (\mathbf{M}_B^T\mathbf{M}_B + \theta_M\mathbf{I}_b)^{-1}\mathbf{M}_B^T\mathbf{x}, \tag{55}$$

where $\mathbf{I}_b$ is a $N_b \times N_b$ identity matrix.

Following the notation in [14], if we perform the singular value decomposition (SVD) on $\mathbf{M}_B$, saying $p < N_b$, we obtain

$$\mathbf{M}_B = \mathbf{U}\mathbf{D}\mathbf{V}^T, \tag{56}$$

where $\mathbf{U}$ and $\mathbf{V}$ are $p \times p$ and $N_b \times N_b$ orthogonal matrices, with columns of $\mathbf{U}$ spanning the column space of $\mathbf{M}_B$ and columns of $\mathbf{V}$ spanning the row space of $\mathbf{M}_B$; and $\mathbf{D}$ is a $p \times N_b$ rectangular diagonal matrix with singular values of $\mathbf{M}_B$ on the diagonal in descending order. Based on the relationship between this SVD and the eigen-decomposition of covariance matrix $\mathbf{C}_B$ in MSSD-a, we have

1) $\mathbf{U} = \mathbf{B}$ ($r_b = p$ in this case) and

2) $\mathbf{D}_p^2 = N_b \mathbf{\Lambda_B}$,

where $\mathbf{D}_p$ is a $p \times p$ diagonal matrix of the first p columns of $\mathbf{D}$. Then the solution of $\mathbf{M}_B \hat{\mathbf{a}}_{iso}$ has the following form:

$$\begin{aligned}
\mathbf{M}_B \hat{\mathbf{a}}_{iso} &= \mathbf{M}_B(\mathbf{M}_B^T \mathbf{M}_B + \theta_M \mathbf{I}_b)^{-1} \mathbf{M}_B^T \mathbf{x} \\
&= \mathbf{U}\mathbf{D}_p(\mathbf{D}_p^2 + \theta_M \mathbf{I}_p)^{-1} \mathbf{D}_p \mathbf{U}^T \mathbf{x} \\
&= \mathbf{B}(N_b \mathbf{\Lambda_B})(N_b \mathbf{\Lambda_B} + \theta_M \mathbf{I}_p)^{-1} \mathbf{B}^T \mathbf{x} \\
&= \mathbf{B}(\mathbf{I}_p + \frac{\theta_M}{N_b} \mathbf{\Lambda_B}^{-1})^{-1} \mathbf{B}^T \mathbf{x} \\
&= \mathbf{B}(\mathbf{I}_p + \theta_0 \mathbf{\Lambda_B}^{-1})^{-1} \mathbf{B}^T \mathbf{x},
\end{aligned} \tag{57}$$

where $\mathbf{I}_p$ is a $p \times p$ identity matrix and $\theta_0 = \frac{\theta_M}{N_b}$. This is indeed the same as the solution of $\mathbf{B}\hat{\boldsymbol{\beta}}_{aniso}$, where $\hat{\boldsymbol{\beta}}_{aniso}$ is given by (28) in the MSSD-a method. Similar derivation can also be obtained for model H_1, which we omit here.

6. Experimental studies

In the experimental studies, we compare the performances of the MSSD-i, MSSD-a and MSD by applying them to a real HSI dataset called Hymap image. To measure the detection performances of the three methods, the receiver operating characteristic (ROC) curve is used, in which a good detection curve should lie near to the top left. In pair with ROC curve, we also employ the area under curve (AUC) statistics to measure the detection results quantitatively.

The Hymap image shown in Figure 1 was captured at the location of a small town of Cook City, USA. This image is published by Rochester Institute of

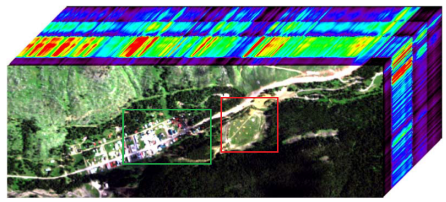
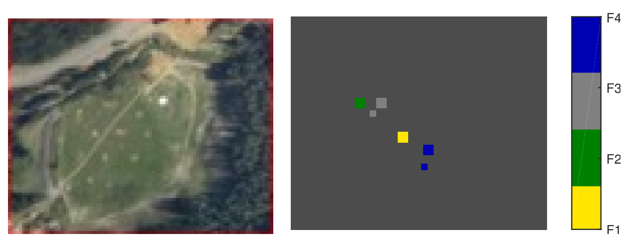


Figure 1: The Hymap scene. Two sub-images are cropped for evaluation.

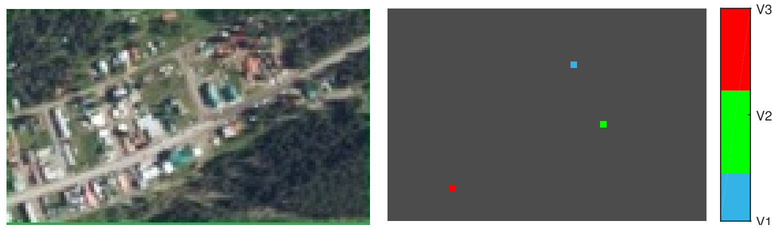
Technology (RIT) [22], which is widely used as a testbed for the algorithms of the HSI target detection. The hyperspectral image of Hymap has a total of 126 spectral bands with a pixel size of 280×800, covering the spectral range of 453nm-2496nm. Seven types of targets including four types of fabric panels (F1, F2, F3 and F4) and three vehicles (V1, V2 and V3) are deployed in the Hymap scene. When one type of target is to be detected, e.g. F1, the other targets, i.e. F2, F3, F4, V1, V2 and V3, are regarded as background pixels. We cropped two regions of interests (ROIs) into two separate HSI cubes, with the pixel size of 100×120 and 100×150, respectively. The ROIs of fabric panels (F1, F2, F3 and F4) and their corresponding target locations are shown in Figure 2, and the ROIs of three vehicles (V1, V2 and V3) and their corresponding target locations are shown in Figure 3.



(a) (b)

Figure 2: Target F1, F2, F3 and F4: (a) Hymap image scene of fabric panels; (b) locations of fabric panels. Pixels in different colours indicate different targets. The pixels sizes of ROIs of F1, F2, F3 and F4 are 25, 25, 34 and 34, respectively.

There are two widely accepted experiment settings regarding the target pix-



(a) (b)

Figure 3: Target V1, V2, V3: (a) Hymap image scene of vehicles; (b) locations if vehicles. Pixels in different colours indicate different targets. The pixels sizes of ROIs of V1, V2,and V3 are 9, 9 and 9, respectively.

els in the Hymap scene: 1) In [23, 24, 25, 26], only one target pixel of each desired target is assumed to be in the HSI; 2) whereas in [27], pixels within the ROIs of desired targets are all regarded as target pixels. In the setting 1), no target pixels are available for training. As a consequence, the parameters of the models have to be manually set. While in the setting 2), the target pixels can be randomly split into a training set and a test set and we can tune parameters for models. The setting 2) is believed to be a tougher condition for target detection than the setting 1). In this paper, we adopt the setting 2) in the evaluation of the compared methods for fair comparison.

We randomly choose 2-3 labelled target pixels for training and the rest target pixels for testing; and randomly choose around 10% background pixels for training and the rest background pixels for testing. Summaries of the numbers of training and test pixels of sub-images, which are used for detecting fabrics and vehicles, are given in Table 1 and Table 2, respectively.

6.1. Parameter settings

In real target detection problems, training examples of background pixels are not available. It is often assumed that the target presence in the scene is so sparse that if we extract neighbourhood pixels around a test pixel but not close to the test pixel, this neighbourhood can be seen as a replacement for

Table 1: Target fabrics: the number of target pixels for training and test in the sub-image shown in Figure 2.

	Target pixels			Background pixels		
Target	training	test	total	training	test	total
F1	2	23	25	1197	10778	11975
F2	2	23	25	1197	10778	11975
F3	3	31	34	1196	10770	11966
F4	3	31	34	1196	10770	11966

Table 2: Target vehicles: the number of target pixels for training and for test in the sub-image shown in Figure 3.

	Target pixels			Background pixels		
Target	training	test	total	training	test	total
V1	2	7	9	1499	13492	14991
V2	2	7	9	1499	13492	14991
V3	2	7	9	1499	13492	14991

background samples. Therefore as with [3, 4, 5, 11, 12, 28], we adopt the double concentric sliding window [11], a local and adaptive approach to extract the background pixels from the neighbourhood of each test pixel. Specifically, the concentric window separates the local area around each pixel into two regions, an inner window region (IWR) and an outer window region (OWR). The IWR is used to enclose the target of interest to be detected. The OWR is used to model the local backgrounds around the target region. An illustration of the double concentric window is shown in Figure 4. The determination of the window sizes is difficult. Since there are no labelled background samples in the Hymap dataset, we adopt the widely-used double concentric sliding window scheme to extract background samples and construct background subspace $\mathbf{B}$. For illustrative purposes and as with most of the state-of-the-art works [5, 11, 12, 28], the window sizes are set empirically in this paper. In our cases, the sizes of OWR and IWR are set as 17×17 and 7×7 for detecting fabrics panels, and 15×15

and 5×5 for detecting vehicles, respectively. Therefore, for each test pixel $\mathbf{x}$ in Figure 2, the number of training background pixels is $N_b = 240$; for each test pixel $\mathbf{x}$ in Figure 3, the number of training background pixels is $N_b = 200$, which are all greater than the dimension of the spectra ($p = 126$).

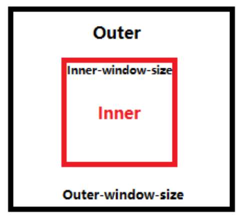


Figure 4: An illustration of the double concentric window.

For each target pixel $\mathbf{x}_i$ in an HSI, we use the mean-centred background samples extracted by double concentric window to compute the covariance matrix $\mathbf{C}_i$, where $i = 1, \ldots, N$ and N is the total number of test pixels in the HSI. Then the columns of the subspace $\mathbf{B}$ are created by the eigen-decomposition of $\mathbf{C}_i$. Since we only have one prior spectrum for each desired target, we subtract the background mean $\boldsymbol{\mu}_i$ of the local adaptive background samples around the test pixel $\mathbf{x}_i$ from the target spectrum $\mathbf{m}_t$, i.e. $\mathbf{m}_t - \boldsymbol{\mu}_i$, then normalise $\mathbf{m}_t - \boldsymbol{\mu}_i$ to have a unit l_2-norm as the target subspace $\mathbf{T}$. As a result, the columns in $\mathbf{B}$ and $\mathbf{T}$ all have unit l_2-norms and are independent of each other.

Regarding the variance σ_T of $\boldsymbol{\gamma}$ defined in MSSD-i (39) and the eigenvalue matrix $\boldsymbol{\Lambda}_{\mathbf{T}}$ of $\boldsymbol{\gamma}$ defined in MSSD-a (48), we set both σ_T and $\boldsymbol{\Lambda}_{\mathbf{T}}$ to be ∞, since we only have one target spectrum to construct $\mathbf{T}$ and there is no variance can be estimated in the target subspace. It means that in the real application of target detection where only one target spectrum is available, we actually do not shrink the size of abundance $\boldsymbol{\gamma}$ corresponding to the target basis vector in the H_1 model in both MSSD-i and MSSD-a, and let the projection of a test pixel onto the target basis vector be as much as possible.

In the conventional MSD to be evaluated on the Hymap image, there is only one unknown parameter to be tuned, which is the number of preserved leading eigenvectors $r_b (r_b \leqslant p)$ for the subspace $\mathbf{B}$, since for each desired target there is only one target spectrum, $N_t = r_t = 1$. In the proposed MSSD-i and MSSD-a, two unknown parameters in (18) and (27) need to be tuned: the shrinkage parameters θ_0 and θ_1. The optimal values of r_b of MSD, θ_{iso0} and θ_{iso1} of MSSD-i and θ_{aniso0} and θ_{aniso1} of MSSD-a tuned by the training data are listed in Table 3.

Table 3: Parameter settings of MSD, MSSD-i and MSSD-a.

	MSD	MSSD-i		MSSD-a	
	r_b	θ_{iso0}	θ_{iso1}	θ_{aniso0}	θ_{aniso1}
F1	2	1e-09	1e-07	1e-03	1e-03
F2	2	1e-09	3e-07	7e-07	1e-09
F3	14	1e-09	1e-08	1e-08	3
F4	2	1e-09	1e-08	3e-03	3e-03
V1	124	1e-09	1e-09	3e-07	1e-09
V2	6	1e-09	1e-07	1e-07	1e-06
V3	124	1e-09	1e-07	3	5e+1

6.2. Detection performance

Table 4: Detection performance of MSD, MSSD-i and MSSD-a measured with the AUC statistics. The best performance is indicated in boldface.

	MSD	MSSD-i	MSSD-a
F1	**0.974**	0.662	0.968
F2	0.706	0.713	**0.888**
F3	0.679	0.506	**0.801**
F4	0.711	0.656	**0.784**
V1	0.673	**0.845**	0.726
V2	0.647	0.752	**0.778**
V3	0.643	0.664	**0.676**

The detection performances of MSD, MSSD-i and MSSD-a are listed in Table 4 and shown in Figure 5 and Figure 6. Firstly, we can observe that both

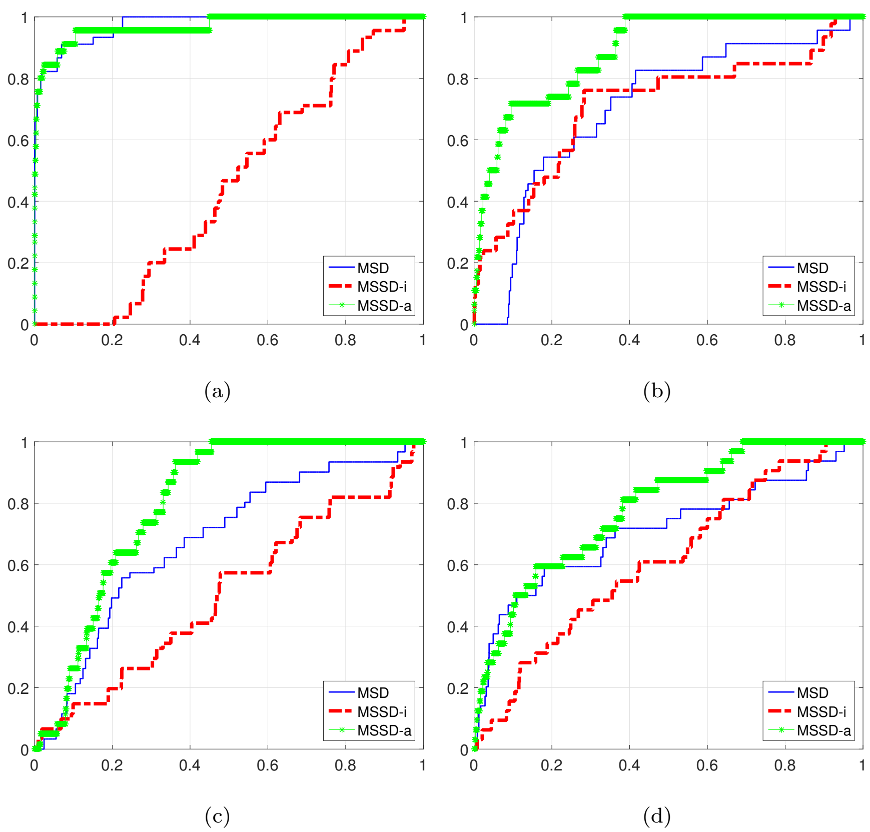
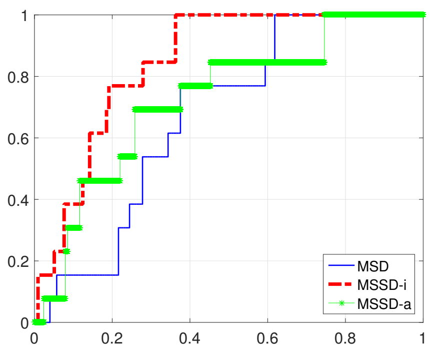
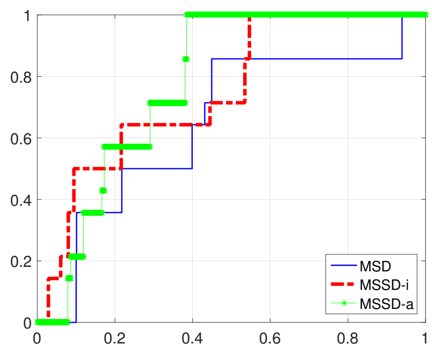


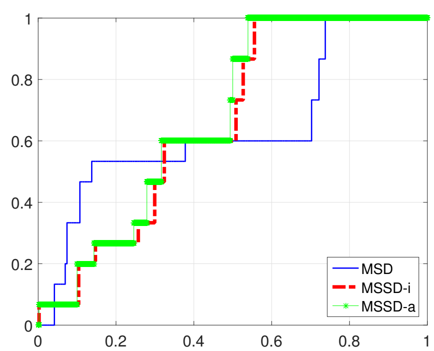
Figure 5: ROC curves of detecting fabric panels: (a) F1; (b) F2; (c) F3; (d) F4.



(a)



(b)



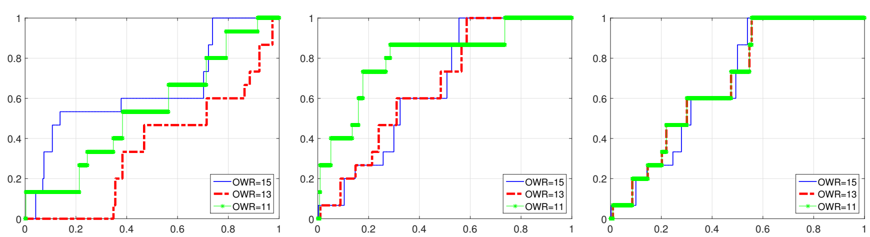
(c)

Figure 6: ROC curves of detecting vehicles: (a) V1; (b) V2; (c) V3.

MSSD-i and MSSD-a can outperform MSD in detecting F2, V1, V2 and V3. Specifically, MSSD-a can improve the detection performance significantly, compared with the conventional MSD method. Among the seven types of targets, MSSD-a improves six of them, F2, F3, F4, V1, V2 and V3, from MSD. Secondly, MSSD-i improves the performance on detecting F2, V1, V2 and V3, compared with MSD. These results suggest that introducing l_2-norm regularisation terms into MSD can improve the detection performance.

We shall note that MSD has better performance on detecting F1 than MSSD-i and MSSD-a. However, MSSD-a still has competitive performance as MSD on detecting F1 (0.9680 vs. 0.9742); it also illustrates that preserving the information from the eigenvalues in the prior distribution of abundance by MSSD-a can have a more stable detection performance than MSSD-i, which assumes an equal variance in the prior distribution.

6.3. Discussion on effects of parameters

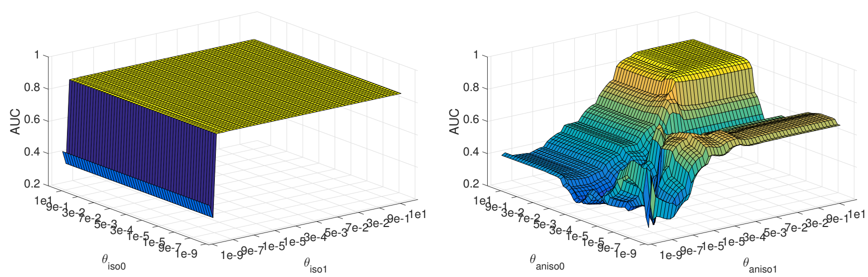


(a) (b) (c)

Figure 7: Effects of window sizes on detecting V3: (a) MSD; (b) MSSD-i; (c) MSSD-a. The IWR size is fixed to be 5×5, and the OWR size varies from 15×15, 13×13 to 11×11.

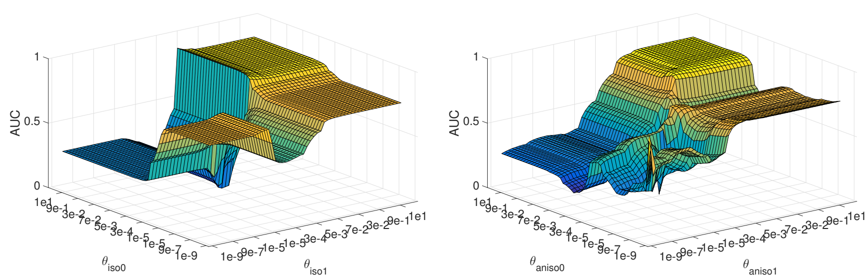
We further investigate the effects of parameters on the performances of detectors.

Firstly, the effects of window sizes on the performances of MSD, MSSD-i and MSSD-a for detecting target V3 are illustrated in Figure 7; the results for detecting other targets are of a similar pattern. It is true that all parameters, such as window sizes of OWR and IWR and shrinkage parameters θ_0 and θ_1, jointly affect the performances of detectors. Here for simplicity of exploring the effect of window sizes alone, we fix the values of other parameters (r_b, θ_0 and θ_1) of corresponding detectors as those in Table 3, and fix the size of IWR. The ROC curves of the detectors under three different sizes of OWR are plotted in Figure 7. We can observe that MSD and MSSD-i are sensitive to OWR, whilst MSSD-a is more stable. This indicates that MSSD-a is more robust to the variation of background samples, and preserving variances of the original data is beneficial in terms of the stability of detection performance.



(a) (b)

Figure 8: For OWR of size 15 × 15 and IWR of size 5 × 5. (a) MSSD-i: effects of θ_{iso0} and θ_{iso1} on detecting V3; (b) MSSD-a: effects of θ_{aniso0} and θ_{aniso1} on detecting V3.



(a) (b)

Figure 9: For OWR of size 11 × 11 and IWR of size 5 × 5. (a) MSSD-i: effects of θ_{iso0} and θ_{iso1} on detecting V3; (b) MSSD-a: effects of θ_{aniso0} and θ_{aniso1} on detecting V3.

Secondly, we investigate the effects of shrinkage parameters by sweeping the parameter spaces of θ_{iso0} and θ_{iso1} of MSSD-i and θ_{aniso0} and θ_{aniso1} of MSSD-

a. Here due to much higher computational complexity for the large number of test pixels, we show the results for the training pixels as illustration. We show the results of MSSD-i and MSSD-a for detecting V3 under two sets of window sizes in Figure 8 and Figure 9, respectively. Again, the results for detecting other targets are of a similar pattern.

We can observe that the AUC surface of MSSD-i is smoother than that of MSSD-a in both sets of window sizes. This pattern is particularly clear in the setting that OWR is of size 15×15 and IWR is of size 5×5, where MSSD-i is not sensitive to θ_{iso0}, as shown in Figure 8(a). Technically, the reason for this 'extreme' pattern is because the number of training background pixels $N_b = 200$ is greater than the pixel dimension $p = 126$, which leads to the result that $r_b = p$ and the $p \times p$ matrix $\mathbf{B}$ represents a full space. Therefore for each pixel $\mathbf{x}_j$, the RSS $e_0^{iso}(\mathbf{x}_j)$ in (21) can be simplified to

$$\begin{aligned} e_0^{iso}(\mathbf{x}_j) &= \mathbf{x}_j^T \left(\mathbf{I} - \mathbf{B}\left((1+\theta_{iso0})\mathbf{I}_0\right)^{-1}\mathbf{B}^T\right)\mathbf{x}_j \\ &= \mathbf{x}_j^T\mathbf{x}_j - \frac{1}{1+\theta_{iso0}}\mathbf{x}_j^T\mathbf{B}\mathbf{B}^T\mathbf{x}_j \\ &= \mathbf{x}_j^T\mathbf{x}_j - \frac{1}{1+\theta_{iso0}}\mathbf{x}_j^T\mathbf{x}_j \\ &= \frac{\theta_{iso0}}{1+\theta_{iso0}}\mathbf{x}_j^T\mathbf{x}_j\ . \end{aligned} \tag{58}$$

In (58), $e_0^{iso}(\mathbf{x}_j)$ is equivalent to scaling the l_2-norm of every pixel $\mathbf{x}_j$ with a scaler $\frac{\theta_{iso0}}{1+\theta_{iso0}}$. The detection ratio (23) is then scaled by $\frac{\theta_{iso0}}{1+\theta_{iso0}}$ as well when θ_{iso1} is fixed. As a result, the AUC of MSSD-i does not depend on θ_{iso0}, as shown in Figure 8(a). However, in Figure 9(a) when the OWR size reduces to 11×11, the number of background samples N_b becomes 96 and thus $N_b < p$, and the AUC becomes dependent on θ_{iso0}, because now $e_0^{iso}(\mathbf{x}_j)$ cannot be simplified to (58) and θ_{iso0} affects the AUC.

As a by-product, the above analysis suggests a guideline on the use of MSSD-i: when $N_b < p$, both shrinkage parameters θ_{iso0} and θ_{iso1} should be tuned during the training phase; when $N_b > p$, only θ_{iso1} needs to be tuned and θ_{iso0} can be arbitrary. For example, the values of θ_{iso0} in Table 3 are in the case of $N_b > p$ and are not necessary to be 1e-09; instead, they can be any values.

For MSSD-a, the detection performance varies with both θ_{aniso0} and θ_{aniso1}, as shown in Figure 8(b) and Figure 9(b).

Finally, it is worth discussing why MSSD-a is more favourable than MSSD-i, as indicated by the test results listed in Table 4. We believe a big reason for this is that MSSD-a considers both eigenvectors and eigenvalues to preserve the information of the data for the shrinkage, while MSSD-i considers only eigenvectors. MSSD-i essentially assumes an equal variance in the prior distribution of each coefficient in the eigenspace, while MSSD-a assumes different variances for different coefficients based on eigenvalues. Hence the latter preserves the variances of the original data and can adapt to the shrinkage in different directions in the eigenspace better than the former.

7. Conclusion

We have proposed a new approached to hyperspectral target detection, called the matched shrunken subspace detector (MSSD), and its two implementations, MSSD-i with isotropic shrinkage and MSSD-a with anisotropic shrinkage. The MSSD introduces the l_2-norm regularisation into the popular matched subspace detector (MSD), seeking more reliable projection for the hypothesis models H_0 and H_1. From the Bayesian perspective, the added regularisation terms preserve non-uniform prior distributions of the coefficient vectors in the models. Both MSSD-i and MSSD-a can reduce the variances of the coefficients and result in more stable estimators. The links among MSD, MSSD-i and MSSD-a have also been discussed in detail, and the two proposed methods have shown superior detection performance compared with the conventional MSD on the real dataset of Hymap.

Acknowledgment

The authors thank the associate editor and the anonymous reviewers for their constructive comments. This work was partially supported by University

College London's Security Science Doctoral Training Centre under Engineering and Physical Sciences Research Council (EPSRC) grant EP/G037264/1.

References

[1] N. M. Nasrabadi, Hyperspectral target detection: An overview of current and future challenges, Signal Processing Magazine, IEEE 31 (1) (2014) 34–44.

[2] W. Li, Q. Du, A survey on representation-based classification and detection in hyperspectral remote sensing imagery, Pattern Recognition Letters 83 (2016) 115–123.

[3] S. Matteoli, M. Diani, G. Corsini, A tutorial overview of anomaly detection in hyperspectral images, Aerospace and Electronic Systems Magazine, IEEE 25 (7) (2010) 5–28.

[4] W. Li, Q. Du, Collaborative representation for hyperspectral anomaly detection, Geoscience and Remote Sensing, IEEE Transactions on 53 (3) (2015) 1463–1474.

[5] Y. Yuan, Q. Wang, G. Zhu, Fast hyperspectral anomaly detection via high-order 2-D crossing filter, IEEE Transactions on Geoscience and Remote Sensing 53 (2) (2015) 620–630.

[6] Y. Yuan, D. Ma, Q. Wang, Hyperspectral anomaly detection by graph pixel selection, IEEE Transactions on Cybernetics 46 (12) (2016) 3123–3134.

[7] L. L. Scharf, B. Friedlander, Matched subspace detectors, Signal Processing, IEEE Transactions on 42 (8) (1994) 2146–2157.

[8] Z. Wang, J.-H. Xue, The matched subspace detector with interaction effects, Pattern Recognition 68 (2017) 24–37.

[9] D. Manolakis, C. Siracusa, G. Shaw, Hyperspectral subpixel target detection using the linear mixing model, Geoscience and Remote Sensing, IEEE Transactions on 39 (7) (2001) 1392–1409.

[10] H. Kwon, N. M. Nasrabadi, A comparative analysis of kernel subspace target detectors for hyperspectral imagery, EURASIP Journal on Advances in Signal Processing 2007 (1) (2006) 1–13.

[11] Y. Chen, N. M. Nasrabadi, T. D. Tran, Sparse representation for target detection in hyperspectral imagery, Selected Topics in Signal Processing, IEEE Journal of 5 (3) (2011) 629–640.

[12] Y. Zhang, B. Du, L. Zhang, A sparse representation-based binary hypothesis model for target detection in hyperspectral images, Geoscience and Remote Sensing, IEEE Transactions on 53 (3) (2015) 1346–1354.

[13] Y. Gu, Y. Wang, H. Zheng, Y. Hu, Hyperspectral target detection via exploiting spatial-spectral joint sparsity, Neurocomputing 169 (2015) 5–12.

[14] J. Friedman, T. Hastie, R. Tibshirani, The Elements of Statistical Learning, Vol. 1, Springer, Berlin, 2001.

[15] A. N. Tikhonov, A. Goncharsky, V. Stepanov, A. G. Yagola, Numerical Methods for the Solution of Ill-posed Problems, Vol. 328, Springer Science & Business Media, 2013.

[16] J. Li, H. Zhang, Y. Huang, L. Zhang, Hyperspectral image classification by nonlocal joint collaborative representation with a locally adaptive dictionary, Geoscience and Remote Sensing, IEEE Transactions on 52 (6) (2014) 3707–3719.

[17] W. Li, Q. Du, M. Xiong, Kernel collaborative representation with Tikhonov regularization for hyperspectral image classification, Geoscience and Remote Sensing Letters, IEEE 12 (1) (2015) 48–52.

[18] W. Li, Q. Du, B. Zhang, Combined sparse and collaborative representation for hyperspectral target detection, Pattern Recognition 48 (12) (2015) 3904–3916.

[19] Y. Zhang, B. Du, L. Zhang, Regularization framework for target detection in hyperspectral imagery, Geoscience and Remote Sensing Letters, IEEE 11 (1) (2014) 313–317.

[20] N. M. Nasrabadi, Regularized spectral matched filter for target recognition in hyperspectral imagery, Signal Processing Letters, IEEE 15 (2008) 317–320.

[21] S. M. Kay, Fundamentals of Statistical Signal Processing: Detection Theory, Vol. 2, Prentice Hall, NJ, USA, 1998.

[22] D. Snyder, J. Kerekes, I. Fairweather, R. Crabtree, J. Shive, S. Hager, Development of a web-based application to evaluate target finding algorithms, in: Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International, Vol. 2, IEEE, 2008, pp. II–915.

[23] L. Zhang, L. Zhang, D. Tao, X. Huang, Sparse transfer manifold embedding for hyperspectral target detection, IEEE Transactions on Geoscience and Remote Sensing 52 (2) (2014) 1030–1043.

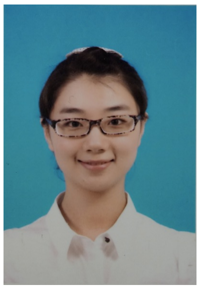
[24] L. Zhang, L. Zhang, D. Tao, X. Huang, B. Du, Hyperspectral remote sensing image subpixel target detection based on supervised metric learning, IEEE Transactions on Geoscience and Remote Sensing 52 (8) (2014) 4955–4965.

[25] L. Gao, B. Yang, Q. Du, B. Zhang, Adjusted spectral matched filter for target detection in hyperspectral imagery, Remote Sensing 7 (6) (2015) 6611–6634.

[26] B. Du, L. Zhang, D. Tao, D. Zhang, Unsupervised transfer learning for target detection from hyperspectral images, Neurocomputing 120 (2013) 72–82.

[27] B. Du, L. Zhang, Target detection based on a dynamic subspace, Pattern Recognition 47 (1) (2014) 344–358.

[28] T. Wang, B. Du, L. Zhang, A background self-learning framework for unstructured target detectors, IEEE Geoscience and Remote Sensing Letters 10 (6) (2013) 1577–1581.



Ziyu Wang received the M.Sc. degree in statistics in 2012 and the M.Res. degree in security science in 2013, both from University College London. She is currently pursuing the Ph.D. degree in the Security Science Doctoral Research Training Centre and the Department of Statistical Science, University College London. Her current research interests include hyperspectral image analysis, sparse representation and statistical classification.

Jing-Hao Xue received the Dr.Eng. degree in signal and information processing from Tsinghua University in 1998 and the Ph.D. degree in statistics from the University of Glasgow in 2008. Since 2008, he has worked in the Department of Statistical Science at University College London as a Lecturer and Senior Lecturer. His current research interests include statistical classification, high-dimensional data analysis, pattern recognition and image analysis.