City Research Online

City St George's, University of London

Citation: Wang, Z., Zhu, R., Fukui, K. & Xue, J-H. (2017). Matched Shrunken Cone Detector (MSCD): Bayesian Derivations and Case Studies for Hyperspectral Target Detection. IEEE Transactions on Image Processing, 26(11), pp. 5447-5461. doi: 10.1109/tip.2017.2740621

This is the published version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/20735/

Link to published version: https://doi.org/10.1109/tip.2017.2740621

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

Matched Shrunken Cone Detector (MSCD): Bayesian Derivations and Case Studies for Hyperspectral Target Detection

Ziyu Wang, Rui Zhu, Kazuhiro Fukui, *Member, IEEE*, and Jing-Hao Xue

***Abstract*— Hyperspectral images (HSIs) possess *non-negative* properties for both hyperspectral signatures and abundance coefficients, which can be naturally modeled using cone-based representation. However, in hyperspectral target detection, cone-based methods are barely studied. In this paper, we propose a new regularized cone-based representation approach to hyperspectral target detection, as well as its two working models by incorporating into the cone representation l_2-norm and l_1-norm regularizations, respectively. We call the new approach the matched shrunken cone detector (MSCD). Also important, we provide principled derivations of the proposed MSCD from the Bayesian perspective: we show that MSCD can be derived by assuming a multivariate half-Gaussian distribution or a multivariate half-Laplace distribution as the prior distribution of the coefficients of the models. In the experimental studies, we compare the proposed MSCD with the subspace methods and the sparse representation-based methods for HSI target detection. Two real hyperspectral data sets are used for evaluating the detection performances on sub-pixel targets and full-pixel targets, respectively. Results show that the proposed MSCD can outperform other methods in both cases, demonstrating the competitiveness of the regularized cone-based representation.**

***Index Terms*— Target detection, hyperspectral image (HSI), cone representation, non-negativity, half-Gaussian distribution, half-Laplace distribution, shrunken estimation.**

I. Introduction

With the help of remote sensors, hyperspectral imaging has become an important scientific tool for various fields of real-world applications. In the analysis of hyperspectral images (HSIs), target detection is a major task, which aims to detect small objects or anomalies in an hyperspectral image. Typical target detection applications include military defence, agricultural management and mineral detection.

Manuscript received January 1, 2017; revised June 6, 2017 and July 18, 2017; accepted August 10, 2017. Date of publication August 16, 2017; date of current version September 1, 2017. This work was supported in part by the University College London's Security Science Doctoral Training Centre under Engineering and Physical Sciences Research Council under Grant EP/G037264/1. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Guoliang Fan. *(Corresponding author: Jing-Hao Xue.)*

Z. Wang is with the Department of Security and Crime Science, University College London, London WC1E 6BT, U.K., and also with the Department of Statistical Science, University College London, London WC1E 6BT, U.K. (e-mail: ziyu.wang.12@ucl.ac.uk).

R. Zhu and J.-H. Xue are with the Department of Statistical Science, University College London, London WC1E 6BT, U.K. (e-mail: r.zhu.12@alumni.ucl.ac.uk; jinghao.xue@ucl.ac.uk).

K. Fukui is with the Department of Computer Science, University of Tsukuba, Tsukuba 305-8577, Japan (e-mail: kfukui@cs.tsukuba.ac.jp).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2017.2740621

Target detection is essentially a binary classification problem, of which the task is to determine if an HSI pixel is a target spectrum or a background spectrum. Hence, target detection can be regarded as a binary hypothesis model with two competing hypotheses: the null hypothesis H_0 for the absence of the target; and the alternative hypothesis H_1 for the presence of the target. Binary hypothesis models for target detection have been nicely reviewed in [1]–[4].

Target objects often appear as sub-pixels in an HSI. That is, the spectrum of an HSI pixel can be a mixture of different component spectra of materials. These component spectra are usually termed endmembers. To model the mixture of an HSI pixel, the linear mixing model (LMM) [5] has been widely adopted. The underlying assumption of LMM is that an HSI pixel can be approximated by a linear combination of endmembers with different fractions. When a target pixel presents, its spectrum is decomposed as a linear combination of background endmembers and target endmembers; in contrast, when a background pixel presents, its spectrum is fully represented by background endmembers.

Within the framework of binary hypothesis modelling, researches have explored a variety of techniques and extensions on the basis of LMM. Since it is difficult to obtain comprehensive spectral libraries to serve as the endmembers for all desired targets, many methods focus on extract endmembers directly from HSIs. On the one hand, provided with a large number of background samples, subspace methods have been widely developed for target detection. Typical methods, such as the orthogonal subspace projection detector (OSP) [6] and matched subspace detector (MSD) [7], adopt the leading eigenvectors (with dominant eigenvalues) as the subspace bases and implicitly the endmembers. On the other hand, sparse representation (SR) techniques [8] originating from compressed sensing have been recently studied in the HSI analysis [9]. For HSI target detection, SR-based methods, such as sparse target detection (STD) [10], sparse representation-based binary hypothesis model (SRBBH) [11] and hybrid sparsity and statistics detector [12], model a test HSI pixel as a linear combination of only a few training samples (aka atoms of an over-complete dictionary). It implicitly regards the atoms as endmembers, hence the SR-based methods can be viewed as being developed in the original sample space.

These methods can be further extended to nonlinear mixing models. The kernel methods, which aim to define a model in a high-dimensional feature space associated with a nonlinear mapping of input data, have also been studied for HSI target

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/

detection [13]–[15]. In [13], subspace methods such as MSD, OSP have been extended to their kernel versions. Kernelisation of the SR-based methods has been also developed, such as kernel-based STD [14] and kernel-based SRBBH [15].

For the sake of physical interpretations, HSIs as instances of natural signals possess *non-negative* properties for both hyperspectral signatures and the abundance coefficients. A number of investigations have focused on the non-negative matrix factorisation (NMF) [16], [17] for HSI unmixing problems. NMF factorises a sample data matrix into two low-dimensional matrices in terms of bases and corresponding coefficients, and explicitly enforces the non-negative constraints on both of them. However, in the past researches of HSI target detection [6], [7], [9]–[15], the non-negativity properties have not been considered yet, particularly for the abundance coefficients. If we use the samples directly from HSIs as endmembers, it is desirable to impose the non-negative constraints on the coefficients. In this way, both endmembers and coefficients are non-negative, such that this physical characteristic of hyperspectral signatures are modelled.

Statistically, the estimation of non-negatively-constrained coefficients in the LMM is often termed non-negative least squares (NNLS) [18]. Geometrically, the NNLS estimation induces a cone-shape representation [19]. Suppose that a hyperspectral spectrum $\mathbf{x}$ is a p-dimensional vector, and that there are K types of materials, i.e. K endmembers potentially constituting an HSI pixel, which are represented by $\mathbf{m}_1, \ldots, \mathbf{m}_K$ with each $\mathbf{m}_k$ also a p-dimensional vector. Then the cone-based representation of pixel $\mathbf{x}$ expresses the spectral signature of $\mathbf{x}$ as a *non-negative* linear combination of endmembers $\mathbf{m}_1, \ldots, \mathbf{m}_K$ with corresponding *non-negative* abundance fractions $a_1, \ldots, a_K$, such that $a_k \geq 0$ for $k = 1, \ldots, K$. More specifically, a *convex cone* $\mathbb{C}$ is defined as

$$\mathbb{C} : \left\{ \mathbf{x} | \mathbf{x} = \sum_{k=1}^{K} a_k \mathbf{m}_k = \mathbf{M}\mathbf{a}, a_k \geq 0 \right\}, \tag{1}$$

where $\mathbf{M}$ is a $p \times K$ matrix whose columns are the K endmembers spectra $\mathbf{m}_k = [m_{k,1}, \ldots, m_{k,p}]^T$; and $\mathbf{a} = [a_1, \ldots, a_K]^T$ denotes the abundance vector. For the non-negative LMM, an additional noise term is also considered:

$$\mathbf{x} = \mathbf{M}\mathbf{a} + \mathbf{n}, a_k \geq 0, \tag{2}$$

where the vector $\mathbf{n}$ is assumed to be the Gaussian white noise, i.e. $\mathbf{n} \sim (\mathbf{0}, \sigma^2 \mathbf{I}_p)$, where $\mathbf{I}_p$ is the $p \times p$ identity matrix.

It is worth noting that, LMM-based methods may suffer from the problem of high variance of coefficients estimations. To this end, shrinkage methods [20] have been developed in statistical learning. Typical shrinkage methods include l_2-norm regularisation, also known as *ridge regression* or Tikhonov regularisation, and l_1-norm regularisation, also known as *lasso*. For the convex cone analysis, these regularisations have also been studied, mainly on the computational efficiency of the algorithms developed based on the NNLS [21]–[24].

In this paper, to account for the non-negativity as well as the shrinkage of the coefficients of the convex cone model (2) for HSI target detection, we propose a new approach called the matched shrunken cone detector (MSCD). Specifically, on the cone representations we propose to shrink the abundance coefficients of target endmembers and background endmembers by imposing constraints; we propose two working models with the l_2-norm and l_1-norm regularisations, respectively. We call these two methods MSCD-l_2 and MSCD-l_1. Equally important, we derive the proposed MSCD from the Bayesian perspective, showing that MSCD-l_2 and MSCD-l_1 can be derived if a *multivariate half-Gaussian distribution* [25] and a *multivariate half-Laplace distribution* [26] are assumed as the prior distributions of the coefficient vectors. To our knowledge, it is the first time that the cone representations with the l_2-norm and l_1-norm regularisations are derived from the Bayesian perspective, as well as the prior distributions identified.

The main novelties and contributions of this paper are summarised as follows.

1) We propose a regularised cone-based representation approach called MSCD for HSI target detection. This is the first time that the cone-based representation and its regularised versions are brought to HSI target detection.

2) We introduce two independent working models of MSCD, namely MSCD-l_2 and MSCD-l_1, by incorporating l_2-norm and l_1-norm regularisations, respectively, into the cone-based representation (2).

3) More importantly, we derive the proposed MSCD-l_2 and MSCD-l_1 from the Bayesian perspective, showing that they imply a multivariate half-Gaussian distribution and a multivariate half-Laplace distribution as the prior distributions for the coefficients. As far as we are concerned, this is the first time that the l_2-norm and l_1-norm regularised cone representations are derived from the Bayesian perspective with their corresponding prior distributions identified.

4) Through illustrating the Bayesian derivations of the proposed MSCD-l_2 and MSCD-l_1, our principled work opens a door to different new regularised models to accommodate various prior knowledge of the practitioners, which provides a valuable direction to further and enrich the research of HSI target detection.

5) Last but not least, we illustrate the competitive detection performance of the proposed models, compared with some classical and state-of-the-art HSI target-detection methods, on two real hyperspectral datasets for sub-pixel and full-pixel target detections, respectively.

It is worth noting that our proposed models are in nature different from the widely-used sparse-representation-based detectors [10], [11], the collaborative-representation-based detector [27] and their hybrids [28], although our two working models also apply the l_2-norm and l_1-norm regularisations. From the modelling perspective, motivated by NMF and physical interpretations, our MSCD introduces the non-negativity constraints into the estimation of the model coefficients. From the geometrical perspective, these non-negativity constraints induce a cone-shaped representation. Furthermore, we provide comprehensive statistical derivations of our proposed models from both frequentist and Bayesian perspectives. In fact, none of the non-negativity, the cone-based representation, or the Bayesian derivation were presented in [10], [11], [27], and [28].

In the rest of the paper, Section II reviews the binary hypothesis model in terms of the likelihood ratio test. Section III introduces the propose MSCD. Section IV shows the derivations of the proposed MSCD-l_2 and MSCD-l_1 from the Bayesian perspective with the prior distributions of the coefficients identified. Section V illustrates the superior performance of MSCD to other subspace and SR-based methods; and section VI gives the conclusion of this work.

II. Binary Hypothesis Testing Model for HSI Target Detection

HSI target detection methods are typically derived from a binary hypothesis testing model [3]. We suppose that an HSI pixel $\mathbf{x}$ is a continuous random vector. A likelihood ratio of the conditional probability density functions (pdfs) on two competing hypotheses is constructed as follows:

$$\begin{aligned} &H_0 : \mathbf{x} \text{ is a background pixel,} \\ &H_1 : \mathbf{x} \text{ is a target pixel.} \\ &\Rightarrow D(\mathbf{x}) = \frac{f_{\mathbf{x}|H_1}(\mathbf{x})}{f_{\mathbf{x}|H_0}(\mathbf{x})} \underset{H_0}{\overset{H_1}{\gtrless}} \nu. \end{aligned} \tag{3}$$

In (3), $D(\mathbf{x})$ is an output detector, which is the ratio of two conditional pdfs of $\mathbf{x}$ under the null hypothesis H_0 and the alternative hypothesis H_1, i.e. $f_{\mathbf{x}|H_0}(\mathbf{x})$ and $f_{\mathbf{x}|H_1}(\mathbf{x})$; ν is a predefined detection threshold, such that when $D(\mathbf{x}) > \nu$ the test HSI pixel $\mathbf{x}$ is identified as a target. The pdfs are usually unknown and estimated parametrically. Specifically, the likelihood ratio is replaced by the generalised likelihood ratio (GLR), using their maximum likelihood estimates (MLEs):

$$\begin{aligned} D_{GLR}(\mathbf{x}) &= \frac{f_{\mathbf{x}|H_1}(\mathbf{x};\hat{\omega}_1)}{f_{\mathbf{x}|H_0}(\mathbf{x};\hat{\omega}_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \nu \\ &= \frac{\max_{\omega_1}\{f_{\mathbf{x}|H_1}(\mathbf{x};\omega_1)\}}{\max_{\omega_0}\{f_{\mathbf{x}|H_0}(\mathbf{x};\omega_0)\}} \underset{H_0}{\overset{H_1}{\gtrless}} \nu. \end{aligned} \tag{4}$$

In (4), we use $\hat{\omega}_0$ and $\hat{\omega}_1$ to denote the MLEs of ω_0 and ω_1, where ω_0 and ω_1 are unknown parameters of conditional pdfs $f_{\mathbf{x}|H_0}(\mathbf{x};\omega_0)$ and $f_{\mathbf{x}|H_1}(\mathbf{x};\omega_1)$, respectively.

A. Formulation of LMM-Based Binary Hypothesis Models

In the framework of LMM [5], a test pixel $\mathbf{x}$ is modelled by a linear combination of target endmembers and background endmembers. Specifically, the LMM for HSI target detection is constructed as follows:

$$\begin{aligned} &H_0 : \mathbf{x} = \mathbf{M}_B\boldsymbol{\beta} + \mathbf{n}_0, \ \mathbf{x} \text{ is a background pixel,} \\ &H_1 : \mathbf{x} = \mathbf{M}_T\boldsymbol{\gamma} + \mathbf{M}_B\boldsymbol{\beta} + \mathbf{n}_1, \ \mathbf{x} \text{ is a target pixel,} \end{aligned} \tag{5}$$

where $\mathbf{M}_T = [\mathbf{t}_1, \ldots, \mathbf{t}_{N_t}]$ is a $p \times N_t$ matrix whose columns $\mathbf{t}_1, \ldots, \mathbf{t}_{N_t}$ are N_t target spectra; $\mathbf{M}_B = [\mathbf{b}_1, \ldots, \mathbf{b}_{N_b}]$ is a $p \times N_b$ matrix whose columns are N_b background spectra; $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are the abundance vectors of $\mathbf{M}_T$ and $\mathbf{M}_B$, respectively; and $\mathbf{n}_0$ and $\mathbf{n}_1$ are assumed to be p-dimensional vectors of Gaussian white noise: $\mathbf{n}_0 \sim \mathcal{N}(\mathbf{0}, \sigma_{H_0}^2\mathbf{I}_p)$ and $\mathbf{n}_1 \sim \mathcal{N}(\mathbf{0}, \sigma_{H_1}^2\mathbf{I}_p)$, where $\mathbf{I}_p$ is the $p \times p$ identity matrix.

For a more convenient representation, we let $\mathbf{M}$ be the concatenated matrix of $\mathbf{M}_T$ and $\mathbf{M}_B$: $\mathbf{M} = [\mathbf{M}_T, \mathbf{M}_B] = [\mathbf{t}_1, \ldots, \mathbf{t}_{N_t}, \mathbf{b}_1, \ldots, \mathbf{b}_{N_b}] \in \mathbb{R}^{p\times(N_t+N_b)}$. Accordingly, we concatenate the abundance vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ of model H_1 into one vector $\boldsymbol{\alpha}$: $\boldsymbol{\alpha} = \begin{bmatrix}\boldsymbol{\gamma}\\ \boldsymbol{\beta}\end{bmatrix} = [\gamma_1, \ldots, \gamma_{N_t}, \beta_1, \ldots, \beta_{N_b}]^T \in \mathbb{R}^{(N_t+N_b)}$. Then model H_1 can be rewritten as

$$\begin{aligned} H_1 : \mathbf{x} &= \mathbf{M}_T\boldsymbol{\gamma} + \mathbf{M}_B\boldsymbol{\beta} + \mathbf{n}_1 \\ &= \begin{bmatrix}\mathbf{M}_T & \mathbf{M}_B\end{bmatrix}\begin{bmatrix}\boldsymbol{\gamma}\\ \boldsymbol{\beta}\end{bmatrix} + \mathbf{n}_1 \\ &= \mathbf{M}\boldsymbol{\alpha} + \mathbf{n}_1, \end{aligned} \tag{6}$$

and the LMM-based binary hypothesis model becomes

$$\begin{aligned} &H_0 : \mathbf{x} = \mathbf{M}_B\boldsymbol{\beta} + \mathbf{n}_0, \ \mathbf{x} \text{ is a background pixel,} \\ &H_1 : \mathbf{x} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{n}_1, \ \mathbf{x} \text{ is a target pixel,} \end{aligned} \tag{7}$$

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

B. Derivations of LMM-Based GLR

The generalised likelihood ratio (GLR) of LMM for target detection is formulated as

$$\begin{aligned} \hat{l}(\mathbf{x}) &= \frac{l(\hat{\boldsymbol{\alpha}}, \hat{\sigma}_{H_1}^2; \mathbf{x})}{l(\hat{\boldsymbol{\beta}}, \hat{\sigma}_{H_0}^2; \mathbf{x})} \\ &= \left(\frac{\hat{\sigma}_{H_1}^2}{\hat{\sigma}_{H_0}^2}\right)^{-p/2} \exp\left\{-\frac{1}{2\hat{\sigma}_{H_1}^2}\left\|\hat{\mathbf{n}}_1\right\|_2^2 + \frac{1}{2\hat{\sigma}_{H_0}^2}\left\|\hat{\mathbf{n}}_0\right\|_2^2\right\}. \end{aligned} \tag{8}$$

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 (8), we have

$$\begin{aligned} L_{LMM}(\mathbf{x}) &= (\hat{l}(\mathbf{x}))^{2/p} \\ &= \left(\frac{\hat{\sigma}_{H_1}^2}{\hat{\sigma}_{H_0}^2}\right)^{-1} = \frac{\hat{\sigma}_{H_0}^2}{\hat{\sigma}_{H_1}^2} \\ &= \frac{\left\|\hat{\mathbf{n}}_0\right\|_2^2}{\left\|\hat{\mathbf{n}}_1\right\|_2^2} = \frac{\left\|\mathbf{x} - \mathbf{M}_B\hat{\boldsymbol{\beta}}\right\|_2^2}{\left\|\mathbf{x} - \mathbf{M}\hat{\boldsymbol{\alpha}}\right\|_2^2}. \end{aligned} \tag{9}$$

The MLEs of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ in (9) are given by

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \underset{\boldsymbol{\beta}}{\text{argmax}}\left\{f_{\mathbf{x}|H_0}(\mathbf{x};\boldsymbol{\beta},\sigma_0^2)\right\} \\ &= \underset{\boldsymbol{\beta}}{\text{argmin}}\left\{\frac{1}{2\sigma_{H_0}^2}\left\|\mathbf{x} - \mathbf{M}_B\boldsymbol{\beta}\right\|_2^2\right\} \end{aligned} \tag{10}$$

and

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= \underset{\boldsymbol{\alpha}}{\text{argmax}}\left\{f_{\mathbf{x}|H_1}(\mathbf{x};\boldsymbol{\alpha},\sigma_{H_1}^2)\right\} \\ &= \underset{\boldsymbol{\alpha}}{\text{argmin}}\left\{\frac{1}{2\sigma_{H_1}^2}\left\|\mathbf{x} - \mathbf{M}\boldsymbol{\alpha}\right\|_2^2\right\}, \end{aligned} \tag{11}$$

and thus

$$\hat{\boldsymbol{\beta}} = (\mathbf{M}_B^T\mathbf{M}_B)^{-1}\mathbf{M}_B^T\mathbf{x} \text{ and} \tag{12}$$

$$\hat{\boldsymbol{\alpha}} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\mathbf{x}, \tag{13}$$

by least square estimates. Based on solutions (12) and (13), the residual sums of squares (RSS) e_0 and e_1 for models H_0 and H_1 are computed as

$$\begin{aligned} H_0 : e_0 &= \left\|\hat{\mathbf{n}}_0\right\|_2^2 = \left\|\mathbf{x} - \mathbf{M}_B\hat{\boldsymbol{\beta}}\right\|_2^2 \\ &= \mathbf{x}^T(\mathbf{I}_p - \mathbf{M}_B(\mathbf{M}_B^T\mathbf{M}_B)^{-1}\mathbf{M}_B^T)\mathbf{x} \end{aligned} \tag{14}$$

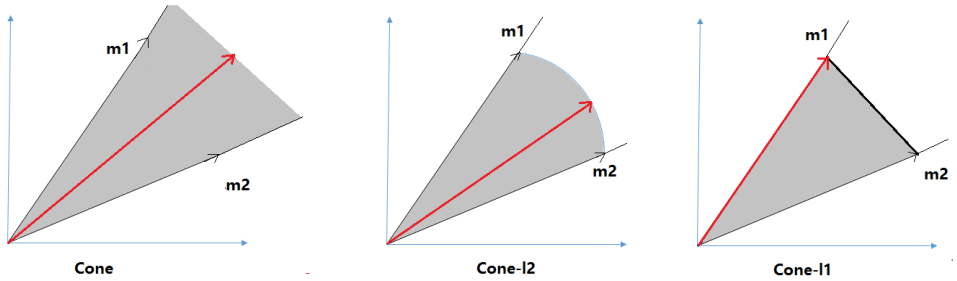


Fig. 1. Illustration of cone-representation methods in a 2D case with different constraints on coefficient vector **a**: (a) cone (17); (b) cone representation with l_2-norm regularisation (18); and (c) cone representation with l_1-norm regularisation (19).

and

$$H_1 : e_1 = \left\|\hat{\mathbf{n}}_1\right\|_2^2 = \left\|\mathbf{x} - \mathbf{M}\hat{\boldsymbol{\alpha}}\right\|_2^2 = \mathbf{x}^T(\mathbf{I}_p - \mathbf{M}(\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T)\mathbf{x}, \tag{15}$$

respectively. The final GLR detector of LMM is then

$$D_{LMM}(\mathbf{x}) = \frac{e_0}{e_1} = \frac{\mathbf{x}^T(\mathbf{I}_p - \mathbf{M}_B(\mathbf{M}_B^T\mathbf{M}_B)^{-1}\mathbf{M}_B^T)\mathbf{x}}{\mathbf{x}^T(\mathbf{I}_p - \mathbf{M}(\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T)\mathbf{x}} \underset{H_0}{\overset{H_1}{\gtrless}} \nu. \tag{16}$$

The value of $D_{LMM}(\mathbf{x})$ is compared to a threshold ν to make the final decision of which hypothesis should be rejected for the test pixel **x**. It is worth noting that the over-fitting problem may happened in (16), and to this end the matched subspace detector (MSD) [7] can be used instead. In MSD, the endmembers of background spectra and target spectra, $\mathbf{M}_B$ and $\mathbf{M}_T$, are represented by the leading eigenvectors of the background and target subspaces, respectively.

III. Matched Shrunken Cone Detector (MSCD)

Rather than using an unconstrained LMM, it is desirable to adopt the non-negative linear model for modelling a mixed HSI pixel, so as for a reasonable physical interpretation. On top of that, we also introduce the regularisation to the non-negative representation to control the variance of estimates, and derive the whole new model from the Bayesian perspective. Particularly, we introduce the popular l_2-norm and l_1-norm regularisations to the cone-based representation. We call the proposed approach matched shrunken cone detector (MSCD) with two specific models MSCD-l_2 and MSCD-l_1.

A. Regularised Cone

The cone representation of a mixed pixel and its l_2-norm and l_1-norm regularised models are formulated as follows.

Cone representation:

$$\underset{\mathbf{a}\geq\mathbf{0}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{M}\mathbf{a}\|_2^2; \tag{17}$$

l_2-norm regularised cone representation:

$$\underset{\mathbf{a}\geq\mathbf{0}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{M}\mathbf{a}\|_2^2 + \lambda \|\mathbf{a}\|_2^2; \tag{18}$$

l_1-norm regularised cone representation:

$$\underset{\mathbf{a}\geq\mathbf{0}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{M}\mathbf{a}\|_2^2 + \lambda \|\mathbf{a}\|_1. \tag{19}$$

To illustrate the relationship among (17), (18) and (19), we show a two-dimensional cone with different constraints in Fig. 1. It is easily to see that the non-negative linear combination of two endmembers $\mathbf{m}_1$ and $\mathbf{m}_2$ will always lie in the cone. With additional l_2-norm or l_1-norm regularisations, the regions of the constructed vectors are down-sized to be a fan or a triangle, respectively. In other words, l_2-norm and l_1-norm regularisations shrink the value of the coefficient vector **a** for the representation of an HSI pixel.

In the following sections, we shall derive the cone-based binary hypothesis models corresponding to the optimisation problems of (17), (18) and (19), respectively.

B. Regularised Cone-Based Estimators of Coefficient Vectors

The cone-based binary hypothesis models for target detection can be formulated as the model in (7) but with additional constraints. Then we call such models corresponding to (17), (18) and (19) matched cone detector (MCD), matched shrunken cone detector with l_2-norm regularisation (MSCD-l_2) and matched shrunken cone detector with l_1-norm regularisation (MSCD-l_1), respectively.

MCD: given the non-negative constraints (17), the MLEs of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ for models H_0 and H_1 of (7) are given by

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\geq\mathbf{0}}{\operatorname{argmin}} \left\{\|\mathbf{x} - \mathbf{M}_B\boldsymbol{\beta}\|_2^2\right\} \text{ and} \tag{20}$$

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}\geq\mathbf{0}}{\operatorname{argmin}} \left\{\|\mathbf{x} - \mathbf{M}\boldsymbol{\alpha}\|_2^2\right\}. \tag{21}$$

MSCD-l_2: given the l_2-norm regularised cone representation in (18), the estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ of (7) are given by

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\geq\mathbf{0}}{\operatorname{argmin}} \left\{\|\mathbf{x} - \mathbf{M}_B\boldsymbol{\beta}\|_2^2 + \lambda_0 \|\boldsymbol{\beta}\|_2^2\right\} \text{ and} \tag{22}$$

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}\geq\mathbf{0}}{\operatorname{argmin}} \left\{\|\mathbf{x} - \mathbf{M}\boldsymbol{\alpha}\|_2^2 + \lambda_1 \|\boldsymbol{\alpha}\|_2^2\right\}. \tag{23}$$

MSCD-l_1: given the l_1-norm regularised cone representation in (19), the estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ of (7) are given by

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\geq\mathbf{0}}{\operatorname{argmin}} \left\{\|\mathbf{x} - \mathbf{M}_B\boldsymbol{\beta}\|_2^2 + \lambda_0 \|\boldsymbol{\beta}\|_1\right\} \text{ and} \tag{24}$$

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}\geq\mathbf{0}}{\operatorname{argmin}} \left\{\|\mathbf{x} - \mathbf{M}\boldsymbol{\alpha}\|_2^2 + \lambda_1 \|\boldsymbol{\alpha}\|_1\right\}. \tag{25}$$

IV. Bayesian Derivations of MSCD

Given the cone representation under the null hypothesis H_0 of (7) and Bayes' theorem

$$f(\boldsymbol{\beta}|\mathbf{x}) = \frac{f(\mathbf{x}|\boldsymbol{\beta})f(\boldsymbol{\beta})}{f(\mathbf{x})}, \tag{26}$$

the *maximum a posteriori* (*MAP*) estimate of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}}\, f(\boldsymbol{\beta}|\mathbf{x}) = \underset{\boldsymbol{\beta}}{\operatorname{argmax}}\, f(\mathbf{x}|\boldsymbol{\beta})f(\boldsymbol{\beta}). \tag{27}$$

As the noise $\mathbf{n}_0 \sim \mathcal{N}(\mathbf{0}, \sigma_{H_0}^2\mathbf{I}_p)$, the likelihood function $f(\mathbf{x}|\boldsymbol{\beta})$ can be formulated as

$$f(\mathbf{x}|\boldsymbol{\beta}) \propto \exp\left\{-\frac{1}{2\sigma_{H_0}^2}\|\mathbf{x} - \mathbf{M}_B\boldsymbol{\beta}\|_2^2\right\}. \tag{28}$$

Similarly, the *MAP* estimate of $\boldsymbol{\alpha}$ in the alternative hypothesis model H_1 is

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmax}}\, f(\boldsymbol{\alpha}|\mathbf{x}) = \underset{\boldsymbol{\alpha}}{\operatorname{argmax}}\, f(\mathbf{x}|\boldsymbol{\alpha})f(\boldsymbol{\alpha}), \tag{29}$$

and as the noise $\mathbf{n}_1 \sim \mathcal{N}(\mathbf{0}, \sigma_{H_1}^2\mathbf{I}_p)$, the likelihood function $f(\mathbf{x}|\boldsymbol{\alpha})$ can be formulated as

$$f(\mathbf{x}|\boldsymbol{\alpha}) \propto \exp\left\{-\frac{1}{2\sigma_{H_1}^2}\|\mathbf{x} - \mathbf{M}\boldsymbol{\alpha}\|_2^2\right\}. \tag{30}$$

In the ordinary cone representations (20) and (21) of the MCD model, improper uniform (non-informative) prior distributions are actually implied for parameters $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, with $\boldsymbol{\beta} \geq \mathbf{0}$ and $\boldsymbol{\alpha} \geq \mathbf{0}$. However, in the proposed regularised MSCD-l_2 and MSCD-l_1, multivariate folded distributions are in fact utilised as the priors for the estimation of $\boldsymbol{\beta}$ in (22) and (24) and $\boldsymbol{\alpha}$ in (23) and (25), as we shall show below.

A. Folded Distributions

Suppose that the pdf of a random variable Y is $g(y)$ with $y \in \mathbb{R}$. The folding of $g(y)$ over to the non-negative line is accomplished via transform

$$X = |Y|, \tag{31}$$

where X is a random variable on the non-negative real line $\mathbb{R}_+ = [0, \infty)$ with pdf $f(x)$ [26]:

$$f(x) = g(x) + g(-x),\ x \in \mathbb{R}_+. \tag{32}$$

If we treat coefficients β_i and α_i in (7) as random variables, then the non-negative constraints on them imply that their pdf are on $\mathbb{R}_+$. We shall identify that a multivariate folded Gaussian distribution and a multivariate folded Laplace distribution are the prior distributions of coefficients in the proposed MSCD-l_2 and MSCD-l_1, respectively.

B. Prior Distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ in MSCD-l_2

A univariate half-Gaussian distribution is defined as follows. If $Y \sim N(0, \sigma^2)$ with mean zero, then $X = |Y|$ follows a half-Gaussian distribution

$$f(x) = \frac{2}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right), x \geq 0, \tag{33}$$

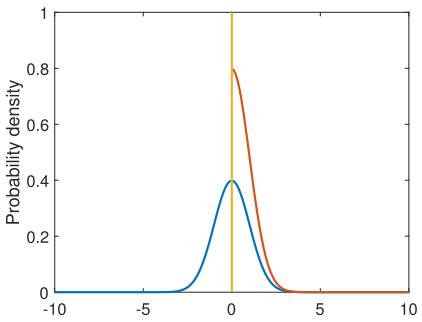


Fig. 2. Illustration of a half-Gaussian distribution.

with mean

$$E(X) = \sqrt{2/\pi}\sigma, \tag{34}$$

and variance

$$var(X) = \sigma^2(1 - 2/\pi). \tag{35}$$

An illustration of the half-Gaussian distribution is shown in Fig. 2. The half-Gaussian distribution is a special case of the folded version of Gaussian distribution $N(\mu, \sigma^2)$ when $\mu = 0$.

We shall identify that, if two *multivariate half-Gaussian distributions* are imposed on the coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively, as the prior distributions, then the estimators (22) and (23) of MSCD-l_2 can be derived in a Bayesian way.

In the model of the null hypothesis H_0 of the proposed MSCD-l_2, let us assume a multivariate half-Gaussian distribution as the prior for the coefficient vector $\boldsymbol{\beta}$. Specifically, suppose that a vector $\mathbf{s} = [s_1, \ldots, s_{N_b}]^T$ follows a multivariate Gaussian distribution $N(\mathbf{0}, \sigma_\beta^2\mathbf{I}_{N_b})$, where $\mathbf{I}_{N_b}$ is the $N_b \times N_b$ identity matrix, then $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_{N_b}]^T$ follows a multivariate half-Gaussian distribution with $\beta_i = |s_i|$ and $\beta_i \geq 0$, where $i = 1, \ldots, N_b$. The expectation of $\boldsymbol{\beta}$ is

$$E(\boldsymbol{\beta}) = \sqrt{2/\pi}\sigma_\beta\mathbf{1}_{N_b} \in \mathbb{R}^{N_b},$$

where $\mathbf{1}_{N_b} = [1, \ldots, 1]^T$ is an N_b-dimensional vector of all ones; the covariance matrix

$$COV(\boldsymbol{\beta}) = \sigma_\beta^2(1 - 2/\pi)\mathbf{I}_{N_b} \in \mathbb{R}^{N_b \times N_b},$$

and the pdf is

$$f(\boldsymbol{\beta}) = \frac{1}{(\frac{1}{2}\pi\sigma_\beta^2)^{N_b/2}}\exp\left(-\frac{||\boldsymbol{\beta}||_2^2}{2\sigma_\beta^2}\right). \tag{36}$$

In MSCD-l_2, placing the likelihood function $f(\mathbf{x}|\boldsymbol{\beta})$ (28) and the prior distribution $f(\boldsymbol{\beta})$ (36) into the *MAP* estimate $f(\boldsymbol{\beta}|\mathbf{x})$ (27) and taking a logarithm, we have

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= \underset{\boldsymbol{\beta}\geq\mathbf{0}}{\operatorname{argmax}}\log\{f(\boldsymbol{\beta}|\mathbf{x})\} \\
&= \underset{\boldsymbol{\beta}\geq\mathbf{0}}{\operatorname{argmax}}\log\{f(\mathbf{x}|\boldsymbol{\beta})f(\boldsymbol{\beta})\} \\
&= \underset{\boldsymbol{\beta}\geq\mathbf{0}}{\operatorname{argmax}}\left\{-\frac{1}{2\sigma_{H_0}^2}\|\mathbf{x} - \mathbf{M}_B\boldsymbol{\beta}\|_2^2 - \frac{1}{2\sigma_\beta^2}\|\boldsymbol{\beta}\|_2^2\right\} \\
&= \underset{\boldsymbol{\beta}\geq\mathbf{0}}{\operatorname{argmin}}\left\{\|\mathbf{x} - \mathbf{M}_B\boldsymbol{\beta}\|_2^2 + \lambda_0\|\boldsymbol{\beta}\|_2^2\right\},
\end{aligned} \tag{37}$$

where $\lambda_0 = \sigma_{H_0}^2/\sigma_\beta^2$. In this way, parameter λ_0 effectively controls the degree of shrinkage via the ratio of two variances $\sigma_{H_0}^2$ and σ_β^2. Equation (37) is exactly the same as model (22).

Similarly, let us assume the prior distribution of coefficients $\boldsymbol{\gamma}$ of the target endmembers in the alternative hypothesis H_1 is a multivariate half-Gaussian distribution, with the expectation

$$E(\boldsymbol{\gamma}) = \sqrt{2/\pi}\sigma_\gamma \mathbf{1}_{N_t} \in \mathbb{R}^{N_t},$$

where $\mathbf{1}_{N_t} = [1, \ldots, 1]^T$ is an N_t-dimensional vector; the covariance matrix

$$COV(\boldsymbol{\gamma}) = \sigma_\gamma (1 - 2/\pi)\mathbf{I}_{N_t} \in \mathbb{R}^{N_t \times N_t},$$

where $\mathbf{I}_{N_t}$ is the $N_t \times N_t$ identity matrix; and the pdf is

$$f(\boldsymbol{\gamma}) = \frac{1}{(\frac{1}{2}\pi\sigma_\gamma^2)^{N_t/2}} \exp\left(-\frac{||\boldsymbol{\gamma}||_2^2}{2\sigma_\gamma^2}\right). \tag{38}$$

Then the concatenated $\boldsymbol{\alpha}$ in model H_1 is actually assumed to follow a half-Gaussian distribution with mean

$$E(\boldsymbol{\alpha}) = \sqrt{2/\pi}[\sigma_\gamma, \ldots, \sigma_\gamma, \sigma_\beta, \ldots, \sigma_\beta]^T \in \mathbb{R}^{(N_t+N_b)}. \tag{39}$$

Let $\boldsymbol{\Sigma}$ denote an $(N_t+N_b)\times(N_t+N_b)$ diagonal matrix equal to diag $\left([\sigma_\gamma, \ldots, \sigma_\gamma, \sigma_\beta, \ldots, \sigma_\beta]^T\right)$. Then the covariance matrix of $\boldsymbol{\alpha}$ is

$$COV(\boldsymbol{\alpha}) = (1 - 2/\pi)\boldsymbol{\Sigma}, \tag{40}$$

which is an $(N_t + N_b) \times (N_t + N_b)$ matrix; and the pdf is

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

where $\sigma_i = \sigma_\gamma$ for $i = 1, \ldots, N_t$ and $\sigma_i = \sigma_\beta$ for $i = N_t + 1, \ldots, N_t + N_b$.

When $\sigma_\gamma = \sigma_\beta$ and we let both of them be σ_α, (41) can be simplified to

$$f(\boldsymbol{\alpha}) = \frac{1}{(\frac{1}{2}\pi\sigma_\alpha^2)^{(N_t+N_b)/2}} \exp\left(-\frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right). \tag{42}$$

Then placing the likelihood function $f(\mathbf{x}|\boldsymbol{\alpha})$ (30) and the prior distribution (42) into the *MAP* estimate $f(\boldsymbol{\alpha}|\mathbf{x})$ (29), we have

$$\begin{aligned}
\hat{\boldsymbol{\alpha}} &= \underset{\boldsymbol{\alpha}\geq\mathbf{0}}{\operatorname{argmax}} \log\{f(\boldsymbol{\alpha}|\mathbf{x})\} \\
&= \underset{\boldsymbol{\alpha}\geq\mathbf{0}}{\operatorname{argmax}} \log\{f(\mathbf{x}|\boldsymbol{\alpha})f(\boldsymbol{\alpha})\} \\
&= \underset{\boldsymbol{\alpha}\geq\mathbf{0}}{\operatorname{argmax}} \left\{-\frac{1}{2\sigma_{H_1}^2}\|\mathbf{x} - \mathbf{M}\boldsymbol{\alpha}\|_2^2 - \frac{1}{2\sigma_\alpha^2}\|\boldsymbol{\alpha}\|_2^2\right\} \\
&= \underset{\boldsymbol{\alpha}\geq\mathbf{0}}{\operatorname{argmin}} \left\{\|\mathbf{x} - \mathbf{M}\boldsymbol{\alpha}\|_2^2 + \lambda_1\|\boldsymbol{\alpha}\|_2^2\right\},
\end{aligned} \tag{43}$$

where $\lambda_1 = \sigma_{H_1}^2/\sigma_\alpha^2$ is the shrinkage parameter. Equation (43) is exactly the same as model (23).

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

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}\geq\mathbf{0}}{\operatorname{argmin}} \left\{\|\mathbf{x} - \mathbf{M}\boldsymbol{\alpha}\|_2^2 + \sum_{i=1}^{N_t+N_b} \lambda_{1i}\alpha_i^2\right\}. \tag{44}$$

In (44), when $i = 1, \ldots, N_t$, we have $\lambda_{1i} = \sigma_{H_1}^2/\sigma_\gamma^2$, and when $i = N_t + 1, \ldots, N_t + N_b$, we have $\lambda_{1i} = \sigma_{H_1}^2/\sigma_\beta^2$.

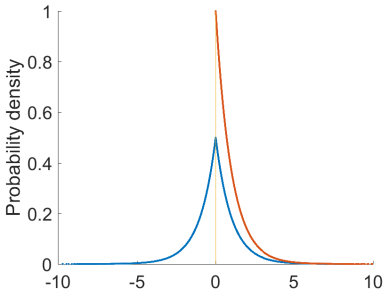


Fig. 3. Illustration of a half-Laplace distribution.

C. Prior Distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ in MSCD-l_1

A Laplace distribution is defined as follows. If a random variable Y has a Laplace distribution $\mathcal{L}(\mu, b)$, then it has mean μ, variance $2b^2$, and pdf

$$g(y) = \frac{1}{2b}\exp -\frac{|y - \mu|}{b}, \ y \in \mathbb{R}. \tag{45}$$

A folded Laplace distribution is also accomplished via transform (31), and the pdf of the transformed random variable X becomes (32). Placing (45) in (32), we have the pdf of a folded Laplace distribution [26]:

$$f(x) = \frac{1}{b}\begin{cases}\exp(-\frac{\mu}{b})\cosh(\frac{x}{b}) & \text{for } 0 \leq x < \mu, \\ \exp(-\frac{x}{b})\cosh(\frac{\mu}{b}) & \text{for } \mu \leq x.\end{cases} \tag{46}$$

Specifically, when $\mu = 0$, (46) reduces to

$$f(x) = \frac{1}{b}\exp\left(-\frac{x}{b}\right), \ x \in \mathbb{R}_+, \tag{47}$$

which is the pdf of a half-Laplace distribution with mean b.

We shall also identify that, if two *multivariate half-Laplace distributions* are imposed on the coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively, as the prior distributions, then the estimators (24) and (25) of MSCD-l_1 can be derived in a Bayesian way.

Let a random multivariate vector $\mathbf{v} = [v_1, \ldots, v_{N_b}]^T$ have a multivariate Laplace distribution $\mathcal{L}(\mathbf{0}, \varphi_\beta\mathbf{I}_{N_b})$. For model (24), coefficient vector $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_{N_b}]^T$ follows a multivariate half-Laplace distribution if $\beta_i = |v_i|$ for $i = 1, \ldots, N_b$. In this case, the mean of $\boldsymbol{\beta}$ is $E(\boldsymbol{\beta}) = \varphi_\beta\mathbf{1}_{N_b}$ and the pdf is

$$f(\boldsymbol{\beta}) = \frac{1}{\varphi_\beta^{N_b}}\prod_{i=1}^{N_b}\exp\left(-\frac{\beta_i}{\varphi_\beta}\right), \ \text{for } \beta_i \geq 0. \tag{48}$$

Then placing the likelihood function $f(\mathbf{x}|\boldsymbol{\beta})$ (28) and the prior distribution $f(\boldsymbol{\beta})$ (48) into the *MAP* function $f(\boldsymbol{\beta}|\mathbf{x})$ (27) and taking the logarithm, we have

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= \underset{\boldsymbol{\beta}\geq\mathbf{0}}{\operatorname{argmax}} \log\{f(\boldsymbol{\beta}|\mathbf{x})\} \\
&= \underset{\boldsymbol{\beta}\geq\mathbf{0}}{\operatorname{argmax}} \log\{f(\mathbf{x}|\boldsymbol{\beta})f(\boldsymbol{\beta})\} \\
&= \underset{\boldsymbol{\beta}\geq\mathbf{0}}{\operatorname{argmax}} \left\{-\frac{1}{2\sigma_{H_0}^2}\|\mathbf{x} - \mathbf{M}_B\boldsymbol{\beta}\|_2^2 - \frac{1}{\varphi_\beta}\sum_{i=1}^{N_b}\beta_i\right\} \\
&= \underset{\boldsymbol{\beta}\geq\mathbf{0}}{\operatorname{argmax}} \left\{-\frac{1}{2\sigma_{H_0}^2}\|\mathbf{x} - \mathbf{M}_B\boldsymbol{\beta}\|_2^2 - \frac{1}{\varphi_\beta}\|\boldsymbol{\beta}\|_1\right\} \\
&= \underset{\boldsymbol{\beta}\geq\mathbf{0}}{\operatorname{argmin}} \left\{\|\mathbf{x} - \mathbf{M}_B\boldsymbol{\beta}\|_2^2 + \lambda_0\|\boldsymbol{\beta}\|_1\right\},
\end{aligned} \tag{49}$$

where $\lambda_0 = 2\sigma^2_{H_0}/\varphi_\beta$ controls the degree of shrinkage through the ratio of $2\sigma^2_{H_0}$ and φ_β. Equation (49) is exactly the same as model (25).

In the same fashion, the prior distribution of coefficients $\boldsymbol{\gamma}$ of the target endmembers in the alternative model H_1 is also assumed to be a multivariate half-Laplace distribution with pdf

$$f(\boldsymbol{\gamma}) = \frac{1}{\varphi_\gamma^{N_t}} \prod_{i=1}^{N_t} \exp\left(-\frac{\gamma_i}{\varphi_\gamma}\right), \text{ for } \gamma_i \geq 0. \tag{50}$$

As a result, the concatenated coefficients $\boldsymbol{\alpha}$ in model H_1 is in fact assumed to follow a multivariate half-Laplace distribution as well, with pdf

$$f(\boldsymbol{\alpha}) = \prod_{i=1}^{N_t+N_b} \frac{1}{\varphi_i} \exp\left(-\frac{\alpha_i}{\varphi_i}\right), \text{ for } \alpha_i \geq 0, \tag{51}$$

where $\varphi_i = \varphi_\gamma$ for $i = 1, \ldots, N_t$ and $\varphi_i = \varphi_\beta$ for $i = N_t + 1, \ldots, N_t + N_b$.

As with the derivations in section IV-B, when we have $\varphi_\gamma = \varphi_\beta$ and let both of them to be φ_α, (51) can be rewritten as

$$f(\boldsymbol{\alpha}) = \frac{1}{\varphi_\alpha^{N_t+N_b}} \exp\left(-\frac{||\boldsymbol{\alpha}||_1}{\varphi_\alpha}\right), \text{ for } \alpha_i \geq 0. \tag{52}$$

Then placing the likelihood function $f(\mathbf{x}|\boldsymbol{\alpha})$ (30) and the prior distribution $f(\boldsymbol{\alpha})$ (52) into the *MAP* estimate of $\boldsymbol{\alpha}$ (29) and taking the logarithm, we have

$$\begin{aligned}
\hat{\boldsymbol{\alpha}} &= \underset{\boldsymbol{\alpha}\geq\mathbf{0}}{\operatorname{argmax}} \log\{f(\boldsymbol{\alpha}|\mathbf{x})\} \\
&= \underset{\boldsymbol{\alpha}\geq\mathbf{0}}{\operatorname{argmax}} \log\{f(\mathbf{x}|\boldsymbol{\alpha}) f(\boldsymbol{\alpha})\} \\
&= \underset{\boldsymbol{\alpha}\geq\mathbf{0}}{\operatorname{argmax}} \left\{ -\frac{1}{2\sigma^2_{H_1}} \|\mathbf{x} - \mathbf{M}\boldsymbol{\alpha}\|_2^2 - \frac{1}{\varphi_\alpha} \|\boldsymbol{\alpha}\|_1 \right\} \\
&= \underset{\boldsymbol{\alpha}\geq\mathbf{0}}{\operatorname{argmin}} \left\{ \|\mathbf{x} - \mathbf{M}\boldsymbol{\alpha}\|_2^2 + \lambda_1 \|\boldsymbol{\alpha}\|_1 \right\},
\end{aligned} \tag{53}$$

where λ_1 is a shrinkage parameter equal to $2\sigma^2_{H_1}/\varphi_\alpha$. Equation (53) is exactly the same as model (25).

Again, (53) can be generalised as

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}\geq\mathbf{0}}{\operatorname{argmin}} \left\{ \|\mathbf{x} - \mathbf{M}\boldsymbol{\alpha}\|_2^2 + \sum_{i=1}^{N_t+N_b} \lambda_{1i}\alpha_i \right\}, \tag{54}$$

where $\lambda_{1i} = 2\sigma^2_{H_1}/\varphi_\gamma$ for $i = 1, \ldots, N_t$ and $\lambda_{1i} = 2\sigma^2_{H_1}/\varphi_\beta$ for $i = N_t + 1, \ldots, N_t + N_b$.

It is worth noting that there is often only one target spectrum available in practice for HSI target detection. In such case, the target training sample $\mathbf{M}_T$ is a $p \times 1$ single vector instead of a $p \times N_t$ matrix. Then the variance σ_γ defined in MSCD-l_2 and the diversity φ_γ in MSCD-l_1 are both have to be set as ∞, since there is no σ_γ and ϕ_γ can be estimated from the target samples. In other words, we actually do not shrink the coefficient $\gamma \in \mathbb{R}$ for the target subset $\mathbf{M}_T$ so long as $N_t = 1$, and let non-negative projection of a test HSI pixel $\mathbf{x}$ onto the target endmember to be as much as possible.

D. Regularisation and Prior Distributions of MSCD

To adjust (and often improve) the performance of a statistical model like MSD or MCD, some prior domain knowledge about the model, particularly the coefficients, can be incorporated by imposing regularisation (a frequentist fashion) or assuming the prior distributions (a Bayesian fashion). These two ways, although from different statistical schools of thinking and inference, can often achieve the same effect, in particular if we can find the pair of a regularisation term and a prior distribution. That is, deriving the corresponding prior distribution to a regularisation term can not only provide a theoretical justification of the latter, but also assist a deeper understanding of the latter; and vice versa. This inspires our derivation of MSCD from the Bayesian perspective.

Specifically, the benefit from proposing MSCD-l_2 and MSCD-l_1 can be understood from both regularisation and Bayesian points of view.

In MSCD-l_2, an l_2-norm regularisation term is added to impose constraints on the combination coefficients in the model of MCD. This will shrink the value of the coefficients and thus reduce the variances of the estimated coefficients, as usually achieved by a shrinkage methods [20]. From the Bayesian perspective, as the coefficients are non-negative, such an l_2-norm regularisation can be derived as corresponding to a multivariate half-Gaussian prior distribution for the coefficients, as we have shown in section IV-B. Equivalently, using such a prior will reduce the posterior variances of the coefficients, in a Bayesian sense. On the one hand, the original MCD models (20) and (21) are equivalent to (37) and (43) when λ_0 and λ_1 are zeros, which implies the use of prior distributions of infinite prior variance. In contrast, the non-zero shrinkage parameters λ_0 and λ_1 in (37) and (43) imply a finite prior variances for the coefficients. On the other hand, with such a prior, the posterior variance of a coefficient will be smaller than the variance of the estimator inferred from the likelihood only. Provided with the lower variance, MSCD-l_2 can provide more stable classification performance than MCD.

The case of MSCD-l_1 is similar to MSCD-l_2, in terms of shrinkage, though the l_1-norm regularisation on the coefficients of the cone representation-based MCD implies a multivariate half-Laplace prior distribution for the coefficients, as we have shown in section IV-C. In fact, as well known, l_1-norm regularisation (like lasso) or a Laplace prior distribution can induce not only shrinkage of the values of the coefficients, but also zero values of some coefficients, i.e. the sparsity of the coefficient vectors. This actually implies an endmember selection in the cone representation for HSI target detection.

V. EXPERIMENTS

We conduct target detection experiments on two real hyperspectral datasets for sub-pixel target detection and full-pixel target detection, respectively. For sub-pixel target detection, a target appearing in an HSI is smaller than an HSI pixel. In this case we compare the target detection methods on the Hymap dataset [29], which was captured at the location of Cook city, USA. For full-pixel target detection, a target appearing in an HSI can occupy more than one HSI pixel. We use the dataset collected by Airborne Visible/Infrared Imaging

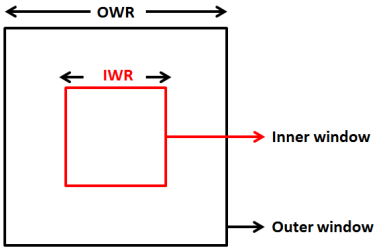


Fig. 4. An illustration of the dual-window scheme adopted for adaptively sampling background. For fair comparison, we set OWR = 15 × 15 and IWR = 9 × 9 for all the compared target detection methods listed in Table IV.

Spectrometer (AVIRIS) from San Diego, CA, USA to evaluate the performance of detecting the full-pixel targets.

We compare the proposed methods MSCD-l_2 and MSCD-l_1 with four types of target detectors: 1) the classical baseline methods ACE and CEM, which are not affected by the dual-window scheme; 2) the cone representation-based detector MCD in (20) and (21); 3) the subspace detectors OSP [6] and MSD [7]; and 4) the sparse representation-based detectors STD [10] and SRBBH [11]. For the proposed MSCD-l_2 and MSCD-l_1, we adopt the MATLAB codes provided by [23] and on http://www.yelab.net/software/SLEP/ to solve the l_2-norm regularised cone model (18) and the l_1-norm regularised cone model (19), respectively.

It is worth noting that, in real target detection problems, it is difficult to obtain training background pixels in a global approach. Instead, most of works on target detection adopt a local and adaptive approach to obtaining the background samples. It is believed that, if the target samples appearing in an HSI scene are sparse enough, we can use the neighbouring HSI pixels around a test HSI pixel as a set of local background samples. Therefore as with [3], [10], [11], and [27], we adopt the dual window scheme. An illustration of a dual window is shown in Fig. 4, which separates a local area of a test HSI 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 but not to be necessarily large. The OWR is set to be outside of the IWR and the HSI pixels lie between the IWR and OWR are used to represent the background samples. However it is often difficult to determine the window sizes in practice. Therefore as with [11], [12], and [30], we empirically set OWR and IWR to be 15 × 15 and 9 × 9 respectively for all compared methods, in order to detect targets appearing in both of Hymap and AVIRIS datasets.

A. The Hymap Dataset

1) Data Description: The Hymap dataset [29] serves as a standard dataset for evaluating hyperspectral target detection, such as in [28] and [31]–[33]. It has a spatial dimension of 280 × 800 and covers 126 spectral bands, as shown in Fig. 5. In this paper, we use the reflectance spectral data, and pre-process the Hymap dataset to remove some bad spectral bands that have negative values in the collected data and finally preserve 119 spectral bands for evaluation.

In the Hymap scene, there exist seven types of targets, including four types of cars (F1, F2, F3 and F4) and three types of fabrics (V1, V2 and V3). In total, nine target samples

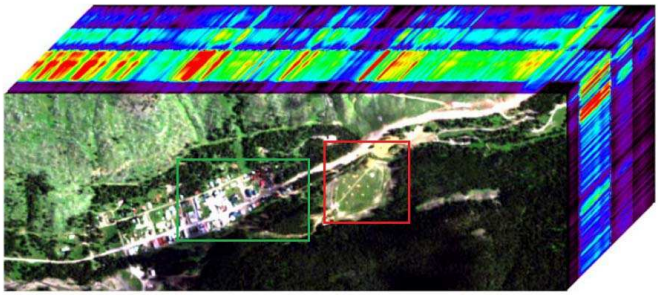
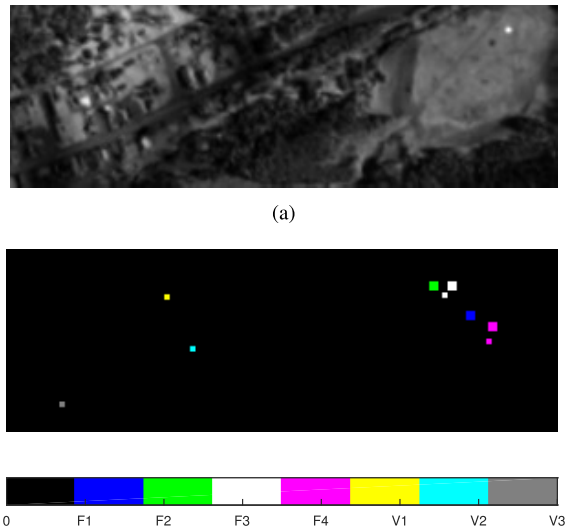


Fig. 5. The Hymap dataset with a spatial size of 280 × 800 and 126 spectral bands [29].



(a)

(b)

Fig. 6. (a) The Hymap sub-image (100 × 300) of the 33rd spectral band; (b) ROIs of seven types of targets (F1, F2, F3, F4, V1, V2 and V3) in the Hymap sub-image. There are two samples of targets F3 and F4 each, termed F3a and F3b, and F4a and F4b, respectively. The pixel sizes of the ROIs of targets F1, F2, F3a, F3b, F4a, F4b, V1, V2 and V3 are 25, 25, 25, 9, 25, 9, 9, 9 and 9, respectively. Different types of targets are shown in different colours. Pixels in black are background pixels.

need to be identified including F1, F2, F3a, F3b, F4a, F4b, V1, V2 and V3. The details of the targets and their corresponding locations in the scene are summarised in Table I.

The detection is performed for each type of targets. For instance, if we are interested in detecting target F1, other targets that are not of interest will be regarded as backgrounds. Note that the spatial resolution of the Hymap dataset is 3m. From the region of interests (ROIs) of each type of targets as shown in Table I, we can infer that only targets F1 and F2 are nearly of full pixels, as their ROIs occupy 3m × 3m. The rest of targets are all smaller than an HSI pixel, as their ROIs are smaller than 3m × 3m. Therefore, the Hymap dataset is a good example for evaluating the sub-pixel target detection performance of all compared methods.

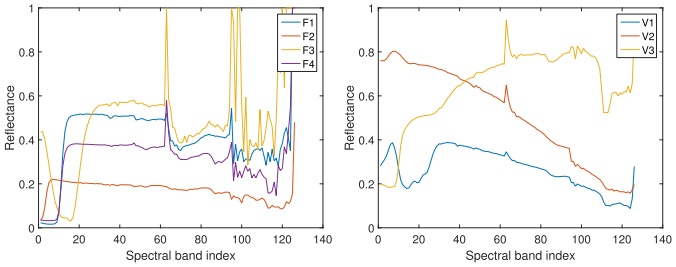
As the targets of interests are mainly located in the central region of the Hymap scene, we spatially crop a 100 × 300 sub-image from the original Fig. 5, as with [27], [28], and [34]. The cropped sub-image as well as the ROIs of all seven types of targets are shown in Fig. 6(a) and Fig. 6(b), respectively.

The ground-truth spectra of seven types of targets (F1-F4 and V1-V3) are given in Fig. 7 and the sample spectral

TABLE I

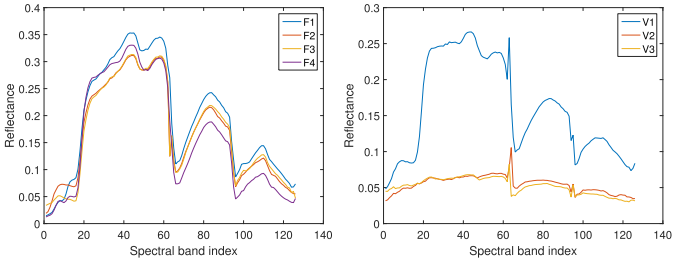
LIST OF THE TARGETS IN THE HYMAP DATASET [29]

Target	F1	F2	F3	F4	V1	V2	V3
Target name	Red cotton	Yellow nylon	Blue cotton	Red Nylon	Green Chevy Blazer	White Toyota T100	Red Subaru GL
ROI	(3m × 3m) (5 × 5 pixels)	(3m × 3m) (5 × 5 pixels)	(2m × 2m & 1m × 1m) (5 × 5 & 3 × 3 pixels)	(2m × 2m & 1m × 1m) (5 × 5 &3 × 3 pixels)	(3 × 3 pixels)	(3 × 3 pixels)	(3 × 3 pixels)
Central coords of ROI	(138, 504)	(122, 484)	(122, 494) & (127, 490)	(144, 516) & (152, 514)	(128, 339)	(156, 353)	(186, 282)
Photo							



(a) (b)

Fig. 7. Rescaled prior spectra of all the targets in the SPL files: (a) fabric panels; (b) vehicles.



(a) (b)

Fig. 8. Rescaled sample spectra of all targets in the Hymap scene: (a) fabric panels; (b) vehicles. The selected sample spectra are located in the central coordinates of the ROIs of F1, F2, F3a, F4a, V1, V2 and V3, respectively.

signatures of the corresponding targets in the Hymap scene are shown in Fig. 8. We can clearly see that target spectra signatures in the scene are very different from those ground-truth spectra, which makes the detections difficult.

2) Experimental Settings: The ROIs mean that a target pixel may appear in any coordinates within the ROIs, and the exact number of pixels of a type of target is unknown. As with the experimental settings in [32] and [34], the criterion for measuring the correct detection is that if at least one pixel in the ROIs is identified as target, then this detection is regarded as a correction detection. Moreover, since the predefined threshold of each compared detector is unknown, we also adopt the *false alarm rate (FAR)* defined in [32] and [34] for measuring the detection performance. The FAR is equal to the number of pixels that are not in the target ROIs but have the test values equal to or greater than the highest test value of pixels within the ROIs, over the total number of pixels in the Hymap HSI, i.e. 30,000 in the example of Fig. 6. Hence we expect to see the lower the FAR, the better the detection performance.

TABLE II

PARAMETER SETTINGS: THE NUMBER r_b OF LEADING EIGENVECTORS OF OSP AND MSD; AND THE SPARSITY LEVEL L OF STD AND SRBBH

Target	r_b		L	
	OSP	MSD	STD	SRBBH
F1	104	116	30	16
F2	119	76	30	30
F3	117	117	10	16
F4	119	116	18	12
V1	119	1	30	10
V2	119	8	10	10
V3	119	7	12	16

TABLE III

PARAMETER SETTINGS: λ_0 AND λ_1 OF MSCD-l_1 AND MSCD-l_2

Target	MSCD-l_1		MSCD-l_2	
	λ_0	λ_1	λ_0	λ_1
F1	1e-04	1e-04	1e-03	1e-01
F2	1e-04	1e-04	1e-02	1e-01
F3	1e-01	1e-03	1e-02	1e-02
F4	1e-00	1e-01	1e-00	1e-00
V1	1e-03	1e-04	1e-02	1e-00
V2	1e-03	1e-04	1e-02	1e-00
V3	1e-01	1e-01	1e+01	1e-00

Parameters of the compared methods should be determined. For the subspace methods OSP and MSD, parameter r_b, which is the number of leading eigenvectors of background subspace, should be determined. For the sparse-representation methods STD and SRBBH, parameter L, which is the sparsity level, should be determined. We shall also determine the parameter λ_0 and λ_1, which are the shrinkage parameters of models H_0 and H_1, respectively, for both the proposed MSCD-l_1 and MSCD-l_2. Due to the limited size of training samples, we are unable to do cross-validation for tuning parameters. Specifically, we have only one ground-truth spectrum of each type target and we do not even have the ground-truth spectra of background samples within the Hymap HSI. Therefore for illustration purposes, we manually tune the parameters of each compared method to their optimal values when the FARs of each method are the lowest, as done by most published works on the Hymap dataset [31], [32], [35]. The range of r_b is [1, 119]; the range of L is [1, 30]. For the proposed MSCD-l_1 and MSCD-l_2, we also manually tune the parameters λ_0 and λ_1 to their optimal values by sweeping the value in [1e-05, 1e-04, 1e-03, 1e-02, 1e-01, 1e-00, 1e+01, 1e+02]. The optimal values of r_b for OSP and MSD and of L for STD and SRBBH are listed in Table II. The optimal values of λ_0 and λ_1 for the proposed MSCD-l_1 and MSCD-l_2 are listed in Table III.

TABLE IV

FALSE ALARM RATE (FAR) OF COMPARED METHODS FOR THE HYMAP DATASET. THE OWR AND IWR ARE SET TO BE 15×15 AND 9×9, RESPECTIVELY, FOR OSP, MSD, STD, SRBBH, MSCD, MSCD-l_1 AND MSCD-l_2. THE MINIMUM FARS ARE IN BOLDFACE

Target	Baseline		Subspace representation		Sparse representation		Cone representation		
	ACE	CEM	OSP	MSD	STD	SRBBH	MCD	MSCD-l_1	MSCD-l_2
F1	1.02e-02	1.19e-02	5.64e-02	1.64e-02	0.04e-02	**0.00e-02**	**0.00e-02**	**0.00e-02**	**0.00e-02**
F2	8.55e-02	1.11e-02	0.56e-02	**0.01e-02**	0.30e-02	0.05e-02	0.24e-02	0.31e-02	0.19e-02
F3	0.57e-02	1.35e-02	2.53e-02	**0.00e-02**	1.84e-02	0.34e-02	0.79e-02	**0.00e-02**	0.02e-02
F4	0.21e-02	0.51e-02	0.13e-02	0.26e-02	0.32e-02	0.11e-02	0.45e-02	0.35e-02	**0.04e-02**
Sum of FARs of detecting F1-F4	10.35e-02	4.16e-02	8.86e-02	1.91e-02	2.5e-02	0.5e-02	1.48e-02	0.66e-02	**0.25e-02**
V1	1.37e-02	1.41e-02	0.52e-02	1.17e-02	0.89e-02	0.36e-02	0.04e-04	0.03e-02	**0.02e-02**
V2	1.34e-02	2.22e-02	3.50e-02	6.73e-02	**1.14e-02**	2.04e-02	5.18e-02	5.22e-02	3.00e-02
V3	19.94e-02	24.87e-02	7.74e-02	2.75e-02	**0.66e-02**	9.56e-02	21.9e-02	7.26e-02	1.85e-02
Sum of FARs of detecting V1-V3	21.42e-02	28.5e-02	11.76e-02	10.65e-02	**2.69e-02**	11.95e-02	27.12-02	12.51e-02	4.87e-02
Sum of FARs	31.77e-02	32.66e-02	20.63e-02	12.56e-02	5.19e-02	12.46e-02	28.60e-02	13.17e-02	**5.12e-02**

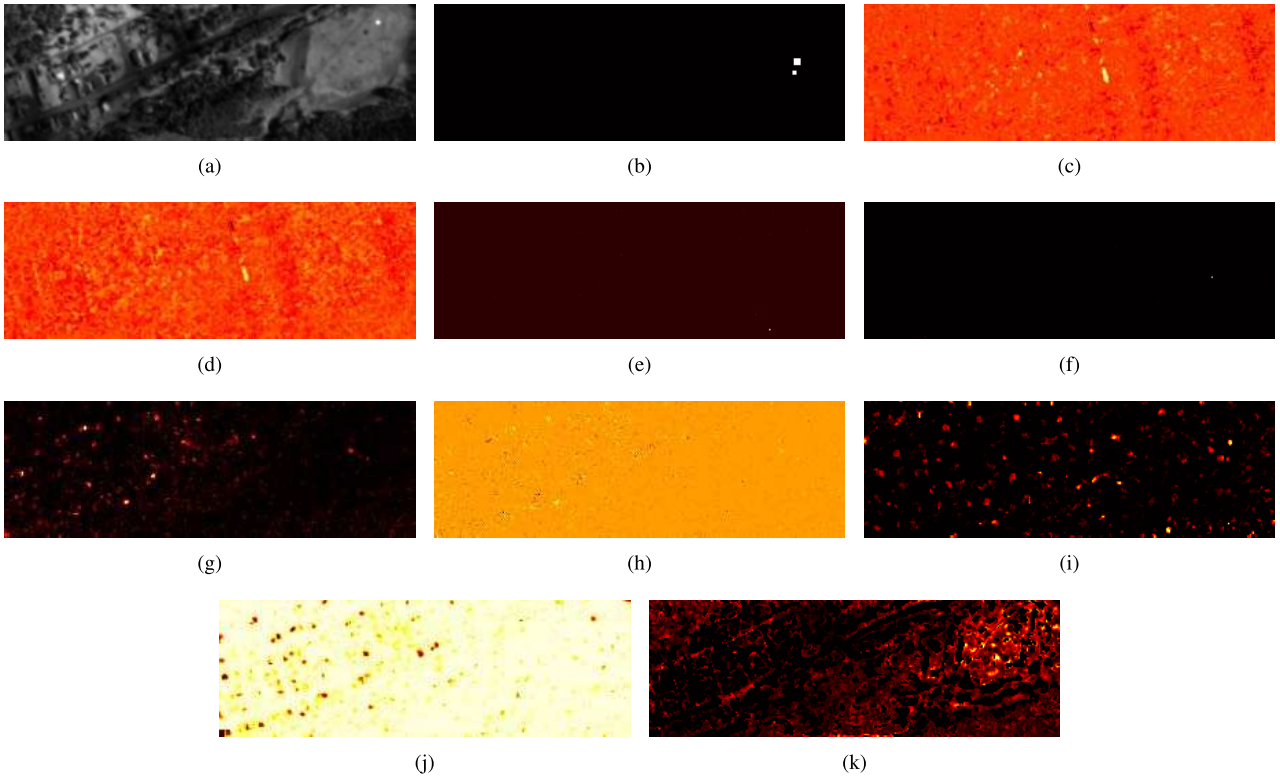
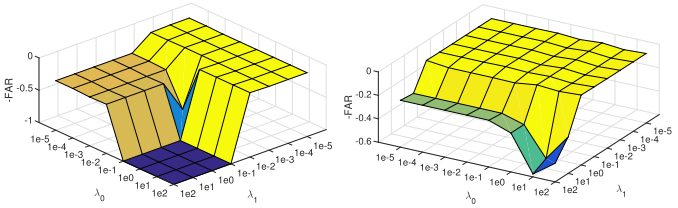


Fig. 9. Prediction maps of test statistics for detecting F4 in the Hymap image. (a) The Hymap HSI of the 33rd spectral band; (b) ground-truth labels of F4; (c) ACE, FAR = 0.21e-02; (d) CEM, FAR = 0.51e-02; (e) OSP, FAR = 0.13e-02; (f) MSD, FAR = 0.26e-02; (g) STD, FAR = 0.32e-02; (h) SRBBH, FAR = 0.11e-02; (i) MCD, FAR = 0.45e-02; (j) MSCD-l_1, FAR = 0.35e-02; (k) MSCD-l_2, FAR = 0.04e-02.

3) Experimental Results and Analysis: The FARs of all compared methods for detecting each type of targets are listed in Table IV. Firstly, for the cone-based detectors, MCD, MSCD-l_2 and MSCD-l_1, we can observe that the proposed MSCD-l_2 (FAR 5.12e-02) and MSCD-l_1 (13.17e-02) outperform MCD (28.60e-02) for detecting different types of targets. This illustrates the effectiveness of incorporating the regularisations into the optimisation of non-negative problems. Furthermore, MSCD-l_2 performs significantly better than MSCD-l_1, which implies that the l_2-norm regularised cone representation is more effective than the l_1-norm regularised cone representation for detecting the targets in the Hymap dataset.

Secondly, comparing all the methods listed in Table IV, we can clearly see that our proposed MSCD-l_2 outperforms ACE, CEM, OSP, MSD, STD, SRBBH, MCD and MSCD-l_1 for detecting targets F1, F4 and V1, and it performs the best in terms of the sum of FARs of detecting fabric targets F1-F4 with FAR as 0.25e-02. More importantly, MSCD-l_2 also outperforms others in detecting all types of targets, i.e. F1-F4 and V1-V3, with the smallest sum of FARs as 5.12e-02. This indicates that the proposed MSCD-l_2 is more competitive than the subspace and sparse-representation methods.

Last but not least, we shall note that, among the compared methods, the subspace method MSD and the sparse-representation method STD perform relatively better than each



(a) (b)

Fig. 10. Effects of λ_0 and λ_1 of (a) MSCD-l_1 and (b) MSCD-l_2 on detecting target F4 in the Hymap dataset. Window sizes: IWR 9×9 and OWR 15×15. Since FAR is the smaller the better, for easier reading, we plot "−FAR" such that the best detection performance occurs at the peak of the surface plot.

of their cohort methods, i.e. MSD is better than OSP and STD is better than SRBBH in terms of the sum of FARs of all targets. STD also has competitive performance for detecting the vehicle targets, particularly V2 and V3. However, both of MSD and STD are not as good as the proposed MSCD-l_2 in terms of the sum of FARs for detecting all targets. This also implies that MSCD-l_2 is more stable than other methods, whatever the types of targets and the sizes of them.

To further illustrate the detection performances of the compared methods, we display the prediction maps of all methods in Fig. 9 for detecting target F4. Fig. 9(b) shows the ground-truth map of target F4. The value of each pixel shown in Fig. 9(c)-9(k) represents the test statistic value of the pixel: the brighter the pixel, the higher the test statistic value, and thus the more likely a target. That is, we expect a good prediction map to show a clear pattern for detecting F4 that the brightnesses of the pixels located within the ROIs of F4 are higher than those outside. From these prediction maps, we can visually observe that 1) ACE (Fig. 9(c)), CEM (Fig. 9(d)), OSP (Fig. 9(e)) and MSD (Fig. 9(f)) have no such a clear pattern; 2) STD (Fig. 9(g)), SRBBH (Fig. 9(h)), MCD (Fig. 9(i)) and MSCD-l_1(Fig. 9(j)) look better, but we can easily spot many outside pixels brighter than the pixels within the ROIs of F4; 3) among all the maps, MSCD-l_2 in Fig. 9(k) looks the best, though it still does not provide a zero FAR (FAR = 0.04e-02 in Table IV), where the bright pixels largely stick around the ground-truth of F4, rather than spread over the scene as in other prediction maps.

4) Discussion on Effects of Parameters: We further investigate the effects of two types of parameters on the performance of our proposed MSCD methods: 1) the shrinkage parameters λ_0 and λ_1; and 2) the window sizes IWR and OWR.

Firstly, from Fig. 10 we can make two observations. 1) The "-FAR" surface of MSCD-l_2 (Fig. 10(b)) is smoother than that of MSCD-l_1 (Fig. 10(a)), which indicates that MSCD-l_2 is less sensitive to the shrinkage parameters λ_0 and λ_1 than MSCD-l_1. 2) For both MSCD-l_2 and MSCD-l_1, the detection performance tends to be stable in a wide range of values of λ_0 and λ_1; that is, the values λ_0 and λ_1 do not have to be exactly the same as used in Table III to achieve a similar performance for MSCD-l_2 and MSCD-l_1.

Secondly, we investigate the effects of window sizes on the performance of the compared detectors: OSP, MSD, STD, SRBBH, MCD, MSCD-l_1 and MSCD-l_2. For a simple and effective exploration, we fix the values of other parameters

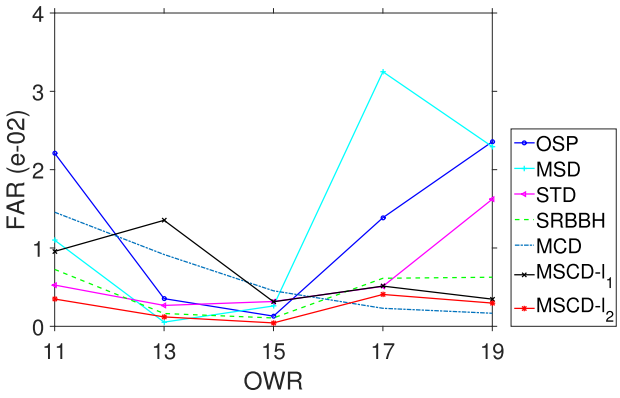
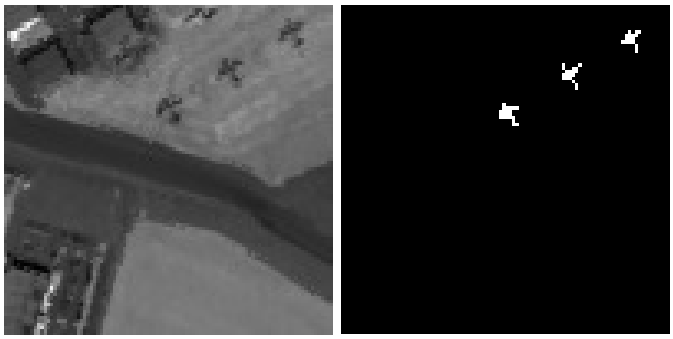


Fig. 11. Effects of window sizes on detecting target F4 in the Hymap dataset. IWR: 9×9; OWR: from 11×11, 13×13, 15×15, 17×17 to 19×19.



(a) (b)

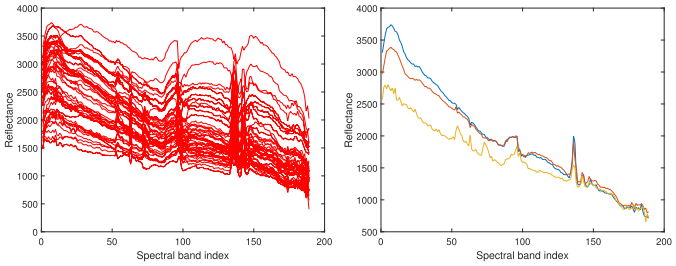
Fig. 12. (a) The AVIRIS sub-image (100×100) of the 45th spectral band; (b) the ground-truth labels of targets including 58 target pixels.

(r_b and L in Table II and λ_0 and λ_1 in Table III), and fix the IWR and tune OWR. From Fig. 11, we can see that, among the detectors, MSCD-l_2 is the most stable with OWR, while the two subspace detectors OSP and MSD are the most sensitive.

B. The AVIRIS Dataset

1) Data Description: The AVIRIS data was captured at an airport in the San Diego, CA, USA with the planes as targets. We select a sub-image that spatially covers a region of 100×100. As with [11] and [12], we remove some bad spectral bands and preserve 189 spectral bands for evaluation. In the AVIRIS scene, there are three planes need to be detected, consisting of 58 HSI pixels that are labelled as target pixels. The hyperspectral image scene and the ground-truth maps are shown in Fig 12(a) and Fig. 12(b), respectively. It is clear that each target plane covers more than one HSI pixel. Hence the AVIRIS dataset adopted here is suitable for evaluating the full-pixel target detection performance of the compared methods.

2) Experimental Settings: Because the labels for individual HSI pixels are available in the AVIRIS dataset, we select the three central HSI pixels of each plane as the prior spectra of target signatures, as with [11] and [12]. The rest of target HSI pixels are used to evaluate the detection performances of methods. The 58 target spectra and the three training target spectra are shown in Fig. 13(a) and Fig. 13(b), respectively. We can observe that the spectra of the target HSI pixels still



(a) (b)

Fig. 13. Spectra of targets in the AVIRIS dataset: (a) all target spectra in the hyperspectral scene; (b) spectra of three training target pixels, which are the central pixels of the three planes, respectively.

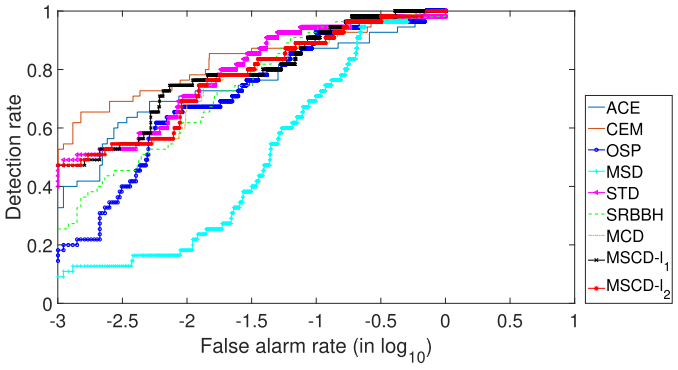


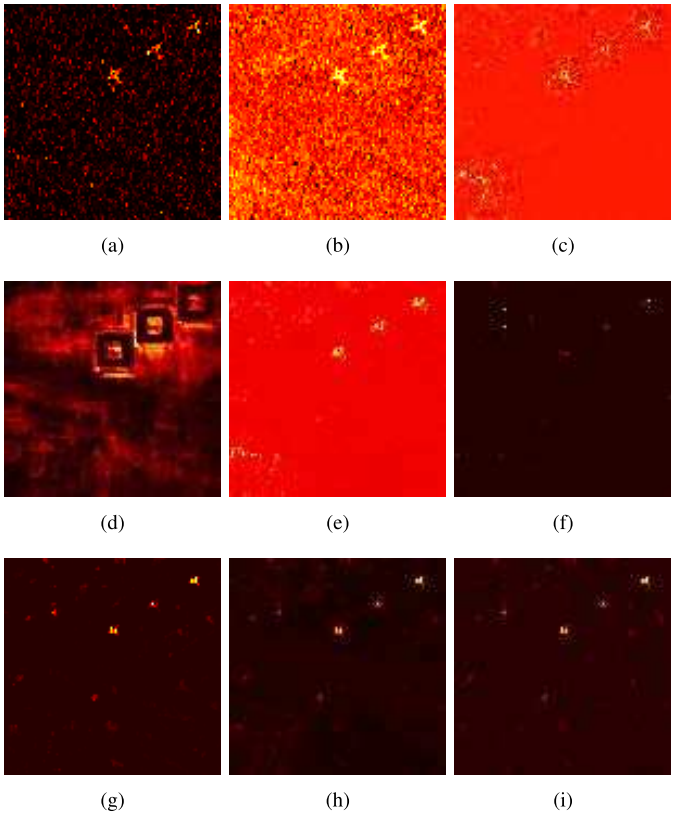
Fig. 14. The ROC curves of the compared methods: ACE, CEM, OSP, MSD, STD, SRBBH, MCD, MSCD-l_1 and MSCD-l_2.

look different from each other. However, compared with Fig. 7 and Fig. 8 for the Hymap dataset, the spectral pattern of the AVIRIS targets may be clearer and the targets may be easier to be detected, as the training target pixels are from the HSI rather than from spectral libraries.

As with [11] and [12], we use the *receiver operating characteristic (ROC)* curves to measure the detection performances for the AVIRIS dataset. The reason of using ROC instead of FAR is that now we have the labelling information for every single target HSI pixel, instead of the only available ROIs in the Hymap dataset. We expect that an ROC curve goes to the top left of the plot, if the detection performance of a method is good. Additionally, we adopt the area under curve (AUC) statistics to quantitatively measure the detection performance in pair with the ROC curves.

Similarly, the parameters of each compare method should be determined: the number of leading eigenvectors r_b for the subspace methods OSP and MSD; the sparsity level L for the SR-based methods; and the shrinkage parameters λ_0 and λ_1 for both of the proposed MSCD-l_1 and MSCD-l_2. Again, for illustration purposes, the parameters are empirically determined and the values are listed in Table V, with the same tuning ranges of values as for the Hymap dataset.

3) Experimental Results and Analysis: The ROC curves of all the compared methods are shown in Fig. 14 and the corresponding AUC statistics are listed in Table V. Once again, we can observe that the proposed MSCD-l_1 and MSCD-l_2 both outperform MCD, which indicates the benefit of incorporating the l_1-norm and l_2-norm regularisations into the cone-based representation for HSI target detection.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 15. Prediction maps for detecting planes in the AVIRIS image. The brighter the pixels, the more likely to be targets. (a) ACE, AUC = 0.9398; (b) CEM, AUC = 0.9596; (c) OSP, AUC = 0.9527; (d) MSD, AUC = 0.9091; (e) STD, AUC = 0.9647; (f) SRBBH, AUC = 0.9547; (g) MCD, AUC = 0.9616; (h) MSCD-l_1, AUC = 0.9713; (i) MSCD-l_2, AUC = 0.9632.

Moreover, the proposed MSCD-l_1 is among the best of all the compared method. This implies that, for detecting full-size target HSI pixels, introducing the sparsity constraints on the coefficients into the MCD can achieve better performance than the l_2-norm constraints on the coefficients. Generally speaking, the cone-representation methods are better than the sparse-representation methods; and the sparse-representation methods are better than the subspace methods for detecting full-size target HSI pixels in the AVIRIS dataset.

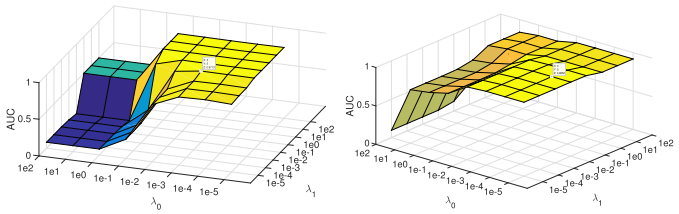
We also plot the prediction maps for all the methods and display them in Fig. 15. It can be seen that the cone-representation methods, i.e. MCD (Fig. 15(g)), MCD-l_1 (Fig. 15(h)) and MCD-l_2 (Fig. 15(i)), look relatively better than the others. The difference among these three prediction maps are not so much. Among the other six methods (ACE, CEM, OSP, MSD, STD and SRBBH), MSD (Fig. 15(d)) looks the worst, as it is badly affected by the dual window scheme (Fig. 4); and STD looks better than ACE, CEM, OSP, MSD and SRBBH. However, the colour contrast in Fig. 15(e) of STD is not as large as those in Fig. 15(g)-15(i) of the cone-representation methods. This means that the test statistics of background pixels and target pixels of MCD, MSCD-l_1 and MSCD-l_2 are more different than those of STD, which further illustrates the stable performances of the cone-based methods for detecting targets in the AVIRIS dataset.

4) Discussion on Effects of Parameters: As with the analysis for the Hymap dataset, we also investigate the effects of

TABLE V

PARAMETERS AND AUC STATISTICS OF THE COMPARED METHODS FOR THE AVIRIS DATASET. THE OWR AND IWR ARE SET TO BE 15×15 AND 9×9, RESPECTIVELY FOR OSP, MSD, STD, SRBBH, MSCD, MSCD-l_1 AND MSCD-l_2. THE MAXIMAL AUC IS IN BOLDFACE

	Baseline		Subspace representation		Sparse representation		Cone representation				
Detector	ACE	CEM	OSP	MSD	STD	SRBBH	MCD	MSCD-l_1		MSCD-l_2	
Parameter	NA	NA	r_b	r_b	L	L	NA	λ_0	λ_1	λ_0	λ_1
	NA	NA	157	7	24	20	NA	1e-03	1e-02	1e-04	1e-02
AUC	0.9398	0.9596	0.9527	0.9091	0.9647	0.9547	0.9616	**0.9713**		0.9632	



(a) (b)

Fig. 16. Effects of λ_0 and λ_1 of (a) MSCD-l_1, (b) MSCD-l_2 on detecting targets in the AVIRIS dataset.

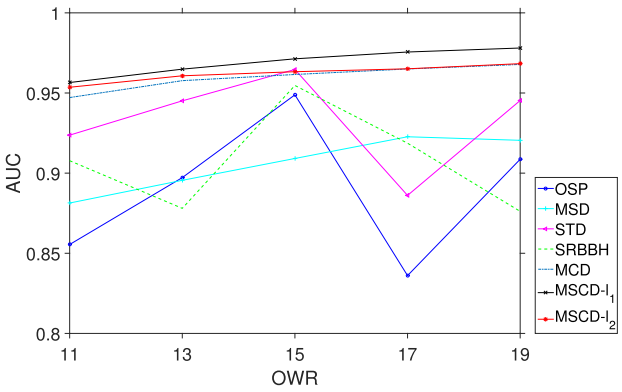


Fig. 17. Effects of window sizes on detecting targets in the AVIRIS dataset. IWR: 9×9; OWR: from 11×11, 13×13, 15×15, 17×17 to 19×19.

TABLE VI

EXECUTION TIME (SEC/PIXEL) SPENT ON THE AVIRIS DATASET

Detector	STD	SRBBH	MCD	MSCD-l_1	MSCD-l_2
Time	5.19e-04	4.95e-04	2.1e-03	1.73e-02	3.51e-02

shrinkage parameters λ_0 and λ_1 and window sizes IWR and OWR on the detection performance on the AVIRIS dataset.

Firstly, from Fig. 16 we can observe two similar patterns to those from Fig. 10: 1) MSCD-l_2 is less sensitive to λ_0 and λ_1; and 2) both MSCD-l_2 and MSCD-l_1 can achieve good performance in a wide range of values of λ_0 and λ_1.

Secondly, Fig. 17 shows that MCD, MSCD-l_1 and MSCD-l_2 are less sensitive to OWR, or say more robust to the variation of background samples, than OSP, MSD, STD and SRBBH.

By analysing the experimental results of the two datasets, we can observe that MSCD-l_2 performs better than MSCD-l_1 for the Hymap dataset, while MSCD-l_1 is better than MSCD-l_2 for the AVIRIS dataset. In line with the debate between choosing sparse representation (lasso) or collaborative representation (ridge regression) in the HSI analysis, both methods have their own advantages and it remains an open question which one is better. We shall also note that MSCD-l_1 and MSCD-l_2 cost more computational resources than ACE, CEM, OSP and MSD, because there is no closed-form solution to the cone-based optimisation. The computational costs of STD, SRBBH, MCD, MSCD-l_1 and MSCD-l_2 are listed in Table VI, as performed on Intel i7-4790 CPU.

VI. CONCLUSION

In this paper, we have proposed a new approach called matched shrunken cone detector (MSCD) for hyperspectral target detection. Two new working models of MSCD, namely MSCD-l_2 and MSCD-l_1, have also been proposed, with the l_2-norm and l_1-norm regularisations incorporated into the MSCD, respectively. Geometrically, we have analysed the underlying effectiveness of MSCD. The values of the coefficients are shrunken within a cone either by the l_2-norm regularisation or the l_1-norm regularisation, which form two different constrained regions for the coefficients. Statistically, we have derived MSCD from the Bayesian perspective. We have shown that if a multivariate half-Gaussian distribution or a multivariate half-Laplace distribution is assumed as the prior distribution of the coefficients, then MSCD-l_2 or MSCD-l_1 can be derived. In our experiments, cases studies on two real hyperspectral datasets have been conducted, with the Hymap dataset to illustrate the sub-pixel target detection and the AVIRIS dataset to illustrate the full-pixel target detection. We have compared four categories detectors including the baseline methods, the subspace methods, the sparse-representation methods and the cone-representation methods. Experimental results on both of the two datasets have showed the competitive performance of the proposed MSCD.

We would like to make two further notes about the Bayesian derivations. One the one hand, in the Bayesian paradigm, the half-Gaussian or half-Laplace prior distribution can be assumed on the basis of our prior knowledge that the model coefficients are positive. In principle any distribution of a positive random variable can be assumed as the prior for such a coefficient; in our case, half-Gaussian and half-Laplace distributions match the l_2-norm and l_1-norm regularisations, respectively. That is, the half-Gaussian and half-Laplace priors provide us with a principled Bayesian interpretation of the two regularised models. On the other hand, if the practitioners hold some specific prior domain knowledge which prefers to be modelled by other positive prior distributions, such as log-normal distributions or gamma distributions, a Bayesian derivation like ours can open a door to different new regularised models, which fit their practice better. This can be an interesting and practically valuable direction to further our principled work presented in this paper.

ACKNOWLEDGMENT

The authors would like to thank Dr. Lefei Zhang, Dr. Bo Du and Dr. Yuxiang Zhang for their help with the datasets and the MATLAB codes of the subspace methods and the sparse-representation methods, and thank the four anonymous reviewers and the associate editor for their constructive comments, in particular on the introduction and the experiments.

REFERENCES

[1] D. Manolakis, E. Truslow, M. Pieper, T. Cooley, and M. Brueggeman, "Detection algorithms in hyperspectral imaging systems: An overview of practical algorithms," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 24–33, Jan. 2014.

[2] N. M. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 34–44, Jan. 2014.

[3] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 25, no. 7, pp. 5–28, Jul. 2010.

[4] S. Matteoli, M. Diani, and J. Theiler, "An overview of background modeling for detection of targets and anomalies in hyperspectral remotely sensed imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2317–2336, Jul. 2014.

[5] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral subpixel target detection using the linear mixing model," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1392–1409, Jul. 2001.

[6] J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.

[7] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, Aug. 1994.

[8] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[9] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

[10] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Sparse representation for target detection in hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 629–640, Jun. 2011.

[11] Y. Zhang, B. Du, and L. Zhang, "A sparse representation-based binary hypothesis model for target detection in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1346–1354, Mar. 2015.

[12] B. Du, Y. Zhang, L. Zhang, and D. Tao, "Beyond the sparsity-based target detector: A hybrid sparsity and statistics-based detector for hyperspectral images," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5345–5357, Nov. 2016.

[13] H. Kwon and N. M. Nasrabadi, "A comparative analysis of kernel subspace target detectors for hyperspectral imagery," *EURASIP J. Appl. Signal Process.*, vol. 2007, no. 1, p. 193, Dec. 2007.

[14] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Kernel sparse representation for hyperspectral target detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2012, pp. 7484–7487.

[15] Y. Zhang, L. Zhang, B. Du, and S. Wang, "A nonlinear sparse representation-based binary hypothesis model for hyperspectral target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2513–2522, Jun. 2015.

[16] Z. Yang, G. Zhou, S. Xie, S. Ding, J.-M. Yang, and J. Zhang, "Blind spectral unmixing based on sparse nonnegative matrix factorization," *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 1112–1125, Apr. 2011.

[17] C. Févotte and N. Dobigeon, "Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4810–4819, Dec. 2015.

[18] C. L. Lawson and R. J. Hanson, "Solving least squares problems," in *Society for Industrial and Applied Mathematics*. Philadelphia, PA, USA: SIAM, 1995. [Online]. Available: http://epubs.siam.org/doi/abs/10.1137/1.9781611971217

[19] D. G. Luenberger, *Optimization by Vector Space Methods*. Hoboken, NJ, USA: Wiley, 1969.

[20] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*, vol. 1. Berlin, Germany: Springer, 2001.

[21] M. Yuan and Y. Lin, "On the non-negative garrotte estimator," *J. Roy. Statist. Soc., B, Statist. Methodol.*, vol. 69, no. 2, pp. 143–161, 2007.

[22] M. Slawski and M. Hein, "Non-negative least squares for high-dimensional linear models: Consistency and sparse recovery without regularization," *Electron. J. Statist.*, vol. 7, pp. 3004–3056, 2013.

[23] R. Zdunek, "Regularized NNLS algorithms for nonnegative matrix factorization with application to text document clustering," in *Computer Recognition Systems*, vol. 4. Berlin, Germany: Springer, 2011, pp. 757–766.

[24] E. Esser, Y. Lou, and J. Xin, "A method for finding structured sparse solutions to nonnegative least squares problems with applications," *SIAM J. Imag. Sci.*, vol. 6, no. 4, pp. 2010–2046, 2013.

[25] D. Aigner, C. K. Lovell, and P. Schmidt, "Formulation and estimation of stochastic frontier production function models," *J. Econometrics*, vol. 6, no. 1, pp. 21–37, 1977.

[26] Y. Liu and T. J. Kozubowski, "A folded Laplace distribution," *J. Statist. Distrib. Appl.*, vol. 2, no. 1, p. 10, 2015.

[27] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.

[28] W. Li, Q. Du, and B. Zhang, "Combined sparse and collaborative representation for hyperspectral target detection," *Pattern Recognit.*, vol. 48, no. 12, pp. 3904–3916, 2015.

[29] D. Snyder, J. Kerekes, I. Fairweather, R. Crabtree, J. Shive, and S. Hager, "Development of a Web-based application to evaluate target finding algorithms," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 2. Jul. 2008, pp. II-915–II-918.

[30] T. Wang, B. Du, and L. Zhang, "A background self-learning framework for unstructured target detectors," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1577–1581, Nov. 2013.

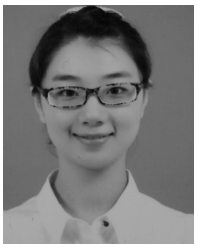
[31] L. Zhang, L. Zhang, D. Tao, and X. Huang, "Sparse transfer manifold embedding for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1030–1043, Feb. 2014.

[32] L. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Hyperspectral remote sensing image subpixel target detection based on supervised metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4955–4965, Aug. 2014.

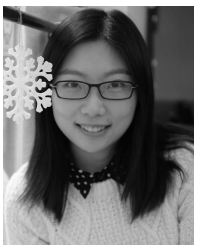
[33] B. Du and L. Zhang, "Target detection based on a dynamic subspace," *Pattern Recognit.*, vol. 47, no. 1, pp. 344–358, 2014.

[34] Y. Zhang, B. Du, L. Zhang, and S. Wang, "A low-rank and sparse matrix decomposition-based Mahalanobis distance method for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1376–1389, Mar. 2016.

[35] L. Gao, B. Yang, Q. Du, and B. Zhang, "Adjusted spectral matched filter for target detection in hyperspectral imagery," *Remote Sens.*, vol. 7, no. 6, pp. 6611–6634, 2015.



Ziyu Wang received the M.Sc. degree in statistics, the M.Res. degree in security science, and the Ph.D. degree in security science and statistics from University College London, in 2012, 2013, and 2017, respectively. Her current research interests include hyperspectral image analysis, sparse representation, and statistical classification.



Rui Zhu received the M.Sc. and Ph.D. degrees in statistics from University College London in 2013 and 2017, respectively. Her research interests include spectral data analysis, hyperspectral image analysis, subspace-based classification methods and metric learning.

Kazuhiro Fukui (M'04) received the Ph.D. degree from the Tokyo Institute of Technology in 2003. He is currently a Professor with the Department of Computer Science, University of Tsukuba. His interests include the theories of computer vision, pattern recognition, and applications of these theories. He is a member of SIAM. He has been serving as a program committee member at many pattern recognition and computer vision conferences, including as an Area Chair of ICPR'12, 14, and 16.

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 been with the Department of Statistical Science, University College London, as a Lecturer and a Senior Lecturer. His current research interests include statistical classification, high-dimensional data analysis, computer vision, and pattern recognition.