GREEDY ALGORITHMS FOR PURE PIXELS IDENTIFICATION IN HYPERSPECTRAL UNMIXING: A MULTIPLE-MEASUREMENT VECTOR VIEWPOINT

Xiao Fu[1]*, Wing-Kin Ma*[1]*, Tsung-Han Chan*[2]*, José M. Bioucas-Dias*[3]*, and Marian-Daniel Iordache*[4]

[1]Dept. Electronic Eng., the Chinese University of Hong Kong, Hong Kong
[2]Advanced Digital Sciences Center, Singapore
[3]Instituto de Telecomunicações, Lisboa, Portugal
[4]Flemish Institute for Technological Research (VITO-TAP), Mol, Belgium

ABSTRACT

This paper studies a multiple-measurement vector (MMV)-based sparse regression approach to blind hyperspectral unmixing. In general, sparse regression requires a dictionary. The considered approach uses the measured hyperspectral data as the dictionary, thereby intending to represent the whole measured data using the fewest number of measured hyperspectral vectors. We tackle this self-dictionary MMV (SD-MMV) approach using greedy pursuit. It is shown that the resulting greedy algorithms are identical or very similar to some representative pure pixels identification algorithms, such as vertex component analysis. Hence, our study provides a new dimension on understanding and interpreting pure pixels identification methods. We also prove that in the noiseless case, the greedy SD-MMV algorithms guarantee perfect identification of pure pixels when the pure pixel assumption holds.

1. INTRODUCTION

Hyperspectral imaging is a sensing technique in which an instrument acquires a set of spectral vectors from a surface of interest. The spectral vector of each pixel carries information of how the reflectance of surface materials interact in the corresponding surface patch. Under the linear mixing model, a measured spectral vector can be represented by a convex combination of the so-called pure spectral vectors — spectral vectors that are purely constituted by a single surface material [1].

Pure pixels identification aims at determining the pure spectral vectors, or simply pure pixels, from the measured hyperspectral data, assuming that pure pixels exist in the measured data. It represents an important class of techniques in blind hyperspectral unmixing, since many algorithms falling in this class are efficient. Numerous pure pixels identification algorithms have been developed for more than a decade, e.g., pure pixel index (PPI) by Boardman *et. al* in 1995 [2], Winter's N-FINDR [3] and its many variants, successive projections algorithm (SPA) [4], automatic target generation process (ATGP) [5], vertex component analysis (VCA) [6], and more recently, successive volume maximization (SVMAX) [7] and the recursive algorithm family in [8]; see [1] for a review. In this paper, we are interested in a very recent development introduced in [9, 10], where a compressive sensing formulation is used to tackle pure pixels identification. The idea is to use the measured spectral vectors as the dictionary to perform multiple-measurement vector (MMV)-based sparse regression [11, 12]. By doing so, one intends to use the fewest number of measured data, which would be the pure pixels, to represent the whole set of measured data. This self-dictionary MMV (SD-MMV) approach was previously studied under convex relaxation [9, 10]. The present work considers SD-MMV under greedy pursuit. We will show that several representative pure pixels identification algorithms are identical or very similar to greedy-based SD-MMV. This result is interesting, and establishes a new connection between the conventional pure pixels identification and relatively novel SD-MMV approaches. The pure pixel identifiability of the greedy-based SD-MMV algorithms will also be studied.

2. REVIEW OF MMV

This section reviews the MMV problem in a general context [11–13]. The connection between MMV and hyperspectral unmixing will be discussed in the next section. In MMV, we consider the following signal model

$$\mathbf{X} = \mathbf{B}\mathbf{C}, \tag{1}$$

where $\mathbf{X} \in \mathbb{R}^{M\times L}$ is a multiple-measurement matrix, in which each column, denoted by $\mathbf{x}_i \in \mathbb{R}^M$, $i = 1, \ldots, L$, is a single-measurement vector (SMV), $\mathbf{B} \in \mathbb{R}^{M\times N}$ is a basis, and $\mathbf{C} \in \mathbb{R}^{N\times L}$ is a coefficient matrix. Here, note that $\mathbf{B}$ is an overcomplete (and given) dictionary. We wish to use the fewest number of columns of $\mathbf{B}$, also known as *atoms*, to do the representation in (1). This problem is the same as finding a $\mathbf{C}$ whose number of nonzero rows is the smallest. Hence,

the MMV problem can be formulated as

$$\begin{aligned} \min_{\mathbf{C}\in\mathbb{R}^{N\times L}} \ & \|\mathbf{C}\|_{\text{row}-0} \\ \text{s.t. } & \mathbf{X}=\mathbf{B}\mathbf{C}, \end{aligned} \tag{2}$$

where $\|\mathbf{C}\|_{\text{row}-0}$ denotes the number of nonzero rows in $\mathbf{C}$, that is, if $\mathbf{c}^i$ denotes the ith row vector of $\mathbf{C}$, then $\|\mathbf{C}\|_{\text{row}-0} = \sum_{i=1}^{M} I(\|\mathbf{c}^i\|_2)$, where $\|\cdot\|_q$ denotes the ℓ_q-norm, and $I(x) = 1$ if $x > 0$ and $I(x) = 0$ if $x = 0$. Notice that when $L = 1$, Problem (2) reduces to the familiarized SMV problem in compressive sensing.

Like its SMV counterpart, the MMV problem in (2) is NP-hard in general. As such, approximation approaches should be sought. One obvious approach is convex relaxation, following the ℓ_1-norm approximation in SMV. Specifically, the approach works by replacing $\|\mathbf{C}\|_{\text{row}-0}$ in Problem (2) by a convex surrogate, such as $\|\mathbf{C}\|_{q,1} = \sum_{i=1}^{N} \|\mathbf{c}^i\|_q$ for a given $q \geq 1$, and then solve the subsequent convex problem to approximate Problem (2). Readers are referred to [11, 13] for the details. Here, our interest lies in the greedy approach. We will review two methods, namely, simultaneous orthogonal matching pursuit (SOMP) and reduced MMV and boost (ReMBo).

2.1. Simultaneous Orthogonal Matching Pursuit

SOMP can be seen as a direct extension of orthogonal matching pursuit (OMP) for SMV [11, 12]. It aims at determining a set of indices, say, $\hat{i}_1, \hat{i}_2, \ldots, \hat{i}_r$ for some $r \geq 1$, such that $\mathbf{B}_S = [\ \mathbf{b}_{\hat{i}_1}, \mathbf{b}_{\hat{i}_2}, \ldots, \mathbf{b}_{\hat{i}_r}\ ]$ forms a sparse basis matrix for MMV. SOMP starts by determining the first index as

$$\hat{i}_1 = \arg \max_{i=1,\ldots,L} \|\mathbf{X}^T\mathbf{b}_i\|_q, \tag{3}$$

for some $q \geq 1$. The intuition behind (3) is that if $\mathbf{X}$ is mostly contributed by one atom, say, $\mathbf{b}_i$, then the value of $\|\mathbf{X}^T\mathbf{b}_i\|_q$ is likely to be the largest among the other $\|\mathbf{X}^T\mathbf{b}_j\|_q$, $j \neq i$. SOMP then follows a similar way to recursively determine $\hat{i}_k$, with an additional process known as successive nulling. To describe it, suppose that we have previously determined a number of $k-1$ atoms, indexed by $\hat{i}_1, \ldots, \hat{i}_{k-1}$, and the next task is to determine a new atom indexed by $\hat{i}_k$. Let $\mathbf{B}_S^{(k-1)} = [\ \mathbf{b}_{\hat{i}_1}, \ldots, \mathbf{b}_{\hat{i}_{k-1}}\ ]$, and consider the residual

$$\mathbf{R}^{(k-1)} = \mathbf{X} - \mathbf{B}_S^{(k-1)}\mathbf{C}^{(k-1)} \tag{4a}$$

$$\mathbf{C}^{(k-1)} = \arg \min_{\mathbf{C}\in\mathbb{R}^{(k-1)\times L}} \|\mathbf{X} - \mathbf{B}_S^{(k-1)}\mathbf{C}\|_2^2 \tag{4b}$$

$$= (\mathbf{B}_S^{(k-1)})^{\dagger}\mathbf{X} \tag{4c}$$

where nulling is applied to the measurement matrix $\mathbf{X}$ to remove signal components related to the previously determined atoms $\mathbf{b}_{\hat{i}_1}, \ldots, \mathbf{b}_{\hat{i}_{k-1}}$. We determine $\hat{i}_k$ by

$$\hat{i}_k = \arg \max_{i=1,\ldots,L} \|\mathbf{R}^{(k-1)^T}\mathbf{b}_i\|_q, \tag{5}$$

which is analogous to (3).

It has been shown that SOMP can perfectly solve the MMV problem (2) under certain sufficient conditions on the mutual coherence of $\mathbf{B}$; see [11].

2.2. Reduced MMV and Boost

Dealing with the MMV problem (2) can be computationally expensive when the number of measurements L is very large. ReMBo tackles this issue by reducing the MMV problem to an SMV problem, and using the subsequent SMV problem to determine a sparse basis matrix for MMV [14]. The idea is to merge the single-measurement vectors $\mathbf{x}_i$ to one. To be specific, let $\boldsymbol{\xi} \in \mathbb{R}^L$ be a randomly generated vector following some continuous distribution (e.g., Gaussian distribution). By considering the merged SMV

$$\bar{\mathbf{x}} = \mathbf{X}\boldsymbol{\xi}$$

and letting $\bar{\mathbf{c}} = \mathbf{X}\boldsymbol{\xi}$, we consider a reduced MMV problem

$$\begin{aligned} \min_{\bar{\mathbf{c}}\in\mathbb{R}^N} \ & \|\bar{\mathbf{c}}\|_0 \\ \text{s.t. } & \bar{\mathbf{x}} = \mathbf{B}\bar{\mathbf{c}}, \end{aligned} \tag{6}$$

where $\|\bar{\mathbf{c}}\|_0$ denotes the number of nonzero elements in $\bar{\mathbf{c}}$. Problem (6) is essentially an SMV problem, and we can determine a sparse basis matrix from Problem (6) by applying standard compressive sensing algorithms such as OMP (the SMV-reduced version of the above introduced SOMP) and ℓ_1-norm relaxation.

In ReMBo we may run the reduced MMV problem (6) multiple times for different randomly generated realizations of $\boldsymbol{\xi}$, and pick the best one. It was empirically shown that such process can boost the atom selection performance.

3. SELF-DICTIONARY MMV FOR BLIND HYPERSPECTRAL UNMIXING

Now, we turn our attention back to the context of blind hyperspectral unmixing. We consider a standard linear mixture model setting, described as follows. The measured hyperspectral pixel vectors are modeled by

$$\mathbf{x}[n] = \mathbf{A}\mathbf{s}[n], \quad n = 1, \ldots, L, \tag{7}$$

where $\mathbf{x}[n]$ denotes the measured hyperspectral pixel vector at pixel n, $\mathbf{A} = [\ \mathbf{a}_1, \ldots, \mathbf{a}_N\ ] \in \mathbb{R}^{M\times N}$ is the endmember signature matrix, in which each $\mathbf{a}_n \in \mathbb{R}^M$ corresponds to the hyperspectral signature vector of a distinct endmember (or material), $\mathbf{s}[n] \in \mathbb{R}^N$ is the abundance vector of the nth pixel, M is the number of spectral bands, N is the number of endmembers and L is the number of hyperspectral pixels. The endmember matrix $\mathbf{A}$ is assumed to have full column rank. Also, the abundance vectors $\mathbf{s}[n]$ satisfy the non-negative and sum-to-one constraints, i.e., $\mathbf{s}[n] \geq \mathbf{0}$ and $\mathbf{1}^T\mathbf{s}[n] = 1$ where

$\geq$ denotes the element-wise inequality, and $\mathbf{1}$ is an all-one vector of appropriate length.

Several mild assumptions are in order. First, to facilitate exposition of the underlying ideas of SD-MMV, we assume the noiseless noise. Note that the resulting algorithms will work in the presence of noise, at least by simulations, although noise sensitivity is not the interest of the present paper. Second, like many works in hyperspectral unmixing, the number of endmembers N is assumed to be known. Third, we assume that the *pure pixel assumption* holds. Specifically, the pure pixel assumption is said to be satisfied if, for each endmember index k, there exists a pixel index n_k such that

$$\mathbf{x}[n_k] = \mathbf{a}_k,$$

or equivalently, $\mathbf{s}[n_k] = \mathbf{e}_k$ where $\mathbf{e}_k$ is a unit vector (i.e., $[\mathbf{e}_k]_i = 1$ for $k = i$ and $[\mathbf{e}_k]_i = 0$ for $k \neq i$).

Our problem here is to identify the unknown endmember matrix $\mathbf{A}$ from the multiple hyperspectral measurement matrix $\mathbf{X} = [\,\mathbf{x}[1], \ldots, \mathbf{x}[L]\,]$. In particular, we are interested in the following criterion which is recently introduced in [9, 10]:

$$\min_{\mathbf{C}\in\mathbb{R}^{L\times L}} \quad \|\mathbf{C}\|_{\text{row}-0}, \tag{8a}$$

$$\text{s.t.} \quad \mathbf{X} = \mathbf{X}\mathbf{C}, \tag{8b}$$

$$\mathbf{C} \geq \mathbf{0},\ \mathbf{1}^T\mathbf{C} = \mathbf{1}^T. \tag{8c}$$

We call Problem (8) the *self-dictionary MMV* (SD-MMV) problem in this paper. The rationale of the SD-MMV problem is as follows: Since each $\mathbf{x}[n]$ can always be represented as a convex combination of $\{\mathbf{a}_1, \ldots, \mathbf{a}_N\}$, or equivalently, a distinct set of pure pixels $\{\mathbf{x}[n_1], \ldots, \mathbf{x}[n_N]\}$, we can see that there exists a $\mathbf{C}$ that satisfies (8b)-(8c) and has many redundant, or all-zero, rows. For example, if $n_1 = 1, n_2 = 2, \ldots, n_N = N$, then such $\mathbf{C}$ is given by

$$\mathbf{C} = \begin{bmatrix} \mathbf{S} \\ \mathbf{0} \end{bmatrix},$$

where $\mathbf{S} = [\,\mathbf{s}[1], \ldots, \mathbf{s}[L]\,]$, cf. (7). Hence, SD-MMV intends to identify the pure pixels by enforcing $\mathbf{C}$ to be as row-sparse as possible. Comparing Problems (8) and (2), we see that SD-MMV takes an MMV form. In particular, the dictionary employed by SD-MMV is the measurement matrix $\mathbf{X}$ itself. Hence, SD-MMV has a flavor of selecting a subset of the measurement vectors as the basis matrix for sparse representation of the whole measurement data.

In the previous works [9, 10], convex relaxation was employed to tackle the SD-MMV problem. It was shown that convex relaxation can lead to the optimal solution of Problem (8) if there are no repeated pure pixels and there is no noise. The greedy alternative has been not considered, however. Our interest in the subsequent subsections will be on the greedy approach.

3.1. SOMP for the SD-MMV Problem

The SD-MMV problem (8) is still not exactly the same as the MMV problem (2). Specifically, the former has additional constraints on $\mathbf{C}$, namely, (8c). In order to employ SOMP introduced in Sec. 2.1, we simply drop the constraints in (8c) and directly apply SOMP. The resulting algorithm, which we call SD-SOMP, is shown in Algorithm 1.

Algorithm 1: SD-SOMP

input : $\mathbf{X}$, and $q \geq 1$;

set $\mathbf{X}_S^{(0)} = \phi$, $\mathbf{R}^{(0)} = \mathbf{X}$ and $k = 1$.

for $k = 1 : N$ **do**

  determine an index of a pure pixel by

$$\hat{n}_k = \arg\max_{n=1,\ldots,L} \|(\mathbf{R}^{(k-1)})^T\mathbf{x}[n]\|_q; \tag{9}$$

  $\mathbf{X}_S^{(k)} = [\,\mathbf{X}_S^{(k-1)}, \mathbf{x}[\hat{n}_k]\,]$;

  $\mathbf{R}^{(k)} = \mathbf{X} - \mathbf{X}_S^{(k)}(\mathbf{X}_S^{(k)})^\dagger\mathbf{X}$;

end

output: $\hat{\mathbf{A}} = \mathbf{X}_S^{(N)}$ (pure pixels).

Now, an interesting question arises—how does SD-SOMP work? We should point out that in the previous work, the constraints in (8c) play a role in identifying the optimality of convex relaxation [9,10] for SD-MMV. Since SD-SOMP does not use (8c), would it affect its pure pixel identifiability? The answer turns out to be no.

Theorem 1 *In the noiseless case and under the pure pixel assumption, SD-SOMP correctly identifies the pure pixels of all the distinct endmembers for any $q > 1$.*

Before proceeding to the proof, we should note that the pure pixels identifiability claim in Theorem 1 is as good as that in convex relaxation. Also, Theorem 1 implies that SOMP perfectly solves the SD-MMV problem when the pure pixel assumption holds.

Proof of Theorem 1: Suppose that after $k-1$ iterations, SD-SOMP has identified $k-1$ distinct pure pixels indexed by $\hat{n}_1, \ldots, \hat{n}_{k-1}$, where $k \leq N-1$. Without loss of generality (w.l.o.g.), we assume $\mathbf{x}[\hat{n}_i] = \mathbf{a}_i$, $i = 1, \ldots, k-1$, and thus $\mathbf{X}_S^{(k-1)} = [\,\mathbf{a}_1, \ldots, \mathbf{a}_{k-1}\,]$. By mathematical induction, it suffices to show that SD-SOMP identifies a new pure pixel at the kth iteration; i.e., we need to show that $\|(\mathbf{R}^{(k-1)})^T\mathbf{x}[n]\|_q$ achieves its maximal value if and only if $\mathbf{x}[n]$ is a pure pixel and $\mathbf{x}[n] \neq \mathbf{a}_i$ for $i = 1, \ldots, k-1$. This can be shown by taking insights from the proof of Property 3 in [7].

To begin with, notice that $\mathbf{R}^{(k-1)} = \mathbf{P}^\perp_{\mathbf{A}_{1:k-1}}\mathbf{X}$, where the notation $\mathbf{P}^\perp_{\mathbf{A}}$ denotes the orthogonal complement projec-

tor of $\mathbf{A}$. Then, we have

$$\left\|(\mathbf{R}^{(k-1)})^T\mathbf{x}[n]\right\|_q = \left\|\sum_{i=k}^{N} s_i[n]\mathbf{X}^T\mathbf{P}^{\perp}_{\mathbf{A}_{1:k-1}}\mathbf{a}_i\right\|_q \quad (10a)$$

$$\leq \sum_{i=k}^{N} s_i[n]\left\|\mathbf{X}^T\mathbf{P}^{\perp}_{\mathbf{A}_{1:k-1}}\mathbf{a}_i\right\|_q \quad (10b)$$

$$\leq \max_{i=k,\ldots,N}\left\|\mathbf{X}^T\mathbf{P}^{\perp}_{\mathbf{A}_{1:k-1}}\mathbf{a}_i\right\|_q \quad (10c)$$

for $q \geq 1$, where (10b) is obtained by the triangle inequality. Assume w.l.o.g. that $\max_{i=k,\ldots,N}\|\mathbf{X}^T\mathbf{P}^{\perp}_{\mathbf{A}_{1:k-1}}\mathbf{a}_i\|_q \leq \|\mathbf{X}^T\mathbf{P}^{\perp}_{\mathbf{A}_{1:k-1}}\mathbf{a}_k\|_q$. We see that if $\mathbf{x}[n]$ is a pure pixel of endmember k, then equality in (10) holds and subsequently $\|(\mathbf{R}^{(k-1)})^T\mathbf{x}[n]\|_q$ achieves the maximum.

We also need to show that equality in (10) does not hold for any non-pure pixels. Let $1 < q < \infty$. By Minkowski's inequality [15], equality in (10b) may be achieved by a non-pure pixel if there exists an index j, $j \neq k$, such that

$$\mathbf{X}^T\mathbf{P}^{\perp}_{\mathbf{A}_{1:k-1}}\mathbf{a}_k = \mathbf{X}^T\mathbf{P}^{\perp}_{\mathbf{A}_{1:k-1}}\mathbf{a}_j. \quad (11)$$

However, this is impossible: it can be shown that $\mathbf{X}^T$ has full column rank if $\mathbf{A}$ has full column rank and the pure pixel assumption holds [16]. Thus, (11) holds only when $\mathbf{P}^{\perp}_{\mathbf{A}_{1:k-1}}\mathbf{a}_k = \mathbf{P}^{\perp}_{\mathbf{A}_{1:k-1}}\mathbf{a}_j$, which does not hold for a full column rank $\mathbf{A}$.

The necessity proof above does not cover the case of $q = \infty$. For $q = \infty$, the right-hand side of (9) can be expressed as

$$\begin{aligned}&\max_{n=1,\ldots,L}\|(\mathbf{R}^{(k-1)})^T\mathbf{x}[n]\|_\infty \\ &\quad= \max_{n=1,\ldots,L}\max_{m=1,\ldots,L}|\mathbf{x}^T[m]\mathbf{P}^{\perp}_{\mathbf{A}_{1:k-1}}\mathbf{x}[n]| \\ &\quad\leq \max_{n=1,\ldots,L}\|\mathbf{P}^{\perp}_{\mathbf{A}_{1:k-1}}\mathbf{x}[n]\|_2^2. \quad (12)\end{aligned}$$

where the inequality in (12) is due to Cauchy-Schwarz inequality. Moreover, it is observed that equality in (12) is achievable. Hence, we can simplify (9) to

$$\hat{n}_k = \arg\max_{n=1,\ldots,L}\|\mathbf{P}^{\perp}_{\mathbf{A}_{1:k-1}}\mathbf{x}[n]\|_2. \quad (13)$$

By the same proof as in (10), $\|\mathbf{P}^{\perp}_{\mathbf{A}_{1:k-1}}\mathbf{x}[n]\|_2$ is shown to be maximized only if $\mathbf{x}[n]$ is a new pure pixel. □

3.2. OMP for the Reduced MMV Problem

Let us consider the ReMBo method introduced in Sec. 2.2. By applying OMP to the reduced MMV problem in ReMBo, and employing only one random realization for simplicity, we obtain a reduced MMV and OMP (ReOMP) algorithm for the SD-MMV problem (8) as presented in Algorithm 2.

For convenience, we will call the above algorithm the *SD-ReOMP algorithm*. Like the previous SD-SOMP algorithm, SD-ReOMP does not incorporate the constraints in (8c). We show that this is not a problem from a viewpoint of pure pixel identifiability:

Algorithm 2: SD-ReOMP

input : $\mathbf{X}$;
randomly generate $\boldsymbol{\xi} \in \mathbb{R}^L$.
set $\mathbf{X}_S^{(0)} = \phi$, $\bar{\mathbf{x}} = \mathbf{X}\boldsymbol{\xi}$ and $\mathbf{r}^{(0)} = \bar{\mathbf{x}}$;
for $k = 1 : N$ **do**
  determine an index of a pure pixel by

$$\hat{n}_k = \arg\max_{n=1,\ldots,L}|(\mathbf{r}^{(k-1)})^T\mathbf{x}[n]|; \quad (14)$$

  $\mathbf{X}_S^{(k)} = [\,\mathbf{X}_S^{(k-1)}, \mathbf{x}[\hat{n}_k]\,]$;
  $\mathbf{r}^{(k)} = \bar{\mathbf{x}} - \mathbf{X}_S^{(k)}(\mathbf{X}_S^{(k)})^{\dagger}\bar{\mathbf{x}}$;
end
output: $\hat{\mathbf{A}} = \mathbf{X}_S^{(N)}$ (pure pixels).

Theorem 2 *Suppose that the random vector $\boldsymbol{\xi}$ in SD-ReOMP is generated following an absolutely continuous distribution. Then, in the noiseless case and under the pure pixel assumption, SD-ReOMP correctly identifies the pure pixels of all the distinct endmembers with probability one.*

Proof of Theorem 2: The proof is similar to that in Theorem 1. Assume w.l.o.g. that after $k-1$ iterations, we have $\mathbf{X}_S^{(k-1)} = [\,\mathbf{a}_1, \ldots, \mathbf{a}_{k-1}\,]$. Now our objective is to show that $|(\mathbf{r}^{(k-1)})^T\mathbf{x}[n]|$ attains its upper bound if and only if $\mathbf{x}[n]$ is a new pure pixel. To show this, re-express $(\mathbf{r}^{(k-1)})^T\mathbf{x}[n]$ by

$$(\mathbf{r}^{(k-1)})^T\mathbf{x}[n] = \sum_{i=k}^{N} s_i[n]\left(\boldsymbol{\xi}^T\mathbf{X}^T\mathbf{P}^{\perp}_{\mathbf{A}_{1:k-1}}\mathbf{a}_i\right).$$

Also, by the triangle inequality, we have

$$|(\mathbf{r}^{(k-1)})^T\mathbf{x}[n]| \leq \sum_{i=k}^{N} s_i[n]\left|\boldsymbol{\xi}^T\mathbf{X}^T\mathbf{P}^{\perp}_{\mathbf{A}_{1:k-1}}\mathbf{a}_i\right|,$$

where the equality holds if and only if $\mathbf{s}[n] = \mathbf{e}_i$, for $i \geq k$ (or, $\mathbf{x}[n]$ is a new pure pixel), given the premise that $\boldsymbol{\xi}^T\mathbf{X}^T\mathbf{P}^{\perp}_{\mathbf{A}_{1:k-1}}\mathbf{a}_i$ for $i = k, \ldots, N$ are non-zero distinct real numbers. We now show that the above premise holds with probability one. Indeed, since $\boldsymbol{\xi}$ is randomly generated following an absolutely continuous probability distribution, it lives in the null space of a given non-zero vector of dimension $1 \times L$ with probability zero [14]. Hence, we have

$$\begin{aligned}&\mathrm{Prob}\{\boldsymbol{\xi}^T\mathbf{X}^T\mathbf{P}^{\perp}_{\mathbf{A}_{1:k-1}}\mathbf{a}_i \neq 0\} = 1, \\ &\mathrm{Prob}\{\boldsymbol{\xi}^T\mathbf{X}^T\mathbf{P}^{\perp}_{\mathbf{A}_{1:k-1}}(\mathbf{a}_i - \mathbf{a}_j) \neq 0\} = 1,\end{aligned}$$

for $k \leq i, j \leq N$ and $i \neq j$, which completes the proof. □

It should be noted that SD-ReOMP guarantees perfect pure pixel identifiability in a probability one sense, rather than deterministically. As a further remark, the study above

does not employ the "boosting" part of ReMBo, that is, running SD-ReOMP multiple times and picking the best solution. While we already show that SD-ReOMP can perfectly identify all the distinct pure pixels, which means that repeating it may not be necessary, our empirical experience is that incorporating the boosting part can improve the pure pixel identification performance in the presence of noise.

4. CONNECTION TO EXISTING ALGORITHMS

Very interestingly, it turns out that the greedy SD-MMV algorithms shown in the previous section are equivalent to some representative pure pixels identification algorithms.

First of all, consider SD-SOMP in Algorithm 1. We are interested in the case of $q = \infty$. As we showed in the proof of Theorem 1, the main step in (9) can be simplified to

$$\hat{n}_k \quad = \arg \max_{n=1,\ldots,L} \|\mathbf{P}^{\perp}_{\mathbf{X}_S^{(k-1)}} \mathbf{x}[n]\|_2;$$

see (13). The resulting SD-SOMP algorithm is identical to SPA [4,8], which is also known to be very similar to ATGP [5] and SVMAX [7] if we neglect some minor algorithmic details. We should note that SPA follows an explicit pure pixel search approach, while SVMAX was derived from Winter's simplex volume maximization criterion. While their underlying principles are different, it is interesting to find that they are very similar [7]. Now, the present result further enriches this previous finding—SPA has three identities, namely, explicit pure pixel search, volume maximization, and SD-MMV.

Next, consider SD-ReOMP in Algorithm 2. Let $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\xi}$, and note that $\boldsymbol{\eta}$ is a random vector. The main step in (14) can be rewritten as

$$\hat{n}_k \quad = \arg \max_{n=1,\ldots,L} |\boldsymbol{\eta}^T \mathbf{P}^{\perp}_{\mathbf{X}_S^{(k-1)}} \mathbf{x}[n]|$$

The above step is very similar to that used in VCA [6], in which the principle is to find the vertices of the convex hull of the measured pixels, which are the pure pixels under the pure pixel assumption, through random projection on a plane.

In summary, we now understand that SPA, ATGP, SVMAX and VCA can be alternatively interpreted as greedy algorithms under the SD-MMV formulation.

5. CONCLUSION AND DISCUSSION

In this paper we studied blind hyperspectral unmixing under a self-dictionary MMV formulation. We considered two greedy algorithms for the formulation, and proved their pure pixel identifiability in the noiseless case. The resulting algorithms turn out to be identical or very similar to some representative pure pixels identification algorithms. Hence, we provide a new perspective on re-interpreting existing pure pixels identification algorithms. Future study should focus on further exploring the potential of SD-MMV, for example, on dealing with unknown number of endmembers.

6. REFERENCES

[1] J.M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 5, no. 2, pp. 354–379, 2012.

[2] J.W. Boardman, F.A. Kruse, and R.O. Green, "Mapping target signatures via partial unmixing of aviris data," in *Proc. Summ. JPL Airborne Earth Sci. Workshop, Pasadena, CA*, Dec 1995, vol. 1, pp. 23 –26.

[3] M.E. Winter, "N-findr: an algorithm for fast autonomous spectral endmember determination in hyperspectral data," in *Proc. SPIE's Inter. Sym. Opt. Sci. Eng. Inst.*, 1999, pp. 266–275.

[4] U.M.C. Araújo, B.T.C. Saldanha, R.K.H Galvão, T. Yoneyama, H.C. Chame, and V. Visani, "The successive projections algorithm for variable selection in spectroscopic multicomponent analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 57, no. 2, pp. 65–73, 2001.

[5] H. Ren and C.I. Chang, "Automatic spectral target recognition in hyperspectral imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1232–1249, 2003.

[6] J.M.P. Nascimento and J.M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, 2005.

[7] T.-H. Chan, W.-K. Ma, A. Ambikapathi, and C.-Y. Chi, "A simplex volume maximization framework for hyperspectral endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4177 –4193, Nov. 2011.

[8] N. Gillis and S.A. Vavasis, "Fast and robust recursive algorithms for separable nonnegative matrix factorization," *arXiv preprint arXiv:1208.1237*, 2012.

[9] E. Esser, M. Moller, S. Osher, G. Sapiro, and J. Xin, "A convex model for nonnegative matrix factorization and dimensionality reduction on physical space," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3239 –3252, Jul. 2012.

[10] E. Elhamifar, G. Sapiro, and R. Vidal, "See all by looking at a few: Sparse modeling for finding representative objects," in *Proc. IEEE CVPR 2012*, 2012, pp. 1600–1607.

[11] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634 –4643, Dec. 2006.

[12] J.A. Tropp, A.C. Gilbert, and M.J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, 2006.

[13] J.A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Process.*, vol. 86, no. 3, pp. 589–602, 2006.

[14] M. Mishali and Y.C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4692–4702, 2008.

[15] D.G. Luenberger, *Optimization by vector space methods*, Wiley-Interscience, 1997.

[16] W.-K. Ma, T.-H. Chan, C.-Y. Chi, and Y. Wang, "Convex analysis for non-negative blind source separation with application in imaging," in *Convex Optimization in Signal Processing and Communications*, D.P. Palomar and Y.C. Eldar, Eds. Cambridge University Press, 2010.