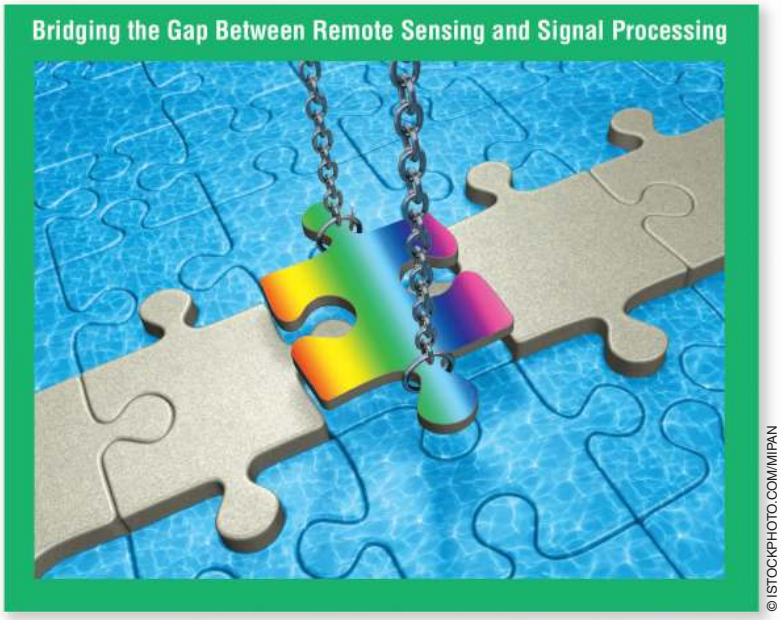
Wing-Kin Ma, José M. Bioucas-Dias, Tsung-Han Chan, Nicolas Gillis, Paul Gader, Antonio J. Plaza, ArulMurugan Ambikapathi, and Chong-Yung Chi

A Signal Processing Perspective on Hyperspectral Unmixing



Insights from remote sensing

Blind hyperspectral unmixing (HU), also known as unsupervised HU, is one of the most prominent research topics in signal processing (SP) for hyperspectral remote sensing [1], [2]. Blind HU aims at identifying materials present in a captured scene, as well as their compositions, by using high spectral resolution of hyperspectral images. It is a blind source separation (BSS) problem from a SP viewpoint. Research on this topic started in the 1990s in geoscience and remote sensing [3]–[7], enabled by technological advances in hyperspectral sensing at the time. In recent years, blind HU has attracted much interest from other fields such as SP, machine learning, and optimization, and the subsequent cross-disciplinary research activities have made blind HU a vibrant topic. The resulting impact is not just on remote sensing—blind HU has provided a unique problem scenario that inspired researchers from different fields to devise novel blind SP methods. In fact, one may say that blind HU has established a new branch of BSS approaches not seen in classical BSS studies. In particular, the convex geometry concepts—discovered by early remote sensing researchers through empirical observations [3]–[7] and refined by later research—are elegant and very different from statistical independence-based BSS approaches established in

Digital Object Identifier 10.1109/MSP.2013.2279731
Date of publication: 5 December 2013

the SP field. Moreover, the latest research on blind HU is rapidly adopting advanced techniques, such as those in sparse SP and optimization. The present development of blind HU seems to be converging to a point where the lines between remote sensing-originated ideas and advanced SP and optimization concepts are no longer clear, and insights from both sides would be used to establish better methods.

This article uses an SP researcher's perspective to review blind HU. We will consider several key developments, which include pure pixel search, convex geometry, dictionary-based sparse regression and nonnegative matrix factorization. We will not cover Bayesian techniques [8], although readers should note that they also represent key developments in blind HU. Our emphasis will be on insights, where we will showcase how each approach fundamentally works, and highlight significant results from a viewpoint of SP theory and methods. Some forefront advances will also be discussed. Note that this article does not aim at survey; please see a recent overview paper [2] that provides a comprehensive coverage of numerous blind HU methods and many other aspects.

Our notations are standard in SP. In addition, given a matrix X, x_i, and x^i denote its ith column and ith row, respectively; "$\geq$" represents elementwise inequality; $\mathbf{1}$ is an all-one vector of appropriate length; $A^\dagger$ is the pseudoinverse of A; $P_A^\perp = I - A(A^TA)^\dagger A^T$ is the orthogonal complement projector of A; $\sigma_{\min}(A)$ and $\sigma_{\max}(A)$ denote the minimum and maximum singular values of A, respectively; $\|\cdot\|_p$ denotes the ℓ_p norm; and $\|\cdot\|_F$ denotes the Frobenius norm.

SIGNAL MODEL

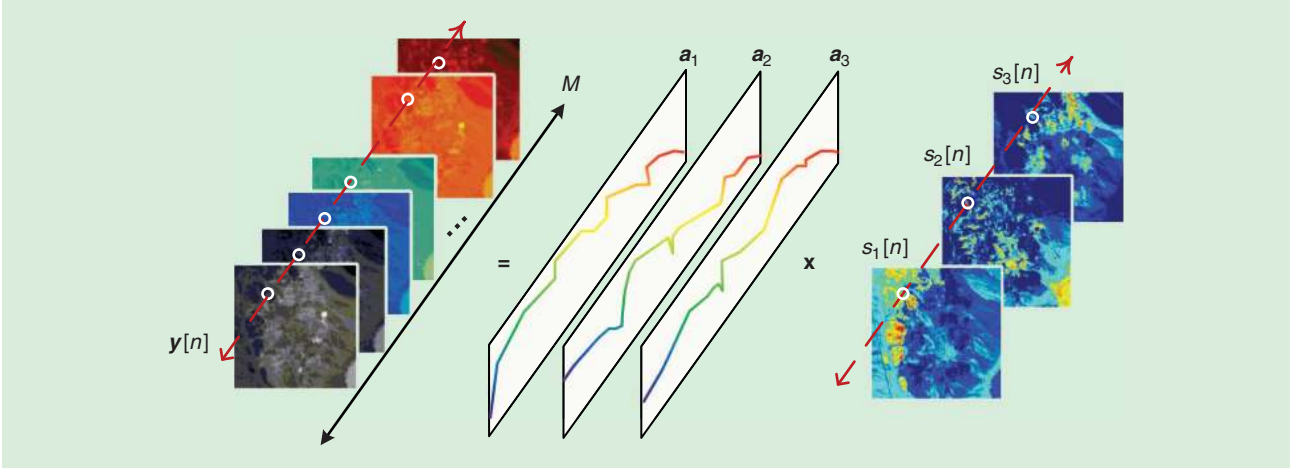
Modeling hyperspectral signals is a difficult problem. It depends on numerous factors; some crucial ones include: the types of materials encountered in the acquired scene, the ways the materials are physically mixed and constitute the scene topologically, the way light interacts with the materials, gets reflected and measured by the hyperspectral instrument, and the measurement environment. Over decades, the geoscience and remote sensing community has devoted tremendous efforts to various modeling aspects, from which we have now significantly improved our understanding of the true problem nature. Nevertheless, modeling can be an overwhelmingly complex process if one wants to treat every aspect very precisely. In particular, while radiative transfer theory (RTT) is well known to be able to provide accurate characterizations of photons' interactions with the materials (see [2] and the references therein), the resulting models are generally too difficult to use for signal analysis and processing. There is a compromise to make between model accuracy and tractability.

We focus on a relatively simplistic but very representative model, specifically, the linear mixing model (LMM). The LMM lies at the center of interest of many important developments in blind HU. Despite the fact that the LMM is not always true, especially under certain scenarios that exhibit strong nonlinearity, it is generally recognized as an acceptable model for many real-world scenarios. The LMM is described as follows. We assume a macroscopic mixing scale in which the incident light interacts with only one material before reflecting off. Let $y_m[n]$ denote the hyperspectral camera's measurement at spectral band m and at pixel n. Letting $y[n] = [y_1[n], y_2[n], \ldots, y_M[n]]^T \in \mathbb{R}^M$ where M is the number of spectral bands, the LMM is given by

$$y[n] = \sum_{i=1}^{N} a_i s_i[n] + \nu[n] = As[n] + \nu[n] \qquad (1)$$

for $n = 1, \ldots, L$, where each $a_i \in \mathbb{R}^M$, $i = 1, \ldots, N$, is called an *endmember signature vector*, which contains the spectral components of a specific material (indexed by i) in the scene; N is the number of endmembers, or materials, in the scene; $A = [a_1, \ldots, a_N] \in \mathbb{R}^{M \times N}$ is called the *endmember matrix*; $s_i[n]$ describes the contribution of material i at pixel n; $s[n] = [s_1[n], \ldots, s_N[n]] \in \mathbb{R}^N$ is called the *abundance vector* at pixel n; L is the number of pixels; and $\nu[n] \in \mathbb{R}^M$ is noise. Figure 1 illustrates the mixing process under the LMM.

There are several important aspects concerning the LMM formulation. First, since hyperspectral cameras have wide spectral ranges and fine spectral resolution, M is often large— typically more than 200. Such large spectral degrees of freedom allow us to distinguish an endmember signature from another, as well as



[FIG1] The linear mixing model.

mixtures of endmember signatures, provided that the materials are sufficiently different from one another. Hence, it is reasonable to assume that $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$ is linearly independent, and we will assume that this condition holds throughout the article. Second, the mixing process in (1) is a consequence of limited spatial resolution of hyperspectral cameras. Specifically, one pixel may not be spatially fine enough to contain one material only. For example, each pixel is about 4 m $\times$ 4 m to 20 m $\times$ 20 m for airborne visible/infrared imaging spectrometer, depending on the altitude of the flight. Third, while the noise vector $\boldsymbol{\nu}[n]$ is commonly used to represent background and instrument noise, one may also use it to incorporate errors arising from modeling inaccuracies. From such a perspective, (1) can serve as a reasonable approximate model when nonlinear effects are not too strong. Fourth, by nature, the abundance vectors $\boldsymbol{s}[n]$ should satisfy

$$s_i[n] \geq 0,\ i = 1, \ldots, N, \text{ and } \sum_{i=1}^{N} s_i[n] = 1, \tag{2}$$

for every $n = 1, 2, \ldots, L$. The second constraint above, commonly referred to as the *abundance sum constraint* or the *sum-to-one constraint*, means that abundances give the fractional proportions, or percentages, of the different materials in a pixel. For convenience, we will write

$$\boldsymbol{s}[n] \in \mathcal{S} = \{\, \boldsymbol{s} \in \mathbb{R}^N \mid \boldsymbol{s} \geq \boldsymbol{0}, \boldsymbol{s}^T\boldsymbol{1} = 1 \,\}, \tag{3}$$

where $\mathcal{S}$ denotes the feasible set of abundance vectors. Note that $\mathcal{S}$ is a unit simplex.

The LMM introduced above is considered standard. That said, there are some hidden complications. Here we briefly mention them; interested readers can find further clarifications in [2]. First, in the model (1), $\boldsymbol{y}[n]$s are actually processed measurements. Raw measurements from the hyperspectral camera usually undergo a series of processing steps, such as radiometric calibration, geometric correction, and atmospheric compensation [9], before arriving at the simple LMM. Second, for simplicity we have associated an endmember with a material, presumably pure. However, an endmember could also be a composition of several materials; i.e., a material made of several materials. The definition of an endmember can be subjective, and is dependent on applications. Third, we have assumed that the sum-to-one constraint in (3) holds. In practice, the sum-to-one constraint may be violated under the so-called endmember variability (EV) effects. Besides modeling issues, it is worth noting that recently there has been growing interest in considering specific but more treatable nonlinear mixture models for HU; the same applies to EV. In these scenarios, insights learned from LMM-based HU remain vital and provide building blocks for non-LMM HU problems there. We refer readers to [10] and [11] in this issue of *IEEE Signal Processing Magazine* for a coverage of nonlinear HU and EV, respectively.

PROBLEM STATEMENT

We are concerned with the HU problem, under the model setting in (1)–(3). Specifically, HU aims at recovering $\boldsymbol{s}[n]$ from $\boldsymbol{y}[n]$, thereby retrieving every material's abundance map $\{s_i[n]\}_{n=1}^L$ from the hyperspectral measurements. Assuming full knowledge of the endmember matrix $\boldsymbol{A}$, we can carry out unmixing by solving constrained linear least squares (LS) problems:

$$\hat{\boldsymbol{s}}[n] = \arg \min_{\boldsymbol{s}[n] \in \mathcal{S}} \| \boldsymbol{y}[n] - \boldsymbol{A}\boldsymbol{s}[n] \|_2^2, \tag{4}$$

for $n = 1, \ldots, L$. Fundamentally, the above problem is considered an "easy" problem—it is a convex optimization problem, and a simple way to obtain a solution is to call some general-purpose convex optimization software, such as the widely used CVX [12]. Alternatively, one can design dedicated algorithms for (4) to have more efficient implementations; this is a more popular option in the field [13]–[15]. What makes HU fundamentally challenging is not (4) (or other variants), but the fact that we often do not have full knowledge of $\boldsymbol{A}$—for if we do, it means that we know exactly all the materials in the scene, which is unlikely in reality.

Blind HU amounts to recovering $\{\boldsymbol{s}[n]\}_{n=1}^L$ from $\{\boldsymbol{y}[n]\}_{n=1}^L$ without knowledge of $\boldsymbol{A}$. The problem can also be stated as that of identifying $\boldsymbol{A}$ from $\{\boldsymbol{y}[n]\}_{n=1}^L$ without knowledge of $\{\boldsymbol{s}[n]\}_{n=1}^L$. At this point, readers who are familiar with BSS may have realized that the problem formulation of blind HU is the same as that of BSS: The endmember matrix $\boldsymbol{A}$ and abundance vectors $\boldsymbol{s}[n]$ are the mixing matrix and true source vectors in BSS, respectively. While this observation is true, and in fact has been noticed for a while [16], classical BSS methods established in the SP field usually do not fall in any of the mainstream blind HU approaches. The key reason is that under the unit simplex constraint (3), the sources $\{\boldsymbol{s}[n]\}_{n=1}^L$ do not satisfy the statistical independence assumption, which is a very essential assumption in many BSS methods, particularly the well-known independent component analysis (ICA). The violation of source independence makes many existing BSS methods an inappropriate choice for blind HU from the outset.

Before delving into blind HU, we should point out that we will generally assume N, the number of endmembers, to be known. As in BSS and sensor array processing in the SP field, where the same aspect has been extensively studied under the name of *model order selection* (see, e.g., [17]), the problem of identifying the number of endmembers can be seen as a separate problem; see [2, Sec. III] for a description. One may also build on an existing blind HU approach to provide joint blind HU and endmember number identification.

PURE PIXELS PURSUIT

Our review begins with a very simple class of methods that hinges on a special model assumption called *pure pixels*.

DEFINITION OF PURE PIXELS

We say that endmember i (or material i) has a pure pixel if for some index denoted by ℓ_i, we have

$$\boldsymbol{s}[\ell_i] = \boldsymbol{e}_i, \tag{5}$$

where $\boldsymbol{e}_i \in \mathbb{R}^N$ is a unit vector with the nonzero element at the ith entry (that is, $[\boldsymbol{e}_i]_j = 0$ for all $j \neq i$, and $[\boldsymbol{e}_i]_i = 1$). Moreover, we say that the pure pixel assumption holds if every endmember has a pure pixel.

Physically, the existence of pure pixels means that while hyperspectral pixels are generally mixtures of several materials, there are certain pixels that are constituted by one material only. This can be seen from the model (1). Assuming pure pixels and no noise, the observed vector at pixel ℓ_i is

$$\boldsymbol{y}[\ell_i] = \boldsymbol{a}_i, \tag{6}$$

for $i = 1, \ldots, N$, which are the endmembers. In practice, there are scenarios where the pure pixel assumption holds. For example, imagine a scene that consists of water and soil. If there exist some local pixel regions that contain either water or soil only, then those regions contain pure pixels. Note that since more than one pure pixel may exist for a particular endmember, ℓ_i may not be unique. However, we should also note that the pure pixel assumption does not always hold, e.g., in a scene consisting of highly mixed minerals, or if the spatial resolution of the hyperspectral camera is too low.

Pure pixels provide a unique opportunity for blind HU. In essence, if we know the pure pixel indices $\ell_1, \ldots, \ell_N$, then $[\boldsymbol{y}[\ell_1], \ldots, \boldsymbol{y}[\ell_N]] = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]$ is the endmember matrix itself—and the problem is solved—in the noiseless case. However, the pure pixel indices are not known a priori, and the problem is to find them.

SUCCESSIVE PROJECTIONS ALGORITHM

We introduce a simple algorithm for finding the pure pixels of all endmembers. The prerequisite required to understand the algorithm is just basic knowledge of linear algebra.

Again, consider the noiseless case and assume that the pure pixel assumption holds. We notice that for any n,

$$\|\boldsymbol{y}[n]\|_2 = \left\| \sum_{i=1}^{N} s_i[n]\boldsymbol{a}_i \right\|_2 \leq \sum_{i=1}^{N} \|s_i[n]\boldsymbol{a}_i\|_2 \tag{7a}$$

$$= \sum_{i=1}^{N} s_i[n]\|\boldsymbol{a}_i\|_2 \tag{7b}$$

$$\leq \max_{i=1,\ldots,N} \|\boldsymbol{a}_i\|_2, \tag{7c}$$

where (7a) is due to the LMM and the triangle inequality, and (7b) and (7c) to the unit simplex constraint (3). It can be seen that equality in (7) holds when $\boldsymbol{s}[n] = \boldsymbol{e}_j$, where $j = \arg\max_{i=1,\ldots,N} \|\boldsymbol{a}_i\|_2$; which holds at $n = \ell_j$, i.e., $\boldsymbol{y}[n]$ is a pure pixel corresponding to the jth endmember [cf. (6)]. Also, equality in (7) cannot be attained by nonpure pixels, by the equality condition of the triangle inequality and the linear independence of $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$. Assuming without loss of generality (w.l.o.g.) that $j = 1$, we can identify the first endmember signature by

$$\hat{\boldsymbol{a}}_1 = \boldsymbol{y}[\hat{\ell}_1], \quad \hat{\ell}_1 = \arg \max_{n=1,\ldots,L} \|\boldsymbol{y}[n]\|_2^2. \tag{8}$$

Note that $\hat{\boldsymbol{a}}_1$ is a perfect estimate of $\boldsymbol{a}_1$ under the aforementioned settings.

The next question is to identify pure pixels corresponding to other endmembers. Suppose that we have previously identified $k-1$ endmember signatures, denoted by $\hat{\boldsymbol{a}}_1, \ldots, \hat{\boldsymbol{a}}_{k-1}$, and that the identification is perfect, i.e., $\hat{\boldsymbol{a}}_i = \boldsymbol{a}_i$ for $i = 1, \ldots, k-1$. The idea to identify the next endmember is to perform nulling—a standard SP trick that has appeared many times (e.g., [17]), but proves very useful in various fields. Let $\hat{\boldsymbol{A}}_{1:k-1} = [\hat{\boldsymbol{a}}_1, \ldots, \hat{\boldsymbol{a}}_{k-1}]$, and construct its orthogonal complement projector $\boldsymbol{P}^{\perp}_{\hat{\boldsymbol{A}}_{1:k-1}}$. Since $\boldsymbol{P}^{\perp}_{\hat{\boldsymbol{A}}_{1:k-1}} \boldsymbol{a}_i = \boldsymbol{0}$ holds for any $i < k$, we have that

$$\|\boldsymbol{P}^{\perp}_{\hat{\boldsymbol{A}}_{1:k-1}} \boldsymbol{y}[n]\|_2 = \left\| \sum_{i=k}^{N} s_i[n] \boldsymbol{P}^{\perp}_{\hat{\boldsymbol{A}}_{1:k-1}} \boldsymbol{a}_i \right\|_2 \tag{9a}$$

$$\leq \max_{i=k,\ldots,N} \|\boldsymbol{P}^{\perp}_{\hat{\boldsymbol{A}}_{1:k-1}} \boldsymbol{a}_i\|_2, \tag{9b}$$

where (9b) is obtained in the same way as (7). And like (7), it can be shown that equality in (9) holds only for a pure pixel corresponding to a previously unidentified endmember, which we can assume w.l.o.g. to be that at $n = \ell_k$. The kth endmember signature can therefore be identified via

$$\hat{\boldsymbol{a}}_k = \boldsymbol{y}[\hat{\ell}_k], \quad \hat{\ell}_k = \arg \max_{n=1,\ldots,L} \|\boldsymbol{P}^{\perp}_{\hat{\boldsymbol{A}}_{1:k-1}} \boldsymbol{y}[n]\|_2^2. \tag{10}$$

Hence, by induction, we can identify all the endmembers.

The algorithm presented on the next page is called the *successive projections algorithm* (*SPA*). Algorithm 1 gives the pseudocode of SPA, which is very simple. From the above algebraic development, we conclude that in the noiseless case and under the pure pixel assumption, SPA perfectly identifies all the endmember signatures $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$.

We should provide a brief historical note on SPA, since it has been repeatedly rediscovered. To our best knowledge, SPA first appeared in chemometrics in 2001 by Araújo et al. [18]. Later, a very similar algorithm, called the *automatic target generation process* (*ATGP*), was proposed by Ren and Chang in 2003 in remote sensing [19]. Curiously, the development we just displayed, which shows why SPA works from an algebraic SP viewpoint and pins down its endmember identifiability, was not seen until recently; see [20, Appendix F]. There are other ways to derive SPA, which will be described later. It is worth pointing out that SPA has been used successfully for rather different purposes. In numerical linear algebra, SPA is closely related to the so-called modified Gram–Schmidt algorithm with column pivoting, used for example to solve linear LS problems [21]. In machine learning, SPA has been used for document classification where the pure pixel assumption is referred to as the *separability* assumption and requires that, for each topic, there exists at least one word used only by that topic; see [22] and the references therein.

The above SPA development is based on the noiseless argument. An interesting question is therefore on sensitivity against noise. A provable performance bound characterizing noise sensitivity has been proposed very recently in [23], and is briefly described here. Let us denote $\sigma = \sigma_{\min}(\boldsymbol{A})$, which is positive since $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$ is linearly independent, and $K = \max_{1 \leq i \leq N} ||\boldsymbol{a}_i||_2$. Let us also denote the noise level $\epsilon = \max_{1 \leq n \leq L} ||\boldsymbol{v}[n]||_2$. Then, under the pure pixel assumption and assuming that the noise

Algorithm 1 SPA.

input $\{\boldsymbol{y}[n]\}_{n=1}^{L}$, N.

1: $\boldsymbol{P}^{\perp} = \boldsymbol{I}$

2: **for** $k = 1, \ldots, N$ **do**

3: $\hat{\ell}_k = \arg\max_{n=1,\ldots,L} \| \boldsymbol{P}^{\perp}\boldsymbol{y}[n] \|_2^2$

4: $\hat{\boldsymbol{a}}_k = \boldsymbol{y}[\hat{\ell}_k]$

5: $\boldsymbol{P}^{\perp} := (\boldsymbol{I} - (\boldsymbol{P}^{\perp}\hat{\boldsymbol{a}}_k)(\boldsymbol{P}^{\perp}\hat{\boldsymbol{a}}_k)^T / \| \boldsymbol{P}^{\perp}\hat{\boldsymbol{a}}_k \|_2^2)\boldsymbol{P}^{\perp}$

6: **end for**

output $\hat{\boldsymbol{A}} = [\hat{\boldsymbol{a}}_1, \ldots, \hat{\boldsymbol{a}}_N]$.

level satisfies $\epsilon \leq O(\sigma^3/(NK^2))$, SPA identifies all the endmember signatures $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$ up to error $O(\epsilon(K^2/\sigma^2))$; more precisely, we have

$$\max_{1\leq i\leq N} \min_{1\leq j\leq N} \| \boldsymbol{a}_i - \hat{\boldsymbol{a}}_j \|_2 \leq O\left(\epsilon \frac{K^2}{\sigma^2}\right). \tag{11}$$

The above analysis result provides significant practical implications. We see in (11) that the noise robustness of SPA depends on the ratio K/σ. It can be shown that [23]

$$\frac{\max_{1\leq i\leq N} \| \boldsymbol{a}_i \|_2}{\min_{1\leq i\leq N} \| \boldsymbol{a}_i \|_2} \leq \frac{K}{\sigma} \leq \frac{\sigma_{\max}(\boldsymbol{A})}{\sigma_{\min}(\boldsymbol{A})}.$$

Thus, the noise robustness of SPA depends on 1) how different the magnitudes of the endmember signatures are and 2) how well the true endmember signatures are spectrally distributed. In particular, the latter implies that challenging scenarios lie in highly similar endmembers.

Let us further point out two notable facts. First, one can generalize SPA by replacing the ℓ_2 norm in (9)–(10) by any continuously differentiable and locally strongly convex function whose minimizer is zero, e.g., any ℓ_p norm with $1 < p < +\infty$. The corresponding algorithm not only works in the noiseless case, it is also shown to possess a similar error bound as in (11) [23]. According to the analysis, the variant using the ℓ_2 norm has the best robustness against noise among all locally strongly convex functions; see also [24] for numerical evidence. Second, it is possible to improve the error bound above to $O(\epsilon(K/\sigma))$ by using the following postprocessing strategy [22]: Let $\{\hat{\boldsymbol{a}}_1, \ldots, \hat{\boldsymbol{a}}_N\}$ be the N endmembers extracted by SPA. Then, for $i = 1, \ldots, N$,

1) Project the original data $\{\boldsymbol{y}[n]\}_{n=1}^{L}$ onto the orthogonal complement of $\{\hat{\boldsymbol{a}}_k\}_{k=1,k\neq i}^{N}$.

2) Replace $\hat{\boldsymbol{a}}_i$ with the column of $\{\boldsymbol{y}[n]\}_{n=1}^{L}$ whose ℓ_2 norm of the projection is maximum.

This iterative refinement strategy is identical to a previously proposed blind HU algorithm (but without a robustness analysis); it will be further discussed in the section "Simplex Volume Maximization."

OTHER ALGORITHMS AND DISCUSSION

There are many other pure pixels search algorithms; see [2, Sec. VI.A] for a review. A representative algorithm in this family is vertex component analysis (VCA), proposed in 2003 [25], [26]. VCA is similar to SPA—it also employs successive nulling, but differs in the way it picks pure pixels. Specifically, in VCA, the right-hand sides (RHSs) of (8) and (10) are replaced by

$$\hat{\ell}_k = \max_{n=1,\ldots,L} | \boldsymbol{w}_k^T \boldsymbol{y}[n] |, \tag{12}$$

for $k = 1, \ldots, N$, where $\boldsymbol{w}_k$ is a randomly generated vector lying on the orthogonal complement subspace of $\hat{\boldsymbol{A}}_{1:k-1}$. Specifically, it is given by $\boldsymbol{w}_k = \boldsymbol{P}_{\hat{\boldsymbol{A}}_{1:k-1}}^{\perp}\boldsymbol{\xi} / \| \boldsymbol{P}_{\hat{\boldsymbol{A}}_{1:k-1}}^{\perp}\boldsymbol{\xi} \|_2$, where $\boldsymbol{\xi}$ is an independent and identically distributed (i.i.d.) zero-mean Gaussian vector. Following the same derivations described above for SPA, one can show that VCA also perfectly identifies all the endmember signatures in the noiseless case and under the pure pixel assumption; this result holds with probability one. Also, we must mention the pixel purity index (PPI) by Boardman et al. in 1995 [6], which is one of the earliest blind HU algorithms. PPI does not have successive nulling. It is analogous to running (12) only for $k = 1$, but for many independent random trials. The number of trials needs to be large enough so to increase the chance of successfully hitting all endmembers' pure pixels. For numerical comparisons of SPA, VCA, and PPI, please see [23] (also [20]).

Some additional comments are in order.

1) To simplify the presentation, we have intentionally skipped a conventional preprocessing procedure, specifically, dimension reduction (DR). In practice, VCA and PPI would apply DR to the observed data $\{\boldsymbol{y}[n]\}_{n=1}^{L}$, prior to pure pixels search. While we have seen that DR is not required in SPA (as well as VCA and PPI), applying DR plays a crucial role in suppressing noise, which in turn helps improve pure pixel identification performance. Readers are referred to [2, Section III] for the state-of-the-art DR methods in HU.

2) SPA can be extended in at least two ways. First, it can be modified to accommodate outliers, which are anomalous pixels that exhibit markedly different behaviors from the nominal model and can cause substantial performance degradation. The idea is to consider outliers as endmembers, identify them together with true endmembers, and discard them from the obtained estimates [23]. Second, one can extend the method for joint blind HU and endmember number identification. We note that if we keep running the SPA step in (10) recursively, then, at stage $k = N + 1$, the projection residuals $\| \boldsymbol{P}_{\hat{\boldsymbol{A}}_{1:k-1}}^{\perp}\boldsymbol{y}[n] \|_2^2$ become zeros. Thus, the projection residuals may serve as an indicator of the number of endmembers. Similar ideas have been considered in [24] and [27].

CONVEX GEOMETRY

We have previously shown how blind HU may be easily handled under the pure pixel assumption. The pure pixel concept actually came from the study of convex geometry (CG) of hyperspectral signals, where remote sensing researchers examined the special geometric structure of hyperspectral signals and looked for automatic methods for endmember determination, i.e., blind HU. In fact, a vast majority of blind HU developments, if

WHO DISCOVERED CONVEX GEOMETRY FOR BLIND UNMIXING?

In geoscience and remote sensing, the work by Craig in the early 1990s [3], [4] is widely recognized to be most seminal in introducing the notion of CG for hyperspectral signal analysis and unmixing. Craig's original work not only described simplex volume minimization, which turns out to become a key CG concept for blind HU, it also inspired other pioneers, such as Boardman who made notable early contributions to CG-based blind HU [5] and introduced pure pixel search [6], and Winter, who proposed the simplex volume maximization concept [7] that results in the popularized N-FINDR algorithm class. What is remarkable in these early studies is that they discovered such beautiful blind SP concepts through sharp empirical observations and strong intuitions, rather than through rigorous SP or mathematics.

CG is also an idea that has been discovered several times in different areas. The introduction of CG can be traced back to as early as 1964 by Imbrie [28]. Imbrie's work belongs to another branch of geoscience studies wherein CG is used for the analysis of compositional data in earth science, such as mineral assemblages, grain-size distribution data, and geochemical and petrological data; see [29] for an overview. In fact, Imbrie's Q-mode analysis and the subsequent QMODEL by Klovan and Miesch [30] are conceptually identical to vertex or pure pixel search, although the methodology is different. Likewise, Full et al. already considered the same simplex volume minimization principle as Craig's in the 1980s [31]. CG has also been independently discovered in other fields such as chemometrics [32] and SP [33], [34]. In all the discoveries or rediscoveries mentioned above, the driving force that led researchers on different backgrounds to devise the same idea seems to be with the geometric elegance of CG and its powerful implications on solving blind unmixing problems.

not all, are directly or intuitively related to concepts introduced in early CG studies, such as simplex volume minimization by Craig [4], simplex volume maximization by Winter [7], and the previously reviewed pure pixel search by Boardman et al. [6]. We give a historical review in the "Who Discovered Convex Geometry for Blind Unmixing?"

PRELIMINARIES

We introduce several mathematical notations and facts in convex analysis, whose physical relevance to blind HU will become clear soon. The affine hull of a set of vectors $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\} \subset \mathbb{R}^M$ is defined as

$$\text{aff}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\} = \left\{ \boldsymbol{y} = \sum_{i=1}^{N} \theta_i \boldsymbol{a}_i \;\middle|\; \boldsymbol{\theta} \in \mathbb{R}^N, \sum_{i=1}^{N} \theta_i = 1 \right\}. \quad (13)$$

An affine hull can always be represented by

$$\text{aff}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\} = \{\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{d} \mid \boldsymbol{x} \in \mathbb{R}^P\} \quad (14)$$

for some $\boldsymbol{C} \in \mathbb{R}^{M \times P}$, $\boldsymbol{d} \in \mathbb{R}^M$, where $\text{rank}(\boldsymbol{C}) = P$ and $P \le N-1$ is the affine dimension of the affine hull. The affine dimension is $P = N-1$ if $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$ is affinely independent.

The convex hull of a set of vectors $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\} \subset \mathbb{R}^M$ is defined as

$$\text{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\} = \left\{ \boldsymbol{y} = \sum_{i=1}^{N} \theta_i \boldsymbol{a}_i \;\middle|\; \boldsymbol{\theta} \ge 0, \sum_{i=1}^{N} \theta_i = 1 \right\}. \quad (15)$$

The set $\text{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$ is called an $(N-1)$-simplex, or simply a simplex, if $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$ is affinely independent. The vertices of a simplex are $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$. Given a full-dimensional simplex, i.e., an $(N-1)$-simplex lying in $\mathbb{R}^{N-1}$ (or $M = N-1$), its volume can be determined by

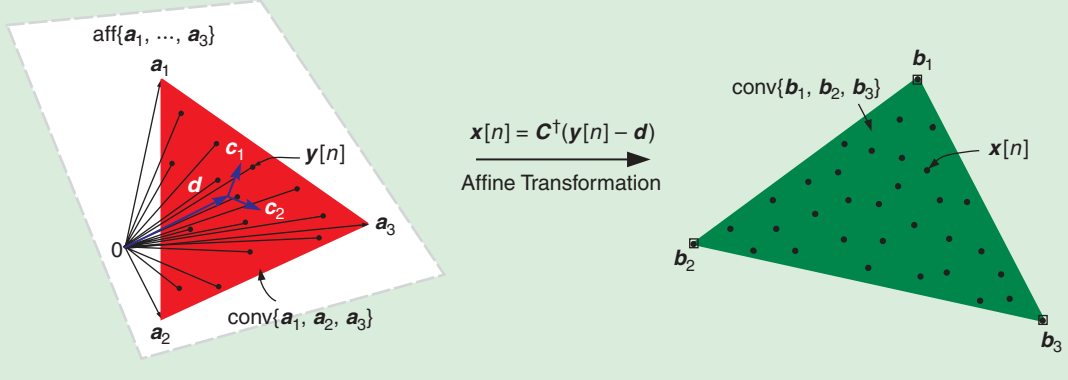
$$\text{vol}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N) = c \left| \det\left(\begin{bmatrix} \boldsymbol{a}_1 & \ldots & \boldsymbol{a}_N \\ 1 & \ldots & 1 \end{bmatrix} \right) \right| \quad (16a)$$

$$= c \left| \det\left([\boldsymbol{a}_1 - \boldsymbol{a}_N, \ldots, \boldsymbol{a}_{N-1} - \boldsymbol{a}_N] \right) \right|, \quad (16b)$$

where $c = 1/(N-1)!$ For the mathematical details of the above concepts, readers are referred to the literature [35].

CONVEX GEOMETRY IN HYPERSPECTRAL SIGNALS

There is a strong connection between convex analysis and hyperspectral signals. To see it, consider the signal model (1)–(3) in the noiseless case. By comparing the model and the definition of convex hull in (15), we observe that



[FIG2] The convex geometry of hyperspectral signals.

$$\boldsymbol{y}[n] \in \text{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}, \quad \text{for all } n = 1, \ldots, L,$$

i.e., each measured hyperspectral pixel $\boldsymbol{y}[n]$ is a convex combination of the endmember signatures $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$. Also, the set $\text{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$ is a simplex, since $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$ is linearly independent (and thus affinely independent). The left-hand side of Figure 2 gives a vector space illustration for the case of $N = 3$. As can be seen, $\text{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$ is a triangle; note that $\text{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$ is a tetrahedron for $N = 4$, and so forth. Also, every $\boldsymbol{y}[n]$ is enclosed by the triangle, and the corners of the triangle, or more formally, the vertices of $\text{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$, are the true endmember signatures $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$. This observation is simple, but gives a very powerful implication—if we can find all the vertices of $\text{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$ from the observation $\{\boldsymbol{y}[n]\}_{n=1}^L$, then blind HU is solved.

Intuitively speaking, CG-based blind HU amounts to finding a set of vectors, say, $\{\hat{\boldsymbol{a}}_1, \ldots, \hat{\boldsymbol{a}}_N\}$, such that the corresponding simplex $\text{conv}\{\hat{\boldsymbol{a}}_1, \ldots, \hat{\boldsymbol{a}}_N\}$ gives a best fitting to the true endmembers' simplex $\text{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$. The previously reviewed pure pixel search algorithms are among one class of such CG solutions; the idea is that pure pixels, if they exist, are also vertices of $\text{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$. Hence, pure pixel search is also vertex search in CG, under the pure pixel assumption. Now, we are interested in a different approach where simplex volume is used as the metric to find the best-fitting simplex. Moreover, the pure pixel assumption will not be assumed during the development. We should nevertheless mention a subtle point that the pure pixel assumption will come back when we discuss endmember identifiability.

Before proceeding to the main developments, it is essential for us to introduce a concept related to the affine nature of $\boldsymbol{y}[n]$. Since $\boldsymbol{y}[n] \in \text{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$, it also holds true that $\boldsymbol{y}[n] \in \text{aff}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$; cf. (13). By the equivalent affine hull representation in (14), we can write

$$\boldsymbol{y}[n] = \boldsymbol{C}\boldsymbol{x}[n] + \boldsymbol{d}, \tag{17}$$

for some $\boldsymbol{C} \in \mathbb{R}^{M \times (N-1)}$, $\text{rank}(\boldsymbol{C}) = N - 1$, $\boldsymbol{d} \in \mathbb{R}^M$, $\boldsymbol{x}[n] \in \mathbb{R}^{N-1}$, $n = 1, \ldots, L$. Suppose that $(\boldsymbol{C}, \boldsymbol{d})$ is known, and consider the inverse of (17) with respect to (w.r.t.) $\boldsymbol{x}[n]$

$$\boldsymbol{x}[n] = \boldsymbol{C}^{\dagger}(\boldsymbol{y}[n] - \boldsymbol{d}). \tag{18}$$

From the signal model (1)–(3), it is easy to show that

$$\boldsymbol{x}[n] = \sum_{i=1}^{N} \boldsymbol{b}_i s_i[n] = \boldsymbol{B}\boldsymbol{s}[n], \tag{19}$$

where $\boldsymbol{b}_i = \boldsymbol{C}^{\dagger}(\boldsymbol{a}_i - \boldsymbol{d}) \in \mathbb{R}^{N-1}$, $i = 1, \ldots, N$, and $\boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N] \in \mathbb{R}^{(N-1) \times N}$. We see that (19) takes exactly the same form as the original model (1), but its vector dimension is $N - 1$, which is less than M. Also, $\text{conv}\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N\}$ is a full-dimensional simplex [36]. Therefore, (19) is a dimension-reduced equivalent model for hyperspectral signals, where the CG structure is preserved. We will employ the equivalent model (19) in our subsequent CG developments. The transformation for the equivalent model is illustrated in Figure 2.

We should discuss how the affine set variable $(\boldsymbol{C}, \boldsymbol{d})$ is obtained in practice. Since there is no prior knowledge on $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$, we must estimate $(\boldsymbol{C}, \boldsymbol{d})$ from the observation $\{\boldsymbol{y}[n]\}_{n=1}^L$. This can be done by solving an affine set fitting (ASF) problem

$$\min_{\substack{\boldsymbol{C}, \boldsymbol{d}, \{\boldsymbol{x}[n]\}_{n=1}^L \\ \text{rank}(\boldsymbol{C}) = N-1}} \sum_{n=1}^{L} \| \boldsymbol{y}[n] - \boldsymbol{C}\boldsymbol{x}[n] - \boldsymbol{d} \|_2^2, \tag{20}$$

where the rationale is to find an affine set that gives the best fitting w.r.t. the measured pixels $\boldsymbol{y}[n]$, given knowledge of N; see [36] for details. The ASF solution is as follows. Let $\boldsymbol{\mu}_y = 1/L \sum_{n=1}^L \boldsymbol{y}[n]$ and $\boldsymbol{\Phi}_y = 1/L \sum_{n=1}^L (\boldsymbol{y}[n] - \boldsymbol{\mu}_y)(\boldsymbol{y}[n] - \boldsymbol{\mu}_y)^T$ be the sample mean and sample covariance of $\boldsymbol{y}[n]$, respectively. Also, let $\boldsymbol{q}_i$ be the ith principal eigenvector of $\boldsymbol{\Phi}_y$. The solution to (20) is given by $\boldsymbol{C} = [\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{N-1}]$, $\boldsymbol{d} = \boldsymbol{\mu}_y$. There is an interesting coincidence here—the ASF solution is exactly the same as that of principal component analysis (PCA), which is a commonly used DR preprocessing procedure. While ASF and PCA turn out to be equivalent, one should note that they were derived from different principles: ASF is deterministic and concerned with CG-preserving transformation, while PCA is statistical and does not exploit CG.

SIMPLEX VOLUME MAXIMIZATION

This subsection focuses on the simplex volume maximization approach. This approach considers the following problem:

$$\begin{aligned} &\max_{\boldsymbol{B}} \ \text{vol}(\boldsymbol{B}) \\ &\text{s.t. } \boldsymbol{b}_i \in \text{conv}\{\boldsymbol{x}[1], \ldots, \boldsymbol{x}[L]\}, \ i = 1, \ldots, N. \end{aligned} \tag{21}$$

We will call (21) VolMax for convenience. A picture is illustrated in Figure 3(a) to help us explain the aim of (21). We intend to find a best-fitting simplex, $\text{conv}\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N\}$, by maximizing its volume while keeping it inside $\text{conv}\{\boldsymbol{x}[1], \ldots, \boldsymbol{x}[L]\}$. One can imagine that if the pure pixel assumption holds, then $\text{conv}\{\boldsymbol{x}[1], \ldots, \boldsymbol{x}[L]\}$ is also the true endmembers' simplex and the maximum volume simplex should perfectly match the latter—this is Winter's intuition when he first introduced VolMax [7].

We are interested in simple optimization schemes for processing VolMax. Two such schemes are described as follows.

SUCCESSIVE VOLUME MAXIMIZATION

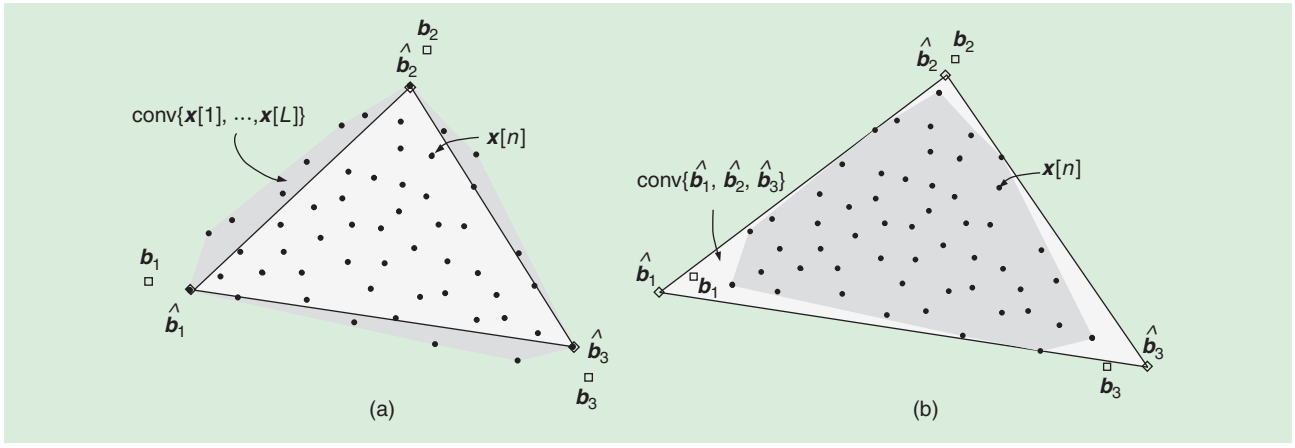
To facilitate our description, let

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{b}_1 & \ldots & \boldsymbol{b}_N \\ 1 & \ldots & 1 \end{bmatrix}, \ \boldsymbol{f}_i = \begin{bmatrix} \boldsymbol{b}_i \\ 1 \end{bmatrix}, \ \bar{\boldsymbol{x}}[n] = \begin{bmatrix} \boldsymbol{x}[n] \\ 1 \end{bmatrix}.$$

It can be shown that [20], [37]

$$|\det(\boldsymbol{F})|^2 = \prod_{k=1}^{N} \left\| \boldsymbol{P}^{\perp}_{\boldsymbol{F}_{1:(k-1)}} \boldsymbol{f}_k \right\|_2^2, \tag{22}$$

where $\boldsymbol{F}_{1:i} \in \mathbb{R}^{N \times i}$ denotes a submatrix of $\boldsymbol{F}$, obtained by picking the first i columns of $\boldsymbol{F}$. We see from the simplex volume formula in (16a) that maximizing $\text{vol}(\boldsymbol{B})$ is the same as maximizing (22). In successive volume maximization (SVMAX) [20] (also [37]), the principle is to exploit the successive structure of (22) to recursively generate an approximate solution to (21). Specifically, we carry out the following heuristic: for $k = 1, \ldots, N$, determine an estimate



[FIG3] (a) Simplex volume maximization. (b) Simplex volume minimization.

$$\hat{\boldsymbol{b}}_k = \arg\max_{\boldsymbol{b}_k} \| \boldsymbol{P}^{\perp}_{\hat{\boldsymbol{F}}_{1:(k-1)}} \boldsymbol{f}_k \|_2^2 \quad \text{s.t. } \boldsymbol{b}_k \in \text{conv}\{\boldsymbol{x}[1], \ldots, \boldsymbol{x}[L]\}, \quad (23)$$

where $\hat{\boldsymbol{F}}_{1:(k-1)}$ is defined in the same way as $\boldsymbol{F}_{1:(k-1)}$, with $\boldsymbol{b}_i$ replaced by $\hat{\boldsymbol{b}}_i$ for all i. Essentially, we estimate one endmember $\hat{\boldsymbol{b}}_k$ based on the previous endmember estimates $\hat{\boldsymbol{b}}_1, \ldots, \hat{\boldsymbol{b}}_{k-1}$ and partial maximization of (22). Let us complete the SVMAX algorithm by giving the solution to (23)

$$\hat{\boldsymbol{b}}_k = \boldsymbol{x}[\hat{\ell}_k], \quad \hat{\ell}_k = \arg\max_{n=1,\ldots,L} \| \boldsymbol{P}^{\perp}_{\hat{\boldsymbol{F}}_{1:(k-1)}} \bar{\boldsymbol{x}}[n] \|_2^2; \quad (24)$$

see [20]. Intriguingly, we have seen this algorithm before—SPA in the previous section. To explain, first note that $\bar{\boldsymbol{x}}[n]$ can be expressed as $\bar{\boldsymbol{x}}[n] = \boldsymbol{F}\boldsymbol{s}[n]$, an LMM form. If we apply SPA to $\{\bar{\boldsymbol{x}}[n]\}_{n=1}^L$ to retrieve $\boldsymbol{F}$, then the resulting SPA is exactly the same as SVMAX. Hence, we conclude that SVMAX is also a pure pixel search algorithm, and SPA has a "dual" identity in VolMax.

SUCCESSIVE N-FINDR

We consider an alternative scheme based on alternating optimization (AO). The idea is to optimize (21) w.r.t. one $\boldsymbol{b}_i$ at a time, while fixing other variables $\{\boldsymbol{b}_j\}_{j \neq i}$. To be specific, given a starting point $\hat{\boldsymbol{B}} = [\hat{\boldsymbol{b}}_1, \ldots, \hat{\boldsymbol{b}}_N]$, we update each $\hat{\boldsymbol{b}}_k$ via

$$\hat{\boldsymbol{b}}_k := \arg\max_{\boldsymbol{b}_k} \text{vol}([\hat{\boldsymbol{B}}_{-k}, \boldsymbol{b}_k]) \quad \text{s.t. } \boldsymbol{b}_k \in \text{conv}\{\boldsymbol{x}[1], \ldots, \boldsymbol{x}[L]\} \quad (25)$$

for $k = 1, \ldots, N$, where $\hat{\boldsymbol{B}}_{-k}$ denotes a submatrix of $\hat{\boldsymbol{B}}$ in which the kth column is removed. Also, we repeat the AO cycle in (25) until some stopping rule (e.g., almost no volume increase) is satisfied. The updates in (25) have a closed form

$$\hat{\boldsymbol{b}}_k = \boldsymbol{x}[\hat{\ell}_k], \quad \hat{\ell}_k = \arg\max_{n=1,\ldots,L} \| \boldsymbol{P}^{\perp}_{\hat{\boldsymbol{F}}_{-k}} \bar{\boldsymbol{x}}[n] \|_2^2, \quad (26)$$

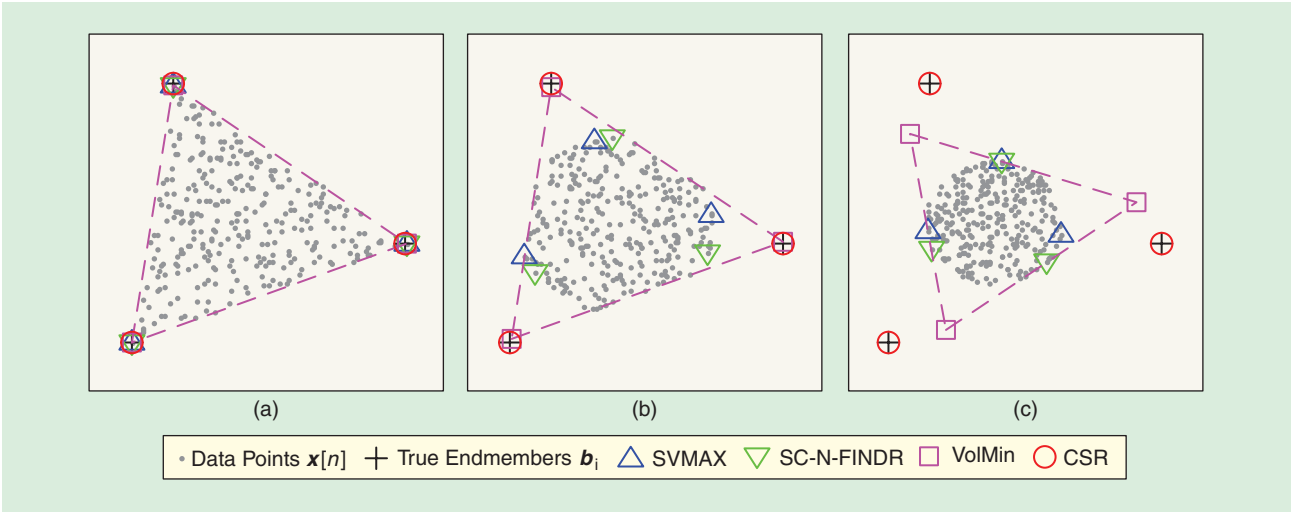
where (26) is obtained by using (22) to turn (25) to (23) (with a proper index reordering), and then applying (24). We call the resulting algorithm successive N-FINDR (SC-N-FINDR) since it is very similar to the SC-N-FINDR proposed in [38]. The pseudocode of SC-N-FINDR is given in Algorithm 2. Note that for initialization, we can use another algorithm, e.g., SVMAX, or do so randomly. There are several interesting connections here. First, SC-N-FINDR performs pure pixel search. Following [20, Prop. 1], it can be shown that in the noiseless case and under the pure pixel assumption, SC-N-FINDR may perfectly identify all the endmembers' pure pixels within one AO cycle. Second, since $\bar{\boldsymbol{x}}[n] = \boldsymbol{F}\boldsymbol{s}[n]$, we see from (26) that SC-N-FINDR is performing nulling—this time for all other endmember estimates $\hat{\boldsymbol{F}}_{-k}$; cf. the nulling in SPA in (9a). Thus, SC-N-FINDR is also a nulling-based algorithm. Third, we notice that each AO cycle in SC-N-FINDR is essentially the same as the SPA postprocessing strategy we briefly discussed in the section "Successive Projections Algorithm," which is provably robust against noise.

Algorithm 2 SC-N-FINDR

input $\{\boldsymbol{x}[n]\}_{n=1}^L$, N, $\hat{\boldsymbol{B}}$ (a starting point)
1: **repeat**
2: **for** $k = 1, \ldots, N$ **do**
3: $\hat{\boldsymbol{F}} := [\hat{\boldsymbol{B}}^T \ \boldsymbol{1}^T]^T$
4: $\hat{\ell}_k := \arg\max_{n=1,\ldots,L} \| \boldsymbol{P}^{\perp}_{\hat{\boldsymbol{F}}_{-k}} \hat{\boldsymbol{x}}[n] \|_2^2$
5: $\hat{\boldsymbol{b}}_k := \boldsymbol{x}[\hat{\ell}_k]$
6: **end for**
7: **until** a stopping rule is satisfied
output $\hat{\boldsymbol{B}} = [\hat{\boldsymbol{b}}_1, \ldots, \hat{\boldsymbol{b}}_N]$.

VolMax-based solutions, such as the SVMAX and SC-N-FINDR algorithms above, are usually simple and efficient to implement. Some further discussions are in order.

1) Historically, Winter mainly used VolMax to devise the N-FINDR concept [7] for pure pixel search. There, the intuition is to update one endmember estimate at a time to iteratively increase the volume. N-FINDR is now a popularized algorithm class in blind HU, where we can find many N-FINDR implementation variants in the literature; see [2], [20], and [38]. The SC-N-FINDR we just illustrated is just among one of the many N-FINDR variants, although we have revealed that SC-N-FINDR has several good characteristics.



[FIG4] **Numerical comparison of VolMax, VolMin, and sparse regression solutions.**

2) VolMax is a provably sound criterion from an endmember identifiability viewpoint. Specifically, the optimal solution of (21) is uniquely the true endmembers' signatures in the noiseless case and under the pure pixel assumption [20]. Also, in this setup, the optimal solution can be easily retrieved by either SC-N-FINDR or SVMAX. However, we should note a fundamental caveat—that SC-N-FINDR and SVMAX are not globally optimal solvers of (21), say, in the presence of noise and/or without pure pixels. In fact, (21) is NP-hard in general [37].

SIMPLEX VOLUME MINIMIZATION

We turn our attention to the simplex volume minimization approach, or simply VolMin, which was first pursued by Craig [4] and Boardman [5] in the blind HU context. VolMin is different from VolMax. It performs simplex fitting by finding a simplex that encloses all the measured pixels, while yielding the minimum volume. This is illustrated in Figure 3(b). Mathematically, VolMin can be formulated as

$$\min_{B} \ \text{vol}(B) \quad \text{s.t. } x[n] \in \text{conv}\{b_1, \ldots, b_N\},\ n = 1, \ldots, L. \tag{27}$$

VolMin is generally recognized as a more powerful approach than VolMax. Let us illustrate this numerically, before describing VolMin optimization schemes. We simulated a noiseless, three endmember case, where the endmembers were taken from a spectral library [39] and the abundances synthetically generated. Figure 4(a) shows a scenario where the pure pixel assumption holds. We see that both VolMax (via SVMAX or SC-N-FINDR) and VolMin perfectly identify the true endmembers. Figure 4(b) shows another scenario where pure pixels are missing. VolMax is seen to fail, while VolMin can still give accurate endmember estimates. Readers are referred to [2], [20], [36], and [40]–[43] for more numerical comparisons and real-data experiments. Simply speaking, VolMin is numerically found to be robust against lack of pure pixels.

Let us now discuss how VolMin is optimized. VolMin does not have simple closed-form schemes as in VolMax, and requires numerical optimization. In fact, the VolMin problem in (27) is more difficult to handle; a major obstacle is with the simplex constraints in (27), which are nonconvex. This issue can be overcome by transforming the simplex to a polyhedron (see, e.g., [35, pp. 32–33]). To help the reader understand the idea, an illustration is given in Figure 5. We see that a simplex can be equivalently represented by an intersection of halfspaces, i.e., a polyhedron. More precisely, the following equivalence holds for an affinely independent $\{b_1, \ldots, b_N\}$ [36]

$$x[n] \in \text{conv}\{b_1, \ldots, b_N\} \iff Hx[n] - g \geq 0, (Hx[n] - g)^T \mathbf{1} \leq 1, \tag{28}$$

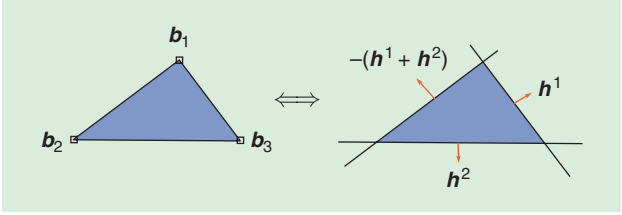
where the RHS is a polyhedron, and

$$H = [b_1 - b_N, \ldots, b_{N-1} - b_N]^{-1}, \quad g = Hb_N. \tag{29}$$

By the change of variables in (29), and noting (28) and (16b), we can recast (27) as

$$\max_{H,g} |\det(H)| \quad \text{s.t. } Hx[n] - g \geq 0, (Hx[n] - g)^T \mathbf{1} \leq 1, n \in \mathcal{L}, \tag{30}$$

where $\mathcal{L} = \{1, \ldots, L\}$. The equivalent VolMin problem in (30) is arguably easier to handle than the original in (27). Specifically, the constraints in (30), which form a data-enclosing polyhedron, are linear (and convex). However, there is still one obstacle—the objective function $|\det(H)|$ is nonconvex. Current



[FIG5] **The transformation of a simplex to a polyhedron.**

state-of-the-art methods for VolMin tackle this issue by successive convex approximation. Specifically, one can apply iterative linear approximation to the objective function [40], [44]. Another alternative is to perform row-by-row AO w.r.t. $(\boldsymbol{H},\boldsymbol{g})$ [36]. These two schemes both operate by solving a sequence of convex optimization problems; see [36], [40], and [44] for the details and comparison.

We complete this subsection by the following comments.

1) As mentioned above, numerical evidence suggests that VolMin may be able to identify the true endmembers accurately in the absence of pure pixels. By analysis, it is known that in the noiseless case, the optimal solution of VolMin is uniquely the true endmembers' signatures if the pure pixel assumption holds [36]. A proof for the no pure pixel case is currently unavailable and is an open fundamental question.

2) While VolMin is deterministic and geometric based, it has a dual identity in stochastic maximum-likelihood (ML) estimation. Specifically, consider the noiseless case, and assume that every abundance vector $\boldsymbol{s}[n]$ is i.i.d. uniformly distributed on the support of unit simplex $\mathcal{S}$. Then, it can be shown that the corresponding ML estimator is the same as the equivalent VolMin problem in (30) [42]. Note that the authors in [42] also consider a generalization where the abundance prior distribution is nonuniform.

FURTHER DISCUSSION

The CG framework presented above is based on exploitation of the simplex $\text{conv}\{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_N\}$. There is an alternative CG formulation where the simplex $\text{conv}\{\boldsymbol{0},\boldsymbol{a}_1,\ldots,\boldsymbol{a}_N\}$ is utilized [40]–[42]; the concepts are identical, though the resulting algorithms exhibit minor differences. Readers should also note other CG interpretations, e.g., [45]. For tutorial purposes, we have focused on the noiseless case only. In the presence of spectrally i.i.d. noise, the ASF preprocessing stage (or equivalently PCA) can be shown to be noise resistant. However, for non-i.i.d. noise, HySime [46] may provide better DR performance. Moreover, both VolMax and VolMin can be modified to improve robustness against noise; e.g., soft constraints [41], chance constraints [43], and robust max-min volume [20]. CG is known to be sensitive to outliers. A robust ASF can be used to identify and discard outliers, before they get into the data [47]. Soft constraints also help "desensitize" VolMin w.r.t. the outliers [41].

DICTIONARY-BASED SEMIBLIND HU

This section describes a relatively new development, where HU is performed by using spectral libraries and techniques arising in compressive sensing (CS). This approach also has a link to sensor array processing in SP, as we will discuss.

SPARSE REGRESSION

When performing blind HU, we generally assume no information on the spectral shapes of the true endmember signatures. The latter is not totally true. In geoscience and remote sensing, a tremendous amount of effort has been spent on measuring and recording spectral samples of many different materials, which has resulted in spectral libraries for various research purposes. For example, the U.S. Geological Survey (USGS) Library, which has taken over 20 years to assemble, contains more than 1,300 spectral samples covering materials such as minerals, rocks, liquids, artificial materials, vegetations, and even microorganisms [39]. Such valuable knowledge base can be turned to blind HU purposes, or more precisely, semiblind HU.

A slight abuse of notations is required to explain the semiblind formulation. We redefine $\boldsymbol{A} = [\boldsymbol{a}_1,\ldots,\boldsymbol{a}_K] \in \mathbb{R}^{M\times K}$ as a dictionary of K hyperspectral samples, where each $\boldsymbol{a}_i$ corresponds to one material (each $\boldsymbol{a}_i$ is also assumed to have been appropriately processed, e.g., atmospherically compensated). We assume that the dictionary $\boldsymbol{A}$ is known, obtained from an available spectral library, and that the true endmembers in each measured pixel $\boldsymbol{y}[n]$ are covered by the dictionary. The measured pixels in the noiseless case (again, for tutorial purposes) can then be represented by

$$\boldsymbol{y}[n] = \sum_{i\in S_n} \boldsymbol{a}_i s_i[n], \tag{31}$$

where $S_n \subseteq \{1,\ldots,K\}$ is an index subset that indicates the materials present in the measured pixel $\boldsymbol{y}[n]$, and $s_i[n] > 0, i \in S_n$ are the corresponding abundances. In this representation, note that the sum-to-one constraint $\sum_{i\in S_n} s_i[n] = 1$ may not hold; the measurement conditions of library samples and the actual scene are often different and this can introduce scaling inconsistencies between the library samples and true endmembers. By also letting $s_i[n] = 0$ for all $i \notin S_n$, (31) can be written as

$$\boldsymbol{y}[n] = \boldsymbol{A}\boldsymbol{s}[n], \tag{32}$$

where $\boldsymbol{s}[n] = [s_1[n],\ldots,s_K[n]]^T \in \mathbb{R}^K$ is now a sparse abundance vector. The problem now is to recover $\boldsymbol{s}[n]$ from $\boldsymbol{y}[n]$. This is not trivial because we often have $K > M$ and the corresponding system in (32) is underdetermined. However, we know beforehand that $\boldsymbol{s}[n]$ have only a few nonzero components, since the number of materials present in one pixel is often very small, typically within five. Hence, a natural formulation for the semiblind HU problem is to find the sparsest $\boldsymbol{s}[n]$ for the representation in (32). This inference problem turns out to be identical to that investigated in CS, where the objective is to recover a sparse representation of a signal on a given frame from compressive measurements [48]. This connection allows us to capitalize on the wealth of theoretical and algorithmic results available in the CS area.

The sparse regression (SR) problem we describe above can be formulated as

$$\min_{\boldsymbol{s}[n]} \|\boldsymbol{s}[n]\|_0 \quad \text{s.t. } \boldsymbol{y}[n] = \boldsymbol{A}\boldsymbol{s}[n], \tag{33}$$

for each $n = 1,\ldots,L$, where $\|\boldsymbol{s}[n]\|_0$ denotes the number of nonzero elements in $\boldsymbol{s}[n]$. The above SR problem possesses provably good endmember identifiability. Specifically, (33) is known to have a unique solution if the true sparse abundance vector $\boldsymbol{s}[n]$ satisfies

$$\| s[n] \|_0 < \frac{1}{2} \cdot \text{spark}(A), \tag{34}$$

where $\text{spark}(A)$ is the smallest number of linearly dependent columns of A [49]. Since every $s[n]$ is highly sparse by nature, (34) should hold in practice. The consequent implication is meaningful—the SR problem (33) can perfectly identify all the true endmembers in general.

While the SR approach sounds promising, there are challenges. Since (33) is NP-hard in general, it is natural to seek approximate solutions. Let us consider the popularized ℓ_1 relaxation solution to (33):

$$\min_{s[n]} \| s[n] \|_1 \quad \text{s.t. } y[n] = As[n], \tag{35}$$

which is convex and has efficient solvers. The CS literature has a series of analysis results telling when (35) gives the same solution as (33), or simply sufficient conditions for exact recovery. Those sufficient conditions usually depend on the conditioning of A. For example, one sufficient exact recovery condition for (35) is $\| s[n] \|_0 < (1/2)(1 + \mu^{-1}(A))$, where

$$\mu(A) = \max_{\substack{1 \le i,j \le K \\ i \ne j}} \frac{|a_i^T a_j|}{\| a_i \|_2 \| a_j \|_2} \tag{36}$$

is called the mutual coherence of A [49]. Unfortunately, spectral libraries in practice are strongly correlated, yielding $\mu(A)$ almost being one [50]. A similar issue also occurs in other sufficient conditions, particularly in the restricted isometry property [48]. Thus, one may not obtain a desirable SR solution from a straight ℓ_1 relaxation application.

However, all is not lost. Recall that every $s[n]$ is, by nature, nonnegative. Let us consider a nonnegative ℓ_1 relaxation problem, which is (35) plus the nonnegative constraint $s[n] \ge 0$. As it turns out, exploiting nonnegativity helps a lot. There is a large amount of experimental evidence that indicates that nonnegative ℓ_1 relaxation can yield useful unmixing results [2], [50], [51]. Also, nonnegative ℓ_1 relaxation is theoretically proven to be able to give rather sparse solutions for certain classes of A [52]. Although the above noted theoretical result does not give a direct answer to exact recovery under highly correlated libraries, it gives good insight on the capability of nonnegative ℓ_1 relaxation.

We can also combat the spectral library mutual coherence issue by using the multiple-measurement vector (MMV) formulation [53], which exploits the fact that in a given data set all the spectral vectors are generated by the same subset of library signatures, corresponding to the endmember signatures. Let $S = [s[1], \ldots, s[L]] \in \mathbb{R}^{K \times L}$ and $Y = [y[1], \ldots, y[L]] \in \mathbb{R}^{M \times L}$, so that we can write $Y = AS$. Also, define $\| S \|_{\text{row}-0}$ to be the number of nonzero rows in S; i.e., $\| S \|_{\text{row}-0} = |\, \text{rowsupp}(S) \,|$, $\text{rowsupp}(S) = \{1 \le i \le K \,|\, s^i \ne 0\}$. We consider a collaborative SR (CSR) problem [54]

$$\min_S \| S \|_{\text{row}-0} \quad \text{s.t. } Y = AS, \tag{37}$$

where the rationale is to use the whole set of measured pixels, rather than one, to strengthen SR performance. It is interesting to note that $\| S \|_{\text{row}-0}$ also represents the number of endmembers. Like the previous SR problem, we can apply a convex relaxation to CSR by replacing $\| S \|_{\text{row}-0}$ in (37) by $\| S \|_{2,1}$, where $\| S \|_{p,q} = \left(\sum_{i=1}^k \| s^i \|_p^q \right)^{1/q}$. In theory, there is no extra benefit in using the CSR or MMV formulation in the worst-case sense (think about a special and rather unrealistic case where $s[1] = \cdots = s[L]$) [53]. However, an average analysis in [55] gives an implication that increasing the number of measurements (or pixels here) can significantly reduce the probability of recovery failure. In practice, this has been found to be so. Also, the nonnegativity constraint $S \ge 0$ can be incorporated in (37) to improve performance.

A practical SR or CSR solution should also cater for the presence of noise. For CSR, the following alternative convex relaxation formulation may be used to provide HU [54]

$$\min_{S \ge 0} \| Y - AS \|_F^2 + \lambda \| S \|_{2,1} \tag{38}$$

for some constant $\lambda > 0$. The rationale is to seek an LS data fitting, rather than exact, with a sparsity-promoting regularizer $\lambda \| S \|_{2,1}$. It is important to note that while (38) is convex, it is a large-scale optimization problem. An efficient solver for (38) is provided in [54], where a divide-and-conquer optimization strategy, specifically, the alternating direction method of multipliers (ADMM), was implemented.

At this point readers may be wondering: How do we compare SR- and CG-based solutions? Simply speaking, CG relies on exploitation of simplex structures, while SR does not. To illustrate, consider the previous numerical example in Figure 4. In Figure 4(c), we generated a heavily mixed (and noiseless) scenario where data do not possess simplex structures expected in CG. It is seen that even VolMin fails in this scenario. However, CSR, which was run under the USGS Library with 498 spectral signatures, is seen to be able to identify the true endmembers perfectly. Note that the true endmember signatures were taken from the same library, which makes the setting slightly ideal. It would not be too surprising that if the library fails to cover all true endmember signatures (e.g., a new material), then SR solutions would fail. For further numerical results and real-data experiments, see [2], [50], [54], [56], and [57].

SENSOR ARRAY PROCESSING MEETS SEMIBLIND HU

MMV is a powerful concept that has been applied to estimation problems in statistical SP and sensor array processing [58]. Curiously, a classical concept originated from sensor array processing, specifically, subspace methods, also finds its way to MMV research [59]—this provides yet another opportunity for semiblind HU [56].

The idea is simple for readers who are familiar with subspace methods or sensor array processing; or, see classical literatures such as [17]. Consider the block model $Y = AS$ (again, assuming no noise). Let $S = \text{rowsupp}(S)$ be the set of indices of active materials in the measured data Y, and A_S be a submatrix

of $\boldsymbol{A}$ whose columns are $\{\boldsymbol{a}_i\}_{i\in S}$. Note that $\boldsymbol{A}_S$ is the true endmember matrix. Let us assume that $\{\boldsymbol{s}^i\}_{i\in S}$, the set of true abundance maps, is linearly independent; in practice this refers to situations where the abundance maps are sufficient different. Then, one can easily deduce that $\mathcal{R}(\boldsymbol{Y}) = \mathcal{R}(\boldsymbol{A}_S)$, where $\mathcal{R}$ denotes the range space of its argument. The above expression implies that

$$\boldsymbol{P}_{\boldsymbol{Y}}^{\perp}\boldsymbol{a}_k = \boldsymbol{0} \Longleftrightarrow k \in S \tag{39}$$

for all $1 \le k \le K$, as far as $\{\boldsymbol{a}_k\} \cup \{\boldsymbol{a}_i\}_{i\in S\setminus\{k\}}$ is linearly independent for any $1 \le k \le K$. Since the latter holds for $|S|+1 < \text{spark}(\boldsymbol{A})$, we have the following endmember identifiability condition for (39):

$$\|\boldsymbol{S}\|_{\text{row}-0} < \text{spark}(\boldsymbol{A}) - 1. \tag{40}$$

Remarkably, with the mild assumption of linear independence of $\{\boldsymbol{s}^i\}_{i\in S}$, we can achieve such provably good endmember identifiability by the simple subspace projection in (39).

In practice, the identification in (39) can be implemented by the classical multiple signal classification (MUSIC) method [17]; see [56] for implementation details.

FURTHER DISCUSSION

There are a few more points to note.

1) As a side advantage, the SR approach does not require knowledge of the number of endmembers N. Note that this does not apply to the subspace approach, which often requires knowledge of N to construct subspace projections.

2) Hyperspectral signals are very often piecewise smooth w.r.t. their three dimensional domain (one spectral dimension plus two spatial dimensions). Therefore, one can exploit such spatial/spectral contextual information for improving SR performance by applying piecewise smooth regularization, such as total variations (TVs) [57], on top of an SR formulation, e.g., (38).

3) An interesting (but also elusive) question is whether a given dictionary can truly cover the true endmembers. From an end user's viewpoint, it depends on the scene and whether one can preselect a reliable library for that scene specifically. Moreover, there are concurrent studies that consider learning the dictionary from the data, thereby circumventing these issues [51], [60], [61]. Dictionary learning is an active research topic. It is also related to NMF, to be described in the next section. In addition, there has been interest in using the measured data $\boldsymbol{Y}$ itself as the dictionary for MMV [62]. This self-dictionary MMV (SD-MMV) approach is related to pure pixel search. For example, SPA and VCA can both be derived from SD-MMV [63].

NONNEGATIVE MATRIX FACTORIZATION

This section turns the attention back to blind HU, where we review a class of algorithms known as nonnegative matrix factorization (NMF).

NMF was originally proposed as a linear DR tool for analyzing environmental data [64] and for data mining applications [65]. It is posed as a low-rank matrix approximation problem where, given a data matrix $\boldsymbol{Y} \in \mathbb{R}^{M\times L}$, the task is to find a pair of nonnegative matrices $\boldsymbol{A} \in \mathbb{R}^{M\times N}$, $\boldsymbol{S} \in \mathbb{R}^{N\times L}$, with $N < \min\{M, L\}$, that solves

$$\min_{\boldsymbol{A}\ge 0,\, \boldsymbol{S}\ge 0} \|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{S}\|_F^2. \tag{41}$$

In blind HU, the connection is that the NMF factors obtained, $\boldsymbol{A}$ and $\boldsymbol{S}$, can serve as estimates of the endmembers and abundances, respectively (note that endmember spectral signatures are nonnegative by nature). However, there are two problems here. First, (41) is NP-hard in general [66]. For this reason, optimization schemes we see in the current NMF-based blind HU developments are rather pragmatic. We should, however, mention that lately, there are new theory-guided NMF developments in optimization [67], [68]. Second, NMF may not guarantee solution uniqueness. This is a serious issue to the blind HU application, since it means that an NMF solution may not necessarily be the true endmembers and abundances, even in the noiseless case.

In blind HU, NMF is modified to fit the problem better. Roughly speaking, we may unify many NMF-based blind HU developments under one formulation

$$\min_{\boldsymbol{A}\ge 0,\, \boldsymbol{S}\in\mathcal{S}^L} \|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{S}\|_F^2 + \lambda\cdot g(\boldsymbol{A}) + \mu\cdot h(\boldsymbol{S}), \tag{42}$$

where $\mathcal{S}^L = \{\boldsymbol{S} \mid \boldsymbol{s}[n] \ge \boldsymbol{0}, \boldsymbol{1}^T\boldsymbol{s}[n] = 1, 1 \le n \le L\}$, g and h are regularizers, which vary from one work to another, and $\lambda, \mu > 0$ are some constants. In particular, the addition of g and h is to make (42) more well posed through exploitation of the problem natures. Also, for the same reason, we incorporate the unit simplex constraints on $\boldsymbol{S}$.

In the literature, one can find a plethora of NMF-based blind HU algorithms—each work may use different g, h, modified constraints for simpler implementations (e.g., no constraints on $\boldsymbol{A}$), and a different optimization algorithm. Our intention here is not to give an extensive coverage of all these developments. Instead, we are interested in several representative NMF-based blind HU formulations, where we will see connections between NMF, CG, and SR. A summary of those formulations is shown in Table 1.

Although we see many choices with the regularizers g and h, the philosophies behind the choices follow a few core principles. For the endmember regularizer g, the principle can be traced back to VolMin in CG. A classical example is minimum volume constrained NMF (MVC-NMF) [69]

$$\min_{\boldsymbol{A}\ge 0, \boldsymbol{S}\in\mathcal{S}^L} \|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{S}\|_F^2 + \lambda\cdot(\text{vol}(\boldsymbol{B}))^2, \tag{43}$$

where $\text{vol}(\boldsymbol{B})$ is the simplex volume corresponding to $\boldsymbol{A}$, in which $\boldsymbol{b}_i = \boldsymbol{C}^{\dagger}(\boldsymbol{a}_i - \boldsymbol{d})$ for all i; cf. the section "Convex Geometry." MVC-NMF is essentially a variation of the VolMin formulation [see (27)] in the noisy case, with endmember nonnegativity incorporated. As mentioned before, $\text{vol}(\boldsymbol{B})$ is nonconvex. Iterated constrained endmember (ICE) [70] and sparsity promoting ICE (SPICE) [73] avoid this issue by replacing $(\text{vol}(\boldsymbol{B}))^2$ with a

[TABLE1] A SUMMARY OF SOME NMF FORMULATIONS.

ALGORITHM	$g(A)$	$h(S)$	OPTIMIZATION SCHEMES AND REMARKS
MVC-NMF [69]	$\text{vol}^2(\boldsymbol{C}^T(\boldsymbol{A}-\boldsymbol{d}\boldsymbol{1}^T))$	0	AO + ONE-STEP PROJECTED GRADIENT
ICE [70]	$\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\|\|\boldsymbol{a}_i-\boldsymbol{a}_j\|\|_2^2$	0	AO; UNCONSTRAINED $\boldsymbol{A}$
DL [60]	0	$\|\|\boldsymbol{S}\|\|_{1,1}$	AO + ONE-STEP PROJECTED GRADIENT FOR $\boldsymbol{A}$; $\boldsymbol{S}\geq\boldsymbol{0}$
$L_{1/2}$-NMF [71]	0	$\|\|\boldsymbol{S}\|\|_{1/2,1/2}^{1/2}$	AO + MULTIPLICATIVE UPDATE
APS [72]	0	$\sum_{n=1}^{L}\sum_{j\in\mathcal{N}(n)}\|\|\boldsymbol{s}[n]-\boldsymbol{s}[j]\|\|_1$ WHERE $\mathcal{N}(n)$ IS THE NEIGHBORHOOD PIXEL INDEX SET OF PIXEL n.	AO + ONE-STEP PROJECTED SUBGRADIENT
SPICE [73]	$\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\|\|\boldsymbol{a}_i-\boldsymbol{a}_j\|\|_2^2$	$\sum_{i=1}^{N}\gamma_i\|\|\boldsymbol{s}^i\|\|_1$	AO; UNCONSTRAINED $\boldsymbol{A}$; ITERATIVELY REWEIGHTED γ_i VIA $\gamma_i:=1/\|\|[\boldsymbol{S}^{(k-1)}]_{i,1:L}\|\|_1$, $1\leq i\leq N$
CoNMF [74]	$\sum_{i=1}^{N}\|\|\boldsymbol{a}_i-\boldsymbol{\mu}_y\|\|_2^2$	$\sum_{i=1}^{N}\|\|\boldsymbol{s}^i\|\|_2^p$, $0<p\leq 1$	AO + ONE-STEP MAJORIZATION MINIMIZATION; UNCONSTRAINED $\boldsymbol{A}$

convex surrogate, specifically, $g(A)=\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\|\boldsymbol{a}_i-\boldsymbol{a}_j\|_2^2$, which is the sum of differences between vertices. A similar idea is also adopted in collaborative NMF (CoNMF) [74]; see Table 1.

As for the abundance regularizer h, the design principle usually follows that of sparsity. A good showcasing example, curiously, lies in dictionary learning (DL) [60]

$$\min_{A\geq 0,S\geq 0}\|Y-AS\|_F^2+\mu\cdot\|S\|_{1,1}; \tag{44}$$

note that $\|S\|_{1,1}=\sum_{n=1}^{L}\sum_{i=1}^{N}|s_i[n]|$. The original idea of (44) is to learn the dictionary A by joint dictionary and sparse signal optimization; cf. the section "Dictionary-Based Semiblind HU" and, in particular, (38). However, (44) can also be seen as an NMF with sparsity-promoting regularization. Following the same spirit, $L_{1/2}$-NMF [71] uses a nonconvex, but stronger sparsity-promoting regularizer based on the $\ell_{1/2}$ quasinorm. Apart from sparsity, exploitation of spatial contextual information via TV regularization may also be used [72].

The aforementioned connection between DL and NMF provides an additional insight. In DL, the dictionary size is often set to be large, and should be larger than the true number of endmembers; the number of endmembers is instead determined by the row sparsity of S, i.e., $\|S\|_{\text{row}-0}$. From an NMF-based blind HU perspective, this means that we can use row sparsity to provide joint endmember number, endmember and abundance estimation. More formally, consider a blind version of the MMV (38)

$$\min_{A\geq 0,S\in\mathcal{S}^L}\|Y-AS\|_F^2+\lambda\cdot g(A)+\mu\cdot\|S\|_{\text{row}-0}, \tag{45}$$

where the number of columns of A, given by N, is now chosen to be a number greater than the true number of endmembers (say, by overestimating the latter), and we use $\|S\|_{\text{row}-0}$ to represent the endmember number. SPICE is arguably the first algorithm that explores such opportunity [73]. In SPICE, the abundance regularizer can be expressed as $h(S)=\sum_{i=1}^{N}\gamma_i\|\boldsymbol{s}^i\|_1$ for some weights $\{\gamma_i\}$ that are iteratively updated; this regularizer is a convex surrogate of $\|S\|_{\text{row}-0}$. CoNMF also aims at row sparsity, using a nonconvex surrogate $h(S)=\sum_{i=1}^{K}\|\boldsymbol{s}^i\|_2^p$, $0<p\leq 1$ [74].

We should also discuss optimization in NMF-based blind HU. Most NMF-based blind HU algorithms follow a two-block AO strategy, although their implementation details exhibit many differences. Two-block AO optimizes (42) w.r.t. either A or S alternatingly. Specifically, it generates a sequence of iterates $\{(A^{(k)},S^{(k)})\}_k$ via

$$A^{(k)}=\arg\min_{A\geq 0}\|Y-AS^{(k-1)}\|_F^2+\lambda\cdot g(A) \tag{46a}$$

$$S^{(k)}=\arg\min_{S\in\mathbb{S}^L}\|Y-A^{(k)}S\|_F^2+\mu\cdot h(S). \tag{46b}$$

Note that if g and h are convex, then (46a)–(46b) are convex and hence can usually be solved efficiently. Moreover, every limit point of $\{(A^{(k)},S^{(k)})\}_k$ is a stationary point of (42) under some fairly mild assumptions [75], [76]. For practical reasons, most algorithms use cheap but inexact updates for (46a) and (46b), e.g., multiplicative update [71], one-step projected gradient or subgradient update [60], [69], [72], and one-step majorization minimization [74]. Convergence to a stationary point of these inexact AO methods has still to be thoroughly analyzed. However, by numerical experience, many NMF-based blind HU algorithms work well under appropriate settings (e.g., using reasonable initializations that can be obtained, e.g., with VCA or N-FINDR).

To summarize, NMF is a versatile approach that has connections to both CG and SR. It leads to a fundamentally hard optimization problem, although practical solutions based on two-block AO usually offer good performance by experience. Also, we should highlight that the more exciting developments of NMF-based blind HU lie in extensions to scenarios such as nonlinear HU [77], EV [78], and multispectral and hyperspectral data fusion [79]. Such extensions may not be easily achieved in other approaches.

CONCLUSIONS

This article provided a tutorial review on blind HU techniques using a fundamental SP perspective. Four major blind HU approaches—pure pixel search, convex geometry, sparse regression, and NMF—have been studied. We briefly compare their advantages and drawbacks. Pure pixel search and VolMax are very simple but require the pure pixel assumption; VolMin is resistant

to lack of pure pixels but still has limitations when data are too heavily mixed; sparse regression holds great potential in unmixing heavily mixed data but one should be aware of its reliance on dictionaries; NMF is a very flexible formulation for blind HU but leads us to a hard optimization problem to solve. Also, real hyperspectral data can be quite elusive at times, where we may be faced with issues such as outliers, modeling errors, and uncertainty in the number of endmembers. Their subsequent effects on the aforementioned approaches could be substantial. On the other hand, the need for meeting these challenges also makes HU continue to be a vibrant and active field of research.

ACKNOWLEDGMENTS

This work was supported in part by the Hong Kong Research Grants Council General Research Fund Project CUHK 415509; Portuguese Science and Technology Foundation, Project PEst-OE/EEI/LA0008/2013; Labex CIMI during visits of J. Bioucas-Dias at the University of Toulouse and National Science Council (R.O.C.) under grant NSC 102-2221-E-007-035-MY2.

AUTHORS

Wing-Kin Ma (wkma@ieee.org) is an associate professor with the Department of Electronic Engineering, the Chinese University of Hong Kong.

José M. Bioucas-Dias (bioucas@lx.it.pt) is an associate professor in the Department of Electronic Engineering, Instituto Superior Técnico, Technical University of Lisbon, Portugal.

Tsung-Han Chan (Th.chan@adsc.com.sg) is a research scientist at the Advanced Digital Sciences Centre, Singapore.

Nicolas Gillis (nicolas.gillis@umons.ac.be) is an assistant professor in the Department of Mathematics and Operational Research, Faculté Polytechnique, Université de Mons, Belgium.

Paul Gader (pgader@cise.ufl.edu) is a professor with the Department of Computer and Information Science and Engineering, University of Florida.

Antonio J. Plaza (aplaza@unex.es) is an associate professor with the Department of Technology of Computers and Communications, University of Extremadura, Cáceres, Spain.

ArulMurugan Ambikapathi (a.arulmurugan@gmail.com) is a postdoctoral research fellow at the Institute of Communications Engineering and Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan.

Chong-Yung Chi (cychi@ee.nthu.edu.tw) is a professor with the Institute of Communications Engineering and Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan.

REFERENCES

[1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Mag.*, vol. 19, no. 1, pp. 44–57, 2002.

[2] J. 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. Select. Topics Appl. Earth Observ. Remote Sensing,* vol. 5, no. 2, pp. 354–379, 2012.

[3] M. D. Craig, "Unsupervised unmixing of remotely sensed images," in *Proc. Australasian Remote Sensing Conf.*, Perth, Australia, 1990, pp. 324–330.

[4] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, no. 3, pp. 542–552, May 1994.

[5] J. W. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Proc. Summary 4th Annual JPL Airborne Geoscience Workshop*, Dec. 1993, vol. 1, pp. 11–14.

[6] J. Boardman, F. Kruse, and R. Green, "Mapping target signatures via partial unmixing of AVIRIS data," in *Proc. Summary JPL Airborne Earth Science Workshop*, Pasadena, CA, Dec. 1995, vol. 1, pp. 23–26.

[7] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral endmember determination in hyperspectral data," in *Proc. SPIE Conf. Imaging Spectrometry*, Pasadena, CA, Oct. 1999, pp. 266–275.

[8] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Processing*, vol. 57, no. 11, pp. 4355–4368, 2009.

[9] B.-C. Gao, M. J. Montes, C. O. Davis, and A. F. Goetz, "Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean," *Remote Sensing Environ.*, vol. 113, pp. S17–S24, Sept. 2009.

[10] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. Bermudez, S. Mclaughlin, and A. Hero, "Nonlinear unmixing of hyperspectral images," *IEEE Signal Processing Mag.*, vol. 31, no. 1, pp. 82–94, Jan. 2014.

[11] A. Zare and K. C. Ho, "Endmember variability in hyperspectral analysis," *IEEE Signal Processing Mag.*, vol. 31, no. 1, pp. 95–104, Jan. 2014.

[12] CVX Research, Inc. (2012, Sept.). CVX: MATLAB software for disciplined convex programming, version 2.0 beta. [Online]. Available: http://cvxr.com/cvx

[13] D. Heinz and C.-I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 3, pp. 529–545, 2001.

[14] J. Bioucas-Dias and M. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing,* Raykjavik, Iceland, 2010, pp. 1–4.

[15] R. Heylen, D. Burazerovi, and P. Scheunders, "Fully constrained least squares spectral unmixing by simplex projection," *IEEE Trans. Geosci. Remote Sensing*, vol. 49, no. 11, pp. 4112–4122, Nov. 2011.

[16] J. Nascimento and J. Bioucas-Dias, "Does independent component analysis play a role in unmixing hyperspectral data?" *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 1, pp. 175–187, 2005.

[17] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. Hoboken, NJ: Wiley, 2002.

[18] U. Araújo, B. Saldanha, R. Galvão, T. Yoneyama, H. Chame, and V. Visani, "The successive projections algorithm for variable selection in spectroscopic multicomponent analysis," *Chemometr. Intell. Lab. Syst.*, vol. 57, no. 2, pp. 65–73, 2001.

[19] 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.

[20] 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 Sensing*, vol. 49, no. 11, pp. 4177–4193, 2011.

[21] G. Golub, "Numerical methods for solving linear least squares problems," *Numerische Mathematik*, vol. 7, no. 3, pp. 206–216, June 1965.

[22] S. Arora, R. Ge, Y. Halpern, D. Mimno, A. Moitra, D. Sontag, Y. Wu, and M. Zhu, "A practical algorithm for topic modeling with provable guarantees," in *Proc. Int. Conf. Machine Learning*, vol. 28, no. 2, 2013, pp. 280–288.

[23] N. Gillis and S. Vavasis, "Fast and robust recursive algorithms for separable nonnegative matrix factorization," [Online]. Available: http://arxiv.org/abs/1208.1237

[24] A. Ambikapathi, T.-H. Chan, C.-Y. Chi, and K. Keizer, "Hyperspectral data geometry-based estimation of number of endmembers using p-norm-based pure pixel identification algorithm," *IEEE Trans. Geosci. Remote Sensing*, vol. 51, no. 5, pp. 2753–2769, 2013.

[25] J. Nascimento and J. Bioucas-Dias, "Vertex component analysis: A fast algorithm to extract endmembers spectra from hyperspectral data," in *Pattern Recognition and Image Analysis: 1st Iberian Conf.*, 2003, pp. 626–635.

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

[27] C.-I. Chang, W. Xiong, H.-M. Chen, and J.-W. Chai, "Maximum orthogonal subspace projection approach to estimating the number of spectral signal sources in hyperspectral imagery," *IEEE J. Select. Topics Signal Processing*, vol. 5, no. 3, pp. 504–520, 2011.

[28] J. Imbrie, "Vector analysis of heavy-mineral data," *Geological Soc. Amer. Bull.*, vol. 75, pp. 1131–1156, Nov. 1964.

[29] R. Ehrlich and W. E. Full, "Sorting out geology—Unmixing mixtures," in *Use and Abuse of Statistical Methods in the Earth Sciences*, W. Size, Ed. New York: Oxford Univ. Press, 1987, pp. 33–46.

[30] J. E. Klovan and A. T. Miesch, "Extended CABFAC and QMODEL computer programs for Q-mode factor analysis of compositional data," *Comput. Geosci.*, vol. 1, no. 3, pp. 161–178, 1976.

[31] W. E. Full, R. Ehrlich, and J. E. Klovan, "Extended QMODEL—Objective definition of external endmembers in the analysis of mixtures," *Math. Geology*, vol. 13, no. 4, pp. 331–344, 1981.

[32] A. Perczel, M. Hollósi, G. Tusnády, and G. D. Fasman, "Convex constraint analysis: A natural deconvolution of circular dichroism curves of proteins," *Protein Eng.*, vol. 4, no. 6, pp. 669–679, 1991.

[33] W. Naanaa and J.-M. Nuzillard, "Blind source separation of positive and partially correlated data," *Signal Process.*, vol. 85, no. 9, pp. 1711–1722, 2005.

[34] T.-H. Chan, W.-K. Ma, C.-Y. Chi, and Y. Wang, "A convex analysis framework for blind separation of non-negative sources," *IEEE Trans. Signal Processing*, vol. 56, no. 10, pp. 5120–5134, 2008.

[35] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[36] T.-H. Chan, C.-Y. Chi, Y.-M. Huang, and W.-K. Ma, "A convex analysis based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Signal Processing*, vol. 57, no. 11, pp. 4418–4432, 2009.

[37] A. Çivril and M. Magdon-Ismail, "On selecting a maximum volume sub-matrix of a matrix and related problems," *Theor. Comput. Sci.*, vol. 410, no. 47, pp. 4801–4811, 2009.

[38] C.-C. Wu, S. Chu, and C.-I. Chang, "Sequential N-FINDR algorithms," in *Proc. SPIE*, vol. 7086, 2008, p. 70860C.

[39] R. Clark, G. Swayze, R. Wise, E. Livo, T. Hoefen, R. Kokaly, and S. Sutley. (2007). USGS digital spectral library splib06a: U.S. Geological Survey, Digital Data Series 231. [Online]. Available: http://speclab.cr.usgs.gov/spectral.lib06.

[40] J. Li and J. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *Proc. IEEE IGARSS*, Aug. 2008, pp. 250–253.

[41] J. Bioucas-Dias, "A variable splitting augmented Lagrangian approach to linear spectral unmixing," in *Proc. IEEE WHISPERS*, Aug. 2009, pp. 1–4.

[42] J. Nascimento and J. Bioucas-Dias, "Hyperspectral unmixing based on mixtures of Dirichlet components," *IEEE Trans. Geosci. Remote Sensing*, vol. 50, no. 3, pp. 863–878, 2012.

[43] A. Ambikapathi, T.-H. Chan, W.-K. Ma, and C.-Y. Chi, "Chance-constrained robust minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sensing*, vol. 49, no. 11, pp. 4194–4209, 2011.

[44] M. B. Lopes, J. C. Wolff, J. Bioucas-Dias, and M. Figueiredo, "NIR hyperspectral unmixing based on a minimum volume criterion for fast and accurate chemical characterisation of counterfeit tablets," *Anal. Chem.*, vol. 82, no. 4, pp. 1462–1469, 2010.

[45] P. Honeine and C. Richard, "Geometric unmixing of large hyperspectral images: A barycentric coordinate approach," *IEEE Trans. Geosci. Remote Sensing*, vol. 50, no. 6, pp. 2185–2195, 2012.

[46] J. Bioucas-Dias and J. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, 2008.

[47] T.-H. Chan, A. Ambikapathi, W.-K. Ma, and C.-Y. Chi, "Robust affine set fitting and fast simplex volume max-min for hyperspectral endmember extraction," *IEEE Trans. Geosci. Remote Sensing*, vol. 51, no. 7, pp. 3982–3997, 2013.

[48] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Processing Mag.*, vol. 24, no. 4, pp. 118–121, 2007.

[49] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," *Proc. Natl. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, 2003.

[50] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sensing*, vol. 49, no. 6, pp. 2014–2039, 2011.

[51] J. B. Greer, "Sparse demixing of hyperspectral images," *IEEE Trans. Image Processing*, vol. 21, no. 1, pp. 219–228, 2012.

[52] D. L. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming," *Proc. Natl. Acad. Sci.*, vol. 102, no. 27, pp. 9446–9451, 2005.

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

[54] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sensing*, to be published.

[55] Y. C. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Trans. Inform. Theory*, vol. 56, no. 1, pp. 505–519, 2010.

[56] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "MUSIC-CSR: Hyperspectral unmixing via multiple signal classification and collaborative sparse regression," *IEEE Trans. Geosci. Remote Sensing*, to be published.

[57] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sensing*, vol. 50, no. 11, pp. 4484–4502, 2012.

[58] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

[59] J. M. Kim, O. K. Lee, and J.-C. Ye, "Compressive MUSIC: Revisiting the link between compressive sensing and array signal processing," *IEEE Trans. Inform. Theory*, vol. 58, no. 1, pp. 278–301, 2012.

[60] A. Charles, B. Olshausen, and C. Rozell, "Learning sparse codes for hyperspectral imagery," *IEEE J. Select. Topics Signal Processing*, vol. 5, no. 5, pp. 963–978, 2011.

[61] A. Castrodad, Z. Xing, J. B. Greer, E. Bosch, L. Carin, and G. Sapiro, "Learning discriminative sparse representations for modeling, source separation, and mapping of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 49, no. 11, pp. 4263–4281, 2011.

[62] 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 Processing*, vol. 21, no. 7, pp. 3239 –3252, 2012.

[63] X. Fu, W.-K. Ma, T.-H. Chan, J. M. Bioucas-Dias, and M.-D. Iordache, "Greedy algorithms for pure pixels identification in hyperspectral unmixing: A multiple-measurement vector viewpoint," in *Proc. EUSIPCO 2013*, to be published.

[64] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111–126, 1994.

[65] D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[66] S. A. Vavasis, "On the complexity of nonnegative matrix factorization," *SIAM J. Optim.*, vol. 20, no. 3, pp. 1364–1377, 2009.

[67] N. Gillis and R. Plemmons, "Sparse nonnegative matrix underapproximation and its application to hyperspectral image analysis," *Lin. Alg. Appl.*, vol. 438, no. 10, pp. 3991–4007, 2013.

[68] S. Arora, R. Ge, R. Kannan, and A. Moitra, "Computing a nonnegative matrix factorization—Provably," in *Proc. 44th Symp. Theory of Computing*, ACM, 2012, pp. 145–162.

[69] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sensing*, vol. 45, no. 3, pp. 765–777, 2007.

[70] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. F. Huntington, "ICE: A statistical approach to identifying endmembers in hyperspectral images," *IEEE Trans. Geosci. Remote Sensing*, vol. 42, no. 10, pp. 2085–2095, 2004.

[71] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via $L_{1/2}$ sparsity-constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sensing*, vol. 49, no. 11, pp. 4282–4297, 2011.

[72] A. Zymnis, S.-J. Kim, J. Skaf, M. Parente, and S. Boyd, "Hyperspectral image unmixing via alternating projected subgradients," in *Proc. Asilomar Conf. Signals, Systems and Computers*, 2007, pp. 1164–1168.

[73] A. Zare and P. Gader, "Sparsity promoting iterated constrained endmember detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 3, pp. 446–450, 2007.

[74] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Collaborative nonnegative matrix factorization for remotely sensed hyperspectral unmixing," in *Proc. IEEE IGARSS*, 2012, pp. 3078–3081.

[75] L. Grippo and M. Sciandrone, "On the convergence of the block nonlinear Gauss-Seidel method under convex constraints," *Oper. Res. Lett.*, vol. 26, no. 3, pp. 127–136, 2000.

[76] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *J. Optim. Theory Appl.*, vol. 109, no. 3, pp. 475–494, 2001.

[77] P. Gader, D. Dranishnikov, A. Zare, and J. Chanussot, "A sparsity promoting bilinear unmixing model," in *Proc. IEEE WHISPERS*, 2012.

[78] A. Zare and P. Gader, "PCE: Piecewise convex endmember detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 6, pp. 2620–2632, 2010.

[79] N. Yokoya, T. Yair, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sensing*, vol. 50, no. 2, pp. 528–537, 2012.

[SP]