ON HYPERSONTICAL UNMIXING

Wing-Kin Ma

Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong SAR of China

ABSTRACT

In this article the author reviews Jose Bioucas-Dias’ key contributions to hyperspectral unmixing (HU), in memory of him as an influential scholar and for his many beautiful ideas introduced to the hyperspectral community. Our story will start with vertex component analysis (VCA)—one of the most celebrated HU algorithms, with more than 2,000 Google Scholar citations. VCA was pioneering, invented at a time when HU research just began to emerge, and it shows sharp insights on a then less-understood subject. Then we will turn to SISAL, another widely-used algorithm. SISAL is not only a highly successful algorithm, it is also a demonstration of its inventor’s ingenuity on applied optimization and on smart formulation for practical noisy cases. Our tour will end with dependent component analysis (DECA), perhaps a less well-known contribution. DECA shows sharp insights on a then less-understood subject. The author’s ingenuity on applied optimization and on smart formulation for dependent component analysis (DECA)—one of the most celebrated HU algorithms, invented at a time when HU research just began to emerge, and it shows sharp insights on a then less-understood subject. Then we will turn to SISAL, another widely-used algorithm. SISAL is not only a highly successful algorithm, it is also a demonstration of its inventor’s ingenuity on applied optimization and on smart formulation for practical noisy cases. Our tour will end with dependent component analysis (DECA), perhaps a less well-known contribution. DECA shows sharp insights on a then less-understood subject.

Index Terms— Hyperspectral unmixing, pure-pixel search, volume minimization, probabilistic simplex component analysis

1. INTRODUCTION

It is well-known within the hyperspectral community that hyperspectral unmixing (HU) is a promising topic, providing a strikingly rich variety of methods for blindly identifying materials’ spectral responses from a hyperspectral image. Since HU is a well-established topic, the author will not elaborate on, or restate, the numerous developments of HU; the reader can easily find overview articles such as [1, 2] and closely-related articles such as [3, 4]. The author wants to give a short tour on José Bioucas-Dias’ original contributions to HU, in memory of him as the undisputedly greatest researcher of our time in hyperspectral signal and image processing. The author’s background is on signal processing, with an emphasis on fundamental aspects. It is inevitable that he will use his lens to view Bioucas-Dias’ inventions and the insights thereof, and the reader should note that his view represents only one of the perspectives to appreciate Bioucas-Dias’ works. In fact, the author and Bioucas-Dias did not necessarily share the same view on every aspect despite the fact that they are good friends and had many discussions for about a decade.

This paper will cover vertex component analysis (VCA) [5], simplex identification via split augmented Lagrangian (SISAL) [6] and dependent component analysis (DECA) [7]. VCA is very widely-used and should be the most cited work in the history of HU. SISAL is arguably the most popularly-used algorithm among the various simplex volume minimization algorithms in HU. DECA is, by comparison, not as well-known, but it shows significant insights and great potential as the author will explain. The author will not cover the model-order estimator HySime and sparse unmixing, which are also Bioucas-Dias’ key contributions.

The author assumes the prerequisite that the reader has knowledge about the signal processing basics of HU (see, e.g., [2–4]). Or, for a reader not from remote sensing, he or she should be equipped with relevant concepts in signal processing or machine learning, e.g., blind source separation, or non-negative matrix factorization.

2. VCA

We begin by hypothesizing that the hyperspectral image we capture obeys a noiseless linear mixture model

\[ y_t = \sum_{i=1}^{N} a_i s_{t,i} = As_t, \quad t = 1, \ldots, T, \]  

(1)

where \( y_t \in \mathbb{R}^M \) collects the reflectances of the image over \( M \) spectral bands and at a pixel indexed by \( t \); each \( a_i \in \mathbb{R}^M \) is the spectral response of a distinct endmember; \( A = [a_1, \ldots, a_N] \); \( s_t = [s_{1,t}, \ldots, s_{N,t}]^\top \) describes the abundances of the different endmembers at pixel \( t \); \( N \) is the number of endmembers; \( T \) is the number of pixels. It is typical to assume, and we will assume, that i) every \( s_t \) lies in the unit simplex \( \mathcal{U} := \{ s \in \mathbb{R}^N \mid s \geq 0, s^\top 1 = 1 \} \); ii) \( A \) has full column rank; iii) \( S = [s_1, \ldots, s_T] \) has full row rank. The problem of HU is to identify \( A \) from the image \( Y = [y_1, \ldots, y_T] \).

VCA uses the pure-pixel assumption—i.e., for each endmember, there exists a pixel that is contributed purely by that endmember. Mathematically, we say that the pure-pixel assumption holds if, for each \( i \), there exists a \( t_i \in \{1, \ldots, T\} \) such that \( y_{t_i} = a_i \). Under the pure-pixel assumption, the problem of identifying \( A \) can be done by finding \( t_1, \ldots, t_N \). VCA has its insight reminiscent of Boardman’s pure-pixel index (PPI) [8], which employs

\[ \hat{i} = \arg\max_{t=1, \ldots, T} \langle [y_t, r], r \rangle, \]

(2)

to identify pure pixels. Here, \( \langle ., . \rangle \) denotes the inner product; \( r \) is a randomly drawn direction. Eq. (2) randomly projects the \( y_t \)’s into a line and then finds an extreme point there. It is easy to show that \( \hat{i} \) is one of the pure-pixel indices \( t_i \’s \), with a high probability. The issue with (2) is that we may need to re-run (2) many times many times to obtain \( t_1, \ldots, t_N \), all endmembers’ pure-pixel indices.

VCA fixes the above issue by orthogonal projection. Suppose we already identified a number of \( k \) pure-pixel indices, each corresponding to a distinct endmember. Without loss of generality (w.l.o.g.), let \( t_1, \ldots, t_k \) be the identified pure-pixel indices. Since \( a_1, \ldots, a_k \) is known (note \( y_{t_i} = a_i \)), consider a randomly drawn direction \( r_k \) such that \( r_k \) is orthogonal to \( a_1, \ldots, a_k \). Since

\[ |\langle y_{t_i}, r_k \rangle| = |\sum_{i=k+1}^{N} \langle a_i, r_k \rangle s_{t,i} | \leq \max_{i=k+1, \ldots, N} |\langle a_i, r_k \rangle|, \]

(3)

where equality holds if \( y_{t_i} = a_i, i \in \arg\max_{i=k+1, \ldots, N} |\langle a_i, r_k \rangle| \), it makes sense to consider

\[ \hat{i}_{k+1} = \arg\max_{t=1, \ldots, T} |\langle y_{t}, r_k \rangle| \]

(4)

The work was supported by a General Research Fund (GRF) of Hong Kong Research Grant Council (RGC) under Project ID CUHK 14205717.
to identify a new pure-pixel index. One may show that $\hat{t}_{k+1}$ is a
pure-pixel index associated with a new endmember $a_i, i \in \{k + 1, \ldots, N\}$, with a high probability. The idea of VCA is to succes-
sively run (4), from $k = 1$ to $k = N$. Compared with PPI, VCA
requires calling the projections $N$ times only.

The merit of VCA is that it is computationally very efficient.
While later developments lead to similar algorithms, with some hav-
ing better fundamental explanations and/or provable identifiability
 guarantees (see, e.g., [3, 4] and the references therein), VCA was in-
troduced at a time when the theory and methods for pure-pixel search
were much less well-understood than what we know today.

3. SISAL

SISAL is an algorithmic realization of simplex volume minimization
(SVMin), an idea first conceived by Craig in HU in 1994 [9]. SVMin
was an intuition that said that the endmembers can be identified by
finding a minimum-volume simplex that encloses all the hyperspec-
tral pixels. This intuition was empirically found to be valid in later
studies, even for heavily mixed pixel (and no-pure-pixel) instances;
and it is recently confirmed to be true in theory, under some assump-
tions [10, 11]. It is now commonly accepted that SVMin can be
mathematically described by an optimization problem

$$\min \text{vol}(A) := (\det(\bar{A}^\top \bar{A}))^{1/2}/(N - 1)!$$

$$s.t. \quad y_t \in \text{conv}(A) := \{y = As \mid s \in \mathcal{U}\}, \quad t = 1, \ldots, T \quad (5)$$

where $\bar{A} := [a_1 - a_N, \ldots, a_{N-1} - a_N]$, and the solution to
problem (5) serves as the endmembers estimate. Here, $\text{conv}(A)$
denotes the convex hull of the set of points $a_1, \ldots, a_N$, which is a
simplex under the assumption of affinely independent $a_i, \ldots, a_N$;
$\text{vol}(A)$ is the volume of the simplex $\text{conv}(A)$. There was an issue
back in the 2000’s—from Craig’s 1994 paper it was not clear how
the SVMin problem was solved, precisely.

3.1. Formulation

Bioucas-Dias was the among the first who seriously studied the
realization of the SVMin problem (5).† Let us first describe the formu-
lation [12]. Consider the following: i) $M = N$ such that $A$ is square;
ii) replace $\det(\bar{A}^\top \bar{A})$ in (5) by $\det(\bar{A}^\top \bar{A}) = |\det(A)|^2$. Then we
can formulate the SVMin problem as

$$\min_{A, s} |\det(A)| \quad s.t. \quad Y = AS, \quad S \geq 0, \quad S^\top 1 = 1. \quad (6)$$

By a change of variable $B = A^{-1}$, we can transform (6) as

$$\max_B |\det(B)| \quad s.t. \quad BY \geq 0, \quad B^\top 1 = (Y^\top)^\dagger 1, \quad (7)$$

where the superscript $\dagger$ denotes the pseudo-inverse. The equivalence
of problems (6) and (7) is shown in Section 6.1.‡ Problem (7) is
easier to handle than problem (6) because the former’s constraints

†The author’s team happened to begin to investigate SVMin around the
same time as Bioucas-Dias’ team, and the co-occurrence of the two indepen-
dent research led the two sides to know each other in 2009 WHISPERS. The
author is grateful to Tsung-Han Chan, the key member of the author’s team
back then. He made bold attempts to study HU—which was then unknown
to the team—and triggered the team’s interest in SVMin.
‡Bioucas-Dias wrote down the problem transformation (7) very concisely
(see [12]). The author has been long wondering if a mathematically precise
proof on the equivalence of the transformation can be provided.

are convex. However the constraint $BY \geq 0$ is a number of non-
separable inequality constraints, and their presence poses limitations
to the development of computationally efficient schemes for (7).
In that regard we should mention that $T$, the number of pixels, is large,
and we are dealing with an optimization problem that has numerous
non-separable inequality constraints. As a compromise, Bioucas-
Dias turned to a soft-constrained variant of (7)

$$\min_B f(B) + \lambda h(BY) \quad s.t. \quad B^\top 1 = (Y^\top)^\dagger 1, \quad (8)$$

where $\lambda > 0$ is given:

$$f(B) := -\log(|\det(B)|); \quad h(X) := \sum_{i,j} \max\{-x_{ij}, 0\}. \quad$$

Here, $h(X)$ is an element-wise hinge function, serving as a penalizer
to discourage $X$ from having negative elements; the incorporation
of log on $|\det(B)|$, which is w.l.o.g., is to make the problem numerically better to tackle—intuitively.

3.2. Optimization

We now turn to the optimization. In SISAL, successive convex opti-
mization is adopted to tackle problem (8), specifically,

$$B^{k+1} \in \arg\min_B g_k(B) + \lambda h(BY) \quad s.t. \quad B^\top 1 = (Y^\top)^\dagger 1 \quad (9)$$

for $k = 0, 1, 2, \ldots$. Here,

$$g_k(B) := f(B_k) + (\nabla f(B_k), B - B_k) + \mu_k ||B - B_k||^2$$

is a quadratic approximation of $f$ at $B_k$, where $\mu_k > 0$; $\nabla f$ is
the gradient of $f$; $\cdot || \cdot$ is the Euclidean norm. Knowledge readers
on optimization may notice that (9) resembles the proximal gradient
method (see, e.g., [13]), although they should also be warned that $f$
does not have Lipschitz continuous gradient, a key assumption with the
use of proximal gradient. The question that remains is to solve the
convex problems in (9). Bioucas-Dias devised a specialized
algorithm for (9) via the variable splitting augmented Lagrangian
method (which is the same as the alternating direction method
of multipliers). The algorithm exploits the problem structure and is
computationally very efficient.

The iterations in (9) is the basic form of SISAL. The actual algo-
rithm is a modification that resembles the gradient projection method
with the limited minimization step-size rule [14, Ch. 2.3].
SISAL shows Bioucas-Dias’ strong application insights with
what we call non-convex large-scale optimization today. Bioucas-
Dias did so in 2009, well ahead of the blooming of the topic in signal
and image processing, machine learning, data science, etc.

3.3. Some Well-Known Advantages of SISAL

SISAL generally runs faster than other SVMin state-of-the-art
schemes, such as those that tackle the hard-constrained SVMin
problems (such as (7)); this is particularly so when $T$ is very large.
SISAL is known to be robust to noise and outliers. In that regard
we should first point out that the SVMin intuition, as well as the for-
mulations (5)–(7), were established on the case of noiseless data. In
the noisy case, using hard constraints to enforce enclosing of all the
pixels can result in sensitivity issues, subsequently causing poor esti-
mates of the endmembers. Using soft constraints allows some pix-
els, particularly the adversarial ones, to lie outside the simplex, and
that tolerance can be beneficial in mitigating the sensitivity effects.
The only issue that is not easy to answer is the selection of $\lambda$ in (8); i.e.,
should we suppress volume more, or encourage non-negativity
more? Usually, the parameter $\lambda$ is manually chosen.
3.4. Further Discussion

In 2014, Guangzhou, Bioucas-Dias and the author discussed the SVMin identifiability which was solved in the noiseless case [10,11]. Bioucas-Dias challenged the author with this difficult question.

Question 1 Ill-conditioned $A$ can cause serious noise sensitivity.

This issue is valid in practice, and noise sensitivity analysis for SVMin is still an unsolved problem. But recently the author has a different way to answer this question, which will be explained later. In 2017, they discussed online the following problem.

Question 2 Is minimizing $\text{vol}(A) \propto (\det(A^T A))^{1/2}$ over data-enclosing constraints, i.e., the SVMin problem (5), the same as the counterpart of minimizing $(\det(A^T A))^{1/2}$, as in (6)?

While $\det(A^T A)$ and $\det(A^T A)$ are similar, it does not mean that they lead to the same minimization result. They studied this aspect and the answer is yes. The proof is shown in Section 6.2.

4. DECA

DECA pursues a probabilistic paradigm for HU. It is different from pure-pixel search and SVMin right from the onset. But researchers may not know that DECA can be viewed as a more powerful form of SVMin, subsuming SVMin as a special case. The author will describe DECA using his derivations [15]. Consider a noisy model

$$y_t = As_t + v_t,$$

where $v_t$ is noise. Assume the following: i) $A$ has affinely independent columns; ii) all the $s_t$'s and $v_t$'s are independent; iii) every $v_t$ follows a Gaussian distribution with mean zero and covariance $\sigma^2 I$; iv) every $s_t$ follows a Dirichlet mixture distribution, whose probability density function (PDF) is given by

$$p(s; \gamma, \alpha_1, \ldots, \alpha_K) = \sum_{k=1}^{K} \gamma_k D(s; \alpha_k)$$

for some parameters $\gamma > 0$, $\gamma \in 1$, $\alpha_1, \ldots, \alpha_K > 0$; $D(s; \alpha)$ denotes the Dirichlet distribution with concentration parameter $\alpha$.

The use of the Dirichlet mixture prior is to provide a general model for accommodating complex phenomena in real-world hyperspectral data. The problem is to estimate $A$, together with the unknown prior parameters $\gamma$ and $\alpha_k$'s, by maximum-likelihood (ML) inference:

$$\theta_{ML} \in \arg \max_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^{T} \log p(y_t; \theta)$$

where $\Theta = \{A, \gamma, \alpha_1, \ldots, \alpha_K\}$; $\Theta$ is the domain of $\theta$; $p(y_t; \theta)$ is the PDF of $y_t$ parameterized by $\theta$ and is given by

$$p(y_t; \theta) = \int p(y_t; A)p(s; \gamma, \alpha_1, \ldots, \alpha_K)ds$$

$$= \sum_{k=1}^{K} \gamma_k \varphi_\alpha(y - As)D(s; \alpha_k)ds$$

in which $\varphi_\alpha(y) = e^{-\|y\|^2/2\sigma^2}/(\sqrt{2\pi}\sigma)^N$, and $\mu$ is the Lebesgue measure on $\{s \in \mathbb{R}^N | 1^T s = 1\}$. In statistical inference, ML estimation is known to have benign traits of properties, such as better estimation accuracy for larger $T$, under some assumptions.

DECA considers the noiseless case, for which the integrals in (11) have closed form, and uses expectation maximization (EM) to realize the ML estimator (10). Instead of describing the EM algorithm, the author wants to draw connection. Let us simplify by adopting the uniform prior model, i.e., $p(s; \gamma, \alpha_1, \ldots, \alpha_K) = D(s; 1)$. Also, assume $M = N - 1$. Then, as shown in [15], the log likelihood for affinely independent $A$ can be expressed as

$$\log p(y; \theta) = -\log \text{vol}(A) + \log \left(\int_{\mathbb{R}^{N-1}} \varphi_\alpha(y - x)\|A^T x\|dx\right),$$

where $\|A^T x\| = 1$ if $x \in X$; $1(x) = 0$ if $x \notin X$; $A = \text{conv}(A)$; $A$ is the relative interior of $A$. The ML problem (10) then reduces to

$$\hat{A}_{ML} \in \arg \min_{A} \log \text{vol}(A) - \frac{1}{T} \sum_{t=1}^{T} r(A, y_t).$$

Problem (12) appears as a soft-constrained SVMin: as $r(A, y_t)$ is large (respectively, small) for $y_t$ lying well inside (respectively, far away) the simplex $A$, it serves a penalizer for outside-the-simplex points in (12). In fact, for the noiseless case $\sigma^2 = 0$ and the ML estimator (12) becomes the SVMin problem (5)! This identity was informally mentioned by Bioucas-Dias in his WHISPERS 2009 presentation. It was also alluded to in his paper [7], though not apparent. In this connection, it is worth recognizing that Dobigeon et al. described a similar result on the above-noted identity around the same time [16, Appendix], although it is also not apparent. The author’s latest study [15] further reveals how SISAL can roughly be seen as an approximation of the ML estimator (12) in the noisy case; the details are omitted. The author also shows that

Theorem 1 Consider $T \to \infty$. The ML estimator (12) can lead to exact identification of the true $A$. The result holds for the general noisy case.

The above result is vital in confirming the strength of ML. It is worth noting that the DECA paper [7] provided an intuitive justification (but not a proof) on the same result. Theorem 1 also gives an impression that the noise effects should be reduced as $T$ increases.

The author now gives a partial answer to Question 1: If we have a large number of pixels (also known as big data), we may mitigate the impact of noise by employing the ML estimator, or a good approximation of it by soft-constrained SVMin.

5. CLOSING REMARK

The author wants to express his very heartfelt gratitude to Bioucas-Dias for his many inspirations, challenges and encouragements, which led him to work on interesting problems.

6. APPENDIX

6.1. Proof of Equivalence of Problems (6) and (7)

Let $(A, S)$ be any feasible point of (6). Redefine the true $(A, S)$ in the data model (1) as $(A_0, S_0)$. Since $A_0, S_0 = Y = AS$, and $A_0$ and $S_0$ have full column rank, one can show that $A$ and $S$.

3The author favors $\text{vol}(A)$ zealously because it is the true simplex volume. Bioucas-Dias prefers $(\det(A^T A))^{1/2}$ because it is simpler.

4The author has been deeply intrigued by that since then.
have full column rank. Consider $M = N$. Let $B = A^{-1}$. We have $Y = AS \iff BY = S$. Applying the above to $S^\top 1 = 1$ yields $Y^\top B^\top 1 = 1 \implies B^\top 1 = (Y^\top)^{-1} 1$. (13)

One would be tempted to think that the converse of (13) $Y^\top B^\top 1 = 1 \iff B^\top 1 = (Y^\top)^{-1} 1$ (14) is also true, but it is not true if we see the problem as a generic matrix analysis problem. But (14) can be shown by incorporating $Y = A_0S_0$. To put into context, consider finding an $x$ such that $Y^\top x = 1$. (15)

If a solution to (15) exists, then $x$ is uniquely given by $x = (Y^\top)^{-1} 1$. Also, as a consequence, $x = (Y^\top)^{-1} 1 \implies Y^\top x = 1$ will be true. One can verify that $x = A_0^{-1} 1$ satisfies (15) (we need $S_0^{-1} 1 = 1$). Thus, $x = (Y^\top)^{-1} 1 \implies Y^\top x = 1$ is true, and (14) is also true.

Using the above results, we can equivalently transform (6) to (7), with an extra constraint that $B$ is invertible. We can discard the inversed constraint from (7) w.l.o.g., as $\det(B) = 0$ for non-invertible $B$. Let us conclude: Under the noiseless data model (1) and the assumptions thereof, problems (6) and (7) are equivalent.

6.2. Equivalence of Minimizing $\det(A^\top A)$ and $\det(A^\top A)$

We use the same convention as above: $(A_0, S_0)$ is the ground truth, and $(A, S)$ is any feasible point of (6). Let $F = [I - 1]^\top \in \mathbb{R}^{N \times (N-1)}$, $G = [F \cdot 1^\top] \in \mathbb{R}^{N \times N}$. It can be shown that $\det(G) = 1$. By letting $b = \frac{1}{N} A1$, and noting $A = AF$.

\[
\det(A^\top A) = |\det(G)|^2 \cdot \det(A^\top A) = \det(G^\top A^\top AG)
\]

(16a)

\[
= \det \left( \frac{A^\top A}{b^\top A} \cdot \left| \frac{b}{b^\top A} \right|^2 \right)
\]

(16b)

\[
= \det(A^\top A) \cdot \left( ||b||^2 - b^\top A(A^\top A)^{-1} A^\top b \right)
\]

(16c)

where (16b) is due to Schur’s determinant identity; $A^\top A$ is invertible because $A$ has full column rank (as shown in Section 6.1). If $\min ||b - A\bar{z}||^2$ is a constant $\gamma > 0$ irrespective of $A$, then we will have the desired result $\det(A^\top A) = C \det(A^\top A)$. To that end, denote $\mathcal{A}(A) := \{y = A\bar{z} \mid x^\top 1 = 1\}$. It can be verified that $\mathcal{A}(A) = \mathcal{A}(A_0)$ [17]. Let $\mathcal{S}(A)$ denote the span of the columns of $A$. Consider the following.

**Lemma 1** Let $A_0 \in \mathbb{R}^{M \times N}$ be any matrix, let $\hat{A}_0 = A_0 F_0$ and let $U \in \mathbb{R}^{N \times (N-1)}$ be a semi-orthogonal matrix such that $U^\top 1 = 0$.

(a) $\mathcal{A}(A_0) = \text{span}(A_0) + d$ for any $d \in \mathcal{A}(A_0)$

(b) $\mathcal{S}(A_0U) = \text{span}(A_0U)$

(c) $\{x \in \mathbb{R}^N \mid x^\top 1 = 1\} = \text{span}(U) + \frac{1}{N} 1$

The proof of Lemma 1 can be found in [15]. Choose $d = \frac{1}{N} A_01$. From $\mathcal{A}(A) = \mathcal{A}(A_0)$, we have $\mathcal{S}(A) = \text{span}(A_0U)$. Also, since $b \in \mathcal{A}(A)$ we can write $b = A_0U\bar{z} + d$ for some $\bar{z}$. Then we have

\[
\min ||b - A\bar{z}||^2 = \min ||b - A_0U\bar{z}\|^2;
\]

(17)

\[
\leq \min \left[ ||A_0(U(\bar{z} - \bar{z} + \frac{1}{N} 1))||^2 + \frac{1}{N} 1 \right]^2 = \min \frac{1}{N} 1 \cdot ||A_0\bar{z}||^2 > 0;
\]

(recall that $A_0$ has full column rank). The proof is done. To summarize, the answer to Question 2 is yes under the noiseless data model (1) and the assumptions thereof.

7. **References**

[1] 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. Sel. Topics Appl. Earth Observ., vol. 5, no. 2, pp. 354–379, 2012.

[2] W. K. Ma, J. M. Bioucas-Dias, T. H. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, and C. Y. Chi, “A signal processing perspective on hyperspectral unmixing,” IEEE Signal Process. Mag., vol. 31, no. 1, pp. 67–81, Jan 2014.

[3] X. Fu, K. Huang, N. D. Sidiropoulos, and W.-K. Ma, “Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications,” IEEE Signal Process. Mag., vol. 36, no. 2, pp. 59–89, 2019.

[4] N. Gillis, Nonnegative Matrix Factorization. SIAM, 2021.

[5] J. Nascimento and J. 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.

[6] J. Bioucas-Dias, “A variable splitting augmented Lagrangian approach to linear spectral unmixing,” in Proc. 2009 First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, Aug. 2009.

[7] J. Nascimento and J. Bioucas-Dias, “Hyperspectral unmixing based on mixtures of Dirichlet components,” IEEE Trans. Geosci. Remote Sens., vol. 50, no. 3, pp. 863–878, 2012.

[8] J. Boardman, F. Kruse, and R. Green, “Mapping target signatures via partial unmixing of AVIRIS data,” in Proc. Summ. JPL Airborne Earth Sci. Workshop, vol. 1, Pasadena, CA, Dec. 1995, pp. 23–26.

[9] M. D. Craig, “Minimum-volume transforms for remotely sensed data,” IEEE Trans. Geosci. Remote Sens., vol. 32, no. 3, pp. 542–552, May 1994.

[10] C.-H. Lin, W.-K. Ma, W.-C. Li, C.-Y. Chi, and A. Ambikapathi, “Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: The no-pure-pixel case,” IEEE Trans. Geosci. Remote Sens., vol. 53, no. 10, pp. 5530–5546, Oct. 2015.

[11] X. Fu, W.-K. Ma, K. Huang, and N. D. Sidiropoulos, “Blind separation of quasi-stationary sources: Exploiting convex geometry in covariance domain,” IEEE Trans. Signal Process., vol. 63, no. 9, pp. 2306–2320, 2015.

[12] J. Li and J. Bioucas-Dias, “Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data,” in Proc. IEEE IGARSS, Aug. 2008.

[13] A. Beck, First-Order Methods in Optimization. Philadelphia, PA, USA: SIAM, 2017, vol. 25.

[14] D. P. Bertsekas, Nonlinear Programming. Athena Scientific, 1999.

[15] R. Wu, W.-K. Ma, Y. Li, A. M.-C. So, and N. D. Sidiropoulos, “Probabilistic simplex component analysis,” submitted to IEEE Trans. Signal Process., Mar. 2021, online available: https://arxiv.org/pdf/2103.10027.pdf.

[16] N. Dobigeon, S. Moussaouai, M. Coulon, J.-Y. Tourneret, and A. O. Hero, “Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery,” IEEE Trans. Signal Process., vol. 57, no. 11, pp. 4355–4368, 2009.

[17] 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 Process., vol. 57, no. 11, pp. 4418–4432, 2009.