ROBUST VOLUME MINIMIZATION-BASED MATRIX FACTORIZATION VIA ALTERNATING OPTIMIZATION

Xiao Fu, Wing-Kin Ma†, Kejun Huang*, and Nicholas D. Sidiropoulos**

*Department of ECE, University of Minnesota, Minneapolis, MN, USA
†Department of EE, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

ABSTRACT

This paper focuses on *volume minimization* (VolMin)-based structured matrix factorization (SMF), which factors a data matrix into a full-column rank basis and a coefficient matrix whose columns reside in the unit simplex. The VolMin criterion achieves this goal via finding a minimum-volume enclosing convex hull of the data. Recent works showed that VolMin guarantees the identifiability of the factor matrices under mild and realistic conditions, which suit many applications in signal processing and machine learning. However, the existing VolMin algorithms are sensitive to outliers or lack efficiency in dealing with volume-associated cost functions. In this work, we propose a new VolMin-based matrix factorization criterion and algorithm that take outliers into consideration. The proposed algorithm detects outliers and suppress them automatically, and it does so in an algorithmically very simple way. Simulations are used to showcase the effectiveness of the proposed algorithm.

Index Terms— Volume minimization, nonnegative matrix factorization, document clustering, hyperspectral unmixing

1. INTRODUCTION

Motivated by the influential paper of Lee and Seung [1], structured matrix factorizations (SMF) such as nonnegative matrix factorization (NMF) have drawn much attention, since they are capable of not only reducing dimensionality of the collected data, but also retrieving loading factors that have physically meaningful interpretations. In addition to NMF, some related types of SMF have attracted considerable interest in recent years. The remote sensing community has spent much effort on a class of factorizations where the columns of one factor matrix are constrained to lie in the unit simplex [2,3]. The same SMF has also been utilized for document clustering [4], and, most recently, array processing and wireless communications [5–7].

The first important question related to SMF lies in identifiability — when does a factorization model or criterion admit unique solution in terms of its factors? In recent years, identifiability conditions have been investigated under the NMF model [8–10]. An undesirable property of NMF highlighted in [10] is that identifiability hinges on an assumption that both the loading factors contain a certain number of zeros. In many applications, however, there is at least one factor that is dense. For example, in a typical matrix factorization application, namely, hyperspectral unmixing (HU), the basis factor (i.e., the spectral signature matrix) is always a dense matrix. On the other hand, recent works [5, 11] showed that the SMF model with the coefficient matrix columns lying in the unit simplex admits much more relaxed identifiability conditions. Specifically, [5, 11] proved that, under some realistic conditions, unique loading factors (up to column permutations) can be obtained by finding a minimum-volume enclosing convex hull of the data vectors. Notably, the identifiability conditions of the so-called *volume minimization* (VolMin) criterion enable us to work with dense (and even complex or negative) basis matrix factors. Since the NMF model can be recast into (viewed as a special case of) the above SMF model [12], such results suggest that VolMin is an attractive alternative to NMF for the wide range of applications of NMF.

To apply VolMin in practice, there are two major challenges. First, dealing with the VolMin cost function is computationally complicated. Prior work in [13] and [14] proposed (log-)determinant minimization-based algorithms, but they work under a noiseless setup and require dimensionality reduction (DR) pre-processing; see also the variants in [15, 16]. Another major branch of work [17, 18] formulated the problem as a volume minimization-regularized data fitting problem. Such formulations are arguably more robust against noise and do not need pre-processing, but are computationally harder to deal with. The second major challenge of VolMin is outlier-sensitivity: It has been noticed in the literature that even a single outlier can make the VolMin criterion fail [3]. An outlier-robust VolMin algorithm has been considered in [19]; it also works in the reduced-dimension domain, and DR itself can be sensitive to outliers.

In this work, we are interested in the volume-regularized fitting-based formulation as in [17, 18], but we take outliers into account. Specifically, we impose an outlier-robust loss function onto the data fitting part, and propose a modified log-determinant loss function as the volume regularizer. By majorizing both functions, the fitting and the volume-regularization terms can be taken care of in a refreshingly easy way. Consequently, a simple three-block alternating optimization algorithm is derived. In addition, we show that every limit point of the solution sequence produced by the algorithm is a stationary point. Simulations are used to showcase the efficacy of our algorithm.

2. VOLMIN-BASED MATRIX FACTORIZATION

Consider the following signal model:

$$\boldsymbol{x}[\ell] = \boldsymbol{A}\boldsymbol{s}[\ell] + \boldsymbol{v}[\ell], \quad \ell = 1, \ldots, L, \tag{1}$$

where $\boldsymbol{x}[\ell] \in \mathbb{R}^M$ denotes the ℓth measured data vector, $\boldsymbol{v}[\ell] \in \mathbb{R}^M$ denotes the corresponding noise vector, $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ denotes an unknown measuring matrix (or a basis), and the coefficient vector $\boldsymbol{s}[\ell] = [s_1[\ell], \ldots, s_N[\ell]]^T \in \mathbb{R}^N$ satisfies

$$\boldsymbol{s}[\ell] \geq \boldsymbol{0}, \; \boldsymbol{1}^T\boldsymbol{s}[\ell] = 1, \tag{2}$$

*E-mail: (xfu,nikos,huang663)@umn.edu
†E-mail: wkma@ee.cuhk.edu.hk

This work was supported in part by ARO STIR W911NF-15-1-0384, in part by by NSF IIS 1247632, and in part by the Hong Kong Research Grants Council (RGC) General Research Fund (GRF), under Project CUHK 14205414.

where $\mathbf{1}$ denotes an all-one vector of appropriate length, and the notation '$\boldsymbol{x} \geq \mathbf{0}$' means that $\boldsymbol{x}$ is element-wise nonnegative. Note that $M \geq N$ is assumed throughout this paper. The signal model in (1)-(2) appears in many applications. In text mining, $\boldsymbol{x}[\ell]$ represents a document, $\boldsymbol{a}_n$ for $n = 1, \ldots, N$ (i.e., the columns of $\boldsymbol{A}$) represent the topics that appear in different documents, and $s_n[\ell]$ denotes the weight of topic n in document ℓ. In hyperspectral unmixing, $\boldsymbol{x}[\ell]$ is a high-dimensional pixel, the $\boldsymbol{a}_n$'s are spectral signatures of the materials contained in the image, and $s_n[\ell]$ denotes the proportion of material n in pixel ℓ. Please see [5, 6, 20–22] for more applications. Notice that the NMF model can also be converted into the model in (1)-(2), with proper normalization [12]. Thus, the applications of NMF can also be considered under the model in (1)-(2).

The task of SMF is to recover the factors $\boldsymbol{A}$ and/or $\boldsymbol{S} = [\boldsymbol{s}[1], \ldots, \boldsymbol{s}[L]]$ from the data matrix $\boldsymbol{X} = [\boldsymbol{x}[1], \ldots, \boldsymbol{x}[L]]$, under the assumption that N is known or has been previously estimated. The VolMin criterion is motivated by the underlying convex geometry of (1)-(2). Specifically, in the noiseless case, every $\boldsymbol{x}[\ell]$ lives in a *convex hull* spanned by $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$, i.e., $\boldsymbol{x}[\ell] \in \operatorname{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$ [23]. If $\operatorname{rank}(\boldsymbol{A}) = N$, then $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ are exactly the vertices of the convex hull; see Fig. 1 (left) for an illustration. Therefore, recovering $\boldsymbol{A}$ amounts to finding these vertices. VolMin is proposed to accomplish this task via identifying a minimum-volume enclosing convex hull. The intuition of VolMin is illustrated in Fig. 1 (left), where we see that if the data points are sufficiently spread on $\operatorname{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$, then the minimum-volume enclosing convex hull is identical the desired convex hull. Formally, the VolMin criterion for the noise-free case is as follows [5]:

$$
\begin{aligned}
\min_{\boldsymbol{A} \in \mathbb{R}^{M \times N}, \boldsymbol{S} \in \mathbb{R}^{N \times L}} \; & \det(\boldsymbol{A}^T \boldsymbol{A}) \\
\text{s.t. } & \boldsymbol{x}[\ell] = \boldsymbol{A}\boldsymbol{s}[\ell], \; \forall \ell, \\
& \boldsymbol{s}[\ell] \geq \mathbf{0}, \; \mathbf{1}^T \boldsymbol{s}[\ell] = 1, \; \forall \ell.
\end{aligned} \tag{3}
$$

where $\det(\boldsymbol{A}^T\boldsymbol{A})$ is a measure of $\operatorname{vol}(\boldsymbol{A})$, i.e., the volume of the convex hull $\operatorname{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$ [1]. Identifiability of the formulation above is proved in [5], thus formally verifying the intuition in Fig. 1 (left). Simply speaking, if $\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{S}) = N$, and if $\boldsymbol{s}[\ell]$'s are sufficiently spread in the unit simplex (so that $\boldsymbol{x}[\ell]$'s are sufficiently spread in $\operatorname{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$), then using VolMin guarantees the identifiability of $\boldsymbol{A}$ and $\boldsymbol{S}$ up to column permutation; see also [11] for similar results. Unlike NMF, where $\boldsymbol{A}$ has to contain some zero entries to guarantee identifiability [10], no specific restriction on $\boldsymbol{A}$ is needed when using the VolMin criterion.

The VolMin criterion is challenging to solve. The works in [13, 14, 19] assumed that $\boldsymbol{A}$ is square and let $\operatorname{vol}(\boldsymbol{A}) = \det(\boldsymbol{A})$. Alternating optimization and successive linearization were employed, respectively. The drawback with the aforementioned works is that noise is not taken into consideration. Plus, these approaches require dimensionality reduction (DR) in a pre-processing stage to make the effective $\boldsymbol{A}$ square – but DR may not be reliable in the presence of outliers or modeling errors. Another major branch of algorithms considers [17, 18]

$$
\begin{aligned}
\min_{\boldsymbol{A} \in \mathbb{R}^{M \times N}, \boldsymbol{S} \in \mathbb{R}^{N \times L}} \; & \|\boldsymbol{X} - \boldsymbol{A}\boldsymbol{S}\|_F^2 + \lambda \cdot \operatorname{vol}(\boldsymbol{A}) \\
\text{s.t. } & \boldsymbol{S} \geq \mathbf{0}, \; \mathbf{1}^T \boldsymbol{S} = \mathbf{1}^T,
\end{aligned} \tag{4}
$$

[1] Note that in the literature such as [17,18,24,25], the definition of $\operatorname{vol}(\boldsymbol{A})$ may be slightly different from one work to another although they all employ determinant.

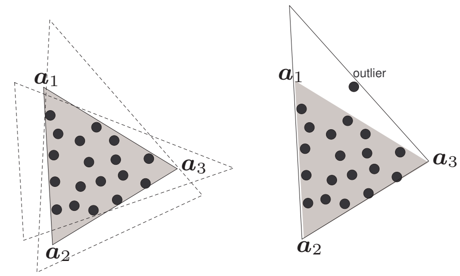

Fig. 1: Left: The intuition of VolMin for an $N = 3$ case. The dots are $\boldsymbol{x}[\ell]$'s; the shaded area is $\operatorname{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$, the triangles with dashed lines are data-enclosing convex hulls, and the one with solid lines is the minimum-volume enclosing convex hull. Right: the impact of outliers.

where $\lambda > 0$ is a parameter that balances data fidelity versus volume minimization. The formulation in (4) avoids DR and takes noise into consideration, but the algorithms are less efficient, since volume-minimization penalties like $\det(\boldsymbol{A}^T\boldsymbol{A})$ are hard to cope with.

Another notable difficulty is that outliers are very damaging to VolMin. In real-world applications, outlying measurements are commonly seen. In HU, 'dead pixels' or pixels that do not obey the linear model in (1) are frequently spotted; and in document clustering, articles not belonging to any known category may exist. In many cases, a single outlier can make the minimum-volume enclosing convex hull very different from the desired one; see Fig. 1 (right) for an illustration. The only work that considers outliers for the VolMin-based factorization is [19], but it takes care of outliers in the dimension-reduced domain. As already mentioned, the DR process itself may be impaired by outliers, and thus dealing with outliers in the original data domain is more appealing.

3. PROPOSED ROBUST VOLMIN ALGORITHM

We are interested in the VolMin-regularized matrix factorization, but we take the outlier problem into consideration. Specifically, we propose to employ the following optimization criterion:

$$
\begin{aligned}
\min_{\boldsymbol{A}, \boldsymbol{S}} \; & \sum_{\ell=1}^{L} \frac{1}{2} \left(\|\boldsymbol{x}[\ell] - \boldsymbol{A}\boldsymbol{s}[\ell]\|_2^2 + \epsilon \right)^{\frac{p}{2}} + \lambda \log \det(\boldsymbol{A}^T\boldsymbol{A} + \epsilon \boldsymbol{I}) \\
\text{s.t. } & \mathbf{1}^T \boldsymbol{s}[\ell] = 1, \; \boldsymbol{s}[\ell] \geq \mathbf{0}, \; \forall \ell,
\end{aligned} \tag{5}
$$

where $0 < p \leq 1$, and $\lambda, \epsilon > 0$. Here, ϵ is a small regularization parameter, which prevents the second term in the cost function from being unbounded; it also keeps the first term smooth, which is desirable from a computation viewpoint. The p-(quasi-) norm-like data fitting criterion is employed to suppress the impact of the outliers: compared to the commonly used Frobenius norm-based fitting criterion (cf. Eq (4)), it is less sensitive to large fitting errors and tends to ignore the associated data points. Notice that we use $\log\det(\cdot)$ instead of $\det(\cdot)$ as in [17, 18, 26] for volume regularization. Log-determinant-based volume minimization has been used in [19, 27], where it has a maximum-likelihood interpretation when $\boldsymbol{s}[\ell]$ follows the uniform Dirichlet distribution. Also, using $\log\det(\cdot)$ will prove handy in our algorithm development.

To deal with the above problem, let us introduce two lemmas:

Lemma 1 *[28] Assume $0 < p < 2$, $\epsilon \geq 0$, and $\phi_p(w) := \frac{2-p}{2}\left(\frac{2}{p}w\right)^{\frac{p}{p-2}} + \epsilon w$. Then, we have $\left(x^2 + \epsilon\right)^{p/2} = \min_{w \geq 0} \; wx^2 +$*

$\phi_p(w)$, *where the solution is uniquely given by* $w_{\rm opt} = \frac{p}{2}\left(x^2+\epsilon\right)^{\frac{p-2}{2}}$.

Lemma 2 *[29] Let* $\boldsymbol{E} \in \mathbb{R}^{N\times N}$ *be any matrix such that* $\boldsymbol{E} \succ \boldsymbol{0}$. *Consider the function* $f(\boldsymbol{F}) = {\rm Tr}\left(\boldsymbol{F}\boldsymbol{E}\right) - \log\det \boldsymbol{F} - K$. *Then, we have* $\log\det\boldsymbol{E} = \min_{\boldsymbol{F}\succeq\boldsymbol{0}}\ f(\boldsymbol{F})$, *and the solution is uniquely given by* $\boldsymbol{F}_{\rm opt} = \boldsymbol{E}^{-1}$.

The above two lemmas provide two functions to 'majorize' both the data fitting term and the volume-regularization term in (5), respectively. Consequently, Problem (5) can be recast as:

$$\begin{aligned}\min_{\boldsymbol{A},\boldsymbol{S},\{w_\ell\},\boldsymbol{F}}\ & \sum_{\ell=1}^{L}\frac{w_\ell}{2}\|\boldsymbol{x}[\ell]-\boldsymbol{A}\boldsymbol{s}[\ell]\|_2^2 + \sum_{\ell=1}^{L}\phi_p(w_\ell) \\ & + \lambda\left({\rm Tr}(\boldsymbol{F}(\boldsymbol{A}^T\boldsymbol{A}+\epsilon\boldsymbol{I})) - \log\det\boldsymbol{F}\right) \\ {\rm s.t.}\ & \boldsymbol{1}^T\boldsymbol{s}[\ell]=1,\ \boldsymbol{s}[\ell]\geq\boldsymbol{0},\ \forall\ell \\ & \boldsymbol{F}\succeq\boldsymbol{0},\ w_\ell\geq 0,\ \forall\ell.\end{aligned} \tag{6}$$

The reformulation in Problem (6) opens a door for handling the problem of interest in an easy way, since the partial minimizations w.r.t. $\boldsymbol{A}$, $\boldsymbol{S}$, $\{w_\ell\}_{\ell=1}^{L}$ and $\boldsymbol{F}$ are all convex problems.

Let us consider the solutions to the aforementioned conditional minimizations. The conditional minimization w.r.t. $\boldsymbol{A}$ is an unconstrained quadratic program, and the solution is

$$\boldsymbol{A} := \boldsymbol{X}\boldsymbol{W}\boldsymbol{S}^T\left(\boldsymbol{S}\boldsymbol{W}\boldsymbol{S}^T+\lambda\boldsymbol{F}\right)^{-1}, \tag{7}$$

where $\boldsymbol{W} = {\rm Diag}(w_1,\ldots,w_L)$. The conditional minimization w.r.t. $\boldsymbol{S}$ is separable w.r.t. its columns $\boldsymbol{s}[\ell]$; specifically, $\boldsymbol{s}[\ell]$ for $\ell=1,\ldots,L$ is updated by solving the following:

$$\boldsymbol{s}[\ell] := \arg\min_{\boldsymbol{1}^T\boldsymbol{s}[\ell]=1,\ \boldsymbol{s}[\ell]\geq\boldsymbol{0}}\ \|\boldsymbol{x}[\ell]-\boldsymbol{A}\boldsymbol{s}[\ell]\|_2^2. \tag{8}$$

Problem (8) is convex and there is one more than one way to solve it. Here, we employ *alternating direction method of multipliers* (ADMM) [30] to solve it. The idea of ADMM is to first rewrite the Problem (8) as

$$\begin{aligned}\min_{\boldsymbol{s}[\ell],\boldsymbol{z}[\ell]}\ & \|\boldsymbol{x}[\ell]-\boldsymbol{A}\boldsymbol{z}[\ell]\|_2^2 \\ {\rm s.t.}\ & \boldsymbol{1}^T\boldsymbol{z}[\ell]=1,\ \boldsymbol{s}[\ell]\geq\boldsymbol{0},\ \boldsymbol{z}[\ell]=\boldsymbol{s}[\ell],\end{aligned} \tag{9}$$

and then solve it using an algorithm that alternately updates $\boldsymbol{s}[\ell]$, $\boldsymbol{z}[\ell]$ and a dual variable; see Algorithm 2 for details. Notice that each update of the ADMM algorithm has a closed-form solution, and thus can be carried out easily.

The conditional minimizations w.r.t. $\boldsymbol{W}$ and $\boldsymbol{F}$ are straightforward by Lemma 1 and Lemma 2; we have

$$w_\ell := \frac{p}{2}\left(\|\boldsymbol{x}[\ell]-\boldsymbol{A}\boldsymbol{s}[\ell]\|_2^2+\epsilon\right)^{\frac{p-2}{2}}, \tag{10}$$

$$\boldsymbol{F} := \left(\boldsymbol{A}^T\boldsymbol{A}+\epsilon\boldsymbol{I}\right)^{-1}. \tag{11}$$

The proposed algorithm updates $\{\boldsymbol{A},\boldsymbol{S},\boldsymbol{W},\boldsymbol{F}\}$ cyclically, until a stopping criterion is satisfied. Notice that $\boldsymbol{W}$ and $\boldsymbol{F}$ can be updated simultaneously. Hence, the algorithm, which is summarized in Algorithm 1, can be considered as three-block alternating optimization algorithm. We show that

Proposition 1 *Let* $\{\boldsymbol{A}^*,\boldsymbol{S}^*,\boldsymbol{F}^*,\boldsymbol{W}^*\}$ *be a limit point of the solution sequence produced by the proposed algorithm. Then,* $\{\boldsymbol{A}^*,\boldsymbol{S}^*\}$ *is a stationary point of Problem* (5).

The proof is omitted due to space limitation. According to Proposition 1, although we have been dealing with Problem (5) indirectly, convergence to a stationary point of Problem (5) is guaranteed.

Algorithm 1: Robust VolMin

input : $\boldsymbol{X}$; $p\in(0,1]$; N; initializations $(\boldsymbol{A},\boldsymbol{S})$; ϵ.

$\boldsymbol{W}=\boldsymbol{I}$; $\boldsymbol{F}=\boldsymbol{I}$;

repeat

- $\boldsymbol{A} := \boldsymbol{X}\boldsymbol{W}\boldsymbol{S}^T\left(\boldsymbol{S}\boldsymbol{W}\boldsymbol{S}^T+\lambda\boldsymbol{F}\right)^{-1}$;
- use Algorithm 2 to update $\{\boldsymbol{s}[\ell]\}_{\ell=1}^{L}$ by solving:
  $$\boldsymbol{s}[\ell] := \arg\min_{\boldsymbol{1}^T\boldsymbol{s}[\ell]=1,\boldsymbol{s}[\ell]\geq\boldsymbol{0}}\ \|\boldsymbol{x}[\ell]-\boldsymbol{A}\boldsymbol{s}[\ell]\|_2^2$$
- $w_\ell := \frac{p}{2}\left(\|\boldsymbol{x}[\ell]-\boldsymbol{A}\boldsymbol{s}[\ell]\|_2^2+\epsilon\right)^{\frac{p-2}{2}}$ for $\ell=1,\ldots,L$;
- $\boldsymbol{F} := \left(\boldsymbol{A}^T\boldsymbol{A}+\epsilon\boldsymbol{I}\right)^{-1}$.

until *Some stopping criterion is reached*;

output: $\{\boldsymbol{s}[\ell]\}$; $\boldsymbol{A}$.

Algorithm 2: ADMM for solving Problem (8)

input : $\boldsymbol{x}[\ell]$, $\boldsymbol{A}$; initialization $\boldsymbol{s}[\ell]$; step size $\rho>0$.

let auxiliary variable $\boldsymbol{z}[\ell]=\boldsymbol{s}[\ell]$;

let dual variable $\boldsymbol{u}[\ell]=\boldsymbol{0}$;

repeat

- $\boldsymbol{z}[\ell] := \arg\min_{\boldsymbol{1}^T\boldsymbol{z}[\ell]=\boldsymbol{1}^T}\ \frac{1}{2}\|\boldsymbol{x}[\ell]-\boldsymbol{A}\boldsymbol{z}[\ell]\|_2^2+\frac{\rho}{2}\|\boldsymbol{z}[\ell]-\boldsymbol{s}[\ell]+\boldsymbol{u}[\ell]\|_2^2$;
- $\boldsymbol{s}[\ell] := \max\{\boldsymbol{z}[\ell]+\boldsymbol{u}[\ell],\boldsymbol{0}\}$;
- $\boldsymbol{u}[\ell] := \boldsymbol{z}[\ell]-\boldsymbol{s}[\ell]+\boldsymbol{u}[\ell]$.

until *Some stopping criterion is reached*;

output: $\boldsymbol{s}[\ell]$

4. SIMULATIONS

In this section, we provide simulations to showcase the effectiveness of the proposed algorithm. We generate the elements of $\boldsymbol{A}\in\mathbb{R}^{M\times N}$ following the uniform distribution between zero and one, and $\boldsymbol{s}[\ell]$ following the uniform Dirichlet distribution. We threshold the maximal value of $\boldsymbol{s}[\ell]$'s elements to be 0.9 so that $\boldsymbol{a}_n$ does not appear in the columns of $\boldsymbol{X}$, resulting in relatively difficult cases for factorization. Zero-mean white Gaussian noise is added to the generated data. Also, we replace some data points by outliers. The outliers are generated following the uniform distribution between zero and $\alpha>0$. Throughout this section, we fix $p=0.5$, $\epsilon=10^{-10}$, and stop the proposed algorithm when the absolute change of the cost function is smaller than 10^{-5} or the number of iterations reaches 1000.

Fig. 2 shows an illustrative example, where $M=30$, $N=3$ and $L=500$. The visualization is by projecting the data and the results onto the ground-truth affine hull of $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_3$. In this case, α is set to be 2, ten outliers are added, and the SNR is 20dB. The benchmarked algorithm is *simplex identification via split augmented Lagrangian* (SISAL) [19], which is a state-of-the-art VolMin algorithm that takes outliers into consideration. We also use SISAL to initialize the proposed algorithm in this section. There is a parameter μ in SISAL; a large μ means that there are no or only a few outliers, and a small μ means the opposite. In Fig. 2, we present the results of SISAL under several μ's and the proposed algorithm with $\lambda=1$, respectively. We see that SISAL fails to find $\boldsymbol{A}$ in this case. The reason is that the DR pre-processing of SISAL already throws the data into a badly estimated subspace because of the presence of

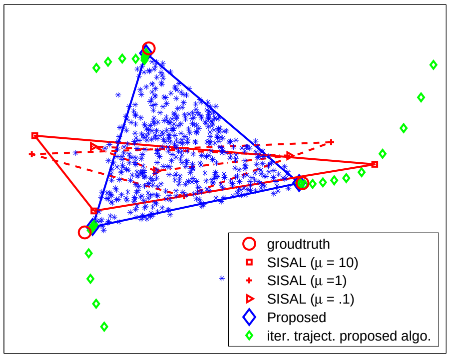

Fig. 2: The estimated $\hat{\boldsymbol{A}}$ by the algorithms for an $N = 3$ case. $(M, N, L) = (30, 3, 500)$; SNR= 20dB; $\alpha = 2$; number of outliers= 10.

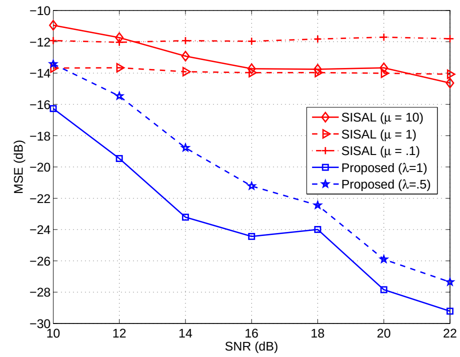

Fig. 3: The MSEs of the estimated $\hat{\boldsymbol{A}}$ by the algorithms under various SNRs. $(M, N, L) = (30, 5, 1000)$; $\alpha = 0.5$; number of outliers= 20.

outliers. On the other hand, we see that the proposed algorithm successfully finds the vertices of the desired convex hull. By looking at the iteration trajectory of the proposed algorithm in Fig. 2, we see that, despite starting from some points far away from the groundtruth vertices, the proposed algorithm manages to approach them.

Fig. 3 presents the mean squared error (MSE) of the estimated $\boldsymbol{A}$ by the algorithms under different SNRs. The results are averaged over 100 independent trials. In each trial, we randomly generate $\boldsymbol{A}$, $\boldsymbol{S}$, noise, and outliers. We let $(M, N) = (30, 5)$, $L = 1000$, and the number of outliers be 20 in this simulation. The parameter α is fixed to be 0.5. We see that the proposed algorithm works better than SISAL for all SNRs under test using both $\lambda = 0.5$ and $\lambda = 1$. Similar results can be seen in Fig. 4, where the results are obtained under different number of outliers and SNR= 20dB. We see that SISAL with $\mu = 1$ yields a reasonable result when the number of outliers is 10. But when the number of outliers increases, SISAL cannot yield accurate estimates of $\boldsymbol{A}$. On the other hand, the proposed algorithm with $\lambda = 1$ consistently gives good estimation results.

We also test the proposed algorithm on a semi-real simulation, where we generate hyperspectral pixels $\boldsymbol{x}[\ell]$ for $\ell = 1, \ldots, 1000$ using real-world hyperspectral signatures. The spectral signatures $\boldsymbol{a}_n$ for $n = 1, \ldots, 4$ correspond to different minerals (namely, *Carnallite*, *Ammonioalunite*, *Biotite*, and *Actinolite*) collected in the U.S.G.S. library [31]. Each signature $\boldsymbol{a}_n$ has $M = 224$ bands. Here, we apply the algorithms to detect the underlying materials, i.e., to estimate $\boldsymbol{A}$. This is considered more challenging than the previous simulations since the signatures are usually highly correlated, resulting in a badly conditioned $\boldsymbol{A}$. In this simulation, we add 30 outliers with $\alpha = 1$ and set SNR= 30dB. The other settings follow the previous simulations. In Fig. 5, we see that SISAL fails to recover at least one spectral signature in this case, but the proposed algorithm gives accurate estimates.

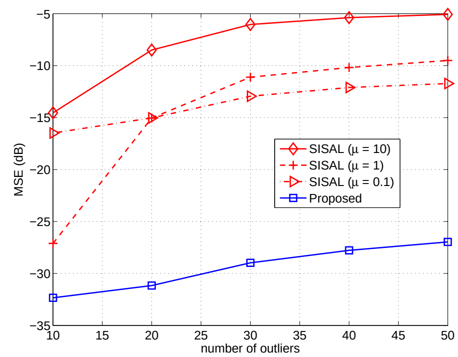

Fig. 4: The MSEs of the estimated $\hat{\boldsymbol{A}}$ by the algorithms under various amount of outliers. $(M, N, L) = (30, 5, 1000)$; $\alpha = 0.5$; SNR= 20dB; $\lambda = 1$.

5. CONCLUSION

We considered the VolMin-based matrix factorization problem by developing an outliers-robust VolMin criterion and algorithm. The proposed algorithm is algebraically simple, and equipped with guarantee of convergence to a stationary point. Numerical results showed that the proposed algorithm outperforms the state-of-the-art in robust volume minimization.

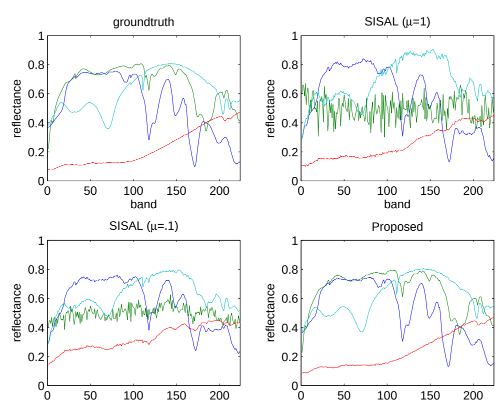

Fig. 5: The groundtruth and the estimated hyperspectral signatures by the algorithms. $(M, N, L) = (224, 4, 1000)$; $\alpha = 1$; SNR= 30dB; $\lambda = 1$.

6. REFERENCES

[1] D.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.

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

[4] N. Gillis, "The why and how of nonnegative matrix factorization," *Regularization, Optimization, Kernels, and Support Vector Machines*, vol. 12, pp. 257, 2014.

[5] 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, May 2015.

[6] X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Tensor-based power spectra separation and emitter localization for cognitive radio," in *Proc. IEEE SAM 2014*, 2014.

[7] X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Power spectra separation via structured matrix factorization," submitted to *IEEE Trans. Signal Process.*, 2016.

[8] D. Donoho and V. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?," in *NIPS*, 2003, vol. 16.

[9] H. Laurberg, M. G. Christensen, M. D. Plumbley, L. K. Hansen, and S. Jensen, "Theorems on positive data: On the uniqueness of NMF," *Computational Intelligence and Neuroscience*, vol. 2008, 2008.

[10] K. Huang, N. D. Sidiropoulos, and A. Swami, "Non-negative matrix factorization revisited: New uniqueness results and algorithms," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 211–224, Jan. 2014.

[11] 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.

[12] N. Gillis and S.A. Vavasis, "Fast and robust recursive algorithms for separable nonnegative matrix factorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 4, pp. 698–714, April 2014.

[13] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data," in *Proc. IEEE IGARSS 2008*, 2008, vol. 3, pp. III–250.

[14] 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, Nov. 2009.

[15] A. Agathos, J. Li, J.M. Bioucas-Dias, and A. Plaza, "Robust minimum volume simplex analysis for hyperspectral unmixing," in *Proc. EUSIPCO 2014*, 2014, pp. 1582–1586.

[16] 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 Sens.*, vol. 49, no. 11, pp. 4194–4209, 2011.

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

[18] G. Zhou, S. Xie, Z. Yang, J.-M. Yang, and Z. He, "Minimum-volume-constrained nonnegative matrix factorization: Enhanced ability of learning parts," *IEEE Trans. Neural Netw.*, vol. 22, no. 10, pp. 1626–1637, 2011.

[19] J. M. Bioucas-Dias, "A variable splitting augmented lagrangian approach to linear spectral unmixing," in *Proc. IEEE WHISPERS'09*, 2009, pp. 1–4.

[20] K. Huang, N.D. Sidiropoulos, E.E. Papalexakis, C. Faloutsos, P.P. Talukdar, and T.M. Mitchell, "Principled neuro-functional connectivity discovery," in *Proc. SIAM SDM 2015*, 2015.

[21] M.B. Lopes, J.-C. Wolff, J.M. Bioucas-Dias, and M.A. Figueiredo, "Near-infrared hyperspectral unmixing based on a minimum volume criterion for fast and accurate chemometric characterization of counterfeit tablets," *Analytical Chemistry*, vol. 82, no. 4, pp. 1462–1469, 2010.

[22] D. Tolliver, C. Tsourakakis, A. Subramanian, S. Shackney, and R. Schwartz, "Robust unmixing of tumor states in array comparative genomic hybridization data," *Bioinformatics*, vol. 26, no. 12, pp. i106–i114, 2010.

[23] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambriadge Press, 2004.

[24] 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 Sens.*, vol. 42, no. 10, pp. 2085–2095, 2004.

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

[26] G. Zhou, Z. Yang, S. Xie, and J.-M. Yang, "Online blind source separation using incremental nonnegative matrix factorization with volume constraint," *IEEE Trans. Neural Netw.*, vol. 22, no. 4, pp. 550–560, 2011.

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

[28] X. Fu, K. Huang, W.-K. Ma, N.D. Sidiropoulos, and R. Bro, "Joint tensor factorization and outlying slab suppression with applications," *IEEE Trans. Signal Process.,* to appear, 2015.

[29] J. Jose, N. Prasad, M. Khojastepour, and S. Rangarajan, "On robust weighted-sum rate maximization in MIMO interference networks," in *Proc. ICC 2011*, 2011, pp. 1–6.

[30] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, pp. 1–122, 2011.

[31] R. N. Clark, G. A. Swayze, R. Wise, K. E. Livo, T. M. Hoefen, R. F. Kokaly, and S. J. Sutley, "USGS digital spectral library splib06a," 2007.