MINIMUM-VOLUME RANK-DEFICIENT NONNEGATIVE MATRIX FACTORIZATIONS

Valentin Leplat, Andersen M.S. Ang, Nicolas Gillis

University of Mons, Rue de Houdain 9, 7000 Mons, Belgium

ABSTRACT

In recent years, nonnegative matrix factorization (NMF) with volume regularization has been shown to be a powerful identifiable model; for example for hyperspectral unmixing, document classification, community detection and hidden Markov models. In this paper, we show that minimum-volume NMF (min-vol NMF) can also be used when the basis matrix is rank deficient, which is a reasonable scenario for some real-world NMF problems (e.g., for unmixing multispectral images). We propose an alternating fast projected gradient method for min-vol NMF and illustrate its use on rank-deficient NMF problems; namely a synthetic data set and a multispectral image.

Index Terms— nonnegative matrix factorization, minimum volume, identifiability, rank deficiency

1. INTRODUCTION

Given a nonnegative matrix $X \in \mathbb{R}^{m \times n}_{+}$ and a factorization rank r, nonnegative matrix factorization (NMF) requires to find two nonnegative matrices $W \in \mathbb{R}^{m \times r}_{+}$ and $H \in \mathbb{R}^{r \times n}_{+}$ such that $X \approx WH$. For simplicity, we will use the Frobenius norm, which is arguably the most widely used, to assess the error of an NMF solution and consider the following optimization problem

$$\min_{W \in \mathbb{R}^{m \times r}, H \in \mathbb{R}^{r \times n}} ||X - WH||_F^2 \text{ s.t. } W \geq 0 \text{ and } H \geq 0.$$

NMF is in most cases ill-posed because the optimal solution is not unique. In order to make the solution of the above problem unique (up to permutation and scaling of the columns of W and rows of H) hence making the problem well-posed and the parameters (W, H) of the problem identifiable, a key idea is to look for a solution W with minimum volume; see [1] and the references therein. A possible formulation for minimum-volume NMF (min-vol NMF) is as follows

$$\min_{W \geq 0, H(:,j) \in \Delta^r \,\forall j} ||X - WH||_F^2 + \lambda \operatorname{vol}(W), \quad (1)$$

where $\Delta^r = \{x \in \mathbb{R}^r_+ \mid \sum_i x_i \leq 1\}$, λ is a penalty parameter, and $\operatorname{vol}(W)$ is a function that measures the volume of the columns of W. Note that H needs to be normalized otherwise W would go to zero since $WH = (cW)(H/c)$ for any $c > 0$. In this paper, we will use $\operatorname{vol}(W) = \operatorname{logdet}(W^T W + \delta I)$, where I is the identity matrix of appropriate dimensions. The reason for using such a measure is that $\sqrt{\det(W^T W)}/r!$ is the volume of the convex hull of the columns of W and the origin. Under some appropriate conditions on $X = WH$, this model will provably recover the true underlying (W, H) that generated X. These recovery conditions require that the columns of X are sufficiently well spread in the convex hull generated by the columns of W [2, 3, 4]; this is the so-called sufficiently scattered condition. In particular, data points need to be located on the facets of this convex hull hence H needs to be sufficiently sparse. A few remarks are in order:

- The ideas behind min-vol NMF have been introduced in the hyperspectral image community and date back from the paper [5]; see also the discussions in [6, 1].
- As far as we know, these theoretical results only apply in noiseless conditions hence robustness to noise of model (1) still needs to be rigorously analyzed (this is a very promising but difficult direction of further research).
- The sufficiently scattered condition is a generalization of the separability condition which requires $W = X(:, \mathcal{K})$ for some index set $\mathcal{K}$ of size r. Separability makes the NMF problem easily solvable, and efficient and robust algorithms exist; see, e.g., [7, 6, 8] and the references therein. Note that although min-vol NMF guarantees identifiability, the corresponding optimization problem (1) is still hard to solve in general; as the original NMF problem [9].

Another key assumption that is used in min-vol NMF is that the basis matrix W is full rank, that is, $\operatorname{rank}(W) = r$; otherwise $\det(W^T W) = 0$. However, there are situations when the matrix W is not full rank: this happens in particular when $\operatorname{rank}(X) \neq \operatorname{rank}_+(X)$ where $\operatorname{rank}_+(X)$ is the nonnegative rank of X which is the smallest r such that X has an exact NMF decomposition (that is, $X = WH$). Here is a simple example:

$$X = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \quad (2)$$

for which $\operatorname{rank}(X) = 3 < \operatorname{rank}_+(X) = 4$. The columns of the matrix X are the vertices of a square in a 2-dimensional subspace; see Fig. 2 for an illustration. A practical situation where this could happen is in multispectral imaging. Let us construct the matrix X such that each column $X(:, j) \geq 0$

Authors acknowledge the support by the European Research Council (ERC starting grant n° 679515) and by the Fonds de la Recherche Scientifique - FNRS and the Fonds Wetenschappelijk Onderzoek - Vlanderen (FWO) under EOS Project no O005318F-RG47.

is the spectral signature of a pixel. Then, under the linear mixing model, each column of X is the nonnegative linear combination of the spectral signatures of the constitutive materials present in the image, referred to as endmembers: we have $X(:,j) = \sum_{k=1}^{r} W(:,k)H(k,j)$, where $W(:,k)$ is the spectral signature of the kth endmember, and $H(k,j)$ is the abundance of the kth endmember in the jth pixel; see [6] for more details. For multispectral images, the number of materials within the scene being imaged can be larger than the number of spectral bands meaning that $r > m$ hence $\text{rank}(W) \leq m < r$.

In this paper, we focus on the min-vol NMF formulation in the rank-deficient scenario, that is, when $\text{rank}(W) < r$. The main contribution of this paper is three-fold: (i) We explain why min-vol NMF (1) can be used meaningfully when the basis matrix W is not full rank. This is, as far as we know, the first time this observation is made in the literature. (ii) We propose an algorithm based on alternating projected fast gradient method to tackle this problem. (iii) We illustrate our results on a synthetic data set and a multispectral image.

2. MIN-VOL NMF IN THE RANK-DEFICIENT CASE

Let us discuss the min-vol NMF model we consider in this paper, namely,

$$\min_{W \geq 0, H(:,j) \in \Delta^r \, \forall j} ||X - WH||_F^2 + \lambda \, \text{logdet}(W^T W + \delta I), \quad (3)$$

which has three key ingredients: the choice of the volume regularizer, that is, $\text{logdet}(W^T W + \delta I)$, the parameters δ and λ. They are discussed in the next three paragraphs.

Choice of the volume regularizer Most functions used to minimize the volume of the columns of W are based on the Gram matrix $W^T W$; in particular, $\det(W^T W)$ and $\text{logdet}(W^T W + \delta I)$ for some $\delta > 0$ are the most widely used measures; see, e.g., [10, 11]. Note that $\det(W^T W) = \Pi_{i=1}^{r} \sigma_i^2(W)$, hence the $\log$ term allows to weight down large singular values and has been observed to work better in practice; see, e.g., [12]. When W is rank deficient (that is, $\text{rank}(W) < r$), some singular values of W are equal to zero hence $\det(W^T W) = 0$. Therefore, the function $\det(W^T W)$ cannot distinguish between different rank-deficient solutions[1]. However, we have $\text{logdet}(W^T W + \delta I)$ = $\sum_{i=1}^{r} \log(\sigma_i^2(W) + \delta)$. Hence if W has one (or more) singular value equal to zero, this measure still makes sense: among two rank-deficient solutions belonging to the same low-dimensional subspace, minimizing $\text{logdet}(W^T W + \delta I)$ will favor a solution whose convex hull has a smaller volume within that subspace since decreasing the non-zero singular values of $(W^T W + \delta I)$ will decrease $\text{logdet}(W^T W + \delta I)$. In mathematical terms, let $W \in \mathbb{R}^{m \times r}$ belong to a k-dimensional subspace with $k < r$ so that $W = US$ where $U \in \mathbb{R}^{m \times k}$ is an orthogonal basis of that subspace and $S \in \mathbb{R}^{k \times r}$ are the coordinates of the columns of W in that subspace. Then, $\text{logdet}(W^T W + \delta I) = \sum_{i=1}^{k} \log(\sigma_i^2(S) + \delta) + (r-k)\log(\delta)$. The min-vol criterion $\text{logdet}(W^T W + \delta I)$ with $\delta > 0$ is therefore meaningful even when W does not have rank r.

Choice of δ The function $\text{logdet}(W^T W + \delta I)$ which is equal to $\sum_{i=1}^{r} \log(\sigma_i^2(W) + \delta)$ is a non-convex surrogate for the ℓ_0 norm of the vector of singular values of W (up to constants factors), that is, of $\text{rank}(W)$ [13, 14]. It is sharper than the ℓ_1 norm of the vector of singular values (that is, the nuclear norm) for δ sufficiently small; see Fig. 1. Therefore, if one wants to promote rank-deficient solutions, δ should not be chosen too large, say $\delta \leq 0.1$. Moreover, δ should not

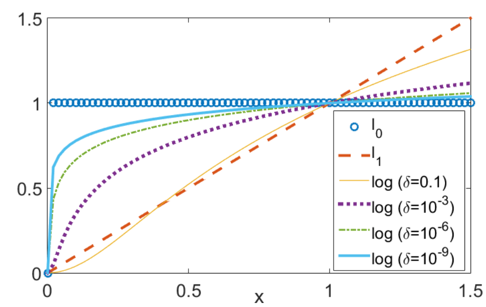

Fig. 1. Function $\frac{\log(x^2+\delta)-\log(\delta)}{\log(1+\delta)-\log(\delta)}$ for different values of δ, ℓ_1 norm ($= |x|$) and ℓ_0 norm ($= 0$ for $x = 0$, $= 1$ otherwise).

be chosen too small otherwise $WW^T + \delta I$ might be badly conditioned which makes the optimization problem harder to solve (see Section 3) –also, this could give too much importance to zero singular values which might not be desirable. Therefore, in practice, we recommend to use a value of δ between 0.1 and 10^{-3}. We will use $\delta = 0.1$ in this paper. Note that in previous works, δ was chosen very small (e.g., 10^{-8} in [11]) which, as explained above, is not a desirable choice, at least in the rank-deficient case. Even in the full-rank case, we argue that choosing δ too small is also not desirable since it promotes rank-deficient solutions.

Choice of λ The choice of δ will influence the choice of λ. In fact, the smaller δ, the larger $|\text{logdet}(\delta)|$, hence to balance the two terms in the objective (3), λ should be smaller. For the practical implementation, we will initialize $W^{(0)} = X(:,\mathcal{K})$ where $\mathcal{K}$ is computed with the successive nonnegative projection algorithm (SNPA) that can handle the rank-deficient separable NMF problem [15]. Note that SNPA also provides the matrix $H^{(0)}$ so as to minimize $||X - W^{(0)}H^{(0)}||_F^2$ while $H^{(0)}(:,j) \in \Delta^r$ for all j. Finally, we will choose

$$\lambda = \tilde{\lambda} \frac{||X - W^{(0)}H^{(0)}||_F^2}{|\text{logdet}(W^{(0)^T} W^{(0)} + \delta I)|},$$

where we recommend to choose $\tilde{\lambda}$ between 1 and 10^{-3} depending on the noise level (the noisier the input matrix, the larger λ should be).

[1] Of course, one could also use the measure $\det(W^T W + \delta I)$ meaningfully in the rank-deficient case. However, it would be numerically more challenging since for each singular value of W equal to zero, the objective is multiplied by δ which should be chosen relatively small.

3. ALGORITHM FOR MIN-VOL NMF

Most algorithms for NMF optimize alternatively over W and H, and we adopt this strategy in this paper. For the update of H, we will use the projected fast gradient method (PFGM) from [15]. Note that, as opposed to previously proposed methods for min-vol NMF, we assume that the sum of the entries of each column of H is smaller or equal to one, not equal to one, which is more general. For the update of W, we use a PFGM applied on an strongly convex upper approximation of the objective function; similarly as done in [11]–although in that paper, authors did not consider explicitly the case $W \geq 0$ (W is unconstrained in their model) and did not write down explicitly a PFGM taking advantage of strong convexity. For the sake of completeness, we briefly recall this approach. The following upper bound for the logdet term holds: for any $Q \succ 0$ and $S \succ 0$, we have

$$\begin{aligned}\operatorname{logdet}(Q) \leq g(Q,S) &= \operatorname{logdet}(S) + \operatorname{trace}\left(S^{-1}(Q-S)\right)\\ &= \operatorname{trace}\left(S^{-1}Q\right) + \operatorname{logdet}(S) - r.\end{aligned}$$

This follows from the concavity of $\operatorname{logdet}(.)$ as $g(Q,S)$ is the first-order Taylor approximation of $\operatorname{logdet}(Q)$ around S–it has also been used for example in [16]. This gives $\operatorname{logdet}(W^TW + \delta I) \leq \operatorname{trace}(YW^TW) + \operatorname{logdet}(Y^{-1}) - r$ for any W and any $Y = (Z^TZ + \delta I)^{-1}$ with $\delta > 0$. Plugging this in the original objective function, and denoting w_i^T the ith row of matrix W and $\langle .,. \rangle$ is the Frobenius inner product of two matrices, we obtain

$$\begin{aligned}\ell(W) &= ||X - WH||_F^2 + \lambda \operatorname{logdet}(W^TW + \delta I)\\ &= ||X||_F^2 - 2\langle XH^T, W\rangle + \langle W^TW, HH^T\rangle\\ &\quad + \lambda \operatorname{logdet}(W^TW + \delta I)\\ &\leq \langle W^TW, HH^T + \lambda Y\rangle - 2\langle C, W\rangle + b\\ &= 2\sum_{i=1}^{n}\left(\frac{1}{2}w_i^TAw_i - c_i^Tw_i\right) + b = \bar{\ell}(W),\end{aligned}$$

where $Y = (Z^TZ + \delta I)^{-1}$ and $A = HH^T + \lambda Y$ are positive definite for $\delta, \lambda > 0$, $C = XH^T$, and b is a constant independent of W. Note that $\bar{\ell}(W) = \ell(W)$ for $Z = W$. Minimizing the upper bound $\bar{\ell}(W)$ of $\ell(W)$ requires to solve m independent strongly convex optimization problems with Hessian matrix A. Using PFGM on this problem, we obtain a linear convergence method with rate $\frac{1-\sqrt{\kappa^{-1}}}{1+\sqrt{\kappa^{-1}}}$ where κ is the condition number of A [17]. Note that the subproblem in variable H is not strongly convex when W is rank deficient in which case PFGM converges sublinearly, in $O(1/k^2)$ where k is the iteration number. In any case, PFGM is an optimal first-order method in both cases [17], that is, no first-order method can have a faster convergence rate. When W is rank deficient, we have $\frac{\lambda}{\delta} \leq L = \lambda_{\max}(A) \leq ||H||_2^2 + \frac{\lambda}{\delta}$, where L is the largest eigenvalue of A. This shows the importance of not choosing δ too small, since the smaller δ, the larger the conditioning of A hence the slower will be the PFGM. Note that L is the Lipschitz constant of the gradient of the objective function and controls the stepsize which is equal to $1/L$. Our proposed algorithm is summarized in Alg. 1. We will use 10 inner iterations for the PFGM on W and H.

Algorithm 1 Min-vol NMF using alternating PFGM

Require: Input matrix $X \in \mathbb{R}_+^{m\times n}$, the factorization rank r, $\delta > 0$, $\tilde{\lambda} > 0$, number of iterations maxiter.
Ensure: (W, H) is an approximate solution of (3).

Initialize (W, H) using SNPA [15].
Let $\lambda = \tilde{\lambda}\frac{||X-WH||_F^2}{\operatorname{logdet}(W^TW+\delta I)}$.
for $k = 1, 2, \ldots,$ maxiter **do**
  % Update W
  Let $A = HH^T + \lambda(W^TW + \delta I)^{-1}$ and $C = XH^T$.
  Perform a few steps of PFGM on the problem $\min_{U\geq 0} \frac{1}{2}\langle U^TU, A\rangle - \langle U, C\rangle$, with initialization $U = W$. Set W as last iterate.
  % Update H
  Perform a few steps of PFGM on the problem $\min_{H(:,j)\in\Delta^r\,\forall j} ||X - WH||_F^2$ as in [15].
end for

4. NUMERICAL EXPERIMENTS

We now apply our method on a synthetic and a real-world data set. All tests are preformed using Matlab R2015a on a laptop Intel CORE i7-7500U CPU @2.9GHz 24GB RAM. The code is available from `http://bit.ly/minvolNMF`.

Synthetic data set. Let us construct the matrix $X \in \mathbb{R}^{4\times 500}$ as follows: W is taken as the matrix from (2) so that $\operatorname{rank}(W) = 3 < r = 4$, and each column of H is distributed using the Dirichlet distribution of parameter $(0.1, \ldots, 0.1)$. Each column of H with an entry larger 0.8 is resampled as long as this condition does not hold. This guarantees that no data point is close to a column of W (this is sometimes referred to as the purity index). Fig. 2 illustrates this geometric problem. As observed on Fig. 2, Alg. 1 is able to perfectly

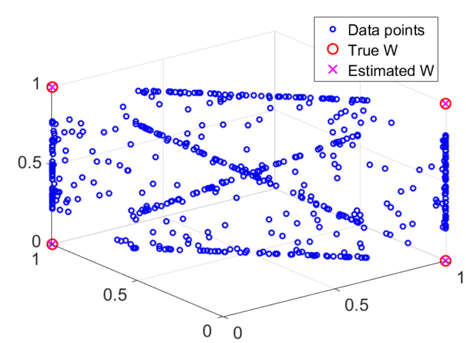

Fig. 2. Synthetic data set and recovery. (Only the first three entries of each four-dimensional vector are displayed.)

recover the true columns of W. For this experiment, we use $\tilde{\lambda} = 0.01$. Fig. 3 illustrates the same experiment where noise is added to $X = \max(0, WH + N)$ where $N = \epsilon$ randn(m,n) in Matlab notation (i.i.d. Gaussian distribution of mean zero and standard deviation ϵ). Note that the average of the entries of X is 0.5 (each column is a linear combination of the columns of W, with weights summing to one). Fig. 3 displays the average over 20 randomly generated matrices X of the relative error $d(W, \tilde{W}) = \frac{||W - \tilde{W}||_F}{||W||_F}$ where $\tilde{W}$ is the solution computed by Alg. 1 depending on the noise level ϵ. This illustrates that min-vol NMF is robust against noise since the $d(W, \tilde{W})$ is smaller than 1% for $\epsilon \leq 1\%$.

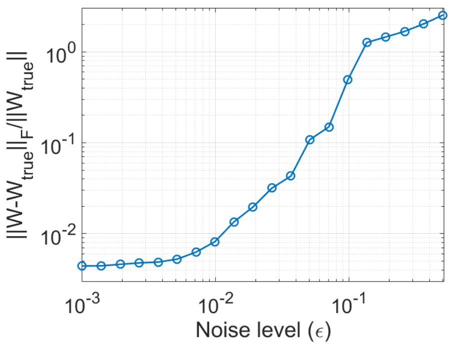

Fig. 3. Evolution of the recovery of the true W depending on the noise $N = \epsilon$ rand(m,n) using Alg. 1 ($\tilde{\lambda} = 0.01$, $\delta = 0.1$, maxiter = 100).

Multispectral image. The San Diego airport is a HYDICE hyperspectral image (HSI) containing 158 clean bands, and 400×400 pixels for each spectral image; see, e.g., [18]. There are mainly three types of materials: road surfaces, roofs and vegetation (trees and grass). The image can be well approximated using r=8. Since we are interested in the case $\text{rank}(W) < r$, we select m=5 spectral band using the successive projection algorithm [19] (this is essentially Gram-Schmidt with column pivoting) applied on X^T. This provides bands that are representative: the selected bands are 4, 32, 116, 128, 150. Hence, we are factoring a 5-by-160000 matrix using a r=8. Note that we have removed outlying pixels (some spectra contain large negative entries while others have a norm order of magnitude larger than most pixels). Fig. 4 displays the abundance maps extracted (that is, the rows of matrix H): they correspond to meaningful locations of materials. Here we have used $\tilde{\lambda}$=0.1 and 1000 iterations. From the initial solution provided by SNPA, min-vol NMF is able to reduce the error $||X - WH||_F$ by a factor of 11.7 while the term $\text{logdet}(W^T W + \delta I)$ only increases by a factor of 1.06. The final relative error is $\frac{||X - WH||_F}{||X||_F} = 0.2\%$.

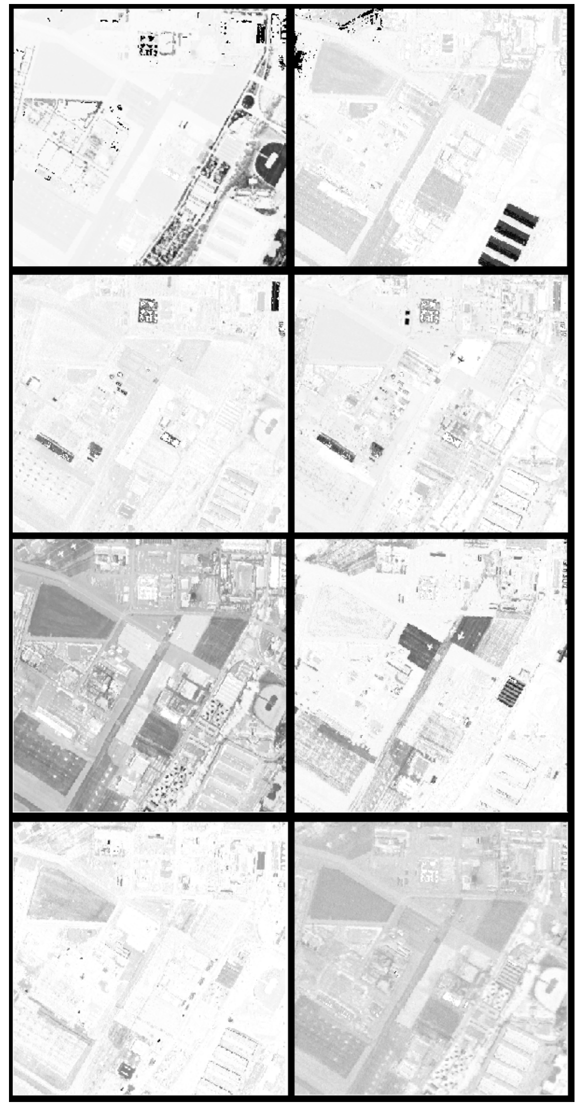

Fig. 4. Abundance maps extract by min-vol NMF using only five bands of the San Diego airport HSI. From left to right, top to bottom: vegetation (grass and trees), three different types of roof tops, four different types of road surfaces.

5. CONCLUSION

In this paper, we have shown that min-vol NMF can be used meaningfully for rank-deficient NMF's. We have provided a simple algorithm to tackle this problem and have illustrated the behaviour of the method on synthetic and real-world data sets. This work is only preliminary and many important questions remain open; in particular

• Under which conditions can we prove the identifiability of min-vol NMF in the rank-deficient case (as done in [2, 3] for the full-rank case)? Intuitively, it seems that a condition similar to the sufficiently-scattered condition would be sufficient but this has to be analysed thoroughly.

• Can we prove robustness to noise of such techniques? (The question is also open for the full-rank case.)

• Can we design faster and more robust algorithms? And algorithms taking advantage of the fact that the solution is rank-deficient?

6. REFERENCES

[1] Xiao Fu, Kejun Huang, Nicholas D Sidiropoulos, and Wing-Kin Ma, "Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications," *IEEE Signal Processing Magazine*, 2018, to appear.

[2] Chia-Hsiang Lin, Wing-Kin Ma, Wei-Chiang Li, Chong-Yung Chi, and ArulMurugan Ambikapathi, "Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: The no-pure-pixel case," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 10, pp. 5530–5546, 2015.

[3] Xiao Fu, Wing-Kin Ma, Kejun Huang, and Nicholas D Sidiropoulos, "Blind separation of quasi-stationary sources: Exploiting convex geometry in covariance domain.," *IEEE Transactions Signal Processing*, vol. 63, no. 9, pp. 2306–2320, 2015.

[4] Xiao Fu, Kejun Huang, and Nicholas D Sidiropoulos, "On identifiability of nonnegative matrix factorization," *IEEE Signal Processing Letters*, vol. 25, no. 3, pp. 328–332, 2018.

[5] Maurice D Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 3, pp. 542–552, 1994.

[6] 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," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 67–81, 2014.

[7] Sanjeev Arora, Rong Ge, Ravindran Kannan, and Ankur Moitra, "Computing a nonnegative matrix factorization–provably," in *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*. ACM, 2012, pp. 145–162.

[8] Nicolas Gillis, "Introduction to nonnegative matrix factorization," *SIAG/OPT Views and News*, vol. 25, no. 1, pp. 7–16, 2017.

[9] Stephen A Vavasis, "On the complexity of nonnegative matrix factorization," *SIAM Journal on Optimization*, vol. 20, no. 3, pp. 1364–1377, 2010.

[10] Lidan Miao and Hairong Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 765–777, 2007.

[11] Xiao Fu, Kejun Huang, Bo Yang, Wing-Kin Ma, and Nicholas D. Sidiropoulos, "Robust volume minimization-based matrix factorization for remote sensing and document clustering," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6254–6268, 2016.

[12] Andersen M.S. Ang and Nicolas Gillis, "Volume regularized non-negative matrix factorizations," in *2018 Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2018.

[13] Maryam Fazel, *Matrix rank minimization with applications*, Ph.D. thesis, Stanford University, 2002.

[14] Maryam Fazel, Haitham Hindi, and Stephen P Boyd, "Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices," in *Proceedings of the 2003 American Control Conference*. IEEE, 2003, vol. 3, pp. 2156–2162.

[15] Nicolas Gillis, "Successive nonnegative projection algorithm for robust nonnegative blind source separation," *SIAM Journal on Imaging Sciences*, vol. 7, no. 2, pp. 1420–1450, 2014.

[16] Kazuyoshi Yoshii, Ryota Tomioka, Daichi Mochihashi, and Masataka Goto, "Beyond NMF: Time-domain audio source separation without phase reconstruction," in *ISMIR*, 2013, pp. 369–374.

[17] Yurii Nesterov, *Introductory lectures on convex optimization: A basic course*, vol. 87, Springer Science & Business Media, 2013.

[18] Nicolas Gillis, Da Kuang, and Haesun Park, "Hierarchical clustering of hyperspectral images using rank-two nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 2066–2078, 2015.

[19] Nicolas Gillis and Stephen A Vavasis, "Fast and robust recursive algorithms for separable nonnegative matrix factorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 4, pp. 698–714, 2014.