Randomized Successive Projection Algorithm

Conference Paper · September 2022

CITATIONS
0

3 authors, including:

Nicolas Nadisic
Université de Mons
11 PUBLICATIONS **14** CITATIONS

SEE PROFILE

Randomized Successive Projection Algorithm

Olivier VU THANH, Nicolas NADISIC, Nicolas GILLIS

Département de Mathématique et Recherche Opérationnelle, Université de Mons
Rue de Houdain 9, 7000 Mons, Belgique.
{olivier.vuthanh,nicolas.nadisic,nicolas.gillis}@umons.ac.be

Résumé – L'algorithme de projections successives (SPA) est un algorithme standard pour la factorisation non-négative de matrices (NMF). Il est basé sur l'hypothèse de séparabilité. En démélange hyperspectral, c'est-à-dire l'extraction des matériaux dans une image hyperspectrale, la séparabilité est équivalente à l'hypothèse du pixel pur et stipule que, pour chaque matériau présent dans l'image, il existe au moins un pixel composé uniquement de ce matériau. SPA est rapide et a été prouvé robuste au bruit, mais il est sensible aux valeurs aberrantes (outliers). Aussi, il est déterministe, et donc pour un problème donné il produit toujours la même solution. Or, il a été démontré empiriquement que l'algorithme non-déterministe vertex component analysis (VCA), lorsqu'il est exécuté un nombre assez grand de fois, produit souvent au moins une solution qui est meilleure que la solution de SPA. Dans cet article, nous cherchons à combiner ces qualités et introduisons une version aléatoire de SPA, appelée RandSPA, qui produit des résultats potentiellement différents à chaque exécution. Il peut être exécuté plusieurs fois pour conserver la meilleure solution, et il est encore garanti robuste au bruit. Des expériences de démélange d'images hyperspectrales montrent que la meilleure solution sur plusieurs exécutions de RandSPA est généralement meilleure que la solution du SPA original.

Abstract – The successive projection algorithm (SPA) is a widely used algorithm for nonnegative matrix factorization (NMF). It is based on the separability assumption. In hyperspectral unmixing, that is, the extraction of materials in a hyperspectral image, separability is equivalent to the pure-pixel assumption and states that for each material present in the image there exists at least one pixel composed of only this material. SPA is fast and provably robust to noise, but it not robust to outliers. Also, it is deterministic, so for a given setting it always produces the same solution. Yet, it has been shown empirically that the non-deterministic algorithm vertex component analysis (VCA), when run sufficiently many times, often produces at least one solution that is better than the solution of SPA. In this paper, we combine the best of both worlds and introduce a randomized version of SPA dubbed RandSPA, that produces potentially different results at each run. It can be run several times to keep the best solution, and it is still provably robust to noise. Experiments on the unmixing of hyperspectral images show that the best solution among several runs of RandSPA is generally better that the solution of vanilla SPA.

1 Introduction

Nonnegative matrix factorization (NMF) is a linear dimensionality reduction technique that became a standard tool to extract latent structures in nonnegative data. Given an input matrix $X \in \mathbb{R}^{m \times n}_+$ and a factorization rank $r < \min(m, n)$, NMF consists in finding two factors $W \in \mathbb{R}^{m \times r}_+$ and $H \in \mathbb{R}^{r \times n}_+$ such that $X \approx WH$. Columns of X are called data points, and if H is column-stochastic then the columns of W can be seen as the vertices of the convex hull containing the data points. Applications of NMF include feature extraction in images, topic modeling, audio source separation, chemometrics, or blind hyperspectral unmixing (HU), see for example [4] and the references therein. Blind HU consists in identifying the materials present in an hyperspectral image as well as their distribution in the pixels of the image.

In general, NMF is NP-hard [12]. However, under the *separability assumption*, it is solvable in polynomial time [2]. This assumption states that for every vertex (column of W), there exists at least one data point (column of X) equal to this vertex. In blind HU, this is known as the *pure-pixel assumption* and means that for each material, there is at least one pixel composed almost purely of this material. Many algorithms have been introduced that leverage this assumption, see for instance [5, Chapter 7] and the references therein. Recently, algorithms for separable NMF that are provably robust to noise have been introduced [2]. One of the most widely used is the successive projection algorithm (SPA) [1].

SPA is robust to noise and generally works well in practice. However, it suffers from several drawbacks, notably a sensitivity to outliers. SPA is deterministic, that is for a given problem it gives the same result at every run. It is also greedy, in the sense that it extract vertices sequentially, so an error at a given iteration cannot be compensated in the following iterations. In this paper, we aim at addressing the sensitivity to outliers by designing a non-deterministic variant of SPA that could be run several times, in the hope that at least one run will not extract outliers.

Let us discuss an observation from [10]. The separable NMF algorithm called vertex component analysis (VCA) [11] includes a random projection, therefore it is non-deterministic and at each run it produces potentially a different result. VCA is simpler and its guarantees are weaker than those of SPA, and the experiments in [10] show that VCA performs worse than

SPA on average, but they also show that the best result of VCA over many runs is in most cases better that the result of SPA in terms of reconstruction error. This observation is our main motivation to design a non-deterministic variant of SPA, that we coin as randomized SPA (RandSPA).

This paper is organized as follows. In section 2 we introduce the general form of recursive algorithm for separable NMF analyzed in [7] which generalizes SPA. In section 3 we present the main contribution of this paper, that is a randomized variant of SPA, called RandSPA. We show the theoretical results on the robustness to noise of SPA still hold for RandSPA, while the randomization allows to better handle outliers by allowing a diversity in the solutions produced. In section 4 we illustrate the advantages of our method with experiments on both synthetic data sets and the unmixing of hyperspectral images.

2 Successive Projection Algorithm

In this section, we discuss the successive projection algorithm (SPA). It is based on the *separability assumption*, detailed below.

Assumption 1 (Separability) *The m-by-n matrix $X \in \mathbb{R}^{m\times n}$ is r-separable if there exist a nonnegative matrix H such that $X = X(:,\mathcal{J})H$, where $X(:,\mathcal{J})$ denotes the subset of columns of X indexed by $\mathcal{J}$ and $|\mathcal{J}| = r$.*

The pseudocode for a general recursive algorithm for separable NMF is given in Algorithm 1. Historically, the first variant of Algorithm 1 has been introduced by Araújo et al. [1] for spectroscopic component analysis with $f(x) = \|x\|_2^2 = x^\top x$, which is the so-called SPA. In the noiseless case, that is, under Assumption 1, SPA is guaranteed to retrieve $\mathcal{J}$ and more generally, the vertices of the set of points which are the columns of X [9]. This particular choice of f is proved to be the most robust to noise given the bounds in [7]. See Theorem 1 with $Q = I$ for the error bounds. The algorithm is iterative and is composed of the following two main steps:

- Selection step: the column that maximizes a given function f is selected (line 3).
- Projection step: all the columns are projected onto the orthogonal complement of the current selected columns (line 5).

These two steps are repeated r times, r being the target number of extracted columns. The drawback with the ℓ_2-norm is its sensitivity to outliers and the fact that it makes SPA deterministic. If some outliers are selected, running SPA again would still retrieve the exact same outliers.

3 Randomized SPA

In this section, we introduce the main contribution of this work, that is a randomized variant of SPA called RandSPA. Its key features are that it computes potentially different solutions at each run, thus allowing a multi-start strategy, and that the theoretical robustness results of SPA still hold.

Algorithm 1: Recursive algorithm for separable NMF [7]. It coincides with SPA when $f(x) = \|x\|_2^2$.

Input: An r-separable matrix $X \in \mathbb{R}^{m\times n}$, a function f to maximize.
Output: Index set $\mathcal{J}$ of cardinality r such that $X \approx X(:,\mathcal{J})H$ for some $H \geq 0$.

1 Let $\mathcal{J} = \emptyset$, $P^\perp = I_m$, $V = [\,]$.
2 **for** $k = 1 : r$ **do**
3 Let $j_k = \operatorname{argmax}_{1\leq j\leq n} f(P^\perp X(:,j))$. (Break ties arbitrarily, if necessary.)
4 Let $\mathcal{J} = \mathcal{J} \cup \{j_k\}$.
5 Update the projector $P^\perp$ onto the orthogonal complement of $X(:,\mathcal{J})$:

$$v_k = \frac{P^\perp X(:,j_k)}{\|P^\perp X(:,j_k)\|_2},$$
$$V = [V\ v_k],$$
$$P^\perp \leftarrow \left(I_m - VV^T\right).$$

RandSPA follows Algorithm 1 with $f(x) = x^\top QQ^\top x$, with $Q \in \mathbb{R}^{m\times\nu}$ being a randomly generated matrix with $\nu \geq r$. To control the conditioning of Q, we generate the columns of Q such that they are mutually orthogonal and such that

$$\|Q(:,1)\|_2 = 1 \geq \cdots \geq \|Q(:,\nu)\|_2 = 1/\sqrt{\kappa}$$

where κ is the desired conditioning of $QQ^\top$. For the columns between the first and the last one, we make the arbitrary choice to fix them also to $1/\sqrt{\kappa}$. If $Q^\top W$ has full column rank, which happens with probability one if $\nu \geq r$, RandSPA is robust to noise with the following bounds:

Theorem 1 *[6, Corollary 1] Let $\tilde{X} = X + N$, where X satisfies Assumption 1, W has full column rank, and N is noise with $\max_j \|N(:,j)\|_2 \leq \epsilon$; and let $Q \in \mathbb{R}^{m\times\nu}$ with $\nu \geq r$. If $Q^\top W$ has full column rank and*

$$\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(W)}{\sqrt{r}\kappa^3(Q^\top W)}\right),$$

then SPA applied on matrix $Q^\top\tilde{X}$ identifies a set of indices $\mathcal{J}$ corresponding to the columns of W up to the error

$$\max_{1\leq j\leq r}\min_{k\in\mathcal{J}} \left\|W(:,j) - \tilde{X}(:,k)\right\|_2 \leq \mathcal{O}\left(\epsilon\kappa(W)\kappa(Q^\top W)^3\right).$$

Theorem 1 is directly applicable to RandSPA since choosing $f(x) = x^\top QQ^\top x$ is equivalent to performing SPA on $Q^\top\tilde{X}$. The only subtlety is that with RandSPA, a random Q is drawn at each column extraction. The error bound for RandSPA is then the one with the highest drawn $\kappa(Q^\top W)$.

Let us note that choosing $\nu = 1$ or $\|Q(:,j)\| = 1/\sqrt{\kappa}$ with $\kappa \to \infty$ for all $j > 1$ retrieves VCA. Choosing $\nu = m$ and $\kappa(Q) = 1$ retrieves SPA. Hence, RandSPA creates a continuum between SPA, with more provable robustness, and VCA, with more solution diversity.

4 Numerical experiments

In this section, we study empirically the performance of the proposed algorithm RandSPA on the unmixing of hyperspectral images. The algorithms have been implemented in Julia [3]. Our codes are available in an online repository[1] along with the data and test scripts used in our experiments. Our tests are performed on 5 real hyperspectral datasets[2] described in Table 1.

Dataset	m	n	r
Jasper	198	$100 \times 100 = 10000$	4
Samson	156	$95 \times 95 = 9025$	3
Urban	162	$307 \times 307 = 94249$	5
Cuprite	188	$250 \times 191 = 47750$	12
San Diego	188	$400 \times 400 = 160000$	8

Table 1: Summary of the datasets, for which $X \in \mathbb{R}^{m \times n}$.

For all the tests, we choose $\nu = r + 1$ and a relatively well conditioned Q with $\kappa(Q) = 1.5$. We then compute $W = X(:, \mathcal{J})$ once with SPA and 30 times with RandSPA. Next, we compute H by solving the nonnegative least squares (NNLS) subproblem $\min_{H \geq 0} \|X - WH\|_F^2$ exactly with an active-set algorithm [8], and we compute the relative reconstruction error $\|X - WH\|_F / \|X\|_F$. For RandSPA, we show the best error and the median error among the 30 runs. Note that in our setting we choose the best solution as the one with the lower reconstruction error, but other methods could be used to choose the best solution among all the computed ones.

The results of the experiments for SPA and RandSPA are presented in Table 2. The median error of RandSPA is on the same order than that of SPA, except for Cuprite where it is higher. It is even slightly smaller for Samson and Urban. On the other hand, the error from the best run of RandSPA is always smaller than that of SPA. Particularly, the error is decreased respectively by 37%, 32% and 27% for Samson, Urban and San Diego. This improvement is quite noticeable.

The resulting false color images for Jasper, Samson, Urban and Cuprite are shown on Figure 2. They represent the repartition of the materials identified by SPA and RandSPA in the image. As we can see for Urban, SPA does not manage to separate well the grass and the trees (both the grass and trees are in green), while with RandSPA, it occurred that some random Q amplified some directions that separate better the grass (in blue) and the trees (in green). Similarly, in the abundance maps from the unmixing of Samson in Figure 2, RandSPA separates the soil (in red), the water (in blue) and the trees (in green) better than SPA where the soil (in blue) is extracted but the water is not clearly identified.

[1] https://gitlab.com/nnadisic/randspa

[2] Downloaded from http://lesun.weebly.com

Dataset	SPA	Med. RandSPA	Best RandSPA
Jasper	8.6869	8.7577	8.0206
Samson	6.4914	6.3114	3.9706
Urban	10.9367	9.6354	6.5402
Cuprite	2.6975	3.526	2.2824
San Diego	12.6845	12.8714	9.2032

Table 2: Relative reconstruction error $\|X - WH\|_F / \|X\|_F$ in percent.

Let us discuss another experiment on the dataset Samson. We add some Gaussian noise such that $SNR = 20dB$, we fix $\kappa = 1$ and vary ν, and then show the average best error in 1,5,10 and 20 runs on Figure 1. As we can see, with a sufficient amount of runs that is 10 in this experiment, the relative error significantly improves for a ν near 10 in comparison to other choices of ν. In particular, it is also better than both $\nu = 1$ (VCA) and a high ν like 50 that should behave like SPA. Without added noise, VCA would perform better than every ν higher than 1 starting from 10 runs. However, when the data is noisy, this experiment highlights that VCA is not robust enough to noise and that the best run from a method between SPA and VCA is better than both SPA and VCA.

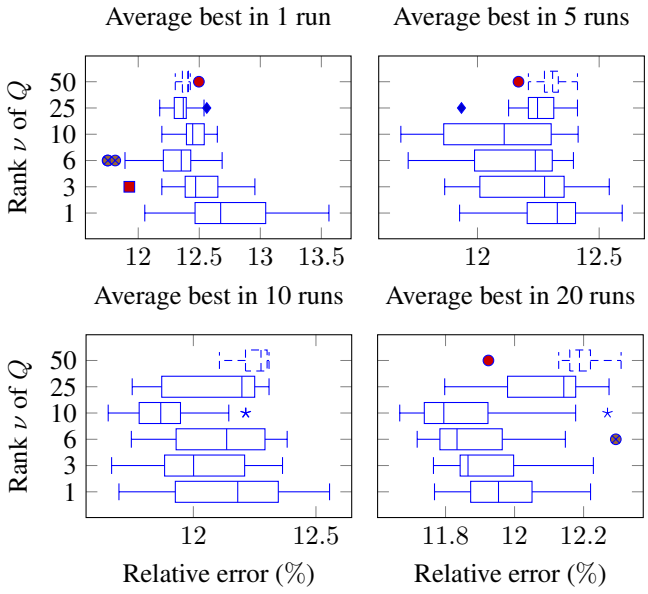

Figure 1: Average best reconstruction error on several runs, depending on ν, with $\kappa = 1$, on the hyperspectral image Samson with added noise such that $SNR = 20dB$.

5 Conclusion

In this paper, we introduced RandSPA, a variant of the separable NMF algorithm SPA that introduces randomness to allow a multi-start strategy. The robustness results of SPA still hold

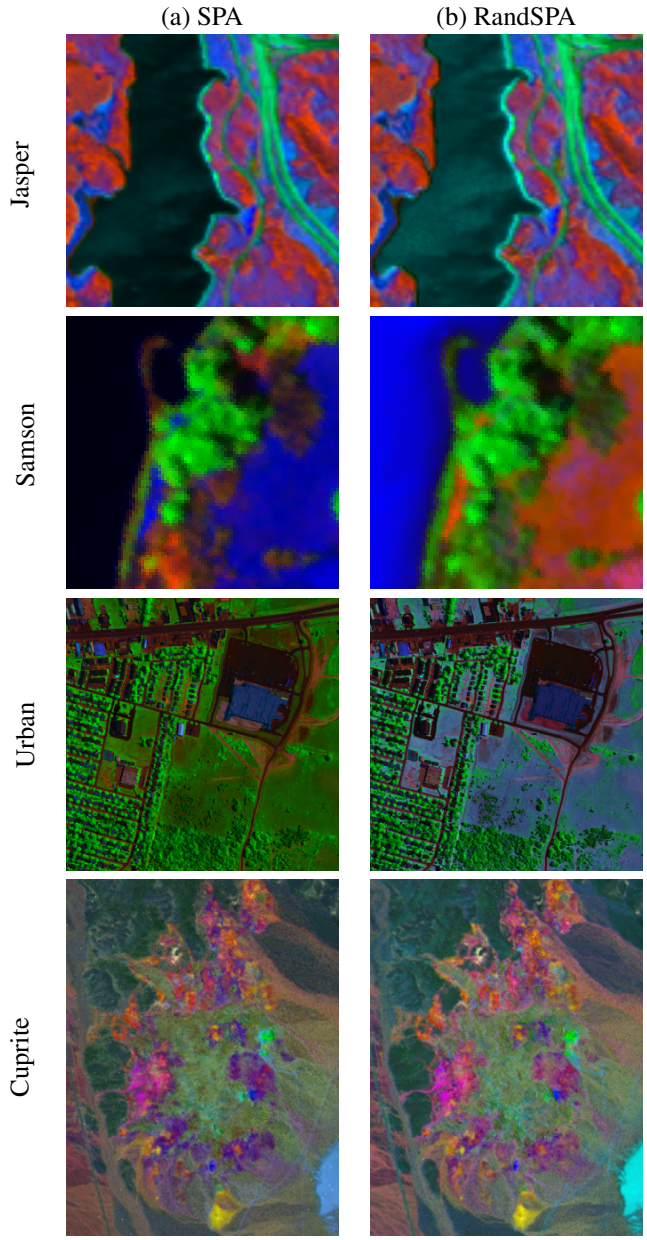

Figure 2: Abundance maps in false color from the unmixing of hyperspectral images.

for RandSPA, provided a bound on the noise that depends on the parameters used. We showed empirically on the unmixing of hyperspectral images that, with sufficiently many runs, the best solution from RandSPA is generally better that the solution from SPA. We also showed that RandSPA creates a continuum between the two algorithms SPA and VCA, as we can recover these algorithms by running RandSPA with some given parameter values.

Acknowledgements The authors acknowledge the support by the F.R.S.-FNRS and the FWO under EOS project O005318F-RG47. NG also acknowledges the Francqui foundation.

References

[1] Mário César Ugulino Araújo et al. “The successive projections algorithm for variable selection in spectroscopic multicomponent analysis”. In: *Chemometrics and intelligent laboratory systems* 57.2 (2001), pp. 65–73.

[2] Sanjeev Arora et al. “Computing a nonnegative matrix factorization-Provably”. In: *44th Annual ACM Symposium on Theory of Computing, STOC’12*. 2012, pp. 145–161.

[3] Jeff Bezanson et al. “Julia: A fresh approach to numerical computing”. In: *SIAM Review* 59.1 (2017), pp. 65–98.

[4] Xiao Fu et al. “Nonnegative Matrix Factorization for Signal and Data Analytics: Identifiability, Algorithms, and Applications.” In: *IEEE Signal processing magazine* 36.2 (2019), pp. 59–80.

[5] Nicolas Gillis. *Nonnegative matrix factorization*. SIAM, 2020.

[6] Nicolas Gillis and Wing-Kin Ma. “Enhancing pure-pixel identification performance via preconditioning”. In: *SIAM Journal on Imaging Sciences* 8.2 (2015), pp. 1161–1186.

[7] Nicolas Gillis and Stephen A Vavasis. “Fast and robust recursive algorithms for separable nonnegative matrix factorization”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36.4 (2013), pp. 698–714.

[8] Jingu Kim and Haesun Park. “Toward faster nonnegative matrix factorization: A new algorithm and comparisons”. In: *2008 Eighth IEEE International Conference on Data Mining*. IEEE. 2008, pp. 353–362.

[9] Wing-Kin Ma et al. “A signal processing perspective on hyperspectral unmixing: Insights from remote sensing”. In: *IEEE Signal Processing Magazine* 31.1 (2013), pp. 67–81.

[10] Nicolas Nadisic, Nicolas Gillis, and Christophe Kervazo. “Smoothed separable nonnegative matrix factorization”. In: *preprint arXiv:2110.05528* (2021).

[11] José MP Nascimento and José MB Dias. “Vertex component analysis: A fast algorithm to unmix hyperspectral data”. In: *IEEE transactions on Geoscience and Remote Sensing* 43.4 (2005), pp. 898–910.

[12] Stephen A Vavasis. “On the complexity of nonnegative matrix factorization”. In: *SIAM Journal on Optimization* 20.3 (2010), pp. 1364–1377.