The successive projection algorithm as an initialization method for brain tumor segmentation using nonnegative matrix factorization

N. Sauwen[1,2], M. Acou[3], H.N. Bharath[1,2], D. Sima[1,2], J. Veraart[4], F. Maes[2,5], U. Himmelreich[6], E. Achten[3] and S. Van Huffel[1,2]

1- KU Leuven - Department of Electrical Engineering (ESAT), STADIUS Centre for Dynamical Systems, Signal Processing and Data Analytics - Leuven, Belgium

2- iMinds - Department Medical Information Technologies - Belgium

3- Ghent University Hospital - Department of Radiology - Ghent, Belgium

4- University of Antwerp - iMinds Vision Lab, Department of Physics - Antwerp, Belgium

5- KU Leuven - Department of Electrical Engineering (ESAT), PSI Centre for Processing Speech and Images - Leuven, Belgium

6- KU Leuven - Department of Imaging and Pathology, Biomedical MRI/MoSAIC - Leuven, Belgium

Abstract

Non-negative matrix factorization (NMF) has become a widely used tool for additive parts-based analysis in a wide range of applications. As NMF is a non-convex problem, the quality of the solution will depend on the initialization of the factor matrices. Random initialization is by far the most commonly used method and is also the benchmark for comparison against other initialization methods, although its results are not reproducible. In this study, the successive projection algorithm (SPA) is proposed as an initialization method for NMF. SPA is a fast and reproducible method, and it aligns well with the assumptions made in near-separable NMF analyses. SPA was applied to a multi-parametric magnetic resonance imaging (MRI) dataset for brain tumor segmentation using different NMF algorithms. Repetitive random (rRandom) initialization served as a reproducible reference for comparison. SPA shows similar performance in terms of segmentation quality of the pathologic tissue classes, it reaches similar or increased convergence speed and it requires much less computation time than rRandom. Whereas SPA was previously applied as a direct endmember extraction tool, we have shown improved segmentation results when using SPA as an initialization method, as it allows further enhancement of the sources during the NMF iterative procedure.

Keywords Non-negative matrix factorization, successive projection algorithm, non-convex optimization, tumor segmentation

1 Introduction

Nonnegative matrix factorization (NMF) has become a widely used tool for multivariate data analysis in the fields of blind source separation and pattern recognition. NMF decomposes a nonnegative input matrix X into the product of 2 nonnegative factor matrices W and H, providing a low-rank (rank r) approximation:

$$X \approx WH \quad \text{with} \quad X \in R_+^{m \times n}, W \in R_+^{m \times r} \quad \text{and} \quad H \in R_+^{r \times n} \tag{1}$$

NMF provides an additive parts-based representation of the input data. As such, it reveals the basic components which are present in the data, the so-called sources, and models each data point (i.e. column of the input matrix) as a weighted sum of the sources. The columns of W represent the sources and each column of H contains the weights, or so-called abundances, of the sources for one particular column of X. The most commonly used similarity measure for X and its factorization is the Frobenius norm. It will also serve as the cost function for solving the optimization problem:

$$\min_{W,H} f(W,H) = \frac{1}{2}\|X - WH\|_F^2 \quad \text{, such that} \quad W \geq 0, \quad H \geq 0 \tag{2}$$

As NMF is a non-convex optimization problem, the obtained factorization will be a local rather than the global minimum of the cost function. The final result therefore depends on the initialization of the factorization matrices, W_0 and H_0. Random initialization is the benchmark and is used in the vast majority of NMF studies. However, the quality and reproducibility of the NMF result is rarely questioned when using random initialization. As was pointed out in [1], one would have to run NMF with random initialization a sufficient number of times, then select the 'best' run based on some criterion (e.g. lowest Frobenius residual error), to warrant robust and reproducible NMF results. This would of course increase computation time dramatically. Alternatively, more advanced initialization strategies have been suggested for NMF. These methods usually provide an initialization which is more appropriate for the data at hand, resulting in faster convergence and/or reproducible NMF results.

Besides random initialization, other randomization based methods have been suggested. These methods obtain the columns of W_0 by averaging a number of random columns of X or a subset of X [2]. Despite having a low computational cost and providing a more realistic first estimate of the sources compared to random initialization, these methods also suffer from a lack of reproducibility. Several initialization methods have been proposed which are based on the singular value decomposition (SVD) [1, 3, 4]. These methods rely on the most

significant singular values and their corresponding singular vectors to initialize W_0 and H_0. One drawback of SVD-based initialization is that the nonnegative structure of the input data is lost, introducing negative values in the singular vectors. Some straightforward ways of dealing with these negative values have been proposed, such as setting them to zero [1], replacing them by a mean value from the input matrix [1] or taking their absolute values [3]. Clustering algorithms have also been suggested for NMF initialization, such as spherical k-means clustering [5], fuzzy C-means clustering [6] and subtractive clustering [7]. Clustering-based initialization schemes will provide more realistic source estimates compared to SVD-based methods, but they can become computationally expensive. Furthermore, clustering methods require some initialization themselves. Most of the proposed initialization methods have been compared to random initialization in terms of speed of convergence and quality of the solution. But in these studies, random initialization is run only once, which makes it a questionnable reference as the results are not reproducible and different results might be obtained with another random initialization.

The Successive Projection Algorithm (SPA) is a forward selection method which minimizes collinearity of the selected variables in vector space. It was introduced by Araújo *et al.* [8] and has been commonly used for feature selection. Several studies have considered SPA as an endmember extraction tool for hyperspectral unmixing [9, 10]. SPA is fast and reproducible, and it takes advantage of the geometrical convexity that is seen in a wide range of NMF problems [9]. The current study proposes SPA as an initialization method for NMF. We illustrate the use of SPA initialization on a multi-parametric magnetic resonance imaging (MRI) dataset, applying NMF for brain tumor segmentation. SPA is compared to repetitive random (rRandom) initialization, i.e. NMF is run a sufficient number of times with random initialization, such that the results are reproducible. Several NMF algorithms are discussed and their sensitivity to the initialization methods is investigated. The paper is further organized as follows: Section 2 introduces the NMF methods and the SPA algorithm. Section 3 describes the multi-parametric MRI dataset and the validation methods for brain tumor segmentation. NMF segmentation results comparing SPA and rRandom initialization are given in Section 4, and an in-depth discussion of the results follows in Section 5.

2 NMF and SPA algorithms

2.1 NMF methods

aHALS NMF Accelerated hierarchical alternating least squares NMF (aHALS NMF) [11] is a member of the family of alternating least squares (ALS) NMF methods. These methods rely on the observation that finding the optimal factor matrix H when W is fixed, and finding the optimal W when H is fixed, are convex problems, as opposed to the original non-convex NMF problem as de-

fined in Equation 2. In the current study we chose to use aHALS because of its computational efficiency and fast convergence compared to other ALS methods [11].

Convex NMF Convex NMF [12] imposes the constraint that the source vectors (i.e. the columns of W) must lie within the column space of X. As such, each source is a weighted sum of the data points. An auxiliary matrix A is introduced to define these weights:

$$W = XA \quad \text{such that} \quad X \approx XAH \tag{3}$$

Multiplicative update rules are defined for A and H. Without additional constraints, Convex NMF results in a sparse H matrix.

GD NMF Gradient-descent NMF (GD NMF) methods are based on nonlinear least squares optimization. As these methods aim at directly minimizing the NMF cost function, they show faster convergence compared to standard NMF algorithms [13]. The GD NMF method applied in this study is based on Gauss-Newton with dogleg trust region [14]. A first-order gradient optimization scheme is usually sufficient for NMF, and it avoids expensive second-order computations such as inverting Hessians.

hNMF Hierarchical NMF (hNMF) has been introduced in previous MRI studies on brain tumor characterization [15, 16]. It consists of a 2-level approach, assigning tissue types which are most similar to the same source after a first level of rank-2 NMF. Tissue-specific sources are obtained after the second level of NMF. The sources are then recombined in a final step to calculate the tissue abundances using nonnegative least squares fitting. hNMF has shown improved differentiation and segmentation of the pathologic tissue types in brain tumors compared to single-level aHALS NMF.

2.2 Successive projection algorithm

NMF aims at finding the sources, i.e. the basic components of an input matrix $X \in R_+^{m \times n}$, such that each data point can be approximated as a weighted sum of the sources. For now, we assume that each source occurs at least once as a column of X, meaning that some data points contain purely one source. This has been referred to as the pure-pixel assumption in the context of hyperspectral unmixing. We further assume that the sum of the abundances is not higher than 1 in any data point, which is called the sum-to-one constraint. For such input data, the NMF problem is said to be *near-separable* [17]. Under the assumptions of near-separable NMF, the data points of X will span a convex hull in m-dimensional space, and the sources correspond to the vertices of this hull. SPA fits well with this geometrical interpretation of NMF, as it aims at finding the vertices [9]. SPA works as follows: in a first step, the data point x_1 with the highest l_2-norm is selected, as it will correspond to a vertex of the

convex hull. Next, all data points are projected onto the orthogonal complement of x_1. The data point x_2 with the highest l_2-norm in this projected subspace will be another vertex. The next vertex x_3 is found as the data point with the highest l_2-norm after projection onto the orthogonal complement of x_1 and x_2, and so on. The columns of W_0 will be formed by x_1, x_2,..., and x_r. SPA does not directly provide an initialization for the abundance matrix, H_0. Based on W_0, we used nonnegative least squares fitting to obtain H_0. For Convex NMF, an initialization of the auxiliary matrix A_0 is required. Solving Equation 3 for A, we find:

$$A \approx H^T(HH^T)^{-1} \tag{4}$$

Since A must be nonnegative, we obtain A_0 as proposed in [12]:

$$A_0 = A^+ + 0.2E\langle A^+ \rangle \tag{5}$$

with A^+ being the positive part of A, $\langle A \rangle = \sum_{m,n} |A_{n,k}| / \|A_{n,k}\|_0$ and $\|A_{n,k}\|_0$ is the number of nonzero elements in A. E is a matrix of ones with the same dimensions as A.

Initialization obtained from SPA is compared to repetitive random initialization (rRandom), where the elements of W_0 and H_0 are set to pseudo-random nonnegative values between 0 and 1 at each run. For rRandom, the final NMF result is selected as the random run with the lowest residual error as defined in Equation 2. Empirically, we found that 30 random runs was sufficient for the NMF problem at hand to obtain reproducible results with all considered NMF methods. For the remainder of this article, NMF analysis using SPA initialization will be denoted as NMF_{SPA} and NMF using rRandom as $NMF_{rRandom}$.

3 Multi-parametric MRI dataset

3.1 Brain tumor segmentation

Brain tumor segmentation aims at outlining the total tumor volume as well as the tumor subregions. The segmentation is crucial for planning surgical resection of the tumor, for radiotherapy planning and to monitor tumor growth or shrinkage during follow-up. The main pathologic tissue types within gliomas are active tumor, necrosis and edema. Active tumor, consisting of the actual proliferating tumor cells, is the main target of treatment. Necrosis results from tumor cell starvation due to a lack of oxygen and blood supply. It is a hallmark of glioblastoma, i.e. the highest grade glioma. Edema is the extracellular fluid accumulation in the peri-tumoral region, which is mostly found in high-grade gliomas. Manual segmentation is the current standard in clinical practice, but it is time-consuming and it suffers from a lack of reproducibility [18]. Several studies have proposed NMF methods for brain tumor segmentation and differentiation of the pathologic tissue types [15, 19, 16]. Using NMF, we expect to find tissue-specific signatures as the columns of W, and the tissue abundances per voxel on the columns of H.

3.2 MRI acquisition and processing methods

21 patients who were diagnosed with a high-grade glioma underwent a multi-parametric MRI acquisition protocol in the Ghent University Hospital. The local ethics committee allowed a retrospective analysis of the data. The acquisition protocol consisted of structural MRI, perfusion-weighted MRI, diffusion-weighted MRI and MR spectroscopic imaging. These MRI modalities provide us with complementary information about the anatomical, haemodynamic, structural and biochemical properties of the brain and tumor tissue. It was shown previously that combining these MRI modalities leads to improved brain tumor segmentation [15]. The MR examinations were performed on a 3T Siemens Trio Tim scanner (Erlangen, Germany), using a standard 12-channel phased array head coil.

Structural imaging included a 3-dimensional T1-weighted gradient-echo sequence (MPRAGE) with isotropic voxels (176 sagittal slices, field of view (FOV) read = 220mm, voxel size $0.9 \times 0.9 \times 0.9\text{mm}^3$, Repetition Time (TR) = 1550ms, Echo Time (TE) = 2.39ms, Inversion Time (TI) = 900ms, matrix size = 256×256, GRAPPA factor 2, flip angle = 9°) and a 3-dimensional T2-weighted inversion recovery sequence (FLAIR) with isotropic voxels (176 sagittal slices, FOV read = 250mm, voxel size $1 \times 1 \times 1\text{mm}^3$, TR = 6000ms, TE = 421ms, TI = 2100ms, matrix size = 256×238, GRAPPA factor 2). The MPRAGE sequence was repeated after administration of gadolinium contrast, namely following the acquisition of the Dynamic Susceptibility Contrast (DSC) perfusion-weighted imaging.

Perfusion-weighted imaging was performed by using a lipid-suppressed, T2*-weighted echo-planar imaging sequence with the following parameters: TR = 1000ms, TE = 29ms, FOV = $230 \times 230\text{mm}^2$, slice thickness = 5.0mm, matrix size = 128×128, in-plane voxel size = $1.8 \times 1.8\text{mm}^2$, 15 slices, distance factor = 30%, GRAPPA factor 2, flip angle = 90°. A series of 90 multi-section acquisitions was acquired at 1 second intervals. The first 10 acquisitions were performed before contrast agent injection to establish a pre-contrast baseline. At the tenth acquisition, a 0.1mmol/kg body weight bolus of gadobutrol (Gadovist, Bayer) was injected with a power injector (Spectris, Medrad Inc., Indianola, PA) at a rate of 4ml/s through a 18-gauge intravenous catheter, immediately followed by a 20 ml bolus of sodium chloride solution at 4ml/s. Relative Cerebral Blood Volume (CBV) maps were derived from the dynamic signal intensity curves in DSCoMAN (Duke University, Durham, NC) using the method proposed by Boxerman et al. [20].

Axial diffusion-weighted images were acquired using a fast single-shot gradient-echo echo-planar imaging sequence with diffusion gradient b-values of 0, 500 and 1000s/mm^2 (voxel size $2.0 \times 2.0 \times 3.0\text{mm}^3$, TR = 5400ms, TE = 80ms, FOV read = 264mm, number of averages = 3). An affine coregistration was applied to account for eddy currents. Apparent Diffusion Coefficient (ADC) maps were

derived from the 3 b-values using weighted linear least squares fitting [21]. The b0 images were also added to the dataset, serving as a T2-weighted reference.

A 3D proton magnetic resonance spectroscopic imaging (MRSI) protocol with long TE was included. In the two-slice MRSI examination, a volume of interest of $80 \times 80\text{mm}^2$ including tumour, perilesional edema and normal brain tissue was positioned on reconstructions of the 3D FLAIR sequence. A Stimulated Echo Acquisition Mode (STEAM) pulse sequence was used for 3D spatial localization. The VOI was completely enclosed within the brain and 8 outer volume suppression slabs were placed outside the volume of interest. Magnetic resonance parameters were TR = 1700ms, TE = 135ms, flip angle = 90°, FOV = $160 \times 160\text{mm}^2$, voxel size $10 \times 10 \times 15\text{mm}^3$, acquisition bandwidth 1200Hz, number of averages = 3. Weak water suppression was used. Automatic shimming with manual fine-tuning of the B_0 magnetic field was used as well as iterative semi-automatic optimization of the transmitter voltage. The SPID software [22] was used to quantify the following metabolites with AQSES-MRSI [23]: lipid (Lip), lactate (Lac), N-acetyl aspartate (NAA), glutamine+glutamate (Glx), total creatine (Cre), total choline (Cho). Maximum phase FIR filtering was applied for residual water suppression [24]. One or two voxel bands at the outer edges of the spectroscopic grid were omitted from the analyses when showing severe chemical shift displacement effects or lipid contamination.

All imaging modalities were coregistered and brought to the same spatial resolution of $1 \times 1 \times 3\text{mm}^3$, to allow for voxel-wise NMF analysis. Only voxels within the MRSI volume of interest were included in the NMF analysis, as only these voxels had a complete set of MRI features. 12 MRI imaging features were obtained from the raw acquired data after pre-processing: T1, T1+C, FLAIR, CBV, ADC, b0, Lip, Lac, NAA, Glx, Cre, Cho. In addition, 3×3 and 5×5 smoothing windows were applied to the 6 non-MRSI image slices, and these averaged features were also added to the feature set to improve robustness of the segmentation results [15]. A total of 24 MRI features was finally obtained for each voxel, making up the columns of the input matrix X.

3.3 Validation

Segmentation of the pathologic tissue types, i.e. actively growing tumor, necrosis and edema, was obtained from NMF by applying k-means clustering to the abundance values of H. The same procedure was also applied to H_0 obtained from SPA, to assess the added value of using SPA as an initialization method rather than a direct source extraction tool. As it is assumed that each source and its associated abundances correspond to one tissue type, we initialized the cluster centroids by setting one abundance value to 1 and all others to 0. The NMF segmentations are compared to manual segmentation performed by an experienced radiologist using MRIcron [25]. The overlap between segmentations was quantified using the Dice-score:

$$Dice_{tissue} = 2 \times \frac{A_{tissue,NMF} \cap A_{tissue,radiol}}{A_{tissue,NMF} \cup A_{tissue,radiol}} \quad (6)$$

with $A_{tissue,NMF}$ the area segmented by NMF and $A_{tissue,radiol}$ the area segmented by the radiologist for the same tissue type. Dice-scores were calculated for 3 tissue classes, as defined in the BraTS challenge [26]: active tumor, the tumor core (i.e. active tumor and necrosis) and the whole tumor (tumor core and edema).

To assess the sensitivity of the NMF methods to the initialization strategy, 2 types of correlation coefficients were calculated. Firstly, correlation coefficients were calculated for each pathologic tissue type between the initialized source obtained directly from SPA and the corresponding final source after running NMF_{SPA}. Secondly, we correlated each pathologic tissue source obtained from NMF_{SPA} to the corresponding tissue source obtained from $NMF_{rRandom}$. Correlation coefficients were calculated as the internal product of the normalized source vectors.

4 Results

Figure 1 shows an example of a coregistered image set of a glioblastoma patient. MRSI data were only available within the region of interest, which is marked in green. The segmentation results obtained with the 4 different NMF methods using SPA are also shown (in blue), and are compared with the manual segmentation by a radiologist (in green). Only active tumor and necrosis were found in the pathologic region of this patient. The single-level NMF methods show some active tumor segmentation outside of the pathologic region, which is less the case for hNMF. All NMF methods obtained good segmentation correspondance for necrosis, except that aHALS NMF showed some underestimation of the necrotic area. Table 1 shows the mean Dice-scores with their standard deviation for the 4 NMF methods with SPA and rRandom initialization. The highest Dice-scores are found for hNMF, which also shows relatively low standard deviations compared to the other NMF methods. For each NMF method, the results are very similar between NMF_{SPA} and $NMF_{rRandom}$ for the 3 tissue classes. Statistical significance of the differences could not be shown in any case using a two-tailed Wilcoxon signed rank test. Segmentation results were also computed using the SPA output matrices W_0 and H_0 directly. The Dice-scores from SPA were lower or at best equal to those from the NMF_{SPA} methods for each tissue class. Statistical significance of the higher Dice-scores for the NMF_{SPA} methods compared to direct SPA was found in about half of the cases.
Table 2 reports the mean correlation coefficients for the pathologic tissue sources. $Corr_{init}$ represents the correlation coefficients between SPA output and NMF_{SPA}, and $Corr_{NMF}$ gives the correlation values between the sources obtained from NMF_{SPA} and $NMF_{rRandom}$. For $Corr_{init}$, the lowest correlation coefficients

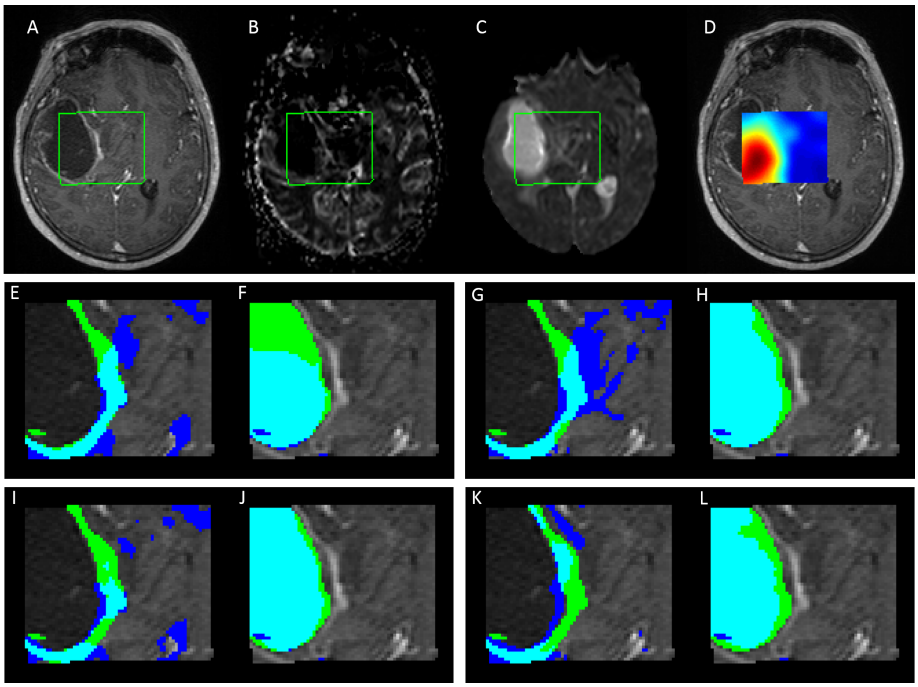

Figure 1: Coregistered set of MRI images, showing T1+C(A), CBV(B), ADC(C) and Lac(D). Segmentation results are shown below for active tumor and necrosis, respectively, for aHALS NMF(E,F), Convex NMF(G,H), GD NMF(I,J) and hNMF(K,L) with SPA initialization. NMF segmentation is shown in blue, manual segmentation in green and segmentation overlap in cyan.

	Dice active tumor [%]		**Dice tumor core [%]**		**Dice whole tumor [%]**	
	SPA	rRandom	SPA	rRandom	SPA	rRandom
aHALS NMF	65 ± 13	64 ± 19	76 ± 11	76 ± 11	78 ± 12	78 ± 14
Convex NMF	64 ± 18	63 ± 20	74 ± 14	73 ± 15	$84 \pm 10^*$	82 ± 13
GD NMF	65 ± 21	63 ± 19	$75 \pm 14^*$	75 ± 11	$80 \pm 11^*$	79 ± 13
hNMF	$69 \pm 15^*$	70 ± 14	$78 \pm 12^*$	78 ± 12	$86 \pm 8^*$	85 ± 8
SPA (W_0,H_0)	64 ± 20	-	73 ± 13	-	77 ± 13	-

Table 1: Comparison of the mean Dice-scores and their standard deviation between SPA and rRandom initialization.* indicates statistically significantly higher Dice-scores compared to direct SPA (bottom row), using a one-tailed Wilcoxon signed rank test ($p < 0.05$)

.

are found for aHALS NMF. The other 3 NMF methods have correlation values which are all close to or above 0.90. Looking at $Corr_{NMF}$, we see that correlation coefficients close to 1 are found for aHALS NMF, Convex NMF and hNMF. Only for GD NMF, $Corr_{NMF}$ values are found below 0.90.

	$\mathbf{Corr_{init}}$			$\mathbf{Corr_{NMF}}$		
	Active tumor	Necro-sis	Edema	Active tumor	Necro-sis	Edema
aHALS NMF	0.84	0.80	0.69	0.98	0.98	0.96
Convex NMF	0.94	0.91	0.88	0.99	0.99	0.97
GD NMF	0.96	0.92	0.89	0.89	0.85	0.83
hNMF	0.90	0.90	0.91	0.98	0.98	0.98

Table 2: Mean correlation coefficients for the pathologic tissue sources. $Corr_{init}$ represents correlation between SPA output and NMF_{SPA}. $Corr_{NMF}$ represents correlation between the sources obtained from NMF_{SPA} and $NMF_{rRandom}$

Table 3 shows the mean number of iterations to convergence as well as the mean time to convergence with both initialization methods. These values could not directly be reported for hNMF, as it consists of 2 levels of aHALS NMF. The same stopping criteria were applied to all NMF methods: a maximum number of iterations of 10000 and a convergence tolerance of 10^{-5} for the relative difference of two subsequent values of the residual error $\|X - WH\|_F$. For *rRandom*, the number of iterations is reported for the selected run, i.e. the run with the lowest final residual error. aHALS NMF reaches fast convergence, with the mean number of iterations below 200 for both initialization methods. Convergence speed is low for Convex NMF, where the mean number of iterations is close to 9000 both with SPA and rRandom initialization. For GD NMF, convergence speed is clearly higher with SPA, with a mean number of 218 iterations compared to a mean of 418 iterations with rRandom. Mean convergence times are clearly lowest for aHALS NMF. With rRandom, convergence times are dramatically higher than with SPA, mainly because for rRandom we run 30 NMF analyses with random initialization. Figure 2 illustrates the convergence behavior of the 3 single-level NMF methods when comparing SPA and rRandom initialization for a particular glioblastoma patient. The main reduction of the residual error occurs within the first 100 iterations in all cases, except for Convex NMF with rRandom. Convergence is particularly fast with SPA compared to rRandom for GD NMF, although a lower final residual error is obtained with rRandom.

5 Discussion

This study illustrated the comparison of SPA to repetitive random initialization for different NMF methods. We have compared both initialization methods in terms of their direct performance on brain tumor segmentation. As was noted

	#Iterations		Convergence time [s]	
	SPA	rRandom	SPA	rRandom
aHALS NMF	179	190	2	46
Convex NMF	8923	8810	69	2051
GD NMF	218	418	48	2070

Table 3: Mean number of iterations to reach convergence and mean time to convergence. Convergence tolerance was set to 10^{-5} and the maximum number of iterations to 10000.

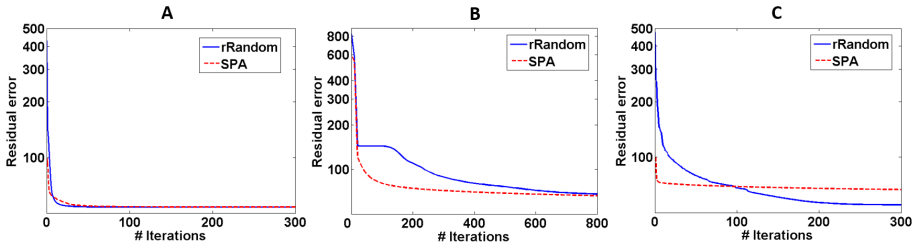

Figure 2: Convergence plots for aHALS NMF (A), Convex NMF (B) and GD NMF (C) with rRandom and SPA initialization. The residual error, $\|X - WH\|_F$, is shown on a log scale. For rRandom, the shown curve corresponds to the selected random run with the lowest residual error.

in [1], there is no consensus on how to assess NMF performance when comparing different initialization methods. Many studies report the final residual error as a measure of performance [3, 2, 7], which has the advantage of being a generally applicable measure. Other studies use a validation which is specific to the problem at hand. Ortega-Martorell et al. calculated correlation coefficients as a quality measure of the obtained tissue sources for brain tumor characterization using single-voxel magnetic resonance spectroscopic data [27]. Wild et al., besides looking at the residual error, assessed the quality of image sources for a facial image reconstruction problem based on visual inspection [5]. We have compared both initialization methods in terms of their direct performance for brain tumor segmentation using the Dice-scores, because a lower residual error does not warrant a better parts-based solution. For instance, one can observe in figure 2 that higher residual errors remain for Convex NMF with both initialization methods compared to aHALS and GD NMF. This is due to the additional constraint imposed by Convex NMF that the sources have to be a weighted sum of the input data. Such sparsity enforcing and other constraints will generally lead to higher residual errors, whereas the quality of the result in terms of interpretability might benefit from it. Therefore we decided to assess the quality of the NMF result using direct validation of the problem at hand, i.e. segmentation of the pathologic tumor regions. For all considered NMF methods, mean Dice-scores were very similar between SPA and rRandom and differences were not higher than 2%. Statistical significance of the differences

has not been found in any of the cases, showing that SPA achieves similar performance compared to rRandom in terms of segmentation quality. For most advanced initialization methods, similar performance has been reported compared to random initialization, although residual errors generally tend to be rather higher than lower. Wild et al. discussed how most advanced methods obtain a head start in terms of residual error compared to random initialization, but this advantage is mostly lost towards the end of the iteration process [5]. In some cases, random initialization will achieve a lower error towards the end of the convergence process, which is also illustrated in [5] and in figure 2 for GD NMF. This is explained by the fact that advanced initialization often pushes the NMF result in a certain direction, thereby imposing restrictions on the factorization. It is therefore important to investigate the ability of different NMF methods to pull the factorization out of a local minimum.

Although a head start in terms of the initial error has been reported for most advanced initialization methods compared to random initialization, mixed results have been found in terms of the number of iterations to reach convergence [7]. As shown in table 3, we found increased speed of convergence with SPA for GD NMF, whereas for aHALS NMF convergence is similarly fast with both initialization methods. For Convex NMF, SPA was not able to speed up convergence, and the mean number of iterations is similarly high with both initialization methods. It has to be pointed out that the number of iterations is not the only factor determining the computational cost of the NMF analysis. The initialization method also requires some processing time, which might become high compared to the NMF analysis itself, such as with several clustering algorithms [7]. SPA is a straightforward and computationally inexpensive method, taking about 0.01 seconds per run with the input data considered in this paper (i.e. input matrix X with 24 rows and approximately 30000 columns for each patient). SPA is also suitable for handling large-scale problems, requiring a total of $\mathcal{O}(mnr)$ operations [9]. However, the initialization proposed in this study is relatively expensive due to the nonnegative least squares fitting to obtain H_0. Many initialization methods only provide W_0, and only few suggestions are available from literature on how to obtain H_0 in that case. Random initialization is sometimes used for H_0 in combination with a more advanced initialization of W_0 [2], although this partially annihilates the advantages of using a fixed and input-specific initialization. Clustering based methods might construct H_0 as the binary partition matrix [7] or the fuzzy partition matrix when using fuzzy C-means [6]. For certain NMF methods, such as most ALS methods, it is however sufficient to have W_0, i.e. no estimation of H_0 is required. Another time-related factor is the selected NMF method itself. The GD NMF method, despite being Gauss-Newton based and therefore having only first order terms, has a high cost per iteration compared to the other NMF methods. As a result, GD NMF shows high mean convergence times in table 3 similar to Convex NMF, even though it requires much less iterations to reach convergence. aHALS NMF, requiring a low computation time per iteration and reaching convergence at a relatively low number of iterations, is clearly the fastest method. Convergence times are dra-

matically higher for rRandom compared to SPA, as we ran NMF 30 times with random initialization. Obviously, convergence times for rRandom will depend on the number of NMF runs. For very large datasets, it might not be feasible to run NMF a high number of times with random initialization. Nevertheless, even with only a few random runs, there will be less variability and a lower probablity of obtaining a poor result due to a particular random initialization.

As some NMF methods might be more prone to getting stuck in a local minimum close to the initialization, we also investigated the sensitivity of different NMF methods to the initialization. Some NMF algorithms might be more prone to getting stuck in a local minimum close to the initialization than others. For instance, in the original multiplicative update NMF algorithm proposed by Lee and Seung [28], any element initialized at zero in W_0 or H_0 will remain zero in the final factorization. Looking at the correlation coefficients in table 2, a different behavior is found for the 3 single-level NMF methods. For aHALS NMF, relatively low values are found for $Corr_{init}$ and values close to 1 for $Corr_{NMF}$. This means that the sources from NMF_{SPA} do not closely correspond to the initial estimates obtained from SPA, whereas the sources obtained from NMF_{SPA} and $NMF_{rRandom}$ are very similar. It shows that aHALS NMF is rather insensitive to the initialization method, and easily moves away from the SPA first estimate. This behaviour was also found for a basic ALS NMF algorithm applied to text document datasets in [7]. For Convex NMF, higher values are obtained for $Corr_{init}$ and high values for $Corr_{NMF}$ as well. As Convex NMF enforces sparsity by requiring the sources to be a weighted sum of the input data, and SPA also selects data points as the vertices of the convex hull, it is to be expected that the NMF_{SPA} sources are more closely related to the initialized sources. However, this does not mean that Convex NMF has difficulty pulling the factorization out of a local minimum, as we can see that $NMF_{rRandom}$ generates sources which are strongly correlating with NMF_{SPA}. Convex NMF naturally generates factorizations that are more closely related to the SPA initialization than aHALS NMF. For GD NMF, high values similar to Convex NMF are found for $Corr_{init}$, but relatively low values are found for $Corr_{NMF}$ compared to aHALS and Convex NMF. The latter means that NMF_{SPA} and $NMF_{rRandom}$ converge to clearly different solutions. This can also be seen in figure 2, where the residual error is higher for GD NMF_{SPA} than for $NMF_{rRandom}$. Considering the high values for $Corr_{init}$, these results suggest that GD NMF, being a local optimization algorithm based on gradient descent, is more sensitive to the initialization method than the other NMF methods. However, we did not see any loss in performance for GD NMF_{SPA} in terms of the mean Dice-scores. hNMF, consisting of 2 levels of aHALS NMF, shows high $Corr_{NMF}$ values just as single-level aHALS, indicating similar solutions for NMF_{SPA} and $NMF_{rRandom}$. On the other hand, $Corr_{init}$ values were clearly higher for hNMF compared to aHALS NMF, probably due to the fact that hNMF produces more sparse factorizations than aHALS [15], which are more closely related to the initialized sources from SPA.

Although SPA has been used as an endmember extraction algorithm in some NMF studies [9, 10], it has so far not been considered as an initialization strategy. We hypothized that there might be some benefit to assigning the SPA sources to W_0 instead of directly using them as the final columns of W, because SPA will not always directly provide the most suitable sources for the NMF problem at hand. One of the main limitations of SPA is its sensitivity to outliers. As SPA looks for the outer vertices of a geometrical hull, outliers might be wrongly assigned as endmembers. Several methods have been proposed to either make SPA less sensitive to outliers [10] or to discard the outliers beforehand [29]. Another important caveat is that SPA assumes the NMF problem to be near-separable [9], which is equivalent to the spatial representation of the input data by a convex hull. The assumptions which are intrinsic to near-separable NMF, i.e. the pure-pixel assumption and the sum-to-one constraint on the abundances, will often not hold, especially in the case of highly mixed NMF problems. But when applying SPA as an initialization method, the NMF procedure might still correct the source estimates, especially when the conditions of near-separability are only mildly violated. For the multi-parametric MRI dataset considered in this study, the pure-pixel assumption holds for most of the MRI modalities due to their high spatial resolution. However, the poor resolution of the MRSI data leads to significant partial volume effects, such that the pure-pixel assumption does not hold for these data. We can further assume that the sum-to-one constraint approximately holds, as each voxel represents a fixed tissue volume. By comparing the Dice-scores of NMF_{SPA} to those obtained directly from SPA output matrices W_0 and H_0, it was shown that we can obtain better segmentation results when using SPA as an initialization method rather than as a direct source extraction tool. Average SPA Dice-scores were lower or at best equal to the Dice-scores obtained with any of the NMF methods and for all tissue classes. Statistical significance of the higher Dice-scores for NMF_{SPA} could be shown in about half of the cases.

6 Conclusion

This study has proposed SPA as an initialization method for NMF. Whereas SPA has previously been applied as a direct endmember extraction algorithm, using it as an initialization method might allow further enhancement of the sources during the NMF iterative procedure. This advantage has been shown on a multi-parametric MRI dataset for brain tumor segmentation. Compared to other initialization methods, SPA is fast and it provides a realistic and reproducible initialization. We have compared SPA performance to repetitive random initialization, such that the reported results are reproducible and unambiguous. SPA shows similar performance in terms of segmentation quality of the pathologic tissue classes, it reaches similar or increased convergence speed and it requires much less computation time than rRandom.

7 Acknowledgements

This work has been funded by the following projects: Flemish Government FWO project G.0869.12N (Tumor imaging); IWT IM 135005; Henri Benedictus Fellowship of the Belgian American Educational Foundation; Interuniversity Attraction Poles Program (P7/11) initiated by the Belgian Science Policy Office. European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013), ERC Advanced Grant: BIOTENSORS (n°339804) and the EU MC ITN TRANSACT 2012 (n°316679).

References

[1] C. Boutsidis and E. Gallopoulos, "SVD based initialization: A head start for nonnegative matrix factorization," *Pattern Recognition*, vol. 41, no. 4, pp. 1350–1362, 2008.

[2] A. N. Langville, C. D. Meyer, R. Albright, J. Cox, and D. Duling, "Initializations for the nonnegative matrix factorization," in *Proceedings of the twelfth ACM SIGKDD international conference on knowledge discovery and data mining*. Citeseer, 2006, pp. 23–26.

[3] H. Qiao, "New SVD based initialization strategy for non-negative matrix factorization," *Pattern Recognition Letters*, vol. 63, pp. 71–77, 2014.

[4] L. Zhao, G. Zhuang, and X. Xu, "Facial expression recognition based on PCA and NMF," in *Intelligent Control and Automation, 2008. WCICA 2008. 7th World Congress on*. IEEE, 2008, pp. 6826–6829.

[5] S. Wild, J. Curry, and A. Dougherty, "Improving non-negative matrix factorizations through structured initialization," *Pattern Recognition*, vol. 37, no. 11, pp. 2217–2232, 2004.

[6] Z. Zheng, J. Yang, and Y. Zhu, "Initialization enhancer for non-negative matrix factorization," *Engineering Applications of Artificial Intelligence*, vol. 20, no. 1, pp. 101–110, 2007.

[7] G. Casalino, N. Del Buono, and C. Mencar, "Subtractive clustering for seeding non-negative matrix factorizations," *Information Sciences*, vol. 257, pp. 369–387, 2014.

[8] M. C. U. Araújo, T. C. B. Saldanha, R. K. H. Galvão, T. Yoneyama, H. C. Chame, and V. Visani, "The successive projections algorithm for variable selection in spectroscopic multicomponent analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 57, no. 2, pp. 65–73, 2001.

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

[10] J. Zhang, B. Rivard, and D. Rogge, "The successive projection algorithm (SPA), an algorithm with a spatial constraint for the automatic search of endmembers in hyperspectral data," *Sensors*, vol. 8, no. 2, pp. 1321–1342, 2008.

[11] N. Gillis and F. Glineur, "Accelerated multiplicative updates and hierarchical ALS algorithms for nonnegative matrix factorization," *Neural Computation*, vol. 24, no. 4, pp. 1085–1105, 2012.

[12] C. Ding, T. Li, and M. Jordan, "Convex and semi-nonnegative matrix factorizations," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 1, pp. 45–55, 2010.

[13] Y.-X. Wang and Y.-J. Zhang, "Nonnegative matrix factorization: A comprehensive review," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 25, no. 6, pp. 1336–1353, 2013.

[14] L. Sorber, M. V. Barel, and L. D. Lathauwer, "Unconstrained optimization of real functions in complex variables," *SIAM Journal on Optimization*, vol. 22, no. 3, pp. 879–898, 2012.

[15] N. Sauwen, D. M. Sima, S. Van Cauter, J. Veraart, A. Leemans, F. Maes, U. Himmelreich, and S. Van Huffel, "Hierarchical non-negative matrix factorization to characterize brain tumor heterogeneity using multi-parametric MRI," *NMR in Biomedicine*, vol. 28, no. 12, pp. 1599–1624, 2015.

[16] Y. Li, D. M. Sima, S. V. Cauter, A. R. Croitor Sava, U. Himmelreich, Y. Pi, and S. Van Huffel, "Hierarchical non-negative matrix factorization (hNMF): a tissue pattern differentiation method for glioblastoma multiforme diagnosis using MRSI," *NMR in Biomedicine*, vol. 26, no. 3, pp. 307–319, 2013.

[17] S. Arora, R. Ge, R. Kannan, and A. 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.

[18] S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies," *Physics in medicine and biology*, vol. 58, no. 13, p. R97, 2013.

[19] P. Sajda, S. Du, T. R. Brown, R. Stoyanova, D. C. Shungu, X. Mao, and L. C. Parra, "Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain," *Medical Imaging, IEEE Transactions on*, vol. 23, no. 12, pp. 1453–1465, 2004.

[20] J. Boxerman, K. Schmainda, and R. Weisskoff, "Relative cerebral blood volume maps corrected for contrast agent extravasation significantly correlate with glioma tumor grade, whereas uncorrected maps do not," *American Journal of Neuroradiology*, vol. 27, no. 4, pp. 859–867, 2006.

[21] J. Veraart, J. Sijbers, S. Sunaert, A. Leemans, and B. Jeurissen, "Weighted linear least squares estimation of diffusion MRI parameters: strengths, limitations, and pitfalls," *Neuroimage*, vol. 81, pp. 335–346, 2013.

[22] J.-B. Poullet, "Quantification and classification of magnetic resonance spectroscopic data for brain tumor diagnosis," Ph.D. dissertation, Leuven, Belgium, 2008.

[23] A. R. Croitor Sava, D. M. Sima, J.-B. Poullet, A. J. Wright, A. Heerschap, and S. Van Huffel, "Exploiting spatial information to estimate metabolite levels in two-dimensional MRSI of heterogeneous brain lesions," *NMR in Biomedicine*, vol. 24, no. 7, pp. 824–835, 2011.

[24] J.-B. Poullet, D. M. Sima, S. Van Huffel, and P. Van Hecke, "Frequency-selective quantitation of short-echo time 1H magnetic resonance spectra," *Journal of Magnetic Resonance*, vol. 186, no. 2, pp. 293–304, 2007.

[25] C. Rorden, H.-O. Karnath, and L. Bonilha, "Improving lesion-symptom mapping," *Journal of cognitive neuroscience*, vol. 19, no. 7, pp. 1081–1088, 2007.

[26] B. Menze, M. Reyes, K. Van Leemput *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *Medical Imaging, IEEE Transactions on*, vol. 34, no. 10, pp. 1993–2024, 2015.

[27] S. Ortega-Martorell, P. J. Lisboa, A. Vellido, M. Julià-Sapé, and C. Arús, "Non-negative matrix factorisation methods for the spectral decomposition of MRS data from human brain tumours," *BMC bioinformatics*, vol. 13, no. 1, p. 38, 2012.

[28] 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.

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