Initializing Nonnegative Matrix Factorization using the Successive Projection Algorithm for multi-parametric medical image segmentation

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. As nonnegative matrix factorization (NMF) represents a non-convex problem, the quality of its solution will depend on the initialization of the factor matrices. This study proposes the Successive Projection Algorithm (SPA) as a feasible NMF initialization method. SPA is applied to a multi-parametric MRI dataset for automated NMF brain tumor segmentation. SPA provides fast and reproducible estimates of the tissue sources, and segmentation quality is found to be similar compared to repetitive random initialization.

1 Introduction

Non-negative matrix factorization (NMF) has become a popular tool for multivariate data analysis in a wide range of applications, such as image analysis, text mining, audio processing, computer vision, and so on. NMF provides a low-rank (rank r) approximation of a nonnegative input matrix X:

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

Factor matrices W and H are found by solving the following non-linear optimization problem:

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

The columns of W correspond to the basic components, or sources, present in the input data. Each row of H contains the weights, or abundances, for one particular source. NMF has the advantage of retaining the nonnegative structure and providing interpretable factor matrices.

Since NMF is a non-convex optimization problem, the NMF result will depend on the initialization of the factor matrices, W_0 and H_0. Random initialization is by far the most commonly used method, although it does not deliver reproducible results and there is no guarantee of a qualitative solution. More advanced initialization strategies have been suggested. Some methods rely on a random selection of columns of the input matrix [1]. These methods have a low computational cost, but don't provide reproducible results. SVD-based initialization methods suffer from the orthogonality constraint imposed on the singular vectors, resulting in negative values. Some straightforward ways of dealing with these negative values have been proposed, such as setting them to zero or replacing them by a mean value from the input matrix [2]. Clustering-based initialization methods often require some initialization themselves, and they can become computationally expensive [3, 4]. In this study, NMF is applied to a set of multi-parametric Magnetic Resonance Imaging (MRI) data for automated image segmentation of brain tumors. We introduce the Successive Projection Algorithm (SPA) [5] as an initialization strategy for NMF. SPA results will be compared to repetitive random initialization, i.e. running NMF with random initialization a sufficient number of times, such that the comparison is unambiguous.

2 Materials and methods

2.1 MRI acquisition and pre-processing

21 high-grade glioma patients underwent a multi-parametric MRI acquisition protocol on a 3T Siemens Trio Tim scanner (Erlangen, Germany). The local ethics committee allowed a retrospective analysis of the data. The acquisition protocol consisted of conventional MRI, perfusion-weighted MRI, diffusion-weighted MRI and MR spectroscopic imaging. A description of the acquisition protocol and the obtained MRI parameters can be found in the *Appendix*. All imaging modalities were coregistered and brought to the same spatial resolution of $1 \times 1 \times 3\text{mm}^3$, for voxel-wise NMF analysis. Only voxels within the spectroscopic volume of interest were included. 12 MRI features were obtained from the acquired data after pre-processing. In addition, 3×3 and 5×5 smoothing windows were applied to the 6 non-spectroscopic image sets, and these averaged features were also added to the feature set to improve robustness of the segmentation [6]. We expect to find tissue-specific signatures on the columns of W, and the abundances per tissue type on the corresponding rows of H.

2.2 NMF methods

Two NMF methods were considered: accelerated hierarchical alternating least squares NMF (aHALS NMF) and hierarchical NMF (hNMF). aHALS NMF [7] is a member of the family of alternating least squares (ALS) NMF methods. These methods iteratively solve the convex subproblems of finding H when W is fixed, and vice-versa, using ALS. aHALS has shown improved computational efficiency compared to other ALS methods [7]. hNMF has been introduced in

previous MRI studies on brain tumor characterization [6]. 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. hNMF has shown improved differentiation and segmentation of the pathologic tissue types in brain tumors compared to aHALS NMF [6].

2.3 Successive projection algorithm

Under the assumptions of near-separable NMF [5], the n data points of $X \in R_+^{m \times n}$ will span a convex hull in m-dimensional space. In the context of NMF, the vertices of the convex hull correspond to the sources. SPA fits well with this geometrical interpretation, as it aims at finding the vertices [5]. The data point x_1 with the highest l_2-norm is selected as the first vertex. All data points are then projected onto the orthogonal complement of x_1. The data point with the highest l_2-norm in this projected subspace will be the second vertex. In a similar way, the third vertex is found as the data point with the highest l_2-norm in the orthogonal subspace of the first 2 vertices, and so on. The algorithm stops when r vertices have been found, with r equal to the NMF rank. The vertices are assigned to the columns of W_0. 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. 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. Empirically, we found that 30 random runs were sufficient for the NMF problem at hand to obtain reproducible results. For rRandom, the final NMF result is selected from the random run with the lowest residual error as defined in Equation 2. For the remainder of this article, NMF analysis using SPA initialization will be denoted as NMF_{SPA} and NMF using rRandom as $NMF_{rRandom}$.

2.4 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. 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. NMF segmentations are compared to manual segmentation by an experienced radiologist. Segmentation overlap was quantified using the Dice-score [8]. Dice-scores were calculated for 3 tissue classes, as defined in the BraTS challenge [8]: active tumor, the tumor core (i.e. active tumor and necrosis) and the whole tumor (tumor core and edema).

To assess the quality of the NMF sources, correlation coefficients were calculated between the normalized source and the averaged feature vector of the corresponding tissue region segmented by the radiologist. Correlation coefficients were calculated for active tumor and necrosis.

3 Results

Table 1 shows the mean Dice-scores and correlation coefficients for both NMF methods with SPA and rRandom initialization. Higher Dice-scores are found for hNMF compared to aHALS. For both NMF methods, the results are very similar between NMF_{SPA} and $NMF_{rRandom}$, with differences in mean Dice-score not exceeding 1% for any of the tissue classes. Statistical significance of the differences could not be shown using a two-tailed Wilcoxon signed rank test. For the correlation coefficients, hNMF shows higher values compared to aHALS. With correlation values close to 0.90, the hNMF sources can be considered as tissue-specific patterns. Differences between NMF_{SPA} and $NMF_{rRandom}$ were also low for the correlation coefficients, and statistical significance could not be shown in any of the cases.

	aHALS NMF		**hNMF**	
	SPA	rRandom	SPA	rRandom
Dice active tumor [%]	65 ± 13	64 ± 19	69 ± 15	70 ± 14
Dice tumor core [%]	76 ± 11	76 ± 11	78 ± 12	78 ± 12
Dice whole tumor [%]	78 ± 12	78 ± 14	86 ± 8	85 ± 8
Corr active tumor	0.78 ± 0.17	0.81 ± 0.09	0.89 ± 0.09	0.90 ± 0.07
Corr necrosis	0.78 ± 0.15	0.77 ± 0.11	0.91 ± 0.07	0.90 ± 0.07

Table 1: Comparison of the mean Dice-scores and correlation coefficients and their standard deviation between SPA and rRandom initialization.

4 Discussion and conclusions

Unlike most studies about NMF initialization, we did not assess NMF performance in terms of the final residual of the cost function. As was pointed out in [4], a lower residual error does not warrant a better parts-based solution. We assessed the quality of the NMF result using direct validation of the problem at hand: the NMF sources were validated by means of the correlation coefficients, and the NMF abundances by means of the Dice-scores. Mean Dice-scores and correlation coefficients were very similar between SPA and rRandom and statistical significance of the differences could not be shown for any tissue class. Several studies have reported similar performance for advanced initialization methods compared to random initialization in terms of the residual error, although the errors were generally slightly higher for the advanced methods [4, 1]. A more advanced initialization method can be more restrictive to the final solution, as the NMF method might not be able to pull the factorization out of

a local minimum [3]. This means that a realistic first estimate of the sources is important when using advanced initialization, and the sensitivity of the NMF method to the initialization will also play a role. In terms of computation time, SPA is fast, taking about 0.01 seconds to estimate the sources for one patient's multi-parametric MRI dataset (i.e. input matrix X containing 24 rows and approximately 30000 columns). In contrast, other initialization methods such as co-occurrence initialization [1] and most clustering-based methods [3] require computation times which are comparable or even higher than for the NMF analysis itself.

It is important to note that SPA assumes the NMF problem to be near-separable [5], which is equivalent to the spatial representation of the input data by a convex hull. When this assumption is only approximately valid, or especially in the case of highly-mixed NMF problems, it is improbable that SPA directly provides the right sources. Therefore, it seems feasible to apply SPA as an initialization strategy rather than as a direct endmember extraction tool. Unlike most previous studies, we have considered repetitive random initialization as a reference, such that the reported results are reproducible. Compared to rRandom, SPA shows similar performance in terms of segmentation quality of the pathologic tissue classes.

5 Acknowledgements

This work has been funded by: Flemish Government FWO project G.0869.12N; IWT IM 135005; Henri Benedictus Fellowship of the Belgian American Educational Foundation; Interuniversity Attraction Poles Program (P7/11) by the Belgian Science Policy Office. ERC 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] 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.

[2] 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.

[3] 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.

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

[5] 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.

[6] 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*, 2015 [In press].

[7] 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.

[8] 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.

6 Appendix

Conventional imaging consisted of a 3-dimensional T1-weighted gradient-echo sequence (MPRAGE) (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 fluid inversion recovery sequence (FLAIR) (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).
Perfusion-weighted imaging was performed using a T2*-weighted echo-planar imaging sequence (TR = 1000ms, TE = 29ms, FOV = $230 \times 230\text{mm}^2$, matrix size = 128×128, voxel size = $1.8 \times 1.8 \times 5\text{mm}^3$, GRAPPA factor 2, flip angle = 90°). 90 multi-section acquisitions were acquired at 1 second intervals. A 0.1mmol/kg body weight bolus of gadobutrol (Gadovist, Bayer) was injected at a rate of 4ml/s, followed by a 20ml bolus of sodium chloride solution at 4ml/s. Cerebral Blood Volume maps were calculated with DSCoMAN (Duke University, Durham, NC).
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, number of averages = 3). Apparent Diffusion Coefficient maps were derived from the 3 b-values using weighted linear least squares fitting.
3D proton magnetic resonance spectroscopic imaging with long TE was included. A volume of interest of $80 \times 80\text{mm}^2$ including tumour and normal brain tissue was positioned based on the FLAIR sequence. 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. The following metabolites were quantified as described in [6]: lipids, lactate, N-acetyl aspartate, glutamine+glutamate, creatine and choline.