

Article

A Review on Initialization Methods for Nonnegative Matrix Factorization: Towards Omics Data Experiments

Flavia Esposito [†] 

Department of Mathematics, University of Bari Aldo Moro, 70125 Bari, Italy; flavia.esposito@uniba.it
[†] Member of INDAM-GNCS research group.

Abstract: Nonnegative Matrix Factorization (NMF) has acquired a relevant role in the panorama of knowledge extraction, thanks to the peculiarity that non-negativity applies to both bases and weights, which allows meaningful interpretations and is consistent with the natural human part-based learning process. Nevertheless, most NMF algorithms are iterative, so initialization methods affect convergence behaviour, the quality of the final solution, and NMF performance in terms of the residual of the cost function. Studies on the impact of NMF initialization techniques have been conducted for text or image datasets, but very few considerations can be found in the literature when biological datasets are studied, even though NMFs have largely demonstrated their usefulness in better understanding biological mechanisms with omic datasets. This paper aims to present the state-of-the-art on NMF initialization schemes along with some initial considerations on the impact of initialization methods when microarrays (a simple instance of omic data) are evaluated with NMF mechanisms. Using a series of measures to qualitatively examine the biological information extracted by a given NMF scheme, it preliminarily appears that some information (e.g., represented by genes) can be extracted regardless of the initialization scheme used.

Keywords: omic data analysis, nonnegative matrix factorization; initialization algorithm; gene extraction; qualitative analysis



Citation: Esposito, F. A Review on Initialization Methods for Nonnegative Matrix Factorization: Towards Omics Data Experiments. *Mathematics* **2021**, *9*, 1006. <https://doi.org/10.3390/math9091006>

Academic Editor: Christophe Guyeux

Received: 26 February 2021

Accepted: 25 April 2021

Published: 29 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Low-rank matrix dimensionality reduction mechanisms represent a class of unsupervised mathematical techniques dedicated to the principle of parsimony, capable of revealing the low-dimensional structure embedded in the original data while preserving as much information as possible. Relevant information stored in data is often non-negative, and this positive sign is strictly related to a physical entity (examples include pixels in images, the probability of a particular topic occurring in a linguistic document, the amount of pollutants emitted by a factory, and so on). Taking into account this non-negativity constraint could bring some advantages in terms of interpretability and visualization of big data, while better preserving physical feasibility. In the biomedical field, the availability of big omic data (genomics, transcriptomics, proteomics, etc.) has led to the emergence of the application of specific numerical mechanisms capable of extracting valuable and interpretable information about complex interactions between data to achieve a better understanding of the underlying biological processes.

Omic data analysis represents a very active research area in the biomedical field and recent advances in technologies have allowed the simultaneous measurement of long sequential molecules, the expression levels of a large numbers of genes and proteins, genetic variants, molecules, bio-markers, cells, tissue samples and individuals [1,2]. One way to manage these data is to investigate their natural representation in terms of a data matrix $X \in \mathbb{R}_+^{n \times m}$, of which the nonnegative elements X_{ij} measure some biological values (expression counts, protein concentrations, gene expression level, etc.) in its i -th row and an individual sample in its j -th column [3–5].

Low rank reduction mechanisms can be fruitfully used to better understand complex biological processes: in the early 2000s, they were first used in the analysis of microarray data [6–9]; gradually, they emerged in the literature as useful mechanisms to explore high-throughput omics data (i.e., uncovering their low-dimensional structure, such as groups of similar genes, interpretable subspaces, critical dimensions) and identifying in them new and known specific biological pathways [3,10–16].

A matrix factorization approximates a given pre-processed omic data matrix X to a low-dimensional space of dimension r as the product of two matrices $W \in \mathbb{R}^{n \times r}$ and $H \in \mathbb{R}^{r \times m}$, i.e., $X \approx WH$, where the first factor W describes the embedded structure between features, and the second matrix H quantifies the structure between samples. Generally speaking, each pair of a column in W and the corresponding row of H represents an ideal source of biological, experimental and technical variation and their relative roles in each sample (which is actually known as a “complex biological process” [3]). From a mathematical point of view, each original sample (one column vector $X(:, j)$ in X) is a linear combination of new column features $W(:, k)$, ($k = 1, \dots, r$) weighted with coefficients H_{kj} so that

$$X(:, j) \approx \sum_{k=1}^r W(:, k) H_{kj}, \quad (1)$$

for $j = 1, \dots, m$ and $r < nm / (n + m)$.

Various matrix factorizations have proven their effectiveness in handling omic data; the best known are: Singular Value Decomposition (SVD) [8,17], Principal Component Analysis (PCA) and its sparse and probabilistic variants [14], Independent Component Analysis (ICA) [18] and Nonnegative matrix factorizations (NMF) [3,19–23]. Each of these techniques is based on different constraints that characterize the final properties of the matrix factors, leading to different optimization problems and numerical algorithms that must be used. PCA is based on a convex quadratic optimization problem with a unique global minimum that leads to orthogonal matrix factors that determine the proportion of explained variance in the omic data. ICA and NMF are based on nonconvex optimization problems, so the numerical algorithms used to solve them provide solutions that depend on the initialization mechanism in question. ICA factors derive from the minimization of mutual information between data components. On the other hand, when dealing with omic data, NMF algorithms minimize the generalized Kullback–Leibler divergence (KL) under the non-negativity of the elements in W and H factors. This constraint produces an approximation of the omic data matrix as a nonnegative linear combination of the columns in W (commonly called *metagenes*), weighted by elements H_{kj} in H , and is compatible with the intuitive notion of combining parts into a whole [24]. Note that each H_{kj} describes the effect that the k th metagene has on the j th sample, so that a low value H_{kj} indicates that the corresponding k th metagene has reduced importance in approximating the j th sample. NMF allows to reveal interpretable latent factors (unlike PCA or ICA, which also have negative entries without obvious biological significance) and to identify genes belonging to multiple pathways or biological processes [20,22,23,25]. As these reasons lead to a much more intuitive and interpretable representation, NMF is quite often preferred to other techniques.

As previously observed, the NMF optimization problem

$$\min_{W, H \geq 0} KL(X; WH), \quad (2)$$

where $KL(X; WH)$ is the KL objective function

$$KL(X, WH) = \sum_{ij} (X_{ij} \log(\frac{X_{ij}}{(WH)_{ij}}) - X_{ij} + (WH)_{ij}) \quad (3)$$

has to be solved. However, this nonlinear objective function is nonconvex in both W and H , so iterative numerical algorithms had to be used. These algorithms guarantee to

converge to some local minima (more precisely, stationary points), but require initialization mechanisms that can greatly affect their convergence rate and often the "biological" quality of the obtained solution factors W and H .

Despite that several different initialization schemes for NMF have appeared in the literature (we direct the reader to [26] for a recent review of theoretical and practical aspects of NMF), very often, only studies on the performance of NMF algorithms with respect to the residual of the cost function are presented. Qualitative behavior of NMF factors with respect to the initialization mechanism has been studied when large text data or images are processed. On the contrary, when NMF algorithms are used to analyze omic data, random initialization is the most commonly used method, although this does not guarantee good quality of the extracted matrix factors from a biological point of view. A very preliminary study regarding the influence of some initialization techniques on a particular NMF algorithm appeared in [27]. Here, it is suggested that there might be some biological information strictly related to the microarray data under analysis that can be extracted independently of the initialization scheme in question.

In this paper, we have proposed a revised taxonomy of initialization methods for NMF, starting from the one presented in [28], highlighting which schemes are used in the analysis of omic data by NMF algorithm, together with their main advantages and disadvantages. Using a specific NMF algorithm and a selected subset of initialization schemes belonging to different types, we also analyzed a pair of benchmark microarray data matrices, showing that the relevance of the extracted information cannot be immediately understood as it is the case for text or image data, indicating that the influence of initialization and its impact on the extracted biological information are worth further investigation.

The remainder of the paper is organized as follows. In Section 2 we briefly highlight the importance of initialization for feeding iterative NMF algorithms. In Section 3 we illustrate the iterative algorithms for NMF, and we describe in detail the initialization schemes that have appeared in the literature panorama to date. Section 4 deals with the brief presentation of some results obtained when a selection of initialization schemes was used to feed a specific NMF algorithm for the analysis of two microarray datasets: the qualitative measurement showed how a core of information strictly related to the analyzed dataset (i.e., some genes) can be extracted independently of the particular initialization scheme used.

2. How Important are Initializations for NMF?

The task of nonnegative low-rank approximation of a given omic data matrix X can be formulated as a constrained (under the condition of nonnegativity of the factor matrices) optimization problem using the nonlinear KL-divergence (3). The lack of convexity of each NMF objective function in W and H simultaneously means that no closed-form solution exists and thus convergence to global optima of (2) is not necessarily achieved [29].

Numerical methods for solving (2) include: multiplicative update rules (MU) [9,24,30,31], which are characterized by good theoretical convergence properties [32], alternating non-negative least squares [33–35], and projected gradient descent methods [36]. The iterative nature of each of these numerical methods requires the use of an initialization of W or both W and H to matrices (of appropriate size) with only nonnegative elements.

However, the choice of initial values affects not only the convergence property of the algorithm but, in particular, the quality of the solution to which it converges. An illustrative example of the influence of initialization on the "quality" of the final factors is the Swimmer dataset presented in [37]. This image dataset consists of 256 black and white images, each representing some stick figures (with a fixed torso of 12 pixels) and four moving parts (four limbs of 6 pixels) positioned in four different ways. An extraction of this dataset is shown in Figure 1.

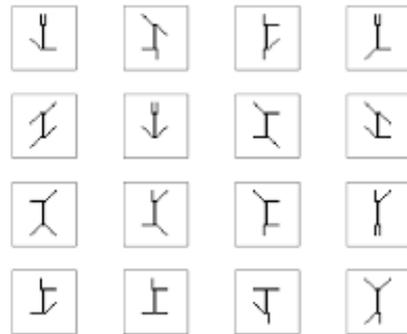


Figure 1. Sample images from the Swimmer database. Each panel depicts a stick figure with a fixed torso and four limbs articulated in different ways (positions) [37].

This dataset highlights the importance of good initial values when solving the NMF task correlated with these images. As discussed in [38], when applying, for example, multiplicative updating rules and the projected gradient algorithm with a random initialization, the final NMF factors fail to properly extract the latent parts (torso and limbs) embedded in the figures, thus mixing the ghostly appearance of the torso with some of the other parts. Due to nonconvexity, iterative algorithms are not guaranteed to converge to the global optimum, and they are equally dependent on initialization. For a simple image dataset satisfying the separability rule theorized in [37], random initialization has been shown to lead to solutions that are not fully satisfactory in terms of interpretability and NMF part-based decomposition [24]. Other types of structured initialization algorithms might give better results in some cases [38].

Nevertheless, comparisons between the results obtained when using different initialization methods have been under-researched. Some initialization schemes have been qualitatively compared to demonstrate the best clustering performance for different face and text datasets [39], others improved face component separation [40], or better separation performance in the audio source separation task [41].

However, when NMFs are used to analyze microarrays or more general omic data, the problem of choosing an appropriate initialization becomes more complicated because the data have special meanings, and very often, only random initialization mechanisms are chosen. In [42], an integrative approach for disease subtype classification based on NMF was proposed. Here, the most appropriate final NMF factors were chosen as those that produce the smallest value of the objective function among numerous local minima obtained by multiple random initializations of W and H . In particular, a random initialization of W was obtained from a uniform distribution or using the SVD-based initialization algorithm [43]. However, final comparisons were made only in terms of the performance of the objective function. In [44], the proposed hierarchical alternating least squares algorithm for solving NMF tasks was initialized with respect to single cell datasets using either a random method and a selection of r columns randomly sampled from the corresponding input data. Comparisons accounting for the variations due to these two different initializations showed that they performed similarly, although the qualitative representations of the final solutions revealed some differences between them. In [45], clustering of multiple types of genomic data was approached via an nNMF algorithm that initialized the factors W s with (i) a uniformly randomly generated matrix and (ii) a nonnegative matrix decomposition technique with respect to each piece of data. However, no comparisons were reported on the final results with respect to the different initializations. Some comparisons of the impact of three specific initialization methods on the solutions obtained by the multiplicative NMF update algorithm applied to problem (2) appeared in [27] and refer to two benchmark cancer datasets. This preliminary study suggests the existence of latent biological information that can be extracted by NMF algorithms regardless of the initialization mechanisms used, but these observations need further investigation.

As the preceding discussion suggests, there are a number of research questions and shortcomings in the existing literature that need to be addressed. In particular, to our knowledge, the question “Do initialization mechanisms affect the final results of NMF for omic data analysis?” still remains open. To pave the way for in-depth research in this context, we present a complete taxonomy of initialization schemes for NMF algorithms, highlighting their mathematical aspects, advantages and possible weaknesses.

3. NMF Iterative Algorithms and a Complete Taxonomy of Initialization Mechanisms

In this section, we present iterative algorithms for NMF of alternating least squares type. These include the majorize–minimize (MM) algorithms, coordinate descent and gradient descent methods, expectation–maximization algorithms, and cone projection approaches [46,47]. We focus on the general constrained nonlinear optimization NMF problem, which is a more comprehensive version of (2) defined as follows:

$$\min_{W, H \geq 0} D_\beta(X; WH) \tag{4}$$

where the cost function is the general β -divergence,

$$D_\beta(X; WH) = \sum_{i=1}^n \sum_{j=1}^m d_\beta(X_{ij}; (WH)_{ij});$$

with d_β is commonly defined for each x, y as

$$d_\beta(x; y) = \begin{cases} \frac{1}{\beta(\beta+1)}(x^\beta + (\beta - 1)y^\beta - \beta xy^{\beta-1}) & \beta \in \mathbb{R} - \{0, 1\}; \\ x \log(\frac{x}{y}) - x + y & \beta = 1; \\ \frac{x}{y} - \log \frac{x}{y} - 1 & \beta = 0. \end{cases}$$

Different values of the parameter β allow to take into consideration the Euclidean distance, the Kullback–Leibler divergence (3) and the Itakura–Saito divergence as special cases ($\beta = 2, 1, 0$, respectively).

Alternating least squares type algorithms sequentially update H given W and then W given H , as

- Set $k = 0$ and W equals to any nonnegative W_0 ;
- With fixed W_k , update $H_{k+1} = \arg \min_{H \geq 0} D_\beta(X; W_k H_k)$;
- With fixed H_{k+1} , update $W_{k+1} = \arg \min_{W \geq 0} D_\beta(X; W_k H_{k+1})$;
- Iterate until a stopping criteria is satisfied.

These two steps are essentially identical because of the symmetry of the factorization. Indeed, from $X \approx WH$ we have $X^T \approx H^T W^T$, where we reverse the roles of W and H and make no assumptions on the data dimensions n and m . Writing the derivative of the cost function $\nabla D_\beta(X; WH)$ with respect to H (resp. W) as the difference of two nonnegative functions (see [48] for details), the update rules of alternating least squares algorithms can be written as functions of the ratio of the nonnegative terms.

For example, the updating rules for Frobenius and KL-divergence (3) for factors H and W can be expressed in the following matrix forms:

$$\begin{aligned} H &\leftarrow H * \frac{(W^T X)}{(W^T WH)}, \\ W &\leftarrow W * \frac{(XH^T)}{(WHH^T)} \end{aligned} \tag{5}$$

and

$$W \leftarrow W \cdot * \frac{(X./(WH))H^T}{\mathbf{1}_n * \left(\sum_{j=1}^m H_{\cdot j}\right)^T} \tag{6}$$

$$H \leftarrow H \cdot * \frac{W^T(X./(WH))}{\left(\sum_{i=1}^n W_{i\cdot}\right)^T * \mathbf{1}_m}$$

where $\cdot *$ denotes the Hadamard product, the ratio is referred to element-wise division and $\mathbf{1}_m$ is a m dimensional vector of ones.

As mentioned earlier, algorithms for computing NMF must be initialized with an initial left factor W (or both factors W and H) to begin with.

We propose to classify the initialization mechanisms for iterative NMF algorithms (which have appeared in the literature panorama so far) into three classes, indicating the main idea on which the initialization is based. In particular, we can identify: (i) random-based schemes, (ii) structured initializations, (iii) evolutionary and nature-based mechanisms. Figure 2 illustrates the three main classes and their subclasses.

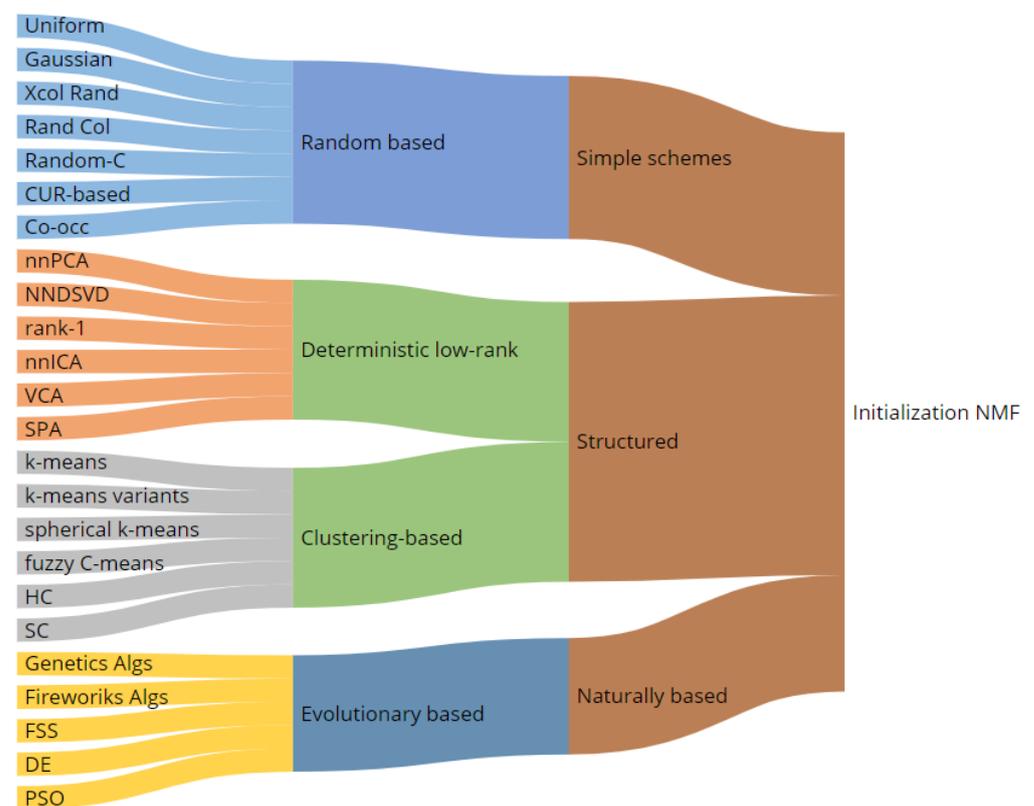


Figure 2. Proposed taxonomy of initialization schemes for NMF appeared up to date in literature panorama. Abbreviations can be found in the relative section.

3.1. Random Based Initializations

The simplest way to choose the initial factors W and H is to construct them as matrices (of dimensions $n \times r$ and $r \times m$, respectively) with random numbers. Random initializations were used early in [30], while a discussion of their goodness in terms of algorithm performance and other variants of randomization methods can be found in [49]. Random-based initializations generally construct dense factors, require low computational cost and processing time for their computations, and are the benchmark mechanisms used in the majority of NMF studies, especially when analyzing omic data. However, the quality and reproducibility of the NMF result are rarely questioned. To ensure robust and reproducible

results, users should run NMF with random initialization a sufficient number of times and finally select the best result based on some quantitative and qualitative criteria [43]. Different mechanisms of random matrix generation are:

- *Uniform random*: the elements in the matrix W (and in H) are chosen as uniformly distributed numbers in the interval $[0, 1]$ or in the same range as the entries of the target matrix.
- *Gaussian random*: the elements in the matrix W (and in H) are chosen as $\max\{g, 0\}$, where the number $g \in \mathbb{R}$ is a Gaussian values.
- *Random Xcol*: the factor matrix W is obtained by averaging over r randomly selected columns of the data matrix X . This scheme is very inexpensive and has the advantage of yielding a sparser factor when the data matrix is already sparse. When dealing with omic data, selected columns of X can be considered as a kind of a priori information that can influence the following update process.
- *Random-C initialization*: this mechanism is based on a double random selection of columns in the data matrix X as follows:
 - Identify p of the longest (in the 2-norm sense) columns of X ;
 - Randomly choose q columns from the previous p longest;
 - Construct each column of W as the average of these q columns.

This scheme is inspired by the C-matrix of the CUR decomposition and produces the densest W , in which column vectors are closed with the centroids of the data matrix (see spherical-k means initialization) [49].

- *CUR-based initialization*: the factor matrix W is constructed as a submatrix (with a small number of actual columns) of the data matrix X . In this way, values are obtained that are more interpretable from a biological point of view (and usually to the same extent as the original data) [50,51]. CUR-based mechanisms differ in the “statistical” way of selecting columns (or rows, if we refer to the initialization of the factor H) from the matrix X . This selection is based on computing an “importance score” for each column (row) of X and randomly selecting r of these columns (rows) using this score as the probability distribution for importance sampling. The scoring values of a column $X_{:,j}$ in X can be calculated as: normalized statistical leverage score [50], spectral angular distance, or symmetrized KL-divergence [52].
- *Co-occurrence initialization*: firstly the data co-occurrence matrix (i.e., XX^T) is formed and then r of its columns are selected by the algorithm proposed in [53] to form the factor W . The data co-occurrence matrix contains information about hidden relationships between columns and rows in the data matrix X , but the computational costs required for this make this initialization mechanism expensive and often impractical for use in the context of biomedical data analysis where large datasets are considered.

3.2. Structured Initialization

Schemes belonging to the structured initialization class are based on the idea of applying low-rank factorizations, well-known clustering mechanisms on the data matrix X to generate initial factors for NMF, or very specific data-driven schemes such as semantic harmonic initialization when audio data are involved ([54–56]). These specifically structured initializations for NMF were mainly designed to improve their performance either in terms of computational complexity or for preserving the particular data structure.

Initialization schemes based on low-rank decomposition algorithms do not require a randomization step, so they can also be classified in the subclass of “deterministic” structured initialization methods. They include schemes using the Singular Value Decomposition and the Nonnegative Double Singular Value Decomposition (NNDSVD) [43,57,58] and its variants [59], rank-1 decomposition [39,60] [61], nonnegative PCA [62,63], nonnegative ICA [64], Vertex Component Analysis [65–67], Successive Projection Algorithm [68].

On the other hand, schemes based on clustering algorithms reduce to the minimum use of random numbers (only some hyperparameters need to be randomly set up before the clustering data matrix [69]) avoiding the application of multi-start random initial-

izations [70]. The subclass of “clustering based” initializations includes: k-means [71] and its variant [72], spherical k-means methods [73,74], fuzzy C-Means clustering [40,75], Hierarchical Clustering [76] and Subtracting Clustering [28]. Table 1 summarises these mechanisms, their related references and if they have or have not been used when microarrays and generally omic data are involved.

Table 1. Structured Initialization schemes for NMF algorithms with their main references and an indication if they have been applied in the context of microarrays or omic data analysis.

	Method	Main References	Omic Data
Deterministic low-rank	NNDSVD	[43,57,58]	yes
	NNDSVD variant	[59]	no
	nonnegative PCA	[62,63,71,75]	no
	rank-1	[39]	no
	nonnegative ICA	[41,64,77]	yes
	Vertex Component Anal.	[65–67]	no
	Successive Projection Alg.	[68]	no
Clustering-based	k-means	[72]	no
	k-means variant	[57,71]	yes
	spherical k-means	[73,74]	no
	fuzzy C-Means	[40,75,78,79]	no
	Hierarchical Clustering	[76]	no
	Subtracting Clustering	[28]	no

In the following, we take a closer look at the structured initializations for NMF, some of which are used for microarrays or omic data in general.

3.2.1. Non-Negative Double Singular Value Decomposition

NNDSVD described in [43] is based on two processes of SVD, and it constructs both factors W and H in NMF as the positive parts of rank-1 matrices obtained by the left and right singular vectors of the SVD decomposition of X .

Let $X = \sum_{i=1}^k \sigma_i u_i v_i^T$ be the SVD representation of the data matrix X , with rank k and σ_i, u_i, v_i as the nonzero singular values and the left and right singular vectors of X , respectively. Then, for every $r < k$, the optimal rank- r approximation of X is given by

$$X \approx \sum_{i=1}^r \sigma_i C_i,$$

where $C_i = u_i v_i^T$ are the rank-1 matrix formed by the first $i = 1, \dots, r$ singular vectors. Each matrix C_i can be written as the sum of two nonnegative components $C_i = C_i^+ - C_i^-$, being the positive and the negative section of C_i , respectively. Approximating each C_i by C_i^+ , by its nonnegativeness from Perron–Frobenius theory, its maximum left and right singular vectors will also be nonnegative. The dominant singular triplets can be used as initial vectors and rows to W and H .

The main phases of this algorithm are described in the following [43]:

- Compute the largest r singular triplets of X (first SVD process);
- Initialize the first column and row vectors in W and H as the nonnegative dominant singular vectors of X weighted by $\sqrt{\sigma_i}$;
- Compute the positive section of each C_i ;
- Compute the largest r singular triplets of C_i^+ (second SVD process);
- Initialize the j -th columns and rows in W and H , for $j = 2, \dots, r$ as the singular dominant vectors of each C_i^+ , weighted by the $\sqrt{\sigma_i(C_i^+)}$ and normalized.

NNDSVD is composed of two processes of SVD and this justifies its name. There are some variants of it, namely NNDSVDA and NNDSVDAR, which involve a replacement

of the zero value on the result (W, H) by the average of all elements in the matrix X and by a very small random value, respectively. This initialization scheme is included in the main tools for NMF analysis of biological data [80,81], moreover, its deterministic properties make NNDSVD the most commonly used initialization scheme together with random approaches.

3.2.2. Nonnegative ICA Initialization

Independent component analysis is a mechanism used to extract a set of statistically independent source variables from a collection of mixed signals without having information about the data source signals or the combination process. It has been used as a knowledge extraction mechanism in various biological and omic data analysis tasks. A modified nonnegative version of ICA (nnICA) was proposed in [64] to feed NMF algorithms, and such initialization has been shown to be effective in speeding up the learning process and obtaining desired solutions.

The main phases of this algorithm are:

- Compute the ICA on the observed data matrix X ;
- Initialized the factor W as the absolute of the independent components source matrix obtained from ICA.

3.2.3. k-Means Initialization

Another structured initialization sometimes used in biological contexts is the (spherical) k-means clustering [74]. In this scheme, the columns of W are initialized with the centroids $\{c_j\}_j$ of the best k (spherical) clusters of data, i.e., the center points of k disjoint subsets $\{\Pi_j\}_{j=1}^k$ of the columns of X that are closest with respect to a distance function (generally the Euclidean norm when spherical-k means clustering is required).

The main phases of this algorithm are described below, adopting the notation introduced in [74].

- Initialize k centroids $\{c^{(0)}\}_{j=1}^k$ randomly choosing some columns in X and set $t = 0$ (this is the initial iteration) ;
- Compute $d_{ij}^{(t)} = \frac{c_j^{(0)} X_{*i}^T}{\|c_j^{(0)}\| \|X_{*i}\|}$, $j = 1, \dots, k$ and $i = 1, \dots, m$;
- Define the new partition of clusters $\Pi_j^{(t+1)} = \{X_{*i} | \operatorname{argmax}_l (d_{il}^{(t)})\}$;
- Recompute each centroid as: $c_{j=1}^{(t+1)} = \frac{\sum_{X_{*i} \in \Pi_j^{(t+1)}} X_{*i}}{\|\sum_{X_{*i} \in \Pi_j^{(t+1)}} X_{*i}\|}$;
- Initialize the columns of the factor W as the final cluster centroid vectors.

This initialization scheme expects the initial matrix H to be chosen randomly or as the matrix in which the elements are the absolute values of the elements in $W^T X$. A variant allows to compute the elements of the factor H as the membership degrees of each data point using the (spherical) k-means clustering [71].

3.3. Evolutionary and Natural Based Initialization

Schemes based on stochastic global search and evolutionary optimization methods have recently appeared in the literature panorama as feeding schemes for NMF algorithms, but they have not been used in the analysis of omic data.

Under this class, we can enumerate a number of different population-based algorithms: Genetic Algorithms, Particle Swarm Optimization, Fish School Search, Differential Evolution and Fireworks Algorithm [82,83], which have been proposed as new initialization variants for NMF multiplicative algorithms. All these population-based methods sequentially initialize individual rows of W or individual columns of H to minimize the NMF objective function before factorization. These methods were compared experimentally, and some of them showed a reduction in the number of NMF iterations required to

achieve a given accuracy. They also allow parallel/distributed computation by splitting the initialization into several partially independent subtasks.

Another Genetic Algorithm approach was proposed in [84] to estimate the factor W by first generating a population of individuals (representing potential solutions for NMF optimization, i.e., estimates of the matrix W), then selecting individuals according to the value of a specific fitness function and reproducing them by applying a specific genetic operator. This process leads to the evolution of individuals within the population that better solve the optimization problem.

Nevertheless, these approaches suffer from difficulties in their applicability due to the large number of hyperparameters that need to be fixed and tuned a priori [69].

4. How Many Initializations Influence NMF Results for Omic Data Analysis?

The preceding literature review illustrates the existence of various initializations that can be used to feed iterative NMF processes and the fact that their impact on the NMF final has been largely considered only in terms of the performances (convergence rates and/or final relative error of the objective function) of the specific algorithm adopted. Some initialization schemes were qualitatively compared to demonstrate the best clustering performance for multiple face and text datasets [39], others improved the separation of face components [40], while initialization based on ICA proved better separation performance in the audio source separation task [41]. In the context of omic data analysis, only some random initializations and NNDSVD are used, but no comparisons have been made on how such a scheme affects the final results from a biological point of view. It should be noted that any initializations can implicitly encode prior knowledge into the NMF that may focus the resulting factors to reflect valid biological information embedded in the data [3]. Because there is a lack of information in the literature about comparing initialization methods and how these can be interpreted concerning the influence of biological results achieved by NMF, we decide to focus on this aspect by testing some techniques on real datasets. We start a preliminary study to inspect how some seeding approaches can affect the results in omic data analysis, hoping to put the basis of new theoretical and experimental studies in this direction.

Here, we briefly illustrate a preliminary study of ours using three different types of initializations with the KL-based MM algorithm to study two benchmark microarray datasets: MCF7 and Golub data derived from breast cancer cells and leukemia microarray studies, respectively. Table 2 contains information about these data.

Table 2. Description of datasets analyzed by the NMF KL-based algorithm fed by a subset of initialization schemes. The MCF7 data matrix consists of 10,331 genes extracted by a cell cycle microarray from breast cancer cells and 434 compounds linked to arachidonic acid. The data matrix was pre-processed, as described in [25]. The Golub data matrix consists of 5000 genes and 38 tumor samples: 27 patients with acute lymphoblastic leukemia and 11 patients with acute myeloid leukemia [9,85].

Dataset	no. Rows (Genes)	no. Columns (Compounds/Patients)	Type	References
MCF7	10,331	434	breast cancer	[25]
Golub	5000	38	leukemia	[9,85]

We performed some numerical experiments (in *R project environment* [81] on an I-7 Core machine with a memory capacity of 12 GB RAM) using the KL-based update rules initialized with: random initialization with randomly generated elements in the interval $[0, 1]$, nnICA and NNDVSD. According to previous studies, the hyperparameter rank value was set to $r = 4$ and a stopping value of 3000 iterations (at most) was considered. Different executions were performed for initialization with randomness (10 random and 10 nnICA initializations have been saved).

We focused our attention on the columns of the basis matrices obtained for each dataset and initialization, relying on the assumption that all the knowledge extracted from the process is hidden into these latent factors. The obtained $W_{dataset} \in \mathbb{R}^{10331 \times 84}$ matrices (concatenating all the basis from experiments for both *MCF7* and *Golub* datasets) were compared in terms of information embedded, exploiting cosine similarity criterion. Particularly, the matrix $cosW_{dataset} \in \mathbb{R}^{84 \times 84}$, defined as:

$$(cosW_{dataset})_{ij} = \frac{W_{dataset}(:,i)^\top W_{dataset}(:,j)}{\|W_{dataset}(:,i)\|_2 \|W_{dataset}(:,j)\|_2}. \quad (7)$$

for $i, j = 1, \dots, 84$ collects the cosine values among columns of the basis matrix $W_{dataset}$, for both *MCF7* and *Golub* datasets.

To demonstrate metagenes strictly related from a geometrical point of view (according to the cosine similarity), we filter values in the cosine similarity matrices, considering only pairs of metagenes with similarity values in the range $[0.8, 1]$. Figures 3 and 4 show a pseudo-binary version of the heatmaps for $cosW_{MCF7}$ and $cosW_{Golub}$, respectively.

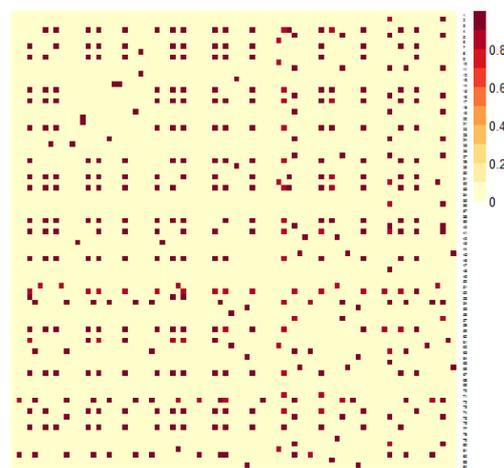


Figure 3. Heatmap of the filtered pseudo-binary cosine similarity matrix derived from $cosW_{MCF7}$: only metagenes with maximum similarity values are reported, while others and diagonal elements are set to 0.

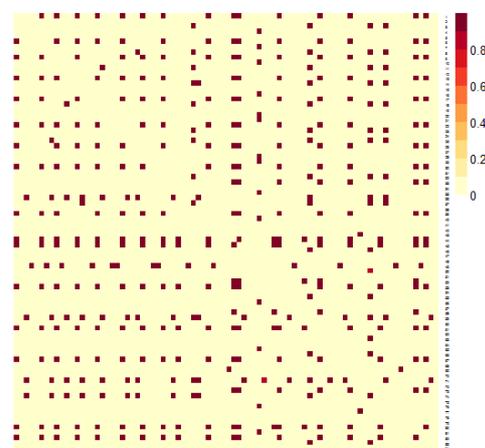


Figure 4. Heatmap of the filtered pseudo-binary cosine similarity matrix derived from $cosW_{Golub}$: only metagenes with maximum similarity values are reported, while others elements and diagonal entries are set to 0.

Despite the large number of relevant metagenes that have appeared, some of them (85.24% and 86.43% from the *MCF7* and *Golub* datasets, respectively) are repeated. Among the latter, those with higher and lower frequency seem to be qualitatively coherent among themselves, and, moreover, they share a common behavior pattern. On the other hand, Table 3 gives the number of metagenes that occur only once in each initialization scheme for the two datasets *MCF7* and *Golub*.

Table 3. Number of metagenes with a high similarity degree for the *MCF7* and *Golub* datasets.

Dataset	Random	Initialization	
		nnICA	NDVSD
MCF7	18	15	10
	24	25	10
	2	4	0
Golub	19	30	10
	10	25	10
	1	3	0

Similarities and common behavior were investigated by also performing PCA on the $W_{dataset}$ matrices for both datasets. Table 4 gives the percentage of variance explained by four principal components. As can be observed for both datasets, the information contained in the data seems to be mainly collected in the first components. High geometric relationships for all extracted metagenes and a high percentage of variance explained in the first components of PCA strongly suggest that metagene groups seem to be quantitatively related. In addition, the similar behavior in the cosine-filtered heatmap and relatively repeated metagenes suggests their qualitative similarity and that metagenes extracted with different initializations share common latent knowledge.

Table 4. Percentage of variance explained for the two microarray datasets.

Percentage of Variance	PC1	PC2	PC3	PC4
<i>MCF7</i>	79.8372	10.5284	4.4372	2.4361
<i>Golub</i>	63.2523	14.9335	12.4118	9.4001

The experimental results in this section are focused on analyzing the information embedded into the column of the basis matrices achieved by several runs of different seeding methodologies. Quantitative and qualitative comparisons have been made on these results to provide evidence and form a geometrical, statistical and visual point of view (with cosine similarities, PCA and explained variances, counting frequencies and heatmap representations) that some shared knowledge is present between the different runs. Furthermore, as explained at the beginning of this section, this is a preliminary study, which needs to be further investigated and extended. We hope that this will provide the basis for future direction and collaboration between the mathematical and biological world.

5. Conclusive Remarks

Various matrix factorizations prove to be effective tools for the analysis of omic data. Among them, Nonnegative Matrix Factorization is able to reveal interpretable latent factors and identify genes belonging to multiple pathways or biological processes. However, the algorithms for calculating NMF need to be initialized. The current literature has not yet considered the question of the possible influence of the particular initialization methods on the final results of NMF for omic data analysis. In order to pave the way for a deeper investigation of this aspect, in this paper, we have completely reviewed the NMF initialization schemes appearing in the literature, pointed out their main characteristics and when they have been used for omic data. Moreover, we briefly illustrated some preliminary

results obtained when three selected initialization schemes were used to feed the KL-based algorithm when two benchmark cancer datasets were considered. The experimental results obtained in this biological context seem to indicate that interpretable information strictly related to the data matrix under analysis can be extracted regardless of the particular initialization scheme used in the iterative NMF algorithm. The results presented in this work are from a preliminary study and part of a future project that aims to provide the basis for a deep collaboration between the theoretical-numerical mathematical aspect of these techniques and the biomedical world. Even if, from a mathematical point of view, the alternate algorithmic nature of NMF achieves local minimum, this is quite often sufficient in data analysis applications to extract useful knowledge from real datasets. For these reasons, future analysis in this direction should be devoted to different aspects: to the comparison of particular objective functions adopted in the minimization process and to construct a biological dataset with some a priori knowledge to better interpret the results. It is of the authors opinion that this could help researchers connect how the particular minimization process adopted is hiddenly related to the extracted biological information embedded in the data. Further analysis is needed, and a massive controlled experimental session with multiple datasets and some a priori biological knowledge of them are required to finally assess the actual influence of initialization produced on the final information extracted by NMF from omic data.

Funding: The author was funded by the REFIN Project, grant number 363BB1F4, Reference project idea UNIBA027 “Un modello numerico-matematico basato su metodologie di algebra lineare e multilineare per l’analisi di dati genomici”.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data could be available from authors.

Acknowledgments: This work has been supported in part by the GNCS (*Gruppo Nazionale per il Calcolo Scientifico*) of Istituto Nazionale di Alta Matematica Francesco Severi, P.le Aldo Moro, Roma, Italy.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

NMF	Nonnegative matrix factorizations
SVD	Singular Value Decomposition
PCA	Principal Component Analysis
KL	Kullback-Leiber
MU	Multiplicative Update
MM	Majorise-Minimize
nnPCA	Nonnegative PCA
ICA	Independent Component Analysis
nnICA	nonnegative ICA
NNDSVD	Nonnegative Double SVD
VCA	Vertex Complex Analysis
SPA	Successive Projection Algorithm
HC	Hierarchical Clustering
SC	Subtracting Clustering
FSS	Fisch School Search
DE	Differential Evolution
Fireworks Algs	Fireworks Algorithms
PSO	Particle Swarm Optimization

References

1. Yamada, R.; Okada, D.; Wang, J.; Basak, T.; Koyama, S. Interpretation of omics data analyses. *J. Hum. Genet.* **2021**, *66*, 93–102. [[CrossRef](#)]
2. Nicora, G.; Vitali, F.; Dagliati, A.; Geifman, N.; Bellazzi, R. Integrated Multi-Omics Analyses in Oncology: A Review of Machine Learning Methods and Tools. *Front. Oncol.* **2020**, *10*, 1030. [[CrossRef](#)] [[PubMed](#)]
3. Stein-O'Brien, G.L.; Arora, R.; Culhane, A.C.; Favorov, A.V.; Garmire, L.X.; Greene, C.S.; Goff, L.A.; Li, Y.; Ngom, A.; Ochs, M.F.; et al. Enter the Matrix: Factorization Uncovers Knowledge from Omics. *Trends Genet.* **2018**, *34*, 790–805. [[CrossRef](#)] [[PubMed](#)]
4. Kossenkov, A.V.; Ochs, M.F. Matrix factorisation methods applied in microarray data analysis. *Int. J. Data Min. Bioinform.* **2010**, *4*, 72–90. [[CrossRef](#)]
5. Devarajan, K. Nonnegative matrix factorization: An analytical and interpretive tool in computational biology. *PLoS Comput. Biol.* **2008**, *4*, e1000029. [[CrossRef](#)] [[PubMed](#)]
6. Moloshok, T.; Klevecz, R.; Grant, J.; Manion, F.; Speier, W.; Ochs, M. Application of Bayesian Decomposition for analysing microarray data. *Bioinformatics* **2002**, *18*, 566–575. [[CrossRef](#)]
7. Saidi, S.A.; Holland, C.M.; Kreil, D.P.; MacKay, D.J.; Charnock-Jones, D.S.; Print, C.G.; Smith, S.K. Independent component analysis of microarray data in the study of endometrial cancer. *Oncogene* **2004**, *23*, 6677. [[CrossRef](#)]
8. Alter, O.; Brown, P.O.; Botstein, D. Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10101–10106. [[CrossRef](#)]
9. Brunet, J.P.; Tamayo, P.; Golub, T.R.; Mesirov, J.P. Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 4164–4169. [[CrossRef](#)]
10. Dai, J.J.; Lieu, L.; Rocke, D. Dimension reduction for classification with gene expression microarray data. *Stat. Appl. Genet. Mol. Biol.* **2006**, *5*, 6. [[CrossRef](#)]
11. Devarajan, K.; Ebrahimi, N. Class Discovery via Nonnegative Matrix Factorization. *Am. J. Math. Manag. Sci.* **2008**, *28*, 457–467. [[CrossRef](#)]
12. Kong, W.; Mou, X.; Hu, X. *Exploring Matrix Factorization Techniques for Significant Genes Identification of Alzheimer's Disease Microarray Gene Expression Data*; BMC bioinformatics; BioMed Central: London, UK, 2011; Volume 12, p. S7.
13. Ochs, M.F.; Fertig, E.J. Matrix Factorization for Transcriptional Regulatory Network Inference. In Proceedings of the IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, San Diego, CA, USA, 9–12 May 2012; pp. 387–396. [[CrossRef](#)]
14. Meng, C.; Zeleznik, O.A.; Thallinger, G.G.; Kuster, B.; Gholami, A.M.; Culhane, A.C. Dimension reduction techniques for the integrative analysis of multi-omics data. *Briefings Bioinform.* **2016**, *17*, 628–641. [[CrossRef](#)] [[PubMed](#)]
15. Liu, J.X.; Wang, D.; Gao, Y.L.; Zheng, C.H.; Xu, Y.; Yu, J. Regularized non-negative matrix factorization for identifying differential genes and clustering samples: A survey. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2017**, *15*, 974–987. [[CrossRef](#)] [[PubMed](#)]
16. Li, Y.; Wu, F.-X.; Ngom, A. A review on machine learning principles for multi-view biological data integration. *Brief Bioinform* **2018**, *19*, 325–340. doi:10.1093/bib/bbw113.
17. Wall, M.E.; Rechtsteiner, A.; Rocha, L.M. Singular Value Decomposition and Principal Component Analysis. In *A Practical Approach to Microarray Data Analysis*; Berrar, D.P., Dubitzky, W., Granzow, M., Eds.; Springer: Boston, MA, USA, 2003; pp. 91–109. [[CrossRef](#)]
18. Sompairac, N.; Nazarov, P.V.; Czerwinska, U.; Cantini, L.; Biton, A.; Molkenov, A.; Zhumadilov, Z.; Barillot, E.; Radvanyi, F.; Gorban, A.; et al. Independent Component Analysis for Unraveling the Complexity of Cancer Omics. *Int. J. Mol. Sci.* **2019**, *20*, 4414. [[CrossRef](#)] [[PubMed](#)]
19. Yang, Z.; Michailidis, G. A Non-negative Matrix Factorization Method for Detecting Modules in Heterogeneous Omics Multi-modal Data. *Bioinformatics* **2015**, *32*. [[CrossRef](#)]
20. Boccarelli, A.; Esposito, F.; Coluccia, M.; Frassanito, M.A.; Vacca, A.; Del Buono, N. Improving knowledge on the activation of bone marrow fibroblasts in MGUS and MM disease through the automatic extraction of genes via a Nonnegative Matrix Factorization approach on gene expression profiles. *J. Transl. Med.* **2018**, *16*, 217. [[CrossRef](#)]
21. Rappoport, N.; Shamir, R. Multi-omic and multi-view clustering algorithms: Review and cancer benchmark. *Nucleic Acids Res.* **2018**, *46*, 10546–10562. [[CrossRef](#)]
22. Esposito, F.; Gillis, N.; Del Buono, N. Orthogonal joint sparse NMF for microarray data analysis. *J. Math. Biol.* **2019**, *79*, 223–247. [[CrossRef](#)]
23. Esposito, F.; Del Buono, N.; Selicato, L. Nonnegative Matrix Factorization models for knowledge extraction from biomedical and other real world data. *PAMM* **2021**, *20*, e202000032. [[CrossRef](#)]
24. Lee, D.D.; Seung, H.S. Algorithms for Non-negative Matrix Factorization. In *Proceedings of the Advances in Neural Information Processing Systems Conference*; MIT Press: Cambridge, MA, USA, 2000; Volume 13, pp. 556–562.
25. Del Buono, N.; Esposito, F.; Fumarola, F.; Boccarelli, A.; Coluccia, M. Breast Cancer's Microarray Data: Pattern Discovery Using Nonnegative Matrix Factorizations. In *International Workshop on Machine Learning, Optimization and Big Data*; Springer: Berlin, Germany, 2016; pp. 281–292.
26. Gillis, N. *Nonnegative Matrix Factorization*; SIAM: Philadelphia, PA, USA, 2020.

27. Del Buono, N.; Esposito, F. Investigating initialization techniques for Nonnegative Matrix Factorization: A survey and a case of study of microarray. In *Molecular and Mathematical Biology, Chemistry, Medicine and Medical Statistics, Bioinformatics and Numerical Analysis (Series in Applied Sciences)*; Carletti, M., Spaletta, G., Eds.; Universitas Studiorum: Mantova, Italy, 2019; Volume 2. [[CrossRef](#)]
28. Casalino, G.; Del Buono, N.; Mencar, C. Subtractive clustering for seeding non-negative matrix factorizations. *Inf. Sci.* **2014**, *257*, 369–387. [[CrossRef](#)]
29. Vavasis, S. On the Complexity of Nonnegative Matrix Factorization. *SIAM J. Optim.* **2010**, *20*, 1364–1377. [[CrossRef](#)]
30. Lee, D.D.; Seung, H.S. Learning the parts of objects by non-negative matrix factorization. *Nature* **1999**, *401*, 788–791. [[CrossRef](#)]
31. Yang, Z.; Oja, E. Unified Development of Multiplicative Algorithms for Linear and Quadratic Nonnegative Matrix Factorization. *IEEE Trans. Neural Netw.* **2011**, *22*, 1878–1891. [[CrossRef](#)] [[PubMed](#)]
32. Zhao, R.; Tan, V.Y.F. A Unified Convergence Analysis of the Multiplicative Update Algorithm for Regularized Nonnegative Matrix Factorization. *IEEE Trans. Signal Process.* **2018**, *66*, 129–138. [[CrossRef](#)]
33. Kim, H.; Park, H. Nonnegative Matrix Factorization Based on Alternating Nonnegativity Constrained Least Squares and Active Set Method. *SIAM J. Matrix Anal. Appl.* **2008**, *30*, 713–730. [[CrossRef](#)]
34. Chen, D.; Plemmons, R. Nonnegativity constraints in numerical analysis. In *Symposium on the Birth of Numerical Analysis*; Bultheel, A., Cools, R., Eds.; World Scientific Press: Singapore, 2009.
35. Gillis, N.; Glineur, F. A multilevel approach for nonnegative matrix factorization. *J. Comput. Appl. Math.* **2012**, *236*, 1708–1723. [[CrossRef](#)]
36. Lin, C. Projected Gradient Methods for Nonnegative Matrix Factorization. *Neural Comput.* **2007**, *19*, 2756–2779. [[CrossRef](#)]
37. Donoho, D.; Stodden, V. When Does Non-negative Matrix Factorization Give a Correct Decomposition into Parts? In *NIPS'03 Proceedings of the 16th International Conference on Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2003; pp. 1141–1148.
38. Fogel, P.; Hawkins, D.M.; Beecher, C.; Luta, G.; Young, S.S. A Tale of Two Matrix Factorizations. *Am. Stat.* **2013**, *67*, 207–218. [[CrossRef](#)]
39. Zhaoqiang, L.; Tan, V.Y.F. Rank-One NMF-Based Initialization for NMF and Relative Error Bounds Under a Geometric Assumption. *IEEE Trans. Signal Process.* **2017**, *65*, 4717–4731.
40. Rezaei, M.; Boostani, R.; Rezaei, M. An Efficient Initialization Method for Nonnegative Matrix Factorization. *J. Appl. Sci.* **2011**, *11*, 354–359. [[CrossRef](#)]
41. Kitamura, D.; Ono, N. Efficient initialization for nonnegative matrix factorization based on nonnegative independent component analysis. In *Proceedings of the 2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*, Xi'an, China, 13–16 September 2016; pp. 1–5. [[CrossRef](#)]
42. Chalise, P.; Fridley, L. Integrative clustering of multi-level omic data based on non-negative matrix factorization algorithm. *PLoS ONE* **2017**, *12*. [[CrossRef](#)] [[PubMed](#)]
43. Boutsidis, C.; Gallopoulos, E. SVD based initialization: A head start for nonnegative matrix factorization. *Pattern Recognit.* **2008**, *41*, 1350–1362. [[CrossRef](#)]
44. Gao, C.; Welch, J.D. Iterative Refinement of Cellular Identity from Single-Cell Data Using Online Learning. In *Research in Computational Molecular Biology*; Schwartz, R., Ed.; Springer International Publishing: Cham, Switzerland, 2020; pp. 248–250.
45. Chalise, P.; Ni, Y.; Fridley, B.L. Network-based integrative clustering of multiple types of genomic data using non-negative matrix factorization. *Comput. Biol. Med.* **2020**, *118*, 103625. [[CrossRef](#)]
46. Hobolth, A.; Guo, Q.; Kousholt, A.; Jensen, J.L. A Unifying Framework and Comparison of Algorithms for Non-negative Matrix Factorisation. *Int. Stat. Rev.* **2020**, *88*, 29–53. [[CrossRef](#)]
47. Kim, J.; He, Y.; Park, H. Algorithms for nonnegative matrix and tensor factorizations: A unified view based on block coordinate descent framework. *J. Glob. Optimization* **2014**, *58*, 285–319. [[CrossRef](#)]
48. Fevotte, C.; Idier, J. Algorithms for Nonnegative Matrix Factorization with the β -Divergence. *Neural Comput.* **2011**, *23*, 2421–2456. [[CrossRef](#)]
49. Langville, A.; Meyer, C.D.; Albright, R. Initializations for the nonnegative matrix factorization. In *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Philadelphia, PA, USA, 20–23 August 2006.
50. Mahoney, M.; Drineas, P. CUR matrix decompositions for improved data analysis. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 697–702. [[CrossRef](#)] [[PubMed](#)]
51. Piwowar, M.; Kocemba-Pilarczyk, K.; Piwowar, P. Regularization and grouping-omics data by GCA method: A transcriptomic case. *PLoS ONE* **2018**, *13*. [[CrossRef](#)]
52. Li, S.; Gengxin, Z.; Xinpeng, D. CUR Based Initialization Strategy for Non-Negative Matrix Factorization in Application to Hyperspectral Unmixing. *J. Appl. Math. Phys.* **2016**, *4*, 614.
53. Sandler, M. On the Use of Linear Programming for Unsupervised Text Classification. In *KDD '05 Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*; ACM: New York, NY, USA, 2005; pp. 256–264. [[CrossRef](#)]
54. Ewert, S.; Muller, M. Using score-informed constraints for NMF-based source separation. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, 25–30 March 2012.

55. Fritsch, J.; Plumbley, M.D. Score informed audio source separation using constrained nonnegative matrix factorization and score synthesis. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 888–891. [[CrossRef](#)]
56. Rohlfing, C.; Becker, J.M. Extended semantic initialization for NMF-based audio source separation. In Proceedings of the 2015 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Nusa Dua, Bali, Indonesia, 9–12 November 2015; pp. 95–100. [[CrossRef](#)]
57. Zdunek, R. Initialization of Nonnegative Matrix Factorization with Vertices of Convex Polytope. In *Artificial Intelligence and Soft Computing*; ICAISC 2012; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7267.
58. Julian Mathias, B.; Matthias, M.; Christian, R. Complex SVD Initialization for NMF Source Separation on Audio Spectrograms. Proc. Deutsche Jahrestagung fur Akustik (DAGA), Nuremberg, Germany, 16–19 March 2015.
59. Atif, S.; Qazi, S.; Gillis, N. Improved SVD-based initialization for nonnegative matrix factorization using low-rank correction. *Pattern Recognit. Lett.* **2019**, *122*. [[CrossRef](#)]
60. Biggs, M.; Ghodsi, A.; Vavasis, S. Nonnegative Matrix Factorization via Rank-One Downdate. In Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland, 5–9 July 2008. [[CrossRef](#)]
61. Xuansheng Wang, X.X.; Lu, L. An Effective Initialization for Orthogonal Nonnegative Matrix Factorization. *J. Comput. Math.* **2012**, *30*, 34–46. [[CrossRef](#)]
62. Zhao, L.; Zhuang, G.; Xu, X. Facial expression recognition based on PCA and NMF. In Proceedings of the 2008 7th World Congress on Intelligent Control and Automation, Chongqing, China, 25–27 June 2008; pp. 6826–6829.
63. Xiu-rui, G.; Lu-yan, J.; Kang, S. Non-negative matrix factorization based unmixing for principal component transformed hyperspectral data. *Front. Inf. Technol. Electron. Eng.* **2016**, *17*, 403–412. [[CrossRef](#)]
64. Oja, E.; Plumbley, M. Blind Separation of Positive Sources by Globally Convergent Gradient Search. *Neural Comput.* **2004**, *16*, 1811–1825. [[CrossRef](#)] [[PubMed](#)]
65. Nascimento, J.M.P.; Dias, J.M.B. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote. Sens.* **2005**, *43*, 898–910. [[CrossRef](#)]
66. Tang, W.; Zhenwei, S.; Zhenyu, A. Nonnegative matrix factorization for hyperspectral unmixing using prior knowledge of spectral signatures. *Opt. Eng.* **2012**, *51*, 1–10. [[CrossRef](#)]
67. Cao, J.; Lilian, Z.; Haiyan, T. An Endmember Initialization Scheme for Nonnegative Matrix Factorization and Its Application in Hyperspectral Unmixing. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 195. [[CrossRef](#)]
68. Sauwen, N.; Acou, M.; Halandur, N.; Bharath, D.M.; Sima, J.V.; Maes, F.; Himmelreich, U.; Achten, E.; Van Huffel, S. The successive projection algorithm as an initialization method for brain tumor segmentation using non-negative matrix factorization. *PLoS ONE* **2017**, *12*. [[CrossRef](#)] [[PubMed](#)]
69. Selicato, L.; Del Buono, N.; Esposito, F. Methods for Hyperparameters Optimization in Learning Approaches: An overview. In *Machine Learning, Optimization, and Data Science. LOD 2020. Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2020; Volume 12514. [[CrossRef](#)]
70. Cichocki, A.; Zdunek, R.; Phan, A.H.; Amari, S. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation*; Wiley: New York, NY, USA, 2009.
71. Gong, L.; Nandi, A.K. An enhanced initialization method for non-negative matrix factorization. In Proceedings of the 2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP), Southampton, UK, 22–25 September 2013; pp. 1–6. [[CrossRef](#)]
72. Xue, Y.; Sze Tong, C.; Chen, Y.; Chen, W.S. Clustering-based initialization for non-negative matrix factorization. *Appl. Math. Comput.* **2008**, *205*, 525–536. Special Issue on Advanced Intelligent Computing Theory and Methodology in Applied Mathematics and Computation. [[CrossRef](#)]
73. Wild, S. Seeding Non-Negative Matrix Factorizations with the Spherical K-Means Clustering. Ph.D. Thesis, University of Colorado, Denver, CO, USA, 2003.
74. Wild, S.; Curry, J.; Dougherty, A. Improving non-negative matrix factorizations through structured initialization. *Pattern Recognit.* **2004**, *37*, 2217–2232. [[CrossRef](#)]
75. Zheng, Z.; Yang, J.; Zhu, Y. Initialization enhancer for non-negative matrix factorization. *Eng. Appl. Artif. Intell.* **2007**, *20*, 101–110. [[CrossRef](#)]
76. Kim, Y.D.; Choi, S. A Method of Initialization for Nonnegative Matrix Factorization. In Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP '07, Honolulu, HI, USA, 15–20 April 2007; Volume 2, pp. 537–540.
77. Djaouad, B.; Shahram, H.; Yannick, D.; Moussa, K.; Abdelkader, H. Modified Independent Component Analysis for Initializing Non-negative Matrix Factorization: An approach to Hyperspectral Image Unmixing. In Proceedings of the International Workshop on Electronics, Control, Modelling, Measurement and Signals (ECMS 2013), Toulouse, France, 24–26 June 2013; pp. 1–6.
78. Alshabrawy, O.S.; Ghoneim, M.E.; Awad, W.A.; Hassanien, A.E. Underdetermined blind source separation based on Fuzzy C-Means and Semi-Nonnegative Matrix Factorization. In Proceedings of the 2012 Federated Conference on Computer Science and Information Systems (FedCSIS), Wroclaw, Poland, 9–12 September 2012; pp. 695–700.
79. Suleman, A. On ill-conceived initialization in archetypal analysis. *Adv. Data Anal. Classif.* **2017**, *11*, 785–808. [[CrossRef](#)]

80. Mejia Roa, E.; Carmona-Saez, P.; Nogales-Cadenas, R.; Vicente, C.; Vazquez, M.; Yang, X.; Garcia, C.; Tirado, F.; Pascual-Montano, A. BioNMF: A web-based tool for nonnegative matrix factorization in biology. *Nucleic Acids Res.* **2008**, *36*, W523–W528. [[CrossRef](#)] [[PubMed](#)]
81. Gaujoux, R.; Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinform.* **2010**, *11*, 367. [[CrossRef](#)]
82. Janecek, A.; Tan, Y. Iterative improvement of the Multiplicative Update NMF algorithm using nature-inspired optimization. In Proceedings of the 2011 Seventh International Conference on Natural Computation, Shanghai, China, 26–28 July 2011; Volume 3, pp. 1668–1672. [[CrossRef](#)]
83. Janecek, A.; Tan, Y. Using Population Based Algorithms for Initializing Nonnegative Matrix Factorization. In *Advances in Swarm Intelligence*; Tan, Y. Shi, Y. Chai, Y., Wang, G., Eds.; ICSI 2011; Lecture Notes in Computer Science: Berlin/Heidelberg, Germany, 2011; Volume 6729.
84. Stadlthanner, K.; Lutter, D.; Theis, F.J.; Lang, E.W.; Tome, A.M.; Georgieva, P.; Puntonet, C.G. Sparse Nonnegative Matrix Factorization with Genetic Algorithms for Microarray Analysis. In Proceedings of the 2007 International Joint Conference on Neural Networks, Orlando, FL, USA, 12–17 August 2007; pp. 294–299. [[CrossRef](#)]
85. Golub, T.R.; Slonim, D.K.; Tamayo, P.; Huard, C.; Gaasenbeek, M.; Mesirov, J.P.; Coller, H.; Loh, M.L.; Downing, J.R.; Caligiuri, M.A.; et al. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* **1999**, *286*, 351–357. [[CrossRef](#)] [[PubMed](#)]