Non-negative Matrix Factorization using posrank-based approximation decompositions

Ana de Almeida
Department of Information Science and Technologies ISCTE-IUL
Av. das Forças Armadas, 1649-026 Lisboa, Portugal
Email: Ana.Almeida@iscte.pt

***Abstract*—The present work addresses a particular issue related to the nonnegative factorisation of a matrix (NMF). When NMF is formulated as a nonlinear programming optimisation problem some algebraic properties concerning the dimensionality of the factorisation arise as especially important for the numerical resolution. Its importance comes in the form of a guarantee to obtain good quality approximations to the solutions of signal processing image problems. The focus of this work lies in the importance of the rank of the factor matrices, especially in the so-called *posrank of the factorisation*. We report computational tests that favor the conclusion that the value of the posrank has an important impact on the quality of the test images recovered from the decomposition.**

I. Introduction

Factorisation of matrices is a fundamental step in various models and systems. The decomposition of nonnegative matrices able to preserve the nonnegative constraints, usually known as *Nonnegative Matrix Factorisation* (NMF), has been used in virtually all application areas of data mining and machine learning such as biomedical applications, face and object recognition, computer vision, activity recognition, social network analysis, and information extraction, among many others. In fact, NMF concerns the analysis of data matrices whose elements are nonnegative, a common occurrence in data sets derived from text and images. Usually, this type of decomposition becomes necessary mainly for processing very large datasets. The possibility of achieving reduction of dimensionality contributes to the preprocessing of high-dimensional spaces, namely by giving a low-dimensional representation of high-dimensional data for effective and efficient data analysis. Dimensionality reduction is also believed to be essential to human perception helping to reveal low dimensional structures of patterns observed in high dimensional spaces. Consider an image as a point in high-dimensional space where each image pixel takes a value in $\{0, 1, \ldots, 255\}$. The number of possible combinations is $256^{m\times n}$, where n and m are the matrix dimensions. Although it is possible to describe a wide variety of visual object classes or patterns, for a specific pattern (such as the human face) the number of feasible combinations is but a small fraction of that huge number [11].

Some models for physical phenomena are only meaningful when the data to be analysed is nonnegative in nature. Consequently, in order to maintain an accurate interpretation of the captured phenomena, we need to apply processing tools over these representations that are able to preserve the nonnegativity of the data. Other decomposition techniques are well known but the only matrix decomposition that insures the needed preservation of nonnegativity in data while achieving low–rank factorisation is NMF. The method can be described as: given a nonnegative data matrix V, find a decomposition $V = WH$ such that W and H are also nonnegative matrices.

There are two major characteristics of NMF that are immediately appealing: a) it provides a lower rank approximation formed by factors whose elements are kept nonnegative; b) it tends to produce sparse representations of the data. This sparseness can be further improved by imposing penalty weights into the objective function [12]. Thus, NMF can produce both object detection and recognition with characterization of a pattern (or classification of different patterns) as well as dimension reduction.

Within video and image signal processing, each data matrix V is made up of several images, showing a composite object in many articulations or poses, or a sequence of video frames. In a simplified way this means that: (a) each column (line) in V can be an image (frame) or a sequence of images; (b) the columns in W represent the basis elements for this image space; (c) the columns of H denote the coefficient sequences representing n images in the basis elements. The importance of NMF within Video Signal Processing is twofold: on one hand, it can be applied as a compression tool (previous to coding) since the factor matrices tend to be sparse and, if a good enough low-rank approximation is produced, there are much less matrix entries to code. Secondly, NMF can also detect scene boundaries or special features within the video, which can be used in video summarisation/segmentation tools or, even, for devising new methods for motion detection [6].

The paper is organised as follows. We begin by presenting the definition of an NMF minimisation problem using the most common formulation and classical algorithmic approaches. Section III offers a brief review of the most important theoretical results concerning the rank dimensionality and its importance for the existence of exact factorisations. In Section IV we present some computational experiments as a proof of concept towards the importance of the parameter that controls the factorisation rank, ending with Section V where conclusions are drawn and trends for future work are discussed.

II. Classical Numerical Optimisation Formulations and Algorithms

The NMF as an optimisation problem can be applied to various and potentially different applications. Nevertheless, our main motivation and examples will come from the area of Video and Image Signal Processing. Given a nonnegative matrix $V \in R^{m \times n}$, the most commonly used and general formulation for NMF as an optimisation problem is the minimisation of the Fröbenius norm:

$$\begin{aligned} \min \quad & \tfrac{1}{2}\|V - WH\|_F^2 \\ \text{s.t.} \quad & W_{m\times r} \geq 0\,, H_{r\times n} \geq 0\,. \end{aligned} \tag{1}$$

Clearly, the product WH is an approximate factorisation of rank r. The choice of the parameter r is, usually, problem dependent. However, it is generally chosen so that $1 < r << \min\{m, n\}$. Therefore, WH can be thought of as a compressed form of V.

Important challenges affecting the numerical optimisation include the existence of local minima due to the non-convexity of the objective function in both W and H. Within Video and Image Signal Processing, if a good enough approximation is obtained it is considered that a (good) solution was found since our eyes cannot detect but a certain level of image distortion.

The numerical approaches found in the related literature apply one of four general techniques: Alternating Constrained Least-Squares (e.g. [7], [14]); Multiplicative Update Rules (Fixed Point approach, namely [17]); Projected Gradient Descent Methods (e.g. [15], [16]); and Unconstrained Newton-type Methods (e.g. [9]). In 2006, Lin proposes the use of Projected Gradient descent (PG) methods for NMF [15]. The author proved that a projected gradient method solving least-squares subproblems leads to faster convergence than the, until then, most popular multiplicative update method by the previous authors. Recently, there has been a considerable growth of interest in PG methods: they are usually highly efficient in solving large-scale optimisation problems subject to linear constraints. The work presented in [15] started a surge of studies that attempt to present faster and better PG and Projected Alternating Least-Squares (PALS) methods (e.g. [7], [14]). In fact, Chichoki, Zudnek and others have been exploring the use of projected gradient descent methods with considerable success in their experiments [21].

III. A brief overview of former results

The numerical optimisation of formulation (1) uses $(n + m)r$ variables, which implies that the bigger the data matrix dimension the larger the scale of the problem resolution. Other issue relates to the existence of local minima due to the non-convexity of the objective function in both W and H. The fact that the norm is not a convex function implies that a stationary point is not necessarily a minimal solution for (1). Nevertheless, this formulation constitutes a global optimisation problem for which the minimum possible value is known: zero.

Another important point to be made is that, although the human eye cannot detect but a certain level of image distortion, still the major unresolved question in NMF within video and image fields is the insurance on the quality of the final approximate solution.

Is has been proven that if the elements of V are strictly positive then there exists an infinity of solutions for the optimisation of problem (1). Moreover, in order to have a unique solution, V must have zero value elements ([1]).

The following definition allows to establish a more useful condition to ensure the existence of an exact factorisation for a given matrix V: Given $V_{m\times n} \geq 0$, the minimum r such that there is $W_{m\times r}, H_{r\times n}$ for the exact factorisation $V = WH$, is the *positive rank* for V, $posrank(V)$ ([1]).

Thus, obtaining exact factorisations is directly dependent on the value of r that is used. On the other hand, we know that given $V_{m\times n} \geq 0$([1]),

$$rank(V) \leq posrank(V) \leq \min\{m, n\}\,.$$

There are some particular matrices and certain values of r for which we know that an exact decomposition exists. That is the case of any positive definite matrix whose Cholesky decomposition is an exact factorisation of rank $r = n$. Moreover, for $\{0, 1\}$ matrices there are particular results that either provide the minimal rank exact factorisation or are able to reduce the possible range of the posrank values ([3], [4], [5]). Nevertheless, for a general nonnegative matrix the exact determination of the value of a matrix $posrank$ is still unknown.

When thinking of global optimisation, the estimation of the $posrank$ value is especially important for the success of the approach since, as previously referred, if we assure that exact decompositions exist then we know the minimal possible value for the objective function. However, the existence of exact factorisations is directly dependent on the value of r that is used. The optimum, that is, the exact factorisation, can obviously be achieved fixing $r \geq posrank(V)$. In fact, the lowest rank exact factorisation is obtained using $r = posrank(V)$. For any value below this limit only approximate solutions exist.

IV. The influence of the $posrank$ value on NMF results

Hereinafter, all the algorithms used are Projected Gradient Descent based. Furthermore, the results presented were obtained either using the commercial optimisation software GAMS, namely, the MINOS solver for non-linear optimisation models, or the initial Alternated Least-Squares Projected Gradient (ALSPG) algorithm proposed by C.Lin [15] in 2006. The initial points used were randomly generated so as to limit the influence of the initialisation in the final result. The final difference outcomes are due only to the value of the chosen parameter r.

The experiments next described use images from known data sets, in particular, a shorter form of the Swimmer database that was introduced by Donoho and Stodden [10], and the ORL database that stores pictures of several faces (www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html).

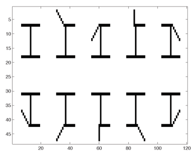


Fig. 1. Original Swimmer image picturing 10 different positions ($n = 48; m = 120$).

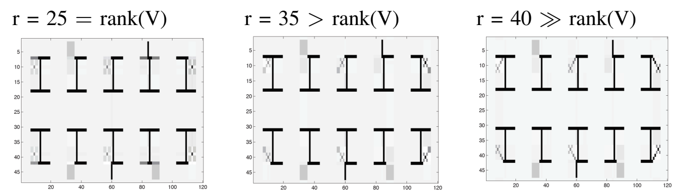
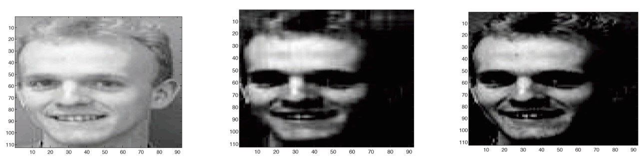


Fig. 2. Images obtained with MINOS model software when $r = 25, 35, 40$ and using $w_{ik} = 0.5 = h_{kj}$ as the initial values of W and H.



(a) (b) (c)

Fig. 3. (a) Original ORL image; Results obtained using LIN's ALSPG algorithm starting with P3 and setting: (b) worst result for $r = 16$ and (c) best result for $r = 92$.

A. Figures from the Swimmer database

So that we can draw conclusions using a known example we experimented to use a reduced *Swimmer* matrix. This example was taken from the Swimmer Database where we can see a stick figure with four limbs in articulated positions, depicting a swimmer (in a kind of) swim (Fig.1).

The sequence of images presented in Figure 2 shows that using MINOS software we were not able to find any good enough approximation for the original image. Nevertheless, setting r equal or bigger than the original data matrix rank, which is 25, the final results showed but an increase in image resolution.

B. Faces from the ORL database

The ORL repository is an archive of the AT&T Laboratories at Cambridge hosted in conjunction with The Digital Technology Group from Cambridge University Computer Laboratory. When compared with the previous images already experimented with these are different in the sense that the data matrices are full rank ones. This means that the $posrank$ value is exactly the rank of the original matrix and no exact factorisation exists with lower rank. Thus, the lower-rank nonnegative decomposition of any of these faces will always be an approximate one.

Hereinafter, and due to the fact that we access GAMS/MINOS software through an academic license, which rather limits the power and the solvers the software can provide, we resorted to use only the ALSPG Lin's approximation algorithm introduced in [15] (ALSPG algorithm).

TABLE I

PERFORMANCE OF ALSPG FOR DIFFERENT STARTING POINTS (P1: $w_{ik} = h_{ik} = 124$, P2: $w_{ik} = 10$; $h_{ik} = 246$ AND P3: $w_{ik} = 1$; $h_{ik} = 255$) AND DIFFERENT PARAMETER VALUES.

r = 92								
P1			P2			P3		
Iter	Tmp	OF	Iter	Tmp	OF	Iter	Tmp	OF
134	1000.00	0.49	143	1000.00	0.55	137	1000.00	0.49
1000	8818.74	0.41	1000	9237.13	0.49	1000	9724.43	0.40
1418	12000.00	0.40	1079	12000.00	0.49	1267	12000.00	0.39
r = 16								
P1			P2			P3		
Iter	Tmp	OF	Iter	Tmp	OF	Iter	Tmp	OF
904	1000.00	2.31	899	1000.00	2.30	929	1000.00	2.29
3830	4032.33	2.31	7002	6000.00	2.29	6681	6000.00	2.28
3830	4032.33	2.31	9106	10000.00	2.29	8746	10000.00	2.28

We chose, at random, one of the faces in the database since it is sufficient for settling our point (Fig.3 (a)). The pixel matrix representation has dimension 112×92 and rank 92. We tested several starting points and rank parameter values. However, due to the similarity between the final results, we chose to present only the ones concerning the lowest and the highest parameter values used, $r = 16$ and $r = 92$, respectively. The latter presents, as expected, the best visual performance. When compared with the other intermediate parameter results, is the only one that consistently presented values around 0.5 for the objective function in (1), i.e, the minimisation of the Fröbenius norm. In Table I we can observe that, with one exception, all the OF values stay below this threshold. The fact that the algorithm failed to converge to the optimum using $r = posrank$ is due to time limitations since an upper limit of 12000 seconds (more than 3h) was set. In fact, for each value of r, each row represents different stopping time bounds (1000s, 6000s, and 12000s, respectively). With $r = 16$ the ALSPG presented the worst–case performance of all. Notice that, although none of the recovered images in Figure 3 is as fair as desirable towards the original picture, the image (c) is more detailed and slightly less fuzzy than (b).

For all the other parameter values used, $r \in]16, 46]$, the final objective function values ranged between the ones shown in Table I. However, the smallest of the remaining values was never inferior to 1.0 thus far from the results obtained by using the $posrank$ parameter $r = 92$ for the decomposition.

V. CONCLUSIONS

The main goal of this work is that of highlighting the importance of the rank value for nonnegative matrix factorisation. The case-study here presented and its numerical experiments help stressing the fact that setting the decomposition parameter r respecting limiting bounds for $posrank$ is very important to assure that good enough quality images can be achieved.

Using posrank–based decompositions might imply that the parameter value r would be larger than desirable when compression is intended. If this is the case, sparsity in W and H is quite important not only for algorithms to efficiently compute the decomposition pair of matrices but also for subsequent image processing (coding) so that it can efficiently decrease the amount of loss of information.

An interesting trend for future work relates to the fact that although active-set methods can find stationary points (local minima) for small values of m, n and r are unable to efficiently process large-scale optimisation problems, which is the case for video and image data. The dimensions of the problems within these areas tend, in practice, to be quite large so Projected Gradient type algorithms can be useful for dealing with this kind of dimensions. There is one variation of PG algorithms that seems, to us, quite promising and that we intend to follow: the use of spectral controlled projected gradient methods.

REFERENCES

[1] de Almeida, On an optimization model for Approximate Nonnegative Matrix Factorization, Computational Intelligence and Decision Making: Trends and Applications, A.Madureira, C.Reis, and V. Marques (eds:), vol. 61, pp. 249-257, 2013.

[2] S. Ambikkumar and S.W. Drury, A reverse Hadamard inequality, Linear Algebra Applications, 247:83-89, 1996.

[3] F. Barioli and A. Berman, The maximal cp-rank of rank k completely positive matrices, Linear Algebra and its Applications 363:17–33, 2003.

[4] A. Berman and C. Xu, $\{0, 1\}$ Completely positive matrices, Linear Algebra and its Applications 399–3551, 2005.

[5] A. Berman and U.G. Rothblum A note on the computation of the CP-rank, Linear Algebra and its Applications 419, 1–7, 2006.

[6] S.S. Bucak and B. Gunsel, Video Content Representation by Incremental Non-negative Matrix Factorization, Proc. of IEEE International Conference on Image Processing (ICIP), II –113 – II –116, 2007.

[7] A. Cichocki and R. Zdunek, Multilayer nonnegative matrix factorization using projected gradient approaches. International Journal of Neural Systems, vol.17, pp431–446, 2007.

[8] A. Cichocki, R. Zdunek, A.H. Phan, and S. Amari, Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation, John Wiley, 2009.

[9] M.T. Chu, F. Diele, R. Plemmons, and S. Ragni, When Does Non-negative Matrix Factorization Give a Correct Decomposition into Parts? Advances in Neural Information Processing 16 (NIPS 2003 Proceedings), 2003.

[10] D. Donoho and V. Stodden, When Does Non-negative Matrix Factorization Give a Correct Decomposition into Parts? Advances in Neural Information Processing 16 (NIPS 2003 Proceedings), 2003.

[11] T. Feng, S.Z. Li, H-Y. Shum, and HJ . Zhang, Local Non-negative Matrix Factorization as a visual Representation IEEE Proc. of the 2nd Intern. Conf. on Development and learning (ICDL'02), 2002.

[12] N. Gillis and F. Glineur, Using underapproximations for sparse nonnegative matrix factorization, Pattern Recognition, 43(3):676–1687, 2010.

[13] N. J. Higham, Matrix Nearest Problems and Applications, In M. Gover and S.Barnett, editors, *Applications of Matrix Theory*, pp.1–27, Oxford University Press, 1989.

[14] J. Kim and H. Park, Toward Faster Nonnegative Matrix Factorization: A New Algorithm and Comparisons, Proc. of ICDM '08, Eighth IEEE International Conference on Data Mining, 2008.

[15] C. Lin, Projected Gradient Methods for Non-negative Matrix Factorization, Neural Computing, 19(10), 2007.

[16] C. Lin, On the Convergence of Multiplicative Update Algorithms for Nonnegative Matrix Factorization, IEEE Transactions on Neural Networks, 18(6): 1589–1596, 2007.

[17] D.D. Lee and H.S. Seung, Learning the parts of objects by non–negative matrix factorization, Nature, 401:788–791, 1999.

[18] D.D. Lee and H.S. Seung, Algorithms for nonnegative matrix factorization, Advances in Neural Information Processing, MIT Press, 13:556–562, 2001.

[19] P. Paatero, U. Tapper, Positive Matrix Factorization: a nonnegative factor model with optimal utilization of error estimates of data values, Environmetrics, 5:111–126, 1994.

[20] A. Pascual-Montano, J.M. Carabo, K. Kochi, D. Lehmann, and R. Pascual-Marqui, Nonsmooth Nonnegative Matrix Factorization (nsNMF), IEEE Trans. Pattern Analysis and Machine Intelligence(3):403–415, 2006.

[21] R. Zdunek and A. Chichoki, Fast nonnegative matrix factorization algorithms using projected gradient approaches for large-scale problems, Computational Intelligence and Neuroscience, vol. 2008, article 939567, 2008.