Technical University of Denmark

DTU

Approximate L0 constrained Non-negative Matrix and Tensor Factorization

Mørup, Morten; Madsen, Kristoffer Hougaard; Hansen, Lars Kai

DTU Library
Technical Information Center of Denmark

Approximate L_0 constrained Non-negative Matrix and Tensor Factorization

Morten Mørup, Kristoffer Hougaard Madsen and Lars Kai Hansen

Informatics and Mathematical Modelling
Technical University of Denmark
2800 Kgs Lyngby, Denmark
Email: {mm,khm,lkh}@imm.dtu.dk

Abstract— **Non-negative matrix factorization (NMF), i.e. $\mathbf{V} \approx \mathbf{WH}$ where both $\mathbf{V}$, $\mathbf{W}$ and $\mathbf{H}$ are non-negative has become a widely used blind source separation technique due to its part based representation. The NMF decomposition is not in general unique and a part based representation not guaranteed. However, imposing sparseness both improves the uniqueness of the decomposition and favors part based representation. Sparseness in the form of attaining as many zero elements in the solution as possible is appealing from a conceptional point of view and corresponds to minimizing reconstruction error with an L_0 norm constraint. In general, solving for a given L_0 norm is an NP hard problem thus convex relaxation to regularization by the L_1 norm is often considered, i.e., minimizing ($\frac{1}{2}\|\mathbf{V}-\mathbf{WH}\|_F^2 + \lambda\|\mathbf{H}\|_1$). An open problem is to control the degree of sparsity λ imposed. We here demonstrate that a full regularization path for the L_1 norm regularized least squares NMF for fixed $\mathbf{W}$ can be calculated at the cost of an ordinary least squares solution based on a modification of the Least Angle Regression and Selection (LARS) algorithm forming a non-negativity constrained LARS (NLARS). With the full regularization path, the L_1 regularization strength λ that best approximates a given L_0 can be directly accessed and in effect used to control the sparsity of $\mathbf{H}$. The MATLAB code for the NLARS algorithm is available for download.**

I. INTRODUCTION

Non-negative matrix factorization (NMF) is a blind source separation algorithm (BSS) given by the decomposition

$$\mathbf{V}_{n,m} \approx \sum_d \mathbf{W}_{n,d}\mathbf{H}_{d,m}, \qquad (1)$$

where $\mathbf{V} \in \mathbb{R}_+^{N \times M}$, $\mathbf{W} \in \mathbb{R}_+^{N \times D}$ and $\mathbf{H} \in \mathbb{R}_+^{D \times M}$, i.e. such that the variables $\mathbf{V}$, $\mathbf{W}$ and $\mathbf{H}$ are non-negative. The decomposition is useful because it results in easy interpretable part based representations [1]. Non-negative decompositions is also named positive matrix factorization [2] but was popularized by Lee and Seung due to a simple algorithmic procedure based on multiplicative updates [3]. The decomposition has proven useful for a wide range of data where non-negativity is a natural constraint. These encompass data for text-mining based on word counts, image data, biomedical data, and spectral data. The algorithm can also be useful even when the data in itself is negative by considering, say, the amplitude of a spectral representation [4]. The model is often estimated alternatingly solving for $\mathbf{W}$ and $\mathbf{H}$. While the estimation of $\mathbf{H}$ for fixed $\mathbf{W}$ is a convex problem, the joint problem solving for $\mathbf{W}$ and $\mathbf{H}$ is not convex and as such a global minimum is not guaranteed. For data with modality more than two, i.e.,

tensors, purely non-negative decompositions are also relevant. The two most widely used decompositions for tensors are the PARAFAC model independently proposed in [5], [6] and the Tucker model [7]. The two model read

PARAFAC:

$$\mathcal{V}_{i_1,i_2,\ldots,i_N} \approx \sum_d \mathbf{A}_{i_1,d}^{(1)}\mathbf{A}_{i_2,d}^{(2)} \cdots \mathbf{A}_{i_n,d}^{(N)}$$

Tucker:

$$\mathcal{V}_{i_1,i_2,\ldots,i_N} \approx \sum_{j_1,j_2,\ldots,j_N} \mathcal{G}_{j_1,j_2,\ldots,j_n}\mathbf{A}_{i_1,j_1}^{(1)}\mathbf{A}_{i_2,j_2}^{(2)} \cdots \mathbf{A}_{i_N,j_N}^{(n)}$$

In [8] it was demonstrated how NMF generalizes well to the PARAFAC model and in [9] how NMF also generalizes to the Tucker model forming fully non-negative decompositions as for regular NMF resulting in part based representation. The generalization of NMF to the two models are based on the observation that by matricizing the array, i.e. $\mathbf{V}_{(n)} \in R^{I_n \times I_1 I_2 I_{n-1} I_{n+1} \cdots I_N}$ the optimization of the loading of each mode can be restated as regular NMF-problem, for details on this see [9]. As such, the PARAFAC and Tucker models can be estimated through algorithms for regular matrix decomposition alternating solutions for each mode keeping the other modes fixed. We will here only treat the matrix case keeping in mind that the approach generalizes to tensors.

The non-negative decomposition is in general not unique [10] which hampers interpretation of the decomposition. Furthermore, the representation is not guaranteed to be part-based for instance if the data resides well inside the positive orthant. To improve the uniqueness of the decomposition as well as enforcing a part based representation sparseness constraints have been suggested for the NMF decomposition. This has been implemented by penalizing with the L_1 norm [11], [12], i.e.

$$C_{SparseNMF} = \frac{1}{2}\|\mathbf{V} - \mathbf{WH}\|_F^2 + \lambda\|\mathbf{H}\|_1 \qquad (2)$$

where $\|\mathbf{H}\|_p$ is the L_p-norm of $\mathbf{H}$ given by $L_p(\mathbf{H}) = \sqrt[p]{\sum_{d,m}|\mathbf{H}_{d,m}|^p}$ (Notice, for the L_0 norm we use $L_0(\mathbf{H}) = \sum_{d,m}|\mathbf{H}_{d,m}|^0$). The L_1 norm is often used as a penalty for sparsity as it is known to mimic the behavior of the L_0 norm [13] while keeping the alternating optimization problem estimating $\mathbf{H}$ for fixed $\mathbf{W}$ convex. Thus, although often the aim is to find a sparse solution with respect to the L_0-norm,

in practice the NP-hard optimization problem is sought solved with the L_1-norm as a proxy to avoid local minima in the optimization problem.

In [14] the sparseness of the decomposition was controlled by the sparseness measure

$$sp_{\text{PH}-\text{L1/L2}}(\mathbf{H}_{d,:}) = \frac{1}{\sqrt{M}-1}(\sqrt{M} - \frac{\|\mathbf{H}_{d,:}\|_1}{\|\mathbf{H}_{d,:}\|_2}), \quad (3)$$

where $\mathbf{H}_{d,:}$ denotes the dth row of $\mathbf{H}$. Notice, $sp_{\text{PH}-\text{L1/L2}}$ takes values between 0 and 1. This was achieved solving a constrained minimization problem forming the Non-negative matrix factorization with Sparseness Constraint (NMFSC) algorithm described in [14]. Solving for a specific sparsity level for each component is a difficult problem. However, solving for a specific sparsity on the full matrix $\mathbf{H}$ mounts to controlling the single parameter λ which we presently demonstrate has a simple solution.

Consider the unconstrained Least squares minimization with L_1-norm constraint also referred to as the Least Absolute Shrinkage and Selection Operator (LASSO) [15] or Basis Pursuit De-noising (BPD) [16], i.e., the problem

$$C_{\text{LASSO}} = \frac{1}{2}\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_F^2 + \lambda\|\boldsymbol{\beta}\|_1. \quad (4)$$

Through the Least Angle Regression and Selection algorithm (LARS) [17] or the equivalent homotopy algorithm [18] the entire regularization path, i.e., solution for all values of λ can be achieved for the cost of an ordinary least squares (OLS) solution. We will in the present paper modify the LARS algorithm such that the entire regularization path can be found for the Sparse NMF decomposition. With this result the sparsity can be directly controlled for the following two new types of sparsity measures on the full matrix $\mathbf{H}$

$$sp_{\text{L0}}(\mathbf{H}) = 1 - \frac{\|\mathbf{H}\|_0}{DM} \quad (5)$$

$$sp_{\text{L1/L2}}(\mathbf{H}) = \frac{1}{\sqrt{DM}-1}(\sqrt{DM} - \frac{\|\mathbf{H}\|_1}{\|\mathbf{H}\|_2}). \quad (6)$$

The measure $sp_{\text{L1/L2}}(\mathbf{H})$ is equivalent to the measure proposed in [14], but for the full matrix rather than for each row of $\mathbf{H}$ and is included to give a qualitative comparison to the measure given in equation 3. sp_{L0} on the other hand gives direct control of the percentage of active elements in $\mathbf{H}$, i.e., the L_0 norm of $\mathbf{H}$. Both measures take values between 0 and 1, for instance $sp_{\text{L0}}(\mathbf{H}) = 1$ implies that all elements of $\mathbf{H}$ are turned off while $sp_{\text{L0}}(\mathbf{H}) = 0$ that all elements are nonzero.

II. METHOD

In the paper on Least Angle Regression and Selection (LARS) [17] an algorithm was proposed for the LASSO by considering normalized data $\mathbf{X}$ and $\mathbf{y}$ formed by an active set algorithm. Let A be the indices of the active set and I the indices of the inactive set. Then the LARS algorithm solves for the LASSO as

LARS
while $\exists j \in I : |\mathbf{c}_j| > 0$
$\mathbf{c} = \mathbf{X}^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$
$j = \arg\max(|\mathbf{c}_I|)$
$A = [A\ j]$
$I = I \setminus j$
$\boldsymbol{\beta}_A = \boldsymbol{\beta}_A + \mu(\mathbf{X}^\top\mathbf{X})_{A,A}^{-1}\operatorname{sign}(\mathbf{c}_A)$
$\mu = \arg\min_\mu \begin{cases} \exists\widetilde{\boldsymbol{\beta}}_{A_k} = 0 \text{ then } I = [I\ A_k], A = A \setminus A_k \\ \exists l \in I : \ |\widetilde{\mathbf{c}}_l| = |\widetilde{\mathbf{c}}_A| \\ \widetilde{\mathbf{c}}_A = \mathbf{0} \end{cases}$

where $\widetilde{\mathbf{c}}$ is the gradient and $\widetilde{\boldsymbol{\beta}}$ the value of $\boldsymbol{\beta}$ at the step of size μ. We make the following observations

1) $\mathbf{c} = \mathbf{X}^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ is identical to the negative gradient of the un-regularized problem.
2) Since the inverse Hessian is given by $(\mathbf{X}^\top\mathbf{X})^{-1}$ the step $(\mathbf{X}^\top\mathbf{X})_{A,A}^{-1}\operatorname{sign}(\mathbf{c}_A)$ will be in the direction such that the amplitude of the gradients in the active set are identical, i.e. $|\mathbf{c}_{A_1}| = |\mathbf{c}_{A_2}| = \ldots = |\mathbf{c}_{A_n}|$.
3) Since the amplitude of the gradients of the active set are identical the update correspond to the Newton-Raphson step $\boldsymbol{\beta}_A = \boldsymbol{\beta}_A + \gamma(\mathbf{X}^\top\mathbf{X})_{A,A}^{-1}\mathbf{c}_A$ where $\gamma = \frac{\mu}{|\mathbf{c}_A|}$.
4) The gradient of the L_1 penalty is $\lambda\operatorname{sign}(\boldsymbol{\beta})$ and as such thresholds the gradient of the un-regularized problem such that the solution for a given regularization λ is found when $|\mathbf{c}_A| = \lambda$.

From observation one there is no reason to normalize the data. This serves only the purpose of interpreting the selection by angles, however, in general the selection is according to the largest absolute gradient. Add 2 and 3 it follows that the update is along the line where the gradients $\mathbf{c}_A$ in the active set are identical and corresponds to a Newton-Raphson step in the active variables. Notice, for least squares problems the inverse Hessian is a constant within each active set forming the piecewise linearity of the LARS path. Since the gradient of the L_1 constraint is $\lambda\operatorname{sign}(\boldsymbol{\beta})$, λ serves as a threshold of the gradient of the LASSO problem given by $g = -\mathbf{c} + \lambda\operatorname{sign}(\beta)$. Thus, does not allow for additional elements to enter the active set when $|\mathbf{c}_j| < \lambda$. By simply replacing the criterion $j = argmax(|\mathbf{c}_I|)$ with $j = argmax(\mathbf{c}_I)\ for\ \mathbf{c}_j > 0$, i.e. considering only negative gradients of the least squares objective - variables are only introduced if they reduce the objective by becoming positive. Once a variable is updated to zero it is taken out of the active set. Thus, the following modification of the above LARS algorithm forms the non-negativity constrained LARS (NLARS) (i.e., for $\boldsymbol{\beta} \geq 0$). For more details on the algorithm see the MATLAB code available from [19].

NLARS
while $\exists j \in I : \mathbf{c}_j > 0$
$\mathbf{c} = \mathbf{X}^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$
$j = \arg\max(\mathbf{c}_I), \ \mathbf{c}_j > 0$
$A = [A\ j]$
$I = I \setminus j$
$\boldsymbol{\beta}_A = \boldsymbol{\beta}_A + \mu(\mathbf{X}^\top\mathbf{X})_{A,A}^{-1}\mathbf{1}$
$\mu = \arg\min_\mu \begin{cases} \exists\ \boldsymbol{\beta}_{A_k} = 0 \text{ then } I = [I\ A_k], A = A \setminus A_k \\ \exists l \in I : \ \widetilde{\mathbf{c}}_l = \widetilde{\mathbf{c}}_A \\ \widetilde{\mathbf{c}}_A = \mathbf{0} \end{cases}$

Where $\mathbf{1}$ is a vector of ones. The sparse NMF problem can

now be separated into N LASSO problems where the nth problem is given by

$$C_{\text{SparseNMF}} = \frac{1}{2}\|\mathbf{v}_n - \mathbf{W}\mathbf{h}_n\|_F^2 + \lambda\|\mathbf{h}_n\|_1 \qquad (7)$$

Notice, the algorithm works even when $\mathbf{W}$ and $\mathbf{V}$ are unconstrained. Having solved for each column of $\mathbf{H}$ we need to order the given LASSO solutions. However, this is easy since we can directly calculate the value of λ corresponding to each part of the regularization path of each given non-negative LASSO solution as

$$\boldsymbol{\lambda} = (\mathbf{W}_{A,:}^\top \mathbf{v}_m - (\mathbf{W}^\top\mathbf{W})_{A,A}\mathbf{h}_{A,m}). \qquad (8)$$

Notice, all entries in $\boldsymbol{\lambda}$ take the same value λ. Hence, the solutions of each column of $\mathbf{H}$ can be sorted according to this value. The sparseness measures given in equation 5 can now be controlled by evaluating the sparseness degree over the entire regularization path. Where several values of λ result in same sparseness degree, the smallest value of λ was chosen as regularization strength. Let $\cdot \frac{\cdot}{\cdot} \cdot$ and $\cdot \bullet \cdot$ denote element wise division and multiplication, respectively. Then, $\mathbf{W}$ was updated according the the normalization invariant multiplicative updates proposed in [11] given by

$$\mathbf{W} \leftarrow \mathbf{W} \bullet \frac{\mathbf{V}\mathbf{H}^\top + \mathbf{W}\operatorname{diag}(\mathbf{1}(\mathbf{W}\mathbf{H}\mathbf{H}^\top \bullet \mathbf{W}))}{\mathbf{W}\mathbf{H}\mathbf{H}^\top + \mathbf{W}\operatorname{diag}(\mathbf{1}(\mathbf{V}\mathbf{H}^\top \bullet \mathbf{W}))}, \qquad (9)$$

such that $\mathbf{W}_d$ is constrained to have unit L_2 norm, i.e., $\mathbf{W}_{n,d} = \frac{\mathbf{W}_{n,d}}{\|\mathbf{W}_d\|_2}$. We solve for $\mathbf{H}$ alternatingly according to the NLARS and return the $\mathbf{H}$ corresponding to the correct degree of the sp_{L0} criterion and update $\mathbf{W}$ as given above. The algorithm was stopped after 100 iterations or when the relative change in λ was less than 10^{-6}.

III. RESULTS

We tested our sp_{L0} constrained NMF algorithm on the 2429 training faces of size 361 pixels obtained from the MIT face database [20] (see figure 1) as well as the US Postal Service (USPS) database of 7291 handwritten digit each of size 256 pixels - the training set (see figure 2). We compared our sparseness measure given in equation 5 to the corresponding decomposition obtained from the same degree of sparseness according to equation 3 using the algorithm given in [14]. As the algorithm given in [14] converges slowly we terminated this algorithm after 2500 iterations. Finally, we compared the performance of the proposed algorithm with the fast Non-Negative Least Squares (fNNLS) algorithm described in [21] which is also based on an active set approach. Contrary, to the present NLARS algorithm essentially progressing each step until inactive elements have same negative gradient as the active elements, the fNNLS takes the step in the active set to the optimal unconstrained solution and backtracks this solution to the positive orthant. Thus the strategies employed by NLARS is entirely different from the strategy of the fNNLS algorithm. The computational time for the two algorithms tested on the USPS and MIT face data sets are given in table 3.
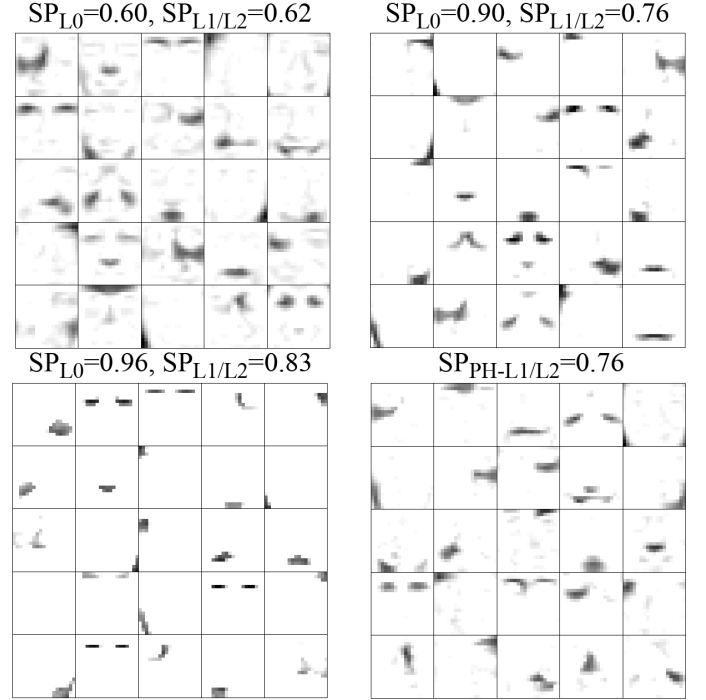

Fig. 1. Result obtained analyzing the MIT face database using the NLARS to control the sparsity of the decomposition. Top left panel: Sparsity set to $sp_{L0} = 0.6$. Top right panel: Sparsity set to $sp_{L0} = 0.9$. Bottom left panel: Sparsity set to $sp_{L0} = 0.96$. Bottom right panel: The NMFSC algorithm with sparsity set to $sp_{PH-L1/L2} = 0.76$ corresponding to the value of sparsity given in the top right panel of $sp_{L1/L2} = 0.76$. The result obtained by the proposed NLARS resembles well the results obtained by the NMFSC algorithm. However, regularizing on the full matrix $\mathbf{H}$ does not fix the degree of sparsity within each components and as such each component has different degree of sparsity. Thus, component 16 is completely turned off when $sp_{LO} = 0.96$.

IV. DISCUSSION

We have proposed an algorithm for non-negative matrix factorization with an L_0 norm constraint based on convex relaxation, i.e, invoking an L_1 norm as a proxy. This has the benefit that the NP-hard problem of constraining $\mathbf{H}$ by a given L_0-norm, sp_{L0}, for a fixed $\mathbf{W}$ can be approximated by a convex problem which is efficiently solved through the also proposed NLARS algorithm. The main benefit is that the entire regularization path can be found for the cost of an ordinary least squares solution. By evaluating the L_0 norm of the solution throughout the path, we can find the appropriate strength of the L_1 regularization term λ that will provide a specific value of the L_0 norm.

In [14] sparseness was controlled by constraining each component to a given degree of sparseness. The present measures of sparseness given by controlling the degree of sparsity on the full matrix $\mathbf{H}$ rather than each row $\mathbf{H}_{d,:}$ has the benefit that the control of regularization reduces to the single parameter λ. This is a much simpler problem in which the entire regularization path formed by the NLARS algorithm can be used to directly control this value. In general, controlling the sparseness componentwise versus controlling

SP$_{L0}$=0.80, SP$_{L1/L2}$=0.72   SP$_{L0}$=0.90, SP$_{L1/L2}$=0.77

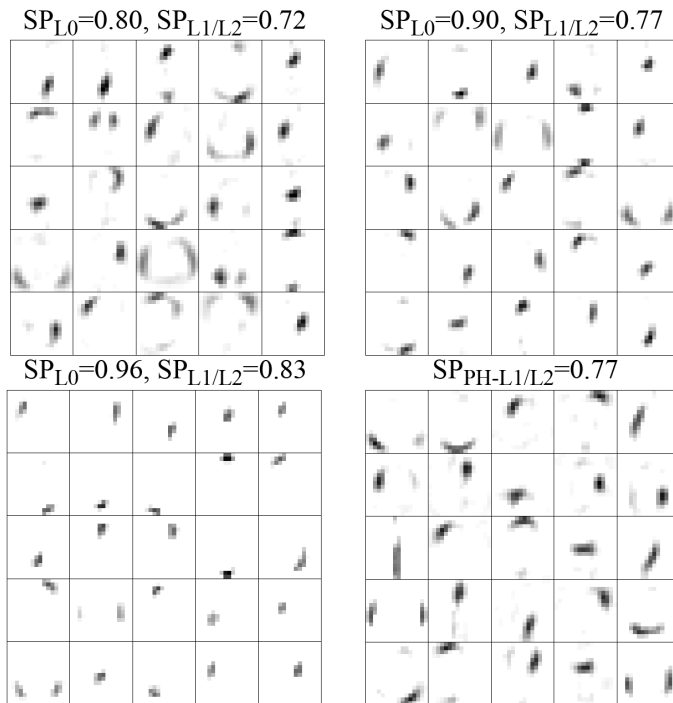SP$_{L0}$=0.96, SP$_{L1/L2}$=0.83   SP$_{PH-L1/L2}$=0.77

Fig. 2. Result obtained analyzing the USPS handwritten digit database using the NLARS to control the sparsity of the decomposition. Top left panel: Sparsity set to $sp_{L0} = 0.8$. Top right panel: Sparsity set to $sp_{L0} = 0.9$. Bottom left panel: Sparsity set to $sp_{L0} = 0.96$. Bottom right panel: The NMFSC algorithm with sparsity set to $sp_{PH-L1/L2} = 0.77$ corresponding to the value of sparsity given in the top right panel of $sp_{L1/L2} = 0.77$. Again the solution for $sp_{L0} = 0.9$ having $sp_{L1/L2} = 0.77$ correspond well to the solution obtained for $sp_{PH-L1/L2} = 0.77$ using the NMFSC algorithm.

	USPS Handwritten Digit	MIT face database
NLARS	3361 sec	2520 sec
fNNLS	2901 sec	2129 sec

Fig. 3. Comparison of the NLARS and fNNLS algorithm for the USPS and MIT face data sets. The solution given is obtained for the regular NMF problem without regularization corresponding to finding the full regularization path for the NLARS algorithm. The fNNLS and NLARS were also used for the update of **W**. The time is given as the time it took each of the two algorithms to perform 100 iterations. Clearly, the two algorithms are of same order of magnitude, hence the full regularization path can be calculated at the cost of the ordinary LS solution given by fNNLS.

the sparsity of the full matrix as here have different properties. It was demonstrated for the MIT face data that sparseness imposed on the full matrix enabled some components to be less sparse than others and also for components to be turned off completely whereas controlling the degree of sparseness of each component separately enabled us to fix the number of components used to represent the data. Both for the MIT face as well as the USPS data the results obtained using the proposed algorithm corresponded well to the results obtained for the corresponding degree of sparsity using the NMFSC algorithm. Thus, controlling the sparseness of the full matrix **H** through the NLARS is a viable approach to control the sparsity of the solutions.

Both the present NLARS algorithm as well as the fNNLS

algorithm given in [21] for non-negativity constrained least squares minimization are based on an active set approach. As seen from figure 3 the two algorithms had about the same computational cost. Thus, the present NLARS algorithm is comparable to 'state of the art' non-negative least squares algorithms such as the fNNLS while also allowing for finding the full regularization path solution.

The sparsity measures proposed can also be used to control the degree of sparsity for sparse coding [22], i.e., unconstrained decompositions with L_1-norm regularization by the regular LARS algorithm rather than the non-negative extension presently derived. Finally, it is worth emphasizing that solving for the sp_{L0} constrained problem using the L_1-norm is only approximate and as such does not necessarily find the optimal combination of elements to be active. The NLARS algorithm can be downloaded from [19].

REFERENCES

[1] D.D. Lee and H. Seung *Learning the parts of objects by non-negative matrix factorization. Nature*, 401(6755):788–91, 1999.
[2] P. Paatero and U. Tapper *Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics*, 5(2):111–126, 1994.
[3] D. D. Lee and H. S. Seung *Algorithms for non-negative matrix factorization.* In *NIPS*, pages 556–562, 2000.
[4] P. Smaragdis and J. C. Brown *Non-negative matrix factorization for polyphonic music transcription. WASPAA*, pages 177–180, 2003.
[5] R. A. Harshman *Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. UCLA Working Papers in Phonetics*, 16:1–84, 1970.
[6] J. D. Carroll and J. J. Chang *Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. Psychometrika*, 35:283–319, 1970.
[7] L. R. Tucker *Some mathematical notes on three-mode factor analysis. Psychometrika*, 31:279–311, 1966.
[8] M. Welling, and M. Weber *Positive tensor factorization. Pattern Recogn. Lett.*, 22(12):1255–1261, 2001.
[9] M. Mørup, L.K. Hansen and S. M. Arnfred *Algorithms for sparse non-negative tucker. Submitted Neural Computation.*
[10] D. Donoho and V. Stodden *When does non-negative matrix factorization give a correct decomposition into parts? NIPS*, 2003.
[11] J. Eggert and E. Korner *Sparse coding and nmf.* In *Neural Networks*, volume 4, pages 2529–2533, 2004.
[12] P. Hoyer *Non-negative sparse coding. Neural Networks for Signal Processing, 2002.*, pages 557–565, 2002.
[13] D. Donoho *For most large underdetermined systems of linear equations the minimal l^1-norm solution is also the sparsest solution. Communications on Pure and Applied Mathematics*, 59(6):797–829, 2006.
[14] P. Hoyer *Non-negative matrix factorization with sparseness constraints. Journal of Machine Learning Research*, 5, 1457–1469.
[15] R. Tibshirani *Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B*, 58(1):267–288, 1996.
[16] S. C. Shaobing and D. Donoho *Basis pursuit. 28th Asilomar conf. Signals, Systems Computers*, 1994.
[17] B. Efron, T. Hastie, I. Johnstone and R. Tibshirani *Least angle regression. Annals of Statistics*, 32(2):407–499, 2004.
[18] M. Osborne, B. Presnell and B. Turlach *A new approach to variable selection in least squares problems. IMA Journal of Numerical Analysis*, 20(3):389–403, 2000.
[19] M. Mørup and K. Madsen *Nlars algorithm.* http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/5523/zip/imm5523.zip, 2007.
[20] http://www.ai.mit.edu/projects/cbcl. Cbcl face database ♯1. MIT Center For Biological and Computation Learning.
[21] R. B. Bro, and S. D. Jong, *A fast non-negativity-constrained least squares algorithm. Journal of Chemometrics*, 11(5):393–401, 1997.
[22] B. A. Olshausen and D. J. Field *Sparse coding of sensorty inputs. Current Opinion in Neurobiology*, 14:481–487, 2004.