CORRELATED NOISE: HOW IT BREAKS NMF, AND WHAT TO DO ABOUT IT

Sergey M. Plis[2*], *Vamsi K. Potluru*[2†], *Vince D. Calhoun*[1,2,3], *Terran Lane*[2]

[1]Electrical and Computer Engineering Department, University of New Mexico,New Mexico, USA, 87131
[2]Computer Science Department, University of New Mexico,New Mexico, USA, 87131
[3]MIND Research Network, New Mexico, USA, 87131

ABSTRACT

Non-negative matrix factorization (NMF) is an algorithm for decomposing multivariate data into a signal dictionary and its corresponding activations. When applied to experimental data, NMF has to cope with noise, which is often highly correlated. We show that correlated noise can break the Donoho and Stodden separability conditions of a dataset and a regular NMF algorithm will fail to decompose it, even when given freedom to be able to represent the noise as a separate feature. To cope with this issue, we present an algorithm for NMF with a generalized least squares objective function (glsNMF) and derive multiplicative updates for the method together with proving their convergence. The new algorithm successfully recovers the true representation from the noisy data. Robust performance can make glsNMF a valuable tool for analyzing empirical data.

Index Terms— NMF, GLS, parts based representation, correlated noise

1. INTRODUCTION

Since the introduction of multiplicative updates for non-negative matrix factorization (NMF) [1], the algorithm has gained general recognition. Simplicity of implementation, an adaptive learning rate and automatic satisfaction of positivity constraints are in part responsible for the wide acceptance of the algorithm. It has been successfully used to analyze functional brain imaging data [2, 3, 4], gene expression [5], and other empirical datasets.

Lee and Seung provide two updates for NMF: one is based on the least squares (LS) criteria and the other on Kullback-Leibler (KL) divergence. In this study we focus on LS updates, for which the data model is:

$$X = WH + \epsilon, \tag{1}$$

where each entry in X, W and H is greater than or equal to zero and ϵ is Gaussian noise. Subsequent sections provide further details.

The LS formulation implicitly assumes that the noise is white. This is a widely used assumption and it is valid in many realistic cases with large number of independent noise sources. However, in many experimental settings noise is more complicated and is not limited to white sensor noise. In these environments, noise represents background activity, which can have complex covariance structure. Ignoring the structure in the noise can change the results of NMF substantially.

Donoho and Stodden introduced the notion of a separable factorial articulation family [6] as a collection of points obeying three conditions: generative model, separability and complete factorial sampling. Datasets satisfying these conditions are guaranteed to be properly factored by any correct NMF algorithm. Presence of correlated noise may, however, violate the conditions and render otherwise separable dataset not factorable by NMF. We demonstrate an example in which a dataset that otherwise satisfies the Donoho and Stodden conditions is not factored properly by the regular NMF when contaminated with correlated noise. We also show that despite a reasonable expectation that an NMF model of sufficient rank can recover correlated components in the noise simply as nuisance features, NMF fails to do so and most features are recovered incorrectly. Undermodeling the noise may lead to a misinterpretation of results when applied to real dataset without known ground truth.

As a solution, we introduce a generalized least squares NMF (glsNMF) algorithm that takes the correlation structure of the noise into account. We derive multiplicative updates for the algorithm providing a proof of their convergence and demonstrate the algorithm's performance on a synthetic dataset. We show that the new algorithm handles the noisy data better than the LS based algorithm and produces the expected unique factors.

2. NMF

NMF is a tool producing a low rank approximation to a non-negative data matrix by splitting it into a product of two non-negative matrix factors. The constraint of non-negativity (all elements are ≥ 0) usually results in a parts-based representation making NMF different from other factorization tech-
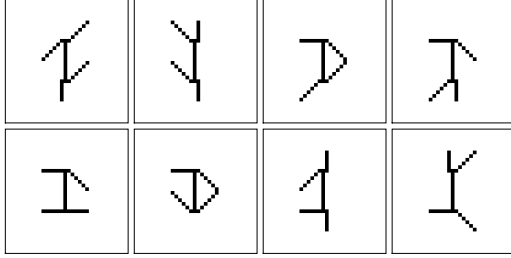
Fig. 1. Randomly selected samples from the swimmer dataset of [6], which consists of 256 total images with all possible combination of limbs positions.

niques yielding more holistic representations, such as principal component analysis (PCA) and vector quantization (VQ).

Using standard notation [1, 6], we formally define NMF task as follows. Given a non-negative $m \times n$ matrix $\boldsymbol{X}$, represent it with a product of two non-negative matrices $\boldsymbol{W}, \boldsymbol{H}$ of sizes $m \times r$ and $r \times n$ respectively:

$$\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H}. \tag{2}$$

The non-negativity constraint corresponds to the intuitive notion of features adding up, without canceling each other, to give the resulting data.

Lee and Seung [1] describe two multiplicative update rules for $\boldsymbol{W}$ and $\boldsymbol{H}$ which work well in practice. The updates correspond to two different cost functions representing the quality of approximation. In this work, we use the Frobenius norm for the cost function:

$$E = \frac{1}{2}\|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}\|_F \tag{3}$$

and the corresponding updates are:

$$\boldsymbol{W} = \boldsymbol{W} \odot \frac{\boldsymbol{X}\boldsymbol{H}^T}{\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^T} \tag{4}$$

$$\boldsymbol{H} = \boldsymbol{H} \odot \frac{\boldsymbol{W}^T\boldsymbol{X}}{\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{H}}, \tag{5}$$

where $\|.\|_F$ denotes the Frobenius norm and the operator $\odot$ represents element-wise multiplication. Division is also element-wise. We have omitted iteration indices for clarity. It should be noted that the cost function to be minimized is convex in either $\boldsymbol{W}$ or $\boldsymbol{H}$ but not in both [1]. As the algorithm iterates using the updates (4) and (5), $\boldsymbol{W}$ and $\boldsymbol{H}$ converge to a local minimum of the cost function.

The slightly mysterious form for the above updates can be understood as follows. A simple additive update for $\boldsymbol{H}$ is given by:

$$\boldsymbol{H} = \boldsymbol{H} + \boldsymbol{\eta} \odot (\boldsymbol{W}^T\boldsymbol{X} - \boldsymbol{W}^T\boldsymbol{W}\boldsymbol{H}) \tag{6}$$

If the learning rate given by the matrix elements of $\boldsymbol{\eta}$ are all set to some small positive number then this is the conventional gradient descent. However, setting the learning rate
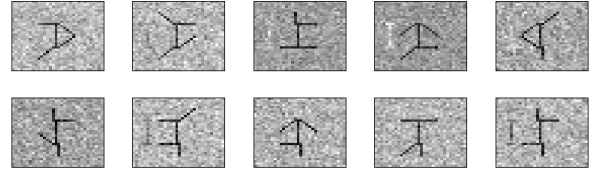
Fig. 2. Random samples from the swimmer dataset contaminated by correlated noise. Note the salient spatial correlation with the shape of the swimmer's torso to the left of the swimmer.
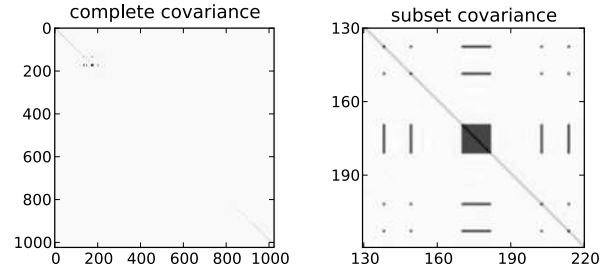
Fig. 3. Noise covariance for all 1024 image pixels shows the expected spatial correlations. The covariance is very close to identity, only the 90 by 90 close up on the right shows the correlated part. Such a covariance matrix is favorable to the conventional least squares analysis because it satisfies the assumptions of the method.

matrix as follows:

$$\boldsymbol{\eta} = \frac{\boldsymbol{H}}{\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{H}}, \tag{7}$$

where division is element-wise, produces the NMF updates.

3. FAILURE MODE

An example of a separable factorial articulation family is the swimmer dataset presented in [6]. The dataset contains 256 32×32 images of a swimmer with all possible combinations of limbs positions, as shown in Figure 1.

In order to study the effect of the correlated noise on the algorithm we have constructed such noise, where a small part of the image is spatially correlated. Figure 2 shows several random samples from the swimmer dataset of Figure 1 contaminated by the correlated noise.

The LS objective function of (3) can be derived from the Gaussian likelihood with noise covariance of the form $\sigma^2\boldsymbol{I}$. Note that correlated noise results in a differently structured covariance matrix. The covariance of the correlated noise is shown in Figure 3. It is clearly very close to a diagonal matrix. For comparison, the figure also shows a close up image of a section of the covariance, where there are correlations. Correlations among the 2% of the image pixels are high as demonstrated by the high contrast of the covariance image in
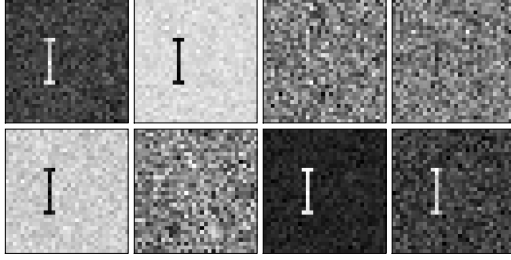
Fig. 4. Correlated noise samples that were added to the swimmer dataset. Note the salient spatial correlation.

the figure. Several samples of the noise are shown in Figure 4. The correlated part has the shape of the swimmer's torso shifted to the left of the original torso position. In summary the noise is mostly white, with a small locally concentrated correlated component.

A reasonable expectation of NMF's behavior on this dataset would be a result that has the correlation torso as a separate feature with the other features correctly recovered. This common-sense behavior would go along with other matrix factorization techniques such as independent component analysis (ICA), which exploit this feature for denoising. Surprisingly, we have observed quite different behavior. A typical result is shown in Figure 5, where it becomes clear that correlated noise affects estimation of all of the features, as opposed to be estimated just as a separate one. For comparison the correct factorization that is obtained by NMF from the noiseless dataset is shown in Figure 6. Note that we have observed similar behavior even when using the KLD divergence based NMF objective, although we do not pursue this finding further in this work.

Introduction of the torso-shaped correlation in the noise violates the *separability* condition from [6]. The condition requires that each part/articulation pair's presence or absence in the image is indicated by a certain pixel associated to that pair. However, the torso-shaped correlation, when present, can overlap with limbs in some positions. Note that conditions of generative model and complete factorial sampling are still satisfied, because the correlation can always be treated as a separate feature.

4. PROPOSED SOLUTION

We argue that in practical applications of NMF one needs to model the noise adequately. Here we propose an NMF algorithm that alleviates the problem caused by the correlated noise.

One of the objective functions for non-negative matrix factorization proposed in [1] is the least squares error (LSE) of (3). After rewriting (3) in the matrix form:

$$E = \frac{1}{2}\text{Tr}((\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H})^T(\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H})), \qquad (8)$$
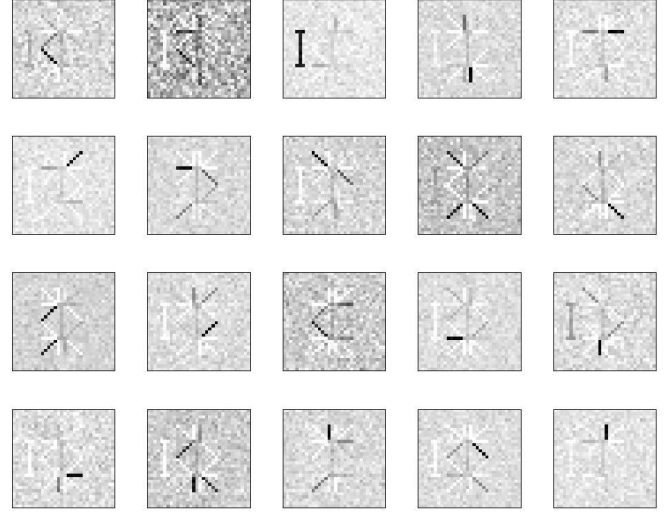
Fig. 5. 20 features (columns of the $\boldsymbol{W}$ matrix) as extracted by the regular NMF algorithm with the LS objective function from the dataset contaminated by correlated noise.

the assumption of zero mean noise with unit standard deviation becomes explicit.

For optimization purposes, the formulation is also valid for noise with covariance structure $\sigma^2 \boldsymbol{I}$. Richer noise structures, including those with diagonal covariance or correlated noise, are not captured by such a formulation. The former problem has been addressed in [7]. In this case the scaling of each dimension by a corresponding positive variance is performed. Scaling by the positive constants does not alter the original formulation of multiplicative updates.

We address the problem of generalized least squares (GLS) of (9), where $\boldsymbol{C}$ is the noise covariance, and derive multiplicative updates for this general form of the objective:

$$E = \frac{1}{2}\text{Tr}((\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H})^T \boldsymbol{C}^{-1}(\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H})) \qquad (9)$$

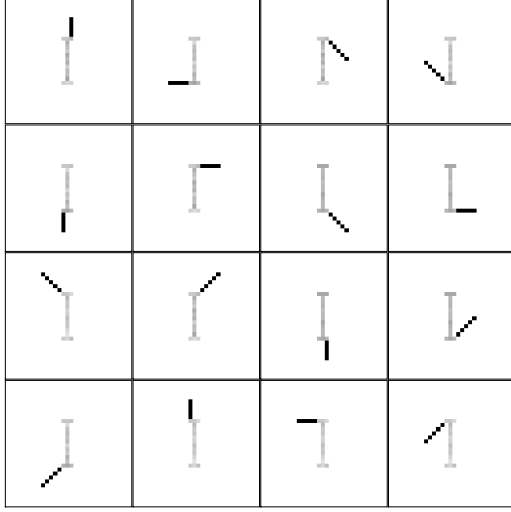### 4.1. Derivation of the updates

To simplify the notation, we define the precision matrix, $\boldsymbol{S} = \boldsymbol{C}^{-1}$. First, we rewrite the objective:

$$E = \frac{1}{2}(\text{Tr}(\boldsymbol{X}^T \boldsymbol{S} \boldsymbol{X}) + \text{Tr}(\boldsymbol{H}^T \boldsymbol{W}^T \boldsymbol{S} \boldsymbol{W} \boldsymbol{H}) - \qquad (10)$$
$$\text{Tr}(\boldsymbol{X}^T \boldsymbol{S} \boldsymbol{W} \boldsymbol{H}) - \text{Tr}(\boldsymbol{H}^T \boldsymbol{W}^T \boldsymbol{S} \boldsymbol{X})).$$

Then find the derivatives:

$$\frac{\partial E}{\partial \boldsymbol{W}} = \boldsymbol{S}(\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^T - \boldsymbol{X}\boldsymbol{H}^T) \qquad (11)$$

$$\frac{\partial E}{\partial \boldsymbol{H}} = \boldsymbol{W}^T \boldsymbol{S}(\boldsymbol{W}\boldsymbol{H} - \boldsymbol{X}), \qquad (12)$$

Fig. 6. Features extracted by NMF from the noiseless swimmer dataset. There only 16 unique features and only they are shown here.

Fig. 7. 20 features extracted from the noisy dataset by glsNMF.

where we have used the fact that S is symmetric. Finally, we obtain the following multiplicative updates for the GLS error function:

$$W = W \odot \frac{S^+ X H^T + S^- W H H^T}{S^- X H^T + S^+ W H H^T} \quad (13)$$

$$H = H \odot \frac{W^T S^+ X + W^T S^- W H}{W^T S^- X + W^T S^+ W H} \quad (14)$$

In these updates, the precision matrix is split into two parts as follows:

$$S_{ij}^+ = \begin{cases} S_{ij} & S_{ij} > 0, \\ S_{ij} + \lambda & i = j, \\ 0 & \text{otherwise,} \end{cases}$$

$$S_{ij}^- = \begin{cases} |S_{ij}| & S_{ij} < 0, \\ |S_{ij}| + \lambda & i = j, \\ 0 & \text{otherwise.} \end{cases}$$

In the matrix representation the split of S is expressed as:

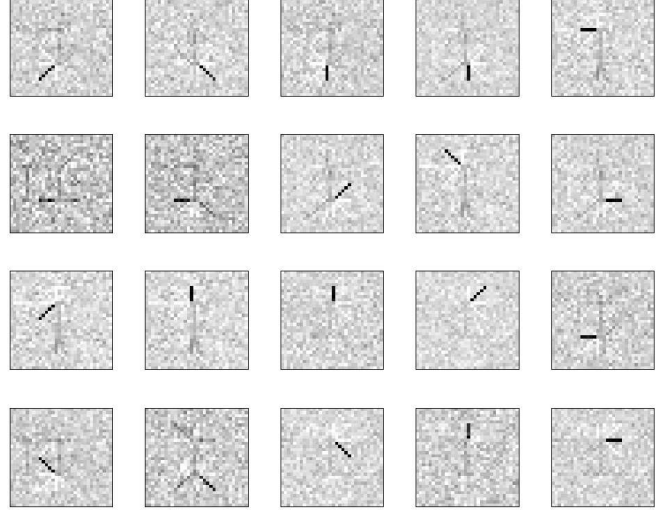$$S^+ = \hat{S}^+ + \lambda I$$
$$S^- = \hat{S}^- + \lambda I$$
$$S = \hat{S}^+ - \hat{S}^-$$

where λ is the minimal negative eigenvalue of $\hat{S}^-$ or 0 if $\hat{S}^-$ is positive semidefinite or empty. This ensures that S^- is positive semidefinite – a condition required for convergence properties of the updates. We defer the proof to the appendix.

4.2. Complexity

Introduction of S^+ and S^- added complexity to the updates, namely, four additional matrix multiplications and two matrix summations. Repeating parts of expressions in the numerator and the denominator of equations (14) and (13) can be precomputed before each respective updates. After that only multiplications by parts of the precision matrix and summation of the result are required.

4.3. Covariance estimation

In order for glsNMF to function properly, a good estimate of the noise covariance is required. This is sometimes possible to obtain as background measurement of the system without the task of interest. This is especially true in functional electromagnetic brain imaging (an area of increasing use of NMF [4]), where sampling rates allow collection of sufficient samples at least for spatial only analysis. Many models that use covariance matrix, like glsNMF, assume that it is computed elsewhere.

5. RESULTS

The glsNMF algorithm was applied to the previously introduces noisy swimmer dataset. As before, the number of features was set to 20.

The features obtained by glsNMF are shown in Figure 7. Compare with the features typically obtained by NMF 5. Note that we have run both algorithms many times changing the starting point. The starting points in the figures are the same.

NMF applied to the noisy dataset produces features spanning several swimmer configurations at once. Thus, it is hard to identify the original noise free image. In addition to that there is a spurious feature – the correlated part of the noise. It is not only present as a feature by itself but also contaminates many other features extracted by the algorithm.

Features extracted by glsNMF are sparse and completely recover the limbs. Furthermore, the correlated part of the noise is not present as a separate feature or as a part of any other features. Although some residual noise still remains even after convergence it does not destroy the parts based representation.

6. DISCUSSION

We have demonstrated a possibility of an unexpected failure mode of NMF in the case of data contaminated by correlated noise. This is almost the only situation when dealing with experimental datasets. Issue of noise has been previously addressed in NMF algorithms by rescaling each of the estimates by the amount of noise variance (uncertainty) [7], or by using Gaussian process priors to smooth out the estimates [8]. Results similar to [7] can probably be achieved using the advances in research on weighted NMF [9]. An approach that has formulation similar to our suggested solution was presented in [10]. However, there the goal was not to target correlated noise and also the novelty of our formulation is the multiplicative updates and their convergence proof. In fact, a solution by a projected gradient [11] method is easily possible and we have also derived it for our method but leave it out.

There are a multitude of extensions to NMF, such as additional sparsity, convolutive algorithm etc. [12, 13, 14]. We believe some of them can benefit from using GLS objective function.

7. CONCLUSIONS

A growing interest in application of NMF to experimental datasets [3, 4, 2, 5] requires special handling of issues introduces by unavoidable presence of noise. We have demonstrated a failure mode of NMF algorithm in which correlated noise can violate the separability assumption of unique factorization and hamper the results hurting interpretability of solutions. We also proposed a solution to correlated noise problem – glsNMF algorithm able to recover features from contaminated data. For this we have derived a multiplicative update and presented a convergence proof. Our future work includes application of the method to functional brain imaging and gene expression datasets as well as extending the method to deal with large dimensionality of the input space which makes the covariance matrix hard to handle. It is also possible to perform glsNMF with simultaneous estimation of the covariance matrix which we also leave for future work.

8. APPENDIX

Convergence proof

Consider the problem for a single column of $\boldsymbol{H}$ denoted by $\boldsymbol{h}$. The corresponding column of $\boldsymbol{X}$ is given by $\boldsymbol{x}$. The objective is now given by:

$$F(\boldsymbol{h}) = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{Wh})^T \boldsymbol{S}(\boldsymbol{x} - \boldsymbol{Wh}) \tag{15}$$

We define an auxiliary function $G(\boldsymbol{h}, \boldsymbol{h}^t)$ with the properties that $G(\boldsymbol{h}, \boldsymbol{h}) = F(\boldsymbol{h})$ and $G(\boldsymbol{h}, \boldsymbol{h}^t) \geq F(\boldsymbol{h})$. The multiplicative update rule is found at each iteration by minimizing the auxiliary function :

$$\boldsymbol{h}^{t+1} = \arg\min_{\boldsymbol{h}} G(\boldsymbol{h}, \boldsymbol{h}^t) \tag{16}$$

Note that this does not increase the objective function F, as we have

$$F(\boldsymbol{h}^{t+1}) \leq G(\boldsymbol{h}^{t+1}, \boldsymbol{h}^t) \leq G(\boldsymbol{h}^t, \boldsymbol{h}^t) = F(\boldsymbol{h}^t) \tag{17}$$

Define G as follows:

$$\begin{aligned} G(\boldsymbol{h}, \boldsymbol{h}^t) = \quad & F(\boldsymbol{h}^t) + (\boldsymbol{h} - \boldsymbol{h}^t)\nabla F(\boldsymbol{h}^t) \\ & + \frac{1}{2}(\boldsymbol{h} - \boldsymbol{h}^t)K(\boldsymbol{h}^t)(\boldsymbol{h} - \boldsymbol{h}^t) \end{aligned} \tag{18}$$

where the diagonal matrix $K(\boldsymbol{h}^t)$ is defined as

$$K_{ab}(\boldsymbol{h}^t) = \delta_{ab} \frac{(\boldsymbol{W}^T \boldsymbol{S}^+ \boldsymbol{W} \boldsymbol{h}^t + \boldsymbol{W}^T \boldsymbol{S}^- \boldsymbol{x})_a}{\boldsymbol{h}_a^t} \tag{19}$$

$G(\boldsymbol{h}, \boldsymbol{h}) = F(\boldsymbol{h})$ holds trivially. $G(\boldsymbol{h}, \boldsymbol{h}^t) \geq F(\boldsymbol{h})$ holds if

$$(\boldsymbol{h} - \boldsymbol{h}^t)^T [K(\boldsymbol{h}^t) - \boldsymbol{W}^T \boldsymbol{SW}](\boldsymbol{h} - \boldsymbol{h}^t) \geq 0 \tag{20}$$

Denote $\boldsymbol{M} = K(\boldsymbol{h}^t) - \boldsymbol{W}^T \boldsymbol{SW}$ and it split into parts as follows:

$$\boldsymbol{M} = \boldsymbol{P}_1 + \boldsymbol{P}_2 + \boldsymbol{P}_3 \tag{21}$$

$$(\boldsymbol{P}_1)_{ab} = \delta_{ab} \frac{(\boldsymbol{W}^T \boldsymbol{S}^+ \boldsymbol{W} \boldsymbol{h}^t)_a}{\boldsymbol{h}_a^t} - (\boldsymbol{W}^T \boldsymbol{S}^+ \boldsymbol{W})_{ab} \tag{22}$$

$$(\boldsymbol{P}_2)_{ab} = \delta_{ab} \frac{(\boldsymbol{W}^T \boldsymbol{S}^- \boldsymbol{x})_a}{\boldsymbol{h}_a^t} \tag{23}$$

$$(\boldsymbol{P}_3)_{ab} = (\boldsymbol{W}^T \boldsymbol{S}^- \boldsymbol{W})_{ab} \tag{24}$$

If each $\boldsymbol{P}_i$ is positive semidefinite then their sum $\boldsymbol{M}$ is also so. $\boldsymbol{P}_2$ is trivially positive semidefinite since it is a diagonal matrix with non-negative entries. $\boldsymbol{P}_3$ is also positive semidefinite since by construction in Section 4.1 we obtain a positive semidefinite $\boldsymbol{S}^- = \boldsymbol{LL}^T$, which gives $\boldsymbol{P}_3 = (\boldsymbol{W}^T \boldsymbol{L})(\boldsymbol{W}^T \boldsymbol{L})^T$ also positive semidefinite.

We show $\boldsymbol{P}_2$ to be positive semidefinite using the proof

structure of [1] which is as follows:

$$\boldsymbol{Q}_{ab}(\boldsymbol{h}^t) = \boldsymbol{h}_a^t(\boldsymbol{P}_2)_{ab}\boldsymbol{h}_b^t \tag{25}$$

$$\boldsymbol{\nu}^T\boldsymbol{Q}\boldsymbol{\nu} = \sum_{ab}\boldsymbol{\nu}_a\boldsymbol{Q}_{ab}\boldsymbol{\nu}_b \tag{26}$$

$$= \sum_{ab}(\boldsymbol{W}^T\boldsymbol{S}^+\boldsymbol{W})_{ab}\boldsymbol{h}_a^t\boldsymbol{h}_b^t[\frac{1}{2}\boldsymbol{\nu}_a^2 + \frac{1}{2}\boldsymbol{\nu}_b^2 - \boldsymbol{\nu}_a\boldsymbol{\nu}_b]$$

$$= \frac{1}{2}\sum_{ab}(\boldsymbol{W}^T\boldsymbol{S}^+\boldsymbol{W})_{ab}\boldsymbol{h}_a^t\boldsymbol{h}_b^t(\boldsymbol{\nu}_a - \boldsymbol{\nu}_b)^2$$

$$\geq 0$$

Setting the gradient of G to zero, we obtain an expression for the minimum of G.

$$\Delta = \frac{\boldsymbol{h}^t}{(\boldsymbol{W}^T\boldsymbol{S}^+\boldsymbol{W}\boldsymbol{h}^t + \boldsymbol{W}^T\boldsymbol{S}^-\boldsymbol{x})} \odot (\boldsymbol{W}^T\boldsymbol{S}(\boldsymbol{W}\boldsymbol{h}^t - \boldsymbol{x}))$$

$$\boldsymbol{h}^{t+1} = \boldsymbol{h}^t - \Delta$$

$$= \boldsymbol{h}^t \odot \frac{\boldsymbol{W}^T\boldsymbol{S}^+\boldsymbol{x} + \boldsymbol{W}^T\boldsymbol{S}^-\boldsymbol{W}\boldsymbol{h}^t}{\boldsymbol{W}^T\boldsymbol{S}^-\boldsymbol{x} + \boldsymbol{W}^T\boldsymbol{S}^+\boldsymbol{W}\boldsymbol{h}}$$

This is the update rule for $\boldsymbol{h}$ and similarly we can derive the update rule for $\boldsymbol{w}$.

9. REFERENCES

[1] Daniel D. Lee and Sebastian H. Seung, "Algorithms for non-negative matrix factorization," in *NIPS*, 2000, pp. 556–562.

[2] Gabriele Lohmann, Kirsten G Volz, and Markus Ullsperger, "Using non-negative matrix factorization for single-trial analysis of fMRI data.," *Neuroimage*, vol. 37, no. 4, pp. 1148–60, 2007.

[3] V. K. Potluru and V. D. Calhoun, "Group learning using contrast NMF : Application to functional and structural MRI of schizophrenia," *Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on*, pp. 1336–1339, May 2008.

[4] M. Mørup, L. K. Hansen, and S. M. Arnfred, "ER-PWAVELAB a toolbox for multi-channel analysis of time-frequency transformed event related potentials," *Journal of Neuroscience Methods*, vol. 161, pp. 361–368, 2007.

[5] Karthik Devarajan, "Nonnegative matrix factorization: An analytical and interpretive tool in computational biology," *PLoS Comput Biol*, vol. 4, no. 7, pp. –1000029, Jul 2008.

[6] David Donoho and Victoria Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?," in *Advances in Neural Information Processing Systems 16*, Sebastian Thrun, Lawrence Saul, and Bernhard Schlkopf, Eds. MIT Press, Cambridge, MA, 2004.

[7] G. Wang, A. V. Kossenkov, and M. F. Ochs, "LS-NMF: a modified non-negative matrix factorization algorithm utilizing uncertainty estimates.," *BMC Bioinformatics*, vol. 7, 2006.

[8] Mikkel N Schmidt and Hans Laurberg, "Nonnegative matrix factorization with Gaussian process priors.," *Computational intelligence and neuroscience*, p. 361705, 2008.

[9] D. Guillamet, J. Vitri, and B. Schiele, "Introducing a weighted non-negative matrix factorization for image classification," *Pattern Recognition Letters*, vol. 24, no. 14, pp. 2447–2454, 2003.

[10] . Andrzej Cichocki and . Rafal Zdunek, "Regularized alternating least squares algorithms for non-negative matrix/tensor factorization," in *ISNN 07: Proceedings of the 4th international symposium on Neural Networks*, Berlin, Heidelberg, 2007, pp. 793–802, Springer-Verlag.

[11] Chih-Jen Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural Comp.*, vol. 19, no. 10, pp. 2756–2779, October 2007.

[12] P. O. Hoyer, "Non-negative sparse coding," 2002, pp. 557–565.

[13] V. K. Potluru, S. M. Plis, and V. D. Calhoun, "Sparse shift-invariant nmf," *Image Analysis and Interpretation, 2008. SSIAI 2008. IEEE Southwest Symposium on*, pp. 69–72, mar 2008.

[14] Paul D. OGrady and Barak A. Pearlmutter, "Convolutive non-negative matrix factorisation with a sparseness constraint," in *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2006)*, Maynooth, Ireland, sep 2006, pp. 427–432.