Correlated Noise: How it Breaks NMF, and What to Do About it

Sergey M. Plis · Vamsi K. Potluru · Terran Lane · Vince D. Calhoun

Abstract Non-negative matrix factorization (NMF) is a problem of decomposing multivariate data into a set of features and their 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.

Keywords Nonnegative matrix factorization · Correlated noise · Separability

S. M. Plis (✉) · V. K. Potluru · T. Lane
Computer Science Department,
University of New Mexico, Albuquerque, NM 87131, USA
e-mail: pliz@cs.unm.edu

V. D. Calhoun
Electrical and Computer Engineering Department,
University of New Mexico, Albuquerque, NM 87131, USA

V. D. Calhoun
Mind Research Network, Albuquerque, NM 87106, USA

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 [2–4], gene expression [5], and other empirical datasets.

Lee and Seung [1] 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. Input data X is encoded in the feature or source matrix W with a column per feature, and corresponding feature activations encoded in H. 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 a 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. The presence of correlated noise may, however, violate the conditions and render otherwise separable dataset not factorable by NMF. We show 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 techniques 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 X, represent it with a product of two non-negative matrices W, H of sizes $m \times r$ and $r \times n$ respectively:

$$X \approx WH. \tag{2}$$

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 W and 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:

$$E = \frac{1}{2} \|X - WH\|_F \tag{3}$$

and the corresponding updates are:

$$W = W \odot \frac{XH^T}{WHH^T} \tag{4}$$

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

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 W or H but not in both [1]. As the algorithm iterates using the updates 4 and 5, W and H converge to a local minimum of the cost function.

The slightly mysterious form for the above updates can be understood as follows. Simple additive updates for H and W are given by:

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

$$W = W + \eta_W \odot (XH^T - WHH^T) \tag{7}$$

If the learning rate given by the matrix elements of $\eta_{\{W,H\}}$ are all set to some small positive number then this is the conventional gradient descent. However, setting the learning rate matrices as follows:

$$\eta_H = \frac{H}{W^T WH}, \tag{8}$$

$$\eta_W = \frac{W}{WHH^T}, \tag{9}$$

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 (encoded in the feature or source matrix W with a column per 32×32
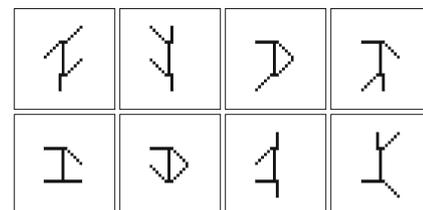

Figure 1 Randomly selected samples from the swimmer dataset of [6], which consists of 256 total images with all possible combination of limbs positions.
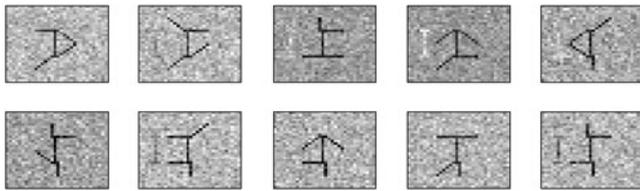
Figure 2 Random samples from the swimmer dataset contaminated by correlated noise. Note the salient spatial correlation with the shape of the swimmer's torso to the left of the swimmer.

Figure 4 Correlated noise samples that were added to the swimmer dataset. Note the salient spatial correlation.

limb, and corresponding feature activations encoded in **H**), as shown in Fig. 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 Fig. 1 contaminated by the correlated noise.

The LS objective function of Eq. 3 can be derived from the Gaussian likelihood with noise covariance of the form $\sigma^2 I$. assuming covariances only between pixels within a data picture. Note that correlated noise results in a differently structured covariance matrix. The covariance of the correlated noise is shown in Fig. 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 the figure. Several samples of the noise are shown in Fig. 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
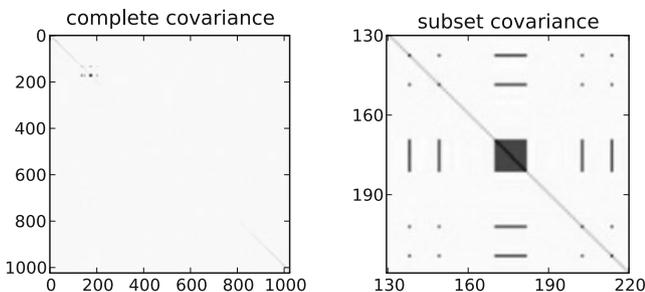
torso as a separate feature with the other features correctly recovered. This common-sense behavior would go along with other matrix factorization techniques such as independent component analysis (ICA), which exploit this feature for denoising. Surprisingly, we have observed quite different behavior. A typical result is shown in Fig. 5, where it becomes clear that correlated noise affects the estimation of all of the features, in addition to being estimated as a separate feature. For comparison the correct factorization that is obtained by NMF from the noiseless dataset is shown in Fig. 6. Note that we have observed similar behavior even when using the KL 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

Figure 3 Noise covariance for all 1,024 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.
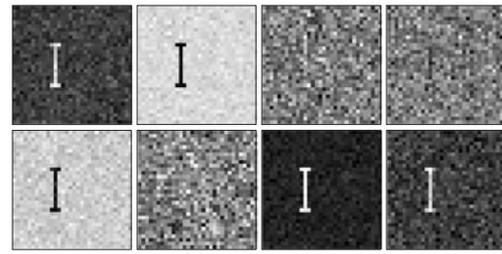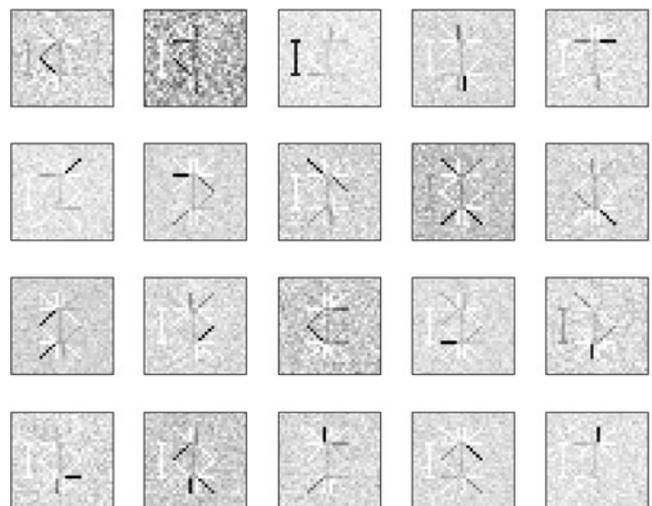

Figure 5 20 features (columns of the **W** matrix) as extracted by the regular NMF algorithm with the LS objective function from the dataset contaminated by correlated noise.
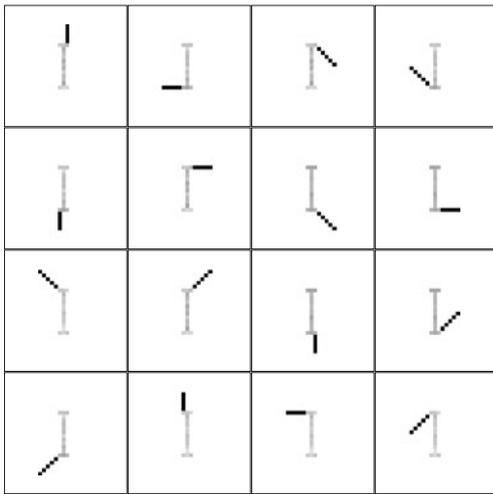
Figure 6 Features extracted by NMF from the noiseless swimmer dataset. Only 16 unique features are shown.

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 Eq. 3. After rewriting Eq. 3 in the matrix form:

$$E = \frac{1}{2}\text{Tr}((X - WH)^T(X - WH)), \tag{10}$$

Minimizing the above objective maximizes the likelihood which is a product of independent Gaussian likelihoods for each column of X, that share a common feature set W and the covariance matrix C (identity in Eq. 10). The following makes it more explicit

$$E = \frac{1}{2}\sum (x_i - Wh_i)^T (x_i - Wh_i). \tag{11}$$

For optimization purposes, the formulation is also valid for noise with covariance structure $\sigma^2 I$. 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 Eq. 12, where C is the noise covariance,[1] and derive multiplicative updates for this general form of the objective:

$$E = \frac{1}{2}\text{Tr}((X - WH)^T C^{-1}(X - WH)) \tag{12}$$

Minimizing this objective is equivalent to maximizing the Gaussian likelihood with full covariance matrix, i.e. correlated noise. The problems that we further solve arise within the NMF context since the covariance matrix is not guaranteed to be nonnegative.

4.1 Derivation of the Updates

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

$$E = \frac{1}{2}(\text{Tr}(X^T SX) + \text{Tr}(H^T W^T SWH) \\ - \text{Tr}(X^T SWH) - \text{Tr}(H^T W^T SX)). \tag{13}$$

Then find the derivatives:

$$\frac{\partial E}{\partial W} = S(WHH^T - XH^T) \tag{14}$$

$$\frac{\partial E}{\partial H} = W^T S(WH - X), \tag{15}$$

where we have used the fact that S is symmetric. Next, we follow the procedure similar to the one used to obtain the updates for the original NMF algorithm [1]. We set the values of the learning rate for both matrices W and H such that the updates become multiplicative. A detailed convergence proof that lead to these values is presented in Appendix. Finally, we arrive at

---

[1]Note, that we use the vector variate definition of the sample noise covariance which results in a matrix using vectors (columns of X) as samples, as opposed to the matrix variate tensor covariance matrix that would have required matrix observations.

the following multiplicative updates for the GLS error function:

$$W = W \odot \frac{S^+ X H^T + S^- W H H^T}{S^- X H^T + S^+ W H H^T} \quad (16)$$

$$H = H \odot \frac{W^T S^+ X + W^T S^- W H}{W^T S^- X + W^T S^+ W H} \quad (17)$$

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. The Appendix also provides details of implementing a projected gradient algorithm for glsNMF.

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 Eqs. 17 and 16 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 a sample covariance of the background measurements of the system without the task of interest. When such background noise is available, it is possible to estimate the covariance using the sample covariance estimation [8]. This is the method we use in this paper. 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 the 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 Fig. 7. Compare with the features typically obtained by NMF in Fig. 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 some other features extracted by the algorithm.

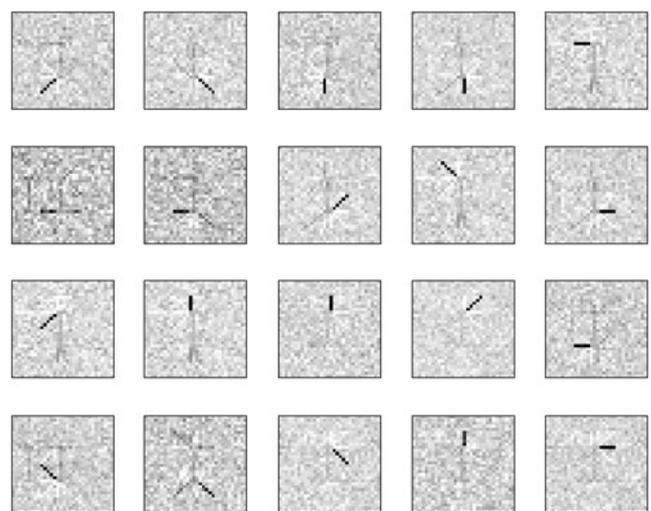Features extracted by glsNMF are sparse and completely recover the limbs. Furthermore, the correlated

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

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 shown that NMF can fail in the case of data which contains correlated noise. This is a common situation when dealing with experimental datasets. The issue of noise has been previously addressed in previously proposed 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 [9]. Results similar to [7] can probably be achieved using the advances in research on weighted NMF [10]. An approach that has formulation similar to our suggested solution was presented in [11]. 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. We show that a solution by a projected gradient [12] method is easily possible and we have also derived it for our method including the derivation in Appendix.

There are a multitude of extensions to NMF like sparse NMF [13, 14] and convolutive NMF [13, 15, 16]. We believe some of them can benefit from using the GLS objective function.

7 Conclusions

A growing interest in application of NMF to experimental datasets [2–5] requires special handling of issues introduces by unavoidable presence of noise. We have demonstrated that the NMF algorithm can fail in the presence of correlated noise which can violate the separability assumption of unique factorization and degrade the results in applications such as feature extraction. We also proposed the glsNMF algorithm as a solution to the correlated noise problem which is able to recover features from data with correlated noise. 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.

Appendix

Convergence Proof

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

$$F(h) = \frac{1}{2}(x - Wh)^T S(x - Wh) \tag{18}$$

We define an auxiliary function $G(h, v)$ with the properties that $G(h, h) = F(h)$ and $G(h, h^t) \geq F(h)$, where h^k is the current estimate with iteration index k and h is the free parameter. The multiplicative update rule is found at each iteration by minimizing the auxiliary function:

$$h^{k+1} = \arg \min_h G(h, h^k) \tag{19}$$

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

$$F(h^{k+1}) \leq G(h^{k+1}, h^k) \leq G(h^k, h^k) = F(h^k) \tag{20}$$

Define G as follows:

$$G(h, h^k) = F(h^k) + (h - h^k)^T \nabla F(h^k)$$
$$+ \frac{1}{2}(h - h^k)^T K(h^k)(h - h^k) \tag{21}$$

where the diagonal matrix $K(h^k)$ is defined as

$$K_{ab}(h^k) = \delta_{ab} \frac{(W^T S^+ W h^k + W^T S^- x)_a}{h_a^k}, \tag{22}$$

where a and b are indices of vector and matrix elements, and

$$\delta_{ab} = \begin{cases} 1, & \text{if } a = b \\ 0, & \text{if } a \neq b \end{cases} \tag{23}$$

It trivially holds that $G(h, h) = F(h)$. For $G(h, h^k) \geq F(h)$ to hold we need

$$(h - h^k)^T[K(h^k) - W^T SW](h - h^k) \geq 0 \tag{24}$$

We denote $M = K(h^k) - W^T SW$, observing that M must be positive semidefinite for (24) to hold, and split it into parts as follows:

$$M = P_1 + P_2 + P_3 \tag{25}$$

$$(P_1)_{ab} = \delta_{ab} \frac{(W^T S^+ W h^k)_a}{h_a^k} - (W^T S^+ W)_{ab} \tag{26}$$

$$(P_2)_{ab} = \delta_{ab} \frac{(W^T S^- x)_a}{h_a^k} \tag{27}$$

$$(P_3)_{ab} = (W^T S^- W)_{ab} \tag{28}$$

If each $\boldsymbol{P}_i$ is positive semidefinite then their sum $\boldsymbol{M}$ is also so. The most difficult is $\boldsymbol{P}_1$ and we show it is positive semidefinite the last. $\boldsymbol{P}_2$ 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 semi-definite $\boldsymbol{S}^-$ which can be written as a square root $\boldsymbol{L}\boldsymbol{L}^T$ giving a positive semidefinite $\boldsymbol{P}_3 = (\boldsymbol{W}^T\boldsymbol{L})(\boldsymbol{W}^T\boldsymbol{L})^T$.

We show $\boldsymbol{P}_1$ to be positive semidefinite using the proof structure of [1] which is as follows:

$$\boldsymbol{Q}_{ab}(\boldsymbol{h}^k) = h_a^k(\boldsymbol{P}_1)_{ab}\, h_b^k \tag{29}$$

$$\boldsymbol{v}^T\boldsymbol{Q}\boldsymbol{v} = \sum_{ab} v_a \boldsymbol{Q}_{ab}\, v_b \tag{30}$$

$$= \sum_{ab}(\boldsymbol{W}^T\boldsymbol{S}^+\boldsymbol{W})_{ab}\, h_a^k h_b^k \left[\frac{1}{2}v_a^2 + \frac{1}{2}v_b^2 - v_a v_b\right]$$

$$= \frac{1}{2}\sum_{ab}(\boldsymbol{W}^T\boldsymbol{S}^+\boldsymbol{W})_{ab}\, h_a^k h_b^k (v_a - v_b)^2$$

$$\geq 0$$

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

$$\Delta = \frac{\boldsymbol{h}^k}{(\boldsymbol{W}^T\boldsymbol{S}^+\boldsymbol{W}\boldsymbol{h}^k + \boldsymbol{W}^T\boldsymbol{S}^-\boldsymbol{x})} \odot (\boldsymbol{W}^T\boldsymbol{S}(\boldsymbol{W}\boldsymbol{h}^k - \boldsymbol{x})) \tag{31}$$

$$\boldsymbol{h}^{k+1} = \boldsymbol{h}^k - \Delta \tag{32}$$

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

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

Projected Gradient

Multiplicative updates suffer the phenomena of exponential slowdown of the convergence rate in the vicinity of the solution, when the initial point was in the positive quadrant. Fortunately, projected gradient solutions do not suffer this problem and have nicer convergence properties, although they are considerably more difficult to implement.

To implement a projected gradient solver for glsNMF we require a modification of the original NMF algorithm [12] that accounts for including the covariance into the Gaussian likelihood objective. Individual objectives for $\boldsymbol{W}$ and $\boldsymbol{H}$ written in the vector form $vec(\cdot)$ (columns of a matrix stacked in a vector) are:

$$E_{\boldsymbol{W}} = vec(\boldsymbol{W})^T\left(\boldsymbol{H}\boldsymbol{H}^T \otimes \boldsymbol{S}\right)vec(\boldsymbol{W})$$
$$+ \boldsymbol{W}\text{'s linear terms} \tag{33}$$

$$E_{\boldsymbol{H}} = vec(\boldsymbol{H})^T\left(\boldsymbol{I} \otimes \boldsymbol{W}^T\boldsymbol{S}\boldsymbol{W}\right)vec(\boldsymbol{H})$$
$$+ \boldsymbol{H}\text{'s linear terms}, \tag{34}$$

where $\otimes$ is the Kronecker product defined as

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix} \tag{35}$$

Hessian matrices in both cases tend to be well conditioned when number of training examples is larger than number of features and dimensionality of the data, and $\boldsymbol{S}$ is well conditioned too. The first is a usual case in the application domain of NMF. The second requires a well conditioned sample covariance, which can be provided by estimation from a sufficient population of noise or adequately modeled.

Equation 34 is structurally equivalent to the one presented by Lin [12]. The original routine for solving the subproblem of taking a step in the direction opposite to the maximal gradient $\nabla_{\boldsymbol{H}} f(\boldsymbol{W}, \boldsymbol{H})$ can be used with a single change: $\boldsymbol{W}^T\boldsymbol{S}\boldsymbol{W}$ needs to be precomputed instead of the plain $\boldsymbol{W}^T\boldsymbol{W}$.

A different situation arises in the case of Eq. 33. The subroutine of $\boldsymbol{H}$ cannot be used directly and a straightforward solution is to have individual subroutines for $\boldsymbol{H}$ and $\boldsymbol{W}$. The sufficient decrease test [12, equation 4.3] becomes:

$$(1 - \sigma)\langle\nabla_{\boldsymbol{W}} f(\boldsymbol{W}, \boldsymbol{H}), d\boldsymbol{W}\rangle + \frac{1}{2}\langle d\boldsymbol{W}, \boldsymbol{S}d\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^T\rangle \leq 0,$$

where $\langle\cdot, \cdot\rangle$ is the sum of element-wise product of two matrices.

The shape of the second term in the sum is derived from the full tensor product of the Hessian with $vec(d\boldsymbol{W})$ taking advantage of the following identity [17]:

$$vec(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}) = \left(\boldsymbol{B}^T \otimes \boldsymbol{A}\right)vec(\boldsymbol{X}) \tag{36}$$

The additional complexity with respect to the original projected gradient algorithm for NMF is the matrix multiplication by $\boldsymbol{S}$. In practice we have found this not be an issue and projected gradient was considerably

faster than multiplicative updates, which aligns well with what has been previously observed for NMF.

References

1. Daniel Lee, D., & Sebastian Seung, H. (2000). Algorithms for non-negative matrix factorization. In *NIPS* (pp. 556–562).
2. Lohmann, G., Volz, K. G., & Ullsperger, M. (2007). Using non-negative matrix factorization for single-trial analysis of fMRI data. *Neuroimage, 37*(4), 1148–1160.
3. Potluru, V. K., & Calhoun, V. D. (2008). Group learning using contrast NMF: Application to functional and structural MRI of schizophrenia. In *IEEE International Symposium on Circuits and Systems, 2008 (ISCAS 2008)* (pp. 1336–1339).
4. Mørup, M., Hansen, L. K., & Arnfred, S. M. (2007). ERPWAVELAB a toolbox for multi-channel analysis of time-frequency transformed event related potentials. *Journal of Neuroscience Method, 161*, 361–368.
5. Devarajan, K. (2008). Nonnegative matrix factorization: An analytical and interpretive tool in computational biology. *PLoS Computational Biology, 4*(7), e1000029.
6. Donoho, D., & Stodden, V. (2004). When does non-negative matrix factorization give a correct decomposition into parts? In S. Thrun, L. Saul, & B. Schölkopf (Eds.), *Advances in neural information processing systems 16*. Cambridge: MIT Press.
7. Wang, G., Kossenkov, A. V., & Ochs, M. F. (2006). LS-NMF: A modified non-negative matrix factorization algorithm utilizing uncertainty estimates. *BMC Bioinformatics, 7*, 175.
8. Stark, H., & Woods, J. W. (2002). *Probability and random processes with applications to signal processing*. Prentice Hall Upper Saddle River.
9. Schmidt, M. N., & Laurberg, H. (2008). Nonnegative matrix factorization with Gaussian process priors. *Computational Intelligence and Neuroscience, 2008*, 361705.
10. Guillamet, D., Vitri, J., & Schiele, B. (2003). Introducing a weighted non-negative matrix factorization for image classification. *Pattern Recognition Letters, 24*(14), 2447–2454.
11. Cichocki, A., & Zdunek, R. (2007). Regularized alternating least squares algorithms for non-negative matrix/tensor factorization. In *ISNN 07: Proceedings of the 4th international symposium on neural networks* (pp. 793–802). Berlin, Heidelberg: Springer.
12. Lin, C.-J. (2007). Projected gradient methods for nonnegative matrix factorization. *Neural Computation, 19*(10), 2756–2779.
13. O'Grady, P. D., & Pearlmutter, B. A. (2006). Convolutive non-negative matrix factorisation with a sparseness constraint. In *Proceedings of the IEEE international workshop on Machine Learning for Signal Processing (MLSP 2006)* (pp. 427–432). Maynooth, Ireland.
14. Hoyer, P. O. (2002). Non-negative sparse coding. In *Proceedings of the 2002 12th IEEE workshop on Neural networks for signal processing, 2002* (pp. 557–565).
15. Potluru, V. K., Plis, S. M., & Calhoun, V. D. (2008). Sparse shift-invariant NMF. In *IEEE Southwest Symposium on Image analysis and interpretation, 2008 (SSIAI 2008)* (pp. 69–72).
16. Smaragdis, P. (2004). Non-negative matrix factor deconvolution; extraction of multiple sound sources from monophonic inputs. In *International Congress on Independent Component Analysis and Blind Signal Separation*.
17. Petersen, K. B., & Pedersen, M. S. (2008). The matrix cookbook. Version 20081110.

Sergey M. Plis received a CAD/CAM engineering degree from the Kovrov State Tech Academy, Kovrov, Russia, in 2001, master's degree in computer science from the University of New Mexico, Albuquerque, in 2004, and the Ph.D. degree in computer science from the University of New Mexico in 2007. During his graduate school he worked as a graduate research assistant at the Los Alamos National Lab and then as a postdoctoral fellow at the University of New Mexico computer science department. Dr. Plis is currently a research scientist at the Mind Research network interested in applications of machine learning to brain imaging.

Vamsi K. Potluru received a bachelor's degree in Electrical Engineering from the Indian Institute of Technology Bombay, Mumbai in 2001 and master's degree in Computer Science from University of New Mexico, Albuquerque, in 2008. Currently, he is a PhD student in the University of New Mexico with interests in machine learning (ML) and ML approach to how the brain works.

Terran Lane is an associate professor at the University of New Mexico department of Computer Science. Dr. Lane received his Ph.D. in 2000 from Purdue University, where he studied machine learning algorithms for computer security applications. Subsequently, he spent two years at the MIT AI Lab (now CSAIL), where he worked on planning and learning techniques for stochastic control domains representable as Markov decision

processes. His current interests include machine learning and artificial intelligence in general, with applications to problems in scientific data analysis and knowledge discovery. He is particularly interested in network modeling and identification and time series analysis for domains such as neuroscience, bioinformatics, sensor networks, and cognitive modeling.

For more information, including contact info and publications, please refer to his web page at http://www.cs.unm.edu/~terran (no photo submitted).

Vince D. Calhoun received a bachelor's degree in Electrical Engineering from the University of Kansas, Lawrence, Kansas, in 1991, master's degrees in Biomedical Engineering and Information Systems from Johns Hopkins University, Baltimore, in 1993 and 1996, respectively, and the Ph.D. degree in electrical engineering from the University of Maryland Baltimore County, Baltimore, in 2002. He worked as a senior research engineer at the psychiatric neuroimaging laboratory at Johns Hopkins from 1993 until 2002. He then served as the director of medical image analysis at the Olin Neuropsychiatry Research Center and as an associate professor at Yale University.

Dr. Calhoun is currently Chief Technology Officer and Director of Image Analysis and MR Research at the Mind Research Network and is a Professor in the Departments of Electrical and Computer Engineering (primary), Neurosciences, Psychiatry and Computer Science at the University of New Mexico. He is the author of more than 150 full journal articles, over 250 technical reports, abstracts and conference proceedings. Much of his career has been spent on the development of data driven approaches for the analysis of brain imaging data. He has won over $18 million in NSF and NIH grants on the incorporation of prior information into independent component analysis (ICA) for functional magnetic resonance imaging, data fusion of multimodal imaging and genetics data, and the identification of biomarkers for disease.

Dr. Calhoun is a senior member of the IEEE, the Organization for Human Brain Mapping, the International Society for Magnetic Resonance in Medicine, and the American College of Neuropsychopharmacology. He is a chartered grant reviewer for NIH. He has organized workshops and special sessions at multiple conferences. He is currently serving on the IEEE Machine Learning for Signal Processing (MLSP) technical committee and previous served as the general chair of the 2005 meeting. He is a reviewer for many journals and is on the editorial board of the Human Brain Mapping and Neuroimage journals.