BLIND BOUNDED SOURCE SEPARATION USING NEURAL NETWORKS WITH LOCAL LEARNING RULES

Alper T. Erdogan

Electrical-Electronics Engineering
Koc University
Istanbul, Turkey

*Cengiz Pehlevan**

School of Engineering and Applied Sciences
Harvard University
Cambridge, MA USA

ABSTRACT

An important problem encountered by both natural and engineered signal processing systems is blind source separation. In many instances of the problem, the sources are bounded by their nature and known to be so, even though the particular bound may not be known. To separate such bounded sources from their mixtures, we propose a new optimization problem, Bounded Similarity Matching (BSM). A principled derivation of an adaptive BSM algorithm leads to a recurrent neural network with a clipping nonlinearity. The network adapts by local learning rules, satisfying an important constraint for both biological plausibility and implementability in neuromorphic hardware.

Index Terms— Similarity Matching, Recurrent Neural Networks, Local Update Rule, Blind Source Separation, Bounded Component Analysis.

1. INTRODUCTION

Blind source separation is a fundamental problem for natural and engineered signal processing systems. In this paper, we show how it can be solved by a class of neural networks important for both neuroscience and machine learning, i.e. those with local learning rules, where the strength of a synapse is updated based on the activations of only the pre- and postsynaptic neurons. Locality of learning rules is a natural constraint for biological neural networks [1, 2], and enables large scale implementations of neural network algorithms with low power in recent neuromorphic integrated circuits [3].

Similarity Matching has been introduced as a gradient-based optimization framework for principled derivation of neural networks with local learning rules [4, 5]. This framework can be used to provide solutions for clustering [6, 7], sparse feature extraction [6] and manifold learning [8]. It was also applied to a nonnegative blind source separation problem [9] which leads to a recurrent neural network with ReLU activation functions.

Similar to nonnegativity, spatial boundedness is a property that can be exploited to separate sources from their linear mixtures. Bounded Component Analysis (BCA) has been introduced as a framework exploiting this property to separate both dependent and independent sources from their linear mixtures [10, 11]. It was successfully applied to separation of dependent natural images, digital communication signals, as well as sparse bounded sources [11, 12, 13].

This article has two main contributions. First, we formulate a new optimization problem, bounded similarity matching (BSM), for blind bounded source separation (BBSS). By using diagonally weighted inner products and bounded outputs, we show that the BBSS problem can be formulated as a minimization of the inner product weights under a weighted and bounded similarity matching constraint. Second, by an online optimization of this problem, we derive a biologically plausible recurrent neural network with clipping activation functions, whose parameters are updated by a local learning rule. The update rules of synaptic weights parallel the plasticity mechanisms observed in biological neurons.

2. BBSS SETTING

We consider the following BBSS scenario:

- There are d sources, represented by the vector sequence $\{\mathbf{s}_t \in \mathbb{R}^d, t \in \mathbb{Z}^+\}$. Sources are uncorrelated,

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathbf{s}} &= E((\mathbf{s}_t - \boldsymbol{\mu}_{\mathbf{s}})(\mathbf{s}_t - \boldsymbol{\mu}_{\mathbf{s}})^\top) \\ &= \mathrm{diag}(\sigma^2_{s^{(1)}}, \sigma^2_{s^{(2)}}, \ldots, \sigma^2_{s^{(d)}}), \end{aligned} \quad (1)$$

where $\boldsymbol{\mu}_{\mathbf{s}} = E(\mathbf{s})$, and bounded around their mean

$$\boldsymbol{\mu}_{\mathbf{s}} - \mathbf{r}/2 \le \mathbf{s}_t \le \boldsymbol{\mu}_{\mathbf{s}} + \mathbf{r}/2, \qquad \forall t \in \mathbb{Z}^+, \quad (2)$$

where $\mathbf{r} = \left[r^{(1)}, r^{(2)}, \ldots, r^{(d)}\right]^\top$, $r^{(i)}$ the range of source $s^{(i)}$. The bounds are unknown to the BBSS algorithm.

- Sources are mixed through a linear memoryless system, $\mathbf{m}_t = \mathbf{A}\mathbf{s}_t$, $\forall t \in \mathbb{Z}^+$, where $\mathbf{A} \in \mathbb{R}^{q \times d}$ is the full rank mixing matrix, and $\mathbf{m}_t \in \mathbb{R}^q$ are mixtures. We assume an (over)determined mixing, i.e., $q \ge d$. We define $\boldsymbol{\mu}_{\mathbf{m}} = E(\mathbf{m}_t)$.

- The mixtures are pre-processed with whitening and mean-removal, which can be implemented through various adaptive mechanisms including biologically plausible neural networks, see e.g. [4]. The pre-processed mixtures can be written as

$$\mathbf{x}_t = \mathbf{W}_{pre}(\mathbf{m}_t - \boldsymbol{\mu_m}) = \boldsymbol{\Theta}\boldsymbol{\Sigma}_{\mathbf{s}}^{-1/2}(\mathbf{s}_t - \boldsymbol{\mu_s}) = \boldsymbol{\Theta}\bar{\mathbf{s}}_t, \quad (3)$$

where $\bar{\mathbf{s}}_t$ is the scaled and mean-removed sources that satisfy $E(\bar{\mathbf{s}}) = \mathbf{0}$, and $E(\bar{\mathbf{s}}\bar{\mathbf{s}}^T) = \mathbf{I}$. Furthermore, $\bar{\mathbf{s}}_t$ is symmetrically bounded around zero satisfying $-\mathbf{b} \leq \bar{\mathbf{s}}_t \leq \mathbf{b}$, where $\mathbf{b} = \boldsymbol{\Sigma}_{\mathbf{s}}^{-1/2}\frac{\mathbf{r}}{2}$. Here, note that since $\sigma_{s^{(i)}} \leq \frac{r^{(i)}}{2}$, bounds $b^{(i)}$'s are greater than or equal to 1. $\boldsymbol{\Theta}$ is a real orthogonal matrix satisfying $\boldsymbol{\Theta}^T\boldsymbol{\Theta} = \mathbf{I}$ and $\mathbf{x}_t$, like $\bar{\mathbf{s}}_t$, is a white signal with unit variance.

3. BSM: A NEW OPTIMIZATION FORMULATION OF BBSS

Our solution to BBSS takes inspiration from the Nonnegative Similarity Matching (NSM) method given in [9], which proposed the following optimization problem to recover nonnegative sources from their whitened mixtures:

$$\mathbf{Y}^* = \underset{\mathbf{Y} \geq \mathbf{0}}{\arg\min} \|\mathbf{X}^\top\mathbf{X} - \mathbf{Y}^\top\mathbf{Y}\|_F^2, \quad (4)$$

where $\mathbf{X} := [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{t-1}, \mathbf{x}_t]$ is the data matrix and $\mathbf{Y} := [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{t-1}, \mathbf{y}_t]$ is the output matrix. NSM amounts to finding nonnegative $\mathbf{y}_n$ vectors which have the same decorrelated and variance-equalized structure as whitened-mixtures.

If we were to adopt a similar approach for bounded but potentially mixed-sign sources, we could instead propose

$$\mathbf{Y}^* = \underset{-\mathbf{b} \leq \mathbf{Y} \leq \mathbf{b}}{\arg\min} \|\mathbf{X}^\top\mathbf{X} - \mathbf{Y}^\top\mathbf{Y}\|_F^2. \quad (5)$$

However, this form requires knowledge of the source bounds, contrary to our setting. Instead, we aim to extract whitened sources up to scale factors, which are defined as $\tilde{\mathbf{s}}_t = \boldsymbol{\Pi}_{\mathbf{b}}^{-1}\bar{\mathbf{s}}_t$, where $\boldsymbol{\Pi} = \text{diag}(b^{(1)}, b^{(2)}, \ldots, b^{(d)})$, and the resulting scaled sources satisfy $-\mathbf{1} \leq \tilde{\mathbf{s}} \leq \mathbf{1}$. Therefore, we restrict outputs to a fixed range, such as the unity ℓ_∞-norm ball defined as $\mathcal{B}_\infty = \{\mathbf{y} : \|\mathbf{y}\|_\infty \leq 1\}$.

In light of these arguments, we pose the following optimization problem, Bounded Similarity Matching (BSM), for bounded source separation:

$$\underset{\mathbf{Y}, D_{11}, \ldots D_{dd}}{\text{minimize}} \quad \sum_{i=1}^{d} D_{ii}^2 \qquad \qquad \textit{BSM}$$

$$\text{subject to} \quad \mathbf{X}^\top\mathbf{X} - \mathbf{Y}^\top\mathbf{D}\mathbf{Y} = 0$$
$$-\mathbf{1} \leq \mathbf{Y} \leq \mathbf{1}$$
$$\mathbf{D} = \text{diag}(D_{11}, \ldots, D_{dd})$$
$$D_{11}, D_{22}, \ldots D_{dd} > 0$$

Note that the similarity metric of BSM is different than that of NSM. In order to confine outputs to a fixed dynamic range, we propose to match their weighted inner products to the inner products of inputs. This enables imposing boundedness of sources without knowing exact bounds. Estimation of individual bounds is achieved by optimizing over the inner product weights $\mathbf{D}$.

The following theorem asserts that the global minima of BSM are perfect separators:

Theorem 1. *Given the BBSS setup described in Section 2, if the vector sequence $\{\tilde{s}_t\}$ contains all corner points of $\mathcal{B}_\infty$, then the global optima for BSM satisfy*

$$\mathbf{Y} = \boldsymbol{\Lambda}\mathbf{P}\mathbf{D}^{-1/2}\bar{\mathbf{S}} = \boldsymbol{\Lambda}\mathbf{P}\tilde{\mathbf{S}}, \quad (6)$$
$$D_{ii} = b_{J_i}^2, \qquad i = 1, \ldots d, \quad (7)$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix with ± 1's on the diagonal (representing sign ambiguity), $\mathbf{P}$ is a permutation matrix, J is a permutation of $\{1, 2, \ldots, d\}$.

Proof. The proof is in Appendix A. $\square$

4. ADAPTIVE BSM VIA A NEURAL NETWORK WITH LOCAL LEARNING RULES

In this section, we derive an adaptive implementation of BSM and show that it corresponds to a biologically plausible neural network with local update rules. For this purpose, our first task is to introduce an online version of the BSM optimization problem introduced in the previous section.

In the online setting, we consider exponential weighting of the signals for dynamical adjustment and define the weighted input data snapshot matrix by time t as,

$$\boldsymbol{\mathcal{X}}_t = \begin{bmatrix} \gamma^{t-1}\mathbf{x}_1 & \gamma^{t-2}\mathbf{x}_2 & \ldots & \gamma\mathbf{x}_{t-1} & \mathbf{x}_t \end{bmatrix}$$
$$= \mathbf{X}_t\boldsymbol{\Gamma}_t, \quad (8)$$

where $\boldsymbol{\Gamma}_t = \text{diag}(\gamma^{t-1}, \gamma^{t-2}, \ldots, 1)$. Similarly, we define the weighted output snapshot matrix by time t as,

$$\boldsymbol{\mathcal{Y}}_t = \begin{bmatrix} \gamma^{t-1}\mathbf{y}_1 & \gamma^{t-2}\mathbf{y}_2 & \ldots & \gamma\mathbf{y}_{t-1} & \mathbf{y}_t \end{bmatrix}$$
$$= \mathbf{Y}_t\boldsymbol{\Gamma}_t \quad (9)$$

We define $\kappa_t = \sum_{k=0}^{t-1} \gamma^{2k} = \frac{1-\gamma^{2t}}{1-\gamma^2}$ as a measure of effective time window length for sample correlation calculations, given the exponential weights. Assuming sufficiently large t, we take $\kappa = \frac{1}{1-\gamma^2}$.

In order to define on online BSM cost function, we replace the hard equality constraint on BSM with a square error minimization, and introduce the weighted similarity matching (WSM) cost function as,

$$J_{WSM}(\boldsymbol{\mathcal{Y}}_t, \mathbf{D}_t) = \frac{1}{\kappa^2}\|\boldsymbol{\mathcal{X}}_t^T\boldsymbol{\mathcal{X}}_t - \boldsymbol{\mathcal{Y}}_t^T\mathbf{D}_t\boldsymbol{\mathcal{Y}}_t\|_F^2. \quad (10)$$

3813

With this definition, we consider a relaxation of *BSM* (rBSM):

$$J_{rBSM}(\boldsymbol{\mathcal{Y}}_t, \mathbf{D}_t) = J_{WSM}(\boldsymbol{\mathcal{Y}}_t, \mathbf{D}_t) + 2\alpha_D \sum_{i=1}^{d} D_{t,ii}^2, \quad (11)$$

to be minimized under the constraint set $-1 \leq \boldsymbol{\mathcal{Y}}_t \leq 1$.

Following a treatment similar to [9], we derive an online cost from J_{rBSM} by considering its optimization with respect to the data already received and only with respect to the current output, $\mathbf{y}_t$, and weights, $\mathbf{D}_t$. This reduces to minimizing,

$$h(\mathbf{y}_t, \mathbf{D}_t) = \kappa \text{Tr}(\mathbf{M}_t \mathbf{D}_t \mathbf{M}_t \mathbf{D}_t) - 2\kappa \text{Tr}(\mathbf{W}_t^T \mathbf{D}_t \mathbf{W}_t)$$

$$+ 2\mathbf{y}_t^T \mathbf{D} \mathbf{M}_t \mathbf{D}_t \mathbf{y}_t - 4\mathbf{y}_t^T \mathbf{D}_t \mathbf{W}_t \mathbf{x}_t + 2\alpha_D \sum_{i=1}^{d} D_{t,ii}^2, \quad (12)$$

where

$$\mathbf{W}_t = \frac{1}{\kappa} \boldsymbol{\mathcal{Y}}_{t-1} \boldsymbol{\mathcal{X}}_{t-1}^T = \frac{1}{\kappa} \sum_{k=1}^{t-1} (\gamma^2)^{t-1-k} \mathbf{y}_k \mathbf{x}_k^T,$$

$$\mathbf{M}_t = \frac{1}{\kappa} \boldsymbol{\mathcal{Y}}_{t-1} \boldsymbol{\mathcal{Y}}_{t-1}^T = \frac{1}{\kappa} \sum_{k=1}^{t-1} (\gamma^2)^{t-1-k} \mathbf{y}_k \mathbf{y}_k^T, \quad (13)$$

subject to $-1 \leq \mathbf{y}_t \leq 1$.

In order to minimize this cost function, we write its gradient with respect to $\mathbf{y}_t$ as

$$\frac{1}{4} \nabla_{\mathbf{y}_t} h(\mathbf{y}_t, \mathbf{D}_t) = \mathbf{D}_t \mathbf{M}_t \mathbf{D} \mathbf{y}_t - \mathbf{D}_t \mathbf{W}_t \mathbf{x}_t, \quad (14)$$

and the gradient with respect to $D_{t,ii}$ as

$$\frac{1}{4} \nabla_{D_{t,ii}} h(\mathbf{y}_t, \mathbf{D}_t) = \frac{\kappa}{2} (\|\mathbf{M}_{t+1_{i,:}}\|_{\mathbf{D}}^2 - \|\mathbf{W}_{t+1_{i,:}}\|_2^2)$$
$$+ \alpha_D D_{t,ii} \quad (15)$$

Note that, as $\mathbf{D}_t > 0$,

$$-\mathbf{D}_t^{-1} \frac{1}{4} \nabla_{\mathbf{y}_t} h(\mathbf{y}_t, \mathbf{D}_t) = \mathbf{W}_t \mathbf{x}_t - \mathbf{M}_t \mathbf{D}_t \mathbf{y}_t \quad (16)$$

provides a local descent direction with respect to $\mathbf{y}_t$. We decompose $\mathbf{M}_t = \bar{\mathbf{M}}_t + \Upsilon_t$ where Υ_t is the diagonal, non-negative matrix containing diagonal components of $\mathbf{M}_t$. As a result the the last term in the descent direction can be rewritten as $\bar{\mathbf{M}}_t \mathbf{D} \mathbf{y}_t + \Upsilon_t \mathbf{D}_t \mathbf{y}_t$. We further define $\mathbf{u} = \Upsilon_t \mathbf{D}_t \mathbf{y}_t$. Since $\mathbf{u}$ and $\mathbf{y}_t$ are monotonically related, (16) is a descent direction with respect to $\mathbf{u}$. Using a gradient search with small step size, we obtain a neural network form for the gradient search algorithm

$$\frac{d\mathbf{u}(\tau)}{d\tau} = -\mathbf{u}(\tau) + \mathbf{W}_t \mathbf{x}_t - \bar{\mathbf{M}}_t \mathbf{D}_t \mathbf{y}_t(\tau),$$

$$\mathbf{y}_{t,i}(\tau) = c \left(\frac{\mathbf{u}_i(\tau)}{\Upsilon_{t,ii} \mathbf{D}_{t,ii}} \right), \quad (17)$$
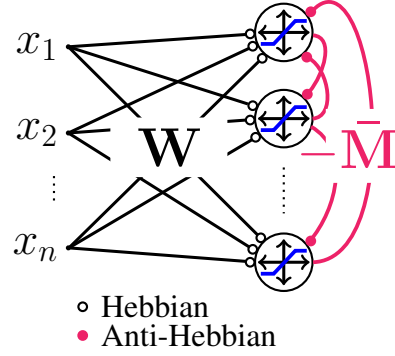

Fig. 1. Recurrent Neural Network for BSM.

where c is the clipping function corresponding to the projection of output components to the constraint set $[-1, 1]$, which can be written as

$$c(z) = \begin{cases} z & -1 \leq z \leq 1, \\ \text{sign}(z) & \text{otherwise.} \end{cases} \quad (18)$$

The corresponding recurrent neural network is shown in Fig. 1. It is interesting to observe that the inverse of the inner product weights act as activation function gains.

After the neural network dynamics of (17) converges, synaptic and inner product weights are updated for the next input by

$$\mathbf{W}_{t+1} = \gamma^2 \mathbf{W}_t + (1 - \gamma^2) \mathbf{y}_t \mathbf{x}_t^T$$

$$\mathbf{M}_{t+1} = \gamma^2 \mathbf{M}_t + (1 - \gamma^2) \mathbf{y}_t \mathbf{y}_t^T$$

$$\mathbf{D}_{t+1,ii} = (1 - \beta) \mathbf{D}_{t,ii} + \eta (\|\mathbf{W}_{t+1_{i,:}}\|_2^2 - \|\mathbf{M}_{t+1_{i,:}}\|_{\mathbf{D}_t^2}),$$

where η is the step size and $\beta = 2\eta\alpha_D$. The synaptic weight updates follow from their definitions in (13). In neuroscience, $\mathbf{W}$ updates are called Hebbian, and $\mathbf{M}$ updates are called anti-Hebbian (because of the minus sign in (17)) [14, 1]. The gain updates turn out to be in the form of a leaky integrator due to the last term in (15). It is interesting to observe that these gain updates are negative or positive depending on the balance between the norms of feedforward and recurrent weights. These weights are the reflectors of the recent excitation and inhibition statistics respectively (as they are corresponding correlation matrices). Relative increase in excitation(inhibition) would cause increase(decrease) in weights, and therefore, decrease(increase) in the corresponding homeostatic gains. This resembles the experimentally observed homeostatic gain adjustment behavior in biological neurons [15].

5. NUMERICAL EXAMPLES

5.1. Source Separation Example

As an illustration of the bounded source separation capability of the recurrent BSM network, we performed the following

3814
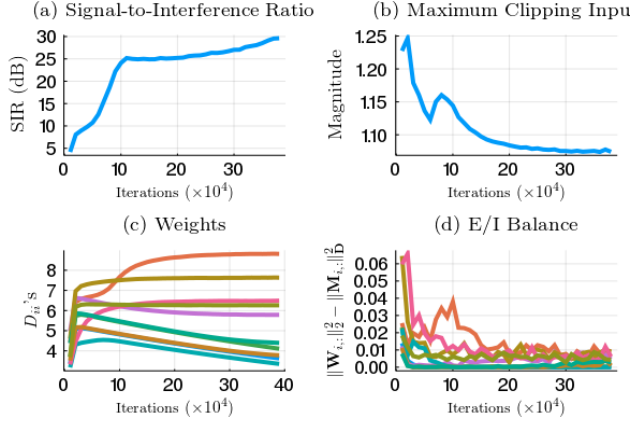
Fig. 2. 10 Uniform Source Example: (a) Signal-to-Interference Ratio (SIR) for outputs, (b) the maximum magnitude input (among all neurons) to the clipping function, (c) Weights(D_{ii}'s), (d) E/I balance for each neuron as measured based on excitation and inhibition synaptic strength norms.

numerical experiment: We generated 10 uniform sources from $\mathcal{U}[0, B]$ where the maximum value B is chosen randomly for each choice from $[2, 7]$ interval randomly (uniformly chosen). The sources were mixed by a random real orthogonal matrix, representing the whitened mixtures case. We selected the synapse update parameter $1 - \gamma^2 = 4 \times 10^{-3}$, the homeostatic gain update parameter as $\eta = 10^{-3}$ and the leaky integral parameter for homeostatic gain parameter as $\beta = 10^{-6}$.

Fig. 2.(a) shows the Signal-to-Interference Ratio energy (SIR), which is the total energy of source signals at the outputs to the energy of interference from other sources. The SIR increases with iterations after an initial transient and the clipping input's peak converges towards the clipping level. Homeostatic gains settle, as can be seen in both weight convergence curves in Fig. 2.(c) and the excitation/inhibition curves in Fig. 2.(d).

5.2. Image Separation

We consider the problem of separating 3 images selected from the database in [16] from their 3 random unitary mixtures. As illustrated in Fig. 3, we obtain outputs that are very similar to original nearly uncorrelated sources, with an SIR level around 30dB.

A. PROOF OF THEOREM I

We start by observing that, since $\mathbf{x}_t = \boldsymbol{\Theta}\bar{\mathbf{s}}_t$ and $\boldsymbol{\Theta}$ is real orthogonal, the equality constraint of *BSM* is equivalent to $\bar{\mathbf{S}}^T\bar{\mathbf{S}} - \mathbf{Y}^T\mathbf{D}\mathbf{Y} = 0$, where $\bar{\mathbf{S}}$ contains the sequence $\bar{\mathbf{s}}_t$ in its columns. Referring to Theorem I in [4], this equality con-
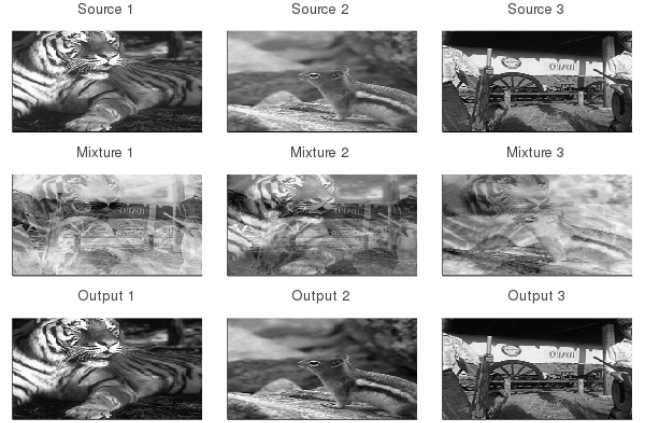
Fig. 3. Image Separation Example.

straint implies

$$\mathbf{D}^{1/2}\mathbf{y}_t \;=\; \mathbf{G}\bar{\mathbf{s}}_t = \underbrace{\mathbf{G}\boldsymbol{\Pi}_\mathbf{b}}_{\boldsymbol{\Phi}}\tilde{\mathbf{s}}_t. \tag{19}$$

The ith component of left hand side is maximized for index m where $\mathbf{s}_m = \operatorname{sign}(\boldsymbol{\Phi}_{i,:})^T$, i.e.,

$$D_{ii}\max_t(y_t^{(i)}) \;=\; \boldsymbol{\Phi}_{i,:}\operatorname{sign}(\boldsymbol{\Phi}_{i,:})^T = \|\boldsymbol{\Phi}_{i,:}\|_1 \tag{20}$$

As assumed in Theorem 1, since the vector sequence $\{\tilde{s}_t\}$ contains the corners of $\mathcal{B}_\infty$, which are all possible sign patterns, we can write

$$D_{ii} = \frac{\|\boldsymbol{\Phi}_{i,:}\|_1}{\max_t(y_t^{(i)})}, \qquad i = 1, \ldots, d. \tag{21}$$

In terms of this expression, we can rewrite the cost function for *BSM*, and obtain a lower bound:

$$\sum_{i=1}^d D_{ii}^2 \;=\; \sum_{i=1}^d \frac{\|\boldsymbol{\Phi}_{i,:}\|_1^2}{\max_t(y_t^{(i)})^2} \tag{22}$$

$$\geq\; \sum_{i=1}^d \|\boldsymbol{\Phi}_{i,:}\|_2^2 \tag{23}$$

$$=\; \sum_{i=1}^d (b^{(i)})^2\|\mathbf{G}(:,i)\|_2^2 = \sum_{i=1}^d (b^{(i)})^2, \tag{24}$$

where the inequality is due to $\max_t(y_t^{(i)}) \leq 1$ and ℓ_p-norm inequality, and the last equality is due to the fact that $\mathbf{G}$ is real orthogonal. The lower bound is achieved if and only if the inequality in (23) is equality. This condition is achieved if all rows of $\boldsymbol{\Phi}$, and therefore, $\mathbf{G}$ have only one nonzero element. Therefore, for the optimal solution, $\mathbf{G}$ has the form $\mathbf{G} = \boldsymbol{\Lambda}\mathbf{P}$, where $\boldsymbol{\Lambda}$ and $\mathbf{P}$ are as defined in the theorem. This implies that optimal D_{ii}s are equal to permuted versions of b_is. As a result, the global optimal solutions to *BSM* can be characterized by the statement of the theorem.

3815

6. REFERENCES

[1] Peter Dayan and LF Abbott, *Theoretical neuroscience: computational and mathematical modeling of neural systems (computational neuroscience)*, The MIT Press, 2005.

[2] Eric R Kandel, James H Schwartz, Thomas M Jessell, Steven Siegelbaum, and AJ Hudspeth, *Principles of Neural Science*, vol. 4, McGraw-hill New York, 2000.

[3] Chit-Kwan Lin, Andreas Wild, Gautham N Chinya, Yongqiang Cao, Mike Davies, Daniel M Lavery, and Hong Wang, "Programming spiking neural networks on intels loihi," *Computer*, vol. 51, no. 3, pp. 52–61, 2018.

[4] Cengiz Pehlevan and Dmitri B Chklovskii, "A normative theory of adaptive dimensionality reduction in neural networks supplementary information," *Proceedings of Advances in Neural Information Processing Systems 28 (NIPS 2015)*, 2015.

[5] Cengiz Pehlevan and Dmitri B Chklovskii, "Neuroscience-inspired online unsupervised learning algorithms: Artificial neural networks," *IEEE Signal Processing Magazine*, vol. 36, no. 6, pp. 88–96, 2019.

[6] Cengiz Pehlevan and Dmitri B Chklovskii, "A hebbian/anti-hebbian network derived from online nonnegative matrix factorization can cluster and discover sparse features," in *2014 48th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2014, pp. 769–775.

[7] Yanis Bahroun, Eugénie Hunsicker, and Andrea Soltoggio, "Neural networks for efficient nonlinear online clustering," in *International Conference on Neural Information Processing*. Springer, 2017, pp. 316–324.

[8] Anirvan Sengupta, Cengiz Pehlevan, Mariano Tepper, Alexander Genkin, and Dmitri Chklovskii, "Manifold-tiling localized receptive fields are optimal in similarity-preserving neural networks," in *Advances in Neural Information Processing Systems*, 2018, pp. 7080–7090.

[9] Cengiz Pehlevan, Sreyas Mohan, and Dmitri B Chklovskii, "Blind nonnegative source separation using biological neural networks," *Neural computation*, vol. 29, no. 11, pp. 2925–2954, 2017.

[10] Sergio Cruces, "Bounded component analysis of linear mixtures: A criterion of minimum convex perimeter," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2141–2154, 2010.

[11] Alper T Erdogan, "A class of bounded component analysis algorithms for the separation of both independent and dependent sources," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5730–5743, 2013.

[12] Eren Babatas and Alper T Erdogan, "An algorithmic framework for sparse bounded component analysis," *IEEE Transactions on Signal Processing*, vol. 66, no. 19, pp. 5194–5205, 2018.

[13] Eren Babatas and Alper T Erdogan, "Sparse bounded component analysis for convolutive mixtures," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2741–2745.

[14] Peter Földiak, "Forming sparse representations by local anti-hebbian learning," *Biological cybernetics*, vol. 64, no. 2, pp. 165–170, 1990.

[15] Gina G Turrigiano, "Homeostatic plasticity in neuronal networks: the more things change, the more they stay the same," *Trends in neurosciences*, vol. 22, no. 5, pp. 221–227, 1999.

[16] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition and Measuring Ecological Statistics*, 2002, vol. 2, p. 416.