Using Sinusoidally-Modulated Noise as a Surrogate for Slow-Wave Sleep to Accomplish Stable Unsupervised Dictionary Learning in a Spike-Based Sparse Coding Model

Yijing Watkins¹, Edward Kim³, Andrew Sornborger¹ and Garrett T. Kenyon¹,²
Los Alamos National Laboratory¹
New Mexico Consortium²
Drexel University³
twatkins@lanl.gov, ek826@drexel.edu, sornborg@lanl.gov, gkeynon@lanl.gov

Abstract

Sparse coding algorithms have been used to model the acquisition of V1 simple cell receptive fields as well as to accomplish the unsupervised acquisition of features for a variety of machine learning applications. The Locally Competitive Algorithm (LCA) provides a biologically plausible implementation of sparse coding based on lateral inhibition. LCA can be reformulated to support dictionary learning via an online local Hebbian rule that reduces predictive coding error. Although originally formulated in terms of leaky integrator rate-coded neurons, LCA based on lateral inhibition between leaky integrate-and-fire (LIF) neurons has been implemented on spiking neuromorphic processors but such implementations preclude local online learning. We previously reported that spiking LCA can be expressed in terms of predictive coding error in a manner that allows for unsupervised dictionary learning via a local Hebbian rule but the issue of stability has not previously been addressed. Here, we use the Nengo simulator to show that unsupervised dictionary learning in a spiking LCA model can be made stable by incorporating epochs of sinusoidally-modulated noise that we hypothesize are analogous to slow-wave sleep. In the absence of slow-wave sleep epochs, the $\|L\|_2$ norm of individual features tends to increase over time during unsupervised dictionary learning until the corresponding neurons can be activated by random Gaussian noise. By inserting epochs of sinusoidally-modulated Gaussian noise, however, the $\|L\|_2$ norms of any activated neurons are down regulated such that individual neurons are no longer activated by noise. Our results suggest that slow-wave sleep may act, in part, to ensure that cortical neurons do not “hallucinate” their target features in pure noise, thus helping to maintain dynamical stability.
ically transpose and normalization operations performed globally on the entire weight matrix, and further requires signed outputs in order to represent the sparse reconstruction error. In our previous work [17], we showed that unsupervised dictionary learning can be accomplished using a modified Spiking LCA (S-LCA) that employs only local computations and uses only unsigned spiking output from all neurons. Although our previous results provide a proof-of-concept for how neuromorphic processors can be configured so as to self-organize in response to natural environmental stimuli without sacrificing efficiency, the issue of stability was not addressed. In this work, we use the Nengo simulator to show that unsupervised dictionary learning in a spiking LCA model can be made stable by incorporating epochs of sinusoidally-modulated noise that we hypothesize are analogous to slow-wave sleep.

2. Methods

2.1. Unsupervised Dictionary Learning with a Non-Spiking LCA

Given an overcomplete basis, non-spiking LCA [9] can be used to find a minimal set of active neurons that represent the input to some degree of fidelity. Each neuron can be thought of as a generator that adds its associated feature vector to the reconstructed input with an amplitude equal to its activation. For any particular input, the optimal sparse representation is given by a vector of neural activations that minimizes the following cost function:

\[
E = \frac{1}{2} ||I - \{\Phi \ast a\}||^2_2 + \frac{1}{2} \lambda^2 ||a||_0 \tag{1}
\]

where \(I\) is the input vector and \(\Phi\) is a dictionary of feature kernels that are convolved with the activation coefficients \(a\). The feature kernels are assumed to be normalized such that \(\Phi \Phi^T = I\). The \(L_0\) norm \(||a||_0\) simply counts the number of non-zero activation coefficients. The factor \(\lambda\) acts as a trade-off parameter; larger \(\lambda\) values encourage greater sparsity (fewer non-zero coefficients) at the cost of greater reconstruction error.

LCA finds a local minimum of the cost function defined in Eq. (1) by introducing the dynamical variables (membrane potentials) \(u\) such that the output \(a\) of each neuron is given by a hard-threshold transfer function, with threshold \(\lambda\), of the membrane potential: \(a = T_\lambda(u) = H(u - \lambda)u\), where \(H\) is the Heaviside function[9].

For a given input, the cost function defined in equation (1) is then minimized by taking the gradient of the cost function with respect to \(a\) and solving the resulting set of coupled differential equations for the membrane potentials \(u\):

\[
\dot{u} \propto -\frac{\partial E}{\partial a} = -u + \Phi^T [I - \Phi T_\lambda(u)] + T_\lambda(u). \tag{2}
\]

where the final term in Eq. (2) removes the self-interaction (a neuron does not inhibit itself).

An update rule for feature kernels can be obtained by taking the gradient of the cost function with respect to \(\Phi\):

\[
\Delta \Phi \propto -\frac{\partial E}{\partial \Phi} = a \otimes [I - \Phi a] = a \otimes R \tag{3}
\]

where we introduced an intermediate residual layer \(R\) corresponding to the sparse reconstruction error.

For non-spiking LCA, online unsupervised dictionary learning is achieved via a two step process: First, a sparse representation for a given input is obtained by integrating Eq. (2), after which Eq. (3) is evaluated to slightly reduce the reconstruction error given the sparse representation of the current input.

As illustrated in Figure 1, the weight update (3) resembles a local Hebbian learning rule for \(\Phi\) with pre- and post-synaptic activities \(a\) and \(R\) respectively. However, the computation of \(\Phi^T\) as well as the normalization constraint renders the overall dictionary learning process a non-local operation.

We have previously shown that our implementation of non-spiking LCA can be used to learn a convolutional dictionary in an unsupervised, self-organizing manner that factors a complex, high-dimensional natural image into an overcomplete set of basis vectors that capture the high-dimensional correlations in the data [11]. In the next section, we show how this implementation can be adapted to an S-LCA model that uses only local computations and unsigned spiking output.

2.2. S-LCA Unsupervised Dictionary Learning

Our S-LCA model for unsupervised dictionary learning is shown in Figure 2, where the superscripts indicate the different layers (e.g. \(a^1\) denotes the input spikes). We replace the non-spiking leaky-integrator model of Eq. 2 with...
We augment Eq. 6 with a firing condition analogous to Eq. 5 with
\[ \Delta \Phi \propto (a^L) \otimes (a^R)^T. \] (9)

3. Results

3.1. Instability of unsupervised learning without sleep

In our previous work [17], we showed that unsupervised dictionary learning can be accomplished using a modified Spiking LCA (S-LCA) that employs only local computations and uses only unsigned spiking output from all neurons. Although our previous work provide a proof-of-concept for how neuromorphic processors can be configured so as to self-organize in response to natural environmental stimuli without sacrificing efficiency, the issue of stability was not addressed.

As indicated in In Figure 3 and Figure 4, in the absence of slow-wave sleep epochs, the \( |L|_2 \) norm of individual features tends to increase over time during unsupervised dictionary learning until the corresponding neurons become continuously activated. The features learned in the absence of sleep epochs are initially reasonable (Figure 5) but ultimately become meaningless (Figure 5) but ultimately become meaningless as the system becomes unstable.

3.2. Stability of learning with slow-wave sleep

As illustrated in In Figure 6, we used sinusoidally-modulated noise that we hypothesize is analogous to slow-wave sleep. Shown is the norm of the stimulus over one slow-wave sleep cycle. Nothing about the model or learning rule changes during slow-wave sleep. Shown is the firing rates of the LIF neurons in the sparse coding layer \( \langle a^L \rangle \) are likewise represented as low-pass filtered versions of the corresponding spike trains. In terms of these local firing rates, the update of \( \Phi \) is given by a local Hebbian learning rule:

\[ \Delta \Phi \propto (a^L) \otimes (a^R). \] (9)

The traces time in the local Hebbian learning rule are 100 msec.

a leaky integrate-and-fire (LIF) model consisting of a membrane potential, \( u \), and a binary spiking output, \( a \):

\[ \dot{u} \propto -u + u_{input} \] (4)

\[ a = \begin{cases} 0 & u < \lambda \\ 1 & u \rightarrow 0 & u \geq \lambda \end{cases} \] (5)

where \( \lambda \) again plays the role of a threshold that controls the level of sparsity and \( u_{input} \) is the sum of the input received from connected neurons. When \( u \) crosses the threshold \( \lambda \) a spike is generated and \( u \) is reset to zero.

As with non-spiking LCA, the residual layer \( R \) in S-LCA is driven by the difference between the input and the reconstructed input generated by the LCA layer, which for S-LCA is given by \( a^I - \Phi a^L \). Values of \( \langle a^R \rangle \) above and below the target baseline firing rate \( \langle a^R \rangle_0 \) encode positive and negative errors, respectively. Eq. 4 for the residual layer \( R \) then becomes

\[ \dot{u}^R \propto -u^R + a^I - \Phi a^L. \] (6)

We augment Eq. 6 with a firing condition analogous to Eq. 5 with \( \lambda \rightarrow \lambda^R \).

The input to the sparse coding layer \( L \) in Figure 2 is denoted by \( u^L_{input} \), given by,

\[ u^L_{input} = \Phi^T a^R. \] (7)

Likewise, in Figure 2, the LCA layer \( L \) is again an LIF layer whose equation of motion is given by:

\[ \dot{u}^L \propto -u^L + \Phi^T a^R + \alpha (a^L), \] (8)

where \( \alpha (a^L) \) is a self interaction involving the low-pass filtered (trace) of the spiking output that helps to maintain firing activity when the residual error is zero. The constant \( \alpha = 0.75 \) and the trace time constant 500 msec were determined empirically to produce improved sparse reconstructions without ringing or overshoot. We again augment Eq. 8 with a firing condition analogous to Eq. 5 with \( \lambda \rightarrow \lambda^L \).

Unsupervised dictionary learning can be used to update the weight matrices \( \Phi \) given only information locally available at each synapse. To compute the weight updates, we introduce the low-pass filtered spike trains, or instantaneous firing rates or traces, of the LIF neurons in the residual layer \( \langle a^R \rangle = \langle a^R \rangle - \langle a^R \rangle_0 \) computed relative to the target baseline firing rate of the residual layer. The firing rates of the LIF neurons in the sparse coding layer \( \langle a^L \rangle \) are likewise represented as low-pass filtered versions of the corresponding spike trains. In terms of these local firing rates, the update of \( \Phi \) is given by a local Hebbian learning rule:

\[ \Delta \Phi \propto (a^L) \otimes (a^R). \] (9)
Figure 3: Spikes (bottom) and spike trace (top) from a representative S-LCA neuron without sleep. In the absence of slow-wave sleep epochs, the $|L_2|$ norm of individual features tends to increase over time during unsupervised dictionary learning until the corresponding neurons are continuously activated. Individual spikes cannot be resolved at this level of resolution.

Figure 4: Representative residual layer neuron spikes (bottom) and spike trace (top) without sleep. As the gain of the feedback loop increases, the residual layer becomes persistently active.

(a) Before training  (b) Some training  (c) More training

Figure 5: Top 64 most active dictionary elements without slow-wave sleep.

features. During slow-wave sleep, however, the stimulus consists of random Gaussian noise that is not learnable.

As indicated in Figure 7 and Figure 8, by inserting epochs of sinusoidally-modulated Gaussian noise, the $|L_2|$ norms of the feature vectors associated with neurons activated during slow-wave sleep are down regulated such that individual neurons are no longer persistently active. In particular, neurons dynamically adjust their gain during slow-wave sleep so as not to be activated by random Gaussian noise. Instead, as in the representative S-LCA neuron depicted, activity is only sparsely induced by a limited subset of natural stimuli. On occasion, S-LCA neurons are activated by sinusoidally-modulated Gaussian noise but any such activations come to be repressed over successive slow-wave sleep cycles. Our approach is consistent with evidence that long term potentiation changes occur during wakefulness due to a net increase in synaptic weights, and slow wave sleep is necessary to promote generalized depression of synapses [15].

Example reconstructions of a natural stimulus over the
Figure 7: Representative S-LCA neuron spikes (bottom) and spike trace (top) with slow-wave sleep. By inserting epochs of sinusoidally-modulated Gaussian noise, the $|L|^2$ norms of any neurons activated during slow-wave sleep are down regulated such that individual neurons are no longer activated by random Gaussian noise. As a result, neurons are sparsely activated by only a subset of natural stimuli and rarely by random noise.

Figure 8: Representative residual neuron spikes (bottom) and spike trace (top) with slow-wave sleep. Although the trace does not always tend to the baseline value indicting zero residual, we observe that average residual does diminish with continued unsupervised training.

course of unsupervised training with slow-wave sleep are shown in Figure 11 and Figure 12. Corresponding learned features are shown in Figure 13. As dictionary elements converge toward more meaningful features, the corresponding reconstructions become more accurate. Computational constraints prevented us from continuing these experiments until the dictionary was fully converged. However, including epochs of sinusoidally-modulated noise hypothesized to be analogous to slow-wave sleep produced stable sparse reconstructions that steadily improved with further unsupervised training whereas the system became dynamically unstable in the absence of slow-wave sleep even when using identical random weight initialization parameters and learning rates.

3.3. Conclusion

Although our previous work [17] provide a proof-of-concept for how neuromorphic processors can be configured so as to self-organize in response to natural environmental stimuli without sacrificing efficiency, the issue of stability was not addressed. In this work, we use the Nengo simulator to show that unsupervised dictionary learning in a spiking LCA model can be made stable by incorporating epochs of sinusoidally-modulated noise that we hypothesize are analogous to slow-wave sleep.

References


Figure 9: Sinusoidally-modulated slow-wave sleep with the noise amplitude set to 0.

Figure 10: Sinusoidally-modulated slow-wave sleep with the noise amplitude set to 1.

Figure 11: Cifar-10 Input.

Figure 12: Reconstructions with sleep.

(a) Before training  (b) Some training  (c) More training

(a) Before training  (b) Some training  (c) More training

Figure 13: Top 64 most active dictionary elements with Sleeps.


