On the Memory Storage Capacity of Hopfield Networks using Löwdin Orthogonalizations

Ramesh Naidu Annavarapu
Department of Physics
School of Physical, Chemical and Applied Sciences
Pondicherry University, Puducherry, 605 014, India.

rameshnaidu.phy@pondiuni.edu.in

Abstract – There are several processes of memory where one does not always handle the patterns sequentially. For such situations, Löwdin's non-sequential orthogonalization methods might play a role where one can handle all the input patterns at the same time. These democratic orthogonalization methods are incorporated in the Hopfield model to circumvent the problem of catastrophic interference in the retrieval of, say, episoidal memories. In this paper, we have numerically tested the condition of associative memory of Hopfield model for large number of synaptic patterns p with the number of firing and not-firing neurons N. We have simulated the memory using Hopfield model by incorporating the sequential Gram-Schmidt and non-sequential Löwdin orthogonalization methods as the storage strategies for large data sets. We have numerically tested how the stored synaptic patterns can be retrieved.

Key Terms: Canonical orthogonalization, Catastrophic interference, Gram-Schmidt method, Hopfield model, Memory plasticity, Symmetric orthogonalization.

I. INTRODUCTION

The spin glass model proposed by [1] is the origin to model the brain's neural network. The activity in neural networks gained considerable momentum after the associative memory model proposed by J. J. Hopfield [2, 3]. But this model has a serious constraint that there is an interference between the stored input patterns which turns *catastrophic* when the number of stored patterns becomes quite large. In the Hopfield model, in a system of *N* firing and quiescent neurons and *p* input patterns, the condition of associative memory fails when $p/N > 0.14$, [4, 5, 6] i.e., this model gives a memory capacity of $0{:}14N$ [7]. The retrieval process of stored patterns breaks down and a memory blackout occurs when the number of patterns that come to be recorded exceeds $0{:}14N$. This is the *catastrophic interference* [8] and is also known as the *stability-plasticity* problem [9]. The cause for this *catastrophe* is the correlation among the patterns (or memories), which makes the system noisy as the number of stored memories increases until retrieval becomes minimal [4, 5, 6]. This catastrophic interference has been a serious constraint also in other connectionist models [10] of neural networks in which the learning of new information beyond a certain limit causes sudden and complete disappearance of previously stored information.

It was proposed [11] that the process of orthogonalization could help to overcome the limitation of catastrophic forgetting on the memory capacity of the stored patterns. V. Srivastava et. al., [11, 12] have proposed models to understand how the brain discriminates and categorises different tasks. The mathematical computation of orthogonalization overcomes the noise among the learnt patterns and it circumvents the problem of catastrophic interference among the memories. The process of orthogonalization essentially compares a new object with those already in memory and it identifies their *similarities* and *differences* so that the new object can be placed in the right category. It is our understanding that the process of orthogonalization can overcome catastrophic interference but at the same time it can perform discrimination and categorization of different objects and tasks.

The Gram-Schmidt orthogonalization does this sequentially by taking patterns one by one and comparing them with the existing patterns in the memory. The sequential Gram-Schmidt orthogonalization process acts like decision making process wherein one object is compared with the stored ones and the nature of correlation with them decides, in an efficient and economical manner, in what form it should be stored. The noise is also eliminated. It is still unclear how the information in the brain is actually stored, processed and retrieved in biological neural networks. In the process of understanding how the brain processes *similar* and *different* tasks simultaneously, a few researches have explored the finer aspects of distinction between the *same* and *different* kind of stimuli [13], how the brain isolates objects from their backgrounds [14], how rats compare scents to do a task efficiently [15, 16] and how the previous awareness affects the sensory learning [17]. But the problem in all these and many more cognitive functions is the mind's ability to *discriminate*. How it is carried out in the brain physiologically is not known. It is possible that the brain could perhaps do orthogonalization, or something similar to it, to discriminate and categorise the information.

There are several other processes of memory where one does not always handle the patterns sequentially.

For such situations, Löwdin's non-sequential orthogonalization methods [18] and other methods [19] might play a role where one can handle all the input patterns at the same time. These *democratic* orthogonalization methods are incorporated in the Hopfield model to circumvent the problem of catastrophic interference in the retrieval of, say, episoidal memories.

In this paper, we have numerically tested the condition of associative memory of Hopfield model for large number of synaptic patterns *p* with the number of firing and not-firing neurons *N*. We have simulated the memory using Hopfield model by incorporating the sequential Gram-Schmidt and non-sequential Löwdin orthogonalization methods as the storage strategies for large data sets. We have numerically tested how the stored synaptic patterns can be retrieved.

II. THE HOPFIELD MODEL

We first outline the Hopfield model briefly to elicit the mathematical representation of the brain functions involved in the process of learning, memory and recognition, etc. [4, 5, 6]. We have studied the Hopfield model in neural networks to understand how the brain stores, retrieves and processes the information and to understand the catastrophic interference. The Hopfield model is numerically tested for large data sets to understand the condition of memory catastrophe in the case of large data sets. In a system of *N* number of firing and not-firing neurons, a pattern is represented by an *N*-dimensional vector $\vec{\xi}$ whose *N* components $\vec{\xi}$'s are either +1 or -1, respectively. A pattern of +1 or -1 represents a stable state of the neural system, and hence a memory, if it minimises the Hamiltonian,

$$H = -\frac{1}{2}\sum_{i,j=1}^{N} W_{ij}\, \xi_i \xi_j \qquad \text{…. (1)}$$

where W_{ij} is the synaptic connection between the two neurons ξ_i and ξ_j . The procedure involved in Hopfield model is outlined here.

Let p be the number of patterns to be memorized and let the memorized patterns be represented by $\overrightarrow{\xi^{\mu}}$, with $\mu = 1,2 \ldots p$= 1. These patterns are randomly generated and each pattern is normalized. The weight matrix is formulated by using the Hebb's auto-association outer product rule,

$$W_{ij} = \frac{1}{N}\sum_{\mu=1}^{p} \xi_i^{\mu}\, \xi_j^{\mu} \quad i,j = 1,2, \ldots N, \qquad \text{….. (2)}$$

where the index μ represents the patterns. It gives the symmetric matrix $W_{ij} = W_{ji}$. Each memory is treated as independent, i.e. a new memory is added without reference to the previous ones. As there is no self-feedback in a neuron, we set

$$W_{ij} = 0 \qquad \text{….(3)}$$

$$\text{so that, } W_{ij} = \frac{1}{N}\sum_{\mu=1}^{p} \xi_i^{\mu} \xi_{j-}^{\mu} \frac{p}{N} I, \qquad \text{..... (4)}$$

where I is identity matrix. Starting with the first pattern, the patterns are picked up one by one sequentially and weights are calculated according to (2) in a cumulative manner. The cumulative changes in the weights store the patterns or the memories. In order to test if the patterns are there in the memory, we present one of the p memories of learnt patterns, say $\mu = \nu$, to the brain for association. It produces local fields h_i^{ν} on all the neurons. This can be expressed as

$$h_i^{\nu} = \frac{1}{\sqrt{N}}\sum_{j=1}^{N} (\xi_i^{\mu}\, \xi_j^{\mu}) \xi_j^{\nu} \qquad \text{…..(5)}$$

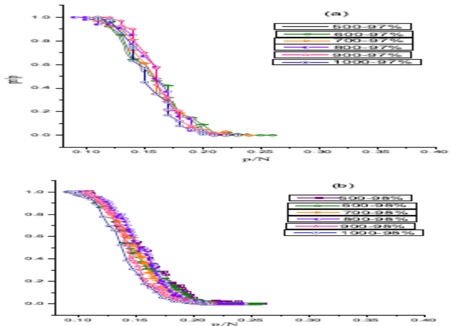
$$or\ \frac{1}{\sqrt{N}}\, h_i^{\nu} \xi_i^{\nu} = \frac{1}{\sqrt{N}}\sum_{j=1}^{N} (\frac{1}{N}\sum_{\mu=1}^{p} \xi_i^{\mu} \xi_j^{\mu}) \xi_j^{\nu}\, \frac{1}{\sqrt{N}} \xi_i^{\nu} \qquad \text{….(6)}$$

$$= \frac{1}{N^2}\sum_{j=1}^{N} (\xi_j^1 \xi_i^1 + \xi_j^2 \xi_i^2 + \cdots + \xi_j^p \xi_i^p) \xi_j^{\nu} \xi_i^{\nu} \qquad \text{....(7)}$$

The success of retrieval is measured by the sign of h_i^{ν} on all the sites i through the requirement $\frac{1}{\sqrt{N}} h_i^{\nu} \xi_i^{\nu} > 0$. If the sign of h_i^{ν} turns out to be the same as that of ξ_i^{ν} for almost all *i*'s, then the presented ξ^{ν} is supposed to associate well with the ν^{th} pattern. Hence the condition for good retrieval of the memorised patterns can be written as

$$sgn(h_i) = sgn(\xi_i). \qquad \text{…..(8)}$$

If this condition is satisfied on 97% or more of neurons, then it is considered to be a good retrieval. We have numerically tested a network consisting of 100 to 1000 neurons storing p patterns for an overlap $(\overrightarrow{h^{\nu}}.\overrightarrow{\xi^{\nu}})$ of 97%, 98% and 99% or more. We have drawn the graphs for these three cases between the fraction of the presented pattern that are retrieved and $p/_N$, the number of presented patterns normalized by the number of neurons, N, for 500 to 1000 neurons considering the overlap between the presented and retrieved patterns to be 97%, 98% and 99% and is shown in Figure 1. The results show that the memory capacity of the Hopfield model is limited to $p/_N \simeq$ 0.14. The fraction of retrieved patterns begins to drop as $p/_N$ approaches 0.12 and then drops rapidly to *zero* around $p/_N \simeq$ *0.18*. The rate of drop does not



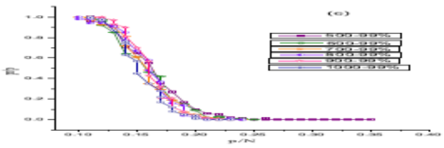


Figure 1: Plots of (no. of retrieved patterns)/(no. of presented patterns) versus (no. of presented patterns)/(no.of. neurons) for retrieval percentages (a) 97%, (b) 98% and (c) 99%.

appear to depend on the *overlap* parameter although one would expect it to be steeper as overlap increases from 97% to 99%. The above constraints in the model are known in the literature as *memory catastrophe* which is quite serious if we wish to have a large number of stored patterns. The reason for this is that there is interference between the stored input patterns because they are correlated, i.e., their vectors have non-zero dot products. The correlation among the memories makes the system noisy as the number of stored patterns increases. As the stored patterns increase in number, at some point the storage capacity will be exhausted, and new patterns will interfere with old patterns and they will prevent the retrieval of stored patterns. The problem of catastrophic interference among the memories could be overcome with the help of orthogonalization without having to resort to sparse coding as is sometimes suggested in the literature. The process of orthogonalization enables the network to compare a new object with those already in memories and to identify their similarities and differences so that the new object can be placed in the right memory.

III. HOPFIELD NETWORKS WITH GRAM-SCHMIDT METHOD

The Hopfield model is tested by incorporating the Gram-Schmidt orthogonalization process before the calculation of weight matrix W_{ij} . The use of Gram-Schmidt orthogonalization was proposed by V. Srivastava and S. F. Edwards to eliminate the noise and to achieve high memory capacity but at the same time to handle the correlated memories. The orthonormalized patterns obtained through the Gram-Schmidt method are used to calculate the weight matrix W_{ij}. The local field h_i is calculated by presenting the normalized raw patterns and the orthonormalized ones. Then the normalized raw patterns and orthonormalized patterns are presented to the weight matrix constructed by using the orthonormalized patterns. It is observed that the number of retrieved patterns is increased when the normalized raw patterns are presented to the weight matrix calculated using the orthonormalized patterns than in the usual Hopfield model. When the orthonormalized patterns are presented to the same weight matrix, all the patterns are completely retrieved, i.e., there is 100% retrieval of the presented patterns

The Gram-Schmidt orthogonalization is incorporated in the Hopfield model as follows.

Let there be p patterns to be memorized in a system of N neurons. Let the memorized patterns be represented by $\overrightarrow{\xi^{\mu}}$, $\mu = 1,2, \dots p$. These patterns are randomly generated and each pattern is normalized. The Gram-Schmidt orthogonalization process is used to calculate the orthonormal patterns. That is the process of exploring the store on the arrival of new information, say $\overrightarrow{\xi^{n}}$, is represented mathematically as modifying the raw $\overrightarrow{\xi^{n}}$ so that it becomes orthogonal to the existing patterns. This process is defined through

$$\eta_k^n = \xi_k^n - \sum_{q=1}^{n-1} \eta_k^q \frac{\sum_{i=1}^{N} \eta_i^q \xi_i^n}{\sum_{i=1}^{N} \eta_i^q \eta_i^q} \quad \text{.....(9)}$$

This amounts to exacting details from the n^{th} as well as the earlier patterns. The brain stores these $\overrightarrow{\eta^{n}}'s$ rather than the raw $\overrightarrow{\xi^{n}}'s$ as in the case of Hopfield model. The second term on the right-hand side represents the sum of projections of $\overrightarrow{\xi^{n}}'s$ on all $\overrightarrow{\eta^{n}}'s$ with $q < n$. Thus $\overrightarrow{\eta^{n}}$ is obtained after subtracting out from $\overrightarrow{\xi^{n}}'s$ its commonalities with all the earlier $\overrightarrow{\eta^{n}}'s$. The normalization $\overrightarrow{\eta^{\mu}}\overrightarrow{\eta^{\mu}} = N$ remains for all μ. Note that

$$\sum_i \eta_i^n \eta_i^q = 0, for\ all\ q < n \quad \text{.....(10)}$$

The weight matrix is obtained as before using $\overrightarrow{\eta^{n}}'s$.

$$W_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \eta_i^{\mu} \eta_j^{\mu}, i, j = 1,2, \dots N, \quad \text{.....(11)}$$

where the index _ represents the patterns. It gives the symmetric matrix $W_{ij} = W_{ji}$. Each memory is independent, i.e., each time a new pattern (or memory) is added without reference to the previous ones. Again for $W_{ii} = 0$, we have the memory matrix

$$W_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \eta_i^{\mu} \eta_j^{\mu} - \frac{p}{N} I \quad \text{.....(12)}$$

where I is the identity matrix.

If we now present $\overrightarrow{\eta^{\mu}}'s$ for retrieval all learnt $\overrightarrow{\eta^{\mu}}'s$ are retrieved perfectly as long as $p \leq N$. Thus the memory capacity is increased from $p/_N = 0.14$ to $p/_N = 1$. Interestingly the raw patterns $\overrightarrow{\xi^{n}}'s$ are also retrieved for $p \leq N$. Figure 2 shows perfect retrieval of orthogonalized as well as raw patterns after Gram-Schmidt scheme is invoked. In an N-dimensional space there can be N orthogonal vectors. This restricts the capacity to $p/_N = 1$.

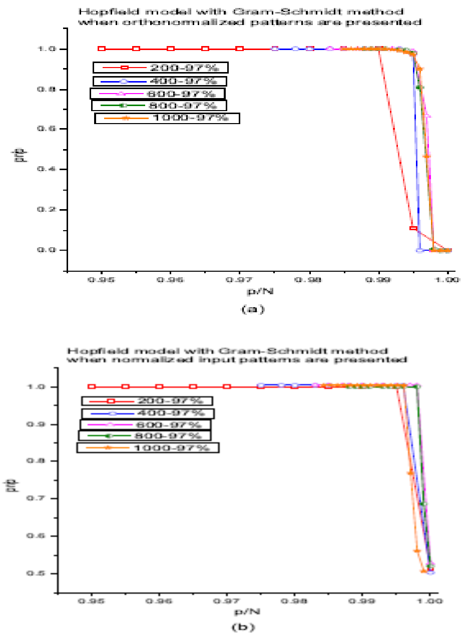


Figure 2: Plots of (no. of retrieved patterns)/(no. of presented patterns) versus (no. of presented patterns)/(no. of. neurons) for Gram-Schmidt method.

IV. HOPFIELD NETWORKS USING LÖWDIN ORTHOGONALIZATIONS

We have tested the Hopfield model by incorporating the Löwdin's symmetric and canonical orthogonalization methods as the storage strategies. As with Gram-Schmidt we expect the patterns orthogonalized by these two schemes to be retrieved perfectly since they will also eradicate the noise in the same manner as Gram-Schmidt does. However we expect the orthogonalized patterns to possess different characters than those orthogonalized with Gram-Schmidt scheme. Since the underlying schemes are democratic in nature and handle the given set of random patterns altogether yet quite differently, we expect that the fine differences between canonical and symmetrical methods and their differences with Gram-Schmidt will be useful in their applications to cognitive phenomena. We randomly generate the raw patterns of firing/not-firing neurons and normalize them. We then calculate the symmetric and canonical orthonormal bases and form the weight matrices using them. Then, we presented the normalized raw patterns to the symmetric and canonical weight matrices and calculated their local field to study retrieval.

A. HOPFIELD NETWORKS WITH SYMMETRIC ORTHOGONALIZATION

The Löwdin's symmetric orthogonalization method can be incorporated in the Hopfield model as follows. Let p patterns $\{\xi^{\mu}\}, \mu = 1,2,\ldots,p$ in N dimensions be written as matrix Ξ, whose columns represent the input patterns. The Gram matrix M is constructed as $M = \Xi\Xi^{T}$ and its eigenvalues d and normalized eigenvectors U are calculated. The symmetric orthonormal bases are obtained using

$$\Phi = \Xi M^{-1/2} \qquad \ldots.(13)$$

The brain stores these $\overrightarrow{\phi^{n}}'s$ rather than the raw $\xi_i's$. The weight matrix is now

$$W_{ij} = \frac{1}{N}\sum_{\mu=1}^{p}\phi_i^{\mu}\phi_j^{\mu} - \frac{p}{N}I, \qquad \ldots.(14)$$

where I is the identity matrix.

B. HOPFIELD NETWORKS WITH CANONICAL ORTHOGONALIZATION

The Löwdin's canonical orthogonalization method is incorporated in the Hopfield model in the same fashion as above. The canonical orthonormal bases are obtained using

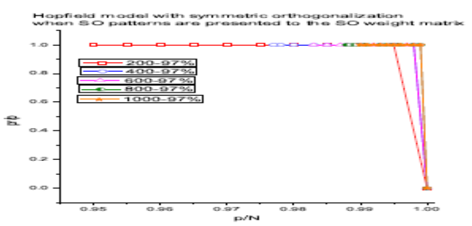
$$\Lambda = \Xi\, U\, d^{-1/2} \qquad \ldots.(15)$$

And the weight matrix is written as

$$W_{ij} = \frac{1}{N}\sum_{\mu=1}^{p}\lambda_i^{\mu}\lambda_j^{\mu} - \frac{p}{N}I \qquad \ldots.(16)$$

where I is the identity matrix. Quite expectedly, the memory capacity in both `symmetric' and `canonical' orthogonalizations is $p/_N = 1$ as shown in Figures 3 and 4.

The major difference between the Gram-Schmidt and the Löwdin methods is that in the latter as one new pattern is added after a set of □ patterns is orthogonalized; the entire lot of □ + *1* patterns is orthogonalized all over again. Thus the sets of □ and □ + *1* orthogonalized patterns differ from each other. The differences between the sets also depend on whether we employ canonical or symmetric orthogonalization and discussed these changes in the following section.



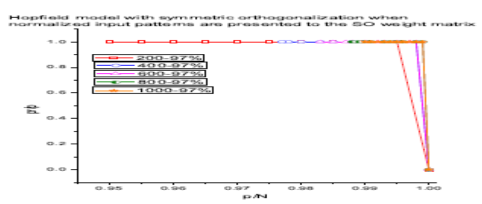
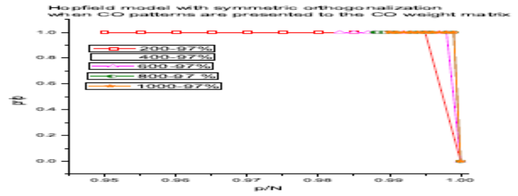


Figure 3: Plots of (no. of retrieved patterns)/(no. of presented patterns) versus (no. of presented patterns)/(no. of. neurons) for symmetric method.



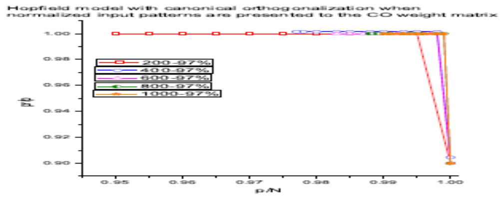
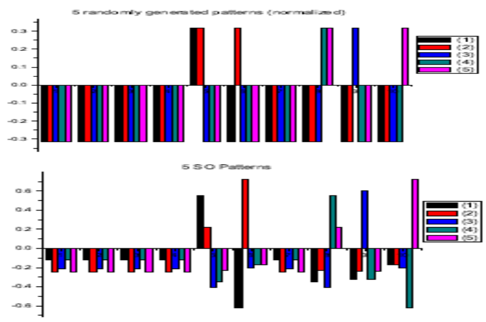


Figure 4: Plots of (no. of retrieved patterns)/(no. of presented patterns) versus (no. of presented patterns)/(no. of. neurons) for canonical method.

V. COMPUTATIONS AND RESULTS

We have numerically simulated and tested the Hopfield model by presenting patterns of firing/not-firing neurons starting with the smaller data sets. We have applied the Löwdin's symmetric and canonical orthogonalizations for the case of 5 randomly generated normalized input patterns of 10 firing or not-firing neurons. When a new pattern is added each time without disturbing the previously stored patterns, it changes all the weights and triggers electric activity among neurons and generates local field on each of them for symmetric and canonical orthogonalizations. The addition of new pattern will change all the previously stored 5 patterns in both symmetric and canonical schemes, and generate new ones. The randomly generated 5 patterns and their corresponding symmetric and canonical patterns are shown in Figure 5. Figure 6 shows the results of symmetric and canonical orthogonalizations when the new pattern is added.



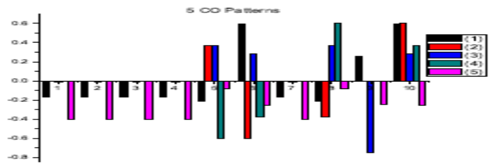


Figure 5: 5 randomly generated patterns and their corresponding symmetrically and canonically orthogonalized patterns.

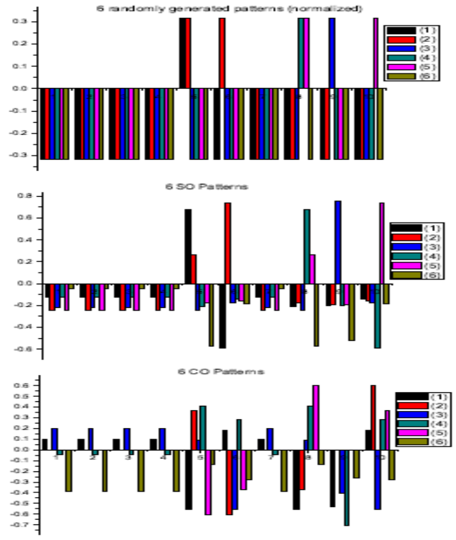


Figure 6: 6 randomly generated patterns and their corresponding symmetrically and canonically orthogonalized patterns.

We have added the 6^{th} pattern, with 70% dissimilarity to the 5^{th} pattern and all the six patterns are orthogonalized using the symmetric and canonical orthogonalizations. We have projected the normalized input patterns onto the symmetric and canonical orthonormal patterns to check how they have changed in a gross sense. In the case of symmetric orthogonalization, the sum of squared projections of any of the normalized input patterns onto any of the symmetric orthonormal patterns is equal to *unity*. When all the normalized input patterns are projected individually onto the canonical patterns picked up one by one, then the sum of squared projections is found to be the largest on the canonical pattern which corresponds to the largest eigenvalue.

Starting from 5 patterns we have added one by one 5 new patterns. Quite surprisingly it turns out that the 2*nd* canonical pattern carries the largest eigenvalue and therefore captures the maximum projection of all the raw patterns. Each of 5 patterns changes as the sixth pattern is added and they are subjected to canonical and symmetric orthogonalizations respectively. The changes are generally vigorous in the case of canonical orthogonalization (even signs change) whereas in the case of symmetric orthogonalization the changes caused by adding the 6^{th} pattern are moderate (the sign does not change).

VI. CONCLUSION

Our numerical simulations of the neural networks model due to Hopfield by incorporating the two democratic orthogonalization schemes could pave the way for studying cognitive learning and memory under special circumstances. The comparison with the earlier work on Hopfield network with Gram-Schmidt orthogonalization incorporated shows significant departure in the results. In the broad terms this work shows that the brain might handle sequential learning through Gram-Schmidt, and the storage of information thus acquired, very differently from the way it might process the information acquired in bunches (i.e. simultaneous intake of a set of disparate information), for example in episodic memories. A lot of work needs to be done in close association with neuroscientists to unravel some mysteries of how the brain functions.

VII. ACKNOWLEDGMENT

This work is financially supported from the grant (No: F. 37-557/2009 (SR)) of University Grants Commission (UGC), New Delhi, India. The author would like to thank the UGC for the financial support.

REFERENCES

[1] Pastur, L.A., Figotin, A.L., (1977). Exactly soluble model of a spin-glass. Soviet J. Low Temperature Phys. 3 (6), 378383.
[2] Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proceedings of the national academy of sciences, 79(8), 2554-2558, doi:10.1073/pnas.79.8.2554.
[3] Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. Proceedings of the national academy of sciences, 81(10), 3088-3092.
[4] Amit, D. J. (1992). Modelling brain function: The world of attractor neural networks. Cambridge University Press.
[5] Hertz, J., Krogh, A., Palmer, R. G. (1982). Introduction to the theory of neural networks, Cambridge University Press, Cambridge.
[6] Peretto, P. (1992). An introduction to the modeling of neural networks (Vol. 2). Cambridge University Press.
[7] The mathematical computation of this value is given in D. J. Amit, H. Gutfreund, H. Sompolinski; (1985). Phys. Rev. Lett. 55, 1530-1533.
[8] McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. The psychology of learning and motivation, 24(109-165), 92.
[9] French, R. M. (1999). Catastrophic forgetting in connectionist networks. Trends in cognitive sciences, 3(4), 128-135.
[10] Ratcliff, R. (1990). Connectionist models of recognition memory: constraints imposed by learning and forgetting functions. Psychological review, 97(2), 285-308.
[11] Srivastava, V., Edwards, S. F. (2000). A model of how the brain discriminates and categorises. Physica A: Statistical Mechanics and its Applications, 276(1), 352-358.
[12] Srivastava, V., Parker, D. J., & Edwards, S. F. (2008). the nervous system might orthogonalizeto discriminate. Journal of theoretical biology, 253(3), 514-517.
[13] Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven month-old infants. Science, 283(5398), 77-80.
[14] Lee, S. H., Blake, R. (1999). Visual form created solely from temporal structure. Science, 284(5417), 1165-1168.
[15] Rawlins, J. N. P. (1999). Neurobiology: A place for space and smells. Nature, 397(6720), 561-563.
[16] Wood, E. R., Dudchenko, P. A., & Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity. Nature, 397(6720), 613-616.
[17] McIntosh, A. R., Rajah, M. N., & Lobaugh, N. J. (1999). Interactions of prefrontal cortex in relation to awareness in sensory learning. Science, 284(5419), 1531-1533.
[18] Naidu, A. R., & Srivastava, V. (2004). Löwdin's canonical orthogonalization: Getting round the restriction of linear independence. International journal of quantum chemistry, 99(6), 882-888.
[19] Ramesh Naidu Annavarapu., (2013), Singular Value Decomposition and the Centrality of Löwdin Orthogonalizations, American Journal of Computational and Applied Mathematics, 3(1), 33-35. doi:10.5923/j.ajcam.20130301.06.