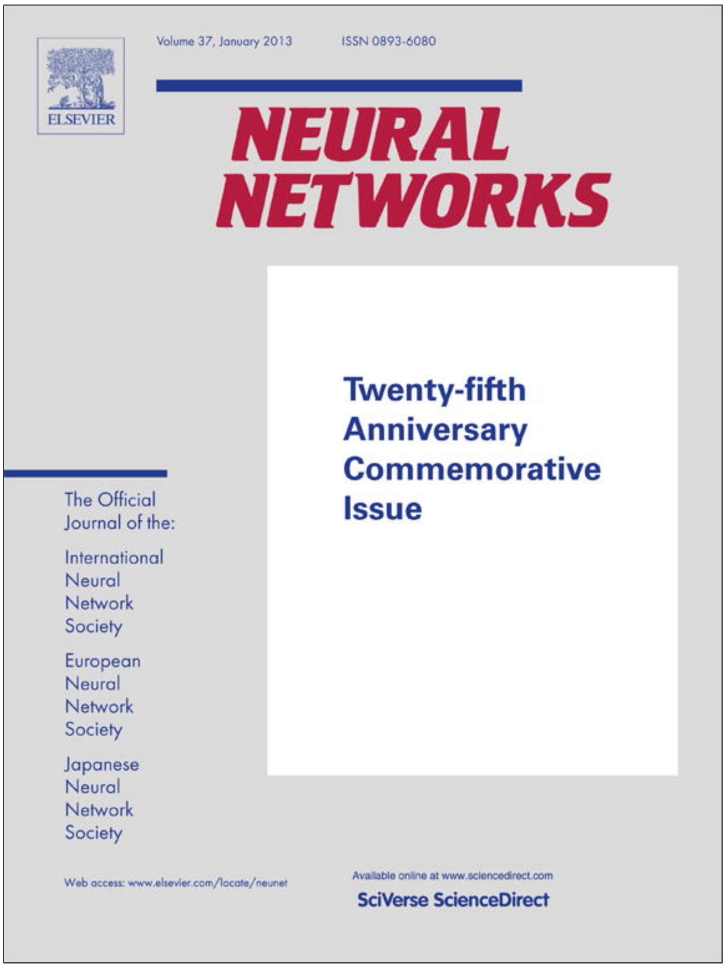
ELSEVIER
Volume 37, January 2013
ISSN 0893-6080
NEURAL NETWORKS
Twenty-fifth Anniversary Commemorative Issue
The Official Journal of the:
International Neural Network Society
European Neural Network Society
Japanese Neural Network Society
Web access: www.elsevier.com/locate/neunet
Available online at www.sciencedirect.com
SciVerse ScienceDirect

Neural Networks 37 (2013) 48–51

Contents lists available at SciVerse ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Dreaming of mathematical neuroscience for half a century

Shun-ichi Amari

RIKEN Brain Science Institute, Japan

ARTICLE INFO

Keywords:
Mathematical neuroscience
Statistical neurodynamics
Neural learning
Dynamics of neural field

ABSTRACT

Theoreticians have been enchanted by the secrets of the brain for many years: how and why does it work so well? There has been a long history of searching for its mechanisms. Theoretical or even mathematical scientists have proposed various models of neural networks which has led to the birth of a new field of research. We can think of the 'pre-historic' period of Rashevski and Wiener, and then the period of perceptrons which is the beginning of learning machines, neurodynamics approaches, and further connectionist approaches. Now is currently the period of computational neuroscience. I have been working in this field for nearly half a century, and have experienced its repeated rise and fall. Now having reached very old age, I would like to state my own endeavors on establishing mathematical neuroscience for half a century, from a personal, even biased, point of view. It would be my pleasure if my experiences could encourage young researchers to participate in mathematical neuroscience.

1. Introduction

Science begins with observations of phenomena. They are then categorized for systematic descriptions and consolidated as knowledge. The final stage is to search for the principles underlying the phenomena that give rise to theoretical science. Mathematical equations are often used for elucidating principles, as is typically seen in physics. We wonder if it is possible to establish mathematical neuroscience. I have been dreaming of this for nearly five decades, and it is still developing.

The brain has emerged via a long history of evolution, which is different from that of universal physical phenomena. Even though the principles of complex phenomena in everyday life are hidden in the world of physics, one may think of extreme situations where the fundamental principles can be directly observed. For example, the principles of Newtonian mechanics are directly observed in celestial motion. Extremely low temperatures are useful for statistical physics. However, the brain is a living organ that cannot survive in extreme situations. As it is a highly complex system, we wonder whether it really is possible to discover mathematical principles.

These arguments look so reasonable that I do not believe that the detailed functions of the brain can be described by simple mathematical equations. Information in the brain is widely distributed over networks of neurons and processed by parallel dynamics. It has learning and self-organizing capabilities. I believe that there are fundamental principles based on which outstanding information processing takes place due to parallel distributed dynamics with self-organization. As our brains have implemented such principles throughout their extremely long history of evolution, it is not easy to understand the principles.

We need to construct ideal models of information processing to elucidate such principles, in which parallel dynamics takes place and information is widely distributed. Mathematical analysis of the models would hopefully make it possible to understand the underlying principles. Although the actual brain is different from simple models, I believe that the same principles would work even in very complex actual brains. As simple forms of principles are insufficient to understand the real brain, we also need to find how the principles are substantialized in the actual brain. Here comes the role of computational neuroscience, which seeks for computational aspects of realistic neural networks.

My research began nearly fifty years ago when I was striving to discover the principles of the brain mathematically by using typical simple models of neural networks. Allow me to take the liberty of retrospectively describing my own research on mathematical neuroscience. Some topics might have been superseded, having been replaced by new developments, but some still remain classics.

2. Multilayer perceptrons as learning machines

When I first read a book on perceptrons, entitled "Principles of Neurodynamics" (Rosenblatt, 1961), I wondered why the hidden layer neurons could not be modified. Obviously, the capabilities of multilayer perceptrons would be greater if all the neurons were modifiable. To overcome the difficulty in learning hidden neurons, I considered using nonlinear analog neurons, instead of McCulloch–Pitts neurons, to model multilayer perceptrons (Amari, 1967). I then defined a differentiable loss function (error function)

E-mail address: amari@brain.riken.jp.

0893-6080/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2012.09.014

for a general learning machine including a modifiable vector parameter $\boldsymbol{\theta}$. When an input $\boldsymbol{x}$ is given, a learning machine with parameter $\boldsymbol{\theta}$ processes it. A loss function is defined, for example, by the squared error

$$l(\boldsymbol{x}, \boldsymbol{\theta}) = \frac{1}{2} |y - f(\boldsymbol{x}, \boldsymbol{\theta})|^2 , \tag{1}$$

where y is the required output and $f(\boldsymbol{x}, \boldsymbol{\theta})$ is the actual output from the machine with parameter $\boldsymbol{\theta}$. A key point of using analog neurons is that it makes f differentiable with respect to $\boldsymbol{\theta}$. I proposed a new learning algorithm, which I called the stochastic gradient descent method, defined by

$$\boldsymbol{\theta}_{t+1} = -\varepsilon \nabla l(\boldsymbol{x}_t, \boldsymbol{\theta}_t) , \tag{2}$$

since $\boldsymbol{x}_t$ is a sequence of randomly chosen inputs. The paper appeared in IEEE Transactions on Computers (Amari, 1967).

I was surprised when I read a galley proof of the well-known back-propagation paper (Rumelhart, McClelland, & Hinton, 1986) twenty years later. The algorithm (2) was called the generalized delta rule in that paper and the beautiful name "back-propagation" was given to the learning algorithm. There were exciting new findings in the paper such as back-propagating errors together with lots of computer simulations. I could not present computer simulated results in my previous paper, because computers were very limited in Japan in the early sixties. Instead, my paper discussed analysis of various aspects of the dynamics of learning, using a stochastic approximation technique, and the dynamical behaviors of learning were mathematically analyzed. The trade-off between the speed and accuracy of learning was clearly demonstrated. Most of my results were rediscovered later by a number of researchers (e.g., Heskes and Kappen (1991)). However, I have found that they are insufficient and deeper analysis is needed because the multilayer perceptron provides a "singular" statistical model. As I will touch upon later, information geometry is necessary to study its dynamical behavior.

It was a surprise that my theory was introduced in a Russian book by Tsypkin (1973). The theory of machine learning had been well developed in Russia, which was then isolated from the Western world. A remark was made that no such research could be found at that time in the Western world except for Japan (Vladimir Vapnik, personal communication).

3. Statistical neurodynamics

It was in the late sixties and through the seventies that I devoted myself to a program of research on the dynamical behaviors of neural networks and learning. It consisted of statistical neurodynamics (randomly connected networks), neural field theory, the associative memory model, learning, and self-organization. I studied the dynamical behaviors of recurrently connected neural networks with random connections in 1969. The dynamical equation I used is the same as that for the Hopfield network

$$\tau_i \frac{du_i(t)}{dt} = -u_i + \sum_j w_{ij} f\left[u_j(t)\right] + s_i, \tag{3}$$

where $u_i(t)$ represents the average activities of the i-th neuron or i-th neuron pool, w_{ij} is the connection weight from neuron j to neuron i, and f is a sigmoid function. I proved that randomly connected networks can have monostable and multi-stable behaviors depending on connection weights and external stimuli. Moreover, I proved that a network consisting of excitatory and inhibitory neurons has stable oscillatory behaviors (the discrete-time case in Amari (1971) and the continuous-time case in Amari (1972a)). The equations are

$$\tau_E \frac{du_E(t)}{dt} = -u_E + w_{EE} f\left[u_E\right] - w_{EI} f\left[u_I\right] + s_E, \tag{4}$$

$$\tau_I \frac{du_I(t)}{dt} = -u_I + w_{IE} f\left[u_E\right] - w_{II} f\left[u_I\right] + s_I. \tag{5}$$

The same model was independently proposed (Wilson & Cowan, 1972). The only difference was that their activation function f was originally not a monotone sigmoid function, while I used a sigmoid function from the beginning. The model with sigmoid f is now widely known as the Wilson–Cowan oscillator. My paper (Amari, 1971) was submitted to the Proceedings of IEEE in 1969, but it took two years for it to be published. I heard that an American reviewer had handed the manuscript to a graduate student who soon dropped out, so that it was lost and not reviewed for over a year. In addition to neural oscillators, it is now well known that chaotic behaviors exist and are one of the fundamental characteristics of these types of networks. However, chaos was not popular at that time.

I naturally had an interest in the foundations of statistical neurodynamics. This was a topic discussed by a Russian researcher Rozonoer (1969), where the validity of annealed approximation was the main subject. I undertook mathematical studies (Amari, 1974; Amari, Yoshida, & Kanatani, 1977), although the problem still remains unsolved. In connection with this, I studied microscopic dynamics of a simple random network of binary neurons, and found, to my amazement, that its state transition graph had scale-free properties (Amari, 1974, 1990). Each state had a unique next state so that it had only one outgoing branch. This implied that the average numbers of incoming branches (or the parent states) of a state was also one. However, its variance diverged to infinity as the number, n, of neurons increased. This suggested that the incoming branches had a long-tailed distribution subject to a power law. We are now studying this topic further, elucidating differences between Boolean logic random networks (Kauffman, 1969) and random majority decision (neural) networks (Amari, 1974). Our results demonstrated the origin of speed and robustness in biological decision systems including neural networks (Amari, Ando, Toyoizumi, & Masuda, submitted for publication).

A neural field is a natural extension of neuron pools, which has the topology of a continuous field. Wilson and Cowan (1973) carried out pioneering work on its dynamics. Being inspired by their work, I provided an exact mathematical theory of neural fields by using the Heaviside activation function (Amari, 1977). This reduced the dynamics of a field to that of the boundaries of excitation, so that we could construct an exact theory. The existence and stability of a bump solution as well as a traveling bump were established. A large number of papers on neural fields are currently appearing and many new interesting phenomena have been discovered. This is because the neural field is useful not only for analyzing working memory in the brain, but also for explaining psychological phenomena and robot navigation.

4. Associative memory

Four papers appeared in 1972 that concerned correlation-type associative memory (Kohonen, Anderson, Nakano, and Amari). Some were linear and some used ternary neurons. The model that was proposed in Amari (1972b) was exactly the same as the so-called Hopfield model of associative memory (Hopfield, 1982):

$$x_i(t+1) = \operatorname{sgn}\left\{\sum_j w_{ij} x_j(t)\right\}, \tag{6}$$

$$w_{ij} = \sum_\mu s_i^\mu s_j^\mu , \tag{7}$$

where $\boldsymbol{x} = (x_i)$ represents binary state vectors ($x_i = \pm 1$), and $\boldsymbol{s}^\mu = \left(s_i^\mu\right)$, $\mu = 1, \ldots, m$, are m patterns to be memorized. The patterns are memorized as equilibrium states of network dynamics. I mathematically analyzed its behaviors. Moreover, I introduced an extended model for dynamically memorizing and recalling sequences of patterns. The connection matrix $\left(w_{ij}\right)$ is asymmetric in that model

$$w_{ij} = \sum_\mu s_i^\mu s_j^{\mu-1}, \tag{8}$$

where $S = \left(\boldsymbol{s}^1, \ldots, \boldsymbol{s}^m\right)$ is a sequence of patterns. This model memorizes sequences of patterns and recalls them sequentially.

The contribution of Hopfield (1982) was overwhelming, because he introduced the notion of memory capacity for the first time by using random patterns. He demonstrated by computer simulations that the model could memorize about $0.14n$ patterns in a model of n neurons. Because of this intriguing finding, statistical physicists entered this field and made great theoretical contributions by using the replica method of calculating the expectation of log probability. Sompolinski made a remark in one of his papers that the same model was proposed earlier by Amari but the contribution by Hopfield was so great that we called it the Hopfield model.

The method of statistical physics is capable of analyzing the capacity of the model. The replica method is further applied to learning machines and information systems. However, it is mostly used to study equilibrium states, and lacks the dynamical process of recalling. Consequently, I proposed a dynamical equation of recalling processes by introducing new macroscopic quantities (Amari & Maginu, 1988). Although the theory is not exact but approximate, it does explain the interesting phenomena of recalling processes such as the hysteresis of recalling and the fractal structure of the basins of attraction. The theory was further extended by Okada (1995). Ton Coolen extended the statistical-mechanical method (Coolen & Sherrington, 1993; During, Coolen, & Sherrington, 1998), and told me that the Amari–Maginu method gives the exact capacity in the case of asymmetric sequence recalling.

5. Learning and self-organization

In my quest for principles of learning systems, I developed a general theory of learning and self-organization (Amari, 1977), and established a unified framework to neural learning and self-organization. It elucidated various types of neural learning, which were both supervised and unsupervised. For example, it was pointed out that neural learning is capable of principal components analysis, which was later analyzed independently and intensively by Oja (1982). My theory of self-organization, inspired by the model of von der Malsburg (1973), pointed out an important role of modifiable inhibitory synapses for the first time. I was surprised to see that the mechanism of self-organization was logically the same as the Bienenstock–Cooper–Munro (BCM) model of the sliding threshold (Bienenstock, Cooper, & Munro, 1982) that appeared later. The only difference is that they used a modifiable (sliding) threshold, while the modifiable inhibitory synapse played the same role in my paper. A detailed mathematical expression was given to the dynamics of self-organization. This made it possible to develop a theory of self-organization of neural fields, proposed by Willshaw and von der Malsburg (1976) that integrated the dynamics of neural fields with self-organization (Takeuchi & Amari, 1979). The formation of a topological map was later further developed by Kohonen (1982).

Other topics in my study on machine learning include the natural gradient method for multilayer perceptrons (Amari, 1998) based on information geometry (Amari & Nagaoka, 2000). The singular structure of the parameter space of multilayer perceptrons was analyzed in particular detail (Amari, Park, & Ozeki, 2006; Wei & Amari, 2008), demonstrating why the retardation of backpropagation learning (plateau phenomenon) appeared and how effective the natural gradient learning method was for overcoming this difficulty. It complements the dynamics of learning studied previously. Machine learning is an interesting subject arising from neural networks. The relation between training error and generalization error has been analyzed in detail (Amari, Fujita, & Shinomoto, 1992). Generalization capabilities depend on the geometry of separating surfaces in the version space of a learning machine (Amari, 1993), which should be further explored.

All these topics form the basic constituents of mathematical neuroscience as well as of machine learning. Obviously, new important contributions have appeared, such as spike-timing dependent plasticity (STDP), support vector machines (SVMs), and methods to boost machine learning, toward which I have made little contribution. I hope all of these will be integrated to formulate mathematical principles of learning.

6. Independent component analysis

The mandatory retirement age of the University of Tokyo was 60 (it has now been extended to 65). I was invited to organize a theoretical group for Brain Science at RIKEN, which is a governmental institute. I invited Andrzej Cichocki, who is an expert on signal processing, in particular on independent component analysis (ICA) to join RIKEN. He showed me a paper by Bell and Sejnowski (1995). This led me to consider the natural gradient method of learning (Amari, 1998; Amari, Cichocki, & Yang, 1996). I enjoyed a new topic of research, because I was able to apply various mathematical tools such as information geometry, statistical inference, Lie group theory, and even non-holonomic analysis. I had thought when I was young that few creative ideas would emerge from those past their sixties. Therefore, I was glad to see one could still make meaningful contributions past their sixties. My paper on a nonparametric statistical method of ICA (Amari & Cardoso, 1997) was given a best paper award by the IEEE Signal Processing Society. The natural gradient method (Amari, 1998) is a consequence of information geometry and can be applied to various other problems. I mentioned that it was applied to machine learning, particularly to multilayer perceptrons, to provide excellent convergence of learning. It was also applied to reinforcement learning under the name of the policy natural gradient. See, e.g., Amari, Kurata, and Nagaoka (1992), Ikeda, Tanaka, and Amari (2004) and Murata, Takenouchi, Kanamori, and Eguchi (2004) for applications of information geometry to machine leaning.

7. Information geometry of neuronal spikes

I proposed information geometry in my forties in which I studied the invariant geometrical structure of a manifold of probability distributions (Amari & Nagaoka, 2000). It used modern differential geometry in which I introduced a new mathematical concept of dual affine connections. It was a different field from neural networks, but I dreamed of applying it even to neural networks. I found many outstanding young researchers, who were better than I, and it appeared that the information-geometrical method might distinguish me from more competent researchers. Indeed, I proposed the natural gradient method based on it (Amari, 1998). I wanted to discover further applications of information geometry to neuroscience.

Neural activities are often stochastic and carried by spikes of neurons. It is important to analyze the joint probability distribution of spikes. Spike sequences include various information such as firing rates, pairwise correlations, and further intrinsic higher order correlations. We need to clearly separate the effects of firing rates, pairwise correlations and higher order correlations. The firing rates of neurons can easily be estimated from data. The correlations of two neuronal spikes are usually calculated by covariances of spikes. However, the two parameters of firing rates and covariance are not orthogonal from the point of view of information geometry (Amari, 2009). This implies that, when firing rates change, the covariances change even if the mechanism for their interactions does not change. Therefore, covariance is not a good measure of the mutual interaction of two neurons. An orthogonal measure of higher-order correlations is also needed.

I proposed orthogonal parameters in a general hierarchical model of probability distributions (Amari, 2001). Here, the orthogonality of these parameters is defined in terms of the Riemannian geometry of Fisher information (Nakahara & Amari, 2002). We have also proposed a mechanism for the emergence of higher order correlations (Amari, Nakahara, Wu, & Sakai, 2003). However, it is not realistic to search for all higher order correlations from experimental data. I am searching for a method of sparse signal analysis that can be applied to this problem.

Information geometry is now one of the most important methods of analyzing neuronal spike data, see, e.g., Miura, Okada, and Amari (2006). When parts of data are missing or when the decoding schemes of spikes use an unfaithful model, we want to know how much information loss is caused (Oizumi, Okada, & Amari, 2011). I am currently engaged in this topic.

8. Concluding remarks

I have taken a retrospective glance covering half a century of my research life. I belong to a miracle generation, because the Japanese economy had dramatically developed through this period. I came from an underprivileged background and never thought of having opportunities of visiting foreign countries, driving a car, or speaking in English. Japan was not economically developed at that time (in the sixties), and I could not even use airmail to submit papers to international journals. Sea mail to America took nearly a month one way. There were no opportunities at that time for me to visit foreign countries to participate in conferences. However, I was lucky to be able to devote myself to research in isolated and serene academic circumstances. The impossible then changed to the possible as time went on. I have led a fortunate life by focusing on research, and I am still working in academia. I am still dreaming of establishing the field of mathematical neuroscience, but young researchers should go further beyond my classic research, which I commend passionately. I am looking forward to hearing about their further developments toward mathematical neuroscience.

My dream would come true if my personal message could encourage young researchers in this emerging field, particularly those from developing countries.

References

Amari, S. (1967). Theory of adaptive pattern classifiers. *IEEE Transactions*, *EC-16*, 299–307.

Amari, S. (1971). Characteristics of randomly connected threshold-element networks and network systems. *Proceedings of the IEEE*, *59*(1), 35–47.

Amari, S. (1972a). Characteristics of random nets of analog neuron-like elements. *IEEE Transactions on Systems, Man and Cybernetics*, *SMC-2*(5), 643–657.

Amari, S. (1972b). Learning patterns and pattern sequences by self-organizing nets of threshold elements. *IEEE Transactions on Computers*, *C-21*(11), 1197–1206.

Amari, S. (1974). A method of statistical neurodynamics. *Kybernetik*, *14*, 201–215.

Amari, S. (1977). Neural theory of association and concept-formation. *Biological Cybernetics*, *26*, 175–185.

Amari, S. (1990). Mathematical foundations of neurocomputing. *Proceedings of the IEEE*, *78*(9), 1443–1463.

Amari, S. (1993). A universal theorem on learning curves. *Neural Networks*, *6*(2), 161–166.

Amari, S. (1998). Natural gradient works efficiently in learning. *Neural Computation*, *10*, 251–276.

Amari, S. (2001). Information geometry on hierarchy of probability distributions. *IEEE Transactions on Information Theory*, *47*(5), 1701–1711.

Amari, S. (2009). Measure of correlation orthogonal to change in firing rate. *Neural Computation*, *21*(4), 960–972.

Amari, S., Ando, H., Toyoizumi, T., & Masuda, N. (2012). State concentration exponent as a measure of quickness in Kauffman-type networks (submitted for publication).

Amari, S., & Cardoso, J.-F. (1997). Blind source separation—semiparametric statistical approach. *IEEE Transactions on Signal Processing*, *45*(11), 2692–2700.

Amari, S., Cichocki, A., & Yang, H. H. (1996). A new learning algorithm for blind signal separation. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), *Advances in neural information processing systems, vol. 8* (pp. 757–763).

Amari, S., Fujita, N., & Shinomoto, S. (1992). Four types of learning curves. *Neural Computation*, *4*, 605–618.

Amari, S., Kurata, K., & Nagaoka, H. (1992). Information geometry of Boltzmann machines. *IEEE Transactions on Neural Networks*, *3*(2), 260–271.

Amari, S., & Maginu, K. (1988). Statistical neurodynamics of associative memory. *Neural Networks*, *1*(1), 63–73.

Amari, S., & Nagaoka, H. (2000). *Methods of information geometry*. AMS and Oxford University Press.

Amari, S., Nakahara, H., Wu, S., & Sakai, Y. (2003). Synchronous firing and higher-order interactions in neuron pool. *Neural Computation*, *15*, 127–142.

Amari, S., Park, H., & Ozeki, T. (2006). Singularities affect dynamics of learning in neuromanifolds. *Neural Computation*, *18*, 1007–1065.

Amari, S., Yoshida, K., & Kanatani, K. (1977). A mathematical foundation for statistical neurodynamics. *SIAM Journal on Applied Mathematics*, *33*, 95–126.

Bell, A. J., & Sejnowski, T. J. (1995). An information maximisation approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1129–1159.

Bienenstock, E. L., Cooper, L., & Munro, P. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, *2*(1), 32–48.

Coolen, A. C. C., & Sherrington, D. (1993). Dynamics of fully connected attractor neural networks near saturation. *Physical Review Letters*, *71*, 3886–3889.

During, A., Coolen, A. C. C., & Sherrington, D. (1998). Phase diagram and storage capacity of sequence storing neural networks. *Journal of Physics*, *A31*, 8607.

Heskes, T., & Kappen, H. (1991). Learning processes in neural networks. *Physical Review A*, *44*, 2718–2726.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, *79*, 2554–2558.

Ikeda, S., Tanaka, T., & Amari, S. (2004). Stochastic reasoning, free energy, and information geometry. *Neural Computation*, *16*(9), 1779–1810.

Kauffman, S. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, *22*, 437–467.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*, 59–69.

Miura, K., Okada, M., & Amari, S. (2006). Estimating spiking irregularities under changing environments. *Neural Computation*, *18*(10), 2359–2386.

Murata, N., Takenouchi, T., Kanamori, T., & Eguchi, S. (2004). Information geometry of U-Boost and Bregman divergence. *Neural Computation*, *16*(7), 1437–1481.

Nakahara, H., & Amari, S. (2002). Information-geometric measure for neural spikes. *Neural Computation*, *14*, 2269–2316.

Oizumi, M., Okada, M., & Amari, S. (2011). Information loss associated with imperfect observation and mismatched decoding. *Frontiers in Computational Neuroscience*, *5*(9), 1–13.

Oja, E. (1982). A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, *15*, 267–273.

Okada, M. (1995). A hierarchy of macrodynamical equations for associative memory. *Neural Networks*, *8*, 833–838.

Rosenblatt, F. (1961). *Principles of neurodynamics*, Spartan.

Rozonoer, L. I. (1969). Random logical nets I–III. *Avtomatika i Telemekhanika*, *5*, *6*, *7*, pp. 137–147, 99–109, 127–136.

Rumelhart, D., McClelland, J., & Hinton, J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, & J. L. McClelland (Eds.), *Parallel distributed processing: explorations in the microstructure of cognition, vol. 1*. MIT Press.

Takeuchi, A., & Amari, S. (1979). Formation of topographic maps and columnar microstructures. *Biological Cybernetics*, *35*, 63–72.

Tsypkin, Y. Z. (1973). *Foundations of the theory of learing systems*. Academic Press (English translation).

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, *14*, 85–100.

Wei, H., & Amari, S. (2008). Dynamics of learning near singularities in radial basis function networks. *Neural Networks*, *21*, 989–1005.

Willshaw, D. J., & von der Malsburg, C. (1976). How patterned neural connections can be set up by self-organization. *Proceedings of the Royal Society*, *B-194*, 431–445.

Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, *12*, 1–24.

Wilson, H. R., & Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, *13*, 55–80.