Complexity of Learning for Networks of Spiking Neurons with Nonlinear Synaptic Interactions*

Michael Schmitt

Lehrstuhl Mathematik und Informatik, Fakultät für Mathematik
Ruhr-Universität Bochum, D–44780 Bochum, Germany
http://www.ruhr-uni-bochum.de/lmi/mschmitt/
mschmitt@lmi.ruhr-uni-bochum.de

Abstract. We study model networks of spiking neurons where synaptic inputs interact in terms of nonlinear functions. These nonlinearities are used to represent the spatial grouping of synapses on the dendrites and to model the computations performed at local branches. We analyze the complexity of learning in these networks in terms of the VC dimension and the pseudo dimension. Polynomial upper bounds on these dimensions are derived for various types of synaptic nonlinearities.

1 Introduction

In many paradigms of neural computation the interaction of synaptic inputs is modeled in terms of linear operations. This is evident for standard neuron models such as the threshold, the sigmoidal, and the linear unit, but it also holds, in most cases, for the biologically closer leaky integrate-and-fire neurons and the spiking neuron models. Linearity is known to be sufficient for capturing the passive properties of the dendritic membrane where synaptic inputs occur in the form of currents that are combined using a summing operation. In neurobiology, however, it is well established that synaptic inputs can interact nonlinearly, for instance, when the synapses are colocalized on patches of dendritic membrane with specific properties (see, e.g., Koch [3]). In contrast to linear interactions, nonlinearities can reflect the spatial grouping of synapses (the "synaptic clustering") on the dendritic tree and the computations performed at local branches.

In this paper we study models for spiking neurons that use nonlinearities to combine synaptic inputs. We analyze the complexity of learning in networks of these model neurons in terms of the Vapnik-Chervonenkis (VC) dimension and the pseudo dimension. We establish bounds for these dimensions that can be used to derive results on the sample complexity and generalization capability of these networks. (For the relationship of the VC dimension and the pseudo dimension to the complexity of learning see, e.g., Anthony and Bartlett [1].)

* Work supported by the ESPRIT Working Group in Neural and Computational Learning II, NeuroCOLT2, No. 27150. This is a longer (8-page) version of a paper that appears in the Proceedings of the International Conference on Artificial Neural Networks ICANN 2001.

2 Spiking Neurons with Nonlinear Synaptic Interactions

In a network of *spiking neurons* each node v receives inputs in the form of spikes through its synaptic connections from other nodes. Assume that v has n input connections, each of them being characterized by a synaptic weight $w_i \in \mathbb{R}$ and a transmission delay $d_i \in \mathbb{R}^+$, where $\mathbb{R}^+ = \{x \in \mathbb{R} : x \geq 0\}$. If the neuron presynaptic to connection i emits a spike at time $t_i \in \mathbb{R}^+$, this generates in v a pulse described by the function $t \mapsto h_i(t - t_i)$ where $h_i : \mathbb{R}^+ \to \mathbb{R}$ is defined as

$$h_i(t) = \begin{cases} w_i & \text{if } d_i \leq t < d_i + 1 \ , \\ 0 & \text{otherwise} \ . \end{cases}$$

Hence, the pulse has unit duration and its height corresponds to the strength, or efficacy, of the synapse. We consider d_i and w_i, for $i = 1, \ldots, n$, as adjustable parameters of the neuron.

A *synaptic cluster* of v is a subset $I \subseteq \{1, \ldots, n\}$ of its synapses spatially grouped together on the dendritic tree. They interact nonlinearly where the nonlinearity is specified for each subset $\{i_1, \ldots, i_l\} \subseteq I$ by a rational function $f_{\{i_1,\ldots,i_l\}}$ in l variables. We call $f_{\{i_1,\ldots,i_l\}}$ the *synaptic interaction* of $\{i_1, \ldots, i_l\}$. A synapse may participate in more than one cluster. Let $F \subseteq \{1, \ldots, n\}$ be the set of synapses that receive a spike during the computation of v and define $J_I : \mathbb{R}^+ \to 2^I$ as

$$J_I(t) = \{i \in I \cap F : h_i(t - t_i) \neq 0\} \ ,$$

that is, $J_I(t)$ is the set of synapses in cluster I that are simultaneously active at time t. The interaction of the synapses in I is then modeled by the function $M_I : \mathbb{R}^+ \to \mathbb{R}$ with

$$M_I(t) = f_{\{i_1,\ldots,i_l\}}(w_{i_1}, \ldots, w_{i_l}), \quad \text{where } \{i_1, \ldots, i_l\} = J_I(t) \ .$$

Assume that v has k synaptic clusters $I_1, \ldots, I_k$. They interact linearly to yield the *membrane potential* $P : \mathbb{R}^+ \to \mathbb{R}$ defined by

$$P(t) = \sum_{j=1}^{k} M_{I_j}(t) \ .$$

Neuron v fires when the membrane potential reaches the threshold θ, that is, the *firing time* t_v of v is defined by $t_v = \min\{t : P(t) \geq \theta\}$. If t_v is undefined then v does not fire. The *networks of spiking neurons* that we study here use temporal coding for computations, that is, they encode information in the timing of single firing events. Precisely, if neuron v fires at time t_v this is assumed to encode the real number t_v. We consider feedforward networks with a given number of input nodes and one output node. Input vectors enter the network in terms of firing times of the input nodes. The output value of the network is the firing time of the output node. If the output node does not fire, the network output is defined to be 0. The *depth* of the network is the length of the longest path from an input node to the output node.

The model introduced here generalizes that of [5] by allowing nonlinear interactions of synaptic inputs in terms of clusters that represent the spatial grouping of synapses. Models of spiking neurons with linear interactions are discussed, e.g., in [2, 4]. We have included synaptic nonlinearities to take into account the well established neurobiological finding that the nature of the interaction of synapses colocalized on the dendritic tree is essentially nonlinear (see [3, Chapter 5]). The nonlinearities we employ here are based on arithmetic operations that can easily be implemented in analog VLSI circuits and thus be used for hardware designs of pulsed neural networks such as described in [6].

In this paper we calculate bounds on the VC dimension and the pseudo dimension of the networks introduced above. These dimensions are defined as follows (see also [1]): A class $\mathcal{F}$ of $\{0,1\}$-valued functions in n variables is said to *shatter* a set of vectors $S \subseteq \mathbb{R}^n$ if $\mathcal{F}$ induces all dichotomies on S (that is, if for every partition of S into two disjoint subsets (S_0, S_1) there is some $f \in \mathcal{F}$ satisfying $f(S_0) \subseteq \{0\}$ and $f(S_1) \subseteq \{1\}$). The *Vapnik-Chervonenkis (VC) dimension* of $\mathcal{F}$ is the cardinality of the largest set shattered by $\mathcal{F}$. If $\mathcal{F}$ is a class of real-valued functions in n variables, the *pseudo dimension* of $\mathcal{F}$ is defined as the VC dimension of the class $\{g : \mathbb{R}^{n+1} \to \{0,1\} \mid \exists f \in \mathcal{F}\, \forall x \in \mathbb{R}^n \forall y \in \mathbb{R} : g(x,y) = \mathrm{sgn}(f(x) - y)\}$ (where $\mathrm{sgn}(z) = 1$ if $z \geq 0$, and 0 otherwise). The VC dimension (pseudo dimension) of a network of spiking neurons is defined to be the VC dimension (pseudo dimension) of the set of functions computed by the network with all possible assignments of values to its adjustable parameters, i.e. its delays, weights, and thresholds. To define the VC dimension for networks with real-valued output we assume that the output values are mapped to $\{0,1\}$ using some (non-trivial) threshold. Clearly, the VC dimension of a network is not larger than its pseudo dimension.

3 Bounded Depth Networks

In the following we provide an upper bound on the pseudo dimension for networks of spiking neurons that is almost linear in the number of network parameters (delays, weights, and thresholds) and the network depth.

Theorem 1. *Suppose $\mathcal{N}$ is a network of spiking neurons with W parameters and depth D where each neuron has at most k synaptic clusters and p is a bound on the degree of the synaptic interactions. Then the pseudo dimension of $\mathcal{N}$ is at most $O(WD\log(WDkp))$.*

The main idea of the proof is first to derive from $\mathcal{N}$ and a given set of input vectors a set of rational functions in the parameter variables of $\mathcal{N}$. Next, the connected components into which the parameter domain is partitioned by the zero-sets of these rational functions are considered. The construction guarantees that the number of dichotomies that $\mathcal{N}$ induces on the set of input vectors is not larger than the number of connected components generated by these functions.

Thus a bound on the number of connected components also limits the number of dichotomies. We require Definitions 7.4 and 7.5 from [1]: A set $\{f_1, \ldots, f_k\}$ of differentiable real-valued functions on $\mathbb{R}^d$ is said to have *regular zero-set intersections* if for every non-empty set $\{i_1, \ldots, i_l\} \subseteq \{1, \ldots, k\}$ the Jacobian (that is, the matrix of the partial derivatives) of $(f_{i_1}, \ldots, f_{i_l}) : \mathbb{R}^d \to \mathbb{R}^l$ has rank l at every point of the set

$$\{a \in \mathbb{R}^d : f_{i_1}(a) = \cdots = f_{i_l}(a) = 0\} \ .$$

A class $\mathcal{G}$ of real-valued functions defined on $\mathbb{R}^d$ has *solution set components bound B* if for every $k \in \{1, \ldots, d\}$ and every $\{f_1, \ldots, f_k\} \subseteq \mathcal{G}$ that has regular zero-set intersections the number of connected components of the set

$$\{a \in \mathbb{R}^d : f_1(a) = \cdots = f_k(a) = 0\}$$

is at most B. Further, a class $\mathcal{F}$ of real-valued functions is said to be *closed under addition of constants* if for every $c \in \mathbb{R}$ and $f \in \mathcal{F}$ the function $z \mapsto f(z) + c$ is also in $\mathcal{F}$. The following Lemma gives a stronger formulation of a bound stated in Theorem 7.6 of [1].

Lemma 2. *Let $\mathcal{F}$ be a class of real-valued functions $(y_1, \ldots, y_d, x_1, \ldots, x_n) \mapsto f(y_1, \ldots, y_d, x_1, \ldots, x_n)$ that is closed under addition of constants and where each function in $\mathcal{F}$ is C^d in the variables $y_1, \ldots, y_d$. If the class $\mathcal{G} = \{(y_1, \ldots, y_d) \mapsto f(y_1, \ldots, y_d, s) : f \in \mathcal{F}, s \in \mathbb{R}^n\}$ has solution set components bound B then for any sets $\{f_1, \ldots, f_k\} \subseteq \mathcal{F}$ and $\{s_1, \ldots, s_m\} \subseteq \mathbb{R}^n$, where $m \geq d/k$, the function $T : \mathbb{R}^d \to \{0,1\}^{mk}$ defined by*

$$T(a) = (\mathrm{sgn}(f_1(a, s_1)), \ldots, \mathrm{sgn}(f_1(a, s_m)), \ldots, \mathrm{sgn}(f_k(a, s_1)), \ldots, \mathrm{sgn}(f_k(a, s_m)))$$

partitions $\mathbb{R}^d$ into at most $B \sum_{i=0}^{d} \binom{mk}{i} \leq B(emk/d)^d$ equivalence classes (where $a_1, a_2 \in \mathbb{R}^d$ are equivalent iff $T(a_1) = T(a_2)$).

Proof. The statement follows from the proofs of Theorems 7.6 and 7.8 in [1]. □

Proof (Theorem 1). Assume that $\mathcal{N}$ is given as supposed, let $\{s_1, \ldots, s_m\}$ be a set of input vectors and $u_1, \ldots, u_m$ be real numbers. The main task is to bound the number of dichotomies induced on $S = \{(s_1, u_1), \ldots, (s_m, u_m)\}$ by functions of the form $(x, y) \mapsto \mathrm{sgn}(f(x) - y)$ where f is computed by $\mathcal{N}$. We proceed by induction on the levels of the network nodes where the level of a node v is defined as the length of the longest path from an input node to v. Thus, input nodes have level 0 and the level of the output node is equal to the network depth.

As induction hypothesis, assume that we have derived a set of functions $\mathcal{G}_{\lambda-1}$, $\lambda \geq 1$, in the network parameters for the nodes of level at most $\lambda - 1$ that partition the parameter domain into equivalence classes such that for parameters from the same class the computations of these nodes do not change on inputs from S in the following sense: The sequence of subsets of simultaneously active synapses remains the same for each node and the firing of a node is triggered by the same set of (starting or ending) synaptic pulses.

Let some $(s,u) \in S$ be given. Assume that v is some node of level λ that receives spikes through its i-th and j-th connection with corresponding delay parameters $d_{v,i}$ and $d_{v,j}$. Consider the function

$$r_{v,i} + d_{v,i} + b_i - r_{v,j} - d_{v,j} - b_j \ , \tag{1}$$

where $r_{v,i}$ and $r_{v,j}$ are the firing times of the nodes that are presynaptic to connection i and j, respectively. (If $\lambda = 1$ then $r_{v,i}$ and $r_{v,j}$ are firing times of input nodes.) Further, b_i and b_j are binary values depending on whether they refer to the starting or ending point of the corresponding synaptic pulse h_i or h_j, respectively. Thus, the sign of the function (1) indicates the temporal order of these points. The function is affine in the delay parameters of the nodes of level at most λ.

Let C_λ denote the total number of synapses of nodes on level λ. Varying over i,j and the possible values for b_i, b_j there arise at most $(2C_\lambda)^2$ functions of the form (1) for the given (s,u). The union of these functions partitions the parameter domain of the delays of the nodes up to level λ into equivalence classes such that for parameters within the same class the temporal order of synaptic pulses in each of these nodes does not change. By taking into account up to $m(2C_\lambda)^2$ functions of the form (1) we can achieve this property for all m elements of S. Then for each of the nodes up to level λ the sequence of subsets of simultaneously active synapses generated on any element from S remains unchanged within the same class of delay parameters.

Next we consider the weights and thresholds on level λ. We partition their domain into equivalence classes such that for each node on level λ the same set of synaptic pulses h_i (some starting, some ending) is responsible for the firing of the node when choosing weights and thresholds within the same class. (This also includes the possible non-firing of the node.) This then completes the induction step.

For given $(s,u) \in S$, node v of level λ, and time $t \in \mathbb{R}^+$ consider the function

$$\sum_{j=1}^{k} M_{I_{v,j}}(t) - \theta_v \tag{2}$$

where $I_{v,1}, \ldots, I_{v,k}$ are the synaptic clusters of v and θ_v is its threshold. If v has n synapses and the delay parameters of level at most λ remain within the same class, the firing of v on (s,u) depends on at most $2n$ functions of the form (2). This follows from the fact that the subset of synapses that are active at a certain time can change only when one of the synaptic pulses h_i starts or ends. Thus, the $2n$ starting and ending points define no more than $2n$ intervals during each of which a particular subset of synapses is active. Hence, considering all nodes of level λ for the given (s,u) and delays from one class, there are at most $2C_\lambda$ functions that partition the parameter domain of weights and thresholds into equivalence classes where for each node the same set of synaptic pulses triggers its firing. Using at most $2mC_\lambda$ functions of the form (2), we obtain a partition such that for all elements from S the firing of the nodes on level λ does not change in this sense. Each function is a sum of at most k rational functions in the weights and thresholds of level λ where each term in the sum has degree at most p.

Clearly, the class of affine functions has solution set components bound $B = 1$. Further, the class of sums of at most k rational functions in d variables with degree p has solution set components bound $B = 2(2kp)^d$. (This is obtained from the bound $2(2l)^d$ for degree l polynomials, see [1, Corollary 8.2]. Without loss of generality we assume that the denominator of the rational functions is never zero.) Both function classes are closed under addition of constants. Thus, applying Lemma 2 twice we obtain for the number of equivalence classes generated by the functions defined in (1) and (2) the upper bound

$$(em(2C_\lambda)^2/W_\lambda)^{W\lambda} \cdot 2(2kp)^{W_\lambda} \cdot (2emC_\lambda/W_\lambda)^{W_\lambda} \tag{3}$$

where W_λ is the total number of parameters of nodes of level at most λ.

Bound (3) is derived assuming a given class in the partition that arises from the function class $\mathcal{G}_{\lambda-1}$. Defining $\mathcal{G}_\lambda$ to include $\mathcal{G}_{\lambda-1}$ and the functions from (1) and (2) it follows by induction that we get an upper bound for the number of equivalence classes for the parameter domain of all nodes of level at most l by forming the product of (3) for $\lambda = 1, \dots, l$. For the output node of the network, which has level D, we include $2mC_D$ additional affine functions

$$r_{v,i} + d_{v,i} + b_i - u \ ,$$

with $b_i \in \{0, 1\}$ and $u \in \{u_1, \dots, u_m\}$, to describe the starting or ending of postsynaptic pulses relative to u. This leads to the upper bound

$$(2emC_D/W_D)^{W_D} \cdot \prod_{\lambda=1}^{D} \left[(em(2C_\lambda)^2/W_\lambda)^{W_\lambda} \cdot 2(2kp)^{W_\lambda} \cdot (2emC_\lambda/W_\lambda)^{W_\lambda}\right]$$

for the number of equivalence classes of the parameter domain of the network such that within each class the network induces the same dichotomy on S via the function $(x, y) \mapsto \mathrm{sgn}(f(x) - y)$ where f is computed by $\mathcal{N}$. Using $C_\lambda \leq W_\lambda$ and $W_\lambda \leq W$ for $\lambda = 1, \dots, D$ we get the bound

$$2^D(2em)^W(16e^2m^2Wkp)^{WD} \ .$$

If S is shattered there must be 2^m dichotomies induced and hence at least this many equivalence classes for the parameter domain of the network. This implies that

$$m \leq D + (W + 2WD)\log m + W\log(2e) + WD\log(16e^2Wkp) \ .$$

Using

$$(W + 2WD)\log m \leq m/2 + (W + 2WD)\log(2(W + 2WD)/(e\ln 2))$$

(see, e.g., [1, Appendix A.1.1]) we conclude that $m = O(WD\log(WDkp))$. This completes the proof of Theorem 1. □

If we restrict the type of functions used to model the synaptic interactions to the class of polynomials, we can derive a bound that improves the previous one to the effect that it does not depend on the number of synaptic clusters.

Theorem 3. *Suppose $\mathcal{N}$ is a network of spiking neurons with W parameters and depth D where the synaptic interactions are polynomials of degree at most p. Then the pseudo dimension of $\mathcal{N}$ is at most $O(WD\log(WDp))$.*

Proof. (Sketch) The proof follows the reasoning of the previous one. The only place where the number of clusters was taken into account was in the solution set components bound. For the class of degree p polynomials in d variables the solution set components bound $2(2p)^d$ can be used instead. This leads to the claimed result. □

If the depth and the degree of synaptic interactions are fixed, the previous theorem yields a tight bound. The same holds for rational interactions if the number of synaptic clusters in each node is fixed. We further note that the following almost linear upper bound improves the quadratic bound for networks with linear interactions given in [5].

Corollary 4. *Consider the class of networks of spiking neurons with fixed depth and with polynomial synaptic interactions of fixed degree. Suppose $\mathcal{N}$ is a network from this class with W parameters. Then the pseudo dimension of $\mathcal{N}$ is $\Theta(W\log W)$. This bound is also valid for the class of networks with rational synaptic interactions where the depth, the degree of synaptic interactions, and the number of synaptic clusters in each neuron are fixed.*

Proof. The upper bound is due to Theorem 1, the lower bound is due to Theorem 2.2 in [5] where it was shown that a single spiking neuron with W delay parameters has VC dimension $\Omega(W\log W)$. □

4 Networks with Arbitrary Depth

Clearly, we obtain from Theorem 1 a bound not depending on the depth of the network using that the depth is not larger than the number of parameters. A direct method yields a better bound. Its proof is omitted here.

Theorem 5. *Let $\mathcal{N}$ be a network of spiking neurons with W parameters and rational synaptic interactions of degree at most p where each neuron has at most k synaptic clusters. Then the pseudo dimension of $\mathcal{N}$ is at most $O(W^2\log(kp))$. For networks with polynomial interactions the bound $O(W^2\log p)$ holds.*

This leads to the following result analogous to Corollary 4. The lower bound is due to Theorem 3.1 in [5].

Corollary 6. *Consider the class of networks of spiking neurons with polynomial synaptic interactions of fixed degree. Suppose $\mathcal{N}$ is a network from this class with W parameters. Then the pseudo dimension of $\mathcal{N}$ is $\Theta(W^2)$. This bound is also valid for the class of networks with rational synaptic interactions where the degree of these interactions and the number of synaptic clusters in each neuron are fixed.*

5 Networks with Unrestricted Order

The networks considered thus far require restricted degrees of synaptic interactions. We next establish a polynomial pseudo dimension bound even for unlimited degree. Let the interaction of synpatic cluster $\{i_1, \ldots, i_l\}$ be modeled by the function $f^{(q_1,\ldots,q_l)}_{\{i_1,\ldots,i_l\}}$ in l variables and parameters $q_1, \ldots, q_l \in \mathrm{I\!R}$ with

$$f^{(q_1,\ldots,q_l)}_{\{i_1,\ldots,i_l\}}(w_1, \ldots, w_l) = w_1^{q_1} w_2^{q_2} \cdots w_l^{q_l} \quad .$$

(We require that q_i is an integer if $w_i < 0$, to keep the network working in the real domain.) A bound on the pseudo dimension is given by the following result based on new solution set components bounds for function classes defined in terms of polynomials, exponentials, and logarithms, which we omit here.

Theorem 7. *Suppose $\mathcal{N}$ is a network of spiking neurons with W parameters and nonlinear synaptic interactions of the form $f^{(q_1,\ldots,q_l)}_{\{i_1,\ldots,i_l\}}$ where each neuron has at most k synaptic clusters. Then $\mathcal{N}$ has pseudo dimension at most $O((Wk)^4)$.*

6 Conclusions

We have derived upper bounds on the pseudo dimension of networks of spiking neurons allowing for nonlinear synaptic interactions. All bounds are polynomials of low degree. The pseudo dimension is well known to be an upper bound for the so-called fat-shattering dimension. Both dimensions give bounds on the so-called covering numbers. The latter can be employed to obtain estimates for the number of training examples required for the networks to have low generalization error. The results presented here show that even when the computational power of networks of spiking neurons increases considerably due to nonlinear interactions, the sample complexity remains close to that of linear interactions.

Not all bounds were shown to be tight, however, and future work could lead to improvements. It might also be possible to obtain better estimates for the sample complexity by deriving bounds on the fat-shattering dimension directly.

References

1. M. Anthony and P. L. Bartlett. *Neural Network Learning: Theoretical Foundations.* Cambridge University Press, Cambridge, 1999.
2. W. Gerstner. Spiking neurons. In W. Maass and C. M. Bishop, editors, *Pulsed Neural Networks*, chapter 1, pages 3–53. MIT Press, Cambridge, Mass., 1999.
3. C. Koch. *Biophysics of Computation.* Oxford University Press, New York, 1999.
4. W. Maass. Computing with spiking neurons. In W. Maass and C. M. Bishop, editors, *Pulsed Neural Networks*, chapter 2, pages 55–85. MIT Press, Cambridge, Mass., 1999.
5. W. Maass and M. Schmitt. On the complexity of learning for spiking neurons with temporal coding. *Information and Computation*, 153:26–46, 1999.
6. A. F. Murray. Pulse-based computation in VLSI neural networks. In W. Maass and C. M. Bishop, editors, *Pulsed Neural Networks*, chapter 3, pages 87–109. MIT Press, Cambridge, Mass., 1999.