Gradient Learning in Spiking Neural Networks by Dynamic Perturbation of Conductances

Ila R. Fiete[1] and H. Sebastian Seung[2]

[1]Kavli Institute for Theoretical Physics, University of California, Santa Barbara, California 93106, USA
[2]Howard Hughes Medical Institute and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

(Received 19 January 2006; published 28 July 2006)

We present a method of estimating the gradient of an objective function with respect to the synaptic weights of a spiking neural network. The method works by measuring the fluctuations in the objective function in response to dynamic perturbation of the membrane conductances of the neurons. It is compatible with recurrent networks of conductance-based model neurons with dynamic synapses. The method can be interpreted as a biologically plausible synaptic learning rule, if the dynamic perturbations are generated by a special class of ‘‘empiric’’ synapses driven by random spike trains from an external source.

DOI: 10.1103/PhysRevLett.97.048104 PACS numbers: 87.18.Sn, 07.05.Mh, 84.35.+i, 87.19.La

Neural network learning is often formulated in terms of an objective function that quantifies performance at a desired computational task. The network is trained by estimating the gradient of the objective function with respect to synaptic weights, and then changing the weights in the direction of the gradient.

If neural and network dynamics and the objective function are all exactly known functions of the weights, such learning can be accomplished by explicitly computing the relevant gradients. A famous example of this approach, used with wide success in nonspiking, deterministic artificial neural networks [1], is the *backpropagation* (BP) algorithm [2].

However, the relevance of BP to neurobiological learning is limited. Biological neural activity can be noisy, and involves the highly nonlinear and often history-dependent dynamics of membrane voltages and conductances: neurons generate voltage spikes, and the efficacy of synaptic transmission varies dynamically on a spike by spike basis [3]. Further, the objective function in neurobiological learning may depend on the dynamics of muscles and external variables of the world unknown to the brain. Similar complications are also present in analog on-chip or robotic implementations of machine learning.

For learning in such systems, alternative strategies are necessary. The method of *weight perturbation* estimates the gradients by independently perturbing synaptic weights, and observing the change in the objective function. Unlike BP, weight perturbation is completely ‘‘model free’’ [4]—it does not depend on knowing anything about the functional dependence of the objective on the network weights—and can be applied to stochastic spiking networks [5]. The disadvantage of a completely model-free approach is the trade-off between generality and learning speed: weight perturbation is far more widely applicable than BP, but BP is much faster when it is applicable.

Here we propose a method that is intermediate between these two extremes, yet is applicable to arbitrary neural networks. Instead of making perturbations to the synaptic weights, it estimates the N^2 weight gradients through dynamic perturbation of the conductances of the N network neurons. Our algorithm does this by exploiting a feature generic to many models of neural networks: that inputs to a neuron combine additively before being subjected to further nonlinearities. Otherwise, the algorithm is model free. Our approach generalizes the concept of *node perturbation*, which has been proposed for training feed forward networks of nonspiking neurons [2,6] and can be much faster than weight perturbation [7]. We show how neural conductance perturbations can be biologically plausibly used to perform synaptic gradient learning in fully recurrent networks of realistic spiking neurons.

Spiking neural networks.—We briefly discuss the mathematical conditions under which our assumption, that the synaptic inputs to a single neuron combine linearly, holds in spiking neural networks. If each neuron i is electronically compact, it can be described by a transmembrane voltage V_i, obeying the current balance equation $C_i dV_i/dt = -I_i^{\mathrm{int}}(t) - I_i^{\mathrm{syn}}(t)$. The intrinsic current I_i^{int} is generally a nonlinear function of voltage and dynamical variables associated with the spike-generating conductances in the membrane. The dynamics of these variables may be arbitrarily complex (e.g. Hodgkin-Huxley model) without affecting our derivations. A simple model for the synaptic current is $I_i^{\mathrm{syn}} = \sum_j W_{ij} s_{ij}(t)[V_i(t) - E_{ij}]$. The time-varying synaptic conductance from neuron j to neuron i is $W_{ij} s_{ij}(t)$, with amplitude controlled by the parameter W_{ij}. Its time course is determined by $s_{ij}(t)$, which could include complex forms of short-term depression and facilitation. If the reversal potentials E_{ij} of the synapses are all the same, then the synaptic current can be written as $I_i^{\mathrm{syn}} = g_i(t)[V_i(t) - E^{\mathrm{syn}}]$, where

$$g_i(t) = \sum_j W_{ij} s_{ij}(t) \tag{1}$$

is the sum of all postsynaptic conductances of the synapses onto neuron i. The linear dependence of $g_i(t)$ on the synaptic weights W_{ij} will be critical below. However, this linear dependence may be embedded inside a nonlinear network, which may be arbitrarily complex without affecting the following derivations. In fact, all networks—whether neural and spiking or neither—that depend on a set of interaction variables $s_{ij}(t)$ and parameters W_{ij} through Eq. (1) satisfy the necessary conditions for our derivation below.

Gradient learning.—We represent the state of the network by a vector $\Omega(t)$, which includes the synaptic variables $s_{ij}(t)$ and all other dynamical variables [e.g., the voltages $V_i(t)$ and all variables associated with the membrane conductances]. Starting from an initial condition $\Omega(0)$ the network generates a trajectory from time $t = 0$ to $t = T$, and in response receives a scalar "reinforcement" signal $R[\Omega]$, which is an arbitrary functional of the trajectory. For now we assume that the network dynamics are deterministic, and present the fully stochastic case in the Appendix. Each trajectory along with its reinforcement is called a "trial," and the learning process is iterative, extending over a series of trials. The signal R depends implicitly on the synaptic weights W_{ij}, and is an objective function for learning. In other words, the goal of learning is to find synaptic weights that maximize R. A heuristic method for doing this is to follow the gradient of R with respect to W_{ij}. Next we derive our gradient learning rule.

Sensitivity lemma.—If W_{ij} were a time-dependent function, a variation in W_{ij} at time t would cause a variation in R given by $\delta R/\delta W_{ij}(t)$; by Eq. (1) and the chain rule, this variation is equivalent to

$$\frac{\delta R}{\delta W_{ij}(t)} = \frac{\delta R}{\delta g_i(t)} s_{ij}(t). \tag{2}$$

Because W_{ij} is a parameter, it is constrained to take on the same value at every time. However, Eq. (2) shows that variations in the neural conductance $g_i(t)$ can act as proxies for computing the effects of hypothetical time-dependent variations in W_{ij}. It follows that

$$\frac{\partial R}{\partial W_{ij}} = \frac{1}{T}\int_0^T dt \frac{\delta R}{\delta W_{ij}(t)} = \frac{1}{T}\int_0^T dt \frac{\delta R}{\delta g_i(t)} s_{ij}(t). \tag{3}$$

We call this the *sensitivity lemma*, because it relates the sensitivity of R to changes in W_{ij} with the sensitivity to changes in $g_i(t)$. The implication of the lemma is that *dynamic* perturbations of the variables $g_i(t)$ can be used to instruct modifications of the *static* parameters W_{ij}.

Gradient estimation.—In order to estimate $\delta R/\delta g_i(t)$ suppose that Eq. (1) is perturbed by a time-varying noise term,

$$g_i(t) = \sum_j W_{ij} s_{ij}(t) + \xi_i(t). \tag{4}$$

The noise satisfies $\langle \xi_i(t)\rangle = 0$ and $\langle \xi_i(t)\xi_j(t')\rangle = \sigma^2 \delta_{ij}\delta(t - t')$, where the angle brackets denote a trial average. For now, let us regard this perturbation as a mathematical device; its biological interpretation will be discussed later.

To show that $\delta R/\delta g_i(t)$ can be estimated from the covariance of R and the perturbation $\xi_i(t)$, use the linear approximation $R - R_0 \approx \int_0^T dt \sum_k [\delta R/\delta g_k(t)]\xi_k(t)$, which is accurate when the perturbations $\xi_i(t)$ are small. Here R_0 is defined as R in the absence of any perturbations, $\xi = 0$. Since the perturbations are spatially and temporally uncorrelated, it follows that

$$\langle (R - R_0)\xi_i(t)\rangle \approx \sigma^2 \frac{\delta R}{\delta g_i(t)}. \tag{5}$$

Because $\langle \xi \rangle = 0$, the baseline R_0 may be replaced by any quantity that is uncorrelated with the perturbations of the current trial. For example, choosing $R_0 = 0$ leaves Eq. (5) valid. However, baseline subtraction can have a large effect on the variance of the estimate (5) when based on a finite number of trials [8]. Thus a good choice of baseline can decrease learning time, sometimes dramatically.

If the covariance relation of Eq. (5) is combined with the sensitivity lemma Eq. (3), it follows that

$$\sigma^2 \frac{\partial R}{\partial W_{ij}} \approx \frac{1}{T}\int_0^T dt \langle (R - R_0)\xi_i(t)\rangle s_{ij}(t). \tag{6}$$

Synaptic learning rule.—Equation (6) suggests the following stochastic gradient learning procedure. At each synapse the purely local *eligibility trace*

$$e_{ij} = \int_0^T dt \xi_i(t) s_{ij}(t) \tag{7}$$

is accumulated over the trajectory. At the end of the trajectory, the synaptic weight is updated according to

$$\Delta W_{ij} = \eta (R - R_0) e_{ij}. \tag{8}$$

The update ΔW_{ij} fluctuates because of the randomness in the perturbations. On average, the update points in the direction of the gradient, because it satisfies $\langle \Delta W_{ij}\rangle \propto \partial R/\partial W_{ij}$, according to Eq. (6). This means that the learning rule of Eq. (8) is stochastic gradient following.

We note one subtlety in the derivation: In the sensitivity lemma, the synaptic variables $s_{ij}(t)$ are noise free ($\xi = 0$), while in Eq. (7) they are defined in the presence of perturbations. However in the linear approximation, this discrepancy leads to a higher order correction that is negligible for small perturbations.

Biological interpretation.—According to the above, synaptic weight gradients of R can be estimated by using conductance perturbations $\xi_i(t)$. Could this mathematical trick be used by the brain? In the actor-critic terminology of reinforcement learning [9], one can imagine that the neurons of one brain area (the "actor") drive actions that are assessed by another brain area (the "critic"), which in response issues a global, scalar reinforcement signal R to the actor (Fig. 1). A novel feature of our rule is that in addition to its regular plastic synapses W_{ij}, the actor would receive a special class of "empiric" synapses from another hypothesized part of the brain (the "experimenter"), which perturb the actor from trial to trial. Each plastic synapse locally computes (by keeping track of its coincident acti-

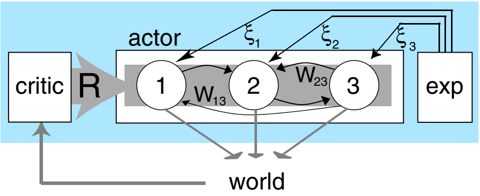

FIG. 1 (color online). Neurons in a recurrent network (actor), connected by plastic (modifiable) weights W. A global reinforcement signal R is broadcast by a critic to all neurons in the network. In addition, each neuron i receives an empiric synapse carrying perturbing input $\xi_i(t)$ from an external experimenter.

vation with the perturbing input) and stores its scalar eligibility, and multiplies this with R to undergo modification. This idea is developed in detail elsewhere in a model of bird song learning [10], resulting in novel, testable predictions for synaptic plasticity in the brain.

Note that if the perturbation $\xi_i(t)$ is a synaptic conductance, its mean value $\langle \xi_i(t) \rangle$ must be positive. Then the linear approximations above are expansions about the mean conductance $\xi_i(t) = \langle \xi_i(t) \rangle$, rather than $\xi_i(t) = 0$. As a result, $\xi_i(t)$ must be replaced by the zero-mean fluctuation $\delta\xi_i(t) = \xi_i(t) - \langle \xi_i(t) \rangle$ in the eligibility trace. In addition, the fluctuations $\delta\xi_i(t)$ will resemble a filtered point process, and will not be truly white, but will have a correlation time set by the time constant of the synaptic currents. However, if this correlation time is short relative to the time scale of variation in $\delta R/\delta g_i(t)$, then the gradient estimate Eq. (5) should still be accurate.

Accurate gradient estimation requires that the eligibility trace filter out the mean conductance $\langle \xi_i(t) \rangle$ of the empiric synapse. If the external experimenter neurons drive the empiric synapses at a constant or slowly varying underlying rate, $\langle \xi_i(t) \rangle \approx \langle \xi_i \rangle$ is a number and this operation is biologically plausible. It can be implemented by a simple time average at every actor neuron.

By contrast, other proposals for stochastic gradient learning typically rely on *intrinsic* sources of noise generated within the actor network [5,11,12]. If the perturbations are intrinsic, their statistics depend on network activity, and may vary rapidly during the actor trajectory. Thus, the expected perturbation cannot be accurately estimated by a time average. But because all these schemes, like ours, depend on the subtraction of expected activity, individual neurons would have to keep track of and filter out $\langle \xi_i(t) \rangle$, a *time-varying* expected perturbation vector within each trial, which seems rather complex. Our algorithm avoids this complexity, because the fluctuations are injected by an *extrinsic* source, and are therefore independent of the network trajectory and may be assumed to be approximately stationary. The extrinsic experimenter approach has the additional advantage that the statistics of exploration can be controlled without manipulation of the actor.

Generalization to excitatory and inhibitory synapses.—Above we assumed that all synapses have the same reversal potential. But neurons may receive both excitatory and inhibitory synapses, which have different reversal potentials. The unmodified learning rule allows both synapse types to perform gradient following if there are two types of empiric synapses per neuron: an excitatory empiric synapse used to train the excitatory synapses, and an inhibitory empiric synapse used to train the inhibitory synapses. But if there is only one empiric synapse per neuron, then for both types of synapses to perform gradient following, the rule must be modified. Let E_{ij} and $E_{\xi,i}$ be the reversal potentials of the regular $i \leftarrow j$ synapse and of the empiric synapse onto the ith actor neuron, respectively. Then we obtain a generalized sensitivity lemma

$$\frac{\partial R}{\partial W_{ij}} = \int dt a_{ij}(t) \frac{\delta R}{\delta g_i(t)} s_{ij}(t), \tag{9}$$

where

$$a_{ij}(t) = \frac{V_i(t) - E_{ij}}{V_i(t) - E_{\xi,i}} \tag{10}$$

is the ratio of the synaptic driving force at the $i \leftarrow j$ synapse to the driving force of the empiric synapse at neuron i. The stochastic gradient learning rule remains $\Delta W_{ij} = \eta(R - R_0)e_{ij}$, but with modified eligibility trace

$$e_{ij} = \int_0^T dt a_{ij}(t) \xi_i(t) s_{ij}(t).$$

For synapses with the same reversal potential as the empiric synapse, $a_{ij}(t) = 1$, returning the original learning rule. Even for synapses of the opposite variety, the sign of a_{ij} does not change with time because neural voltage is constrained to stay between the inhibitory and excitatory reversal potentials V_I and V_E ($V_I \leq V_i(t) \leq V_E$), and $E_{\xi,i}$, $E_{ij} \in \{V_I, V_E\}$. Nevertheless, for these synapses of the opposite variety, the term $a_{ij}(t)$ adds complexity to the simple learning rule and reduces its biological plausibility.

Generalization to multicompartmental model neurons.—Suppose the model neuron is not isopotential, but has several dendritic compartments. Then it can be trained without modification of the learning rule by using a separate empiric synapse for each compartment. Alternatively, a single empiric synapse could be used for the whole neuron, but with the introduction of complexities in the learning rule similar to the $a_{ij}(t)$ factor of Eq. (10).

Technical issues.—Our synaptic learning rule performs stochastic gradient following, and therefore shares the virtues and deficiencies of all methods in this class [13]. For example, it is possible to become stuck at a local optimum of the objective function. The stochasticity of the gradient estimation may allow some small probability of escape, but there is no guarantee of finding a global optimum. Like all gradient-based learning techniques, our algorithm works when the gradient of the expected value of R with respect to the synaptic weights W_{ij} is well defined.

Comparison with previous work.—If the perturbation $\xi_i(t)$ is *Gaussian* white noise, our learning rule can be derived exactly using the REINFORCE formalism, where eligibility is defined as the derivative of the log probability of the network state, rather than by using linear approximations. Thus, for Gaussian white noise our rule can be included as a member of the REINFORCE class of algorithms [12]. More generally, however, our perturbative methodology is complementary to REINFORCE, and unlike REINFORCE for most non-Gaussian perturbations, will always produce learning rules that are linear in the perturbation. Further, REINFORCE requires that the part of the probability density function (PDF) of the network state involving the weights W_{ij} be smooth, a condition that can seriously affect the applicability of REINFORCE theory: for example, a network of neurons perturbed by a noise ξ generated from a point process by filtering a random spike train cannot be treated by REINFORCE. By contrast, our linear response-based derivation can handle nonsmooth PDF's involving W_{ij} so long as the injected perturbations are small. In short, our linear response approach is applicable whenever REINFORCE is, and generally results in a different plasticity rule; it is also often applicable when REINFORCE is not.

The sensitivity lemma permits us to derive rules for synaptic gradient learning based on perturbations of quantities not directly related to the synaptic parameters. Versions of the sensitivity lemma have appeared in the literature for nonspiking feed forward networks, and been used to estimate the gradient by serially perturbing one neuron at a time (node perturbation) [6,14]. Our version of the sensitivity lemma is more general, because it is applicable to learning trajectories in recurrent networks, via parallel perturbation of multiple neurons.

We are grateful to Xiaohui Xie for helpful discussions. For thoughtful comments on the manuscript, we thank H. Sompolinsky, Y. Loewenstein, and U. Rokni. I. F. acknowledges funding from NSF PHY 99-07949.

Appendix: Stochastic networks.—Above the network dynamics and reinforcement R were assumed to be deterministic. Both elements can be made stochastic, as outlined below. Consider the case of discrete time (continuous time is a limiting case). The network generates a trajectory $\Omega = \{\Omega(0), \Omega(1), \ldots, \Omega(T)\}$ from a probability density $P_W(\Omega)$. Suppose each trajectory is generated by drawing an initial condition $\Omega(0)$ from some probability density and then drawing $\Omega(1)$ through $\Omega(T)$ from a Markov process with transition probability $P_W[\Omega(t)|\Omega(t-1)]$. The assumption of Markov transition probabilities is compatible with most spiking neural network models. The network receives reinforcement R from the conditional density $P(R|\Omega)$. Since the network is parametrized by W, the expected reward

$$\langle R\rangle = \int RP(R|\Omega)P_W(\Omega)dR\mathcal{D}\Omega \tag{A1}$$

is a function of W. We assume that the transition probability depends on the weights W through

$$P_W[\Omega(t)|\Omega(t-1)] = f(g_1(t), \ldots, g_N(t)), \tag{A2}$$

where as before

$$g_i(t) = \sum_j W_{ij}s_{ij}(t-1). \tag{A3}$$

The transition probability depends on all the dynamical variables in $\Omega(t)$, although they have been suppressed for notational simplicity in Eq. (A2). As before, the important mathematical property here is the linearity of Eq. (A3), which is embedded inside a nonlinear system. The sensitivity lemma takes the form

$$\frac{\partial\langle R\rangle}{\partial W_{ij}} = \sum_{t=1}^{T}\frac{\partial}{\partial g_i(t)}\langle Rs_{ij}(t-1)\rangle. \tag{A4}$$

The sensitivity lemma shows that the appropriate change in the weight of a synapse is not given by the covariance of its activity with reinforcement (as might be naively expected), but is instead given by the derivative with respect to $g_i(t)$ of this covariance. As before, the proof of the sensitivity lemma involves comparing derivatives of the reinforcement—synaptic activity correlations convolved with the transition probabilities, taken with respect to W_{ij} and $g_i(t)$, without actually performing either differentiation. [Note that REINFORCE requires the stronger condition that the log probability $\ln P_W(\Omega)$ be differentiable]. For small perturbations $\xi_i(t)$, this sensitivity lemma leads us again to the gradient learning rule of Eqs. (7) and (8), now valid for fully stochastic networks.

[1] Y. LeCun *et al.*, Proc. IEEE **86**, 2278 (1998).

[2] B. Widrow and M. Lehr, Proc. IEEE **78**, 1415 (1990); D. Rumelhart, in *Parallel Distributed Processing*, edited by D. Rumelhart and J. McClelland (MIT, Cambridge, MA, 1986).

[3] H. Markram and M. Tsodyks, Nature (London) **382**, 807 (1996); A. Thomson and J. Deuchars, Trends Neurosci. **17**, 119 (1994).

[4] A. Dembo and T. Kailath, IEEE Trans. Neural Nets. **1**, 58 (1990).

[5] H. S. Seung, Neuron **40**, 1063 (2003).

[6] Y. LeCun *et al.*, in *Advances in Neural Information Processing*, edited by D. Touretzky (MIT, Cambridge, MA, 1989), Vol. 1, p. 141.

[7] J. Werfel, X. Xie, and H. S. Seung, Neural Comput. **17**, 2699 (2005).

[8] P. Dayan, in *Proceedings of the Connectionist Models Summer School*, edited by D. Touretzky *et al.* (Morgan Kaufmann, San Diego, 1990).

[9] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction* (MIT, Cambridge, MA, 1998).

[10] I. Fiete, M. Fee, and H. S. Seung (to be published).

[11] A. Barto and P. Anandan, IEEE Trans. Syst. Man Cybern. **15**, 360 (1985); X. Xie and H. S. Seung, Phys. Rev. E **69**, 041909 (2004).

[12] R. Williams, Mach. Learn. **8**, 229 (1992).

[13] B. Pearlmutter, IEEE Trans. Neural Nets. **6**, 1212 (1995).

[14] D. Andes *et al.*, in *International Joint Conference on Neural Networks* (Lawrence Erlbaum Associates, Washington, DC, 1990).