

The Variance of Covariance Rules for Associative Matrix Memories and Reinforcement Learning

Peter Dayan

Terrence J. Sejnowski

Computational Neurobiology Laboratory, The Salk Institute,
P. O. Box 85800, San Diego, CA 92186-5800 USA

Hebbian synapses lie at the heart of most associative matrix memories (Kohonen 1987; Hinton and Anderson 1981) and are also biologically plausible (Brown *et al.* 1990; Baudry and Davis 1991). Their analytical and computational tractability make these memories the best understood form of distributed information storage. A variety of Hebbian algorithms for estimating the covariance between input and output patterns has been proposed. This note points out that one class of these involves stochastic estimation of the covariance, shows that the signal-to-noise ratios of the rules are governed by the variances of their estimates, and considers some parallels in reinforcement learning.

Associations are to be stored between Ω pairs $[a(\omega), b(\omega)]$ of patterns, where $a(\omega) \in \{0, 1\}^m$ and $b(\omega) \in \{0, 1\}^n$, using the real-valued elements of an $m \times n$ matrix W . Elements of $a(\omega)$ and $b(\omega)$ are set independently with probabilities p and r , respectively, of being 1. A learning rule specifies how element W_{ij} changes in response to the input and output values of a particular pair—the model adopted here (from Palm 1988a,b) considers local rules with additive weight changes for which:

$$W_{ij} = \sum_{\omega=1}^{\Omega} \Delta_{ij}(\omega), \quad \text{where } \Delta_{ij}(\omega) = f[a_i(\omega), b_j(\omega)]$$

and f can be represented as $[\alpha, \beta, \gamma, \delta]$ based on

		Output	
		$b = b_j(\omega)$	
		0	1
Input	0	α	β
	1	γ	δ
$a = a_i(\omega)$			

One way to measure the quality of a rule is the signal-to-noise ratio (S/N) of the output of a single "line" or element of the matrix, which is a measure of how well outputs that should be 0 can be discriminated from outputs that should be 1. The larger the S/N, the better the memory will perform (see Willshaw and Dayan, 1990 for a discussion). A wide variety of Hebbian learning rules has been proposed for hetero-

and autoassociative networks (Kohonen 1987; Sejnowski 1977a; Hopfield 1982; Perez-Vincente and Amit 1989; Tsodyks and Feigel'man 1988). The covariance learning rule $f_{\text{cov}} = [pr, -p(1-r), -(1-p)r, (1-p)(1-r)]$, has the highest S/N (Willshaw and Dayan 1990; Dayan and Willshaw 1991); however, both it and a related rule, $f_{\text{prd}} = [-pr, -pr, -pr, 1-pr]$ (Sejnowski 1977a), have the drawback that $\alpha \neq 0$, that is, a weight should change even if both input and output are silent. Note the motivations behind these rules:

$$\begin{aligned} f_{\text{cov}} &\sim (\text{input} - p) \times (\text{output} - r) \\ f_{\text{prd}} &\sim \text{input} \times \text{output} - pr \end{aligned}$$

Alternative rules have been suggested that better model the physiological phenomena of long-term potentiation (LTP) and depression (LTD) in the visual cortex and hippocampus, including the heterosynaptic rule $f_{\text{het}} = [0, -p, 0, 1-p]$ (Stent 1973; Rauschecker and Singer 1979), and the homosynaptic rule $f_{\text{hom}} = [0, 0, -r, 1-r]$ (Sejnowski 1977b; Stanton and Sejnowski 1989), motivated as

$$\begin{aligned} f_{\text{het}} &\sim (\text{input} - p) \times \text{output} \\ f_{\text{hom}} &\sim \text{input} \times (\text{output} - r) \end{aligned}$$

These have been shown to have lower S/Ns than the covariance rule (Willshaw and Dayan 1990); however, for the sparse patterns, that is, low values of p and r , this difference becomes small. The sparse limit is interesting theoretically, because many more patterns can be stored, and empirically, because the cortex has been thought to employ it (see, for example, Abeles *et al.* 1990).

All of these rules are effectively stochastic approximations of the covariance between input and output $\langle (a_i(\omega) - \bar{a}_i) (b_j(\omega) - \bar{b}_j) \rangle_{\omega}$ where the averages $\langle \rangle_{\omega}$ are taken over the distributions generating the patterns; they all share this as their common mean.¹ If inputs and outputs are independent, as is typically the case for heteroassociative memories, or autoassociative ones without the identity terms, then their common expected value is zero. However, the rules differ in their variances as estimates of the covariance. Since it is departures of this quantity from its expected value that mark the particular patterns the matrix has learned, one would expect that the lower the variance of the estimate the better the rule. This turns out to be true, and for independent inputs and outputs the S/N of the rules:

$$\begin{aligned} \text{S/N} [f_{\text{prd}}] &= \frac{mp(1-p)}{\Omega} \frac{1}{pr(1-pr)} & \text{S/N} [f_{\text{het}}] &= \frac{mp(1-p)}{\Omega} \frac{1}{pr(1-p)} \\ \text{S/N} [f_{\text{hom}}] &= \frac{mp(1-p)}{\Omega} \frac{1}{pr(1-r)} & \text{S/N} [f_{\text{cov}}] &= \frac{mp(1-p)}{\Omega} \frac{1}{pr(1-p)(1-r)} \end{aligned}$$

¹This is true for all rules $[\gamma r, -\gamma(1-r), -(1-\gamma)r, (1-\gamma)(1-r)]$ and $[p\eta, -p(1-\eta), -(1-p)\eta, (1-p)(1-\eta)]$ for any γ or η .

are inversely proportional to their variances. f_{cov} is the best, f_{prd} the worst, but the ordering of the other two depends on p and r . Circumstances arise under which the optimal rule differs, as for instance if patterns are presented multiple times but input lines can fail to fire on particular occasions—this would favor the homosynaptic rule.

Exactly the same effect underlies the differences in efficacy between various comparison rules for reinforcement learning. Sutton (1984) studied a variety of two-armed bandit problems, which are conventional tasks for stochastic learning automata. On trial ω , a system emits action $y(\omega) \in \{0, 1\}$ (i.e., pulls either the left or the right arm) and receives a probabilistic reward $r(\omega) \in \{0, 1\}$ from its environment, where

$$\rho_0 = \mathcal{P}[r(\omega) = 1 \mid y(\omega) = 0] \quad \rho_1 = \mathcal{P}[r(\omega) = 1 \mid y(\omega) = 1]$$

In the supervised learning case above, the goal was to calculate the covariance between the input and output. Here, however, the agent has to measure the covariance between its output and the reward in order to work out which action it is best to emit (i.e., which arm it is best to pull). Sutton evaluated $r(\omega)[y(\omega) - \langle y(\omega) \rangle]$ and an approximation to $\{r(\omega) - \langle r(\omega) \rangle\}[y(\omega) - \langle y(\omega) \rangle]$, where $\langle y(\omega) \rangle$ averages over the stochastic process generating the outputs and

$$\langle r(\omega) \rangle = \mathcal{P}[y(\omega) = 0]\rho_0 + \mathcal{P}[y(\omega) = 1]\rho_1$$

is the expected reinforcement given the stochastic choice of $y(\omega)$. These are direct analogues of f_{het} or f_{hom} [depending on whether $y(\omega)$ is mapped to $a(\omega)$ or $b(\omega)$] and f_{cov} , respectively, and Sutton showed the latter significantly outperformed the former. There is, however, an even better estimator. In the previous case, a and b were independent; here, by contrast, $r(\omega)$ is a stochastic function of $y(\omega)$. The learning rule that minimizes the variance of the estimate of the covariance is actually

$$[r(\omega) - \hat{r}][y(\omega) - \langle y(\omega) \rangle]$$

where $\hat{r} = \mathcal{P}[y(\omega) = 0]\rho_1 + \mathcal{P}[y(\omega) = 1]\rho_0$ pairs the probability of emitting action 0 with the reinforcement for emitting action 1. Williams (personal communication) suggested \hat{r} on just these grounds and simulations (Dayan 1991) confirm that it does indeed afford an improvement.

Four previously suggested Hebbian learning rules have been shown to be variants of stochastic covariance estimators. The differences between their performances in terms of the signal-to-noise ratio they produce in an associative matrix memory may be attributed to the differences in the variance of their estimates of the covariance. The same effect underlies the performance of reinforcement comparison learning rules, albeit suggesting a different optimum.

Acknowledgments

We are very grateful to Steve Nowlan and David Willshaw for helpful comments. Support was from the SERC and the Howard Hughes Medical Institute.

References

- Abeles, M., Vaadia, E., and Bergman, H. 1990. Firing patterns of single units in the prefrontal cortex and neural network models. *Network* 1, 13–25.
- Anderson, J. A., and Rosenfeld, E., eds. 1988. *Neurocomputing: Foundations of Research*. MIT Press, Cambridge, MA.
- Baudry, M., and Davis, J. L. 1991. *Long-Term Potentiation: A Debate of Current Issues*. MIT Press, Cambridge, MA.
- Brown, T. H., Kairiss, E. W., and Keenan, C. L. 1990. Hebbian synapses: Biophysical mechanisms and algorithms. *Annu. Rev. Neurosci.* 13, 475–512.
- Dayan, P. 1991. Reinforcement comparison. In *Connectionist Models: Proceedings of the 1990 Summer School*, D. S. Touretzky, J. L. Elman, T. J. Sejnowski and G. E. Hinton, eds. Morgan Kaufmann, San Mateo, CA.
- Dayan, P., and Willshaw, D. J. 1991. Optimal synaptic learning rules in linear associative memories. *Biol. Cybernet.* 65, 253–265.
- Hinton, G. E., and Anderson, J. A., eds. 1981. *Parallel Models of Associative Memory*. Lawrence Erlbaum, Hillsdale, NJ.
- Hopfield, J. J. 1982. Neural networks and physical systems with emergent computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* 79, 2554–2558.
- Kohonen, T. 1987. *Content-addressable Memories*, 2nd ed. Springer-Verlag, Berlin.
- Palm, G. 1988a. On the asymptotic information storage capacity of neural networks. In *Neural Computers. NATO ASI Series*, R. Eckmiller and C. von der Malsburg, eds., Vol. F41, 271–280. Springer-Verlag, Berlin.
- Palm, G. 1988b. Local synaptic rules with maximal information storage capacity. In *Neural & Synergetic Computers, Springer Series in Synergetics*, H. Haken, ed., Vol. 42, 100–110. Springer-Verlag, Berlin.
- Perez-Vincente, C. J., and Amit, D. J. 1989. Optimised network for sparsely coded patterns. *J. Phys. A: Math. General* 22, 559–569.
- Rauschecker, J. P., and Singer, W. 1979. Changes in the circuitry of the kitten's visual cortex are gated by postsynaptic activity. *Nature (London)* 280, 58–60.
- Sejnowski, T. J. 1977a. Storing covariance with nonlinearly interacting neurons. *J. Math. Biol.* 4, 303–321.
- Sejnowski, T. J. 1977b. Statistical constraints on synaptic plasticity. *J. Theoret. Biol.* 69, 385–389.
- Stanton, P., and Sejnowski, T. J. 1989. Associative long-term depression in the hippocampus: Induction of synaptic plasticity by Hebbian covariance. *Nature (London)* 339, 215–218.
- Stent, G. S. 1973. A physiological mechanism for Hebb's postulate of learning. *Proc. Natl. Acad. Sci.* 70, 997–1001.

- Sutton, R. S. 1984. *Temporal Credit Assignment in Reinforcement Learning*. Ph.D. Thesis, University of Massachusetts, Amherst, MA.
- Tsodyks, M. V., and Feigel'man, M. V. 1988. The enhanced storage capacity in neural networks with low activity level. *Europhys. Lett.* **6**, 101–105.
- Willshaw, D. J., and Dayan, P. 1990. Optimal plasticity in matrix memories: What goes up MUST come down. *Neural Comp.* **2**, 85–93.

Received 20 January 1992; accepted 14 September 1992.