ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 207-212

Regularization in Oculomotor Control

John A. Bullinaria & Patricia M. Riddell

Department of Psychology, The University of Reading
Reading, RG6 6AL, UK

Abstract. In modelling the development of the oculomotor control system using neural networks, it is important to determine the appropriate cost function on which to train the models. Whilst blur and disparity are fairly obvious error components, choosing the regularization component is less easy. In this paper we explore the consequences of a number of the most reasonable possibilities and investigate the extent to which other factors may dominate their influence.

1. Introduction

The human oculomotor control system adjusts the accommodation (focusing) and vergence (relative directions) of our eyes so that we can see objects clearly at different distances. It is able to generate quick and efficient transitions between targets at different locations in the visual field. Numerous control systems models already provide a good account of the performance of the adult system for unpredictable target sequences [1]. Neural network models have an advantage over these systems in that, rather than being set up by hand to simulate adult performance, they can be set up to *learn* to perform the given task as best they can [2, 3]. The pattern of learning in the model should then correspond to the developmental changes found in children and the fully trained network should match the adult behaviour.

In principle, the modelling process is straightforward. The basic neural network architecture is given by known physiology and/or the existing systems models. We know what the network is meant to be learning to do, namely minimizing blur and disparity. So we 'simply' need to use some form of gradient descent weight learning to minimize an appropriate cost function, and then compare the resulting network performance with empirical child and adult data. In practice, the choice of cost function is not so straightforward. Whilst blur and disparity are fairly obvious error components, choosing an appropriate regularization component is less easy. This paper considers a range of reasonable possibilities and examines their consequences.

2. The Neural Network Model

The control systems for accommodation and vergence are very similar [1, 3, 4], so for clarity of analysis we shall concentrate on a simplified version of the vergence system as shown in Figure 1. Each neuron and the plant in the model are leaky integrators with empirically determined time constants. They represent the action of assemblies of real components, and activation flows between them via weighted connections. There is a direct correspondence with the equivalent systems models [3, 4]. The first pair of fast and slow neurons correspond to the standard phasic sub-system, the slow tonic and bias correspond to the standard tonic sub-system, and there is a time lag of 0.15s in the feedback loop. The systems models are linear and generally set up with

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 207-212

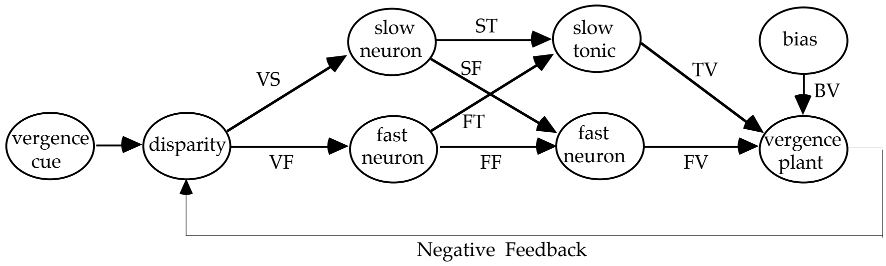

Figure 1: Simplified neural network model of the vergence system.

weights SF = ST = FF = FT = FV = 1. Assuming our neurons are to be linear over their operating ranges allows us to conveniently normalise VS = VF = FV = TV = 1. If our network's free parameters SF, ST, FF, FT and BV satisfy ST.FF = SF.FT, we have mathematical equivalence with the systems model. The bias BV always tends to assume a mean output value and has little effect on the discussion that follows.

As noted earlier, a crucial feature of our models is that, rather than setting their parameters by hand so that their performance matches adult human performance, we allow them to learn appropriate parameters to perform as best they can by carrying out gradient descent on an appropriate cost function. There are various maturational factors, such as the quality of the vergence cues changing with age, that affect the learning process [3]. Clearly, if we always minimize the cost function to the global minimum, such details will not matter, but in practice we often end up in local minima. In fact, one of the main reasons for formulating these models is to understand the causes of abnormal developmental trajectories in children with view to identifying precursors and designing remedial actions. The aim of this paper is to determine an appropriate cost function and explore how significant the choice of regularization is compared with other choices which may potentially lead to different local minima.

3. Regularization

The standard regularization approach [5] attempts to recover a function $f(\mathbf{x})$ from a set of data points $\{(\mathbf{x}_i, y_i) \in \mathrm{R}^d \times \mathrm{R}\}_{i=1}^{N}$ obtained by random sampling with noise. This is done by minimizing, e.g. by some form of gradient descent, a cost function such as

$$E[f] = \sum_{i=1}^{N} (f(\mathbf{x}_i) - y_i)^2 + \lambda \Phi[f]$$

with a parameterized trade-off between a sum-squared error term keeping f close to the data and a regularization term $\Phi[f]$ that enforces some form of smoothness.

Our problem is somewhat different, but has a similar solution. We have a fully dynamical system with feedback and training data parameterized by the time t. We require the network outputs (vergence responses) $f(x(t))$ to match the inputs (vergence cues) $x(t)$ as closely as possible given the network architecture and the constraint that $f(x(t))$ must be suitably smooth despite the time lag in the feedback loop and $x(t)$ frequently being discontinuous. We again have a trade-off between error and regularization components, so we can re-use the above cost function with the

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 207-212

summation over *i* replaced by an integral over *t*,

$$E[f] = \int (f(x(t)) - x(t))^2 \, dt \; + \; \lambda \Phi[f] .$$

In this paper we shall consider three natural forms of regularization functional:

a) $\Phi[f] = 0$

b) $\Phi[f] = \int \left| \frac{\partial f(t)}{\partial t} \right|^m dt$

c) $\Phi[f] = \int \omega^{2m} \, |F(\omega)|^2 \, d\omega \quad \text{where} \quad F(\omega) = \int f(t) \, e^{-i\omega t} \, dt$.

In practice we use discrete approximations to these integrals over finite ranges, so the simple relation between the m = 2 case *b* and m = 1 case *c* is broken. Case *a* is clearly identical to the λ = 0 limits of cases *b* and *c*, but it is worthy of separate consideration since it reveals the problem of output oscillations and the need for regularization in the first place. Case *b* attaches cost to the velocity of the eyes' movement as has been done previously [3]. Case *c* deals with the output oscillations more directly. Working with the Fourier transform $F(\omega)$ and the power $|F(\omega)|^2$ at frequency ω, allows us to penalize the high frequency components and reduce the oscillations in that way [5, 6]. For completeness, we also consider using an alternative error term based on the L1 norm $|f(x(t)) - x(t)|$ rather than the traditional sum-squared error. The remainder of this paper presents explicit simulation results that explore the consequences of these choices as a function of the trade-off parameter λ and the extent to which their differences are significant compared with those caused by other factors.

4. Simulation Results

The model was repeatedly trained to asymptote on random sequences of natural vergence values. Figure 2a shows the model's response to a step change of input when trained without regularization (case *a*). There is considerable overshoot and oscillation, which is not observed in humans. Figure 2b shows the more human-like response produced by a typical regularized model. We now compare the effect on performance of the different cost functions discussed above. To ease comparison

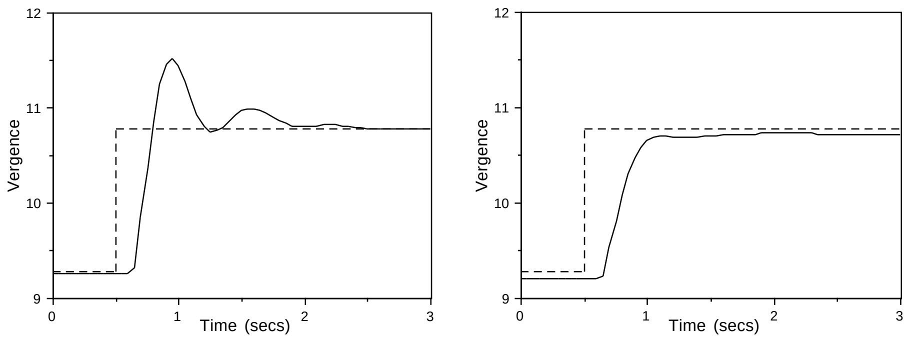

Figure 2: Model responses for (a) un-regularized and (b) regularized training.

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 207-212

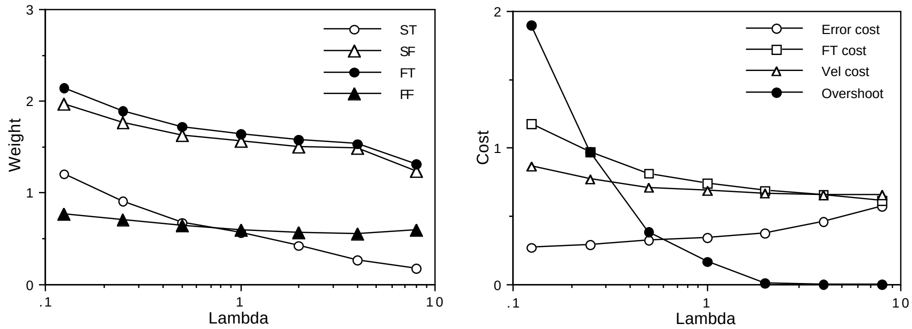

Figure 3: λ dependence of weights and costs for L1 error and velocity regularization.

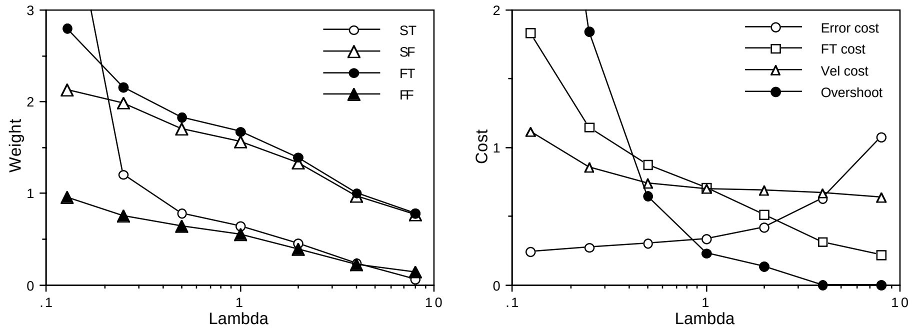

Figure 4: λ dependence of weights and costs for L1 error and Fourier regularization.

between the models, in each case we shall plot the L1 error, both $m = 1$ regularization costs, and an overshoot measure defined as the total vergence change (summed over oscillations) in the direction opposite to the standard step producing it. Figure 3 shows the effect of λ on the final weights and costs for L1 error and simple velocity regularization (case *b*, $m = 1$). We see, as expected, a clear trade-off between error and over-shoot as we increase λ. Figure 4 shows the equivalent plots for the simple Fourier transform regularization (case *c*, $m = 1$). In this case we must suffer a much larger error to remove the overshoot. Figure 5 shows what happens with sum squared error and sum squared velocity regularization (case *b*, $m = 2$). The error required for zero overshoot is larger again. Finally, Figure 6 shows that no better results are obtained for sum squared error with simple Fourier transform regularization (case *c*, $m = 1$). The remaining permutations and values of *m* perform even less well.

Taken together, Figures 3 to 6 show that, whilst an increase in trade-off parameter λ for any cost function results in a reduction in each of the velocity cost, Fourier cost and overshoot at the expense of increased error, there is considerable variation between cases. It follows that choosing a convenient cost function and showing that it can result in responses with human-like smoothness is not sufficient to support a claim that we have found the unique accurate model of human performance.

One common feature of the trained models in all four cases plotted is the tendency for SF ~ FT to be several times ST ~ FF. This is very different from the structure of typical systems models which, we noted above, are set up with ST.FF = SF.FT. For

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 207-212

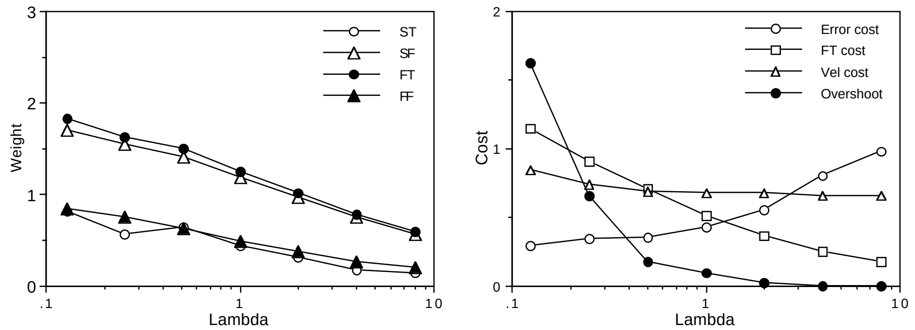

Figure 5: λ dependence for Squared error and Squared velocity regularization.

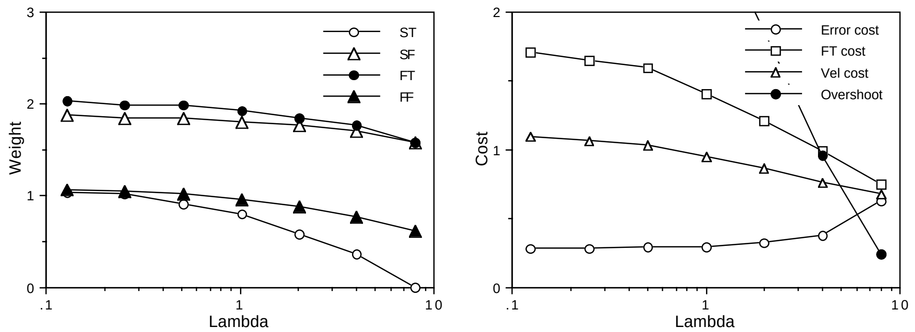

Figure 6: λ dependence for Squared error and Fourier regularization.

comparison, Figure 7 shows what happens when the models of Figure 3 are trained under this systems model constraint. We get a radically different pattern of weights, yet the costs and output response curves are hardly distinguishable. Moreover, if we remove the constraint on the weights and continue the training, we find that the weights are stable, suggesting that we have at least two roughly equivalent local minima considerably separated in weight space.

5. Conclusions and Discussion

We have explored the use of different cost functions in neural network models of oculomotor control and found that both velocity and Fourier transform regularization give good, but slightly different, final performance. Perhaps a more surprising result is the discovery that our network models naturally learn a somewhat different structure to that assumed in traditional engineering style systems models [1, 4]. The networks' output responses are very similar though. This means that it is no longer so obvious that the existing systems models should be considered a sensible starting point for our network models simply because they already provide a good account of human responses [3]. Rather, we should start again from known physiology, and if our models learn different structures to the systems models, we must either find fault with the performance of those systems models, or think more carefully about our modelling assumptions.

The above results suggest that, as long as we use near optimal values of the

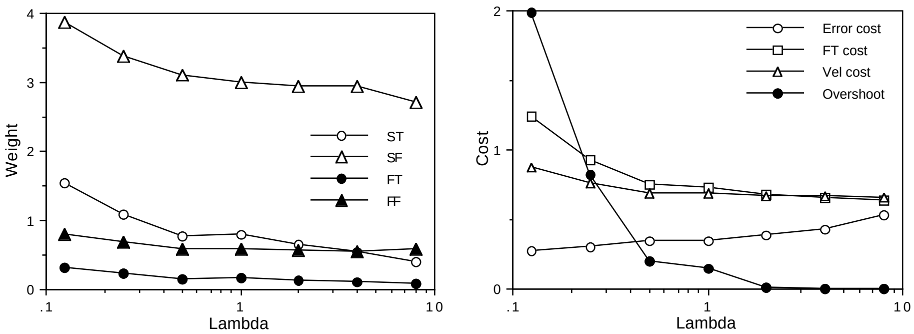

Figure 7: λ dependence for Systems Model parameterised version of Figure 3.

regularization parameters λ and *m*, factors other than the details of the regularization will have a more significant influence on the weights that are learnt. We have seen above that constraints on the weight patterns (such as might, for example, arise from innate brain layout) have a big effect. Further simulations suggest that allowing different weights to have different starting values and/or different learning rates can also result in the model ending up in different local minima. These features, as well as λ and the regularization function, have presumably been fine tuned by human evolution. This is probably a fruitful direction to pursue for future research in this area and neural network control systems more generally.

Finally, we have previously shown that regularization in the form of a simple polynomial weight cost can account for the empirical pattern of response gains found in vertical disparity adaptation [2]. It is natural to ask if the regularization functions used in this paper to smooth the transitions between vergence responses can also account for the vergence adaptation data. Unfortunately, preliminary simulations suggest that the answer is ‘no’.

References

1. Eadie, A.S. & Cardin, P.J. (1995). Evolution of control system models of ocular accommodation, vergence and their interaction. *Medical & Biological Engineering & Computing*, **33**, 517-524.
2. Bullinaria, J.A., Riddell, P.M. & Rushton, S.K. (1999). Regularization in oculomotor adaptation. In *Proceedings of the European Symposium on Artificial Neural Networks*, 159-164. Brussels: D-Facto.
3. Riddell, P.M. & Bullinaria, J.A. (1999). Incorporating developmental factors into models of accommodation and vergence. Submitted for publication.
4. Schor, C.M., Alexander, J., Cormack, L. & Stevenson, S. (1992). Negative feedback control model of proximal convergence and accommodation. *Ophthalmic and Physiological Optics*, **12**, 307-318.
5. Girosi, F., Jones, M. & Poggio, T. (1995). Regularization theory and neural network architectures. *Neural Computation*, **7**, 219-269.
6. Duchon, J. (1977). Spline minimizing rotation-invariant semi-norms in Sobelov spaces. In Schempp, W. & Zeller, K. (Eds), *Constructive Theory of Functions of Several Variables, Lecture Notes in Mathematics, 571*. Berlin: Springer-Verlag.