Tuning positive feedback for signal detection in noisy dynamic environments

Anders Johansson a,c, Kai Ramsch b, Martin Middendorf b, David J.T. Sumpter a,*

* Mathematics Department, Uppsala University, Uppsala, Sweden
b Institut für Informatik, Universität Leipzig, Leipzig, Germany
c Mathematics Department, University of Gävle, Gävle, Sweden

Abstract

Learning from previous actions is a key feature of decision-making. Diverse biological systems, from neuronal assemblies to insect societies, use a combination of positive feedback and forgetting of stored memories to process and respond to input signals. Here we look how these systems deal with a dynamic two-armed bandit problem of detecting a very weak signal in the presence of a high degree of noise. We show that by tuning the form of positive feedback and the decay rate to appropriate values, a single tracking variable can effectively detect dynamic inputs even in the presence of a large degree of noise. In particular, we show that when tuned appropriately a simple positive feedback algorithm is Fisher efficient, in that it can track changes in a signal on a time of order \( t \approx (|h|/\sigma)^{-2} \), where \(|h|\) is the magnitude of the signal and \( \sigma \) the magnitude of the noise.

1. Introduction

Many biological systems have evolved to detect and respond to noisy signals (Green and Swets, 1966; van Drongelen, 2006; Couzin, 2009; Smith and Ratcliff, 2004; Brandman and Meyer, 2008; Brandman et al., 2005; Sumpter and Pratt, 2009; Marshall et al., 2009). Neuronal assemblies integrate noisy sensory information to make decisions about how to respond to stimuli (Gold and Shadlen, 2007; Feng et al., 2009; Bogacz et al., 2006; Brown et al., 2005; Simen et al., 2006). Ant colonies use pheromones to integrate information about the location of food and other resources (Detrain and Deneubourg, 2008; Bonabeau et al., 1997). These systems accumulate information based on past experience and continuously learn the best action to take. While one can argue whether, for example, biochemical systems can be said to be ‘learning’ or ‘making decisions’, the function of these systems is similar at many different levels of biological organization (Trewavas, 2009). Given weak, noisy evidence for one of a number of options these systems attempt to choose the best available option.

A striking observation about the above systems is the similarities in the mechanisms underlying decision-making (Couzin, 2009; Marshall et al., 2009). The process by which a system learns can be described in terms of one or more tracking variables which encode the system’s propensity to choose a particular option. Furthermore, the dynamics of these tracking variables usually involve three key components: (1) an increase in response to the noisy input signal; (2) a sharp increase at a threshold of the input signal or the tracking variable itself; and (3) an exponential decay in the absence of external signal. We now describe three examples with these components and to which the models now developed in this paper should apply: ant pheromone trails, neuronal assemblies and biochemical reactions.

Many species of ants when foraging for food leave pheromone trails which track the existence and quality of food at various locations. When offered the choice between \( n \) options the amount of pheromone at time step \( t + 1 \) leading to option \( i \) can be modelled as a difference equation:

\[
C_{t+1}(i) = (1 - \lambda)C_t(i) + \lambda \left( \frac{C_t(i) + K^2}{\sum_{j=1}^{n} (C_t(j) + K)} + h \right)
\]

where \( \lambda \) is the evaporation rate, \( h \) is the input signal indicating the quality of option \( i \), and \( \lambda \) and \( K \) determine the shape of the
response (Nicolis and Deneubourg, 1999). When \( \alpha > 2 \) the second term in Eq. (1) provides a sigmoidal-like response, whereby if one of the \( C(i) \) exceeds all the other \( C(j) \) then it becomes disproportionally likely to be chosen on future time steps. Similar responses, sometimes referred to as quorums, have been observed in the decision-making of a wide range of other group-living animals (Ame et al., 2006; Sumpter et al., 2008; Sumpter and Pratt, 2009). This model has been widely studied and validated against experimental data (Deneubourg et al., 1990; Goss et al., 1989; Dussutour et al., 2005; Bonabeau et al., 1997). Here the variables \( C(i) \) are deterministic, but a corresponding stochastic model is easily constructed (Nicolis et al., 2003). This model forms the basis for the feedback in the ACO metaheuristic (Dorigo and Stützle, 2004) and other applications in swarm intelligence (Bonabeau et al., 1999).

The leaky competing accumulator model for two competing neuronal assemblies \( x_1 \) and \( x_2 \) is a stochastic differential equation with the form (Usher and McClelland, 2001; Feng et al., 2009; Brown et al., 2005):

\[
\begin{align*}
\dot{x}_1 &= -\lambda x_1 - \beta f(x_1) + h_1 + \sigma dW_1 \\
\dot{x}_2 &= -\lambda x_2 - \beta f(x_2) + h_2 + \sigma dW_2
\end{align*}
\]

(2)

(3)

where \( \lambda \) is the rate of memory decay, \( h_i \) is the input signal, and \( \beta \) is a factor controlling the influence of the competing neuron. The noise is now explicitly represented as standard Brownian motion or a Wiener process with standard deviation of size \( \sigma \). As in Eq. (1), the function \( f(x) \) is defined to be sigmoidal, but instead of \( x_1 \) having an excitatory effect on itself, \( x_1 \) laterally inhibits \( x_2 \) and vice-versa. As a result the tracking variable for which \( h_1 \) is greater becomes disproportionally stronger. For parameter values where decision accuracy is optimised, this model fits experimental data in which monkeys attempt to ascertain the direction of a noisy random dot stimuli (Feng et al., 2009).

Sigmoidal positive feedback responses are a ubiquitous feature of biochemical systems (De Jong, 2002). Indeed, it is difficult to write down a model of gene regulation or molecular dynamics that does not involve some form of sigmoidal or Hill response function. A basic function of cells during, for example, development is to track and respond to signals emitted by other cells (Lewis, 2008). When interacting with each other and their environment, these systems employ quorum-like responses to filter out noise in the signals they receive (Tanouchi et al., 2008; Andrews et al., 2006).

Here we investigate tracking variables for decision-making in a dynamic noisy environment. We first present a dynamic problem of tracking the difference between two variable and propose a learning system based on a tracking variable which encodes an agent’s memory. We identify the optimal parameters for this model and show that they depend on the strength of the signal, implying that the system is scale dependent. We extend this result to the case where there are two alternatives given and we wish to determine which gives the highest expected reward at the present point in time. We show how this result may be used to tune positive feedback algorithms to respond optimally in dynamic environments.

2. Tracking reward differences

We consider the following setup for decision-making in a noisy environment. On each time step \( t \) an agent chooses one of two actions, \( A_t \in \{-1,1\} \), on the basis of observations of a single signal (all symbols used in the paper are summarised in Table 1). For example, in chemotaxis some bacteria species set tumbling frequency on the basis of noisy information about the nutrient gradient they experience (Andrews et al., 2006; Barkai and Leibler, 1997). For some hidden signal \( h_t \), the observed signal is

\[ x_t = h_t + N_t \]

(4)

where \( N_t \sim N(0,\sigma^2) \) are Gaussian random variables of mean zero and variance \( \sigma^2 \).

We make the problem of learning the correct action difficult by making the magnitude \( |h_t| \) small relative to the magnitude of the noise \( \sigma \). Furthermore, over time \( h_t \) changes its sign and we allow the agent no a-priori knowledge of the pattern of these changes. Our aim is to design a system which detects changes in \( h_t \) of a particular magnitude as rapidly as possible. Typically, we assume \( \sigma \) to be known and constant. We will also generally assume that the signal \( h_t \) is a piecewise constant process as in Fig. 1. These assumptions are not essential to what follows, but it does allow us to simplify the discussion. We note that if \( h_t \) is equal to a constant \( h \) in the interval \( [0,t] \) then the average \( x_t = \frac{1}{t} \sum_{s=0}^{t} x_s \) has distribution \( N(h, \frac{h^2}{2t}) \).

There are proven limits to how fast one can detect a signal subjected to a certain level of noise. The Cramer (1946) and Radhakrishna Rao (1945) inequality and its generalizations, such as the Chapman and Robbins (1951) bound give a fundamental lower bound on the time necessary to sample in order to detect the sign of \( h_t \). In our scenario the Chapman–Robbins bound shows that in order

<table>
<thead>
<tr>
<th>Table 1</th>
<th>List of parameter, variables and symbols used and their interpretation in the tracking systems.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Param.</td>
<td>Description</td>
</tr>
<tr>
<td>( h_t )</td>
<td>Hidden signal to be detected at time ( t )</td>
</tr>
<tr>
<td>( N_t )</td>
<td>Gaussian noise at time ( t )</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>Standard deviation of Gaussian noise</td>
</tr>
<tr>
<td>( x_t )</td>
<td>Observed signal at time ( t )</td>
</tr>
<tr>
<td>( h )</td>
<td>Constant absolute value used for ( h_t )</td>
</tr>
<tr>
<td>( l(h) )</td>
<td>Lower bound for detection of a signal ( h )</td>
</tr>
<tr>
<td>( y_t )</td>
<td>Tracking variable at time ( t )</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>Step size of memory decay</td>
</tr>
<tr>
<td>( \beta )</td>
<td>Strength of non-linearity</td>
</tr>
<tr>
<td>( y_s )</td>
<td>Stable fixed point of ( y_t )</td>
</tr>
<tr>
<td>( \lambda_{op} )</td>
<td>Operational value of ( \lambda )</td>
</tr>
<tr>
<td>( s )</td>
<td>Performance of simulation run</td>
</tr>
<tr>
<td>( r(A_t) )</td>
<td>Reward an agents obtains from action ( A_t ) at time ( t )</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>Constant reward obtained</td>
</tr>
<tr>
<td>( z_{+1,-1,f} )</td>
<td>Pheromone values for action +1 or -1 at time ( t )</td>
</tr>
<tr>
<td>( \beta_{op} )</td>
<td>Operational value of ( \beta )</td>
</tr>
<tr>
<td>( \lambda_{op} )</td>
<td>Parameter values for ( \lambda ) maximising the performance ( s )</td>
</tr>
</tbody>
</table>

![Fig. 1.](image-url)
to infer, with a small error probability, the sign of \( h = E(x_t) \), we always need to observe the process for a time of order \( L(h) \), where \( L(h) = (|h|/\sigma)^{-2} \). We call the ratio \( |h|/\sigma \) the signal-to-noise ratio since \( h_t \) and \( \sigma \) can be identified with the signal and noise amplitudes, respectively. Amari (1998) uses the term Fisher efficient to describe systems that learn within a time given by the Cramér–Rao bounds. In this paper, we use the term Fisher efficient to describe systems that learn within a small constant multiplied by \( L(h) \). In the dynamic setting the lower bound of Fisher efficiency implies that we must always expect a delay in our knowledge of the sign of \( h_t \) in order of a small constant multiplied by \( L(h_t) \).

When \( h_t = h \) does not change in time there is a Fisher efficient and scale-robust methodology for estimating \( h \) for all \( h > 0 \). Under the classical stochastic approximation paradigm one would take the sign of the average input signal \( A_{t-1} = \text{sign}(x_t) \), and obtain that \( A_t \to \text{sign}(h) \) (see, e.g., Borkar, 2008). Calculating the log-likelihood of the hypotheses \( h > 0 \) and \( h < 0 \) to the sequence \( \text{sign}(x_t) \) provides confidence intervals for this estimate. However, these methods cannot be applied when \( h_t \) is changing dynamically. For example, if \( h_t \) oscillates symmetrically around zero then the stochastic approximation will, due to a failure to unlearn its previous samples, get locked in to the one alternative chosen first. Various extensions of the stochastic approximation paradigm attempt to adjust the time-scale (i.e., the constant step-size \( \lambda \) in (5)) to the case of slowly varying signals (see Kushner and Yin, 1997, Chapter 3 for references).

We now take inspiration from the models of biological systems discussed in Introduction, to construct a means of tracking the signal using a single ‘pheromone’ or memory variable. To avoid being too specific about the type of biological system we refer to, we call this ‘pheromone’ or memory a tracking variable and denote it by \( y_t \). The variable \( y_t \) is updated on the basis of the difference between the rewards, i.e. \( x_t = h_t + N_t \), \( s < t \), and transmits information about the sign of \( h_t \), although not necessarily information about the exact value of \( h_t \). We stipulate that \( y_t \) has threshold-like positive feedback on itself, so that when \( y_t \) is positive it increases and when \( y_t \) is negative it decreases. Specifically,

\[
y_{t+1} = \frac{f(y_t)}{\exp(\hat{y}_t)} \left[ \exp(\hat{y}_t) + \exp(-\hat{y}_t) \right] - \frac{\exp(-\hat{y}_t)}{\exp(\hat{y}_t) + \exp(-\hat{y}_t)} + x_t
\]

\[
= (1-\lambda)y_t + \lambda \text{sign}(y_t) + x_t
\]

This tracking variable is designed to have the same sign as \( h_t \). The term \((1-\lambda)y_t\), the tracking signal decay, the function \( tanh \) determines the degree to which a signal is reinforced and \( x_t \) is the signal itself. \( \beta \) is the strength of the non-linearity. The action or choice of the agent can be obtained by the tracking variable simply as \( \text{sign}(y_t) \).

We now investigate properties of our proposed tracking variable, \( y_t \). When \( \beta > 1 \) and \( h > 0 \) the function \( f(y) \) in (5) has two stable fixed points and one unstable fixed point at \( 0 \) (see Fig. 2). In this case, the addition of a bias, i.e., \( h > 0 \), does not necessarily lead to a switch to the steady state associated with the higher rewards. For example, if \( \beta = 1.1 \) and \( h < 0.019 \) then the suboptimal choice remains a stable fixed point when the bias is added.

For \( \beta < 1 \) there exists a single stable fixed point \( y_* \) for \( F(y_*, h) = y_* \). In order to track a signal \( h > 0 \) with maximum effectiveness this fixed point should be as far away as possible from zero. The fixed point is the (in this case unique) solution to

\[
y_* = \text{tanh}(\beta y_*) + h
\]

Since \( \text{tanh}(y) \) is monotone increasing, \( y_* \) is maximised for \( \beta = 1 \). Furthermore, for all \( y < y_* \), \( F(y, h) \) is maximised when \( \beta = 1 \). We thus conclude that, for arbitrary \( h \), \( \beta = 1 \) is the only parameter which guarantees a single fixed point, while providing the strongest tracking of that signal.

When \( \beta = 1 \) and \( h > 0 \) is small we, using a Taylor expansion, find an approximate solution to Eq. (6),

\[
y_* \approx 3^{1/3} h^{1/3}
\]

The rate of convergence to this fixed point is determined by

\[
F(y_*, h) = 1-\lambda \text{tanh}(y_*) - 1-\lambda(y_* - h)2 \approx r = 1(y_* - h)^2 - 3^{2/3} h^{2/3}
\]

For \( y \) near to \( y_* \) we can approximate Eq. (5), this time accounting for the noise \( N_t \), with the linear stochastic process:

\[
y_{t+1} = F(y_t, h + N_t) - y_* = F(y_t, h) - y_* + \Delta N_t
\]

\[
= F(y_t, h) + \lambda \text{tanh}(y_t) - y_* + \Delta N_t
\]

In its linearised form this model is equivalent to the standard Uhlenbeck and Ornstein (1930) process. Fluctuations around \( y_* \) thus have variance

\[
\frac{\lambda^2 \text{Var}(N_t)}{1-r^2} \approx \frac{\lambda^2 \sigma^2}{2 \cdot 3^{2/3} h^{2/3}} \leq \frac{\lambda^2 \sigma^2}{4 h^{1/3}}
\]
**Lemma 1.** Assume that \( h_t = h > 0 \) and that \( y_0 \in [-1, 1] \). For every \( \epsilon > 0 \) and \( \delta > 0 \), there is a \( c = c(\delta, \epsilon) > 0 \), such that

\[
\Pr[y_t > (1 - \delta)y_s] > 1 - \epsilon
\]

if

\[
\lambda = c \cdot \sigma^2 h^{4/3}
\]

and

\[
t \geq \frac{1}{c} \frac{1}{\lambda h^{2/3}} = \frac{L(h)}{c^2}
\]

for a sufficiently small \( c < 1 \).

This lemma is proved in Appendix. We make no attempt to compute the best possible bound on \( c \).

To test these predictions against numerical simulations, we simulated Eq. (5) for different levels of a fixed bias \( |h| \) and decay rate \( \lambda \). We made 500 runs for three different levels of \( |h| \), each run consisting of \( 2L(h) \) iterations of Eq. (5). Since the theoretical minimum for signal detection is \( L(h) \), this gives the algorithm a comparatively

---

**Fig. 3.** Simulation of the signal tracking system (Eq. (5)) for \( |h| = 0.0156 \) and (a) \( \lambda = 0.0338 \), (b) \( \lambda = 0.0085 \) and (c) \( \lambda = 0.0021 \). The + and − indicate the sign of \( h \) for that particular part of the simulation. When \( \text{sign}(y_t) = \text{sign}(h_t) \) then the signal is successfully tracked. The sign of \( h \) is switched every \( 2\sigma^2h^2 \) time steps.
short time to detect the signal. The first run had $h_0 = h$ and consequent runs changed the sign of $h$, i.e., even runs had $h_t = -h$ and odd runs had $h_t = h$. At the start of a new run $y_0$ was initialised with the value $y_t$ established at the end of the previous run, so as to simulate dynamic changes in the bias. The performance of the system was measured by the mean performance over the last $L(h)$ steps of each run. Specifically,

$$s = \frac{1}{L(h)} \sum_{t = L(h) + 1}^{2L(h)} \text{sign}(y_t \cdot h_t)$$

measures the performance of a run.

Fig. 4(a–c) shows the performance of the first 40 runs as a function of $\lambda$. As in Fig. 3, large values of $\lambda$ produce a sequence of $y_t$ which are oversensitive to noise and thus fluctuate around zero (seen as yellow/green on the right hand side of panels a–c). Very small values of $\lambda$ produce a situation where $y_t$ fails to detect when $h_t$ changes, so that $y_t$ is correct every second time $h_t$ changes sign, but nearly always incorrect for just a single change (seen as alternating rows of red and blue on the left hand side of the figure). The optimal value of $\lambda$ is determined by Eq. (9). Here, the signal is correctly identified 75% of the time (Fig. 4d–f).

### 3. Exploration vs exploitation

In the above setup the agent can observe $x_t$ directly on every time step. In many of the examples discussed in the Introduction, decision-making agents can only sample one of the available options on each time step. This case is relevant to situations where an agent must balance exploration with exploitation. For example, the agent might establish that one of the options is more profitable, but if it does not sample the other one from time to time then it cannot detect changes in relative profitability. Social insect colonies face such challenges in deciding where to forage for food (Seeley, 1995; Bonabeau et al., 1997). In this case, the agent only has information about the reward chosen on the previous time step.

To model this we stipulate that dependent on its action the agent obtains the reward:

$$r_t(A_t) = V + \frac{1}{2} A_t h_t + \frac{1}{\sqrt{2}} N_{t,A_t}$$

where $V$ is a constant. The underlying signal $h_t$ again takes positive, or respectively negative, values this time corresponding to action 1, or respectively $-1$, being the most rewarding choice. $N_{t,A_t} \sim N(0, \sigma^2)$ are independent and identically distributed.
Gaussian random variables of mean zero and variance $\sigma^2$. The agent’s goal is to maximise its long term reward. Since $E(N_{i+1}) = E(N_{i}) = 0$, the reward is maximised when $A_i$ has the same sign as $h_i$. We measure agent success with the score, $A_i \cdot \text{sign}(h_i)$, which is one whenever the signal has the same sign as $h_i$. The question for the agent is how to choose the actions so as to maximise this score.

To investigate such a setup we let $z_{1,t}$ and $z_{2,t}$ be the “pheromone” placed on taking actions $-1$ and $1$ respectively. On each time step we set the probability of taking action $1$ to

$$P(A_{t+1} = 1 | z_{1,t}, z_{2,t}) = \frac{\exp(\beta z_{1,t})}{\exp(\beta z_{1,t}) + \exp(\beta z_{2,t})} \tag{15}$$

and $P(A_{t+1} = -1 | z_{1,t}, z_{2,t}) = 1 - P(A_{t+1} = 1 | z_{1,t}, z_{2,t})$. The update rules for $z_{1,t}$ and $z_{2,t}$ are

$$z_{1,t+1} = (1-\lambda)z_{1,t} + \lambda \left(1 + r_{t+1}(1)\right) \tag{16}$$

and

$$z_{2,t+1} = (1-\lambda)z_{2,t} + \lambda \left(1 - r_{t+1}(1)\right) \tag{17}$$

so that each pheromone is only increased if the action is taken. Here $r_{t+1}(A)$ refers to the reward of action $A$ in Eq. (14).

We can approximate the above system to a one dimensional system based on the difference $y_t = z_{1,t} - z_{2,t}$ between the pheromones:

$$y_{t+1} = (1-\lambda)y_t + \lambda(1 + r_{t+1})h_t + N_t + A_{t+1}V$$ \tag{18}

From this we can use the results in the previous section to set the parameters. First, we note that

$$E(A_{t+1}) = \frac{\exp(\beta y_t)}{\exp(\beta y_t) + 1} = \tanh\left(\frac{\beta}{2} y_t\right).$$

Thus, based on the argument illustrated in Fig. 2, $\beta_{op} = 2$. Second, we note that Eq. (18) is identical to Eq. (5) apart from an additional term $A_{t+1}V$. Thus provided $V$ in Eq. (14) is small ($V \ll 1$), then the sampling presented here should work in a manner similar to the system for the difference tracking scenario (i.e., Eq. (5)).

We thus predict that for $\beta = 2$, $\lambda_{op}$ should be proportional to $h^{4/3}$. To test this prediction, we simulated the single variable tracking system for the sampling scenario with $\beta = 2$, $V = 0$ and $\lambda \in \{4^{-2}, 4^{-2}, \ldots, 4^{-20}\}$ for 50 runs, each of length $20h$. We examined the performance:

$$s = \frac{1}{20h} \sum_{t=1}^{20h} \text{sign}(y_t, h_t)$$

averaged over all simulation runs for differing combinations of $\lambda$ and $h$. Fig. 5 shows the values $\lambda$ which gave the largest $s$ averaged over all $20 \cdot (1/h^2)$ iterations. These observations confirm our prediction for each bias $h$ tested. Furthermore $\lambda$ is also robust with respect to changes of the interval of iterations over which performance was evaluated. For instance, selecting only the iterations of a subinterval, e.g.,

$$s = \frac{1}{10h} \sum_{t=20h+1}^{120h} \text{sign}(y_t, h_t)$$

yields a vertical shift of the $\lambda$ values in Fig. 5, but the relationship to the bias ($\lambda \approx h^{4/3}$) remains unchanged.

Are there other combinations of $\lambda$ and $\beta$ which outperform those for $\beta = 2$? Fig. 6 shows the mean performance $s$ for different combinations of $\beta$ and $\lambda$ for two biases. Here again we see that $\beta = 2$ corresponds to high success rates for suitable values of $\lambda$ and there is a local maximum close to $\lambda_{op} \approx h^{4/3}$. However, there exist other maxima for $\beta$ slightly greater than $2$. In particular, $\beta = 2.3$ and $\lambda = 2^{-7} \approx 0.0078$ perform best when $h = 10^{-2}$, and $\beta_{op} = 2.2$ and $\lambda_{op} = 2^{-3} \approx 0.0019$ perform best when $h = 10^{-3}$. This pattern is repeated for other values of $h$ (results not shown). These parameter values correspond to the existence of two stable states (as in Fig. 2a). However, since $\beta$ is not much greater than $2$ when $h$ gets closer to $0$, the stability of these states is marginal and fluctuations usually allow an escape to the better option. The distribution of decisions shows that the system converges less often to the correct choice for $\beta > 2$, but when it does it makes correct choices more often.

There are several ways of improving the performance of the above tracking system if one would like to apply this problem to optimisation problems. For example, we can set

$$z_{1,t+1} = (1-\lambda)z_{1,t} + \lambda\frac{1+A_{t+1}}{2}$$

and

$$z_{2,t+1} = (1-\lambda)z_{2,t} + \lambda\frac{1-A_{t+1}}{2}$$

with

$$r_{t+1} = \begin{cases} r_{t+1} : A_{t+1} = i \\ r_{t+1} : A_{t+1} = -i \end{cases}$$

and $r_{t+1}$ being the mean of the last $k$ rewards sampled from decision $i$. This eliminates a sampling bias whereby the reward of an action degrades simply because it has not been chosen.

In order to concentrate on our method for tracking $h_t$ in the presence of large noise, we have skipped over two problems that will arise in practical applications signal tracking. First, as mentioned above, the technique depends on having $V$ small, otherwise the agent must determine the magnitude of the two rewards, determined by $V$ in Eq. (14). This can be done by standard methods of gain control, removing the average of $r_t$ from the signal. Second, in setting $\lambda$ the agent must also estimate...

---

**Fig. 5.** Optimal decay rate $\lambda$ of the tracking system for the sampling scenario (Eqs. (15)–(17)) having $\beta = 2$ fixed and $h \in \{10^{-4}, 10^{-3}, \ldots, 10^{-1}\}$, simulated for 50 runs of length $20 \cdot (1/h^2)$ each having alternating bias.
σ (which we have set to 1). This can be addressed by setting $\hat{\sigma}_t$ to be the standard deviation of the recent rewards.

4. Discussion

We have shown how systems used to model pheromone trails, neuronal interactions and biochemical kinetics can be tuned to solve dynamic learning problems. By setting the steepness of the reaction threshold $\beta$ and the stochastic evaporation or forgetting rate $\lambda$ to appropriate values we can tune these systems to quickly react to a particular input $h$ even in the presence of a large degree of noise, i.e., $\sigma \gg |h|$. The tuning of $\beta$ has already been appreciated under the rather grand title of tuning to the “edge of chaos” (Bonabeau, 1996, 1997).

The tuning of the stochastic forgetting rate $\lambda$ proportional to $h^{4/3}$ is a more subtle result. It can only be recovered through the stochastic analysis presented here. The key result is that the optimal $\lambda$ depends on the magnitude of the signal to be detected. Equations similar to (16) and (17) are widely used, for example, in ant algorithms, to solve optimisation problems (Dorigo and Stützle, 2004). It is notoriously difficult to set parameters of these algorithms to values which allow them to search effectively and account for aspects, such as, $V \gg 1$ and the existence of more than three alternatives. Even for three alternatives, the system corresponding to the function $F_y(x)$ is not immediately identifiable. Addressing these limitations is not straightforward, but we believe the techniques outlined here are a valuable start.

The relationship between $\lambda$ and $h^{4/3}$ implies that any fixed value of $\lambda$ will not be optimal for all possible signal strengths, and thus fails to be robust to changes in scale. Biological systems which track signals are likely to be either tuned to particular signal to noise ratios or contain an additional aspect, not captured in previous models, to allow tuning to signal strength. One possible way to allow scale-free tuning to signal strength is to couple different tracking variables, each of which decays on a different time scale. There may be parallels to be drawn here to the idea of biochemical and genetic networks which allow the production of robust patterns in spite of intrinsic noise (Barkai and Leibler, 1997; Brandman et al., 2005). Future work will develop this multiple tracking variable system, in which one particular tracking variable catches the input signal and passes it up to higher-level variables.

Acknowledgements

This work was funded by the Human Frontiers Science Program grant RGP51/2007.

Appendix

Proof of Lemma 1. The typical fluctuations $y_t - y_\star$ are determined by the magnitude of the noise $\sigma$ and the rate of contraction to the fixed point. Consider the Taylor-approximation:

$$(\tanh(xh^{4/3}) - xh^{4/3} + h)/h \approx 1 - \frac{1}{2}x^2 + O(h^{2/3}x^5)$$

A simple calculation shows that for $y < y_\star$ and as long as this approximation is valid, i.e., for $y = o(1)$, we have

$$y_{t+1} - y_t < r(y_t - y_\star) + \lambda N_t, \quad y < y_\star$$

where $r \leq 1 - 2/3h^{2/3}$, say. Thus we can deduce that $y_{t+1} - y_t$ is majorised by a discrete linear diffusion process:

$$u_{t+1} = ru_t + \lambda N_t, \quad u_t \geq 0$$

with a reflecting barrier in $u = 0$. Then $u_t = |W_t + r^t u_0|$, where $W_t$ is Gaussian with zero mean and variance $v^2_t$ bounded by

$$\lambda^2 \sigma^2 / (1-r)2 \leq \frac{\lambda^2 \sigma^2}{4h^{4/3}} \leq \frac{\lambda^2 \sigma^2}{4h^{4/3}}$$

for $c < 1$. We can choose $c$ to determine the proportion of time which $y_t$ spends within some distance of $y_\star$. Thus, $\Pr[y_t > (1-\epsilon) y_\star] < 1 - \epsilon/2$ provided that $u_0 = y_\star - y_0 = O(h^{1/3})$ and that $c$ is chosen small enough.

By the Markov property of $y_t$, it remains to show that for any initial state $y_0 = O(1)$, we have $y_t > -h^{1/3}$, with probability greater than $1 - \epsilon$ for $t = O(L)$. For the reason for this is that $y \rightarrow F(y,h)$ is a contractive mapping for $y < -h^{1/3}$. Let $\bar{y}_0 = y_0$ and define the deterministic process

$$\bar{y}_{t+1} = F(\bar{y}_t, h) + \lambda (\tanh(\bar{y}_t) - \bar{y}_t + h)$$

and let $z_t = y_t - \bar{y}_t$. The convexity of $F(y,h)$, $y < 0$, implies that

$$y_{t+1} = F(y_t + z_t, h) + \lambda N_t \geq \bar{y}_{t+1} + \lambda N_t + r(y_t) z_t$$

(21)
where, for $y < h^{1/3}$,

$$r(y) = \frac{\partial^2}{\partial y^2}(y/h) = 1 - \lambda \left(1 - \frac{1}{\cosh^2 y}\right) < r = 1 - \frac{1}{2} h^{2/3}.$$

Since

$$\tan(u) - u + h \geq \frac{2}{\sqrt{3}} \cdot u^3, -1 < u < 0,$$

we can approximate $\tilde{y}_t \approx u(s)$ with the differential equation:

$$\frac{ds}{dt} \leq \frac{2}{\sqrt{3}} \cdot u(s)^3$$

where $s = \lambda t$. The solution

$$u(s) \geq -\frac{1}{\sqrt{2\pi} - u(0)}$$

shows that for the deterministic system $\tilde{y}_t \leq -h^{1/3}/2$ if, say, $s \geq 2 h^{-2/3}$, i.e., $t \geq 2 h^{2/3}$.

It thus remains to show that $Pr(z_t > -h^{1/3}/2) > 1 - \epsilon$, for some $t = L/\epsilon^2$. But, from (21) above, it follows that $z_t$ is bounded below by the linear diffusion $u_{t+1} = u_t + \lambda N$, $u < 0$, reflected in $u_{t+1} = 0$. Hence, $z_t \geq -W_t$, where $W_t$ is Gaussian mean zero with variance bounded by $\lambda^2 \sigma^2/(1-r) \geq 2 \lambda \sigma^2 h^{2/3}$. Thus it is clear that adjusting $c$ small enough gives the sought for bound on $Pr(z_t > -h^{1/3}/2)$.

References