ELSEVIER

Journal of Neuroscience Methods 83 (1998) 73–88

JOURNAL OF NEUROSCIENCE METHODS

On the optimal control of behaviour: a stochastic perspective

Christopher M. Harris *

Department of Ophthalmology, Great Ormond Street Hospital for Children NHS Trust and Institute of Child Health, University College London, London WC1N 3JH, UK

Received 30 November 1997; received in revised form 9 February 1998; accepted 19 February 1998

Abstract

Evolution is a closed stochastic optimisation process driven by the interaction between behaviour and environment towards local maxima in fitness. It is inferred that nervous systems are selected to provide optimal control of behaviour (the 'assumption of optimality'), such that for some behaviours, the expectation of future hazards to survival are minimised. This is illustrated by goal-directed saccades in which minimising total flight-time of primary and secondary movements provides a better fit to observations than simply minimising the error of the primary movement. This optimisation is extended to intra-movement trajectories, where low-bandwidth (smooth) velocity profiles provide a more satisfactory description of observations than simple bang-bang control. Since minimum-time behaviours cannot be controlled by error feedback, it is concluded that the cerebellum must be executing a real-time unreferenced optimisation process. This requires explorative as well as exploitative behaviour. Stochastic gradient descent is discussed as a possible means by which the cerebellum may optimise behaviour.

Keywords: Saccades; Optimal control; Adaptive control; Reinforcement learning; Cerebellum; Infant development; Arm movements; Reaching

1. Introduction

> Dr Pangloss taught metaphysico-theologico-cosmo-codology. He would say, 'For everything having been made for a purpose, everything is necessarily for the best purpose. Observe how noses were made to bear spectacles, and so we have spectacles. Legs are evidently devised to be clad in breeches, and breeches we have… and since pigs were made to be eaten, we eat pork all year round. Consequently, those who have argued that all is well have been talking nonsense. They should have said that all is for the best'. (from Candide, Voltaire, 1759.) (Voltaire, 1990)

The notion that the form and function of organisms (especially humans) are ideal, or tend towards an ideal, has a very long history in human thought. The Darwinian revolution dispelled the belief that organisms were *designed* by some external agent, and replaced it with the concept of *evolution* through natural selection made possible by *random* mutations (of genes), thus sweeping away many centuries of teleology.

The engine of this stochastic evolutionary process is the interaction between behaviour and the environment, which through reproduction (or its lack), modifies the gene pool. In a world of limited resources, competition within and between species leads to fitter individuals, or 'survival of the fittest', where fitness is the ability of the individual to orchestrate its behaviour to survive, procreate, and parent offspring, contingent on the vagaries of the environment (including other organisms). Since fitness is determined by behaviour, there will be a selection for those nervous systems that

* Tel.: +44 171 4059200; fax: +44 171 8298647; e-mail: chris@vissci.ion.ucl.ac.uk

0165-0270/98/$19.00
PII S0165-0270(98)00063-6

can provide fitter behavioural control. Thus we expect an evolutionary trend towards 'optimal control', where optimality is gauged in fitness.

The purpose of this article is to clarify the issues in understanding behaviour as an optimal process by examining saccadic eye movements as a simple example. We emphasise that this is not an engineering question of how to design an optimal system, but about asking whether, or in what sense, an evolved biological behaviour, or class of behaviour, is optimal. Hence, this is not a prescription of analytic or numerical optimisation techniques because there is no 'correct' technique; different approaches are needed depending on the particular behavioural parameter under consideration.

For the natural scientist there are two fundamental unknowns in examining whether, or in what sense a performance is optimal. First, we need to know what is being optimised in a behaviour; that is, what parameter(s) affect fitness. In most texts on optimal control, the quantity that is being optimised (actually minimised) is called the performance index (PI), or cost function, which is specified by the design problem. Our problem is that we have an existing *evolved* system, but do not know the PI. Second, optimality only makes sense when there are constraints. In optimal design they are usually determined by the limitations of the components (weight, fuel capacity, finite control signals, etc.). In biological systems we are usually unsure as to which limitation is the constraining influence. For a given PI, different constraints lead to different optimal solutions, thus there may not be a unique solution to the problem.

Obviously, we will not be able to *prove* that a behaviour is optimal, since we would be unable to *prove* the validity of a particular set of PI and constraints. Likewise we cannot prove that a behaviour is not optimal since failure to match theoretically 'optimal' behaviour with observed behaviour may merely indicate that we have chosen an inappropriate PI and–or constraints. In reality, the more interesting question is to take a Darwinistic viewpoint and assume that the behaviour is optimal (the *assumption of optimality*) and find a set of PI and constraints that yield an optimal solution that agrees with observations, and use this to make further predictions. This is not simple, but if it can be found, then we have a theory of *what* the brain (evolution) is trying to achieve with the behaviour. We can then ask *how* this might be attained.

The remainder of this article is divided into four sections. In section 2 we consider the broad problem of optimising fitness from the viewpoint of survival. It becomes readily obvious that just reducing the immediate cost of a behaviour may not be optimal in the long-run, but future consequences need to be taken into account. To limit the scope, we shall focus on the issue of optimal *performance*, which is the way in which a behaviour is executed (speed, accuracy, timing etc.), rather than optimal *strategy*, which is the decision whether to make a behaviour or not. The latter belongs to the realm of Game Theory and will not be discussed here.

Many of the issues of optimal performance are well illustrated by open-loop reaching behaviours, such as saccadic eye movements and fast arm movements. These behaviours have been considered as 'goal-directed' and there is a small but quantitative literature on optimality. Our premise is that these movements are evolutionary important to human fitness, and consequently they have evolved to be optimal. In Section 3, we consider the accuracy of human goal-directed saccadic eye movements. Our quest is to find biological plausible PI. We show that the PI of total saccade flight-time fits observations better than the more intuitive PI of target error. In the longer Section 4, we assume the PI to be total saccadic flight-time, and show how different constraints lead to different optimal trajectories. We will compare bang-bang control to the less intuitive optimisation of variance and bandwidth. Here we will provide a plausible explanation for the long-standing question of why movements are smooth.

In Section 5, we ask the question of how the brain may adaptively control behaviour when the desired goal, or reference, of the control cannot be explicitly coded. We will examine real-time stochastic gradient descent as a very simple model of cerebellar-mediated optimal control. Here we take a leaf from evolution and describe how random noise can be utilised to find an optimum.

It is emphasised at the outset that the schemes outlined in the following sections are hypotheses, and as such need yet to be substantiated by experiment.

2. Survival analysis

How a specific behaviour affects overall fitness of an individual is in general difficult to estimate due to the high-dimensionality of behaviour and the dependence on unpredictable environmental events. Nevertheless, survival analysis can provide a broad overview of the problem.

Denote the probability that the individual survives up to, or beyond, age T as $F(0, T)$ (the 'survivor function'), where T is an age when the individual has contributed to the gene pool (i.e., when offspring have been produced and are viable without the individual remaining alive). From the theory of survival analysis we have

$$F(0, T) = \mathrm{e}^{-\int_0^T h(t)\,\mathrm{d}t}$$

where $h(t)$ is the 'hazard function' and $h(t)\delta(t)$ is the probability or risk of dying in the small interval $(t, t+$

δt) given that the individual is still alive at time t. (Note that $h(\cdot)$ is not a probability but a probability rate, where $0 \leq h(\cdot) < \infty$). Clearly, to maximise survival to age T, it is necessary to minimise the integral of the hazard function:

$$H(0, T) = \int_0^T h(t)\, \mathrm{d}t$$

that is $\max\{F(0, T)\} \Leftrightarrow \min\{H(0, T)\}$. Assume that a behavioural event occurs over the time interval $\{A, B\}$, then the probability of surviving until B is $F(0, B) = F(0, A)F(A, B)$, where $F(A, B)$ is the probability of the individual surviving the event, given that it was alive at the beginning of the event. The total hazard is $H(0, B) = H(0, A) + H(A, B)$. We can extend this to N discrete non-overlapping behaviours:

$$H(0, T) = \sum_{i=1}^{N} H_i.$$

Thus, it may be just as important to optimise behaviours that have only a small affect on fitness but are very frequent, as it is to optimise life-endangering behaviours that are very occasional.

Let us assume that the individual produces a second behaviour contiguous with but not overlapping the first over the period $\{B, C\}$. Then $F(A, C) = F(A, B)F(B, C)$ and the integrated hazard is $H(A, C) = H(A, B) + H(B, C)$. Now for some behaviours, it may be true that:

$$\min\{H(A, B) + H(B, C)\}$$
$$= \min\{H(A, B)\} + \min\{H(B, C)\} \quad (1)$$

that is, overall survival is maximised if the risk during each behavioural event is separately minimised. This holds if $H(A, B)$ has no effect on $H(B, C)$, or if lowering $H(A, B)$ enhances the survival to C given being alive at B by lowering $H(B, C)$ (thus it is not just a question of independence). For example, maintaining optimal nutrition will not only increase the chance of survival from A to B, but it may provide additional benefit in surviving an event from B to C, such as fleeing a predator. We can extend this argument to many behaviours that are not necessarily contiguous in time. For convenience, will call behaviours that obey Eq. (1) 'type-1' behaviours.

If all behaviour were type-1, then nervous systems would be relatively simple. What makes behaviour non-trivial is that some behaviours that lower $H(A, B)$ may have the deleterious effect of raising $H(B, C)$ to such an extent that:

$$\min\{H(A, B) + H(B, C)\}$$
$$< \min\{H(A, B)\} + \min\{H(B, C)\}. \quad (2)$$

Here, the consequences of a behaviour need to be included in choosing the optimal behaviour. We will call behaviours that obey Eq. (2) 'type-2'. Behaviours which are fatigueable or rely on limited resources tend to be type-2, since excessive work in the short term may have serious long-term consequences. For example, if a predator impulsively initiates a chase when the probability of success is low, it may deplete its own energy reserves and make a subsequent chase less successful, thereby raising integrated hazard. A more appropriate behaviour might be to initiate a chase when the probability of success is estimated to be high, provided the predator does not need to wait too long! Type-2 strategies are complicated and are the subject of Game Theory, and will not be discussed here. With respect to performance, behaviours with a trade-off between speed and accuracy are a prime example that will be examined in more detail in Section 4 (see below).

Since the behaviour BC occurs after AB, optimisation requires that the first behaviour be executed and optimised in anticipation of predicted future events and later behaviour so that the total integrated hazard is minimised. Environmental events tend to be uncertain, so optimal type 2 behaviours are based on the expectation of the future consequences (rather than actual consequences). Since most expectations are not fixed (nor even stationary), type-2 behaviours cannot be coded directly genetically, but must be learnt–adapted from previous experience. That is, it is the adaptive controller or learning mechanism that is coded genetically for type-2 behaviour. We now consider a specific example.

3. Saccade gain

Saccades are the fastest type of eye movement and appear to have the function of redirecting the fovea to visual objects of interest ('oculomotor reaching'). Saccadic trajectories tend to be stereotyped with larger movements having both longer durations and higher peak velocities in a more-or-less fixed relationship, sometimes called the 'main sequence' (Bahill et al., 1975b) (see Fig. 1). In everyday viewing most saccades have an amplitude under 20° (Bahill et al., 1975a), and for this natural range, velocity profiles tend to be nearly symmetric and similar in shape (except for scaling in time and in velocity) (Collewijn et al., 1988; see Fig. 2), and are similar to some wrist movement trajectories (Abrams et al., 1989). Since normal saccades reach speeds of hundreds of degrees per second and are completed in tens of milliseconds there is no possibility that a saccade can be guided by visual feedback during the saccade. Thus, a saccade is a fast open-loop behaviour in which the amplitude is under adaptive control. Optican and Robinson (1980) have shown that the cerebellum is necessary for this adaptation.

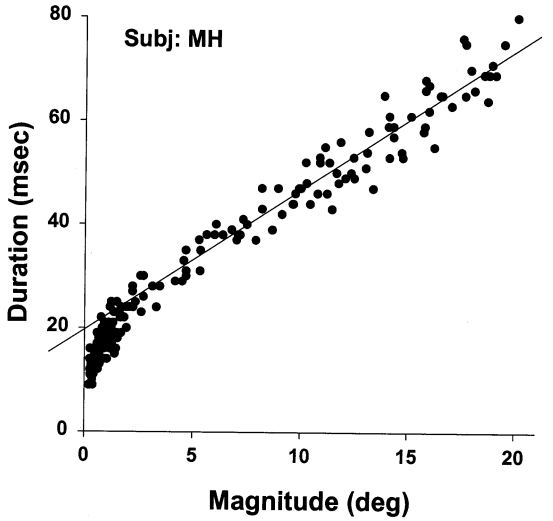

Fig. 1. Typical duration vs magnitude plot of horizontal human saccades. Note roughly constant slope beyond 5° with non-zero intercept, shown by line.

3.1. Undershoot bias

In experimental laboratories, saccades are most often elicited by presenting a peripheral visual target to a subject, who then executes a saccade towards the target. It might be thought that the most likely PI would be target error, so that the goal of the adaptive controller would be to adjust the amplitude of the saccade so that, at the end of a saccade, the fovea is pointing at the target. Biological noise from sensory and motor sources would lead to a distribution of errors across trials, but with an error feedback controller, we would expect that the average saccade would fall roughly symmetrically on the target, with an equal proportion of over- and under-corrections (within the steady-state error of the controller). However, empirically this does *not* happen. Saccades to suddenly appearing peripheral visual targets show a clear bias to undershoot the target, typically reaching about 90% of the target distance (i.e., a gain of 0.9) (Becker, 1989). Even if visual targets are constantly illuminated, undershoot bias still occurs, although it may be less (Lemij and Colewijn, 1989).

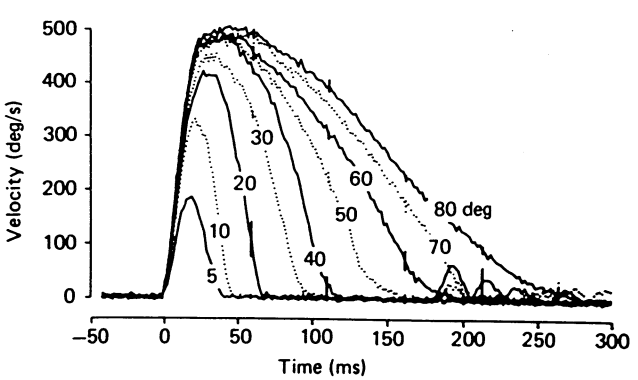

Fig. 2. Velocity trajectories of human horizontal saccades showing increase in duration and peak velocity with amplitude. Note similar symmetrical shape for saccades below 20°. Each curve represents an average of four saccades recorded with the magnetic search coil technique. From Collewijn et al., 1988 (reproduced with the permission of the authors and the Physiological Society).

This undershoot bias appears not to be a steady-state error of an error feedback adaptive controller. Henson (1978) showed that adaptive change in gain from an initially high gain did not asymptote to small overshoot bias (as would be expected for a steady state error), but asymptoted to undershoot. This is not consistent with a simple error feedback adaptive controller. Thus, despite its intuitive appeal, target error appears not to be the PI of goal-oriented saccadic eye movements. We remark that an undershoot bias also occurs in arm movements (Toni et al., 1996).

3.2. Saccade flight-time minimisation hypothesis

When a primary goal-directed saccade misses the target, secondary saccades are needed to correct the error. It can be seen that a PI of target error does not acknowledge any subsequent corrective behaviour, i.e., it is type-1. Instead, Harris (1995) proposed that the secondary saccades should be included in the PI, and hypothesised that the PI is *total* saccadic flight-time, so that the adaptive controller attempts to minimise total flight-time of the primary *and* secondary saccades. The rationale is that vision is very degraded during fast eye movements and keeping flight-time to a minimum would minimise integrated hazard. Thus, it is the consequences of targeting error that are also important, i.e., type 2.

The key to this hypothesis is that larger movements take longer to execute. For saccades over about 5°, the duration is roughly a linear function amplitude with a non-zero intercept. Below 5°, there is a more curvilinear relationship (see example in Fig. 1). From this we deduce, first, that reaching a target in one saccade always takes less time than reaching the target in two or more saccades; and second, that if there is an error in reaching the target on the first saccade, the total flight-time will be less for an undershoot than an overshoot error of the same magnitude (Fig. 3).

As with any hypothesis of optimality, we need to know theoretically what the optimal behaviour should be given our hypotheses. Unfortunately, finding the theoretical optimum in closed form is usually impossible. Using Monte-Carlo simulations, Harris (1995) showed that the optimal gain would be about 0.93 for parameters similar to those used in real experiments, but that the optimum would depend on the end-point variability of the primary saccade. Less variability would permit a higher gain (but still undershooting), as seen in saccades to stationary targets (Lemij and Colewijn, 1989). Whereas, a lower gain would be opti-

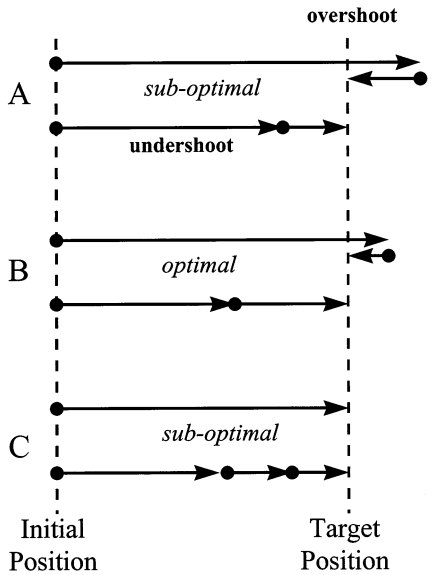

Fig. 3. (a) The overall distance travelled is greater for an overshoot error than for an undershoot error of the same magnitude. (b) Since duration of movement depends on distance travelled there is always a greater time penalty for overshoot. Thus, on average it pays to deliberately undershoot the target. (c) Excessive undershoot will lead to too many secondary saccades which is also sub-optimal.

mal when there is more variability, as is observed in predictive saccades (Bronstein and Kennard, 1987), and in saccades made by human infants (Harris et al., 1993).

A very similar hypothesis has also been proposed for the control of rapid arm movements (Meyer et al., 1988), which has more recently been applied to infant reaching (Berthier, 1996). In summary, a PI of total flight-time can account for undershoot bias. If there were no error, then the shortest flight-time would occur when the saccade was perfectly accurate, thus producing the same result as if error were minimised. The distinction occurs when an error occurs *and* the error requires a correction—overshoot is more expensive than undershoot. Thus, this PI obeys Eq. (2). The hypothesis of saccade flight-time minimisation leads directly to two further questions: how does the adaptive controller find the optimum gain? (discussed in Section 5) and what is the optimal trajectory for the individual saccades?

4. Trajectories

Assuming that overall flight-time is the PI of the saccade gain controller, it would seem to be a logical deduction that individual movements (sub-movements) should be as fast as possible, which has been claimed by some (Clark and Stark, 1975; Lehman and Stark, 1979; Enderle and Wolfe, 1987). We now consider this important issue, but it should be recognised that demonstrating such optimality is far from easy because we now encounter the complexities of largely unknown biological *constraints*.

First, consider the problem of trying to make a movement in the shortest possible time given a second order over-damped low-pass plant: $O(s) = 1/[(sa + 1)(sb + 1)]$ (s = Laplace variable). Now, it would be quite feasible to build a compensating pre-motor system that exactly cancelled the plant by having a transfer function that is the reciprocal of the plant, i.e., $1/O(s) = (sa + 1)(sb + 1)$ (inverting dynamics). With such compensation, the output would faithfully follow the input, and the plant would become completely transparent. The problem appears when we wish to move the output rapidly, say, with a step. The compensator would need to differentiate a signal of infinite slope, which is physically impossible. All biological systems have limits to the magnitudes of signals that they can transmit. So for fast movements, low-pass plants cannot be functionally inverted. In general, finite (saturated) control signals place a limit on the speed of the output of low-pass plants, which in turn places a minimum duration on the movement.

4.1. Bang-bang control

It is well-known that for saturated control of a stable *linear* system, the optimal control signal that minimises the time to reach the final position (and remain at that position) is given by 'bang-bang' control (based on Pontryagin's Minimum Principle, see Bryson and Ho, 1975). Indeed, it has been claimed that saccades are under time-optimal bang-bang control (Enderle and Wolfe, 1987). In bang-bang control, the control signal is switched instantaneously at precise times between the extreme saturated limits permitted by the system. The number of switches, or 'bangs', depends on the complexity of the system (but is always less than the number of poles in the plant). This is illustrated in Fig. 4 (middle panel), where the optimal bang-bang control needed to drive our hypothetical second order system as fast as possible has been computed (for illustration, a and b have been set to be roughly equal to the human ocular motor plant). The optimal duration has been computed for different amplitudes of movement (Fig. 4 bottom panel) and is qualitatively similar to observed durations (Fig. 1). This kind of relationship is a common feature of the bang-bang control of simple low pass plants. However, although bang-bang control is intuitively appealing and apparently unifying, it suffers from considerable problems.

The optimal bang-bang trajectory depends crucially on the plant dynamics, or more pertinently, on one's model of the plant dynamics. For example, if we include a zero in the plant, i.e., $O(s) = (sz + 1)/[(sa + 1)(sb + 1)]$ as originally proposed by Robinson (1964),

then quite a different trajectory becomes optimal but this produces saccades that are far too fast. Using a complex sixth-order model, Enderle and Wolfe (1987) claimed that saccades are time-optimal with a bang-on bang-off (rectangular) control signal, but of course, this depends on the plant model.

Another problem with bang-bang control is that it tends not to yield symmetrical velocity profile. With current models of the plant, a symmetrical bang-on bang-off input can never produce a symmetrical output. In addition, Fourier analysis of actual saccades reveal sharp minima in the energy spectra (Harris et al., 1990). The frequency response of linear models of the plant are smooth, so these minima should arise from the driving signal. If we accept the bang-on–bang-off hypothesis (i.e., a rectangular driving signal), the energy spectrum of this shape must have minima at frequencies

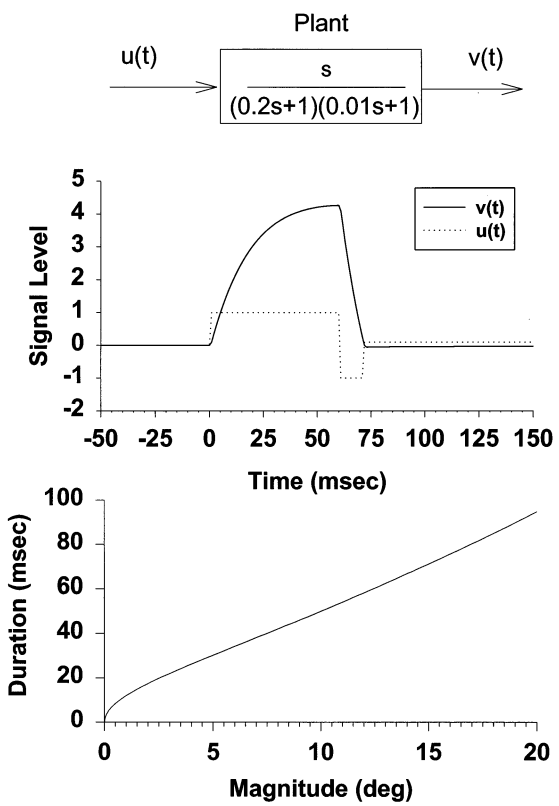

Fig. 4. Example of bang-bang control of a 2-pole low-pass plant $1/[(0.2s+1)(0.01s+1)]$, with time-constants similar to the oculomotor plant, where $u(t)$ is the input control signal that saturates at 1 and -1, and $v(t)$ is the derivative of the output of the plant, hence the s in the numerator (top panel). Middle panel shows the optimal bang-bang trajectory, where $u(t)$ (dashed line) is switched as fast as possible between limits at optimal switching times so as to bring output velocity (solid line) to zero in the shortest possible time for a given amplitude of output movement (i.e., for given area under $v(t)$). (Note $u(t)$ does not return to zero but to a tonic level so that the new output position can be maintained.) Bottom panel shows the duration vs amplitude of the optimal movement computed for different durations of $u(t)$. Note similarity to observed data in Fig. 1, but does not provide reasonable fits to observed near-symmetric profiles for real saccades with amplitude $< 20°$ (see Fig. 2).

that are harmonically related at $1/T$, $2/T$, $3/T\ldots$ where T is the duration of the rectangular pulse. However, the spectra of most saccades have minima not at harmonics but approximately at $1.5/T$, $2.5/T$, $3.5/T$ (see Fig. 5). Thus, bang-on–bang-off control of a linear plant cannot account for observed trajectories, regardless of one's (linear) model of the plant! However, this does *not* mean that trajectories are sub-optimal (as has been claimed by some) because there are other types of constraints besides simple saturation of signals.

4.2. Minimising square derivatives

The question of what constitutes an optimal trajectory has frequently been posed in the arm movement literature. It is also interesting to note that there is a similarity between the trajectories of saccades and the symmetrical trajectories of fast arm movement, and wrist movements (Abrams et al., 1989). Although many different functions have been fitted to arm movement trajectories, considerable attention has been paid to trajectories that minimise the square of a time derivative of the position profile. It has been shown that fast arm movements are well described by the trajectory that minimises the square of 'jerk' (the rate of change of acceleration) over the duration of the trajectory (Hogan, 1984; Flash and Hogan, 1985), although minimum square 'snap' (rate of change of jerk) may also provide a good fit (Wiegner and Wierzbicka, 1992). Minimum square derivative (MSD) profiles of movement velocity are symmetrical and seem at least qualitatively similar to saccade velocity trajectories, although goodness of fit studies are still needed. The implication is that the velocity profiles of trajectories are optimal with a PI of the square of a derivative.

To clarify this approach, let us assume that we wish to minimise the total square of the nth derivative of the position profile of the trajectory, then the PI is given by

$$H(0, T) = \int_0^T [x^{(n)}(t)]^2 \, \mathrm{d}t$$

where the parenthetical superscript indicates the order of the time derivative [$x^{(2)}(t) = \mathrm{d}^2x/\mathrm{d}t^2 =$ acceleration, $x^{(3)}(t) =$ 'jerk', $x^{(4)}(t) =$ 'snap', etc.] The trajectory that minimises $H(0, T)$ can be found from the Euler equation of calculus of variations (see Hogan (1984)), which requires that

$$\frac{\mathrm{d}^n}{\mathrm{d}t^n}\left(\frac{\partial h}{\partial x^{(n)}}\right) = 0.$$

Applying this yields the differential equation: $x^{(2n)} = 0$, which can be solved simply by successive integration. For example, minimising squared acceleration ($n = 2$) yields a cubic polynomial as the general solution: $x(t) = a_0 + a_1t + a_2t^2 + a_3t^3$. Imposing the boundary condition that velocity is zero at the beginning and end

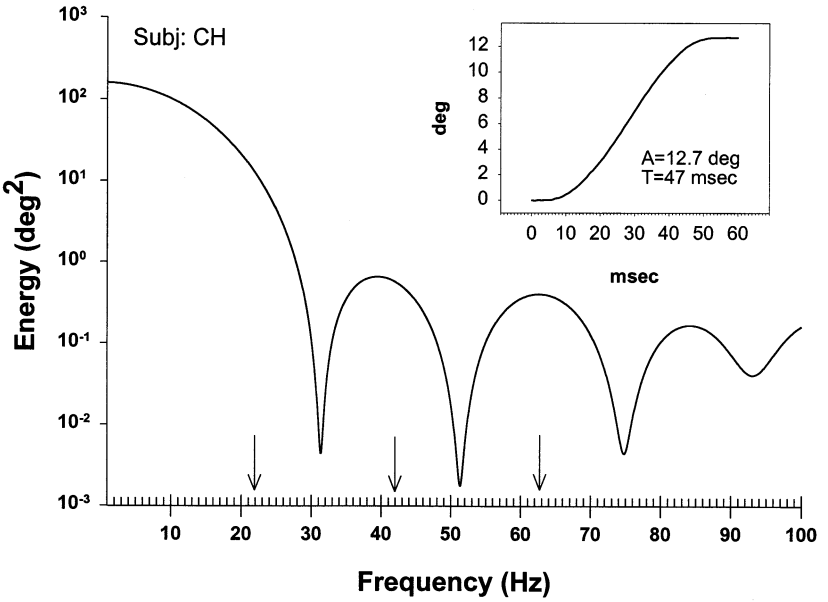

Fig. 5. Typical Fourier energy spectrum of the velocity profile of a horizontal human saccade (insert) with amplitude 12.7°, duration $T = 47$ ms ($1/T \sim 21$ Hz). Log energy is plotted against linear frequency. Note sharp minima in spectrum occurring at non-harmonic frequencies, at approximately $3/2T$, $5/2T$ and $7/2T$. If the saccadic plant were linear and driven by a rectangular pulse, minima should occur at $1/T$, $2/T$, and $3/T$ (arrows). Therefore, saccades cannot be described by a linear plant driven by bang-on bang-off control.

of the movement, the velocity of the optimal trajectory is given by the parabola (Fig. 6): $v(6) = 6A[(t/T) - (t/T)^2]/T$, where A is the amplitude of the movement. The trajectory is finite and its shape is symmetrical and independent of the duration or amplitude of the movement. However, this trajectory has instantaneous onset and offset of acceleration, which could be considered biologically implausible.

Minimising squared jerk ($n = 3$) yields $x^{(6)} = 0$, which has a solution given by a quintic polynomial. Applying the boundary conditions that velocity and acceleration are zero at the beginning and end of the movement, yields $v(t) = 30A[(t/T)^2 - 2(t/T)^3 + (t/T)^4]/T$, which is the so-called 'minimum jerk' trajectory. Again this shape is symmetrical and invariant to the amplitude and duration of the movement. It could be argued that this too is biologically implausible since it requires instantaneous jumps in jerk at the beginning and end of the movement, so perhaps we should consider minimum snap ($n = 4$), and so on. However, the differences between higher order MSD profiles becomes small, as they asymptote to a Gaussian (Fig. 6).

Although, to our knowledge, experimental fits of saccades by MSD profiles have not been performed, there is an obvious similarity, which suggests that trajectories of saccades and arm movements may reflect some underlying principle of movement. However, the fundamental difficulty with MSD profiles is that there is no *a priori* reason for choosing any particular squared derivative as a biologically plausible PI, other than MSD profiles are 'smooth'. Moreover, the constraints that derivatives up to a rather arbitrary order are zero at the beginning and end of the trajectories are also of dubious biological relevance (see Slepian (1983)). The question becomes, therefore, why is smoothness so important for goal-directed movements (and most other movements for that matter)?

4.3. Optimising movement bandwidth

Let us return to bang-bang control. Its failure to fit observations is really not surprising because it is quite implausible that all pre-motor burst neurons, ocular motor neurons, and muscle fibres could be switched on or off at precisely the same time. We expect neuromuscular impulses to be probabilistic in time. Moreover, if we assume that neural innervation signals are stochastic with Poisson-like statistics then noise must be signal-dependent. That is, as signal level increases, the noise in the signal also increases. This has fundamental implications for movement trajectories, because for low-pass muscle plants, to change the output quickly requires a large input signal (e.g., the saccadic pulse) which will induce more noise. Thus, attempting to move fast leads to more variance, which requires corrective behaviour. Whereas moving slowly reduces variance but, of course, takes longer to reach the goal. Signal-dependent noise leads, therefore, directly to a speed-accuracy trade-off (which is ignored by bang-bang control).

To clarify this approach, we examine a 'simple' case where all the noise occurs on the input to the plant, and the plant is linear with fixed deterministic dynamics. Denote the input signal as a stochastic signal $u(t)$ which we write as the sum of a mean signal, $\bar{u}(t)$, and a noise

component, $n(t)$, which has zero mean and autocorrelation–autocovariace of $R_{nn}(t_1, t_2)$: $u(t) = \bar{u}(t) + n(t)$. Similarly, we can write the output of the plant as a stochastic eye position signal, $x(t) = \bar{x}(t) + y(t)$, where $\bar{x}(t)$ is the mean position trajectory, and $y(t)$ is a random process with zero mean and autocorrelation–autocovariace of $R_{yy}(t_1, t_2)$. Given a *linear* plant with impulse response $p(t)$, the output mean and noise are determined by the convolution of the plant with the input mean and noise:

$$\bar{x}(t) = \int_0^t \bar{u}(\alpha)p(t-\alpha)\,\mathrm{d}\alpha \tag{3a}$$

and

$$y(t) = \int_0^t n(\alpha)p(t-\alpha)\,\mathrm{d}\alpha. \tag{3b}$$

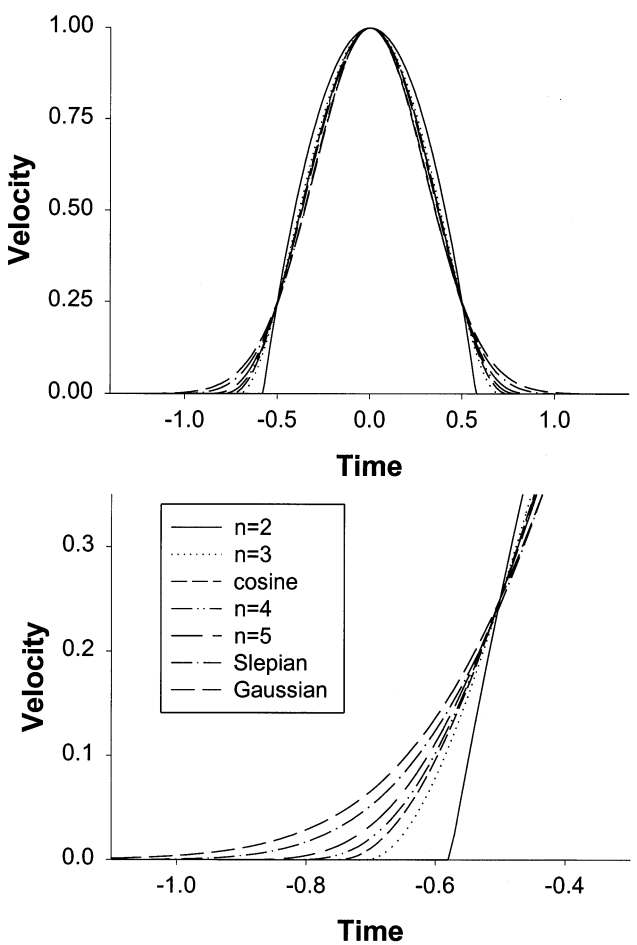

Fig. 6. Comparison of low bandwidth trajectories. *Upper panel*: velocity profiles with time-origin centred at peak velocity, which has been normalised to unity. Trajectories have been scaled in time so that velocity is 0.25 at ± 0.5 time units. Curves show Minimum Square Derivative (MSD) trajectories for $n = 2, 3, 4, 5$ (2 = minimum acceleration; 3 = minimum jerk etc.) and Gaussian ($n \to \infty$). These are compared to standard low spectral leakage profiles: a 100% cosine window (Hanning) and a Slepian taper (prolate spheroidal wave function with argument = 4). *Lower panel:* magnified view of tail.

From the standard theory of stochastic processes (see Papoulis (1991)), the autocorrelation of the output noise is:

$$R_{yy}(t_1, t_2) = \int_0^{t_1}\int_0^{t_2} R_{nn}(\beta, \alpha)p(t_2-\alpha)p(t_1-\beta)\,\mathrm{d}\alpha\,\mathrm{d}\beta \tag{4}$$

for an input signal that starts at zero time. The variance in eye position at time t is $\sigma^2(t) = R_{yy}(t, t)$.

To model signal-dependent (non-stationary noise) we assume the input noise to be the product $n(t) = w(t)r(t)$. $r(t)$ is a zero-mean *stationary* white noise with an autocorrelation of $R_{rr}(t_1, t_2) = \delta(t_1 - t_2)$, where $\delta(t)$ is the Dirac delta function, and $w(t)$ is a non-negative deterministic function. The autocorrelation of the input noise is then $R_{nn}(t_1, t_2) = E\{n(t_1)n(t_2)\} = w(t_1)w(t_2)\delta(t_1 - t_2)$. Substituting into Eq. (4) and setting $t_1 = t_2 = t$, we obtain

$$\sigma^2(t)\int_0^t\int_0^t w(\alpha)w(\beta)\delta(\alpha-\beta)p(t-\alpha)p(t-\beta)\,\mathrm{d}\alpha\,\mathrm{d}\beta,$$

which reduces to:

$$\sigma^2(t) = \int_0^t w^2(\alpha)p^2(t-\alpha)\,\mathrm{d}\alpha, \tag{5}$$

and is itself a convolution. Eqs. (3a) and (5) describe how output position and variance of position change in time given a mean control input $\bar{u}(t)$ with additive white noise scaled by an arbitrary non-negative deterministic function $w(t)$. Our interest is in minimising the variance after the end of a movement $\sigma^2(T')$ $(T' > T)$, *given* that after the end of the movement at time T, eye position *remains* steady at some amplitude $x(t) = A$ for $t \geq T$.

If the noise were independent of the signal so that $w(t)$ were a constant, output variance would be minimised by making the movement as brief as possible (i.e., bang-bang control). However, this is not the case for signal dependent noise, i.e., when $w(t)$ depends on $\bar{u}(t)$. To illustrate consider the case when the standard deviation is proportional to the modulus of mean input signal during the movement. We set $w(t)$ to be zero for $t > T$, and ignore the effects of noise in the gaze-holding signals needed to maintain the final position (what happens after the movement is very interesting but beyond the scope of this article); thus:

$$w(t) = \begin{cases} 0 & T < t < \infty \\ k|\bar{u}(t)| & 0 \leq t \leq T \end{cases}$$

where k is a constant of proportionality. Then the optimum trajectory minimises.

$$\sigma^2(T') = k^2\int_0^{T'} |\bar{u}(t)|^2 p^2(T'-t)\,\mathrm{d}t.$$

To simplify, consider brief movements with a low-pass plant such that $p(T'-t)$ is roughly constant over the movement, then

$$\sigma^2(T') \approx k^2 \int_0^{T'} |\bar{u}(t)|^2 \, \mathrm{d}t,$$

so the optimal trajectory will minimise

$$\int_0^{T'} |\bar{u}(t)|^2 \, dt.$$

For a low-pass plant with only poles, a neural signal is needed after the movement ($t > T$) to maintain steady final position (the role of the famous eye position 'neural' integrator), but this signal depends only on position at the end of the movement—not on how the end point was reached. Therefore since

$$\int_0^{\infty} |\bar{u}(t)|^2 \, \mathrm{d}t = \int_0^{T'} |\bar{u}(t)|^2 \, dt + \int_{T'}^{\infty} |\bar{u}(t)|^2 \, \mathrm{d}t$$

the optimum trajectory must also minimise

$$\int_0^{\infty} |\bar{u}(t)|^2 \, \mathrm{d}t.$$

However, the input signal $\bar{u}(t)$ is constrained by Eq. (3a) to yield an output in which final position is not only attained: $\bar{x}(T) = A$, but also *maintained*. Unfortunately, this is difficult to solve, but Fourier analysis yields some insight.

Denoting $\bar{U}(\omega)$ as the Fourier transform of $\bar{u}(t)$, we can write the Fourier transform of the mean output position trajectory as $\bar{X}(\omega) = \bar{U}(\omega)P(\omega)$, where $P(\omega)$ is the transfer function of the plant. Then from Parseval's theorem, the optimum trajectory will minimise:

$$\int_{-\infty}^{\infty} \frac{|\bar{X}(\omega)|^2}{|P(\omega)|^2} \, \mathrm{d}\omega. \tag{6}$$

Thus, end-point variance is minimised by minimising the total energy in $\bar{X}(\omega)$ (or $\bar{x}(t)$). However, $\bar{x}(t)$ is constrained to remain stationary after T. Since the energy spectrum of a step of amplitude A is A^2/ω^2, then for frequencies below $1/T$, $|\bar{X}(\omega)|^2 \to A^2/\omega^2$. Therefore, the energy in $\bar{X}(\omega)$ can only be reduced by lowering the bandwidth of $\bar{X}(\omega)$. We conclude that *the end-point variance of the trajectory increases with the bandwidth of the trajectory*. This is a fundamental result that tells us that, to minimise end-point variance, it is necessary to choose a trajectory with as low a bandwidth as possible, or equivalently as *smooth* as possible.

Low bandwidth requires an increase in the spread in the time domain, that is an increase in the duration of the movement. Thus, on the one hand, moving fast inflates variance because of the increased bandwidth, which in turn, requires costly corrective movements. Whereas, moving slowly reduces bandwidth and the need for corrective movements, but at the expense of a long primary movement. In essence, movement trajectories would be limited by the *uncertainty principle*.

There should be some optimum trajectory that balances the time-width and bandwidth of the trajectory. This is a well-studied problem and many profiles have been suggested to minimise bandwidth (see Percival and Walden (1993)) (sometimes called 'windowing'). For example, a raised cosine window (a Hanning window) has good spectral leakage properties and is similar to MSD profiles (Fig. 6). It has been shown that the trajectory that simultaneously minimises the variance of trajectory profile in the frequency and time domains is given by the zero order prolate spheroidal wave function (sometimes called a 'Slepian taper') (see Percival and Walden (1993)). This trajectory is smooth and symmetrical, and is also similar to MSD profiles (Fig. 6). We can also easily see why there is a similarity to MSD profiles. The Fourier transform of the kth derivative of $x(t)$ is $(-i\omega)^k X(\omega)$. Clearly, derivatives emphasise high frequencies, and in the limit $k \to \infty$, ω^k tends to a step function. Thus, minimising a square derivative is equivalent to keeping bandwidth low MSD profiles are, however, special cases of low bandwidth profiles. For very fast movements the dynamics of the plant will be dominated by the highest order term, so that $\bar{u}(t) \approx c\bar{x}^{(n)}(t)$, where c is a constant of proportionality and n is the order of the plant. Clearly, minimising $\int_0^T |\bar{u}(t)|^2 \, dt$ will then yield a MSD profile.

From Eq. (6) the optimal profile will depend on the weighting function $1/|P(\omega)|^2$, but provided $P(\omega)$ is a smoothly decreasing function we expect optimal profiles to be similar in shape (although not necessarily identical). Practically, the differences among low bandwidth trajectories are small since they are all (by definition) smooth and featureless in the time-domain. Whether they can be distinguished in biological trajectory data remains to be seen.

4.4. The plant

An important aspect of this approach is that the trajectory shape is not determined directly by the dynamic response of the plant. Rather, it is the transfer of variance that becomes the critical factor. For a fixed plant, a larger movement can be made by either increasing peak velocity (velocity scaling) or by increasing duration (time-scaling), or both. The precise optimal peak velocity and duration for a given amplitude would be a complex function of $v(t)$, the end-point distribution, the undershoot bias, and the strategy of corrective saccades, and is beyond the scope of this discussion. However, we can see that if the noise were only weakly dependent on the trajectory then it would pay to scale velocity and keep duration constant, leading to (fixed-plant) isochronous movements. It is interesting to note that some arm movements are nearly isochronous (Viviani and Flash, 1995) (although this isochrony could also reflect a variable plant). If the noise were strongly dependent on the velocity trajectory then it would pay to increase duration as well as velocity, as seen in saccades.

4.5. Summary

Even though we have considered the same PI (minimising flight-time), different constraints lead to quite different optimal trajectories. For a linear plant with no noise (or signal-independent noise) the optimal trajectory is given by bang-bang control, which is highly dependent on the plant dynamics. For signal-dependent noise, the optimal trajectory is more symmetrical with low bandwidth (smooth), and not very sensitive to the plant dynamics. For saccades below about 20° and for fast arm movements, observed trajectories are clearly similar to low-bandwidth trajectories.

5. Stochastic gradient descent and the cerebellum

How the nervous system optimises performance is open to debate, although there is little doubt that the cerebellum is intimately involved in the process. The most common view is that performance is adapted to reach some future goal or explicit desired state. If eventual outcome deviates from the desired state then performance parameters are adjusted according to the error, which we call error-feedback. Here we take error to be a signed quantity with a defined zero point, that is, it is a vector. There are many models for accomplishing this task, and they usually have some adjustable predictive sub-unit. These models may be quite complicated, especially if the relationship between desired output and the effectors is not simple, as in multi-joint arm movements. The implicit assumption in these models is that the desired output state or reference is known explicitly by the brain ahead of time, and moreover, that it can be coded in the brain so that the error can be computed to update the performance. For some adaptively controlled behaviours this is plausible (whether type 1 or 2). For example, the desired output state in the vestibulo-ocular reflex is presumably no slippage of the retinal image during natural head movements. This can be coded via image motion receptors, where directional sensitivity permits a zero slip to be coded. For other behaviours it is not clear how the reference could be coded.

An important reason for examining minimum-time problems is that they *cannot* be controlled by simple error feedback. The desired minimum time is not known ahead of time and cannot be coded as a reference. Even if it could, there could be no signed error because, by definition, flight-time could never be shorter than the minimum. Thus, optimal flight-time can be found only by unreferenced control (sometimes called 'reinforcement learning', see Barto (1992)). If we accept flight-time as the PI, and that performance optimisation is carried out by the cerebellum, then we are forced to reject the many models of cerebellar function that have depended on error feedback, in which the desired behaviour is the reference (Fujita, 1982; Ito, 1984; many others). We must look for an alternative approach.

In error feedback, a single measurement of the output of the system gives sufficient information about how to change the gain; i.e., an overshoot error requires a lower gain, and undershoot error requires a higher gain. A zero error (should it occur) signifies perfect (optimal) gain. In unreferenced control, a single measurement does not indicate which direction to proceed, it could be due to too low gain or too high gain. For guided unreferenced control (as opposed to random trial and error), the controller needs to extract gradient information about how the cost (total flight-time) changes with parameter (gain) changes, so that the direction of parameter change can be found to improve performance. (Note that this is not a trivial problem because it is the derivative of cost with respect to the parameter(s)—not time!) Thus, an intrinsic feature of unreferenced control is that the parameter space needs to be *explored* to find gradient information. Inevitably exploratory behaviours will sometimes not be in the direction of the optimum. Thus, there is always a trade-off between sub-optimal exploratory behaviour and behaviour that exploits gradient information (see Thrun (1992)). A related aspect of unreferenced control is that there is no way of knowing when the optimum has been reached, so unreferenced control is a continuous process of exploration even when the behaviour is around optimal. Therefore, variability is an intrinsic part of unreferenced control, and is not simply 'noise'. We now consider the possibility that the adaptive control may be a process of stochastic gradient descent, which we illustrate as a simple hypothesised scheme of cerebellar function.

We assume that behaviours occur in discrete events–epochs, such as a saccade and its corrective saccades, trial, etc., and denote these epochs by the index $k = 1, 2,\ldots$ (there is no implication that these events occur at regular intervals). Based on Fujita (1982) classic model, we assume that a command $c(t)$ is transmitted along a mossy fibre (*mf*) to granule-Golgi complexes that distribute the command into an array of signals with different transfer functions $x_i(t)$ (to form the basis set), which are then transmitted to Purkinje cells (*Pc*'s) via parallel fibres (*pf*'s). Thus the ith *pf* will carry a signal given by $C(s)X_i(s)$, (s = Laplace variable), and the response of the ith *Pc* to this signal depends on the ith *pf*–*Pc* synaptic weight, $w_i(k)$. Assuming linear summation of *Pc* outputs on the target nucleus, the output $T(s)$ will be given by

$$T_k(s) = C(s) \sum_{i=1}^{N} w_i(k) X_i(s).$$

This output will lead to a motor behaviour, $o_k(t)$, that depends on the muscle plant, $P(s)$, so that $O_k(s) = P(s)T_k(s)$. This behaviour will then interact with the environment in some way to ultimately yield a signal that represents the actual cost to the organism (PI), $J_k = J(o_k(t))$ where $J(\cdot)$ is generally an unknown non-linear function. Clearly, we can summarise the whole cascade of signals as simply $J_k = J[w_1(k), w_2(k),\ldots, w_N(k)]$, where the adaptive control problem is to find the weights $w_i(k)$ that minimise the expectation of the unknown cost function $\mathrm{E}\{J(\cdot)\}$. The physiological relationship between a synaptic weight and the ensuing behaviour is not important for our purposes.

5.1. Stochastic gradient descent

If we perturb each weight by some small amount, Δw_i, from epoch to epoch, then to a first order, the change in cost ΔJ will be a linear function of the weight changes:

$$\Delta J = \frac{\partial J}{\partial w_1}\Delta w_1 + \frac{\partial J}{\partial w_2}\Delta w_2 + \cdots + \frac{\partial J}{\partial w_N}\Delta w_N \tag{7}$$

If the weight perturbations are made to be statistically uncorrelated, so that the expectation $E[\Delta w_i, \Delta w_j] = \sigma_{ij} = 0$ for $i \neq j$, and σ_i^2 for $i = j$, then the slope of the cost function with respect to any particular weight can be recovered (to a first order) by correlating the cost change with the weight change:

$$E[\Delta J, \Delta w_i] = \sigma_i^2 \frac{\partial J}{\partial w_i}$$

where it is assumed that there is negligible change in the gradient during the correlation period. Thus, the gradient can be estimated without the need to know the absolute synaptic weights *per se*, but only by knowing only how much a weight is changed. Moreover, correlation is a biological plausible neural function. One 'inelegance' is that the change in cost between successive epochs, ΔJ, needs to be computed. This can be avoided, however, as follows. We write the value of the ith synaptic weight as $w_i = \mu_i + \Delta w_i$, where μ_i is the mean and assumed to change slowly, Δw_i is the perturbation random variable with zero mean. A Taylor expansion of the cost function about synaptic means yields to a first order:

$$J = J(\mu_1, \mu_2,\ldots,\mu_n) + \frac{\partial J}{\partial w_i}\Delta w_1 + \frac{\partial J}{\partial w_2}\Delta w_2 + \cdots + \frac{\partial J}{\partial w_n}\Delta w_n + \cdots$$

If we now perform the expectation of the product of cost and weight change we obtain:

$$E[J, \Delta w_i] = \sigma_i^2 \frac{\partial J}{\partial w_i}.$$

This implies the simple updating rule for synaptic weights:

$$\Delta w_i(k+1) = \Delta_i - \kappa E_T[J(k), \Delta w_i(k)] \tag{8}$$

where $\Delta w_i(n) = w_i(n) - w_i(n-1)$, and κ is a small scalar that ensures that the changes in weights are sufficiently slow to maintain convergence on the minimum. Δ_i is a random variable with zero mean and $E[\Delta_i, \Delta_j] = 0$ for $i \neq j$. The expectation $E_T[\cdot]$ indicates that the operation must take place over a finite number of T epochs. This technique can find the minimum of smooth cost functions even when there are hundreds of weights, although full-scale simulations are still needed.

On the right-hand side of Eq. (8) we can readily identify the exploratory random process, Δ_i, to extract gradient information, and the exploitation of the gradient information given by the expectation term. When the process reaches the minimum, the expectation term becomes zero (within sampling error) and the weights will fluctuate around the optimum (as permitted by the basis set). It is important to recognise that these weights represent the function (in the basis set space) that transfers the command $c(t)$ to the optimal motor behaviour $o(t)$, and generally, they are *not* an internal representation of either the forward or inverse dynamics of the plant or the environment (except when cost is genuinely the squared error from a reference).

This approach depends on the orthogonality of the noise processes, namely that $E[\Delta w_i, \Delta w_j] = \sigma_{ij} = 0$ for $i \neq j$, and σ_i^2 for $i = j$, and a sufficiently low learning time-constant, κ, to descend down the steepest slope. Really, the expectation must be over a finite number of epochs, so that there will be some non-zero correlation (cross-talk) among the Δw_i due to sampling error, but we do not expect this non-orthogonality to be a serious problem since gradient descent is usually robust. However, there will also be unwanted correlations between the Δw_i and other sources of biological variablity in output from epoch to epoch [for a given $o(t)$]. For example if targeting uncertainty were very large, very small exploratory fluctuations in Δw_i would be swamped and unable to detect the gradient. This could be overcome by either taking expectation over a longer time, or by increasing the variance in the Δw_i to extract more accurate gradient information. Both, of course, induce their own additional cost, reflecting the ever-present trade-off between exploratory and exploitative behaviour.

5.2. Correlation with cerebellar physiology

An important consideration in this stochastic gradient descent hypothesis is the relatively straightforward way in which it can be physiologically represented (Fig. 7). As with many previous authors, we identify the *pf–Pc* synapses as holding the weights, w_i, which are

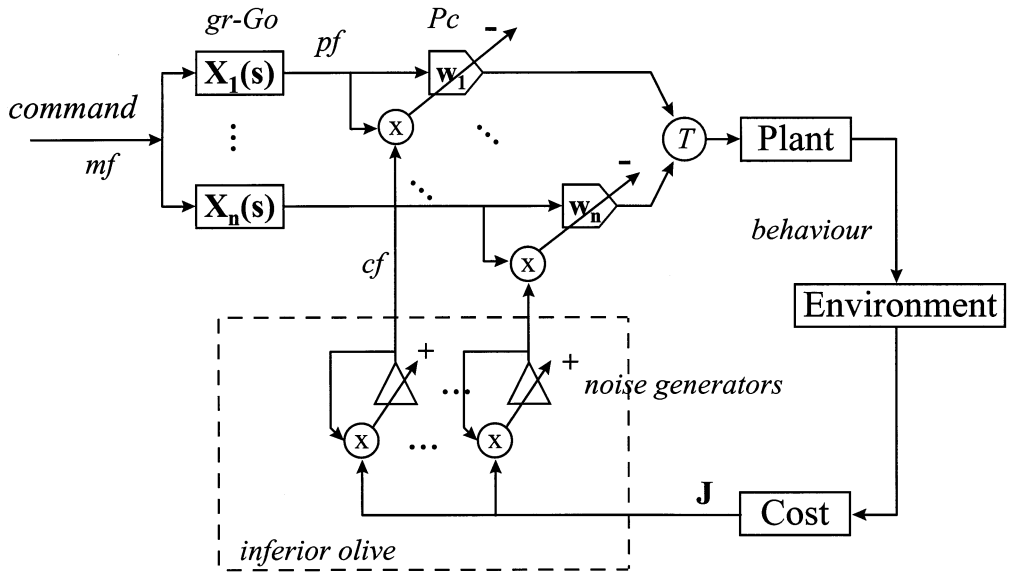

Fig. 7. Outline of cerebellar-inferior olive model of stochastic gradient descent (other brainstem connections not shown). Behaviour commands are introduced on mossy fibres (*mf*) which are distributed by granule-Golgi complexes (*gr-Go*) into a basis set array, $X_i(s)$ transmitted along parallel fibres (*pf*). Each element of the basis set is weighted (w_i) at a Purkinje cell (*Pc*) and then summed at the corresponding target nucleus (*T*) (usually a deep cerebellar nucleus), which is relayed to effectors that perform the behaviour (after Fujita (1982)). Thus, the shape of the behaviour command is determined by the w_i and the choice of the basis set (assumed fixed). The behaviour ultimately gives rise to a scalar cost, *J*, which is determined by the environment in some unknown way. Each w_i is modified with heterosynaptic plasticity according to the correlation between the parallel fibre (*pf*) activity and random signal of a single corresponding climbing fibre (*cf*). The *cf* signals is assumed to be spontaneous activity of an inferior olive neuron which depends on the correlation between itself and the cost function (see text). Optimal behaviour (within the space of the basis set) is found without any reference to a 'desired' behaviour.

plastically changed according to the correlation between *pf* and climbing-fibre (*cf*) activity. If a *cf* spike occurs coincident with a particular phasic *pf* signal, the efficacy of the *pf–Pc* synapse will decrease by some small amount Δw_i^- by heterosynaptic long term depression (LTD). Conversely, (we assume), that if a signal appears on a *pf* but there is no correlating *cf* spike, synaptic efficacy will increase (long term potentiation (LTP)) by some small amount, Δw_i^+. The precise dynamics of this change in efficacy are unknown and may involve hetero- and mono-synaptic plasticity and other cells (basket and stellate cells). Since *cf* spontaneous rates are low, the probability of a *cf* spike occurring is much lower than a *cf* spike not occurring simultaneously with a *pf* event, and therefore, to maintain an equilibrium in w_i, LTD would have to be much stronger than LTP, which is consistent with the profuse ramifications of a *cf* on its target *Pc* dendrites.

The right-hand side of Eq. (8) could be represented by *cf* activity, where the low background *cf* activity can be identified with the exploratory random process component Δ_i, which is always present and is determined by spontaneous activity of neurons in the inferior olive (*IO*). Superimposed on this background is an activity reflecting the gradient of the cost function. This would be obtained by correlating the spontaneous activity itself with a cost signal *J*, arising from sensory, proprioceptive, or efference copy signals relayed to the *IO* from various sources. Thus, an increase in the correlation between cost signal at *IO* input and the spontaneous output of the *IO* leads to an increase in spontaneous rate, which in turn leads to a decrement in w_i; vice versa, a decrease in correlation leads to a decrease in spontaneous rate and an increase in w_i. Therefore, the net effect will be that w_i will change to minimise *J* with respect to w_i. To minimise *J* with respect to all weights, each *pf–Pc* synapse would require its own stochastic signal that is statistically independent of the others. This would be fulfilled by each *cf* carrying a random spontaneous rate and projecting to only one *Pc*. Since spontaneous rates are low (which would be necessary to keep Δ_i low), there would be very little dynamic range to code gradients of different signs. Thus, positive and negative gradients would need to be represented separately, possibly by cerebellar lateralisation.

5.3. Unrefenced learning vs error feedback

Although we have only outlined this scheme with broad strokes, it is immediately clear how different such as system would appear to the electrophysiologist from a system with error feedback.

In stochastic gradient descent, the *cf* signals represent gradient information about the cost function rather than error (except of course when squared error really

is the cost function). When the controlled behaviour is optimal, the gradients will be minimal so that only the low spontaneous random signals would be recorded. If the environment changed (either naturally or through some experimental manipulation) the weights would no longer be optimal and cost gradients would become significantly different from zero causing some *cf*'s to increase and others to decrease their firing rate (Fig. 8). Gradually, weights would be adjusted until the new optimim was found and the *cf*'s return to their quiescent background rate. Thus the optimum would be tracked in time. *Cf*'s would carry neither motor nor sensory signals, but signals responding to unexpected sensory consequences of the behaviour (see Gellman et al. (1985)). It must be emphasised that the randomness of *cf* activity is central to this scheme, as it reflects an active process for exploring the cost space, rather than being simply neural noise.

Another aspect of gradient descent is that the cost function is a scalar, not a vector, and could be in quite a different domain than the behaviour that is being controlled. For example, in saccade flight-time minimisation, the cost is in the domain of time yet it is position that is being controlled. Thus, we would not expect a vectorial representation of the cost as in error feedback. The actual internal signal representing cost could be any monotonic function of actual cost, since it is only the minimum that needs to be preserved. Thus, there is no requirement for linearity in the system.

Clearly, the above scheme can be elaborated so that the dynamics of the optimisation could also be adjusted. Thus, the Δ_i could be changed during descent, which could speed up convergence or reduce the chance of becoming trapped in a local minima ('annealing'). A similar effect could be obtained by changing the correlation among the Δ_i by changing σ_{ij}. It is tempting to speculate that electrotonic coupling among *IO* neurons (Llinas et al., 1974) could achieve this.

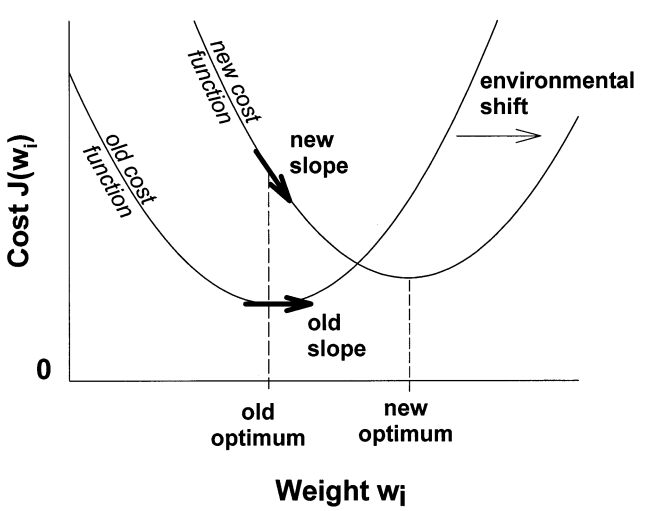

Fig. 8. A change in the environment shifts the cost function, which requires a change in synaptic weights to produce the new optimal behaviour. The shift causes a change in gradient from nominally zero at the old optimum to a new non-zero value, which is used to guide the optimisation process to the new minimum.

Although this gradient descent method would find the minimum of an aribitrary costs function (provided Eq. (8) holds), an important limitation is that the change in cost and the change in weights need to be contemporaneous (within an epoch). Delays of more than one epoch between the behaviour and its cost would prevent convergence on the optimum. For delayed 'reinforcement' the process would need to be driven not by actual cost but by *predicted* cost, which would also need to be learnt, sometimes called the 'adaptive critic' (see Barto (1992), Werbos (1992)).

6. Discussion

6.1. Assumption of optimality

When we observe the natural world, we see patterns of morphological invariance, which we categorise taxonomically into various species etc. From a Darwinistic viewpoint, natural selection has favoured a group of individuals with similar genotype and phenotype that cluster around a local maximum in fitness space for a given environment. Likewise, we argue, when we observe invariant patterns in behaviours, we infer that some limit has been reached reflecting a local optimum in fitness. If this were not the case, natural selection would select a fitter behaviour and there would be no invariance. It is likely that this deduction is neither provable nor disprovable, hence we have called it the *assumption of optimality*. The value of this assumption is that it is a rich source of hypotheses and potentially may unify a wide range of behavioural phenomenology.

In order to understand behaviour, the assumption of optimality focuses on the need to know the Performance Index (PI), that is, what is being achieved by the behaviour, and the constraints, that is, the biophysical limitations in reducing the PI as low as possible. This is in contrast to the often implicit neurophysiological view that a behaviour can be understood by completely mapping out in detail the entire motor, neuromuscular, and muscular pathway. Instead, we argue, that whatever the plant dynamics, the neural signals (and the neural architecture) will evolve–adapt–learn the appropriate innervation for the optimal output. Put simply, to understand behaviour we need to recognise that the activity of neurons is ultimately determined by the behaviour and its interaction with the environment.

This Darwinistic viewpoint is most pertinent in developing systems. Infant reaching behaviour is quite different from the normal adult behaviour; reaching movements (whether with saccades or the arm) are

composed of multiple sub-movements that are highly variable (Aslin and Salapatek, 1975; von Hofsten, 1991). Both Harris (1995), Berthier (1996) have argued that the immature reaching results from the uncertainty in generating motor behaviour, but they are nevertheless optimal (or near-optimal) with the PI of minimal time. Thus, the development of motor performance reflects changes in the constraints (e.g., reduction in motor and sensory uncertainty) but not in the PI.

This view may prove to be useful in understanding behaviour in some clinical conditions. For example, patients with homonymous hemianopia (loss of vision in one half field) produce saccades that have considerable undershoot (as opposed to being scattered all over the blind hemifield), as well as large variability when made to targets in the blind field (Mezey et al., 1998). This is what we would expect if there were very large target uncertainty. Again, we propose, the constraints have changed but the PI remains the same.

6.2. PI and constraints

Given that we can identify invariance in behaviour patterns, our task is to find a *biologically plausible* PI and constraints. This is not easy. Although we may sometimes need the tools of the optimal control engineer to test a hypothesis, the problem is not a design problem *per se*. We must not fall into the trap that when observations cannot be reconciled with our theoretical optimal behaviour, we conclude the behaviour to be not optimal. To the contrary, we should assume that the behaviour is optimal (the null hypothesis), but that we have not discovered the appropriate PI and–or the correct constraints. In particular, the assumptions in finding the theoretical optimum need to be carefully examined. Such assumptions may be subtle but can lead to wildly different 'optimal' behaviour, as illustrated in Section 4.

Since optimality refers to fitness, and a major component of fitness is survival, optimal behaviour depends inevitably on future events. Neither the individual nor 'evolution' can foretell the future, and the optimisation must depend on the statistical expectation of the future (for type-2 behaviours). We have attempted to show that even for simple goal-directed saccades, the trajectory (and hence all the usual parameters of the movement) may be optimised with reference to the future outcome of the behaviour, which we have hypothesised to be total flight-time of the corrective as well as the primary movements. A similar hypothesis for the control of arm movements has been proposed by Meyer et al. (1988).

One obvious difficulty in finding the PI and constraints is determining the epoch of the behaviour; that is, how far into the future is a behaviour optimised? We have chosen goal-directed saccades and arm movements because it seems plausible that the adaptive horizon ends once the target has been reached. However, this can only be partly correct. We make many saccades of different magnitudes, and the optimal gain for one magnitude may not be optimal for a different magnitude. One way to optimise a range of saccades would be to optimise separately for every possible magnitude. This would be inefficient because it would require a long time to learn, and does not capitalise on the fact that the curve of cost vs gain for one amplitude is similar (but not precisely the same) as the cost vs gain for a saccade of different magnitude. Thus, it would pay to optimise in gain rather than in absolute amplitude, as seen empirically (Deubel et al., 1986). However, this could not be perfect so the gain for a saccade would depend to some extent on previous distribution of executed saccades. Whether this can explain the so-called 'range effect' remains to be seen.

It is also not clear whether saccades are always goal-directed. Most saccades occur during visual scanning in which fixations are distributed over large areas of the visual scene, not necessarily on individual point targets. Although such scanning may minimise flight-time (see Harris (1993)), the epoch is unclear.

6.3. Unreferenced control

An important aspect of minimal time is that the goal of the behaviour cannot be coded ahead of time and cannot be used as a reference for error feedback control. Instead, the optimal behaviour must be 'discovered'. To find the optimum, it is necessary to find how the PI changes with the control parameters (e.g., gain), that is, the gradient of the cost surface. This, in turn, requires some degree of exploratory behaviour, which is not necessarily cognitive or deterministic, but may be quite random as we have described. Thus, variability in behaviour is intrinsic to the adaptive process rather than being just biological noise. Consequently, in order to reduce cost in the long-run, it is necessary to introduce variability which raises cost in the short-run. This is an important issue in understanding development. If correct, then we would expect that some of this variability to be exploratory. Indeed, in the face of large variability (e.g., due to poor target specification), substantial exploratory variability would be needed to extract gradient information.

There is a qualitative leap in understanding and modelling biological neural circuits that can find an unreferenced optimum state, rather than control circuits that reach a desired or pre-determined state (with or without prediction). A circuit 'designed' to minimise an arbitrary cost function can easily be used for error control by simply assigning squared error as the cost. However, a system designed for error control requires a reference and cannot be used to minimise an arbitrary

cost function—referenced control is a subset of unreferenced control, not vice versa. It has often been noted that the basic cerebellar circuit is remarkably stereotyped (see Bloedel (1992)). If we accept that the control of at least some behaviours must be unreferenced, then we must reject the notion that the basic cerebellar circuit is 'wired' specifically for error feedback.

6.4. Optimal optimisation

Finally, if we take our argument to its logical conclusion, not only would evolution have evolved structures for optimally controlling behaviour, but would not natural selection favour the optimal optimisation process? Instead of trying to fit the cerebellum and its functions into a standard control engineering design, perhaps we could stand to learn from the grand optimiser.

Acknowledgements

I am indebted to Professor Marcus Pembrey, Dr Daniel Wolpert, Dr Susan Goodbody, and Peter West for many informative discussions and suggestions. This work has been supported by The Medical Research Council, grant G9316292N, and the charities 'Iris Fund', 'Help Child to See' and 'Child Health Research Appeal Trust'.

References

Abrams RA, Meyer DE, Kornblum S. Speed and accuracy of saccadic eye movements: characteristics of impulse variability in the oculomotor system. J Exp Psychol Hum Percept Perf 1989;15:529–43.

Aslin R, Salapatek P. Saccadic localization of visual targets by the very young human infant. Percept Psychophys 1975;17:293–302.

Bahill TA, Adler D, Stark L. Most naturally occurring human saccades have magnitudes of 15° or less. Invest Ophthalmol 1975a;14:468–469.

Bahill AT, Clark MR, Stark L. The main sequence, a tool for studying human eye movements. Math Biosci 1975b;24:191–204.

Barto AG. Reinforcement learning and adaptive critic methods. In: White DA, Sofge DA, editors. Handbook of Intelligent Control: Neural, Fuzzy and Adaptive Approaches. New York: Van Nostarnd Reinhold, 1992:469–491.

Becker W. Metrics. In: Wurtz RH, Goldberg ME, editors. The Neurobiology of Saccadic Eye Movements. Amsterdam: Elsevier, 1989:13–67.

Berthier NE. Learning to reach: a mathamatical model. Dev Psychol 1996;32:811–23.

Bloedel JR. Functional heterogeneity with structural homogeneity: how does the cerebellum operate? Behav Brain Sci 1992;15:666–78.

Bronstein AM, Kennard C. Predictive eye saccades are different from visually triggered saccades. Vis Res 1987;27:517–20.

Bryson AE, Ho Y. Applied Optimal Control: Optimization, Estimation, and Control. New York: Hemisphere, 1975.

Clark MR, Stark L. Time optimal behavior of human saccadic eye movement. IEEE Trans Automat Contr 1975;AC-20:345–348.

Collewijn H, Erkelens CJ, Steinman RM. Binocular coordination of human horizontal saccadic eye movements. J Physiol (London) 1988;404:157–82.

Deubel H, Wolf W, Hauske G. Adaptive control of saccadic eye movements. Hum Neurobiol 1986;5:245–53.

Enderle JD, Wolfe JW. Time-optimal control of saccadic eye movements. IEEE Trans Biomed Eng BME 1987;34:43–55.

Flash T, Hogan N. The coordination of arm movements: an experimentally confirmed mathematical model. J Neurosci 1985;5:1688–703.

Fujita M. Adaptive filter model of the cerebellum. Biol Cybern 1982;45:195–206.

Gellman RS, Gibson AR, Houk JC. Inferior olivary neurones in the awake cat: Detection of contact and passive body displacement. J Neurophysiol 1985;54:335–48.

Harris CM, Wallman J, Scudder C. Fourier analysis of primate saccades. J Neurophysiol 1990;63:877–86.

Harris CM, Jacobs M, Shawkat F, Taylor D. The development of saccadic accuracy in the first 7 months. Clin Vis Sci 1993;8:85–96.

Harris CM. On the reversibility of Markov scanning in free viewing. In: Brogan D, Gale A, Carr K, editors. Vision Search 2. London: Taylor and Francis, 1993:123–135.

Harris CM. Does saccadic under-shoot minimize saccadic flight-time? a Monte-Carlo study. Vis Res 1995;35:691–701.

Henson DB. Corrective saccades: effects of altering visual feedback. Vis Res 1978;18:63–7.

Hogan N. An organizing principle for a class of voluntary movements. J Neurosci 1984;4:2745–54.

Ito M. The cerebellum and neural control. New York: Raven, 1984.

Lehman S, Stark L. Simulation of linear and nonlinear eye movement models: sensitivity analyses and enumeration studies of time optimal control. J Cybern Inf Sci 1979;2:21–43.

Lemij HG, Colewijn H. Differences in accuracy of human saccades between stationary and jumping targets. Vis Res 1989;12:1737–48.

Llinas R, Baker R, Sotelo C. Electrotonic coupling between neurons in cat inferior olive. J Neurophysiol 1974;37:541–9.

Meyer DE, Abrams RA, Kornblum S, Wright CE, Smith JEK. Optimality in human motor performance: ideal control of rapid aimed movements. Psychol Rev 1988;98:340–70.

Mezey L, Harris CM, Shawkat FS, Timms C, Kriss A, West P, Taylor DSI. Saccadic strategies in hemianopic children. Dev Med Child Neurol 1998; in press.

Optican LM, Robinson DA. Cerebellar-dependent adaptive control of primate saccadic system. J Neurophysiol 1980;44:1058–76.

Papoulis A. Probability, Random Variables and Stochastic Processes. 3rd ed. New York: McGraw–Hill, 1991.

Percival DB, Walden AT. Spectral Analysis for Physical Applications. Cambridge: Cambridge University Press, 1993.

Robinson DA. The mechanics of human saccadic eye movements. J Physiol (London) 1964;174:245–64.

Slepian D. Some comments on Fourier analysis, uncertainty, and modeling. SIAM Rev 1983;25:379–93.

Thrun SB. The role of exploration in learning control. In: White DA, Sofge DA, editors. Handbook of Intelligent Control: Neural, Fuzzy and Adaptive Approaches. New York: Van Nostrand Reinhold, 1992:527–559.

Toni I, Gentilucci M, Jeannerod M, Decety J. Differential influence of the visual framework on end point accuracy and trajectory specification of arm movements. Exp Brain Res 1996;111:447–54.

Viviani P. Flash, T. (1995) Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning. J Exp Psychol 1995;21:32–53.

Voltaire (1759) Candide and other stories. Translation by R. Pearson. Oxford: Oxford University Press, 1990.

von Hofsten C. Structuring of early reaching movements: a longitudinal study. J Motor Behav 1991;23:280–92.

Werbos PJ. Approximate dynamic programming for real-time control and neural modeling. In: White DA, Sofge DA, editors. Handbook of Intelligent Control: Neural, Fuzzy and Adaptive Approaches. New York: Van Nostrand Reinhold, 1992:493–525.

Wiegner AW, Wierzbicka MM. Kinematic models of human elbow flexion movements: quantitative analysis. Exp Brain Res 1992;88:665–73.