12 Optimal feedback control

Suppose you would like to retire wealthy and live in some tropical island. Currently, however, you are a young student taking classes at some university. What actions should you take today so that in 30 years, you will have reached your goal? Well, you have data which suggest a relationship between actions and consequences. This data comes from what you have read about other people's lives, from what you have experienced about your own actions and their consequences, and from what you have observed in your friends and family. This is your forward model: it gives you a way to predict what might be the result of any potential action. Given your goal (retire in Bahamas), this forward model (actions and their predicted consequences), and your current state (young student taking classes), you compute an optimum set of actions that describe your plan to get from your current state to your goal. (Perhaps, the fact that you are reading this book is part of this optimum sequence of actions!) However, your goal is very far away in time and it would seem rational to re-evaluate your plan occasionally. That is, as your state changes, it makes sense for you to reconsider what might be the best sequence of actions to get you to your goal.

Because the state you are in is not only influenced by your actions, but also by unforeseen and possibly unforeseeable events, it is probably not very productive for you to compute a specific sequence of actions that you would want to do in the future. Instead, you need to devise a policy that specifies what actions should be performed for each possible state that you might find yourself in. Ideally, for each time point in the future, the policy specifies the action that you would perform for each possible state. This is called a *feedback control policy*, as the actions that are performed at any given time depend on your belief[1] about your state at that time. In effect, feedback control policies are instructions about how to get to the goal from any state, at any time in the future.

It may seem hard to believe, but the problem of generating motor commands, say to move your arm, is really not that different from the problem of retiring in Bahamas. The objectives and timeframes are different, but the basic problem is similar:

[1] The idea that feedback depends upon a belief sounds like an oxymoron, because in the ordinary use of the word, beliefs are somewhat antagonistic to the concept of evidence or sensory information. One believes in a God regardless of any external evidence. Similarly, ideological belief is a construct that is not much affected by external inputs. However, here we use the term 'belief' in the Bayesian sense, i.e., a belief is a combination of prior knowledge and current evidence.

- Reaching the goal has a cost: The money you need to spend for retiring to the Bahamas or the effort that your muscles must make to reach a target.
- A forward model specifies what to expect from your actions: Where you will end up retiring or the sensory consequences of your motor commands, and
- In both cases you have a sensory system that provides you with information about your state.

The basic framework is summarized in Fig. 12.1. Given the goal of the task, and our belief about our current state, we generate some motor commands. These commands change the state of our body and our environment. Using our forward model, we predict what these state changes should be. We observe the state changes via our sensors. We combine what we predicted with what we observed to form a belief about the current state of the body/environment. Based on our current belief about our state, we apply our policy and form a new set of motor commands. The process continues until we reach our goal. Note that the sensory information may also lead to changes in the forward model. We may see, for example, that investing money on a particular stock may not be a good idea for retiring to the Bahamas and this will lead us to revise our policy. This kind of change, however occurs over a slower time scale, compare to the scale of our instantaneous decisions.

Whereas in the previous chapter we computed a sequence of motor commands while ignoring the sensory feedback, here we want to consider forming feedback control policies that continuously adjust the motor commands in response to the sensory feedback. Our rationale is that our movements can be perturbed, and we want to have in place a policy that provides us with the best possible motor commands, regardless of where we end up during the movement. That is, we want to compute a time-varying sensory-motor transformation that, given an estimate of state - based on the integration of predictions with observations, - we can compute the motor commands that are optimal in the sense that they will provide us with the best possible way to get to the goal.

12.1 Examples of feedback dependent motor control

In the previous chapter we focused on saccadic eye movements partly because these movements are so brief that sensory feedback appears to play no role in their control (Keller and Robinson, 1971;Guthrie et al., 1983). However, this does not mean that the state of the eye is not monitored during a saccade. Indeed, there is data suggesting that the motor commands that move the eyes during a saccade benefit from an internal feedback system, possibly a forward model that monitors the motor commands and predicts their sensory consequences. Let us consider some of the evidence regarding this state and goal-dependent process of generating motor commands.

Saccades are sometimes accompanied by blinks. In a blink, the eyelids close and reopen in a movement that takes about 100ms (Fig. 12.2). However, during a blink the brain sends motor commands to not just the lids, but the eyes as well: as the brain sends the commands to the lid muscles, it also sends commands to the extra-ocular muscles, causing a displacement of the eyes. (Indeed, the motion of the eyes during a blink is not due to a mechanical interaction between the lid and the eyes, but rather it is due to the specific motor commands to the eyes.) Thus, a blink during a saccade is a natural disturbance that affects the states of the eyes. Klaus Rottach, John Leigh, and colleagues (1998) used a ‘search coil’ technique (which relies on a contact lens that is placed on the eyes) to measure eye and lid motion during a horizontal saccadic eye movement. They noted that a blink that occurred during a saccade significantly altered the kinematics of the saccade, slowing it initially and then producing an over-shoot in the trajectory of the eyes (Fig. 12.2). Remarkably, they found that at saccade end the eyes were accurately stopped at the target: “accuracy is almost unaffected” by the blink, they wrote. A more recent study in monkeys confirmed that despite the absence of visual feedback, blink-disturbed saccades were corrected mid-flight, producing near normal endpoint accuracy (Goossens and Van Opstal, 2000). Importantly, the motion of the eyes during a blink-disturbed saccade was not simply a combination of commands that produced a blink and commands that produced a normal saccade. Rather, the data suggested that when the motor commands that initiated a saccade were corrupted by additional commands that were associated with a blink, the motor commands that followed responded to the state of the eye in a process resembling feedback control, effectively steering the eyes to the target.

A second example of this mid-flight correction of saccades comes from an experiment in which transcranial magnetic stimulation (TMS) was used to perturb an ongoing movement. TMS is usually placed over a specific part of the brain in order to briefly disturb the activity of a small region directly beneath the stimulating coil. However, Minnan Xu-Wilson, Jing Tian, Reza Shadmehr, and David Zee (2011) discovered that no matter where they placed the coil on the head, it always disturbed an ongoing saccade (Fig. 12.3). The discharge of the TMS coil appeared to engage a startle-like reflex in the brain that sent inhibitory commands to the saccadic system, resulting in a disturbance to the eye trajectory at a latency of around 60ms (with respect to TMS pulse), and lasting around 25ms. Interestingly, despite the fact that the task was performed in the dark and the target was extinguished after saccade onset, the TMS-induced disturbance was immediately corrected with additional motor commands that guided the eyes to the target. This is consistent with a view that as the saccadic motor commands are generated by one part of the brain, another part receives a copy and estimates the resulting state of the eye. When a TMS pulse is given to the head, it engages a startle-like neural system that inhibits the ongoing motor commands. The system that monitors the oculomotor commands responds to this perturbation and corrects for the motor commands so the eye will arrive near the target. Therefore, the motor commands to the eyes during a saccade are not pre-programmed in some open-loop form, but depend on internal monitoring and feedback.

A third example of this mid-flight correction of saccades comes from an experiment in which the target was predictably moved at saccade onset. Haiyin Chen-Harris, Wilsaan Joiner, Vincent Ethier, David Zee, and Reza Shadmehr (Chen-Harris et al., 2008) performed an experiment in which people were shown a visual stimulus at 15^o on the horizontal meridian, i.e., at (15,0). As soon as the saccade started, the target was removed and a new target appeared at 5^o vertical displacement, i.e., at (15,5) (Fig. 12.4). Therefore, the saccade completed with an endpoint error. Trial after trial, the motor commands adapted in response to this endpoint error, moving the eyes away from the stimulus at (15,0) and toward (15,5). However, the result of adaptation was not a straight trajectory toward (15,5). (A straight trajectory is the normal response to a control target that appears at 15,5 or elsewhere, as shown by the trajectories labeled control in Fig. 12.4.) Rather, in this target-jump condition, saccades developed a curvature. (The saccades remained curved whether or not the target jumped on a particular trial, so we can be sure that the curvature was not due to visual input during the saccade.) The motor commands that moved the eyes appeared to be corrected as they were executed. The authors interpreted this data as evidence for a forward model that learned from endpoint errors: it learned to predict that a consequence of the motor commands was a displacement of the target. In effect, this internal feedback acted as a mechanism that steered the eyes toward the predicted position of the target.

In addition to internal feedback via a possible forward model, there are also instances in which sensory feedback affects the control of eye movements. A good example of this is in the case of natural eye movements in which both the eyes and head are free to move. Indeed, most eye movements are not done in isolation, but accompany head movements. An example of a natural (i.e., head-free) gaze shift is shown in Fig. 12.5A. The gaze shift begins with the eyes making a saccade, and the head soon follows. For a target at 40^o, the gaze (sum of head and eye positions) is on the target by around 90ms and maintains the target on the fovea, yet the head continues to rotate toward the target while the eyes rotate back to re-center with respect to the head. These natural gaze shifts are a good example of a coordinated motion in which multiple body parts cooperate in order to achieve a common goal: maintain the position of the target on the fovea. Emilio Bizzi, Ronald Kalil, and Vinenzo Tagliasco (1971) used this simple movement to answer a fundamental question: were the motor commands to the eyes pre-programmed and open-loop, or did these commands depend on the sensory feedback that measured the state of the head? To answer this question, they devised an apparatus that on random trials held the head stationary by applying a brake. They found that if the head was not allowed to move, the eyes made a saccade to the target, but did not rotate back. This was the case even if the visual stimulus was removed at saccade onset and the gaze shift took place in darkness. The experiment was later repeated by Daniel Guitton and Michel Volle (1987), whose data are shown in Fig. 12.5B. On a randomly selected trial the target was shown at 40^o but the head was not allowed to rotate. The eyes made a saccade, but because the head was not allowed to move, the eyes did not rotate back. When the target was shown at 80^o, normally the eyes make a 30^o saccade as the head rotates toward the target. However, when on a randomly selected trial the head was not allowed to rotate (brake condition, Fig. 12.5B), the eyes made a larger amplitude saccade as compared to when the head was free (Fig. 12.5A), and the eyes did not rotate back until the brake was released and the head was allowed to move. The fact that the eyes exhibited a larger displacement during head-braked trials is summarized in Fig. 12.5C. Together, these data demonstrate that during head-free eye movements, the motor commands to the eyes are not ‘open-loop’ but depend on the state of the head.

These examples are consistent with the idea that the motor commands that move our body rely on two forms of feedback: internal predictions regarding state of the body/environment (Figs. 12.2-4), and sensory observations (Fig. 12.5). Let us now show that the feedback gains are also a reflection of available resources and expected rewards. That is, motor commands that are produced in response to sensory feedback are optimized with respect to some cost function. Joern Diedrichsen (Diedrichsen, 2007) considered a reaching task in which the two arms cooperated and shared a common goal. In one version of the task (one-cursor condition, Fig. 12.6A), there was a single cursor that reflected the average position of the left and the right hand. In this one-cursor condition, the goal was to move the cursor to the single target. There was also a two-cursor condition in which there was a cursor associated with each arm. In the two-cursor condition, the goal was to move each cursor to its own target. Joern’s idea was that the feedback gains associated with how each arm responded to a given perturbation should be different in these two conditions. In the two-cursor condition, the right arm should not respond to a perturbation that displaced the left arm, whereas in the one-cursor condition, the right arm should respond to this perturbation. The reason is that in the one-cursor condition (but not the two-cursor condition), the two arms have a common goal. This common goal should translate into cooperative behavior in which a perturbation that affects one arm should be handled by a reaction by both arms. In Fig. 12.6B we see that following a left-ward perturbation to the left hand, the right hand moved slightly to the right in the one-cursor condition but not the two-cursor condition. This is illustrated better in the velocities of the left and the right arms in Fig. 12.6C and 12.6D: about 200ms after move onset, the right arm responded to the perturbation to the left arm (Fig. 12.6D). That is, the right arm takes up some of the correction when the left arm is perturbed, but only if the two arms are working together in a task in which they share a common goal.

Suppose that in the one cursor condition, a perturbation is given to the left arm. Suppose that this perturbation would require 2 N to compensate. Further suppose that motor commands cost us proportional to the squared magnitude. So if the left arm alone compensated for this perturbation, it would cost us 4 N^2. The optimum thing to do for this particular cost is to have the left and right arms each produce 1N. Now the cost is 1 N^2 + 1 N^2, or 2 N^2. Cooperation leads to a smaller total cost. If we think of the force that each arm produces in response to the perturbation (a displacement) as a feedback gain, these results provide us with two ideas: 1) in the one-cursor task the feedback gain of the left arm should be smaller than in the two-cursor task, and 2) in the one-cursor task the feedback gain of the right arm should depend on the state of the left arm, whereas in the two-cursor task this feedback gain should depend only on the state of the right arm. What we need is a principled way to set these feedback gains so that the limbs cooperate to generate motor commands and bring the cursor to the goal in some efficient way.

In summary, during a movement the motor commands depend on the state of the body part that is being controlled, as well as the overall goal of the task. How does one generate motor commands that depend on the state of the body while simultaneously optimizing some long-term goal? We will consider this question in this chapter.

12.2 A brief history of ideas in biological control

Early theories of motor control stressed feedback mechanisms as the dominant circuits for voluntary movements. Sir Charles Sherrington (1923) looked at movements as “chains” of reflexes. He wrote: “Coordination, therefore, is in part the compounding of reflexes… This compounding of reflexes with orderliness of coadjustment and of sequence constitutes coordination, and want of it incoordination.” His idea was to explain all the richness of motor behavior as a combination of simpler and automatic transformations of senses into actions. As we discussed in Chapter 3, developments in robotics shifted the focus of motor neuroscientists from feedback to preprogrammed, or “feedforward”, control. The nonlinear dynamics of our limbs and the relatively long delays of sensory feedback would make it hard or impossible to generate stable movements without some form of anticipatory mechanism to compensate for inertial forces. Therefore the brain must be able to prepackage motor commands based on implicit knowledge of the body’s mechanics. As it often happens in science one extreme view was replaced by another. First, it was all reflexes, and then it became all open-loop control. But while there is evidence for the brain preprogramming movement patterns, there is also equally strong evidence for our ability to correct movements “on the fly” in response to incoming information. This is common sense and it has also been demonstrated by numerous studies of reaching movements, where the target is suddenly changed after the movement starts or the limb perturbed during the movement.

So, while movements are executed based on prior beliefs, our brain also pays attention to the incoming stream of sensory information. But how does it respond to sensory information? The idea of simple fixed reflexes is clearly inadequate. For example, recall that Cordo and Nashner (1982) showed how the muscles at the ankle became active when subjects are pulling on a handle, but did not show activity with the same pulling action if a bar insured stability to the body (Fig. 4.1). The ides is that the response to sensory information is modulated by prior knowledge about the environment in which we move.

Optimal feedback control provides a framework to move beyond the antagonism of feedback vs. feedforward toward a view in which both prior beliefs and sensory-based actions coexist. In this view, a fundamental outcome of motor learning is to shape feedback, by establishing, on the basis of experience, how the brain must respond to incoming sensory input. Through learning, we acquire knowledge of the statistical properties of the environment; we learn how sensory inputs as well as motor commands are affected by uncertainty and how uncertainty itself has structure, for instance being larger in some directions than others. This knowledge is essential to tune future motor responses to sensory information. How can this tuning of feedback parameters be done optimally, so that we can reach our goals with minimal error and effort? Unfortunately the mathematical tools at our disposal are limited and the answer to this question can be only be given for simple systems and simple forms of uncertainty. Nevertheless, answers in these simple cases can guide us toward the development of methods with broader range of applications. In the following sections we show how the combination of optimal control and optimal state estimation provide us with a way to relate feedback gains to the dynamical properties of a linear control system and to the statistical properties of signal dependent noise in motor commands and sensory signals.

12.3 Bellman optimality principle

Our aim is to produce motor commands so that they not only achieve a goal in some optimal sense (e.g., minimize a cumulative cost), but also respond to feedback, i.e., we want the motor commands to be the best that they can be no matter which state we find ourselves in. A framework that is appropriate for solving this problem is optimal feedback control: we have a goal that we wish to achieve, and the goal is specified as a cost per unit of time (e.g., accuracy cost and effort cost). For example, suppose that we have a state specified by vector $\mathbf{x}$, and motor commands specified by vector $\mathbf{u}$. For simplicity, we assume that effort is a linear function of the command vector. Engineers use the term “control cost” because machinery is not yet endowed with a sense of effort. We have a cost per unit of time (i.e., cost per step) that depends on our state (which includes the goal state), and effort:

$$\alpha^{(k)} = \mathbf{u}^{(k)T} L\mathbf{u}^{(k)} + \mathbf{x}^{(k)T} T^{(k)}\mathbf{x}^{(k)} \tag{12.1}$$

The above cost is quadratic, but our discussion at this point does not require any specific form for the cost per step. So without loss of generality, suppose that the cost per step is of the form shown in Eq. (12.1). We will assume a finite horizon to our goal, which means that we wish to achieve the goal within some specified time period p. Our objective is to find a *policy* $\mathbf{u}^{(k)} = \pi\left(\mathbf{x}^{(k)}\right)$ such that at each time step k, we can transform our state $\mathbf{x}^{(k)}$ into motor commands $\mathbf{u}^{(k)}$. If this policy is optimal, depicted by the term π^*, then it will minimize the sum total of costs $\sum_{k=0}^{p} \alpha^{(k)}$ from our initial time step $k = 0$ to the end step $k = p$.

To find this policy, we will rely on a fundamental observation of Richard Bellman (1957) (his chapter 3.3), who wrote the following:

> *The Principle of Optimality*. An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.

To explain his idea, suppose that we start at the last time point p. We find ourselves at state $\mathbf{x}^{(p)}$. What is the optimum action $\mathbf{u}^{(p)}$ that we can perform? Because we are at the last time point, we have run out of time, and nothing that we can do now will have any bearing. Therefore, the optimum thing to do from a cost standpoint (Eq. 12.1) is nothing, $\mathbf{u}^{(p)} = 0$. So the optimal policy for the last time point is $\pi^*\left(\mathbf{x}^{(p)}\right) = 0$. Using Eq. (12.1), let us assign a value to the state that we find ourselves at time point p, given that we are using policy π^*:

$$v_{\pi*}\left(\mathbf{x}^{(p)}\right) = \mathbf{x}^{(p)T} T^{(p)} \mathbf{x}^{(p)} \tag{12.2}$$

The function $v_{\pi*}$ assigns a number to each state, typically implying that the closer we are to the goal state (which is a part of the vector $\mathbf{x}^{(p)}$), the better. [In a bit of confusing terminology, the smaller the value v for some state, the more valuable that state is for us. At this point, the concept of value seems identical to the concept of cost. The difference will become clearer at the next step.] Now let us move back one step to time point $p-1$. If we find ourselves at state $\mathbf{x}^{(p-1)}$, what is the best action that we can perform? Suppose our policy $\pi\left(\mathbf{x}^{(p-1)}\right)$ instructs us to perform action $\mathbf{u}^{(p-1)}$. How good is this policy? Well, given that we are at state $\mathbf{x}^{(p-1)}$ and have performed action $\mathbf{u}^{(p-1)}$, we will have incurred a cost specified by $\alpha^{(p-1)}$ for that time step. Furthermore, the action $\mathbf{u}^{(p-1)}$ will have taken us from state $\mathbf{x}^{(p-1)}$ to $\mathbf{x}^{(p)}$ with probability $p\left(\mathbf{x}^{(p)} \middle| \mathbf{x}^{(p-1)}, \mathbf{u}^{(p-1)}\right)$. This probability is specified by the dynamics of the system that we are acting on. The state $\mathbf{x}^{(p)}$ has a value. It seems rational that the goodness of our policy should be related to both the cost it incurs on the current time step $p-1$, and the value of the state it takes us to in the next time step p. More precisely, according to Bellman, the value of any given state is the cost that we incurred in reaching that state (by whatever policy) plus the expected value produced by the optimal policy from that state to the goal. The reader may want to pause for a moment and see how this elegant concept is a reformulation of the optimality principle. More importantly, this idea can be directly translated into an equation. Thus, we assign a value to state $\mathbf{x}^{(p-1)}$ as follows:

$$v_{\pi}\left(\mathbf{x}^{(p-1)}\right)=\alpha^{(p-1)}+\int v_{\pi^*}\left(\mathbf{x}^{(p)}\right)p\left(\mathbf{x}^{(p)}\middle|\mathbf{x}^{(p-1)},\mathbf{u}^{(p-1)}\right)d\mathbf{x}^{(p)} \tag{12.3}$$

The right most term in Eq. (12.3) is the mean value of the state that the optimal motor command takes us to:

$$v_{\pi}\left(\mathbf{x}^{(p-1)}\right)=\alpha^{(p-1)}+E\left[v_{\pi^*}\left(\mathbf{x}^{(p)}\right)\middle|\mathbf{x}^{(p-1)},\mathbf{u}^{(p-1)}\right] \tag{12.4}$$

If our policy $\pi\left(\mathbf{x}^{(p-1)}\right)$ was optimum, then it would produce motor commands $\mathbf{u}^{(p-1)}$ that minimize the sum of current cost $\alpha^{(p-1)}$, plus the value of the state that it takes us to, i.e., minimize Eq. (12.4). [This is because the value of the next state is the smallest cost that we can incur if we were to perform the optimal policy at the next step.] Therefore, the optimum policy at time step $p-1$ has the following property:

$$\pi^*\left(\mathbf{x}^{(p-1)}\right)=\underset{\mathbf{u}(p-1)}{\arg\min}\left\{\alpha^{(p-1)}+E\left[v_{\pi^*}\left(\mathbf{x}^{(p)}\right)\middle|\mathbf{x}^{(p-1)},\mathbf{u}^{(p-1)}\right]\right\} \tag{12.5}$$

And so, if our policy was optimal, then the value of the state $\mathbf{x}^{(p-1)}$ would be related to the value of the state $\mathbf{x}^{(p)}$ as follows:

$$v_{\pi^*}\left(\mathbf{x}^{(p-1)}\right)=\min_{\mathbf{u}(p-1)}\left\{\alpha^{(p-1)}+E\left[v_{\pi^*}\left(\mathbf{x}^{(p)}\right)\middle|\mathbf{x}^{(p-1)},\mathbf{u}^{(p-1)}\right]\right\} \tag{12.6}$$

To help de-mystify the nomenclature, in Fig. 12.7 we have plotted what we mean by the terms min and argmin. The term min refers to the minimum value of a function, and the term argmin refers to the argument of a function for which the function has a minimum value. In general, the optimal policy $\pi^*\left(\mathbf{x}^{(k)}\right)$ has the property that it produces the smallest value possible for state $\mathbf{x}^{(k)}$, where 'value' refers to the cost that is accumulated if we were to apply the motor commands specified by this policy starting from time step k to the last time point p. If we write the expression in Eq. (12.6) for arbitrary time step k, we arrive at what is called the *Bellman Equation*:

$$v_{\pi^*}\left(\mathbf{x}^{(k)}\right)=\min_{\mathbf{u}(k)}\left\{\alpha^{(k)}+E\left[v_{\pi^*}\left(\mathbf{x}^{(k+1)}\right)\middle|\mathbf{x}^{(k)},\mathbf{u}^{(k)}\right]\right\} \tag{12.7}$$

Eq. (12.7) implies that if we knew the optimal value associated with states at time point $k+1$, i.e. $v_{\pi^*}\left(\mathbf{x}^{(k+1)}\right)$, then we could find the optimum motor command $\mathbf{u}^{(k)}$, and this would provide us with the value associated with $v_{\pi^*}\left(\mathbf{x}^{(k)}\right)$. That is, we could recursively solve our problem by starting at some state (usually the end state $\mathbf{x}^{(p)}$), form an optimal value function at this state, and then work backwards one step at a time. Using the Bellman equation (Eq. 12.7), we have a tool to break down the problem into smaller sub-problems.

12.4 Control policy

We are going to imagine that in order to make a movement, our brain formulates a goal in terms of a rewarding state that it wants to achieve. It has a model of how the motor commands influence the state of the body (i.e., a forward model), and if the movement is slow enough, the brain has access to sensory feedback during the movement. In reality, we do not know how the cost of achieving the goal is represented, or what costs may be involved in formulating the motor commands. However, there is some evidence that the rewarding state is discounted in time, i.e., it is better to get to the rewarding state sooner than later (Shadmehr et al., 2010). There is also some evidence that the efforts associated with the motor commands carry a cost, and the form of this effort cost is approximately quadratic (O'Sullivan et al., 2009). In the last chapter we showed that for control of very simple movements like saccades, motor commands that optimize the following cost function are fairly successful in accounting for movement kinematics:

$$J(p) = \mathbf{x}^{(p)T} T \mathbf{x}^{(p)} + \sum_{k=0}^{p} \mathbf{u}^{(k)T} L \mathbf{u}^{(k)} + \lambda \left(1 - \frac{1}{1+\beta p} \right) \tag{12.8}$$

In the above expression, the first term describes a cost for the distance to the goal state at the end of the movement, the second term describes a cost for the accumulated effort, and the last term describes a cost of time (a function that grows with movement duration). In our previous chapter we assumed that movements were open-loop (i.e., no feedback of any kind). For a given movement duration p we computed the sequence of motor commands that minimized the above cost. We then found the optimum movement duration by searching among the various total costs $J(p)$ for the one with the minimum cost. Here, we will assume that there is feedback during the movement. *We are no longer interested in computing the optimum motor commands. Rather, we wish to compute the optimum policy, something that transforms each state that we might encounter during the movement into the motor commands*. For a given movement duration p, our cost per step is:

$$\alpha^{(k)} = \mathbf{x}^{(k)T} T^{(k)} \mathbf{x}^{(k)} + \mathbf{u}^{(k)T} L \mathbf{u}^{(k)} + \frac{\lambda\beta}{1+\beta p} \tag{12.9}$$

The last term in Eq. (12.9) is the average cost of time per step, i.e., the last term in Eq. (12.8) divided by p. [The term $T^{(k)}$ may be zero for all time steps except the last.] We will consider a system in which the motor commands produce signal-dependent noise, and the sensory observations also suffer from noise. We will estimate the state of the system using a Kalman filter. Our objective is to apply Bellman's theory in order to formulate a control policy in which the motor commands depend on our current estimate of the state of the system. Emanuel Todorov (2005) considered this problem and was first to describe a solution for the case in which the noise was signal dependent (which appears to be the case for biological systems). The derivations in this chapter are based on his work. Once we derive the solution, we will apply it to some of the movements that we presented in Figs. 12.2-6.

Suppose that we have a system of the form:

$$\begin{aligned}\mathbf{x}^{(k+1)} &= A\mathbf{x}^{(k)} + B\left(\mathbf{u}^{(k)} + \boldsymbol{\varepsilon}_u^{(k)}\right) + \boldsymbol{\varepsilon}_x \\ \mathbf{y}^{(k)} &= H\left(\mathbf{x}^{(k)} + \boldsymbol{\varepsilon}_s^{(k)}\right) + \boldsymbol{\varepsilon}_y\end{aligned} \qquad (12.10)$$

where $\boldsymbol{\varepsilon}_x$ and $\boldsymbol{\varepsilon}_y$ are zero mean Gaussian noise vectors with variance Q_x and Q_y:

$$\begin{aligned}\boldsymbol{\varepsilon}_x &\sim N\left(\mathbf{0}, Q_x\right) \\ \boldsymbol{\varepsilon}_y &\sim N\left(\mathbf{0}, Q_y\right)\end{aligned} \qquad (12.11)$$

and $\boldsymbol{\varepsilon}_u$ and $\boldsymbol{\varepsilon}_s$ are zero mean signal dependent noise terms, meaning that noise depends on the motor commands $\mathbf{u}$ and state $\mathbf{x}$, respectively:

$$\begin{array}{ll}\boldsymbol{\varepsilon}_u^{(k)} \equiv \begin{bmatrix} c_1 u_1^{(k)} \phi_1^{(k)} \\ c_2 u_2^{(k)} \phi_2^{(k)} \\ \vdots \\ c_m u_m^{(k)} \phi_m^{(k)} \end{bmatrix} & \boldsymbol{\varepsilon}_s^{(k)} \equiv \begin{bmatrix} d_1 x_1^{(k)} \mu_1^{(k)} \\ d_2 x_2^{(k)} \mu_2^{(k)} \\ \vdots \\ d_n x_n^{(k)} \mu_n^{(k)} \end{bmatrix} \\ \phi \sim N(0,1) & \mu \sim N(0,1) \\ c_i \geq 0 & d_i \geq 0 \end{array} \qquad (12.12)$$

The signal dependent motor noise $\boldsymbol{\varepsilon}_u$ affects the state $\mathbf{x}$ and the signal dependent sensory noise $\boldsymbol{\varepsilon}_s$ affects the observation $\mathbf{y}$. It is useful to express the signal dependent noise terms as a linear function of $\mathbf{u}$ and $\mathbf{x}$. To do so, we define:

$$
\begin{aligned}
&C_1 \equiv \begin{bmatrix} c_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \ddots \end{bmatrix} \quad C_2 \equiv \begin{bmatrix} 0 & 0 & 0 \\ 0 & c_2 & 0 \\ 0 & 0 & \ddots \end{bmatrix} \\
&D_1 \equiv \begin{bmatrix} d_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \ddots \end{bmatrix} \quad D_2 \equiv \begin{bmatrix} 0 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & \ddots \end{bmatrix}
\end{aligned}
\tag{12.13}
$$

and so we have:

$$
\begin{aligned}
\boldsymbol{\varepsilon}_u^{(k)} &= \sum_{i=1}^{m} C_i \mathbf{u}^{(k)} \phi_i^{(k)} \\
\boldsymbol{\varepsilon}_s^{(k)} &= \sum_{i=1}^{n} D_i \mathbf{x}^{(k)} \mu_i^{(k)}
\end{aligned}
\tag{12.14}
$$

In Eq. (12.14), m is the dimension of the vector $\mathbf{u}$ and n is the dimension of the vector $\mathbf{x}$. Because ϕ and μ are Gaussian random variables, $\boldsymbol{\varepsilon}_u$ and $\boldsymbol{\varepsilon}_s$ are also Gaussian with the following distribution:

$$
\begin{aligned}
\boldsymbol{\varepsilon}_u^{(k)} &\sim N\left(\mathbf{0}, \sum_{i=1}^{m} C_i \mathbf{u}^{(k)} \mathbf{u}^{(k)T} C_i\right) \\
\boldsymbol{\varepsilon}_s^{(k)} &\sim N\left(\mathbf{0}, \sum_{i=1}^{n} D_i \mathbf{x}^{(k)} \mathbf{x}^{(k)T} D_i\right)
\end{aligned}
\tag{12.15}
$$

And so our system has the following dynamics:

$$
\begin{aligned}
\mathbf{x}^{(k+1)} &= A\mathbf{x}^{(k)} + B\mathbf{u}^{(k)} + \boldsymbol{\varepsilon}_x + B\sum_i C_i \mathbf{u}^{(k)} \phi_i^{(k)} \\
\mathbf{y}^{(k)} &= H\mathbf{x}^{(k)} + \boldsymbol{\varepsilon}_y + H\sum_i D_i \mathbf{x}^{(k)} \mu_i^{(k)} \\
\boldsymbol{\varepsilon}_x &\sim N\left(\mathbf{0}, Q_x\right) \quad \boldsymbol{\varepsilon}_y \sim N\left(\mathbf{0}, Q_y\right) \\
\phi &\sim N(0,1) \qquad \mu \sim N(0,1)
\end{aligned}
\tag{12.16}
$$

We estimate of the state of our system by combining our predictions with our sensory observations using the Kalman framework:

$$
\begin{aligned}
\hat{\mathbf{x}}^{k|k} &= \hat{\mathbf{x}}^{k|k-1} + K^{(k)}\left(\mathbf{y}^{(k)} - H\hat{\mathbf{x}}^{k|k-1}\right) \\
\hat{\mathbf{x}}^{k+1|k} &= A\hat{\mathbf{x}}^{k|k} + B\mathbf{u}^{(k)}
\end{aligned}
\tag{12.17}
$$

Our estimate of state at time point $k+1$ is simply our prior estimate at time point $k+1$, $\hat{\mathbf{x}}^{(k+1)} \equiv \hat{\mathbf{x}}^{k+1|k}$. In general, our estimate of state at any time point $k+1$ is related to our observations on the previous trial $\mathbf{y}^{(k)}$, our prior beliefs $\hat{\mathbf{x}}^{(k)}$, and motor commands $\mathbf{u}^{(k)}$ as follows:

$$
\hat{\mathbf{x}}^{(k+1)} = A\hat{\mathbf{x}}^{(k)} + AK^{(k)}\left(\mathbf{y}^{(k)} - H\hat{\mathbf{x}}^{(k)}\right) + B\mathbf{u}^{(k)}
\tag{12.18}
$$

At this point we have described a procedure for estimating the state $\hat{\mathbf{x}}^{(k)}$. What motor command should we produce at this time point? We need to compute the policy $\pi^*\left(\hat{\mathbf{x}}^{(k)}\right)$ that transforms our estimate $\hat{\mathbf{x}}^{(k)}$ into motor command $\mathbf{u}^{(k)}$. Let us start at the last time point and consider what the value function $v_{\pi^*}\left(\mathbf{x}^{(p)},\hat{\mathbf{x}}^{(p)}\right)$ might look like. Our cost at this last time point is specified by Eq. (12.9). At the last time point the best thing to do is nothing, that is, we will set $\mathbf{u}^{(p)}=\mathbf{0}$. When we do so, the cost at this last time point is a quadratic function of $\mathbf{x}^{(p)}$, and so the value function at this last time point is:

$$v_{\pi^*}\left(\mathbf{x}^{(p)},\hat{\mathbf{x}}^{(p)}\right)=\mathbf{x}^{(p)T}T^{(p)}\mathbf{x}^{(p)}+\frac{\lambda\beta}{1+\beta p} \tag{12.19}$$

In general, for linear systems for which the cost per step (Eq. 12.9) is quadratic in state, the value of the states under the optimal policy is also quadratic. For this reason, Todorov (2005) hypothesized that the value function for any given time step has the following form:

$$v_{\pi^*}\left(\mathbf{x}^{(k)},\hat{\mathbf{x}}^{(k)}\right)=\mathbf{x}^{(k)T}W_x^{(k)}\mathbf{x}^{(k)}+\left(\mathbf{x}^{(k)}-\hat{\mathbf{x}}^{(k)}\right)^T W_e^{(k)}\left(\mathbf{x}^{(k)}-\hat{\mathbf{x}}^{(k)}\right)+w^{(k)} \tag{12.20}$$

For the last time point, Eq. (12.20) certainly seems reasonable because if we set $W_x^{(p)}=T^{(p)}$, $W_e^{(p)}=0$, and $w^{(p)}=\dfrac{\lambda\beta}{1+\beta p}$, we simply get the cost $\alpha^{(p)}$ under the optimal policy of $\mathbf{u}^{(p)}=\mathbf{0}$. If our policy was optimal, then the value function in one step would be related to the value function in the next step via the Bellman equation:

$$v_{\pi^*}\left(\mathbf{x}^{(k)},\hat{\mathbf{x}}^{(k)}\right)=\min_{\mathbf{u}(k)}\left\{\alpha^{(k)}+E\left[v_{\pi^*}\left(\mathbf{x}^{(k+1)},\hat{\mathbf{x}}^{(k+1)}\right)\middle|\mathbf{x}^{(k)},\hat{\mathbf{x}}^{(k)},\mathbf{u}^{(k)}\right]\right\} \tag{12.21}$$

To find the optimal control policy we will proceed in the following three steps:

1. Starting at time step $k+1$, we will assume that the value function $v_{\pi^*}\left(\mathbf{x}^{(k+1)},\hat{\mathbf{x}}^{(k+1)}\right)$ has the form specified by Eq. (12.20). We will apply the Bellman equation (Eq. 12.21) and compute $v_{\pi^*}\left(\mathbf{x}^{(k)},\hat{\mathbf{x}}^{(k)}\right)$.
2. Next, we will find the motor commands $\mathbf{u}^{(k)}$ that minimize $v_{\pi^*}\left(\mathbf{x}^{(k)},\hat{\mathbf{x}}^{(k)}\right)$.
3. By finding these motor commands, we will be able to check whether $v_{\pi^*}\left(\mathbf{x}^{(k)},\hat{\mathbf{x}}^{(k)}\right)$ has the form specified by Eq. (12.20). That is, by having $\mathbf{u}^{(k)}$, we will check whether the

value function at time step k is in fact a quadratic function of states. If it is, then we will have found the optimal motor commands for one time step.

By stepping back another step and so on, we will have found the optimal policy for all time steps. In the end, because we assume that we are controlling a linear system (Eq. 12.10), the value function will turn out to be quadratic both in the estimated state and in the motor command. Furthermore, the optimal policy will turn out to be linear in the state estimated at each step:

$$\mathbf{u}^{(k)} = -G^{(k)}\hat{\mathbf{x}}^{(k)}$$

Our goal is to derive the gain matrix $G^{(k)}$ from the statistical properties of the motor and sensory processes, both of which are affected by signal dependent noise (Eq. 12.14). At the end of our story, our policy will be the sequence of gain matrices $G^{(k)}$ for time steps $k = 0,\cdots, p$, allowing us to produce motor commands for whatever state $\hat{\mathbf{x}}$ we happen to find ourselves at.

Step 1.

To simplify the notation, we define the difference between our estimate of state and actual state as an estimation error:

$$\mathbf{e}^{(k)} \equiv \mathbf{x}^{(k)} - \hat{\mathbf{x}}^{(k)} \tag{12.22}$$

By combining Eqs. 12.16 and 12.18, we write the dynamics of error in our estimation of state as:

$$\begin{aligned}\mathbf{e}^{(k+1)} &= \left(A - AK^{(k)}H\right)\mathbf{e}^{(k)} + \boldsymbol{\varepsilon}_x + B\sum\nolimits_i C_i\mathbf{u}^{(k)}\phi_i^{(k)} \\ &\quad -AK^{(k)}\boldsymbol{\varepsilon}_y - AK^{(k)}H\sum\nolimits_i D_i\mathbf{x}^{(k)}\mu_i^{(k)}\end{aligned} \tag{12.23}$$

We re-write Eq. (12.20) as:

$$v_{\pi^*}\left(\mathbf{x}^{(k+1)},\hat{\mathbf{x}}^{(k+1)}\right) = \mathbf{x}^{(k+1)T}W_x^{(k+1)}\mathbf{x}^{(k+1)} + \mathbf{e}^{(k+1)T}W_e^{(k+1)}\mathbf{e}^{(k+1)} + w^{(k+1)} \tag{12.24}$$

To minimize Eq. (12.21), given that we are at state $\mathbf{x}^{(k)}$ and $\hat{\mathbf{x}}^{(k)}$, and have produced motor command $\mathbf{u}^{(k)}$, we need to compute the expected value of the above value function. To do so, we will need the expected value and variance of $\mathbf{x}^{(k+1)}$:

$$\begin{aligned}E\left[\mathbf{x}^{(k+1)}\middle|\mathbf{x}^{(k)},\hat{\mathbf{x}}^{(k)},\mathbf{u}^{(k)}\right] &= A\mathbf{x}^{(k)} + B\mathbf{u}^{(k)} \\ \operatorname{var}\left[\mathbf{x}^{(k+1)}\middle|\cdots\right] &= Q_x + \sum\nolimits_i BC_i\mathbf{u}^{(k)}\mathbf{u}^{(k)T}C_i^T B^T\end{aligned} \tag{12.25}$$

We will also need the expected value and variance of $\mathbf{e}^{(k+1)}$:

$$\begin{aligned} E\left[\mathbf{e}^{(k+1)}\middle|\cdots\right] &= \left(A - AK^{(k)}H\right)\mathbf{e}^{(k)} \\ \operatorname{var}\left[\mathbf{e}^{(k+1)}\middle|\cdots\right] &= Q_x + \sum\nolimits_i BC_i\mathbf{u}^{(k)}\mathbf{u}^{(k)T}C_i^T B^T + AK^{(k)}Q_y K^{(k)T}A^T \\ &\quad + \sum\nolimits_i AK^{(k)}HD_i\mathbf{x}^{(k)}\mathbf{x}^{(k)T}D_i^T H^T K^{(k)T}A^T \end{aligned} \tag{12.26}$$

With the above expressions we can compute the expected value of Eq. (12.24):

$$\begin{aligned} E\left[v_{\pi^*}\left(\mathbf{x}^{(k+1)},\hat{\mathbf{x}}^{(k+1)}\right)\middle|\mathbf{x}^{(k)},\hat{\mathbf{x}}^{(k)},\mathbf{u}^{(k)}\right] &= E\left[\mathbf{x}^{(k+1)T}W_x^{(k+1)}\mathbf{x}^{(k+1)}\middle|\cdots\right] \\ &\quad + E\left[\mathbf{e}^{(k+1)T}W_e^{(k+1)}\mathbf{e}^{(k+1)}\middle|\cdots\right] + w^{(k+1)} \end{aligned} \tag{12.27}$$

Recall that the expected value of a scalar quantity that depends on the quadratic form of a random variable $\mathbf{x}$ is:

$$E\left[\mathbf{x}^T A\mathbf{x}\right] = E[\mathbf{x}]^T AE[\mathbf{x}] + tr\left[A\operatorname{var}[\mathbf{x}]\right]$$

Therefore, the expected value terms on the right side of Eq. (12.27) are:

$$\begin{aligned} E\left[\mathbf{x}^{(k+1)T}W_x^{(k+1)}\mathbf{x}^{(k+1)}\middle|\cdots\right] &= \left(A\mathbf{x}^{(k)} + B\mathbf{u}^{(k)}\right)^T W_x^{(k+1)}\left(A\mathbf{x}^{(k)} + B\mathbf{u}^{(k)}\right) \\ &\quad + tr\left[W_x^{(k+1)}Q_x\right] + \mathbf{u}^{(k)T}\left(\sum\nolimits_i C_i^T B^T W_x^{(k+1)}BC_i\right)\mathbf{u}^{(k)} \end{aligned} \tag{12.28}$$

$$\begin{aligned} E\left[\mathbf{e}^{(k+1)T}W_e^{(k+1)}\mathbf{e}^{(k+1)}\middle|\cdots\right] &= \mathbf{e}^{(k)T}\left(A - AK^{(k)}H\right)^T W_e^{(k+1)}\left(A - AK^{(k)}H\right)\mathbf{e}^{(k)} \\ &\quad + tr\left[W_e^{(k+1)}\left(Q_x + AK^{(k)}Q_y K^{(k)T}A^T\right)\right] \\ &\quad + \mathbf{u}^{(k)T}\left(\sum\nolimits_i C_i^T B^T W_e^{(k+1)}BC_i\right)\mathbf{u}^{(k)} \\ &\quad + \mathbf{x}^{(k)T}\left(\sum\nolimits_i D_i^T H^T K^{(k)T}A^T W_e^{(k+1)}AK^{(k)}HD_i\right)\mathbf{x}^{(k)} \end{aligned} \tag{12.29}$$

We can now write the Bellman equation in a form that we can use to find the optimal motor command at time point k:

$$v_{\pi^*}\left(\mathbf{x}^{(k)},\hat{\mathbf{x}}^{(k)}\right) = \min_{\mathbf{u}^{(k)}}\left\{\begin{aligned} &\mathbf{u}^{(k)T}L\mathbf{u}^{(k)} + \mathbf{x}^{(k)T}T^{(k)}\mathbf{x}^{(k)} + \frac{\lambda\beta}{1+\beta p} \\ &+E\left[\mathbf{x}^{(k+1)T}W_x^{(k+1)}\mathbf{x}^{(k+1)}\middle|\mathbf{x}^{(k)},\hat{\mathbf{x}}^{(k)},\mathbf{u}^{(k)}\right] \\ &+E\left[\mathbf{e}^{(k+1)T}W_e^{(k+1)}\mathbf{e}^{(k+1)}\middle|\cdots\right] + w^{(k+1)} \end{aligned}\right\} \tag{12.30}$$

Step 2

To find the optimal motor command at time point k , we find the derivative of the sum on the right side of Eq. (12.30) with respect to $\mathbf{u}^{(k)}$ and set it equal to zero. Using Eq. (12.28) and Eq. (12.29), we can write Eq. (12.30) as:

$$v_{\pi*}\left(\mathbf{x}^{(k)},\hat{\mathbf{x}}^{(k)}\right)=\min_{\mathbf{u}^{(k)}}\left\{\begin{array}{l}\mathbf{u}^{(k)T}\left(L+C_x^{(k+1)}+C_e^{(k+1)}+B^TW_x^{(k+1)}B\right)\mathbf{u}^{(k)}+\dfrac{\lambda\beta}{1+\beta p}\\ +\mathbf{x}^{(k)T}\left(A^TW_x^{(k+1)}A+T+D_e^{(k+1)}\right)\mathbf{x}^{(k)}+2\mathbf{x}^{(k)T}A^TW_x^{(k+1)}B\mathbf{u}^{(k)}\\ +\mathbf{e}^{(k)T}\left(A-AK^{(k)}H\right)^TW_e^{(k+1)}\left(A-AK^{(k)}H\right)\mathbf{e}^{(k)}\\ +tr\left[W_x^{(k+1)}Q_x\right]+tr\left[W_e^{(k+1)}\left(Q_x+AK^{(k)}Q_yK^{(k)T}A^T\right)\right]+w^{(k+1)}\end{array}\right\}\quad(12.31)$$

with the following shortcuts:

$$\begin{aligned}C_x^{(k+1)}&\equiv\sum\nolimits_i C_i^TB^TW_x^{(k+1)}BC_i\\ C_e^{(k+1)}&\equiv\sum\nolimits_i C_i^TB^TW_e^{(k+1)}BC_i\\ D_e^{(k+1)}&=\sum\nolimits_i D_i^TH^TK^{(k)T}A^TW_e^{(k+1)}AK^{(k)}HD_i\end{aligned}\qquad(12.32)$$

After we find the derivative of sum in Eq. (12.31) with respect to $\mathbf{u}^{(k)}$ and set it equal to zero, we have:

$$\begin{aligned}G^{(k)}&\equiv\left(L+C_x^{(k+1)}+C_e^{(k+1)}+B^TW_x^{(k+1)}B\right)^{-1}B^TW_x^{(k+1)}A\\ \mathbf{u}^{(k)}&=-G^{(k)}\mathbf{x}^{(k)}\end{aligned}\qquad(12.33)$$

The above expression is our control policy for time step k . In practice, the state $\mathbf{x}^{(k)}$ is not observable and can only be estimated, and therefore the best that we can do is replace it with $\hat{\mathbf{x}}^{(k)}$:

$$\begin{aligned}\hat{\mathbf{x}}^{(k)}&=A\hat{\mathbf{x}}^{(k-1)}+AK^{(k-1)}\left(\mathbf{y}^{(k-1)}-H\hat{\mathbf{x}}^{(k-1)}\right)+B\mathbf{u}^{(k-1)}\\ \mathbf{u}^{(k)}&=-G^{(k)}\hat{\mathbf{x}}^{(k)}\end{aligned}\qquad(12.34)$$

Notice that the gain $G^{(k)}$ is inversely proportional to L , a variable that penalizes the effort expended on the task. Eq. (12.33) implies that the greater the effort cost, the smaller the feedback gain. A second point to notice is that the feedback gain $G^{(k)}$ depends on $C_e^{(k+1)}$, which in turn depends on $W_e^{(k+1)}$. This matrix is a weight that penalizes the 'squared' state estimation error $\mathbf{x}^{(k)}-\hat{\mathbf{x}}^{(k)}$ in Eq. (12.20). As we will see below, the estimation error depends on the Kalman gain $K^{(k+1)}$, implying that the feedback gain $G^{(k)}$ will also depend on the Kalman gain. The Kalman gain depends on the noise properties of the system, describing the uncertainty regarding our estimate of states. In summary, our policy, described as a time-dependent feedback gain, transforms our estimate of state into motor commands. This policy depends on the cost function that we are trying to minimize, as well as the uncertainties that we have regarding our estimate of state.

Step 3.

As our final step, we need to show that when we apply the policy in Eq. (12.33), the resulting value function $v_{\pi^*}\left(\mathbf{x}^{(k)},\hat{\mathbf{x}}^{(k)}\right)$ remains in the quadratic form that we assumed for $v_{\pi^*}\left(\mathbf{x}^{(k+1)},\hat{\mathbf{x}}^{(k+1)}\right)$. We insert $\mathbf{u}^{(k)}=-G^{(k)}\hat{\mathbf{x}}^{(k)}$ from Eq. (12.33) into Eq. (12.31) and we have:

$$\begin{aligned}v_{\pi^*}\left(\mathbf{x}^{(k)},\hat{\mathbf{x}}^{(k)}\right) &= \hat{\mathbf{x}}^{(k)T}G^{(k)T}B^TW_x^{(k+1)}A\hat{\mathbf{x}}^{(k)} - 2\hat{\mathbf{x}}^{(k)T}G^{(k)T}B^TW_x^{(k+1)}A\mathbf{x}^{(k)} \\ &+\mathbf{e}^{(k)T}\left(A-AK^{(k)}H\right)^T W_e^{(k+1)}\left(A-AK^{(k)}H\right)\mathbf{e}^{(k)} \\ &+\mathbf{x}^{(k)T}\left(T^{(k)}+A^TW_x^{(k+1)}A+D_e^{(k+1)}\right)\mathbf{x}^{(k)} \\ &+tr\left[W_x^{(k+1)}Q_x+W_e^{(k+1)}\left(Q_x+AK^{(k)}Q_yK^{(k)T}A^T\right)\right]+\frac{\lambda\beta}{1+\beta p}\end{aligned} \tag{12.35}$$

Using the identity $\hat{\mathbf{x}}^TZ\hat{\mathbf{x}}-2\hat{\mathbf{x}}^TZ\mathbf{x}=\left(\mathbf{x}-\hat{\mathbf{x}}\right)^T Z\left(\mathbf{x}-\hat{\mathbf{x}}\right)-\mathbf{x}^TZ\mathbf{x}$, we can simplify the first line of the above expression and remove the dependence on the interaction between $\hat{\mathbf{x}}^{(k)}$ and $\mathbf{x}^{(k)}$. When we do so, we arrive at the observation that the value function at time step k is quadratic:

$$\begin{aligned}v_{\pi^*}\left(\mathbf{x}^{(k)},\hat{\mathbf{x}}^{(k)}\right) &= \mathbf{x}^{(k)T}W_x^{(k)}\mathbf{x}^{(k)}+\mathbf{e}^{(k)T}W_e^{(k)}\mathbf{e}^{(k)}+w^{(k)} \\ W_e^{(k)} &\equiv \left(A-AK^{(k)}H\right)^T W_e^{(k+1)}\left(A-AK^{(k)}H\right)+G^{(k)T}B^TW_x^{(k+1)}A \\ W_x^{(k)} &\equiv T^{(k)}+A^TW_x^{(k+1)}A+D_e^{(k+1)}-G^{(k)T}B^TW_x^{(k+1)}A \\ w^{(k)} &\equiv tr\left[W_x^{(k+1)}Q_x+W_e^{(k+1)}\left(Q_x+AK^{(k)}Q_yK^{(k)T}A^T\right)\right]+\frac{\lambda\beta}{1+\beta p}\end{aligned} \tag{12.36}$$

With Eq. (12.36), we have used induction to prove that our policy is optimal as it satisfies the Bellman equation.

In summary, we start at time point p and set $W_x^{(p)}=T^{(p)}$, $W_e^{(p)}=0$, and $w^{(p)}=\dfrac{\lambda\beta}{1+\beta p}$.

From this we compute $G^{(p-1)}$ (Eq. 12.33). We then use Eq. (12.36) to compute $W_x^{(p-1)}$,

$W_e^{(p-1)}$, and $w^{(p-1)}$. From these weights we compute $G^{(p-2)}$, etc. As a result, we have a recipe to compute the feedback gains. Our feedback control policy is:

$$\pi^*\left(\hat{\mathbf{x}}^{(k)}\right) = -G^{(k)}\hat{\mathbf{x}}^{(k)}$$

12.5 The interplay between state-estimation and control policy

In Chapter 4 we considered the problem of state estimation for a system that had signal dependent noise, as in the system of Eq. (12.16). We found that the Kalman gain on step $k+1$ was affected by the size of the motor commands on step k. Let us briefly review that result, as it has an impact on our ability to compute an optimal control policy. We found that if on step k our prior state uncertainty is $P^{(k|k-1)}$, then the Kalman gain $K^{(k)}$ has the following form:

$$K^{(k)} = P^{(k|k-1)}H^T\left(HP^{(k|k-1)}H^T + Q_y + \sum_i HD_i\hat{\mathbf{x}}^{(k)}\hat{\mathbf{x}}^{(k)T}D_i^T H^T\right)^{-1} \qquad (12.37)$$

The state uncertainty has the following form:

$$\begin{aligned} P^{(k|k)} &= P^{(k|k-1)}\left(I - H^T K^{(k)T}\right) \\ P^{(k+1|k)} &= AP^{(k|k)}A^T + Q_x + \sum_i BC_i\mathbf{u}^{(k)}\mathbf{u}^{(k)T}C_i^T B^T \end{aligned} \qquad (12.38)$$

It is because of signal dependent noise that the Kalman gain is a function of the state estimate $\hat{\mathbf{x}}$. Furthermore, because state uncertainty depends on the motor commands, as motor commands increase in size so does state uncertainty. Therefore, the state uncertainty increases with the size of the motor commands, and the Kalman gain decreases with the size of the state. The implication being that if we are pushing a large mass (producing relatively large motor commands), then we will have a larger uncertainty regarding the consequences of these commands (as compared to pushing a small mass with a smaller amount of force). As a result, when we are producing large forces, the Kalman gain will be large, and we should rely more on the sensory system and our observations and less on our predictions.

Note that according to Eq. (12.37), the Kalman gain $K^{(k)}$ depends on $\hat{\mathbf{x}}^{(k)}$, which according to Eq. (12.34) depends on $\mathbf{u}^{(k-1)}$. When we minimized the value function in Eq. (12.31), we took the derivative of the sum with respect to $\mathbf{u}^{(k)}$. That sum had the term $K^{(k)}$ in it. Because $K^{(k)}$ does not depend on $\mathbf{u}^{(k)}$, our derivative is valid. However, we face a practical issue: in order to compute the sequence of feedback gains $G^{(0)}, G^{(1)}, \cdots, G^{(p)}$, we need to know the sequence of Kalman gains $K^{(0)}, K^{(1)}, \cdots, K^{(p)}$.

Emo Todorov (2005) considered this issue and suggested the following. Say that we compute a sequence of control gains $G^{(0)}, G^{(1)}, \cdots, G^{(p)}$, which are optimal for a sequence of *fixed* Kalman gains $K^{(0)}, K^{(1)}, \cdots, K^{(p)}$. These Kalman gains would not depend on the motor commands. Rather, $K^{(k)}$ will be computed so that for a given sequence of control gains, it will minimize the expected value of the future state $E\left[v_{\pi*}\left(\mathbf{x}^{(k+1)}, \hat{\mathbf{x}}^{(k+1)}\right)\right]$. Once a sequence of fixed Kalman gains had been computed, we then re-compute the sequence of control gains. He showed that this iterative approach was guaranteed to converge. Indeed, in practice the approach converges within a few iterations. The recipe for computing the Kalman gain is as follows. We begin at time step $k = 0$, set $S_e^{(0)}$ to be the prior uncertainty, and $S_x^{(0)} = \hat{\mathbf{x}}^{(0)}\hat{\mathbf{x}}^{(0)T}$, and compute the following sequence of Kalman gains:

$$\begin{aligned}
\hat{\mathbf{x}}^{(k+1)} &= A\hat{\mathbf{x}}^{(k)} + AK^{(k)}\left(\mathbf{y}^{(k)} - H\hat{\mathbf{x}}^{(k)}\right) - BG^{(k)}\hat{\mathbf{x}}^{(k)} \\
K^{(k)} &= S_e^{(k)}H^T\left(HS_e^{(k)}H^T + Q_y\right)^{-1} \\
S_e^{(k+1)} &= Q_x + \left(A - AK^{(k)}H\right)S_e^{(k)} + \sum\nolimits_i C_i G^{(k)} S_x^{(k)} G^{(k)T} C_i^T \\
S_x^{(k+1)} &= AK^{(k)}HS_e^{(k)}A^T + \left(A + BG^{(k)}\right)S_x^{(k)}\left(A + BG^{(k)}\right)^T
\end{aligned} \tag{12.39}$$

To summarize, one begins by computing a set of Kalman gains $K^{(0)}, K^{(1)}, \cdots, K^{(p)}$. These gains can be computed from Eqs. (12.37) and (12.38) with the signal dependent noise component set to zero. One then computes a set of control gains $G^{(0)}, G^{(1)}, \cdots, G^{(p)}$ using Eqs. (12.33) and (12.35). One then re-computes the Kalman gains using Eq. (12.39), and the re-computes the control gains. The Kalman and control gains converge within a few iterations.

12.6 Example: control of eye and head during head-free gaze changes

In many laboratory experiments on eye movements, the head of the subject is kept at rest and only the eyes are allowed to move. But in more natural, unconstrained conditions the head participates by redirecting the gaze. When we look around a room, searching for our keys, our eyes and head move in fairly complex and well coordinated patterns. As we shift the gaze from one point to another, the eyes tend to start the movement with a saccade (Fig. 12.5A). During the saccade, the head starts rotating. While the head is still moving, the saccade ends and the eyes roll back in the head. If the motion of the head is perturbed, the saccade is altered in mid-flight (Fig. 12.5B). For example, if a brake is applied to the head for 100ms at saccade onset, the saccade is longer in duration and amplitude as compared to when the head is free to rotate, implying that the motion of the eyes during the saccade is affected by the state of the head. Let us show that all of these behaviors are consistent with a very simple goal: keep the target on the fovea, and keep the eyes centered in the head. We will express this goal as a cost function, and then use optimal feedback control theory to produce movements that best achieve the goal, i.e., minimize the cost. We will perturb the motion of the simulated system and test whether it produces movements that resemble the recorded data.

Our model of the eye dynamics is identical to the one that we used in the previous chapter. We have a third order system where x_e represents eye position in the orbit, and f_e is the torque on the eye. We also assume that when $x_e = 0$ the eye is in the central neutral position in its orbit. The third order system has three states $x_1 \equiv x_e$, $x_2 \equiv \dot{x}_e$, and $x_3 \equiv f_e$. We have

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -\frac{k_e}{m_e} & -\frac{b_e}{m_e} & \frac{1}{m_e} \\ 0 & 0 & -\frac{\alpha_2}{\alpha_1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \frac{1}{\alpha_1} \end{bmatrix} u_e \qquad (12.40)$$

The parameters of the eye plant are set so that the resulting system has time constants of 224, 13, and 4 ms. So we set $k_e = 1$, $b_e = \tau_1 + \tau_2$, $m_e = \tau_1 \tau_2$, $\alpha_2 = 1$, and $\alpha_1 = 0.004$, where $\tau_1 = 0.224$ and $\tau_2 = 0.013$. We represent Eq. (12.40) as:

$$\dot{\mathbf{x}}_e = A_e \mathbf{x}_e + \mathbf{b}_e u_e \qquad (12.41)$$

Our head model is similar to the eye model, but with time constants of 270, 15, and 10 ms. The head position, x_h, is an angle that we measure with respect to a stationary frame in the environment. Therefore, the direction of gaze in the same stationary frame is $x_e + x_h$. The noise-free version of our dynamical system in continuous time is:

$$\begin{aligned} A_c &\equiv \begin{bmatrix} A_e & \mathbf{0} \\ \mathbf{0} & A_h \end{bmatrix} \\ \begin{bmatrix} \dot{\mathbf{x}}_e \\ \dot{\mathbf{x}}_h \end{bmatrix} &= A_c \begin{bmatrix} \mathbf{x}_e \\ \mathbf{x}_h \end{bmatrix} + \begin{bmatrix} \mathbf{b}_e & \mathbf{0} \\ \mathbf{0} & \mathbf{b}_h \end{bmatrix} \begin{bmatrix} u_e \\ u_h \end{bmatrix} \end{aligned} \tag{12.42}$$

Suppose that g represents the position of our target. We translate our continuous time model into discrete time with time step Δ as follows. Set the state of the system to be:

$$\mathbf{x} \equiv \begin{bmatrix} \mathbf{x}_e & \mathbf{x}_h & g \end{bmatrix}^T \tag{12.43}$$

In Eq. (12.43), g represents the goal location, described as an angle in the same coordinate system in which we measure position of our head. Define the following matrices using matrix exponentials:

$$\begin{aligned} A &\equiv \begin{bmatrix} \exp\left(A_c \Delta\right) & \mathbf{0}_{6\times1} \\ 0 & 1 \end{bmatrix} \\ B &\equiv \begin{bmatrix} \mathbf{b}_e & \mathbf{0}_{3\times1} \\ \mathbf{0}_{3\times1} & \mathbf{b}_h \\ 0 & 0 \end{bmatrix} \end{aligned} \tag{12.44}$$

The discrete version of our model is:

$$\begin{aligned} \mathbf{x}^{(k+1)} &= A\mathbf{x}^{(k)} + B\mathbf{u}^{(k)} + \boldsymbol{\varepsilon}_x^{(k)} + B\sum_i C_i \mathbf{u}^{(k)} \phi_i^{(k)} \\ \mathbf{y}^{(k)} &= H\mathbf{x}^{(k)} + \boldsymbol{\varepsilon}_y^{(k)} + H\sum_i D_i \mathbf{x}^{(k)} \mu_i^{(k)} \\ H &\equiv \begin{bmatrix} -1 & 0 & 0 & -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\ \boldsymbol{\varepsilon}_x &\,\square\, N\left(\mathbf{0}, Q_x\right) \quad \boldsymbol{\varepsilon}_y \,\square\, N\left(\mathbf{0}, Q_y\right) \\ \phi &\,\square\, N\left(0,1\right) \quad\quad \mu \,\square\, N\left(0,1\right) \end{aligned} \tag{12.45}$$

The *H* matrix in Eq. (12.45) implies that we can sense the position of the eye x_e, as well as the position of the target on the retina. The position of the target on the retina is the difference between the goal location and the sum of the eye and head positions: $g - \left(x_e + x_h\right)$. What we want is to place the target at the center of the fovea, so we assume a cost per step that penalizes the distance of the target to the fovea. We also want to keep the eyes centered in their orbit, and so we will penalize eccentricity of the eye. Finally, we want to minimize the motor commands to eye and head. Our cost per step is:

$$\begin{aligned} \alpha^{(k)} &= \mathbf{y}^{(k)T} T^{(k)} \mathbf{y}^{(k)} + \mathbf{u}^{(k)T} L \mathbf{u}^{(k)} \\ &= \mathbf{x}^{(k)T} H^T T^{(k)} H \mathbf{x}^{(k)} + \mathbf{u}^{(k)T} L \mathbf{u}^{(k)} \end{aligned} \tag{12.46}$$

Suppose that we want gaze position $x_e + x_h$ to arrive at target at time step k_1. The state cost matrix T is zero until time step k_1, and then is kept constant until end of simulation period p. The motor cost L is diagonal with equal costs for eye and head motor commands. We assumed non-zero signal dependent motor noise $c_1 = c_2 = 0.01$ (as in Eq. (12.13)), but zero signal-dependent sensory noise $d_i = 0$.

Fig. 12.8A shows the behavior of our model for a target at 40$^{\text{o}}$. The gaze arrives at target at around 100ms, as our cost function had implied, and this is accomplished through a cooperation of the eye and head systems. The gaze change begins with motion of the eye, making a 25$^{\text{o}}$ saccade, and is accompanied with motion of the head. The saccade amplitude is consistent with the data in Fig. 12.5C. Upon saccade completion the eyes roll back in the head as the head continues to move toward the target. The gaze change begins with motion of the eye because the dynamics of the eye present a lighter system to move, and therefore it costs less in terms of effort (squared motor commands) than moving the head. When the gaze is at the target, i.e., target is on the fovea, the eyes roll back. This is because we incur a cost for not having the eyes centered with respect to the head. These two costs, gaze on target, eyes centered, are sufficient to produce the coordinated motion of a natural gaze change.

Now let us consider a situation in which head motion is perturbed during the gaze change. To simulate this, we prevented the head from moving for 50, 100, or 200ms at start of the gaze change (Fig. 12.8B). We find that the saccade made by the eye is lengthened so that the gaze still arrives near the target at around 100ms, but that the eyes no longer roll back unless the head is allowed to move toward the target. The increased saccade amplitude is consistent with the data in Fig. 12.5C, in which we see that in the head-braked condition saccade amplitudes are increased. In the experimental data, the head-brake (Fig. 12.5B) also prevents the roll back of the eyes after saccade completion, consistent with our simulation results.

12.7 Limitations

The real advantage of using a cost function to describe the goal of a movement is that it removes the burden of having to describe the desired kinematics of the task in some *a priori* format. That is, we do not need to describe how we wish for the eyes and the head to move in order to produce a gaze change. We do not have a desired trajectory that specifies a sequence of positions and velocities that we wish for the eyes and the head to achieve. Our control system does not have feedback gains that are arbitrary attractors around this desired trajectory. Rather, we have a fairly simple goal (put the target on the fovea, keep the eye centered), and then we leave it up to the optimization process to find the controller that achieves this goal as well as it can be done.

However, there are potential problems with our simulations and the general framework. Let us highlight some of the limitations here.

In our present simulations we assumed a desired time at which gaze should arrive on target. Where does this desired time, i.e., desired movement duration, come from? One possibility is that movement duration carries a cost because passage of time discounts rewards of the task. We used this idea in the previous chapter to show that head-fixed saccade durations reflect an optimization process in which passage of time discounts reward. A recent work showed that the duration of gaze changes in head-free movements can also be explained by a cost function like that of Eq. (12.9) in which movement durations carry a cost that has a hyperbolic form (Shadmehr et al., 2010). So in principle, a cost that involves three components (keep the target on the fovea, keep the eye centered, and get to the target as soon as possible) can account for the kinematics and timing of head-fixed and head-free gaze changes.

Following this line of thinking implies that before a movement begins, our brain considers all possible movement durations, and given the current cost and reward conditions per step, arrives at a control policy that carries a minimum total expected cost. Putting aside the fact that we have no idea how such an optimization process might biologically take place, our framework arrives at a specific movement duration for which we implement a control policy. However, if our movement is perturbed, does the perturbation alter our desired movement duration? It certainly seems like it should. For example, say that during a movement a perturbation alters the position of the target. Our control policy has no trouble with this and will respond to the altered goal position. However, it will do so approximately in the same desired movement period as before. This is inconsistent with experimental data (Liu and Todorov, 2007). People allow more time to correct for a perturbation than is expected in our finite-horizon formulation of the problem. This hints that the framework of finite-horizon optimization is probably inappropriate and will need to be abandoned for a more general approach.

A second, more troubling problem is with the general theme of optimality devoid of evolutionary history. Our framework has proceeded with the assumption that biological behavior can be understood as minimization of a cost function in which we usually do not consider the constraints that are imposed by the history of the animal. To illustrate the problem, consider the task of trying to explain swimming behavior in seagoing mammals. Marine mammals swim by beating their horizontal tail flukes up and down. However, fishes swim by moving their tail flukes from side to side. How could there be two seemingly orthogonal but still ‘optimum’ ways of swimming in water? In considering this question, Stephen J. Gould (1995) writes: “the explanation probably resides in happenstances of history, rather than in abstract predictions based on universal optimality.” In particular, marine mammals evolved from animals that used to run on land. Locomotion of these animals involved flexing their spinal column up and down. He writes:

> Thus, horizontal tail flukes may evolve in fully marine mammals because inherited spinal flexibility for movement up and down (rather than side to side) directed this pathway from a terrestrial past. The horizontal tail fluke evolved because whales carried their terrestrial system of spinal motion to the water.

Clearly, the mathematics can be expanded to represent constraints that are imposed by neuroanatomy and biomechanics. For example, we can incorporate models of local controllers as may be present in the spinal cord, and more detailed models of muscles and biomechanics. What is unclear, however, is whether our cost per step (Eq. 12.9), or our description of the system’s dynamics (Eq. 12.45), need to incorporate some measure of the history of the species.

The brain finds a better way to clear a barrier

We have been spending the entirety of this book examining relatively simple movements like saccades and reaching, movements that probably have not changed much in the past few million years of evolution. During the last century, however, the human brain has made fundamental breakthroughs in finding better ways to clear a barrier. The new ways of performing this task are real-world examples of better solutions to an optimal control problem.

In the high-jump you are asked to clear the highest hurdle possible. The task has been a part of track and field competition since the advent of such competitions in the modern Olympics. The earliest techniques were to hurdle over the bar: approach at an angle, scissoring the legs over the bar and land on the feet (Fig. 12.9A). In the late part of the 19th century the technique was improved by taking off like the scissor but extending the back and flattening out over the bar. By mid 20th century, the most successful approach was a straddle technique called the “western roll” or the “straddle”, in which one begins by running toward the bar at a diagonal, and then kicks the outer leg over the bar and crosses the bar face down (Fig. 12.9B). In all these cases the feet cross the bar first, and there is at least a theoretical possibility of landing on your feet. Starting in the early 20th century all competitions included a landing area made of sand, sawdust, or woodchips. In the late 1960s, however, there was a revolutionary change in the way athletes jumped over the bar, and the change occurred because of the brain of one kid, Richard Fosbury, a high school sophomore in Medford, Oregon. [Our source for this story is an article by Richard Hoffer, Sports Illustrated, Sept. 14, 2009.]

Fosbury was a tall kid who in eight and ninth grade was still using the scissor jump to clear the bar. When he entered high school in the tenth grade, his coach insisted that he try the more modern western roll. However, Fosbury had trouble adapting his style. Rather than improving on his previous records, at his first meet as a sophomore he did not clear the opening height and failed on all three chances. The repeated failures continued that year until something amazing happened in a meet of a dozen schools near the end of the sophomore year in 1963 at Grants Pass, Oregon. Feeling desperate, Fosbury’s coach had given him permission to revert back to the scissor technique. On his first jump that day he cleared 5’4”, the height that he had achieved a year earlier. Richard Hoffer writes:

> The other jumpers were still warming up, waiting for the bar to be set at an age-appropriate height, while Fosbury continued to noodle around at his junior high elevations. If they, or anyone else, had been interested though, they might have seen an odd transformation taking place. Fosbury was now arching backward ever so slightly as he scissored the bar, his rear end now coming up, his shoulder going down. He cleared 5’6”. He didn’t even know what he was doing, his body reacting to desperation. His third round, his body reclined even more, and he made 5’8”. On his fourth attempt Fosbury took a surprisingly leisurely approach to the bar and … he was completely flat on his back now, cleared 5’10”. The high jump was an event that measured advancement by fractions of an inch, sometimes over a year. Fosbury, conducting his own defiance, had just improved a half foot in one day.

Fosbury was crossing the bar head first, and landing on his neck and back on the sawdust and woodchips (Fig. 12.9C). Fortunately, in his junior year the school replaced the pit with foam and he reached 6’3” by the end of that year and 6’5.5” by the end of his senior year. After graduation, he enrolled in Oregon State University in the engineering program. By his second year, he had improved his record to 6’10”. He was still not a world class jumper, but he would become one by his third year, consistently crossing the 7 foot barrier. In the 1968 Olympics in Mexico City he was the only jumper to go over the bar head first. He won by crossing 7’4.25”, an Olympic record.

In looking back at the change that Fosbury brought, we see two important ideas. First, the scissor technique had evolved into the western roll and the straddle, techniques in which you cross the bar face down. These are control policies that are now viewed as sub-optimal in the sense that they are local minima. Fosbury, starting from the scissor evolved it into a jump in which you cross the bar facing the sky. By effectively re-starting the search process for a better control policy, he was able to stumble upon a better solution. Second, his solution was made possible because there was a critical change in the cost function: kids were no longer jumping into piles of sawdust and woodchips, but had the luxury of landing on foam. The margin of safety and comfort was significantly increased if you happened to land on your neck. As a result of a healthy disrespect for history, and being present when technology altered the cost of the task, an engineering student found a fundamentally new solution to an old optimal control problem, and took home the Olympic gold for his genius.

Summary

We make movements in order to improve the state of our body and the environment. To make the movement, we must spent effort. We can express the purpose of the movement as obtaining a more rewarding state, while spending as little effort as possible. Mathematically this is formulated as a cost function in which the state at the end of the movement is afforded a value (e.g., the closer we end up to the rewarding state, the better), discounted by the accumulated effort. Together this describes a cost function. The best movement is one in which the motor commands achieve the least cost. To produce these motor commands, we do not wish to pre-program them at movement start, because movements can be perturbed and we wish to be able to respond to sensory feedback. Rather, we wish to form a policy that produces motor commands based on our current belief about our state. This belief comes from our ability to make predictions about sensory consequences of our motor commands, and combine them to the actual sensory feedback. An optimal feedback controller describes a feedback dependent policy that produces motor commands that minimize a cost regardless of the state that we may find ourselves during the movement.

In this chapter we used Bellman's optimality principle to compute the policy that minimized a quadratic state and effort cost for a linear system with signal dependent motor and sensory noise. Our work followed the approach described by Todorov (2005). We applied the result to control of head-free gaze changes in which the eyes and the head cooperate to keep a target stimulus on the fovea. The motor commands to these two systems respond to sensory feedback, and we simulated conditions in which the head was perturbed, demonstrating how it affects the ongoing saccade of the eye.

Figure Legends

Figure 12.1. A general framework for goal directed behavior. Motor commands are generated in order to achieve a goal. The goal is expressed as a cost function. The motor command generator is a feedback controller that produces motor commands as a function of the current belief regarding the state of the body and the environment. This feedback controller is optimum in the sense that it produces motor commands that achieve the goal, i.e., minimize the cost function. The state of the body and environment comes from a combination of sensory measurements (via sensory feedback) and predictions about the outcome of motor commands (via forward models).

Figure 12.2. An eye blink that takes place during a saccadic eye movement disturbs the trajectory of the eyes, yet the eyes accurately arrive at the target. (From (Rottach et al., 1998)).

Figure 12.3. TMS applied to anywhere on the head inhibits the ongoing oculomotor commands, but the saccade is corrected mid-flight with subsequent motor commands that bring the eye to the target. **A**. Experimental setup and example of TMS perturbed saccades. The arrow near saccade onset marks the time of a single TMS pulse. **B**. TMS applied at various times after saccade onset inhibit the ongoing motor commands at a latency of around 60ms, as reflected in the perturbation to saccade velocity.

Figure 12.4. Volunteers were shown a target on the horizontal meridian at (15,0) deg. Upon saccade initiation, the target was jumped vertically to (15,5) deg. The resulting adaptation produced curved saccades, marked here with ‘target jump’. In contrast, control saccades to targets that do not jump are generally straight. (From (Chen-Harris et al., 2008)).

Figure 12.5. The coordinated response of the eye and head to visual targets. **A.** Eye and head motion under head unrestrained condition. Volunteers were shown a target on the horizontal meridian, with amplitude marked at start of the traces. Head position is measured with respect to straight ahead. Eye position reflects the position of the globe with respect to a centered location in the head. Gaze is head position plus eye position. **B.** Eye and head motion under head restrained condition. On random trials, a brake prevented motion of the head. The period of the break is indicated by the heavy black line. Motion of the eyes depends on the state of the head. **C**. Amplitude of the eye saccade as a function of target amplitude in head free and head-braked condition. In the head-braked condition, the head was restrained for 100ms at start of the gaze

shift. Eye saccades are generally larger when head movement is prevented. (From (Guitton and Volle, 1987)).

Figure 12.6. Response to a perturbation depends on the cost function. **A.** In the two-cursor condition, each hand controlled a cursor and the goal was to place the cursors at the two targets. In the one-cursor condition, the cursor position represented the average position of the left and right hands and the goal was to place the cursor at the single target. Perturbations were applied to the left arm only. B. Gray dots indicate left and right hand (LH, RH) positions in the condition in which the left hand was perturbed. The black dots reflect the un-perturbed condition. In the two-cursor condition, the perturbation to the left hand produces no response in the right hand. In the one-cursor condition, the perturbation to the left hand produces a small response in the right hand. **C**. Velocity of the left hand in the perturbed (gray line) and un-perturbed conditions. **D**. Velocity of the right hand. In the two-cursor condition, right hand velocity is nearly identical whether or not the left hand was perturbed. In the one-cursor condition, right hand velocity shows a corrective response when the left hand is perturbed. (From (Diedrichsen, 2007))

Figure 12.7. Schematic description of $\min f(u)$ and $\arg\min f(u)$.

Figure 12.8. Simulation results for a head-free eye/head movement with target at 40^{o}. **A**. An unperturbed movement. **B**. Perturbed movements. The head was fixed and not allowed to move for 50, 100, and 200ms. Parameter values: The state cost matrix T is zero until time step 110ms, and then is kept constant at $T = \left[10^5 \quad 0; \quad 0 \quad 300\right]$. Motor signal dependent noise standard deviation is $c_1 = c_2 = 0.01$. Motor costs are $L = \left[10 \quad 0; \quad 0 \quad 10\right]$. State noise is $Q_x = 0.1 \times I_7$ where the term I_7 represents an identity matrix of size 7×7. Observation noise is $Q_y = I_2$.

Figure 12.9. Techniques used in the high jump competition. **A**. Scissor technique, as illustrated by Platt Adams during the 1912 Summer Olympics. **B.** Straddle techniques, as illustrated by Rolf Beilschmidt in 1977. **C.** Fosbury technique. (Images from Wikimedia Commons)

Reference List

Bellman RE (1957) Dynamic Programming. Princeton, NJ: Princeton University Press.

Bizzi E, Kalil RE, Tagliasco V (1971) Eye-head coordination in monkeys: evidence for centrally patterned organization. Science 173:452-454.

Chen-Harris H, Joiner WM, Ethier V, Zee DS, Shadmehr R (2008) Adaptive control of saccades via internal feedback. J Neurosci 28:2804-2813.

Diedrichsen J (2007) Optimal task-dependent changes of bimanual feedback control and adaptation. Curr Biol 17:1675-1679.

Goossens HH, Van Opstal AJ (2000) Blink-perturbed saccades in monkey. I. Behavioral analysis. J Neurophysiol 83:3411-3429.

Gould SJ (1995) Dinosaur in a haystack: reflections in natural history. Harmony Books.

Guitton D, Volle M (1987) Gaze control in humans: eye-head coordination during orienting movements to targets within and beyond the oculomotor range. J Neurophysiol 58:427-459.

Guthrie BL, Porter JD, Sparks DL (1983) Corollary discharge provides accurate eye position information to the oculomotor system. Science 221:1193-1195.

Keller EL, Robinson DA (1971) Absence of a stretch reflex in extraocular muscles of the monkey. J Neurophysiol 34:908-919.

Liu D, Todorov E (2007) Evidence for the flexible sensorimotor strategies predicted by optimal feedback control. J Neurosci 27:9354-9368.

O'Sullivan I, Burdet E, Diedrichsen J (2009) Dissociating variability and effort as determinants of coordination. PLoS Comput Biol 5:e1000345.

Rottach KG, Das VE, Wohlgemuth W, Zivotofsky AZ, Leigh RJ (1998) Properties of horizontal saccades accompanied by blinks. J Neurophysiol 79:2895-2902.

Shadmehr R, Orban de Xivry JJ, Xu-Wilson M, Shih TY (2010) Temporal discounting of reward and the cost of time in motor control. J Neurosci 30:10507-10516.

Sherrington C (1923) The integrative action of the nervous system. Cambridge University Press.

Todorov E (2005) Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system. Neural Comput 17:1084-1108.

Xu-Wilson M, Tian J, Shadmehr R, Zee DS (2011) TMS induced startle perturbs saccade trajectories and unmasks the internal feedback controller. J Neurosci in press.