UNIVERSITY OF CALIFORNIA, SAN DIEGO

Rethinking Optimal Control of Human Movements

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in
Physics

by

Dongsung Huh

Committee in charge:

Professor Terrence J. Sejnowski, Chair
Professor Tatyana O. Sharpee, Co-chair
Professor Garrison W. Cottrell
Professor David Kleinfeld
Professor Herbert Levine
Professor Emanuel Todorov

2012

The dissertation of Dongsung Huh is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-chair

Chair

University of California, San Diego

2012

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank all of my committee members, in particular Terry Sejnowski, for his support, guidance, tolerance, and patience throughout my graduate career, and Emo Todorov, for encouraging me to ground my thinking in consideration of optimality. In addition, I thank Howard Hughes Medical Institute for their support during my tenure as a graduate student. This work would not have been possible without the extensive help of the staff of the Salk Institute, including the Machine Shop for construction of equipment, and the CNL computer support personnel. Finally I thank my mother, Jangwol Choi, without whose dedication and wisdom I would not have been possible.

Chapter 2, in full, is currently being prepared for submission for publication of thematerial. Huh, Dongsung; Sejnowski, Terrence. The dissertation author was the primaryinvestigator and author of this material.

Chapter 3, in full is currently being prepared for submission for publication of thematerial. Huh, Dongsung; Todorov, Emanuel; Sejnowski, Terrence. The dissertation author was the primary investigator and author of this material.

Chapter 4, in full, is a reprint of the material as it appears in International Federation of Automatic Control - World Congress, 2011. Huh, Dongsung; Sejnowski, Terrence. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in full, is an edited reprint of the material as it appears in IEEE Adaptive Dynamic Programming and Reinforcement Learning, 2009. Huh, Dongsung; Todorov, Emanuel. The dissertation author was the primary investigator and author of this paper.

VITA

2000–2004	T.J. Park Scholarship
2004	B. S. in Physics, Massachusetts Institute of Technology
2004	B. S. in Brain and Cognitive Sciences, Massachusetts Institute of Technology
2004–2006	Graduate Research Fellowship Program, National Science Foundation
2006–2007	Visiting scholar, Harvard University
2010	M. S. in Physics, University of California, San Diego
2012	Ph. D. in Physics, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Rethinking Optimal Control of Human Movements

by

Dongsung Huh
Doctor of Philosophy in Physics
University of California, San Diego, 2012
Professor Terrence J. Sejnowski, Chair
Professor Tatyana O. Sharpee, Co-chair

The complex bio-mechanics of human body is capable of generating an unlimited repertoire of movements, which on one hand yields highly versatile motor behavior but on the other hand presents a formidable control problem for the brain. Understanding the computational process that allows us to easily perform various motor tasks with a high degree of coordination is of central interest to both neuroscience and robotics control. In recent decades, it became widely accepted that the observed movement features can be understood as a result of the brain's effort of optimizing the motor behaviors. However, a major challenge for the optimal control models is deciding how the motor task should be represented.

In this thesis, (1) I compare the existing representations in literature, (2) derive a novel formulation using the principles and analytic tools from physics, (3) predict previously unreported features of human movements, and (4) confirm them in psychophysics experiments. I also explore implementing the optimal controller with artificial neural networks, which may potentially lead to a deeper understanding of neural computation. The obtained controller generates realistic target-directed hand movements in real-time, and reproduces many of reported neuro-physiology data in primary motor cortex.

Chapter 1

Introduction

1.1 Background

Biological organisms generate complex motor behaviors whose performance is unmatched by any man-made robots in its versatility and robustness. The ultimate goal of motor control field is to understand the computational principle of the brain's motor system that underlies this challenging task.

In most cases, the motor goal one faces is highly under-constrained — a desired behavioral goal can typically be accomplished by infinitely many different movements. However, the movements we actually generate tend to form a much smaller subset and exhibit highly repeatable regularities [Bernstein 1967]. Some of the well known regularities of human movements are: straight-paths and bell-shaped speed profiles of reaching movements [Abend, Bizzi, Morasso, et al. 1982]; the relation between speed, distance and accuracy of reaching movements (Fitts's law) [Fitts 1954]; the relation between speed and curvature of curve-drawing movements [Lacquaniti, Terzuolo, and Viviani 1983].

These observations have inspired a number of theoretical hypothesis for the underlying structure of the brain's motor system. In early modeling works, the popular concept was that the motor system chooses rather simplified control schemes that are "easy to execute", which led to the idea of equilibrium control hypothesis, muscle synergies, motor primitives, etc., and the observed regularities were interpreted as the evidence of the underlying simplicity [Merton 1953; Bernstein 1967; Morasso and Mussa Ivaldi 1982; Viviani and Flash 1995].

However, it became clear that such models were too restricted to fully embrace the flexibility in biological movements, as observed in perturbation and adaptation experiments. More recently, a different type of models have been proposed based on optimality principles [Flash and Hogan 1985; Uno, Kawato, and Suzuki 1989; Harris and Wolpert 1998a; Todorov and Jordan 2002]. In these models, the motor system is assumed to solve the control problem by optimizing a behaviorally relevant cost function under the constraint given by the plant and the motor goal at hand. According to a more generalized notion which incorporates feedback correction, the optimality principle can explain both the stereotyped behavior as well as the flexible adaptation to perturbations [Todorov 2004].

The optimality-based approach presents two potential challenges: (1) identifying the behaviorally relevant cost function, and (2) the appropriate representation of the motor goal. Most of the effort in computational models have focused on the first issue, which produced a number of cost functions, including jerk, torque-change, muscle-discharge, etc., [Flash and Hogan 1985; Uno, Kawato, and Suzuki 1989; Stein, Oguztoreli, and Capaday 1986; Pandy and Garner 1995]. Qualitatively, however, these costs can be summarized as penalizing abrupt changes in movements, and therefore they effectively promote smooth and economic characteristics of movement trajectories, differing in the details of the model body dynamics.

In contrast, the issue of correctly representing the motor goals, such as where to reach to, and how fast to do so, has not been investigated in a principled manner yet. The importance of specifying such information is apparent, since without any motor goals in mind, the most economic strategy is not to move at all!

1.2 Organization of the Thesis

In this thesis, I show that choosing how to represent the motor goal fundamentally affects prediction power of the model and provide a new perspective for understanding the computational process.

In the second chapter, I investigate the issue of how fast we move. Most movements we generate can be executed with different degree of speed. In most theoretical models, such information is assumed to be represented in terms of the duration of the movement, which has two serious drawbacks. First, it implies that the entire movement trajectory must be preplanned until the very end — which is when the duration can be defined — which is highly unlikely especially if the movements are long and complex. Moreover, such models do not make any prediction regarding how fast we move. The generated movement will have whatever duration it was told to be. Here, I propose a new, dual formulation of the original optimal control model that directly represents how fast we are willing to move, which I call *drive*. At a constant level of drive, the model predicts that the overall-speed of movements will generally increase with the 2/3 power of the spatial size of the movement. I have confirmed the prediction in motor psychophysics experiments where subjects made various types of arm movements (e.g. straight, circle drawing). The constant level of drive also reproduces the time-varying speed-profile shape, such as the bell-shaped speed profile of straight reaching movements. The analysis also predicts that for spiral movements the speed-profile would scale with local radius of curvature by the 2/3 power, which we also confirmed experimentally. Drive level is likely to be represented In the brain as the tonic dopamine level of neurons in the brainstem, particularly those that projects to regions that control movements. Patients with disease in these areas exhibit abnormal movement speed — Parkinson's disease is associated with very slow movements, while Tourette's syndrome and Huntington's disease with very fast movement.

In the third chapter, I investigate representation of spatial goals for straight reaching movements. Earlier models included constraints that specify both where and when movement trajectories should terminate. The state-constraints make the models incompatible in realistic settings since the stochasticity in sensory-motor dynamics and the environment makes it impossible for the constraint to be satisfied. The later extended models instead specify the target state by an extra cost function that penalizes deviation between the desired state and the actual state at the terminal time. This relaxed representation allowed incorporating feedback correction mechanism in stochastic settings. However, this model has a number of problems arising from the fact that the timing of reaching is constrained. Here, I propose a novel formulation that contains neither the time nor the state constraints: Instead, movements are assumed to happen without any time-limit, and they are subjected to a state-cost function that continuously penalizes states that are away from the target-state. The new, infinite-horizon model has a number of important features. First, the cost function must have a saturating shape: it is quadratic near the target, but saturates at a constant value when far away. A similar cost function has been measured in a recent psychophysics experiment. Secondly, this is the first model that allows integration of stochastic feedback control as well as duration prediction, which is necessary for explaining one of the well known phenomenon — speed-accuracy trade-off known as Fitts's law. Thirdly, the model unifies many of the existing models in a unified framework: Each model is just a special limiting case. Moreover, this formulation is fully compatible with the notion of drive from the previous chapter: drive is represented as the relative weight between the state-cost and the control-cost.

In the fourth chapter, I further investigate infinite-horizon optimal control under multiplicative noise. Such noise model has been found relevant for a number of physical systems, including neuromuscular system — larger muscle tension leads to larger level of noise — which was used in the dynamic model of chapter 3. The analysis reveals a control law that is very different from the conventional control scheme.

In the fifth chapter, I investigate neuro-implementation of optimal controllers. Most theoretical approaches can only characterize motor control on the algorithmic/representational level [David 1982], which is sufficient for descriptions on the behavioral or muscle level, but not on the neural level. Instead, the optimal controller is merely represented as an abstract sensori-motor mapping. Here, we implement an optimal controller with a recurrent neural network, which is connected to a realistic limb dynamics model in a sensory-motor loop. The connection weight matrix is trained for performing target-directed reaching movements using the (modified version of) infinite-horizon optimal control formulation described in chapter 3. The trained controller can generate realistic reaching movements in real-time without requiring a separate planning stage., reproduces many of reported neuro-physiology data in primary motor cortex (not shown here), and therefore it may potentially lead to deeper understanding of the neural computation. Although the resulting neural-controller is far from the complexity and flexibility of the motor system in the brain, this success is a positive step toward it.

References

Abend, W., E. Bizzi, P. Morasso, et al. (1982). "Human arm trajectory formation." In: *Brain: a journal of neurology* 105.Pt 2, p. 331.

Bernstein, N.A. (1967). *The coordination and regulation of movements*.

David, M. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. WH Freeman.

Fitts, P.M. (1954). "The information capacity of the human motor system in controlling the amplitude of movement". In: *Journal of experimental psychology* 47.6, pp. 381–391.

Flash, T. and N. Hogan (1985). "The coordination of arm movements: an experimentally confirmed mathematical model". In: *Journal of neuroscience* 5.7, p. 1688.

Harris, C.M. and D.M. Wolpert (1998a). "Signal-dependent noise determines motor planning". In: *Nature* 394.6695, pp. 780–784.

Lacquaniti, F., C. Terzuolo, and P. Viviani (1983). "The law relating the kinematic and figural aspects of drawing movements". In: *Acta psychologica* 54.1-3, pp. 115–130.

Merton, PA (1953). "Speculations on the servo-control of movement". In: *The Spinal cord: a Ciba Foundation symposium*. Little, Brown, p. 247.

Morasso, P. and FA Mussa Ivaldi (1982). “Trajectory formation and handwriting: a computational model”. In: *Biological Cybernetics* 45.2, pp. 131–142.

Pandy, MG and BA Garner (1995). “Optimal Control ot Non-ballistic Muscular Movements: A Constraint-Based Performance Criterion for Rising From a Chair”. In: *Journal of biomechanical engineering* 117, p. 15.

Stein, RB, MN Oguztoreli, and C. Capaday (1986). “What is optimized in muscular movements”. In: *Human muscle power*, pp. 131–150.

Todorov, E. (2004). “Optimality principles in sensorimotor control (review)”. In: *Nature neuroscience* 7.9, p. 907.

Todorov, E. and M.I. Jordan (2002). “Optimal feedback control as a theory of motor coordination”. In: *Nature neuroscience* 5.11, pp. 1226–1235.

Uno, Y., M. Kawato, and R. Suzuki (1989). “Formation and control of optimal trajectory in human multijoint arm movement”. In: *Biological cybernetics* 61.2, pp. 89–101.

Viviani, P. and T. Flash (1995). “Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning.” In: *Journal of Experimental Psychology: Human Perception and Performance* 21.1, p. 32.

Chapter 2

Conservation of Drive Motivates a New 1/3 Power-Law for Human Arm Movements

Abstract When asked to reach for an object, we can do so over a wide range of speeds, depending on our degree of motivation. We have derived a novel mathematical framework for a quantity we call drive, which represents the internal motivation level for how fast a person is willing to move. This theory predicts that humans tend to move with a constant level of drive, based on an optimality principle, which in turn predicts a specific, 2/3 power-law relationship between the spatial scale and movement speed. This prediction was confirmed in psychophysics experiments where subjects made unconstrained, free hand movements in which the overall-speed scaled with movement size according to the predicted power law. Independent sources of evidence suggests that the biological basis for drive may be the tonic dopamine level in the brainstem. The drive conservation principle provides a new viewpoint for understanding motor control in the brain.

2.1 Introduction

Biological organisms exhibit highly stereotyped regularities in their movement repertoire [Bernstein 1967]. The well known examples include: straight-paths and bell-shaped speed profiles of reaching movements [Abend, Bizzi, Morasso, et al. 1982]; the relation between speed, distance and accuracy of reaching movements (Fitts's law) [Fitts 1954]; the relation between speed and curvature of curve-drawing movements [Lacquaniti, Terzuolo, and Viviani 1983]; and more.

While some of the behavioral characteristics may be partially attributed to the musculo-skeletal mechanics of the body [Gribble and Ostry 1996], a growing body of evidence suggests that they in fact may be a result of the brain's effort of optimizing the motor behaviors [Todorov 2004]. Indeed, the optimality principle based approach have explained a wide range of observed motor phenomena in a unifying theoretical framework.

Despite the glorious success, however, optimal control models have failed to characterize one of the most prominent features — the overall-speed and duration of movements. For most of the models, movement duration is instead something that needs to be pre-specified by the modeler. Although several amended models have been proposed, they generally introduce some hypothetical criteria to determine movement duration, whose validity is yet to be confirmed. (e.g. constraining the total cost, introducing additional time-costs, or assuming the desired accuracy level) [Guigon, Baraduc, and Desmurget 2008; Hoff 1994; Harris and Wolpert 1998a].

In this paper, we show that duration and overall-speed of movements can be predicted from the first principle. Moreover, our result presents a very different view point to understand the problem.

2.2 Drive and Duration

2.2.1 Optimal control

Optimal control models describe human movements as the optimal solutions that minimize the total cost of movement,

$$\mathcal{A}(T, \cdots) \equiv \min \int_0^T \mathcal{L}\, dt, \tag{2.1}$$

where the minimization process considers all potential movement trajectories that can be generated by the body dynamics, and the instantaneous cost $\mathcal{L}$ generally penalizes abrupt changes of the trajectories [Flash and Hogan 1985; Uno, Kawato, and Suzuki 1989; Todorov and Jordan 2002]. As a result, the model reproduces smooth and economical trajectories as in observed human movements. The left-hand-side of eq (2.1), the *minimum total-cost*, is of central interest here and will be referred to as *action.*[1]

What is often not explicitly mentioned in such a brief description, however, is that there are certain motor behavioral characteristics that fundamentally cannot be predicted by the optimality principle itself, such as the distance of reaching movements or the path shape of curve-drawing movements, etc.. Instead, such information must be pre-specified for the model. In other words, they appear as independent variables for the action function. (Here, they are suppressed for notational simplicity.) Note that in the absence of such specifications the most economic movement would be a trivial one — not moving at all.

In previous models, the duration of movements, T, has been also regarded as one of the independent variables. Indeed, if the duration were not specified, the optimal solution of eq (2.1) would predict a movement of infinite duration (zero-speed).[2] However, there are a number of implausibilities in treating duration as an independent variable. As we present later, movement duration (and overall-speed) exhibits

[1]The terminology is borrowed from Lagrangian mechanics, which uses a closely related formulation.

[2]This is because action is general a monotonically decreasing function of duration T, without a well defined minimum (see discussion).

high correlation with other independent variables (also see [Fitts 1954; Gordon et al. 1994; Lacquaniti, Terzuolo, and Viviani 1983; Viviani and McCollum 1983; Viviani and Schneider 1991]). Moreover, in many natural movements, such as in doodling, the subjects do not plan out the entire future movement, and therefore, the movement duration cannot be pre-determined.

In fact, it is quite easy to understand why the traditional formulation fails to predict movement duration — eq (2.1) optimizes only one instance of movement, and therefore it can only characterize *within-movement-regularities* (e.g. how speed-profile fluctuates within a movement). On the other hand, the regularity concerning duration and overall-speed is a *cross-movement relationship*: *i.e.* how speed of one movement compares with another. Therefore, a natural extension of eq (2.1) would be to consider multiple movements as a single optimization problem[3]

$$\begin{aligned} \mathcal{A}_{\text{all}}(T_{\text{all}}, \cdots) &= \min \left[\sum_{\substack{\text{multiple} \\ \text{movements}}} \int \mathcal{L}\, dt \right] \\ &= \min_{T_1, \cdots, T_N} \left[\sum_{i=1}^{N} \mathcal{A}_i(T_i, \cdots) \right], \end{aligned} \tag{2.2}$$

where the minimization process considers all possible N movements, which may or not be contiguous in time, and each movement has its own set of independent variables. The second equality results from eq (2.1), where $\mathcal{A}_i$ are the individual actions. Note that in this extended formulation, duration of individual movements T_i are determined from the optimization process. However, the formulation requires the total duration ($T_{all} = \sum_{i=1}^{N} T_i$) to be pre-specified, *i.e.* as an independent variable, since otherwise eq (2.2) would predict all movements being infinitely slow. Therefore, the duration prediction problem is now reduced to an optimal allocation problem of limited total resource.

The solution to this optimization problem is a conservation relation (see

[3] To be more precise, eq (2.2) considers *equally-paced movements*. But the definition of pace is fuzzy.

Appendix 1),

$$\mathcal{D}_1 = \cdots = \mathcal{D}_N \quad (= \mathcal{D}_{\text{all}}), \tag{2.3}$$

where the conserved quantity is the (rate of change/derivative) of action with respect to duration (while all other independent variables are fixed),

$$\mathcal{D}_i(T_i, \cdots) \equiv -\frac{\partial \mathcal{A}_i}{\partial T_i}, \tag{2.4}$$

[4] which we call the *drive*. In other words, the movements are related to one another by sharing the same level of drive. Note that if the relation between duration and drive (eq (2.4)) is monotonic (*i.e.* one-to-one mapping), then the conservation relation eq (2.3) uniquely determines the duration of individual movements according to the given total duration. This turns out to be true — in general, smaller drive level is associated with longer duration (i.e. slower movements) — as we show later.

2.2.2 Duality between Drive and Duration

While the extended formulation eq (2.2) provides some crucial insights and may accurately characterize the observed regularity of movement duration and overall-speed, however, it also presents a number of implausibilities: First, it is unclear how many future movements N need to be considered in the formulation. Moreover, the formulation requires the total duration to be pre-specified. Therefore, the above approach requires preplanning of an indefinite number of future movements even for generating just one movement at the moment. The brain would have to perform preplanning for an arbitrarily large N. The cause of this limitation fundamentally tracks down to the *extensive* (or *additive*) nature of duration — duration of each movement adds up to compose the total duration. As a result, the total duration can be defined only at the end of the last movement, which leads to the pre-planning problem in the end.

---

[4] Similarly, the overall drive-level is $\mathcal{D}_{\text{all}} = -\partial\mathcal{A}_{\text{all}}/\partial T_{\text{all}}$,

Drive, on the contrary, is an *intensive* variable — the overall drive-level is identical to those of individual movements (eq (2.3)).[5] Therefore, the preplanning problem can be resolved by reformulating the control problem such that the drive becomes an independent variable in place of duration. Such process is formally described by *Legendre transform*, which defines the *dual action* and duration of individual movements as

$$\begin{aligned}
\mathcal{A}_i^*(\mathcal{D}, \cdots) &\equiv \min_T \left[\mathcal{A}_i(T, \cdots) + \mathcal{D} \cdot T \right] \\
T_i(\mathcal{D}, \cdots) &\equiv \frac{\partial \mathcal{A}_i^*}{\partial \mathcal{D}},
\end{aligned} \tag{2.5}$$

[6] where the subscript for drive is now dropped since its the same. The relationship between drive and duration described by the new Legendre transformed formulation is numerically identical to the previous formulation eq (2.4); only their dependence relationship is inverted. Note that the duration of individual movements gets uniquely determined by equation (2.5) given the drive-level, and importantly separately from other (future) movements, i.e. no pre-planning required (See Appendix).[7][8]

[5] Other examples of extensive-intensive pair include thermal-energy vs temperature (thermodynamic systems), volume vs pressure (gaseous systems), and charge vs voltage (capacitors in parallel), etc..

[6] Legendre transformation is well defined if action is a convex function of duration. Once again, the other motor-task variables are suppressed for notational simplicity.

[7] This property is reflected on the fact that the dual-action of multiple movements is simply a sum of the individual dual-actions, i.e. without involving any further optimization process, since

$$\begin{aligned}
\mathcal{A}_{\text{all}}^*(\mathcal{D}, \cdots) &\equiv \min_{T_{\text{all}}} \left[\mathcal{A}_{\text{all}}(T_{\text{all}}, \cdots) + \mathcal{D} \cdot T_{\text{all}} \right] \\
&= \min_{T_1, \cdots, T_N} \left[\sum_{i=1}^{N} \mathcal{A}_i(T_i, \cdots) + \mathcal{D} \cdot T_i \right] \\
&= \sum_{i=1}^{N} \mathcal{A}_i^*(\mathcal{D}, \cdots).
\end{aligned}$$

where the order of minimization and summation is switched for the last equality. Therefore, in the dual formulation, the optimization problem over a large number of movements is broken down to smaller problems of individual movements.

[8] Speed-profile of circle-drawing movements has a central portion of a constant speed (peak-speed) surrounded by the initial accelerating and the final deceleration portions. The analytical result in Table 1 is obtained from analyzing only the central portion of circle-drawing movements. This is well justified since the portion shares the same drive-level with the whole movement segment (See appendix for details.)

Table 2.1: Comparison between the original (duration-based) v.s. the Legendre-transformed (drive-based) formulations for straight and circle-drawing movements. Squared-jerk cost is used ($\mathcal{L} = \|\ddot{\boldsymbol{v}}\|^2$, where $\boldsymbol{v}$ is hand velocity). ($\bar{v}$: peak speed)

Type	Original Formulation	Legendre Dual Formulation
Straight	$\mathcal{A} = \frac{720L^2}{T^5}$, $\mathcal{D} = \frac{3600L^2}{T^6}$, $\bar{v} = \frac{30L}{16T}$	$\mathcal{A}^* = \frac{6}{5}\sqrt[6]{3600\mathcal{D}^5L^2}$, $T = \sqrt[6]{\frac{3600L^2}{\mathcal{D}}}$, $\bar{v} = \frac{30}{16}\sqrt[6]{\frac{\mathcal{D}L^4}{3600}}$
Circle[5]	$\mathcal{A} = \frac{R^2\Theta^6}{T^5}$, $\mathcal{D} = \frac{5R^2\Theta^6}{T^6}$, $\bar{v} = \frac{R\Theta}{T}$	$\mathcal{A}^* = \frac{6}{5}\sqrt[6]{5\mathcal{D}^5R^2\Theta^6}$, $T = \sqrt[6]{\frac{5R^2\Theta^6}{\mathcal{D}}}$, $\bar{v} = \sqrt[6]{\frac{\mathcal{D}R^4}{5}}$

2.2.3 Predictions for squared-jerk cost

Let us illustrate the above result by using the squared-jerk cost ($\mathcal{L} = \|\ddot{\boldsymbol{v}}\|^2$, where $\boldsymbol{v}$ is the hand velocity), which has been widely used for describing the *within-movement-regularities* of various hand movement types [Flash and Hogan 1985; Todorov and Jordan 1998]. Here, we analyze straight movements and circle-drawing movements which yield analytic solutions. The main results are shown in Table 2.1. In the original (duration-based) formulation, the independent variables for straight movements are duration and length of movements (T, L), and for circle-drawing movements, the corresponding variables are duration, radius and angular distance (T, R, Θ). Note that duration and peak speed are not predicted in this formulation, since duration must be pre-specified.

In the dual (drive-based) formulation, on the other hand, the role of duration is substiuted by drive $\mathcal{D}$ — the independent variables for straight and circular movements are $(\mathcal{D}, L)$ and $(\mathcal{D}, R, \Theta)$, respectively. For a given level of drive, the movement duration is predicted to scale with the one-third power of the length scale ($L^{1/3}$ for straight and $R^{1/3}$ for circle-drawing movements), and the peak speed to scale with the two-thirds power of the length scale ($L^{2/3}$ and $R^{2/3}$, respectively); and the peak speed should not depend on the angular distance of circle-drawing movements.[9] Moreover, another surprising prediction is that if one makes a straight movement of

[9] This statement is true for movement of large enough angular distance $\Theta \gtrapprox \pi$. In the limit of small θ, the results of straight movements would apply (with $L \approx R\theta$).

length L and a circular movement of radius $R = L$ under the same drive-level, then their peak speed would be related by $\bar{v}_{\text{circle}} \approx 1.6\ \bar{v}_{\text{straight}}$.

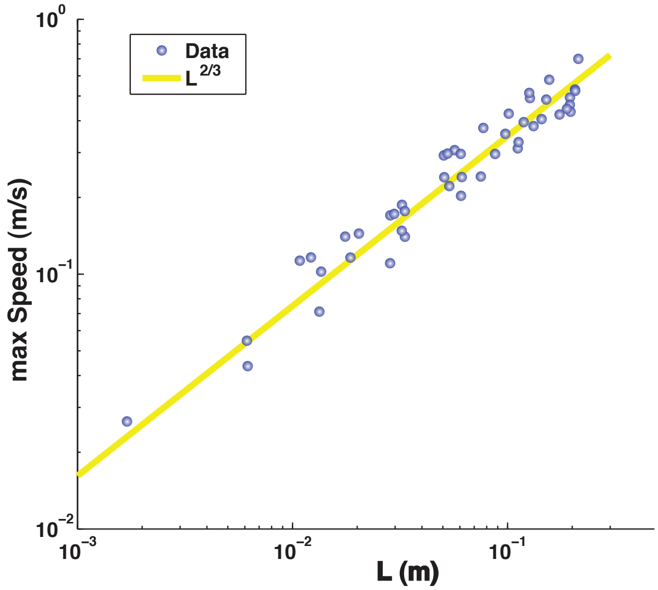
2.3 Experimental Results

We recorded movements from 13 subjects while they generated straight, circular and spiral hand movements using a Wacom Intuos4 digitizing tablet. The recordings had a sampling frequency of 200Hz and a spatial resolution of 0.05mm. The results were then analyzed using customized scripts in Matlab.

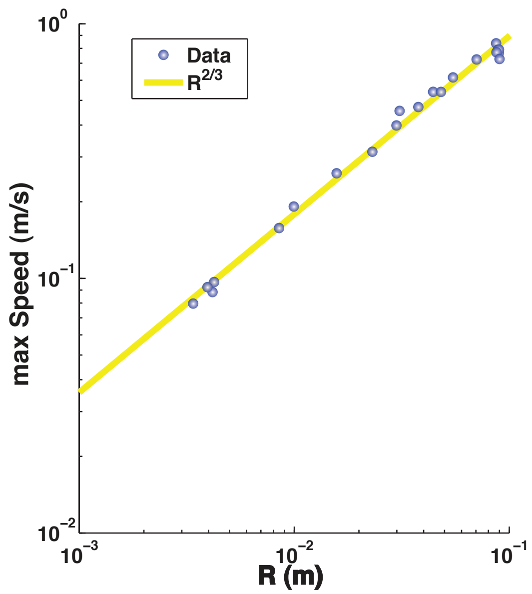
The subjects were simply asked to draw one of the three shapes using the wand on the tablet at a comfortable, fast, or slow pace. There was no target for the subject to hit or any time limit, which are the constraints often used in other studies. This allowed us to study how overall speed is regulated in a naturalistic mode.

For straight hand movements, the speed profiles exhibited symmetric bell-shape curves (Fig1a, consistent with previous studies [Viviani and McCollum 1983; Viviani and Schneider 1991]. There were no secondary or corrective movements near the end, which is typical when a target is not present. For circle-drawing movements, the speed profiles accelerated at the beginning, maintained an approximately constant speed (a central plateau-zone) and decelerated at the end of the continuous movement (Fig2a). The transients at the beginning and end of the circular movements lasted for about 1 or 2 radians. In both types of movements, the speed-profile can be accurately reconstructed by the optimal control models eq (2.1) based on the measured durations.

The scaling of the peak speed with the size of the movements confirmed the predictions made in Table 2.1: For straight movements Fig1b, the peak speed scaled as a power law with movement length with the exponent close to 2/3, $v_{\max} \propto L^{2/3}$. Similarly, for circular movements Fig2b, The peak speed was found to scales with the radius in a power-law manner with the exponent close to 2/3, $v_{\max} \propto R^{2/3}$.



(a) Speed-distance relationship for straight hand movements.



(b) Speed-radius relationship for circle-drawing movements

Figure 2.1: Human movement data for straight and circle-drawing movements.

2.4 Continuous-time Conservation Principle

Here we generalize the concept of drive eq (2.4), which is defined for individual movement segment, to drive-profile in continuous time. That is, at every moment in time, it can be written

$$\mathcal{D}(t) \equiv -\frac{\partial \mathcal{A}(t)}{\partial t}, \qquad \text{and} \qquad \frac{d\mathcal{D}}{dt} = 0, \tag{2.6}$$

where $\mathcal{A}(t)$ is the partial action up to an intermediate time t (see Appendix). Therefore, drive is conserved within through out an movement segment in time, as well as across multiple segments.

For example, the drive profile for the squared-jerk cost is directly related to the velocity profile:

$$\mathcal{D}(t) = \|\ddot{\boldsymbol{v}}\|^2 - 2\dot{\boldsymbol{v}} \cdot \dddot{\boldsymbol{v}} + 2\boldsymbol{v} \cdot \ddddot{\boldsymbol{v}} \tag{2.7}$$

(See Appendix).

The solution of eq (2.7) for a constant drive-profile yields the well known minimum-jerk speed-profile for straight movements: [Flash and Hogan 1985], Thus, the continuous drive profile $\mathcal{D}(t)$ is intimately related to many details of a movement trajectory as it is unfolding in real-time.

2.4.1 Spiral and ellipses

Another class of movements that has analytic solution is spiral movements. Consider an exponential spiral whose radius of curvature is exponentially related to

Table 2.2: Drive-conserving speed-profiles for straight and spiral movements. Squared-jerk cost is used ($\mathcal{L} = \|\ddot{\boldsymbol{v}}\|^2$). Note that in the limit $a \to 0$, the result for spiral reduces to the case of circle-drawing movements (Table 2.1).

Type	Speed-profile	Note
Straight	$v(t) = \frac{30L}{T}\frac{t^2(1-t)^2}{T^4}$,	$T(\mathcal{D}, L) = \sqrt[6]{\frac{3600L^2}{\mathcal{D}}}$
Spiral ($R(\theta) \propto e^{a\theta}$)	$v(t) = \sqrt[6]{\frac{\mathcal{D}\,R(t)^4}{g(a)}}$,	$g(a) = 5 + \frac{5a^2}{3} + \frac{4a^4}{81}$

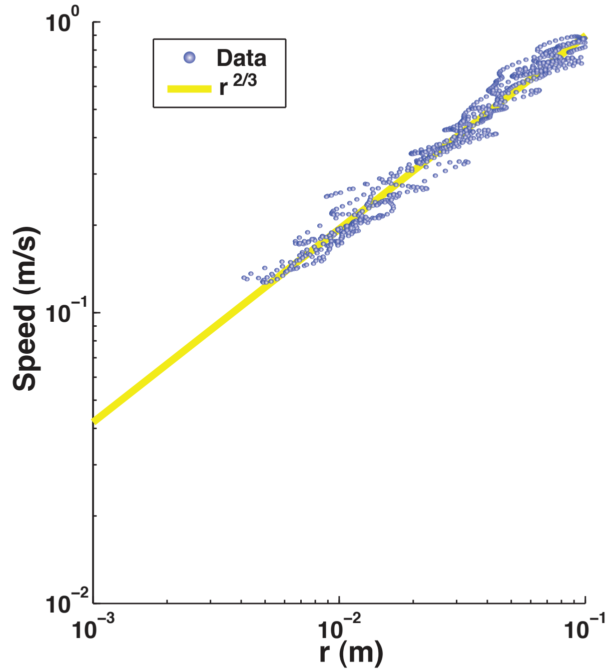


Figure 2.2: Instantaneous relationship between speed and local radius of curvature for spiral-drawing movements.

the angle of rotation: $R(\theta) \propto e^{a\theta}$. (Such spirals are said to be self-similar: Scaling the spiral by a factor α is equivalent to rotation by an angle $\Delta\theta = \alpha/a$.) For the spiral trajectories, the speed profile that corresponds to a constant level of drive can be shown to satisfy the following relationship with the local radius of curvature:

$$v(t) \propto R^{2/3}(t).$$

That is, the instantaneous speed is proportional to the two-thirds power of local radius of curvature (See Table 2.2). Our experimental data indeed confirms human movements this predicted power-law in spiral-drawing movements (Fig 2.2).

2.5 Discussion

2.5.1 A New Conservation Law

In modern physics, the fundamental role in understanding numerous physical phenomena is played by the energy conservation principle, which is based on the time-invariant property of Lagrangian. According to Noether's theorem (1904) such an invariance property always leads to a conservation law [Landau and Lifshitz 1960].

In this work, we derived an analogous principle of drive conservation based on the time-invariant property of both the dynamic system and the cost function of optimal control models for numerous human movements. (This fact implies that the control process of such movements must be also time-invariant.) This novel principle provides a new perspective for understanding previously known phenomena as well as predictions for novel observations.

2.5.2 A new $^1/_3$ Power Law for Spatio-Temporal Scaling of Human Movements

In this work, we demonstrated that a power-law of the same exponent holds for straight, circular, and spiral movements. This results can be further generalized as the following:

Consider two natural hand movements of a same path shape except that one is a spatially magnified version of another by a factor α (a spatial scaling factor). Then, their velocity profiles would be related by a factor $\alpha^{2/3}$ (or equivalently, the temporal scale of the movements would be related by a factor $\alpha^{1/3}$). That is, if the velocity profile of one movement is denoted by $\boldsymbol{v}_o(t)$, then the velocity profile of the scaled movement would be $\boldsymbol{v}(t) = \alpha^{2/3} \cdot \boldsymbol{v}_o(\alpha^{-1/3}t)$. This result, which applies to arbitrary hand trajectories, can be confirmed for squared-jerk cost function from the fact that the corresponding drive-profile eq (2.7) is invariant under this scaling transformation. In fact, however, the origin of the scaling law is more directly related to the drive conservation principle and it holds for a more general class of optimal control models. By the definition (eq (2.4)), the drive has the same dimensions as the cost function, $\mathcal{L}$, which in the case of the squared-jerk cost is

$$[\mathcal{D}] = [\mathcal{L}] = \frac{[length]^2}{[time]^6}. \tag{2.8}$$

Since drive is conserved, the time-scale of a movement must scale with the one-third power of its length-scale. This dimensional analysis based derivation is not restricted to the choice of squared jerk for the cost function, but applies to a more general class of optimal control models, such as the *minimum-joint-torque-change* [Uno, Kawato, and Suzuki 1989] and the *minimum-muscular-force-change* [Pandy and Garner 1995; Stein, Oguztoreli, and Capaday 1986] models. Also see [Todorov and Jordan 2002; Todorov 2005a], which uses squared control cost with a third-order dynamical system. Any cost function related to third-order derivatives of movement trajectories would lead to the same one-third power-law (although approximately, because of the non-linearity of the assumed system dynamics).

It have been argued that the speed-profile shape of minimum-squared-jerk trajectories cannot be distinguished from those of minimum-squared-derivative trajectories of higher order, such as the minimum-snap (4th order) [Harris 1998]. However, according to the above analysis, those models would predict power-laws with different exponents (in general, a *1/n* power law for the *n*th order derivative), which does not match with experimental data. Therefore, the observed scaling law provides a strict criterion for distinguishing between cost functions.

2.5.3 Size Scaling of Spirals vs Ellipses

Because the speed-curvature relationship for spiral movements results from the self-similar property — any two segments of equal angular width within the spiral are scaled versions of each other — therefore, the $1/3$ power-law should apply within a single segment of the spiral movement.

The best studied curved movement have been elliptical trajectories. Within each ellipse, the speed profile varies with local curvature as, $v(t) \propto R(t)^{1/3}$ [Lacquaniti, Terzuolo, and Viviani 1983; Viviani and Flash 1995]. However, the speed profile also depends on the overall size of the ellipse [Viviani and McCollum 1983; Viviani and Schneider 1991], and approximately follows:

$$v(t) \propto \mathcal{D}^{1/6} \bar{R}^{2/3} \left(\frac{R(t)}{\bar{R}} \right)^{1/3} \text{(approximately)}$$

where the term inside of the parenthesis is the normalized radius of curvature and $\bar{R}$ ($\approx \sqrt{R_{\max} R_{\min}}$) is the mean radius of curvature.

Thus, we predict that for elliptic-spiral movements of increasing size, the speed-profile should exhibit both the 1/3 power law for each loop and the 2/3 power law over the entire spiral trajectory.

2.5.4 Comparisons with Previous Models

This new formulation of control theory predicted a new power-law relationship between curvature and speed for spiral-drawing movements: $v(t) \propto R(t)^{2/3}$, which was confirmed with experimental measurements. Previous theoretical papers have predicted that minimum-jerk trajectories should instead follow a one-third power-law [Viviani and Flash 1995; Todorov and Jordan 1998; Richardson and Flash 2002]: $v(t) \propto R(t)^{1/3}$, which was seemingly confirmed by numerous experimental studies [lacquaniti1983law,], and widely thought to apply generally to all curved movements. The resolution of this conflict can be found in Fig xxx, which shows that for a single 360o arc, the slope of the trajectory in the speed versus curvature plot is approximately 1/3, but the average slope over all the arcs in the spiral is closer to 2/3. A similar multiscale structure can be seen in previous measurements, but was ignored.

Another approach to explaining the relationship between duration and overall speed proposed that a lower bound for duration could be set by imposing a desired level of accuracy cyan for the motor task due to the speed-accuracy trade-off (Fitts's Law) [Harris and Wolpert 1998a]. Although this is a reasonable assumption, it cannot be the fundamental mechanism for speed control, which occurs even in the absence of accuracy requirements or targets. Accuracy is only one of many factors that affects the drive-level. Another approach to this problem suggested that movement duration could be predicted by assuming the total-cost (or *action*, $\mathcal{A}$) be a fixed parameter [Guigon, Baraduc, and Desmurget 2008]. However, action is not an intensive property, like drive, and is not conserved.

Other models have been proposed that include a time cost in the action (total cost) to penalize movement duration [Hoff 1994]. Indeed, models with a linear time-cost term make the same predictions shown in Table 2.1, along the lines of the linear time-cost in the Legendre-transformation eq (2.5). However, our analysis reveals something more fundamental, namely that the origin of the predictive power is in fact due to the drive conservation principle, which derives from the time-invariance of the cost-function $\mathcal{L}$. The linear time-cost is simply a side-effect of the Legendre-transform. Moreover, the movement duration T in their approach is an independent parameter that must be pre-specified before the detailed speed-profile is determined by the *optimality principle*.

2.5.5 Biological control

From the perspective of the motor system, the drive-based description has another advantage over the traditional, duration-based description. Consider the planning of a movement with several parts. In the duration-based framework, each segment must be separately planned each with a fixed duration. This becomes problematic for movements that are long and complicated, such as drawing many loops of circle or random doodling that are essentially open ended. In contrast, with a drive-based controller, only the current drive-level and curvature are needed to execute a complex sequence of movements. This raises the question of how much information about the future is needed for generating curved hand movements. We will show in a future study that humans only need on the order of ≈ 1 radian, which corresponds to the angular distance it takes for a circle-drawing movement to reach the plateau speed.

The successful fitting by a model with constant drive to straight, circular and spiral movements suggests that the default mode in the human is to generate movements that conserve constant drive-level during and across tasks regardless of the movement path shape. This also suggests that drive is likely to be used as a high-level control variable in the brain to modulate the overall-speed of movements.

Since drive is a scalar, a natural candidate to represent it in the brain is a neuromodulatory system that can maintain a constant level of activity over the duration of a movement. The tonic level of dopamine neurons in the substantia nigra pars compacta projects diffusely through the cerebral cortex and basal ganglia, particularly to regions that control movements. Patients with Parkinsons disease have low levels of dopamine and move abnormally slowly; in contrast, patients with Tourette syndrome and Huntingtons disease are characterized by uncontrolled ballistic movements. To confirm this prediction it would be fruitful to compare our model prediction with the relationship between tonic dopamine and movement speed.

Chapter 2, in full, is currently being prepared for submission for publication of the material. Huh, Dongsung; Sejnowski, Terrence. The dissertation author was the primary investigator and author of this material.

Appendices

2.A Derivation of the Conservation Law

For illustration, consider the two movement case ($N = 2$):

$$\mathcal{A}_{all}(T_{\text{all}}) = \min_{T_1,T_2}[\ \mathcal{A}_1(T_1) + \mathcal{A}_2(T_2)\ ],$$

where the total duration is considered as fixed: $T_1 + T_2 = T_{\text{all}}$. Then, according to the usual mathematical condition for a minimum, the total action does not change by infinitesimal changes in T_1, T_2, i.e.

$$\begin{aligned} d\mathcal{A}_{all} &= \frac{\partial \mathcal{A}_1}{\partial T_1} dT_1 + \frac{\partial \mathcal{A}_2}{\partial T_2} dT_2 = 0 \\ &= \left(\frac{\partial \mathcal{A}_1}{\partial T_1} - \frac{\partial \mathcal{A}_2}{\partial T_2} \right) dT_1 = 0, \end{aligned}$$

where the actions are differentiated with respect to durations while the other motor-task variables are held constant (partial derivatives). The second equality is due to the condition $dT_1 + dT_2 = 0$. Since dT_1 is an arbitrary variation, we obtain the following conservation relation

$$\frac{\partial \mathcal{A}_1}{\partial T_1} = \frac{\partial \mathcal{A}_2}{\partial T_2}.$$

Now let us consider varying the total duration. According to the above result, the resulting change of total action is

$$d\mathcal{A}_{all} = \frac{\partial \mathcal{A}_1}{\partial T_1} (dT_1 + dT_2) = \frac{\partial \mathcal{A}_1}{\partial T_1} dT_{\text{all}},$$

where $dT_1 + dT_2 = dT_{\text{all}}$ is used. Therefore, we obtain

$$\frac{\partial \mathcal{A}_{\text{all}}}{\partial T_{\text{all}}} = \frac{\partial \mathcal{A}_1}{\partial T_1} = \frac{\partial \mathcal{A}_2}{\partial T_2}.$$

This derivation trivially generalizes to the conservation relationship eq (2.3)(2.4) for general N cases.

2.B Generalized Noether's theorem

This derivation is inspired by Hamilton-Pontryagin principle [Pontryagin, 1962].

An optimal control problem (of/with) a dynamics system $\dot{q} = f(q, u, t)$ and a cost function $\mathcal{L}(q, u, t)$, (q : state vector, u: control signal vector), can be solved by defining *augmented-action* $\tilde{\mathcal{A}}$ for an arbitrary smooth trajectory $\{\tilde{p}, \tilde{q}, \tilde{u}\}$ within a time-window $[0, t]$ as

$$\tilde{\mathcal{A}}(\{\tilde{p}, \tilde{q}, \tilde{u}\}, t) = \int_0^t \left(\tilde{p} \cdot \dot{\tilde{q}} - \tilde{\mathcal{H}}(\tilde{p}, \tilde{q}, \tilde{u}, \tau) \right) d\tau \tag{2.9}$$

where $\tilde{p}$ is the auxiliary or *co-state* vector (or *momentum*), which can be interpreted as Lagrange multipliers, and $\mathcal{H}$ is called *Hamiltonian*, defined as

$$\mathcal{H}(p, q, u, t) \equiv p \cdot f(q, u, t) - \mathcal{L}(q, u, t). \tag{2.10}$$

Then, for given boundary states $q(0)$, $q(t)$, the *extremal-trajectory* for $\tilde{\mathcal{A}}$ can be obtained by applying an arbitrary variation $\{\delta p, \delta q, \delta u\}$ and calculating the changes in $\tilde{\mathcal{A}}$

$$\begin{aligned}\delta\tilde{\mathcal{A}} &= \tilde{\mathcal{A}}(\{\tilde{p}+\delta p, \tilde{q}+\delta q, \tilde{u}+\delta u\}, t) - \tilde{\mathcal{A}}(\{\tilde{p}, \tilde{q}, \tilde{u}\}, t) \\ &= [\tilde{p}\cdot\delta q]_0^t + \int_0^t \big(\delta p\cdot(\dot{\tilde{q}} - \mathcal{H}_p) - (\dot{\tilde{p}} + \mathcal{H}_q)\cdot\delta q - \mathcal{H}_u\cdot\delta u\big)\, d\tau,\end{aligned} \tag{2.11}$$

where integration by parts is used for integrating $\tilde{p}\cdot\delta\dot{q}$, and subscript notation is used to represent differentiation. The *extremal-trajectory* $\{p, q, u\}$ is the one that satisfies

$$\dot{q} = \mathcal{H}_p = f \tag{2.12}$$

$$-\dot{p} = \mathcal{H}_q = p\cdot f_q - \mathcal{L}_q \tag{2.13}$$

$$0 = \mathcal{H}_u = p\cdot f_u - \mathcal{L}_u, \tag{2.14}$$

such that eq (2.11) reduces to

$$\delta\tilde{\mathcal{A}} = p(t)\cdot\delta q(t) - p(0)\cdot\delta q(0), \tag{2.15}$$

which vanishes for boundary-constrained variations (i.e. $\delta q(0) = \delta q(t) = 0$) — a necessary condition for an extremum.

This extremal trajectory is the optimal solution for the original control problem as well as the augmented problem, since the extremal state-trajectory $\{q\}$ satisfies the system dynamics (eq (2.12)), in which case the *augmented-action* eq (2.9) simply reduces to the total cost $\int_0^T \mathcal{L}(q, u, \tau)d\tau$. Therefore, the extremized *augmented-action* is identical to the *action*

$$\mathcal{A}(q(t), t) = \tilde{\mathcal{A}}(\{p, q, u\}, t)]. \tag{2.16}$$

Now come two important identities:

$$p(t) = \left.\frac{\partial \mathcal{A}}{\partial q}\right|_t, \quad \text{and} \quad \mathcal{H}(t) = -\left.\frac{\partial \mathcal{A}}{\partial t}\right|_t. \tag{2.17}$$

The former identity is easily shown from the fact that the variation in action caused by altering the final state is $\delta\mathcal{A} = \delta\tilde{\mathcal{A}} = p(t) \cdot \delta q(t)$, according to eq (2.15)(2.16). The latter is true because on the extremal trajectory

$$\mathcal{H} = p \cdot f - \mathcal{L} = \frac{\partial \mathcal{A}}{\partial q} \cdot \dot{q} - \frac{d\mathcal{A}}{dt} = -\frac{\partial \mathcal{A}}{\partial t}. \tag{2.18}$$

Note that the RHS of eq (2.18) is precisely our continuous-time definition of drive eq (2.6). Therefore, drive and Hamiltonian share the same numerical value along the extremal trajectory (but they have different functional representation), and their rate of change is

$$\begin{aligned} \frac{d\mathcal{D}}{dt} &= \frac{d}{dt}\mathcal{H}(p(t), q(t), u(t), t) \\ &= \mathcal{H}_p \cdot \dot{p} + \mathcal{H}_q \cdot \dot{q} + \mathcal{H}_u \cdot \dot{u} + \mathcal{H}_t \\ &= \mathcal{H}_t \\ &= p \cdot f_t - \mathcal{L}_t, \end{aligned} \tag{2.19}$$

where eq (2.12)(2.13)(2.14) are used for the third equality.

Therefore, drive is conserved for time-invariant problems, i.e. $f_t = 0$, $\mathcal{L}_t = 0$. Note that this result generalizes energy conservation principle of classical mechanics to optimal control problems. (p and $\mathcal{H}$ are closely related to canonical momentum and energy in classical mechanics.)

2.C Instantaneous Jerk-Drive Calculation

Consider a kinematic cost — a function of a generalized coordinate $q \in \mathbb{R}$ and its time derivatives — of up to the 3rd-order derivatives: $\mathcal{L}(q, \dot{q}, \ddot{q}, \dddot{q}, t)$. Generalization to a multi-coordinate problem is trivial. Then, total time differentiation of

$\mathcal{L}$ is

$$\begin{aligned}\frac{d\mathcal{L}}{dt} &= \dot{q}\frac{\partial\mathcal{L}}{\partial q} + \ddot{q}\frac{\partial\mathcal{L}}{\partial\dot{q}} + \dddot{q}\frac{\partial\mathcal{L}}{\partial\ddot{q}} + \ddddot{q}\frac{\partial\mathcal{L}}{\partial\dddot{q}} + \frac{\partial\mathcal{L}}{\partial t}\\ &= \dot{q}\left(\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\dot{q}} - \frac{d^2}{dt^2}\frac{\partial\mathcal{L}}{\partial\ddot{q}} + \frac{d^3}{dt^3}\frac{\partial\mathcal{L}}{\partial\dddot{q}}\right) + \ddot{q}\frac{\partial\mathcal{L}}{\partial\dot{q}} + \dddot{q}\frac{\partial\mathcal{L}}{\partial\ddot{q}} + \ddddot{q}\frac{\partial\mathcal{L}}{\partial\dddot{q}}\\ &\quad + \frac{\partial\mathcal{L}}{\partial t}\end{aligned} \tag{2.20}$$

where *Euler-Lagrange equation* is used for the substitution:

$$\frac{\partial\mathcal{L}}{\partial q} = \frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\dot{q}} - \frac{d^2}{dt^2}\frac{\partial\mathcal{L}}{\partial\ddot{q}} + \frac{d^3}{dt^3}\frac{\partial\mathcal{L}}{\partial\dddot{q}}.$$

Rearranging the terms, we obtain

$$\frac{d}{dt}\left(\mathcal{L} - \dot{q}\frac{\partial\mathcal{L}}{\partial\dot{q}} - \ddot{q}\frac{\partial\mathcal{L}}{\partial\ddot{q}} + \dot{q}\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\ddot{q}} - \dddot{q}\frac{\partial\mathcal{L}}{\partial\dddot{q}} + \ddot{q}\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\dddot{q}} - \dot{q}\frac{d^2}{dt^2}\frac{\partial\mathcal{L}}{\partial\dddot{q}}\right) = \frac{\partial\mathcal{L}}{\partial t}.$$

According to eq (2.19), the expression within the parenthesis is identified as the drive with a minus sign ($-\mathcal{D}$). (In a kinematic problem, the system dynamics is time-invariant, $f_t = 0$.) Therefore,

$$\mathcal{D} \equiv -\mathcal{L} + \dot{q}\frac{\partial\mathcal{L}}{\partial\dot{q}} + \ddot{q}\frac{\partial\mathcal{L}}{\partial\ddot{q}} - \dot{q}\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\ddot{q}} + \dddot{q}\frac{\partial\mathcal{L}}{\partial\dddot{q}} - \ddot{q}\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\dddot{q}} + \dot{q}\frac{d^2}{dt^2}\frac{\partial\mathcal{L}}{\partial\dddot{q}}, \tag{2.21}$$

(In fact, this result is an extension of the usual expression for energy from classical mechanics, which is $\mathcal{E}_{mech} \equiv -\mathcal{L} + \dot{q}\cdot\partial\mathcal{L}/\partial\dot{q}$.)

For the squared-jerk cost ($\mathcal{L} = \dddot{q}^2$, where q is the hand position), the drive equation (2.21) becomes

$$\mathcal{D} = \dddot{q}^2 - 2\ddot{q}\ddddot{q} + 2\dot{q}q^{(5)}. \tag{2.22}$$

2.D Spiral movements

Consider movements along an exponential spiral path represented as $r(\theta) = r_o e^{a\theta}$ in polar coordinate. Then, velocity of the spiral movement is expressed as $\boldsymbol{v}(\theta,\dot{\theta}) = (a\boldsymbol{e}_r + \boldsymbol{e}_\theta)\dot{\theta}r(\theta)$, where $\boldsymbol{e}_r, \boldsymbol{e}_\theta$ are the basis vectors, and radius of curvature is $R(\theta) = \sqrt{a^2+1}\,r(\theta)$.

The (optimal/drive-conserving) solution can be found by assuming a similar exponential form for the angular speed, $\dot{\theta}(\theta) = \omega_o e^{-b\theta}$, or equivalently $\theta(t) = \log(1 + bw_o t)^{1/b}$, which yields $r(t) = r_o(1 + b\omega_o t)^{a/b}$, $\dot{\theta}(t) = \omega_o(1 + b\omega_o t)^{-1}$.

Then, the drive profile can be found from eq (2.7) as

$$\mathcal{D}(t) \propto (1 + b\omega_o t)^{2(a/b-3)},$$

which becomes a constant function only if $b = a/3$. Then, the drive profile can be expressed in terms of local radius of curvature and angular speed as

$$\mathcal{D}(t) = g(a) \cdot R^2(t)\dot{\theta}^6(t) \ = const,$$

where $g(a) = (5 + {}^{5a^2}/_3 + {}^{4a^4}/_{81})$ is a constant gain, or equivalently, the speed-profile is

$$v(t) = R(t)\dot{\theta}(t) = \left(\frac{\mathcal{D}(t)}{g(a)}\right)^{1/6} R^{2/3}(t).$$

References

Abend, W., E. Bizzi, P. Morasso, et al. (1982). “Human arm trajectory formation.” In: *Brain: a journal of neurology* 105.Pt 2, p. 331.

Bernstein, N.A. (1967). *The coordination and regulation of movements.*

Fitts, P.M. (1954). “The information capacity of the human motor system in controlling the amplitude of movement”. In: *Journal of experimental psychology* 47.6, pp. 381–391.

Flash, T. and N. Hogan (1985). “The coordination of arm movements: an experimentally confirmed mathematical model”. In: *Journal of neuroscience* 5.7, p. 1688.

Gordon, J. et al. (1994). “Accuracy of planar reaching movements”. In: *Experimental Brain Research* 99.1, pp. 112–130.

Gribble, P.L. and D.J. Ostry (1996). “Origins of the power law relation between movement velocity and curvature: modeling the effects of muscle mechanics and limb dynamics”. In: *Journal of Neurophysiology* 76.5, pp. 2853–2860.

Guigon, E., P. Baraduc, and M. Desmurget (2008). “Optimality, stochasticity, and variability in motor behavior”. In: *Journal of computational neuroscience* 24.1, pp. 57–68.

Harris, C.M. (1998). “On the optimal control of behaviour: a stochastic perspective”. In: *Journal of neuroscience methods* 83.1, pp. 73–88.

Harris, C.M. and D.M. Wolpert (1998a). “Signal-dependent noise determines motor planning”. In: *Nature* 394.6695, pp. 780–784.

Hoff, B. (1994). “A model of duration in normal and perturbed reaching movement”. In: *Biological Cybernetics* 71.6, pp. 481–488.

Lacquaniti, F., C. Terzuolo, and P. Viviani (1983). “The law relating the kinematic and figural aspects of drawing movements”. In: *Acta psychologica* 54.1-3, pp. 115–130.

Landau, L.D. and EM Lifshitz (1960). “Classical mechanics”. In: *Course of Theoretical Physics. Butterworth–Heinemann.*

Pandy, MG and BA Garner (1995). “Optimal Control ot Non-ballistic Muscular Movements: A Constraint-Based Performance Criterion for Rising From a Chair”. In: *Journal of biomechanical engineering* 117, p. 15.

Richardson, M.J.E. and T. Flash (2002). “Comparing smooth arm movements with the two-thirds power law and the related segmented-control hypothesis”. In: *Journal of Neuroscience* 22.18, p. 8201.

Stein, RB, MN Oguztoreli, and C. Capaday (1986). “What is optimized in muscular movements”. In: *Human muscle power*, pp. 131–150.

Todorov, E. (2004). “Optimality principles in sensorimotor control (review)”. In: *Nature neuroscience* 7.9, p. 907.

— (2005a). “Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system”. In: *Neural computation* 17.5, pp. 1084–1108.

Todorov, E. and M.I. Jordan (1998). “Smoothness maximization along a predefined path accurately predicts the speed profiles of complex arm movements”. In: *Journal of Neurophysiology* 80.2, p. 696.

— (2002). “Optimal feedback control as a theory of motor coordination”. In: *Nature neuroscience* 5.11, pp. 1226–1235.

Uno, Y., M. Kawato, and R. Suzuki (1989). “Formation and control of optimal trajectory in human multijoint arm movement”. In: *Biological cybernetics* 61.2, pp. 89–101.

Viviani, P. and T. Flash (1995). “Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning.” In: *Journal of Experimental Psychology: Human Perception and Performance* 21.1, p. 32.

Viviani, P. and G. McCollum (1983). “The relation between linear extent and velocity in drawing movements”. In: *Neuroscience* 10.1, pp. 211–218.

Viviani, P. and R. Schneider (1991). “A developmental study of the relationship between geometry and kinematics in drawing movements.” In: *Journal of Experimental Psychology: Human Perception and Performance* 17.1, p. 198.

Chapter 3

Infinite-Horizon Optimal Feedback Control Formulation of Target-directed Reaching Movements

Abstract Humans can rapidly and accurately reach for objects in the environment with smooth trajectories. In a finite horizon model, the time to reach the target is fixed, and trajectories are found that optimize the smoothness cost. Here we analyze an infinite horizon optimal control framework, including control dependent noise, which generalizes previous models of reaching. Three scaling parameters emerge for distance, speed and noise level. Within this framework a wide range of human reaching data can be explained, including Fitts's law for the tradeoff between speed and accuracy. The speed parameter can be interpreted as the motivational level and has biological correlates.

3.1 Introduction

The complex bio-mechanics of human body is capable of generating an unlimited repertoire of movements, which on one hand yields highly versatile motor behavior but on the other hand presents a formidable control problem for the brain. Yet, we can easily perform various motor tasks with a high degree of coordination, and interestingly we tend to utilize only a narrow range of movements out of the much larger set of possibilities (i.e. our movements are highly stereotyped) Bernstein 1967. Understanding the computational process behind the biological motor systems is of central interest to both neuroscience and robotics control.

In recent decades, the principle of optimality has been widely accepted as an adequate framework for understanding biological motor control. For various types of movements, elaborate predictions of behavior have been obtained from optimizing parsimonious performance criteria (see review [Todorov 2004]).

However, there are two main challenges associated with implementing optimal control models, one of which is identifying the appropriate intrinsic cost of movement generation. Indeed, this issue has been extensively considered in earlier investigations, which proposed a number of candidate cost functions, including *squared jerk* [Flash and Hogan 1985], *squared torque-change* [Uno, Kawato, and Suzuki 1989], *squared muscular-force-change* [Pandy and Garner 1995; Stein, Oguztoreli, and Capaday 1986], and *squared control-signal* cost [Todorov and Jordan 2002; Todorov 2005a]. Except for some minor differences, however, these cost functions tend to generate trajectories of similar, smooth spatio-temporal characteristics. Moreover, all of them can be re-expressed as the squared control-signal costassociated with different body dynamics models.

Besides the intrinsic cost, another crucial component of an optimal control model is the intended goal of movement; After all, without a specified motor goal, the most economic motor strategy is a trivial one — not to move at all! Despite its importance, however, the problem of identifying the appropriate representation of motor goal has not been treated in a principled manner.

On the other hand, the choice of motor goal representations may lead to much more diverse variety of qualitative differences in the model predictions, yet has been relatively under investigated.

For a target-directed movement, for example, which is the simplest and the most extensively studied type of movements, the motor goal only includes information of target-location and how fast to reach for the target.

Even such simplest goal information can be represented in a number of ways, each of which has strengths and limitations. In many of the early optimal control models, the goal information is represented as explicit constraints that designate the terminal state of the limb and the terminal time of the movement (*fully-constrained* models) [Flash and Hogan 1985; Uno, Kawato, and Suzuki 1989; Pandy and Garner 1995; Stein, Oguztoreli, and Capaday 1986].

Even such simplest goal information, however, can be represented in alternative ways, each of which has strengths and limitations.

For example, in *state-relaxed* models, the state-constraint is replaced by a terminal-state-cost function that penalizes the error between the desired target-state and the actual limb-state at the end of the movement [Todorov and Jordan 2002; Todorov 2005a]. In *time-relaxed* models, the time-constraint is substituted by a time-cost term which penalizes long movement duration [Hoff 1994].

Because they consider a more general class of movements as their solutions, the relaxed formulations can be used to predict a wider variety of movement features. Time-relaxed models can predict movement duration and overall speed as functions of other movement features (e.g. distance to target), while the state-relaxed formulation can incorporate stochasticity in the body dynamics and environment, which is crucial for understanding the effects of sensory feedback correction in motor control.

There are, however, several major drawbacks to current models with relaxed formulations. First, in a state-relaxed model the predicted trajectories systematically undershoot instead of accurately achieving the desired target-states. This occurs because accuracy is being traded off with other terms in the cost function and the optimal solution tolerates some error in the endpoint of the movement. Second, relative weights must be assigned for each of the terms in the cost function and there is no good method for choosing these values. Introducing an additional time-cost term to the cost function adds another coefficient and exacerbates this problem.

In this paper, we propose an alternative method based on infinite-horizon control for describing target-directed reaching movements in a constraint-free way. In the results section, the new formulation is analyzed and further integrated with stochastic feedback control. We then show how previously proposed models of reaching movements can be derived from the infinite-horizon model as special, limiting cases, which illuminates their relationship with each other and reveals the origin of their limitations.

3.2 Results

Reaching movements toward a stationary target have smooth spatiotemporal characteristics: a nearly straight path and a smooth bell-shaped speed profile. Subjects systematically trade-off speed for accuracy, known as Fitts's law [Fitts 1954]. In a finite-horizon optimal control framework, trajectories are defined within a time-window of finite duration, T, which may be fixed or variable, and the duration of a movement is assumed to coincide with the duration of the time-window. The control cost is applied throughout a finite time window, but the target-information is provided only at the terminal time as a state constraint or as a state cost.

In this paper, we propose an alternative formulation based on infinite-horizon optimal control, which minimizes the total expected cost over an infinitely long time window:

$$\min_{\{u_t\}} \mathbb{E}\left[\int_0^\infty \mathcal{L}(\boldsymbol{x}_t, u_t)\, dt\right] \tag{3.1}$$

where $\boldsymbol{x}$ is the state and u is the control signal).

The instantaneous cost function consists of a control cost term and a state-dependent cost term, called accuracy cost, that penalizes the deviation between the

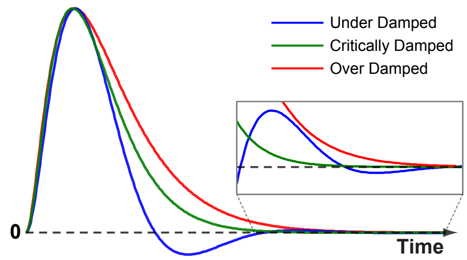


Figure 3.1: Optimal speed-profiles for quadratic accuracy-cost eq (3.2). Three different types are shown: under-damped (i.e. oscillations due to time constants with imaginary components) over-damped (i.e. slowly decaying tail) and critically-damped (i.e. fastest convergence without oscillations due to degenerate time-constant).

actual state, $\boldsymbol{x}$, and the target-state, $\boldsymbol{x}^*$:

$$\mathcal{L}(\boldsymbol{x}, u) = q(\boldsymbol{x} - \boldsymbol{x}^*) + r\, u^2, \quad (r > 0) \tag{3.2}$$

where the target-state corresponds to a state with zero velocity and zero acceleration at the target location.

The solution of this optimal control problem is composed of linear mixtures of exponentially converging dynamics. The cost-coefficients Q and r determine the time constants for the convergence and the speed profile (Fig 3.1). For some parameters, the solutions may oscillate (underdamped) or have a slowly decaying tail (over-damped). The most symmetric speed profile without oscillation is the critically-damped condition when the time-constants are all real, corresponding to the cost function:

$$\mathcal{L}_c(\boldsymbol{x}, u) = x^2 + 3\tau^2\dot{x}^2 + 3\tau^4\ddot{x}^2 + \tau^6 u^2, \tag{3.3}$$

where τ is a free parameter corresponding to the time constant that determines the overall magnitude of speed profile (See Appendix). However, the shape of the speed profile is skewed compared to human movements: the initial acceleration phase is much faster than the deceleration phase (Fig 3.2). The challenge is to find an accuracy-cost that does not distort the speed profile in the infinite-horizon formulation. We therefore explored more general, non-quadratic accuracy-cost functions. Previous optimal control models did not provide guidance since they generally considered only quadratic cost functions. The space of possible non-quadratic cost functions is vast and most lack analytic solutions.

We consider a restricted class of tractable non-quadratic cost functions:

$$\mathcal{L}(\boldsymbol{x}, u) = \Theta\left(y\right) + \frac{\tau^6 u^2}{\delta^2}, \qquad (3.4)$$
$$y = \frac{x^2 + 3\tau^2\dot{x}^2 + 3\tau^4\ddot{x}^2}{\delta^2}$$

where the length constant δ and the time constant τ were introduced to make the arguments of the cost dimensionless; $\Theta(\cdot)$ is a nonlinear, monotonically growing envelope function with a unique local minimum at the origin around which it is asymptotically linear for small inputs: $\lim_{y\to 0}\Theta(y) = y$. (See Methods/Appendix). The coefficients were chosen so that the cost eq (3.4) reduces to the critically-damped quadratic cost eq (3.3) near the origin.

The parameters δ and τ in eq (3.4) set the spatial and temporal scales of movements, respectively, since they modulate quantitative behavioral features, such as overall magnitude of speed profiles, but not the shape of the movement, which is determined solely by the envelope function $\Theta(\cdot)$.

Fig 3.2 illustrates the relationship between a range of envelope functions and the optimal speed profile. In all cases the trajectories are critically-damped near the target as the accuracy cost becomes approximately quadratic. Comparing the trajectories of the different envelope functions, the speed profile becomes more symmetric as the steepness of the envelope function decreases: The initial acceleration phase is broadened and the deceleration phase is shortened (Fig 3.2). The envelope function with the most symmetric speed profile is $\Theta(y) = 1-\exp(-y))$, which becomes completely flat far away from target.

The duration and peak speed of movements in the infinite-horizon formulation are determined during the optimization process in a natural way that also depends on the accuracy cost function. The saturating envelope yields the closest prediction of movement duration. Steeper accuracy-cost functions, such as the quadratic cost, predict much shorter durations for large movement amplitudes (See Supplementary Materials). This result is closely related to Hoff's time-relaxed optimal control

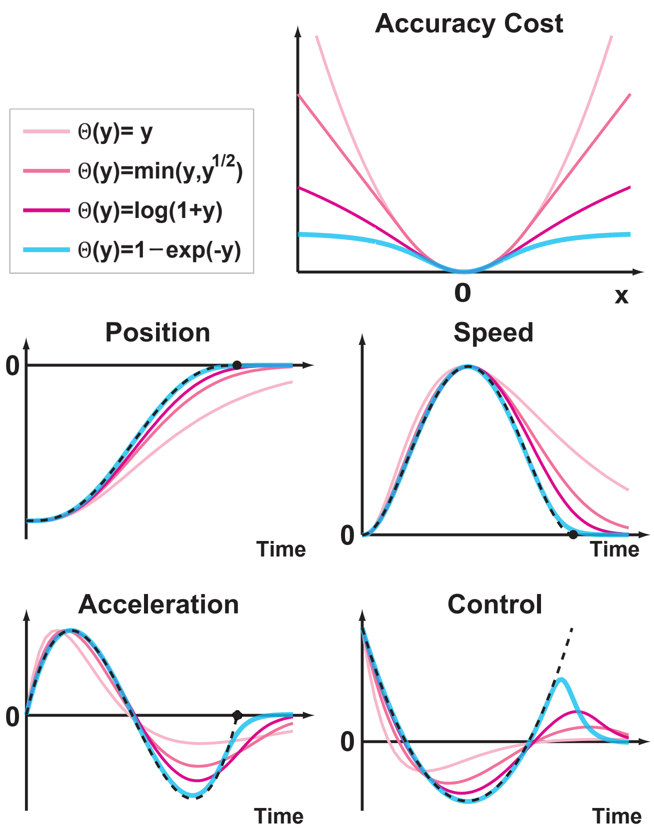


Figure 3.2: Comparison of various accuracy costs (upper right) and corresponding optimal movement kinematics (lower plots). The envelope functions $\Theta(\cdot)$ that generates the non-quadratic cost functions eq (3.4) are shown in legend. Upper Right: Cross section of accuracy-costs along the position (x) axis. For a vivid qualitative comparison, the magnitudes of speed-profiles and the decay-time constant τ have been adjusted for each envelope function, such that the magnitude and time of peak-speed all coincide. The prediction of the deterministic finite-horizon model (with constraints on the final state and time) is also plotted for comparison (black dashed curves).

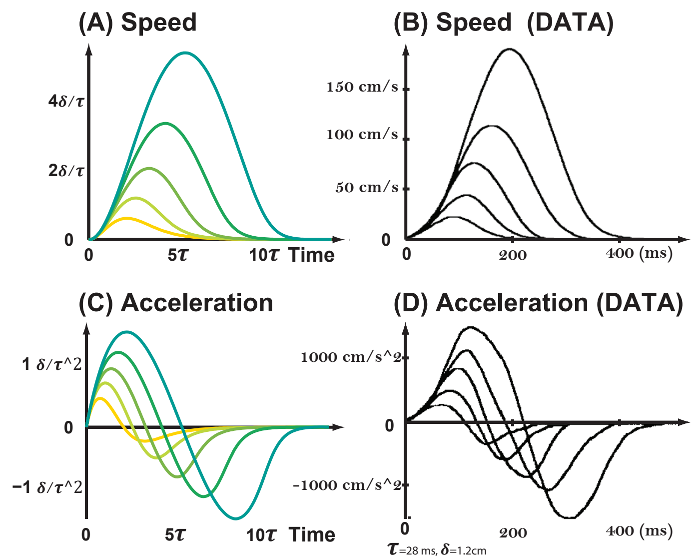


Figure 3.3: Speed- and acceleration-profiles for movements over various distances. (B,D) Average speed- and acceleration-profiles of human reaching movements for various movement amplitudes in the absence of visual feedback of hand position [Gordon et.al. 1994]. (A,C) Predicted kinematic-profiles of the deterministic infinite-horizon model with a saturating cost function. The predicted the speed-profile shape as well as the predicted pattern of how movement duration and peak-speed scales with movement amplitude match well with the experimental data. Note that the parameters δ and τ simply set the scales on the axes. ($\delta \approx 1.2$ cm and $\tau \approx 28$ ms)

model for duration in reaching movements [Hoff and Arbib 1993].

The close match between the predicted and observed trajectories is primarily due to the structure of the cost function and the infinite-horizon framework, not a result of parameter fitting. The only parameters in the model, δ and τ, set the scale of the spatiotemporal axes (Fig 3.3) without influencing the shapes of the trajectories. These parameters will be discussed in more detail in the next section.

3.2.1 Stochastic feedback control model

The deterministic, open-loop infinite-horizon model accurately describes averaged ballistic behavior, but biological movements are continuously affected by internal neuromuscular noise and external perturbations, which require closed-loop feedback control to achieve high accuracy movements. In this section, we generalize the infinite-horizon framework to incorporate a realistic noise model of limb dynamics, solve a feedback-control model and analyze how the motor noise and feedback correction affects the average and variance of reaching movements.

Internal noise in biological motor systems is not additive and instead is multiplicative and control-dependent: When the muscles are relaxed, the passive limb dynamics has low noise and is deterministic, but during active control the motor noise measured by the standard deviation of muscle force grows linearly with the mean force level

[Sutton & Sykes, 1967; Todorov, 2002; Schmidt, Zelaznick, Hawkins, Frank, & Quinn, 1979]. Motor noise can be explained in part by intrinsic variability of neural activity combined with the size principle of motoneuron recruitment (Jones, Hamilton, & Wolpert, 2002). Multiplicative noise can be formally incorporated into the limb dynamics model by adding a stochastic component to the control signal during an infinitesimal interval dt: $u \cdot dt + \sqrt{\sigma} u \cdot d\xi$, where $u \cdot dt$ is the average effect of the control, ξ_t is the standard Wiener process, and $\sqrt{\sigma} u$ is the control-dependent multiplicative factor. Consequently, the variance of the control noise during the interval dt is $\sigma\, u^2 dt$ (See eq (3.13) in Methods). Similar noise models were previously proposed for discrete-time stochastic models [Todorov 2005a; Todorov and Jordan 2002; Harris and Wolpert 1998a].

Multiplicative noise has nonintuitive consequences for motor control. First, although the motor noise is typically measured as positional inaccuracy, neither the noise process ξ_t nor the noise rate σ are spatial variables. Rather, σ has the dimensions of $[time]$ and thus is best understood as a time scale parameter. Moreover, the noise magnitude depends not on σ alone, but on the dimensionless ratio σ/τ. This is because even when σ is large the noise level can be reduced by moving slowly, which requires a smaller control signal. Increasing both τ and σ by a common factor α rescales the duration of the movements without altering qualitative features such as noise magnitude and the shape of the speed profile (See Appendix). Thus, one of time constants can be the unit of time. Here, we choose σ to scale time since it is an intrinsic property of the neuromuscular system and is relatively constant during movements, whereas τ can be quite variable. The ratio τ/σ controls the overall speed and can be considered the level of motivation of a planned movement.

Due to the non-quadratic cost function eq (3.4), the optimal feedback control law becomes nonlinear even though the limb dynamics model is linear; therefore, the resulting closed-loop dynamics of the solution is nonlinear. Moreover, the control-dependent noise adds extra complexity to the problem, which cannot be solved analytically. The optimal solution of the stochastic model was therefore found by discretizing the state-space and treating it as a standard Markov Decision Process problem using value iteration.

Movements over a range of τ/σ for the stochastic model are shown in Fig 3.4. Faster movements are noisier because of the multiplicative noise, which could explain the well-known tradeoff between speed and accuracy [Sutton and Sykes 1967; Schmidt et al. 1979; Todorov and Jordan 2002].

Optimal movement trajectories in the limit of maximum motivation are shown in Fig 3.5a and Fig 3.7. The variability for a single movement increases sharply in the initial accelerating phase due to the control-dependent motor noise, and decreases in the decelerating phase due to the corrective feedback and at the target, the positional variance reduces to δ Fig 3.7a. Thus δ, the width of the accuracy-cost, can be naturally interpreted as the target size, or the desired level of accuracy.

Another well investigated property of reaching movements is the speed-accuracy trade-off when subjects are required to move as fast as possible toward targets of various sizes, called Fitts's Law [Fitts 1954]. Under this condition, the duration of a movement was found to increase with the dimensionless ratio between initial distance from the target and target-size. The movement duration (T) depends the ratio between movement amplitude (A) and the width of the accuracy-cost (δ), which provides the [$length$]-scale for the movement. The model prediction matches well with the Fitts's data when $\sigma \approx 40ms$ Fig 3.6A.

3.3 Comparison with Finite-Horizon models

The infinite-horizon model introduced here incorporates stochastic feedback control and accurately predicts a wide range of observed phenomena. Moreover, movement-related features — target-size (δ), noise-rate (σ), and tempo (τ) — are model parameters in a simple, easily interpretable framework without terminal-time or terminal-state constraints. The model also provides a context for comparing previous, finite-horizon models and explaining the source of their limitations.

To make the comparison concrete, we first introduce an alternative formulation of the infinite-horizon model, which minimizes the following cost function:

$$\min_{\{u_t\},\tilde{T}} \left[\int_0^{\tilde{T}} \left(1 + \frac{\tau^6 u_t^2}{\delta^2} \right) dt + \int_{\tilde{T}}^{\infty} \mathcal{L}_c(\boldsymbol{x}_t, u_t)\, dt \right], \tag{3.5}$$

where $\mathcal{L}_c$ refers to the quadratic cost function eq (3.3). Here, the accuracy cost is a constant value of 1 and become a quadratic state-cost at some time $\tilde{T}$, which is to be optimized. The following condition holds for optimal solutions at the target boundary:

$$x_{\tilde{T}}^2 + 3\tau^2 \dot{x}_{\tilde{T}}^2 + 3\tau^4 \ddot{x}_{\tilde{T}}^2 = \delta^2, \tag{3.6}$$

where $\tilde{T}$ is the moment when the optimal movement trajectory crosses the ellipsoid

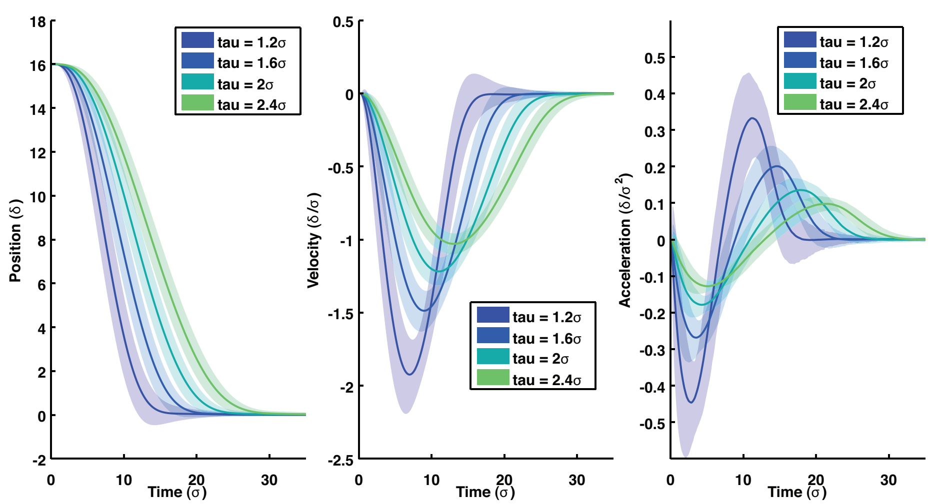


Figure 3.4: Movements of various τ. Note that lower value of τ produces faster and noisier movements.

surface described by eq (3.6) (See appendix).

This alternate formulation is equivalent to the original infinite-horizon model in a deterministic setting with a sharply saturating envelope function, $\Theta(y) = \min(y, 1)$. The target boundary represents the states with values equal to 1 on the inflection of the envelope function. This re-expressed infinite-horizon model can now be compared with the finite-horizon models.

3.3.1 Comparison with Relaxed Terminal-State Models

The relaxed terminal-state models [Todorov and Jordan 2002] minimize the following cost function:

$$\min_{\{u_t\}} \left[\int_0^T r\, u_t^2 \, dt + \boldsymbol{x}_T^\intercal W \, \boldsymbol{x}_T \right], \tag{3.7}$$

where the last term is called the terminal-state cost, r and W are positive-definite coefficients and the terminal time, T, is a fixed parameter. The terminal state $\boldsymbol{x}_T$, however, does not necessarily hit the target. Motivation to achieve the target state is provided by the terminal-state cost. These models can be generalized to stochastic control, but cannot predict movement duration.

The relaxed terminal-state models have several limitations: (1) The optimal movement trajectories do not reach the target-state, but instead exhibit a systematic undershooting bias. This occurs because the terminal state is subject to the trade-off between the control cost and terminal-state cost. A residual error is optimal in this formulation because of the trade-off between control cost and terminal-state cost.

(2) The models include many free parameters (6 coefficients in W, 1 in r and the terminal time T), without a procedure for specifying them. Moreover, the influence of some parameters is unclear: The coefficients W, r do not even contribute to overall speed (when they are sufficiently large), since the duration is pre-determined.

We now show how the terminal-state-relaxed formulation is related to the infinite-horizon model. The reformulated infinite-horizon problem eq (3.5) can be simplified by dividing the total cost at time $\tilde{T}$ into two parts and by optimizing the later part: $[\tilde{T}, \infty]$ (Bellman's principle of optimality). This leads to the following reduced problem for the interval $[0, \tilde{T}]$:

$$\min_{\{u_t\},\tilde{T}} \left[\tilde{T} + \int_0^{\tilde{T}} r\, u_t^2 \, dt + \boldsymbol{x}_{\tilde{T}}^{\intercal} V \, \boldsymbol{x}_{\tilde{T}} \right], \tag{3.8}$$

where $r = \tau^6/\delta^2$ and the last term

$$\boldsymbol{x}_{\tilde{T}}^{\intercal} V \, \boldsymbol{x}_{\tilde{T}} = \min \left[\int_{\tilde{T}}^{\infty} \mathcal{L}_c \, dt \right], \quad V = \frac{1}{\delta^2} \begin{pmatrix} 1\tau^1 & 3\tau^2 & 3\tau^3 \\ 3\tau^2 & 8\tau^3 & 3\tau^4 \\ 3\tau^3 & 3\tau^4 & 1\tau^5 \end{pmatrix} \tag{3.9}$$

is the optimal cost-to-go function at time $\tilde{T}$, which describes the expected future cost from beyond $\tilde{T}$, assuming that the continuation of the trajectory is optimally controlled by solving the algebraic Riccati equation. Therefore, if we interpret $\tilde{T}$ as the terminal time and the optimal cost-to-go function as the terminal-state cost, in effect truncating the trajectory at $\tilde{T}$, then the reduced formulation eq (3.8) is indeed closely related to the state-relaxed formulation eq (3.7). The reduced infinite-horizon model has several advantages over the related finite-horizon model: (1) It can predict movement duration. (2) The coefficients of the cost function are determined by just two parameters (target-width δ and tempo τ) whose influence on the movement trajectories are intuitively understood. (3) It explains that the systematic undershooting bias problem is due to the truncation at the target boundary at time T, which is an unavoidable shortcoming that arises from approximating an infinite-horizon process with the finite-horizon model by prematurely truncating the trajectory at the target boundary at time T, or equivalently, at the target boundary eq (3.6).

3.3.2 Comparison with relaxed terminal-time models

The undershoot bias problem with time-relaxed finite-horizon optimal control can be minimized by reducing the size of the target-boundary, δ. Here, we consider the limit $\delta \to 0$ in eq (3.8), while decreasing τ such that the control-cost coefficient, $r = \delta^2/\tau^6$ remains finite. In this limit, the terminal-state-cost eq (4.8) effectively becomes a hard constraint on terminal-state, since all the elements of the V matrix diverge to infinity, and the reduced problem becomes equivalent to the relaxed terminal-time formulation [Hoff 1994], which minimizes the following cost

$$\min_{\{u_t\},\tilde{T}} \left[\tilde{T} + \int_0^{\tilde{T}} r\, u_t^2 \; dt \right], \tag{3.10}$$

with the constraint that the trajectories must terminate at the target-state: $\boldsymbol{x}_{\tilde{T}} = \mathbf{0}$.

Although this formulation can predict movement duration, it cannot incorporate feedback-control in stochastic settings or explain the speed-accuracy relationship in Fitts law because the target-width parameter has vanished. State-constraints are incompatible with stochastic control because the arrival at the target is governed by probabilities and cannot be guaranteed.

3.3.3 Comparison with Minimum-Variance models

The reduced infinite-horizon formulation can also be considered in a stochastic setting in the limit $\tau \to 0$, in which the expectation of the total cost eq (3.5) reduces to:

$$\min_{\{u_t\},\tilde{T}} \left[\tilde{T} + \frac{1}{\delta^2} \int_{\tilde{T}}^{\infty} \mathbb{E}[x_t^2] \; dt \right]. \tag{3.11}$$

This is equivalent to the *minimum-variance model* that minimizes the positional variance after the movement has ended. An early minimum-variance used a fixed terminal-time [Harris and Wolpert 1998a], but a later version of the model relaxed this constraint and instead included a time-cost term [Harris and Wolpert 2006], as in the above formulation. These are open-loop control models with a rigid terminal-state constraint, but they can be applied in stochastic settings because the constraint is imposed on the average trajectory rather than on individual trajectories.

This class of models is distinguished from other models in that the cost function comprises time cost and the positional-variance cost but does not include a control cost. The variance-cost is integrated only for a finite duration, which is a free-variable, since the positional variance never decays to zero. Perhaps the most important departure is that the positional variance cost is integrated only after the terminal time T. This latter point is inconsistent with finite-horizon frameworks in which the cost-function can only reflect events that occur within the fixed time-window $[0, T]$.

The origin of these differences arises from the infinite-horizon formulation in the limit $\tau \to 0$. Although minimum variance approaches can be used in explaining Fitts's law for the fastest movements under accuracy requirements, they cannot explain how the speed is modulated at lower rate, because the motivation parameter τ cannot be relaxed from maximum motivation at zero.

3.4 Discussion

Motor performance is constrained by the complexity of the anatomical structure, physiological variability computational resources available to the nervous system. A complex model that included all of these complexities would be difficult to specify and analyze. Our goal was to start with an idealized model that avoids extra assumptions whenever possible, introducing them only when some aspect of observed behavior is suboptimal in the idealized sense.

We used a simple, linear dynamics model for the plant. Nonetheless, the model generated many interesting features of reaching movements. Other realistic features of reaching movements may be reproduced with more realistic limb dynamics. Another feature is the inclusion of noise and delay in sensory signals. We used a nonspecific model of the lumped effects of all unknown internal noise sources and adjusted.

The infinite-horizon model of reaching movements can accommodate stochastic feedback control and predicts movement duration as well as other critical features of movements. The model reproduces all features of reaching movements in an integrated manner, such as accurate reaching (no offset), speed-accuracy trade-off, speed-motivation trade-off and feedback control under sensorimotor noise. Several previously studied finite-horizon models as well as the minimum variance model, although seemingly unrelated, can be derived from this infinite-horizon formulation as special limiting cases.

3.4.1 Accuracy Cost

Target information in the infinite-horizon model is included in the accuracy cost function. The saturating shape of the accuracy cost around the target is essential for reproducing the symmetric speed profile and the amplitude-duration scaling of ballistic reaching movements without visual feedback. The same saturating accuracy cost, necessary for reproducing the symmetric speed-profile, correctly predicts the observed scaling relationship between movement amplitude and duration (Fig 3.3). The accuracy cost also links the infinite-horizon model to several well known finite-horizon models.

Additional support for the saturating accuracy term comes from a psychophysics experiment showing that humans tend to perceive accuracy error in a similar, saturating way. [Koerding and Wolpert 2004]. In these experiments, subjects in a reaching experiment internally represented both the statistical distribution of the task and their sensory uncertainty. The human data were best fit by an inverted Gaussian saturating-accuracy cost function. The infinite-horizon model provides a theoretical explanation for why this shape may be necessary for generating motor behavior.

The width, δ, of the accuracy-cost function is related to the target size, or equivalently, the required accuracy level, which appears as a scaling parameter in Fitts's Law for speed-accuracy trade-off. In the relaxed terminal-state formulation, δ is related to the size of the undershoot bias. The saturating cost is also consistent with Linear-Exponential-Quadratic-Gaussian (LEQG) family of widely used risk-sensitive control models [Bryson and Ho 1975], in which δ scales the degree of risk-seeking. In contrast, the quadratic state-cost of state-relaxed model eq (3.7) cannot adequately represent target-size, because since quadratic functions are scale-invariant.

The plateau region of the accuracy cost can be interpreted as the motivation level for generating the movements. The movement is initiated under a constant level of motivation until the vicinity of the target state is reached. The accuracy cost of the infinite-horizon model can also be considered as the inverse of a reward function, which is maximized when the target is reached. This makes the infinite-horizon model compatible with biological reward systems, which motivate behaviors based on internal homeostatic state variables such as hunger and thirst. This explanation is consistent with recent findings, which show that the slow movements of Parkinsonian patients are due to decreased motivation, not increased motor noise [Mazzoni, Hristova, and Krakauer 2007]. The motivation level (τ) may be related to the tonic level of dopamine, which is low in patients with Parkinson's disease and high in patients with Huntingtons disease and Tourettes syndrome, who make fast, ballistic movements.

3.4.2 Feedback Control

The solution of the infinite-horizon optimal feedback model is a sensorimotor transformation that computes the control signal as a function of the estimated state of the limb from sensory signals. Patterns of trial-to-trial variability and perturbation experiments have shown that the motor system takes advantage of sensory information during movement generation rather than simply executing a pre-determined sequence of control signals [Todorov and Jordan 2002]. Motor adaptation experiments have revealed that the learned feedback correction is strongly associated with limb positions and velocities rather than time, even when the applied perturbation was a function of time [Conditt and Mussa-Ivaldi 1999; Conditt, Gandolfo, and Mussa-Ivaldi 1997; Diedrichsen, Criscimagna-Hemminger, and Shadmehr 2007; Sing et al. 2009].

In human reaching movements, the speed-profile has a symmetric bell shape if the movements are performed without visual feedback, but the speed profile becomes skewed when the hand is visible [Chua and Elliott 1993]. Apparently, the motor noise is ignored when a visual signal is unavailable, despite the presence of motor noise in both cases. The deterministic model without visual feedback describes human movements remarkably accurately, which might be a default mode for planning movements.

There is no practical method for solving finite-horizon optimal feedback control problems with a variable terminal-time. The source of the difficulty is that a different feedback control strategy is required at each instance of time because of the explicit dependence of the control cost on time. In contrast, the optimal feedback control law for infinite-horizon formulation is time-independent, which can be solved with numerical optimization methods. We used dynamic programming (value iteration), a standard method that uses discretized state-space. For a more realistic, high-dimensional limb-dynamics model, however, such discretized state-space based methods are not practical due to the rapid increase in computational complexity with dimensionality.

An alternative to numerical optimization is to approximate the feedback control law by training an artificial neural network controller using the cost function of the infinite-horizon. For example, a recurrent neural network was trained to control a realistic, two-link limb model to generate reaching movements [Huh, Todorov 2009]. The structure of the neural network model can be further constrained by anatomy to provide intuition for understanding the neural basis for motor control, which is not possible in the finite-horizon model.

Chapter 3, in full is currently being prepared for submission for publication of the material. Huh, Dongsung; Todorov, Emanuel; Sejnowski, Terrence. The dissertation author was the primary investigator and author of this material.

Appendices

3.A Model Formulation

Our optimal control model consists of a dynamic model of the limb and a cost function. In general, the limb (arm) can be modeled as a stochastic dynamical system

$$d\boldsymbol{x} = \boldsymbol{a}(\boldsymbol{x})\, dt + \boldsymbol{b}(\boldsymbol{x}, \boldsymbol{u})\; dt + \Sigma(\boldsymbol{x}, \boldsymbol{u})\, d\boldsymbol{\xi}, \tag{3.12}$$

where $\boldsymbol{x}_t$ is the limb state vector, $\boldsymbol{u}_t$ the control signal, $\boldsymbol{a}(\boldsymbol{x})$ the passive dynamics, and $B(\boldsymbol{x})$ the control-dependent dynamics. Motor noise is represented by $\boldsymbol{\xi}_t$, a vector of independent Wiener processes (i.e. Brownian motion) of zero mean and unit diffusion rate ($\mathbb{E}[d\xi_i] = 0$, $\mathbb{E}[d\xi_i^2] = dt$, $\mathbb{E}[d\xi_i\, d\xi_{j\neq i}] = 0$), and $\Sigma(\boldsymbol{x}, \boldsymbol{u})$ is the noise magnitude matrix.

More specifically, we focus our analysis on a simple linear model

$$d\boldsymbol{x} = A\,\boldsymbol{x}\; dt + B\,u\; (dt + \sqrt{\sigma}\; d\xi), \quad (u, \xi, \sigma \in \mathbb{R}) \tag{3.13}$$

where u_t, ξ_t, $\sigma \in \mathbb{R}$ (they are scalars) and

$$\boldsymbol{x} = \begin{pmatrix} x \\ \dot{x} \\ \ddot{x} \end{pmatrix}, \; A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \; B = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

and $x \in \mathbb{R}$ represents the 1-D end-effector (i.e. hand) position. But, generalizations to other limb models eq (3.12) are straight-forward.

Note that the control signal affects both the average behavior ($\mathbb{E}[d\ddot{x}] = u\, dt$) as well as the trial-to-trial variability ($\mathbb{E}[d\ddot{x}^2] = \sigma\; u^2 dt$). Therefore, the noise-scaling parameter, σ, has the dimensions of $[time]$.

The control-dependent multiplicative noise model reflects the empirical observations that the standard deviation of muscle force grows linearly with the mean force in isometric force tasks Sutton and Sykes 1967; Schmidt et al. 1979; Todorov and Jordan 2002 . Similar noise models have been used in Todorov 2005a; Todorov and Jordan 2002; Harris 1998.

(a) Numerical Optimization

For deterministic cases ($\sigma = 0$), optimal solutions can be obtained by using the Pontryagin's minimum principle. For stochastic cases, we applied a dynamic programming method (called value iteration) on discretized state space to obtain the cost-to-go function and the optimal control policy.

3.B Time-constant analysis for quadratic cost

Here, we analyze the optimal solutions for the generic quadratic cost function eq (3.2). Consider

$$\mathcal{L} = x^2 + a_1\dot{x}^2 + a_2\ddot{x}^2 + a_3u^2 \qquad (x, u, a_i \in \mathbb{R}), \tag{3.14}$$

where $u \equiv \dddot{x}, a_i > 0$. (The coefficient for x^2 can be set to 1 without loss of generality.) Optimal solutions for this problem can be found via the *Euler-Lagrange* equation:

$$\begin{aligned} 0 &= \frac{\partial \mathcal{L}}{\partial x} - \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{x}} + \frac{d^2}{dt^2}\frac{\partial \mathcal{L}}{\partial \ddot{x}} - \frac{d^3}{dt^3}\frac{\partial \mathcal{L}}{\partial u} \\ &= 2\left(1 - a_1\frac{d^2}{dt^2} + a_2\frac{d^4}{dt^4} - a_3\frac{d^6}{dt^6}\right)x, \\ &= 2\left(1 - \tau_1^2\frac{d^2}{dt^2}\right)\left(1 - \tau_2^2\frac{d^2}{dt^2}\right)\left(1 - \tau_3^2\frac{d^2}{dt^2}\right)x, \end{aligned}$$

where the last factorized equation explicitly implies that the optimal solutions are made of exponentially converging terms (the diverging terms are clearly non-optimal and therefore do not appear in the solution). Therefore, the optimal speed-profile will be expressed as

$$\dot{x}_t = \sum_{i=1}^{3} c_i\, e^{-t/\tau_i}, \quad \mathrm{Re}\,\tau_i > 0, \tag{3.15}$$

where the time constants are related to the cost-coefficients as: $a_1 = \tau_1^2 + \tau_2^2 + \tau_3^2$, $a_2 = (\tau_1\tau_2)^2 + (\tau_2\tau_3)^2 + (\tau_3\tau_1)^2$, $a_3 = (\tau_1\tau_2\tau_3)^2$. The control-cost coefficient can also be expressed as $a_3 = \overline{\tau}^6$, where $\overline{\tau} \equiv \sqrt[3]{\tau_1\tau_2\tau_3}$ is the geometric mean. For trajectories that begin from a stationary state (i.e. $\dot{x}_{(0)} = 0, \ddot{x}_{(0)} = 0$), which is typical for reaching movements, the coefficients are:

$$c_1 = \frac{-\tau_1 l}{(\tau_3 - \tau_1)(\tau_1 - \tau_2)}, c_2 = \frac{-\tau_2 l}{(\tau_1 - \tau_2)(\tau_2 - \tau_3)}, c_3 = \frac{-\tau_3 l}{(\tau_2 - \tau_3)(\tau_3 - \tau_1)},$$

(for $\tau_1 \neq \tau_2 \neq \tau_3$), where l is the movement distance (i.e. $x_{(0)} = -l$). (Note that / Then, it can be easily shown that) the initial jerk simply determined by the mean time-constant: $\dddot{x}_{(0)} = {}^{l}/_{\overline{\tau}^3}$. As shown in Fig 3.1, the exponentially converging trajectories may involve oscillatory components and/or have highly skewed speed-profiles — fast accelerating phase followed by slow decelerating phase — depending on the configuration of time-constants. The oscillation disappears if the time-constants are all real. The degree of skewness can be quantified by comparing the slowest time constant with the mean time-constant (i.e. the initial jerk). Then, it can be shown that the speed-profile becomes most symmetric without oscillations when the time constants are all equal: $\tau_1 = \tau_2 = \tau_3 = \overline{\tau}$.

The time-constants affect movement trajectories in both quantitative as well as qualitative ways. The overall temporal pace of movements is modulated by the geometric mean of the time-constants, $\overline{\tau} \equiv \sqrt[3]{\tau_1\tau_2\tau_3}$: larger $\overline{\tau}$ = slower movements. On the other hand, different qualitative convergence patterns may emerge depending on the distribution of time-constants relative to the mean (${}^{\tau_1}/_{\overline{\tau}}$, ${}^{\tau_2}/_{\overline{\tau}}$, ${}^{\tau_3}/_{\overline{\tau}}$). For examples, a wide distribution of time-constants results in a highly skewed speed-profile with a strong initial acceleration phase followed by a weak deceleration phase near the target (over-damped condition), whereas tremor-like oscillations can be caused by large imaginary components (under-damped condition), which may have important implications for certain movement disorders (see Fig 3.1).

The fastest convergence for given initial jerk (i.e. fixed mean time constant) is achieved when the time constants are all equal: $\tau_1 = \tau_2 = \tau_3 = \overline{\tau}$.

Such degenerate case is called *critically-damped*, and is achieved when the cost-coefficients are $a_1 = 3\overline{\tau}^2$, $a_2 = 3\overline{\tau}^4$, and $a_3 = \overline{\tau}^6$ (eq (3.3)). Then, the speed-profile eq (3.15) reduces to

$$\dot{x}_t = \frac{t^2}{2\overline{\tau}^3} e^{-t/\overline{\tau}}. \tag{3.16}$$

3.C Nonlinear Envelope function

$\Theta(y)$ is a monotonic nonlinear envelope function that initially grows linearly for small y: $\Theta(y) \approx y$ for $y \ll 1$.

We found that smallest distortion is achieved if the envelope function satisfies the following *saturating property*,

$$\frac{d\Theta}{dy} = 1 - \Theta(y)^m, \tag{3.17}$$

where $\Theta(0) = 0,\ y \geq 0,\ m \geq 1$.

Examples: $\Theta_{m=1}(y) = 1 - \exp(-y)$, $\Theta_2(y) = \tanh(y)$, $\Theta_\infty(y) = \min(y, 1)$. ($y$ must be a dimensionless input.)

Note that a length constant δ is introduced in order to neutralize the dimensions of the cost eq (3.4), which is necessary since the transcendental envelope function $\Theta(\cdot)$ only allows dimensionless input and output.

Lemma For initial states outside the *target boundary*, optimal solutions of the VD-OFC model are the same as the IHOC solutions (in infinite saturation-rate limit: Θ_∞) with truncation at *target boundary*[1].

In order to prove this lemma, we introduce a generalized accuracy-cost:

$$\Psi_p[y] \equiv \begin{cases} 1 & \text{if } y > p \\ y & \text{otherwise} \end{cases} \tag{3.18}$$

whose minimum with respect to p is the saturating cost $\min_p \Psi_p[y] = \Theta_\infty[y]$, and $p^* = 1$.

Proof Consider optimizing an infinite-horizon problem involving the following total cost

$$\min_{p,\{u\}} \int_0^\infty \left(s_p(x, \dot{x}, \ddot{x}) + u^2\right) dt \tag{3.19}$$

where p is also variable. Full optimization reproduces the IHOC solutions. Alternatively, consider taking minimization in several steps. Notice that eq

[1]For initial states inside the *target boundary*, optimal strategy is to terminate immediately: $T^* = 0$.

(3.19) can be re-expressed as[2]

$$T + \int_0^T u^2 dt + \int_T^\infty \mathcal{L}_c \, dt \,, \tag{3.20}$$

where T is the time when the trajectory crosses the boundary

$$x_T^2 + 3\dot{x}_T^2 + 3\ddot{x}_T^2 = p \ . \tag{3.21}$$

Then, we apply optimization only inside the boundary, which reduces the latter integration in eq (3.21) to $V_{quad}\left(x_T, \dot{x}_T\right)$. Such partial optimization is possible because optimal trajectories that initiate inside the boundary never leaves the boundary. Next, the constraint eq (3.21) is removed during optimization over Q. □■

References

Bernstein, N.A. (1967). *The coordination and regulation of movements.*

Bryson, A.E. and Y.C. Ho (1975). *Applied optimal control.* wiley New York.

Chua, R. and D. Elliott (1993). "Visual regulation of manual aiming* 1". In: *Human Movement Science* 12.4, pp. 365–401.

Conditt, M.A., F. Gandolfo, and F.A. Mussa-Ivaldi (1997). "The motor system does not learn the dynamics of the arm by rote memorization of past experience". In: *Journal of Neurophysiology* 78.1, pp. 554–560.

Conditt, M.A. and F.A. Mussa-Ivaldi (1999). "Central representation of time during motor learning". In: *Proceedings of the National Academy of Sciences* 96.20, p. 11625.

Diedrichsen, J., S.E. Criscimagna-Hemminger, and R. Shadmehr (2007). "Dissociating timing and coordination as functions of the cerebellum". In: *The Journal of neuroscience* 27.23, pp. 6291–6301.

Fitts, P.M. (1954). "The information capacity of the human motor system in controlling the amplitude of movement". In: *Journal of experimental psychology* 47.6, pp. 381–391.

[2]This proof considers only the trajectories that pass through the boundary $x^2 + 2\dot{x}^2 = Q$ exactly once. This is based on two assumptions: (1) It is optimal to enter the boundary at least once, and (2) once entered, it is optimal not to leave the boundary. Apparently, the first assumption is true. The second assumption is true for most intial states, except for a small area near $[x, \dot{x}] = \left[\pm\sqrt{Q}, 0\right]$. This can be resolved by adding a cross term, $1/2 x\dot{x}$, to the quadratic expression in eq 3.18, which has unnoticeable effect on optimal solutions.

Flash, T. and N. Hogan (1985). "The coordination of arm movements: an experimentally confirmed mathematical model". In: *Journal of neuroscience* 5.7, p. 1688.

Harris, C.M. (1998). "On the optimal control of behaviour: a stochastic perspective". In: *Journal of neuroscience methods* 83.1, pp. 73–88.

Harris, C.M. and D.M. Wolpert (1998a). "Signal-dependent noise determines motor planning". In: *Nature* 394.6695, pp. 780–784.

— (2006). "The main sequence of saccades optimizes speed-accuracy trade-off". In: *Biological cybernetics* 95.1, pp. 21–29.

Hoff, B. (1994). "A model of duration in normal and perturbed reaching movement". In: *Biological Cybernetics* 71.6, pp. 481–488.

Hoff, B. and M.A. Arbib (1993). "Models of trajectory formation and temporal interaction of reach and grasp". In: *Journal ol Moior Behauior* 2.3, pp. 175–192.

Koerding, K.P. and D.M. Wolpert (2004). "The loss function of sensorimotor learning". In: *Proceedings of the National Academy of Sciences* 101.26, p. 9839.

Mazzoni, P., A. Hristova, and J.W. Krakauer (2007). "Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation". In: *Journal of Neuroscience* 27.27, p. 7105.

Pandy, MG and BA Garner (1995). "Optimal Control ot Non-ballistic Muscular Movements: A Constraint-Based Performance Criterion for Rising From a Chair". In: *Journal of biomechanical engineering* 117, p. 15.

Schmidt, R.A. et al. (1979). "Motor-output variability: A theory for the accuracy of rapid motor acts." In: *Psychological Review* 86.5, p. 415.

Sing, G.C. et al. (2009). "Primitives for motor adaptation reflect correlated neural tuning to position and velocity". In: *Neuron* 64.4, pp. 575–589.

Stein, RB, MN Oguztoreli, and C. Capaday (1986). "What is optimized in muscular movements". In: *Human muscle power*, pp. 131–150.

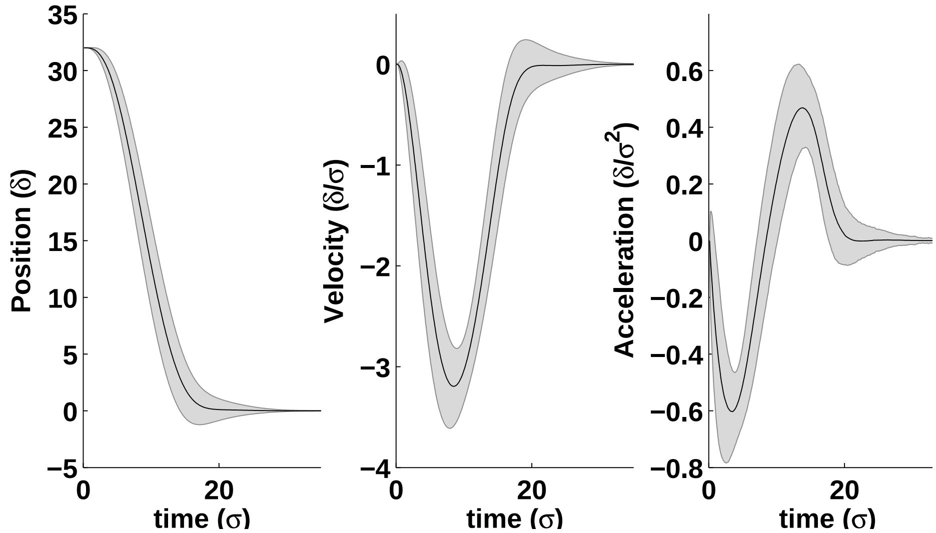
Sutton, GG and K. Sykes (1967). "The variation of hand tremor with force in healthy subjects". In: *The Journal of physiology* 191.3, pp. 699–711.

Todorov, E. (2004). "Optimality principles in sensorimotor control (review)". In: *Nature neuroscience* 7.9, p. 907.

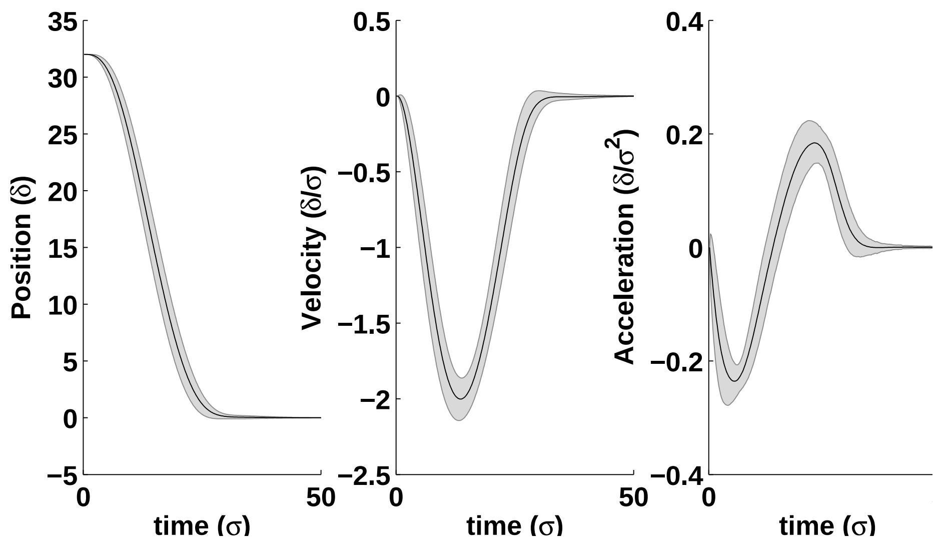
— (2005a). "Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system". In: *Neural computation* 17.5, pp. 1084–1108.

Todorov, E. and M.I. Jordan (2002). "Optimal feedback control as a theory of motor coordination". In: *Nature neuroscience* 5.11, pp. 1226–1235.

Uno, Y., M. Kawato, and R. Suzuki (1989). "Formation and control of optimal trajectory in human multijoint arm movement". In: *Biological cybernetics* 61.2, pp. 89–101.

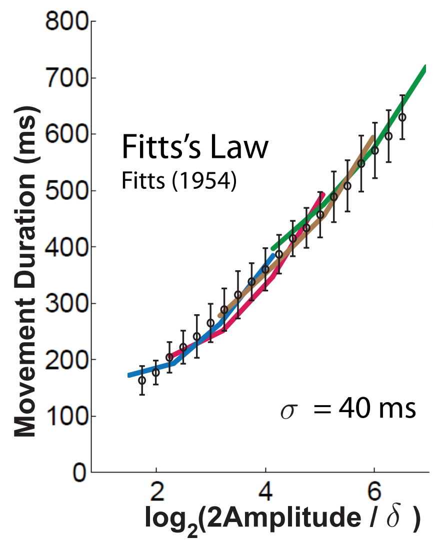


(a) For $\tau = 1.2\sigma$.

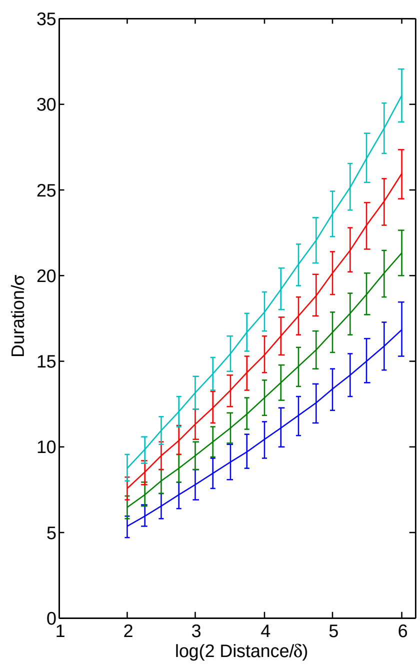


(b) For $\tau = 2\sigma$.

Figure 3.5: Mean and variance of movement trajectories from the initial position $x_0 = 32\delta$.

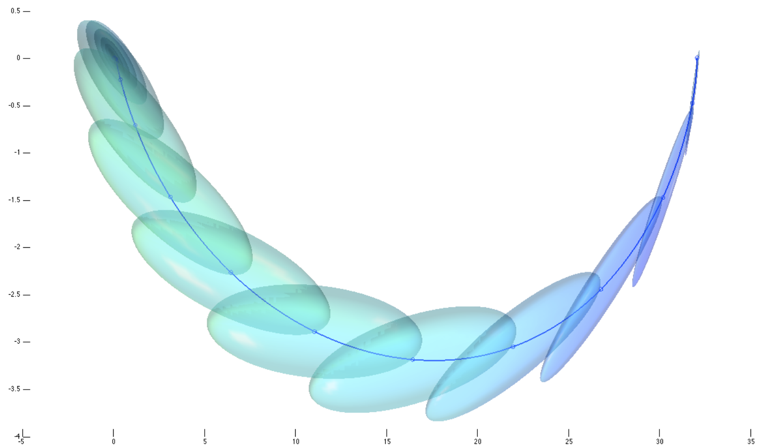


(a) Duration as a function of amplitude and target size. Solid-lines: Data from [Fitts 1954].

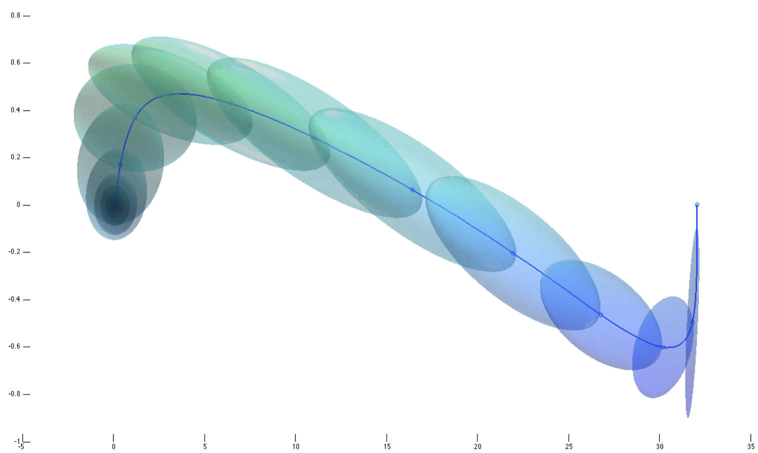


(b) Longer duration is predicted for larger values of τ.

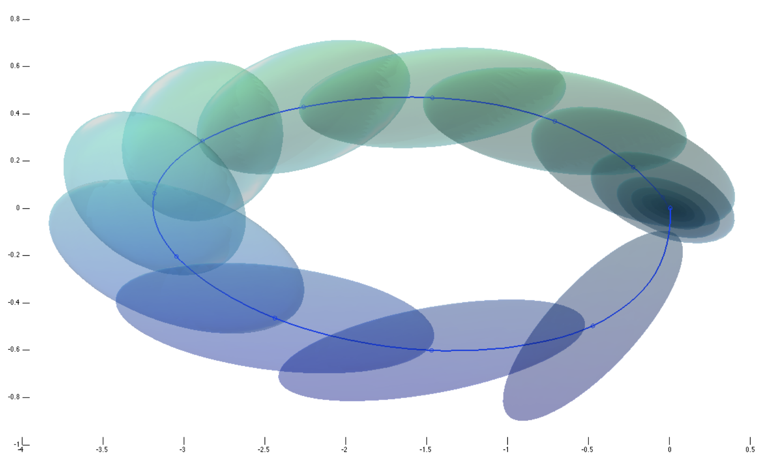
Figure 3.6: Comparison with Fitts' law.



(a) In position-velocity plane.

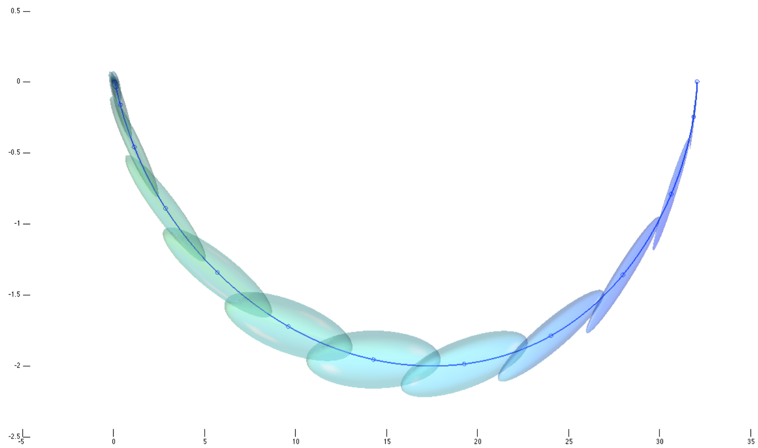


(b) In position-acceleration plane.

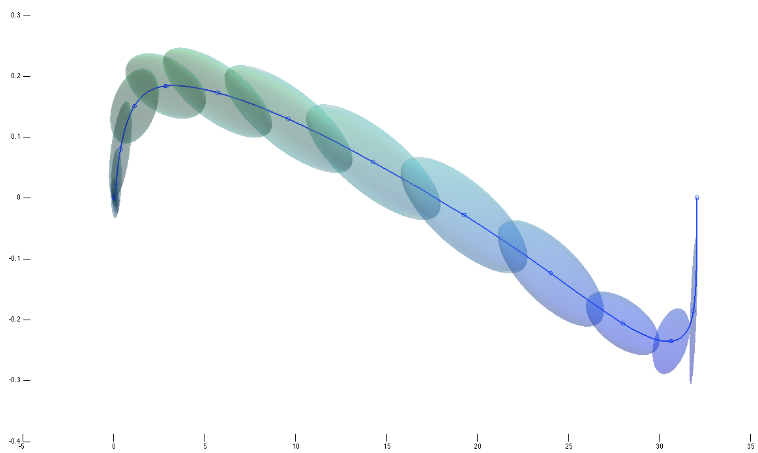


(c) In velocity-acceleration plane.

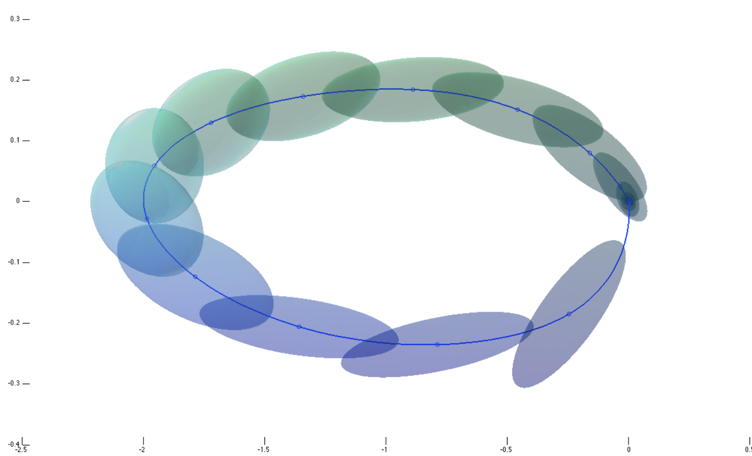
Figure 3.7: 3-D representation of Fig 3.5a



(a) In position-velocity plane.



(b) In position-acceleration plane.



(c) In velocity-acceleration plane.

Figure 3.8: 3-D representation of Fig 3.5b

Chapter 4

Exact Analytic Solutions for Optimal Control Problems Under Multiplicative Noise

Abstract Control-dependent (multiplicative) noise makes it difficult to achieve optimal control because large control signals amplify noise. This paper considers a minimal (one-dimensional) system that includes multiplicative noise and solves the optimal control problem for arbitrary cost functions. In a limit when the control-cost approaches zero, this formulation becomes analytically solvable. The analysis reveals several important properties that are absent in traditional LQ analysis. In particular, multiplicative noise makes optimal solutions depend on the global features of cost functions.

4.1 Introduction

For most physical systems with noisy dynamics, the noise statistics generally depends to some extent on the system state and the control signal rather than being strictly constant (additive noise). Often, such signal dependent noise has significant effects in control systems such as aerospace thrust control, digital feedback control, chemical reactors, analog mechanical actuators, biological muscles, and synaptic transmission of neural systems (see McLane 1971, Ibrahim 1985, Wingerden and De Koning 1984, Wagenaar 1989, Bolotin 1984, Ruan and Choudhury 1993 and Harris and Wolpert 1998b). In these models, noise is often modeled as being amplified by the control signal in a multiplicative manner.

Most control problems with multiplicative noise have been anlayzed on the basis of linear control model with quadratic cost functions on state and control signals (see Wonham 1967, Willems and Willems 1976, Phillis 1985, Kubrusly and Costa 1985, Ruan and Choudhury 1993, Yaz and Skelton 1994, Boyd et al. 1994, El Ghaoui 1995, Lu and Skelton 2000, Todorov 2005b).

Multiplicative noise imposes an additional challenge in optimal control because large control signals amplify noise. For convex state-costs, larger noise means increased average cost, which effectively penalizes large control (effective regularization). Therefore, convex state-costs promotes smaller, risk-aversive control, which is in direct conflict with the original role of state-costs: promoting fast transition and stabilization toward a desired state.

This naturally leads to the following questions:

1. Is it better to use larger or smaller control signals when the state-cost gets steeper? [1]

2. What if the state-cost is not everywhere convex? Is it ideal to be risk-seeking and use larger control where the state-cost is locally concave? [2]

3. Can the effective-regularization due to the state-cost completely substitute the control-cost? [3]

These questions cannot be answered in the tranditional LQ settings (linear-control, quadratic-cost).

In this paper, we investigate a minimal system that includes multiplicative noise and compute optimal solutions for arbitrary state-cost functions. Especially,

[1] *Quick Answer: Smaller*
[2] *Quick Answer: No*
[3] *Quick Answer: Depends on the state-cost.*

we focus on a limiting case when the control-cost approaches zero, which reveals the trade-off between the state-cost and the effective regularization in a most clear manner.

Organization of the paper

Section 4.2 describes the optimal control problem. An uncomplete solution is given in section 4.3. Section 4.4 provides additional infomation necessary for completing the solution. Sections 4.5–4.9 describe complete solutions for various classes of state-costs. Section 4.11 generalizes the results to the cases when control-cost exists. Section 4.12 concludes the paper.

4.2 Problem Description

4.2.1 Deterministic problem

Let us begin with describing a deterministic control problem. Consider a one-dimensional dynamical system

$$\dot{x} = f(x, u), \tag{4.1}$$

which is assumed to have a unique inverse-dynamics

$$u = f^{-1}(x, \dot{x}), \tag{4.2}$$

where $x \in \mathbf{R}$ is the state, and $u \in \mathbf{R}$ is the control. Then, the cost function, which is typically represented as a function of x and u, can equally well be represented as a function of x and $\dot{x}$:

$$\mathcal{L}(x, \tau\dot{x}), \tag{4.3}$$

where τ — a constant with dimension of *time* — is shown explicitly. Such time constant is required for dimensional consistency within the cost function. τ can be regarded as a weight parameter for the control-cost, because it turns out to inversely scale the optimal state-transition-rate. (See section 4.11.)

4.2.2 Stochastic problem with multiplicative noise

In this paper, we consider a stochastic problem in which control-dependent (multiplicative) noise is introduced to the dynamics:

$$dx = f(x,u) \cdot (dt + \sqrt{\sigma} \cdot d\xi), \tag{4.4}$$

where $\xi(t) \in \mathbf{R}$ is a standard Wiener process ($E[d\xi] = 0$, $E[d\xi^2] = dt$), and $\sqrt{\sigma}$ is the diffusion-rate constant — with dimension of $\sqrt{time}$. We call

$$\mu \equiv f(x,u) \tag{4.5}$$

the average state-transition-rate, or simply *drift*.

The goal is to minimize the expectation of total integrated cost (without temporal discount)

$$E_{\xi(\cdot)}\left[\int_0^\infty \mathcal{L}(x(t), \tau\mu(t))dt\right]. \tag{4.6}$$

In this stochastic setting, the effect of control at one state tends to diffuse out to a range of nearby states. The extent of the diffusion is charaterized by the ratio of time constants: $\beta \equiv \sigma/\tau$. In deterministic limit ($\beta \ll 1$), the control problem becomes highly localized. For larger β, the problem gets much harder to solve, as it requires consideration of longer-range effects of control. (See section 4.11.)

However, the interesting case is the limit $\beta \gg 1$ (or $\tau \to 0$), in which the characteristics of control-dependent noise is most emphasized. Ironically, this limit turns out to simplify the control problem, because the cost-function reduces to a pure state-cost (without control-cost):

$$Q(x) \equiv \mathcal{L}(x, 0). \tag{4.7}$$

In following sections, we analytically solve the control problem for arbitrary state-costs in the zero-control-cost limit ($\tau \to 0$), and extend the analysis to finite β cases in section 4.11.

4.3 Hamilton-Jacobi-Bellman Equation

We assume the state-cost $Q(x)$ does not diverge to infinity for finite x and that it has a unique global minimum (which will be removed in section 4.9). Without loss of generality, we set the global minimum to be zero at $x = 0$:

$$\min_x Q(x) = Q(0) = 0.$$

Evidently, optimal drift at $x = 0$ will be zero, since it stops accumulation of any further cost. Therefore, the problem is equivalent to a *first-exit* problem whose *exit-state* — where state-transition permanently ends — is $x = 0$.

Solving stochastic optimal control problems typically involves *cost-to-go*, the minimum total expected cost:

$$V(x) \equiv \min_{\mu(\cdot)} E_{\xi(\cdot)} \left[\int_0^\infty Q(\tilde{x}(t))dt \right], \tag{4.8}$$

where $\tilde{x}(t)$ is a stochastic trajectory with initial state $\tilde{x}(0) = x$.

For stochastic first-exit problems, cost-to-go satisfies the following Hamilton-Jacobi-Bellman (HJB) equation

$$0 = Q(x) + \min_\mu \left[\mu(x)V_x(x) + \frac{1}{2}\sigma\mu(x)^2 V_{xx}(x) \right], \tag{4.9}$$

where subscripts represent differentiations. This equation is non-trivial to solve because the minimum operation makes it a non-linear differential equation.

Once cost-to-go is obtained, optimal drift can be found in a closed form (from the minimum operation of (4.9)):

$$\mu^*(x) = -\frac{V_x(x)}{\sigma V_{xx}(x)}, \tag{4.10}$$

and optimal control law follows from inverse-dynamics

$$u^*(x) = f^{-1}(x, \mu^*(x)). \tag{4.11}$$

Note that cost-to-go must be convex ($V_{xx} \geq 0$), since otherwise the minimum in (4.9) does not exist.

4.3.1 Solution of the nonlinear HJB equation

Substituting (4.10) to HJB (4.9), we obtain a nonlinear differential equation

$$Q = \frac{{V_x}^2}{2\sigma V_{xx}}. \tag{4.12}$$

However, if we take the reciprocal of (4.12), this becomes a linear differential equation

$$\frac{1}{2\sigma Q} = \frac{V_{xx}}{{V_x}^2} = -\frac{d}{dx}\left(\frac{1}{V_x}\right) \quad \text{for } x \neq 0, \tag{4.13}$$

and $1/V_x$ can be calculated by a simple integration

$$\frac{1}{V_x(x)} = \int_x^\infty \frac{dx'}{2\sigma Q(x')} + c \quad \text{for } x > 0, \tag{4.14}$$

where c is an integration constant, yet undetermined.

Then, according to (4.10) and (4.12), optimal drift can be calculated as the ratio between cost and cost-to-go derivative:

$$\mu^*(x) = -\frac{2Q(x)}{V_x(x)} \tag{4.15}$$

$$= -\frac{Q(x)}{\sigma}\left(\int_x^\infty \frac{dx'}{Q(x')} + 2c\right) \qquad x > 0. \tag{4.16}$$

Solution for $x < 0$ can be obtained in a similar manner. In this paper, however, we shall only consider solutions for $x > 0$ for simplicity. Readers may assume state-costs being symmetric in x, while $V_x(x)$ and $\mu^*(x)$ anti-symmetric.

The solution obtained above will be complete once the integration constant is determined. Initially, we guessed this could be done by further analyzing the HJB eq (4.9). Indeed a useful result can be derived from the convexity property implied in (4.9):

Lemma 4.3.1 *c must be non-negative ($c \geq 0$).*

Proof 4.3.1 *Remind that cost-to-go must be convex. According to (4.14), however, $1/V_x$ is a monotonically decreasing function that approaches c as $x \to \infty$. For the convexity condition to be satisfied, V_x must be non-negative for $x > 0$. Therefore, $c \geq 0$.*

□

Unfortunately, this is as far as one can go with HJB equation alone. The truth is, it does not contain sufficient information to uniquely determinine the cost-to-go function. To fill in the missing information, other properties of cost-to-go must be investigated as well.

4.4 Fundamental Properties of $V(x)$

In this section, we enumerate several essential properties of cost-to-go that follow from the definition (4.8). Some are very simple, and do not require proofs.

Proposition 4.4.1 *(Axes rescaling)*
If two cost functions are related by

$$Q_2(x) = Q_1\left(\frac{x}{b}\right) \qquad b > 0, \tag{4.17}$$

then the same holds for cost-to-go functions, $V_2(x) = V_1(x/b)$, and their derivative are

$$V_{2x}(x) = \frac{1}{b}V_{1x}\left(\frac{x}{b}\right). \tag{4.18}$$

Similarly, if two cost functions are related by multiplication by a positive constant

$$Q_2(x) = aQ_1(x) \qquad a > 0. \tag{4.19}$$

then the same holds for cost-to-go functions, $V_2(x) = aV_1(x)$, and theirs derivative are

$$V_{2x}(x) = aV_{1x}(x). \tag{4.20}$$

These properties derives from the fact that a control problem does not change when the units change (*e.g.* from meters to feet). In other words, a control problem is invariant to axis re-scaling.

Lemma 4.4.1 *If two cost functions are related by*

$$Q_1(x) \leq Q_2(x) \qquad \forall x > 0, \tag{4.21}$$

then the same holds for cost-to-go functions: , i.e.

$$V_1(x) \leq V_2(x) \qquad \forall x > 0. \tag{4.22}$$

Proof 4.4.1 *Given a drift function $\mu(x)$ and a cost function $Q_i(x)$, define $\tilde{V}_i(x;\mu)$ as the expected total cost:*

$$\tilde{V}_i(x;\mu) \equiv E\left[\int_0^\infty Q_i(\tilde{x}(t))dt\right] \qquad \tilde{x}(0) = x. \tag{4.23}$$

Then, $\tilde{V}_1(x;\mu) \leq \tilde{V}_2(x;\mu)$, since

$$E\left[\int_0^\infty \left(Q_1(\tilde{x}(t)) - Q_2(\tilde{x}(t))\right) dt\right] \leq 0. \tag{4.24}$$

Then, $V_1(x) = \min_\mu \tilde{V}_1(x;\mu) \leq \tilde{V}_1(x;\mu_2^*) \leq \tilde{V}_2(x;\mu_2^*) = V_2(x)$, where μ_2^* is the optimal drift for Q_2. □

Proposition 4.4.2 *(Shifting and zero-padding)*
Consider a cost $Q_1(x)$ and its cost-to-go $V_1(x)$. Then, for the shifted and zero-padded cost

$$Q_2(x) = \begin{cases} 0 & 0 \leq x \leq x_0 \\ Q_1(x - x_0) & x_0 < x \end{cases} \tag{4.25}$$

the resulting cost-to-go gets shifted and zero-padded by the same amount — x_0:

$$V_2(x) = \begin{cases} 0 & 0 \leq x \leq x_0 \\ V_1(x - x_0) & x_0 < x \end{cases} \tag{4.26}$$

Lemma 4.4.2 *If $c > 0$, $V(x)$ is upper-bounded by x/c.*

Proof 4.4.2 *Since $V_x(x)$ is a monotomically increasing function and $\lim_{x\to\infty} V_x(x) = 1/c$, $V_x(x)$ is upper-bounded by $1/c$. And $V(0) = 0$.* □

4.5 Monomial Cost

For a (superlinear) monomial cost-function

$$Q_1(x) = x^{1+\epsilon} \qquad \epsilon > 0,\ x > 0, \tag{4.27}$$

cost-to-go derivative (4.14) is

$$\frac{1}{V_{1x}(x)} = \frac{1}{2\sigma\epsilon x^{\epsilon}} + c\ . \tag{4.28}$$

Here, we use the invariant property (proposition 4.4.1) to determine the integration constant.

Lemma 4.5.1 *For monomial costs, $c = 0$.*

Proof 4.5.1 *Consider two positive constants a, b that are related by $a = b^{1+\epsilon}$. Then, a re-scaled cost $Q_2(x) \equiv aQ_1(x/b)$ is identical to the original cost $Q_1(x)$:*

$$Q_2(x) = a(x/b)^{1+\epsilon} = x^{1+\epsilon}. \tag{4.29}$$

Then, it follows the cost-to-go's should also be identical: $V_2(x) = V_1(x)$. According to (4.18) and (4.20), however,

$$\begin{aligned} \tfrac{1}{V_{2x}(x)} &= \tfrac{b}{a}\left(\tfrac{1}{2\sigma\epsilon(x/b)^{\epsilon}} + c\right) \\ &= \tfrac{1}{V_{1x}(x)} + c\left(\tfrac{1}{b^{\epsilon}} - 1\right). \end{aligned} \tag{4.30}$$

Since b is arbitrary, $\therefore c = 0$. □

Therefore, we obtain following cost-to-go

$$V_1(x) = \frac{2\sigma\epsilon}{1+\epsilon} x^{1+\epsilon}, \tag{4.31}$$

and its derivative

$$V_{1x}(x) = 2\sigma\epsilon x^{\epsilon}, \tag{4.32}$$

Intriguingly, optimal drift is linear for a monomial cost of any (superlinear) power:

$$\mu_1^*(x) = -\frac{2Q_1(x)}{V_{1x}(x)} = -\frac{x}{\sigma\epsilon}. \tag{4.33}$$

Moreover, the slope of optimal drift is inversely proportional to the power ϵ. This is the first evidence that a steeper state-cost leads to a slower drift.

Note that in the limit $\epsilon \to 0$ (linear monomial), cost-to-go and its derivative (4.31,4.32) collapse to zero and optimal drift (4.33) diverges to infinity.

4.6 Superlinear Cost

In this section, we generalize lemma 4.5.1 to a more general class of cost-functions.

Definition: A cost-function $Q_1(x)$ is *superlinear*, if
$\exists a, \epsilon > 0$, $\exists x_0 \geq 0$ s.t.

$$Q_1(x) \geq Q_2(x) \qquad \forall x > 0, \tag{4.34}$$

where $Q_2(x)$ is a shifted monomial

$$Q_2(x) = \begin{cases} 0 & 0 \leq x \leq x_0 \\ a(x - x_0)^{1+\epsilon} & x_0 < x \end{cases} . \tag{4.35}$$

Lemma 4.6.1 *For superlinear costs, $c = 0$.*

Proof 4.6.1 *If $c > 0$, $V_1(x)$ is upper-bounded by a linear function — x/c (lemma 4.4.2). Since $Q_1(x) \geq Q_2(x)$, however, $V_1(x)$ is lower-bounded by $V_2(x)$ (lemma 4.4.1), which grows superlinearly — eq (4.31). $\therefore c = 0$ by contradiction.*
$\square$

With lemma 4.6.1, the results from monomial cost-functions readily generalize to "polynomial" cost-functions:

$$\frac{1}{Q(x)} = \sum_i \frac{a_i}{x^{1+\epsilon_i}} \qquad \epsilon_i > 0, \tag{4.36}$$

which covers a much broader class of cost-functions. It is not necessary for all coefficients a_i to be positive, but $Q(x)$ must be positive for all $x > 0$. Cost-to-go derivative is

$$\frac{1}{V_x(x)} = \frac{1}{2\sigma} \sum_i \frac{a_i}{\epsilon_i x^{\epsilon_i}} \tag{4.37}$$

and optimal drift

$$\mu^*(x) = -\frac{\sum_i a_i/(\epsilon_i x^{\epsilon_i})}{\sigma \sum_i a_i/x^{1+\epsilon_i}}. \tag{4.38}$$

More generally, the following compositionality theorem holds for all superlinear cost-functions.

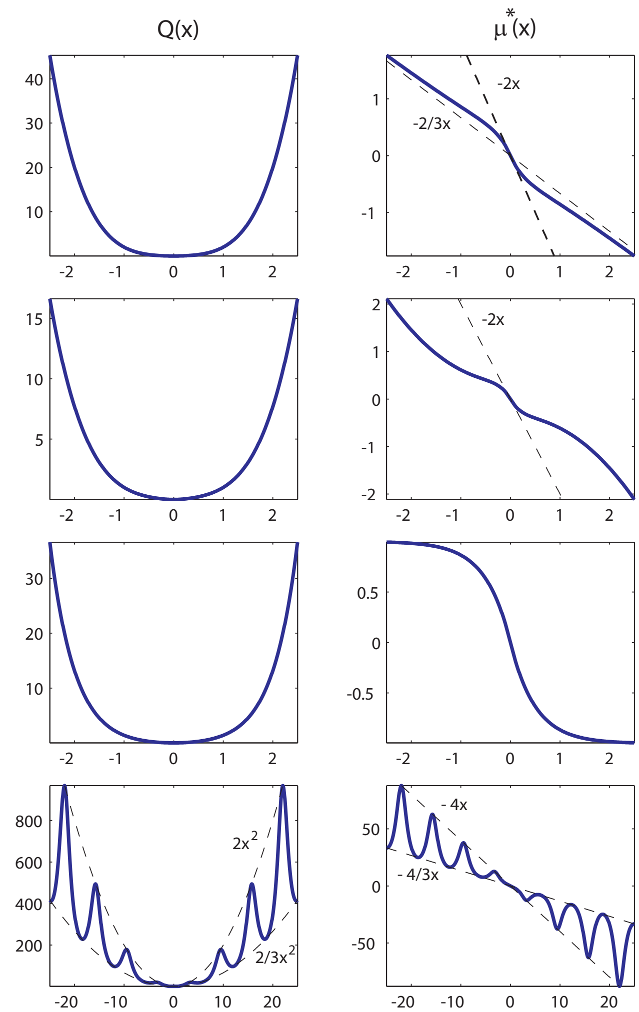


Figure 4.1: Examples of superlinear costs. The cost functions and the optimal drift functions are shown in the same order as in Table 4.1. $\sigma = 1$

Theorem 4.6.1 *(Compositionality Rule)*

If three superlinear cost functions are related by

$$\frac{1}{Q_3(x)} = \frac{a_1}{Q_1(x/b_1)} + \frac{a_2}{Q_2(x/b_2)} \ , \tag{4.39}$$

then cost-to-go derivative are related by

$$\frac{1}{V_{3x}(x)} = \frac{a_1 b_1}{V_{1x}(x/b_1)} + \frac{a_2 b_2}{V_{2x}(x/b_2)} \ . \tag{4.40}$$

Proof 4.6.2 *Because (4.13) is linear and integration constants are identically zero for all three cost functions.* □

Of course, a cost function need not be represented in a "polynomial" form to be solved. Table 4.1 and Fig 4.1 shows some examples. Note that optimal drift functions are no longer linear. Instead, they reflect the shape of state-cost. Also note that non-convex state-costs do not lead to infinite, risk-seeking control. All superlinear state-cost leads to risk-aversive control.

4.7 Sublinear Cost

At the end of section 4.5, we showed that for linear state-costs, cost-to-go collapses to zero and optimal drift diverges to infinity: a linear cost is not steep

Table 4.1: Examples shown in Fig 4.1

$Q(x)$	$\frac{V_x(x)}{2\sigma}$	$\sigma\mu^*(x)$
$x^2 + x^4$	$\frac{1}{\frac{1}{x} + \tan^{-1}(x) - \frac{\pi}{2}}$	$-\frac{(x+x^3)}{(1+x(\tan^{-1}(x)-\frac{\pi}{2}))^{-1}}$
$\frac{(x+x^3)^2}{1+3x^2}$	$x + x^3$	$-\frac{x+x^3}{1+3x^2}$
$\sinh(x)^2$	$\frac{\tanh(x)}{1-\tanh(x)}$	$-\frac{1}{2}(1-\exp(-2x))$
$\frac{(x+\sin(x)/2)^2}{1+\cos(x)/2}$	$x + \sin(x)/2$	$-\frac{x+\sin(x)/2}{1+\cos(x)/2}$

enough to provide sufficient self-regularization. The same is true for all sublinear state-costs.

Definition: A cost function $Q_1(x)$ is *sublinear*, if
$\exists a > 0, \exists b \geq 0$ (yet finite), s.t.

$$Q_1(x) \leq ax + b \qquad \forall x \geq 0. \tag{4.41}$$

(As one can see, this includes linear costs as well.)

Lemma 4.7.1 *For sublinear costs, cost-to-go is zero.*

Proof 4.7.1 *Define $Q_2(x)$ as the shifted and zero-padded version of $Q_1(x)$ by $x_0 = b/a$ (as in (4.25)). Then according to proposition 4.4.2*

$$Q_2(x) \leq ax \qquad \forall x \geq 0. \tag{4.42}$$

Since cost-to-go for a linear cost is zero, $V_2(x)$ is upperbounded by zero (lemma 4.4.1). $\quad \therefore V_1(x) = V_2(x) = 0.$ $\square$

The following example shows the process of cost-to-go collapsing as the state-cost becomes sublinear.

4.7.1 Example: Saturating cost

Consider the following cost

$$Q(x) = \frac{x^2}{(a+1) - a\left(\frac{\sqrt{3}x/\sqrt{a}}{\sinh(\sqrt{3}x/\sqrt{a})}\right)^2} \qquad a \geq 0, \tag{4.43}$$

which is in fact composed of $Q_1(x) = x^2$ and $Q_2(x) = \sinh^2(x)$ as in (4.39). In the limit $a \to \infty$, this becomes a saturating function

$$Q(x) \approx x^2/(1 + x^2). \tag{4.44}$$

According to theorem 4.6.1, the cost-to-go derivative is

$$V_x(x) = \frac{2\sigma x}{(a+1) - a\left(\frac{\sqrt{3}x/\sqrt{a}}{\tanh(\sqrt{3}x/\sqrt{a})} - \frac{\sqrt{3}x}{\sqrt{a}}\right)}\ . \tag{4.45}$$

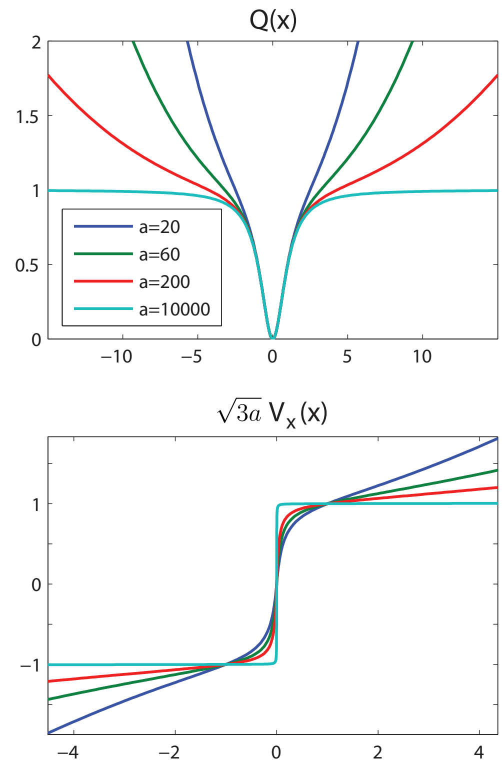


Figure 4.2: Saturating cost example in section 4.7.1: Plot of $Q(x)$ (4.43) and $V_x(x)$ derivative (4.45) for $a = 20, 60, 200, 10000$. Note that $V_x(x)$ is multiplied by $\sqrt{3a}$.

In the limit $a \to \infty$, this becomes

$$V_x(x) \approx 2\sigma x/(1 + \sqrt{3a}x) \approx 0. \tag{4.46}$$

Notice that as a grows large, $V_x(x)$ becomes a step function with amplitude $1/\sqrt{3a}$. This is shown in Fig 4.2. Also note that although $Q(x)$ remains unchanged in the region $-2 \le x \le 2$, $V_x(x)$ in the same region goes through a drastic change as a increases. This is because the optimal solution depends not only on local features but on global shape of state-cost.

4.8 Almost-linear Cost

Previous sections showed that superlinear state-costs provide sufficient self-regularization, but sublinear (and linear) state-costs do not. The difference originates from the convergence property of the integration (4.14).

However, there exists a small class of cost-functions for which such distinction is impossible.

Definition: A cost $Q(x)$ is *almost-linear*, if
$\exists \alpha, \beta > 0$ s.t.

$$\alpha x < Q(x) < \beta x^{1+\epsilon} \qquad \forall x, \epsilon > 0 \tag{4.47}$$

(after appropriate shifting and zero-padding).

4.8.1 Examples

Consider two *almost-linear* cost-functions

$$Q_1(x) = (1+x) \cdot \log(1+x) \tag{4.48}$$

$$Q_2(x) = (1+x) \cdot \log^2(1+x). \tag{4.49}$$

Note that integration of $1/Q_1$ diverges

$$\int_x^\infty \frac{dx'}{Q_1(x')} = \Big[\log(\log(1+x)) \Big]_x^\infty = \infty. \tag{4.50}$$

Despite the subtle difference, however, integration of $1/Q_2$ converges and returns the following optimal solution,

$$\begin{aligned} V_x(x) &= 2\sigma^2 \log(1+x), \\ \mu\,(x) &= -\tfrac{1}{\sigma^2}(1+x) \cdot \log(1+x). \end{aligned} \tag{4.51}$$

where we assumed $c = 0$, yet without a formal proof.

Therefore, convergence property of *almost-linear* cost-functions can only be confirmed on case-by-case basis.

4.9 Cost with Multiple Global Minima

In this section, we extend our results to state-costs with multiple global minima. Let us begin with an example.

Consider the state-cost

$$Q(x) = (x^2 - 1)^2,$$

which is zero at $x = \pm 1$. Integrating $1/Q$ returns the following cost-to-go derivative

$$V_x(x) = \frac{8\sigma}{\frac{2x}{x^2-1} + \log\left(\frac{x-1}{x+1}\right)}, \tag{4.52}$$

which is a well-behaved real function for $|x| > 1$, but becomes complex for $-1 \leq x \leq 1$. In fact, this results follows from setting $c = 0$, which lacks justification for $-1 \leq x \leq 1$. A new principle is needed for state-space between global minima.

Lemma 4.9.1 *If a state-cost has multiple global minima $x_1, \cdots, x_n$ s.t. $Q(x_i) = 0$, then*

$$V(x) = 0 \quad x_1 \leq x \leq x_n \tag{4.53}$$

Proof 4.9.1 *(1) $V(x_1) = V(x_n) = 0$, (2) $V(x) \geq 0, \forall x$, and (3) $V(x)$ must be convex everywhere.* □

This lemma implies absence of self-regulation between global minima. In this case, optimal drift diverges to infinity in the direction of the nearby global minimum.

4.10 Uniqueness of Non-smooth Solutions

All the examples shown up to now have smooth, differentiable cost-to-go solutions. However, solutions can be non-smooth for some cost functions.

4.10.1 Example

Consider the cost-function

$$Q(x) = x^2 + \sqrt{|x|}. \tag{4.54}$$

The corresponding cost-to-go derivative is

$$V_x(x) = \frac{3}{\log\left(\frac{x-\sqrt{x}+1}{x+2\sqrt{x}+1}\right) + 2\sqrt{3}\left(\pi/2 - \tan^{-1}\left(\frac{2\sqrt{x}-1}{\sqrt{3}}\right)\right)} \tag{4.55}$$

for $x > 0$. Notice that the cost-to-go derivative does not become zero in the limit $x \to 0^+$, but instead $V_x(x) \to \sqrt{27}/4\pi$. Since $V_x(x)$ is anti-symmetric in x, this solution is non-smooth at $x = 0$.

In general, one must be cautious when admitting non-smooth functions as potential solutions of a HJB equation, because this relaxed notion may introduce an infinite number of possible solutions. The problem of choosing a unique non-smooth solution is usually solved by introducing *viscosity condition*. For the control problems here considered, it can be shown that with all the conditions mentioned in section 4.4, uniqueness of solution is indeed guaranteed by allowing non-differentiability only at $x = 0$, and that these are indeed *viscosity solutions*.

4.11 Generalization to Finite β using Perturbation Theory

In this section, we extend the analysis to the cases when control-cost is non-zero. However, the analysis becomes very difficult when β is finite, and in fact impossible for most forms of control-cost.

Here we focus on quadratic control-cost. Consider

$$\mathcal{L}(x, \mu) = Q(x) + \frac{1}{2}(\tau\mu)^2. \tag{4.56}$$

for which the HJB equation is

$$0 = Q(x) + \min_{\mu} \left[\mu V_x(x) + \frac{1}{2}\left(\tau^2 + \sigma V_{xx}(x)\right)\mu^2 \right]. \tag{4.57}$$

After explicit minimization, this simplifies to

$$\frac{1}{2Q} = \frac{\tau^2}{{V_x}^2} - \sigma \frac{d}{dx}\left(\frac{1}{V_x}\right). \tag{4.58}$$

where optimal drift is

$$\mu^*(x) = -\frac{V_x(x)}{\tau^2 + \sigma V_{xx}(x)} = -\frac{2Q(x)}{V_x(x)}. \tag{4.59}$$

Equation (4.58) is non-linear, and therefore it does not allow closed form solutions in general. If β is either very small or very large, however, the solution can be approximated as a perturbation series.

4.11.1 Perturbation Series I

Here we expand the solution near the deterministic limit. Equation (4.58) can be expressed as

$$\frac{1}{2Q} = U^2 - \beta\frac{dU}{dx}, \tag{4.60}$$

where $U \equiv \tau/V_x$ and $\beta \equiv \sigma/\tau$. We express U as a power series of β:

$$U(x) = U_0(x) + \beta U_1(x) + \beta^2 U_2(x) + \cdots ,$$

where U_0 represents the deterministic limit solution, and other U_k terms describe higher-order correction. When this is substituted in (4.60), we obtain

$$0 = \left(U_0^2 - \frac{1}{2Q}\right) + \beta\left(2U_0U_1 - {U_0}'\right)$$

Table 4.2: Summary of Perturbation Series

Expansion I :	Expansion II :
$U(x) = \sum_{k=0}^{\infty}\left(\frac{\sigma}{\tau}\right)^k U_k(x)$ $\mu^*(x) = -2Q(x)U(x)/\tau$	$W(x) = \sum_{k=0}^{\infty}\left(\frac{\tau}{\sigma}\right)^{2k} W_k(x)$ $\mu^*(x) = -2Q(x)W(x)/\sigma$
$U_0 = \frac{1}{\sqrt{2Q}}$	$W_0(x) = \int_x^{\infty} \frac{1}{2Q(x')}dx'$
$U_1 = \frac{{U_0}'}{2U_0}$	$W_1(x) = \int_x^{\infty} -{W_0}^2(x')dx'$
$U_2 = \frac{{U_1}' - {U_1}^2}{2U_0}$	$W_2(x) = \int_x^{\infty} -2W_0W_1dx'$
$U_3 = \frac{{U_2}' - 2U_1U_2}{2U_0}$	$W_3(x) = \int_x^{\infty} -\left({W_1}^2 + 2W_0W_2\right)$
$U_{k+1} = \frac{{U_k}' - \sum_{i,j} U_iU_j}{2U_0}$ $1 \leq i,j \leq k,\ i+j = k+1$	$W_{k+1}(x) = \int_x^{\infty} -\sum_{i,j} W_iW_jdx'$ where $0 \leq i,j \leq k,\ i+j = k$

$$+\,\beta^2({U_1}^2 + 2U_0U_2 - {U_1}') + \cdots$$

Each term in parenthesis must vanish repectively, which leads to a series differential equations that can be solved iteratively. The solution is summarized in Table 4.2.

Note that the higher order terms introduce non-local features of state-cost, since U_i depends on the nth order differentiation of $Q(x)$. Therefore, as β gets larger, the optimal solution will depend more on global features of state-cost.

4.11.2 Perturbation Series II

The solution can also be expanded near the zero-control-cost limit. We express (4.58) as

$$\frac{1}{2Q} = \alpha W^2 - \frac{dW}{dx}, \tag{4.61}$$

where $W \equiv \sigma/V_x$ and $\alpha \equiv 1/\beta^2 = \tau^2/\sigma^4$. and look for the solution of the form

$$W(x) = W_0(x) + \alpha W_1(x) + \alpha^2 W_2(x) + \cdots$$

When this is substituted in (4.61), we find

$$0 = (W_0' + \frac{1}{2Q}) + \alpha(W_1' - {W_0}^2) + \alpha^2(W_2' - 2W_0W_1) + \cdots$$

which gives out a series of 1st-order differential equations that can be easily integrated interatively (see Table 4.2).

Note that in this expansion, it is the lowest order term (W_0) that introduces the global-feature dependence of the solution, while the higher order terms gradually cancel out the global dependence. This is apparent in (4.61), since the αW^2 term acts as a non-linear filter that decays over distance.

Note that $c = 0$ is assumed for each integration, which remains to be proved.

Example: Quadratic State-Cost

For the quadratic cost function,

$$\mathcal{L}(x, \mu) = x^2 + \frac{1}{2}(\tau\mu)^2,$$

the expansion terms are $W_0 = \frac{1}{2x}$, $W_1 = -\frac{1}{4x}$, $W_2 = \frac{1}{4x}$, $W_3 = -\frac{5}{16x}$, $\cdots$, or in general

$$W_k = \frac{(-1)^k}{2x} \frac{(2k-1)!!}{(k+1)!} \ .$$

This perturbation series converges to a closed form

$$\begin{aligned} W(x) &= \tfrac{1}{2x}\left(1 + \textstyle\sum_{k=1}^{\infty} \tfrac{(2k-1)!!}{(k+1)!}(-\alpha)^k\right) \\ &= \tfrac{\sqrt{1+2\alpha}-1}{2\alpha x} \end{aligned} \tag{4.62}$$

which is indeed the solution of (4.61). Optimal drift is

$$\mu^*(x) = -\frac{2x}{\sigma + \sqrt{\sigma^4 + 2\tau^2}}. \tag{4.63}$$

This solution is valid for all range of β.

4.12 Conclusion

This paper considered optimum control of systems with multiplicative noise. Our analysis is vastly different from the traditional linear-quadratic (LQ) analysis: Rather than analyzing a high dimensional system with a simple quadratic cost function, we analyzed a one dimensional system with complex, arbitrary cost functions. As a result, our analysis revealed exotic, even counter-intuitive, relationships between cost functions and the corresponding optimal feedback-control strategies.

Most importantly, we showed that multiplicative noise makes optimal solutions depend on the global features of cost functions, and that the extent of such global dependence is determined by β — the ratio of time constants. Other interesting discoveries include (1) feedback gain must decrease as the state-cost becomes steeper, and (2) optimality of risk-aversion vs risk-seeking strategies depends on superlinearlity of state-cost and not convexity. Traditional LQ anlysis cannot find these intriguing facts because the quadratic cost function is too simple.

Previously, it was suggested that a genearlized, iterative-LQG (linear-quadratic-gaussian) method could approximately solve non-LQG problems with multiplicative noise (Todorov 2005b). However, it was not clear when such methods would fail. As mentioned above, the determining parameter is β. These methods would eventually fail as β gets sufficiently large, since LQG approximation only collects at most second order local information.

We also showed that analysis becomes simple in the large β limit. This limit occurs when the cost of control is negligible compared to the need for system stabilization. The iterative expansion method (expansion II) provides high accuracy solutions in such situations.

Unfortunately, the analytic approach presented here is limited to one-dimensional problems. However, many of the implications revealed by this analysis are expected to generalize to higher-dimensional problems and they could provide insights when solving practical problems.

Chapter 4, in full, is a reprint of the material as it appears in International Federation of Automatic Control - World Congress, 2011. Huh, Dongsung; Sejnowski, Terrence. The dissertation author was the primary investigator and author of this paper.

References

Bolotin, V.V. (1984). *Random vibrations of elastic systems*. Springer. ISBN: 9024729815.

Boyd, S. et al. (1994). *Linear matrix inequalities in system and control theory*. Society for Industrial Mathematics. ISBN: 0898714850.

El Ghaoui, Laurent (1995). "State-feedback control of systems with multiplicative noise via linear matrix inequalities". In: *Syst. Control Lett.* 24 (3), pp. 223–228. ISSN: 0167-6911. DOI: http://dx.doi.org/10.1016/0167-6911(94)00045-W. URL: http://dx.doi.org/10.1016/0167-6911(94)00045-W.

Harris, C.M. and D.M. Wolpert (1998b). "Signal-dependent noise determines motor planning". In: *Nature* 394.6695, pp. 780–784. ISSN: 0028-0836.

Ibrahim, R.A. (1985). *Parametric random vibration*. John Wiley & Sons, London.

Kubrusly, C. and O. Costa (1985). "Mean square stability conditions for discrete stochastic bilinear systems". In: *Automatic Control, IEEE Transactions on* 30.11, pp. 1082–1087. ISSN: 0018-9286.

Lu, J. and R.E. Skelton (2000). “Robust variance control for systems with finite-signal-to-noise uncertainty”. In: *Automatica* 36.4, pp. 511–525. ISSN: 0005-1098.

McLane, P. (1971). “Optimal stochastic control of linear systems with state-and control-dependent disturbances”. In: *Automatic Control, IEEE Transactions on* 16.6, pp. 793–798. ISSN: 0018-9286.

Phillis, Y. (1985). “Controller design of systems with multiplicative noise”. In: *Automatic Control, IEEE Transactions on* 30.10, pp. 1017–1019. ISSN: 0018-9286.

Ruan, M. and A.K. Choudhury (1993). “Steady-state optimal controller with actuator noise variance linearly related to actuator signal variance”. In: *Automatic Control, IEEE Transactions on* 38.1, pp. 133–135. ISSN: 0018-9286.

Todorov, E. (2005b). “Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system”. In: *Neural computation* 17.5, pp. 1084–1108. ISSN: 0899-7667.

Wagenaar, TJ (1989). “Stability and stabilizability of chemical reactors modelled with stochastic parameters”. In: *International Journal of Control* 49.1, pp. 33–44. ISSN: 0020-7179.

Willems, J.L. and J.C. Willems (1976). “Feedback stabilizability for stochastic systems with state and control dependent noise”. In: *Automatica* 12.3, pp. 277–283. ISSN: 0005-1098.

Wingerden, A. van and W. De Koning (1984). “The influence of finite word length on digital optimal control”. In: *Automatic Control, IEEE Transactions on* 29.5, pp. 385–391. ISSN: 0018-9286.

Wonham, W.M. (1967). “Optimal stationary control of a linear systems with state-dependent noise”. In: *SIAM Journal of Control* 5, pp. 486–500.

Yaz, E. and R.E. Skelton (1994). “Parametrization of all linear compensators for discrete-time stochastic parameter systems”. In: *Automatica* 30.6, pp. 945–955. ISSN: 0005-1098.

Chapter 5

Real-Time Motor Control using Recurrent Neural Networks

Abstract Currently, the field of sensory-motor neuroscience lacks a computational model that can replicate real-time control of biological brain. Due to incomplete neural and anatomical data, traditional neural network training methods fail to model the sensory-motor systems. Here we introduce a novel modeling method based on stochastic optimal control framework which is well suited for this purpose. Our controller is implemented with a recurrent neural network (RNN) whose goal is approximating the optimal global control law for the given plant and cost function. We employ a risk-sensitive objective function proposed by Jacobson (1973) for robustness of controller. For maximum optimization efficiency, we introduce a step response sampling method, which minimizes complexity of the optimization problem. We use conjugate gradient descent method for optimization, and gradient is calculated via Pontryagins maximum principle. In the end, we obtain highly stable and robust RNN controllers that can generate infinite varieties of attractor dynamics of the plant, which are proposed as building blocks of movement generation. We show two such examples, a point attractor based and a limit-cycle based dynamics.

5.1 Introduction

A sensory-motor system of the brain implements a global control law (or strategy) which transforms body state and task condition into an appropriate control signal in general behavioral conditions. The global control law is best described in the optimal control framework as an optimal solution for some performance criteria. Most advancement in the field has focused on finding a local linear approximation of the global control law which only solves a single instantiation of control problem at a time. Such local analysis has been very successful in explaining behavioral data in Deltail with minimal number of principles, such as energy minimization and multiplicative motor noise [Todorov 2004; Todorov and Jordan 2002]. When it comes to neural data, however, the local analysis method loses its predictive power. The missing links are the neural network models of sensory-motor systems, which are poorly understood in motor cortex. Once we have a dynamic model of the underlying neural network, the local analysis method can explain the neural data as well as the behavioral data.

Previously, small scale neural networks was used to model simple sensory-motor reflex systems [Fang and Sejnowski 1990; Lockery and Sejnowski 1992]. Training was based on partially known synaptic connectivity and detailed record of input-output (stimulus-neural response) data. However, as we model more complex systems the number of neurons involved and the repertoire of movement grow very large, so

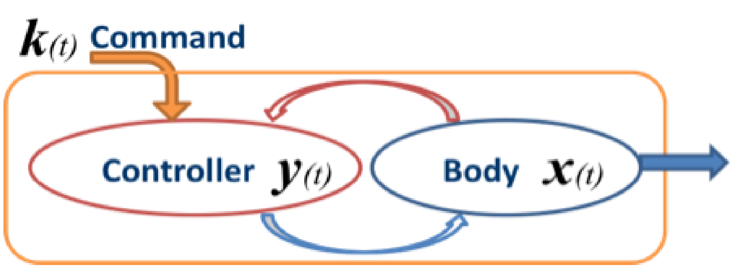


Figure 5.1: Illustration of the combined body-controller dynamics. Body is a dynamical system with state $\boldsymbol{x}(t)$. Controller receives time-varying command (task-information) $\boldsymbol{k}(t)$ and controls body accordingly. The combined body-controller system can be seen as performing a mapping from command sequence $\{\boldsymbol{k}(t)\}$ to body-state sequence $\{\boldsymbol{x}(t)\}$, i.e. movement.

that obtaining detailed neural activity and synaptic connectivity data becomes unrealistic.

Here, we take a radically different approach. Instead of separately obtaining the neural network model and the optimal behavior and later combining them, we directly implement the global control law of the sensory-motor system with a recurrent neural network (RNN) model. The global control law indeed carries enough information to predict both neural activity and underlying synaptic connectivity. This is because the sensory-motor system itself is the product of processes that continuously act to improve behavioral performance.

In the conventional optimal control framework, motor tasks are formulated in finite-horizon setting. This reflects the usual setting of psychophysical experiments, where for each trial a subject must accomplish a fixed motor task during a finite time interval. However, the low-level sensory-motor systems, as well as our RNN controllers, must work in the infinite-horizon setting where the motor task changes over time. The time varying task information (or command) arises internally from high-level decision making areas of the brain. Therefore, it is appropriate to model the low-level sensory-motor area as performing a mapping from command sequence to movement (Fig 5.1).

In section 5.2, we introduce a modified optimal control framework which properly deals with the modified infinite-horizon setting. It shares similarities with model reference control, which we briefly mention in section 5.3. In section 5.4, we explain the ergodic principle, which is used for calculating complexity of the learning problem in section 5.5. In section 5.6, we introduce step response sampling method which minimizes the complexity.

We use conjugate gradient descent method for optimization, where gradient is calculated via Pontryagins maximum principle (section 5.7). In section 5.8, we show two successful applications of the method. Section 5.9 concludes our paper.

5.2 Modified Optimal Control Formulation

In this section, we formulate the mapping problem in the stochastic optimal control framework. It takes a dynamical equation of the body-controller system, a cost function which describes the desired movement, and a probability distribution which describes the stochastic processes. We have modified the framework so that the desired motor task may change as function of high-level command, which is considered as stochastic event. Here, we choose a RNN as a model controller, but this framework applies to any parametric model in general. Note that a mapping problem is translation invariant in time, and therefore our cost function, dynamic equation and probability distribution do not have explicit time dependence.

5.2.1 General Model of Dynamical System

The model system is composed of a body and a controller connected in closed loop. A body is a stochastic dynamical system described by the following set of equations,

$$\begin{aligned} \dot{\boldsymbol{x}} &= \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}) + \text{noise} && \text{state update} \\ \boldsymbol{s} &= \boldsymbol{s}(\boldsymbol{x}) + \text{noise} && \text{output,} \end{aligned} \tag{5.1}$$

and a controller is also a stochastic dynamical system

$$\begin{aligned} \dot{\boldsymbol{y}} &= \boldsymbol{h}(\boldsymbol{y}, \boldsymbol{s}, \boldsymbol{k}; \boldsymbol{W_1}) + \text{noise} && \text{state update} \\ \boldsymbol{u} &= \boldsymbol{u}(\boldsymbol{y}, \boldsymbol{s}, \boldsymbol{k}; \boldsymbol{W_2}) + \text{noise} && \text{output,} \end{aligned} \tag{5.2}$$

where $\boldsymbol{x}_{(t)}$ is the body-state, $\boldsymbol{y}_{(t)}$ is controller-state, $\boldsymbol{s}_{(t)}$ is sensory measurement from body, $\boldsymbol{k}_{(t)}$ is high-level command, $\boldsymbol{u}_{(t)}$ is low-level control signal, and W_1,W_2 are matrices of free parameters. For our RNN implementation of the controller, (5.2) has the following detailed form

$$\begin{bmatrix} \tau\dot{\boldsymbol{y}} + \boldsymbol{y} \\ \boldsymbol{u} \end{bmatrix} = \sigma\left(\begin{bmatrix} \\ \boldsymbol{W} \\ \\ \end{bmatrix} \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{s} \\ \boldsymbol{k} \end{bmatrix} \right), \tag{5.3}$$

where $\boldsymbol{W} = [\boldsymbol{W}_1; \boldsymbol{W}_2]$ is a matrix of free parameters with appropriate dimension. $\boldsymbol{W}$ is referred to as the synaptic weight matrix. The non-linear transfer function $\sigma(\cdot)$ is applied component-wise to vector elements. Note that this structure assumes all-to-all connections between input nodes, neurons, and output nodes.

Since body and RNN always interact in closed loop, it helps to view them as a combined dynamical system with the following dynamical equation,

$$\begin{aligned} \dot{\boldsymbol{q}} = \begin{bmatrix} \dot{\boldsymbol{x}} \\ \dot{\boldsymbol{y}} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}(\boldsymbol{y}, \boldsymbol{s}, \boldsymbol{k}; \boldsymbol{W_2})) \\ \boldsymbol{h}(\boldsymbol{y}, \boldsymbol{s}(\boldsymbol{x}), \boldsymbol{k}; \boldsymbol{W_1}) \end{bmatrix} + \text{noise} \\ &= \boldsymbol{f}(\boldsymbol{q}, \boldsymbol{k}, \boldsymbol{\omega}, \boldsymbol{W}), \end{aligned} \tag{5.4}$$

where $\boldsymbol{q}_{(t)} = [\boldsymbol{x}_{(t)}; \boldsymbol{y}_{(t)}]$ is the combined state vector, and $\boldsymbol{\omega}_{(t)}$ is noise. Note the unusual notation of expressing noise as an explicit input for the dynamics equation. This expression emphasizes the fact that a movement trajectory is determined by the following variables: initial state $\boldsymbol{q}_o$, command sequence $\{\boldsymbol{k}\}$, and noise sequence $\{\boldsymbol{\omega}\}$. The system operates in stochastic environment where these variables are randomly drawn from a distribution $P(\boldsymbol{q}_o, \{\boldsymbol{k}\}, \{\boldsymbol{\omega}\})$.

5.2.2 Value Function and Objective Function

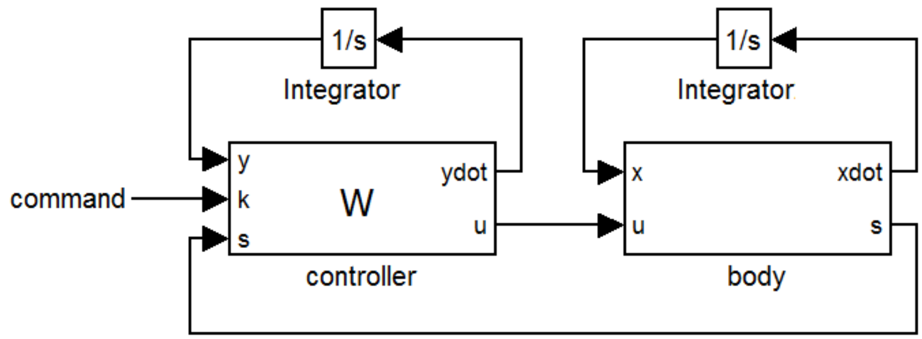
In optimal control framework, a state trajectory $\{\boldsymbol{q}\}$ is evaluated by time-average of instantaneous cost function along the trajectory:

$$V(\boldsymbol{q}_o, \{\boldsymbol{k}\}, \{\boldsymbol{\omega}\}; \boldsymbol{W}) = \frac{1}{T} \int_{\{\boldsymbol{q}\}} c(\boldsymbol{q};\ \boldsymbol{k}, \boldsymbol{W}) dt, \tag{5.5}$$

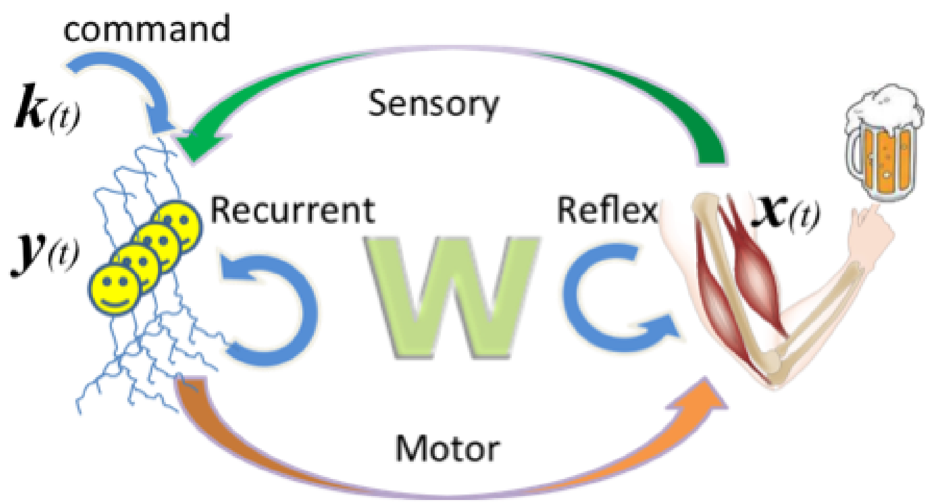
where T is duration of the trajectory, and V is called the value function.

The true objective function of the optimal control problem is the value function of an infinitely long trajectory $V_\infty = \lim_{T\to\infty} V$. However, according to the ergodic principle in section 5.4, the same value function can be obtained by an ensemble average of finite duration value functions $E_P[V]$, if the duration is long enough. Here, we instead use the risk-sensitive objective function introduced by [Jacobson 1973].

$$A(\boldsymbol{W}; \beta) = \frac{1}{\beta} \log E_P[\exp(\beta V)], \tag{5.6}$$



(a) Block diagram of the combined dynamic system: A RNN controller and a body are connected in a closed loop.



(b) Information flow diagram. Information from command signal, sensory signal and neural acvitity gets sent to all neurons and muscles. This information flow is determined by the synaptic weight matrix, W, which we optimize.

Figure 5.2: Illustration of the combined dynamic system.

where β is a positive number that provides a link between optimality and robustness. In statistical physics, A is the Helmholtz free energy and $-1/\beta$ is temperature. In the limit $\beta \to 0$, A reduces to the average cost $E_P[V]$, and for the limit $\beta \to \infty$, A reduces to $\max_P[V]$, the worst case cost. For a finite β, a movement with high cost gets exponentially increasing emphasis. Empirically, we have found that setting β adaptively during optimization leads to both faster optimization and robustness. We describe the adaptive method in section 5.7.1.

For general nonlinear stochastic problems, however, the expectation in (5.6) cannot be computed analytically. Therefore, we approximate it with sample average

$$\tilde{A}(\boldsymbol{W};\beta) = \frac{1}{\beta}\log\left[\frac{1}{N}\sum_{i=1}^{N}\exp(\beta V_i)\right], \tag{5.7}$$

which implies batch training. Here each $V_i(\boldsymbol{W}) \equiv V(\boldsymbol{q}_o i, \{\boldsymbol{k}\}_i, \{\boldsymbol{\omega}\}_i; \boldsymbol{W})$ is a randomly drawn sample from $P(\boldsymbol{q}_o, \{\boldsymbol{k}\}, \{\boldsymbol{\omega}\})$. Depending on the complexity of the problem, a large number of samples may be required. We discuss this issue in section 5.5.

5.3 Comparison with Model Reference Control

Application of RNNs for real-time feedback control of dynamical systems has been already suggested by others. Among several such approaches, a training strategy that shares many similarities with our approach is the *model-reference control* (MRC) [Narendra and Annaswamy 1989; Prokhorov, Puskorius, and Feldkamp 2001; Prokhorov 2007].

In usual MRC settings, there are two non-interacting dynamical systems: a reference model and a combined RNN controller-plant system. They are both driven by a common stream of input called the *reference signal* (which is our command signal) and generate output. The goal of MRC problems is to tune the controller such that the combined RNN-plant dynamics would approximate the reference model dynamics via matching the plant output to the reference model output. Therefore, MRC is a traditional of neural network training method that uses desired input/output pair data, where input is the reference signal and desired output is the reference model output. This approach is appropriate for many engineering situations when the desired dynamics is explicitly known beforehand.

Our optimal control based method can be viewed as a generalization of the MRC method that works even when the reference model dynamics is not directly available in an explicit form, such as modeling biological movements. Here, the reference model dynamics is implicitly expressed as the optimal dynamics for the given cost function and plant dynamics. Therefore, our method combines optimal control and model reference control in a generalized framework.

From now on, the term reference model refers to the optimally controlled body dynamics.

5.4 Ergodic Property of Reference Model

In order for the problem to be solvable in an infinite-horizon setting, the reference models should possess the following properties:

1. Translation invariance in time, which is required in infinite-horizon setting. In our optimal control framework, the followings should be time-invariant: body-RNN dynamics, cost function, and statistical property of random processes. Therefore, we assume that command sequence $\boldsymbol{k}$ and noise sequence $\boldsymbol{\omega}$ are generated by stationary processes. It implies that auto-covariance depends on time difference only. For noise, we simply use zero-mean Brownian white noise $C_{ij}(t_1, t_2) = \delta_{ij}\delta(t_1 - t_2)$. Command sequence is assumed to have finite correlation time constant τ_c, such that its auto covariance decays roughly exponentially $C_{ij}(t_1, t_2) \approx \exp(|t_1 - t_2|/\tau_c)$.

2. Stability, which implies that when the command signal is held fixed at some constant value ($\boldsymbol{k}(t) = \boldsymbol{k}_o$ for $t > t_o$) for long enough time, the reference model state will converge to an attractor, or a set of states with low dimensional structure. Attractors may be zero dimensional (point), one dimensional periodic orbit (limit-cycle), higher dimensional (manifold), or fractal dimensional (chaotic). In this paper, we deal with point attractors and limit-cycle attractors only.

3. Uniqueness, which means that an attractor is uniquely determined by the current value of command input, regardless of past history. Uniqueness property is often violated for higher-level decision-making control problems where the

action space is discrete. But our main focus is in low level sensory-motor transformation with continuous dynamics, and uniqueness property holds here.

The stability and the uniqueness properties, can be used to characterize the *dynamic memory* property of the reference model. According to [Ganguli, Huh, and Sompolinsky 2008], all dynamical systems can store information of recent input signal sequences in their dynamic state vector. Using their Fisher-information based definition, dynamic memory is measured as the past signal's ability to perturb the current state of the system. Since stability and uniqueness properties imply that current state of the reference model becomes insensitive to input signal far enough in the past, the reference model therefore has *decaying memory.* The memory time constant τ_m can be defined to denote the maximum memory duration. The decaying memory property and translation invariance in time imply *ergodic property.* Due to their combined effect, the reference model state $\boldsymbol{q}(t)$ becomes statistically independent from the previous state $\boldsymbol{q}(t')$ if the time difference is greater than $\tau_m + \tau_c$. Then, according to the ergodic principle from statistical physics, time average of any function of $\boldsymbol{q}$ may be substituted with ensemble average. Therefore, the infinite-horizon value function can be obtained from an ensemble average of finite duration value functions as long as their duration is longer than $\tau_m + \tau_c$:

$$V_\infty(\boldsymbol{W}) = E_P[V(\boldsymbol{q}_o, \{\boldsymbol{k}\}, \{\boldsymbol{\omega}\})]$$

where expectation is taken over the stationary distribution $P(\boldsymbol{q}_o, \{\boldsymbol{k}\}, \{\boldsymbol{\omega}\})$. This property is used for the definition of the objective function in (5.6).

5.5 Complexity of the Identification Problem

For the content of this section, it helps to view the optimization problem as a system identification problem, whose goal is identifying the reference model dynamics with the combined dynamical system $\dot{\boldsymbol{q}} = \boldsymbol{f}(\boldsymbol{q}, \boldsymbol{k}, \boldsymbol{\omega}, \boldsymbol{W})$ by tuning the parameters W. Except for simple, analytically solvable problems, identification process should involve revealing the dynamic property of the reference model by driving it with many samples of input command sequences. The number of samples required for quality identification depends on the size of space covered by $P(\boldsymbol{q}_o, \{\boldsymbol{k}\}, \{\boldsymbol{\omega}\})$. In fact, it is the same problem as approximating the objective function (5.6) with sample average (5.7). In this section, we estimate complexity of the identification problem. This will provide the basis for evaluating our step response sampling method in the next section. For simplicity, we consider zero noise cases.

Now, we measure cardinality of instantaneous command signal space as exponential of entropy,

$$|K| = \exp\left(-\int d\boldsymbol{k} \log(P(\boldsymbol{k}))\right), \tag{5.8}$$

where the instantaneous command signal distribution $P(\boldsymbol{k})$ is obtained by marginalizing the command sequence distribution $P(\{\boldsymbol{k}\})$. Then, cardinality of command sequence space roughly grows exponentially as $|K|^{(1+T/\tau_c)}$, where T is the trajectory duration and τ_c is correlation time constant. According to the ergodic principle, we take $T = \tau_m + \tau_c$, which makes cardinality of command sequence be $|K|^{(2+\tau_m/\tau_c)}$. Therefore, complexity of the identification problem is

$$|Q| \cdot |K|^{(2+\tau_m/\tau_c)}, \tag{5.9}$$

where $|Q|$ is the cardinality of initial state distribution. This crude analysis yields a crucial insight of the problem. The complexity depends on the ratio between memory time constant of the reference model and correlation time of command signal in a highly sensitive manner, especially when $|K|$ is large. The minimum complexity $|Q| \cdot |K|^2$ is achieved if the command signal fluctuates in much slower time scale than τ_m. For fast time varying command signal, however, complexity may become indefinitely large. In such cases, taking a sample-average from $P(\{\boldsymbol{k}\})$ may require an enormous number of samples. For efficient identification, we need a different distribution to sample from.

5.6 Step Response Sampling

In this section, we introduce a sampling method based on step response analysis. For linear systems, step response analysis means observing output of the systems at an operational state (usually at the origin) in response to a step change of the control input. If the command signal has the same dimensionality as the state vector and the state vector is fully observable, step responses can identify the full dynamic property of the systems in the most efficient manner.

For nonlinear, partially observable and partially controllable systems, step responses do not reveal the full dynamics. We can bother less with the observablity and controllability issues, since our goal is identifying only the observable and controllable part of the model dynamics.

More important problem is that dynamic property of the system differs at different operational states, so step responses at one state is not enough to reveal the full dynamics. We solve this problem by measuring step responses at all possible states. Following is Deltails of the step response sampling method (see fig 3)

1. Phase 0 ($0 < t < t_0$) : Set command at the first constant value $k = k_1$, and let the system run for τ_m. The reference model state will converge to the first attractor, which will be used as the initial state for phase 1.

2. Phase 1 ($t_0 < t < t_1$) : Keep the command at the same value $k = k_1$, and let the system run for Δt. If the reference model has limit-cycle attractor dynamics, Δt is the limit-cycle period. For point attractor dynamics, use $\Delta t = \tau_m$.

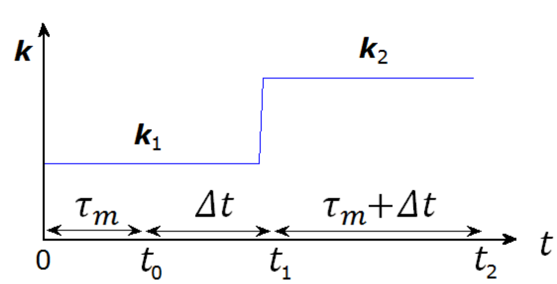


Figure 5.3: Illustration of the step response sampling method.

3. Phase2 ($t_1 < t < t_2$): Change command to the second constant value $k = k_2$, and let the system run for $\tau_m + \Delta t$. The reference model will make a transition movement from the first attractor to the second attractor.

where k_1 and k_2 are randomly drawn from the marginal distribution . Movement cost is evaluated during phase 1 and phase 2 only.

Evidently, cardinality of step response sampling is

$$|Q| \cdot |K|^2 \tag{5.10}$$

which is the minimal complexity of the identification problem. $|Q|$ is cardinality of remaining uncertainty of the first attractors which is not significant. For point attractors $|Q| = 1$ (no uncertainty), and for limit-cycle attractors $|Q|$ is proportional to length of the limit-cycle orbit (1 dimensional). Therefore, step response sampling allows efficient identification at minimum complexity. It correctly captures the long-time response during the constant input, as well as the high frequency response during the step. There are also some technical reasons for using step response sampling.

1. Availability of $P(\{\boldsymbol{k})$: Due to high dimensionality of sequences, their distribution is often too complicated to describe. On the other hand step response sampling only require the marginal distribution $P(\boldsymbol{k})$.

2. The cost functions we obtain from behavioral experiments are usually more suitable for discrete changes of task condition (e.g. reaching movements), rather than continuous change (e.g. tracking movements). For continuously changing task condition, the cost function should have an explicit description of desired delay in response. Model reference control may provide a better description for continuous case.

5.7 Optimization

For optimization, we use Rasmussens matlab code *minimize.m* [*minimize.m*], which implements *conjugate gradient descent method*. However, any other optimization method may be used as well, such as Kalman filter based methods [Prokhorov 2007]. In order to use gradient based methods, we need to differentiate the objective function. Because of the extra free parameter β, there are more than one ways to do partial differentiation.

5.7.1 Optimization with Adaptive β

The simplest is partial differentiation with fixed β. For convenience, we rewrite the objective function in (7) as

$$\tilde{A} = \frac{1}{\beta} \log Z, \tag{5.11}$$

$$\text{where} \quad Z(\beta, \{V_i\}) = \frac{1}{N} \sum_{i=1}^{N} \exp(\beta V_i)$$

is called the partition function. For a fixed β, gradient of the objective function is

$$\left. \frac{\partial \tilde{A}}{\partial \boldsymbol{W}} \right|_{\beta} = \left. \frac{1}{\beta Z} \frac{\partial Z}{\partial \boldsymbol{W}} \right|_{\beta} = \sum_{i=1}^{N} \frac{\exp(\beta V_i)}{Z} \frac{dV_i}{d\boldsymbol{W}} \tag{5.12}$$

Note that the gradient is a weighted average of $dV_i/d\boldsymbol{W}$ with the weights growing exponentially with the values V_i, which yields robust optimization. However, the effect of the exponential weights will eventually fade away if a constant β is used. As optimization progresses, magnitude of all values will decrease, and the exponential weights will all decay to be flat:

$$\frac{\exp(\beta V_i)}{Z} \approx \frac{1}{N}, \qquad \forall V_i.$$

What matters is the relative magnitude between β and V_i. In order to maintain the robust effect of exponential weights throughout the optimization process, β should increase in proportion to $1/\tilde{A}$. This is achieved by fixing Z at a constant value. In this case, the objective function can be rewritten as

$$\tilde{A} = \frac{1}{\beta} \log Z, \tag{5.13}$$

$$\text{where} \quad \beta(Z, \{V_i\}) \text{ satisfies } Z = \frac{1}{N} \sum_{i=1}^{N} \exp(\beta V_i)$$

In this expression, Z is defined as the free parameter and is a function of Z. For a fixed Z, the gradient of the objective function is

$$\left.\frac{\partial \tilde{A}}{\partial \boldsymbol{W}}\right|_Z = \log Z \frac{\partial}{\partial \boldsymbol{W}}\left.\left(\frac{1}{\beta}\right)\right|_Z = \frac{A}{U}\sum_{i=1}^{N}\frac{\exp(\beta V_i)}{Z}\frac{dV_i}{d\boldsymbol{W}}. \tag{5.14}$$

Here $U = \sum_{i=1}^{N}\frac{\exp(\beta V_i)}{Z}V_i$ is the weighted average of values. In this method, β increases as $1/\tilde{A}$ because $\beta\tilde{A}$ $(= \log Z)$ is constant. Therefore, the exponential weights remains effective throughout the optimization process.

5.7.2 Modified Pontryagin's Maximum Principle

Next, we compute $dV_i/d\boldsymbol{W}$ in order to complete the gradient calculation. However, this is not a simple calculation, because altering W has a complicated effect on the state trajectory $\{\boldsymbol{q}\}$. Here, we use a modified version of *Pontryagin's maximum principle* [Iyanaga and Kawada 1980]. For simplicity, the subscript i is suppressed.

Define Hamiltonian as

$$H(\boldsymbol{q},\boldsymbol{p};\boldsymbol{W}) = \boldsymbol{p}^T \cdot \boldsymbol{f}(\boldsymbol{q},\boldsymbol{k},\boldsymbol{\omega};\boldsymbol{W}) - c(\boldsymbol{q},\boldsymbol{k};\boldsymbol{W}) \tag{5.15}$$

Then, the gradient of value function is

$$\frac{dV}{d\boldsymbol{W}} = \int_{\{\boldsymbol{q}^*,\boldsymbol{p}^*\}} -\frac{\partial H}{\partial \boldsymbol{W}}dt = \int_{\{\boldsymbol{q}^*,\boldsymbol{p}^*\}} \left[-\boldsymbol{p}^T \cdot \boldsymbol{f}_{\boldsymbol{W}} + c\right] dt, \tag{5.16}$$

which is integrated along the *extremal trajectory* $\{\boldsymbol{q}^*,\boldsymbol{p}^*\}$, defined as

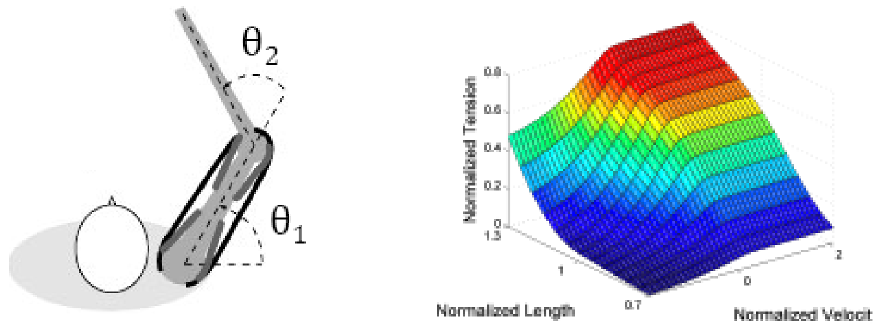
$$\boldsymbol{q}^* = \frac{\partial H}{\partial \boldsymbol{p}} = \boldsymbol{f} \qquad \boldsymbol{q}^*(t_i) = \boldsymbol{q}_o \tag{5.17}$$

$$-\boldsymbol{p}^{*T} = \frac{\partial H}{\partial \boldsymbol{q}} = \boldsymbol{p}^{*T} \cdot \boldsymbol{f}_{\boldsymbol{q}} - c_{\boldsymbol{q}} \qquad \boldsymbol{p}^*(t_f) = \boldsymbol{0} \tag{5.18}$$

Note that $\boldsymbol{q}^*$ is the actual trajectory generated by the dynamics. $\boldsymbol{p}^*$ is called the co-state vector and it represents sensitivity of the value function with respect to change of the state trajectory $\boldsymbol{q}^*$. The co-state trajectory $\boldsymbol{p}^*$ is calculated by integrating (5.18) backward in time.

This method is a generalization of the *Back-Propagation-Through-Time* method [Werbos 1994] for control problems.

5.8 Examples



(a) Simplified arm dynamics model: The muscles are connected around the joints on to the bones.

(b) Muscle tension generation function: Tension is a non-linear function of length and speed, and has linear dependence with muscle activation level. Activation level was set to 1 for this plot.

Figure 5.4: Arm dynamics model and muscle tension generation function.

We trained a recurrent neural network to control a realistic human arm-model for two different movement types: (1) straight, point-to-point reaching movement, which is made of transitions between point attractors, and (2) circle drawing movement, which consists of transitions between limit-cycle attractors.

5.8.1 Model System - Specification

Two-Link Arm Model

The body plant we use is a simulated, two degree-of-freedom arm restricted to horizontal movements (fig 4a). Actuation of the arm is performed by three pairs of opposing muscles (flexor/extensor). Two pairs of muscles individually actuate the shoulder and elbow joints, while the third pair actuates both joints simultaneously. Control of the arm is achieved by activating each muscle.

The arm state is represented by a 10 dimensional vector:
$\boldsymbol{x} = [\theta_1, \theta_2, \dot{\theta}_1, \dot{\theta}_2, a_1, a_2, a_3, a_4, a_5, a_6]^T$, where θ_1 is the shoulder joint angle, θ_2, the elbow angle, and a_i the activation level of each muscle. The activation dynamics is

approximated by a first order model: $\tau \dot{a}_i + a_i = \boldsymbol{u}_i$, where $\tau = 50\,ms$, and $\boldsymbol{u}_i$ is the control signal for each muscle.

The tension on each muscle is approximated to grow linearly with its activation level, but it also has a non-linear dependence on the length and the contraction speed of the muscle $T_i = a_i \cdot T(l_i, \dot{l}_i)$ (see fig 4b).

Sensory Signal and Command

In this model, sensory signal includes length, speed, and tension of each muscle: $\boldsymbol{s}(\boldsymbol{x}) = [\boldsymbol{l}; \dot{\boldsymbol{l}}; \boldsymbol{T}]$.

Process Noise

In this model, process noise is modeled as multiplicative noise acting on sensory, command, and control signals. That is, each element of $\boldsymbol{u}$ is replaced with its noisy version $\boldsymbol{u}_i \cdot (1 + \boldsymbol{\omega}_{ui})$, and similarly for $\boldsymbol{s}_i$, $\boldsymbol{k}_i$ as well. We modeled $\boldsymbol{\omega}_i$ as white Gaussian noise with zero mean and 0.2 standard deviation (20% signal to noise level).

5.8.2 Two types of movements

Reaching: Point-Attractor dynamics

We trained our system to perform reaching movements to randomly appearing target positions. The *RNN* receives a desired target location (in Cartesian coordinate) as the command signal ($\boldsymbol{k} = \boldsymbol{p}_{target}$). This is consistent with the experimental observation that neurons in motor cortex have cosine-tuning for target location [Georgopoulos et al. 1982], which implies Cartesian representation.

The instantaneous cost function is

$$c(\boldsymbol{x}, \boldsymbol{k}, \boldsymbol{u}) = 1 - \exp\left(-\frac{d^2}{2d_0^2}\right) + \frac{\alpha}{2}\frac{\|v\|^2}{v_0^2}\exp\left(-\frac{d^2}{2d_0^2}\right) + \frac{\gamma}{2}\|u\|^2, \tag{5.19}$$

where $d = \|\boldsymbol{p} - \boldsymbol{p}_{target}\|$ is the Euclidean distance between hand position $\boldsymbol{p}$ and the target, $d_o = 0.04\,m$ is the *target size*, v is the hand velocity, $v_o = 0.3\,m/s$

is the typical movement speed, and α, γ are appropriate mixing coefficients. For the distance cost, we used a saturating function (inverse of Gaussian) rather than the popular quadratic. We used the saturating cost for converting the distance cost in a time-invariant form. The velocity cost is active only near the target.

Targets are uniformly distributed in two dimensional space within the range

(a)
(b)
(c)
(d)

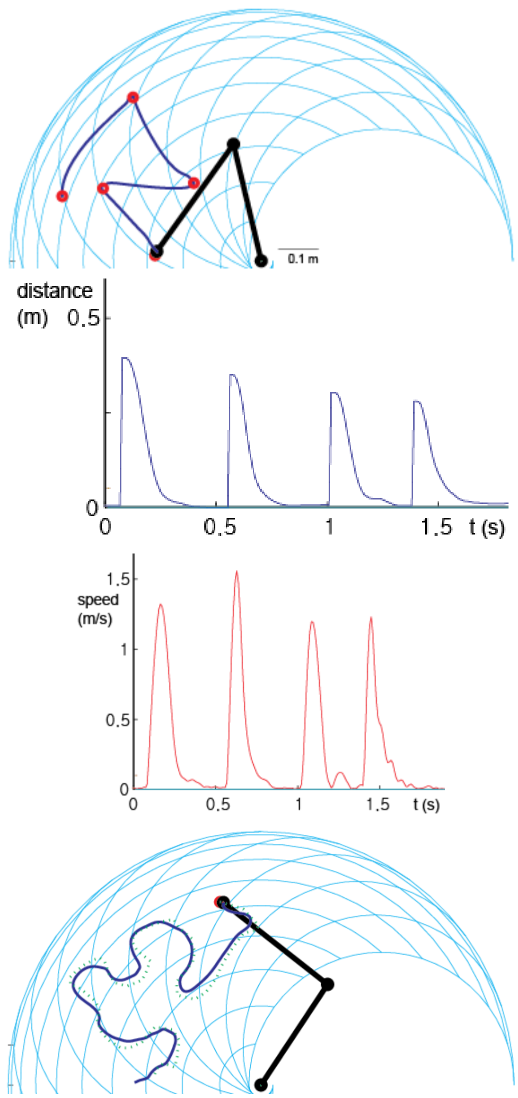


Figure 5.5: (a) Three consecutive reaching movements for discretely jumping target locations. Open circles represent target locations. Thick lines represent the arm. Background curves are joint-coordinate grids. (b) Distance between hand and target (c) Speed of hand movement (d) A tracking movement for a continuously moving target. A dotted line is the target trajectory and a solid line is the hand trajectory.

of the arm, which implies that complexity of reaching movements is four dimensional when step response sampling is used.

For batch training, 200 movement samples are used with RNN of 12 neurons. For larger RNNs, more training samples are needed in order to prevent over-fitting.

We first tested the system with a series of discrete target jumps (fig 5(a,b,c)). The arm successfully generated a series of reaching movements to a given target locations with near straight movement. The curvature is due to arm dynamics and non-isotropic distribution of inertia, which is also observed in human and monkey behaviors. Distance to target indeed decreases to near zero, and the speed plot matches well with the bell-shaped profile recorded from human reaching movements. When the target was allowed to make slow continuous movement, the arm-RNN system naturally generalized to follow the trajectory (fig 5(d)).

Circle Drawing: limit-cycle

We also applied our method to circle drawing movements whose desired location, radius and rotation speed are given by the command signal: $\boldsymbol{k} = [\boldsymbol{p}_{center}; r; v_d]$. We define positive v_d as clockwise rotation and negative v_d as counter-clockwise.

Instantaneous cost function is

$$c(\boldsymbol{x}, \boldsymbol{k}, \boldsymbol{u}) = 1 - \exp\left(-\frac{d^2}{2d_0^2}\right)\left(1 + \frac{\alpha}{2}\frac{v_d^2}{v_0^2}\right) + \frac{\alpha}{2}\frac{\|v - v_d\|^2}{v_0^2}\exp\left(-\frac{d^2}{2d_0^2}\right) + \frac{\gamma}{2}\|u\|^2, \tag{5.20}$$

where $d = \|\boldsymbol{p} - \boldsymbol{p}_{center}\| - r$ is the distance between hand and the desired circular path, and $v_d \, (= v_d \hat{\boldsymbol{n}})$ is the desired velocity in the tangent direction ($\hat{\boldsymbol{n}}$) of the path. Notice that this cost function reduces to the reaching cost when both the radius and v_d are set to zero. Also note that this cost function can easily be generalized to limit-cycles of any orbit shapes, simply by redefining d to be the distance between the orbit and hand, and v_d to be tangent to that orbit.

The command signal is distributed on four dimensional space, which leads to nine dimensional complexity for the circle-drawing movements. This is much more difficult task than the reaching-task. 600 movement samples are used with RNN size of 36 neurons.

The arm-RNN system was tested for discrete jumps of the center location (fig 6(a)), and for discrete changes of radius and rotation speed (fig 6(b)). In both cases, the system successfully generated correct circle drawing movements.

The most fascinating property of the system is its stability. The system never fails to converge to the correct limit-cycle dynamics in all harsh conditions such as abrupt changes of command sequence and sudden perturbation of body state or neural state. The system is also very robust. It draws circles of acceptable quality for all values of command signal.

5.9 Conclusion

We presented a novel framework for approximating an optimal global control law with a RNN controller. It combines stochastic optimal control framework with model reference control in infinite-horizon setting. We developed step response sampling method to achieve maximum efficiency. This framework well suited for modeling biological sensory-motor systems and for robotics application.

A RNN controller was trained generate infinite varieties of attractor dynamics for both point-attractors and limit-cycle attractors. At the end of optimization, we obtained highly stable and robust RNN controllers for real-time control of non-linear plants.

The RNN controller for reaching movements shares similarities with the arm control area of the primary motor cortex. On the behavioral level, it generates realistic reaching movements. In our primitive analysis, it also predicted several neurophysiological data such as distribution of preferred movement direction on population level. In the primary motor cortex of monkeys, micro-injection of step-function current evoked realistic reaching movements toward certain targets whose locations were uniquely determined by the sites of stimulation [Graziano, Taylor, and Moore 2002]. This is clear evidence that the primary motor cortex indeed employs point-attractor dynamics, and that reaching movements are generated in response to step-like changes of input command.

The RNN controller for circle-drawing movements does not have biological relevance. However, the same training paradigm can be applied to locomotive systems to approximate the locomotive CPGs (central pattern generators) in the spinal cord.

Passive-walkers have been suggested as a model of biological locomotion [Collins et al. 2005]. However, the main fallacy of the claim has been that biological legs are not passive-walkers. Our RNN controllers can be applied on the biological leg models and turn them into passive-walkers, or stable limit-cycle attractors. This will provide a link that bridges the gap between passive-walkers and biological locomotion.

Our RNN controllers model the lowest-level-controllers of the motor control hierarchy that maps a high-level command sequence to an appropriate body movement. The mapping is guaranteed to be unambiguous by the fading memory property. The stable attractor nature of our RNN controller dynamics makes it easy for the higher-level controller to plan complex movements. For example, figure-eight drawing movements can easily be generated with inputs that alternate between two center locations and opposite rotational directions at appropriate phase (fig 7). Therefore, it works as an interpreter between a high-level brain area and the body, such that the high-level areas can focus on abstract planning without having to deal with the complex physical Deltails of body dynamics. These RNN controllers can be the first step toward building hierarchical controllers.

Chapter 5, in full, is an edited reprint of the material as it appears in IEEE Adaptive Dynamic Programming and Reinforcement Learning, 2009. Huh, Dongsung; Todorov, Emanuel. The dissertation author was the primary investigator and author of this paper.

References

Collins, S. et al. (2005). "Efficient bipedal robots based on passive-dynamic walkers". In: *Science* 307.5712, pp. 1082–1085.

Fang, Y. and T.J. Sejnowski (1990). "Faster learning for dynamic recurrent backpropagation". In: *Neural Computation* 2.3, pp. 270–273.

Ganguli, S., D. Huh, and H. Sompolinsky (2008). "Memory traces in dynamical systems". In: *Proceedings of the National Academy of Sciences* 105.48, p. 18970.

Georgopoulos, A.P. et al. (1982). "On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex". In: *The Journal of Neuroscience* 2.11, pp. 1527–1537.

Graziano, M.S.A., C.S.R. Taylor, and T. Moore (2002). "Complex movements evoked by microstimulation of precentral cortex". In: *Neuron* 34.5, pp. 841–851.

Iyanaga, S. and Y. Kawada (1980). "Pontrjagin's Maximum Principle". In: *Encyclopedic Dictionary of Mathematics*, pp. 295–296.

Jacobson, D. (1973). "Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games". In: *Automatic Control, IEEE Transactions on* 18.2, pp. 124–131.

Lockery, SR and TJ Sejnowski (1992). "Distributed processing of sensory information in the leech. III. A dynamical neural network model of the local bending reflex". In: *The Journal of neuroscience* 12.10, pp. 3877–3895.

Narendra, K.S. and A.M. Annaswamy (1989). *Stable adaptive systems*. Prentice-Hall, Inc.

Prokhorov, D., G. Puskorius, and L. Feldkamp (2001). *Dynamical neural networks for control*.

Prokhorov, D.V. (2007). "Training recurrent neurocontrollers for real-time applications". In: *Neural Networks, IEEE Transactions on* 18.4, pp. 1003–1015.

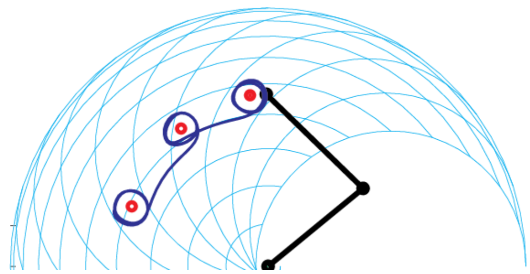
Rasmussen, C.E. *minimize.m*.

Todorov, E. (2004). "Optimality principles in sensorimotor control (review)". In: *Nature neuroscience* 7.9, p. 907.

Todorov, E. and M.I. Jordan (2002). "Optimal feedback control as a theory of motor coordination". In: *Nature neuroscience* 5.11, pp. 1226–1235.

Werbos, P.J. (1994). *The roots of backpropagation: from ordered derivatives to neural networks and political forecasting*. Vol. 1. Wiley-Interscience.

(a)



(b)

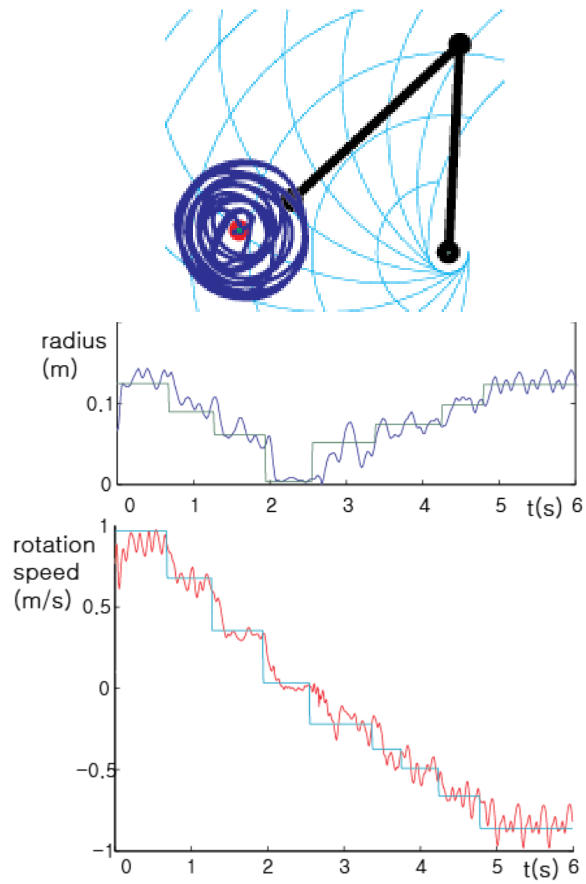


Figure 5.6: (a) A trajectory of circle drawing movement for the condition where the center location makes two two consecutive discrete jumps (b) Top: A trajectory of circle drawing movement for time varying radius and rotation speed. Middle: Radius plot. Bottom: Rotation speed plot. A command sequence is overlaid on the actual body movement.

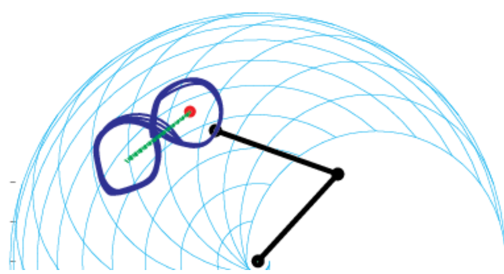


Figure 5.7: A figure-eight drawing movement. Such movements can be generated with command inputs that alternate between two center locations and opposite rotational directions at appropriate phase.