



Optimal control and reinforcement learning in simple physical systems

Citation

Davenport, Ian. 2023. Optimal control and reinforcement learning in simple physical systems. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

Link

<https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37375821>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#)

HARVARD UNIVERSITY
Graduate School of Arts and Sciences



DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the
Department of Physics
have examined a dissertation entitled

Optimal control and reinforcement learning in simple physical systems

presented by Ian Christopher Davenport

candidate for the degree of Doctor of Philosophy and hereby
certify that it is worthy of acceptance.

Signature Mahadevan

Typed name: Professor Lakshminarayanan Mahadevan, Chair

Signature Cora Dvorkin

Typed name: Professor Cora Dvorkin

Signature Douglas Finkbeiner

Typed name: Professor Douglas Finkbeiner

Date: May 2, 2023

Optimal control and reinforcement learning in simple physical systems

A DISSERTATION PRESENTED
BY
IAN CHRISTOPHER DAVENPORT
TO
THE DEPARTMENT OF PHYSICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
PHYSICS

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
MAY 2, 2023

©2023 – IAN CHRISTOPHER DAVENPORT
ALL RIGHTS RESERVED.

Optimal control and reinforcement learning in simple physical systems

ABSTRACT

Motor control is a ubiquitous theme in the natural world and its richness demands we approach it from several directions. Alongside the challenge of understanding perception and coordinated muscular activation we need to account for the fact that all motion takes place in a body embedded in a dynamic environment. Inspired by extreme feats of athletic and animal performance we bring together ideas from physics and optimization to approach problems in motor control. First we study the Estonian sport Kiiking and show how to augment basic physical models to account for physiological limitations. We use methods from both optimal control theory and reinforcement learning to understand optimal strategies in the sport and compare our findings to observations. Next we describe an approach for the control of gliding flight by dynamically changing the shape of the gliding body. We compare strategies that are time optimal to those which require the least control input. Finally we use elasticity theory to understand the mechanics of snakes as they traverse large vertical gaps. After describing our data analysis procedure we detail the optimal control problem we formulated to understand the posture of the animals as they complete their ascent.

Contents

TITLE	i
COPYRIGHT	ii
ABSTRACT	iii
CONTENTS	iv
LISTING OF FIGURES	vi
AUTHOR LIST	x
DEDICATION	xi
ACKNOWLEDGEMENTS	xii
1 INTRODUCTION	1
1.1 Optimal Control Theory	3
1.2 Reinforcement Learning	9
2 OPTIMAL STRATEGIES FOR KIIKING: PUMPING A SWING TILL INVERSION	14
2.1 Experimental data analysis	17
2.2 Mathematical model	17
2.3 Time-optimal control	22
2.4 Reinforcement learning	24
2.5 Results	26
2.6 Discussion	27
3 PLANAR CONTROLLED GLIDING THROUGH SHAPE CHANGE	29
3.1 Mathematical model	30
3.2 The control problem	33
3.3 Results	35

3.4	Discussion	41
4	VERTICAL GAP BRIDGING IN SNAKES	42
4.1	Data Analysis	43
4.2	Mathematical model	46
4.3	Optimal control	49
4.4	Results	51
4.5	Conclusion	55
5	CONCLUSION	57
	APPENDIX A SUPPLEMENTARY INFORMATION FOR KIIKING	59
A.1	Data analysis	59
	REFERENCES	61

Listing of figures

1.1	How an agent interacts with its environment: The agent is presented with a state S_t and reward R_t at time t by the environment. Based on the state S_t the agent decides on an action A_t using its policy π . The action affects both the agent and the environment leading to a new state S_{t+1} . As a result of its action the agent also receives a reward R_{t+1} . The goal of the RL agent is to learn a policy of actions which maximizes the accumulated reward it receives.	10
2.1	(a) An athlete on a tall kiiking swing. The arms of the swing are rigid, and the athlete is attached to the base by their feet. Adapted from Estonian Kiiking Association ⁸ , ©2014 Estonian Kiiking Association. (b) A schematic of an athlete pumping the kiiking swing by squatting and standing. The swing-athlete system is modeled as a simple (point-mass) pendulum with variable length $l(t)$. Over the course of one cycle, the athlete quickly stands up when the swing is near its lowest point, and squats down as the swing reaches its highest point. This stand-squat cycle is repeated twice over each oscillation of the swing's motion. (c) The angle θ measured from experiments as a function of time for a successful kiiking attempt. The length of the swing is approximately 7 meters. (d) Nondimensionalized energy $E/mgl_0 \approx -\cos \theta$ after the n th cycle, i.e. at the n th extremum of θ	16
2.2	Analysis of the standing phase of a stand-squat cycle. Starting at the switching time $t = t_s$, the athlete decreases the swing length l at the maximum allowed rate $u = u^-$ (Eq. (2.4)) until l reaches the minimum value l_- . (a) The angle θ and control $u = \dot{l}$ as functions of (nondimensionalized) time. Left: The switching time t_s is varied in the absence of power constraints ($P_m = \infty$). The greedy switching time t_s^g (solid green lines) maximizes the energy gained by the system. Right: Solutions with $t_s = t_s^g$ and varying power bounds P_m . In both cases, $\Delta l = 0.115$, $u_m = 0.5$. (b) Energy E_f at the end of a cycle as a function of the switching time t_s . Each solid curve corresponds to a different initial angle $\theta_i = \theta(0)$. The dashed black curve indicates the greedy switching time t_s^g for each θ_i . When the system is not power-constrained (left), t_s^g increases monotonically with θ_i . In the presence of power constraints (right), the relationship is more complicated. Here, $\Delta l = 0.1$, $u_m = 1.0$	21

2.3	Time-optimal solution of the kiiking problem. The angle θ and control $u = \dot{l}$ as functions of (nondimensionalized) time for the numerically computed optimal control solution (OC, black lines) and the solutions obtained from reinforcement learning (RL, blue lines). Left: When the system is not power-constrained, the OC solution consists of several interior arcs where $u = \pm u_m$, separated by boundary arcs where $ l-1 = \Delta l$ and $u = 0$. Right: In the presence of power constraints, u takes the value $u^\pm$ (equation (2.4)) on the interior arcs. In both cases, the RL solution has the same form but with slightly different switching times. For both simulations, $\theta_0 = \pi/4$, $\Delta l = 0.05$, $u_m = 0.1$	24
2.4	Nondimensionalized energy E/mgl at the end of each stand-squat cycle for five trials by different kiiking trials. The gray ribbon shows the envelope of the error bars for all five trials. In the two longest trials (circles and triangles) the length of the swing arms is approximately 7m. For the other trials, the length is smaller but not precisely known. Two of the series (diamonds and pentagons) show failed attempts, which ended without reaching $\theta = \pi$. A simulation with parameters $\Delta l = 0.05$, $u_m = 0.125$, $P_m = 0.125$, $d = 2.75 \times 10^{-2}$ is shown for comparison, and captures the sigmoidal shape of the $E(n)$ curves.	27
3.1	Model of the glider: u is the velocity along the longitudinal axis of the body. v is the velocity along the transverse axis. w is the angular velocity. a and b are the radii of the semi major and semi minor axes. The x and y axes are shown in the bottom and are fixed while the u and v axes rotate with the body. We consider the problem of how such a body can glide through a quiescent fluid and perch at a given location. We assume control of the shape of the ellipse characterized by the aspect ratio $\beta = \frac{b}{a}$	31
3.2	Maximum glide distance as a function of density: For each density ratio ρ_s/ρ_f the glider starts with $(u, v, w, x, y, \theta, \beta) = (0.1m/s, 0.1m/s, 0, 0, 0, 0, 1)$ and glides as far as it can until its center of mass hits the ground located at $y_f = -10m$. To encourage greater gliding distance we set the running cost $L = -x^2$ so that the body is penalized over time for its proximity to the y axis. When the solid is not much denser than the liquid, added mass effects are more significant and gliding performance is improved.	36
3.3	Optimal perching with $\rho_s/\rho_f = 200$, $\alpha = ab = 0.1m^2$. In both tasks we start with $(u, v, w, x, y, \theta, \beta) = (0.1m/s, 0.1m/s, 0, 0, 0, 0, 1)$ and land with $(x_f, y_f, \theta_f, \beta_f) = (25m, -75m, \pi/4, 1)$. Left: Trajectories for the time and energy optimal task. Right: The value of β over time. Note how the time optimal solution has much more change in β (i.e. control input) despite the final trajectories looking very similar. The time optimal solution achieves the perching task about 4% faster than the energy optimal. The time axis is scaled by t_f , the time to land, in each case for a more clear comparison of control strategies.	37

3.4	Optimal perching with $\rho_s/\rho_f = 2, \alpha = ab = 0.1m^2$. In both tasks we start with $(u, v, w, x, y, \theta, \beta) = (0.1m/s, 0.1m/s, 0, 0, 0, 0, 1)$ and land with $(x_f, y_f, \theta_f, \beta_f) = (75m, -75m, \pi/4, 1)$. Since the density ratio ($\rho_s/\rho_f = 2$) is much smaller, and we know the body won't have trouble gliding to greater distances, we use a larger target x in the desired perching location than in the $\rho_s/\rho_f = 200$ case. Left: Trajectories for the time and energy optimal task. Right: The value of β over time. Again the time optimal solution shows much more control input and uses the extra control to perch about 15% faster than in the energy optimal case. The time axis is scaled by t_f , the time to land, in each case for a more clear comparison of control strategies.	38
3.5	Glider shape and orientation over time for the energy optimal perching trajectory $\rho_s/\rho_f = 200$. Since the body is much denser than the fluid β must be changed significantly to alter the course of the glider. We also see a significant variation in the orientation of the glider as it tumbles towards the ground.	39
3.6	Glider shape and angle over time for the energy optimal perching trajectory with $\rho_s/\rho_f = 2$. Note how β is very near 1 for most of the trajectory indicating that significant control input is not required as long as it is appropriately timed (cf. Fig 3.5).	40
4.1	A time lapse of a python as it completes a gap bridging task. The gap is about 70% of the snake's body length. Repeated trials of this kind of experiment were performed in the lab with a python and 3 different brown tree snakes. We analyzed these videos to understand how the snake's shape evolved over time as it completed it's task. . .	44
4.2	A sample of how the image analysis process works. We take an image from the video (left), crop it and edit out the top and bottom perches. Starting from the cropped image (center) we use our image analysis software to extract the front of the body (right). In red are points representing the pixel values of points on the front of the snake. The blue line is an interpolation through the pixel values. We use the interpolated line to compute curvature because the raw pixel data is too noisy.	45
4.3	Top panels: A snapshot of the curvature along the snakes when it is near the top of the perch. Bottom panel: A time lapse of the curvature at each frame over the course of the video. Negative curvature is dorsiflexion. Left: Python 71 cm gap. Middle: Python 50 cm gap. Right: Tree snake 60 cm gap	46
4.4	We model the snake as an elastic filament. The center line is parameterized by s . The tangent vector $\mathbf{t} = (dx/ds, dy/ds)$ describes the local orientation of the snake. . .	47
4.5	Left: A snapshot of a brown tree snake in the middle of a climbing task. Right: Outline found from solving the optimization problem. There is a qualitative agreement between the shapes.	52
4.6	Left: The angle θ the snake makes with the horizontal along its body. Right: The curvature along the snake scaled by the maximum curvature. Note we are using $\kappa = -d\theta/ds$ to be consistent with curvature conventions (dorsiflexion should correspond to negative curvature) used in the data analysis portion of the work.	53

4.7	The control profile for reaching to $(x_f, y_f) = (0.1, 0.9)$ with $\alpha = 5$. The active moment (a proxy for the muscular activity) is greatest in magnitude near the base of the snake. This is consistent with the intuition that muscular activity is maximal where the curvature along the snake is the greatest cf. fig. 4.6	54
A.1	Measuring angles with image analysis software Fiji ⁸⁵ . We approximate the mass of the swing structure and the athlete to create a weighted average of the system center of mass. The angular location of the true center of mass is somewhere in between the post the athlete is holding and the athletes torso. The difference in measured angles provides an upper bound on the measurement error. (a) Measuring the angle the swing makes with the vertical. (b) Measuring the angle between the approximate center of mass of the athlete and the vertical.	60

THE FOLLOWING AUTHORS CONTRIBUTED TO CHAPTER 2: I. DAVENPORT, P. BRYDE, L. MAHADEVAN

THE FOLLOWING AUTHORS CONTRIBUTED TO CHAPTER 3: I. DAVENPORT, L. MAHADEVAN

THE FOLLOWING AUTHORS CONTRIBUTED TO CHAPTER 4: I. DAVENPORT, P. BRYDE, G. PRASATH, B. JAYNE, L. MAHADEVAN.

FOR MY FAMILY AND FRIENDS.

Acknowledgments

Thank you to everyone in the Harvard community who gave me support along the way. Thank you to my advisor Dr. L Mahadevan and dissertation committee members Dr. Cora Dvorkin and Dr. Douglas Finkbeiner. Thank you to everyone in the Mahadevan research group. Thank you to Dr. Bruce Jayne for providing us with experimental data on snake locomotion. I'd also like to especially thank Dr. Ilarion Melnikov for giving me an early example of how a great scientist approaches research.

1

Introduction

In physics we generally consider passive systems and try to understand or predict the properties of matter as it interacts with its surroundings. The rapidly expanding field of “active matter”⁷⁹ broadens the purview of traditional physics by studying systems of active particles which consume energy and generate forces^{39,96}. A major motivating factor for such an extension is that living systems fit in the active matter framework. In active matter phenomena from flocking in birds¹² to collective transport in ants³⁴ have been addressed and new studies are published every week. The lesson we

draw from such work is that tools from physics combined with a proper accounting of biological activity can teach us a lot about the physical world. Nowhere is the importance of activity more apparent than in the striking feats of individual organisms as they move themselves through their environments.

Motor control is an inescapable theme in both evolution and every day life^{17,53,54,82}. Nearly everything we do, even seemingly mundane tasks like speaking⁸⁸ or walking, requires a tremendous amount of coordination. We often take for granted the ability to process diverse sensory inputs³⁷, decide on goals, and send out precise commands to our muscles until a link in that chain is broken. One way of appreciating the difficulties involved is to consider the limitations of current robotic technology. Despite the societal relevance and decades of development, robots are still highly constrained to do specific kinds of tasks^{33,35}. Not is it until we observe striking successes of motor control that we come to appreciate the challenges involved. Fortunately such extreme feats abound.

In our work we bring together modeling of agents, their environment, and their decision making processes to make sense of the natural world^{68,93}. Our starting point is often observation of an impressive athletic or animal behavior. We first make our description of the behavior more quantitative by analyzing data captured in a lab or by other means. With that quantitative description in hand we develop physical models to see if we can understand the data. Since we are considering living systems we need to augment traditional physical models which only provide descriptions of passive matter. With a model in place we then formulate the objective that the agent is trying to achieve in mathematical language. We assume those objectives are achieved optimally and compare the solutions we find with observed data. Evolutionary or sporting pressures drive behavior towards near optimal in some manner⁹⁴, so the presumption of optimality is not far fetched.

In this thesis I will present three studies. First I will describe our project on the Estonian swinging sport Kiiking. In that project we show how to incorporate physiological limitations into a physical model and use optimization methods to match model predictions to observed performance. Next I

describe a project on the control of gliding flight by changing the shape of the glider. Finally I detail our project on the vertical gap bridging behavior of snakes. There we perform a novel analysis of video data and propose a model of snakes as active elastic filaments.

Before describing our work I will give an introduction to the optimization and control methods we use throughout. In particular I will describe the methods of optimal control theory and reinforcement learning.

1.1 OPTIMAL CONTROL THEORY

Optimal control theory (OCT) is a powerful tool for controlling complex dynamical systems^{48,20}. It is widely applied in both industrial^{49,100} and academic^{94,91} settings. Physical systems are highly constrained to satisfy the governing equations at all times. The framework of optimal control allows us to naturally account for these constraints and guide the system through it's state space as efficiently as possible. Here we walk through the general formulation of an OCT problem and highlight some major results. Throughout the remainder of this thesis we use these results and the examples presented will highlight the range of problems the theory can address.

Let $\mathbf{x}(t)$ be the state of some physical system at time t . Let $\mathbf{u}(t)$ be the control inputs used to manipulate the system. The dynamics are encapsulated in the equations

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), t) \quad (1.1)$$

$$\mathbf{x}(0) = \mathbf{x}_0 \quad (1.2)$$

which says that $\mathbf{x}$ starts in some $\mathbf{x}_0$ and evolves in time according to equations laid out by the function $\mathbf{f}$. Note we are assuming the system dynamics can be described as a set of first order differential equations. Redefinition of variables allows higher order equations to be reduced to a set of first order equations⁹⁰ so this formulation does not limit the complexity of systems we consider. OCT

methods can be extended to the control of partial differential equations⁹⁵ but we do not require that formalism in our work. The goal in optimal control theory is to find a control $\mathbf{u}^*(t)$ that causes the system in (1.1) to follow a trajectory $\mathbf{x}^*(t)$ which minimizes the performance measure

$$J(\mathbf{u}) = \int_{t_0}^{t_f} L(\mathbf{x}(t), \mathbf{u}(t), t) dt + K(\mathbf{x}(t_f), t_f) \quad (1.3)$$

while satisfying all constraints on $\mathbf{x}$ and $\mathbf{u}$. The function L is a “running cost” while K is known as a “terminal cost.” Suppose we want to land a spaceship on the moon. The dynamics $\mathbf{f}$ would then capture the competing effects of gravity, thrust, air resistance, and any other forces affecting the spaceship. The state $\mathbf{x}$ will likely describe the position, orientation, and velocity of the body. The control $\mathbf{u}$ will represent the thrust the rockets can generate. The running cost L will represent fuel burn or distance from a desired trajectory, and the terminal cost K will be proportional to the distance between the real and desired landing locations. Of course the specifics depend on the constraints present but we hope the example gives intuition for the meaning of the different variables.

Now that we have the problem set up we need to find the optimal controls $\mathbf{u}^*(t)$. There are two methods to choose from. One way to proceed is to use dynamic programming developed by Richard Bellman and others¹⁵. The second way is to use the calculus of variations and the minimum principle developed by Lev Pontryagin⁷⁷ and collaborators. We will briefly describe each. See for example^{48,56} for more detail.

1.1.1 THE HAMILTON-JACOBI-BELLMAN EQUATION

Dynamic programming methods are fundamentally based on subdividing the stated optimization problem into a sequence of smaller ones. Optimal solutions are easier to find on sub-intervals and in the end we combine them to solve the full optimization problem. Let’s walk through how this process leads to a nonlinear partial differential, the Hamilton-Jacobi-Bellman equation, which we

use to solve for the optimal controls $\mathbf{u}^*$.

Consider the cost in eq (1.3) from time t to t_f .

$$J(\mathbf{x}(t), t, \mathbf{u}(\tau)) = \int_t^{t_f} L(\mathbf{x}(\tau), \mathbf{u}(\tau), \tau) d\tau + K(\mathbf{x}(t_f), t_f) \quad (1.4)$$

Note J depends on the values $\mathbf{x}(t)$, t , and the control history from t to t_f . Let's write the minimum possible J (assuming all constraints are satisfied throughout the interval) as

$$J^*(\mathbf{x}(t), t) = \min_{\mathbf{u}} \left[\int_t^{t_f} L(\mathbf{x}(\tau), \mathbf{u}(\tau), \tau) d\tau + K(\mathbf{x}(t_f), t_f) \right] \quad (1.5)$$

If we subdivide the interval into $[t, t + \Delta t]$ and $[t + \Delta t, t_f]$, we get

$$J^*(\mathbf{x}(t), t) = \min_{\mathbf{u}} \left[\int_t^{t+\Delta t} L(\mathbf{x}(\tau), \mathbf{u}(\tau), \tau) d\tau + \int_{t+\Delta t}^{t_f} L(\mathbf{x}(\tau), \mathbf{u}(\tau), \tau) d\tau + K(\mathbf{x}(t_f), t_f) \right] \quad (1.6)$$

The principle of optimality states that for optimal performance the sequence of controls is optimal on each sub-interval¹⁶. In particular that means we can rewrite (1.6) as

$$J^*(\mathbf{x}(t), t) = \min_{\mathbf{u}} \left[\int_t^{t+\Delta t} L(\mathbf{x}(\tau), \mathbf{u}(\tau), \tau) d\tau + J^*(\mathbf{x}(t + \Delta t), t + \Delta t) \right] \quad (1.7)$$

since we have assumed we are following the optimal trajectory and $J^*(\mathbf{x}(t + \Delta t), t + \Delta t)$ is the minimal cost from $t + \Delta t$ to t_f . Let's assume Δt is small and Taylor expand J^* . That gives

$$J^*(\mathbf{x}(t), t) = \min_{\mathbf{u}} \left[L(\mathbf{x}(t), \mathbf{u}(t), t) \Delta t + J^*(\mathbf{x}(t), t) + \frac{\partial J^*}{\partial t} \Delta t + \frac{\partial J^*}{\partial \mathbf{x}}^T (\mathbf{x}(t + \Delta t) - \mathbf{x}(t)) + \text{h.o.t} \right] \quad (1.8)$$

With higher order terms in Δt that we ignore going forward. We know $\mathbf{x}(t + \Delta t) - \mathbf{x}(t) =$

$\mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), t)\Delta t$ from (1.1) so

$$J^*(\mathbf{x}(t), t) = \min_{\mathbf{u}} \left[L(\mathbf{x}(t), \mathbf{u}(t), t)\Delta t + J^*(\mathbf{x}(t), t) + \frac{\partial J^*}{\partial t} \Delta t + \frac{\partial J^*}{\partial \mathbf{x}}^T \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), t)\Delta t \right] \quad (1.9)$$

Since $J^*(\mathbf{x}(t), t)$ does not depend on $\mathbf{u}$ we can pull those terms outside of the minimization and we are left with

$$-\frac{\partial J^*}{\partial t} = \min_{\mathbf{u}} \left[L(\mathbf{x}(t), \mathbf{u}(t), t) + \frac{\partial J^*}{\partial \mathbf{x}}^T \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), t) \right] \quad (1.10)$$

and the boundary condition $J^*(\mathbf{x}(t_f), t_f) = K(\mathbf{x}(t_f), t_f)$ which we get by substituting $t = t_f$ into (1.4). Finally we define a Hamiltonian $\mathcal{H}(\mathbf{x}(t), \mathbf{u}(t), \frac{\partial J^*}{\partial \mathbf{x}}, t)$

$$\mathcal{H} = L(\mathbf{x}(t), \mathbf{u}(t), t) + \frac{\partial J^*}{\partial \mathbf{x}}^T \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), t) \quad (1.11)$$

and note that the optimal control $\mathbf{u}^*(\mathbf{x}(t), \frac{\partial J^*}{\partial \mathbf{x}}, t)$ must minimize this Hamiltonian (eq. 1.10). The Hamilton-Jacobi-Bellman (HJB) equation is then

$$-\frac{\partial J^*}{\partial t} = \mathcal{H} \left(\mathbf{x}(t), \mathbf{u}^*(\mathbf{x}(t), \frac{\partial J^*}{\partial \mathbf{x}}, t), \frac{\partial J^*}{\partial \mathbf{x}}, t \right) \quad (1.12)$$

$$J^*(\mathbf{x}(t_f), t_f) = K(\mathbf{x}(t_f), t_f) \quad (1.13)$$

and we use this equation to solve for $\mathbf{u}^*$. Unsurprisingly, in the vast majority of cases, this equation is intractable and we must rely on numerical methods³⁰. Nonetheless we have a necessary condition that any optimal control $\mathbf{u}^*$ must satisfy, namely it must minimize $\mathcal{H}$. Although we will not show it here any cost J which solves the HJB equation is actually the optimal cost so eq. (1.12) is a sufficient condition for optimality. See⁵⁶ for a more detailed discussion.

1.1.2 THE MINIMUM PRINCIPLE

Another approach to solving optimal control problems is to use the calculus of variations and the Pontryagin minimum principle. Consider again the cost in eq. (1.3) with a fixed initial condition $\mathbf{x}_0$ and variable final condition $\mathbf{x}(t_f)$ (for simplicity)

$$J(\mathbf{u}) = \int_{t_0}^{t_f} L(\mathbf{x}(t), \mathbf{u}(t), t) dt + K(\mathbf{x}(t_f), t_f) \quad (1.14)$$

To directly incorporate the constraint from the state dynamics into eq. (1.14) use Lagrange multipliers $\lambda(t)$ to arrive at

$$J(\mathbf{u}) = \int_{t_0}^{t_f} \left(L(\mathbf{x}(t), \mathbf{u}(t), t) + \lambda^T(\mathbf{f}(\mathbf{x}, \mathbf{u}, t) - \frac{d\mathbf{x}}{dt}) \right) dt + K(\mathbf{x}(t_f), t_f) \quad (1.15)$$

Let $\mathcal{H}(\mathbf{x}, \mathbf{u}, \lambda, t) \equiv L(\mathbf{x}(t), \mathbf{u}(t), t) + \lambda^T \mathbf{f}$ so that

$$J(\mathbf{u}) = \int_{t_0}^{t_f} \left(\mathcal{H}(\mathbf{x}, \mathbf{u}, \lambda, t) + \frac{d\lambda^T}{dt} \mathbf{x} \right) dt + [\lambda^T(t_f) \mathbf{x}(t_f) - \lambda^T(t_0) \mathbf{x}(t_0)] + K(\mathbf{x}(t_f), t_f) \quad (1.16)$$

by substitution and integrating by parts. If the control $\mathbf{u}^*$ is optimal, then a first variation of J around $\mathbf{u}^*$ must vanish. Note that $\mathbf{x}$ and the final state $\mathbf{x}(t_f)$ will change as a result of a new control history. Doing this variation gives

$$\delta J(\mathbf{u}) = \int_{t_0}^{t_f} \left(\delta \mathbf{x}^T \left(\frac{\partial \mathcal{H}}{\partial \mathbf{x}} + \frac{d\lambda}{dt} \right) + \delta \mathbf{u}^T \frac{\partial \mathcal{H}}{\partial \mathbf{u}} \right) dt + \delta \mathbf{x}(t_f)^T \left[\lambda(t_f) + \frac{\partial K(\mathbf{x}(t_f), t_f)}{\partial \mathbf{x}} \right] \quad (1.17)$$

where we have used the symmetry of the dot product to rearrange terms. Using the definition of $\mathcal{H}$ and since δJ must vanish for all variations around the optimal $\mathbf{u}^*$, we arrive at the following set of

necessary equations for $\mathbf{u}$ to be an optimal control

$$\frac{\partial \mathcal{H}}{\partial \mathbf{u}} = 0 \quad (1.18)$$

$$\frac{\partial \mathcal{H}}{\partial \lambda} = \frac{d\mathbf{x}}{dt} \quad (1.19)$$

$$-\frac{\partial \mathcal{H}}{\partial \mathbf{x}} = \frac{d\lambda}{dt} \quad (1.20)$$

$$\frac{\partial K(\mathbf{x}(t_f), t_f)}{\partial \mathbf{x}} = -\lambda(t_f) \quad (1.21)$$

The minimum principle is captured in eq. (1.18) which says $\mathcal{H}$ is extremal for the optimal control. Intuitively we know $\mathcal{H}$ must be a minimum by considering eq. (1.16) over a short time (i.e $t_f = t_0 + dt$). In such a case it is clear that if we want to minimize J , then $\mathcal{H}$ must be minimal. Again, it is typically too difficult to solve these equations and we must resort to numerical methods.

1.1.3 NUMERICAL METHODS

The results laid out in the previous two sections lay the foundation on how to solve optimal control problems. In practice the equations described are too complicated to solve by hand and we must resort to numerical methods⁸⁰. Broadly there are two classes of methods to solve OCT problems numerically: direct and indirect. Indirect methods use the necessary conditions in eqs. (1.18 - 1.21) and solve a boundary value problem for the trajectories $\mathbf{x}(t)$ and costates $\lambda(t)$ (from which the optimal controls can be constructed). Indirect methods are elegant but difficult to implement because it's often hard to use the minimum principle to simplify the form of the differential equations and integrate them forward in time³⁰. Direct methods start from the basic statement of the control problem (minimize $J(\mathbf{u})$) and discretize both the state and control over time. Starting from the discretized system we then directly minimize J subject to the constraints using non linear programming¹⁰. In our work we use direct methods implemented in the CasADI⁶ software package.

We especially make use of direct collocation methods which typically approximate the control as a piecewise constant function and the state as a piecewise polynomial. These methods determine polynomial coefficients by requiring continuity and that eq. 1.1 be satisfied at a number of collocation points inside the discretization intervals. Collocation methods are very powerful³⁰ but the art of getting them to work generally requires the user provide a good initial guess for the solution.

1.2 REINFORCEMENT LEARNING

Reinforcement learning (RL) is a paradigm for solving sequential decision making problems⁹². It has become a much more popular tool in recent years after researchers combined it with deep learning⁴⁰ to achieve human level performance in video game play⁶⁶. An enormous number of successful applications of the technology have followed ranging from robotic manipulation⁷ to optimized industrial cooling⁶⁰. In RL agents interact with their environment and receive feedback in the form of a reward (fig 1.1). The reward encapsulates prescribed goals and over the long term the agent learns to maximize the reward it receives. Aside from its use as an engineering tool RL provides a computational framework to address questions in behavioral neuroscience regarding how animals and humans make decisions in the face of reward or punishment⁶⁹. Specifically, RL is useful when thinking about the development of skills that require several intermediate steps to complete (e.g. learning to pump a swing). In such cases error based learning methods, where gradients of the final error are computed with respect to motor commands, are not the likely means by which learning occurs⁹⁹. Because of its power as an optimization technique and connection to learning in nature, we incorporate techniques from RL into our work. Next we describe how to set up and solve a reinforcement learning problem.

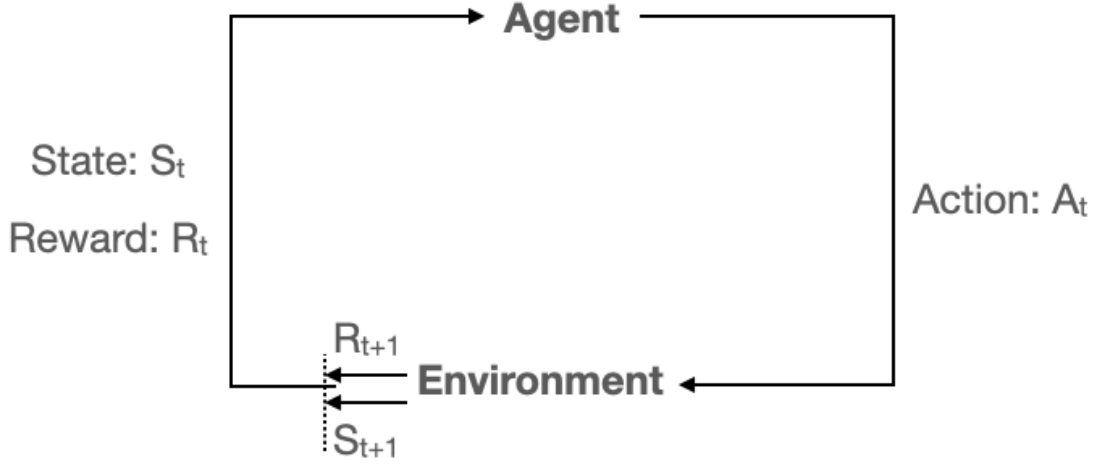


Figure 1.1: How an agent interacts with its environment: The agent is presented with a state S_t and reward R_t at time t by the environment. Based on the state S_t the agent decides on an action A_t using its policy π . The action affects both the agent and the environment leading to a new state S_{t+1} . As a result of its action the agent also receives a reward R_{t+1} . The goal of the RL agent is to learn a policy of actions which maximizes the accumulated reward it receives.

1.2.1 MARKOV DECISION PROCESSES

Problems in RL are posed in the language of Markov Decision Processes (MDP)⁹². In an MDP we have an agent which is learning and doing, and an environment which is everything outside the agent that it interacts with. After each of its interactions the agent receives a reward from the environment and over time adjusts its behavior to maximize accumulated rewards. Mathematically we have a set of states $\mathcal{S}$, actions $\mathcal{A}$, and rewards $\mathcal{R}$. The specific form of the states and actions will vary depending on the problem, but $\mathcal{R}$ is always a set of scalars. The dynamics of the environment are captured by a function $p : \mathcal{S} \times \mathcal{R} \times \mathcal{S} \times \mathcal{A} \rightarrow [0, 1]$ which tells us the probability of transitioning from one state to another, and receiving a particular reward, after taking an action.

$$p(s', r | s, a) \equiv \Pr(S_{t+1} = s', R_{t+1} = r | S_t = s, A_t = a) \quad (1.22)$$

This formulation allows the environments to be quite general and naturally accounts for uncertainty in the consequences of actions. Deterministic environments are handled by setting all probabilities to either 0 or 1. The agent makes decisions based on its policy $\pi : \mathcal{S} \times \mathcal{A} \rightarrow [0, 1]$

$$\pi(a \mid s) \equiv \Pr(A_t = a \mid S_t = s) \quad (1.23)$$

which gives the probability that action $a \in \mathcal{A}$ is chosen given that the agent is in state $s \in \mathcal{S}$ at time t . As with the dynamics, deterministic policies are handled by setting all probabilities to 0 or 1. A useful concept in reinforcement learning is that of discounted future returns. The discounted future return G_t tells us the accumulated reward from time t onward.

$$G_t = R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + \dots \quad (1.24)$$

γ is called the discount rate and determines how much to value present rewards against those in the future. Using the returns we can define what is known as a value function $V_\pi : \mathcal{S} \rightarrow \mathbb{R}$ for a given policy π . The value function tells us the expected future return if the agent starts in state s and follows policy π for the rest of its interactions with the environment.

$$V_\pi(s) = \mathbb{E}[G_t \mid S_t = s] \quad (1.25)$$

Ultimately reinforcement learning algorithms are used to find the optimal policy π_* which maximizes the value of each state. There are a wide variety of methods to choose from and in our work we use policy optimization algorithms.

1.2.2 POLICY OPTIMIZATION

We describe the simplest policy optimization algorithm to give a sense of how it works. Our derivation follows² and we suggest looking there for a helpful discussion of extensions.

Suppose we have a finite horizon (all agent-environment interactions terminate at some point), undiscounted ($\gamma = 1$) MDP. Let $\tau = (s_0, a_0, \dots, s_{T+1})$ be a trajectory of states and actions an agent follows in its interaction with the environment. Also suppose that rewards are deterministic so we can write $p(s', r \mid s, a) = p(s' \mid s, a)$. In general the policy will be some complicated function of the states and actions so we parameterize it with parameters θ and write $\pi = \pi_\theta$ to indicate that. Our goal is to maximize the expected return

$$J(\pi_\theta) = \mathbb{E}_{\tau \sim \pi_\theta} [G(\tau)] \quad (1.26)$$

where (with a slight abuse of notation) we have written the return along trajectory τ as $G(\tau)$. Note

$\mathbb{E}_{\tau \sim \pi_\theta}$ means we are taking an expected value over trajectories found by following policy π_θ .

We start out not knowing what the optimal policy looks like. The perspective in reinforcement learning is to use interaction with the environment (i.e. experience) to update the parameters θ towards those of the optimal policy. To find the parameters we simply want to use gradient ascent

$$\theta_{k+1} = \theta_k + \alpha \nabla_\theta J(\pi_\theta) \quad (1.27)$$

where α is known as a learning rate and dictates the size of the parameter updates made at each step.

The problem is then how to estimate $\nabla_\theta J(\pi_\theta)$ through experience. We know the probability of a trajectory given the current parameters is

$$P(\tau \mid \theta) = \prod_{t=0}^T p(s_{t+1} \mid s_t, a_t) \pi_\theta(a_t \mid s_t) \quad (1.28)$$

which just states that the probability of a trajectory is the probability of taking the observed actions times the probability of making the observed state transitions. From there we can use the facts that $\mathbb{E}_{\tau \sim \pi_\theta} [G(\tau)] = \int P(\tau | \theta) G(\tau) d\tau$ and state transition probabilities don't depend on θ to derive the following

$$\nabla_\theta J(\pi_\theta) = \nabla_\theta \int P(\tau | \theta) G(\tau) d\tau \quad (1.29)$$

$$= \int \nabla_\theta P(\tau | \theta) G(\tau) d\tau \quad (1.30)$$

$$= \int P(\tau | \theta) \nabla_\theta \log(P(\tau | \theta)) G(\tau) d\tau \quad (1.31)$$

$$= \mathbb{E}_{\tau \sim \pi_\theta} [\nabla_\theta \log(P(\tau | \theta)) G(\tau)] \quad (1.32)$$

$$= \mathbb{E}_{\tau \sim \pi_\theta} \left[\sum_{t=0}^T \nabla_\theta \log(\pi_\theta(a_t | s_t)) G(\tau) \right] \quad (1.33)$$

where in eq. (1.31) we have used $\frac{df(x)}{dx} = f(x) \frac{d}{dx} \log(f)$ as long as $f(x)$ is greater than zero (which it is in our case since we are integrating over observed trajectories making all $P(\tau | \theta) > 0$). The important part about the above derivation is that we can estimate eq. (1.33) from experience in the environment. We specified the policy and can compute the gradient with respect to its parameters, and we can determine $G(\tau)$ by simply observing the returns along a trajectory. By repeatedly observing the agent follow policy π_θ in the environment we can estimate eq. (1.33) using a sample average and update the parameters using eq. (1.27).

In practice we need to ensure our parameterization π_θ is robust enough to represent complicated policies. Most recent development in RL has come from using deep neural networks⁴⁰ to approximate π_θ because of their ability represent general functions. This branch of RL is known as deep reinforcement learning (DRL). We use StableBaselines3⁷⁸, a thoroughly tested and benchmarked software package, for implementations of DRL algorithms.

2

Optimal strategies for Kiiking: pumping a swing till inversion

Kiiking, an extreme version of swinging, is a competitive sport invented in Estonia. It uses an especially long swing (Estonian: *kiik*), where a platform is connected to the swing fulcrum by rigid bars of adjustable length, Fig. 2.1(a). The athlete stands on the platform and pumps the swing by appropriately timed standing and squatting, with the goal of inverting the swing and passing over

the fulcrum, thereby completing a full rotation. Athletes compete to achieve a full rotation on the longest swing possible, potentially within the shortest time. As in any sport, we see a natural coming together of neurobiology and physics that raises a number of questions: what is the relative importance of timing and power to invert the swing as a function of the underlying physical constraints and the biological constraints? What does the optimal pumping strategy look like? How can an individual learn this strategy from experience? In this letter, we address some of these questions.

Conventional playground swings modeled as pendula serve as models of one of the simplest models in elementary physics with connections to everything from Josephson junctions to lasers and active matter. A child sitting on a swing actively drives it by leaning back and forth, and thus pumps it to reach but modest amplitudes^{25,24} but unlimited pleasure. In contrast, kiiking consists of standing and squatting, i.e. raising and lowering the center of mass, Fig. 2.1(a), as this method is more efficient at high amplitudes⁹⁸. This corresponds well to a model of a pendulum with time-varying length l , a well studied model especially in the linear (low amplitude) regime, as a prototypical example of a parametric oscillator^{21,18}. In this limit, if the length is periodically varied with a frequency which is approximately twice the natural frequency of the pendulum, the oscillation amplitude grows exponentially. These approaches do not address the interesting question of optimal strategies for modulating the length or how this might be learnt.

A step in this direction takes the perspective of optimal control theory, e.g.⁵⁵, where the authors consider a half-period and maximize the angle θ of the swing at its highest point, or minimizing the time needed to reach a given target angle or target potential energy^{74,51,75,59}. Assuming that the length l can change discontinuously leads to the following intuitive result: stand up at the lowest point of the swing, and squat at the highest point.

But the rate at which the athlete can stand and squat is limited, especially at higher angular speeds when the athlete can be subject to 3-4 g of acceleration, limiting the applicability of these simple models to reality. A natural way forward is suggested by combining the analysis of Youtube videos

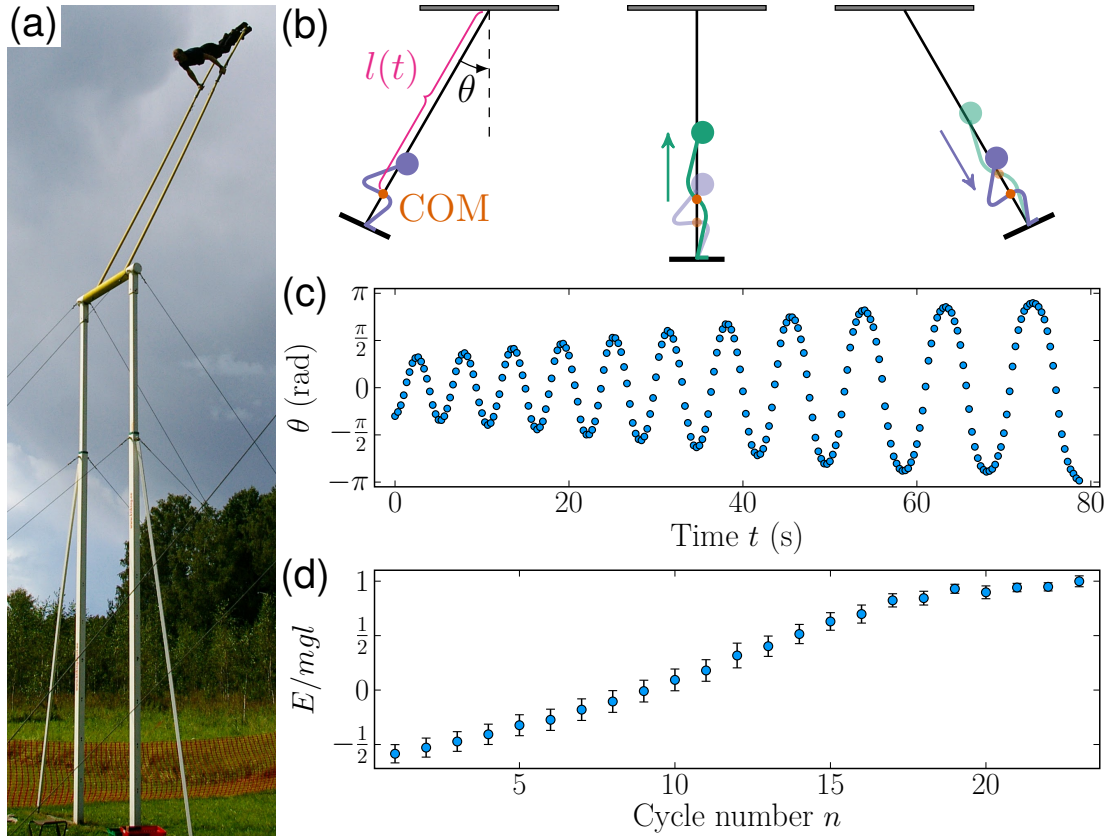


Figure 2.1: (a) An athlete on a tall kiiking swing. The arms of the swing are rigid, and the athlete is attached to the base by their feet. Adapted from Estonian Kiiking Association⁸, ©2014 Estonian Kiiking Association. (b) A schematic of an athlete pumping the kiiking swing by squatting and standing. The swing-athlete system is modeled as a simple (point-mass) pendulum with variable length $l(t)$. Over the course of one cycle, the athlete quickly stands up when the swing is near its lowest point, and squats down as the swing reaches its highest point. This stand-squat cycle is repeated twice over each oscillation of the swing's motion. (c) The angle θ measured from experiments as a function of time for a successful kiiking attempt. The length of the swing is approximately 7 meters. (d) Nondimensionalized energy $E/mgl_0 \approx -\cos \theta$ after the n th cycle, i.e. at the n th extremum of θ .

of kiiking to extract time series, simple estimates of constraints on human athletic performance, e.g. maximum power exerted, and a minimal model of the sport in terms of an extensible controllable pendulum. By posing the problem in terms of optimal control theory, we compute the time-optimal control for the system in the form of a feedforward (open-loop) controller. Complementing this from the perspective of reinforcement learning (RL), we consider how an athlete or other intelligent agent might learn to complete the kiiking task from repeated attempts.

2.1 EXPERIMENTAL DATA ANALYSIS

In order to develop a quantitative understanding of kiiking strategies we found videos of kiiking online and took snapshots of the videos at a rate of 3 Hertz. For each frame we used the Fiji software platform⁸⁵ to measure the angle the swing makes with the vertical. Figure 2.1(c) shows how the angle grows over time as the athlete makes a full swing up to 180° (see appendix A for more detail).

Our model requires we make a few key estimates of human performance. In particular we need to estimate the maximum rate at which a person can stand under a load and the maximum power an athlete can output. We estimate the maximum rate of standing from the video data and find it to be around 140 cm/s (we assume the athlete is squatting half their height and count the number of frames it takes to do so). We use a maximum power value of around 850 Watts taken from measurements on leg power generated during dynamic exercise⁸³. We use the power and rate of standing estimates to set appropriately set simulation parameters as we search for optimal swinging strategies.

2.2 MATHEMATICAL MODEL

We model the kiiking system as an idealized pendulum consisting of a point mass m connected to a pivot by a massless rigid rod of variable length l . In this model, m represents the combined mass of the swing arms and the person riding on the swing, while l corresponds to the distance from the

pivot to the center of mass. We denote by θ the angle between the pendulum and the downward vertical direction. We neglect any motion of the center of mass perpendicular to the swing arms, as well as friction and air resistance.

The system is shown schematically in Figure 2.1(b). The pendulum length l is bounded between some minimum length l_- and maximum length l_+ ($0 < l_- < l_+$), corresponding to squatting and standing, respectively.

The kinetic and potential energy of the system are $T = \frac{1}{2}m[(\dot{l}\theta)^2 + \dot{l}^2]$ and $V = -mgl \cos \theta$ respectively. Defining the conjugate momentum $p = ml^2\dot{\theta}$, the evolution of the state $\mathbf{x} = (\theta, p, l)$ is given by Hamilton's equations

$$\begin{aligned}\dot{\theta} &= \frac{1}{ml^2}p, \\ \dot{p} &= -mgl \sin \theta, \\ \dot{l} &= u,\end{aligned}\tag{2.1}$$

where u is the rate of change of l , which we take as the control variable. We nondimensionalize the system (2.1) by choosing units so that $m = l_0 = t_0 = 1$ where $l_0 = (l_+ + l_-)/2$ is a characteristic length scale and $t_0 = \sqrt{l_0/g}$ is a corresponding time scale. We write the nondimensionalized equations in the form

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}, u) = \mathbf{F}(\mathbf{x}) + \mathbf{G}(\mathbf{x})u, \quad \mathbf{x}(0) = \mathbf{x}_0, \\ \mathbf{F}(\mathbf{x}) &= \begin{pmatrix} p/l^2 \\ -l \sin \theta \\ 0 \end{pmatrix}, \quad \mathbf{G}(\mathbf{x}) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\end{aligned}\tag{2.2}$$

to emphasize that the system is affine in the control u . For simplicity, we only consider initial conditions of the form $\mathbf{x}_0 = (\theta_0, 0, l_+)$ where $\theta_0 \neq 0$.

Our goal is to find a control u so that the system reaches the target set $S = \{(\theta, p, l) \mid \theta = \pm\pi\}$ in minimum time, subject to certain constraints on u and the trajectory $\mathbf{x}(t)$, which will be detailed

below. The control may be given in the form of an open-loop control $u = u(t)$, or else as a feedback control policy $u = \pi(\mathbf{x})$. Importantly, the bound $l_- \leq l \leq l_+$ means that, for most initial conditions, any control which steers the system to S must involve several *cycles* of squatting followed by standing. These individual cycles will be analyzed first before turning to the full optimal control problem.

We can gain some intuition about the system by inspecting its total energy $E = T + V$. The rate of change of the (nondimensionalized) energy is

$$\dot{E} = u\dot{u} - (p^2/l^3 + \cos \theta)u. \quad (2.3)$$

We see that the integral of the first term vanishes over a cycle, so we consider only the second term. For the system to gain energy, we must take $u < 0$ (corresponding to standing up) when the mass is near its lowest point, so that $p^2/l^3 + \cos \theta$ is maximal. Conversely, we should take $u > 0$ near the highest point in order to minimize energy losses during the squatting phase. In our chosen units, the minimum energy needed to reach the target set S is simply $E = l_-$.

We make the natural assumption that the rate $u = \dot{l}$ at which the length of the pendulum can be changed is bounded by some maximum, $|u| \leq u_m$. Motivated by equation (2.3), we also impose a *power bound* of the form $-(p^2/l^3 + \cos \theta)u \leq P_m$ for some $P_m > 0$. Note that this bound does not take into account the contribution $u\dot{u}$ to $\dot{E}$. This is done for the sake of simplicity; however, this term is usually small compared to the other two terms when the constraint is active, except on very short times when u has a jump discontinuity. The two constraints imply that $u^- \leq u \leq u^+$ where

$$u^\pm = u^\pm(\mathbf{x}) = \pm \min \left(u_m, \frac{\mp P_m}{p^2/l^3 + \cos \theta} \right) \quad (2.4)$$

Finally, we define the dimensionless parameter $\Delta l := (l_+ - l_-)/2$ so that l is bounded between

$1 - \Delta l$ and $1 + \Delta l$. For kiiking athletes, typical ranges for the dimensionless parameters are $0.04 \leq \Delta l \leq 0.08$, $0.1 \leq u_m \leq 0.15$, and $0.1 \leq P_m \leq 0.25$. We remark that in the limit $P_m \rightarrow \infty$, $u_m \rightarrow \infty$, we obtain the problem considered in ^{59,75}. Figure 2.2 shows numerical simulation results for the standing phase of a single cycle. At $t = 0$, the swing starts from rest at an initial angle θ_i and remains in the lowest position ($l = l_+$) until a *switching time* t_s , when u is taken to be the minimum value allowed by the rate and power constraints, i.e. $u(t) = u^-$ where u^- is defined in equation (2.4). This $u(t)$ is maintained until l reaches the minimum length l_- , after which $u(t)$ is identically zero. In particular, $u(t)$ is piecewise constant in the absence of power constraints ($P_m = \infty$). Figure 2.2 (a) shows some representative solutions and the corresponding controls. The angle θ at the end of the standing phase, and thus the energy gained by the system, is seen to depend sensitively on the switching time t_s . The dependence of the final energy E_f on the switching time is shown explicitly in Figure 2.2 (b). We denote by t_s^g the *greedy switching time*, which maximizes the energy at the end of the standing phase. In the limiting cases $P_m \rightarrow \infty$, $u_m \rightarrow \infty$, the greedy switching time is precisely when $\theta = 0$, i.e. when the swing reaches its lowest point. This is in agreement with previous results ^{59,75}. Since the period of a pendulum increases with amplitude, t_s^g increases monotonically with θ_i in the case when $P_m = \infty$. In the power-constrained case, the dependence is more complicated due to a competing effect: As θ_i increases, the power constraint becomes more strict, leading to smaller $|u|$ and earlier switching times t_s^g .

The preceding discussion suggests the following *greedy* control strategy: Let $u(t) = u^{(i)}(t)$ for $t^{(i)} \leq t < t^{(i+1)}$ ($i = 1, \dots, N$) where the $u^{(i)}$ switch cyclically between the $u = u^-$, $u = 0$ and $u = u^+$ and the switching times $t^{(i)}$ are chosen greedily, i.e. in order to maximize the energy at the end of each stand-squat cycle. By construction, this strategy achieves the task of reaching S in the fewest number of cycles. Moreover, it approximates the behavior observed from kiiking athletes. As we will show, the time-optimal solution has the same structure as the greedy strategy, but with slightly different switching times. We will now precisely pose the time-optimal control problem.

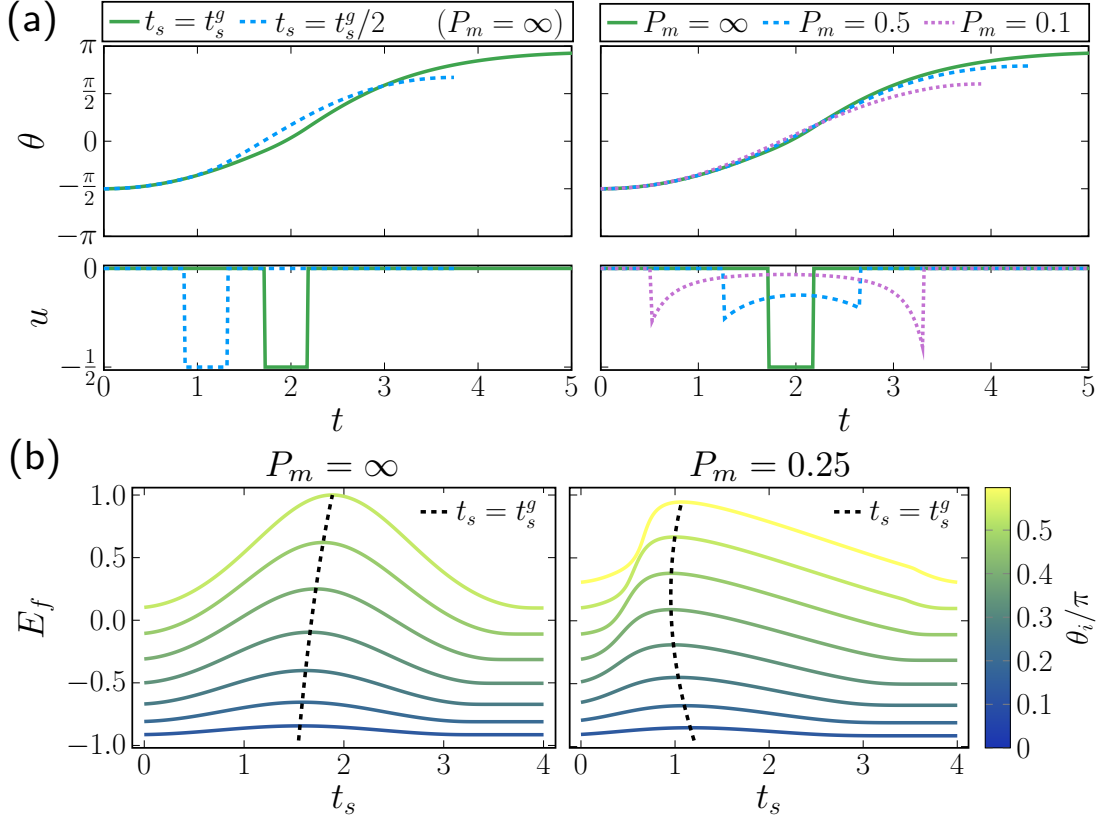


Figure 2.2: Analysis of the standing phase of a stand-squat cycle. Starting at the switching time $t = t_s$, the athlete decreases the swing length l at the maximum allowed rate $u = u^-$ (Eq. (2.4)) until l reaches the minimum value l_- . (a) The angle θ and control $u = \dot{l}$ as functions of (nondimensionalized) time. Left: The switching time t_s is varied in the absence of power constraints ($P_m = \infty$). The greedy switching time t_s^g (solid green lines) maximizes the energy gained by the system. Right: Solutions with $t_s = t_s^g$ and varying power bounds P_m . In both cases, $\Delta l = 0.115$, $u_m = 0.5$. (b) Energy E_f at the end of a cycle as a function of the switching time t_s . Each solid curve corresponds to a different initial angle $\theta_i = \theta(0)$. The dashed black curve indicates the greedy switching time t_s^g for each θ_i . When the system is not power-constrained (left), t_s^g increases monotonically with θ_i . In the presence of power constraints (right), the relationship is more complicated. Here, $\Delta l = 0.1$, $u_m = 1.0$.

2.3 TIME-OPTIMAL CONTROL

We say that a measurable function $u: \mathbb{R}_+ \rightarrow U$, $U = [-u_m, u_m]$, is an *admissible* control if u and the corresponding state trajectory x , solution of (2.1), satisfy the inequality constraints

$$g(\mathbf{x}(t), u(t)) \geq 0, \quad \mathbf{h}(\mathbf{x}(t)) \geq 0, \quad (2.5)$$

where

$$g(\mathbf{x}, u) = P_m + (p^2/l^3 + \cos \theta)u, \quad \mathbf{h}(\mathbf{x}) = \begin{pmatrix} l_+ - l \\ l - l_- \end{pmatrix}.$$

Define the terminal time $t_f = t_f(u)$ as the first time that the corresponding trajectory $\mathbf{x}$ hits the target set $S = \{(\theta, p, l) \mid \theta = \pm\pi\}$. Then the time-optimal control problem (TOCP) for the kiiking system can be stated as follows: Given an initial state $\mathbf{x}_0$, find an admissible control $u \in \mathcal{U}$ which minimizes t_f subject to the state constraints (2.5).

The presence of the pure state constraints, represented by $\mathbf{h}(\mathbf{x})$ in equation (2.5), makes this a difficult problem to solve by variational methods. A set of necessary conditions for a control-trajectory pair to solve the optimal control problem are provided by the Pontryagin Maximum Principle (PMP). A sufficiently general version of the PMP, which allows for both mixed control-state inequality constraints as well as pure state constraints, is given by Sethi⁸⁷. We will only need a subset of the conclusions, which we summarize now. Define the control Hamiltonian associated with the TOCP as

$$H(\mathbf{x}, \lambda, u) = \lambda \cdot (\mathbf{F}(\mathbf{x}) + u\mathbf{G}(\mathbf{x})).$$

The time derivative of the constraint function $\mathbf{h}$ is

$$\mathbf{h}^1(\mathbf{x}, u) = \nabla \mathbf{h}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}, u) = (-u, u)^T$$

and we define the restricted control set

$$\begin{aligned}\tilde{U}(\mathbf{x}) &:= \{u \in U \mid g(\mathbf{x}, u) \geq 0 \text{ and} \\ b_i^1(\mathbf{x}, u) &\geq 0 \text{ if } b_i(\mathbf{x}) = 0, \ i = 1, 2\}.\end{aligned}\tag{2.6}$$

The PMP states that if $u^*: [0, t_f] \rightarrow U$ is an optimal control and $\mathbf{x}^*$ is the corresponding trajectory, then there exists a nowhere vanishing function $\lambda^*: [0, t_f] \rightarrow \mathbb{R}^3$ (the *costate*) so that at any time t the function $u \mapsto H(\mathbf{x}^*(t), \lambda^*(t), u)$ attains its maximum on the restricted control region $\tilde{U}(\mathbf{x}^*(t))$ at $u = u^*(t)$.

In our case, the control Hamiltonian $H(\mathbf{x}, (\lambda), u) = \lambda \cdot \mathbf{F}(\mathbf{x}) + u\lambda_3$ depends on u only through the term $u\lambda_3$. The coefficient $\phi(t) = \lambda_3(t)$ multiplying u is called the *switching function* and determines the structure of the optimal control. On *interior arcs*, i.e. at times t when the constraint $\mathbf{h}$ is inactive, the PMP states that $u^*(t) = u^+$ if $\phi(t) > 0$ and $u^*(t) = u^-$ if $\phi(t) < 0$. In other words, u^* changes between the maximum and minimum value at *switching times* when ϕ switches sign. This completely determines $u^*(t)$ on interior arcs unless a *singular arc* occurs, i.e. if $\phi(t)$ vanishes on a nontrivial interval. We find that singular arcs do not appear in our problem, so we do not consider them further. A time interval $[t_1, t_2]$ on which the constraint $\mathbf{h}$ is active is called a *boundary arc* and its endpoints t_1, t_2 are called *junction times*. We see that in the interior of a boundary arc, we necessarily have $u^*(t) = 0$.

Computing the costate λ (and thus the switching function) requires solving a difficult nonlinear multi-point boundary value problem. Alternatively, the above observations completely determine $u^*(t)$ as long as we know the switching times and junction times. Following Maurer et al.⁶⁴, we transcribe the time-optimal control problem into a finite-dimensional optimization problem over the switching and junction times, which can be solved using standard nonlinear programming methods. By using a modern numerical ODE solver with support for automatic differentiation⁶¹,

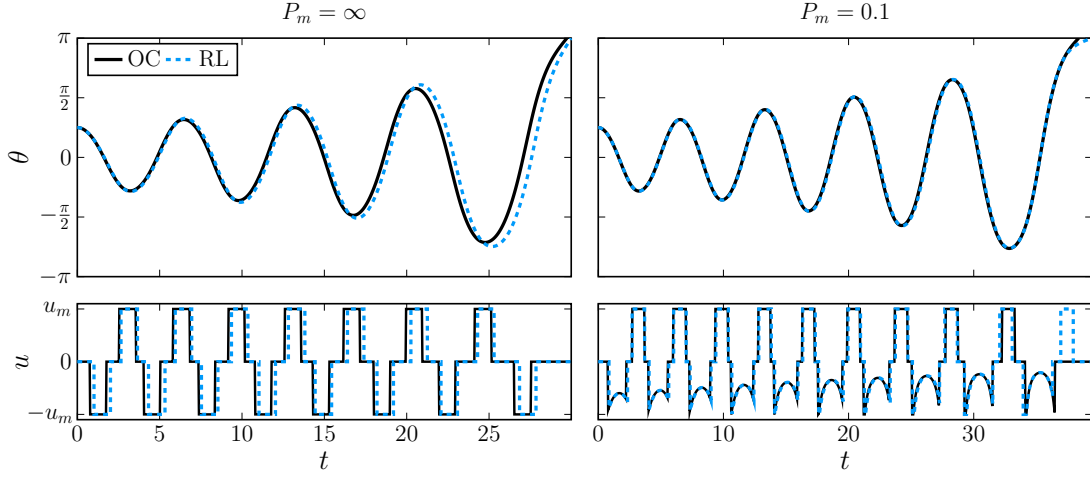


Figure 2.3: Time-optimal solution of the kiiking problem. The angle θ and control $u = \dot{l}$ as functions of (nondimensionalized) time for the numerically computed optimal control solution (OC, black lines) and the solutions obtained from reinforcement learning (RL, blue lines). Left: When the system is not power-constrained, the OC solution consists of several interior arcs where $u = \pm u_m$, separated by boundary arcs where $|l - 1| = \Delta l$ and $u = 0$. Right: In the presence of power constraints, u takes the value $u^\pm$ (equation (2.4)) on the interior arcs. In both cases, the RL solution has the same form but with slightly different switching times. For both simulations, $\theta_0 = \pi/4$, $\Delta l = 0.05$, $u_m = 0.1$.

we can compute the gradient of the terminal time t_f with respect to the switching times, and apply an efficient gradient-based method. Using this approach, we are able to determine the switching and junction times to very high precision. We check the validity of the solution by comparing with a direct collocation method¹⁹. The numerical solution codes are provided in the supplemental information²⁹.

The optimal control and corresponding $\theta(t)$ are shown in Figure 2.3 (black lines). We find that the optimal control has the same structure as the greedy strategy defined in the previous section, and differs only slightly in the switching and junction times.

2.4 REINFORCEMENT LEARNING

Reinforcement learning (RL) provides a set of tools to solve sequential decision making problems. In our case reinforcement learning provides an expedient method of formulating kiiking as a feed-

back control problem. RL methods do not provide the same optimality guarantees as control theory methods, so it is important we compare the solutions to see how close to optimal the learned solution is.

To formulate a reinforcement learning problem⁹² we need to define a state space $\mathcal{S}$, an action space $\mathcal{A}$, and a reward signal $\mathcal{R}$. The state space be the state of the system is the same as in 2.1 $S_t = (\theta_t, p_t, l_t)$.

We assume that athletes either try to stand or squat as quickly as possible and otherwise do nothing. The action space is then $\mathcal{A} = \{-u_m, 0, u_m\}$. Where u_m is the maximum rate the athlete can stand or squat. At every time step the RL agent chooses an action and we integrate the dynamics in equation 2.1 forward in time according to that action. In the power bounded version of the problem we ensure that the bounds are not exceeded at each time step.

The goal is to train the agent to swing up to $\theta = \pi$ as quickly as possible. In principle we should give a reward that is proportional to the time it takes to reach the goal. In practice we supplement the time based reward with a reward which encourages incremental progress. In particular the reward R_t at time t is

$$R_t = \begin{cases} 1 & \text{if } \theta = \pi \\ -1 + E_t/E_{max} & \text{otherwise} \end{cases} \quad (2.7)$$

where E_{max} is the gravitational potential energy of the swing at $\theta = \pi$ and E_t is the energy at time t . Energy increases as the swing cycles become larger in amplitude so an energy based reward is similar, but not identical, to a time based reward (see Fig 2.2). Using informative intermediate rewards is a practice known as reward shaping and is frequently used in reinforcement learning problems⁶⁷. We found no improvement in performance after starting with the hybrid time-energy optimal agent and retraining it with the time based reward. As such we present results using the hybrid reward.

For this problem we parameterized the policy as a neural network and use the Proximal Policy

Optimization (PPO)⁸⁶ algorithm, a variant of traditional policy optimization, to update the network weights towards the optimal policy. We used the PPO implementation from Stable Baselines 3⁷⁸, a thoroughly tested software package with implementations of various reinforcement learning algorithms.

2.5 RESULTS

Figure 2.3 compares solutions obtained by the optimal control (OC)-based method to solutions found by the reinforcement learning (RL) algorithm. The solutions are identical apart from minor differences in switching and junction times. In particular, both solutions agree qualitatively with the general strategy of kiiking athletes, namely standing up as quickly as possible as θ passes through 0 and squatting near zeros of $\dot{p}$ (the extrema of θ).

To validate our model, we compare the computed OC solutions to data collected from trials by five different kiiking athletes. Since the length of the swing arms and the mass of the athletes are unknown, the $\theta(t)$ time series cannot be directly compared to trajectories of our nondimensionalized model. Instead, we compute an estimate of the nondimensionalized energy $E(n)$ at the end of each stand-squat cycle n . The energy is estimated when the swing is at its highest point, so that the kinetic energy vanishes and the nondimensionalized energy is simply $E = -\cos \theta$. The results are shown in figure 2.4. We see that for the longer trials, the $E(n)$ curves have a distinctive sigmoidal shape. Recalling equation (2.3), we see that the energy gain is initially limited by the small term $(p^2/l^3 + \cos \theta)u \approx p^2u$. At larger velocities, the energy lost to air resistance over the course of a cycle becomes comparable to the energy gained by standing and squatting, and the $E(n)$ curve levels off. In order to capture these qualitative features, we modify our model Eq. (2.1) by adding a quadratic drag term:

$$\dot{p} = -l \sin \theta - d|p|p/l, \quad (2.8)$$

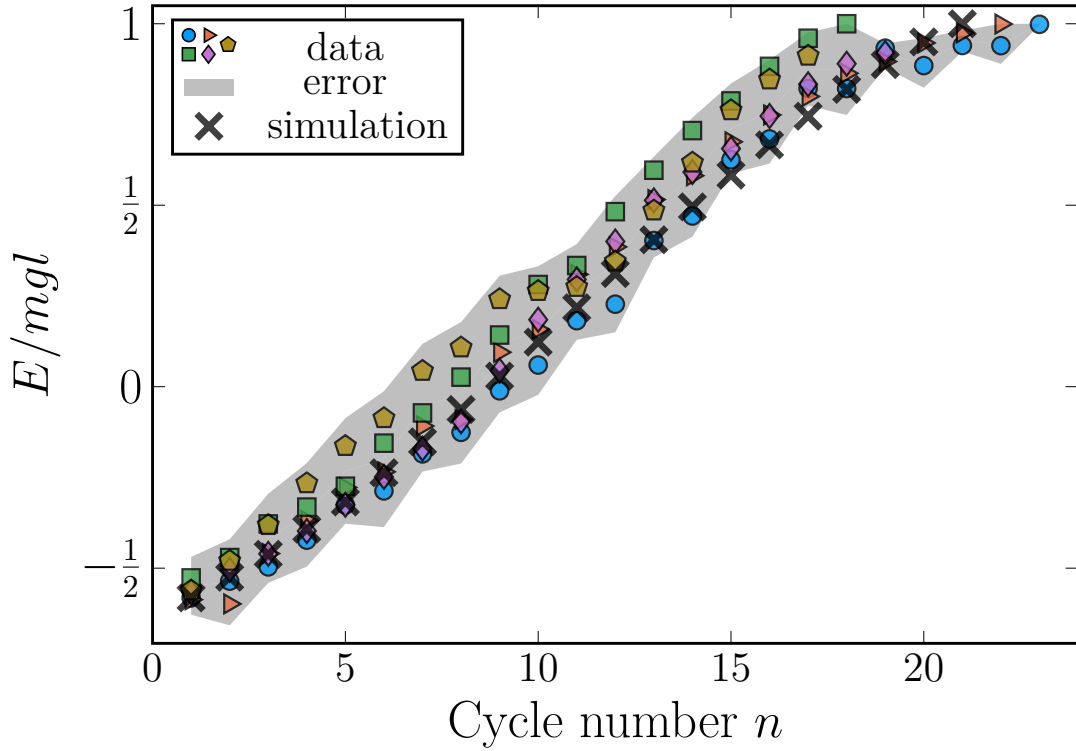


Figure 2.4: Nondimensionalized energy E/mgl at the end of each stand-squat cycle for five trials by different kiiking trials. The gray ribbon shows the envelope of the error bars for all five trials. In the two longest trials (circles and triangles) the length of the swing arms is approximately 7m. For the other trials, the length is smaller but not precisely known. Two of the series (diamonds and pentagons) show failed attempts, which ended without reaching $\theta = \pi$. A simulation with parameters $\Delta l = 0.05$, $u_m = 0.125$, $P_m = 0.125$, $d = 2.75 \times 10^{-2}$ is shown for comparison, and captures the sigmoidal shape of the $E(n)$ curves.

where d is a dimensionless constant. Figure 2.4 shows a typical simulation result. The values of the parameters were chosen to correspond to a challenging kiiking exercise, requiring over twenty stand-squat cycles. The corresponding $E(n)$ curve has a sigmoidal shape similar to the data.

2.6 DISCUSSION

In this work we studied the Estonian sport of kiiking. The problem of optimal swinging has been studied before but we go beyond prior work by explicitly including physiological limitations, de-

veloping both closed and open loop controllers, and by providing quantitative comparisons with data.

3

Planar controlled gliding through shape change

Change of configuration is ubiquitous in both powered and gliding flight⁴¹. The incredible maneuverability of birds²⁷ is due to their ability to adeptly change their aerodynamic and inertial properties in the air^{43,97}. Manned and unmanned aerial vehicles (UAVs) can not fly as dynamically as birds largely due to their relatively fixed shape. There is a long history¹³ of trying to bring together the

proper materials and control systems to extend aircraft capabilities through shape change^{26,42,50,89}. Inspired by this prior work and natural gliders^{11,65,101,102,103,104}, we examine the usefulness of dynamic shape change in the simplified context of wingless gliding bodies. In particular we study how to use shape change to complete perching (i.e. landing at a specific spot in a prescribed orientation) tasks. We study the problem in a two dimensional setting and introduce a method for varying aerodynamic properties. We show how our method allows us to turn what would be a haphazard and chaotic descent into precisely controlled flight. Our work may inspire atmospheric or underwater vehicles which more fully incorporate the shape of the body into their controls.

First we introduce the glider model which is motivated by the dynamics of falling bodies at intermediate Reynolds number^{36,52,62} (e.g leaves or cards). After introducing the model we show precisely how shape change affects the inertial and aerodynamic properties of the falling body. We then formulate the problem of perching as an optimal control problem. Posing the problem in the language of optimal control theory allows us to select the strategies, of which there may be many, which are optimal according to some objectives that are secondary to the goal of perching³. After formulating the problem we present the solutions and compare strategies with different optimization objectives. We end by pointing out some natural extensions of the work.

3.1 MATHEMATICAL MODEL

Following^{4,5} we obtain a model for the glider by considering a long and thin cylinder falling due to gravity in an inviscid, incompressible fluid. We ignore the motion along the long axis and characterize the dynamics by the position and orientation of the cylinder head. Figure 3.1 shows the dynamic variables. The speed along the semi-major and semi-minor axes are u and v respectively. The angle θ measures the angle between the semi-major axis and the fixed x axis. The aspect ratio $\beta = b/a$ measures the ratio of the semi-minor and semi-major axis radii. We change the shape of the body by

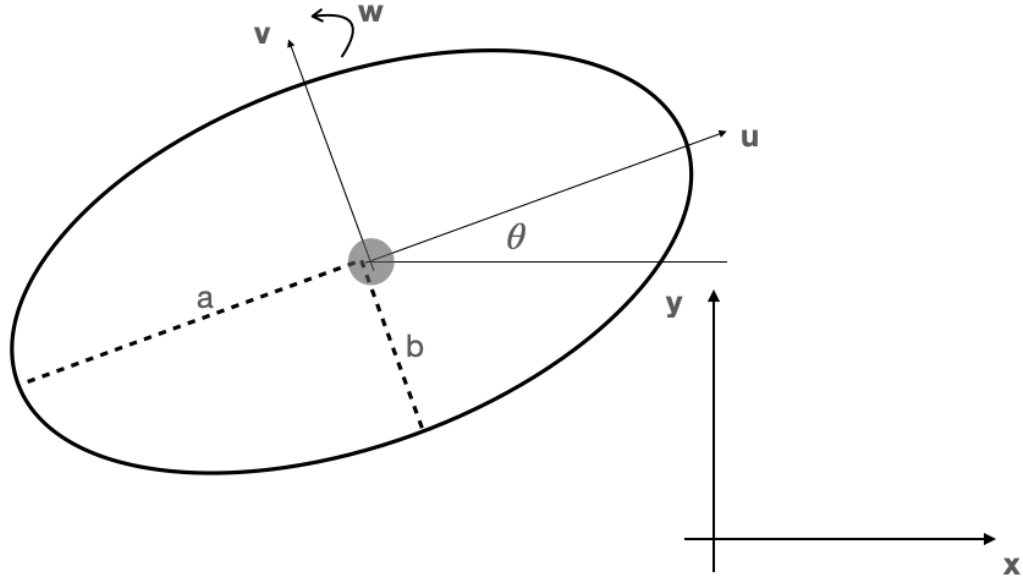


Figure 3.1: Model of the glider: u is the velocity along the longitudinal axis of the body. v is the velocity along the transverse axis. w is the angular velocity. a and b are the radii of the semi major and semi minor axes. The x and y axes are shown in the bottom and are fixed while the u and v axes rotate with the body. We consider the problem of how such a body can glide through a quiescent fluid and perch at a given location. We assume control of the shape of the ellipse characterized by the aspect ratio $\beta = \frac{b}{a}$.

changing β .

The motion of the glider and the fluid are coupled so in general we need to solve the Navier-Stokes⁷³ equation to understand the gliders' motion over time. Instead of solving the Navier Stokes equations Andersen et. al⁵ presented a reduced order model which captures the dynamics and makes the model more realistic by accounting for viscous drag. We use the same model here. The equations describing the system are given below (for convenience we use notation where $\dot{u} = du/dt$).

$$(m + m_1)\dot{u} = (m + m_2)vw - \rho_f \Gamma v - \pi(\rho_s - \rho_f)abg \sin \theta - F \quad (3.1)$$

$$(m + m_2)\dot{v} = -(m + m_1)uw + \rho_f \Gamma u - \pi(\rho_s - \rho_f)abg \cos \theta - G \quad (3.2)$$

$$(I + I_a)\dot{w} = (m_1 - m_2)uv - M \quad (3.3)$$

$$\dot{x} = u \cos \theta - v \sin \theta \quad (3.4)$$

$$\dot{y} = u \sin \theta + v \cos \theta \quad (3.5)$$

$$\dot{\theta} = w \quad (3.6)$$

With the mass and moment of inertia terms given by

$$m = \rho_s \pi ab \quad (3.7)$$

$$m_1 = \rho_f \pi b^2 \quad (3.8)$$

$$m_2 = \rho_f \pi a^2 \quad (3.9)$$

$$I = \frac{\rho_s \pi ab}{4} (a^2 + b^2) \quad (3.10)$$

$$I_a = \frac{\rho_f \pi}{8} (a^2 - b^2)^2 \quad (3.11)$$

F , G , Γ , and M are given by

$$\Gamma = \frac{2}{\pi} \left[-C_T \frac{uv}{\sqrt{u^2 + v^2}} + C_R w \right] \quad (3.12)$$

$$F = \frac{1}{\pi} \left[A - B \frac{u^2 - v^2}{u^2 + v^2} \right] \sqrt{u^2 + v^2} u \quad (3.13)$$

$$G = \frac{1}{\pi} \left[A - B \frac{u^2 - v^2}{u^2 + v^2} \right] \sqrt{u^2 + v^2} v \quad (3.14)$$

$$M = [\mu_T + \nu_T |w|] w \quad (3.15)$$

and constants $A = 1.4$, $B = 1.0$, $C_T = 1.2$, $C_R = \pi$, $\mu_T = \nu_T = 0.2$. Note m_1, m_2 and I_a are added mass and moment of inertia terms arising from the fact that in order to accelerate a body through a fluid some of the fluid must be displaced.

The first term in the $\dot{u}$ and $\dot{v}$ equations account for acceleration due to u and v being rotating axes. Viscous forces are modeled by F and G . Lift is due to fluid circulation Γ around the body. Lift is proportional to Γ and generated by motion orthogonal to a given axis (e.g lift in the u direction is generated by motion transverse to u)¹. Finally there are gravitational forces that determine whether the body will rise or fall in the absence of any other forces. M characterizes fluid torques on the body. Andersen et al⁴ found expressions for these forces and torques using a combination of experimental and numerical studies.

This model captures the dynamics of falling plates at intermediate Reynolds number⁵. There are two main modes of motion seen in these system. First is the oscillatory side to side motion known as fluttering. Second is tumbling where the body rotates rapidly and experiences some net horizontal displacement. Prior work^{71,70} on the evolution of flight^{31,32} studied controlled gliding with this model by adding in a torque control τ in the angular acceleration equation. Instead of supposing the glider can directly effect some torque, we presume the body can change shape and study the dynamics arising from this new flight control.

3.2 THE CONTROL PROBLEM

The goal is to determine how to control the shape of the glider so it arrives at a specific point and orientation in space. We control the shape by changing $\beta = b/a$. Figure 3.1 demonstrates that a small β corresponds to a cylinder elongated on the u axis and increasing β elongates the glider on the v axis.

Let $\alpha = ab$. Since we require that the mass of the glider is unchanged, then $a = \sqrt{\frac{\alpha}{\beta}}$ and $b =$

$\sqrt{\alpha\beta}$. With these definitions

$$\begin{aligned} m &= \rho_s \pi \alpha \\ m_1 &= \rho_f \pi \alpha \beta \\ m_2 &= \frac{\rho_f \pi \alpha}{\beta} \\ I &= \frac{\rho_s \alpha^2}{4} \left(\frac{1 + \beta^2}{\beta} \right) \\ I_a &= \frac{\rho_f \pi \alpha^2}{8} \left(\frac{1 - \beta^2}{\beta} \right)^2 \end{aligned}$$

We see that β influences the dynamics through its affect on the added masses, added moment of inertia, and moment of inertia. As the glider becomes elongated along either of its axes the moments of inertia grow and it becomes more difficult to rotate. At the same time the added mass along the axis parallel to elongation shrinks and the glider more readily accelerates in that direction. In essence, by appropriately timing added mass changes, we can make motion in a given direction much easier and use it to glide to a desired spot. With that intuition we now describe exactly how we set up the control problem and walk through solutions.

OPTIMAL CONTROL THEORY

Optimal control theory⁴⁸ is a technique used to control the state $\mathbf{x}$ of a dynamical system while minimizing a specified cost $\mathcal{C}$. In particular our goal is to minimize

$$\mathcal{C} = \int_0^T L(\mathbf{x}(t), U(t), t) dt$$

for some function (a “running cost”) L subject to the constraints that the state $\mathbf{x}(t) = (u, v, w, x, y, \theta, \beta)$ satisfies the dynamics at all times, the system reaches a specified final $(x_f, y_f, \theta_f, \beta_f)$, and that the con-

trol $U(t)$ is within some prescribed bounds. We supplement the dynamics given in equations 1-6 with equation 3.16 describing how β changes in response to a control input U .

$$\dot{\beta} = -(\beta - \beta_0) + U(t) \quad (3.16)$$

We assume the glider has a rest shape with $\beta_0 = 1$ and the first term models relaxation towards that equilibrium. We constrain $\beta \in [0.1, 10]$ and let $|U_{max}| = 5$ so that the aspect ratio can vary an order of magnitude from it's rest shape. In practice the range of β and $|U_{max}|$ will be set by experimental constraints. Note we do not control the shape directly but rather its derivative. This ensures that even if discontinuous controls are needed, the shape will change continuously. In the perching task we let L have the general form $L = \gamma + (1 - \gamma)U(t)^2$. When $\gamma = 1$ the goal is to minimize the time it takes for the glider to perch. When $\gamma = 0$ the goal is to reach the desired point with the least control input, a proxy for energy expenditure. Given the complexity of the dynamical equations we solved the optimal control problem numerically using the CasADI software package⁶. We present the solutions in the next section.

3.3 RESULTS

To evaluate gliding performance at different density ratios we first did a study of how far a glider can reach horizontally when dropped from a given height. We find that the smaller ρ_s/ρ_f is (increasing the importance of added mass effects) the farther the body can glide. In order to find the maximal gliding distance we set up the optimal control task minimizing $L = -x^2$ to penalize the glider for being near the origin. The results are shown in fig. 3.2.

With a sense of maximal glide distance for each density ratio, we then posed the perching problem with a horizontal distance we know the body can reach. We found both time ($L = 1$) and energy ($L = U(t)^2$) optimal solutions. Figures 3.3 and 3.4 show center of mass trajectories and

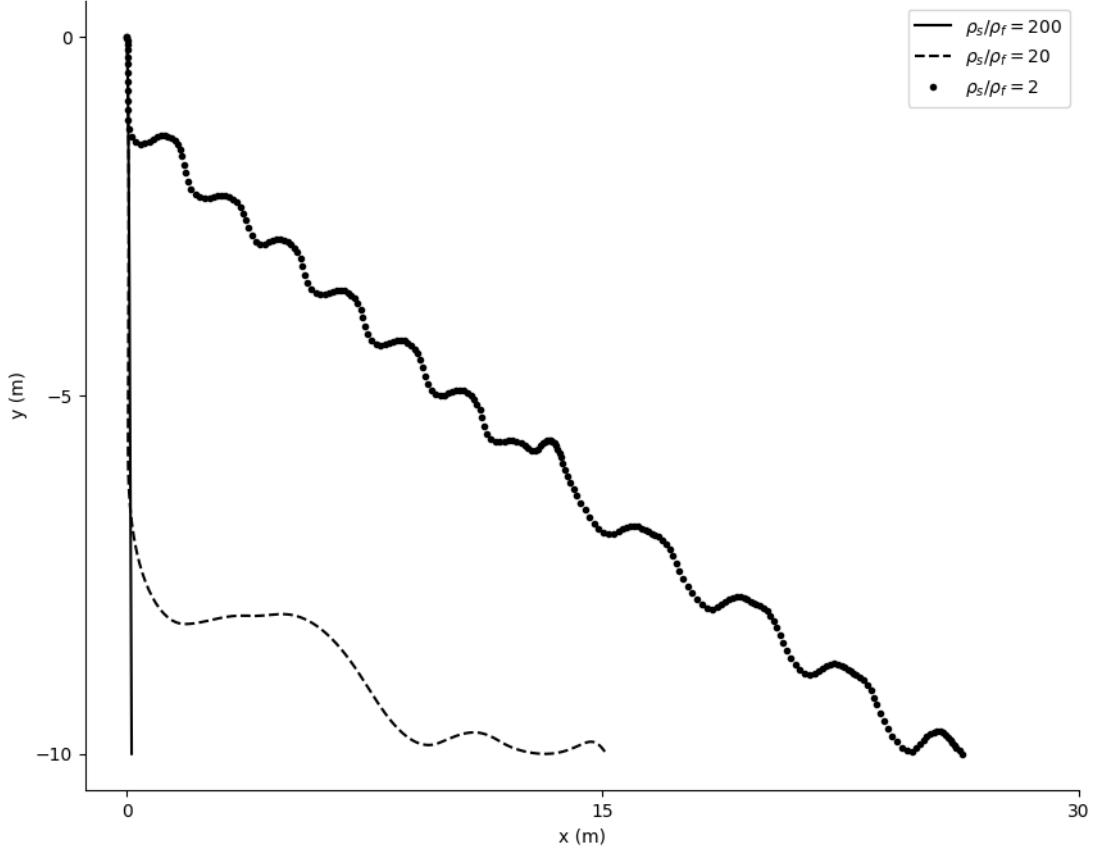


Figure 3.2: Maximum glide distance as a function of density: For each density ratio ρ_s/ρ_f the glider starts with $(u, v, w, x, y, \theta, \beta) = (0.1m/s, 0.1m/s, 0, 0, 0, 0, 1)$ and glides as far as it can until its center of mass hits the ground located at $y_f = -10m$. To encourage greater gliding distance we set the running cost $L = -x^2$ so that the body is penalized over time for its proximity to the y axis. When the solid is not much denser than the liquid, added mass effects are more significant and gliding performance is improved.

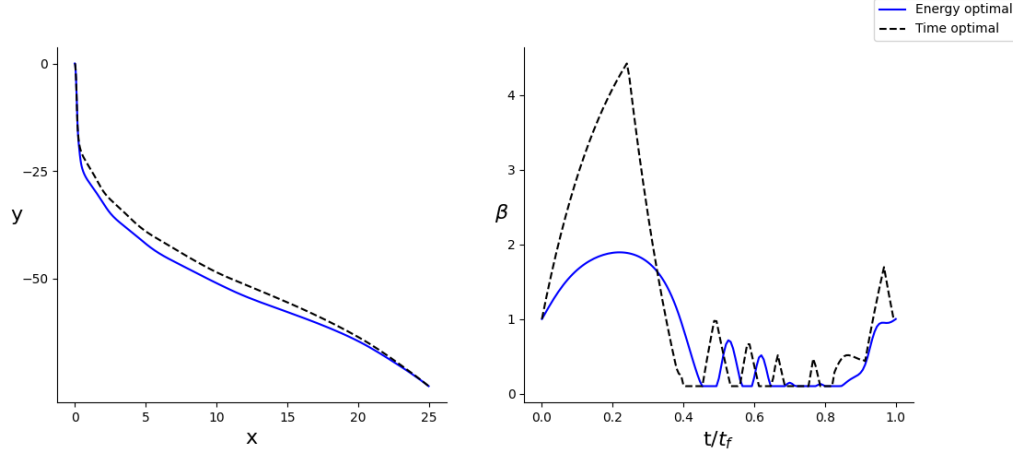


Figure 3.3: Optimal perching with $\rho_s/\rho_f = 200, \alpha = ab = 0.1m^2$. In both tasks we start with $(u, v, w, x, y, \theta, \beta) = (0.1m/s, 0.1m/s, 0, 0, 0, 0, 1)$ and land with $(x_f, y_f, \theta_f, \beta_f) = (25m, -75m, \pi/4, 1)$. Left: Trajectories for the time and energy optimal task. Right: The value of β over time. Note how the time optimal solution has much more change in β (i.e. control input) despite the final trajectories looking very similar. The time optimal solution achieves the perching task about 4% faster than the energy optimal. The time axis is scaled by t_f , the time to land, in each case for a more clear comparison of control strategies.

shape over time. The solutions where the solid is much denser than the fluid follow very similar trajectories with the time optimal one completing the task about 4% faster than the energy optimal. When $\rho_s/\rho_f = 2$ we see similar trajectories but the time optimal solution is 15% faster. In each case the glider starts with a small launch ($u > 0, v > 0$) from a level orientation so u and v are nearly parallel with the x and y axes respectively.

For the high density ratio ($\rho_s/\rho_f = 200$) case each strategy goes through an initial period of free fall with increasing β to pick up speed in the y direction. Eventually β is decreased and the glider rotates down so it's elongated at an angle with respect to the direction of motion. From this unsteady edge on position the fluid torque causes the glider to enter the tumbling phase where the bulk of the horizontal displacement is achieved. While tumbling we see regular changes in β that match with the rotation of the glider itself (fig 3.5). These changes are timed to increase the rotation rate w and lead to a larger displacement in a given time. The time optimal solution has more dramatic oscilla-

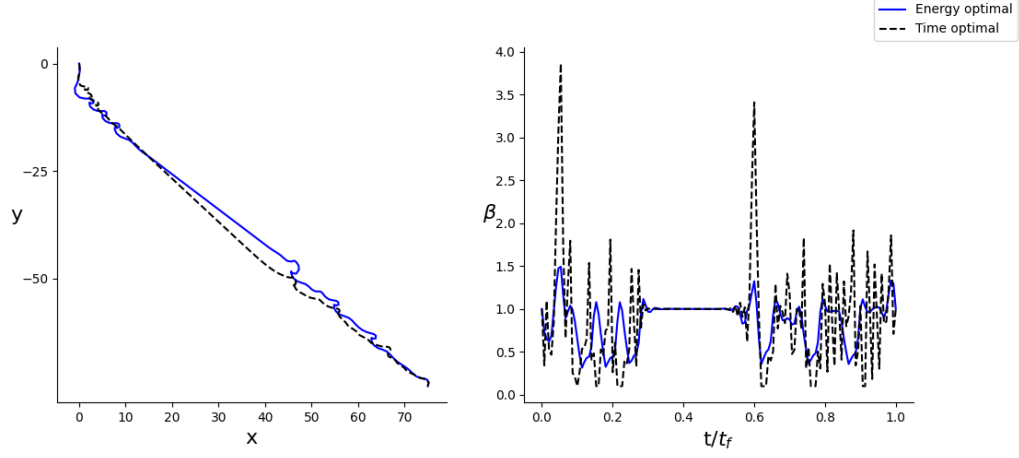


Figure 3.4: Optimal perching with $\rho_s/\rho_f = 2$, $\alpha = ab = 0.1m^2$. In both tasks we start with $(u, v, w, x, y, \theta, \beta) = (0.1m/s, 0.1m/s, 0, 0, 0, 0, 1)$ and land with $(x_f, y_f, \theta_f, \beta_f) = (75m, -75m, \pi/4, 1)$. Since the density ratio ($\rho_s/\rho_f = 2$) is much smaller, and we know the body won't have trouble gliding to greater distances, we use a larger target x in the desired perching location than in the $\rho_s/\rho_f = 200$ case. Left: Trajectories for the time and energy optimal task. Right: The value of β over time. Again the time optimal solution shows much more control input and uses the extra control to perch about 15% faster than in the energy optimal case. The time axis is scaled by t_f , the time to land, in each case for a more clear comparison of control strategies.

tions in β with relatively little gain in performance. Eventually the glider nears the perching location and needs to exit the tumble to land properly. During the end stage changes in β are used to increase fluid torque in the direction of rotation and land at the desired spot with $\theta_f = \pi/4, \beta = 1$.

When the solid is twice as dense as the fluid β changes until the body finds a long glide slope that it can follow with no control input 3.4. Once the body nears the ground it needs to reconfigure to ensure it perches as desired. Again we see that the energy optimal solution has significantly less control input (which we can see in the change in β over time) compared to the time optimal. Figures 3.5 and 3.6 show the shape and angle of the glider over energy optimal trajectories.

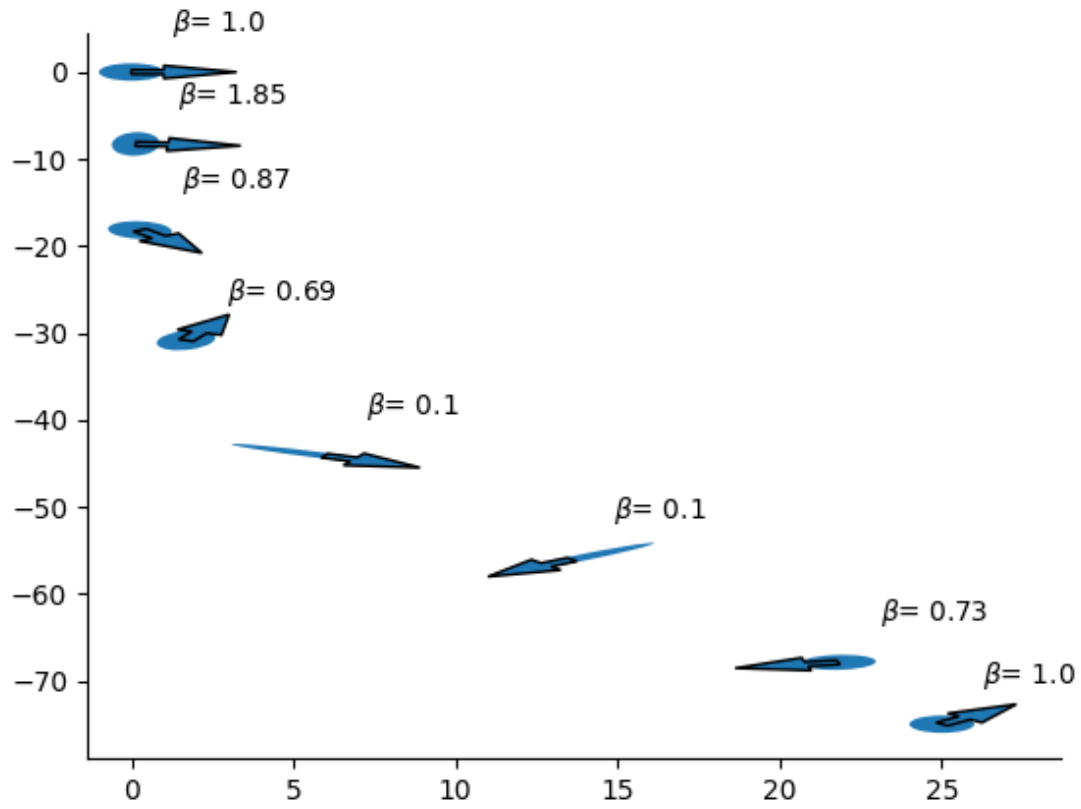


Figure 3.5: Glider shape and orientation over time for the energy optimal perching trajectory $\rho_s/\rho_f = 200$. Since the body is much denser than the fluid β must be changed significantly to alter the course of the glider. We also see a significant variation in the orientation of the glider as it tumbles towards the ground.

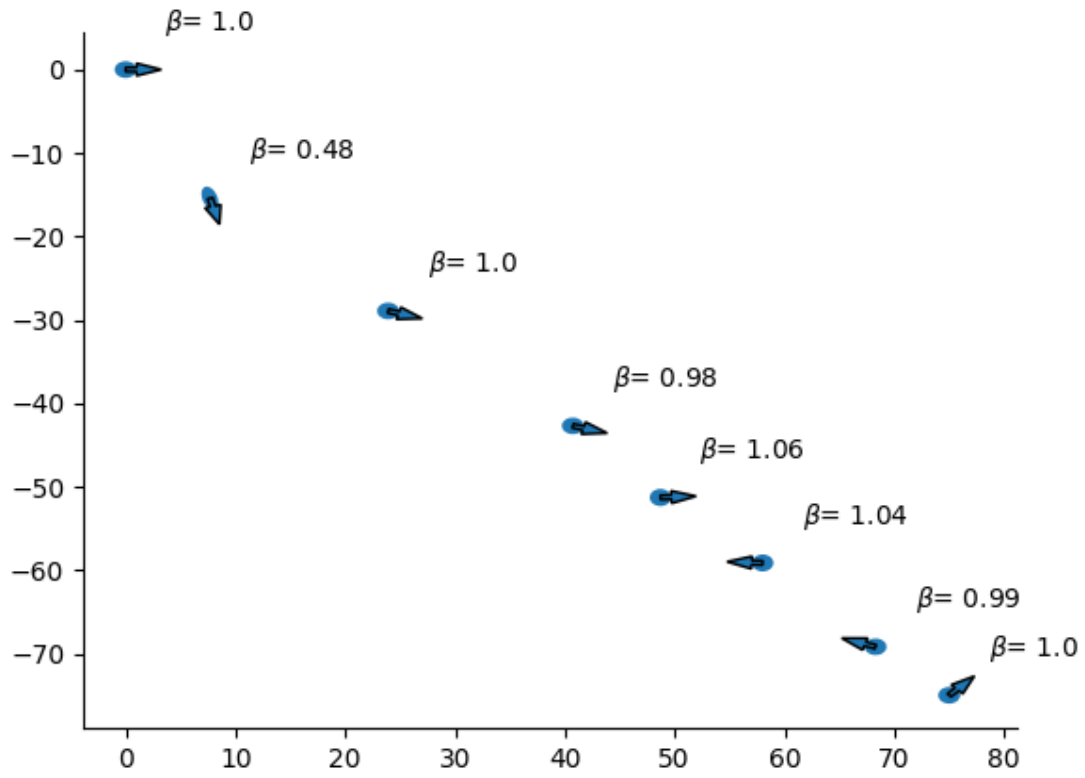


Figure 3.6: Glider shape and angle over time for the energy optimal perching trajectory with $\rho_s/\rho_f = 2$. Note how β is very near 1 for most of the trajectory indicating that significant control input is not required as long as it is appropriately timed (cf. Fig 3.5).

3.4 DISCUSSION

We have considered both energy and time optimal methods of gliding to a specific point. Our solutions show that the chaotic seeming motion of tumbling can be harnessed to achieve optimal perching. We also see that in situations (like underwater) where the effects of added mass are significant, gliding range increases measurably and tumbling is not needed to achieve large displacements.

There are several clear paths for future work. First the model can be extended to include the true three dimensional nature of the gliding body. By considering a more realistic 3D motion we also introduce another length parameter which we can use to control flight characteristics. Another avenue to pursue is the addition of a more robust fluid dynamics model by directly simulating the Navier-Stokes equation coupled to the glider. With a more widely applicable model we can study the effects of shape change in different media. Finally we can add in external winds or turbulent effects and study how robust the shape change strategy is in an unpredictable environments¹⁴. Instead of relying on classical control theory methods we can also take advantage of developments in reinforcement learning^{92,70} to develop feedback based controls that allow for more robust and generalizable strategies.

4

Vertical gap bridging in snakes

This project is joint work with Dr. Bruce Jayne, Dr. Ganga S. Prasath, Dr. Petur Bryde, and Dr. L. Mahadevan. All experimental data was provided by Dr. Bruce Jayne.

Snakes demonstrate a wide range of impressive locomotive behavior⁴⁵. Particularly surprising is their skill in climbing trees and overcoming large vertical gaps^{47,84} (figure 4.1). Life in arboreal habitats and the presence of branches spread far apart likely led to the evolution of such prowess^{22,76}. As human habitats further expand into natural ones, snakes climbing up on to power lines have caused

significant economic damage⁸¹ and motivated studies into the behavior⁴⁶. Relatively little is known about the mechanical^{57,58,63} aspects of snake locomotion out of the plane. Recent work in the development of robotics inspired by snake locomotion has spurred modeling effort³⁸ but the field is still nascent.

In this work we make two contributions to the study of snake locomotion. First we analyze videos of a gap bridging task to attain a quantitative understanding of snake shape over time. We do this by developing image analysis software which extracts the two dimensional outline of the snake frame by frame. Second we introduce a physical model which accounts for the fact that snakes are not purely passive physical systems. In particular we add a term to the equations of elasticity⁹ to capture the muscular activation along the animal. Using methods from optimal control theory we solve for the profile of muscular activity which leads to the snake tip reaching a desired location, a simplified form of the gap bridging task in the videos.

First we introduce the data and provide a description of the data analysis procedure. Next we describe the mathematical model and the necessary simplifications made to manage model complexity. After introducing the model we explain how we formulate the task of bridging a vertical gap as an optimal control problem. Finally we compare simulation results to experimental observations.

4.1 DATA ANALYSIS

We analyze videos of 4 snakes taken by Dr. Bruce Jayne and his students in their lab at the University of Cincinnati. There were three brown tree snakes (*Boiga Irregularis*) and one python (*Morelia Amethistina*). Each trial consisted of a snake trying to cross a vertical gap as pictured in figure 4.1. The experiments were repeated several times for a variety of gap heights ranging from 30-70% of the snake's body length. We include data from 12 different trials in this work.

Our approach to making the visual data quantitative is to compute the curvature along the body

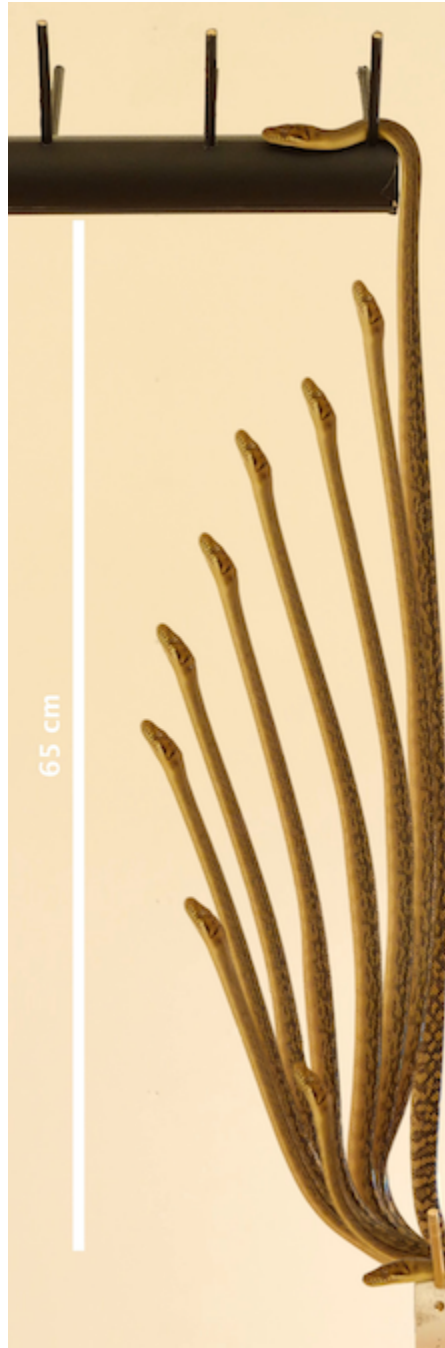


Figure 4.1: A time lapse of a python as it completes a gap bridging task. The gap is about 70% of the snake's body length. Repeated trials of this kind of experiment were performed in the lab with a python and 3 different brown tree snakes. We analyzed these videos to understand how the snake's shape evolved over time as it completed it's task.



Figure 4.2: A sample of how the image analysis process works. We take an image from the video (left), crop it and edit out the top and bottom perches. Starting from the cropped image (center) we use our image analysis software to extract the front of the body (right). In red are points representing the pixel values of points on the front of the snake. The blue line is an interpolation through the pixel values. We use the interpolated line to compute curvature because the raw pixel data is too noisy.

of the snakes frame by frame throughout the trials. To calculate curvature we need to extract the outline of the front of the snake. Figure 4.2 shows our process for going from videos to values of the curvature. First we sample the video to create a sequence of images. Then we crop the images to edit out the supports. For the editing task we use the Fiji software package⁸⁵. Starting from a cropped image (fig 4.2 center) we developed software to get (x, y) coordinates for points along the front of the body. Since we know the distance between the top and bottom perch in each trial we use that to set a conversion between pixel distances and physical distances. We repeat this analysis across trials to develop an understanding of snake shape with different gap heights and species.

For a parameterized two dimensional curve $\gamma(t) = (x(s), y(s))$ the curvature is given by

$$\kappa = \frac{x'y'' - y'x''}{(x'^2 + y'^2)^{3/2}} \quad (4.1)$$

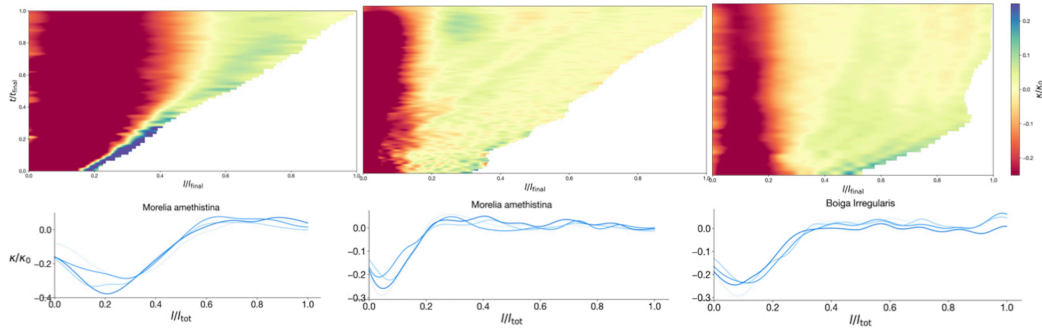


Figure 4.3: Top panels: A snapshot of the curvature along the snakes when it is near the top of the perch. Bottom panel: A time lapse of the curvature at each frame over the course of the video. Negative curvature is dorsiflexion. Left: Python 71 cm gap. Middle: Python 50 cm gap. Right: Tree snake 60 cm gap

Where x' denotes the derivative of x with respect to the parameter s . We interpolate a line through the snake outline described above and use the distance along the snake as the parameter s . The interpolation fits the outline well but is not as jagged so we compute curvature using the interpolated curve. Samples of those curvature calculations are shown in 4.3. The top row of plots, what we call kymographs, show the curvature along the snake throughout a particular trial. The kymographs demonstrate that the curvature of the snake is concentrated at the base where it departs the lower perch. The bottom plots show the curvature along the snake at the moment before it contacts the upper perch and begins to climb up. The curvature line plots show that the curvature profile along the snakes is similar across heights and species.

Now that we have a more thorough understanding of the data, our next goal is to develop a physical model of the snake.

4.2 MATHEMATICAL MODEL

Considered as purely passive objects snakes are very similar to elastic filaments. They have long, thin bodies and present some natural resistance to bending even without activating their muscles.

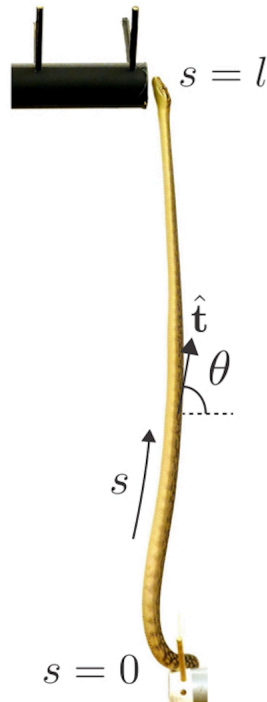


Figure 4.4: We model the snake as an elastic filament. The center line is parameterized by s . The tangent vector $\mathbf{t} = (dx/ds, dy/ds)$ describes the local orientation of the snake.

Therefore we choose to model the snake as such. We supplement the usual equations of elasticity theory with a new term intended to account for muscular activity.

Although the real motion of the snake is 3 dimensional we only consider the vertical and horizontal displacements of the body. The center-line is given by the coordinates $\mathbf{r}(s) = (x(s), y(s))$ where s is the arc length along the body. The local orientation along the snake is $d\mathbf{r}(s)/ds = \mathbf{t}(s) = (\cos\theta(s), \sin\theta(s))$ (see fig 4.4). The snake experiences both forces and torques along its body as it completes the gap bridging task. We study the static problem to estimate the posture of the snake using the equations of torque and force balance⁹ given below

$$\frac{d\mathbf{f}}{ds} + \rho\mathbf{g} = 0 \quad (4.2)$$

$$\frac{dm}{ds}\hat{\mathbf{z}} + \mathbf{t}(s) \times \mathbf{f}(s) = 0 \quad (4.3)$$

where ρ is the mass density of the snake body and $\mathbf{g}$ is gravity pointing downwards. These equations simply state that the snake generates some forces $\mathbf{f}(s) = (H(s), V(s))$ and torques $m(s)$ which cancel out the effects due to gravity. We write the moment m as a sum of a passive and active torque: $m(s) = B\kappa(s) + m^a(s)$ where $\kappa(s) = d\theta(s)/ds$ is the curvature along the body and B is the bodies bending stiffness. The passive torque $B\kappa(s)$ stems from the fact that the snakes body resists curvature to an extent proportional to the bending stiffness B . The active torque $m^a(s)$ is the new addition to the model which simulates muscle activity. For now we have no way of determining the form of $m^a(s)$ but we will address that issue momentarily.

We need to specify boundary conditions to be able to solve the equilibrium equations. We choose

the following

$$\theta(l) = \pi/2 \quad (4.4)$$

$$\frac{d\theta}{ds} \big|_{s=l} = \kappa(l) = 0 \quad (4.5)$$

$$\mathbf{f}(l) = 0 \quad (4.6)$$

$$(4.7)$$

Setting $\theta(l) = \pi/2$ and $\kappa(l) = 0$ simply means the snake is flat and pointed vertically upwards at the top. The condition $\mathbf{f}(l) = 0$ indicates the snake does not need to generate forces at its tip to counteract gravity since there is no body weight to support above that point.

Combining the equilibrium equations we get

$$B \frac{d^2 \theta(s)}{ds^2} + \frac{dm^a(s)}{ds} - \rho g(s-l) \cos \theta(s) = 0. \quad (4.8)$$

where we have substituted in the solution to the force balance equation for $H(s)$ and $V(s)$.

Now that we have the equation to describe the snake shape, we need to decide on a method for determining the active moment $m^a(s)$. We turn to that next.

4.3 OPTIMAL CONTROL

There are many different profiles of muscular activity that can solve the gap bridging task. Of all the successful methodologies for reaching the target we presume the one chosen by the animals is optimal in some way³. We have no way of knowing exactly what the snakes are trying to optimize (if they are optimizing anything at all) and there may be many different optimization criteria that lead to the same results. Nonetheless we continue with this method and look for qualitative agreement with the observed data. Future work with more detailed measurements of muscular activity along-

side a robust model of the snake's musculature will be invaluable in resolving some of the questions raised above.

The complete set of equations describing the system are

$$\frac{dx}{ds} = \cos\theta \quad (4.9)$$

$$\frac{dy}{ds} = \sin\theta \quad (4.10)$$

$$\frac{d^2\theta(s)}{ds^2} + \frac{da}{ds} - \alpha(s-l) \cos\theta(s) = 0. \quad (4.11)$$

Where $a = \frac{m^a}{B}$, $\alpha = \frac{\rho g}{B}$ and $x(0) = y(0) = 0$ set the origin of the x, y coordinate system. The parameter $\alpha = \rho g/B$ gives an approximate length of the snake when gravitational torques overcome the resistance due to the bodies natural bending stiffness. In particular, a passive rod roughly of length greater than $\alpha^{-1/3}$ will buckle under it's own weight²⁸. In terms of these equations the gap bridging task is simply a question of how to choose $u = \frac{da}{ds}$ so that

$$x(l) = x_0 \quad (4.12)$$

$$y(l) = y_0 \quad (4.13)$$

$$\theta(l) = \pi/2 \quad (4.14)$$

$$\frac{d\theta}{ds} \big|_{s=l} = 0 \quad (4.15)$$

where (x_0, y_0) is the location of the target. Note that $u = da/ds$ is the control variable instead of $a(s)$ directly. This choice is made because optimal control theory requires that the control, and not its derivatives, appear in the equations of motion. More importantly choosing $u = da/ds$ ensures that the active moment along the snake is continuous, a reasonable physical assumption. Optimality

constrains the choice of u so that the above equations are satisfied and a cost

$$\mathcal{C} = \int_0^l L(x, y, \theta, u, s) ds \quad (4.16)$$

is minimized. The choice of $L(x, y, \theta, u, s)$ encapsulates the goals of the animal. Our assumption is that the snake completes the task while minimizing the work done by its muscles. In other words it minimizes the work done by the active moment and $L = m^a \kappa$. In the next section we present results with this choice of L .

4.4 RESULTS

We choose units where the length $l = 1$. We compare results over a range of α , which corresponds to different values of the bending stiffness of the snake, a parameter we do not yet have experimental measurements of. Figure 4.5 compares a snapshot of a brown tree snake completing the gap bridging task with the solution found from optimal control theory. We set the target at $(x_0, y_0) = (0.1, 0.9)$ (the solution is reflected around the y axis for ease of comparison).

The results from optimal control theory allow us to develop a qualitative understanding of the posture (fig 4.5) but quantitative comparison requires more experimental measurement and model development. In particular precise measurements of α are needed for more quantitative comparisons. Figure 4.6 shows the curvature results from optimal control theory. More model complexity will likely be required if we aim to capture the inflection in curvature observed in experimental data. Figure 4.7 shows the profile of muscular activity from optimal control theory. The intuition that muscular actuation is largest when curvature is largest is borne out in fig. 4.7 and we anticipate a more detailed comparison to experimental electromyography data⁴⁷ will aid us in further model development.

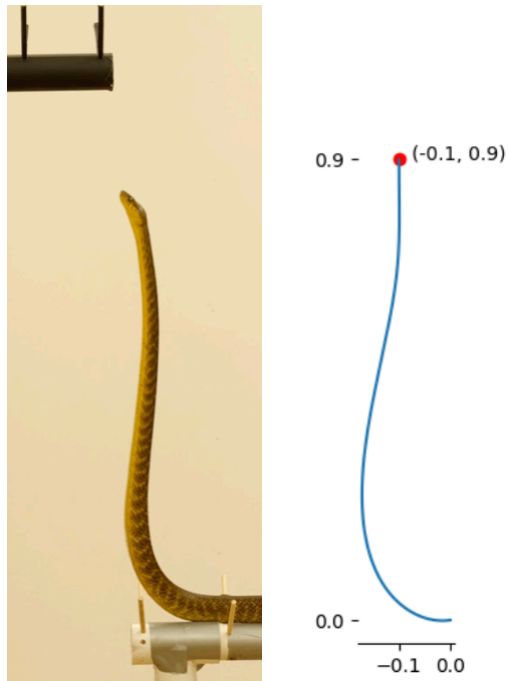


Figure 4.5: Left: A snapshot of a brown tree snake in the middle of a climbing task. Right: Outline found from solving the optimization problem. There is a qualitative agreement between the shapes.

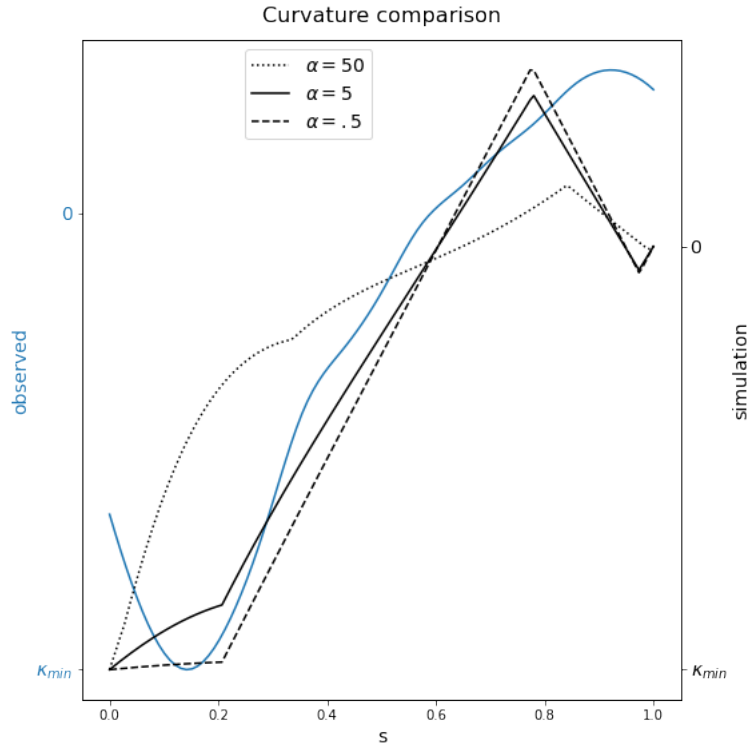


Figure 4.6: Left: The angle θ the snake makes with the horizontal along its body. Right: The curvature along the snake scaled by the maximum curvature. Note we are using $\kappa = -d\theta/ds$ to be consistent with curvature conventions (dorsiflexion should correspond to negative curvature) used in the data analysis portion of the work.

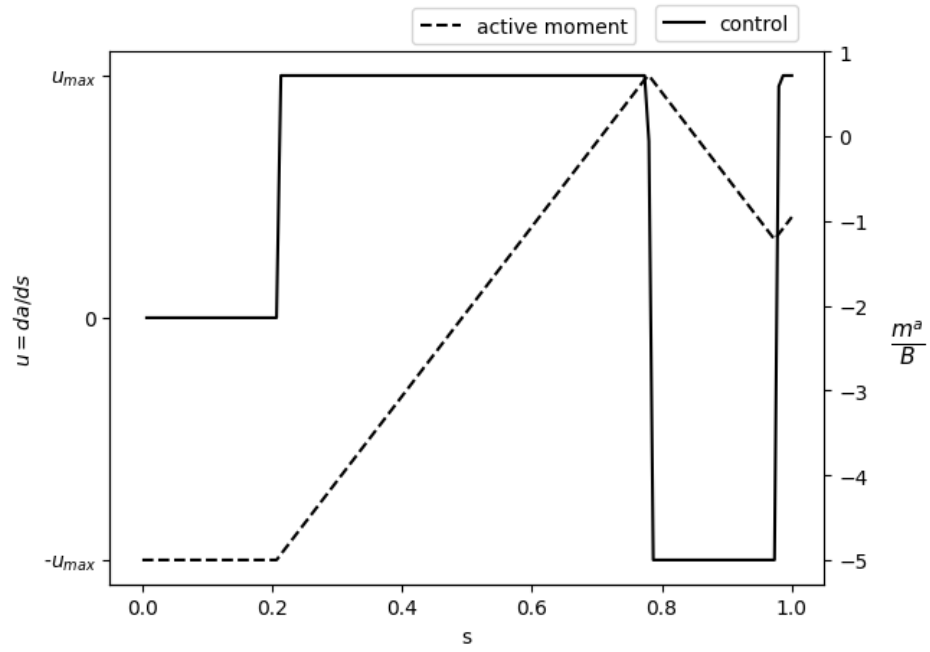


Figure 4.7: The control profile for reaching to $(x_f, y_f) = (0.1, 0.9)$ with $\alpha = 5$. The active moment (a proxy for the muscular activity) is greatest in magnitude near the base of the snake. This is consistent with the intuition that muscular activity is maximal where the curvature along the snake is the greatest cf. fig. 4.6

4.5 CONCLUSION

In this work we analyzed video data of snakes completing a vertical gap bridging task. The data consisted of trials for 4 different snakes covering vertical gaps from 20–70% of their body length. From the videos we computed a curvature profile along the body of the animals and show how it changes over time. As far as we know such an analysis has never been done and constitutes an addition to the literature on snake locomotion.

After analyzing the videos we developed a mathematical model to understand the observed curvature profiles. We chose to model the snakes as active elastic filaments. With such a model we can not resolve many complexities associated with the real muscular layout^{72,44} of snakes, but gain a tractable way of beginning to address the issues at hand. The active bending moment encapsulates the muscular activity of the snakes. To constrain the form the active moment can take we assume the snake performs the gap bridging task in some optimal manner. This assumption leads us to formulate an optimal control problem where the control determines the value of the active moment. The control is chosen so that it minimizes the amount of work done by the active moment (our proxy for the work done by the snake) while reaching to a specified target. We then compare the solutions found from control theory with the observed solutions and find qualitative agreement.

There are a few clear directions for future work. On the experimental side having detailed muscular activity data like that collected in⁴⁷ would give more guidance in developing a model of muscular activity. As the snakes climb the gap they periodically lose balance and sway back and forth to ready themselves for continued ascent. In this work we only studied static stability (no accelerations), but the original aim was to fully understand the dynamic stability properties of the snake as it climbs. From a data analysis perspective a relatively straightforward next step is to quantify the observed swaying behavior. Understanding more about the frequency and magnitude of swaying will be essential for future studies of the dynamic stability of the animals. There are many more directions to

pursue and we hope our work can serve as a starting point for further exploration.

5

Conclusion

In this thesis we brought together ideas from physics and optimization to study problems in motor control. Using extreme behaviors as a natural starting point we examined how augmenting physical models with active parameters allows us to understand movement in the natural world. We studied swinging sports, gliding flight, and snake posture, demonstrating how our approach allows us to tackle a wide variety of behaviors. Using the dual approaches of optimal control theory (OCT) and reinforcement learning (RL) gives us extra freedom in exploring solutions. Optimal control

theory can provide provably optimal solutions but incorporating feedback controls is much more straightforward in the reinforcement learning framework. While convergence issues are more difficult to overcome in RL, we can use it to explore feedback based learning using different sensory modalities. For example, the work on Kiiking can be directly extended by considering different state observations for the RL agent. While Kiiking, athletes do not have precise measurement of their instantaneous position, angular momentum, and length as we have assumed. Instead they use some more complicated mixture of visual information and cues from their vestibular system²³ to make decisions. Presenting RL agents with observations more closely related to those made by real athletes is a promising way to see what information is required for task learning.

There are an endless number of fascinating motor control tasks and we hope our work demonstrates a fruitful method of approaching these problems. With simple physical models and modern optimization techniques we can learn a surprising amount about the world.



Supplementary information for Kiiking

A.1 DATA ANALYSIS

We found several videos of kiiking online and through the Estonian Kiiking Association. To develop a quantitative understanding of swinging trajectories we needed to measure swing angle over time.

We also measured the angle between the athlete's center of mass and the vertical post. To measure the angles we used the image analysis tool Fiji⁸⁵. Figure A.1 shows how we analyze a typical frame in



Figure A.1: Measuring angles with image analysis software Fiji⁸⁵. We approximate the mass of the swing structure and the athlete to create a weighted average of the system center of mass. The angular location of the true center of mass is somewhere in between the post the athlete is holding and the athletes torso. The difference in measured angles provides an upper bound on the measurement error. (a) Measuring the angle the swing makes with the vertical. (b) Measuring the angle between the approximate center of mass of the athlete and the vertical.

a kiiking video. After taking these two measurements we compute a weighted average to estimate the angular position of the true center of mass. We use

$$\theta = r\theta_{swing} + (1 - r)\theta_{person} \quad (\text{A.1})$$

where r is chosen so that the sequence of swing energies at the oscillation peaks is monotonically increasing. The maximum error is bounded by the angle between the athlete's torso and the swing strut they are holding on to.

References

- [1] Acheson, D. J. (1991). *Elementary fluid dynamics*.
- [2] Achiam, J. (2018). Spinning up in deep reinforcement learning.
- [3] Alexander, R. M. (1996). *Optima for animals*. Princeton university press.
- [4] Andersen, A., Pesavento, U., & Wang, Z. J. (2005a). Unsteady aerodynamics of fluttering and tumbling plates. *Journal of Fluid Mechanics*, 541, 65–90.
- [5] Andersen, A., Pesavento, U., & Wang, Z. J. (2005b). Analysis of transitions between fluttering, tumbling and steady descent of falling cards. *Journal of Fluid Mechanics*, 541, 91–104.
- [6] Andersson, J. A. E., Gillis, J., Horn, G., Rawlings, J. B., & Diehl, M. (2019). CasADi – A software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*, 11(1), 1–36.
- [7] Andrychowicz, O. M., Baker, B., Chociej, M., Jozefowicz, R., McGrew, B., Pachocki, J., Petron, A., Plappert, M., Powell, G., Ray, A., et al. (2020). Learning dexterous in-hand manipulation. *The International Journal of Robotics Research*, 39(1), 3–20.
- [8] Association, E. K. (2014). Kiiking. https://commons.wikimedia.org/wiki/File:Kiiking_07.JPG. Accessed: 01-10-23.
- [9] Audoly, B. & Pomeau, Y. (2010). *Elasticity and geometry: from hair curls to the non-linear response of shells*. Oxford university press.
- [10] Avriel, M. (2003). *Nonlinear programming: analysis and methods*. Courier Corporation.
- [11] Azuma, A. & Okuno, Y. (1987). Flight of a samara, alsomitra macrocarpa. *Journal of Theoretical Biology*, 129(3), 263–274.
- [12] Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., Lecomte, V., Orlandi, A., Parisi, G., Procaccini, A., et al. (2008). Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *Proceedings of the national academy of sciences*, 105(4), 1232–1237.

- [13] Barbarino, S., Bilgen, O., Ajaj, R. M., Friswell, M. I., & Inman, D. J. (2011). A review of morphing aircraft. *Journal of intelligent material systems and structures*, 22(9), 823–877.
- [14] Bellemare, M. G., Candido, S., Castro, P. S., Gong, J., Machado, M. C., Moitra, S., Ponda, S. S., & Wang, Z. (2020). Autonomous navigation of stratospheric balloons using reinforcement learning. *Nature*, 588(7836), 77–82.
- [15] Bellman, R. E. (2010). *Dynamic programming*. Princeton university press.
- [16] Bellman, R. E. & Dreyfus, S. E. (2015). *Applied dynamic programming*, volume 2050. Princeton university press.
- [17] Bernstein, N. A. (2014). *Dexterity and its development*. Psychology Press.
- [18] Berry, M. (2018). Pumping a swing revisited: minimal model for parametric resonance via matrix products. *European Journal of Physics*, 39(5), 055007.
- [19] Betts, J. T. (2010). *Practical methods for optimal control and estimation using nonlinear programming*. SIAM.
- [20] Bryson, A. E. (1996). Optimal control-1950 to 1985. *IEEE Control Systems Magazine*, 16(3), 26–33.
- [21] Burns, J. A. (1970). More on pumping a swing. *American Journal of Physics*, 38(7), 920–922.
- [22] Byrnes, G. & Jayne, B. C. (2012). The effects of three-dimensional gap orientation on bridging performance and behavior of brown tree snakes (*Boiga irregularis*). *Journal of Experimental Biology*, 215(15), 2611–2620.
- [23] Casale, J., Browne, T., Murray, I., & Gupta, G. (2018). Physiology, vestibular system.
- [24] Case, W. B. (1996). The pumping of a swing from the standing position. *American Journal of Physics*, 64(3), 215–220.
- [25] Case, W. B. & Swanson, M. A. (1990). The pumping of a swing from the seated position. *American Journal of Physics*, 58(5), 463–467.
- [26] Chanzy, Q. & Keane, A. (2018). Analysis and experimental validation of morphing uav wings. *The Aeronautical Journal*, 122(1249), 390–408.
- [27] Cone, C. D. (1962). Thermal soaring of birds. *American Scientist*, 50(1), 180–209.
- [28] Cox, S. J. & McCarthy, C. (1998). The shape of the tallest column. *SIAM Journal on Mathematical Analysis*, 29(3), 547–554.
- [29] Davenport, I., Bryde, P., & Mahadevan, L. (2023). Supplementary material for "optimal control and learning for kiiking, an estonian swinging sport".

- [30] Diehl, M. & Gros, S. (2011). Numerical optimal control. *Optimization in Engineering Center (OPTEC)*.
- [31] Dudley, R. (2002). *The biomechanics of insect flight: form, function, evolution*. Princeton University Press.
- [32] Dudley, R., Byrnes, G., Yanoviak, S. P., Borrell, B., Brown, R. M., & McGuire, J. A. (2007). Gliding and the functional origins of flight: biomechanical novelty or necessity? *Annu. Rev. Ecol. Evol. Syst.*, 38, 179–201.
- [33] Dynamics, B. (2023). Atlas partners in parkour. <https://www.youtube.com/watch?v=tF4DML7FIWk>. Accessed: 04-24-23.
- [34] Feinerman, O., Pinkoviezky, I., Gelblum, A., Fonio, E., & Gov, N. S. (2018). The physics of cooperative transport in groups of ants. *Nature Physics*, 14(7), 683–693.
- [35] Feng, S., Whitman, E., Xinjilefu, X., & Atkeson, C. G. (2014). Optimization based full body control for the atlas robot. In *2014 IEEE-RAS International Conference on Humanoid Robots* (pp. 120–127).: IEEE.
- [36] Field, S. B., Klaus, M., Moore, M., & Nori, F. (1997). Chaotic dynamics of falling disks. *Nature*, 388(6639), 252–254.
- [37] Franklin, D. W. & Wolpert, D. M. (2011). Computational mechanisms of sensorimotor control. *Neuron*, 72(3), 425–442.
- [38] Fu, Q. & Li, C. (2020). Robotic modelling of snake traversing large, smooth obstacles reveals stability benefits of body compliance. *Royal Society open science*, 7(2), 191192.
- [39] Ginelli, F. (2016). The physics of the vicsek model. *The European Physical Journal Special Topics*, 225, 2099–2117.
- [40] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
- [41] Harvey, C., Baliga, V., Wong, J., Altshuler, D., & Inman, D. (2022a). Birds can transition between stable and unstable states via wing morphing. *Nature*, 603(7902), 648–653.
- [42] Harvey, C., Gamble, L. L., Bolander, C. R., Hunsaker, D. F., Joo, J. J., & Inman, D. J. (2022b). A review of avian-inspired morphing for uav flight control. *Progress in Aerospace Sciences*, 132, 100825.
- [43] Harvey, C. & Inman, D. J. (2022). Gull dynamic pitch stability is controlled by wing morphing. *Proceedings of the National Academy of Sciences*, 119(37), e2204847119.
- [44] Jayne, B. C. (1988). Muscular mechanisms of snake locomotion: an electromyographic study of the sidewinding and concertina modes of crotalus cerastes, nerodia fasciata and elaphe obsoleta. *Journal of Experimental Biology*, 140(1), 1–33.

- [45] Jayne, B. C. (2020). What defines different modes of snake locomotion? *Integrative and comparative biology*, 60(1), 156–170.
- [46] Jayne, B. C., Lehmkuhl, A. M., & Riley, M. A. (2014). Hit or miss: branch structure affects perch choice, behaviour, distance and accuracy of brown tree snakes bridging gaps. *Animal Behaviour*, 88, 233–241.
- [47] Jorgensen, R. M. & Jayne, B. C. (2017). Three-dimensional trajectories affect the epaxial muscle activity of arboreal snakes crossing gaps. *Journal of Experimental Biology*, 220(19), 3545–3555.
- [48] Kirk, D. E. (2004). *Optimal control theory: an introduction*. Courier Corporation.
- [49] Kleis, D. & Sachs, E. W. (2000). Optimal control of the sterilization of prepackaged food. *SIAM Journal on Optimization*, 10(4), 1180–1195.
- [50] Kota, S., Osborn, R., Ervin, G., Maric, D., Flick, P., & Paul, D. (2009). Mission adaptive compliant wing–design, fabrication and flight test. In *RTO Applied Vehicle Technology Panel (AVT) Symposium* (pp. 1–18).: RTO-MP-AVT-168, Evora, Portugal.
- [51] Kulkarni, J. E. (2003). Time-optimal control of a swing. In *42nd IEEE International Conference on Decision and Control (IEEE Cat. No. 03CH37475)*, volume 2 (pp. 1729–1733).: IEEE.
- [52] Lamb, H. (1924). *Hydrodynamics*. University Press.
- [53] Latash, M. L. (2012). *Fundamentals of motor control*. Academic Press.
- [54] Latash, M. L. & Lestienne, F. (2006). *Motor control and learning*, volume 78. Springer.
- [55] Lavrovskii, E. & Formal'skii, A. (1993). Optimal control of the pumping and damping of a swing. *Journal of Applied Mathematics and Mechanics*, 57(2), 311–320.
- [56] Liberzon, D. (2011). *Calculus of variations and optimal control theory: a concise introduction*. Princeton university press.
- [57] Lillywhite, H., LaFrentz, J., Lin, Y., & Tu, M.-C. (2000). The cantilever abilities of snakes. *Journal of Herpetology*, (pp. 523–528).
- [58] Lin, Y.-C., Hwang, J.-C., & Tu, M.-C. (2003). Does the saccular lung affect the cantilever ability of snakes? *Herpetologica*, 59(1), 52–57.
- [59] Luo, J. & Lee, E. (1998). Time-optimal control of the swing using impulse control actions. In *Proceedings of the 1998 American Control Conference. ACC (IEEE Cat. No. 98CH36207)*, volume 1 (pp. 200–204).: IEEE.

- [60] Luo, J., Paduraru, C., Voicu, O., Chervonyi, Y., Munns, S., Li, J., Qian, C., Dutta, P., Davis, J. Q., Wu, N., Yang, X., Chang, C.-M., Li, T., Rose, R., Fan, M., Nakhosht, H., Liu, T., Kirkman, B., Altamura, F., Cline, L., Tonker, P., Gouker, J., Uden, D., Bryan, W. B., Law, J., Fatiha, D., Satra, N., Rothenberg, J., Waraich, M., Carlin, M., Tallapaka, S., Witherspoon, S., Parish, D., Dolan, P., Zhao, C., & Mankowitz, D. J. (2022). Controlling commercial cooling systems using reinforcement learning.
- [61] Ma, Y., Dixit, V., Innes, M. J., Guo, X., & Rackauckas, C. (2021). A comparison of automatic differentiation and continuous sensitivity analysis for derivatives of differential equation solutions. In *2021 IEEE High Performance Extreme Computing Conference (HPEC)* (pp. 1–9).
- [62] Mahadevan, L., Ryu, W. S., & Samuel, A. D. (1999). Tumbling cards. *Physics of Fluids*, 11(1), 1–3.
- [63] Mansfield, R. H. & Jayne, B. C. (2011). Arboreal habitat structure affects route choice by rat snakes. *Journal of Comparative Physiology A*, 197, 119–129.
- [64] Maurer, H. & Vossen, G. (2009). Sufficient conditions and sensitivity analysis for optimal bang-bang control problems with state constraints. In *System Modeling and Optimization: 23rd IFIP TC 7 Conference, Cracow, Poland, July 23–27, 2007, Revised Selected Papers 23* (pp. 82–99).: Springer.
- [65] McCAY, M. G. (2001). Aerodynamic stability and maneuverability of the gliding frog polypedates dennysi. *Journal of Experimental Biology*, 204(16), 2817–2826.
- [66] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.
- [67] Ng, A. Y., Harada, D., & Russell, S. (1999). Policy invariance under reward transformations: Theory and application to reward shaping. In *ICML*, volume 99 (pp. 278–287).: Citeseer.
- [68] Nishikawa, K., Biewener, A. A., Aerts, P., Ahn, A. N., Chiel, H. J., Daley, M. A., Daniel, T. L., Full, R. J., Hale, M. E., Hedrick, T. L., et al. (2007). Neuromechanics: an integrative approach for understanding motor control. *Integrative and comparative biology*, 47(1), 16–54.
- [69] Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3), 139–154.
- [70] Novati, G., Mahadevan, L., & Koumoutsakos, P. (2019). Controlled gliding and perching through deep-reinforcement-learning. *Phys. Rev. Fluids*, 4, 093902.

- [71] Paoletti, P. & L, M. (2011). Planar controlled gliding, tumbling and descent. *Journal of Fluid Mechanics*, 689, 489–516.
- [72] Penning, D. A. (2018). Quantitative axial myology in two constricting snakes: *Lampropeltis holbrooki* and *pantherophis obsoletus*. *Journal of anatomy*, 232(6), 1016–1024.
- [73] Pesavento, U. & Wang, Z. J. (2004). Falling paper: Navier-stokes solutions, model of fluid forces, and center of mass elevation. *Physical review letters*, 93(14), 144501.
- [74] Piccoli, B. (1995). Time-optimal control problems for the swing and the ski. *International Journal of Control*, 62(6), 1409–1429.
- [75] Piccoli, B. & Kulkarni, J. (2005). Pumping a swing by standing and squatting: do children pump time optimally? *IEEE Control Systems Magazine*, 25(4), 48–56.
- [76] Pizzatto, L., Almeida-Santos, S. M., & Shine, R. (2007). Life-history adaptations to arboreality in snakes. *Ecology*, 88(2), 359–366.
- [77] Pontryagin, L. S. (1987). *Mathematical theory of optimal processes*. CRC press.
- [78] Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., & Dormann, N. (2021). Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, 22(268), 1–8.
- [79] Ramaswamy, S. (2010). The mechanics and statistics of active matter. *Annu. Rev. Condens. Matter Phys.*, 1(1), 323–345.
- [80] Rao, A. V. (2009). A survey of numerical methods for optimal control. *Advances in the Astronautical Sciences*, 135(1), 497–528.
- [81] Rodda, G. H., Sawai, Y., & Tanaka, H. (2020). *Problem snake management: the habu and the brown treesnake*. Cornell University Press.
- [82] Rosenbaum, D. A. (2009). *Human motor control*. Academic press.
- [83] Sargeant, A., Hoinville, E., & Young, A. (1981). Maximum leg force and power output during short-term dynamic exercise. *Journal of applied physiology: respiratory, environmental, and exercise physiology*, 51(5), 1175–1182.
- [84] Savidge, J. A., Seibert, T. F., Kastner, M., & Jayne, B. C. (2021). Lasso locomotion expands the climbing repertoire of snakes. *Current Biology*, 31(1), R7–R8.
- [85] Schindelin, J., Arganda-Carreras, I., Frise, E., & et al (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9, 676–682.
- [86] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

- [87] Sethi, S. P. (2021). *The Maximum Principle: Pure State and Mixed Inequality Constraints*, (pp. 115–144). Springer International Publishing.
- [88] Simonyan, K. & Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *The Neuroscientist*, 17(2), 197–208.
- [89] Sofla, A., Meguid, S., Tan, K., & Yeo, W. (2010). Shape morphing of aircraft wing: Status and challenges. *Materials & Design*, 31(3), 1284–1292.
- [90] Strogatz, S. H. (2018). *Nonlinear dynamics and chaos with student solutions manual: With applications to physics, biology, chemistry, and engineering*. CRC press.
- [91] Sun, E. D., Michaels, T. C., & Mahadevan, L. (2020). Optimal control of aging in complex networks. *Proceedings of the National Academy of Sciences*, 117(34), 20404–20410.
- [92] Sutton, R. S. & Barto, A. G. (2018). *Reinforcement learning: An introduction 2nd Ed.* MIT Press Ltd.
- [93] Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature neuroscience*, 7(9), 907–915.
- [94] Todorov, E. et al. (2006). Optimal control theory. *Bayesian brain: probabilistic approaches to neural coding*, (pp. 268–298).
- [95] Tröltzsch, F. (2010). *Optimal control of partial differential equations: theory, methods, and applications*, volume 112. American Mathematical Soc.
- [96] Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., & Shochet, O. (1995). Novel type of phase transition in a system of self-driven particles. *Physical review letters*, 75(6), 1226.
- [97] Wickenheiser, A. M. & Garcia, E. (2008). Optimization of perching maneuvers through vehicle t. *Journal of Guidance, Control, and Dynamics*, 31(4), 815–823.
- [98] Wirkus, S., Rand, R., & Ruina, A. (1998). How to pump a swing. *The College Mathematics Journal*, 29(4), 266–275.
- [99] Wolpert, D. M., Diedrichsen, J., & Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature reviews neuroscience*, 12(12), 739–751.
- [100] Wu, O. Q. & Chen, H. (2010). Optimal control and equilibrium behavior of production-inventory systems. *Management Science*, 56(8), 1362–1379.
- [101] Yanoviak, S. P., Dudley, R., & Kaspari, M. (2005). Directed aerial descent in canopy ants. *Nature*, 433(7026), 624–626.
- [102] Yanoviak, S. P., Kaspari, M., & Dudley, R. (2009). Gliding hexapods and the origins of insect aerial behaviour. *Biology letters*, 5(4), 510–512.

- [103] Yanoviak, S. P., Munk, Y., Kaspari, M., & Dudley, R. (2010). Aerial manoeuvrability in wingless gliding ants (*cephalotes atratus*). *Proceedings of the Royal Society B: Biological Sciences*, 277(1691), 2199–2204.
- [104] Yeaton, I. J., Ross, S. D., Baumgardner, G. A., & Socha, J. J. (2020). Undulation enables gliding in flying snakes. *Nature physics*, 16(9), 974–982.