Our non-feathered friends: sensing and control in insect-scale flight

Michelle Hickner

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Steven L. Brunton, Chair

Bingni W. Brunton, Chair

Thomas L. Daniel

Program Authorized to Offer Degree:
Mechanical Engineering

University of Washington

Abstract

Our non-feathered friends: sensing and control in insect-scale flight

Michelle Hickner

Co-Chairs of the Supervisory Committee:
Steven L. Brunton
Department of Mechanical Engineering

Bingni W. Brunton
Department of Biology

Insects are capable of many tasks that insect-scale flying robots cannot yet perform, such as dodging waving grasses or branches, catching other insects as aerial predators, and taking off from and landing on a wide variety of surfaces. There has been significant progress in getting insect-scale flying robots off the ground in recent years, but to perform at the same level as insects, many advances will need to be made. These advances will come from both applying and adapting existing engineering techniques, and from interdisciplinary efforts towards gaining a better understanding of biological insects. Here, we present three ways to approach insect-scale flight. First, we develop a system identification method that is well suited to the unsteady nature, flexible wings, and viscous Reynolds number regime of insect-scale flight. The method builds a linear state-space model from a quasi-impulse response of an aeroelastic system. Second, we seek to better understand possible tradeoffs associated with aerodynamic force production and gyroscopic sensing in insects. In particular, we look at how wing and haltere shapes can affect the relative forces and strain associated with flapping and body rotation. Third, we provide a resource for bringing together animal movement and linear control theory. Feedback control is a fundamental concept in animal movement, and to translate advances in biological understandings to engineered vehicles, we must be able to describe and compare the dynamics of biomechanics and nervous systems into control architectures.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure Number Page

ACKNOWLEDGMENTS

First of all, I'd like to thank my wonderful advisors, Steve Brunton and Bing Brunton, for sharing their knowledge, supporting my professional goals, introducing me to colleagues and collaborators, and including me in both their labs.

My committee members, labmates, writing group, and collaborators have been generous with their time to offer encouragement and answer many questions about computing, insects, math, and more, with special thanks to Tom Daniel, Katie Stanchak, Ali Weber, and Urban Fasel. My biomechanics students have reminded me to remain curious about the natural world and all the amazing things animals can do.

I could not have completed this PhD without the support of my family, especially during the pandemic. My husband, Brian Polagye, provided many pep talks as well as technical advice and Matlab tips.

Chapter 1

INTRODUCTION

Small-scale flapping flight is an open problem in both the natural and engineering sciences. Flying insects demonstrate impressive agility and versatility with the limited computation available in their relatively small nervous systems. Insects do this in small and lightweight bodies that contain not just the muscles, joints, and nervous system necessary for locomotion, but also breathing and digestive systems for powering movement, and systems for non-locomotion tasks such as reproduction. Engineers have made progress in the last few decades in building insect-scale flapping robots, but many challenges remain in understanding insect flight and design and control of engineered flapping flyers.

Obtaining a better understanding of insect biology can help guide flapping robot design choices. As with all living things, insect bodies and functions have been shaped by evolution, and it is worth considering which aspects have been exquisitely tune through evolution for the purposes of flight, and which aspects of insect form and function are byproducts or have trade-offs associated with other bodily functions. For example, the veins in wings serve multiple purposes: they stiffen and strength the thin wing structure, they circulate hemolymph through the wings, and they serve as paths for the axons of sensory neurons embedded in the wings [15, 23, 24, 28].

1.1 Gyroscopic sensing

Insect-scale flapping robots have been developed that have one set of wings, two sets of wings, and more recently, four sets of wings. Most flying insects have two sets of wings, but insects the order Diptera, which includes flies, crane flies, and mosquitoes, have one set of wings. The tradeoffs associated with number of wings are not well understood for either biological or robotic insects. Having more wings increases complexity, but is there increased maneuverability or stability in return? One advantage to more wings is that they allow more wing area for aerodynamic force generation, without the structural limitations of having long, thin structures that are undergoing large accelerations from flapping and need to be able to withstand some collisions. However, it is not clear that Dipteran insects are limited by the aerodynamic force generated by one set of wings instead of two. Many Dipteran species are highly maneuverable, which indicates that they have no deficiencies regarding aerodynamic force production.

Insects wings serve more functions that just aerodynamic force production: they are surfaces used for warning or cryptic coloring, they may be used for heat transfer, they are used in mating and threat displays, and most importantly for flight, they have sensory neurons embedded in the cuticle of the wings [1, 10, 21]. While Dipteran insects no longer have hindwings, they have retained hindwing-derived structures called halteres. Both wings and halteres contribute to gyroscopic sensing and sensing of body accelerations, in concert with antennae and visual systems [20, 21]. Body rotations interact with the flapping kinematics of the wings and halteres to generate Coriolis forces and Euler (angular acceleration) forces. These forces cause the flexible wings to deform, and this deformation is sensed by strain-sensitive neurons called Campaniform sensilla.

For the purposes of maneuverable flight control, sensing and force generation are both key functions. Assuming that increased maneuverability increases evolutionary fitness, how has this dual function of wings affected their morphology? Which aspects of wing morphology enhance both the aerodynamic and sensing functions, and which aspects result in trade-offs? These questions are particularly interesting in the context of halteres. Because Dipteran insects are descended from four-winged insects, this implies that there is a fitness advantage to having halteres rather than hindwings. Beyond that, Dipteran insects are not an example of a morphological curiosity in a single species or a narrow environment: the order Diptera is hyper-diverse, and Dipteran insects are present in widespread geographically and ecologically. The success of both two-winged and four-winged insects shows that both are successful strategies.

One reason that it may be advantageous to have halteres is that they might be better gyroscopic sensors than hindwings, or the dissimilar structures of forewings and halteres may detect different information, serving as a more robust sensory system in combination with each other. In order to make progress towards answering the question of whether halteres provide better gyroscopic sensing than wings alone, we need metrics to describe what “better” means in this context. One approach is to do comparative behavioral studies with flying insects to study performance of tasks requiring gyroscopic sensing [2, 7, 9, 11, 18, 21, 25]. That approach takes advantage of the full complexity of insect wing structure, kinematics, and control, but also has many confounding variables associated with differences between species. A different approach is to study the underlying physics, which allows comparisons of gyroscopic sensing across many wing and haltere morphologies with relatively little information, such as wing shape and flapping kinematics [14, 17, 19].

In Chapter 3, we present several ways of evaluating wing and haltere shapes and mass distributions for the purpose of gyroscopic sensing. We approach the topic of wing and haltere shape from the perspective of the evolutionary divergence between wings and halteres, however our metrics and conclusions are also useful for the design of strain-based gyroscopes such as bio-inspired halteres [8, 13]. We simulate the strain due to flapping and body rotation for forty-nine shapes and mass distributions ranging from wing-like to haltere-like.

1.2 Aeroelastic modeling for insect-scale flight

Automated insect-scale robotic flight requires control algorithms and system models that are appropriate for thin, flexible micro-fabricated wings. As the degrees of freedom or dimensions of a system increase, it becomes more difficult to derive a model of the system dynamics from first principles like Newtonian mechanics or the Navier-Stokes equations. An alternative way to describe the system dynamics is to discover them from system response data. Aeroelastic systems, including flying insects and insect-scale flapping robots, are high-dimensional systems, but for feedback control, we can approximate the dynamics well with a low-dimensional model.

The development of aeroelastic models, from both first principles and from data, is a large field. Insect flight occurs at Reynolds numbers 10-100,000 and so requires models that include the effect of viscosity [16]. This is true not just because of the small size scales and low flight velocities of insects, but because of the added mass effects that result from the accelerations inherent in back and forth motion. For these reasons, aeroelastic models designed for fast rigid-wing aircraft are not well suited to the problem of modeling insect-scale flight. A better approach is to build models based on frameworks developed for oscillating airfoils, including those originally designed for aeroelastic flutter, rather than flapping flight.

We design a system identification algorithm based on three previous models. The first is the Eigensystem Realization Algorithm (ERA), which uses impulse response data from a system to build a reduced order linear state-space model [12]. It does this by forming several time-delay (Hankel) matrices of the response data, and then forms the state-space model from the singular values and singular vectors of the time-delay matrices. The next model that we incorporate is Theodorsen's model, which is equivalent to Wagner's model, which describes the coefficient of lift of a harmonically pitching airfoil [26, 27]. An aspect of Theodorsen's model that makes it well suited for flapping flight, as well as for rotorcraft, is that it builds upon thin-airfoil theory to model the effects of an idealized wake due to an oscillating foil as well as the effects of added mass due to acceleration and velocity. The third model that we build upon is one developed a decade ago by Brunton et al., which incorporates both of the previously mentioned models to build an interpretable data-driven system identification technique for rigid wings [3–5].

1.3 Control theory for biomechanics

To develop bio-inspired control algorithms and architectures, we need to have a framework for translating biological control to computer control. This includes having a common language that is shared between biologists and engineers, so that scientists from both fields can describe and compare various control strategies. Control theory is also a useful tool for purely biological studies. In particular, biologists are often trying to *understand* movement, rather than trying to *command* movement. The language and mathematical frameworks of control theory are used in many biological studies. For example, using control block diagrams to design experiments that "open the loop" to study sensory and motor neural signals and computations [6, 22].

We identified a need for a control theory primer aimed at biologists who study animal movement, to help with understanding of studies that use control theory, and to enable more biologists to use control theory. Many control theory books oriented at an engineering audience assume a mathematical strong background knowledge of differential equations and linear algebra that is less common among biologists, making these texts inaccessible to some natural scientists. There also exist many excellent articles about the utility of control theory for biology, however they often do not build up the concepts from scratch. In the primer in Chapter 4, we aim to provide sufficient mathematical background and worked examples that someone with a background of undergraduate calculus can translate the movement control of their organism of interest into a control theoretic language, and can implement feedback control loops for biological studies.

1.4 Contributions

In Chapter 2, we develop and demonstrate a data-driven system identification method for aeroelastic flight. The system ID method builds on previous algorithms, including the Eigensystem Realization Algorithm (ERA) [12], Theodorsen's model [26, 27], and other data-driven models of pitching and plunging wings [3–5]. Our method combines and extends these previous methods by modeling wing deformation in addition to coefficient of lift. Modeling wing deformation enables the controller to limit deformation, for example to stay within structural limits or to increase the service life of a wing by decreasing fatigue. The model is low-rank and linear, which decreases computation time, and has interpretable coefficients, which aid in controller design. Unlike many aeroelastic modeling techniques, this system identification technique can model the viscous effects on quasi-steady lift, added mass, and wing bending that affect insect-scale aerodynamics.

In Chapter 3, we develop several metrics for describing how wing shape, flapping kinematics, and body rotation affect the inertial forces and strain on the wing. We demonstrate these inertial effects by simulating flapping and body rotation for a range of wing shapes and mass distributions using a finite element model that we developed in COMSOL 6.1. Our metrics show that the way morphology affects the relative effects of flapping and rotating depends on body rotation direction. We also provide evidence that both wing-like shapes and long slender haltere-like shapes have strong body rotation signals.

For Chapter 4, we address a need for describing classical control theory through the lens of biology and biomechanics. This work aims to provide biologists with undergraduate-level knowledge of calculus the tools to use classical time-domain control theory for investigations of animal movement. We provide worked examples, including descriptions of how to set up the equations necessary to implement these techniques. In particular, we build up and work through several control scenarios and strategies for the spring-loaded inverted pendulum

(SLIP) model. We do this not just to enable the use of control theory in biology, but also to bridge the gap between biologists and engineers working together to study locomotion and biomechanics.

Bibliography

[1] Sweta Agrawal, David Grimaldi, and Jessica L Fox. Haltere morphology and campaniform sensilla arrangement across Diptera. *Arthropod Structure and Development*, 46(2):215–229, 2017.

[2] John A. Bender and Michael H. Dickinson. A comparison of visual and haltere-mediated feedback in the control of body saccades in Drosophila melanogaster. *Journal of Experimental Biology*, 209(23):4597–4606, 2006.

[3] Steven L Brunton, Scott T.M. Dawson, and Clarence W Rowley. State-space model identification and feedback control of unsteady aerodynamic forces. *Journal of Fluids and Structures*, 50:253–270, 2014.

[4] Steven L Brunton and Clarence W Rowley. Empirical state-space representations for Theodorsen's lift model. *Journal of Fluids and Structures*, 38:174–186, 2013.

[5] Steven L. Brunton, Clarence W. Rowley, and David R. Williams. Reduced-order unsteady aerodynamic models at low Reynolds numbers. *Journal of Fluid Mechanics*, 724:203–233, 2013.

[6] Noah J Cowan, Mert M Ankarali, Jonathan P Dyhr, Manu S Madhav, Eatai Roth, Shahin Sefati, Simon Sponberg, Sarah A Stamper, Eric S Fortune, and Thomas L Daniel. Feedback control as a framework for understanding tradeoffs in biology. *American Zoologist*, 54(2):223–237, 2014.

[7] Tanvi Deora, Siddharth S. Sane, and Sanjay P. Sane. Wings and halteres act as coupled dual oscillators in flies. *eLife*, 10:1–20, 2021.

[8] H. Droogendijk, R. A. Brookhuis, M. J. De Boer, R. G.P. Sanders, and G. J.M. Krijnen.

Towards a biomimetic gyroscope inspired by the fly's haltere using microelectromechanical systems technology. *Journal of the Royal Society Interface*, 11(99), oct 2014.

[9] J L Fox and T L Daniel. A neural basis for gyroscopic force measurement in the halteres of Holorusia. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 194(10):887–897, 2008.

[10] G. Fraenkel and J.W.S Pringle. Biological Sciences: Halteres of Flies as Gyroscopic Organs of Equilibrium. *Nature*, 141:919–920, 1938.

[11] Joshua M Hall, Dane P. McLoughlin, Nicholas D Kathman, Alexandra M Yarger, Shwetha Mureli, and Jessica L Fox. Kinematic diversity suggests expanded roles for fly halteres. *Biology Letters*, 11(11), 2015.

[12] Jer-Nan Juang and Richard S. Pappa. An eigensystem realization algorithm for modal parameter identification and model reduction. *Journal of Guidance, Control, and Dynamics*, 8(5):620–627, sep 1985.

[13] Chulhong Kim, Junghyun Park, Taeyup Kim, Jee-Seong Kim, Jeongmo Seong, Hyungbo Shim, Hyoungho Ko, and Dong-il Cho. Development and evaluation of haltere-mimicking gyroscope for three-axis angular velocity sensing using a haltere-mimicking structure pair. *Bioinspiration & Biomimetics*, 18(1), 2023.

[14] T L Mohren, T L Daniel, A L Eberle, P G Reinhall, and J L Fox. Coriolis and centrifugal forces drive haltere deformations and influence spike timing. *Journal of the Royal Society Interface*, 16(153), 2019.

[15] Andrew M. Mountcastle and Stacey A. Combes. Biomechanical strategies for mitigating collision damage in insect wings: Structural design versus embedded elastic materials. *Journal of Experimental Biology*, 217(7):1108–1115, 2014.

[16] Werner Nachtigall. Some aspects of Reynolds number effects in animals. *Mathematical Methods in the Applied Sciences*, 24(17-18):1401–1408, 2001.

[17] G Nalbach. The halteres of the blowfly Calliphora I. Kinematics and dynamics. *J Comp Physiol A*, 173:293–300, 1993.

[18] G. Nalbach and R. Hengstenberg. The halteres of the blowfly Calliphora II. Three-dimensional organization of compensatory reactions to real and simulated rotations. *Journal fo Comparative Physiology A*, 175(6):695–708, 1994.

[19] Rizuwana Parween, Rudra Pratap, Tanvi Deora, and Sanjay P. Sane. Modeling strain sensing by the gyroscopic Halteres, in the Dipteran soldier fly, Hermetia illucens. *Mechanics Based Design of Structures and Machines*, 42(3):371–385, 2014.

[20] Brandon Pratt, Tanvi Deora, Thomas Mohren, and Thomas Daniel. Neural evidence supports a dual sensory-motor role for insect wings. *Proceedings of the Royal Society B: Biological Sciences*, 284, 2017.

[21] J.W.S Pringle. The gyroscopic mechanism of the halteres of Diptera. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 233(602):347–384, 1948.

[22] E Roth, S Sponberg, and N J Cowan. A comparative approach to closed-loop computation, 2014.

[23] Mary K Salcedo, Jordan Hoffmann, Seth Donoughe, and L Mahadevan. Computational analysis of size, shape and structure of insect wings. *Biology Open*, 8(10), 2019.

[24] Mary K Salcedo and John J Socha. Circulation in Insect Wings. *Integrative and Comparative Biology*, 60(5):1208–1220, nov 2020.

[25] Alana Sherman and Michael H. Dickinson. A comparison of visual and haltere-mediated equilibrium reflexes in the fruit fly Drosophila melanogaster. *Journal of Experimental Biology*, 206(2):295–302, 2003.

[26] Theodore Theodorsen. General theory of aerodynamic instability and the mechanism of flutter. *Technical Report, NACA*, pages 291–311, 1935.

[27] Herbert Wagner. Über die Entstehung des dynamischen Auftriebes von Tragflügeln. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 5(1):17–35, jan 1925.

[28] Jieliang Zhao, Mengdi Xu, Youjian Liang, Shaoze Yan, and Wendong Niu. Influence of hydrodynamic pressure and vein strength on the super-elasticity of honeybee wings. *Journal of Insect Physiology*, 109:100–106, 2018.

Chapter 2

DATA-DRIVEN UNSTEADY AEROELASTIC MODELING FOR CONTROL

Aeroelastic structures, from insect wings to wind turbine blades, experience transient unsteady aerodynamic loads that are coupled to their motion. Effective real-time control of flexible structures relies on accurate and efficient predictions of both the unsteady aeroelastic forces and airfoil deformation. For rigid wings, classical unsteady aerodynamic models have recently been reformulated in state-space for control and extended to include viscous effects. Here we further extend this modeling framework to include the deformation of a flexible wing in addition to the quasi-steady, added mass, and unsteady viscous forces. We develop low-order linear models based on data from direct numerical simulations of flow past a flexible wing at low Reynolds number. We demonstrate the effectiveness of these models to track aggressive maneuvers with model predictive control while constraining maximum wing deformation. This system identification approach provides an interpretable, accurate, and low-dimensional representation of an aeroelastic system that can aid in system and controller design for applications where transients play an important role.

Nomenclature

α	angle of attack of plate
κ	curvature near leading edge of plate
ν	kinematic viscosity
τ	convective time (dimensionless), $\tau \equiv tU_\infty/c$
$\Delta\tau$	time step
$\Delta\tau_c$	coarse time step, duration of impulse
a	pitch axis location relative to the 1/2 chord
c	chord length of plate
Δc	length of discretized plate element
r	model rank
t	time
$\mathbf{x}$	state vector of state-space model
$\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$	state-space matrices identified by ERA
C_α	angle of attack (lift slope) coefficient
$C_{\dot{\alpha}}$, $C_{\ddot{\alpha}}$	added mass coefficients
C_L	coefficient of lift
EI	plate flexural rigidity
$\mathbf{H}$	Hankel matrix, time-delay matrix of measurements
K_B	plate bending stiffness
M_ρ	mass ratio
$\mathbf{Q}$	MPC weights for outputs
$\mathbf{R}$	MPC weights for inputs
Re	Reynolds number, $Re \equiv U_\infty c/\nu$
$\mathbf{R}_{\Delta u}$	MPC weights for input rate of change
T_c	MPC control horizon
T_p	MPC prediction horizon
U_∞	free stream velocity
$\mathbf{Y}$	matrix of measurements
$\tilde{\mathbf{Y}}$	Markov parameters

2.1 Introduction

Wing flexibility plays a major role in aerodynamic systems, both natural and engineered, by increasing efficiency and mitigating structural damage from sudden loading [30, 42, 68, 77]. The effect of wing flexibility has been long studied in the context of bio-locomotion [6, 16, 19, 20, 22, 23, 31–33, 43, 53, 54, 57, 71, 74, 78–80, 94, 99, 101], where it is known that flexibility has favorable aerodynamic and hydrodynamic properties [21, 67, 85, 95].

There is an increasing need to develop unsteady aeroelastic models to realize these benefits for engineering systems, for example to control autonomous flight vehicles, including at sizes comparable to insects and birds. The benefits of wing flexibility are often the consequences of subtle aeroelastic effects, and for many applications it is important to capture unsteady viscous fluid transients [76, 93]. This work develops accurate and efficient models to capture these effects in design and control efforts.

Aeroelastic modeling is challenging because of the strongly coupled fluid-structure interactions, which are difficult to capture in reduced order models. Existing high-fidelity models can simulate the coupled fluid flow fields and the deformation profile of a structure [38, 60, 65, 75], but they are computationally intensive and do not fit readily into model-based control design. Medium fidelity models, such as doublet-lattice methods [1], strip theory [50, 87], and panel methods [98] are widely-used approaches that can be combined with optimization to aid in aeroelastic design [29], but they neglect viscous effects and are not well suited for real-time control. The unsteady vortex-lattice method provides some balance between accuracy and tractability for control; however, it has limitations for wings with significant camber or viscous effects [45, 69]. Other nonlinear modeling approaches include Volterra kernels [4, 7, 62] and modal models [3, 34, 56, 63]. Linear models are often sufficient for even large amplitude motions [10] until leading edge separation and stall [28]. Considerable attention has been paid to development of aeroelastic models at or near transonic speeds [37, 41, 58, 61, 73, 91, 102], while techniques for small flexible wings at low Reynolds number have been gaining more interest recently [11, 52, 59, 86, 92]. Use of these low Reynolds number modeling techniques is necessary for characterizations of micro air vehicles (MAVs) or investigation of biological flight [81].

There are many existing techniques for developing reduced order models of aeroelastic systems that take advantage of data-driven system identification methods [51], such as those based on reduced order bases from eigen-decompositions [25, 26], proper orthogonal decomposition (POD) [2, 5, 58, 62], dynamic mode decomposition (DMD), or the eigensystem realization algorithm (ERA) [47, 59]. The use of ERA with aeroelastic models has been demonstrated both in the context of aircraft and with canonical benchmark problems such as the vortex-induced vibration of a sphere or bluff body [82, 83, 103, 104]. These fully unsteady nonlinear flow configurations provide examples of how linear modeling techniques may be applied more broadly to biologically relevant separated flows over wings. Data-driven techniques avoid the simplifying assumptions built into some analytical models, such as inviscid or quasi-steady flow, allowing models to be used in a wider range of flow conditions, including low Reynolds number flows and those resulting from rapid, large-amplitude maneuvers.

Reduced order state-space models have been developed for the control of lift, including those with significant viscous and unsteady effects, extending the classic Theodorsen [90] and Wagner models [97] based on viscous flow data [8–10, 44, 70]. These linear unsteady aerodynamic models use techniques such as ERA and quasi-linear parameter varying techniques (qLPV) to develop minimal realizations that are both accurate and computationally efficient enough for real-time control. A similar approach has been used for the fluid-structure interaction problem of vortex-induced vibration. This approach generates an analytical model for the structural system and separately constructs a fluid model using ERA. The structural and fluid models can then be combined into a single reduced-order model [103, 104]. An advantage of this technique is increased interpretability of the fluid modes. However, it requires knowledge of structural properties which are difficult to obtain as complexity increases, especially for biological specimens where wings are composed of multiscale, anisotropic, and heterogeneous materials [17, 100].

In this work, we extend these Theodorsen-inspired rigid low-order unsteady aerodynamic models to include the effect of wing flexibility and demonstrate these models for the control of aggressive maneuvers. The resulting linear state-space models are purely data-driven, so that they may be tailored to specific wing geometries and flow conditions. In particular,

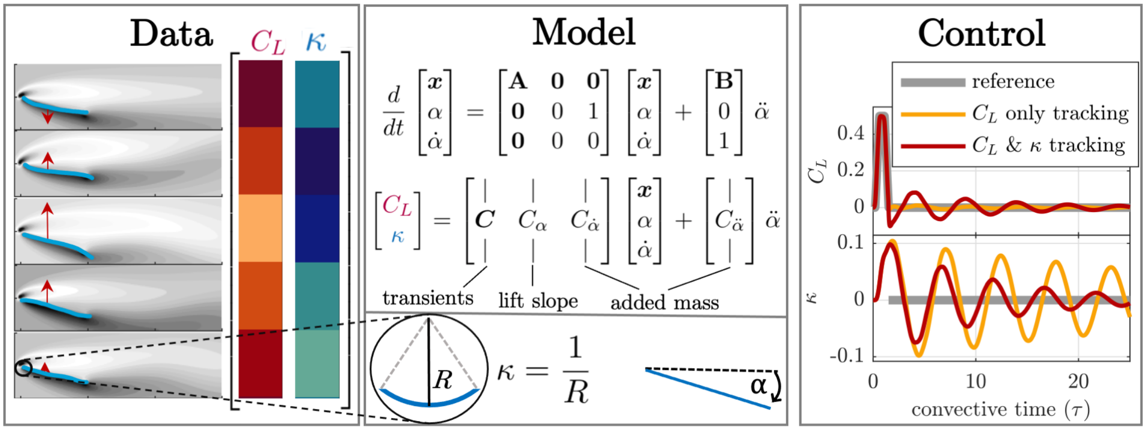


Figure 2.1: A linear state-space model of an aeroelastic wing is developed from lift and deformation observations. The model separates the contributions from transients, quasi-steady pitch effects, and added mass. The model is used to track reference C_L and curvature, κ.

time series data of lift and wing deformation are used to build a model without requiring information about the material properties or mass distribution of the wing. Extending the model to predict wing deformation is key for control in aeroelastic applications, to reduce material fatigue due to vibrations or structural failures due to large stresses. The ERA algorithm is used to augment the limited measurement data with time delayed information, enabling a reduced order model for fast computation in real-time control. We demonstrate this approach on data from a high-fidelity direct numerical simulation of a flexible wing in a low Reynolds number flow. Specifically, this model is used for model predictive control (MPC) of an aggressive lift trajectory, while simultaneously minimizing deflection. Animals achieve impressive maneuverability and stability with limited computation [89], and one goal of low-order models is to enable similar performance in engineered vehicles. This modeling framework fills a need for control-oriented models for flying or swimming animals and for MAVs, both of which rely on viscous and unsteady forces.

2.2 Background

Here we provide a brief overview of some of the most relevant background material on data-driven unsteady aerodynamic modeling and control. Section 2.2.1 introduces Theodorsen's model and describes a data-driven extension for a rigid plate [8], which serves as the foundation for the aeroelastic modeling procedure developed in this chapter. This modeling procedure uses the eigensystem realization algorithm, which is described in section 2.2.2. Finally, model predictive control is introduced in section 2.2.3 because it is well suited to aeroelastic control due to its ability to constrain deformation.

2.2.1 Empirical Theodorsen model: a data-driven approach

Two of the earliest and most important unsteady aerodynamic models for the lift on a flat plate in response to motion are those developed by Wagner [97] in 1925 and Theodorsen [90] in 1935. These models, which would later be shown to be equivalent [35, 46], describe the unsteady lift on a two-dimensional pitching and/or plunging flat plate in an incompressible fluid. Both models analytically represent the unsteady lift as a function of angle of attack, added mass effects, and the vorticity of an idealized planar wake.

Theodorsen's model for thin airfoils describes the forces and moments on an airfoil for harmonic motions. Theodorsen's model assumes attached flow over the wing with an idealized planar wake in an irrotational, incompressible flow, with no structural damping in the wing. For pure pitching motion with dimensional frequency f, the contributions to the lift coefficient C_L from both circulation and added mass can be determined as a function of the reduced frequency, $k = \pi f c/U_\infty$, the pitch axis, a, and the angle of attack α and its derivatives $\dot{\alpha}$ and $\ddot{\alpha}$:

$$C_L = \frac{\pi}{2}\left(\dot{\alpha} - \frac{a}{2}\ddot{\alpha}\right) + 2\pi\left(\alpha + \frac{\dot{\alpha}}{2}\left(\frac{1}{2} - a\right)\right)C(k). \tag{2.1}$$

Lengths are nondimensionalized by the chord length, c, velocities are nondimensionalized by the free stream velocity, U_∞, and a represents the pitch axis as a number from -1 (leading edge) to 1 (trailing edge). Theodorsen's transfer function $C(k)$ is defined as

$$C(k) = \frac{H_1^{(2)}(k)}{H_1^{(2)}(k) + iH_0^{(2)}(k)}; \tag{2.2a}$$

$$H_v^{(2)} = J_v - iY_v, \tag{2.2b}$$

where J_v and Y_v are Bessel functions of the first and second kind, respectively. See Leishman [55] for a more complete description of Theodorsen's model.

For control applications, state-space models in terms of ordinary differential equations are preferable [24, 84]. There are numerous examples in the literature of state-space representations of Theodorsen's and Wagner's models [9, 46, 96]. The coefficients in Theodorsen's model, in particular the 2π coefficient for the quasi-steady lift, become inaccurate in regimes with significant viscous effects [10]. More accurate coefficients can be obtained from experimental data, as is done in the state-space modeling method developed for rigid airfoils [8, 9]:

$$\frac{d}{dt}\begin{bmatrix}\mathbf{x}\\ \alpha\\ \dot{\alpha}\end{bmatrix} = \begin{bmatrix}\mathbf{A} & \mathbf{0} & \mathbf{B}\\ \mathbf{0} & 0 & 1\\ \mathbf{0} & 0 & 0\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \alpha\\ \dot{\alpha}\end{bmatrix} + \begin{bmatrix}\mathbf{0}\\ 0\\ 1\end{bmatrix}\ddot{\alpha}, \tag{2.3a}$$

$$C_L = \begin{bmatrix}\mathbf{C} & C_\alpha & C_{\dot{\alpha}}\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \alpha\\ \dot{\alpha}\end{bmatrix} + C_{\ddot{\alpha}}\ddot{\alpha}. \tag{2.3b}$$

The model coefficients are empirically determined, with C_α replacing the 2π quasi-steady coefficient and the coefficients $C_{\dot{\alpha}}$ and $C_{\ddot{\alpha}}$ representing the contributions from added mass. The dynamics of the model states, $\mathbf{x}$, are found using the eigensystem realization algorithm (ERA). These dynamics take the place of Theodorsen's transfer function, describing the unsteady transient behavior of the fluid wake. The effect of the pitch axis, a, is captured in the pitch velocity coefficient, $C_{\dot{\alpha}}$.

2.2.2 Eigensystem realization algorithm

ERA generates a linear state-space model from an impulse response, without requiring prior knowledge of the model. If enough data is taken so that all transients decay, ERA produces a balanced model for which the observability and controllability Gramians are equal [64]. To develop an impulse response for a pitching airfoil in direct numerical simulations, a fast smoothed linear step maneuver is implemented in α, which may be viewed approximately as a discrete-time impulse in $\dot{\alpha}$. The output measurements of interest are denoted by $\mathbf{Y}$. These measurements of the observables of interest, such as lift and drag, are sampled at the same time scale as the length of the impulse, and formed into two Hankel matrices, $\mathbf{H}$ and $\mathbf{H}'$:

$$\mathbf{H} = \begin{bmatrix} \mathbf{Y}(t_1) & \mathbf{Y}(t_2) & \cdots & \mathbf{Y}(t_{N/2-1}) \\ \mathbf{Y}(t_2) & \mathbf{Y}(t_3) & \cdots & \mathbf{Y}(t_{N/2}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Y}(t_{N/2-1}) & \mathbf{Y}(t_{N/2}) & \cdots & \mathbf{Y}(t_{N-1}) \end{bmatrix}; \tag{2.4a}$$

$$\mathbf{H}' = \begin{bmatrix} \mathbf{Y}(t_2) & \mathbf{Y}(t_3) & \cdots & \mathbf{Y}(t_{N/2}) \\ \mathbf{Y}(t_3) & \mathbf{Y}(t_4) & \cdots & \mathbf{Y}(t_{N/2+1}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Y}(t_{N/2}) & \mathbf{Y}(t_{N/2+1}) & \cdots & \mathbf{Y}(t_N) \end{bmatrix}, \tag{2.4b}$$

shown here for a square Hankel matrix. Each $\mathbf{Y}(t_i)$ may include measurements of multiple observables, such as lift and drag, at time t_i, and they are sometimes referred to as Markov parameters. Next, $\mathbf{H}$ is decomposed using a truncated singular value decomposition (SVD),

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^* = \begin{bmatrix} \tilde{\mathbf{U}} & \mathbf{U}_T \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{\Sigma}} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_T \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{V}}^* \\ \mathbf{V}_T^* \end{bmatrix}, \tag{2.5}$$

where $\mathbf{\Sigma}_T$ are the truncated singular values, and $\tilde{\mathbf{\Sigma}}$ are the largest singular values. The rank of the model can be chosen based on the singular value spectrum $\mathbf{\Sigma}$, or based on the model error on a validation data maneuver. The rank-reduced SVD decomposition is used, along with $\mathbf{H}'$, to determine the reduced order, discrete-time state-space model of the system,

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k, \tag{2.6a}$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k, \tag{2.6b}$$

where the system matrices are given by

$$\mathbf{A} = \tilde{\mathbf{\Sigma}}^{-1/2}\tilde{\mathbf{U}}^{*}\mathbf{H}'\tilde{\mathbf{V}}\tilde{\mathbf{\Sigma}}^{-1/2}; \tag{2.7a}$$

$$\mathbf{B} = \tilde{\mathbf{\Sigma}}^{1/2}\tilde{\mathbf{V}}^{*}\begin{bmatrix}\mathbf{I}_q \\ \mathbf{0}\end{bmatrix}; \tag{2.7b}$$

$$\mathbf{C} = \begin{bmatrix}\mathbf{I}_p & \mathbf{0}\end{bmatrix}\tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}^{1/2}, \tag{2.7c}$$

and $\mathbf{D}$ is the first Markov parameter in $\mathbf{H}$. Here, q is the number of control inputs and $\mathbf{I}_q$ is the $q \times q$ identity matrix; p is the number of system outputs and $\mathbf{I}_p$ is the $p \times p$ identity matrix.

2.2.3 Model predictive control

In this work, we will use our unsteady aeroelastic models for model predictive control (MPC) to track aggressive lift maneuvers while minimizing wing bending. Model predictive control is a widely-used control optimization technique, due to the ability to handle uncertain and nonlinear dynamics and to include constraints. Each control action is determined based on minimizing a highly customizable cost function over a receding predictive horizon. Both input and output constraints can be incorporated into the control optimization. In the context of flight, this means that actuation can be constrained to match motor specifications, angle of attack can be constrained to avoid stall, and wing deformation can be constrained to avoid structural damage.

A drawback of MPC is an increase in on-board computation, which can be mitigated by the use of low-order and linear models. As with all model-based control, accurate models enhance performance. Some key parameters for achieving accurate control while limiting the computational cost are the prediction and control horizons, which determine how many time steps in the future the system output and control inputs are calculated, respectively. The control horizon, $T_c = m_c\Delta t$, is generally less than or equal to the prediction horizon, $T_p = m_p\Delta t$. The cost function used to find the optimal control sequence, $\mathbf{u}(\mathbf{x}_j) = \{\mathbf{u}_{j+1}, ..., \mathbf{u}_{j+m_c}\}$, for the current state estimate or measurement, $\mathbf{x}_j$, over the prediction horizon is

$$J = \sum_{k=0}^{m_p-1} \|\mathbf{x}_{j+k} - \mathbf{r}_k\|^2_{\mathbf{Q}} + \sum_{k=1}^{m_c-1} \left(\|\mathbf{u}_{j+k}\|^2_{\mathbf{R}_u} + \|\Delta\mathbf{u}_{j+k}\|^2_{\mathbf{R}_{\Delta u}}\right), \tag{2.8}$$

subject to constraints. The weighting matrices $\mathbf{Q}$, $\mathbf{R}_u$, and $\mathbf{R}_{\Delta u}$, are used to prioritize the outputs, inputs, and rate of change of the inputs, respectively.

2.3 Low-dimensional aeroelastic modeling framework

The model in equation (2.3) is formulated for a rigid wing and does not provide predictions of wing deformation. Prediction of deformation is essential for active control of flexible wing shapes. Flutter and wing vibrations have been a recognized problem since the early days of flight, and was a key reason for the development of Theodorsen's model. The aerodynamics community now recognizes the benefits of flexible structures, especially for increased efficiency, and advances in sensors, actuators, and algorithms are leading to successes in active flutter and vibration control.

We extend the rigid model in equation (2.3) to include wing deformation, resulting in a low-order, linear aeroelastic model. Because the rigid model has been widely studied, we seek to minimally modify this structure so that our model can fit naturally into existing model-based control efforts. This modeling method does not require any information about the structure, such as the flexural rigidity or mass distribution, as an input. The aeroelastic model developed here is designed for wings that passively bend, without any direct control over camber. Instead, wing vibrations can be attenuated using only pitch control in concert with a controller designed based on the linear model.

2.3.1 Aeroelastic model identification from an integrated impulse in $\dot{\alpha}$

To formulate the model, time series data of C_L and wing deformation are collected after an impulse response in $\dot{\alpha}$. In this work, wing deformation is described for a two-dimensional plate using the local curvature, κ, near the leading edge. Curvature at other locations could be chosen, based on any areas of particular structural concern or expectations about bending mode shapes, and higher order structural modes may be learned with higher-dimensional measurements along the chord. In an experimental setting or with a 3D wing, strain would generally be used instead of curvature, generally with at least one strain sensor near the wing root.

As discussed earlier, because of added mass forces, it is impractical to command an actual impulse in $\dot{\alpha}$ in direct numerical simulations or in experiments. Instead, a smoothed linear ramp-up in α is commanded over a short time $\Delta\tau_c$, which approximately corresponds to a discrete-time delta input to $\dot{\alpha}$. An example of the smoothed linear ramp-up maneuver can be seen in the appendix, Fig. 2.10, which is modified from the Eldredge maneuver [27, 72]. The measurements are integrated to obtain the response to an approximate impulse in $\ddot{\alpha}$. Using C_L and κ data from the response to an impulse in pitch velocity, a model of the following form is built using a similar procedure as for the rigid unsteady aerodynamic models [8]:

$$\frac{d}{dt}\begin{bmatrix}\mathbf{x}\\ \alpha\\ \dot{\alpha}\end{bmatrix} = \begin{bmatrix}\mathbf{A} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & 0 & 1\\ \mathbf{0} & 0 & 0\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \alpha\\ \dot{\alpha}\end{bmatrix} + \begin{bmatrix}\mathbf{B}\\ 0\\ 1\end{bmatrix}\ddot{\alpha}, \tag{2.9a}$$

$$\begin{bmatrix}C_L\\ \kappa\end{bmatrix} = \begin{bmatrix}| & | & |\\ \mathbf{C} & C_\alpha & C_{\dot{\alpha}}\\ | & | & |\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \alpha\\ \dot{\alpha}\end{bmatrix} + \begin{bmatrix}|\\ C_{\ddot{\alpha}}\\ |\end{bmatrix}\ddot{\alpha}. \tag{2.9b}$$

In particular, the coefficients C_α, $C_{\dot{\alpha}}$, and $C_{\ddot{\alpha}}$ are identified by isolating specific components of the output response, and the remaining transient lift is modeled by the ODE in $\mathbf{x}$ using ERA. The vector of coefficients associated with the quasi-steady lift and deformation, $C_\alpha = \mathbf{Y}_N/\Delta\alpha$, is found by dividing the steady state outputs by the magnitude of the step change in angle of attack, $\Delta\alpha$. $\mathbf{Y}_N$ is the last measurement, taken when transients have largely decayed. $C_{\dot{\alpha}} = \mathbf{Y}_m/\dot{\alpha}_m$ is found from the moment of maximum impulse in $\dot{\alpha}$, when $\ddot{\alpha} = 0$, at time τ_m. $C_{\ddot{\alpha}}$ and the Markov parameters for ERA are found from the response of an impulse in $\ddot{\alpha}$, which is achieved with integration, $\tilde{\mathbf{Y}} = \int \mathbf{Y} d\tau_c$, starting at the point of maximum impulse, τ_m. This integration step has benefits for noise filtering and suppression, which is discussed further in the appendix. The added mass from acceleration is $C_{\ddot{\alpha}} = \Delta\tau_c\tilde{\mathbf{Y}}_0/\Delta\alpha$, where $\Delta\tau_c$ is the time length of the impulse maneuver. The accuracy of the identified coefficients is improved by using sampling with a finer time step than the Markov parameters during the maneuver, to determine the empirical coefficients. Once the Theodorsen-like coefficients have been determined, ERA is used to identify the remaining transients using the coarsely-sampled Markov parameters. Details of the method described in this section can be found in the appendix as well as a comparison with an alternate method which omits the integration step.

2.4 Results

We demonstrate the modeling approach above, specifically the formulation in Section 2.3.1, using data from a high fidelity fluid-structure interaction numerical simulation [38]. The data is generated at $Re = 100$, with plate length and stiffness parameters that are similar to those of a insect wing. These parameters were chosen to model a system with significant effects from viscosity and wing deformation [13].

We show that the the low-rank linear model matches the high fidelity simulation for a rapid test maneuver, as long as the model rank is sufficient to capture the important plate bending modes. For model interpretation and control planning purposes, the predicted lift and deformation can be separated into contributions from quasi-steady and added mass effects, as well as the transient contributions from the viscous wake.

An important use of this aeroelastic model is to enable simultaneous control of unsteady aerodynamic forces and plate bending by actuation with only pitching motions. This is demonstrated in the high-fidelity FSI simulation by tracking an aggressive reference lift, followed by fast attenuation of plate vibrations, with control planning done with the low-rank linear model and model predictive control (MPC).

2.4.1 Direct numerical simulation

We generate training data to develop the low-rank model by performing direct numerical simulation of a flow over a two-dimensional thin deforming plate with a strongly-coupled immersed boundary projection method [38]. The fluid-structure interaction system is governed by three dimensionless parameters: Reynolds number $Re = cU_\infty/\nu$, mass ratio $M_\rho = \frac{\rho_s h}{\rho_f c}$, and bending stiffness $K_B = \frac{EI}{\rho_f U_\infty^2 c^3}$. Here, ρ_s and ρ_f are the density of the plate and fluid, respectively, c is the chord length of the two-dimensional plate, h is the thickness of the

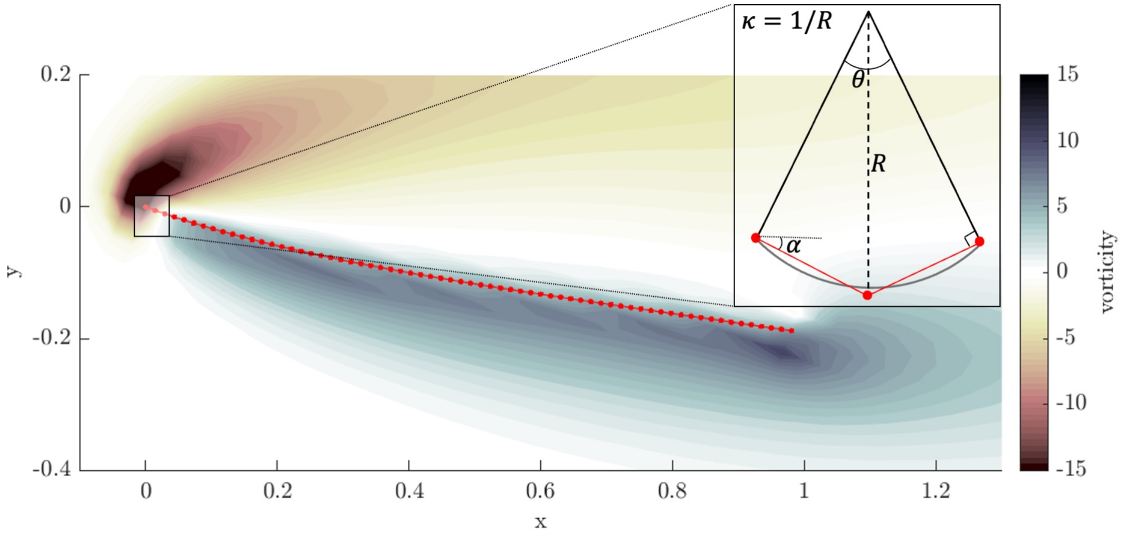


Figure 2.2: The angle of attack, α is measured from the leading edge; in the case illustrated $\alpha = -20°$. Curvature, κ, is measured from the angle formed by two discrete segments, in this case at the leading edge. There are 66 segments in the plate; red dots indicate the joints between segments. Plate curvature in the inset is exaggerated for illustration.

plate, and ν is the kinematic viscosity. We fix $Re = 100$ and $M_\rho = 3$ for all simulations in this work. Three different bending stiffnesses of varying orders of magnitude are considered with $K_B = \{0.3125, 3.125, 31.25\}$. The bending stiffness values were chosen to be similar to those of the leading and trailing edges of an insect wing [17], to model a system with significant wing deformation. Further details of the numerical simulation can be found in the appendix.

Plate deformation is described using curvature, κ, $0.015c$ downstream of the leading edge, across an area $0.03c$ in length, shown in Fig. 2.2. This location was chosen in this study because the two-dimensional plate is pinned at the leading edge, at the pitch point, which results in maximum curvature near this region. The curvature is defined as

$$\kappa = \frac{1}{R} = \frac{\theta}{2\Delta c}, \tag{2.10}$$

where $\Delta c = 0.015c$ is the length of one discrete segment of the plate.

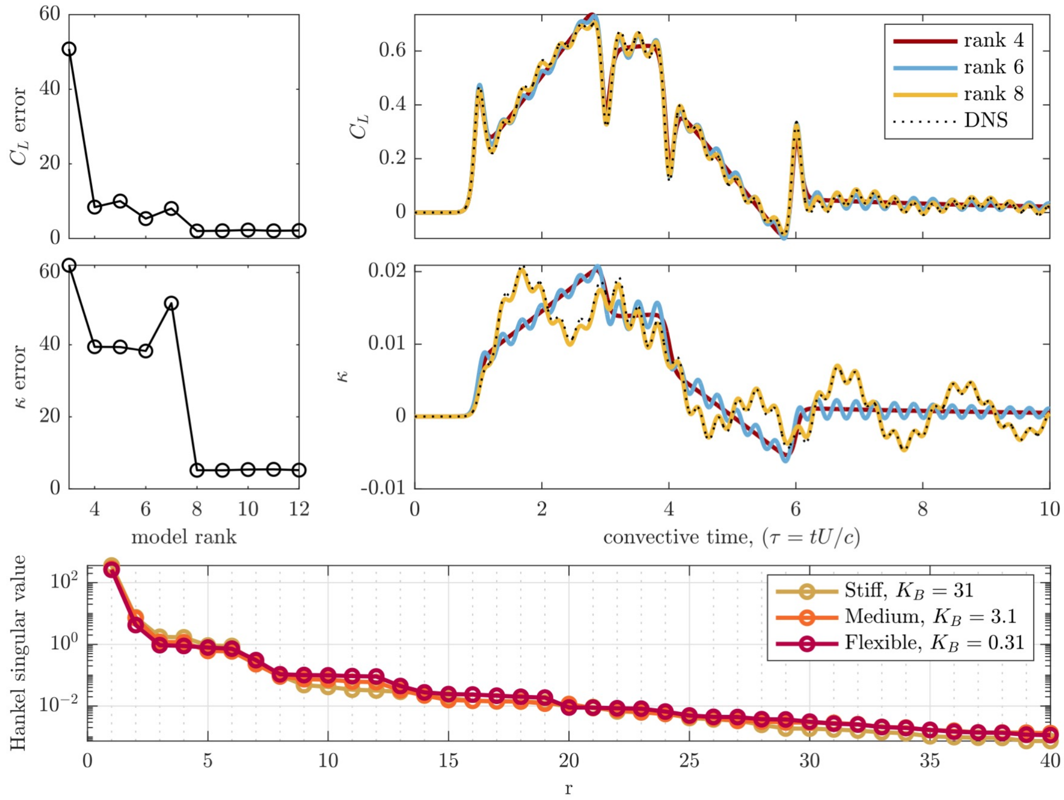


Figure 2.3: Model rank is chosen using a test maneuver, or using the Hankel singular values. (top and middle) Using an aggressive test maneuver, the error can be compared quantitatively (left) or qualitatively (right). When the model order is too low, the quasi-steady and added mass effects are captured, but not the transients associated with the viscous wake and plate bending modes. The models have one state for α, one for $\dot{\alpha}$, and the remaining states are associated with transients. The maneuver used for this test can be seen in the bottom panels of Fig. 2.5. The data shown in the upper plots is for bending stiffness $K_B = 3.1$. (bottom) If ground-truth data is difficult to obtain, the Hankel singular values from ERA can also be used to choose the model rank.

2.4.2 Model accuracy and comparison with Theodorsen's model

For the systems described in this chapter, the low-order linear model shows excellent agreement with the high fidelity simulation with a model rank of roughly eight. The model rank can be chosen from the Hankel singular values, which are the diagonal entries of Σ from Eq. 2.5. While there exist theoretically optimal thresholds for rank truncation using singular values [36], comparing with a test maneuver can provide a more interpretable picture of which physics are lost, such as high or low frequency flutter or wake vorticity. In this work, the model rank was chosen by comparison to high fidelity simulation data from an aggressive pitch-up, pitch-down test maneuver [8, 27, 72]. Both the Hankel singular values and test maneuver rank comparison are shown in Fig. 2.3. This maneuver, and the inclusion of several convective times with no wing actuation, are chosen to balance the effects of errors from fast maneuvers and from transients. The error, e, of the output of the reduced order model, $\mathbf{Y}_{ROM}$, compared to the output of the high fidelity numerical simulation, $\mathbf{Y}_{DNS}$, is

$$e = 100\frac{\sqrt{\sum(\mathbf{Y}_{DNS} - \mathbf{Y}_{ROM})^2}}{\sum \mathbf{Y}_{DNS}^2}. \tag{2.11}$$

The rank is chosen to capture the most energetic bending modes, which is achieved in the example shown in Fig. 2.3 with a rank of eight: six states to represent fluid transients and plate bending, and one state each for α and $\dot{\alpha}$. A frequency response plot, shown in Fig. 2.4, indicates that two bending modes are captured for this system. The models shown in the frequency response plot were developed using the method in Section 2.3.1. Models developed using the alternate method, described in the appendix, can capture additional bending modes if they are present in the signal, but are prone to overfitting if the chosen rank is too high, due to the higher-variance frequency content of the signal prior to integrating.

The empirically determined model captures resonance responses that are not captured with Theodorsen's model. For context, the first two natural frequencies of an undamped beam matching the plate properties were analytically calculated and are shown in the vertical lines in Fig. 2.4. The analytically calculated natural frequencies do not exactly match the model generated from data, likely due to fluid damping, although they are quite close. An advantage of the empirical models shown is they do not require knowledge of the fluid damping or the wing's natural frequencies to generate an accurate model.

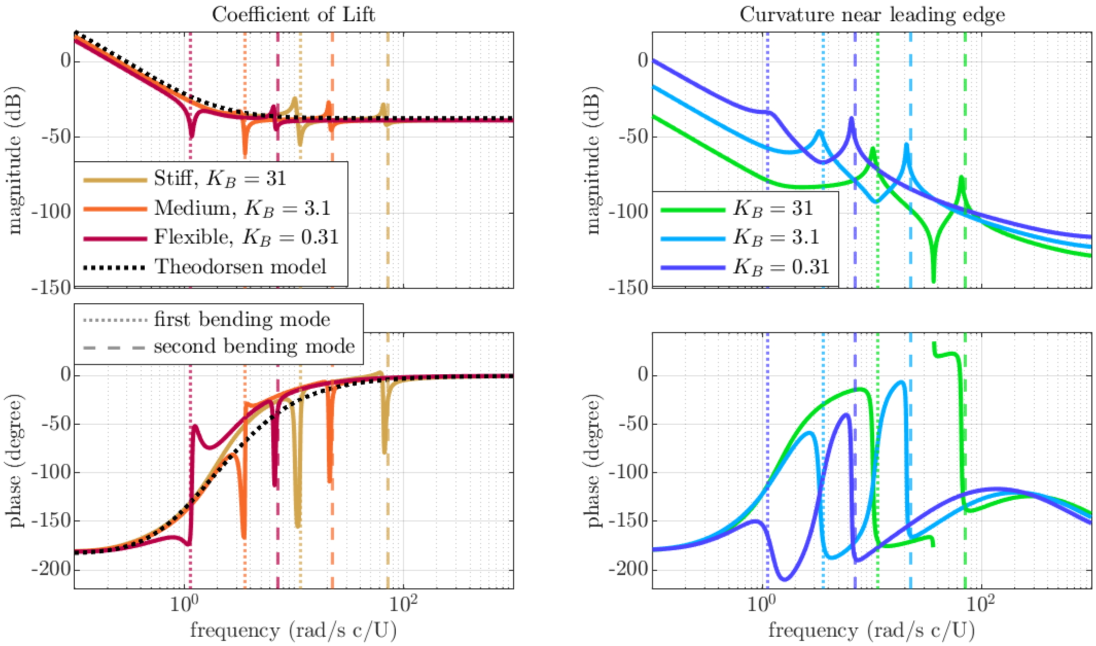


Figure 2.4: Frequency response of lift and deformation from forcing with $\ddot{\alpha}$. The Theodorsen model does not capture resonance behavior from bending modes or vortex shedding. The vertical lines indicating natural frequencies of the bending modes are based on analytically calculated, undamped, bending modes based on the plate dimensions and flexural rigidity.

2.4.3 Model interpretation

This modeling method allows the contributions from i) quasi-steady effects from α, ii) added mass effects due to $\dot{\alpha}$ and $\ddot{\alpha}$, and iii) plate vibrations, wake vorticity, and transients to be analyzed independently of each other. This information can be used to understand the physics driving the system behavior, and to design controllers that take advantage of these physical phenomena.

Fig. 2.5 shows these contributions for the test case described in Section 2.4.2. The most important contributions to leading edge curvature are α and the latent states representing transients and bending modes, while $\dot{\alpha}$ and $\ddot{\alpha}$ contribute relatively little. This can also be intuited by looking directly at the sign and magnitudes of the empirically determined **C** matrix, shown below for the model used in Figs. 2.3 and 2.5, with α in degrees:

$$\begin{bmatrix} C_L \\ \kappa \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{C_L} & 7.1 \times 10^{-2} & 7.9 \times 10^{-2} \\ \mathbf{C}_{\kappa} & 1.6 \times 10^{-3} & 4.0 \times 10^{-4} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \alpha \\ \dot{\alpha} \end{bmatrix} + \begin{bmatrix} 1.2 \times 10^{-2} \\ -6.1 \times 10^{-7} \end{bmatrix} \ddot{\alpha}. \tag{2.12}$$

For C_L, all of the components have important contributions. In this example, the contributions from $C_{\dot{\alpha}}$ may be inflated, but the modeling procedure is robust and adjusts for this in the full model by decreasing the contribution due to transients and wing bending. The overestimate of $C_{\dot{\alpha}}$ may be due to lift enhancement from transient plate bending at the point of maximum $\dot{\alpha}$ in the data used to develop the model.

2.4.4 Model demonstration with feedback control

To demonstrate how the reduced order aeroelastic models can aid in controller design, feedback control was used to track an aggressive reference lift trajectory while minimizing wing deformation. Model predictive control is chosen because it is straightforward to implement constraints, for example on κ or α, and to prioritize either reference tracking of lift or attenuation of wing structural oscillations. In this section, we demonstrate that modeling κ is essential for damping structural oscillations, and discuss how the model's empirical coefficients and structure can be used to interpret errors in reference tracking.

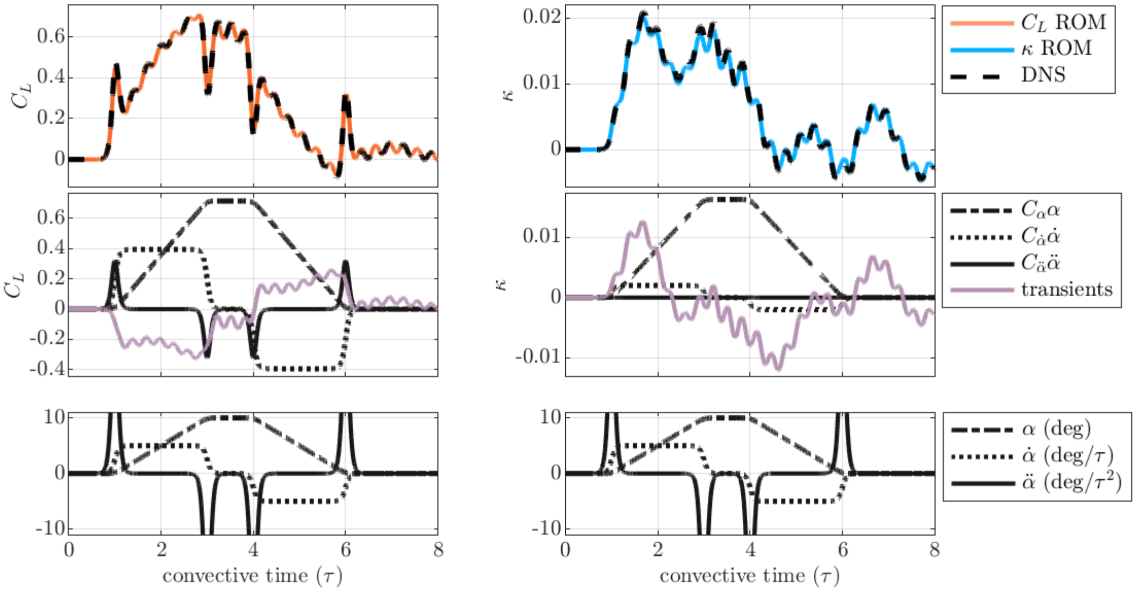


Figure 2.5: Contributions to lift and curvature from pitch angle, velocity, acceleration, and transients. In the top panels, the complete model is shown, and agrees well with the DNS. The middle panels show the contributions from each component of the model. The bottom panels show the test maneuver used. The data shown is for $K_B = 3.1$, with reduced order model rank of 9 (7 transient states).

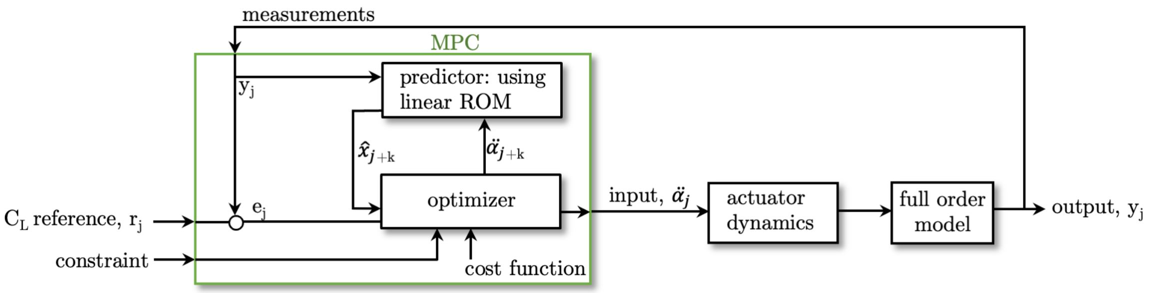


Figure 2.6: Schematic of model predictive control implementation. The predictions inside the control optimization loop are done using the linear reduced order model. The control actions are then applied to a full order fluid-structure interaction simulation.

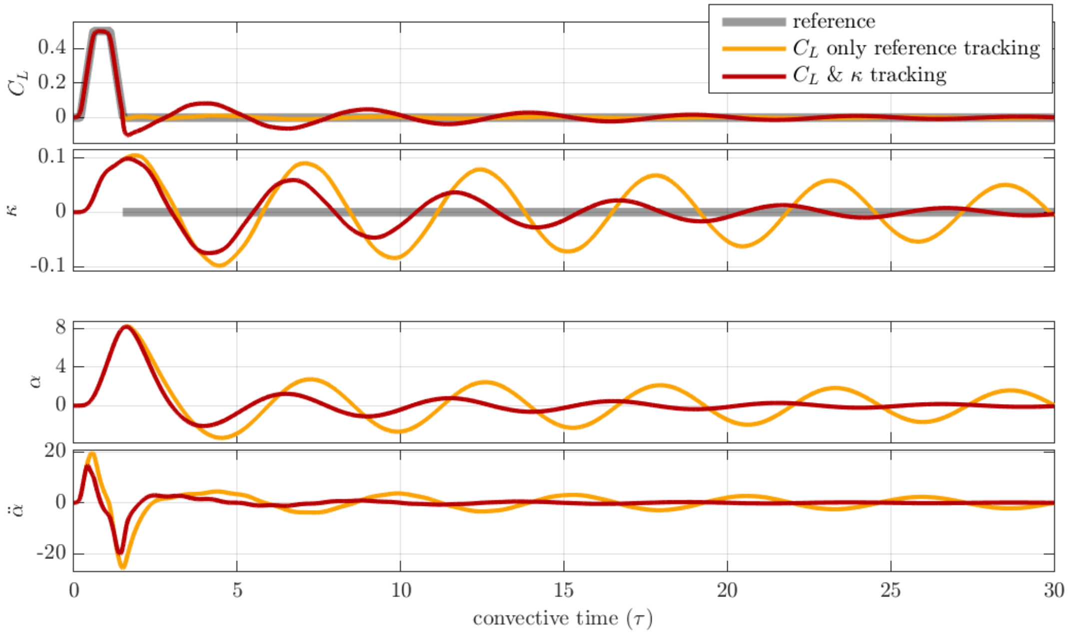


Figure 2.7: Lift and deformation reference tracking. The MPC actions are planned based on linear, reduced order state-space models, and are applied to a full-order direct numerical simulation. Two control cases are shown: i) tracking a reference value for only C_L, and ii) tracking reference C_L and κ. The reference κ shown is only applied to the second case. Angle of attack in degrees, α, and actuation, $\ddot{\alpha}$, are shown for reference.

A linear state-space model was first built from step-response data from a numerical simulation, using the procedure described in Section 2.3.1, resulting in the most flexible model shown in Fig. 2.4, for $K_B = 0.31$ and $r = 9$. The control actions were planned based on linear, reduced order state-space models, and were applied to the full-order fluid structure interaction simulation, shown schematically in Fig. 2.6. The MPC code used is based on that developed by Kaiser et al. [49]. The prediction and control horizons were both 20 time steps, $T_p = T_c = 20\Delta\tau$, with $\Delta\tau = 0.02$. A Kalman filter was used to estimate the states. Because real motors cannot respond instantaneously, actuator dynamics $G_a = 500/(s+500)$ were included in the feedback loop, where s is a Laplace domain variable.

A reference value for the wing deformation can only be specified and tracked if the wing deformation is predicted by the model. Fig. 2.7 shows two cases: i) tracking only

C_L, with no penalty for κ oscillations, and ii) tracking a reference C_L and κ, with the κ tracking starting after an impulse in C_L. The cost function in equation (2.8) was used with $\mathbf{R} = 0.001$ and

$$\mathbf{Q}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{Q}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 10 \end{bmatrix}, \tag{2.13}$$

where $\mathbf{Q}_1$ is used during periods with non-zero reference C_L, and $\mathbf{Q}_2$ is applied for the second case to minimize structural vibrations when the magnitude of the reference C_L is zero. The first case, where only a reference for C_L is specified, shows that wing vibrations from rapid maneuvers persist for a long time after the maneuver, despite the well-controlled lift. When a reference deformation is specified, vibrations are quickly attenuated. There is some trade off in lift tracking performance when deformation is included in the MPC cost function. This model makes it possible to tune the optimization weights and control constraints to decide the right trade-off for a given control application, rather than relying only on lift tracking.

Limiting the maximum wing deformation is another goal of deformation control, to avoid damage from large stresses. Similar to the model described in Fig. 2.5 and equation (2.12), the model used in this section has significant contributions to C_L from $\ddot{\alpha}$, due to added mass; however, the largest contribution to κ is from α. Shown in Fig. 2.8 (left), this results in small deformations when the reference C_L is stepped-up and then stepped back down in a short time, because the MPC optimization is able to use added mass forces to generate most of the increase in C_L. However, if the high reference C_L is held for longer, there is no longer added mass due to acceleration, and the angle of attack must increase to sustain the C_L at the reference value. When the reference maneuver is too rapid, the controller is unable to respond quickly enough.

An understanding of the relative contributions from added mass and angle of attack can be used to balance the need for constraining deformation with desired trajectories. In Fig. 2.8 (right), the κ constraint does not affect shorter maneuvers, but leads to significant error for a longer maneuver. In this regime, constraining κ effectively also constrains α. As constraints on deformation become more aggressive, the lift of the controlled system falter earlier, as shown in Fig. 2.9.

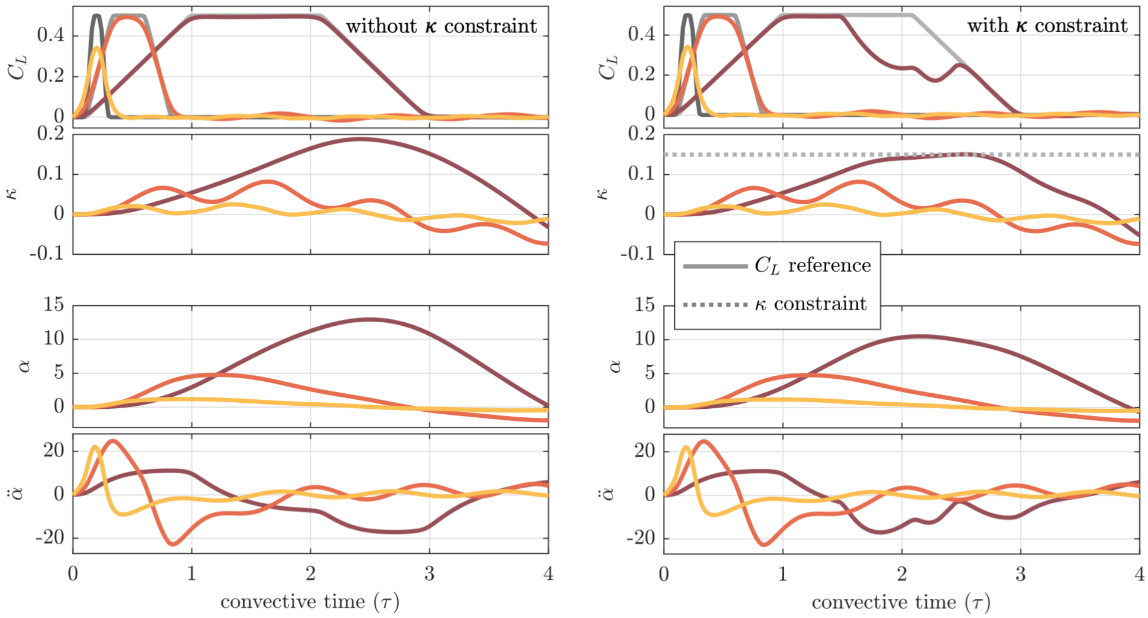


Figure 2.8: The maximum wing deformation depends strongly on the duration of the reference C_L step. (left) Reference C_L is increased, held, and decreased with three different rates of increase and hold time. If the hold time is too short, the system is unable to respond quickly enough. As the hold time increases, the angle of attack, α, must also increase, which results in higher maximum curvature, κ. (right) By choosing reference C_L with short enough hold times, moderate constraints on maximum κ do not affect reference tracking accuracy.

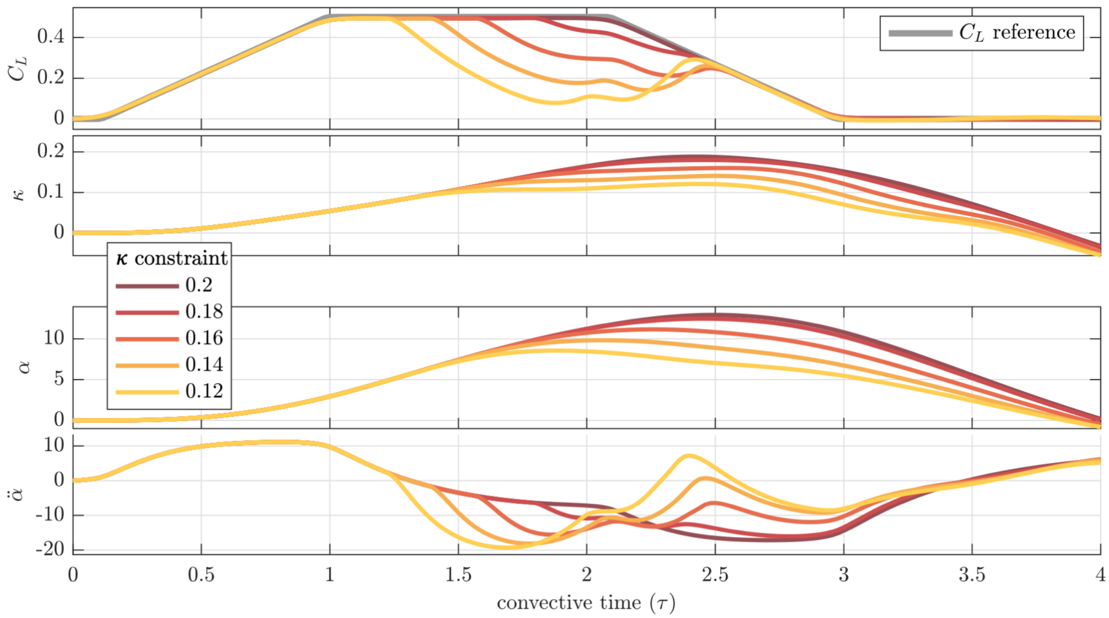


Figure 2.9: For systems with significant added mass effects on C_L, constraints on deformation decrease performance when C_L is held constant, but does not affect accuracy at the beginning of maneuvers.

2.5 Discussion

In this work, we describe a method for obtaining accurate low order, linear state-space aeroelastic models for control. The method uses lift and deformation data from an impulse response to construct the model, providing accurate predictions without requiring information about the surrounding flow field or wing structural properties. Interpretable coefficients relating to added mass, lift slope, and transient effects are built in to the model, providing insights about the underlying physics. The remaining dynamics are modeled using ERA, which accurately captures the transients dynamics due to the viscous wake. The resulting model is low dimensional and captures the dominant dynamics, extending rigid state-space aerodynamic models [8] to account for wing flexibility.

These models are well-suited for use with standard control techniques, which is demonstrated using MPC to track an aggressive reference lift while attenuating oscillations in leading edge curvature for an insect-inspired two-dimensional wing at $Re = 100$. These state-space models allow analysis and control design in both the time and frequency domains, including identification of resonant frequencies of the flexible structure. Because the wing deformation can be predicted, wing vibrations or flutter can be actively controlled by modulating only the angle of attack. These models could be used for stability analysis of wing designs, similar to work that has been done on vortex-induced vibration of stationary structures [12, 14].

Curvature was used as the measure of deformation in this chapter because the data was based on a simulation without a well-defined plate thickness. In real-world and experimental settings, strain may be a more appropriate deformation observable due to ease of measurement. Expanding the modeling algorithm to include a preprocessing step of determining the optimal strain sensor location may be of value [66]. Another possible extension of this modeling method is to include coefficient of thrust as an observable, giving a more complete picture of the forces on the wing, making the work more relevant for flapping and hovering flight.

The demonstration of this modeling framework at $Re = 100$ and mass ratio of $M_\rho = 3$ was particularly motivated by development of MAVs and to investigate control strategies used in animal flight, due to the inclusion of viscous effects. Managing wing loading while increasing efficiency at small scales is an ongoing challenge, which these models are designed to address. Further work is necessary to demonstrate the method presented for higher Re, applicable to wind turbines or aircraft, and for realistic 3D wings. For these applications, inclusion of multiple deformation or strain sensors is likely to be advantageous to capture relevant wing bending modes. For aeroelastic structures with low stiffness and high mass ratio, the wake may be irregular and transient wing deformations may no longer be periodic, limiting the effectiveness of this modeling procedure. It will also be important to extend these models to handle larger amplitude maneuvers, either by combining multiple linear models generated from maneuvers at several angles of attack using gain scheduling or LPV models [44], or by including nonlinear terms in the model.

2.6 Aeroelastic Modeling Appendix

2.6.1 Detailed procedure for obtaining model

Here we provide a detailed step-by-step procedure to identify unsteady aeroelastic models from data, roughly following the procedure for rigid unsteady aerodynamic modeling [8, 10]. Training data is in the form

$$\mathbf{Y} = \begin{bmatrix} | & | \\ C_L & \kappa \\ | & | \end{bmatrix}. \tag{2.14}$$

An example of the impulsive maneuver and training data is show in Fig. 2.10. The initial impulse data is shown, which will generally be only a small portion of the overall time series. Measurements should continue to be taken until the system reaches steady state. To obtain an approximately linear response, a small step amplitude, $\Delta\alpha \in [0.1^\circ,\ 1^\circ]$, and short time, $\Delta\tau \in [0.01,\ 0.1]$, should be used. For experimental data, it may be necessary to employ the observer-Kalman filter identification (OKID) approach to obtain the impulse response [48], which has been demonstrated for the identification of rigid unsteady aerodynamic models from data [8, 10].

The steps to obtain the coefficients in equation 2.9b are:

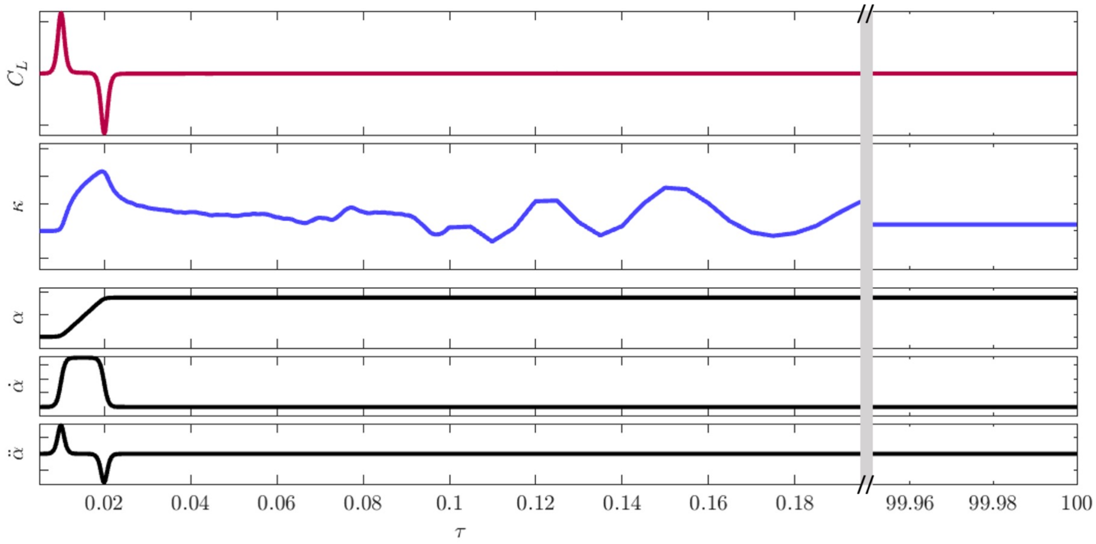


Figure 2.10: An impulse in $\dot{\alpha}$ is used to generate training data used for the system identification algorithm. The data shown is for $K_B = 0.31$, for a short time window during and after the impulse, as well as the end of the time series, when transients and wing bending have ceased.

1. Form the data matrix $\mathbf{Y}$ from a step response in α. Update $\mathbf{Y}$ by subtracting the initial values, $\mathbf{Y}_0 = \begin{bmatrix} C_L(\tau = 0) & \kappa(\tau = 0) \end{bmatrix}$ of each observable from the rest of the time series.

2. $C_\alpha = \mathbf{Y}_N / \Delta\alpha$, is found by dividing the steady state by the magnitude of the step change in angle of attack, $\Delta\alpha$. $\mathbf{Y}_N$ is the last measurement.

   Update $\mathbf{Y}$ by subtracting the steady-state contribution, $C_\alpha \alpha_k$ from each $\mathbf{Y}_k$.

3. $C_{\dot{\alpha}} = \mathbf{Y}_m / \dot{\alpha}_m$, found from the moment of maximum impulse in $\dot{\alpha}$, when $\ddot{\alpha} = 0$, at time τ_m. Update $\mathbf{Y}$ by subtracting the pitch velocity contribution, $C_{\dot{\alpha}} \dot{\alpha}_k$ from each $\mathbf{Y}_k$.

4. $C_{\ddot{\alpha}}$ and the Markov parameters for ERA are found from the response of an impulse in $\ddot{\alpha}$, which is achieved with integration, $\tilde{\mathbf{Y}} = \int \mathbf{Y} d\tau_c$, or in practice for a discrete signal,

with a cumulative sum, starting at the point of maximum impulse, τ_m.

5. $C_{\ddot{\alpha}} = \Delta\tau_c \tilde{\mathbf{Y}}_0 / \Delta\alpha$, where $\Delta\tau_c$ is the time length of the impulse maneuver.
6. The discrete-time state-space matrices, $\mathbf{A}_d$, $\mathbf{B}_d$, $\mathbf{C}$, $\mathbf{D}$, and the states, $\mathbf{x}$, are found using ERA, with the reduced order rank as an additional parameter. The Markov parameters used for ERA are the coarse-time integrated signal, $\tilde{\mathbf{Y}}$.
7. The discrete time matrices must be converted to continuous time before assembling the model in the form in equations 2.9b & 2.9a. If there were non-zero initial conditions, they should be added to the model as a constant term.

All models shown here were generated from an impulse magnitude of $\Delta\alpha = 0.1°$. The size of the Hankel matrix, $\mathbf{H}$, was roughly 5,000 x 5,000. The size of the Hankel matrix will depend on the duration of the impulse, and overall duration required for transients to die out.

2.6.2 Aeroelastic model identification from an impulse in $\dot{\alpha}$

Rather than integrating to get an impulse in $\ddot{\alpha}$, the model can be obtained directly from the impulse in $\dot{\alpha}$. This is done by omitting the signal integration step, as in [8], which results in a model of the form

$$\frac{d}{dt}\begin{bmatrix} \mathbf{x} \\ \alpha \\ \dot{\alpha} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{B} \\ \mathbf{0} & 0 & 1 \\ \mathbf{0} & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \alpha \\ \dot{\alpha} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ 0 \\ 1 \end{bmatrix} \ddot{\alpha}, \tag{2.15}$$

with the observables in the same form as equation (2.9b). This alternate method is problematic if there are several orders of magnitude between the spectral power contained in the first several bending modes, or significant noise at coherent frequencies, including due to sampling frequency. This is shown in Fig. 2.11, for $K_B = 3.1$, with the same test maneuver shown in Fig. 2.3. The spectral power contained in the C_L signal for the lowest frequency mode is two orders of magnitude smaller than the power contained in the second mode, and is similar to the power contained in high frequency noise. For the rank 9 model shown, the modeling procedure described in this section fails to capture the lowest frequency mode, resulting in an inaccurate model. By integrating the signal to obtain the Markov parameters, as described in the modeling procedure in Section 2.3.1, the low power bending mode is captured accurately, and the high frequency noise is smoothed, resulting in an accurate model. Increasing the model rank is not a solution to this problem, because the model rank required to capture the missing low amplitude mode generates spurious resonance peaks. However, an advantage to generating the model without integrating is that the model may capture higher frequency bending modes in a low noise signal, which may be lost when integrating the signal. Both methods were previously shown to produce accurate models for rigid systems, where the power spectrum of the Markov Parameters is dominated by vortex shedding rather than structural bending modes.

To obtain a model in the form of equation 2.15, the procedure above in Section 2.6.1 is followed through step 2. The remaining steps to obtain the method are described below.

3. Identify the time with maximum acceleration, τ_n. $C_{\ddot{\alpha}} = \ddot{\alpha}_n^{\dagger}\mathbf{Y}_n$, where † indicates the pseudo-inverse. Subtract the pitch acceleration contribution, $C_{\ddot{\alpha}}\ddot{\alpha}_k$ from each $\mathbf{Y}_k$.

4. The Markov parameters,$\bar{\mathbf{Y}}$, are the signal $\mathbf{Y}$ sampled at intervals of the coarse time step, $\Delta\tau_c$, starting at the time of maximum impulse, τ_m.

5. Because the impulse in this version of the method is from $\dot{\alpha}$, $C_{\dot{\alpha}}$ is found from the first Markov parameter; $C_{\dot{\alpha}} = \Delta\tau_c\bar{\mathbf{Y}}_0/\Delta\alpha$.

6. $\mathbf{A}_d$, $\mathbf{B}_d$, $\mathbf{C}$, $\mathbf{D}$, and $\mathbf{x}$, are found using ERA with $\bar{\mathbf{Y}}$, and then converted to continuous time and assembled into the form in equations 2.15 & 2.9b. If there were non-zero initial conditions, they should be added to the model as a constant term.

2.6.3 Fluid-structure interaction model

We performed direct numerical simulation of a flow over a two-dimensional thin deforming plate with a strongly-coupled immersed boundary projection method. The incompressible Navier–Stokes equation for the fluid was discretized in the vorticity-streamfunction form [88]

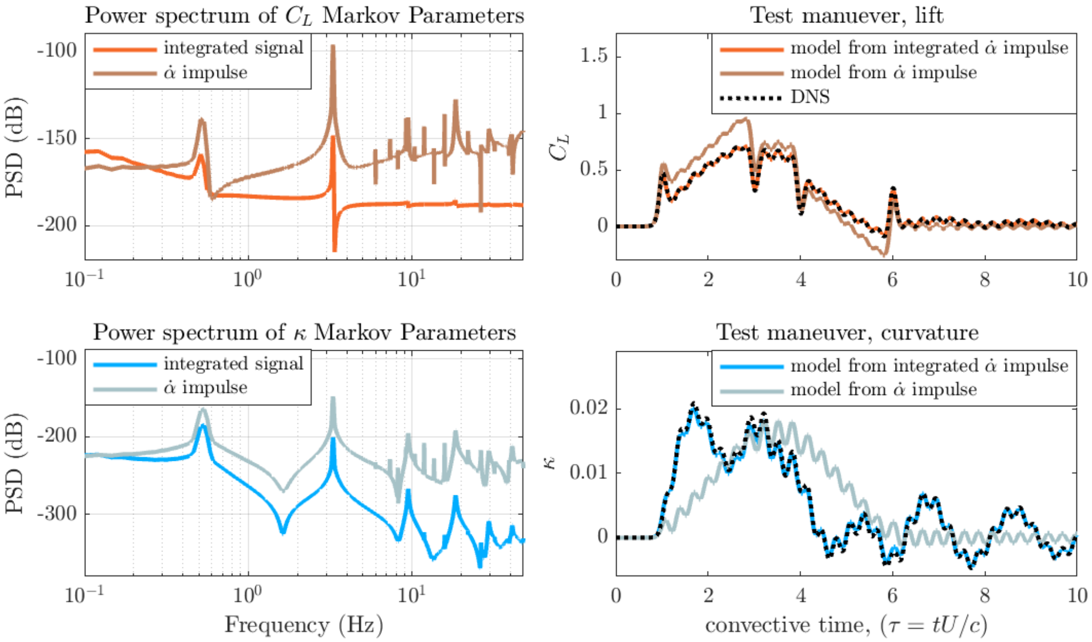


Figure 2.11: The procedure described in Section 2.3.1 (integrated signal) generates a more accurate model than the procedure using an $\dot{\alpha}$ impulse for some cases. For the case shown, $K_B = 3.1$ with rank 9 models, the lowest bending mode is not captured by ERA, resulting in an inaccurate model. (left) Power spectral density (PSD) of the Markov Parameters, $\tilde{\mathbf{Y}}$. Note the low power in C_L for the $\dot{\alpha}$ impulse for the peak at 0.5 Hz, the lowest frequency bending mode, compared to the power contained in high frequency noise. (right) The test maneuver shown in Fig. 2.3 is shown again here for each of the two modeling procedure, with rank 9 models, compared with data from the high fidelity DNS model.

with accurate surface stresses and forces to enforce the boundary condition at the plate [39]. At the far-field boundaries, uniform flow with free stream velocity U_∞ was prescribed. The solver uses an explicit Adam-Bashforth method and an implicit Crank-Nicolson scheme for discretization of the advective and viscous terms of the Navier–Stokes equation, respectively. This method was validated by Goza and Colonius for a flapping flag in [38].

To speed up the computations, a multi-domain technique with five grid levels was implemented [15]. The finest domain was fixed at $-0.2 \leq x/c \leq 1.8, -1 \leq y/c \leq 1$ with a grid spacing of $\Delta x/c \approx 0.0077$, represented by the red dot in Fig. 2.12. Here c is the length of the plate, x is the spatial domain location in the direction of the chord for a plate with no deformation, and y is the spatial domain perpendicular to the chord. The leading edge of the chord, which is the pitch axis, is located at $(0, 0)$. This grid spacing was chosen as a compromise between error and computational time, T_s, with an error within 0.3% of the steady-state lift coefficient. T_s is the time in seconds to run a single time step. The Reynolds number for all the simulations in this work is $Re \equiv U_\infty c/\nu = 100$, where ν is the kinematic viscosity.

The Euler-Bernoulli equation for the plate was discretized using a co-rotational finite element formulation [18]. This formulation enables arbitrary large displacements and rotations by attaching a local coordinate frame to each element. The plate was discretized into 65 elements with the leading edge placed at $(x/c, y/c) = (0, 0)$.

Data accessibility

The code used in this work is available at https://github.com/mhickner/aeroelastic-ss-model, including sample data which can be used to generate the models shown in section 2.4.

Authors and contributions

The published paper that forms the basis of this chapter was co-authored with Urban Fasel, Aditya Nair, Bingni Brunton, and Steven Brunton. Michelle Hickner wrote the chapter and generated the figures. This chapter was published as an article in the AIAA journal [], and was revised based on comments from several anonymous reviewers. Bingni Brunton and Steven Brunton supervised the work. The project was initiated by Aditya

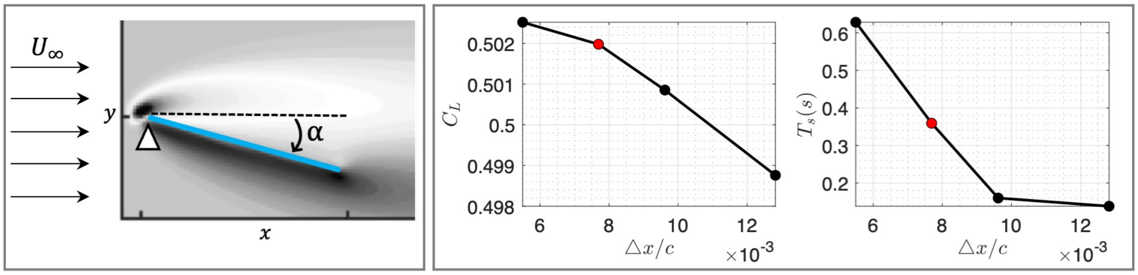


Figure 2.12: Schematic and convergence of direct numerical simulation, shown for an undeformed wing. (left) The angle of attack, α, is measured between the incoming free stream flow, and the plate at the leading edge. The pitch axis of the plate is also at the leading edge, where the angle of attack of a deformed or undeformed plate are the same. The full computational domain is not shown in this image. (right) The grid spacing of the finest mesh, shown in red, was chosen as a compromise between accuracy and computational time, T_s. A grid spacing of $\Delta x/c \approx 0.0077$, shown in red, error within 0.3% of the steady-state coefficient of lift, C_L.

Nair, Steven Brunton, and Bingni Brunton. Aditya Nair and Urban Fasel generated the fluid-structure interaction data, by using and modifying Fortran modelling code provided by Andres Goza [38, 40]. Michelle Hickner and Aditya Nair developed and refined the system identification method and wrote the system identification Matlab code, in consultation with all authors. Urban Fasel contributed to the model predictive control Matlab code, which is modified from work by Eurika Kaiser [49].

The authors acknowledge support for the work in this chapter from the Air Force Office of Scientific Research (AFOSR FA9550-19-1-0386) and the National Science Foundation AI Institute in Dynamic Systems (Grant No. 2112085).

Bibliography

[1] Edward Albano and William P. Rodden. A doublet-lattice method for calculating lift distributions on oscillating surfaces in subsonic flows. *AIAA Journal*, 7(11):2192a–2192a, 1969.

[2] David Amsallem, Julien Cortial, and Charbel Farhat. Toward real-time computational-fluid-

dynamics-based aeroelastic computations using a database of reduced-order information. *AIAA Journal*, 48(9):2029–2037, 2010.

[3] Marc Artola, Norberto Goizueta, Andrew Wynn, and Rafael Palacios. Aeroelastic Control and Estimation with a Minimal Nonlinear Modal Description. *AIAA Journal*, 59(7):2697–2713, 2021.

[4] Maciej Balajewicz and Earl Dowell. Reduced-Order Modeling of Flutter and Limit-Cycle Oscillations Using the Sparse Volterra Series. *Journal of Aircraft*, 49(6):1803–1812, 2012.

[5] Matthew Franklin Barone and Jeffrey L Payne. Methods for simulation-based analysis of fluid-structure interaction. *Sandia Report*, (SAND2005-6573), oct 2005.

[6] J. Birch and M. Dickinson. Spanwise flow and the attachment of the leading-edge vortex on insect wings. *Nature*, 412:729–733, 2001.

[7] Roger W Brockett. Volterra Series and Geometric Control Theory. *Automatica*, 12:167–176, 1976.

[8] Steven L Brunton, Scott T.M. Dawson, and Clarence W Rowley. State-space model identification and feedback control of unsteady aerodynamic forces. *Journal of Fluids and Structures*, 50:253–270, 2014.

[9] Steven L Brunton and Clarence W Rowley. Empirical state-space representations for Theodorsen's lift model. *Journal of Fluids and Structures*, 38:174–186, 2013.

[10] Steven L. Brunton, Clarence W. Rowley, and David R. Williams. Reduced-order unsteady aerodynamic models at low Reynolds numbers. *Journal of Fluid Mechanics*, 724:203–233, 2013.

[11] Matthew Bryant, Juan Carlos Gomez, and Ephrahim Garcia. Reduced-Order Aerodynamic Modeling of Flapping Wing Energy Harvesting at Low Reynolds Number. *AIAA Journal*, 51(12):2771–2782, 2013.

[12] S. R. Bukka, A. R. Magee, and R. K. Jaiman. Stability analysis of passive suppression for vortex-induced vibration. *Journal of Fluid Mechanics*, 886, 2020.

[13] Bo Cheng, Xinyan Deng, and Tyson L. Hedrick. The mechanics and control of pitching manoeuvres in a freely flying hawkmoth (Manduca sexta). *Journal of Experimental Biology*, 214(24):4092–4106, 2011.

[14] Amir Chizfahm and Rajeev Jaiman. Data-driven stability analysis and near-wake jet control for the vortex-induced vibration of a sphere. *Physics of Fluids*, 33(4):44104, 2021.

[15] T. Colonius and K. Taira. A fast immersed boundary method using a nullspace approach and multi-domain far-field boundary conditions. *Computer Methods in Applied Mechanics and Engineering*, 197:2131–2146, 2008.

[16] S. A. Combes and T. L. Daniel. Shape, flapping and flexion: wing and fin design for forward flight. *The Journal of Experimental Biology*, 204:2073–2085, 2001.

[17] S. A. Combes and T. L. Daniel. Flexural stiffness in insect wings II. Spatial distribution and dynamic wing bending. *Journal of Experimental Biology*, 206(17):2989–2997, 2003.

[18] MA Criesfield. Non-linear finite element analysis of solids and structures, vol. 1, 1991.

[19] J. O. Dabiri. Optimal vortex formation as a unifying principle in biological propulsion. *Annual Review of Fluid Mechanics*, 41:17–33, 2009.

[20] Thomas L Daniel. Forward flapping flight from flexible fins. *Canadian journal of zoology*, 66(3):630–638, 1988.

[21] Thomas L Daniel and Stacey A Combes. Flexible wings and fins: Bending by inertial or fluid-dynamic forces? In *Integrative and Comparative Biology*, volume 42, pages 1044–1049, 2002.

[22] M. H. Dickinson and K. G. Götz. The wake dynamics and flight forces of the fruit fly *Drosophila melanogaster*. *The Journal of Experimental Biology*, 199:2085–2104, 1996.

[23] M. H. Dickinson, F. O. Lehmann, and S. P. Sane. Wing rotation and the aerodynamic basis of insect flight. *Science*, 284(5422):1954–1960, 1999.

[24] E H Dowell. A modern course in aeroelasticity, 2016.

[25] Earl H Dowell and Kenneth C Hall. Modeling of Fluid-Structure Interaction. *Annual Review of Fluid Mechanics*, 33:445–90, 2001.

[26] Earl H. Dowell, Kenneth C. Hall, and Michael C. Romanowski. Eigenmode Analysis in Unsteady Aerodynamics: Reduced Order Models. *Applied Mechanics Reviews*, 50(6):371–386, jun 1997.

[27] J. D. Eldredge, C. Wang, and M. V. OL. A computational study of a canonical pitch-up, pitch-down wing maneuver. AIAA Paper 2009-3687, 39th Fluid Dynamics Conference, June 2009.

[28] Jeff D Eldredge and Anya R Jones. Leading-edge vortices: mechanics and modeling. *Annual Review of Fluid Mechanics*, 51:75–104, 2019.

[29] Urban Fasel, Paolo Tiso, Dominic Keidel, and Paolo Ermanni. Concurrent design and flight mission optimization of morphing airborne wind energy wings. *AIAA Journal*, 59(4):1254–1268, 2021.

[30] David Fernández-Gutiérrez and Wim M Van Rees. Effect of leading-edge curvature actuation on flapping fin performance. *J. Fluid Mech*, 921:22, 2021.

[31] Frank E Fish. Transitions from drag-based to lift-based propulsion in mammalian swimming. *American Zoologist*, 36(6):628–641, 1996.

[32] Frank E Fish and Clifford A Hui. Dolphin swimming–a review. *Mammal Review*, 21(4):181–195, 1991.

[33] Daniel Floryan and Clarence W Rowley. Distributed flexibility in inertial swimmers. *Journal of Fluid Mechanics*, 888, 2020.

[34] N. Fonzi, S. L. Brunton, and U. Fasel. Data-driven nonlinear aeroelastic models of morphing wings for control: Data-driven nonlinear aeroelastic models. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 476(2239), 2020.

[35] I. E. Garrick. On Some Reciprocal Relations in the Theory of Nonstationary Flow. *NACA Report 629*, pages 347–350, 1938.

[36] Matan Gavish and David L. Donoho. The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Transactions on Information Theory*, 60(8):5040–5053, 2014.

[37] Philippe Geuzaine, Gregory Brown, Chuck Harris, and Charbel Farhat. Aeroelastic dynamic analysis of a full f-16 configuration for various flight conditions. *AIAA journal*, 41(3):363–371, 2003.

[38] Andres Goza and Tim Colonius. A strongly-coupled immersed-boundary formulation for thin elastic structures. *Journal of Computational Physics*, 336:401–411, 2017.

[39] Andres Goza, Sebastian Liska, Benjamin Morley, and Tim Colonius. Accurate computation of surface stresses and forces with immersed boundary methods. *Journal of Computational Physics*, 321:860–873, 2016.

[40] Andres Goza, Sebastian Liska, Benjamin Morley, and Tim Colonius. Accurate computation of surface stresses and forces with immersed boundary methods. *Journal of Computational Physics*, 321:860–873, 2016.

[41] Kenneth C. Hall, Jeffrey P. Thomas, and Earl H. Dowell. Proper orthogonal decomposition technique for transonic unsteady aerodynamic flows. *AIAA journal*, 38(10):1853–1862, may 2000.

[42] Haotian Hang, Sina Heydari, John H Costello, and Eva Kanso. Active tail flexion in concert with passive hydrodynamic forces improves swimming speed and efficiency. *Journal of Fluid Mechanics*, 932, 2022.

[43] Anders Hedenström, LC Johansson, Marta Wolf, R Von Busse, Y Winter, and GR Spedding. Bat flight generates complex aerodynamic tracks. *Science*, 316(5826):894–897, 2007.

[44] Maziar S. Hemati, Scott T.M. Dawson, and Clarence W. Rowley. Parameter-varying aerodynamics models for aggressive pitching-response prediction. *AIAA Journal*, 55(3):693–701, 2017.

[45] Henrik Hesse and Rafael Palacios. Reduced-order aeroelastic models for dynamics of maneuvering flexible aircraft. *AIAA Journal*, 52(8):1717–1732, 2014.

[46] Robert T Jones. Operational treatment of the nonuniform-lift theory in airplane dynamics. *NACA*, (Technical Note 667):0–11, 1938.

[47] Jer-Nan Juang and Richard S. Pappa. An eigensystem realization algorithm for modal parameter identification and model reduction. *Journal of Guidance, Control, and Dynamics*, 8(5):620–627, sep 1985.

[48] Jer-Nan Juang, Minh Phan, Lucas G Horta, and Richard W Longman. Identification of Observer/Kalman Filter Markov Parameters: Theory and Experiments. *NASA Technical Memorandum*, (104069), 1991.

[49] E Kaiser, J N Kutz, and S L Brunton. Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474(2219), 2018.

[50] Dae-Kwan Kim, Jun-Seong Lee, Jin-Young Lee, and Jae-Hung Han. An aeroelastic analysis of a flexible flapping wing using modified strip theory. In *Active and Passive Smart Structures and Integrated Systems 2008*, volume 6928, 2008.

[51] Jiaqing Kou and Weiwei Zhang. Data-driven modeling for unsteady aerodynamics and aeroelasticity. *Progress in Aerospace Sciences*, 125(June):100725, 2021.

[52] Andrew J Kurdila, Bruce F Carroll, Toshi Nishida, and M Sheplak. Reduced order modeling for low Reynolds number flow control. In Vasundara V Varadan, editor, *Smart Structures and Materials 1999: Mathematics and Control in Smart Structures*, volume 3667, pages 68–79. International Society for Optics and Photonics, SPIE, 1999.

[53] George V Lauder and Eric D Tytell. Hydrodynamics of undulatory propulsion. *Fish physiology*, 23:425–468, 2005.

[54] Megan C Leftwich, Eric D Tytell, Avis H Cohen, and Alexander J Smits. Wake structures behind a swimming robotic lamprey with a passively flexible tail. *The Journal of Experimental Biology*, 215(3):416–425, 2012.

[55] J. Gordon Leishman. *Principles of Helicopter Aerodynamics*. Cambridge University Press, 2006.

[56] Guojun Li, Gaël Kemp, Rajeev Kumar Jaiman, and Boo Cheong Khoo. A high-fidelity numerical study on the propulsive performance of pitching flexible plates. *Physics of Fluids*, 33(5):051901, may 2021.

[57] James C Liao, David N Beal, George V Lauder, and Michael S Triantafyllou. Fish exploiting vortices decrease muscle activity. *Science*, 302(5650):1566–1569, 2003.

[58] T. Lieu, C. Farhat, and M. Lesoinne. Reduced-order fluid/structure modeling of a complete aircraft configuration. *Computer Methods in Applied Mechanics and Engineering*, 195(41-43):5730–5742, aug 2006.

[59] Haojie Liu, Xiumin Gao, Zhaolin Chen, and Fan Yang. Efficient reduced-order aerodynamic modeling in low-Reynolds-number incompressible flows. *Aerospace Science and Technology*, 119:107199, 2021.

[60] Jie Liu, Rajeev K. Jaiman, and Pardha S. Gurugubelli. A stable second-order scheme for fluid–structure interaction with strong added-mass effects. *Journal of Computational Physics*, 270:687–710, aug 2014.

[61] David J Lucia, Philip S Beran, and Paul I King. Reduced-order modeling of an elastic panel in transonic flow. *Journal of Aircraft*, 40(2):338–347, 2003.

[62] David J. Lucia, Philip S. Beran, and Walter A. Silva. Reduced-order modeling: New approaches for computational physics. *Progress in Aerospace Sciences*, 40(1-2):51–117, 2004.

[63] David J. Lucia, Philip S. Beran, and Walter A. Silva. Aeroelastic system development using proper orthogonal decomposition and volterra theory. *Journal of Aircraft*, 42(2):509–518, 2005.

[64] Zhanhua Ma, Sunil Ahuja, and Clarence W. Rowley. Reduced-order models for control of fluids using the eigensystem realization algorithm. *Theoretical and Computational Fluid Dynamics 2010 25:1*, 25(1):233–247, feb 2010.

[65] Rajat Mittal and Gianluca Iaccarino. Immersed boundary methods. *Annual Review of Fluid Mechanics*, 37:239–261, jan 2005.

[66] Thomas L Mohren, Thomas L Daniel, Steven L Brunton, and Bingni W Brunton. Neural-inspired sensors enable sparse, efficient classification of spatiotemporal data. *Proceedings of the National Academy of Sciences*, 115(42):10564–10569, 2018.

[67] Andrew M Mountcastle and Stacey A Combes. Wing flexibility enhances load-lifting capacity in bumblebees. *Proceedings of the Royal Society B: Biological Sciences*, 280(1759), 2013.

[68] Andrew M. Mountcastle and Stacey A. Combes. Biomechanical strategies for mitigating collision damage in insect wings: Structural design versus embedded elastic materials. *Journal of Experimental Biology*, 217(7):1108–1115, 2014.

[69] Joseba Murua, Rafael Palacios, and J. Michael R. Graham. Applications of the unsteady vortex-lattice method in aircraft aeroelasticity and flight dynamics. *Progress in Aerospace Sciences*, 55:46–72, 2012.

[70] Massimiliano Nardini, Simon J Illingworth, and Richard D Sandberg. Reduced-order modeling and feedback control of a flexible wing at low Reynolds numbers. *Journal of Fluids and Structures*, 79:137–157, 2018.

[71] J. C. Nawroth, H. Lee, A. W. Feinberg, C. M. Ripplinger, M. L. McCain, A. Grosberg, J. O. Dabiri, and K. K. Parker. A tissue-engineered jellyfish with biomimetic propulsion. *Nature Biotechnology*, 30:792–797, 2012.

[72] Michael V. OL, Aaron Altman, Jeff D. Eldredge, Daniel J. Garmann, and Yongsheng Lian. Résumé of the AIAA FDTC Low Reynolds Number Discussion Group's canonical cases. *48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition*, (January), 2010.

[73] Max M.J. Opgenoord, Mark Drela, and Karen E. Willcox. Physics-based low-order model for transonic flutter prediction. *AIAA Journal*, 56(4):1519–1531, jan 2018.

[74] J. Peng and J. O. Dabiri. The 'upstream wake' of swimming and flying animals and its correlation with propulsive efficiency. *The Journal of Experimental Biology*, 211:2669–2677, 2008.

[75] Charles S. Peskin. The immersed boundary method. *Acta Numerica*, 11:479–517, 2002.

[76] D Poirel, Y Harris, and A Benaissa. Self-sustained aeroelastic oscillations of a NACA0012 airfoil at low-to-moderate Reynolds numbers. *Journal of Fluids and Structures*, 24:700–719, 2008.

[77] Heidi E Reid, Ryan K Schwab, Miles Maxcer, Robert K D Peterson, Erick L Johnson, and Mark Jankauski. Wing flexibility reduces the energetic requirements of insect flight. *Bioinspiration & Biomimetics*, 14(5):056007, 2019.

[78] Daniel K Riskin, David J Willis, José Iriarte-Díaz, Tyson L Hedrick, Mykhaylo Kostandov, Jian Chen, David H Laidlaw, Kenneth S Breuer, and Sharon M Swartz. Quantifying the complexity of bat wing kinematics. *Journal of Theoretical Biology*, 254(3):604–615, 2008.

[79] S. P. Sane and M. H. Dickinson. The control of flight force by a flapping wing: lift and drag production. *The Journal of Experimental Biology*, 204:2607–2626, 2001.

[80] M. J. Shelley and J. Zhang. Flapping and bending bodies interacting with fluid flows. *Annual Review of Fluid Mechanics*, 43:449–465, 2011.

[81] Wei Shyy, Hikaru Aono, Chang-kwon Kang, and Hao Liu. *Introduction to Flapping Wing Aerodynamics*. Cambridge University Press, 2013.

[82] W. A. Silva and D. E. Raveh. Development of unsteady aerodynamic state-space models from CFD-based pulse responses. In *42nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, page 1213, 2001.

[83] Walter A Silva. Simultaneous excitation of multiple-input/multiple-output CFD-based unsteady aerodynamic systems. In *Journal of Aircraft*, volume 45, pages 1267–1274, 2008.

[84] Sigurd. Skogestad and Ian Postlethwaite. Multivariable feedback control : analysis and design, 2005.

[85] Arnold Song, Xiaodong Tian, Emily Israeli, Ricardo Galvao, Kristin Bishop, Sharon Swartz, and Kenneth Breuer. Aeromechanics of membrane wings with implications for animal flight. *AIAA Journal*, 46(8):2096–2106, 2008.

[86] Bret Stanford, Michael Sytsma, Roberto Albertani, Dragos Viieru, Wei Shyy, and Peter Ifju. Static aeroelastic model validation of membrane micro air vehicle wings. *AIAA Journal*, 45(12):2828–2837, 2007.

[87] Haithem E Taha, Muhammad R Hajj, and Philip S Beran. State-space representation of the unsteady aerodynamics of flapping flight. *Aerospace Science and Technology*, 34:1–11, 2014.

[88] K. Taira and T. Colonius. The immersed boundary method: a projection approach. *Journal of Computational Physics*, 225:2118–2137, 2007.

[89] Graham K. Taylor and Holger G. Krapp. Sensory Systems and Flight Stability: What do Insects Measure and Why? *Advances in Insect Physiology*, 34(07):231–316, 2007.

[90] Theodore Theodorsen. General Theory of Aerodynamic Instability and the Mechanism of Flutter. *Technical Report, NACA*, (496):291–311, 1935.

[91] Jeffrey P. Thomas, Earl H. Dowell, and Kenneth C. Hall. Three-dimensional transonic aeroelasticity using proper orthogonal decomposition-based reduced-order models. *Journal of Aircraft*, 40(3):544–551, may 2003.

[92] Sonya Tiomkin and Daniella E. Raveh. A review of membrane-wing aeroelasticity. *Progress in Aerospace Sciences*, 126(May):100738, 2021.

[93] L Tregidgo, Z Wang, and I Gursul. Unsteady fluid-structure interactions of a pitching membrane wing. *Aerospace Science and Technology*, 28:79–90, 2013.

[94] E. D. Tytell and G. V. Lauder. The hydrodynamics of eel swimming i. wake structure. *The Journal of Experimental Biology*, 207:1825–1841, 2004.

[95] Eric D. Tytell, Megan C. Leftwich, Chia Yu Hsu, Boyce E. Griffith, Avis H. Cohen, Alexander J. Smits, Christina Hamlet, and Lisa J. Fauci. Role of body stiffness in undulatory swimming: Insights from robotic and computational models. *Physical Review Fluids*, 1(7):073202, nov 2016.

[96] R. Vepa. Finite state modeling of aeroelastic systems. *NASA Contracter Report*, CR-2779(February 1977), 1977.

[97] Herbert Wagner. Über die Entstehung des dynamischen Auftriebes von Tragflügeln. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 5(1):17–35, jan 1925.

[98] Yinan Wang, Xiaowei Zhao, Rafael Palacios, and Keisuke Otsuka. Aeroelastic Simulation of High-Aspect Ratio Wings with Intermittent Leading-Edge Separation. *AIAA JOURNAL*, 2021.

[99] Z. J. Wang. Dissecting insect flight. *Annual Review of Fluid Mechanics*, 37:183–210, 2005.

[100] J Wootton. Support and deformability in insect wings. *Journal of Zoology*, 193:447–468, 1981.

[101] T. Y. Wu. Fish swimming and bird/insect flight. *Annual Review of Fluid Mechanics*, 43:25–58, 2011.

[102] Zhijun Yang, Rui Huang, Haojie Liu, Yonghui Zhao, and Haiyan Hu. An improved nonlinear reduced-order modeling for transonic aeroelastic systems. *Journal of Fluids and Structures*, 94:102926, apr 2020.

[103] W. Yao and R. K. Jaiman. Feedback control of unstable flow and vortex-induced vibration using the eigensystem realization algorithm. *Journal of Fluid Mechanics*, 827:394–414, 2017.

[104] W. Yao and R. K. Jaiman. Model reduction and mechanism for the vortex-induced vibrations of bluff bodies. *Journal of Fluid Mechanics*, 827:357–393, 2017.

Chapter 3

TRADEOFFS BETWEEN AERODYNAMIC FORCE PRODUCTION AND SENSING IN INSECT WINGS AND HALTERES

Insect wings serve both a sensory and aerodynamic purpose. The hindwings in the order Diptera have become halteres, which serve a sensory purpose, but are not large enough to generate meaningful aerodynamic force, while the order Lepidoptera has specialized to have large hindwing area. We hypothesize that this functional specialization has been driven by a fundamental tradeoff between strain sensing and force production in flapping wings. We investigate this question using a computational finite element model of stereotyped wing and haltere shapes with realistic flapping kinematics and body rotations. This modeling approach allowed us to determine how varying morphology across many wing-like and haltere-like shapes results in changes in strain sensing of body rotations, and which directions of strain are most useful for gyroscopic sensing. We show that wing-like shapes and haltere-like shapes with a bulb on a stalk are both good shapes for sensing body motions using shear strains. Intermediate cases, such as very small wings, or thin short stalks without a mass on the end, have a weaker body rotation signal relative to the wing signal.

Nomenclature and Abbreviations

A	wing area
b	span length
c	chord length
C_L	coefficient of lift
Cr	Coriolis ratio
CS	campaniform sensilla
F	force
L	lift
PSD	power spectral density
$\mathbf{r}$	position vector
R	center of mass ratio
SR	strain ratio
t	time
U	air velocity
α	wing pitch angle
ϵ_{XY}	shear strain
ϵ_{XX}	chordwise normal strain
ϵ_{YY}	spanwise normal strain
θ	wing sweep angle
ρ	air density
ϕ	body roll angle
$\boldsymbol{\omega}$	body angular velocity

3.1 Introduction

Control of maneuverable flight depends on two essential components: lift force production from aerodynamic surfaces, and sensing for feedback control. In insects, wings serve both of these purposes, with sensory neurons distributed throughout the wing. This raises the question; is there a tradeoff between optimizing wing morphology for aerodynamic purposes and for sensory purposes? This possible tradeoff is of particular note when considering the order Diptera, which encompasses true flies, mosquitoes, and crane flies. Unlike most flying insects, which have two pairs of wings, flyers in the order Diptera have a single pair of wings. Derived from hindwings, Dipterans have halteres, which are slender club-shaped structures that serve a sensory purpose, but are too small to produce

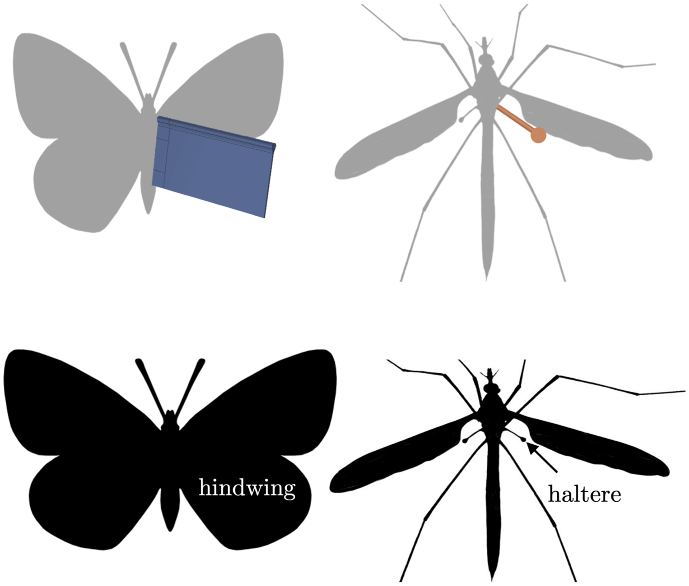


Figure 3.1: Flapping hindwings and halteres of various sizes are simulated using a finite element model. The haltere bulb, shown here on the orange haltere model, is simulated as a point mass.

any significant lift.

We hypothesize that tradeoffs between sensing and force production have led to the functional specialization of halteres. Among the hyper-diverse orders in Holometabola, both two-winged (Diptera) and four-winged (Hymenoptera and Lepidoptera) strategies have been hugely evolutionarily successful regarding abundance, number of species, and presence in most ecosystems around the globe. Coleoptera represents a third strategy, with elytra instead of forewings, which we do not investigate here. Among the hyper-diverse insect orders, Lepidoptera is the most closely related to Diptera. Many of the moths and butterflies of Lepidoptera have taken this wing specialization in a different direction, with large wing areas compared to their body mass [13].

For insects that spend considerable time flying during their adult stage, maneuverability is important for fitness, aiding in predator avoidance, collision avoidance in cluttered environments, and landing on moving surfaces such as bobbing flowers [13]. Insect flight guidance, navigation, and control use feedback from vision and mechanosensory neurons, including antennae, joint proprioceptors, and wing sensory neurons.

The mechanosensors we focus on are the strain-sensitive Campaniform sensilla (CS). We are most interested in strain near the base of wings or halteres, since that is where insects have large clusters of campaniform sensilla, packed into dense fields with distinct orientations, though there are also individual CS neurons distributed sparsely on veins throughout the wing surface [1, 2, 34]. By detecting strain signals, the CS on wings and halteres act as inertial sensors, similar to an accelerometer or gyroscope [17, 23, 28, 30, 31]. This information about how the insect is moving in space is key for feedback and feedforward control of steering and posture [3, 4, 10, 11, 24]. Campaniform sensilla on halteres and wings contribute to reflex circuits that are useful for fast body rotations, while visual feedback contributes primarily to slower adjustments [5, 33]. CS likely have directional sensitivity based on the shape of the cuticular dome around the neuron, and so some CS may be able to structuraly filter for only shear strain, while others filter for only normal strains in chordwise or spanwise directions [1, 17, 31].

3.1.1 Haltere evolution

While total wing area is correlated with body mass for flying insects, there is staggering diversity in the shapes and sizes of wings, even within orders [13, 15, 35]. While most flying insects have four wings, both wings are not always used primarily for aerodynamic purposes. For example, Coleoptera have replaced forewings with elytra, which primarily serve as protective covers while also sometimes playing a role in generating aerodynamic or inertial forces [16]. And of course, some insects no longer have wings. Something that makes the order Diptera interesting is their impressive maneuverability, despite not having hindwings. Halteres are significantly shorter than forewings and are shaped like clubs which narrow at the mid-span, and have asymmetric “bulbs” at the tips [1]. Halteres are flapped at the same frequency as but out of phase with the forewing [22].

Halteres and severely reduce hindwings have evolved multiple times. In extant animals, this includes primarily Diptera, as well as Strepsiptera. Strepsiptera (twisted wing parasites) have halteres instead of forewings. In this study, we focus on Diptera because Strepsiptera are less diverse than Diptera, and have a life cycle less dependent on flight. In Hemiptera, Coccomorpha males often have “hamulohalteres” which are severely reduced hindwings [21]. There is at least one report of a parasitic wasp in Hymenoptera, Mymarommatoidea in amber lacking hindwings [32]. There is at least one specimen of a mid-Cretaceous neuropteran, *Mantispidiptera henryi*, with a hindwing haltere-like structure [20].

3.1.2 Wing morphology affects maneuverability through sensing and flight forces

Stress and strain in a moving object depend on the shape and material properties of the object and the forces acting on it. As discussed in more detail in Section 3.5, as mass shifts towards the outer tip of a beam supported on its inner edge, we would expect more deflection of the beam, and therefore higher stresses and strains. Similarly, an asymmetric mass distribution about the long axis of the beam would be expected to result in more twisting, or torsion, of the beam, which increases shear strain. These effects have been demonstrated in computation studies of halteres with varying mass distributions [27]. The forces on wings and halteres are primarily inertial forces [8], which come from accelerating a mass. Wings and halteres accelerate due to both flapping and body motion. From the reference frame of the sensory neurons on the wing, the flapping kinematics interact with body accelerations resulting in Coriolis forces.

Sensing is just one element of aerial maneuverability; generating aerodynamic force is the other crucial component. To initiate or stop a turn, an animal must be able to both determine its position, velocity, and acceleration, and to generate enough force to recover from unwanted motions and follow a trajectory. Unlike passenger planes, insects fly in a transitional Reynolds number regime, which means that viscous forces and flow structures such as vortices play an important role in generating lift and drag. The wing accelerations from fast flapping and wing flexibility also contribute to generation of lift. Despite these unsteady forces dominating for insect flight, both passenger planes and insects must have large enough wings to generate and distribute their weight [13], as wing area affects how much aerodynamic force is generated. Even if sufficient force could be generated with a wing that is small relative the the size of the animal, the high wing loading, or weight supported per wing area, may result in broken wings.

3.1.3 Metrics for comparing wing and haltere morphologies

In this chapter, we present several methods for comparing wing and haltere morphologies for the purposes of gyroscoping sensing and production of flight forces. We first discuss the use of wing area as a straightforward proxy for the ability to generate aerodynamic force, despite the complexity of insect-scale unsteady aerodynamics. We then turn to comparisons of wing and haltere shapes for the specific task of differentiating between flapping and body rotation strain signals at the base of the wing. Our first two approaches for considering sensing involve calculating the forces due to flapping and body rotation, rather than calculating the wing deformation. One way to do this is to analytically derive these forces over time for the specific flapping kinematics, at a chosen location on the wing, using classical mechanics. These derivations yield time series which enable a nuanced investigation of the effect of kinematics, but are less well suited to summarizing the effect of morphology. To capture the effects of both kinematics and morphology in a compact way, we develop a scalar metric that we call *Coriolis ratio* (Cr). We look at whether the Coriolis ratio can be used in for solid mechanics in a similar way to the Rossby number is used in fluid mechanics; as a ratio between Coriolis forces and other inertial forces. Our final metric for investigating the effects of morphology goes beyond just looking at forces, and uses strain signals. This metric, called *strain ratio* (SR), compares the strain signal power from body rotation to the strain signal power from flapping. Evaluating this metric requires high-fidelity strain signals, which we obtain using a finite element model of a variety of flapping wing-like and haltere-like shapes. We then discuss how the metrics above contribute to evidence regarding whether tradeoffs between sensing and flight may have led to functional specialization of halteres and the divergence between four-winged and two-winged insects.

3.2 Aerodynamic forces and wing area

When we describe the lift force on a wing at a given time with a single number, what we are really describing is the sum, or integral, of the pressure across the entire surface of the wing. When the lift force points up on a horizontal wing, that indicates that the average pressure on the underside of the wing is lower than the average pressure on the top side of the wing. While there are several important factors that determine how the pressure is distributed across the wing, such as pitch relative to the incoming air and camber (bending in the chordwise direction), once those conditions are met, there must be sufficient wing area for those pressures to act on. This relatiionship between lift force and wing area can be seen with this equation,

$$L = AC_L\frac{\rho U^2}{2}, \tag{3.1}$$

which relates lift L, to wing area A, coefficient of lift C_L, air density ρ, and incoming air velocity U. The coefficient of lift is a dimensionless quantity captures effects such the pitch angle relative to the wind (angle of attack) and the camber (curve) of the wing. The incoming air velocity includes both the wind speed and the speed of the insect itself, including wing motion along the direction of motion such as sweeping back and forth. A similar equation describes drag force.

In this study, the largest affect on the lift is the area of the wings, because we vary the shape of the wings, but not the kinematics or the material properties. For that reason, we report wing area as a proxy for total aerodynamic force.

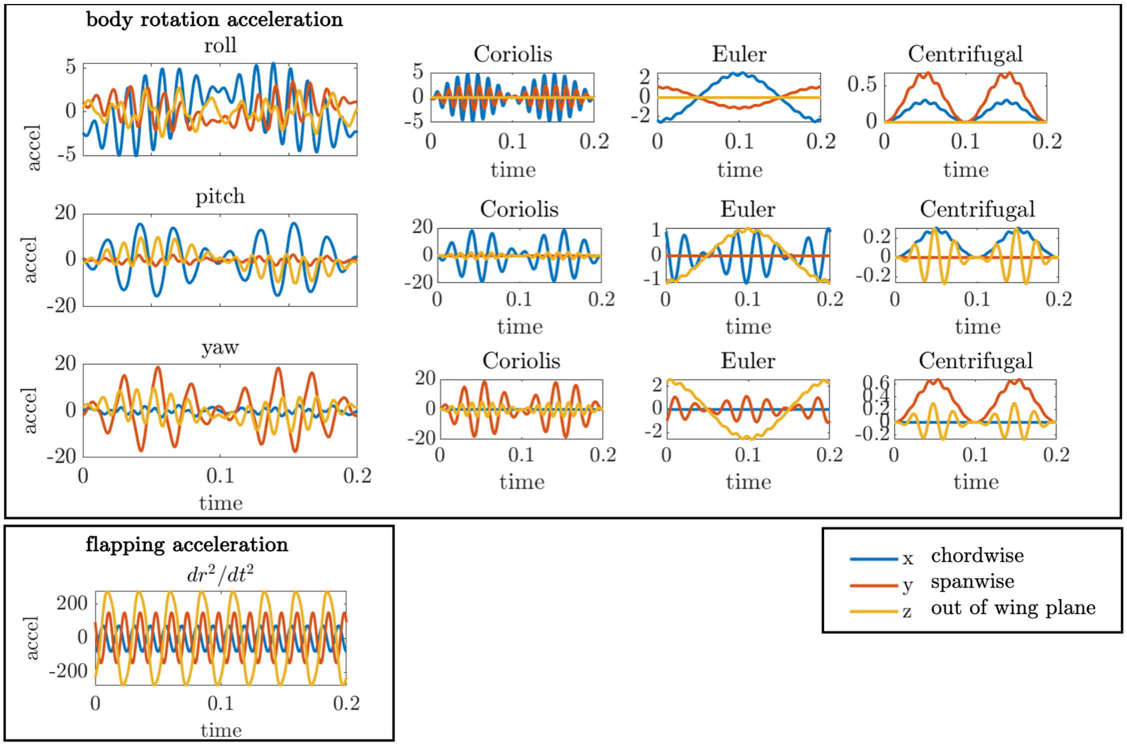


Figure 3.2: Flapping accelerations and body rotation accelerations for a single point on a wing, 1 cm spanwise and 4.5 mm chordwise from the pivot point at the base of the wing.

3.3 Gyroscopic forces

From the reference frame of the insect body, the forces on a given point on the wing are those from flapping, $F_{flapping}$, and those from rotating, $F_{rotation}$. The body rotation forces include centrifugal, Coriolis, and Euler forces:

$$F_{total} = F_{flapping} + F_{rotation} + F_{elastic} + F_{damping}, \tag{3.2a}$$

$$F_{flapping} = m\frac{d^2\mathbf{r}}{dt^2}, \tag{3.2b}$$

$$F_{rotation} = -\underbrace{m\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r})}_{\text{Centrifugal Force}} - \underbrace{2m\boldsymbol{\omega}\times\frac{d\mathbf{r}}{dt}}_{\text{Coriolis Force}} - \underbrace{m\frac{d\boldsymbol{\omega}}{dt}\times\mathbf{r}}_{\text{Euler Force}}. \tag{3.2c}$$

In these simulations, there is no linear body acceleration along the flight path, and gravity is not included in the model. The wing flapping kinematics interact with the rotating frame forces through the instantaneous position, $\mathbf{r}$, and velocity, $d\mathbf{r}/dt$ of a given position on the wing. This can be clearly seen in Figure 3.2, where the faster beating due to flapping affects the rotating frame forces, shown for a single body roll oscillation cycle. Note that the accelerations in Figure 3.2 are only those due to body rotation; wing flapping accelerations are not included. The largest accelerations due to body rotation come from the Coriolis force. The Euler force, which is associated with angular acceleration, also contributes, to a lesser extent, especially in the z-direction. Centrifugal accelerations from body rotation are a less significant contributor. Wing flapping accelerations are considerably larger than body rotation accelerations for this and most realistic cases.

To approach an understanding of how wing morphology affects gyroscopic sensing, we can look at the ratio of the rotating frame accelerations to the flapping accelerations, shown in Figure 3.3. For these accelerations that are calculated at a single point at the furthest spanwise and chordwise locations, at the trailing edge tip, differences in wing morphology only result in small changes in this ratio. However, when the forces across the entire wing are integrated and translated into strain, differences in wing morphology are magnified, which is later shown in Section 3.5. It is interesting to note that the way morphology affects this ratio varies depending on the direction of body rotation. This is because of the way that the rotating frame forces depend on the flapping kinematics, which in this case include sweeping (rotating about the x axis) and pitching (rotating about the y axis), in the same trajectories as are prescribed in Section 3.5.

3.4 Coriolis ratio

In fluid mechanics, there is a nondimensional number called the Rossby number, which describes the ratio of Coriolis to fluid-inertial effects. Here we take this concept from fluid dynamics, and derive

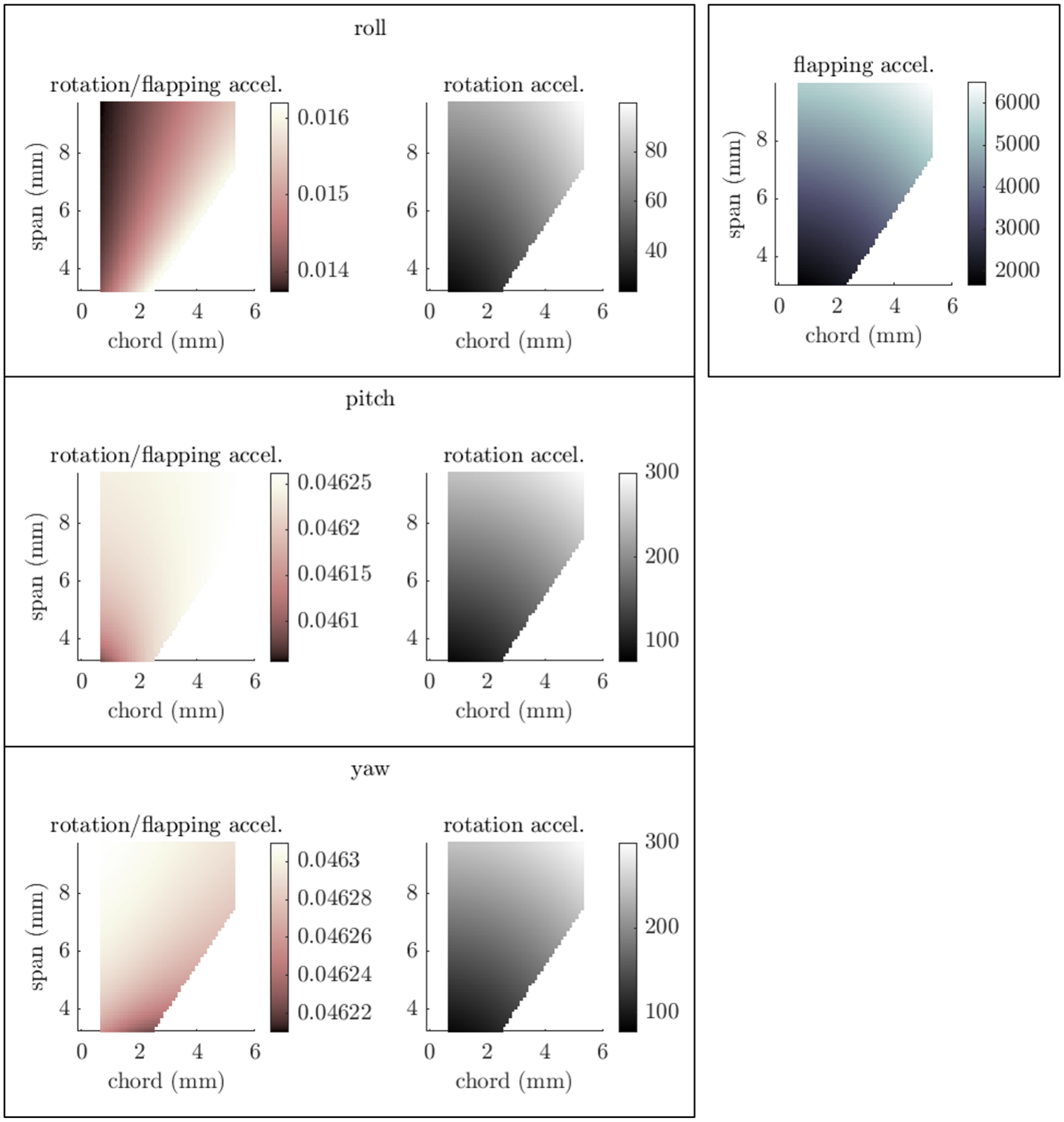


Figure 3.3: The ratio of body rotation acceleration to flapping acceleration, shown on the left in pink for a single point on the tip and trailing edge at various wing shapes, is the metric we are interested in.

a version of this number for the solid mechanics of an oscillating body, such as a flapping wing in a rotating (body) frame. The Rossby number is defined in fluid mechanics

$$Ro = \frac{U}{2L\omega sin(\phi)}, \tag{3.3}$$

where U is the fluid velocity, ω is the angular velocity of the rotating object or frame, and when used for planetary motion, ϕ is the latitude. L is a characteristic length scale, for example the chord length, or the diameter of a spinning cylinder or sphere. Among other applications, the Rossby number has been used to describe aerodynamics of flapping flight [6, 26].

Here we take this concept from fluid dynamics, and derive a version of this number for the solid mechanics of an oscillating body, such as a flapping wing in a rotating (body) frame. The most general form of this number, which we call Coriolis ratio (Cr), is

$$Cr = \frac{f}{2R\omega sin(a)}, \tag{3.4}$$

where ω is the magnitude of peak angular velocity of the rotating (body) frame, a is the amplitude of the wing flapping, and f is the frequency of the wing flapping. R is a non-dimensional number describing how far the center of mass is from the pivot location or base of the wing. For a flat rectangular wing with the base of the wing located at the leading edge, R would be equal to the aspect ratio of a single wing, b/c, because the center of mass would be at the center of the wing, at $\bar{\mathbf{r}} = (c/2, b/2)$. For a wing shape with nonuniform mass distribution or a non-rectangular shape, the center of mass is at

$$\bar{\mathbf{r}} = (\bar{x}, \bar{y}) = \frac{1}{M}(\int x\, dm, \int y\, dm), \tag{3.5}$$

with the coordinate system origin at the pivot point (base) of the wing. M is the total mass of the wing. R is the ratio of the center of mass coordinates, and could be considered a measure of how symmetric the wing is about the axis of rotation:

$$R = \frac{\bar{x}}{\bar{y}}. \tag{3.6}$$

In the above equation, the y direction is in the spanwise direction, and x direction is in the chordwise direction. Asymmetry out of the plane of the wing, in the z direction, would also contribute with certain flapping and body motions, though for simplicity we have initially formulated this with a wing that is symmetric in the z direction.

To derive Cr, we start by simplifying the problem by treating the flexible wing as a rigid body, with forces acting on the center of mass. This simplification is what allows us to do our calculations based only on the location of the center of mass, $\bar{\mathbf{r}}$. We next calculate the velocity at the center of mass for use in calculating the Coriolis force. For a pitching and sweeping wing with pitch amplitude a_α, sweep amplitude a_θ, and flap frequency f in radians, the wing velocity is

$$\frac{d\mathbf{r}}{dt} = \begin{bmatrix} -a_\theta f\sin(ft) \\ a_\alpha f\cos(ft) \\ 0 \end{bmatrix} \times \begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{bmatrix}. \tag{3.7}$$

To find the peak Coriolis force, we maximize the quantity:

$$||F_{cor}|| = 2m\|\boldsymbol{\omega}\|\|\frac{d\mathbf{r}}{dt}\|sin(\gamma), \tag{3.8}$$

where γ is the angle between the body angular velocity vector and the wing linear velocity vector.

3.5 Strain ratio

In this study, the sensing modality of interest is strain sensing. Because of the nearly planar shape of wings, we use just three of the six elements of the strain tensor: spanwise normal strain ϵ_{YY}, chordwise normal strain ϵ_{XX}, and shear strain ϵ_{XY}. For an element with an undeformed chordwise length of X_0 and spanwise length of Y_0, these strains are usually defined as engineering strains,

$$\epsilon_{YY} = \frac{\Delta Y}{Y_0}, \tag{3.9a}$$

$$\epsilon_{XX} = \frac{\Delta X}{X_0}, \tag{3.9b}$$

$$\epsilon_{XY} = \frac{1}{2}\left(\frac{\Delta X}{X_0} + \frac{\Delta Y}{Y_0}\right). \tag{3.9c}$$

Because our finite element model includes nonlinear terms and the prescribed kinematics result in significant rotation, we used Green-Lagrange strain,

$$\epsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial X_j} + \frac{\partial u_j}{\partial X_i} + \frac{\partial u_k}{\partial X_i}\cdot\frac{\partial u_k}{\partial X_j}\right), \tag{3.10}$$

which is a nonlinear measure of the change in length of a small segment of the wing. This formulation is roughly equivalent to engineering strain in situations with rigid body rotations and small strains.

A metric we use to describe how easily body rotation can be sensed is the ratio of the power spectral density (PSD) of the strain due to body rotation to the PSD of strain due to flapping,

$$\text{strain ratio} = \frac{PSD(\epsilon_{body})}{PSD(\epsilon_{flap})}. \tag{3.11}$$

The power spectral density is calculated in the frequency domain, after performing a fast fourier transform on the strain signal at a given location on the wing. $PSD(\epsilon_{body})$ is the power at the body rotation frequency, 5 Hz and its first two harmonics, 10 Hz and 15 Hz. $PSD(\epsilon_{flap})$ is the power at the flapping frequency, 40 Hz, and its first two harmonics at 80 Hz and 120 Hz.

There are two key reasons why we use this strain ratio metric rather than the magnitude of the strain: i) we expect campaniform sensilla neurons to fire based on time-varying patterns in strain rather than magnitude of strain, and ii) we are interested in the task of differentiating strain due to body movement from strain due to wing flapping. This is analogous to the task of interpreting sensory differences between a repetitive stepping pattern, and disturbances such as rough ground or shifts in body weight.

The strains on the wing are due to primarily to inertial forces. These are forces that arise from the mass of the wing undergoing acceleration and deceleration from flapping and body movements. We have modeled flapping as rotational motion in two different directions, pitching and sweeping. The wing is in a rotating reference frame, which simulates body rotation. For the data we present in this section, the body rotation direction is roll, which results in forces due to rotation in three different orthogonal directions: wing sweep about an x-axis, wing pitch about a y-axis, and body roll about a z-axis.

In this study, we are ultimately interested in strains, not forces or accelerations. The accelerations in Figure 3.2 provide an overly simplistic view of forces on the wing, which is why the finite element model is necessary. Each element of the model has different amplitudes of these forces acting on it, because of the dependence on position and velocity. Even if an element near the base experiences small forces due to acceleration, the effects of forces near the tip of the wing propagate through the flexible wing due to material properties, such as the elasticity and damping.

We model wing deformation and strain for flapping wings and halteres, with and without body rotation. Based on previously reported insect wing kinematics [12, 14, 18, 25], the wings trace a roughly elliptical path. A rigid knob protruding past the base of the wing drives the base of the wing at a prescribed pitch angle α and sweep angle θ in radians of

$$\alpha = \frac{\pi}{6}\sin(2\pi 40t), \tag{3.12a}$$

$$\theta = \frac{\pi}{8}\cos(2\pi 40t). \tag{3.12b}$$

Body motion in the roll direction is simulated using a rotating frame, with the rotation angle in radians of

$$\phi = \frac{pi}{12}\cos(2\pi 5t). \tag{3.13}$$

For a real insect in flight, this body rotation could come from a planned turn by the animal, or from a disturbance such as a wind gust. Note that the wing flapping frequency is 40 Hz, while the body rotation is slower, at 5 Hz. This flapping frequency was chosen based on the scaling relationships between wing area and flapping frequency presented by Dudley [13] for a 1 cm^2 wing. In this study, the hindwing morphology is varied, while the forewing size is assumed to be fixed, which is why the flapping frequency is held constant, rather than scaling with faster wingbeats as the hindwing area shrinks. Both hindwings and halteres flap at the same frequency as the forewings, though halteres flap out of phase with the forewing.

The simplified geometry, shown in Figure 3.1, has a thick leading edge, which represents the costal vein. The remaining wing area is a 30 µm thick sheet with a Young's modulus of 3 GPa, representing the membrane. For longer or larger wing and haltere shapes, the leading edge vein has a 0.5 mm diameter, while for the smallest haltere shapes the leading edge vein diameter is 0.1 mm. The leading edge vein and the tapered fillet connecting the vein to the thin membrane have a Young's modulus of 0.3 GPa. The density of the wing is $1180\,\text{kg}\,\text{m}^{-3}$ and Poisson's ratio is 0.35. The wing is modeled as a linear elastic material. These properties were chosen to be within the range of previously reported wing shapes and material properties [7, 19, 29].

The purpose of the fillet is to smooth the transition between the leading edge vein and the membrane, which reduces numerical artifacts associated with sharp corners. It also ensures a slightly asymmetrical mass distribution, which has been shown to be relevant to the deformations of halteres, especially in torsion [27]. There is a point mass at the tip of the leading edge, which serves as a proxy for the bulb at the end of a haltere. By varying the chord (by shrinking the membrane), the span, and the mass of the bulb, the model goes from being hindwing-like to haltere-like.

The finite element model used is a purely structural model; it does not directly include effects from the fluid, as a fluid-structure interaction simulation would. Instead, the effect of the fluid is accounted for using a body damping force. The accuracy of this approach of using a damping term to approximate the effect of the fluid was demonstrated by Combes and Daniel [8, 9]. The body damping is applied to all elements of the model, using Rayleigh damping with frequencies $f_1 = 30.2\,\text{Hz}$ and $f_2 = 125.3\,\text{Hz}$, and damping ratios $\zeta_1 = \zeta_2 = 2$.

The finite element model is developed and solved using COMSOL version 6.1. The solution includes geometric nonlinear terms, to capture the deformation associated with rotation. The wing mesh for each shape is scaled based on the span, b. The cross-sections of the wings were meshed with a free triangular mesh with a maximum element size of b/80 and a curvature factor of 0.5. The surface is a swept mesh with a maximum element size of b/40 near the base of the leading edge vein, and b/20 elsewhere. An example of a meshed wing is shown in Figure 3.4.

Strains were simulated using generalized alpha, which is an implicit solver. Generalized alpha has

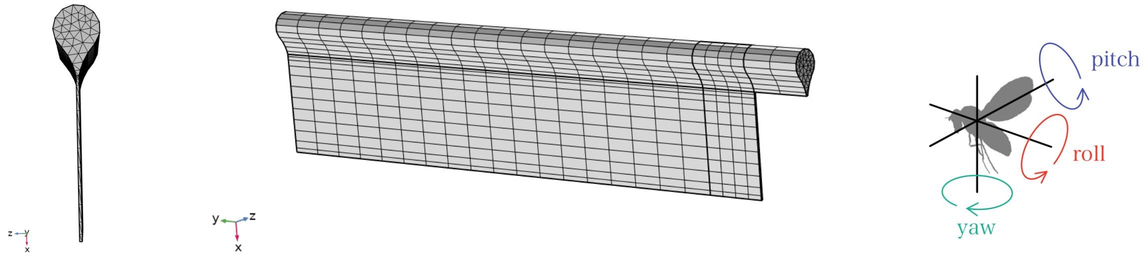


Figure 3.4: Finite element model mesh and coordinate system.

a parameter associated with damping of high frequencies, which was set to 0.25, which is moderate-low damping. Time steps were adaptive, with at least one solver step taken every $dt = 0.000\,125\,\mathrm{s}$. The relative tolerance was 0.001, and the absolute tolerance was scaled based on chordwise displacement with a scale factor of 0.1. Simulations include geometric nonlinearity, due to the prescribed rotations.

3.5.1 Wing morphology comparisons using strain ratio

In this section, we use the strain ratio metric described in the section above to compare a range of wing- and haltere-like shapes. This measure compares the power associated with the body rotation frequency to the power at the flapping frequency of the strain signal. We look at strains in several directions, based on the assumption that the fields of campaniform sensilla near the base of wings and halteres will include some cuticular domes that are oriented to be sensitive to each of the directions we focus on. The strains shown are measured at a distance of $0.15b$ from the base of the wing, on the dorsal side of the vein.

Examples of strain signals with high and low strain ratio are shown in Figure 3.5, for four cycles of body rotation. The faster frequency seen is at 40 Hz, the wingbeat frequency. On the example with a high strain ratio, a slower frequency is visible. The tops of the peaks are oscillating at 5 Hz, the body rotation frequency. The minimums of the peaks are varying at 10 Hz, likely due to the characteristic $2f$ frequency from Coriolis forces. In the low strain ratio example, the 5 Hz body rotation frequency is not clearly visible. These examples also illustrate how the strain signals are not purely sinusoidal or even symmetric between the up and down strokes, despite the sinusoidal and symmetric prescribed flapping at the base of the wing.

Because strain ratio is calculated from the frequency-domain power spectral densities, in Fig-

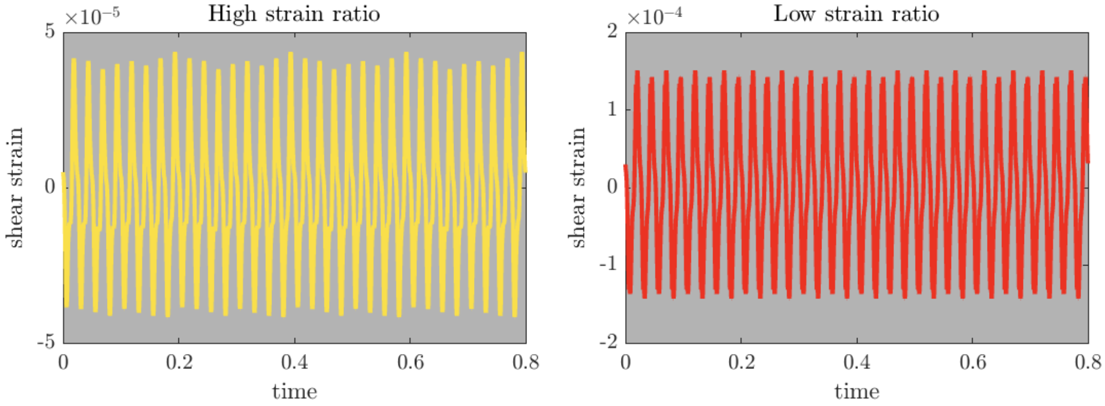


Figure 3.5: Examples of shear strain strain with high strain ratio (strong body rotation signal), and low strain ratio (low body rotation signal).

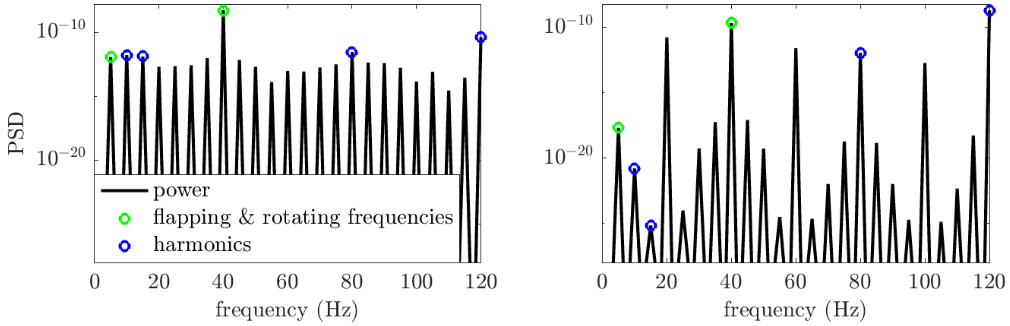


Figure 3.6: Examples of frequency-domain power spectra of shear strain strain for two morphologies with high strain ratio but different spectral qualities. The 5 Hz body rotation and the 40 Hz flapping frequencies are highlighted in red. left: b=8mm, c=3.5mm, the broadband content of the signal is the same order of magnitude as the body rotation signal. right: b=9mm, c=3.5mm, the body rotation signal is clearly distinguishable from the broadband spectral content.

ure 3.6, we show the PSD spectra for the two example time series in Figure 3.5. The overall power in any given peak is small, because strains are generally small quantities. For the example with high strain ratio, the flapping and rotating peaks are only a few orders of magnitude apart, but for the low strain ratio case, those peaks are separated by several orders of magnitude.

Large and mid-sized wings have larger shear strain ratios. Haltere-like shapes perform poorly if they are short and thick, but do well for longer shapes, and when mass is added to the tip. In Figure 3.7, we show the strain ratios for shear, spanwise, and chordwise strains, alongside some example wing-like and haltere-like shapes. It is worth noting that a large maximum strain magnitude, which are shown for comparison, does not directly translate to a large or small strain ratio. In the right plots, we vary both the chord and the span of the wing, and do not add a point mass at the tip of the leading edge vein. The shapes with the longest chords are the most wing-like, with the right-most point on each row having an aspect ratio of roughly 2. The left-most points on each row consist mostly of the leading edge vein. The left panels have halteres with leading edge (costal) veins that are the same diameter as the wings in the top panel. The middle panels are slender halteres, with leading edge veins that have a diameter five times smaller.

Some of the shapes we test would have structural limitations. In particular, the longest, thinnest shapes, such as those with a chord of 0.5 mm and a span of 7-10 mm.

3.5.2 Limitations of strain ratio

An insect would not need to detect a specific body rotation frequency, such as the 5 Hz here, nor would it detect frequency content through a fourier transform. Frequency domain analysis is a useful tool that allows us to generalize without considering the specific features or precision timing that is encoded by campaniform sensilla on a given animal, but it does have its limitations. Analysis of strain signals in the frequency domain has the limitation that fourier transforms interpret signals as sums of sines and cosines. For signals that repeat in time but are not shaped like sines and cosines, for example having sharp peaks or asymmetric shapes, fourier transforms will sum sines and cosines of many frequencies to match the signal. This results in significant power being shifted to broad frequency ranges, rather than concentrated in a single frequency, which can interfere with interpretation of the strain ratio.

3.6 Conclusion

Our hypothesis was that wing-like and haltere-like shapes would have a stronger body rotation signal relative to the flapping strain signal than intermediate shapes, and that halteres would have

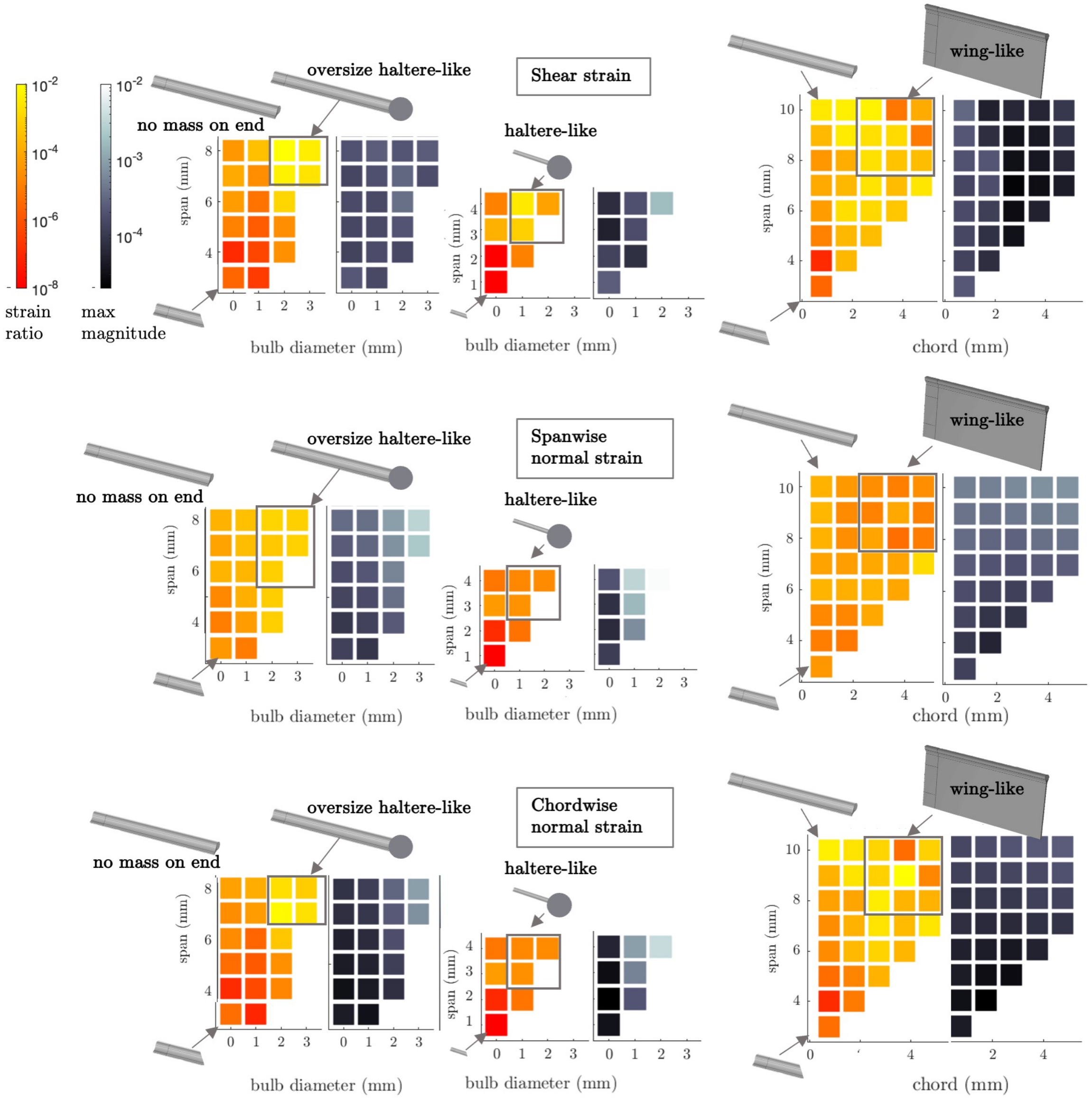


Figure 3.7: Strain ratio and maximum strain magnitude for shear, spanwise, and chordwise strains. In the left column, oversize halteres with the same leading edge vein thickness as the wings, with span and bulb mass varied. In the center, haltere-like shapes with a thinner leading edge vein, to scale with the shorter spans. In the right column, the chord length and span are varied.

the strongest body rotation signal. Previous work had indicated that shear strain from torsion is a key deformation mode for halteres. In this study, we found that wing shapes also exhibit strong body rotation signals with shear strain. Halteres also had good body rotation signals in shear, as long as they were slender and there was a mass at the tip of the stalk, filling the role of the haltere bulb. Without this additional mass, the halteres had poor performance, and our hypothesis that the haltere-like shapes would in general be better than wing shapes was not borne out. A point of interest is that the flapping frequency was chosen to be similar to that of a cranefly (family *Tipulidae*, and for this frequency and kinematics, the halteres most similar to the long, slender ones found in craneflies performed best.

At first, it may not seem like a surprise that larger wing shapes or haltere shapes with more mass towards the tip have the strongest signals, but in this case, "strongest signals" do not mean the largest magnitude. Rather, we use it to refer to the power in the strain signal due to body rotation, *relative to* the power in the strain signal due to flapping. That the range of shapes from wing-like to haltere-like results in relative signal strengths that vary by more than six orders of magnitude is in itself interesting. This shows us that the deformation due to Coriolis and Euler forces do not change at the same rate as the deformations due to flapping as size and shape change. This brings up the question whether some of the details of haltere shapes and kinematics that weren't captured in this study could be evolutionarily tuned to decrease forces from flapping acceleration and increase forces from body rotation.

While shear strains were of greater interest, normal strains, along the spanwise and chordwise directions, are also relevant to our hypothesis. Trends in normal strains were more dependent on the measurement location than trends in shear strain. Spanwise strain trends were straightforward: most shapes had body rotation signals near the middle of the range seen in this study, except for the shortest halteres, which had poor body rotation signal.

Contributors

Steven Brunton, Bingni Brunton, Urban Fasel, Thomas Daniel, and Michael Dickinson have contributed to the work in this chapter. Michelle Hickner wrote the text and generated the figures, except where otherwise noted, built the COMSOL finite element model, wrote the Matlab code for analyzing strain, and derived the Coriolis ratio metric. Bingni Brunton and Steven Brunton supervised the work. The hypothesis originated with Michael Dickinson. Urban Fasel provided input on the strain ratio metrics and finite element model. Thomas Daniel provided comments on insect biology and neuroscience. We also thank Sweta Agrawal, Tanvi Deora, Alison Weber, Mahnoush

Babaei, and Brian Polagye for illuminating conversations regarding this work.

The work in this chapter was supported by funding from the Air Force Office of Scientific Research (AFOSR FA9550-19-1-0386).

Bibliography

[1] Sweta Agrawal, David Grimaldi, and Jessica L Fox. Haltere morphology and campaniform sensilla arrangement across Diptera. *Arthropod Structure and Development*, 46(2):215–229, 2017.

[2] Brett R Aiello, Kathryn E Stanchak, Alison I Weber, Tanvi Deora, Simon Sponberg, and Bingni W Brunton. Spatial distribution of campaniform sensilla mechanosensors on wings: form, function, and phylogeny. *Current Opinion in Insect Science*, 48:8–17, 2021.

[3] Claire N. Balint and Michael H. Dickinson. The correlation between wing kinematics and steering muscle activity in the blowfly Calliphora vicina. *Journal of Experimental Biology*, 204(24):4213–4226, 2001.

[4] Jan Bartussek and Fritz Olaf Lehmann. Proprioceptive feedback determines visuomotor gain in Drosophila. *Royal Society Open Science*, 3(1):150562, jan 2016.

[5] John A. Bender and Michael H. Dickinson. A comparison of visual and haltere-mediated feedback in the control of body saccades in Drosophila melanogaster. *Journal of Experimental Biology*, 209(23):4597–4606, 2006.

[6] Shantanu S. Bhat, Jisheng Zhao, John Sheridan, Kerry Hourigan, and Mark C. Thompson. Uncoupling the effects of aspect ratio, Reynolds number and Rossby number on a rotating insect-wing planform. *Journal of Fluid Mechanics*, 859:921–948, 2019.

[7] S. A. Combes and T. L. Daniel. Flexural stiffness in insect wings: I. Scaling and the influence of wing venation. *Journal of Experimental Biology*, 206(17):2979–2987, sep 2003.

[8] S. A. Combes and T. L. Daniel. Into thin air: Contributions of aerodynamic and inertial-elastic forces to wing bending in the hawkmoth Manduca sexta. *Journal of Experimental Biology*, 206(17):2999–3006, 2003.

[9] Thomas L Daniel and Stacey A Combes. Flexible wings and fins: Bending by inertial or fluid-dynamic forces? *Integrative and Comparative Biology*, 42:1044–1049, 2002.

[10] Bradley H Dickerson. Timing precision in fly flight control: integrating mechanosensory

input with muscle physiology. *Proceedings of the Royal Society B: Biological Sciences*, 287(1941):20201774, 2020.

[11] Bradley H. Dickerson, Alysha M. de Souza, Ainul Huda, and Michael H. Dickinson. Flies Regulate Wing Motion via Active Control of a Dual-Function Gyroscope. *Current Biology*, 29(20):3517–3524.e3, 2019.

[12] Robert Dudley. Biomechanics of flight in neotropical butterflies: Morphometrics and kinematics. *Journal of Experimental Biology*, 150:37–53, 1990.

[13] Robert Dudley. *The Biomechanics of Insect Flight; Form, Function, Evolution.* Princeton University Press, 2000.

[14] C P Ellington. The aerodynamics of hovering insect flight. iii. kinematics. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 305:41–78, 1984.

[15] C P Ellington. The aerodynamics of hovering insect flight. II. Morphological parameters. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 305(1122):17–40, feb 1984.

[16] Sergey E Farisenkov, Dmitry Kolomenskiy, Pyotr N Petrov, Thomas Engels, Nadezhda A Lapina, Fritz Olaf Lehmann, Ryo Onishi, Hao Liu, and Alexey A Polilov. Novel flight style and light wings boost flight performance of tiny beetles. *Nature*, 602(7895):96–100, 2022.

[17] G. Fraenkel and J.W.S Pringle. Biological Sciences: Halteres of Flies as Gyroscopic Organs of Equilibrium. *Nature*, 141:919–920, 1938.

[18] Steven N. Fry, Rosalyn Sayaman, and Michael H. Dickinson. The aerodynamics of hovering flight in drosophila. *Journal of Experimental Biology*, 208:2303–2318, 6 2005.

[19] R Ganguli, S Gorb, F. O. Lehmann, S Mukherjee, and S Mukherjee. An Experimental and Numerical Study of Calliphora Wing Structure. *Experimental Mechanics*, 50(8):1183–1197, 2010.

[20] David A Grimaldi and Michael S Engel. Evolution of the insects, 2005.

[21] Penny J Gullan and Michael Kosztarab. Adaptations in scale insects. *Annual review of entomology*, 42:23–50, 1997.

[22] Joshua M Hall, Dane P. McLoughlin, Nicholas D Kathman, Alexandra M Yarger, Shwetha Mureli, and Jessica L Fox. Kinematic diversity suggests expanded roles for fly halteres. *Biology*

Letters, 11(11), 2015.

[23] Roland Hengstenberg. Gaze control in the blowfly calliphora : a multisensory, two-stage integration process. *Seminars in Neuroscience*, 3:19, 1991.

[24] Mark Jankauski, T.L. Daniel, and I.Y. Shen. Asymmetries in wing inertial and aerodynamic torques contribute to steering in flying insects. *Bioinspiration & Biomimetics*, 12(4):046001, jun 2017.

[25] Dmitry Kolomenskiy, Sridhar Ravi, Ru Xu, Kohei Ueyama, Timothy Jakobi, Thomas Engels, Toshiyuki Nakata, Jörn Sesterhenn, Kai Schneider, Ryo Onishi, and Hao Liu. The dynamics of passive feathering rotation in hovering flight of bumblebees. *Journal of Fluids and Structures*, 91, 2019.

[26] David Lentink and Michael H. Dickinson. Biofluiddynamic scaling of flapping, spinning and translating fins and wings. *Journal of Experimental Biology*, 212(16):2691–2704, 2009.

[27] T L Mohren, T L Daniel, A L Eberle, P G Reinhall, and J L Fox. Coriolis and centrifugal forces drive haltere deformations and influence spike timing. *Journal of the Royal Society Interface*, 16(153), 2019.

[28] G Nalbach. The halteres of the blowfly Calliphora I. Kinematics and dynamics. *J Comp Physiol A*, 173:293–300, 1993.

[29] Rizuwana Parween, Rudra Pratap, Tanvi Deora, and Sanjay P. Sane. Modeling strain sensing by the gyroscopic Halteres, in the Dipteran soldier fly, Hermetia illucens. *Mechanics Based Design of Structures and Machines*, 42(3):371–385, 2014.

[30] Brandon Pratt, Tanvi Deora, Thomas Mohren, and Thomas Daniel. Neural evidence supports a dual sensory-motor role for insect wings. *Proceedings of the Royal Society B: Biological Sciences*, 284, 2017.

[31] J.W.S Pringle. The gyroscopic mechanism of the halteres of Diptera. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 233(602):347–384, 1948.

[32] Alexandr P Rasnitsyn, Ekaterina A Sidorchuk, Haichun Zhang, and Qi Zhang. Dipterommatidae, a new family of parasitic wasps (Hymenoptera: Mymarommatoidea) in mid-Cretaceous Burmese amber: The first case of morphological diptery in flying Hymenoptera. *Cretaceous Research*, 104(May), 2019.

[33] Alana Sherman and Michael H. Dickinson. A comparison of visual and haltere-mediated equilibrium reflexes in the fruit fly Drosophila melanogaster. *Journal of Experimental Biology*, 206(2):295–302, 2003.

[34] Kathryn E Stanchak, Tanvi Deora, Alison I Weber, Michelle K Hickner, Abna Moalin, Laila Abdalla, Thomas L Daniel, and Bingni W Brunton. Intraspecific variation in the placement of campaniform sensilla on the wings of the hawkmoth Manduca sexta. *bioRxiv*, 2023.

[35] Maximillian P.T.G. Tercel, Fabio Veronesi, and Tom W. Pope. Phylogenetic clustering of wingbeat frequency and flight-associated morphometrics across insect orders. *Physiological Entomology*, 43(2):149–157, 2018.

Chapter 4

PRIMER: CONTROL THEORY FOR ANIMAL MOVEMENT

Feedback is a fundamental concept for understanding how animals precisely control their bodies in order to accomplish ecologically-relevant tasks that determine fitness. Sensory systems provide this critical feedback that is then integrated into a motor control system that dictates motor output and thus animal movement. Control theory provides a framework for modeling such systems—systems with variable applied forces, sensory feedback, and inherent mechanics. Consequently, control theory is increasingly used to study various aspects of sensorimotor control. However, the terminology and methods of control theory can be opaque to those without a background in applied mathematics or engineering. In this primer, we provide an overview of what we think are salient aspects of control theory for organismal biologists, including fully-worked examples. We aim to provide readers with sufficient background to model linear feedback control systems (linear optimal control). We also briefly touch on nonlinear control methods (model predictive control). Accompanying each topic are suggestions for further reading and learning.

Nomenclature

b	damping coefficient
k	spring constant
m	mass
$\mathbf{d}$	external disturbances that affect the system
$\mathbf{e}$	error between reference and state
$\mathbf{n}$	sensor noise
$\mathbf{p}$	proportional control gains
r	length of leg-spring in SLIP model
ref	reference, or desired, states
$\mathbf{x}$	state vector of state-space model
$\hat{\mathbf{x}}$	estimate of state vector
$\mathbf{x}_0$	initial condition
$\mathbf{y}$	measurements or observables in state-space model
$\mathbf{A}$	state-space matrix of the system dynamics, or how the states change in time
$\mathbf{B}$	state-space matrix of how the actuation affects the system dynamics
$\mathbf{C}$	state-space matrix of the transformation between the states and the sensor measurements
$\mathbf{D}$	state-space matrix of the direct effect of actuation on the sensor measurements
$\mathbf{K}_f$	Kalman estimator gains
$\mathbf{Q}_x$	optimization weights for minimizing state error
$\mathbf{Q}_u$	optimization weights for actuation
θ	angle of leg from foot-pivot in SLIP model

4.1 Introduction

Learning objectives:

- Understand the essential concepts and language of feedback control theory.
- Model simple physical systems in linear state-space form.
- Implement linear optimal controllers and observers to model motor outputs and sensory inputs.
- Understand the basics of how model predictive control can be used to model nonlinear feedback systems.

Animals exhibit extraordinarily precise and rapid control over their movements. This movement control is critical for capturing prey, evading predators, grooming, social communication, and many

Figure 4.1: (Clockwise from upper left.) Giant clams (genus *Tridacna*), like other bivalves, open and close in response to external stimuli (CC-BY-SA 3.0, Christoph Specjalski). Fish rely on external cues to maintain schooling behavior (CC-BY-SA 4.0, Anna Varona). A cheetah runs at high speeds over uneven terrain to catch prey (CC-BY-SA 3.0, Malene Thyssen). A hummingbird hawk moth can precisely align its proboscis with a flower while hovering (CC-BY-SA 4.0, Charles J. Sharp). A hovering kestral keeps its head still to watch prey while its body adjusts to wind (CC-BY-SA 2.0, William Warby). A yoga practicioner balances on one leg (CC-BY-SA 0.0). Greater flamingos balance on one leg while resting (CC-BY-SA 4.0, Charles J. Sharp).

other behaviors important for fitness (Fig. 4.1).

Animal movement is generally controlled through a feedback mechanism, via coordination of the sensory and motor aspects of the neuromuscular system (Fig. 4.2). Sensory neurons provide a feedback signal for motor control, allowing for constant adjustments as animal bodies interact with the outside world. Studying sensorimotor control requires a tractable but rigorous way to describe feedback cycles within the motion of physical systems—which is exactly the role of control theory. Control theory provides a flexible conceptual and mathematical framework for describing and analyzing the control of systems that change with time.

There is a rich history of control theorists contributing to biology research. For example, studies of central pattern generators in cats [9], processing of antenna mechanosensory information in cockroach wall following [12], motor control in reaching and other trajectories [17].

There are a wide variety of reasons why biologists might elect to co-opt control theory either themselves or through collaboration to study animal movement. However, to do this, biologists must be able to translate the mathematical language of control theory into biologically relevant concepts. Here, we demonstrate how control theory provides a flexible conceptual and mathematical framework that enables organismal biologists to describe, predict, and simulate many aspects of animal movement. First, we explain the basics of control theory, including the commonly encountered jargon. Then, we use examples to demonstrate the application of optimal control theory, in particular, to biologically relevant model systems. Importantly, we write out the math so that readers can fully follow along. Control theory is a huge field of study within applied mathematics and engineering; our examples specifically focus on how sensor information is incorporated into feedback control systems, especially when sensor information is incomplete or unreliable. Thus, we hope to enable further investigation of how neural sensors contribute to motor control.

4.1.1 Introduction summary

- Sensorimotor control of animal movement is a feedback control system.
- Control theory can provide a conceptual and mathematical framework for understanding the control of animal movements.

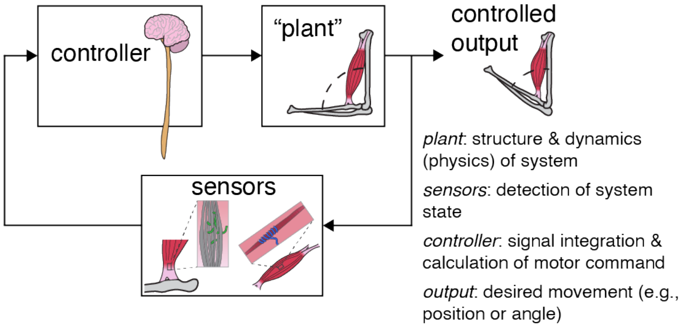
4.1.2 Further Reading

A longer perspective on why control theory is a useful tool for biology, including several examples.

Noah J Cowan, Mert M Ankarali, Jonathan P Dyhr, Manu S Madhav, Eatai Roth, Shahin Sefati, Simon Sponberg, Sarah A Stamper, Eric S Fortune, and Thomas L Daniel. Feedback control as a framework for understanding tradeoffs in biology. *American Zoologist*, 54(2):223–237, 2014

4.2 Background - Feedback Control Systems

Control theory applies to *dynamical systems.* A dynamical system is a system that changes its *state* with time. That is, at different points in time the state of the system (e.g., the position of a point) will have different values. How these states change with time are described by mathematical models. The movement of animal bodies can be mathematically modeled as a dynamical system, where the states are often positions and velocities of a body or limb.



plant: structure & dynamics (physics) of system

sensors: detection of system state

controller: signal integration & calculation of motor command

output: desired movement (e.g., position or angle)

Figure 4.2: Control theory is a valuable framework for describing and testing sensorimotor control of body positions, such as a bicep joint angle. The *plant* encompasses the model of the physical system–in this case, the structural properties of the bones, muscles, and ligaments, and their connections to each other. The controller computes the necessary signal given to the *actuators* (the muscles) to control the output (the joint angle). Sensors like the proprioceptive muscle spindle and Golgi tendon organ provide feedback on the current joint angle (or *state*) so that the controller can compute a proper response.

Control theory is a tractable way to describe complex state changes because it involves dividing the larger system into smaller systems and clearly defining how information flows between the different components. A typical control system consists of the system model (or *plant*), which comprises the physics (for animal movement, this would be the positions and velocities of parts of the body, including passive components such as spring-like dynamics); the actuators (muscles); sensors (neural sensors); and controller (e.g., central nervous system). Different models or equations can then be used to describe each of these components, with information flowing between them. Block diagrams like the conceptual version in Fig. 4.2 and the technical version in Fig. 4.3 illustrate this flow of information.

4.2.1 Open-loop and Closed-loop control

There are two broad classes of control systems: "open-loop" and "closed-loop", illustrated in Fig. 4.3a. In an open-loop system, there is no feedback loop; that is, there is no sensor that "tells" the controller whether to increase or decrease its action based on a desired setting. An everyday example is that of a classic toaster: the heating elements turn on when the appliance is turned on

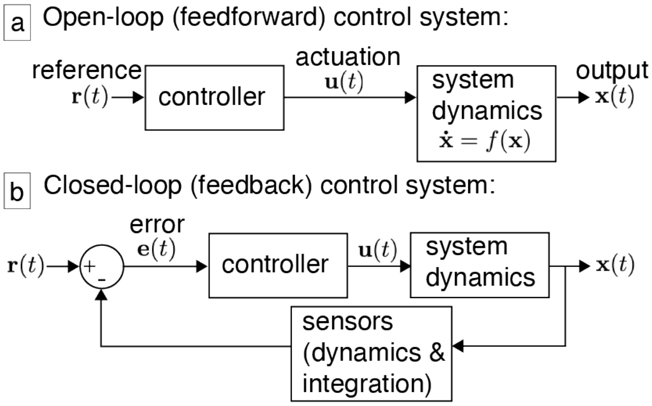


Figure 4.3: Block diagrams can be useful ways of diagramming the inputs, outputs, and the interactions between different subsystems of a control system, especially because they correspond directly with the mathematics of those interactions. In an open-loop control system (a), there is no feedback loop; a reference command is given to the controller, which computes an actuation that is provided to the plant. Because there is no feedback loop, the controller receives no information about the state of the plant and cannot correct for unanticipated disturbances or incorrect planning. In a closed-loop control system (b), there is a feedback loop that provides information on the current state of the system to the controller through the error signal. Then, in the next time step, the controller provide a different actuation signal to the plant to hopefully (if designed correctly!) compensate for the error between the output state and the reference.

and only turn off when the appliance reaches the end of a pre-determined time. There is nothing that tells the toaster when the bread is properly browned (unless you have a fancy, closed-loop control toaster), or even when proper temperatures are reached. Another open-loop example is the passive dynamic walker, which will walk down a gentle slope without any outside input [10]. If the passive dynamic walking machine isn't well tuned, or if it encounters disturbances such as slippery or uneven ground, it will likely fall over. While the original passive dynamic walker was actuated only by placing it an initial condition in a carefully balanced position at the top of a ramp, later models built on this idea by adding actuators to walk on flat ground, and eventually feedback control was added [5]. Open-loop systems can be highly advantageous because they minimize delays between the initiation of an action and the desired end result. Protective reflexes (e.g., blinking, vomiting) and escape responses can be thought of as open-loop—although the initial stimulus is sensory, the completion of the action is not modulated by continued sensory feedback. Sometimes, the decision to model a biological system as open-loop or closed-loop depends on the time scale or biological scale of interest. For instance, in the vestibulo-ocular reflex, a contraction of an ocular muscle (and the inhibition of the opposite) can be thought of as an open-loop reflex based on an initial input, while the stabilization of gaze over time is a closed-loop feedback-control problem. The bulk of control theory (and this tutorial) concerns itself with the more complicated "closed-loop" feedback control systems, but even when studying sensory feedback, it can be advantageous to "open the loop" to conduct experiments [14].

A closed-loop control system is defined by the presence of feedback, which is sensory information on the current state of the system that is "fed back" to the controller to determine the next action (Fig. 4.3b). A feedback control system can enable reliable performance even in the cases of poor internal models (i.e., models that inaccurately represent the system physics), noisy or incomplete sensory information, and external disturbances such as uneven terrain or wind. Feedback is also key for learning and optimization. The rest of this Background section will introduce the parts of a feedback control system, and the remainder of this tutorial will introduce different ways of implementing feedback controllers.

4.2.2 Describing the System Dynamics

The first step in formulating a control system model is to decide on the "state" variables. In organismal-scale biomechanics problems, joint or limb positions and velocities are often convenient state variables, because they are helpful for describing the system dyanmics and are often the outputs that you would like to model or measure. For example, in a system that includes an elbow joint, the vector of state variables might include the joint angle and the joint angular velocity, for a total of two states. Then, actuators are the variables that the controller can manipulate to affect the states. In the elbow joint example, this might be a muscle that produces a torque about the joint, or a flexor-extensor pair that together act as a single combined actuator to generate both positive and negative torque.

The next step is to describe the system dynamics (i.e., the physics of the system) using an equation that models how the state and actuation variables relate to each other. This model might be derived from first principles (e.g., Newtonian mechanics, or "force equals mass times acceleration") or from experimental data. Because of the time-varying nature of dynamical systems, this equation is generally some form of a differential equation,

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

where $\mathbf{x}$ is a vector of states and $\mathbf{u}$ is a vector of actuations. We use dot notation to denote time derivatives in a compact way, and vectors and matrices are bold. One dot above a variable is the first derivative, and two dots indicate a second derivative, e.g.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad \dot{\mathbf{x}} = \frac{d\mathbf{x}}{dt}, \tag{4.2}$$

where x_1 and x_2 are two different state variables. A "differential equation" is simply an equation in which the derivative of a variable is a function of that variable. Some differential equations can be solved by hand, but most cannot and are instead solved using one of a variety of numerical methods available in computation packages (and what we will use here).

Linear State space

The function $f(\mathbf{x}, \mathbf{u})$ is also a vector function, which could be written out as a system of equations, with the same number of equations as the number of states. In linear control theory, the system dynamics are expressed in the linear "state space", which is the set of equations that describe the input and output dynamics of the system. The vectors and matrices in these linear sets of equations correspond directly to the parts of a block diagram (Fig 4.3).

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}, \tag{4.3a}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u}. \tag{4.3b}$$

If we consider the bicep joint angle example, $\mathbf{x}$ is a set of states (joint angle and joint velocity), while $\dot{\mathbf{x}}$ is the derivatives of those states (joint velocity and joint acceleration). And, $\mathbf{u}$ is the vector of actuators, or inputs to the system. The matrices $\mathbf{A}$ and $\mathbf{B}$ are then derived from first principles or data to define how the system and actuation variables interact. This first equation is the "input equation". A worked example is provided in the next section for the classic spring-mass-damper system, and another example is provided in a later section for the grounded, stance phase of the spring-loaded inverted pendulum (SLIP) model that is often used to model terrestrial locomotion.

The second equation is the "output equation", where $\mathbf{y}$ is the vector of output variables, and $\mathbf{x}$ and $\mathbf{u}$ are the same state and input vectors as in the first equation. The matrix $\mathbf{C}$ describes which of the states are returned as outputs, and the matrix $\mathbf{D}$ is the "feedthrough" matrix which defines any control inputs that are directly "fed-through" to the output, and the associated dynamics. The $\mathbf{D}$ matrix is often, but not always, zero. This output equation will be discussed in more detail in later sections that cover *sensors* and *observers*, which allow for choosing outputs and modeling sensory systems.

4.2.3 Control and actuation

How a controller influences the system dynamics is called a *control law.* Control laws are based on the current states of the systems (or measurements of those states) and the target–or reference–states. The controller determines which actuators and how much actuation to use at a given time. The difference between the current state $\mathbf{x}(t)$ and the reference state $ref(t)$ is the system error.

$$\mathbf{e}(t) = \mathrm{ref}(t) - \mathbf{x}(t). \tag{4.4}$$

Feedback controllers calculate this error and apply the control law based on it, and are thus able to compensate for error in the system due to unexpected disturbances, previous actuation that led the system off track, or an updated reference. A block diagram of a feedback control system is often a convenient way to diagram the mathematics of a reference, error calculated from feedback, and the control law (Fig. 4.3b).

The control law is a function of this error. One simple kind of control law is proportional control,

$$\mathbf{u}(t) = \mathbf{pe}, \tag{4.5}$$

where $\mathbf{p}$ is a matrix of constants (or gains) by which the error is multiplied in an effort to drive the system to the desired state. Larger control gains generally mean that the system responds faster to errors, for example an animal making a sharp turn (large gain) to regain a desired heading, instead of a gradual return to that heading. Larger proportional control gains *converge* to the reference (desired) state faster, but are also associated with larger overshoots (increasing the error, before converging to the reference state) and require higher maximum control forces. In some cases, overshooting the reference value might be acceptable, such as turning to regain a flight heading. But in other tasks, overshooting might result in actions that are hard to recover from such as tipping over while attempting to balance on one leg. Determining the control gains of an animal during a task is one way to investigate animal behavior and abilities.

Proportional control is demonstrated with a worked example in the next section. Another simple control law is an on/off control system, which turns on if the states rise or fall above a set threshold. An example of this is a standard thermostat, which turns on only when the temperature goes outside a certain range. Another example is shivering when body temperature decreases too much.

4.2.4 Background summary

- Using control theory, animal movement can be modeled as a closed-loop feedback control system where the *plant* consists of the body dynamics, the *actuators* are the muscles, the *sensors* are the neural sensors, and the *controller* is the central nervous system.

- The first steps in formulating a control system model are to describe the state variables (e.g., joint positions and velocities) and the system dynamics (e.g., the equations of motion).

- A *control law* dictates how an actuator responds to a change in the state of the system.

4.2.5 Further Reading

This is a classic textbook on state-space feedback control:

Karl J Astrom and Richard M Murray. Feedback systems : an introduction for scientists and engineers, 2008

This book has several systems biology examples of linearizing a system and building state-space equations.

Carlo Cosentino and Declan Bates. *Feedback Control in Systems Biology.* CRC Press, Baton Rouge, 2011

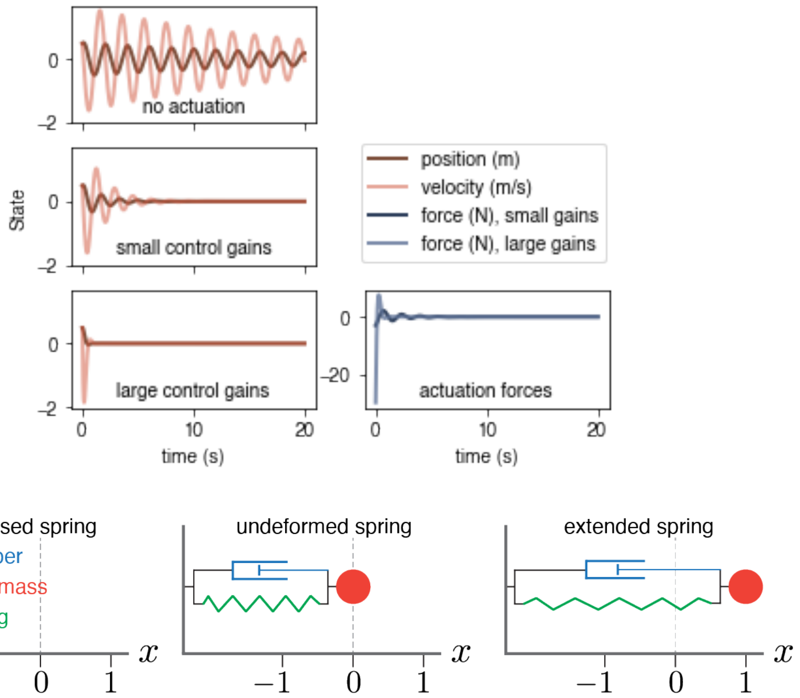


Figure 4.4: Top: Mass-spring-damper system response with no actuation (free response) and with proportional control. When proportional gains are larger, the system settles at the reference value faster, but requires larger actuation forces. Bottom: Negative $\mathbf{x}$ indicates a compressed spring, positive $\mathbf{x}$ indicates a stretched spring.

4.2.6 Worked Example - Proportional Feedback Control of a mass-spring-damper

Mass Spring Damper

In this section, we demonstrate feedback control using a mass on a spring with damping. The mass-spring-damper is a building block for many biomechanical models [2, 8, 13], including the spring-loaded inverted pendulum (SLIP) model that is used as an example in the next section. The spring constant and the damping often represent the elastic properties of tendons and muscles, while the mass represents total mass of the body or a limb. The mass-spring-damper is a useful starting point for understanding feedback control, because it is a linear model, with no sine or cosine terms associated with joint rotation. In later sections, we will show how to linearize models with rotation (or other nonlinear functions), and some more advanced techniques for nonlinear control.

In the absence of any forcing beyond the initial displacement, the length of the spring will continue to oscillate for many cycles if it starts from a stretched or compressed position. With an actuator that can push and pull on the mass, the system can be brought to rest more quickly. A diagram of the system is shown in Fig. 4.4.

The differential equation that describes the motion of the mass can be derived from Newton's second law. The spring force, damping force, and actuation force must add up to the mass times the acceleration;

$$\underbrace{-kx}_{\text{spring}} \overbrace{-b\dot{x}}^{\text{damping}} \underbrace{+F}_{\text{actuation}} = m\ddot{x}, \tag{4.6}$$

where x is the change in length of the spring from an undeformed position, and F is a force aligned with the spring. With dot notation indicating time derivatives, $\dot{x}$ is velocity, and $\ddot{x}$ is acceleration. In this case, the actuation force F takes the place of the more general u variable described in the previous section. The constants are the spring constant, (in this case we'll set $k = 10\text{kg}\,\text{s}^{-2}$) the damping coefficient ($b = 0.1\text{kg}\,\text{s}^{-1}$) and the mass ($m = 1\text{kg}$). This equation can be rearranged into the linear state-space form introduced in Equation 4.3:

$$\begin{bmatrix} \dot{x} \\ \ddot{x} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -k/m & -b/m \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \end{bmatrix} + \begin{bmatrix} 0 \\ 1/m \end{bmatrix} F. \tag{4.7}$$

In this example, the system starts with a stretched spring and a small velocity, and the goal is to bring the mass to rest at the undeformed spring position using a controller with a proportional control law. The initial condition $\mathbf{x}_0$ and reference $ref(t)$ are:

$$\mathbf{x}_0 = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}, \qquad \text{ref}(t) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{4.8}$$

Effect of control gains

In Fig. 4.4, we can see the effect of different gains. First we look at the *free response*, where there is no actuation. The mass bounces back and forth through many cycles of the spring stretching and compressing, and will gradually come to rest because energy is being dissipated by the damper. (You might recall from physics courses that damping is a dissipative force. That is, energy is lost in the action of the damper. Friction is another dissipative force, while an ideal spring is not–it is conservative. If this were just a mass on a spring, it would oscillate forever, in a clearly unrealistic simulation.) With our first set of control gains, the mass is brought to rest after a few cycles. The proportional control law we chose through simply a "trial and error" method follows the form in Eq. 4.5, with proportional weights of

$$\mathbf{p} = \begin{bmatrix} 5 & 1 \end{bmatrix}. \tag{4.9}$$

This means that the controller responds to errors in the position with five times more force than is used for errors in velocity. This system only has a single actuator, so **p** has a single row, with one column per state. If bouncing the mass back and forth a few times is undesirable, we can get more aggressive with the control and increase the controller gains by a factor of ten. With these larger gains, the mass has one small wobble past the equilibrium point, and then comes to rest. Despite smaller errors, there is a downside to the larger controller gains; the larger gains result in a maximum actuation force that is ten times larger.

For more complicated models or experimental setups, a more principled way of finding control weights is often necessary, rather than our "trial-and-error" approach. In the next section, we show a more systematic way of finding control weights, with considerations for the costs of actuation. Later we will discuss a nonlinear method (model predictive control) that can place limits on the maximum actuation forces.

4.2.7 Recovering from external disturbances

Closed-loop feedback control is necessary for two main reasons: correcting for incorrect or simplified plant models, and correcting for external disturbances. One example of a disturbance would be a control loop where the plant model is for walking on flat ground, but the true walking surface is rough, slippery, or has hills. Another example is for a flying animal, a gust of wind could push the animal off course or change its heading. We can represent these disturbances with a vector **d** which has one entry per state, and we can include the effects of the disturbances in the system dynamics with a small modification of Equation 4.3:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{d}. \tag{4.10}$$

The error will include the effects of disturbances, and so the controller will compensate for the disturbances to steer the system back towards the reference.

4.3 Linear Optimal Control and Estimation

Historically, control theorists and practitioners have focused on linear systems. This is because linear systems are much easier to deal with, but also because linear systems can be very effective approximations of the inherently nonlinear natural world. That is, linear models are very useful, as they can quite accurately model nonlinear systems if the questions asked of them are thoughtfully constrained. The strict definition of a linear system is beyond the scope of this primer, but essentially, the output of a linear system is proportional to its input (as in the linear state space we introduced in the previous sections). You can recognize nonlinearity in a system if it is modeled with nonlinear functions (e.g., trigonometric). Even if your system of interest cannot be modeled well using a linear system, linear control methods are still worth learning, as they form the foundation for many nonlinear methods.

In this section, we will cover "linear optimal control" methods for designing controllers, which are useful for considering energy trade-offs and are especially helpful when sensors in the feedback loop are noisy. We note that there are many other ways to design controllers (or to model control systems) such as the entire field of frequency domain techniques, which are also beyond the scope of this primer.

4.3.1 Balancing accuracy with effort: Linear quadratic regulator

In both engineered and biological systems, there are costs associated with generating forces, like energy use and food or fuel consumption. In the first example, we chose control gains with a focus on reaching the reference state quickly, without considering the actuation costs associated with applying large forces. We can, however, choose optimal controller gains to balance accuracy and actuation costs using a *linear quadratic regulator* (LQR) formulation.

LQR controller gains are designed by specifying matrices in *cost function.* A cost function is any equation used for optimization; the optimization problem is to minimize this cost. In this case, the cost function is used to combine matrices of accuracy weights for each state and costs for each actuator into a single vector of controller gains. The square matrix $\mathbf{Q}_x$ contains cost function weights associated with accuracy, and has as many rows and columns as there are states. It "tells" the cost function how important it is to achieve the intended states. The cost function weights associated with actuator cost go into another square matrix, $\mathbf{Q}_u$, which has as many rows and columns as there are actuators. This can be thought of as fuel or energy cost requirements. So, essentially the cost function can be thought of as balancing these requirements of accuracy and cost. There are some technical requirements on the properties of the $\mathbf{Q}_x$ and $\mathbf{Q}_u$ matrices. One way to be sure to meet these requirements is to use positive numbers on the diagonal elements and zeros elsewhere. Each diagonal element corresponds to a single state or single actuator. More details of how this optimization is done are in Section 4.5.2. The minimization of the cost function results in optimal proportional controller gains in the same form as Equation 4.5. There is an analytical solution to minimizing the cost function, which can be used to find the controller gains, but most people do this calculation using existing software packages.

4.3.2 Worked Example: LQR control of a spring-loaded inverted pendulum for single leg balance

This section demonstrates the use of LQR to determine the gains for a proportional controller for a spring-loaded inverted pendulum (SLIP) model. The SLIP is a relatively simple system commonly used to model legged locomotion, including walking, running, and upright balance. This inverted pendulum model consists of a point mass, m, that models the mass of the body, and a massless spring of undeformed length l and linear spring constant k. The spring can be thought of as a simplified combined knee and ankle joint; changes in length represent joint flexion and extension. The spring stores elastic potential energy like the connective tissues, and, in our model, the muscles are modeled by a force can be applied to lengthen or shorten the spring. The classic SLIP model has two phases: a grounded phase and a ballistic phase, where the SLIP leaves the ground and the mass enters a parabolic trajectory. Upon landing, the SLIP enters the grounded phase, where the spring contracts, the pendulum rocks forward, and the spring extends again into the next launch. Thus the combination of these two phases can model bouncing-type trajectories, like running. Variations of this basic SLIP model have been used to model different aspects of legged locomotion [3, 4, 11, 15].

In this example, we limit the model to the grounded phase, with a goal of balancing in an upright position. Biological analogs might be an animal balancing on one leg, or a bird balancing on a perch. This leaves two directions in which the system is free to move: the angle θ around the foot-pivot point can change, and the leg-spring can be compressed or stretched along its length. Here, the undeformed length of the leg-spring is a constant, l, and changes in that length are denoted by r.

These two degrees of freedom, θ and r, and their associated velocities, $\dot{\theta}$ and $\dot{r}$, make up the state vector, $\mathbf{x}$.

This will be a fully actuated system, so there are two actuators (one for each degree of freedom): a torque τ about the foot-pivot, and a force F along the direction of the leg-spring. The derivation of the nonlinear dynamics and the subsequent linearization about the upright equilibrium position are shown in the Section 4.5.1.

In state-space form, the linearized dynamics of the system are

$$\begin{bmatrix} \dot{\theta} \\ \dot{r} \\ \ddot{\theta} \\ \ddot{r} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{g}{l-mg/k} & 0 & 0 & 0 \\ 0 & -\frac{k}{m} & 0 & 0 \end{bmatrix} \begin{bmatrix} \theta \\ r \\ \dot{\theta} \\ \dot{r} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \frac{1}{(l+mg/k)^2} & 0 \\ 0 & \frac{1}{m} \end{bmatrix} \begin{bmatrix} \tau \\ F \end{bmatrix}. \tag{4.11}$$

This linearized model is a good representation of the system near the upright equilibrium position, but it will be increasingly inaccurate further from this position, especially at further angles. Small model inaccuracies are handled easily by a well-designed feedback controller; however, if the leg-spring-mass system leans too far from upright (e.g., due to a disturbance or overshoot), a controller designed using this linearized model may not be able to regain the equilibrium state due to increasing model inaccuracy.

Next, for each state we specify how much we value fast convergence to the reference value (accuracy). While the optimization cost matrices can have cross-terms in the off-diagonal positions, for simplicity we'll choose to only use diagonal elements, which each correspond to a single state. We'll choose accuracy weights of

$$\mathbf{Q}_x = \begin{bmatrix} 1000 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{4.12}$$

where the top left entry of 1000 corresponds with the first state θ, the second row diagonal corresponds with the second state r, and the next two rows with their velocities. Note that we vary the weights by factors of ten, to clearly prioritize θ and $\dot{\theta}$. We prioritize these variables because we are choosing our control gains based on a linearized model, and if the angle θ gets too far from the location we linearized about, the controller won't work well, or may even fail to stabilize the system in an upright position. Unlike the gravitational force that swings the leg about the pivot, the spring force in the leg is linear. So long as θ is maintained near the linearization point, controlling the

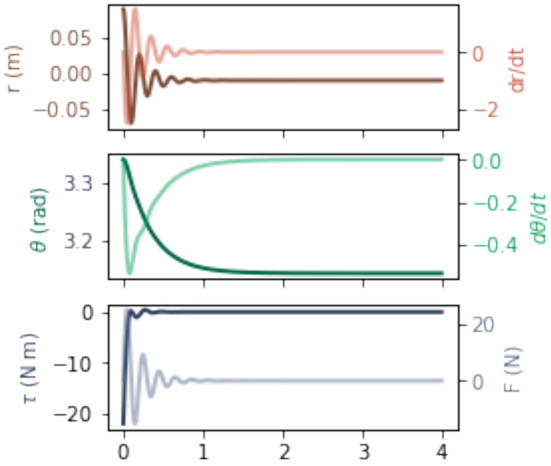


Figure 4.5: LQR control balancing a spring-loaded pendulum model of a leg and body in an upright position. The darker colors are for leg-spring length, angle, and actuation torque. The lighter colors, on the right axis scales, are leg-spring length velocity, angular velocity, and actuation force.

leg-spring length using gains derived using a linearized model is less difficult. One important part of control theory is studying these types of model sensitivities, which is beyond the scope of this tutorial. The most straightforward way to gain intuition into model sensistivity is to simply do a lot of simulation experiments with different weights.

Similar to the accuracy cost matrix, for the actuator cost matrix we use only diagonal elements, such that each term only affects one of the actuators:

$$\mathbf{Q}_u = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}. \tag{4.13}$$

The top left term is the weight associated with the first actuator, the torque τ about the foot-pivot. The bottom right term is the cost weight for the force along the leg-spring F. In this example, we choose to put a lower priority, and so a lower cost weight, on the actuation than was used in the accuracy cost weights. For a specific organism, we might want to run a few numerical simulations with control to check that the actuator forces and torques remain within reasonable limits for the organism of interest.

Using the cost matrices above and the LQR optimization (described in the Appendix), we find optimal proportional control weights of

$$\mathbf{p} = \begin{bmatrix} 110 & 0 & 35 & 0 \\ 0 & 0.05 & 0 & 10 \end{bmatrix}. \tag{4.14}$$

The first row corresponds to the weights for torque, and in with these weights, only errors in θ and $\dot{\theta}$ will result in nonzero torque. Similarly, the second row corresponds to motor force, and only errors in the leg-spring length and leg-spring velocity will result in motor force.

We test these weights by using them to control the nonlinear system; recall that the control weights are determined using the linearized state-space system, but ultimately it is the nonlinear system that we are interested in controlling. We can see how the system responds using initial conditions that are a small angle from upright, and a slightly stretched leg-spring:

$$\mathbf{x}_0 = \begin{bmatrix} \theta \\ r \\ \dot{\theta} \\ \dot{r} \end{bmatrix} = \begin{bmatrix} \pi + 0.2 \\ -mg/k + 0.1 \\ 0 \\ 0 \end{bmatrix}. \tag{4.15}$$

We used a constant reference value,

$$\mathrm{ref}(t) = \begin{bmatrix} \pi \\ -mg/k \\ 0 \\ 0 \end{bmatrix}, \tag{4.16}$$

which means that we are trying to stabilize the system in the upright position with the leg-spring slightly shortened due to gravity, which is an unstable equilibrium position. With these optimized control gains and initial conditions, the system stabilizes to the reference value within one second, which is shown in Figure 4.5.

4.3.3 Incorporating noisy measurements and disturbances: Linear quadratic gaussian

We need a way to handle for noisy sensory measurements, or noisy experimental measurements, to avoid actuating based on incorrect error estimates. To do this, we can use an *estimator*, which estimates the state $\mathbf{x}$ from the noisy measurements $\mathbf{y}$. One of the most common estimators is the *Kalman filter*. The Kalman filter is used in many signal processing contexts to filter noise, but don't let the word "filter" confuse you. The Kalman filter is a *linear quadratic estimator*, so it would perhaps be more fitting in a linear control context for us to call it the Kalman estimator. Using

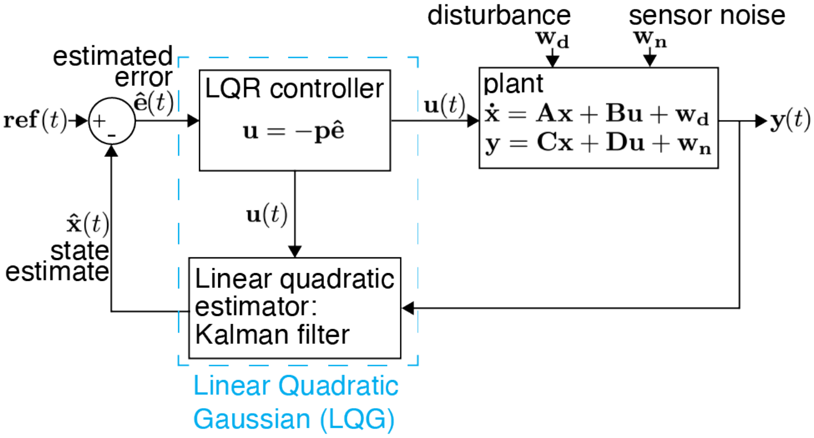


Figure 4.6: The Linear Quadratic Gaussian (LQG) consists of a linear quadratic regulator (LQR) controller design with a linear quadratic estimator (a Kalman filter) in the feedback loop. The estimator takes the output from the plant—which may be computed from a noisy plant or contain incomplete state information—to produce an estimate of full state feedback. This state estimate is then used to compute an estimated error, and this estimated error is what is used to compute the actuation.

a similar optimization framework as finding the control gains with LQR, we can find a vector of Kalman filter gains $\mathbf{K}_f$ which are an estimate of how much to trust each measurement in $\mathbf{y}$. This gain is used to estimate the true state $\mathbf{x}$ from the measurements $\mathbf{y}$ in combination with the expected state of the system, given the system dynamics and control actuation. Because external disturbances can also cause the current state to be different than the expected state based on the dynamics, the Kalman filter gains also take into account some information about the expected disturbances.

For a system with $\mathbf{D} = 0$, the linearized system dynamics with disturbances and noise are

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{d}, \tag{4.17a}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{n}, \tag{4.17b}$$

where $\mathbf{n}$ is the noise, and $\mathbf{d}$ is the disturbance, both of which will vary in time. Because each sensory or sensor channel may have a different amount of noise, $\mathbf{n}$ is a vector with the same dimensions as $\mathbf{y}$.

The Kalman filter uses a modified version of the system dynamics to estimate the current state:

$$\frac{d}{dt}\hat{\mathbf{x}} = \mathbf{A}\hat{\mathbf{x}} + \mathbf{B}\mathbf{u} + \mathbf{K}_f(\mathbf{y} - \hat{\mathbf{y}}), \tag{4.18a}$$

$$\hat{\mathbf{y}} = \mathbf{C}\hat{\mathbf{x}}. \tag{4.18b}$$

We use hat notation for estimates, so $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are the estimated (or predicted) values. The $\mathbf{y} - \hat{\mathbf{y}}$ factor is how much different the actual sensor measurements are from the expected measurements. The Kalman gains $\mathbf{K}_f$ then weights this sensory estimate error; in this way the Kalman gains are a measure of how much to trust the modeled dynamics versus how much to trust the measurements. Equation 4.18 can be used with a computational differential equation solver (integrator) to solve for the estimated state, $\hat{\mathbf{x}}$, in much the same way as a state-space system can be simulated. In experimental settings, this would often be done with a discrete-time state space system, which is not covered in this primer.

LQR control gains and a Kalman filter for estimation are often combined, and together they are called the *Linear Quadratic Gaussian* (LQG).

4.3.4 Partial-state feedback

In the examples we've shown so far, the systems have had *full-state feedback*, where $\mathbf{y} = \mathbf{x}$ in Equation 4.3. This means that the $\mathbf{C}$ matrix is the identity matrix, with ones on the diagonal and zeros elsewhere, and that the $\mathbf{D}$ matrix is composed of zeros.

It is often the case that not every state is directly measured or sensed; instead, only some states are sensed or some linear combination of states are sensed. In an experiment, there may be an effort to limit the use of some sensory pathways, for example by using a dark environment or optogenetically knocking out some sensory pathways.

4.3.5 Linear Optimal Control summary

- Although most systems—especially biological systems—are inherently nonlinear, nonlinear systems systems can often be described quite accurately over small ranges using linear approximations. Consequently, controllers designed using linear approximations are often adequate for controlling even the full nonlinear system.
- Even if a system or experiment ultimately is not well described with a linear model, it is worth learning a bit about linear control as a stepping stone to using nonlinear control.
- The *linear quadratic regulator* computes a controller gain by optimizing according to a cost function that balances controller accuracy with actuator cost (e.g., energy requirements).
- The *linear quadratic gaussian* is a linear quadratic regulator combined with a linear estimator. This is method for calculating controller gains with noisy or incomplete sensory feedback. The estimator—generally a Kalman filter—estimates the full state feedback based on the incomplete information.

4.3.6 Further Reading

A classic text on Optimal Control and Estimation is the homonymous book by Stengel, which covers many of the topics in this primer with full mathematical explanations that do not assume a high level of prior knowledge of linear algebra.

Robert F Stengel. *Optimal control and estimation.* Dover books on mathematics. Dover Publications, New York, 1994

4.4 Planning ahead and adding constraints: Model Predictive Control

In its most basic form, model predictive control (MPC) is a straightforward extension of any other model-based control, such as the linear control strategies described in previous sections. The predicitve aspect of MPC is that control actions are planned based on future expected states. That is, it is the control actuation and future trajectory that is being predicted. The model itself is not being predicted, except in advanced forms of MPC with "learning in the loop".

This simple predictive extension results in significant flexibility in the forms of the models, and can perform well even with models that are not good representations of the real physics. Both linear and nonlinear models can be used with MPC.

In Sections 4.3.1 and 4.3.3, we used optimization to choose proportional control gains, and to estimate the current state when there are noisy measurements. With LQR and LQG, we are limited to a particular form of the cost function. In contrast, with MPC there is considerable flexibility with the cost function that is used to optimize and plan the actuation. For example, the cost function can be formulated to capture the energetic costs in a biologically relevant way. We can include constraints such as limiting a joint angle such as a knee or elbow to a realistic range of motion, for example 20° to 180°. We could also model joint actuation with both a flexor and an extensor muscle, constraining the flexor to negative torque and the extensor to positive torque.

A drawback of MPC is an increase in computation time during control planning. While MPC can be used with nonlinear models, control for linear models with few states will generally be faster optimize. The prediction and control horizons are parameters that determine how many time steps in the future the system output and control inputs are calculated, respectively. The control horizon, $T_c = m_c \Delta t$, is generally less than or equal to the prediction horizon, $T_p = m_p \Delta t$. Model predictive control uses optimization to determine what actuation to use, and so a cost function used to find the optimal control sequence, $\mathbf{u}(\mathbf{x}_j) = \{\mathbf{u}_{j+1}, ..., \mathbf{u}_{j+m_c}\}$, where the subscript indicates the discrete time step. An example of an MPC cost function is

$$J = \sum_{k=0}^{m_p-1} \|\mathbf{e}_k\|^2_{\mathbf{Q}} + \sum_{k=1}^{m_c-1} (\|\mathbf{u}_{j+k}\|^2_{\mathbf{R}_u} + \|\Delta\mathbf{u}_{j+k}\|^2_{\mathbf{R}_{\Delta u}}). \tag{4.19}$$

Similar to LQR and LQG, the weighting matrices $\mathbf{Q}$, $\mathbf{R}_u$, and $\mathbf{R}_{\Delta u}$, are used to prioritize the outputs, inputs, and rate of change of the inputs, respectively.

4.5 Control Primer Appendix

4.5.1 Derivation of SLIP model dynamics

Here we derive the equations of motion (EOMs) of the spring-loaded inverted pendulum model that is used throughout the manuscript to explain controller and observer techniques.

The Spring-loaded inverted pendulum (SLIP) model in its grounded phase is a combination of a mass-on-a-spring and a pendulum. The planar SLIP has two degrees of freedom: rotation about the base due to pendulum action and radial extension/contraction due to spring action. In the

Euler-Lagrange formulation of equations of motion, these degrees of freedom can be taken as the generalized coordinates. In this case the difference between the spring extension/contraction and the resting length of the spring is the radial generalized coordinate).

The generalized coordinates:

$$q = \begin{bmatrix} r \\ \theta \end{bmatrix} \tag{4.20}$$

The equations of motion for a fully actuated system (applied torque to control the angular coordinate and applied force to control the linear, radial coordinate):

$$m(l+r)^2\ddot{\theta} + 2m\dot{r}(l+r)\dot{\theta} + mg(l+r)\sin(\theta) = \tau \tag{4.21}$$

$$m\ddot{r} - m(l+r)\dot{\theta}^2 - mg\cos(\theta) + kr = F \tag{4.22}$$

The equations of motion in a nonlinear, dynamical systems format, where x is the vector of states and u is the control input:

$$x = \begin{bmatrix} \theta \\ r \\ \dot{\theta} \\ \dot{r} \end{bmatrix} \tag{4.23}$$

$$\dot{x} = f(x,u) \tag{4.24}$$

$$\dot{x} = \begin{bmatrix} \dot{\theta} \\ \dot{r} \\ \ddot{\theta} \\ \ddot{r} \end{bmatrix} = \begin{bmatrix} \dot{\theta} \\ \dot{r} \\ \frac{-2\dot{r}}{(l+r)}\dot{\theta} - \frac{g}{(l+r)}\sin(\theta) + \frac{1}{m(l+r)^2}\tau \\ (l+r)\dot{\theta}^2 + g\cos(\theta) - \frac{k}{m}r + \frac{1}{m}F \end{bmatrix} \tag{4.25}$$

In the “up” position, both the force applied by the spring and gravitational acceleration are towards the ground, so

$$r = -\frac{mg}{k} \tag{4.26}$$

Thus, the equilibrium values of the states and control input are:

$$x^* = \begin{bmatrix} 0 \; or \; \pi \\ \pm\frac{mg}{k} \\ 0 \\ 0 \end{bmatrix} \tag{4.27}$$

$$u^* = 0 \tag{4.28}$$

And the resulting linearized equations of motion in state-space form (linearized about the unstable equilibrium):

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \tag{4.29}$$

$$\mathbf{A_{up}} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{g}{l-\frac{mg}{k}} & 0 & 0 & 0 \\ 0 & -\frac{k}{m} & 0 & 0 \end{bmatrix} \tag{4.30}$$

$$\mathbf{B} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \frac{1}{(l+\frac{mg}{k})^2} & 0 \\ 0 & \frac{1}{m} \end{bmatrix} \tag{4.31}$$

4.5.2 Optimization and cost functions

LQR uses a quadratic cost function, which is a form that we also use for handling noisy measurements with Linear Quadratic Gaussian, and for optimizing control while including constraints with Model Predictive Control:

$$J = \int_0^\infty (\mathbf{x}^T \mathbf{Q}_x \mathbf{x} + \mathbf{u}^T \mathbf{Q}_u \mathbf{u})\, dx. \tag{4.32}$$

Optimizing the proportional gains for the LQR cost function involves numerically solving the below equation, called the algebraic Riccati equation, for $\mathbf{P}$:

$$\mathbf{P}\mathbf{A} + \mathbf{A}^T\mathbf{P} - \mathbf{P}\mathbf{B}\mathbf{Q}_u^{-1}\mathbf{B}^T\mathbf{P} + \mathbf{Q}_x = 0, \tag{4.33}$$

where an exponent of -1 indicates the matrix inverse, and an exponent of T indicates a matrix transpose. The proportional weights are then calculated from this solution;

$$\mathbf{p} = \mathbf{Q}_u^{-1}\mathbf{B}^T\mathbf{P}. \tag{4.34}$$

Contributors

Other contributors to this work are Kathryn Stanchak, David Perkel, Thomas Daniel, Steven Brunton, and Bingni Brunton. Steven Brunton and Bingni Brunton supervised the work. The text and figures were written by both Michelle Hickner and Kathryn Stanchak: the bulk of the introduction and diagrams were done by Kathryn Stanchak, and the bulk of the examples and plots were done by Michelle Hickner. Michelle Hickner wrote the python code for the examples.

The work in this chapter was supported by funding from the Air Force Office of Scientific Research (AFOSR FA9550-19-1-0386).

Bibliography

[1] Karl J Astrom and Richard M Murray. Feedback systems : an introduction for scientists and engineers, 2008.

[2] Andrew A. Biewener and Monica A. Daley. Unsteady locomotion: Integrating muscle function with whole body dynamics and neuromuscular control. *Journal of Experimental Biology*, 210(17):2949–2960, 2007.

[3] Giovanni A. Cavagna, Norman C. Heglund, and C. R. Taylor. Mechanical work in terrestrial locomotion: two basic mechanisms for minimizing energy expenditure. *American Journal of Physiology - Regulatory Integrative and Comparative Physiology*, 233(5):R243–R261, 1977.

[4] Chanwoo Chun, Tirthabir Biswas, and Vikas Bhandawat. Drosophila uses a tripod gait across all walking speeds, and the geometry of the tripod is important for speed control. *eLife*, 10:1–47, feb 2021.

[5] Steve Collins, Andy Ruina, Russ Tedrake, and Martijn Wisse. Efficient bipedal robots based on passive-dynamic walkers. *Science*, 307(5712):1082–1085, 2005.

[6] Carlo Cosentino and Declan Bates. *Feedback Control in Systems Biology.* CRC Press, Baton Rouge, 2011.

[7] Noah J Cowan, Mert M Ankarali, Jonathan P Dyhr, Manu S Madhav, Eatai Roth, Shahin Sefati, Simon Sponberg, Sarah A Stamper, Eric S Fortune, and Thomas L Daniel. Feedback control as a framework for understanding tradeoffs in biology. *American Zoologist*, 54(2):223–237, 2014.

[8] Michael H. Dickinson, Claire T. Farley, Robert J. Full, M. A.R. Koehl, Rodger Kram, and Steven Lehman. How animals move: An integrative view. *Science*, 288(5463):100–106, apr 2000.

[9] K. A. Mazurek, B. J. Holinski, D. G. Everaert, R. B. Stein, R. Etienne-Cummings, and V. K. Mushahwar. Feed forward and feedback control for over-ground locomotion in anaesthetized

cats. *Journal of Neural Engineering*, 9(2), 2012.

[10] Tad McGeer. Passive Dynamic Walking. *The International Journal of Robotics Research*, 9(2):62–82, 1990.

[11] Alberto E. Minetti. The biomechanics of skipping gaits: a third locomotion paradigm? *Proceedings of the Royal Society B: Biological Sciences*, 265:1227–1235, 1998.

[12] Jean Michel Mongeau, Simon N. Sponberg, John P. Miller, and Robert J. Full. Sensory processing within cockroach antenna enables rapid implementation of feedback control for high-speed running maneuvers. *Journal of Experimental Biology*, 218(15):2344–2354, aug 2015.

[13] D A Robinson. The Mechanics of Human Saccadic Eye Movement. *Journal of Physiology*, 174:245–264, 1964.

[14] E Roth, S Sponberg, and N J Cowan. A comparative approach to closed-loop computation, 2014.

[15] Manoj Srinivasan and Andy Ruina. Computer optimization of a minimal biped model discovers walking and running. *Nature*, 439(7072):72–75, 2006.

[16] Robert F Stengel. *Optimal control and estimation*. Dover books on mathematics. Dover Publications, New York, 1994.

[17] Emanuel Todorov and Michael I. Jordan. Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, 5(11):1226–1235, 2002.