



Probabilistic error bounds on constraint violation for empirical-analytical Lagrangian models of motion

J. Guo · S. Dadashi · M. Bender · S. T. Paruchuri · N. Powell · Y. Sharma · H. Kurdila · J. W. McGlothlin · A. J. Kurdila

Received: 17 December 2018 / Accepted: 2 August 2019 / Published online: 23 August 2019
© Springer Nature B.V. 2019

Abstract In contrast to many systems studied in the field of classical mechanics, models of animal motion are often distinguished in that they are both highly uncertain and evolve in a high-dimensional configuration space Q . Often it is either suspected or known that a particular motion regime evolves on or near some smaller subset $Q_0 \subseteq Q$. In some cases, Q_0 may itself be a submanifold of Q . A general strategy is presented in this paper for constructing empirical-analytical Lagrangian (EAL) models of the mechanics of such systems. It is assumed that the set $Q_0 \subseteq Q$ is defined by a collection of unknown holonomic constraints on the full configuration space. Since the analytic form of the holonomic constraints is unknown, EAL models are defined that use experimental observations $\{z_1, \dots, z_N\} \subseteq Q^N$ to ensure that the approximate system models evolve near the underlying submanifold Q_0 . This paper gives a precise characteriza-

tion of a probabilistic measure of the distance from the EAL model to the underlying submanifold.

Keywords Lagrangian system · Empirical-analytical model · Biomotion

1 Introduction

For a large class of example problems of animal biomotion models, it is often true that the governing equations of Lagrangian mechanics define evolutions in a high-dimensional configuration manifold. For the sake of argument, suppose that we are only interested in animal motion models that take the form of multibody systems comprised of a rigid skeleton connected by ideal joints. These systems are now a classical topic in standard texts on multibody dynamics or robotics [1, 2]. Our interest in such models has arisen since they are critical in constructing Bayesian filters to predict the motion state from the current state when measurements of animal motion are made via motion capture [3–5].

Motion capture is a popular technique used in biomotion studies to quantify the often complex kinematics of animal locomotion. One study [6] seeks to determine the kinematics of the cheetah spine and tail using both GPS/IMU tracking collars along with rear facing cameras, which helps to build an understanding of the whole-body motion. The authors of [7] examine a gecko body and limb kinematics parameterized

J. Guo (✉) · S. Dadashi · M. Bender · S. T. Paruchuri · N. Powell · Y. Sharma · A. J. Kurdila
Department of Mechanical Engineering, Virginia Tech,
Blacksburg, VA 24060, USA
e-mail: jguo18@vt.edu

H. Kurdila
Department of Mechanical Engineering, New York
University, New York 11205, USA

J. W. McGlothlin
Department of Biological Sciences, Virginia Tech,
Blacksburg,
VA 24060, USA

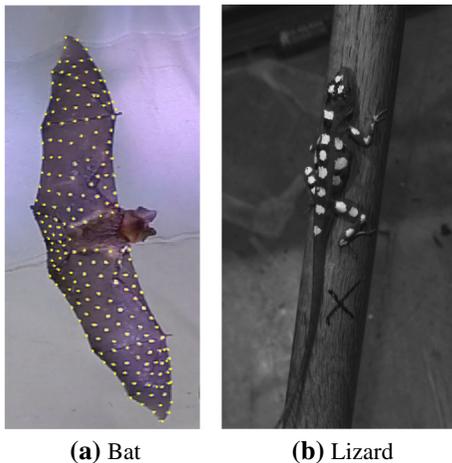


Fig. 1 Motion capture experiments for construction of (left) bat flight and (right) lizard motion models

by the slope of the surface on which a gecko strides. The researchers in [8] use motion capture to examine the kinematics of horse appendages to determine the effects of different surface conditions on a horse's stride. Another study [9] examines the effects of an orthosis on the three-dimensional limb kinematics of dogs for differing types of gaits. These experiments track physical landmarks on the dog limbs as well as markers placed on the orthosis itself. The researchers in [10] use motion capture to determine the kinematics of frog joints during locomotion after freezing in order to study frog muscle activation patterns post-freezing.

Despite these impressive studies, among animal motion researchers it is known that some predictive models can be quite poor [3]. This fact is noted in [11] for human walking models, and it is dealt with systematically in [3–5] in studies of bat flapping flight.

Animal motion models can contain substantial uncertainties. Specifically, the observations from motion capture contain no information about the forces, moments, or parameters that determine the dynamics. This requires researchers often to make very ideal assumptions and can induce significant uncertainties. Here a few examples can illustrate how this substantial uncertainty arises. Figures 1, 3, and 6 depict the results of motion studies by the authors for a bat in flapping flight and a lizard walking along a rod. For either of the examples shown in Figs. 3 and 6, the uncertainties can include: geometric, inertial, and mass properties of the skeleton; energy loss mechanisms in connective tissues or joints; or suitable force-

displacement models of muscle mechanics. Perhaps most important to this paper is the fact that joint kinematics in biomotion models are far from the ensemble of classical ideal joints such as revolute, prismatic, spherical, universal, etc. Animal joints have limited ranges of motion, as well as elastic or inelastic properties, for example. The exploration of the underlying physical mechanics furthermore is restricted by the lack of feasible *in vivo* measuring techniques. Observing the internal mechanisms of biomotion without harming the animal is beyond current technology.

Because of these challenges, investigators have more recently studied and developed data-driven models of animal motion. This strategy has arisen in several papers that seek to develop low-dimensional models of human walking [12], as well as in recent studies of bat flapping flight by the authors in [3–5]. Perhaps the most popular of the general approaches to dimensionality reduction are based on the non-parametric manifold learning or the estimation methods such as Gaussian process (GP) techniques, local linear embedding (LLE) [13], or Isomap [14, 15]. Popular variants of the GP methods include Gaussian process latent variable model (GPLVM) [16–18] and the Gaussian process dynamical model (GPDM) [19–21] technique. In the studies by the authors in [3–5, 22], it is desired to construct a low-dimensional model that can be used in the prediction step of a Bayesian filter that tracks fiducial markers on an animal during motion. It is important to note that all of the methods, LLE, Isomap, GPLVM and GPDM are non-parametric: there is no underlying model of the physics of the system under study.

In this paper, we present a principled, data-driven approach to the generation of Lagrangian models for such high-dimensional, nonlinear, and uncertain systems. We refer to a representation generated by the approach as an empirical-analytical Lagrangian (EAL) model. The goal of the approach is to (1) exploit as much knowledge as possible from the geometric structure of the Lagrangian evolution as encoded in the known kinetic T and potential energy U , but also to (2) ensure that the derived model is approximately faithful to experimental measurements of the systems. In this sense, the EAL method can be viewed as a hybrid modeling approach that exploits known information that is associated with the Lagrangian geometry, but supplements that information with empiri-

cal experimental data. In contrast to Isomap, LLE, GPLVM and GPDM that are non-parametric estimations methods, EAL method does construct a physics-based model.

The approach presented here synthesizes results from learning theory in reproducing kernel Hilbert spaces (RKHS), penalty approximations of Lagrangian mechanics, and approximation theory. See [23] for a discussion of RKHS in Bayesian estimation for a detailed background. A rigorous treatment of penalty approximations, viewed as a problem of singular perturbation, can be found in [24,25]. Penalty approximations in Lagrangian dynamics have also been studied in [26–28].

The EAL method introduced in this paper uses recent advances in learning theory to construct an empirical potential function from observations of the Lagrangian evolution. The potential force generated by empirical potential function can be seen as an approximation of unknown constraint force. Essentially, the introduction of the empirical potential in the Lagrangian equations of motion has the effect of driving the model trajectory toward the subset or sub-manifold on (or near) which observations have been collected. In particular, it can be viewed as a means of addressing one source of uncertainty: the uncertainty in the feasible motions of a particular locomotion regime.

We have elected in this paper to employ empirical potentials in a Lagrangian formulation since there is a general, cohesive theory for enforcing the associated constraints approximately. The theoretical study of penalty methods for Lagrangian formulations has a long history [26,27,29]. There is in principle no restriction to develop such approaches using a Lagrangian formulation of course. It is conceivable that a similar strategy that employs constraint forces derived from empirical potentials could be based on Newtonian methods. The details of such an approach would exceed the limitations of the current paper, and we leave such a study as an open problem for future research.

2 Empirical potentials and biomotion joint models

The discussion above summarizes the general strategy and goals of this paper. But there are also appealing, intuitive interpretations of the empirical potential that

is used in the EAL method. In this section, before discussing the theoretical details of the EAL formulation, we show in some detail that the approach holds promise for novel approaches in robot joint design based on experimental observations of animals. The following subsections give a concise description of the experimental setup as well as a description of some empirical potentials estimated from experimental data.

2.1 Bat motion studies

Bat flight motions are highly nonlinear, and they can exhibit significant self-occlusion during flight. For that reason, motion study of bats can require an order of magnitude more cameras than other motion studies. The bat motion studies in [3–5] use an array of 30 GoPro Hero 3+ cameras. The cameras are configured in a bat flight measurement tunnel at the SDU-VT International Laboratory depicted in Fig. 2. These cameras are inexpensive and provide sufficient sampling rate and resolution for motion study. Further, their size allows much better handling and easier placement when compared to bulkier high-speed cameras. However, the GoPro lenses have a high level of distortion, and it is crucial to remove the distortion in frames before the estimation of trajectories of the various marker points on the bat. After removing the lens distortion, a pinhole projection camera model has proven sufficient for the sensor model. Finally, a standard unscented Kalman filter is used for the estimation of the motion trajectory of 34 marker points that are placed on the bat as shown in Fig. 1. Further details on the study can be found in [3–5].



Fig. 2 The tunnel equipped with an array of GoPro cameras for recording the flight motion of bats

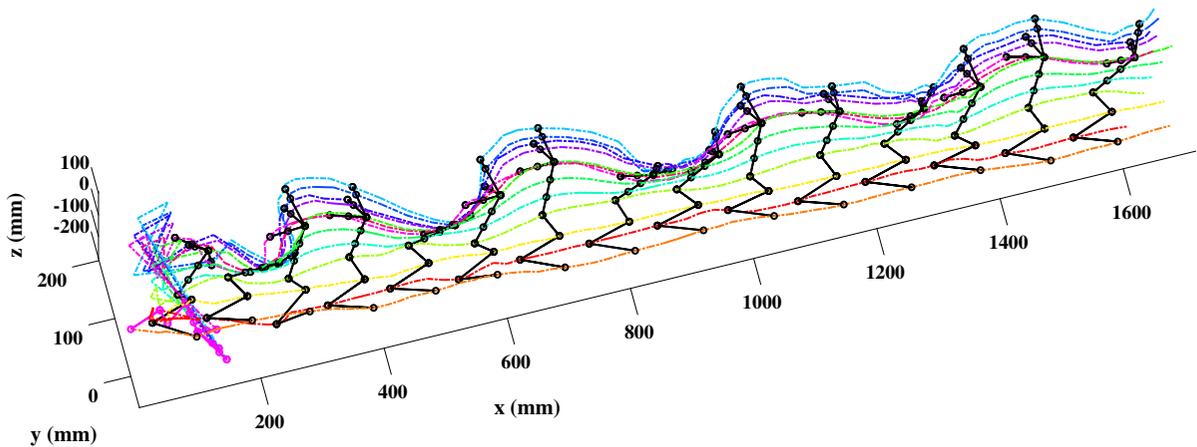


Fig. 3 The trajectories of all the joints on the left wing of a bat during a straight level flight motion

Figure 3 illustrates a typical reconstruction of the marker inertial trajectories obtained from an experiment. As we discuss shortly, such reconstructed motions are used to identify empirical potential functions.

2.2 Reptile motion studies

Unlike the large array of cameras used to study bat flight, the lizard motion study in [22] employs only four Photron FASTCAM cameras. These cameras are placed surrounding a central tank where the lizard motion is recorded. Each time the cameras are switched on to begin recording, a calibration process is performed to determine both the intrinsic parameters of each camera and its extrinsic parameters, i.e., position and orientation of each camera relative to a global frame. Calibration is achieved using a recording of a checkerboard pattern of known dimension in different positions and orientations (Fig. 4).

Before each run, an adult brown anole (*Anolis sagrei*), from a captive laboratory population, is marked at various points of interest with the intention of tracking the trajectory of the points during the lizard's gait. Then the lizard is placed on a wooden dowel that rests diagonally across the central tank. Recording begins as soon as the lizard starts to crawl and ends once it reaches the top of the rod. The motion during a typical experiment is illustrated in Fig. 5. In contrast with the high occlusion of markers during bat flight, the limbs of the lizard are consistently exposed throughout the



Fig. 4 Motion capture experiment for a bat performing level straight flight [30]



Fig. 5 Motion capture experiment for a lizard walking along a rod

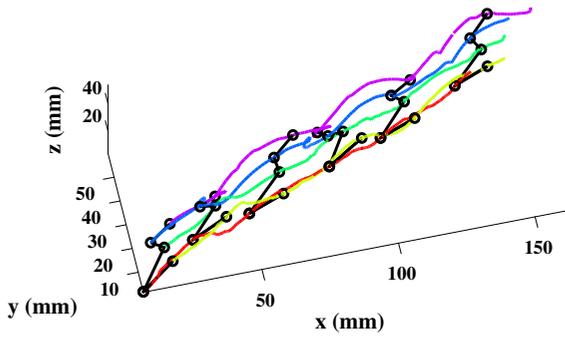


Fig. 6 The trajectories of all the joints on the left arm of a lizard during the walking motion

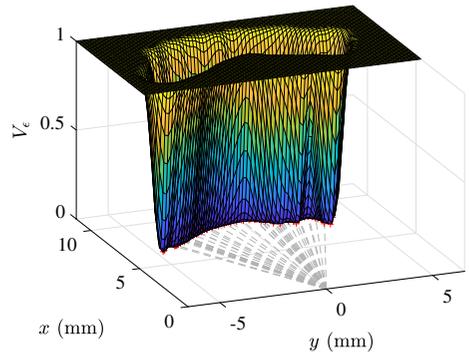


Fig. 8 The normalized empirical potential representing the elbow joint of a lizard during crawling motion

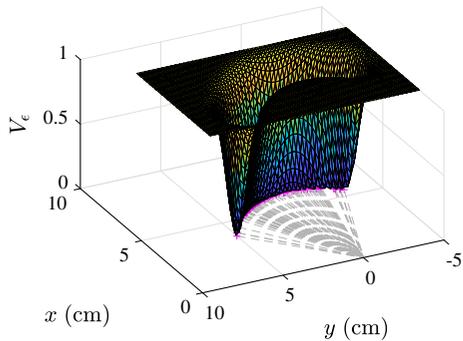


Fig. 7 The normalized empirical potential representing the shoulder joint of a bat during straight level flight motion

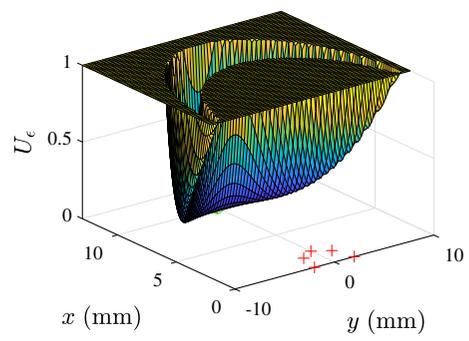


Fig. 9 The normalized strain energy of a prototype nonlinear joint designed to (approximately) realize the experimentally derived potential in Fig. 7

entire motion study to two cameras at minimum. Consequently, stereo triangulation is used to easily determine most marker locations in inertial space. Triangulation of each marker is carried out using Hedrick’s DLTdv digitizing tool [31]. A representative trajectory of the markers during an experiment is depicted in Fig. 6.

2.3 Empirical potential functions

Figure 7 depicts the empirical potential V_ϵ derived from video observations of a bat during a straight and level flight. The figure depicts $V_\epsilon(q)$ when the elbow is fixed at the origin and $q := (q_1, q_2)$ is the location in \mathbb{R}^2 of the wrist. The normalized empirical function V_ϵ depicts both the range and preferred configurations of the bat elbow during this motion regime. Similarly, Fig. 8 depicts a typical empirical potential function derived from the inertial trajectories of the markers on the anole lizard. The qualitative resemblance, and sub-

tle differences, of the two empirical potential functions for the bat and lizard elbows are striking.

Figure 9 illustrates the strain energy of a large displacement, nonlinear truss structure that has been designed to emulate the motion permitted in Fig. 7. The strain energy depicted is $U := U(q)$, where again $q = (q_1, q_2)$ is the position of the wrist and the elbow is fixed at the origin. Our work in progress includes optimization of the strain energy of physical systems that define a joint to approximate the empirical potentials derived directly from experimental observations of motion. These examples emphasize that the empirical potentials V_z that characterize joint motions illustrate the mechanics of the feasible motions of a joint graphically. As studied in much more details in reference [3–5], they also can serve as the foundation for the synthesis of robot joints, even those that can be realized by 3D printing, as illustrated in Fig. 10. In the rest of this paper, we study the theoretical sense in which Lagrangian models based on the empiri-

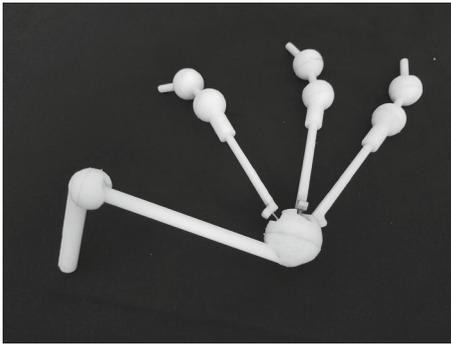


Fig. 10 The prototype of the nonlinear joint manufactured with 3D-printing to realize the empirical potential in Fig. 7 [30]

cal potentials V_z can approximate holonomically constrained Lagrangian systems.

3 Overview of results

3.1 Constrained Lagrangian mechanics and penalty methods

We assume that the motion of a Lagrangian system in the configuration manifold $Q := \mathbb{R}^d$ is subset to m holonomic constraints $\phi_i : Q \rightarrow \mathbb{R}$ for $i = 1, \dots, m$. We define the submanifold $Q_0 = \{q \in Q : \phi_i(q) = 0, i = 1, \dots, m\}$ and refer to Q and Q_0 as the full and constrained motion manifolds, respectively.

Lagrange’s equations for the dynamics of the holonomically constrained mechanical system are commonly written as

$$\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}} \right) - \frac{\partial T}{\partial q} + \frac{\partial U}{\partial q} + \frac{\partial \phi^T}{\partial q} \lambda_j = \Theta, \tag{1}$$

$$\phi(q) = 0, \tag{2}$$

with the generalized coordinates $q(t) \in \mathbb{R}^d$, the constraints $\phi(q(t)) \in \mathbb{R}^m$, the kinetic energy $T = T(q, \dot{q})$, the potential energy $U = U(q)$, and the generalized force $\Theta(t) \in \mathbb{R}^d$. In the form given above, this system constitutes a collection of $(d + m)$ second-order differential-algebraic equations in the $(d + m)$ variables (q, λ) [24]. These equations may also be understood as one coordinate realization of Lagrange’s constrained dynamics on a smooth manifold, as described in [32–34]. From the perspective of differential geometry, the Lagrange’s equations of motion of dynamical systems with constraints can be established on the configura-

tion manifold in the coordinate-free manner. Let Q denote the configuration manifold. It has been proven that the equations of motion expressed in any coordinate chart of Q are equivalent to the ones established with $q(t) \in \mathbb{R}^d$, the generalized coordinates. A fair amount of technical machinery is required to give a full account of the associated coordinate-free formulations of the Lagrangian mechanics on the tangent bundle TQ , or of the Hamiltonian mechanics on the cotangent bundle T^*Q . For the purposes of this paper, we simply assume that the families of approximate trajectories discussed below remain in the domain of a single coordinate chart, which simplifies the analysis substantially.

Equations 1 and 2 must be solved simultaneously to obtain the motion trajectories $t \mapsto q(t)$ and multiplier trajectories $t \mapsto \lambda(t)$. An alternative means to obtain approximations of the constrained motion trajectories is based on penalty methods for Lagrangian mechanics. Penalty approximations of such constrained systems have been studied for many years [25, 27–29]. These methods assume that there is a description of the admissible motion submanifold

$$Q_0 = \{q \in Q \mid \phi(q) = 0\} \tag{3}$$

in terms of the level set of a nonnegative function $V : \mathbb{R}^d \rightarrow \mathbb{R}^+$,

$$Q_0 := \{q \in Q \mid \phi(q) = 0\} = \{q \in Q \mid V(q) = 0\}. \tag{4}$$

The penalized approximation of the constrained Lagrangian system described by Eqs. 1 and 2 is then given by the solutions q_ε of the equations

$$\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}} \right) - \frac{\partial T}{\partial q} + \frac{\partial U}{\partial q} + \frac{1}{\varepsilon^2} \frac{\partial V}{\partial q} = \Theta. \tag{5}$$

The function $\frac{1}{\varepsilon^2} V$ can be thought of as a “effective constraining potential” that drives the system trajectories toward the submanifold as $\varepsilon \rightarrow 0$. Under some non-trivial sufficient conditions that we discuss more fully below, it can be argued that the trajectories q_ε of Eq. 5 converge to the trajectories q_λ of Eq. 1. A typical convergence result has the form

$$\sup_{t \in [0, T]} \|q_\lambda(t) - q_\varepsilon(t)\| \lesssim \varepsilon, \tag{6}$$

that is, the error is uniformly bounded by ε over the time interval $[0, T]$. In this paper, we write $a \lesssim b$ if there is a constant c (that we do not particular care about) such that $a \leq c \cdot b$.

Ensuring this state error convergence is hardly easy, and we summarize some sufficient conditions briefly here for completeness. The proof of the error bound in 6 can be subtle, and the interested reader is referred to [24, 29] for the details. It is perhaps simplest to summarize the analysis in [29], which is presented in Hamiltonian form. Suppose that a $d \times d$ mass matrix $m(q)$ is given, and it is uniformly positive definite, smooth, and symmetric. When $T = \frac{1}{2} \dot{q}^T m(q) \dot{q}$, reference [29] shows that the trajectories of the penalty formulation converge to those of a system that has the effective Hamiltonian $H_e = \frac{1}{2} p^T m^{-1} p + U(q) + W(I, q)$ and is subject to the constraints $\phi(q) = 0$. Note that the effective Hamiltonian H_e includes a correction term, $W(I, q) = \sum_{k=1}^m I_k \omega_k(q)$ where ω_k for $k = 1, \dots, m$ are referred to as frequencies and $I = (I_1, \dots, I_m)$ are the associated actions. The term $W(I, q)$ does not appear in the Lagrangian form with multipliers for holonomically constrained system. Suppose the following assumptions hold,

- (A1) the initial conditions are sufficiently close to the constrained submanifold,
- (A2) the frequencies ω_k satisfy non-resonance conditions, and
- (A3) the mass matrix satisfies tangency conditions defined in terms of the submanifold.

Under these conditions, references [24, 29] show that $\|q_\lambda(t) - q_\varepsilon(t)\| \approx \mathcal{O}(\varepsilon)$.

3.2 Empirical-analytical Lagrangian models

In this paper, we assume that the system of interest is subject to the governing Eqs. 1 and 2, *but we do not have an analytic form of the constraint $\phi(q) = 0$* . Thus, it is not possible in this case to implement Eqs. 1 and 2 in simulation. Instead, we are given observations $\{z_1, \dots, z_N\} \in Q^N$ from experiment of the state q collected during a motion regime of interest. That is, we assume that there is a Lagrangian submanifold that underlies a particular motion regime (such as walking, running, straight and level flight, etc) and collect observations of the system trajectories as they evolve. These observations are taken to be independent and

identically distributed random samples having probability distribution $\rho : \Omega \rightarrow [0, 1]$. In practice Ω is taken to be a compact subset of the configuration space $\Omega \subseteq Q$. Instead of Eq. 4, we assume that there is a potential function $V_\rho = V_\rho(q)$, which is parameterized in terms of the measure ρ , whose level set characterizes the submanifold of admissible configurations in

$$Q_0 = \{q \in Q \mid \phi(q) = 0\} = \{q \mid V_\rho(q) = 0\}.$$

The subset Q_0 is the support of the measure ρ . Now a model of the system, one that does not depend on explicit knowledge of the constraints, can be constructed by choosing $V = V_\rho$ in Eq. 4.

Unfortunately, such a strategy is not yet realizable in computations since the probability measure ρ is unknown. This is the defining feature of distribution-free learning theory [35], a framework that is essential to this paper. We instead construct an empirical approximation $V_z := V_z(q)$ of V_ρ that depends on the samples $z = \{z_1, \dots, z_N\}$. The final, realizable approximation of the unknown system then is given by solutions of the unconstrained equations

$$\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}} \right) - \frac{\partial T}{\partial q} + \frac{\partial U}{\partial q} + \frac{1}{\varepsilon^2} \frac{\partial V_z}{\partial q} = \Theta. \tag{7}$$

The construction of the empirical potential V_z is somewhat lengthy and discussed in Sect. 6. The primary theoretical results in this paper are found in Theorems 1, 2 and 3. The first theorem starts with the analysis of the accuracy confidence function in Eq. 8. Theorem 2 provides a succinct and useful measure of error in expectation over samples. Both of these theorems yield error characterizations that differ significantly from those in reference [36]. Finally, in Theorem 3, we show that V_ρ and V_z can be interpreted as measures of the deviation of the trajectories $q_{\varepsilon, \rho}$ and $q_{\varepsilon, z}$ from the submanifolds underlying their associated penalized Lagrangian systems. We say that these quantify the constraint violation, or in other words, describe the distance of a trajectory to the underlying submanifold of constrained motions. In both cases, these measures of constraint violation are $\mathcal{O}(\varepsilon^2)$.

Theorem 1 *Let ρ^N be the product measure $\rho \otimes \dots \otimes \rho$ on Ω^N , and suppose V_ρ is defined in terms of a reproducing kernel K that satisfies the approximation property in Definition 1. Then we have an accuracy confidence bound*

$$\rho^N \left\{ \sup_{q \in Q} |V_j(q) - V_z(q)| > \eta \right\} \leq \begin{cases} 1 & \text{if } \eta \leq \eta_{\text{cr}} \\ 2e^{-\frac{1}{32}N(\lambda_{M_j} - \lambda_{M_{j+1}})^2 \eta^2} & \text{if } \eta > \eta_{\text{cr}} \end{cases} \tag{8}$$

with N the number of samples, $\eta_{\text{cr}} := 8\sqrt{N}(\lambda_{M_j} - \lambda_{M_{j+1}})$, $\{\lambda_j\}_{j=1, \dots, \infty}$ the sequence of eigenvalues associated with V_ρ in Eq. 13, and λ_{M_j} denoting the eigenvalue at which the spectral approximation V_j of V_ρ is truncated.

Admittedly, while this bound takes a moment to unpack, it is general and useful in generating other bounds that have a simpler form. In Theorem 2, we find how the accuracy confidence function is used to derive other bounds having a simpler form.

Theorem 2 *Suppose that the hypotheses of Theorem 1 hold. Then we have*

$$\mathbb{E}_{\rho^N} \left(\sup_{q \in Q} |V_j(q) - V_z(q)| \right) \leq \frac{C}{\sqrt{N}(\lambda_{M_j} - \lambda_{M_{j+1}})} \tag{9}$$

for some constant $C > 0$.

Also, the accuracy confidence function is a key result that enables decomposition of the error into approximation error (or bias) and probabilistic error (or variance). The use of accuracy confidence functions for such an error decomposition is studied in detail in [35], among other places.

The probability bound encoded in the accuracy confidence function has another important pragmatic interpretation. It relates the measures of constraint violation summarized in Theorem 3

Theorem 3 *Let $q_{\varepsilon, \rho}$ be the solution of the penalized Lagrange’s equation in Eq. 1 when we choose $V = V_\rho$, and let $q_{\varepsilon, z}$ be the solution of the penalized Lagrange’s equation in Eq. 7 with $V = V_z$. If the system is natural, we have the uniform bounds on constraint violation*

$$V_\rho(q_{\varepsilon, \rho}(t)) \lesssim \varepsilon^2 \tag{10}$$

$$V_z(q_{\varepsilon, z}(t)) \lesssim \varepsilon^2 \tag{11}$$

Since these potentials are nonnegative, they can be viewed as defining measures of the distance of the approximate solution to the underlying manifold of admissible configurations.

The above bound can be used to study in what sense the empirical solutions $q_{\varepsilon, z}$ approximate solutions $q_{\varepsilon, \rho}$. However, it remains an open question to investigate when it is true that $q_{\varepsilon, z}$ approximates the solutions q_λ of the constrained system in Eq. 5. The assumptions (A1),(A2), and (A3) that are known to be sufficient for the convergence of solutions q_ε of Eq. 5 to q_λ are complex and difficult to guarantee in the simplest of cases *when an analytic expression for the constraints are known*. This paper is predicated on the assumption that we do not know the constraints. Additionally, the probability measure ρ that dictates concentration of samples in Ω is assumed unknown. This is a standard assumption that underlies distribution-free learning theory, one of the primary theoretical underpinnings of our approach. Since V_ρ is expressed via an infinite summation of unknown eigenfunctions that depend on ρ , it is impossible to verify conditions (A1), (A2), and (A3), at least in their usual form.

The remainder of this paper provides the background and details to derive Theorem 1. We carry out detailed simulations that investigate the convergence properties and stability of the approximate penalized system. In Sect. 4, we review the fundamental definitions from the theory of reproducing kernel Hilbert spaces (RKHS) that are used to construct our approximations. The role of the operators T_K and T_ρ in constructing our potential function V_ρ is discussed. While V_ρ cannot be computed in practice, its empirical approximation V_z is discussed in Sect. 6 in terms of the empirical operator T_z . Section 6 introduces the limiting potential V_ρ , its eigenfunction approximation V_j , and its empirical approximation V_z in terms of the operators T_K , T_ρ , and T_z .

4 Reproducing kernel Hilbert spaces

The approximation of empirical potentials, which are used to modify the analytic expression for the governing equations, are obtained using learning theory in reproducing Kernel Hilbert Spaces. A Hilbert space H defined over a domain Ω is defined in terms of a continuous kernel function $K : \Omega \times \Omega \rightarrow \mathbb{R}$ that is symmetric and positive definite in the sense that

$$\sum_{i,j=1}^I K(x_i, x_j)\alpha_i\alpha_j > 0$$

for all $\{x_i\}_{i \in I} \subseteq \Omega$ and nonzero vectors $\alpha := \{\alpha_1, \dots, \alpha_I\}^T \neq \mathbf{0}$. By convention, the function $K_x(\cdot) := K(x, \cdot)$ is referred to as the kernel function centered at x , and the inner product of any two such kernel functions is defined to be $(K_x, K_y)_H := K(x, y)$. The RKHS H is then defined to be the completion of the finite linear span of the set of functions $\{K_x\}_{x \in \Omega}$ with respect to this inner product. With this setup, it is known for any $f \in H$, we have

$$(K_x, f)_H = f(x)$$

for all $x \in \Omega$. This identity is known as the reproducing property of the kernel K in the RKHS H . There are many well-known examples of possible kernels K on Ω a subset of \mathbb{R}^d and a good description of the most popular choices can be found in [37].

In our applications, we will assume that we collect a sample of observations $z = \{z_1, \dots, z_N\} \subseteq \Omega$ that are independent and identically distributed according to some unknown probability measure p on Ω . The probability measure p is understood intuitively as a description of where samples in z concentrate in Ω . For example, in a human walking experiments, it is pictured that samples are concentrated on a small set or submanifold Q_0 of the full configuration space Q . Under some fairly simple hypothesis (see [38] or [39]), it can be shown that every function in H is continuous and

$$H \subset C(\Omega) \subset L^2_\mu(\Omega).$$

For instance, the rightmost inclusion above holds if the measure $\mu(\Omega)$ is finite, and the left inclusion holds if there is a $c > 0$ such that

$$\sup_{x \in \Omega} |K(x, x)| \leq c < \infty.$$

In fact more can be said about these inclusions. Since $H \subset C(\Omega)$, we define the injection mapping $I_K : H \rightarrow C(\Omega)$ as $I_K : f \in H \rightarrow I_K(f) = f \in C(\Omega)$.

$$\sup_{x \in \Omega} |K(x, x)| \leq c < \infty$$

The requirement above implies that the norm of the operator I_K is uniformly bounded, that is, $\|I_K f\|_{C(\Omega)} \leq \|f\|_H$. In addition, the operator I_K is compact in this case [36]. We can calculate the adjoint I_K^* of I_K from the identity

$$\begin{aligned} (I_K K_q, g)_{L^2(\Omega, \rho)} &= (K_q, I_K^* g)_H \\ &= (I_K^* g)(q) \\ &= \int_\Omega K_q(y)g(y)\rho(dy). \end{aligned}$$

We then introduce the mapping $T_K : L^2(\Omega, \rho) \rightarrow L^2(\Omega, \rho)$ to be $T_K := I_K I_K^*$, which yields

$$\begin{aligned} T_K g &:= I_K I_K^* g \\ &= I_K \underbrace{\int_\Omega K(\cdot, r)g(r)\rho(dr)}_{\in H} \\ &= \underbrace{\int_\Omega K(\cdot, r)g(r)\rho(dr)}_{\in L^2(\Omega, \rho)}. \end{aligned}$$

We next define $T_\rho := I_K^* I_K$ which maps $T_\rho : H \rightarrow H$. Both T_K and T_ρ are self-adjoint compact operators since I_K is a compact embedding. By virtue of the spectral theory for self-adjoint compact operator [11,40], the eigenvalues of T_K and T_ρ are the same can only accumulate at zero. By convention, they are ordered in a nonincreasing extended enumeration that includes multiplicities, $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$. When we denote by $\{\psi_j\}_{j=1}^\infty$ and $\{\phi_j\}_{j=1}^\infty$, the eigenvalues of T_ρ and T_K , respectively, spectral theory guarantee that

$$\begin{aligned} T_\rho g &= \sum_{j=1}^\infty \lambda_j (g, \psi_j)_H \psi_j && \text{in } H, \\ T_K f &= \sum_{j=1}^\infty \lambda_j (f, \phi_j)_{L^2(\Omega, \rho)} \phi_j && \text{in } L^2(\Omega, \rho), \\ I_K^* f &= \sum_{j=1}^\infty \sigma_j (f, \phi_j)_{L^2(\Omega, \rho)} \psi_j && \text{in } H, \end{aligned}$$

for each $f \in L^2(\Omega, \rho)$ and $g \in H$. The j th singular value of I_K^* is define by $\sigma_j := \sqrt{\lambda_j}$, and it can be shown that $\psi_j = \sigma_j \phi_j$ almost everywhere in Ω [36].

5 The discrete operator T_z

In our problem, T_ρ and T_K are integral operators. They map $T_\rho : H \rightarrow H$ and $T_K : L^2(\Omega, \rho) \rightarrow L^2(\Omega, \rho)$, so they act between infinite-dimensional Hilbert spaces in general. Our goal is to construct a suitable *computable* approximation T_z of T_ρ . The construction of T_z takes several steps. We define a sampling operator

$$S_z : H \rightarrow \mathbb{R}^N, \quad S_z(f) := \begin{Bmatrix} f(x_1) \\ \vdots \\ f(x_N) \end{Bmatrix} \in \mathbb{R}^N.$$

Many of the properties of this sampling operator are studied in [38]. We can easily compute the adjoint of the sampling operator by its definition $(S_z f, x)_{\mathbb{R}^N} = (f, S_z^* x)_H$ when H is a reproducing kernel Hilbert space. It follows that the adjoint $S_z^* : \mathbb{R}^N \rightarrow H$ with

$$S_z^* \begin{Bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{Bmatrix} = \sum_{i=1}^N y_i K_{z_i},$$

when the samples $z = \{z_1, z_2, \dots, z_N\} \subseteq \Omega \subseteq Q$. The discrete operator T_z is defined as

$$T_z = \frac{S_z S_z^*}{N} = \frac{1}{N} K(z_i, z_j) = \frac{\mathbb{K}}{N}. \tag{12}$$

The matrix \mathbb{K} is the learning matrix with $\mathbb{K}_{i,j} = K(z_i, z_j)$ and z_i represents the i th data point in the data set.

6 Empirical potential functions

The discrete evolution laws studied in this paper are constructed by using an empirical potential function V_z that drives the discrete state to evolve near the subset $\Omega \subset Q$. Since the analysis in this paper extends the error bounds in [36], we adopt its notation in this section. We define the empirical potential function V_z ,

its ‘‘infinite sample’’ limit V_ρ , and its approximation V_j due to spectral filtering via the expressions

$$V_\rho(q) := 1 - (T_\rho^\dagger T_\rho K_q, K_q)_H, \tag{13}$$

$$V_z(q) := 1 - (g_\lambda(T_z) T_z K_q, K_q)_H, \text{ and} \tag{14}$$

$$V_j(q) := 1 - (g_\lambda(T_\rho) T_\rho K_q, K_q)_H. \tag{15}$$

for all $q \in \Omega \subseteq Q$ where the filter function g_λ is selected to be the spectral filtering function such that $g_\lambda(\xi) = 1$ if $\xi > \lambda$, and $g_\lambda(\xi) = 0$ otherwise. Terms such as $g_\lambda(T_\rho)$ are interpreted in the sense of the functional calculus of operators, see [41]. More general choices of g_λ are discussed in [36]. This reference introduces the following smoothness prior that is used to establish Theorem 1 in this paper.

Definition 1 We say that the kernel K and probability measure ρ satisfy the *smoothness prior* provided

$$\sup_{q \in \Omega} \|T_\rho^{-s/2} P_\rho K_q\|_H^2 \leq C_s,$$

for some constant $C_s > 0$ and $s > 0$ where P_ρ is the projection operator onto the range of T_ρ .

The remainder of this section proves Theorem 1.

Proof of Theorem 1 We begin by using the triangle inequality to derive a bound for the difference between the empirical and limiting potential function V_ρ and V_z .

$$\begin{aligned} & \left| V_z(q) - V_\rho(q) \right| \\ & \leq \left| V_z(q) - V_j(q) \right| + \left| V_j(q) - V_\rho(q) \right| \\ & = \left| \underbrace{\left((g_\lambda(T_z) T_z - g_\lambda(T_\rho) T_\rho) K_q, K_q \right)_H}_{\text{term 1}} \right| \end{aligned} \tag{16}$$

$$+ \left| \underbrace{\left((g_\lambda(T_\rho) T_\rho - T_\rho^\dagger T_\rho) K_q, K_q \right)_H}_{\text{term 2}} \right|. \tag{17}$$

that yields expressions for two terms. We set up our analysis of each term in this section by recalling several results from [36]. The first term in Eq. 17 satisfies

$$\begin{aligned} & \left| \left((g_\lambda(T_z)T_z - g_\lambda(T_\rho)T_\rho)K_q, K_q \right)_H \right| \\ & \leq \frac{2}{\lambda_{M_j} - \lambda_{M_{j+1}}} \|T_z - T_\rho\|_{HS}, \end{aligned} \tag{18}$$

which is found in [36]. The same reference shows that

$$\|T_\rho - T_z\|_{HS} \leq \frac{2(\delta \vee \sqrt{2\delta})}{\sqrt{N}} \tag{19}$$

with probability at least $1 - 2e^{-\delta}$ from Lemma 1 of [36].

holds for all sample z that lie in a set having probability (or measure) at least $1 - 2e^{-\delta}$. When $g_\lambda(\cdot)$ denotes the spectral truncation between eigenvalues λ_{M_j} and $\lambda_{M_{j+1}}$, which is $\lambda_{M_{j+1}} < \lambda_j < \lambda_{M_j}$, the Schmidt expansion of $g_\lambda(T_\rho)$ implies that

$$\begin{aligned} & \left| \left((g_\lambda(T_\rho)T_\rho - T_\rho^\dagger T_\rho)K_q, K_q \right)_H \right| \\ & = \sum_{j=\lambda_{M_{j+1}}}^{\infty} (P_\rho K_q, \psi_j)_H^2 \leq \lambda_{M_{j+1}}^s C_s. \end{aligned} \tag{20}$$

The rightmost inequality above is a result of the smoothness prior in Definition 1. We obtain an upper bound on the inequality in Eq. 17 by combining the expression in Eqs. 17, 6, 19, and 20.

$$\sup_{q \in \Omega} |V_z(q) - V_\rho(q)| \leq \frac{4(\delta \vee \sqrt{2\delta})}{\sqrt{N}(\lambda_{M_j} - \lambda_{M_{j+1}})} + C_s \lambda_{M_{j+1}}^s. \tag{21}$$

The inequality illustrates the dependency of the right hand side on a probabilistic term that is proportional to $\frac{1}{\sqrt{N}}$ and a deterministic term that depends on $\lambda_{M_{j+1}}$. We discuss the balance of these terms in more detail after this section.

Now we show how the inequality in Eq. 21 can be expressed in terms of an accuracy confidence function in the style of approximation theory as in reference [35]. To this end, we begin by writing

$$\sup_{q \in Q} |V_\rho(q) - V_z(q)| \leq C_s \lambda_{M_{j+1}}^s + \eta \tag{22}$$

which holds for all the samples z that are not contained in the “set of bad samples”.

It is easily shown that this inequality is equivalent to that in Eq. 21 when we set

$$\eta = \frac{4(\delta \vee \sqrt{2\delta})}{\sqrt{N}(\lambda_{M_j} - \lambda_{M_{j+1}})}. \tag{23}$$

We then have

$$\eta = \begin{cases} \frac{4\delta}{\sqrt{N}(\lambda_{M_j} - \lambda_{M_{j+1}})} & \text{if } \delta \leq 2 \\ \frac{4\sqrt{2\delta}}{\sqrt{N}(\lambda_{M_j} - \lambda_{M_{j+1}})} & \text{if } \delta > 2 \end{cases}$$

which is identical to

$$\delta = \begin{cases} \frac{1}{4}\sqrt{N}(\lambda_{M_j} - \lambda_{M_{j+1}})\eta & \text{if } \eta \leq \eta_{cr} \\ \frac{1}{32}N(\lambda_{M_j} - \lambda_{M_{j+1}})^2\eta^2 & \text{if } \eta > \eta_{cr} \end{cases}$$

with the definition $\eta_{cr} := 8\sqrt{N}(\lambda_{M_j})$.

We therefore conclude that

$$\begin{aligned} & \rho^N \left\{ \sup_{q \in Q} |V_j(q) - V_z(q)| > \eta \right\} \\ & \leq \begin{cases} 1 & \text{if } \eta \leq \eta_{cr} \\ 2e^{-\frac{1}{32}N(\lambda_{M_j} - \lambda_{M_{j+1}})^2\eta^2} & \text{if } \eta > \eta_{cr} \end{cases} \end{aligned} \tag{24}$$

□

7 Interpretation of empirical potential energy error in terms of bias and variance

Equation 22 is important to interpret the use of the accuracy confidence function. This inequality states that

$$\sup_{q \in Q} |V_\rho(q) - V_z(q)| \leq \underbrace{C_s \lambda_{M_{j+1}}^s}_{\text{Approximation Error (bias)}} + \underbrace{\eta}_{\text{Probabilistic Error (variance)}} \tag{25}$$

for all samples z outside the set of $A_{j,N}(\eta)$ of bad samples. The accuracy confidence function measures the size of the set of bad samples in a probabilistic sense. The decomposition of the error into bias and variance is a classical topic in learning and approximation theory, see [35] or [42]. It gives rise to the typical trade-off that appears in Fig. 11 in the form of a “V”-shaped plot. On this plot of trajectory error versus $\frac{1}{\epsilon^2}$, the right is dominated by the approximation error that decreases

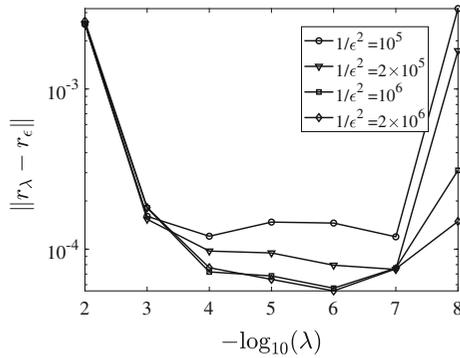


Fig. 11 Double logarithmic plot: spectral filtering parameter versus the norm of constraint violation. The curves are calculated from cases with different values of penalty factor, ϵ , with noise imposed. A typical “V”-shape curve appears when the learning model becomes more complicated

(super)linearly on the log-log plot. The left side of the plot increases linearly due to the dominance of the variance. Optimal estimates are obtained by equilibrating these two contributions, a topic discussed in detail in [35], which yields errors located near the minimum of the “V”-shape.

As emphasized in [35], the utility in estimates of the form in Eq. 24 is that this expression gives a succinct description of the interplay between the number of samples and approximation error. It is quite useful in deriving bounds for the expectation of the error that is given in Theorem 2. We establish this inequality next.

Proof of Theorem 2 We calculate the expectation by integrating the accuracy confidence function: it depends on the distribution function of the error $\|V_j - V_z\|_{C(Q)}$.

$$\begin{aligned} & \mathbb{E}_{\rho^N} \left(\sup_{q \in Q} |V_j(q) - V_z(q)| \right) \\ &= \int_Q \rho^N \left\{ z : \sup_{q \in Q} |V_j(q) - V_z(q)| > \eta \right\} d\eta \\ &= \int_0^{\eta_{cr}} 1 d\eta + \int_{\eta_{cr}}^\infty 2e^{-N(\lambda_{M_j} - \lambda_{M_{j+1}})^2 \eta^2} d\eta \\ &\leq \frac{C}{\sqrt{N}(\lambda_{M_j} - \lambda_{M_{j+1}})} \end{aligned}$$

for some constant $C > 0$. □

Proof of Theorem 3 In general, the kinetic energy of a finite-dimensional Lagrangian system (when expressed

in terms of a specific choice of coordinate chart) can be decomposed into three terms.

$$T = T(t, q, \dot{q}) = T_2(t, q, \dot{q}) + T_1(t, q, \dot{q}) + T_0(t, q, \dot{q}) \tag{26}$$

that are quadratic in, linear in, and independent of \dot{q} , respectively. A system is said to be natural, or T_2 , if $T_1 = T_0 = 0$. When a system is unforced and natural, it is well-known that the Hamiltonian H is conserved. That is, for our systems

$$\begin{aligned} H(t) &= T(q(t), \dot{q}(t)) + U(q(t)) + \frac{1}{\epsilon^2} V_\rho(q(t)) \\ &= T(q(0), \dot{q}(0)) + U(q(0)) + \frac{1}{\epsilon^2} V_\rho(q(0)) \\ &= H_0 \end{aligned}$$

Under the assumption that $q(0)$ is on the submanifold, which holds only if $V_\rho(q(0)) = 0$, it follows that

$$\begin{aligned} \frac{1}{\epsilon^2} V_\rho(q(t)) &\leq T(0) + U(0) \\ V_\rho(q(t)) &\lesssim C\epsilon^2 \end{aligned} \quad \square$$

8 Numerical experiments

In this section, we present a detailed numerical study for some prototype examples to verify and validate the major theoretical results of this paper. These examples are based on a pendulum on a plane, a spherical pendulum in the 3D space, and the simulation of the motion of the elbow joint of an anole lizard. We consider first motion of a pendulum in \mathbb{R}^2 shown in Fig. 12a. A mass m is connected to the origin by a massless bar whose length is $r = 1\text{m}$. Additionally, the mass is subject to gravity. In this prototype example, the only constraint

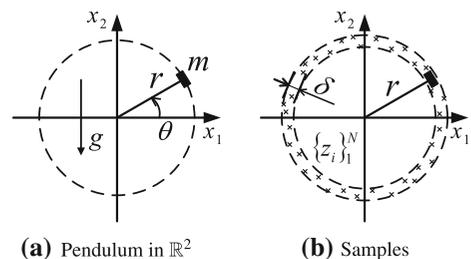


Fig. 12 Diagrams of pendulum and samples

is that the distance between the mass and the origin is constant. Now assume that the analytical form of constraints is unknown (or partially unknown) to us, which is the common situation in biomotion studies. In this way, it is impossible to derive the dynamical equations from the classical Lagrange’s equations with multipliers. However, we suppose that we have observations of motions available from experiments. The EAL model provides a form for incorporating the experimental data into the formulation of Lagrange’s equations for the uncertain system. The empirical potential function is constructed from the data and serves as an additional term in the equations of motion. This term provides an effective potential driving the model such that the trajectory evolves near the submanifold of motion.

8.1 EAL model of pendulum

For the simplest planar pendulum, the constraint on distance between the mass and the origin is denoted by $\phi = \sqrt{x^2 + y^2} - 1 = 0$. Following the convention in the homogenization theories of singular perturbation [25], the constraint is used to define the generalized coordinates $q_1 = r = \sqrt{x^2 + y^2}$. The angular coordinate θ is defined as the “free” coordinate. In this way, the coordinates of the system are $q = [r, \theta]^\top$. We write the Lagrangian and constraints of the system as follows in Eqs. 27 and 28,

$$L = T - V = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) - mgr \sin \theta, \tag{27}$$

$$\phi(q) = r - 1 = 0. \tag{28}$$

The dynamical model of the pendulum can be derived by substituting the kinetic energy, potential energy, and the constraint into Lagrange’s equations. The actual trajectory q_λ is obtained by integrating this system of differential-algebraic equations (DAEs). The mass starts moving at the initial condition of $q_0 = [r_0, \theta_0]^\top$, which is assumed to lie on the admissible configuration submanifold. The samples $\{z_1, \dots, z_N\}$ used for constructing the empirical potential function are taken along the actual trajectory, as is shown in Fig. 12b. In order to study the performance of the EAL model under different noise distributions, two sets of samples are simulated: (1) measurements free from noise, and (2) measurements with uniformly distributed additive noise. Specifically, the measurements are of the following form.

$$z(t_i) = [1 + \Delta r(t_i), \theta(t_i)]^\top \in \mathbb{R}^2, \\ \Delta r \sim \mathcal{U}\left(-\frac{\delta}{2}, \frac{\delta}{2}\right), \quad \delta \geq 0,$$

with $\mathcal{U}(a, b)$ denoting the uniform distribution over $[a, b]$.

As mentioned in Sect. 6, these measurements are used for constructing the empirical potential function $V_z(q)$. Here we demonstrate the construction for this specific example. Reference [36] contains more details about the properties of the constructed function $V_z(q)$. The Gaussian kernel function is defined by the expression $K(x, y) = \exp(-\beta\|x - y\|^2)$, where β is chosen and fixed in what follows. The empirical potential function follows from Eq. 14, which is written as

$$V_z(q) = 1 - K_z(q)^\top g_\lambda(\mathbb{K}) K_z(q) \tag{29}$$

where $K_z(q) = [K(z_1, q), \dots, K(z_N, q)]^\top$ and $\mathbb{K}_{i,j} = [K(z_i, z_j)]$ for $i, j = 1, \dots, N$. The expression of $g_\lambda(\cdot)$ denotes the spectral filtering operator that acts on matrices. We write the eigenvalue decomposition of matrix \mathbb{K} as $\mathbb{K} = VDV^\top$, where matrix D denotes the diagonal matrix of eigenvalues $\{\sigma_i\}$ of \mathbb{K} , and the matrix V is the corresponding orthonormal matrix of eigenvectors. Then the spectral filtering operator $g_\lambda(\cdot)$ can be written explicitly in terms of the scalar function.

$$\bar{g}_\lambda(\sigma) = \begin{cases} 1/\sigma & \sigma \geq \lambda \\ 1/\lambda & 0 < \sigma < \lambda \end{cases} \tag{30}$$

With $\bar{g}_\lambda(\sigma)$ defined as above, the spectral filtering operator is given as

$$g_\lambda(\mathbb{K}) = \frac{1}{N}V\left(\frac{\bar{D}}{N}\right)V^\top \tag{31}$$

with $\bar{D} = \text{diag}(\bar{g}(\sigma_1), \dots, \bar{g}(\sigma_N))$.

Figure 13 shows how the coordinates in the EAL model and the actual model propagate under the same initial conditions, which is set as $r_0 = 1$ and $\theta_0 = \pi/6$. The results in Fig. 13 are typical of EAL methods in several respects. The response trajectory $q_{\varepsilon,z}$ contains both slowly and rapidly oscillating components. The coordinate $q_1 = r$ is nearly a constant since the ideal constraint is $q_1 = r = 1$, or $\phi = q_1 - 1$. The approximate trajectory can be viewed as oscillating rapidly about the underlying submanifold.

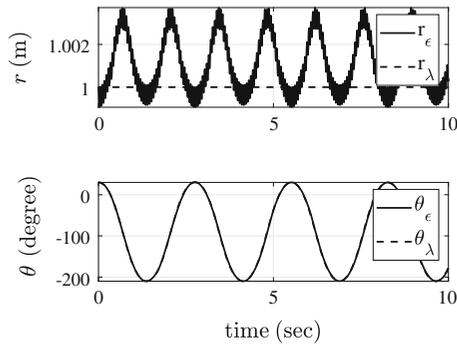


Fig. 13 The trajectories of both the actual and homogenization models within 10s. The spectral filtering parameter, number of samples, and penalty factors are set as $\lambda = 10^{-6}$, $N = 400$, and $1/\varepsilon^2 = 2 \times 10^5$

8.2 Convergence rate of EAL model

Since the actual dynamic model of the pendulum is known clearly, we can compare the trajectories $q_{\varepsilon,z}(t)$ generated by the EAL model and the trajectory $q_\lambda(t)$ satisfying the actual dynamic model to verify the fidelity of the EAL method. The norm of the difference of the two trajectories in the space of continuous functions, $\|q_\lambda - q_\varepsilon\|_{C[0,T]}$, serves as the error metric. In general, the size of error is due to several parameters that arise in the empirical model, which will be discussed shortly. Another important attribute that needs to be examined is the effect of integration error, because the trajectories of both models are obtained via numerical integration. Our numerical study provides insights about how these combinations of attributes affect the accuracy of the model.

We have defined $q_{\varepsilon,\rho}(t)$ as the trajectory generated by Lagrange’s equations that are expressed in terms of $\frac{1}{\varepsilon^2} V_\rho$ and $q_{\varepsilon,z}(t)$ as the trajectory of the corresponding equation in terms of $\frac{1}{\varepsilon^2} V_z$. The empirical potential V_z approximates V_ρ , which is unknown since the distribution ρ is unknown. If ρ were known, and the assumptions A1, A2 and A3 in Sect. 3 hold, it is established in homogenization methods for singular perturbation that [25]

$$\|q_\lambda - q_{\varepsilon,\rho}\|_{C[0,T]} = \mathcal{O}(\varepsilon) \tag{32}$$

with q_λ the trajectory of Lagrange’s equations with multipliers in Eqs. 1 and 2. Ideally, we would hope in the best possible scenario that

$$\begin{aligned} \|q_\lambda - q_{\varepsilon,z}\|_{C[0,T]} &\leq \|q_\lambda - q_{\varepsilon,\rho}\|_{C[0,T]} \\ &\quad + \|q_{\varepsilon,\rho} - q_{\varepsilon,z}\|_{C[0,T]} \\ &\approx \mathcal{O}(\varepsilon) \end{aligned} \tag{33}$$

Our numerical experiments will predominantly study two measures of convergence. From Eq. 33, we can derive the relation between the logarithm of error bound and the penalty factor $1/\varepsilon^2$, which is shown in the following equation. This relation will be useful for the following illustration.

$$\log_{10} \|q_\lambda - q_{\varepsilon,z}\|_{C[0,T]} \leq -\frac{1}{2} \log_{10} \left(\frac{1}{\varepsilon^2} \right) + C \tag{34}$$

Therefore, if we run the simulation for a same case multiple times with different penalty factors, and plot the constraint violation as a function of the penalty factor in a log–log figure, the slope of the curve will reflect the speed of convergence with respect to the penalty. We will see in the following figures that the constraint violation converges to a curve below the reference line whose slope equals to $-\frac{1}{2}$, as predicted by Eq. 34.

The analysis above describes a best possible performance, but in practice the approximation V_z of V_ρ depends on several parameters. The parameters determining the learning process are the spectral filtering parameter λ , number of samples N , the final integration time T , and the distribution of the additive noise. We perform several synthetic experiments in which only one of these parameters is varied. By examining the convergence rate with respect to ε , we are able to tell how a specific parameter qualitatively affects the error bounds of the EAL model. In the remainder of this section, numerical results are presented to illustrate the effects of various simulation parameters on the order of error. The example of planar pendulum described in the previous subsection is used to illustration. In each simulation, a single parameter is varied based on a general case, in which the parameters are $\lambda = 10^{-6}$, $N = 400$, $\beta = 3$, and $T = 5 \times 10^{-2}$ s.

We first study the influence of the spectral filtering operator, which is parameterized by λ . By the definition specified in Eq. 30, a smaller λ implies more of the original eigenvalues of a modal expansion of V_ρ are retained. From Fig. 14a, b, we can see that to the left all the curves are below the reference line with the slope $\frac{1}{2}$. This indicates that the order of constraint violation

Fig. 14 Double logarithmic plot: penalty factor versus the norm of constraint violation. The curves are calculated from cases with different values of spectral filtering parameter, λ . The slope of reference line is $-\frac{1}{2}$

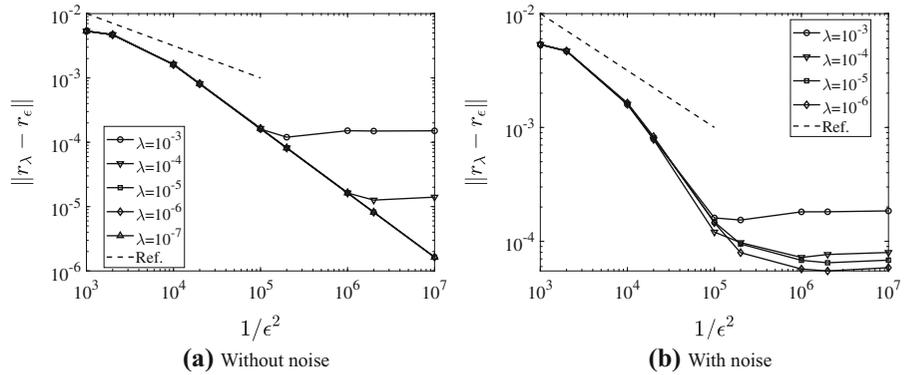
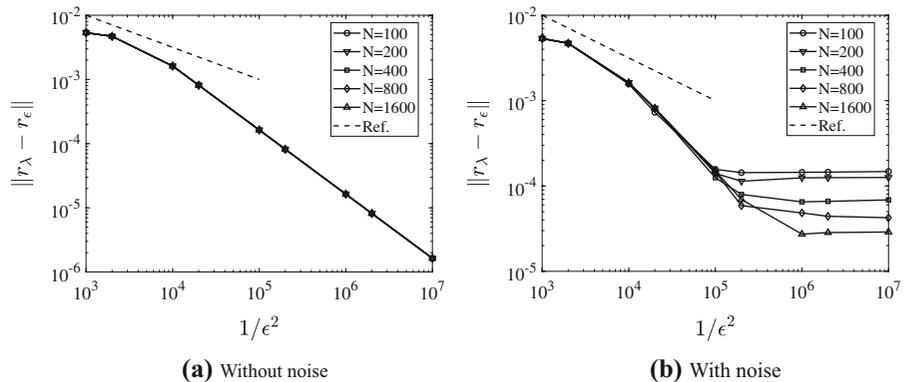


Fig. 15 Double logarithmic plot: penalty factor versus the norm of constraint violation. The curves are calculated from cases with different number of samples, N . The slope of reference line is $-\frac{1}{2}$



is bounded by at least the same order of ϵ guaranteed in Eq. 34.

On the other hand, with the filtering operator g_λ employed to learn the empirical functions, some plateaus appear to the right in both Fig. 14a, b as ϵ approaches zero, i.e., as the penalty factor approaches infinity. That means the penalty model with empirical potential reaches a limit of accuracy. The limit depends on both the noise and the choice of λ . Increasing the penalty factor does not result in a smaller error bound in some cases. In the cases shown in Fig. 14a, the empirical potential functions are learned from samples without noise. As λ decreases, i.e., more information from V_ρ is retained, the limit on accuracy gets smaller accordingly. While in the cases where measurements are noisy, as is shown in Fig. 14b, the decrease in error as λ decreases reaches a limit that depends on the noise. This phenomenon can be explained by the conclusions from approximation theory. As discussed following Eq. 25, the error has the characteristic “V”-shape associated with decomposition of the error into bias and variance. For a range of ϵ to the right in each plot, the convergence rate is the same as in Eq. 34. As

the penalty $\frac{1}{\epsilon^2}$ increases to the right in each plot, the effect of probabilistic error in the sampling operator becomes significant.

Next we study the influence of the number of samples N . According to the probabilistic error bound given in Eq. 9, the expectation of error bound decreases as more samples are used in the empirical potential. Figure 15 shows the results from cases with different numbers of samples. When exact, noise-free measurements are assumed, as is shown in Fig. 15a, the error curves of all cases overlap with each other. The curve is below the reference line, which implies the convergence rate follows the trend observed in given by homogenization theory for singular perturbation. On the other hand, all the curves reach similar limits when the measurements are affected by noise. As predicted in Eq. 9, the limiting error bound is smaller in the cases with more samples.

Integrator error is another important topic in this discussion. In many problems, there exists no closed-form solution, so a numerical integrator is usually the only way to study such dynamical systems. Essentially, a numerical integrator approximates the continuous dynamical systems with a set of discrete difference

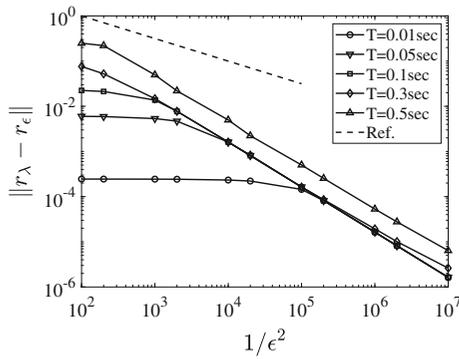


Fig. 16 Double logarithmic plot: penalty factor versus the norm of constraint violation. The curves are calculated from cases with different integration final time, T , without noise. The slope of reference line is $-\frac{1}{2}$

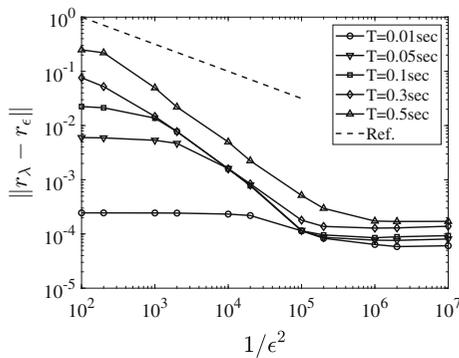


Fig. 17 Double logarithmic plot: penalty factor versus the norm of constraint violation. The curves are calculated from cases with different integration final time, T , with noise. The slope of reference line is $-\frac{1}{2}$

equations. The numerical solution might not propagate exactly on the submanifold generated by constraints, which introduces cumulative error in the calculated trajectory. Therefore, the error in numerical integration needs to be considered and examined in simulations. Figures 16 and 17 show the effect of final integration time on convergence rate. A non-variational integrator, specifically a 4th-order Runge–Kutta method, is employed with the step size set as $h = 10^{-4}$ s. With the same integrator and same integration step size, we conduct simulations for cases with different final times. The integration error accumulates as the final time increases of course. The qualitative trends for both cases with and without noise imposed are quite close to those in which the spectral filtering parameter λ is decreased.

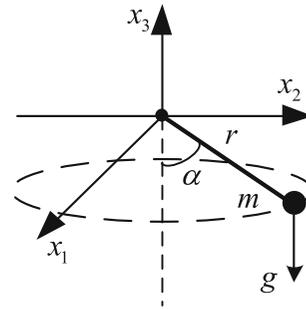


Fig. 18 Diagram of spherical pendulum

So far, we have studied in detail the errors associated with several parameters that arise in the EAL method for a two-dimensional system. We have likewise studied the EAL model of a spherical pendulum, which is shown in Fig. 18. In this example, the Cartesian coordinates of the mass are chosen as the generalized coordinates. Specifically $q = [x_1, x_2, x_3]^T$.

The Lagrangian and constraints of the spherical system are given as follows.

$$L = T - V = \frac{1}{2}m(\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2) - mgx_3 \quad (35)$$

$$\phi(q) = x_1^2 + x_2^2 + x_3^2 - 1 = 0 \quad (36)$$

In all simulations, we observe the same qualitative behavior of the dependence of the error on the parameters ϵ , λ , T , and noise distribution. For example, Fig. 19 depicts the cases when different spectral filtering parameters λ are applied. The measurements in Fig. 19a are exact, and the measurements in Fig. 19b are noisy. Figure 19c shows the typical “V”-shape curve when more original eigenvalues of modal expansions of V_ρ are retained. We see that these plots are qualitatively similar to the simulations for the planar pendulum in Figs. 14 and 11.

In our final example, we use experimental observations of the inertial motion of markers on the anole lizard in Figs. 1b, 5, and 6 to construct an EAL model of the lizard’s wrist joint with respect to the elbow joint.

In this example, only the relative motion is modeled. That is, an empirical potential function is constructed by first calculating the difference between the motion markers attached at the elbow and wrist during a typical motion regime. The dynamics of the relative motion are generated using an effective Lagrangian that has the form $L = T - V$ with $T = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2)$ and $V = 0$. As depicted in Fig. 20, the family of penalized models indexed by ϵ , λ give a good representation of the

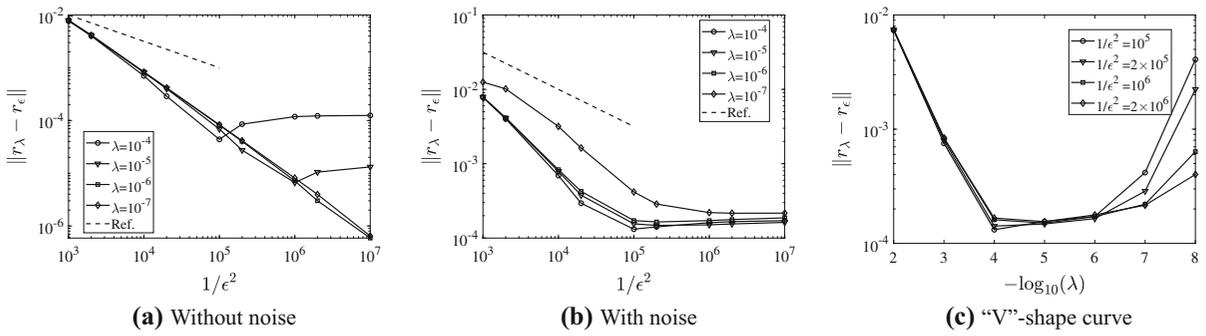


Fig. 19 The simulation results of three-dimensional spherical pendulum. Different spectral filtering parameters λ are applied in the simulations. Other parameters used in the cases are $N = 400$, $\beta = 15$, $T = 5 \times 10^{-2}$ s. (a) and (b) depict the convergence

rates of the cases with exact measurements and noisy measurements. The slope of reference line is $-\frac{1}{2}$. (c) depicts the typical “V”-shape curve that appears when the learning model becomes more complicated

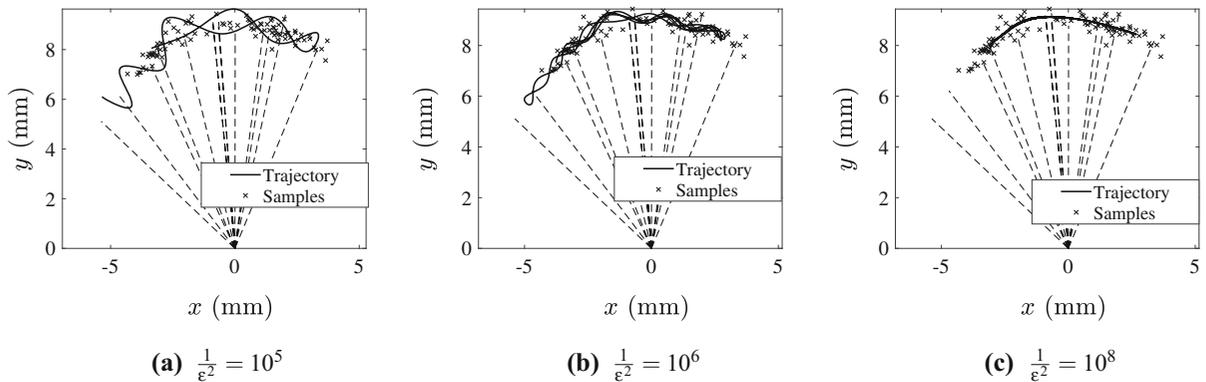


Fig. 20 The positions of a lizard’s wrist joint with respect to the elbow joint in the shoulder-elbow-wrist plane. The samples are taken from a short clip of the lizard’s crawling motion. The

curve is the trajectory of the EAL model constructed from the displayed samples with different penalty factors

relative motion. As ε is decreased, the motion remains close to an underlying submanifold. While this model constitutes a preliminary study of just the relative joint motion, it conveys important information regarding the range of motion and distribution of motion about the submanifold. An expanded study of the inertial motion of the full bioskeletal system is in development by the authors.

9 Conclusions

This paper has introduced a novel, data-driven strategy for the construction of empirical-analytical Lagrangian (EAL) models. This class of models has been motivated by high-dimensional, uncertain systems such as

those that often arise in constructing predictors for Bayesian estimation of animal motion. The method is based on using experimental observations $\{z_1, \dots, z_N\}$ of the system to construct penalty formulations that approximate dynamics of holonomically constrained Lagrangian models. The primary theoretical contribution of the paper is the derivation of an accuracy confidence function that is used to measure the distance of the approximate trajectories to the admissible configuration submanifold in terms of the number of samples N and penalty parameter ε . The form of the accuracy confidence function enables the interpretation of the error between approximated trajectories and the admissible submanifold as a combination of bias and variance terms. The numerical studies carefully assess a num-

ber of properties of the formulations including rates of convergence, modal truncation effects, spectral filtering, and accumulation of integration error.

An important open question suggested by this paper is to extend the convergence results of singular perturbation, which depend only on the penalty parameter ε , to the case in which the penalization term depends on ε and the number of samples N . So far, the numerical studies suggest that the empirical trajectories follow some of the same trends as the trajectories associated with homogenization of singular perturbation problems. Also, the current paper limits consideration to animal motion models that consist of rigid bioskeletal models. Many animal motion studies observe significant flexible body effects. While the theory in this paper can be extended in principle to this case, the details of such a formulation remain an important topic for future study.

Acknowledgements The authors would like to thank the SDU-VT International Laboratory for research collaboration on the bat flight motion capture project.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest concerning the publication of this manuscript.

Ethical approval All applicable international, national and institutional guidelines for the care and use of animals were followed in our studies about biomotion. All procedures performed in studies involving animals were approved by the Institutional Animal Care and Use Committee in Virginia Tech.

References

- Shabana, A.A.: Dynamics of Multibody Systems. Cambridge University Press, Cambridge (2013)
- Spong, M.W., Hutchinson, S., Vidyasagar, M., et al.: Robot Modeling and Control, vol. 3. Wiley, New York (2006)
- Bender, M.J., McClelland, M., Bledt, G., Kurdila, A., Furukawa, T., Mueller, R.: Trajectory estimation of bat flight using a multi-view camera system. In: AIAA Modeling and Simulation Technologies Conference, p. 1806 (2015)
- Bender, M.J., McClelland, H.M., Kurdila, A., Müller, R.: Recursive Bayesian estimation of bat flapping flight using kinematic trees. In: AIAA Modeling and Simulation Technologies Conference, p. 0945 (2016)
- Bender, M., Tian, L., Fan, X., Kurdila, A., Müller, R.: Spatially recursive estimation and Gaussian process dynamic models of bat flapping flight. *Nonlinear Dyn.* **95**, 217–237 (2018)
- Patel, A., Stocks, B., Fisher, C., Nicolls, F., Boje, E.: Tracking the cheetah tail using animal-borne cameras, GPS, and an IMU. *IEEE Sens. Lett.* **1**(4), 1–4 (2017)
- Birn-Jeffery, A.V., Higham, T.E.: Geckos decouple fore- and hind limb kinematics in response to changes in incline. *Front. Zool.* **13**(1), 11 (2016)
- Walker, V., Tranquille, C., Newton, J., Dyson, S., Brandham, J., Northrop, A., Murray, R.: Comparison of limb kinematics between collected and lengthened (medium/extended) trot in two groups of dressage horses on two different surfaces. *Equine Vet. J.* **49**(5), 673–680 (2017)
- Torres, B.T., Fu, Y.C., Sandberg, G.S., Budsberg, S.C.: Pelvic limb kinematics in the dog with and without a stifle orthosis. *Vet. Surg.* **46**(5), 642–652 (2017)
- Santos-Santos, J.H., Culbert, B.M., Standen, E.M.: Kinematic performance and muscle activation patterns during post-freeze locomotion in wood frogs (*Rana sylvatica*). *Can. J. Zool.* **96**, 728–738 (2018)
- Gohberg, I., Goldberg, S., Kaashoek, M.A.: Classes of Linear Operators. Birkhäuser Basel (1990). <https://doi.org/10.1007/978-3-0348-7509-7>
- Elgammal, A., Lee, C.S.: The role of manifold learning in human motion analysis. In: Rosenhahn, B., Klette, R., Metaxas, D. (eds.) *Human Motion*, pp. 25–56. Springer, Berlin (2008)
- Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**(5500), 2323–2326 (2000)
- Tenenbaum, J.B., De Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. *Science* **290**(5500), 2319–2323 (2000)
- Blackburn, J., Ribeiro, E.: Human motion recognition using isomap and dynamic time warping. In: Elgammal, A., Rosenhahn, B., Klette, R. (eds.) *Human Motion—Understanding, Modeling, Capture and Animation*, pp. 285–298. Springer, Berlin (2007)
- Lawrence, N.D.: Gaussian process latent variable models for visualisation of high dimensional data. In: *Advances in Neural Information Processing Systems*, vol. 17, pp. 329–336 (2004)
- Ek, C.H., Torr, P.H., Lawrence, N.D.: Gaussian process latent variable models for human pose estimation. In: *International Workshop on Machine Learning for Multimodal Interaction*, pp. 132–143. Springer (2007)
- Fan, G., Zhang, X., Ding, M.: Gaussian process for human motion modeling: a comparative study. In: *2011 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6. IEEE (2011)
- Wang, J.M., Fleet, D.J., Hertzmann, A.: Gaussian process dynamical models for human motion. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(2), 283–298 (2008)
- Wang, J., Hertzmann, A., Fleet, D.J.: Gaussian process dynamical models. In: *Advances in Neural Information Processing Systems 18*, pp. 1441–1448. MIT Press, Cambridge (2006)
- Urtasun, R., Fleet, D.J., Fua, P.: 3D people tracking with Gaussian process dynamical models. In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 238–245. IEEE (2006)
- Kurdila, H.R.: Gappy pod and temporal correspondence for lizard motion estimation. Ph.D. thesis, Virginia Tech (2018)
- Rasmussen, C.E.: Gaussian processes in machine learning. In: Bousquet, O., von Luxburg, U., Rätsch, G. (eds.) *Advanced Lectures on Machine Learning. ML 2003. Lecture Notes in Computer Science*, vol. 2840. Springer, Berlin (2003)

- ture Notes in Computer Science, vol. 3176, pp. 63–71. Springer, Berlin, Heidelberg (2004). https://doi.org/10.1007/978-3-540-28650-9_4
24. Hairer, E., Lubich, C., Wanner, G.: Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations. Springer Science & Business Media, Berlin (2006)
 25. Bornemann, F.: Homogenization in Time of Singularly Perturbed Mechanical Systems. Springer, Berlin (2006)
 26. De Jalon, J.G., Bayo, E.: Kinematic and Dynamic Simulation of Multibody Systems: The Real-Time Challenge. Springer Science & Business Media, Berlin (2012)
 27. Rubin, H., Ungar, P.: Motion under a strong constraining force. *Commun. Pure Appl. Math.* **10**(1), 65–87 (1957)
 28. Kurdila, A.J., Narcowich, F.J.: Sufficient conditions for penalty formulation methods in analytical dynamics. *Comput. Mech.* **12**(1–2), 81–96 (1993)
 29. Lubich, C., Weiss, D.: Numerical integrators for motion under a strong constraining force. *Multisc. Model. Simul.* **12**(4), 1592–1606 (2014)
 30. Bender, M.J.: Modeling and estimation of bat flight for learning robotic joint geometry from potential fields. Ph.D. thesis, Virginia Tech (2018)
 31. Hedrick, T.L.: Software techniques for two-and three-dimensional kinematic measurements of biological and biomimetic systems. *Bioinspir. Biomim.* **3**(3), 034001 (2008)
 32. Abraham, R., Marsden, J.E., Marsden, J.E.: Foundations of Mechanics, vol. 36. Benjamin/Cummings Publishing Company, Reading, MA (1978)
 33. Bullo, F., Lewis, A.D.: Geometric Control of Mechanical Systems: Modeling, Analysis, and Design for Simple Mechanical Control Systems, vol. 49. Springer Science & Business Media, Berlin (2004)
 34. Monforte, J.C.: Geometric, Control and Numerical Aspects of Nonholonomic Systems. Springer, Berlin (2004)
 35. DeVore, R., Kerkyacharian, G., Picard, D., Temlyakov, V.: Approximation methods for supervised learning. *Found. Comput. Math.* **6**(1), 3–58 (2006)
 36. De Vito, E., Rosasco, L., Toigo, A.: Learning sets with separating kernels. *Appl. Comput. Harmon. Anal.* **37**(2), 185–217 (2014)
 37. Wendland, H.: Scattered Data Approximation, vol. 17. Cambridge University Press, Cambridge (2004)
 38. Smale, S., Zhou, D.X.: Learning theory estimates via integral operators and their approximations. *Constr. Approx.* **26**(2), 153–172 (2007)
 39. Bobade, P., Majumdar, S., Pereira, S., Kurdila, A.J., Ferris, J.B.: Adaptive estimation in reproducing kernel hilbert spaces. In: 2017 American Control Conference (ACC), pp. 5678–5683 (2017). <https://doi.org/10.23919/ACC.2017.7963839>
 40. Weidmann, J.: Linear Operators in Hilbert Spaces, vol. 68. Springer Science & Business Media, Berlin (2012)
 41. Dunford, N., Schwartz, J.T.: Linear Operators Part I: General Theory, vol. 7. Interscience publishers, New York (1958)
 42. Vapnik, V.: The Nature of Statistical Learning Theory. Springer Science & Business Media, Berlin (2013)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.