Nonlinear Dyn
https://doi.org/10.1007/s11071-018-4466-9

ORIGINAL PAPER

Geometrically optimal gaits: a data-driven approach

Brian Bittner · **Ross L. Hatton** · **Shai Revzen**

Received: 28 December 2017 / Accepted: 3 July 2018

Abstract The study of optimal motion of animals or robots often involves seeking optimality over a space of cyclic shape changes, or *gaits*, specified using a large number of parameters. We show a data-driven method for computing the gradient of a cost functional with respect to a large number of gait parameters by employing geometric properties of the dynamics to efficiently construct a local model of the system, and then using this model to rapidly compute the gradients. Our modeling step specifically applies to systems governed by connection-like models from geometric mechanics, which encompass a number of high-friction regimes. We demonstrate using our method for optimizing gaits under noisy, experiment-like conditions by simulating planar multi-segment serpent-like swimmers in a low Reynolds number (viscous friction) environment. Our optimization results recover known results for 3-segment swimmers with a 66 dimensional gait parameterization, and extend to optimizing the motion of a 9 segment swimmer with a 264 dimensional gait space, using only 30 simulation trials of 30 gait cycles each. The data-driven geometric gait optimization approach we present is designed to operate on noisy, stochastically perturbed dynamics—as noisy and variable as experimental data—and efficiently optimize a large number of parameters. We believe this approach has the potential to significantly advance our ability to optimize robot gaits with hardware in the loop and to study the optimality of animal gaits with respect to hypothesized cost functions.

Keywords Gait optimization · Locomotion · Geometric mechanics · Oscillator · Data-driven floquet analysis

RLH thanks the National Science Foundation for support under Civil, Mechanical and Manufacturing Innovation grant 1653220. SR and BB were funded by Army Research Office grant W911NF-14-1-0573 and the Rackham Merit Fellowship.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11071-018-4466-9) contains supplementary material, which is available to authorized users.

B. Bittner (✉)
Robotics Department, University of Michigan, Ann Arbor, USA
e-mail: babitt@umich.edu

R. L. Hatton
Collaborative Robotics and Intelligent Systems (CoRIS) Institute and School of Mechanical, Industrial, and Manufacturing Engineering, Oregon State University, Corvallis, USA
e-mail: ross.hatton@oregonstate.edu

S. Revzen
Electrical Engineering and Computer Science Department, Ecology and Evolutionary Biology Department, Robotics Institute, University of Michigan, Ann Arbor, USA
e-mail: shrevzen@umich.edu

1 Introduction

The ability to move effectively through the environment is both a defining property of animals and a

Published online: 26 July 2018

highly desirable capability for man-made systems such as robots and vehicles. Locomotion (aquatic, terrestrial, and aerial) is most commonly achieved by having a moving body change shape in a way that produces reaction forces from the environment; these reaction forces in turn propel the body. A key question in both robotics and animal research is thus "does a given gait cycle optimally exploit this propulsive relationship, and if not, what changes to the gait would improve its performance?"

This paper details a new approach to answering these questions, by presenting a practical extension of geometric gait optimization theory that incorporates techniques from the data-driven modeling of gaits as oscillators. By efficiently producing a local geometric mechanics model of the observed motion, we can then employ this model to rapidly evaluate the gradient of a goal function with respect to gait parameters. Because this performance simulation is very fast, the number of gait parameters being optimized can be so large that estimating such a gradient by direct experimentation is nigh impossible; below we give an example with 264 parameters optimized in 30 trials of 30 cycles each.

The framework presented in this paper is made possible by combining work in two fields that have developed largely in parallel. In the field of geometric mechanics, Hatton has developed a framework for characterizing gait efficiency in terms of the length and area of the cycle in the shape space [8,11,13,23,24]. Applying these principles to systems that lack an analytical model remains an open area of investigation, especially when high-dimensionality makes the exhaustive exploration of system dynamics from [5,12] infeasible, or when considering an animal whose motions we cannot directly command.

In the field of oscillator theory, Revzen developed a set of tools for extracting oscillator-like motion models from noisy and irregularly spaced data [26,28]. In addition to the method's robustness to the intrinsic system noise of biological and physical systems, it extends well to high dimensional shape spaces. A limitation, however, is the lack of insight that these models provide for gait improvements.

Applying the data-driven oscillator and geometric approaches together enhances their respective capabilities: the data-driven oscillator tools can provide the geometric models with the specific information needed for evaluating a performance criterion and its gradient, improving their predictive power relative to the quantity of available data. Conversely, by viewing the system as a mechanical connection (as a opposed to a general second-order dynamical system) the data-driven oscillator models can ignore certain aspects of the system dynamics that are irrelevant to the optimality of the gait, thus significantly reducing the algorithmic complexity of model extraction.

Here, we lay out a framework for combining the geometric insights from Hatton's work with the data-driven oscillator model construction from Revzen's work. Our combined approach uses noise in the dynamics of a system that follows a nominal gait cycle to build a model of the system dynamics in the neighborhood of this gait cycle. Inserting this model into our geometric tools then provides estimates of both how optimal this gait is relative to nearby cycles and what perturbations can be applied to the cycle to best improve its performance. This estimation technique has two primary use cases.

- The first is as a tool that allows for verification of postulated goal functions for observed animal locomotion.
- The second is in field robotics, where an efficient, noise resistant gait optimization algorithm can potentially enable learning effective gaits without requiring precise analytical models of the robot or its interactions with the environment.

2 Contributions

The approach presented in this paper offers a collection of advantages in speed, scalability, and model reduction for the estimation of motion models and subsequent optimization of gaits. These advantages derive from the use of geometric mechanics models governed by a connection (the "principal kinematic case" in the language of [21]). The absence of a momentum term in the equation of motion implies that the contribution of different segments of motion do not strongly depend on each other, allowing motion models to be integrated in parallel instead of sequentially in time. With multiprocessor computing becoming cheaper, this offers the opportunity for dramatic speedups in the computation of motion plans. Additionally, connections expose the fact that the systems they govern are, for practical purposes, half the dimension of general mechanical systems. As the very name "geometric mechanics" suggests, in these systems, the geometry of motion in body

shape space governs the outcome of motions, admitting a description with only one dimension per degree of freedom, instead of the two needed in conventional Newtonian mechanics. Despite this great promise of geometric mechanics models, little work has been done on producing them in a data-driven way. Developing a tool for the data-driven creation of such models allows us to explore their value for both scientific and engineering applications.

Our most significant contribution is the demonstration of an algorithm for producing data-driven connection models in the neighborhood of a (noisy) rhythmic motion. Because it estimates linear models as a function of phase, our algorithm requires a number of cycles proportional to the state-space dimension, rather than being exponential in the state-space dimension. No previous method for data-driven connection estimation has less than exponential data requirements.

Our second contribution is the demonstration of a gradient ascent algorithm that can utilize the data-driven connection models we identify to optimize a goal function. Our algorithm constructs a local model and estimates the gradient with respect to this fast-to-integrate connection model. It then steps within the volume supported by the local model it used. The power of our algorithm is that it replaces gradient estimation against the full model or physical system with gradient estimation against a fast, easy to parallelize model evaluation. This allows policies with orders of magnitude more parameters to be optimized. In our example, a 264 parameter policy, optimized with 30 trials of 30 movement cycles each, finding strong solutions from a broad collection of seeded initial gaits.

Below we review the geometric and data-driven approaches, and then synthesize them into a tool for simultaneously estimating and optimizing locomotion models. Using simulated mechanical swimming platforms, we illustrate the precision of these data-driven geometric mechanics models and demonstrate that optimal gaits can be learned with very few trials. Finally, we discuss the utility of the new methods for both system identification and field robotics.

3 Geometry of locomotion

The first thread of prior work that this paper draws upon is geometric modeling of locomotion. When analyzing a mobile deformable system, it is convenient to separate its configuration space Q (i.e., the space of its generalized coordinates q) into a position space G and a shape space R, such that the position $g \in G$ locates the system in the world, and the shape $r \in R$ gives the relative arrangement of the particles that compose it.[1]

During locomotion, changes in the system's shape provoke reaction forces from the environment that in turn drive changes in the system's position. For the purposes of this paper, we adopt a (geometric) locomotion model

$$\overset{\circ}{g} = \mathbf{A}(r)\dot{r}, \tag{1}$$

where $\mathbf{A}$, the *local connection*, linearly maps the shape velocity $\dot{r}$ to the body velocity $\overset{\circ}{g} = g^{-1}\dot{g}$ (i.e., the position velocity in the body frame's current forward, lateral, and rotational directions). The local connection acts similarly to the Jacobian map of a kinematic mechanism—it takes the velocity of joints to the position velocity (here, of the body frame instead of end effector) that they generate under the constraints imposed on the system.

We model the cost of changing shape as corresponding to the length s of the trajectory through the shape space,

$$s = \int \sqrt{\mathrm{d}r^{\mathsf{T}}\mathcal{M}(r)\,\mathrm{d}r} = \int_0^T \sqrt{\dot{r}^{\mathsf{T}}\mathcal{M}(r)\,\dot{r}}\,\mathrm{d}t, \tag{2}$$

where $\mathcal{M}$ is a Riemannian metric on the shape space that weights the costs of changing shape in different directions.

This connection-and-metric model applies to systems that move by pushing directly against their environment with negligible accumulated momentum in "gliding" modes, and whose energetic costs are dominated by internal or external dissipative effects. This model has been analytically derived for swimmers in viscous fluids [1,13], and experimentally validated for several robots in dry granular media [5,12,19].

The meaning of the cost encoded by the metric $\mathcal{M}$ depends on the system physics, but at a high level it can be considered as the time it will take the system to execute the motion given a unit power budget. For systems

[1] In the parlance of geometric mechanics, this assigns Q the structure of a (trivial, principal) *fiber bundle*, with G the *fiber space* and R the *base space*.

moving in dry-friction environments, s can be specifically taken the energy dissipated while executing the motion [5]; for the viscous friction model we consider in this paper, s is the time-integral of the square root of power dissipated [13].

3.1 Extremal and optimal gaits

Locomoting systems typically move by repeatedly executing *gaits*—cyclic changes in shape that produce characteristic net displacements in position. Such cycles can be chained together to produce larger motions through the world.

Geometrically, a gait θ is a cyclic trajectory through the shape space with period T,

$$\theta : [0, T] \to R \tag{3}$$

$$\theta(0) = \theta(T), \tag{4}$$

and the system shape at any time t while executing the gait is $r = \theta(t)$.

Under the locomotion model in (1), the net displacement over one cycle of a gait is equal to the path integral of the local connection $\mathbf{A}$ over that trajectory. By an extension of Stokes' theorem, this displacement can be approximated[2] as the integral of the *curvature* of $\mathbf{A}$ over a surface θ_a bounded by the gait,

$$g_\theta = \oint_\theta g\mathbf{A}(r)\,\mathrm{d}r \approx \iint_{\theta_a} \overbrace{\mathbf{d}\mathbf{A} + \textstyle\sum[\mathbf{A}_i, \mathbf{A}_{j>i}]}^{\text{curvature } D\mathbf{A}}. \tag{5}$$

The curvature $D\mathbf{A}$ (formally, the total Lie bracket or covariant exterior derivative of $\mathbf{A}$ [11]) measures how much the coupling between shape and position motions changes across the cycle, and thus how much displacement the system can extract from a cyclic motion. Its components $\mathbf{d}\mathbf{A}$ and $[\mathbf{A}_i, \mathbf{A}_j]$ are the exterior derivative (curl) and local Lie bracket of the system constraints and, respectively, capture the net forward-minus-backward motion and parallel-parking motion available to the system. They are calculated as

$$\mathbf{d}\mathbf{A} = \sum_{j>i} \left(\frac{\partial \mathbf{A}_j}{\partial r_i} - \frac{\partial \mathbf{A}_i}{\partial r_j} \right) \mathrm{d}r_i \wedge \mathrm{d}r_j \tag{6}$$

and

$$[\mathbf{A}_i, \mathbf{A}_j] = g^{-1} \left(\frac{\partial (g\mathbf{A}_j)}{\partial g}\mathbf{A}_i - \frac{\partial (g\mathbf{A}_i)}{\partial g}\mathbf{A}_j \right) \mathrm{d}r_i \wedge \mathrm{d}r_j \tag{7}$$

$$= \begin{bmatrix} \mathbf{A}_i^y \mathbf{A}_j^\theta - \mathbf{A}_j^y \mathbf{A}_i^\theta \\ \mathbf{A}_j^x \mathbf{A}_i^\theta - \mathbf{A}_i^x \mathbf{A}_j^\theta \\ 0 \end{bmatrix} \mathrm{d}r_i \wedge \mathrm{d}r_j, \tag{8}$$

where the wedge product $\mathrm{d}r_i \wedge \mathrm{d}r_j$ is the basis area spanned by the ith and jth basis vectors.

For systems with two shape variables, $\mathbf{d}\mathbf{A}$ and $[\mathbf{A}_i, \mathbf{A}_j]$ have only a single component (on the $\mathrm{d}r_1 \wedge \mathrm{d}r_2$ plane), and (5) reduces to a simple area integral whose integrand is the magnitude of $D\mathbf{A}$. Extremal gaits for these systems (maximizing net displacement per cycle) lie along zero-contours of $D\mathbf{A}$, maximizing the sign-definite region they enclose.

As a general rule, these extremal gaits are more interesting mathematically than as a motion for a robot or animal to follow. With the exception of sports such as basketball that explicitly count steps, displacement-per-cycle is not a useful quantity to optimize, as it leads to wasting time or energy eking out all the available displacement in the cycle, instead of executing smaller but more productive cycles more times. When considering the optimality of a gait, it is thus typically more useful to measure its efficiency by dividing the displacement the gait induces over each cycle by the effort or time required to execute it.

In our model, we take the efficiency γ as the ratio between net displacement g_θ it induces and the path-length cost s calculated in (2), $\gamma := \frac{g_\theta}{s}$. The path-length cost, s, which in the viscous case determines the energy dissipated over a cycle under optimal pacing, can be computed independently of the pacing of the gait. This enables γ to represent the proper notion of efficiency for such systems [13]. Note that maximizing this efficiency is equivalent to maximizing speed at a given power (or minimizing power for a given desired speed), and so gaits with this property are always the most desirable for effective locomotion, even when the goal is "move fast" instead of "move efficiently."

As discussed in [13], optimally-efficient gaits are contracted versions of extremal gaits: they give up low-

[2] The quality of this approximation depends on the choice of body frame for the system, which can be optimally selected once $\mathbf{A}$ is calculated in an arbitrary convenient frame. See [8,9,11] for further discussion of this point.

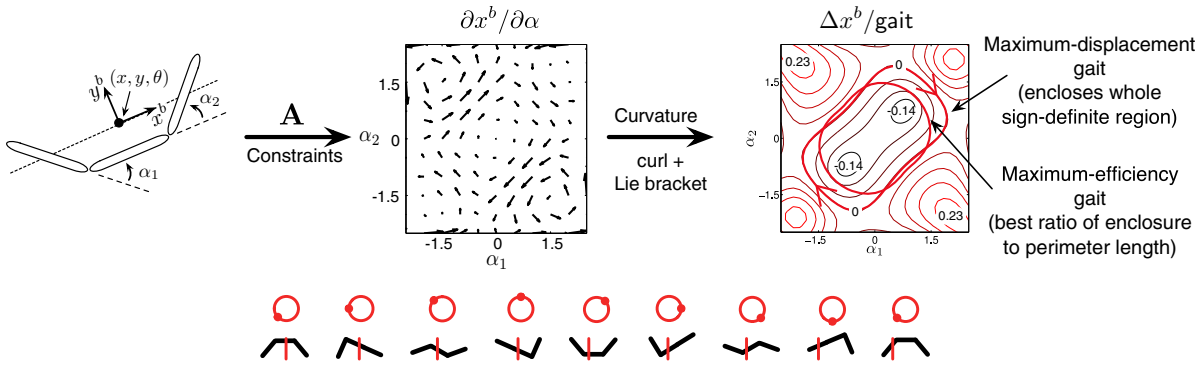

Fig. 1 Key elements of our geometric paradigm, drawn from to [8,13]. Given a locomoting system (left), the system dynamics and constraints produce a relationship between changes in shape and changes in position (center). When the system executes a cyclic change in shape (bottom), the net displacement induced by this gait corresponds to how much curvature of the constraints the gait encompasses, and the time-effort cost of executing this gait is the length of the path it traces out in the shape space (right). In the bottom animation, the top row indicates the phase of the swimmer in a gait, with the shape at each phase shown in the bottom row. The motion can be viewed from left-to-right or right-to-left, with the vertical bar (red) serving as a static reference point. (Color figure online)

yield regions of $D\mathbf{A}$ in exchange for a shorter path length, and thus, a smaller expenditure of power or time. These gaits lie along curves where the gradient of efficiency with respect to changes in the gait parameters p,

$$\nabla_p \gamma = \nabla_p \frac{g_\theta}{s} = \frac{1}{s}\nabla_p g_\theta - \frac{g_\theta}{s^2}\nabla_p s \tag{9}$$

is equal to zero. As further discussed in [23,24], the gradient terms in (9) can be expanded in terms of $D\mathbf{A}$, $\mathcal{M}$ and $\nabla\mathcal{M}$ evaluated along the gait. Given these expansions, the dynamics of optimizing γ resemble those seen in a soap bubble, with an inflating pressure provided by $D\mathbf{A}$ balanced against a surface tension corresponding to s.

3.2 Empirical geometric models

The geometric approach described above was originally developed for systems whose equations of motion can be shown from first principles to have the form in (1). Building on these results, we demonstrated [5,12] that the constraint curvature $D\mathbf{A}$ is also a useful tool for understanding the locomotion of systems whose dynamics are less "clean," and are only tractable through numerical modeling or empirical observation.

In these previous works, we first used nonlinear models [12] or experimental measurements [5] to sample the relationship between $\mathring{g}$ and $\dot{r}$ across the tangent bundle TR. We fit a linear form to this relationship on a grid of tangent space base-points $T_r R$, giving $\mathbf{A}$ on a sampling of the shape space, from which we then calculated the components of $D\mathbf{A}$ as per (6) and (7). Plotting the curvature over the shape space then allowed us to directly identify effective gaits for translation and rotation for three-link and serpenoid system geometries, following the procedure illustrated in Fig. 1.

4 Oscillators and data-driven modeling

The second thread of prior work that this paper draws upon is a robust theory of gaits as oscillators, combined with a statistical approach to data-driven model construction. For an observed physical system, it is not always known a priori what this limit cycle is, what the dynamics of attraction to the limit cycle are, or even what the precise period of oscillation is. These properties of the gait can be extracted using techniques of Data-Driven Floquet Analysis (DDFA) [26,29], the key elements of which we review below.

For simplicity of exposition, we will assume all observations come in a single regularly sampled time-series consisting of (g_n, r_n) position and shape samples, which can be numerically differentiated (e.g., with a second-order Kalman smoother [25,30]) to augment the samples with velocities $\dot{g}_n$, $\dot{r}_n$, and $\mathring{g}_n = g_n^{-1}\dot{g}_n$. From oscillator theory [7,29], we know that every exponentially stable oscillator (which we assume this

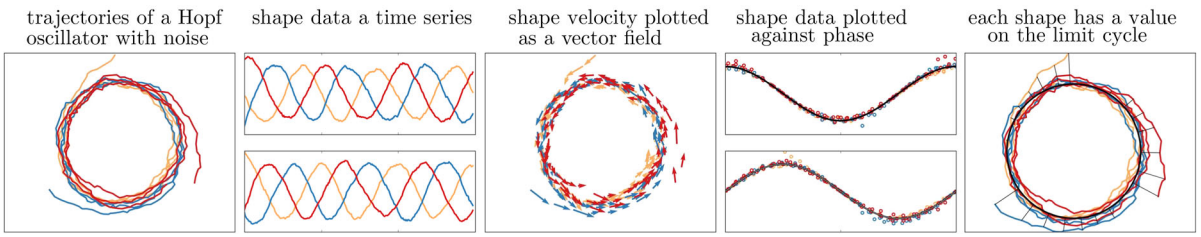

Fig. 2 Data-Driven Floquet Analysis applied to a Hopf oscillator with system noise. Trajectories of the oscillator converge to a (noisy) cycle (left plots, three colors, one per trajectory). This cycle appears as a circle in state-space (extreme left) and as sinusoidal time-series (second from left). By differentiation, we obtain vector field samples at the data points (middle). We estimate the limit cycle as a function of phase (second from right) computed using the phase estimator from [28] providing a canonical map from every trajectory point to a point with identical phase on the limit cycle (thin black lines, right plots). Such surfaces of constant phase—*isochrons*—form radial lines in the (particularly simple) case of the Hopf oscillator. (Color figure online)

to be) can be parameterized with a phase coordinate $\varphi : R \to [0, T) \subset \mathbb{R}$ based on the following rules:

1. Each point on the limit cycle has a unique phase value, spaced such that limit cycle trajectories advance in phase at rate $\dot{\varphi} = 1$.
2. Each point not on the limit cycle inherits its phase value from a corresponding point on the limit cycle, selected such that trajectories starting at the two points ultimately converge. The set of all points sharing a value of φ are called an *isochron* of the oscillator, and the trajectories of the oscillator advance across isochrons such that $\dot{\varphi} = 1$ everywhere.

Our modeling process was as follows: we assigned each sample n a phase φ_n via a *phase estimator* such as that presented in [27], which takes the multivariate time-series of measured data from an oscillator and gives a phase estimate for every data point. Figure 2 provides a visual example.

Once we grouped the samples by phase, we modeled the limit cycle (nominal-gait-as-executed) by computing a pair of Fourier series θ_0 and ω with respect to the phase: $\theta_0(\varphi_n) \approx r_n$ was fitted to the sampled shapes, and $\omega(\varphi_n) \approx \dot{r}_n$ was fitted to the shape velocities. Because each of θ_0 and ω is computed from its own noisy dataset, the condition $\dot{\theta}_0 = \omega$ need not be satisfied after this fitting procedure. We create a self-consistent model θ of the limit cycle by producing the analytical integral of ω, and using a matched filter to combine this integral with the θ_0 estimate to obtain a single self-consistent cyclic trajectory. Past experience [26] has shown that this estimation procedure provides a better representation of the limit cycle than the directly fitted shape model θ_0.

5 Data-driven modeling of the connection

The gait analysis methods described in Sect. 3 provide a powerful link between gaits' optimality and their geometry. Their utility, however, depends on having a model for how small shape changes induce body motion changes. For systems that experience complex interactions with their environments, such models are not readily available from first principles [even if their net effect can be modeled as the linear relationship in (1)], and exhaustive empirical evaluations [5] become infeasible as we move to system with many shape variables and/or limited control affordances.

Conversely, the data-driven methods described in Sect. 4 are able to extract a meaningful model of a system from noisy measurements. This model, however, is limited to the specific gait being executed and does not provide context for comparing the gait against other motions the system could execute, or for optimizing the motion.

Our key innovation in this paper is based on the observation that the data-driven modeling approach can allow us to quickly build up a first-order model of the connection in a tube around a given gait cycle. In turn, the first-order model allows us to rapidly compute the influence of any gait change within the model's domain of validity. Such computations allow us to numerically approximate, at this given gait cycle, the gradient of any goal function computed from a gait with respect to any

parameterization of gaits—even when this parameterization is fairly high dimensional and requires a great many "simulations" of gaits.

In this innovation, we exploit two properties of the geometric model: (1) the variational optimizer/definition of optimality described in (9) only needs to know $D\mathbf{A}$ along the given gait to identify the direction in which that gait can be perturbed to best improve performance. (2) $D\mathbf{A}$, being a two-form and thus a linear map, can be reconstructed at every point along a gait cycle using regressions applied to the relationship between g and r collected from experiments.

5.1 Analytic approximation of the connection near a gait

In this section, we introduce an approximation of the mechanical connection and the cost metric, both centered about a nominal gait. We then construct a procedure to estimate the local model elements from data. As discussed in Sect. 4, a gait cycle $\theta(\cdot)$ can be extracted from shape data r via Data-Driven Floquet Analysis. Perturbations from this phase-averaged behavior are written as $\delta(t) := r(t) - \theta(t)$. These terms can be used to construct a first-order approximation of $\mathbf{A}(\cdot)$ in a neighborhood of the point-set $\operatorname{Im}\theta$ using its Taylor series,

$$\mathbf{A}_i^k(r)\dot{r}^i = \mathbf{A}_i^k(\theta+\delta)\dot{r}^i \approx \left[\mathbf{A}_i^k(\theta) + \frac{\partial \mathbf{A}_i^k}{\partial r_j}(\theta)\delta^j\right]\dot{r}^i, \tag{10}$$

where, as per Einstein index notation, $\mathbf{A}_i^k$ corresponds to the element in the k-th row and i-th column of $\mathbf{A}$. Including the derivative of the connection across the shape space allows us to estimate the connection for behaviors that are not on the current gait cycle $\theta(\cdot)$, and that thus provide velocity samples in nearby, but not identical, tangent spaces of the shape space.

It is important to stress that $\frac{\partial \mathbf{A}^k}{\partial r}$ is not simply the Hessian matrix of g^k with respect to r around points on the gait, i.e., it is not the gradient of a gradient. The Hessian could only be computed if g were a function of r, but it is not. In fact, locomotion via gaits would be impossible if it were such a function since a cyclical change in r could not induce a net change in g. In particular, Hessians are symmetric operators, and the difference term that appears when calculating $\mathbf{dA}^k$ in (6) directly measures the system's ability to locomote along the k-th direction in terms of the asymmetry of $\frac{\partial \mathbf{A}^k}{\partial r}$. Similarly, the $[\mathbf{A}_i, \mathbf{A}_j]$ term from (7) measures the covariant asymmetry of $\frac{\partial g\mathbf{A}}{\partial g}$ when the connection is expanded from local to global coordinates.

5.2 Estimating $\mathbf{A}(\theta)$ and $\mathbf{DA}(\theta)$ from data

Our input data was time series of the system shape r_n, shape velocity $\dot{r}_n$, and observed body velocity $\mathring{g}_n$, at sufficiently many time points $n = 1 \ldots N$. We begin our system identification process by applying the gait extraction algorithm described in Sect. 4, producing Fourier series models of $\theta(\cdot)$ and of $\dot{\theta}(\cdot)$. We then select M evenly spaced values of phase, $\varphi_1 \ldots \varphi_M$, to obtain $\theta_m := \theta(\varphi_m)$ and $\dot{\theta}_m := \dot{\theta}(\varphi_m)$—the shapes and shape velocities of a system that is following the gait cycle precisely. We use these as the points at which we estimate the connection and its derivative.

For each cycle point θ_m, we collect all shapes r_n that are sufficiently close, i.e., n such that $\|r_n - \theta_m\| < \delta_{\max}$. For notational simplicity, when both index n and index m appear in an equation below, we take the values of n to be restricted to only those sufficiently close time-series points. We now define the offset between the shape sample and its current on-gait reference point as $\delta_n := r_n - \theta_m$.

Within each θ_m neighborhood, we now estimate the local connection and its derivatives by using a linear regression to find the slopes of the relationship between $\mathring{g}$, $\dot{r}$, and δ. Naively, this regression is the solution to the Generalized Linear Model formed by placing the Taylor-series expansion of $\mathbf{A}$ from (10) into the locomotion model from (1):

$$\mathring{g}_n^k \sim \left(\mathbf{A}_i^k\right)\dot{r}_n^i + \left(\frac{\partial \mathbf{A}_i^k}{\partial r_j}\right)\delta_n^j \dot{r}_n^i, \tag{11}$$

where $\left(\mathbf{A}_i^k\right)$ are the M separate estimates of $\mathbf{A}_i^k(\theta_m)$ and $\left(\frac{\partial \mathbf{A}_i^k}{\partial r_j}\right)$ are the M separate estimates of $\frac{\partial \mathbf{A}_i^k}{\partial r_j}(\theta_m)$.

When applied to samples generated from an oscillator as illustrated in Fig. 2, this straightforward regression is biased by the shape velocity samples being centered around $\dot{r} = \dot{\theta}_m$ rather than $\dot{r} = 0$. We correct for this bias by re-centering the regression around $\mathbf{A}(\theta_m)\dot{\theta}_m$. We separate the perturbations of the shape

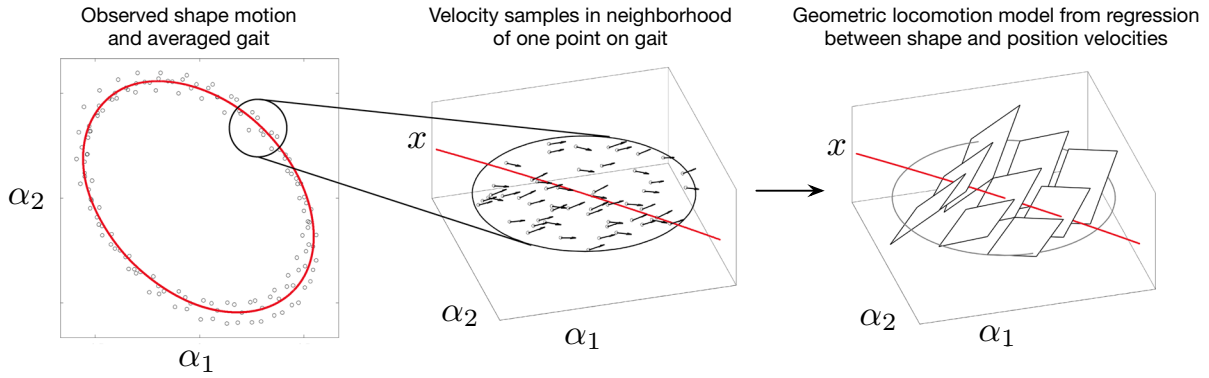

Fig. 3 Illustration of the connection estimation process. We take the rhythmic data, group it by phase, and average using a Fourier series to obtain a periodic gait (left; red cycle). We collect shape velocity and body velocity data (middle; rooted arrows) within the neighborhood of point on the gait (black oval, left; zoomed in area, middle). Using these data, we fit a first-order approximation of the connection model (black planes; right). We repeat this process for a collection of points on the gait cycle at fixed phase intervals, and fit the parameters of the estimated models with a Fourier Series to obtain a model of the connection that smoothly varies with phase. Further detail in Sect. 5.2. (Color figure online)

velocity away from the gait cycle velocity from the influence of the gait cycle velocity itself by defining $\dot{\delta}_n := \dot{r}_n - \dot{\theta}_m$, and re-writing the GLM of (11) as (for velocity component k and each value of m):

$$\overset{\circ}{g}{}^k_n \sim \mathbf{C}^k + \mathbf{B}^k_j \delta^j_n + \left(\mathbf{A}^k_i\right) \dot{\delta}^i_n + \left(\frac{\partial \mathbf{A}^k_i}{\partial r_j}\right) \delta^j_n \dot{\delta}^i_n \tag{12}$$

where $\mathbf{C}^k := \mathbf{A}^k_i \dot{\theta}^i$ is the connection applied to the (unmodified) gait cycle shape velocity, and $\mathbf{B}^k_j := \frac{\partial \mathbf{A}^k_i}{\partial r_j} \dot{\theta}^i$ is the interaction effect of shape offset and shape velocity applied to the (unmodified) gait cycle shape velocity. Here, $\mathbf{C}^k$ is a constant (with k, m fixed); and $\mathbf{B}^k$ is a ("co-")vector that acts on shape offsets from the gait, rather than on conventional tangent vectors. The $\left(\mathbf{A}^k_i\right)$ element is a true co-vector that acts on velocity offsets away from the typical gait velocity, and $\left(\frac{\partial \mathbf{A}^k_i}{\partial r_j}\right)$ is the interaction matrix of shape offsets and shape velocity offsets, offset being taken relative to the nominal gait cycle θ (Fig. 3).

We compute the regression by writing it in matrix form and thereby posing the least-squares problem (for each k and m; indices k and m elided below for clarity):

$$\begin{bmatrix} \overset{\circ}{g}_1 \\ \vdots \\ \overset{\circ}{g}_N \end{bmatrix} = \begin{bmatrix} 1, & \delta_1, & \dot{\delta}_1, & \dot{\delta}_1 \otimes \delta_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1, & \delta_N, & \dot{\delta}_N, & \dot{\delta}_N \otimes \delta_N \end{bmatrix} \cdot \left[\widehat{\mathbf{C}}, \widehat{\mathbf{B}_j}, \widehat{\mathbf{A}_i}, \widehat{\frac{\partial \mathbf{A}_i}{\partial r_j}} \right]^{\mathsf{T}} \tag{13}$$

where $\widehat{\ }$ indicates "estimated" and $\otimes$ is the outer product. For a d dimensional shape space, the row of unknowns on the right consists of $1 + d + d + d^2$ elements.

Once we have the model for every m, we construct a Fourier series model of each of the matrices of the GLM, allowing them to be smoothly interpolated at any phase value.

5.3 Estimating the metric

In the same manner as we estimate $\mathbf{A}$, we can estimate a Riemannian effort-metric $\mathcal{M}$ on the shape space by recording the differential cost of motion $\dot{s}$ along with the system kinematics, and then fitting these costs to a linearized expansion of (2) taken at the points θ_m using the matching n indices,

$$\dot{s}^2_n \sim \dot{r}_n{}^{\mathsf{T}} \left[\mathcal{M} + \left(\frac{\partial \mathcal{M}}{\partial r_j} \right) \delta^j_n \right] \dot{r}_n. \tag{14}$$

This regression suffers from the afore-mentioned bias stemming from the $\dot{r}$ values being centered around $\dot{\theta}$ instead of 0;, so we recenter it in a similar manner as in (13). Additionally, because $\mathcal{M}$ is a symmetric tensor, only $\binom{d}{2}$ elements need to be estimated, reducing by about half the number of quantities to estimate. Details of this regression calculation are in "Appendix A".

5.4 Comparison of estimates to previous work

This process is analogous to the processes we described for empirically estimating $\mathbf{A}$ and its derivatives in [5, 12], but offers some distinct advantages.

In our previous work, the shape velocity samples to identify $\mathbf{A}$ at a point all had to be in the tangent space of that point. Here, we have relaxed that requirement by fitting to a linearized expansion of (1) instead of (1) itself.

Furthermore, since our regression here uses intrinsic noise in the system, it provides an estimate of the average behavior under noise. The average behavior of a system when noise is added depends also on the variance of the noise. In the analysis here, we account for the actual noise present in the system, rather than treating it as mere measurement error of a deterministic system.

The presence of system noise and the form of the linearized expansion allow for collection of data over a singular repeated gait cycle, rather than collection over the whole shape space (as was done for the prior model estimation methods).

5.5 Assumptions for the modeling estimation

We make the following assumptions for modeling: (1) the deterministic part of the systems time evolution is governed by a connection; (2) the dynamics are subject to sufficient IID (independent and identically distributed) system noise to allow them to be identified; (3) noise is sufficiently small to allow a distinct rhythmic motion to be observed and modeled as a limit cycle oscillator representing a gait. For gait optimization, we further assume that the system is fully actuated and able to follow (on average) any trajectories we command.

6 Performance of the data-driven models

To benchmark the accuracy of our data-driven geometric modeling process, we compared its prediction of the body velocity for a test system against three system models that had various levels of knowledge about the "true" system dynamics used in the simulation. The test system had a geometric locomotion model of the form in (1), and its shape trajectories were generated via a noisy oscillator like that illustrated in Fig. 2.

6.1 Reference models

As described in Sect. 5, we used a data-driven process to construct a phase varying first-order model of $\mathbf{A}$ at points θ_m along our observed gait cycle. Each r_n data point from the (noisy) trial was associated with a corresponding (phase-matched) point θ_n on the gait cycle,[3] which allowed us to compare several different models for the body velocity:

1. The *ground truth model*

$$\mathring{g}_{G,n} = \mathbf{A}(r_n)\dot{r}_n, \tag{15}$$

   in which each $(r_n, \dot{r}_n)$ pair is passed directly to the simulator dynamics, giving
2. The fully *data-driven model*, where the regression estimates of the Taylor expansion of $\mathbf{A}$ are used to approximate $\mathbf{A}$ at points off of the gait cycle, and $\mathring{g}_{D,n}$ is given by (12), used with the quantities estimated from (13).
3. An *analytic model*

$$\mathring{g}_{A,n} = \mathbf{A}(\theta_n)\dot{r}_n + \frac{\partial \mathbf{A}}{\partial r}(\theta_n)\delta_n\dot{r}_n \tag{16}$$

   that uses a Taylor-series expansion of the simulator model computed at the same point as the data-driven model, without using any regression or simulation data. This model tests the correctness of the regression in the data-driven model.
4. A *template projection model*

$$\mathring{g}_{T,n} = \mathbf{A}(\theta_n)\dot{\theta}_n. \tag{17}$$

   that projects each $(r_n, \dot{r}_n)$ data point onto its corresponding $(\theta_n, \dot{\theta}_n)$ values for the gait cycle that was used to derive the data-driven model. This approximation tests how much additional information is gained from the higher order term in the Taylor expansion.

Note that the template approximation in (17) can be considered as the leading term of the analytical approximation (after separating $\dot{r}_n$ into $\dot{\theta}_n$ and $\dot{\delta}_n$ components),

[3] These phase-matched θ_n points can be individually computed for each r_n, and so are not restricted to the previously-sampled θ_m values. Similarly, the estimates of $\mathbf{A}$ and its derivative from Sect. 5.1 are computed as Fourier series, and can thus be interpolated to any θ_n.

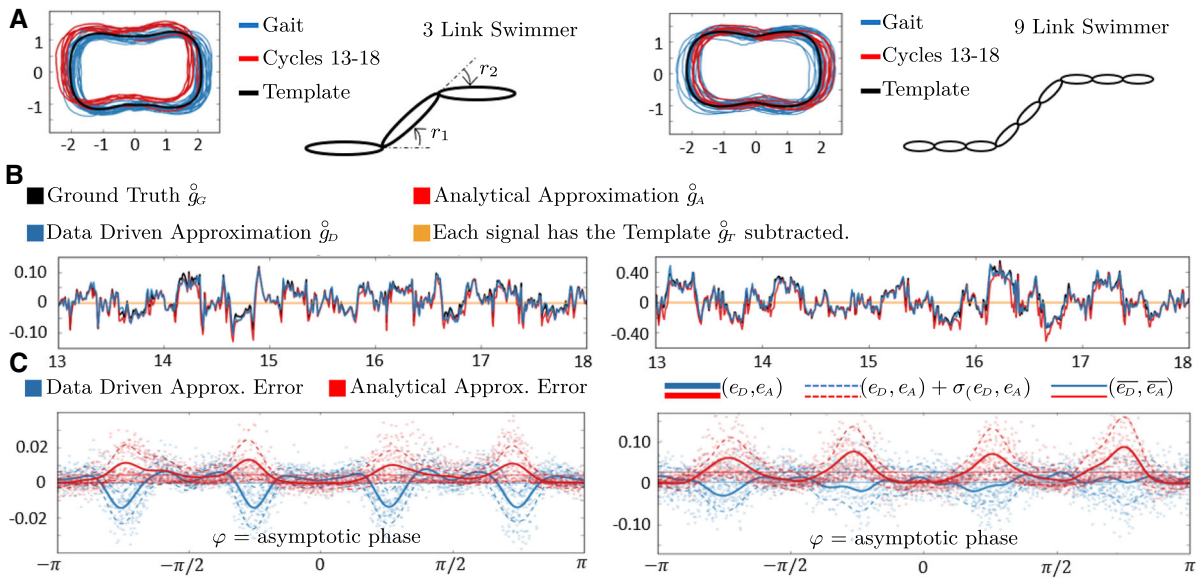

Fig. 4 Comparison of model accuracy for 3 link and 9 link swimmers. **a** We drove each platform to follow the extremal gait for the three-link swimmer (black) generating 30 strokes (blue and red; plotted on first two principal components). Of these, we plotted Cycles 13–18 (red) in the time domain (**b**), showing the additional motion predicted beyond the template model by the ground truth model (black), the data-driven model (teal), and the analytic model (red). Because both analytic and data-driven models follow the ground truth closely, we also plotted a scatter plot of their errors as a function of phase (**c**), showing that the data-driven model (teal) has zero average error, unlike the analytic (red) model. As the number of DOF grows (right; 9 link plots) the mean (solid) and variance (dashed) of the data-driven model (teal) become smaller than those of the analytic model (red). (Color figure online)

and that the partial-derivative terms in (12) and (16) contain the information required to predict the effect of modifying the gait limit cycle.

6.2 Simulation setup: swimming with system noise

For our baseline system model, we used a three-link Purcell swimmer [22] modeled as described in [10]. This system moves through a viscous fluid with linear drag, which we take as having a 2:1 lateral/longitudinal ratio. To demonstrate the ease with which we can extend our approach to systems with higher-dimensional shape spaces, we also considered a nine-link swimmer. Both are pictured in Fig. 4 part A.

To simulate the effects of noise in the shape dynamics (e.g., weak or imprecise shape control), we generated the shape trajectories from sample paths of a (Stratonovich) stochastic differential equation, injected into the shape space:

$$d\varphi = 1\,dt + \eta \circ dW_\theta$$

$$d\delta = -(\alpha\,\delta)\,dt + \eta \circ dW_\delta,$$
$$r(t) := \theta_{\text{REF}}(\varphi(t)) + \delta(t). \tag{18}$$

where $\theta_{\text{REF}}(\cdot)$ was a reference motion we specified as a Fourier series; α was the coefficient of attraction bringing the system back to the reference gait cycle; and η was a noise magnifier for the Weiner processes dW driving both phase noise and shape noise.

For all simulations in this paper $\alpha = 0.05$ and $\eta = 0.025$, chosen based on the superficial similarity the noisy trajectory ensembles have to experimental data we have worked with.

6.3 Model accuracy results

To illustrate the performance of our data-driven models, we examined the differences between motion predicted by the models in Sect. 6.1 when the reference gait was the extremal gait maximizing motion in the x direction, known from [9,33]. We specifically chose an extremal gait as our example because non-extremal

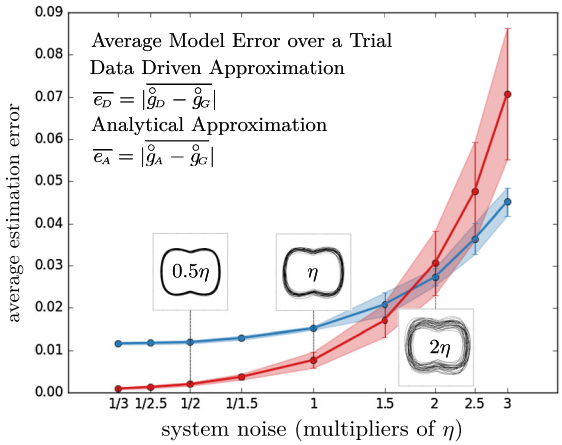

Fig. 5 Comparing analytic and data-driven approximations. Given the same input gait and attraction laws of Fig. 4, we plotted the accuracy of both models (data-driven in red; analytic in teal) over a range of system noise values (0.5η, η and 2η example trajectories in insets) and indicated the range of estimation error observed for x velocity over an ensemble of 20 trials at each noise level. System noise can be seen to strongly degrade the accuracy of the analytic model, whereas the data-driven model retains accuracy at high levels of noise, at the expense of accuracy at low-noise levels. (Color figure online)

gaits should be even easier to model—perturbations around them have first-order effects. These results are shown in Fig. 4.

In high-noise regimes, the data-driven approach yields better models than the analytic Taylor expansion of the dynamics around the gait cycle. This effect is illustrated in Fig. 5, which shows estimation error as a function of noise level for our example gait. The data-driven model also outperforms the analytic model when the system dynamics are very nonlinear, as for the nine-link swimmer at the right of Fig. 4c, and when the system noise is large.

These differences stem from the fact that the analytic model is a linearization of the system dynamics that extrapolates the system dynamics from their values on the gait cycle, whereas the data-driven approximation acts like a secant approximation to a curve, and averages the rate of change of the system dynamics across the neighborhood of the cycle. At the limit of large samples and small noise, the data-driven model approaches the analytic model. Thus, at the limit for many samples and finite noise, the data-driven model should always out-perform the analytic model—giving the best linearization for prediction over the available data, rather than the linearization locally at the gait cycle. However, with finite sample sizes, the accuracy of the estimated linearization can suffer, allowing the analytic model to out-perform the data-driven one.

7 Data-driven geometric gait optimization

Given both the model of the connection (from Sect. 5.2), and the model of the cost metric (from Sect. 5.3), we can evaluate the efficiency of gait cycles in a Sobolev neighborhood of an initial gait cycle. In particular, this allows us to compute the gradient of efficiency [as in (9)] and use a gradient ascent optimization scheme to optimize gait efficiency.

In implementing the gait optimization procedure, two key choices to be made: (1) How is the space of gaits to be represented? (2) How big a step should the optimizer take along the gradient each time it is computed?

7.1 Gait parametrization

For our reference implementation of the gait optimization process, we constructed gaits in which the motion of each shape variable (here, joint angle) is the sum of a set of compactly supported bump functions added to first-order Fourier series. Each shape-space coordinate of the gait $\theta_i(t)$ is thus given, for an order N_o parameterization, as:

$$r_i(t) := c_i + a_i \sin(\Omega t + \phi_i) + \sum_{k=0}^{N_o} u_{i,k}\, w\left(t - k\frac{2\pi}{N_o}\right) \tag{19}$$

$$w(x) := \begin{cases} 1 + \cos(xN_o) & |xN_o| < \pi \\ 0 & |xN_o| \geq \pi \end{cases}, \tag{20}$$

with gait parameters

$$p_i = (c_i, a_i, \phi_i, \{u_{i,k}\}). \tag{21}$$

By construction, only 30 window functions from the sum in Eq. (19) can be non-zero at any time in the gait—making the sum fast to compute, and restricting the influence of each $u_{i,k}$ to only $1/N_o$ of the gait cycle. The expressiveness of this representation in a two-dimensional shape space is illustrated in Fig. 6.

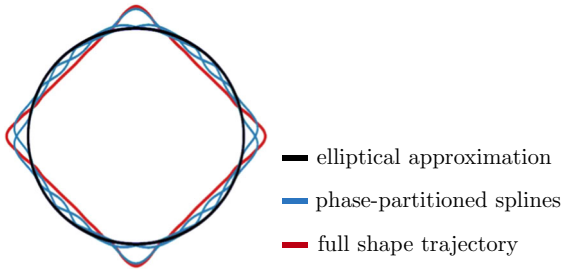

Fig. 6 Illustration of gait parameterization as an ellipse with bump functions. In this parameterization, each gait starts out as an ellipse—the image of a first-order Fourier series (black). To this we added a collection of overlapping, compactly supported, cosine window bumps. The number of bumps is the only order parameter for complexity of the model (here order 30). The sum of the circle and the plotted individual bumps (teal) combine together to give a diamond shaped gait cycle (red). (Color figure online)

7.2 Choosing a step size

Once we have identified an efficiency gradient vector $\hat{p} = \nabla_p \gamma$ on the parameter space, our optimizer must decide how large a step α to take along the gradient. This step size should be informed by the size of the neighborhood around the current gait which was sampled in the most recent trial, which is in turn determined by the level of system noise in the trial (Fig. 7).

To compute the step size, we first measure the noise at each phase bin m as the covariance matrix of off-cycle displacements in the neighborhood of that bin, $\hat{C}_m := \mathbb{E}\left[\delta_n \delta_n{}^{\mathsf{T}}\right]$, which we can interpolate to any phase φ via a Fourier series. We then take:

- θ_0 as the current gait;
- p_0 as the parameters of this gait;
- $p_\alpha = p_0 + \alpha \hat{p}$ as the parameters reached by stepping along the efficiency gradient by α; and
- θ_α as the gait defined by these parameters,

and calculate the Mahalanobis distance with respect to the sampling noise [18] between the updated and current gait cycles as

$$Z(\alpha) := \frac{1}{2\pi} \int_{S^1} (\theta_\alpha - \theta_0)^{\mathsf{T}} C^{-1} (\theta_\alpha - \theta_0)\, d\varphi, \qquad (22)$$

where θ_α, θ_0 and C are all functions of a phase parameter φ which we elide for clarity.

This distance measures the uncertainty of our data-driven modeling process and, generally speaking, grows with α. Presuming this growth to be monotonic,

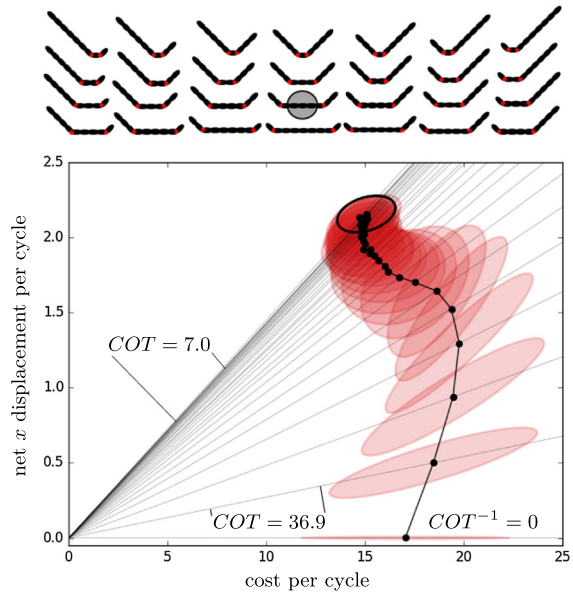

Fig. 7 Optimization is insensitive to initial gait. We provided 28 different initial gaits (cartoons top) each with a different pair of joints (red dots in cartoon) following $r_i(\varphi) = \sin(\varphi)$, with all other joints set to constant angle 0. We optimized each initial gait 3 times, for a total of 84 optimization runs, and plotted the mean (black dots) and covariance ellipsoid (red) of the ensemble of gaits at every simulation iteration on axes of cost and displacement. In these axes, cost of transport (COT) corresponds to a slope. The initial gaits hardly move, giving a distribution along the horizontal axis, which improves to COT 36.9 after one iteration. As optimization progressed, all gaits moved toward the $COT = 7.0$ line, with the final (30th) iteration showing almost no progress and a fairly tight clustering of cost and displacement (black ellipse). Each optimization procedure converged to a serpenoidal motion, although these were not identical and retained some hint of the original choice of active joints. We used the initial gait highlighted (gray circle) for the noise regime testing in Figs. 8 and 9. Our choice of initial gaits was motivated by the following factors. By using reciprocal motions, we ensured (using the Scallop Theorem; see e.g., [17,22]) that all initial gaits have zero net displacement. By having two moving joints—effectively a 3-link swimmer—we ensured that non-reciprocal motions are easily reachable. With only one moving joint, the motion is by definition reciprocal. (Color figure online)

we can use a line search (e.g., a bisection search) to locate (to a user-selectable relative error tolerance; we used 5%) the α value at which $Z(\alpha)$ crosses some threshold value. For the experiments in the next section, we used a threshold constant of 9.5.

The gait optimization framework can be summarized as a gradient ascent algorithm with careful considerations for the parametrization of the gait and step selection. Given an initial parametrization (detailed in

Sect. 7.1), we collect experimental data (30 cycles in our results section) and compute the local motion-and-metric models. We extract a gradient on the efficiency of a motion with respect to the gait parameters by sampling many gaits in the neighborhood of the current policy, using the estimated local model to predict the performance of each sampled gait. We then determine the magnitude of the step size as described in Sect. 7.2. This allows the next gait parametrization to represent a behavior that is reliably informed by the data of the prior trial. Once the next gait is selected, we collect experimental data, repeating the above process. The termination criterion for the gradient ascent algorithm is a pre-specified number of iterations. A more advanced termination criterion will be explored in future work.

8 Swimming gait optimization results

As a demonstration of our gait optimization framework, we applied our algorithm to a 9-link chain "swimmer". All swimming behaviors shown were optimized with respect to the efficiency metric $\gamma = \frac{g_\theta}{s}$, which we report in units of body lengths per unit time at unit power. For any given power budget, this efficiency is inversely proportional to the mechanical cost of transport.

8.1 Optimization is robust to choice of initial condition

One important test of an optimization algorithm is its ability to achieve good outcomes irrespective of initial conditions. To test this ability, we provided the system with gaits in which two selected joints follow identical sinusoidal inputs (no phase offset and amplitude of 1), the other joints attempt to hold at zero angle, and all joints are subjected to noise as discussed in (18). The power costs of these gaits depended on the lengths of the segments between the active joints, and, as illustrated in Fig. 9, as reciprocal motions, they produced no net displacement.

From each of these initial conditions, our optimizer consistently converged (within 30 trials at 30 cycles per trial) to gaits with a cost of transport of 7.0 ± 0.7. As illustrated in Fig. 8, these resulting motions were very close to ellipses embedded in the eight-dimensional shape space, and produced serpenoid undulations traveling along the length of the swimmer. Qualitatively,

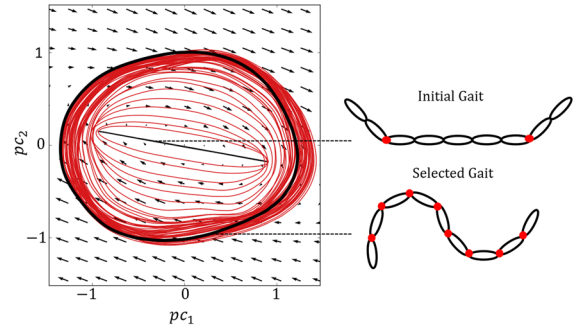

Fig. 8 Visualization of gaits throughout an optimization. We projected all gaits onto the first two principal components of the final gait (viewed as embedded in $\mathbb{R}^8$) and plotted the projection of the x motion connection on that subspace (arrows). The initial gait (top cartoon), allowing only two joints to move (red dots in cartoon swimmer), is a line in the shape space coordinates (black line). The following iterations expand this contour as an ellipse and eventually embellish the ellipse with bumps (red closed ovals) leading to the final gait (black oval) and the serpenoid shape (bottom cartoon). (Color figure online)

the motions are in agreement with the conclusions about optimal swimming behavior in [31], with the exception of maximizing the amplitude of the undulations at the mid-body of the swimmer. In this case, the amplitudes of the discovered gait are typically maximized near the joints that are excited in the initial gait. The reason for the discovery of this family of gaits and their relation to the global optimum in [31] will be the subject of future work.

8.2 Robustness to noise level

A second test of optimizer performance, which is of particular importance to hardware-in-the-loop optimization, is its ability to tolerate a variety of noise levels and produce comparably good results. To demonstrate this ability, we took a single starting gait (in which the active joints are each set two links in from the end, as illustrated in Fig. 8) and optimized its motion under different levels of system noise.

For all four noise regimes tested, the system converged to serpenoidal motions with geometrically similar shapes (similar ratio of wavelength to amplitude), but with different numbers of waves along the body. As illustrated in Fig. 8, the gaits found at different noise levels have similar costs of transport (with mean values ranging from 6.9 to 7.7), but the systems at higher noise levels tended toward gaits that were high-cost/high-displacement, at the expense of some efficiency.

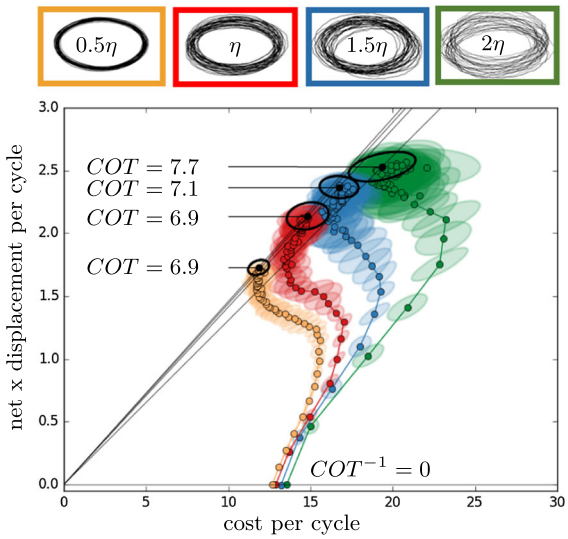

Fig. 9 Course of optimization under different levels of noise. We started with the same initial gait (gray circle highlight in Fig. 7 and top cartoon in Fig. 8), but multiplied the noise level η of Eq. (18) by 0.5, 1., 1.5, 2 (colors yellow, red, teal, and green, respectively). For each noise level, we plotted an example simulation to illustrate the noise level (ovals framed in color; top). We ran 48 optimizations at each noise level, allowing 60 iterations of 30 swimming cycles each, and plotted the mean (circle marker) and covariance (translucent ellipses) of these trials at every iteration of the algorithm, highlighting the final mean (black dot) and covariance (black ellipse). All gaits started unable to move, and reached COT 7.3 ± 0.4 with high-noise optimal gaits being slightly less efficient than low-noise gaits (COT of mean 7.7 vs. 6.9). The two lower noise level achieved indistinguishable cost. It is notable that at higher noises, optimization moved away from the origin, producing larger motions with larger cost. (Color figure online)

Additionally, we note that at all noise levels, the systems initially modified their gait to increase their net displacement, then "pulled left" on the graph to reduce the cost of producing this displacement. The step sizes between trials are smaller on the low-noise systems, as they experience smaller perturbations during the trials, and thus have a lower bound on step size as discussed in Sect. 7.2.

9 Conclusions, limitations, and future work

We have presented two main contributions: (1) a method for locally modeling a connection and a cost metric in the neighborhood of a gait cycle, based solely on the observation of noisy trajectories; (2) an algorithm for gait optimization that employs this method for gradient climbing.

Our modeling relied strongly on system noise to produce sufficient excitations to allow us to employ regression and identify the structure of the dynamics at every phase of the cycle. In this, there is both a strength and a weakness. The strength comes from exploiting noise and being able to model systems with levels of noise comparable to those we have observed in animal and robot data. The weakness comes from relying on noise to be "system" noise—i.e., arising from true changes in the system state rather than from measurement errors. Measurement noise could mask some of the structure we expose by regression. It could also suggest to the optimization to move in a direction that is not achievable by the actual hardware.

The great strength of our gait optimization algorithm is that it decouples the dimension of the gait parameter space from the dimension of the shape space and the number of trials needed. For a shape space of dimension d, order d cycles are needed to identify the model. Once the model is identified for a gait, numerical evaluations of gait perturbations are very quick, and allow the goal function to be differentiated with respect to hundreds of variables with little effort.

Some natural extensions of our work include expanding to a broader class of data-driven models outside those systems which admit connection-like models [2,4,20]. This could enable comparison between analysis of discovered gaits from the data-driven optimizer with geometric analysis on articulated swimming in more complex fluids [14–16]. One natural question which arises is that of systems that are "nearly" geometric—is there a useful and easy to identify notion of "nearly" geometric that translates to good predictive ability of geometric tools?

Improvements to regression, phase estimation, and state estimation could enhance our results even more; in particular, other approaches to gradient ascent step-size choice seem to hold some promise. Elaborating further the relationship between noise level and which gaits are optimal may provide new insights into biological mechanisms of robust locomotion. Expanding to broader notions of asymptotic phase of normally hyperbolic attractors other than limit cycles [6], our methods could perhaps be extended to optimize control policies on more elaborate attractors.

Direct extensions of this work are to test our algorithm on physical hardware-in-the-loop optimization problems in robotics. We have also applied our algorithm to *C. elegans* swimming data kindly provided by

[32], allowing us to asses optimality of low Reynolds number swimming gaits of these animals [3].

Compliance with ethical standards

Conflicts of interest Conflict of Interest: The authors declare that they have no conflict of interest.

Appendix A: Regression for estimating the cost metric

The regression for computation of the metric $\mathcal{M}$ is centered about θ and $\dot{\theta}$, similarly to the construction of the regression for the connection $\mathbf{A}$. The metric approximation now takes the form:

$$\dot{s}_n^2 \sim (\dot{\theta}_n + \dot{\delta}_n{}^{\mathsf{T}}) \left[\mathcal{M} + \left(\frac{\partial \mathcal{M}}{\partial r_j} \right) \delta_n^j \right] (\dot{\theta}_n + \dot{\delta}_n) \tag{23}$$

leading to the regression:

$$\begin{bmatrix} \dot{s}_1^2 \\ \vdots \\ \dot{s}_N^2 \end{bmatrix} = \begin{bmatrix} 1, & \dot{\delta}_1, & \dot{\delta}_1 \hat{\otimes} \dot{\delta}_1, & \delta_1, & \delta_1 \otimes \dot{\delta}_1, & \delta_1 \otimes \dot{\delta}_1 \hat{\otimes} \dot{\delta}_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ 1, & \dot{\delta}_N, & \dot{\delta}_N \hat{\otimes} \dot{\delta}_N, & \delta_N, & \delta_N \otimes \dot{\delta}_N, & \delta_N \otimes \dot{\delta}_N \hat{\otimes} \dot{\delta}_N \end{bmatrix} \cdot R^{\mathsf{T}}, \tag{24}$$

$$R = \left[\widehat{\mathcal{M}_{i,j} \dot{\theta}^i \dot{\theta}^j}, \ \widehat{\mathcal{M}_{i,j} \dot{\theta}^i}, \ \widehat{\mathcal{M}_{i,j}}, \ \widehat{\frac{\partial \mathcal{M}_{i,j}}{\partial r_k} \dot{\theta}^i \dot{\theta}^j}, \ \widehat{\frac{\partial \mathcal{M}_{i,j}}{\partial r_k} \dot{\theta}^j}, \ \widehat{\frac{\partial \mathcal{M}_{i,j}}{\partial r_k}} \right] \tag{25}$$

Here the modified exterior product $\hat{\otimes}$ includes only the upper triangular elements, i.e., $x \hat{\otimes} y = [\ldots, x_i y_j, \ldots]$ s.t. $i \leq j$. Following (14), at each m we solved for $1 + d + \binom{d}{2} + d + d^2 + d\binom{d}{2} \approx \frac{1}{2} d^3$ unknowns to construct our model of the metric.

References

1. Avron, J.E., Raz, O.: A geometric theory of swimming: Purcell's swimmer and its symmetrized cousin. New J. Phys. **9**(437), (2008). https://doi.org/10.1088/1367-2630/10/6/063016
2. Bazzi, S., Shammas, E., Asmar, D., Mason, M.T.: Motion analysis of two-link nonholonomic swimmers. Nonlinear Dyn. **89**(4), 2739–2751 (2017). https://doi.org/10.1007/s11071-017-3622-y
3. Bittner, B., Revzen, S.: What do nematode swimming gaits optimize? In: The Society for Integrative and Comparative Biology Annual Meeting (poster) (2018)
4. Bloch, A.M., Krishnaprasad, P., Marsden, J.E., Murray, R.M.: Nonholonomic mechanical systems with symmetry. Arch. Ration. Mech. Anal. **136**(1), 21–99 (1996). https://doi.org/10.1007/bf02199365
5. Dai, J., Faraji, H., Gong, C., Hatton, RL., Goldman, DI., Choset, H.: Geometric swimming on a granular surface. In: Proceedings of the Robotics: Science and Systems Conference, Ann Arbor, Michigan, https://doi.org/10.15607/rss.2016.xii.012 (2016)
6. Eldering, J., Kvalheim, M., Revzen, S.: Global linearization and fiber bundle structure of invariant manifolds. arXiv preprint arXiv:1711.03646 (2017)
7. Guckenheimer, J., Holmes, P.: Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields. Springer, New York (1983). https://doi.org/10.1007/978-1-4612-1140-2
8. Hatton, R.L., Choset, H.: Geometric motion planning: the local connection, Stokes' theorem, and the importance of coordinate choice. Int. J. Rob. Res. **30**(8), 988–1014 (2011). https://doi.org/10.1177/0278364910394392
9. Hatton, R.L., Choset, H.: Geometric swimming at low and high reynolds numbers. IEEE Trans. Rob. **29**(3), 615–624 (2013a). https://doi.org/10.1109/tro.2013.2251211
10. Hatton, R.L., Choset, H.: Geometric swimming at low and high reynolds numbers. IEEE Trans. Rob. **29**(3), 615–624 (2013b). https://doi.org/10.1109/tro.2013.2251211
11. Hatton, R.L., Choset, H.: Nonconservativity and noncommutativity in locomotion. Eur. Phys. J. Spec. Top. Dyn. Anim. Syst. **224**(17–18), 3141–3174 (2015). https://doi.org/10.1140/epjst/e2015-50085-y
12. Hatton, R.L., Choset, H., Ding, Y., Goldman, D.I.: Geometric visualization of self-propulsion in a complex medium. Phys. Rev. Lett. **110**(078), 101 (2013). https://doi.org/10.1103/physrevlett.110.078101
13. Hatton, R.L., Dear, T., Choset, H.: Kinematic cartography and the efficiency of viscous swimming. IEEE Trans. Rob. (2017). https://doi.org/10.1109/tro.2017.2653810
14. Kanso, E., Newton, P.: Locomotory advantages to flapping out of phase. Exp. Mech. **50**(9), 1367–1372 (2010). https://doi.org/10.1007/s11340-009-9287-9
15. Kanso, E., Marsden, J.E., Rowley, C.W., Melli-Huber, J.B.: Locomotion of articulated bodies in a perfect fluid. J. Nonlinear Sci. **15**(4), 255–289 (2005). https://doi.org/10.1007/s00332-004-0650-9
16. Kelly, SD., Pujari, P., Xiong, H.: Geometric mechanics, dynamics, and control of fishlike swimming in a planar ideal fluid. In: Natural Locomotion in Fluids and on Surfaces, Springer, pp. 101–116, https://doi.org/10.1007/978-1-4614-3997-4_7 (2012)
17. Lauga, E.: Life around the scallop theorem. Soft Matter **7**(7), 3060–3065 (2011). https://doi.org/10.1039/c0sm00953a
18. Mahalanobis, P.: On the generalized distance in statistics. Proc. Natl. Inst. Sci. India **2**(1), 49–55 (1936)
19. McInroe, B., Astley, HC., Gong, C., Kawano, SM., Schiebel, PE., Rieser, JM., Choset, H., Blob, RW., Goldman, DI.: Tail use improves performance on soft substrates in models of early vertebrate land locomotors. Science 353(6295):154–158, https://doi.org/10.1126/science.aaf0984, http://science.sciencemag.org/content/353/6295/154.full.pdf (2016)
20. Ostrowski, J.: Computing reduced equations for robotic systems with constraints and symmetries. IEEE Trans.

Robot. Autom. **15**(1), 111–123 (1999). https://doi.org/10.1109/70.744607
21. Ostrowski, J., Burdick, J.: The geometric mechanics of undulatory robotic locomotion. Int. J. Rob. Res. **17**(7), 683–701 (1998). https://doi.org/10.1177/027836499801700701
22. Purcell, E.M.: Life at low Reynolds numbers. Am. J. Phys. **45**(1), 3–11 (1977). https://doi.org/10.1063/1.30370
23. Ramasamy, S., Hatton, RL. Soap-bubble optimization of gaits. In: The Proceedings of the IEEE Conference on Decision and Control, Las Vegas, NV, https://doi.org/10.1109/cdc.2016.7798407 (2016)
24. Ramasamy, S., Hatton, RL.: Geometric gait optimization beyond two dimensions. In: American Control Conference, https://doi.org/10.23919/acc.2017.7963025 (2017)
25. Rauch, H.E., Tung, F., Striebel, C.T.: Maximum likelihood estimates of linear dynamic systems. AIAA J. **3**(8), 1445–1450 (1965). https://doi.org/10.2514/3.3166
26. Revzen, S.: Neuromechanical control architectures of arthropod locomotion. PhD thesis, University of California, Berkeley (2009)
27. Revzen, S., Guckenheimer, J.M.: Estimating the phase of synchronized oscillators. Phys. Rev. E **78**(5), 051,907 (2008). https://doi.org/10.1103/PhysRevE.78.051907
28. Revzen, S., Guckenheimer, J.M.: Finding the dimension of slow dynamics in a rhythmic system. J. R. Soc. Lond. Interface **9**(70), 957–971 (2012). https://doi.org/10.1098/rsif.2011.0431
29. Revzen, S., Kvalheim, M.: Data driven models of legged locomotion. In: SPIE Defense+ Security, International Society for Optics and Photonics, pp. 94,671V–94,671V, https://doi.org/10.1117/12.2178007 (2015)
30. Roweis, S., Ghahramani, Z.: A unifying review of linear gaussian models. Neural Comput. **11**(2), 305–345 (1999). https://doi.org/10.1162/089976699300016674
31. Shapere, A., Wilczek, F.: Efficiencies of self-propulsion at low reynolds number. J. Fluid Mech. **198**, 587–599 (1989)
32. Sznitman, J., Purohit, P.K., Krajacic, P., Lamitina, T., Arratia, P.E.: Material properties of caenorhabditis elegans swimming at low reynolds number. Biophys. J. **98**(4), 617–626 (2010). https://doi.org/10.1016/j.bpj.2009.11.010
33. Tam, D., Hosoi, A.E.: Optimal stroke patterns for Purcell's three-link swimmer. Phys. Rev. Lett. **98**(6), 068,105 (2007). https://doi.org/10.1103/physrevlett.98.068105