Optimizing snake locomotion in the plane

Silas Alben

Department of Mathematics, University of Michigan, Ann Arbor, MI, USA

We develop a numerical scheme to determine which planar snake motions are optimal for locomotory efficiency, across a wide range of frictional parameter space. For a large coefficient of transverse friction, we show that retrograde traveling waves are optimal. We give an asymptotic analysis showing that the optimal wave amplitude decays as the -1/4 power of the coefficient of transverse friction. This result agrees well with the numerical optima. At the other extreme, zero coefficient of transverse friction, we propose a triangular direct wave which is optimal. Between these two extremes, a variety of complex, locally optimal motions are found. Some of these can be classified as standing waves (or ratcheting motions).
1. Introduction

Snake locomotion has a long history of study by biologists, engineers, and applied mathematicians [1–6]. A lack of appendages makes snake motions simpler in some respects than those of other locomoting animals [7]. However, snakes can move through a wide range of terrestrial [1,8,9], aquatic [10], and aerial [11] environments, with different kinematics corresponding to the different dynamical laws which apply in each setting. Here we focus on terrestrial locomotion, with motions restricted to two dimensions, and a Coulomb frictional force acting on the snake. Even with these simplifying assumptions, there are many possible snake kinematics to consider. One way to organize our understanding of a diversity of locomotory kinematics is to propose a measure of locomotory performance (here, efficiency), and determine which motions are optimal, and how their performance varies with physical parameters. Well-known examples are optimization studies of organisms moving in low- [12–17] and high-Reynolds-number fluid flows [18–23].

In this work, we use a recently-proposed model for snake locomotion in the plane which has shown good agreement with biological snakes [3,6]. The snake is a slender body whose curvature changes as a prescribed function of arc length distance along the backbone and time. Its net rotation and translation are then determined by Newton’s laws, with Coulomb friction between the snake and the ground. This model is perhaps one of the simplest which can represent arbitrary planar snake motions. Previous studies have used the model to find optimal snake motions when the motions are restricted to sinusoidal traveling waves [6], or the bodies are composed of two or three links [24]. These motions are represented by a small number of parameters (2-10); at the upper end of this range, the optimization problem is difficult to solve using current methods. In [2], Guo and Mahadevan developed a model which included the internal elasticity and viscosity of the snake, and studied the effects of these and other physical parameters on locomotion for a prescribed sinusoidal shape and sinusoidal and square-wave internal bending moments. In the present work, we neglect the internal mechanics of the snake, and consider only the work done by the snake on its environment, the ground. However, we consider a more general class of snake motions. Hu and Shelley assumed a sinusoidal traveling-wave snake shape, and computed the snake speed and locomotory efficiency across the two-parameter space of traveling-wave amplitude and wavelength for two values of frictional coefficients [6]. They also considered the ability of the snake to redistribute its weight during locomotion. Jing and Alben considered the locomotion of two- and three-link snakes, and found analytical and numerical results for the scaling of snake speed and efficiency [24]. Among the results were the optimal temporal function for actuating the internal angles between the links, expressed as Fourier series with one and two frequencies.

Here we address the optimization problem for planar snakes more generally, by using a much larger number of parameters (45 in most cases) to represent the snake’s curvature as a function of arc length and time. Although our snake shape space is limited, it is large enough to represent a wide range of shapes. To solve the optimization problem in a 45-dimensional shape space, and simultaneously across a large region of the two-dimensional friction parameter space, we develop an efficient numerical method based on a quasi-Newton method. The simulations show clearly that two types of traveling-wave motions, retrograde and direct waves, are optimal when the coefficient of friction transverse to the snake is large and small, respectively. An asymptotic analysis shows how the snake’s traveling wave amplitude should scale with the transverse coefficient of friction when it is large. The analysis gives a $-1/4$ power law for the amplitude which agrees well with the numerics. In this limit, the power required to move a snake optimally is simply that needed to tow a straight snake forward. Another analytical solution gives an optimally-efficient direct wave motion when the transverse coefficient of friction is zero. Between these extremes, the numerics show a more complicated set of optima which include standing-wave (or ratcheting) motions, among a wide variety of other locally-optimal motions. Taken
together, these results show which planar snake motions are optimal within a large space of possible motions and frictional coefficients.

2. Model

We use the same frictional snake model as [3,6,24], so we only summarize it here. The snake’s position is given by \( X(s,t) = (x(s,t), y(s,t)) \), a planar curve which is parametrized by arc length \( s \) and varies with time \( t \). A schematic diagram is shown in figure 1.

\[
\begin{align*}
\theta(s,t) &= \theta_0(t) + \int_0^s \kappa(s',t) \, ds', \\
x(s,t) &= x_0(t) + \int_0^s \cos \theta(s',t) \, ds', \\
y(s,t) &= y_0(t) + \int_0^s \sin \theta(s',t) \, ds'.
\end{align*}
\]

The unit vectors tangent and normal to the curve are \( \hat{s} \) and \( \hat{n} \) respectively. The tangent angle and curvature are denoted \( \theta(s,t) \) and \( \kappa(s,t) \), and satisfy \( \partial_s x = \cos \theta \), \( \partial_s y = \sin \theta \), and \( \kappa = \partial_s \theta \). We consider the problem of prescribing the curvature of the snake in order to obtain efficient locomotion. When \( \kappa(s,t) \) is prescribed, the tangent angle and position are obtained by integration:

\[
\begin{align*}
\theta(s,t) &= \theta_0(t) + \int_0^s \kappa(s',t) \, ds', \\
x(s,t) &= x_0(t) + \int_0^s \cos \theta(s',t) \, ds', \\
y(s,t) &= y_0(t) + \int_0^s \sin \theta(s',t) \, ds'.
\end{align*}
\]

The trailing-edge position \( (x_0, y_0) \) and tangent angle \( \theta_0 \) are determined by the force and torque balance for the snake, i.e. Newton’s second law:

\[
\begin{align*}
\int_0^L \rho \partial_{tt} x \, ds &= \int_0^L f_x \, ds, \\
\int_0^L \rho \partial_{tt} y \, ds &= \int_0^L f_y \, ds, \\
\int_0^L \rho X^\perp \cdot \partial_{tt} X \, ds &= \int_0^L X^\perp \cdot f \, ds.
\end{align*}
\]

Here \( X^\perp = (-y, x) \), \( \rho \) is the snake’s mass per unit length and \( L \) is the snake length. The snake is locally inextensible, \( \rho \) is uniform in space and constant in time, and \( L \) is constant in time. \( f \) is the force per unit length on the snake due to Coulomb friction with the ground:

\[
f(s,t) = -\rho g \mu_t \left( \partial_t X \cdot \hat{n} \right) \hat{n} - \rho g \left( \mu_f H \left( \partial_t X \cdot \hat{s} \right) + \mu_b \left( 1 - H \left( \partial_t X \cdot \hat{s} \right) \right) \right) \left( \partial_t X \cdot \hat{s} \right) \hat{s}.
\]

Here \( H \) is the Heaviside function and the hats denote normalized vectors. When \( \| \partial_t X \| = 0 \) we define \( \partial_t X \) to be \( 0 \). According to (2.7) the snake experiences friction with different coefficients for motions in different directions. The frictional coefficients are \( \mu_f \), \( \mu_b \), and \( \mu_t \) for motions in the
forward ($\hat{s}$), backward ($-\hat{s}$), and transverse (i.e. normal, $\pm \hat{n}$) directions, respectively. In general the snake velocity at a given point has both tangential and normal components, and the frictional force density has components acting in each direction. A similar decomposition of force into directional components occurs for viscous fluid forces on slender bodies [25].

We assume that the snake curvature $\kappa(s,t)$ is a prescribed function of $s$ and $t$ that is periodic in $t$ with period $T$. Many of the motions commonly observed in real snakes are essentially periodic in time [3]. We nondimensionalize equations (2.4)-(2.6) by dividing lengths by the snake length $L$, time by $T$, and mass by $\rho L$. Dividing both sides by $\mu g T^2$ we obtain:

$$
\frac{L}{\mu g T^2} \int_0^1 \partial_t xds = \int_0^1 f_x ds, 
$$

(2.8)

$$
\frac{L}{\mu g T^2} \int_0^1 \partial_t yds = \int_0^1 f_y ds, 
$$

(2.9)

$$
\frac{L}{\mu g T^2} \int_0^1 X_\perp \cdot \partial_t Xds = \int_0^1 X_\perp \cdot f ds, 
$$

(2.10)

In (2.8)-(2.10) and from now on, all variables are dimensionless. For most of the snake motions observed in nature, $L/\mu g T^2 \ll 1$ [3], which means that the snake’s inertia is negligible. By setting this parameter to zero we simplify the problem considerably while maintaining a good representation of real snakes. (2.8)-(2.10) become:

$$
b = (b_1, b_2, b_3)^T = 0; \quad b_1 \equiv \int_0^1 f_x ds,
$$

(2.11)

$$
b_2 \equiv \int_0^1 f_y ds, 
$$

(2.12)

$$
b_3 \equiv \int_0^1 X_\perp \cdot f ds. 
$$

(2.13)

In (2.11)-(2.13), the dimensionless force $f$ is

$$
f(s,t) = -\frac{\mu_h}{\mu_f} \left( \partial_t \dot{X} \cdot \hat{n} \right) \hat{n} - \left( H(\partial_t \dot{X} \cdot \hat{s}) + \frac{\mu_h}{\mu_f} (1 - H(\partial_t \dot{X} \cdot \hat{s})) \right) \left( \partial_t \dot{X} \cdot \hat{s} \right) \hat{s} 
$$

(2.14)

The equations (2.11)-(2.13) thus involve only two parameters, which are ratios of the friction coefficients. From now on, for simplicity, we refer to $\mu_t/\mu_f$ as $\mu_t$ and $\mu_h/\mu_f$ as $\mu_h$. Without loss of generality, we assume $\mu_h \geq 1$. This amounts to defining the backward direction as that with the higher of the tangential frictional coefficients, when they are unequal. $\mu_t$ may assume any nonnegative value. The same model was used in [3,6,24], and was found to agree well with the motions of biological snakes in [3].

Given the curvature $\kappa(s,t)$, we solve the three nonlinear equations (2.11)-(2.13) at each time $t$ for the three unknowns $x_0(t)$, $y_0(t)$ and $\theta_0(t)$. Then we obtain the snake’s position as a function of time by (2.1)-(2.3). The distance traveled by the snake’s center of mass over one period is

$$
d = \sqrt{\left( \int_0^1 x(s,1) - x(s,0) \, ds \right)^2 + \left( \int_0^1 y(s,1) - y(s,0) \, ds \right)^2}. 
$$

(2.15)

The work done by the snake against friction over one period is

$$
W = \int_0^1 \int_0^1 f(s,t) \cdot \partial_t X(s,t) \, ds \, dt. 
$$

(2.16)

We define the cost of locomotion as

$$
\eta = \frac{W}{d} 
$$

(2.17)

and our objective is to find $\kappa(s,t)$ which minimize $\eta$ at $(\mu_h, \mu_t)$ values which range widely over $[1, \infty) \times [0, \infty)$. In our computations, instead of $\eta$, we minimize a different function which is more
convenient for physical and numerical reasons:

\[ F = -\frac{d}{W} e^{2\cos(\theta(0,1) - \theta(0,0))}. \]  

(2.18)

To obtain \( F \) from \( \eta \) we have substituted \(-d/W\) for \( W/d \), to avoid infinities in the objective function for a common class of motions with \( d \to 0 \) and \( W \) of order 1. The exponential term in (2.18) penalizes rotations over a period, to ensure that the snake moves in a straight path rather than a circular path over many periods. As will be described in the results section, for \( \mu_0 \geq 6 \) and all \( \mu_n \), we find optima which have an analytical traveling wave form, and in the limit of large \( \mu_n \), rotations tend to zero for these solutions. These optima are found even with no rotation penalty, so the rotation penalty plays a negligible role in this region of parameter space. For smaller \( \mu_n \), however, in the absence of the rotation penalty, many of the optima have significant rotations. In Appendix A we give a more detailed explanation for the choice of \( F \).

3. Numerical Minimization Method

Our goal is to determine the snake shape, \( \kappa(s, t) \), which minimizes \( F \), for a given \( (\mu_b, \mu_t) \). Since \( \kappa(s, t) \) has \( T \)-period one, we represent it via a double series expansion:

\[ \kappa(s, t) = \sum_{j=0}^{m_1-1} \sum_{k=0}^{n_1-1} \left( \alpha_{jk} \cos(2\pi jt) + \beta_{jk} \sin(2\pi jt) \right) T_k(s). \]  

(3.1)

Here \( T_k \) is a Chebyshev polynomial in \( s \):

\[ T_k(s) = \cos(k \arccos(s)). \]  

(3.2)

The expansion (3.1) converges as \( m_1, n_1 \to \infty \) for a large class of \( \kappa \) which includes differentiable functions, and with a convergence rate that increases rapidly with the number of bounded derivatives of \( \kappa \) [26]. It is reasonable to hope that minimizing \( F \) over a class of functions (3.1) with small \( m_1 \) and \( n_1 \) gives a good approximation to the minimizers with infinite \( m_1 \) and \( n_1 \). Since the terms in (3.1) with coefficients \( \beta_{jk} \) are automatically zero, the functions given by (3.1) are a \((2m_1-1)n_1\)-dimensional space. In most of this work we use \( m_1 = n_1 = 5 \), so we are minimizing \( F \) in a 45-dimensional space. We give a few results at large \( \mu_T \) with \( m_1 = n_1 = 10 \) for comparison, and find similar minimizers in this region of parameter space. The minimization algorithm converges in a smaller portion of \((\mu_b, \mu_t)\) space as \( m_1 \) and \( n_1 \) increase, and convergence is considerably slowed by the increased dimensionality of the search space. We find that \( m_1 = n_1 = 5 \) is small enough to allow convergence to minima in a large portion of \((\mu_b, \mu_t)\) space, but large enough to approximate the gross features of a wide range of \( \kappa(s, t) \). Another reason to expect that good approximations to minimizers can be obtained with small \( n_1 \) is that the dynamical equations (2.11)–(2.13) depend only on spatial integrals of body position and velocity.

Using the dynamical equations (2.11)–(2.13), we show in Supplementary Material section A that \( d, W \), and \( F \) are the same for any reparametrization of time. Such a reparametrization can be used to reduce the high-frequency components of a motion while keeping the efficiency the same. Therefore, it is reasonable to expect that a good approximation to any minimizer of \( F \) can be obtained with low temporal frequencies, that is, with an expansion in the form of (3.1) with small \( m_1 \).

We use the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [27] to minimize \( F \) numerically, by computing a sequence of iterates which converge to a local minimizer of \( F \). The algorithm requires routines for computing the values and gradients of \( F \) as well as a line search method for finding the next iterate at each step, using the search direction provided within the BFGS algorithm. The BFGS algorithm is a quasi-Newton algorithm, and therefore forms an approximation to the Hessian matrix of second derivatives, without the expense and difficulty of computing the Hessian matrix explicitly.

We have developed an efficient method for evaluating \( F \) and its gradient with respect to the \((2m_1-1)n_1\) shape parameters with just a single simulation of the snake trajectory over one
period. Computing the value of $F$ requires computing the trajectory of the snake over one period. We discretize the period interval uniformly with $m$ time points: $\{0, 1/m, 2/m, \ldots, 1 - 1/m\}$. At each time point we write (2.11)–(2.13) as a nonlinear algebraic system of three equations in three unknowns: $\{\partial_t x_0, \partial_t y_0, \partial_t \theta_0\}$. We also define an artificial frame for each time point in which the snake’s trailing edge coordinates are $\{\tilde{x}_0, \tilde{y}_0, \tilde{\theta}_0\} = \{0, 0, 0\}$ at the time in question; each such frame is fixed in time. We then solve (2.11)–(2.13) for the trailing edge velocities in each of the artificial frames, $\{\tilde{\partial}_t x_0, \tilde{\partial}_t y_0, \tilde{\partial}_t \theta_0\}$. Since each artificial frame is obtained from the lab frame by a fixed translation and rotation, the tangent angle at the snake’s trailing edge changes at the same rate in the two frames: $\tilde{\partial}_t \theta_0 = \partial_t \theta_0$. We thus obtain the trailing edge tangent angle in the lab frame by integrating:

$$\theta_0(t) = \int_0^t \tilde{\partial}_t \theta_0 \, d\tau. \tag{3.3}$$

We then obtain the trailing edge position in the lab frame by integrating the trailing edge translational velocities, rotated to the lab frame:

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \int_0^t \begin{pmatrix} \cos \theta_0(\tau) & -\sin \theta_0(\tau) \\ \sin \theta_0(\tau) & \cos \theta_0(\tau) \end{pmatrix} \begin{pmatrix} \tilde{\partial}_t & \tilde{\partial}_t \end{pmatrix} X_0 \, d\tau. \tag{3.4}$$

We then integrate $\kappa(s,t)$ using $\{x_0(t), y_0(t), \theta_0(t)\}$ to obtain the body trajectory a posteriori. Using $\{\partial_t x_0, \partial_t y_0, \partial_t \theta_0\}$ as the unknowns instead of $\{x_0, y_0, \theta_0\}$ has two advantages: we avoid cancellation error associated with computing discrete time derivatives of $\{x_0, y_0, \theta_0\}$, and we solve $m$ decoupled systems of three equations in three unknowns instead of a larger coupled system of $3m$ equations in $3m$ unknowns ($\{x_0, y_0, \theta_0\}$ at all times points), which is more expensive to solve.

We solve the system (2.11)–(2.13) using Newton’s method with a finite-difference Jacobian matrix. We solve it at each time sequentially from $t = 0$, using a random initial guess for $\{\partial_t x_0, \partial_t y_0, \partial_t \theta_0\}$ at $t = 0$ and an extrapolation from previous time points as an initial guess at other $t$. An important element of our method is the quadrature used to compute the integrals in (2.11)–(2.13) using (2.14). Here we introduce notation for the tangential and normal components of the snake velocity:

$$u_s = \partial_s X \cdot \hat{s}; \quad u_n = \partial_s X \cdot \hat{n}. \tag{3.5}$$

In (2.14) we can write

$$\partial_s X \cdot \hat{s} = \frac{u_s}{\sqrt{u_s^2 + u_n^2}}; \quad \partial_s X \cdot \hat{n} = \frac{u_n}{\sqrt{u_s^2 + u_n^2}}. \tag{3.6}$$

These terms have unbounded derivatives with respect to $u_s$ and $u_n$ when $u_s, u_n \to 0$. The local error when integrating (2.14) using the trapezoidal rule (for example) with uniform mesh size $h$ is $O\left(h^3 \max(u_s, u_n)^{-2}\right)$. To obtain convergence in this case it is necessary to have $h = o(\max(u_s, u_n)^{2/3})$, so the trapezoidal rule (and other classical quadrature rules) need to be locally adaptive when $(u_s, u_n)$ tends to zero.

We use a different approach which allows for a uniform mesh even when $(u_s, u_n) \to 0$. We perform the quadrature analytically on subintervals using locally linear approximations for $u_s$ and $u_n$. Using $\{\partial_t x_0, \partial_t y_0, \partial_t \theta_0\}$ and $\kappa(s,t)$, we compute $\partial_s X, X$, and then $u_s$ and $u_n$ on a uniform mesh. On each subinterval, we approximate $u_s$ and $u_n$ as linear functions of the form $As + B$ using their values at the interval endpoints. We approximate $X^\perp, \hat{s},$ and $\hat{n}$ as constants using their values at the midpoints of the intervals. Then the integrals in (2.11)–(2.13), using (2.14), are of the form

$$C_0 \int_a^b \frac{As + B}{\sqrt{(As + B)^2 + (Cs + D)^2}} \, ds. \tag{3.7}$$
on each subinterval. We evaluate such integrals analytically using

\[ I_1 \equiv \int_0^1 \frac{ds}{\sqrt{s^2 + a_2 s + a_3}} = -\log \left( a_2 + 2\sqrt{a_3} \right) + \log \left( 2 + a_2 + 2\sqrt{1 + a_2 + a_3} \right), \tag{3.8} \]

\[ I_2 \equiv \int_0^1 \frac{s \, ds}{\sqrt{s^2 + a_2 s + a_3}} = -\sqrt{a_3} + \sqrt{1 + a_2 + a_3} - \frac{a_2}{2} I_1. \tag{3.9} \]

When used to evaluate (3.7), the logarithms in (3.8)–(3.9) have canceling singularities when the linear approximations to \( u_s \) and \( u_n \) are proportional (including the case in which one is zero), i.e. when \( AD - BC \to 0 \). We use different (and simpler) analytical formulae on these subintervals because the integrands are constants or sign functions, which are easy to integrate analytically. This quadrature method is \( O \left( h^2 \right) \) for mesh size \( h \), due to the use of linear and midpoint approximations of the functions in the integrands.

With this method for evaluating (2.11)–(2.13), we compute \( \{ \partial_t x_0, \partial_t y_0, \partial_t \theta_0 \} \) by Newton’s method as described above, and obtain the snake trajectory and \( d, W, \) and \( F \) for the prescribed curvature function. The BFGS method also requires the gradient of \( F \) with respect to the \((2m_1 - 1)n_1\) curvature coefficients \( \{ \alpha_{jk}, \beta_{jk} \} \). In Appendix B, we describe how the gradient of \( F \) is computed with a cost comparable to that of computing \( F \) alone.

4. Results

![Figure 2](image-url)

**Figure 2.** Result of one snake optimization run for \( \mu_b = 1 \) and \( \mu_t = 30 \). (a) Map of curvature versus arc length and time. (b) Values of the objective function (dashed line) and \( \log_{10} \) of the 2-norm of its gradient (solid line) with respect to the iteration count in the optimization routine. (c) Snake trajectory over one period.
In Figure 2 we show the result of one run of the optimization routine with ${\mu_b = 1}$ and ${\mu_t = 30}$, starting with a randomly chosen initial vector of curvature coefficients. In panel a we plot a map of curvature values $\kappa(s, t)$ over one period, and find that the curvature approximates a sinusoidal traveling wave with wavelength approximately equal to the snake body length. In this case the waves appear to travel with nearly constant speed, shown by the straightness of the curvature bands. The bands deviate from straightness near the ends, where the largest curvatures occur. In panel b we plot the objective function and $\log_{10}$ of the 2-norm of its gradient over the iteration count of the optimization routine. We find that the gradient norm decreases by four orders of magnitude within the first 100 iterations, and the objective function reaches a plateau at the same time. In panel c we give snapshots of the position of the snake $X(s, t)$ at equal instants in time, moving from left to right, and see that the snake approximately follows a sinusoidal path.

Figure 3 shows three different converged results for the same parameters ${\mu_b = 1}$ and ${\mu_t = 30}$ but with different randomly-chosen initial iterates. Again we have traveling waves of curvature shown by bands, but unlike in Figure 2a, the bands are not straight. Also, the bands travel from right to left instead of left to right. The bands are not straight because the traveling wave speed changes over time. However, at any given time, the distances in $s$ from the maximum to the minimum of curvature are nearly the same in Figure 3a–c and Figure 2a, and so are the shapes. In all four cases, the snake passes through the same sequence of shapes, but at different speeds, and different (integer) numbers of times in one period. Thus the four motions are essentially the same after a reparametrization of time. Correspondingly, the values of $F$ for all four are in the range $-5.4418 \pm 0.0024$. The distances traveled, $d$, are within 0.2% of integer multiples of 0.597, with the multiples 4, 3, 2, and 1 for the motions in Figure 2a and Figures 3a–c, respectively, which correspond to the number of times they move through the sequence of shapes represented by Figure 3c. The values of $F$ are the same whether the waves move from left to right or right to left in $s$, because ${\mu_b}$ (or ${\mu_b/\mu_f}$) = 1, so there is a symmetry between forward and backward motions.

Over an ensemble of 20 random initializations, this sinusoidal motion yields the lowest value of $F$ among the local optima found. We have repeated the search with twice as many spatial and temporal modes ($m_1 = 10$, $n_1 = 10$) and 48 random initializations. About three-quarters of the searches converge, and the majority of the converged states are very close to those shown in Figures 2 and 3. The values of $F$ for these states lie in the range $-5.577 \pm 0.026$. The mean is systematically lower than the mean for $m_1 = 5$, $n_1 = 5$ by about 2.5%. The decrease in the objective function is slight, and the optimal shape is fairly smooth, which indicates that the $m_1 = 5$, $n_1 = 5$ minimizer is representative of the minimizer for much larger $m_1$ and $n_1$, at least for ${\mu_b = 1}$ and ${\mu_t = 30}$. With $m_1 = 10$ and $n_1 = 10$, the curvature bands of Figures 2a and 3 become straighter near
the ends, showing the effect of the truncation in the Chebyshev series. However, it is more difficult
to obtain convergence with four times as many modes, because of the higher dimensionality of
the space. Since a primary goal here is to obtain the optima across as broad a range of \((\mu_b, \mu_t)\)
space as possible, we proceed with \(m_1 = 5\) and \(n_1 = 5\) except when noted.

![Diagram](image)

**Figure 4.** Traveling-wave optima for ten different \(\mu_t\) on a logarithmic scale. a) Values of \(F\). b) Curvature versus arc length at the instant when a curvature maximum crosses the midpoint, for the ten \(\mu_t\) values in (a). c) Snake shapes corresponding to (b).

Performing searches with \(\mu_b = 1\) and a wide range of \(\mu_t\), we find that similar sinusoidal
traveling-wave optima occur for \(\mu_t \geq 6\). In Figure 4a we plot the values of \(F\) corresponding to
the optima over a range of \(\mu_t\) on a logarithmic scale, and find that the optima are more efficient at
higher \(\mu_t\). In panel b, the curvatures of the optimal shapes are plotted for the same \(\mu_t\), at the times
when the curvatures have a local maximum at \(s = 1/2\). The corresponding snake body positions
are shown in panel c, and are vertically displaced to make different bodies easier to distinguish.
In all cases the bodies approximate sinusoidal shapes with slightly more than one wavelength of
curvature on the body. The curvature amplitudes decrease monotonically with increasing \(\mu_t\).

Performing searches with a wide range of \(\mu_t\) and \(\mu_b\) now, we find that similar sinusoidal
traveling-wave optima are found for \(\mu_t \geq 6\) and all \(\mu_b\) used \((1 \leq \mu_b \leq 3000)\). For these motions,
the curvature traveling waves move backward in \(s\), and the tangential motion is purely forward,
so they have the same \(F\) at all \(\mu_b\). However, for \(\mu_b > 1\), a separate class of (local, not global) optima
are also found. These have curvature traveling waves which move forward in \(s\), and for which the
tangential motion is purely backward. Since friction is higher in the backward direction, these
local optima have higher \(F\) (are less efficient) than the oppositely-moving optima, and the values
of \(F\) for these optima decrease with increasing \(\mu_b\), which makes physical sense.

In figure 5a we plot the values of \(F\) for these "Reverse" sinusoidal motions at various \(\mu_b\) and
\(\mu_t\), with leftward-pointing triangles. The triangles connected by a given solid line are data for
various \(\mu_t\) at a particular value of \(\mu_b\), labeled to the right of the line. When \(\mu_b = 1\), the reverse
sinusoidal motions have the same \(F\) values as the forward sinusoidal motions shown in figure 4.
These have the lowest \(F\) values across all \(\mu_b\) for \(\mu_t \geq 6\), and are thus the most efficient motions
we have found in this region of parameter space.

Three other sets of symbols are used to denote other classes of local optima found by our
method. All are less efficient than the forward sinusoidal motions, but we present examples of
them to show some of the types of local optima that exist for this problem. For these three sets,
shown by circles, downward pointing triangles, and plus signs, we do not join the optima with
Figure 5. Four categories of locally optimal motions: reverse sinusoidal motion, sinusoidal motion with curls, reverse sinusoidal motion with curls, and other motions, for ten different $\mu_t$ on a logarithmic scale. a) Values of $F$. Corresponding values of $\mu_b$ are labeled only for the reverse sinusoidal motions. b) Snapshots of a reverse sinusoidal motion with curls. Here $\mu_b = 1$ and $\mu_t = 10$. c,d) Snapshots of motions in the “Other” category, with $\mu_b = 10$ and $\mu_t = 60$ for (c) and $\mu_b = 6$ and $\mu_t = 20$ for (d).

For a given $\mu_b$ by lines, to reduce visual clutter. The circles, labeled “Curl,” give $F$ values for shapes which are like sinusoidal traveling waves, but they have a curl—the snake body winds around itself once in a closed loop. Snapshots of the snake motion for such a case are shown in figure 5b. Here the snake moves horizontally, but the snapshots are displaced vertically to make them distinguishable. The arrow shows the direction of increasing time. The tail of the snake is marked by a circle. The snake body moves rightward while the curvature wave and curl both move leftward. Because the snake overlaps itself, this motion is not physically valid in 2D. However, a large number of such optima are found numerically, with various small-integer numbers of curls (and with both directions of rotation). The efficiency of such motions decreases as the number of curls increases. The downward-pointing triangles, labeled “Curl, Reverse,” are for traveling waves with curls which pass backward along the snake. These have a yet smaller efficiency for the same reason that the reverse sinusoidal motions are less efficient than the forward sinusoidal motions, in the absence of curls. The plusses, labeled “Other,” are for other local optima, which have a variety of motions. Panel c shows snapshots of a curling (but not self-intersecting) motion in which the snake moves horizontally leftward. The horizontal displacement is increased by $8/9$ of a snake length between snapshots to make them distinguishable. Panel d shows a different curling motion in which the snake moves horizontally rightward, and the snapshots are displaced vertically to make them distinguishable. These alternative minima increase greatly in number as $\mu_t$ decreases below 6, and some of them outperform the forward sinusoidal motions at these transitional $\mu_t$ values. We will discuss this regime further after giving an analytical traveling-wave solution to the problem in the large-$\mu_t$ regime.
5. Large-μt analysis

Hu and Shelley [3,6] found that when the curvature of the snake backbone is prescribed as a sinusoidal traveling wave, high speed and efficiency is obtained when the coefficient of transverse friction is large compared to the coefficient of forward friction. Jing and Alben [24] found that the same holds for two- and three-link snakes. In biological snakes, the ratio of transverse to forward friction is thought to be greater than one [3,6], although how much greater is not clear. In [3,6], a ratio of only 1.7 was measured for anaesthetized snakes. These works noted that active snakes use their scales to increase frictional anisotropy, so the ratio in locomoting snakes is potentially much larger, and indeed, a ratio of 10 was found to give better agreement between the Coulomb friction model and a biological snake [6]. Wheeled snake robots have been employed very successfully for locomotion [28,29], and for most wheels the frictional coefficient ratio (with forward friction defined by the rolling resistance coefficient of the wheel) is much greater than 10 [30].

We now analytically determine the optimal snake motion in the limit of large μt, which helps us understand the numerical results in this regime. We give the position of the snake in terms of its components:

\[ \mathbf{X}(s,t) = (x(s,t), y(s,t)). \]  

(5.1)

If μt is large, the snake moves more easily in the tangential direction than in the transverse direction for a given shape dynamics. Furthermore, the most efficient shape dynamics will give motion mainly in the tangential direction, to avoid large work done against friction. We assume that the mean direction of motion is aligned with the x-axis, and that the deflections from the x-axis are small—that is, \(|y|, |\partial y_x|, |\partial y_y|, \text{ and higher derivatives are } O(\mu_t^\alpha) \text{ for some negative } \alpha\). This has been observed in the numerical solutions, and makes intuitive sense. Small deflections allow the snake to move along a nearly straight path, which is more efficient than a more curved path, which requires more tangential motion (and work done against tangential friction) for a given forward motion, i.e. \( d \).

We expand each of the terms in the force and torque balance equations in powers of \(|y|\) and retain only the terms which are dominant at large μt. We expand the cost of locomotion similarly, and find the dynamics which minimize it, in terms of \(y(s,t)\).

We first expand \(x(s,t)\). We decompose \(x(s,t)\) into its \(s\)-average \(\bar{x}\) and a zero-\(s\)-average remainder:

\[ x(s,t) = \bar{x}(s,t) + \int_s^0 \cos \theta(s',t) \, ds'. \]  

(5.2)

where \(\int_s^0\) means the constant of integration is chosen so that the integrated function has zero \(s\)-average. Hence

\[ \partial_t x(s,t) = \partial_t \bar{x}(s,t) + \int_s^0 -\partial_t \theta(s',t) \sin \theta(s',t) \, ds'. \]  

(5.3)

We denote the \(s\)-averaged horizontal velocity (the horizontal velocity of the center of mass) by

\[ U(t) \equiv \partial_t \bar{x}(s,t). \]  

(5.4)

We expand the integrand in (5.3) in powers of \(|y|\) and its derivatives:

\[ -\partial_t \theta(s,t) \sin \theta(s,t) = -\partial_s y(s,t) \partial_s y(s,t) + O(|\partial_s y|^4). \]  

(5.5)

The quartic remainder term in (5.5), \(O(|\partial_s y|^4),\) actually includes time derivatives of \(\partial_s y\) also, but we assume that \(|y|\) and all its derivatives are of the same order of smallness with respect to \(\mu_t\). We define

\[ h(s,t) = \int_s^0 -\partial_s y(s',t) \partial_s y(s',t) \, ds'. \]  

(5.6)

and with the small-deflection assumption, (5.3) becomes

\[ \partial_t x(s,t) = U(t) + h(s,t) + O(|\partial_s y|^4), \]  

(5.7)

and

\[ \partial_t \mathbf{X}(s,t) = (U(t) + h(s,t), \partial_t y(s,t)) + O(|\partial_s y|^4). \]  

(5.8)
We will see that for a quite general class of shape dynamics, \( U(t) \) is \( O(1) \), i.e. of the same order as one snake length per actuation period, even in the limit of large \( \mu_t \). \( U \) is therefore the dominant term in (5.7) and (5.8). We use this assumption on \( U \) to expand the velocity 2-norm:

\[
\| \partial_t \mathbf{X}(s,t) \| = \sqrt{\partial_t x^2 + \partial_t y^2} = U \left( 1 + \frac{h}{U} + \frac{1}{2} \frac{\partial_y y^2}{U^2} \right) + O(\|y\|^4). \tag{5.9}
\]

In writing \( O(\|y\|^4) \) we have again assumed \( |y| \) has the same order as its derivatives. Dividing (5.8) by (5.9), we obtain the expansion for the normalized velocity:

\[
\frac{\partial_t \mathbf{X}(s,t)}{U} = \left( \begin{array}{c} 1 - \frac{1}{2} \frac{\partial_y y^2}{U^2} + O(\|y\|^4) \\ \partial_y y \left( \frac{1}{2} \frac{\partial_y y^2}{U^2} + O(\|y\|^4) \right) \end{array} \right).
\tag{5.10}
\]

We next expand the unit tangent and normal vectors:

\[
\hat{s} = \left( \begin{array}{c} \partial_s x \\ \partial_s y \end{array} \right) = \left( \begin{array}{c} \sqrt{1 - \partial_s y^2} \\ \partial_s y \end{array} \right) = \left( \begin{array}{c} 1 - \frac{1}{2} \partial_s y^2 \\ \partial_s y \end{array} \right) + O(\|y\|^4)
\tag{5.11}
\]

\[
\hat{n} = \left( \begin{array}{c} -\partial_s y \\ 1 + \frac{1}{2} \partial_s y^2 \end{array} \right) + O(\|y\|^4)
\tag{5.12}
\]

Using (5.10) – (5.12),

\[
\partial_t \mathbf{X} \cdot \hat{s} = 1 - \frac{1}{2} \left( \partial_s y - \frac{\partial_y y}{U} \right)^2 + O(\|y\|^4),
\tag{5.13}
\]

\[
\partial_t \mathbf{X} \cdot \hat{n} = \left( \frac{\partial_y y}{U} - \partial_s y \right) \left( 1 - \frac{1}{2} \frac{\partial_y y^2}{U^2} \right) + \partial_y y \left( \frac{h}{U} + \frac{1}{2} \frac{\partial_y y^2}{U^2} \right) + O(\|y\|^5).
\tag{5.14}
\]

Using (5.13) and (5.14) we can write the force and torque balance equations (2.11)–(2.13) and the cost of locomotion \( \eta \) in terms of \( U \), \( h_t \), and derivatives of \( y \). We revert to \( \eta \) instead of \( F \) here because it is analytically simpler and the rotation penalty in \( F \) does not play a significant role at large \( \mu_t \), so \( \eta \) and \( F \) have the same optima.

We first expand \( \eta \) and use it to argue that the optimal shape dynamics is approximately a traveling wave. This allows us to neglect certain terms, which simplifies the formula for the rest of the analysis. If the power expended at time \( t \) is \( P(t) \), the cost of locomotion is

\[
\eta = \int_0^1 \left. P(t) \right| dt \int_0^1 U(t) dt.
\]

\[
= \int_0^1 U(t) \int_0^1 \left( \frac{\partial_t \mathbf{X} \cdot \hat{s}}{U(t)} \right)^2 + \mu_t \left( \frac{\partial_t \mathbf{X} \cdot \hat{n}}{U(t)} \right)^2 \left. \frac{\| \partial_t \mathbf{X} \|}{U(t)} \right| ds \int_0^1 U(t) dt.
\tag{5.15}
\]

In (5.16) we have assumed that the tangential motion of the snake is entirely forward, with no portion moving backward, so the coefficient in front of the tangential term is unity (i.e. \( \mu_t \), rather than the Heaviside term that switches between \( \mu_t \) and \( \mu_b \)). This assumption is not required but it simplifies the equations from this point onward and will be seen to be correct shortly. The numerator of (5.16) can be separated into a part involving tangential velocity and a part involving normal velocity with a factor of \( \mu_t \). Using (5.9), (5.13), and (5.14), we can list the terms with the lowest powers of \( y \) in each part:

\[
\left( \frac{\partial_t \mathbf{X} \cdot \hat{s}}{U(t)} \right)^2 = 1 + O(\|y\|^2)
\tag{5.17}
\]

\[
\mu_t \left( \frac{\partial_t \mathbf{X} \cdot \hat{n}}{U(t)} \right)^2 = \mu_t \left( \partial_s y - \frac{\partial_y y}{U} \right)^2 + O(\mu_t \|y\|^4).
\tag{5.18}
\]
The optimal shape dynamics is a $y(s,t)$ that minimizes $\eta$ subject to the constraints of force and torque balance. The leading order term in (5.17) is 1, which is independent of the shape dynamics. At the next (quadratic) order in $y$, terms appear in both (5.17) and (5.18); only that in (5.18) is given explicitly. It is multiplied by $\mu t$, so it appears to be dominant in the large $\mu t$ limit. Thus, to a first approximation, minimizing $\eta$ is achieved by setting

$$\partial_s y - \frac{\partial_t y}{U} = 0,$$

so a first guess for the solution is a traveling wave, i.e. a function of $s + Ut$ with $U$ constant. However, such a function does not satisfy the $x$-component of the force balance equation (2.11), which is:

$$\int h'' d\partial_t X \cdot \hat{s} = 1 + \mu t \int h'' d\partial_t X \cdot \hat{n} = 0.$$

(5.21)

Inserting (5.13) and (5.14) and retaining only the lowest powers in $y$ from each expression,

$$\int_0^1 1 + \mu t \left( \partial_s y - \frac{\partial_t y}{U} \right) \partial_s y ds = 0.$$

(5.21)

which cannot be satisfied by a function of $s + Ut$. Physically, the first term (unity) in (5.21) represents the leading-order drag force on the body due to forward friction. For a function of $s + Ut$, in which the deflection wave speed equals the forward speed, the snake body moves purely tangentially (up to this order of expansion in $y$) along a fixed path on the ground, with no transverse forces to balance the drag from tangential frictional forces. Also, when a function of $s + Ut$ is inserted into (5.21) all factors of $U$ cancel out and $U$ is left undetermined. This is related to a more fundamental reason why posing the shape dynamics as a function of $s + Ut$ is invalid: the problem of solving for the snake dynamics given in section 2 is not then well-posed.

In the well-posed problem we provide the snake curvature and curvature velocity versus time as inputs, and solve the force and torque balance equations to obtain the average translational and rotational velocities as outputs. Thus $U(t)$ is one of the three outputs (or unknowns) that is needed to solve the three equations.

Instead we pose the shape dynamics as

$$y(s,t) = g(s + U_w t),$$

(5.22)

which is a traveling wave with a prescribed wave speed $U_w$, different from $U$ in general. Since $g$ is periodic in $t$ with period 1, $g$ is periodic with a period of $U_w$. Inserting (5.22) into (5.21), we obtain an equation for $U$ in terms of $U_w$ and $g$:

$$1 + \mu t \left( 1 - \frac{U_w}{U} \right) \int_0^1 g'(s + U_w t)^2 ds = 0.$$

(5.23)

In (5.23), as $\mu t \to \infty$, $(1 - U_w/U) \to 0^-$. In the solution given by (5.22) and (5.23), the shape wave moves backwards along the snake at speed $U_w$, which propels the snake forward at a speed $U$, with $U$ slightly less than $U_w$. Therefore the snake slips transversely to itself, in the backward direction, which provides a forward thrust to balance the backward drag due to forward tangential friction. We refer to $(1 - U_w/U)$ in (5.23), a measure of the amount of backward slipping of the snake, as the “slip.”

We now solve (5.23) for $(1 - U_w/U)$ and insert the result into (5.16) to find $g$ and $U_w$ which minimize $\eta$. Using the lowest order expansions (5.17) and (5.18) with (5.23) we obtain:
\[
\eta = \left. \int_0^1 \frac{P(t)\,dt}{\int_0^1 U(t)\,dt} \right| \int_0^1 U(t)\,dt.
\]

\[
= \left. \int_0^1 U(t) \left( 1 + \mu_t \left( \frac{\partial_y y - \partial_y y}{U} \right)^2 + O(|y|^2, \mu_t |y|^4) \right)\,dsdt \right| \int_0^1 U(t)\,dt.
\]

\[
= \left. \frac{1}{\mu_t \left( g'(s + U_w t)^2 \right) + O(|g|^2, \mu_t |g|^4)} \right| \int_0^1 dt.
\]

where
\[
\langle g'(s + U_w t)^2 \rangle \equiv \int_0^1 g'(s + U_w t)^2 \,ds.
\]

If we minimize (5.26) over possible \( g \) for a given \( \mu_t \), using only the terms in the expansion that are given explicitly, we find that \( \eta \) tends to a minimum of 1 as the amplitude of \( g \) diverges. However, in this limit our small-\( y \) power series expansions are no longer valid. We therefore add the next-order terms in our expansions of \( \eta \) and the \( F_\nu \) equation and search again for an \( \eta \)-minimizer.

We now expand to higher order the components of \( \nu_t \), (5.17) and (5.18),

\[
\left( \partial_y X \cdot s \right)^2 \left( \frac{\partial \partial \Phi}{U(t)} \right) = 1 - \left( \partial_s y - \partial_y y \right)^2 + \frac{h}{U} + \frac{1}{2} \partial_s y^2 + O(|y|^4)
\]

\[
= \mu_t \left( \partial_y \left( \partial_y y \right) - \partial_y y \right)^2 + \partial_s y \left( - \frac{h}{U} + \frac{1}{2} \partial_s y^2 \right) \right)^2 \left( 1 + O(|y|^2) \right),
\]

and equation (2.11),

\[
\left( 1 + \mu_t \left( \partial_s y - \partial_y y \right) + \frac{h}{U} + \frac{1}{2} \partial_s y^2 \right) \partial_s y \,ds = 0.
\]

We again assume the traveling wave form of \( y \), and use this to evaluate \( h \) given by (5.6) in terms of \( g \):

\[
h(s, t) = \frac{U_w}{2} g'(s + U_w t)^2 + \frac{U_w}{2} g'(s + U_w t)^2.
\]

Inserting the traveling-wave forms of \( y \) and \( h \) into (5.30) we obtain

\[
1 + \mu_t \left( \frac{1 - U_w}{U} \right) \left( g'(s + U_w t)^2 + \frac{1}{2} \left( g'(s + U_w t)^2 \right)^2 \right) = 0.
\]

To obtain (5.32) we’ve used the fact that the slip \( 1 - U_w/U \rightarrow 0 \) as \( \mu_t \rightarrow \infty \), which will be verified subsequently. One can also proceed without this assumption, at the expense of a lengthier version of (5.32). We insert the traveling-wave forms of \( y \) and \( h \) into (5.28) and (5.29) to obtain \( \eta \), again with the assumption that \( 1 - U_w/U \rightarrow 0 \) as \( \mu_t \rightarrow \infty \) to simplify the result:

\[
\eta = 1 + \int_0^1 \frac{1}{2} \left( g'(s + U_w t)^2 \right) + \mu_t \left( 1 - \frac{U_w}{U} \right) + \frac{1}{2} \left( g'(s + U_w t)^2 \right)^2 \right) \left( g'(s + U_w t)^2 \right) dt
\]

Solving for the slip \( 1 - U_w/U \) from (5.32) and inserting into (5.33), we obtain an improved version of (5.26):

\[
\eta = 1 + \left. \int_0^1 \frac{1}{1 + \frac{1}{2} \left( g'(s + U_w t)^2 \right) + \frac{1}{\mu_t \left( g'(s + U_w t)^2 \right) + O(|g|^4, \mu_t |g|^8)}} \right| \int_0^1 dt.
\]

(5.34) is similar to (5.26), but with an additional term in the denominator of the integrand which penalizes large amplitudes for \( g \). If we approximate \( \left( g'(s + U_w t)^2 \right) \) as constant in time, we obtain

\[
\eta = 1 + \frac{1}{2} \left( g'^2 \right) + \frac{1}{\mu_t \left( g'^2 \right) + O(|g|^4, \mu_t |g|^8)}
\]

(5.35)
which is minimized for

$$
\langle g'^2 \rangle^{1/2} = 2^{1/4} \mu_t^{-1/4}.
$$

(5.36)

Thus the RMS amplitude of the optimal $g$ should scale as $\mu_t^{-1/4}$. (5.36) represents a balance between competing effects. At smaller amplitudes, the forward component of normal friction is reduced in comparison to the backward component of tangential friction, so the snake slips more, which reduces its forward motion and does more work in the normal direction. At larger amplitudes, as already noted, the snake’s path is more curved, which requires more work done against tangential friction for a given forward distance traveled. (5.36) is only a criterion for the amplitude of the motion. To obtain information about the shape of the optimal $g$, we now consider the balances of the $y$-component of the force and the torque.

So far we have searched for the optimal shape dynamics in terms of $y(s, t)$, but in fact it is the curvature $\kappa(s, t)$, which is prescribed. We obtain $y(s, t)$ from the curvature by integrating twice in $s$:

$$
y(s, t) = y(0, t) + \int_0^s \sin \theta(s', t) ds' = y(0, t) + \int_0^s \theta(s', t) ds' + O(y^3)
$$

(5.37)

$$
y(s, t) = y(0, t) + \int_0^s \theta(0, t) + \int_0^s \kappa(s'', t) ds'' ds' + O(y^3).
$$

(5.38)

$$
y(s, t) = y(0, t) + \theta(0, t) s + \int_0^s \kappa(s'', t) ds'' ds' + O(y^3).
$$

(5.39)

$$
y(s, t) = Y(t) + s R(t) + k(s, t) + O(y^3).
$$

(5.40)

where $Y$ and $R$ are defined for notational convenience. Prescribing the curvature is equivalent to prescribing $k(s, t)$. We set

$$
k(s, t) = g(s + U_w t)
$$

(5.41)

so we have the same form for $y$ as before, with an additional translation $Y(t)$ and rotation $R(t)$. $Y$ and $R$ are determined by the $y$-force and torque balance equations:

$$
\int_0^1 \left[ \left( \partial_t \mathbf{X} \cdot \hat{s} \right) s y + \mu_t \left( \partial_t \mathbf{X} \cdot \hat{n} \right) n y \right] ds = 0.
$$

(5.42)

$$
\int_0^1 \left[ \left( \partial_t \mathbf{X} \cdot \hat{s} \right) (x s y - y s x) + \mu_t \left( \partial_t \mathbf{X} \cdot \hat{n} \right) (x n y - y n x) \right] ds = 0.
$$

(5.43)

We expand these to leading order in $y$ and obtain

$$
\int_0^1 \partial_3 y - \frac{\partial_1 y}{U} + \frac{\partial_2 y}{U} \left( \frac{h}{U} + \frac{1}{2} \partial_3 y^2 \right) ds = 0.
$$

(5.44)

$$
\int_0^1 \left( \partial_3 y - \frac{\partial_1 y}{U} + \frac{\partial_2 y}{U} \left( \frac{h}{U} + \frac{1}{2} \partial_3 y^2 \right) \right) ds = 0.
$$

(5.45)

We insert (5.41) with (5.42) into the three equations (5.30), (5.45), and (5.46) to solve for the three unknowns, $U$, $Y$, and $R$ in terms of $g$ and $U_w$. We obtain:

$$
\int_0^1 1 + \mu_t \left[ \left( \frac{U_w}{U} + 1 + \frac{1}{2} g'^2 \right) g'^2 + \left( - \frac{Y'}{U} - \frac{R'}{U} + R \right) g \right] ds = 0.
$$

(5.47)

$$
\int_0^1 \left( \frac{U_w}{U} + 1 + \frac{1}{2} g'^2 \right) g' - \frac{Y'}{U} - \frac{R'}{U} + R ds = 0.
$$

(5.48)

$$
\int_0^1 \left( \frac{U_w}{U} + 1 + \frac{1}{2} g'^2 \right) g' - \frac{Y'}{U} - \frac{R'}{U} + R \right] ds = 0.
$$

(5.49)
We solve (5.48) and (5.49) for $Y$ and $R$ in terms of $U$:

\[
\frac{Y'}{U} - R = 4B - 6A
\]  
(5.50)

\[
\frac{R'}{U} = 12A - 6B
\]  
(5.51)

\[
A \equiv \left( \frac{U_w}{U} + 1 + \frac{1}{2}(g'{}^2) \right) \langle sg' \rangle
\]  
(5.52)

\[
B \equiv \left( \frac{U_w}{U} + 1 + \frac{1}{2}(g'{}^2) \right) \langle g' \rangle
\]  
(5.53)

We then solve (5.47) for $U$ in terms of $g$:

\[
\frac{U_w}{U'} = -1 - \frac{1}{2}(g'{}^2) + \frac{1}{\mu_t \left( (g')^2 - (g')^2 - 3((g')^2 - 2(2gg'))^2 \right) + O(|g|^4, \mu_t|g|^8))
\]  
(5.54)

We now form $\eta$, (5.16) with terms given by (5.28) and (5.29), using the updated from of $g$ ((5.41) with (5.42) and (5.54). We obtain a modified version of (5.34):

\[
\eta = 1 \int_0^1 \frac{1}{1 + \frac{1}{2}(g'{}^2) + \frac{1}{\mu_t \left( (g')^2 - (g')^2 - 3((g')^2 - 2(2gg'))^2 \right) + O(|g|^4, \mu_t|g|^8))} dt.
\]  
(5.55)

We can find $\eta$-minimizing $g$ in a few steps. First, let $\{L_k\}$ be the family of orthonormal polynomials with unit weight on $[0, 1]$ (essentially the Legendre polynomials):

\[
\int_0^1 L_i L_j ds = \delta_{ij}; \quad L_0 \equiv 1, \; L_1 = \sqrt{12} (s - 1/2), \ldots
\]  
(5.56)

At a fixed time $t$ we expand $g'$ in the basis of the $L_k$:

\[
g'(s + U_w t) = \sum_{k=0}^{\infty} c_k(t) L_k(s).
\]  
(5.57)

Then we have:

\[
\langle g'^2 \rangle = \sum_{k=0}^{\infty} c_k(t)^2,
\]  
(5.58)

\[
\langle g' \rangle^2 = c_0(t)^2,
\]  
(5.59)

\[
3 \left( (g')^2 - 2(2gg') \right)^2 = (L_1 g')^2 = c_1(t)^2.
\]  
(5.60)

Inserting into (5.55), $\eta$ becomes

\[
\eta = 1 \int_0^1 \frac{1}{1 + \frac{1}{2} \left( c_0(t)^2 + c_1(t)^2 + \sum_{k=2}^{\infty} c_k(t)^2 \right) + \frac{1}{\mu_t \left( \sum_{k=2}^{\infty} c_k(t)^2 \right) + O(|g|^4, \mu_t|g|^8)})} dt.
\]  
(5.61)

The denominator of the integrand in (5.61) is minimized when

\[
c_0(t) = 0; \; c_1(t) = 0; \; \sum_{k=2}^{\infty} c_k(t)^2 = \sqrt{2\mu_t}^{-1/2}.
\]  
(5.62)

The function $g(s + U_w t)$ that minimizes $\eta$ is that for which (5.62) holds for all $t$. Then the integral in (5.61) is maximized, so $\eta$ is minimized. Recall that $g$ is a periodic function with period $U_w$.

The relations (5.62) hold for any such $g$ in the limit that $U_w \to 0$, as long as $g$ is normalized.
appropriately. Define

$$A \equiv \left( \frac{1}{U_w} \int_0^{U_w} g'(x)^2 \, dx \right)^{1/2}. \quad (5.63)$$

Then

$$\int_0^1 g'(s + U_w t)^2 \, ds = A^2 + O(U_w) \quad (5.64)$$
$$\int_0^1 g'(s + U_w t) \, ds = O(U_w) \quad (5.65)$$
$$\int_0^1 sg'(s + U_w t) \, ds = O(U_w). \quad (5.66)$$

(5.64) – (5.66) are fairly straightforward to show, and we show them in Supplementary Material section B for completeness. If (5.64) – (5.66) hold, then in the limit that $U_w \to 0$, (5.62) holds with $A = 2^{1/4} \mu_t^{-1/4}$, by (5.58) – (5.60). A physical interpretation of the small-wavelength limit is that the net vertical force and torque on the snake due to the traveling wave alone ($g$) become zero in this limit, so no additional heaving motion ($Y$) or rotation ($R$) are needed, and thus the additional work associated with these motions is avoided. Similar “recoil” conditions were proposed as kinematic constraints in the context of fish swimming [18]. We also note two other scaling laws.

For a given $g \sim \mu_t^{-1/4}$, by (5.54) the slip $(1 - U_w/U) \sim \mu_t^{-1/2}$ and by (5.50)–(5.53), $Y, R \sim \mu_t^{-3/4}$.

Now to obtain an $\eta$-minimizing function $g^*$, let $g$ be any periodic function with period $U_w$. Define

$$g^*(s + U_w t) \equiv \frac{2^{1/4} \mu_t^{-1/4}}{A} g(s + U_w t) \quad (5.67)$$

where $A$ is given in (5.63). Then in the limit $U_w \to 0$,

$$\eta(g^*) \to 1 + \sqrt{2} \mu_t^{-1/2} + O(\mu_t^{-1}), \quad (5.68)$$

by the estimate in (5.55). Thus the optimal traveling wave motions become more efficient as $\mu_t \to \infty$, which agrees with the decrease in $F$ for the numerical solutions in figure 4a.

We consider the relationship of the numerically-found optima to the theoretical results further in figure 6. Here we take the plots of figure 4 and transform them according to the theory. Figure 6a transforms $F$ from figure 4a into $\eta$ by the formula $\eta = -c^2 F^{-1}$, since rotations over one period are negligible. In figure 6a we plot $\eta - 1$ versus $\mu_t$ for the numerical solutions (squares) from figure 4a, along with the leading order terms in the theoretical result (5.68). The numerics follow the same scaling as the theory, but with a consistent upward shift (as expected from the finite-mode truncation). In particular, the numerically-determined optima do not have very small wavelengths, due to the finite number of modes used. This is explained further in Supplementary Material section C. The uniformity of the shift may be due to the shape consistently assumed by the numerical optima for various $\mu_t$. For $\mu_t = 300$ we have also computed $\eta$ via the dynamical simulation for a much shorter-wavelength shape given by $\kappa(s, t) = A_1 \sin(2\pi s/U_w + 2\pi t)$ with wavelength $U_w = 1/8$ and $A_1$ chosen so that $\int_0^1 \kappa \approx g'$ satisfies (5.36). The result is given by the open circle, which outperforms the optimum identified in the truncated-mode space, as expected.

The circle is 0.013 above the solid line. This deviation is due in part to the neglect of the next term in the asymptotic expansion, proportional to $\mu_t^{-1}$ or $1/300$ here. Numerically, it is difficult to simulate motions with $U_w$ much smaller than 1/8. Figure 6b shows the curvature, rescaled by $\mu_t^{1/4}$ to give a collapse according to (5.36). The data collapse well compared to the unscaled data in 4b. Figure 6c shows the body shapes from 4c with the vertical coordinate rescaled, and the shapes plotted with all centers of mass located at the origin. We again find a good collapse, consistent with the curvature collapse.
6. Moderate to small $\mu t$: from retrograde waves to standing waves to direct waves

Near $\mu t = 6$ there is a qualitative change in the numerically computed optima. At values of $\mu t \gtrsim 6$, the global minimizers of $F$ are all backward-moving traveling waves. In addition, there are local minimizers in which the waves move forward along the snake (and the snake moves backward), and overlapping solutions with integer numbers of curls appearing in the traveling waves. The values of $F$ at the local minima, shown in figure 5a, are generally well-separated. In addition, there is a small number of local minima with ratcheting types of motions. For $\mu t \leq 6$ in figure 5a, there is an increasing number of local minimizers, denoted “Other,” quite close to the global minimizer. These motions are perturbed traveling waves. Thus the number of local minima increases greatly as $\mu t$ decreases below 6, and the spacings between their $F$ values decrease greatly.

Figure 7 shows the values of $F$ for all computed local optima at four values of $\mu t$ near the transition: 2, 3, 6, and 10. For each $\mu t$, the $F$ values are plotted against $\mu b$. Also, symbols are used to distinguish whether the local maximum in curvature travels forward or backward on average, and hence gives the direction of the curvature traveling wave (or perturbed traveling wave). At $\mu t = 10$ we find that the smallest $F$ is nearly constant with respect to $\mu b$. The same is true for $\mu t = 6$ except at $\mu b = 2000$, where there are several closely spaced values of $F$ which are slightly below the minimal $F$ at smaller $\mu b$. These values mark the first appearance of the perturbed traveling waves as $\mu t$ decreases. For $\mu t = 3$, the distribution of values is quite different, with a much stronger decrease in $F$ as $\mu b$ increases. There are many clusters of closely-spaced but
distinct values of $F$, showing a large increase in the number of local equilibria. At each $\mu_b$, the smallest values of $F$ are spread over a band of about 0.5 in width. At $\mu_t = 2$, the same qualitative trend holds, except that the decrease in $F$ with $\mu_b$ is much greater. There is more scatter of values overall and of those representing forward-moving curvature maxima. There is randomness in the sample of $F$ values, particularly evident for $\mu_t \leq 3$, because they are values for a finite sample of local minimizers starting from random initial $\kappa$ coefficients.

Intuitively, the minima for $\mu_t = 3$ represent perturbations of traveling wave motions towards “ratcheting”-type motions which involve small amounts of backward motion in a portion of the snake to “push” the rest of the snake forward. The appearance of backward motion explains the variation of $F$ with $\mu_b$. As $\mu_t$ decreases from 2000 to 6, $F$ for the traveling waves increase, and for $\mu_t = 3$, other motions outperform the traveling wave motions at sufficiently large $\mu_b$. For $\mu_t = 2$, the motions tend more toward ratcheting-type motions, and we will investigate this further subsequently.

Another way of viewing the transition is to consider the large-$\mu_t$ solution as $\mu_t$ decreases. A given traveling wave motion will slip more as $\mu_t$ decreases, unless the amplitude increases. The component of the transverse force in the forward direction reaches a maximum when the slope of the snake becomes vertical. Thus there is a limit to how much the slip can be decreased by increasing the amplitude of the traveling wave. When $\mu_t$ decreases below 6, the slope of the snake becomes nearly vertical at its maximum.

**Figure 7.** Values of $F$ for all computed local optima with various random initial iterates, across $\mu_b$ at four $\mu_t$ (10 (a), 6 (b), 3 (c) and 2 (d)) showing the emergence of non-traveling-wave optima.
Figure 8. (a) At each $(\mu_t, \mu_b)$, the maximum value of the function $\psi$ over all optima is plotted as a shaded dot. (b) The values of $F$ corresponding to the optima in (a) are plotted.

In figure 8a we plot a function of the shape dynamics which quantifies the transition away from traveling wave motions near $\mu_t = 6$. The quantity is

$$\psi[\kappa] \equiv \frac{\min_t \max_{0.2 \leq s \leq 0.8} |\kappa(s, t)|}{\max_t \max_{0.2 \leq s \leq 0.8} |\kappa(s, t)|}. \quad (6.1)$$

For $\kappa(s, t)$ which is a traveling wave with a wavelength sufficiently small that a maximum or minimum in curvature occurs in $0.2 \leq s \leq 0.8$ at all times, and for which the maximum and minimum are equal in magnitude, $\psi$ equals unity. For $\mu_t \gtrsim 6$, the numerical $\kappa(s, t)$ are approximately traveling waves with local extrema in $0.2 \leq s \leq 0.8$; we restrict $s$ away from the ends to avoid slight intensifications in curvatures near the ends. In figure 8a we plot the maxima of $\psi$ over all equilibria located at a given $(\mu_t, \mu_b)$ pair. We find $\psi > 0.8$ for $\mu_t \gtrsim 6$. $\psi$ increases slightly from about 0.8 to 0.9 as $\mu_t$ increases from 6 to 200. For $\mu_t = 6$, $\psi$ decreases slightly but noticeably, from 0.8 to 0.7, as $\mu_b$ exceeds 200, showing that at transitional $\mu_t$, the perturbations away from the traveling wave solutions appear first at the largest $\mu_t$. At $\mu_t = 3$, the values of $\psi$ have decreased to around 0.6 for most $\mu_b$; $\psi$ is as large as 0.9 at smaller $\mu_b$ and as small as 0.4 at larger $\mu_b$. For $\mu_t < 3$, the values of $\psi$ are generally much smaller. There is considerable randomness in the values of $\psi$ plotted for $\mu_t < 3$ because they are the maxima over samples of 10-20 local equilibria starting from random initial $\kappa$ coefficients.

In figure 8b we plot the values of $F$ for the optima whose $\psi$-values are shown in panel a. We find that $F$ varies smoothly over the region of transitional $\mu_t$, $2 \leq \mu_t \leq 10$, attaining its highest values (i.e. lowest efficiency) there. The retrograde traveling-wave motions and other motions are equally efficient at a transitional $\mu_t$, and the other motions become optimal as $\mu_t$ decreases.

A global picture of the optimal states is presented in the “phase diagram” of figure 9. At large $\mu_t$, the retrograde traveling waves are optimal. At small $\mu_t$, all of the optima have the form of direct traveling waves. Unlike for the large-$\mu_t$ waves, these direct waves all have large amplitudes, and their specific form varies with $\mu_b$. Near $\mu_b = 1$ there is a region where standing waves with non-zero mean deflection (or ratcheting motions) are seen. There is considerable randomness in the values of $\psi$ plotted for $\mu_t < 3$ because they are the maxima over samples of 10-20 local equilibria starting from random initial $\kappa$ coefficients.

In figure 8b we plot the values of $F$ for the optima whose $\psi$-values are shown in panel a. We find that $F$ varies smoothly over the region of transitional $\mu_t$, $2 \leq \mu_t \leq 10$, attaining its highest values (i.e. lowest efficiency) there. The retrograde traveling-wave motions and other motions are equally efficient at a transitional $\mu_t$, and the other motions become optimal as $\mu_t$ decreases.

A global picture of the optimal states is presented in the “phase diagram” of figure 9. At large $\mu_t$, the retrograde traveling waves are optimal. At small $\mu_t$, all of the optima have the form of direct traveling waves. Unlike for the large-$\mu_t$ waves, these direct waves all have large amplitudes, and their specific form varies with $\mu_b$. Near $\mu_b = 1$ there is a region where standing waves with non-zero mean deflection (or ratcheting motions) are seen. Examples will be given shortly. These optima co-exist with retrograde-wave optima at larger $\mu_t$. We note that there are other optima which do not fit easily into one of these categories, throughout the phase plane. These other optima are generally found less often by the computational search, and are generally suboptimal to the traveling waves, but there are exceptions. For the sake of brevity we do not
Figure 9. Phase diagram of optimal motions in $(\mu_t, \mu_b)$-space.

discuss such motions here. We also note that with twice the spatial and temporal modes ($m_1 = 10$ and $n_1 = 10$), we again find a transition from retrograde travelling waves to standing waves near $\mu_t = 6$, so the transition does not appear to be strongly affected by the mode truncation.

To obtain a clear picture of the snake kinematics across the transitional region shown in figure 9, we give figures showing snapshots of snake kinematics, together with spatiotemporal curvature plots, at $\mu_b = 1$ and 3 in this section and at $\mu_b = 2, 5,$ and 20 in Supplementary Material section D. Each of these figures represents one horizontal “slice” through the phase diagram in figure 9.

In figure 10, we show the optimal snake kinematics for $\mu_b = 1$ and eight $\mu_t$ spanning the transition. Each set of snapshots is arranged in a single column, moving downward as time moves forward over one period, from top to bottom (shown by arrows). Similarly to figure 5b and d, the snapshots are shown in a frame in which the net motion is purely horizontal, from left to right. However, the snapshots are displaced vertically to make them visible (i.e. to prevent overlapping). Below the snapshots, the corresponding curvature plot is shown, in the same format as in figures 2 and 3, with $s$ increasing from 0 to 1 on the horizontal axis, from left to right, and $t$ increasing from 0 to 1, from bottom to top. We have omitted the axes, and the specific curvature values for each plot, instead using a single color bar (shown at bottom) with curvature values which range from the maximum to the minimum for each plot. By omitting this information we are able to show several of the optimal motions side-by-side. By comparing a large number of motions in a single figure, it is easier to see the general trends.

We begin by discussing the rightmost column in figure 10, with $\mu_t = 6$. A retrograde traveling wave is clearly seen, so this optimum is a continuation of the large-$\mu_t$ solutions. At this $\mu_t$, just above the transition, the wave amplitude is large, and the wave clearly moves backwards in the lab frame. Thus the snake slips backward considerably even though it moves forward,
Figure 10. Snapshots of optimal snake motions at various $\mu_t$ (one per column, labeled at the top), for $\mu_b = 1$.

Because $\mu_t$ is not very large. Moving one column to the left, the optimum at $\mu_t = 3$ shows a clear difference. At certain instants, the curvature becomes locally intensified near the snake midpoint. This occurs when the local curvature maximum reaches the snake midpoint. The motion is still that of a retrograde traveling wave, but now the wave shape changes over the period. Hence the displacement can no longer be written in the form $g(s + U_w t)$ for a single periodic function $g$, as in the previous section. Comparing the curvature plots for $\mu_t = 6$ and 3, we see that the bands have nearly constant width across time for $\mu_t = 6$. At $\mu_t = 3$, by contrast, the curvature bands become modulated. The bands show bulges near the curvature intensifications. Moving leftward again, to $\mu_t = 2$ and then to 1.5, the curvature intensifications increase further, showing a general trend. In all cases, however, the curvature peaks move backward along the snake, even as the peak values change, so these motions are retrograde waves. At $\mu_t = 1.2$, a novel feature is seen: the endpoints of the snake curl up (but do not self-intersect), forming “anchor points”. Between these points, the remainder of the snake moves forward. In the curvature plot, bands can still be seen corresponding to retrograde waves. The three remaining columns, moving leftward, all show direct waves. That is, the wave moves in the same direction as the snake. No motion is shown for $\mu_t = 1$. The point $\mu_b = 1, \mu_t = 1$, in which friction is equal in all directions, is in some sense a singular point for our computations. Near this point, it is difficult to obtain convergence to optimal motions, and optima are difficult to classify clearly as retrograde or direct waves (or other simple motions). Nonetheless, we have computed one optimum which can be classified as a
direct wave, but it has complicated and special features which are not easy to classify. In the three left columns, the direct waves can be recognized by the bands in the curvature plots. The bands have the direction of direct waves. However, the bands are not nearly as simple in form as those for the retrograde waves, at large $\mu_t$.

Figure 11. Snapshots of optimal snake motions at various $\mu_t$ (one per column, labeled at the top), for $\mu_b = 3$.

Figure 11 shows optima for $\mu_b$ now increased to 3. The same transition is seen, but now in two middle columns, with $\mu_t = 1$ and 1.5, standing waves are seen. The curvature plots are nearly left-right symmetric. In the corresponding motions, the snake bends and unbends itself. Because $\mu_b > 1$, there is a preference for motion in the forward direction even though the curvature is nearly symmetric with respect to $s = 1/2$. These optima did not occur for $\mu_b = 1$ (figure 10) because symmetric motions with equal forward and backward friction would yield zero net distance traveled. It is somewhat surprising that such motions can be completely ineffective for $\mu_b = 1$ and yet represent optima for $\mu_b$ as small as 2, as shown in figure 9.
We have presented only a limited selection of optima for moderate $\mu_t$, and largely with pictures, because it is difficult to describe and classify many of the motions we have observed. Future work may consider this rich region of parameter space further.

We now move on to the limit of zero $\mu_t$. In this limit, we find a direct-wave motion with zero cost of locomotion. This explains the preference for direct waves at smaller $\mu_t$ in this section.

7. Zero $\mu_t$: direct-wave locomotion at zero cost

When $\mu_t = 0$, it is possible to define a simple shape dynamics with zero cost of locomotion. The shape moves with a nonzero speed while doing zero work. It is a traveling wave with the shape of a triangle wave which has zero mean vertical deflection:

$$y(s, t) = A_0 \int \text{sgn}(\sin(2\pi(s - t))) ds.$$  \hspace{1cm} (7.1)

The vertical speed corresponding to (7.1) is a square wave:

$$\partial_t y(s, t) = -A_0 \text{sgn}(\sin(2\pi(s - t))).$$  \hspace{1cm} (7.2)

The wave of vertical deflection moves forward with constant speed one, and the horizontal motion is also forward, with a constant speed $U$:

$$\partial_t x(s, t) = U$$  \hspace{1cm} (7.3)

We determine $U$ by setting the net horizontal force on the snake to zero:

$$\int \hat{s} \cdot \partial_t \mathbf{X} ds = 0.$$  \hspace{1cm} (7.4)

with

$$\hat{s} = \left( \frac{\sqrt{1 - A_0^2}}{A_0 \text{sgn}(\sin(2\pi(s - t)))} \right).$$  \hspace{1cm} (7.5)

Solving (7.4) for $U$ we obtain

$$U = \frac{A_0^2}{\sqrt{1 - A_0^2}}.$$  \hspace{1cm} (7.6)

With this choice of $U$, $\partial_t \mathbf{X} \cdot \hat{s}$ is identically zero on the snake, so the motion is purely in the normal direction. Therefore, there are no forces and torques on the snake, and no work done against friction. However, the snake moves forward at a nonzero speed $U$. This solution has infinite curvature at the peaks and troughs of the triangle wave, and we may ask if zero cost of locomotion ($\eta$) can be obtained in the limit of a sequence of smooth curvature functions which tend to this singular curvature function. To answer this question, we have set $y$ to the Fourier series approximation to (7.1) with $n$ wave numbers, for $n$ ranging from 2 to 1000. At each time $t$, we have then calculated a net horizontal speed $U(t)$, net vertical speed $V(t)$ and a net rotational speed $dR(t)/dt$ such that $F_x(t) = 0$, $F_y(t) = 0$ and $\tau(t) = 0$. We find very clear asymptotic behaviors for $U(t)$, $V(t)$, $dR(t)/dt$, and $\eta$:

$$U(t) = \frac{A_0^2}{\sqrt{1 - A_0^2}} + O(n^{-1}), \quad V(t) = O(n^{-1}), \quad \frac{dR}{dt} = O(n^{-1}), \quad \eta = O(n^{-1}).$$  \hspace{1cm} (7.7)

Therefore zero cost of locomotion does in fact arise in the limit of a sequence of smooth shape dynamics. This type of traveling wave shape dynamics, which yields locomotion in the direction of the wave propogation, has been called a “direct” wave in the context of crawling by snails [31] and worms [32]. A triangular wave motion was previously considered (in less quantitative detail) by [33].

The triangular traveling wave motion is particularly interesting because it can represent an optimal motion at both zero and infinite $\mu_t$, and its motion can be computed analytically for all
8. Conclusion

We have studied the optimization of planar snake motions for efficiency, using a fairly simple model for the motion of snakes using friction. Our numerical optimization is performed in a space of limited dimension (usually 45, increased to 180 at selected \((\mu_0, \mu_1)\) values), but which is large enough to allow for a wide range of motions. The optimal motions have a clear pattern when the coefficient of transverse friction is large. Our numerical results and asymptotic analysis indicate that a traveling-wave motion of small amplitude is optimal. When this coefficient is small, another traveling-wave motion, of large amplitude, is optimal. When the coefficient of transverse friction is neither large nor small, the optimal motions are far from simple, and we have only begun to address this regime.

A natural extension of this work is to include three-dimensional motions of snakes. In other words, one would give the snake the ability to lift itself off of the plane [2,3,6], as occurs in many familiar motions such as side-winding [9]. One can consider more general frictional force laws, perhaps involving a nondiagonal frictional force tensor. Incorporating internal mechanics of the snake as in [2] would penalize bending with high curvature. Our optima at large and small \(\mu\) can be approximated well with shapes of small and moderate curvature, respectively. Some of the more exotic optima at moderate \(\mu\) which involved curling at the ends would be significantly modified, however. Another interesting direction is to consider the motions of snakes in the presence of confining walls or barriers [5,34].

Acknowledgment

We would like to acknowledge helpful discussions on snake physiology and mechanics with David Hu and Hamidreza Marvi, helpful discussions with Fangxu Jing during our previous study of two- and three-link snakes, and the support of NSF-DMS Mathematical Biology Grant 1022619 and a Sloan Research Fellowship.

A. Appendix: Objective function

To understand our choice of \(F\), we first describe the class of motions \(X(s,t)\) which result from periodic-in-time \(\kappa(s,t)\). Let \(X(s,t)\) and \(\partial_t X(s,t)\) be given for the moment. Let us rotate \(X(s,t)\) by an angle \(\alpha\) and translate \(X(s,t)\) by a vector \(v\), uniformly over \(s\), so we have a rigid body motion. Then \(s\) and \(n\) are also uniformly rotated by \(\alpha\). Let us also rotate \(\partial_t X(s,t)\) uniformly by \(\alpha\). Then by (2.14), \(f\) is uniformly rotated by \(\alpha\). Hence, if \(X(s,t)\) and \(\partial_t X(s,t)\) are such that the force and torque balances (2.11)–(2.13) hold, they will continue to hold if we apply a rigid-body rotation and translation to \(X(s,t)\) and rotate \(\partial_t X(s,t)\) by the same angle. Now, since \(\kappa(s,t+1) = \kappa(s,t)\), \(X(s,t+1)\) is obtained from \(X(s,t)\) by a rigid-body rotation by an angle \(\alpha_1\) and translation by a vector \(v_1\). In fact, the trajectory of the body during each period is the same as during the previous period but with a rigid translation by \(v_1\) and rotation by \(\alpha_1\). Over many periods such a body follows a circular path when \(\alpha_1 \neq 0\), and because of the rotation, the distance traveled over \(n\) periods may be much smaller than \(n\) times the distance traveled over one period. To obtain a motion which yields a large distance over many periods, we restrict to body motions for which \(X(s,t+1)\) is obtained from \(X(s,t)\) by a translation only, with rotation angle \(\alpha_1 = 0\).

We implement this constraint approximately in our optimization problem, using a penalization term. We multiply \(d\) in (2.17) by a factor which penalizes rotations over a single period:

\[
d \to de^{2\cos(\theta(0,1) - \theta(0,0))}
\]

(A 1) involves the change in angle at the snake’s trailing edge over a period, which is the same as \(\alpha_1\). The exponential factor has a maximum of \(e^2\) when \(\alpha_1 = 0\), and a minimum of \(e^{-2}\) when...
Having previously computed the work in the objective function. Decreased from 2, some of the optima begin to involve nontrivial amounts of rotation. In short, the factor of 2 represents a choice of how much to weight lack of rotation relative to distance and work in the objective function.

A second modification to \( \eta \) is made to avoid large values of \( \eta \). There are snake motions (i.e. with the \( \kappa(s,t) = \kappa(1-s,t) \) symmetry and \( \mu_b = 1 \)) which have \( d = 0 \) and \( W \neq 0 \), giving \( \eta \rightarrow \infty \). The minimization search commonly encounters motions which are nearly as ineffective throughout shape space, and \( \eta \) has a large gradient in the vicinity of such motions. It is challenging to minimize such a function numerically. We therefore substitute \( -d/W \) for \( W/d \). The minima are the same for both because each is an increasing function of the other, but \( -d/W \) varies much more smoothly in most of shape parameter space. There are cases where \( -d/W \) tends to \( -\infty \), but these only occur in a small region of parameter space which is rarely approached by our computations.

Combining these ideas, the function we minimize in place of \( \eta \) is

\[
F = -\frac{d}{W} e^{\cos(\theta(0,1) - \theta(0,0))}.
\] (A 2)

**B. Appendix: Computing \( \nabla F \)**

The BFGS method requires the gradient of \( F \) with respect to the \((2m_1 - 1)n_1\) curvature coefficients \( \{\alpha_{jk}, \beta_{jk}\} \). For notation, let \( \mathbf{a} \) be a \((2m_1 - 1)n_1 \times 1\) vector whose entries are the curvature coefficients. We compute a finite-difference gradient using

\[
\nabla F_j = \frac{F(\mathbf{a} + 10^{-8}\|\mathbf{a}\|\mathbf{e}_j) - F(\mathbf{a})}{10^{-8}\|\mathbf{a}\|}, \quad j = 1 \ldots (2m_1 - 1)n_1,
\] (A 1)

where \( \mathbf{e}_j \) is the unit basis vector in the \( j \)th coordinate direction. (A 1) requires \((2m_1 - 1)n_1 + 1\) evaluations of \( F \), or \((2m_1 - 1)n_1 + 1\) computations of the snake motion over a period. We employ a very accurate approximation which requires only a single computation of the snake motion over a period.

Computing \( F(\mathbf{a} + 10^{-8}\|\mathbf{a}\|\mathbf{e}_j) \) requires solving \( \mathbf{b} = 0 \) from (2.11)–(2.13) for \( \{\partial_x x_0, \partial_y y_0, \partial_\theta \theta_0\} \) at each time \( t_k = k/m \in [0, 1/m, \ldots 1 - 1/m] \) using the vector of curvature coefficients \( \mathbf{a} + 10^{-8}\|\mathbf{a}\|\mathbf{e}_j \), \( \mathbf{b} \) is a function of \( \{\partial_x x_0, \partial_y y_0, \partial_\theta \theta_0\}^T \) and the curvature coefficients, so we write it as \( \mathbf{b}(\partial_x x_0, \partial_y y_0, \partial_\theta \theta_0)^T, \mathbf{a} + 10^{-8}\|\mathbf{a}\|\mathbf{e}_j \) and solve \( \mathbf{b}(\partial_x x_0, \partial_y y_0, \partial_\theta \theta_0)^T, \mathbf{a} + 10^{-8}\|\mathbf{a}\|\mathbf{e}_j = 0 \) at each time \( t_k \). The value of \( \{\partial_x x_0, \partial_y y_0, \partial_\theta \theta_0\}^T \) which satisfies these equations at each \( t_k \) may be written \( \mathbf{e}_k(\mathbf{a} + 10^{-8}\|\mathbf{a}\|\mathbf{e}_j) \). Because \( \mathbf{e}_k(\mathbf{a}) \) solves a nearby set of equations, it provides a very good initial guess for a Newton iteration to find \( \mathbf{e}_k(\mathbf{a} + 10^{-8}\|\mathbf{a}\|\mathbf{e}_j) \). We define \( \mathbf{e}_j \equiv 10^{-8}\|\mathbf{a}\|\mathbf{e}_j \) for our description. We start with the usual Taylor series expansion from which Newton’s method is obtained:

\[
0 = \mathbf{b}(\mathbf{c}_k(\mathbf{a} + \mathbf{e}_j), \mathbf{a} + \mathbf{e}_j) = \mathbf{b}(\mathbf{c}_k(\mathbf{a}), \mathbf{a} + \mathbf{e}_j) + \mathbf{J}(\mathbf{c}_k(\mathbf{a}), \mathbf{a} + \mathbf{e}_j)(\mathbf{c}_k(\mathbf{a} + \mathbf{e}_j) - \mathbf{c}_k(\mathbf{a})) + O\left(\|\mathbf{e}_j\|^2\right).
\] (A 2)

Here \( \mathbf{J} \) is the 3-by-3 Jacobian matrix of partial derivatives of \( \mathbf{b} \) with respect to its leading 3-by-1 vector argument:

\[
\mathbf{J}(\mathbf{x}, \mathbf{y})_{ij} = \partial_{x_i} b_j(\mathbf{x}, \mathbf{y}), \quad i = 1, 2, 3, \quad j = 1, 2, 3.
\] (A 3)

Having previously computed \( \mathbf{c}_k(\mathbf{a}) \) via Newton’s method, we have also calculated and stored \( \mathbf{J}(\mathbf{c}_k(\mathbf{a}), \mathbf{a}) \) at each \( t_k \). We therefore modify \( \mathbf{J} \) in (A 2) while retaining the same order of accuracy in \( \mathbf{e}_j \):

\[
0 = \mathbf{b}(\mathbf{c}_k(\mathbf{a} + \mathbf{e}_j), \mathbf{a} + \mathbf{e}_j) = \mathbf{b}(\mathbf{c}_k(\mathbf{a}), \mathbf{a} + \mathbf{e}_j) + \mathbf{J}(\mathbf{c}_k(\mathbf{a}), \mathbf{a})(\mathbf{c}_k(\mathbf{a} + \mathbf{e}_j) - \mathbf{c}_k(\mathbf{a})) + O\left(\|\mathbf{e}_j\|^2\right).
\] (A 4)

We solve (A 4) to obtain \( \mathbf{c}_{jk} \) our approximation to \( \mathbf{c}_k(\mathbf{a} + \mathbf{e}_j) \):

\[
\mathbf{c}_{jk} \equiv \mathbf{c}_k(\mathbf{a}) - \mathbf{J}(\mathbf{c}_k(\mathbf{a}), \mathbf{a})^{-1}\mathbf{b}(\mathbf{c}_k(\mathbf{a}), \mathbf{a} + \mathbf{e}_j) = \mathbf{c}_k(\mathbf{a} + \mathbf{e}_j) + O\left(\|\mathbf{e}_j\|^2\right).
\] (A 5)
\[ (A5) \] requires only a single computation of the snake motion over a period, which gives \( c_k(a) \) and \( J(c_k(a), a)^{-1} \) at each \( t_k \). It also requires \((2m_1 - 1)n_1\) evaluations of the function \( b \). These are much less expensive than even a single computation of the snake motion over a period. The \( O\left(\|e_j\|^2\right) \) error in \( c_{jk} \) is of the same order of magnitude as machine precision. The value of the cost function \( \tilde{F}_j \approx F(a + e_j) \) corresponding to \( c_{jk} \) has an error which is also comparable to machine precision, and is therefore subdominant to the \( 10^{-8} \) error in \( (A1) \). This level of error leads to essentially no deterioration in the convergence of BFGS up to machine precision relative to the version with the exact gradient [35]. Therefore we obtain the gradient of \( F \) for approximately the cost of computing \( F \). Each computation of a search direction in BFGS requires only one gradient computation. These ingredients give an efficient optimization method, which is critical to finding optima in the high-dimensional space of curvature coefficients.

References

A. Supplementary Material: Invariance under reparametrization of time

Here we show that \( d, W, \) and \( F \) are the same for any reparametrization of time—in other words, for \( \kappa(s, \gamma(t)) \) with any nondecreasing differentiable mapping \( \gamma \) from \([0, 1]\) to \([0, 1]\) with \( \gamma(0) = 0 \) and \( \gamma(1) = 1 \). Such a reparametrization can be used to reduce the high-frequency components of a motion while keeping the efficiency the same. Therefore, it is reasonable to expect that a good approximation to a minimizer of \( F \) can be obtained with low temporal frequencies, that is, with an expansion in the form of (3.1) with small \( m_1 \). To show that \( d, W, \) and \( F \) are the same for a reparametrization of time, let \( \kappa(s, t) \) be given and \( \gamma(t) \) be the aforementioned parametrization function. Let \( X(s, t) \) be the motion obtained by integrating \( \kappa(s, t) \) from \( s = 0 \) as in (2.1)–(2.3) with the integration constants \( X_0(t) = X(0, t) \) and \( \theta_0(t) \) chosen to satisfy the dynamical equations (2.11)–(2.13). Now we show that \( \tilde{X}(s, t) = X(s, \gamma(t)) \) also satisfies (2.11)–(2.13) at each time \( t \). By the chain rule of differentiation,

\[
\frac{\partial_t \tilde{X}(s, t)}{\|d\tilde{X}(s, \gamma(t))\|/dt} = \frac{\gamma'(t) \partial_t X(s, \gamma(t))}{\| \gamma'(t) \partial_t X(s, \gamma(t)) \|} = \frac{\partial_t X(s, \gamma(t))}{\| X(s, \gamma(t)) \|} \quad (A 1)
\]

using \( \gamma'(t) \geq 0 \). Let \( \tilde{f}(s, t) \) be the force distribution corresponding to \( \tilde{X}(s, \gamma(t)) \). It is given by (2.14) with \( \partial_t \tilde{X} \) in place of \( \partial_t X \) and \((\tilde{s}, \tilde{n})\) in place of \((s, n)\), where \((\tilde{s}, \tilde{n})\) are the unit vectors tangent and normal to \( \tilde{X}(s, t) \). Using (A 1) we obtain \( \tilde{f}(s, t) = f(s, \gamma(t)) \). Since (2.11)–(2.13) hold for \( f(s, t) \) for all \( t \in [0, 1] \), they also hold for \( \tilde{f}(s, t) \) for all \( t \in [0, 1] \) since \( \gamma(t) \in [0, 1] \). Thus they hold for \( \tilde{f}(s, t) \) in place of \( f(s, t) \), so \( \tilde{X}(s, t) \) satisfies the dynamical equations (2.11)–(2.13). The distance traveled \( \tilde{d} \) for this motion is

\[
\tilde{d} = \left( \int_0^1 \left( \int_0^1 \partial_t \tilde{x}(s, t)dsdt \right)^2 + \left( \int_0^1 \partial_t \tilde{y}(s, t)dsdt \right)^2 \right)^{1/2} \quad (A 2)
\]

\[
= \left( \int_0^1 \left( \int_0^1 \partial_t x(s, \gamma)dsd\gamma \right)^2 + \left( \int_0^1 \partial_t y(s, \gamma)dsd\gamma \right)^2 \right)^{1/2} \quad (A 3)
\]

\[
= d \quad (A 4)
\]

using (2.15). The work done is

\[
\tilde{W} = \int_0^1 \int_0^1 \tilde{f}(s, t) \cdot \partial_t \tilde{X}(s, t)dsdtdt = \int_0^1 \int_0^1 f(s, \gamma) \cdot \partial_t X(s, \gamma)dsd\gamma = W \quad (A 5)
\]

using (2.16). The net rotation over a period is

\[
\tilde{\theta}(0, 1) - \tilde{\theta}(0, 0) = \int_0^1 \partial_t \tilde{\theta}(0, t)dt = \int_0^1 \partial_t \theta(0, \gamma)d\gamma = \theta(0, 1) - \theta(0, 0) \quad (A 6)
\]

so \( F \) (2.18) is unchanged by the reparametrization of time.

B. Supplementary Material: Small-wavelength integrals

Since \( g \) is periodic, \( g' \) is periodic with zero average over a period. (5.64) – (5.66) are similar to results from an analysis course but we show them briefly for completeness. To show (5.64), we decompose the integral in (5.64) into two intervals:

\[
\int_0^1 g'(s + U_w t)^2 ds = \int_0^{U_w \lfloor 1/U_w \rfloor} g'(s + U_w t)^2 ds + \int_{U_w \lfloor 1/U_w \rfloor}^1 g'(s + U_w t)^2 ds. \quad (A 1)
\]

The first integral on the right of (A 1) is over an integral number of periods of \( g'^2 \), and is thus \([1/U_w]U_w A^2 = A^2 + O(U_w)\). The second integral is over an interval of length \( < U_w \), and is thus
\(O(U_w)\), which shows (5.64). To show (5.65), we use (A 1) again but with \(g'\) in place of \(g'^2\):

\[
\int_0^1 g'(s + U_w t) ds = \int_0^{U_w[1/U_w]} g'(s + U_w t) ds + \int_{U_w[1/U_w]}^1 g'(s + U_w t) ds. \tag{A 2}
\]

The first integral on the right of (A 2) is over an integral number of periods of \(g'\), and is thus zero. The second integral is over an interval of length \(U_w\), and is thus \(O(U_w)\). To show (5.66), we decompose the integral into three parts:

\[
\int_0^1 sg'(s + U_w t) ds = \int_0^{U_w[1/U_w]} U_w[s/U_w] g'(s + U_w t) ds + \int_{U_w[1/U_w]}^1 g'(s + U_w t) ds.
\]

The first integral on the right of (A 3) substitutes a constant approximation to \(s\) on each subinterval of period length. It is identically zero since \(g'\) has mean zero. The second integral is the error in the constant approximation to \(s\), \(O(U_w)\) on each subinterval, of which there are \(O(1/U_w)\), each of length \(U_w\). Thus the second integral is \(O(U_w)\). The third integral is also \(O(U_w)\), as explained for (A 1) and (A 2), so (5.66) holds.

C. Supplementary Material: Effect of finite mode truncation on large-\(\mu_t\) optimum

The theory of section 5 predicts that the wavelengths of the optimal shapes should tend to zero, but the wavelengths for the computed optima shown in figure 4b and c are all in the neighborhood of one body length, not particularly small. The reason is that it is not possible to represent traveling waves with very small wavelengths with the value of \(n_1\) used here (5), which yields curvatures which are polynomials in \(s\) of up to fourth degree.

A general periodic curvature function can be represented as a Fourier series in \(x = s + U_w t\).

For any such function, with amplitude scaled appropriately, the theory of section 5 predicts that the efficiency should tend to the optimum as \(U_w \to 0\). For simplicity we focus on a single mode \(\sin(2\pi x/U_w) = \sin(2\pi s/U_w + 2\pi t)\) and ask: how well can it be approximated by the truncated Chebyshev series we have used in the numerics? At any time \(t\), \(\sin(2\pi s/U_w + 2\pi t)\) can be decomposed into a linear combination of \(\sin(2\pi s/U_w)\) and \(\cos(2\pi s/U_w)\), so we compute the best approximations to these two functions in \(L^\infty\) norm by Chebyshev series with \(n_1\) terms. In figure 12 we show the resulting approximation errors (in \(L^2\) and \(L^\infty\)) for \(n_1 = 5\) and 10, and for \(\sin(2\pi s/U_w)\) (panel a) and \(\cos(2\pi s/U_w)\) (panel b). We find the error transitions from small to large for the wavelength \(U_w\) in the neighborhood of 1 for \(n_1 = 5\) and 1/2 for \(n_1 = 10\). For comparison, in panels c-e we plot snapshots of the curvatures from the optimal snake motions identified previously for \(\mu_t = 30\) with \(m_1 = n_1 = 5\) and \(m_1 = n_1 = 10\). The panels show the snapshots at instants when the curvatures have local maxima at \(s = 0.25\) (c), 0.5 (d) and 0.75 (e). These shapes are approximately sinusoidal, and those with \(m_1 = n_1 = 10\) have a somewhat smaller wavelength than those with \(m_1 = n_1 = 5\). The wavelengths for \(m_1 = n_1 = 5\) appear to be somewhat smaller than 1, while those for \(m_1 = n_1 = 10\) appear to be approximately 2/3. Based on panels a and b, we would have expected the \(m_1 = n_1 = 10\) cases in c-e to have still smaller wavelengths, but in the space spanned by the Chebyshev polynomials of small degree, the smallest \(U_w\) is not the only consideration: certain functional forms for the curvature should be expected to perform better than others, and the results here may tend towards such forms.
Figure 12. The approximation error in representing (a) sine and (b) cosine functions with different wavenumbers using either five or ten Chebyshev modes (corresponding curves are labeled). Panels c-e show snapshots of curvature versus arc length in the optimal numerical traveling wave motions with either five (dashed line) or ten (solid line) Fourier and Chebyshev modes. The snapshots are at times when there are local curvature maxima at (c) $s = 0.25$, (d) $s = 0.5$, and (e) $s = 0.75$.

D. Supplementary Material: Optimal snake motions across the $\mu_t$ transition

Figure 13 shows optima for $\mu_b = 2$, in the same format as figure 10, at a different set of seven $\mu_t$. The motions occur on a horizontal line in figure 9 which touches the bottom of the region in which standing waves occur. Starting in the right column, which has $\mu_t = 3$, we see a retrograde wave, with modulations (or bulges) in the bands of the curvature plot. In the next five columns moving leftward, we see the transition from retrograde to direct waves. However, the curvature plots begin to resemble a checkerboard pattern. This resemblance is incomplete, and the motions are still retrograde or direct waves, but not standing waves.

Figure 14 shows optima in for $\mu_b$ increased to 5. The retrograde wave is somewhat difficult to discern at $\mu_t = 3$, without consulting the curvature plot. The motions are qualitatively similar to those for $\mu_b = 3$, shown in figure 11.
Figure 13. Snapshots of optimal snake motions at various $\mu_t$ (one per column, labeled at the top), for $\mu_b = 2$.

Figure 15 shows optima for $\mu_b$ increased to 20. For $\mu_t = 0.2$ we show two local optima. The second column shows a “swing-and-anchor” motion. The snake forms a loop at one end, the “anchor,” and then the rest of the body swings forward, rotating about the anchor. The motion also shows, briefly, a direct wave. This example shows that optima can combine multiple types of motions within a single period of motion. We have seen many other such “mixed” optima at moderate $\mu_t$, but they are not reproduced here for brevity. The other $\mu_t = 0.2$ optimum, shown in the first column, is essentially a direct wave. The optima for $\mu_t = 0.3, 0.7,$ and 3 also involve
the formation of “anchors” near the endpoints, somewhat similar to the concertina motion of biological snakes [5]. The optimum at $\mu_t = 2$ is a very symmetrical ratcheting motion.

### E. Supplementary Material: Triangular wave motion across $\mu_t$

We again consider the triangular wave motion (7.1)–(7.3), but now $\mu_t$ is nonzero, so (7.4) takes the more general form

$$\int H \frac{\partial \hat{s}}{\partial X} \cdot \hat{s} x + \mu_t \frac{\partial \hat{n}}{\partial X} \cdot \hat{n} n x \ ds = 0,$$

(A 1)

$$\mathcal{H} \equiv \left( \mu_f H(\frac{\partial \hat{s}}{\partial X} \cdot \hat{s}) + \mu_b (1 - H(\frac{\partial \hat{s}}{\partial X} \cdot \hat{s})) \right)$$

(A 2)
Figure 15. Snapshots of optimal snake motions at various $\mu_t$ (one per column, labeled at the top), for $\mu_b = 20$.

with $s$ as in (7.5) and

$$\hat{n} = \left( -A_0 \text{sgn}(\sin(2\pi(s-t))) \right).$$

We solve (A 1) for $U$:

$$U = -\frac{(-\mathcal{H} + \mu_t)A_0^2 \sqrt{1 - A_0^2}}{\mathcal{H} + (\mu_t - \mathcal{H})A_0^2}.$$ 

(A 4)
When $\mu_t = 0$, (A 4) becomes (7.6). The tangential speed is

$$u_s = U s_x + \partial_t y y_g = -\frac{A_0^2 \mu_t}{\mathcal{H} + (\mu_t - \mathcal{H}) A_0} \leq 0,$$  \hspace{1cm} (A 5)$$

so $\mathcal{H} = \mu_b$. The snake speed (A 4), as well as $u_s$ and $u_n$, are functions of $A_0$ and $\mu_t/\mu_b$ only. By (2.16) and (2.17), so is $\eta/\mu_b$. In figure 16 we give a contour plot of $\log_{10} \eta/\mu_b$ as a function of $q = \arcsin A_0$ and $\log_{10} \mu_t/\mu_b$. $q$ is the angle that the triangle wave makes with the horizontal axis.

![Contour plot of $\log_{10} \eta/\mu_b$ as a function of the triangular wave angle $q = \arcsin A_0$ and the friction coefficient ratio $\log_{10} \mu_t/\mu_b$.](image)

Figure 16. Contour plot of $\log_{10} \eta/\mu_b$ as a function of the triangular wave angle $q = \arcsin A_0$ and the friction coefficient ratio $\log_{10} \mu_t/\mu_b$.

The forward speed $U$ is zero, and the cost of locomotion infinite, when $\mu_t = \mu_b$, represented by a horizontal line near the middle of figure 16. This is an example of the “singular point” mentioned in the previous section, when $\mu_t = \mu_b = \mu_f = 1$. When $\mu_t > \mu_b$, above the horizontal line, $U < 0$, and the snake moves as a retrograde wave. $\eta/\mu_b$ decays to 1 as $\mu_t/\mu_b \to \infty$. In this regime, for fixed $\mu_t$, $\eta$ itself is the product of $\mu_b$ with a function that decays more slowly than $\mu_b$ as $\mu_b \to 0$. Thus $\eta$ is minimized when $\mu_b \to 1$, the minimal value of $\mu_b$. We plot the optimal amplitude in the limit of large $\mu_t/\mu_b$ (derived for a wave of any shape in section 5) as a dashed line at the upper left of the figure, with the formula for the optimal amplitude in this limit. In the lower portion of figure 16, $\mu_t < \mu_b$, $U > 0$, and the snake moves as a direct wave. Here $\eta/\mu_b$ decays to 0 as $\mu_t/\mu_b \to 0$. In this regime, for fixed $\mu_t$, $\eta$ itself is the product of $\mu_b$ with a function that decays more rapidly than $\mu_b^{-1}$ as $\mu_b \to \infty$. Thus $\eta$ is minimized by taking $\mu_b \to \infty$ and $\mu_t \to 0$. The wave
angle which minimizes $\eta$ can be found by directly minimizing $\eta$ with respect to $q$. We find that in the limit $\mu_t/\mu_b \to 0$, the optimal wave angle tends to $\pi/2$, i.e. the optimal wave becomes steeper.