CrossMark

JM PAPERS

# Optimizing low-Reynolds-number predation via optimal control and reinforcement learning

**Guangpu Zhu[1],‡, Wen-Zhen Fang[2],[1],‡ and Lailai Zhu[1],†**

[1]Department of Mechanical Engineering, National University of Singapore, 117575, Republic of Singapore

[2]Key Laboratory of Thermo-Fluid Science and Engineering, MOE, Xi'an Jiaotong University, Xi'an 710049, PR China

We seek the best stroke sequences of a finite-size swimming predator chasing a non-motile point or finite-size prey at low Reynolds number. We use optimal control to seek the globally optimal solutions for the former and reinforcement learning (RL) for general situations. The predator is represented by a squirmer model that can translate forward and laterally, rotate and generate a stresslet flow. We identify the predator's best squirming sequences to achieve the time-optimal (TO) and efficiency-optimal (EO) predation. For a point prey, the TO squirmer executing translational motions favours a two-fold *L*-shaped trajectory that enables it to exploit the disturbance flow for accelerated predation; using a stresslet mode expedites significantly the EO predation, allowing the predator to catch the prey faster yet with lower energy consumption and higher predatory efficiency; the predator can harness its stresslet disturbance flow to suck the prey towards itself; compared to a translating predator, its compeer combining translation and rotation is less time-efficient, and the latter occasionally achieves the TO predation via retreating in order to advance. We also adopt RL to reproduce the globally optimal predatory strategy of chasing a point prey, qualitatively capturing the crucial two-fold attribute of a TO path. Using a numerically emulated RL environment, we explore the dependence of the optimal predatory path on the size of prey. Our results might provide useful information that help in the design of synthetic microswimmers such as *in vivo* medical microrobots capable of capturing and approaching objects in viscous flows.

**Key words:** micro-organism dynamics, swimming/flying, microscale transport

† Email address for correspondence: lailai_zhu@nus.edu.sg
‡ G. Zhu and W.-Z. Fang contributed equally to this work.

## 1. Introduction

Approaching or chasing a moving target via optimal control has been a common task in natural and human settings, when e.g. animals like lions and sharks forage for prey, phagocytes chase and kill bacteria, predatory bacteria feed on other bacteria (Dashiff *et al.* 2011; Pérez *et al.* 2016), missiles intercept invading aircraft, and shooters aim for running targets. The optimal foraging of natural creatures may remain elusive; however, similar chasing applications in defence and robotic systems have become mature thanks to the optimal control theory.

Among these scenarios, the controlled agent and the target that it approaches are in a dry environment such as air or in a liquid-filled wet environment. In the air, the agent and target, unless closely gapped or in a specific configuration (e.g. a missile in the wake of a high-speed aircraft), may not affect the motion of each other effectively by disturbing the air flow; that is, they can sense weakly the additional aerodynamic force induced by the motion of the other. In a liquid environment, the interaction between the agent and target is no longer weak because the viscosity of liquid is larger than that of air by three orders. Hence a moving agent in liquid such as water can disturb its surrounding flow that exerts a considerable hydrodynamic force on its target nearby. This feature results in hydrodynamic interactions between the agent and target, which can influence significantly the chasing dynamics and the associated optimal chasing strategies. The effects of hydrodynamic interactions become especially pronounced and more long-ranged when the agent approaches/chases its targets in a low-Reynolds-number flow. This scenario occurs commonly for microorganisms or millimetre-scaled organisms swimming to approach motile particulate objects (e.g. bacteria or phytoplankton cells) and non-motile counterparts such as organic debris (Kiørboe *et al.* 2009). Both type of swimmer and size of target span a wide range (Jabbarzadeh & Fu 2018): a typical planktonic grazer is much larger than its prey (Hansen, Bjornsen & Hansen 1994; Kiørboe 2016); an organism swims towards a similarly sized member of the same species during bacterial conjugation (Clark & Adelberg 1962) or mating of copepods (Strickler 1998); the target can also be much larger than the approacher, exemplified by a spermatozoon swimming towards the egg or marine microbes targeting biological debris for nutrients uptake and habitation (Kiørboe 2003). Besides such natural events, similar situations might arise in the applications of future medical microrobots, which need to approach targets such as bacteria and human cells of varying sizes (Nelson, Kaliakatsos & Abbott 2010; Ceylan *et al.* 2019). In these low-Reynolds-number flow configurations featuring important, long-ranged hydrodynamic interactions, prior explored predator–prey dynamics of territory/aerial animals or high-Reynolds-number aquatic animals together with the optimal chasing/approaching strategies would not apply. These tiny swimmers have evolved a variety of unique strategies suited for the viscous environment. For instance, zooplankton achieve feeding by means of ambushing (Kiørboe *et al.* 2009), generating currents (Fenchel 1980), cruising (Kiørboe *et al.* 2009) and colonizing marine snow aggregates.

A decent understanding of the predator–prey dynamics in viscous low-Reynolds-number flows would benefit analysing the predatory and evasive behaviours of microorganisms, and exploring their evolutionary advantages. Also, designing the optimal predatory and evasive strategies will be potentially useful in manipulating future medical microrobots to capture bacteria or escape from hostile immune cells. Apart from a substantial amount of studies on a related topic – nutrients uptake and feeding of swimming microorganisms

(Magar, Goto & Pedley 2003; Langlois *et al.* 2009; Michelin & Lauga 2011; Tam & Hosoi 2011; Lambert *et al.* 2013; Kiørboe *et al.* 2014; Dölger *et al.* 2017; Andersen & Kiørboe 2020) – work has addressed the interaction between a swimming predator and an individual particle or prey nearby. Without considering swimming-induced hydrodynamic effects, Sengupta, Kruppa & Löwen (2011) proposed and investigated a discrete chemotactic predator–prey model that describes a chasing predator and an escaping prey, which sense the diffused chemicals released from each other. Pushkin, Shum & Yeomans (2013) studied theoretically and numerically the advection of a tracer and a material sheet of tracers when a microswimmer moves along an infinite straight path. Mathijssen, Jeanneret & Polin (2018) combined experiments, theory and simulations to perform a deep analysis of hydrodynamic entrainment of a particle by a swimming microorganism. Using a bispherical coordinate system, Jabbarzadeh & Fu (2018) studied analytically the scenario of a forced spherical particle approaching another; they also investigate numerically the head-on approach of a self-propelling swimmer to another passive particle. Słomka *et al.* (2020) conducted a modelling study on the ballistic encounter between elongated model bacteria and a much larger marine snow particle that is sedimenting. Very recently, Borra *et al.* (2022) have studied a pair of point predator and prey considering their hydrodynamic interactions; they used a multi-agent reinforcement learning scheme to explore efficient, physically explainable predatory and evasive strategies. Besides these works, Visser & Kiørboe (2006) has discussed the influence of ballistic or diffusive mobility patterns of planktonic organisms on the encounter rates between them and their prey, mates or predators.

In this work, we explore the optimal strategies of a finite-size swimming predator chasing a non-motile prey represented by a tracer point or a finite-size sphere. The motion of the tracer is driven purely by the propulsion-induced disturbance flow of the predator, whereas for the spherical prey, we consider the two-way hydrodynamic coupling between the predator and prey based on numerical simulations. To seek the most time-saving or energy-efficient pursuing strategies of the predator, we adopt a numerical optimal control approach for a point prey and reinforcement learning (RL) for general cases. The RL-based optimal solutions agree qualitatively with and capture the essential features of the globally optimal solutions identified by the former. We will demonstrate the emergence of non-intuitive optimal solutions in the seemingly simple configurations. We will also interpret physical mechanisms of the optimal strategies and discuss their implication on developing synthetic microrobots designed for capturing moving objects.

## 2. Problem set-up, assumptions and methods

### 2.1. *Problem set-up*

We consider a microscale predator that swims to approach a prey in low-Reynolds-number flows (see figure 1*a*). A spherical squirmer of radius *A* is adopted to model the predator, which attains propulsion and rotation based on its surface actuation described by a slip velocity $\bar{\boldsymbol{u}}_s(\theta', \phi')$, where $\theta' \in [0, \pi]$ and $\phi' \in [0, 2\pi]$ are the polar and azimuthal angles with respect to the squirmer's swimming orientation $\boldsymbol{e}_s$; here, $\boldsymbol{e}_s$ coincides with $\boldsymbol{e}_{z'}$ of the reference coordinates system $\boldsymbol{e}_{x'y'z'}$ translating and rotating with the squirmer. The prey is modelled by a passively moving point tracer or a finite-size spherical particle of radius $\mathcal{A}$. The ratio $\chi = \mathcal{A}/A$ is defined to indicate the size of prey compared to that of predator, which is zero for a point prey. From here on, ‾ is used to denote dimensional variables. It is worth noting that the squirmer model was proposed by Lighthill (1952) and Blake (1971) for ciliary propulsion of microorganisms such as *Paramecium* and *Volvox*. This model
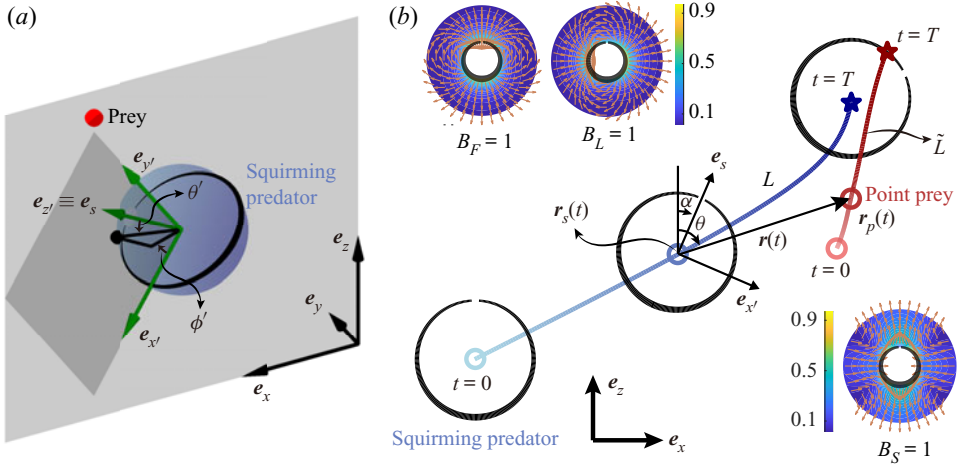
Figure 1. (*a*) A squirming predator chases a point prey on the $y = 0$ plane (finite-size spherical prey will be investigated in § 4.2). Here, $e_{x'y'z'}$ denotes the local coordinate system translating and rotating with the squirmer, while $e_{xyz}$ is the global counterpart. (*b*) The predator captures the prey at time $t = T$, when the latter is within a cut-off distance $\varepsilon \ll 1$ away from the surface of the former. The coordinates of the squirmer centre and prey are $r_s$ and $r_p$, respectively; $r = r_p - r_s = r(\sin\theta\, e_x + \cos\theta\, e_z)$ indicates their relative displacement. The orientation $e_s$ of the squirmer deviates from $e_z$ by an angle $\alpha$. The travelling distances of the predator and prey are denoted by $L$ and $\tilde{L}$, respectively. The insets show the disturbance velocity fields in the $y = 0$ plane (in the lab frame) generated by a squirmer using the $B_F = 1$, $B_L = 1$ and $B_S = 1$ modes. The variables here are all dimensionless.

has been used successfully to study microscale propulsion in the context of rheological complexity (Datt *et al.* 2015; Lintuvuori, Würger & Stratford 2017; Li, Lauga & Ardekani 2021), stratified fluids (More & Ardekani 2020), viscosity gradients (Datt & Elfring 2019), effects of boundaries (Spagnolie & Lauga 2012; Ishimoto & Gaffney 2013; Zhu, Lauga & Brandt 2013), suspension of active particles (Ishikawa, Brumley & Pedley 2021), and so on.

Now we describe our predator–prey problem. At time $\bar{t} = 0$, the orientation $e_s$ of the squirming predator is in the $e_z$ direction; the predator's centre located at $\bar{r}_s$, and the prey's centre at position $\bar{r}_p$, are in the $\bar{y} = 0$ plane. To simplify the setting, we assume a symmetric surface actuation velocity $\bar{u}_s$ about the plane, resulting in zero $\bar{y}$ component of the predator' swimming velocity and of its disturbance velocity in this plane. The latter implies that the prey will remain in the plane, as shown in figure 1(*b*); the angle $\alpha$ between the squirmer's orientation $e_s$ and the $e_z$ axis indicates that $e_s = \sin\alpha\, e_x + \cos\alpha\, e_z$. Also, we assume that the predator detects the instantaneous position of the prey, following the previous works on intelligently controlled swimmers (Gazzola *et al.* 2016; Mirzakhanloo, Esmaeilzadeh & Alam 2020). This assumption is made here to minimize the complexity of the problem, but we should note that most natural predators or prey do not 'see' the nearby motile organisms but detect their presence based on the hydrodynamic (Jakobsen 2001; Visser 2001) or chemical (Svensen & Kiørboe 2000) cues. The predator is actuated by four squirming modes, $\bar{B}_F(\bar{t})$, $\bar{B}_L(\bar{t})$, $\bar{B}_R(\bar{t})$ and $\bar{B}_S(\bar{t})$; the first two modes allow it to translate forward and laterally, respectively, with respect to its orientation $e_s$, corresponding to the $e_{z'}$ and $e_{x'}$ directions; $\bar{B}_R$ enables its rotation about the $e_y$ axis; $\bar{B}_S$ introduces a stresslet flow. We bound the strength of the surface actuations as $\bar{B}_i \in [-\bar{B}_i^{max}, \bar{B}_i^{max}]$, $i = F, L, R, S$. These modes will be described in their dimensionless form below.

We choose $A$ and $4\bar{B}_F^{max}/A^3$ as the characteristic length and velocity, respectively. To ease the calculation, we define the dimensionless displacement $\boldsymbol{r} = \boldsymbol{r}_p - \boldsymbol{r}_s$ between the predator and prey, which remains on the $y = 0$ plane and can be described by its magnitude $r = |\boldsymbol{r}|$ and the angle $\theta$ between $\boldsymbol{r}$ and $\boldsymbol{e}_z$ as $\boldsymbol{r} = r(\sin\theta\,\boldsymbol{e}_x + \cos\theta\,\boldsymbol{e}_z)$. The dimensionless slip velocity $\boldsymbol{u}_s(\theta', \phi')$ on the surface of the squirmer in its own reference frame $\boldsymbol{e}_{x'y'z'}$ reads (Pak & Lauga 2014)

$$\boldsymbol{u}_s = -\left(\frac{3\sin\theta'}{2}B_F + \frac{3\cos\theta'\cos\phi'}{2}B_L + \frac{3\cos\phi'}{4}B_R + \frac{9\sin 2\theta'}{8}B_S\right)\boldsymbol{e}_{\theta'}$$
$$+ \left(\frac{3\sin\phi'}{2}B_L + \frac{3\cos\theta'\sin\phi'}{4}B_R\right)\boldsymbol{e}_{\phi'}. \tag{2.1}$$

We will seek the best time sequences of the predator's surface actuations $[\bar{B}_F, \bar{B}_L, \bar{B}_R, \bar{B}_S](\bar{t})$ leading to the optimal predation. A standard optimum goal is to minimize the predating time $\bar{T}$. We call such an optimization the time-optimal (TO) optimization. In addition to the capture time, we are also concerned about and will optimize the predatory energy efficiency $\eta$ defined as

$$\eta = \frac{6\pi\mu A(\bar{d}_0/\bar{T}) \times \bar{d}_0}{\bar{E}}, \tag{2.2}$$

where $\mu$ is the dynamic viscosity of the fluid, $\bar{d}_0 = |\bar{\boldsymbol{r}}(\bar{t} = 0)| - A - \mathcal{A}$ denotes the initial surface-to-surface distance between prey and squirmer, and $\bar{T}$ and $\bar{E}$ represent the time and energy used by the predator to capture the prey, respectively. The numerator of (2.2) indicates the energy consumption for dragging a dead predator over a distance $\bar{d}_0$ to reach the prey within the capture time $\bar{T}$. Using $4\pi\mu\bar{B}_F^{max}/3A$ as the characteristic energy scale, we write

$$\eta = \frac{6d_0^2}{TE}, \tag{2.3}$$

where $E = \int_0^T P(t)\,\mathrm{d}t$, and $P$ denotes the dimensionless power consumption of the squirmer. Accordingly, maximizing the energy efficiency $\eta$ is termed the efficiency-optimal (EO) optimization problem.

## 2.2. *Optimal control for a point prey*

We adopt a numerical optimal control approach when the prey is modelled by a point tracer, as described here. The point prey is advected passively by the flow, hence the translational and rotational velocities of the predator and the translational velocity of the prey in dimensionless form can be derived as

$$\frac{\mathrm{d}\boldsymbol{r}_s}{\mathrm{d}t} = (B_L\cos\alpha - B_F\sin\alpha)\,\boldsymbol{e}_x + (-B_F\cos\alpha - B_L\sin\alpha)\,\boldsymbol{e}_z, \tag{2.4a}$$

$$\frac{\mathrm{d}\alpha}{\mathrm{d}t} = \frac{3}{4}B_R, \tag{2.4b}$$

$$\frac{\mathrm{d}\boldsymbol{r}_p}{\mathrm{d}t} = \left(\frac{1}{r^3}[-B_F\cos(\theta - \alpha) + B_L\sin(\theta - \alpha)] + \frac{9}{8}B_S\frac{r^2 - 1}{r^4}\left[3\cos^2(\theta - \alpha) - 1\right]\right)\boldsymbol{e}_r$$
$$+ \left(\frac{1}{2r^3}[-B_L\cos(\theta - \alpha) - B_F\sin(\theta - \alpha)] - \frac{9}{8}B_S\frac{1}{r^4}\sin 2(\theta - \alpha)\right)\boldsymbol{e}_\theta. \tag{2.4c}$$

Also, the dimensionless power consumption of the squirmer reads

$$P = 12 \left( B_F^2 + B_L^2 \right) + \tfrac{27}{2} B_S^2. \tag{2.5}$$

Since $d_0 = |r(t = 0)| - 1 - \chi = |r(t = 0)| - 1$ (realizing that $\chi = \mathcal{A}/A = 0$ for a point prey) does not change over time, maximizing the predating efficiency $\eta$ is equivalent to minimizing the product of time $T$ and energy $E$. The optimal control problem for a point prey becomes: given an initial relative displacement $r(t = 0) = r_0 = r_0(\sin\theta_0\, e_x + \cos\theta_0\, e_z)$ with $|r_0| > 1$ between the predator and the prey, the squirming predator will capture the prey at time $t = T$ when $|r(t = T)| - 1 \le \varepsilon$; that is, the point prey is within a small cut-off distance $\varepsilon \ll 1$ from the predator's surface. The small parameter $\varepsilon$ is introduced here for theoretical convenience: when the prey moves very close to the squirmer's surface, $|r| \to 1$, the relative velocity between them will approach zero because our squirmer adopts only the tangential but no radial surface actuation; hence the prey will never exactly touch the squirmer's surface mathematically. In real situations, when they are sufficiently close, other physical ingredients would come into play; for example, diffusion via Brownian motion would allow them to touch each other (Jabbarzadeh & Fu 2018). In this work, we will use a fixed value $\varepsilon = 4 \times 10^{-3}$ unless otherwise specified. We have checked that varying $\varepsilon$ in the range $[10^{-3}, 10^{-1}]$ would not alter the optimal chasing paths qualitatively, though the capture time will increase with decreasing $\varepsilon$ as anticipated. Without losing generality, initially the predator is oriented in the $e_z$ direction, i.e. $\alpha = 0$ when $t = 0$. This predatory process corresponds to the evolution of $r$ described by $r$ and $\theta$. Using $dr/dt = dr_p/dt - dr_s/dt$ and (2.4), we obtain the dynamical system characterizing the predatory process:

$$\frac{dr}{dt} = \frac{(-1 + r^3)\cos(\alpha - \theta)\, B_F + (-1 + r^3)\sin(\alpha - \theta)\, B_L}{r^3}$$
$$+ \frac{9(-1 + r^2)(1 + 3\cos[2(\alpha - \theta)])\, B_S}{16 r^4}, \tag{2.6a}$$

$$\frac{d\theta}{dt} = \frac{\sin(\alpha - \theta)\left[2r(1 + 2r^3)\, B_F + 9\cos(\alpha - \theta)\, B_S\right] - \cos(\alpha - \theta)(2r + 4r^4)\, B_L}{4 r^5}. \tag{2.6b}$$

We will seek the optimal sequences of the bounded actuation modes $[B_F, B_L, B_R, B_S](t)$ to minimize the capture time $t = T$ or to maximize the predating efficiency $\eta$ with these modes subject to

$$B_F \in [-1, 1], \tag{2.7a}$$

$$B_i \in [-B_i^{max}, B_i^{max}], \quad i = L, R, S. \tag{2.7b}$$

Unless otherwise specified, $B_L^{max} = B_R^{max} = B_S^{max} = 1$. This optimal control problem is solved numerically by an open source library 'FALCON.m' (Rieck *et al.* 1999) implemented in MATLAB. The state variables are discretized in time by the trapezoidal collocation method. The nonlinear optimization problem is solved by the built-in open source library IPOPT (Wächter & Biegler 2006).

## 2.3. *Reinforcement learning for a point or finite-size prey*

Besides considering a passively moving point prey, we will also model the prey as a finite-size spherical particle of a dimensionless radius $\chi = \mathcal{A}/A$ that would interact

hydrodynamically with the squirmer. The velocities of predator and prey cannot be derived analytically in the closed form as in (2.4), and thus will be solved numerically. Accordingly, it is inconvenient to use the numerical optimal control approach as for the point prey, and we will instead adopt a deep RL scheme to identify the optimal predatory strategy. Naturally, the RL scheme can also be applied for the point prey model, as we will demonstrate in § 4.

We will extend an extensively validated solver using the boundary integral method (BIM) to emulate the hydrodynamic scenario of a swimming squirmer approaching a spherical prey. Different variants of the solver have been developed to study the microlocomotion inside a tube (Zhu *et al.* 2013) or a droplet (Reigh *et al.* 2017), dynamics of a particle-encapsulated droplet in shear flow (Zhu & Gallaire 2017), and a sedimenting sphere near a corrugated wall (Kurzthaler *et al.* 2020). A brief description of the BIM implementation in its dimensionless form is provided below.

In the spirit of the BIM, we express the dimensionless velocity $u(x)$ at position $x$ everywhere in the domain as

$$u(x) = -\frac{1}{8\pi} \int_{S_s+S_p} q(x') \cdot G(x, x') \, dS(x'),$$ (2.8)

where $q$ is the the density of the so-called single-layer potential on the surface of the squirmer $S_s$ and that of the prey $S_p$. Here, $G(x, x') = I/|x - x'| + (x - x')(x - x')/|x - x'|^3$ is the free-space Green's function, which is also known as the Stokeslet. Both the squirmer and finite-size prey are spherical, which can be discretized by zero-order quadrilateral elements. For either of the two, the hydrodynamic force and torque exerted on it are zero. This condition is used to determine their translational and rotational velocities.

Having introduced the BIM implementation, we now describe the RL algorithm. Compared to the optimal control theory that requires the predator–prey dynamics (2.6), RL does not rely on prior knowledge of the dynamics but allows the squirmer as the predating agent to learn the dynamics, and adapt and optimize its chasing strategy (or policy in the language of RL) via interacting continuously with the environment. It is worth noting that RL algorithms have been used recently in similar swimming-involved scenarios, e.g. to optimize the swimming gaits or navigation routes of microswimmers at low Reynolds number (Colabrese *et al.* 2017; Schneider & Stark 2019; Mirzakhanloo *et al.* 2020; Tsang *et al.* 2020; Muiños-Landin *et al.* 2021; Nasiri & Liebchen 2022; Qiu *et al.* 2022) and in turbulent flows (Alageshan *et al.* 2020; Qiu *et al.* 2020), or macroscopic swimmers such as fish in viscous flows (Gazzola *et al.* 2016; Verma, Novati & Koumoutsakos 2018) or in the potential flow (Jiao *et al.* 2021). In particular, recent pioneering experiments (Muiños-Landin *et al.* 2021) have demonstrated using RL for real-time navigation of micron-sized thermophoretic particles, opening a new horizon for developing swimming microrobots endowed with artificial intelligence.

In this work, we adopt the open-source deep RL framework 'Tensorforce' (Kuhnle, Schaarschmidt & Fricke 2017) and use a policy-based RL scheme – proximal policy optimization (PPO) (Schulman *et al.* 2017) – to train the agent. The general idea behind the policy-based RL methods consists in parametrizing the policy function $\pi_\Theta$ by an artificial neural network (ANN) with a set of weights $\Theta$. The agent equipped with the parametrized policy identifies certain characteristic information, the state $s$ of the environment, as the input to the ANN, and then selects an action $a$ according to the ANN's output. For the predator–prey system considered here, the state that the predator can observe is the

position of the prey relative to itself ($r$ and $\theta$), and the actions of the agent are the bounded squirming modes $B_i$. The selected action advances the environment from the current to the next state, and its effectiveness is quantified by an instantaneous reward $R$. An appropriate reward function will favour the actions allowing the predator to approach the prey.

To achieve the TO predation, we choose the reward function

$$R = -r + \Gamma, \quad \text{where } \Gamma = \begin{cases} -1000, & r > 4, \\ 200/(t - \beta_T), & r \le (1 + \varepsilon), \\ 0, & \text{otherwise,} \end{cases} \quad (2.9)$$

where $t$ is equivalent effectively to the capture time $T$, and $\beta_T$ is the lower-bound estimation of $T$. Here, $-r$ promotes the predator to take only the necessary actions to approach the prey because any unnecessary ones will decrease the accumulated reward. The term $\Gamma$ contributes in two ways: first, it penalizes the agent for wandering far from the prey ($r > 4$) by activating a substantial negative reward; second, it stimulates the predator to expedite the capture by offering a positive reward $\propto 1/(t - \beta_T)$ upon a successful capture. Note that we choose $\beta_T \approx d_0$ as the initial surface-to-surface distance between predator and prey. For the EO setting, the reward function is

$$R = -r - \hat{\alpha}P + \Gamma, \quad \text{where } \Gamma = \begin{cases} -1000, & r > 4, \\ 20\,000/\left[t\,E(t) - \beta_E\right], & r \le (1 + \varepsilon), \\ 0, & \text{otherwise,} \end{cases} \quad (2.10)$$

where $\hat{\alpha}$ is a positive weight introduced to reduce the instantaneous power consumption $P$, and $\hat{\alpha} = 0.1$ is chosen here. Also, $\beta_E$ is a lower-bound estimation of $ET$, chosen as half the value for a squirmer with $B_F = 1$ travelling a distance of $d_0$.

Having defined the reward function, we describe the training process. The objective of RL here is to equip the agent with the optimal policy maximizing the expected cumulative reward. After the agent executes the current ($k$th) policy $\pi_{\Theta_k}$, we collect a set of trajectories $\{v_i\}$ (a trajectory is a sequence of states and actions, $v = \{s_0, a_0, \ldots, s_n, a_n, \ldots\}$) to determine the new $((k + 1)$th) policy $\pi_{\Theta_{k+1}}$ via

$$\Theta_{k+1} = \arg\max_{\Theta} \mathop{\mathbb{E}}_{s,a \sim \pi_{\Theta_k}} [L_{clip}(s, a, \Theta_k, \Theta)]. \quad (2.11)$$

Here, $L_{clip}$ is the so-called clipped surrogate advantage (Achiam 2018) that measures the performance of a general policy $\pi_\Theta$ relative to the current one $\pi_{\Theta_k}$:

$$L_{clip}[s, a, \Theta_k, \Theta] = \min\left\{p(\Theta)\hat{A}^{\pi_{\Theta_k}}(s, a), \text{clip}\left[p(\Theta), 1 - \hat{\epsilon}, 1 + \hat{\epsilon}\right]\hat{A}^{\pi_{\Theta_k}}(s, a)\right\}. \quad (2.12)$$

Here, $\mathbb{E}$ denotes the mathematical expectation, and $p(\Theta) = \pi_\Theta(a\,|\,s)/\pi_{\Theta_k}(a\,|\,s)$ is the probability ratio, where its numerator and denominator represent the probabilities of taking action $a$ in state $s$ at the $\pi_\Theta$ and $\pi_{\Theta_k}$ policies, respectively; $\hat{A}^{\pi_{\Theta_k}}(s, a)$ is the advantage function, describing the advantage of choosing a specific action $a$ in state $s$ over that of a random choice according to the current policy $\pi_{\Theta_k}$. The function clip[] indicates that $p(\Theta)$ is bounded in the range $[1 - \hat{\epsilon}, 1 + \hat{\epsilon}]$ via necessary clipping; here, the hyperparameter $\hat{\epsilon}$ indicates how much the new policy is allowed to deviate from the current one, and is fixed as $\hat{\epsilon} = 0.2$ in this work.

## 3. Results: optimal control for a point prey ($\chi = 0$)

We start to investigate a point prey with $\chi = 0$. In all cases, initially the squirming predator is aligned with the $e_z$ axis, i.e. $\alpha(t = 0) = 0$; the initial distance between predator and prey
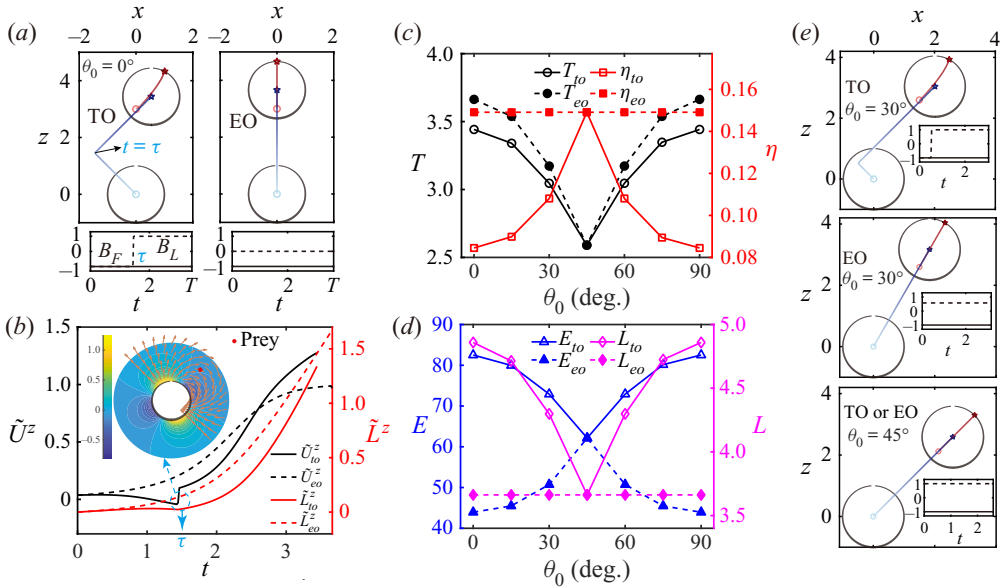
**944** A3-8

Figure 2. Time-optimal (TO) and efficiency-optimal (EO) predating strategies using combined forward and lateral squirming modes of magnitudes $B_F$ and $B_L$, respectively. (*a*) Trajectories of the predator (blue) and point prey (red), where the latter is initially ahead of the former ($\theta_0 = 0°$); circles and stars denote their initial and ending positions. The bottom images show how the two modes evolve in time. The left- and right-hand images depict the TO and EO strategies, respectively. (*b*) The prey's velocity $\tilde{U}^z$ and accumulative travelling distance $\tilde{L}^z$ in $e_z$; $\tau$ indicates when the TO predator takes the sharp turn. The inset illustrates its disturbance flow field in the $y = 0$ plane right before $t = \tau$. (*c*) Capture time $T$ and predating efficiency $\eta$ of the TO and EO predators versus the initial bearing $\theta_0$ of the prey with respect to the predator. (*d*) Similar to (*c*), but for the predator's consumed energy $E$ and travelling distance $L$. (*e*) The TO and EO predating behaviours when $\theta_0 = 30°$ and $45°$.

is $|r_0| = 3$, and $d_0 = |r_0| - 1 = 2$. We vary the initial bearing of the prey with respect to the predator, namely, the angle $\theta - \alpha$ at $t = 0$ between the predator–prey displacement vector $r$ and the squirmer's orientation $e_s$ (see figure 1). This initial angle recovers to $\theta_0$ by realizing $\alpha(t = 0) = 0$. Here, $\theta_0 = 0°$, $90°$ and $180°$ correspond to when the prey is in front, on the right side and in the rear of the predator, respectively.

### 3.1. *A predator combining forward and lateral motions*

We first consider a predator swimming forward and laterally, respectively, via the $B_F$ and $B_L$ squirming modes. This combination is denoted '$F + L$'. Without combining them, the predator with either of them alone can swim only vertically (in $e_z$) or horizontally (in $e_x$) and thus cannot reach the prey located in an arbitrary bearing $\theta_0$. Figure 2(*a*) compares the predator and prey trajectories of the TO and EO cases. The EO predator swims directly towards the prey by attaining the maximum forward movement ($B_F = -1$) and zero lateral movement ($B_L = 0$), which follows an intuitive predatory strategy by taking the straight thus shortest path. In contrast, the TO predator chooses an L-shaped route oriented first in the north-western direction then followed by a sharp $90°$ turn towards the north-eastern direction. Correspondingly, $B_F = -1$ producing the maximum forward motion holds during the whole course, while $B_L$ jumps sharply from $-1$ to $1$ at the turning time $t = \tau$. We now discuss the mechanism behind this peculiar strategy. Initially, the

predator lags behind the prey by a distance 3 in $e_z$. Both the TO and EO predators adopt $B_F = -1$ to maintain the same maximum movement in that direction; the difference in the capture time $T$ then depends on the prey's velocity $\tilde{U}^z$ in $e_z$. Figure 2(*b*) compares $\tilde{U}^z$ of the two prey and their accumulative travelling distances $\tilde{L}^z$ along the $e_z$ direction. An important observation is that the EO prey's $\tilde{U}^z$ remains positive, indicating its consistent motion away from the predator, whereas the TO prey's velocity $\tilde{U}^z$ becomes negative when $t < \tau$, implying its motion towards the predator. The inset of figure 2(*b*) depicts the instantaneous flow field around the squirmer right before $t = \tau$, reflecting the negative velocity $\tilde{U}^z$ experienced by the prey (red dot). To sum up, the EO predator swimming straight towards the prey generates a flow field that always repels the prey away from it; while the TO predator adjusts its position (with respect to the prey) and surface actuation for best exploiting its disturbance flow field to attract the prey, leading to the initially left-upward movement. In addition to the $\theta = 0°$ orientation, similar trajectories of the TO and EO predators are found for an arbitrary orientation $\theta_0 \neq 45°$ of the prey, as shown in figure 2(*e*).

We then examine how the initial orientation $\theta_0 \in [0, 90]°$ of the prey with respect to the predator affects the predating dynamics; the results for $\theta_0 \in [90, 180]°$ are not shown, by realizing the fore–aft symmetry. We depict the capture time $T$ and efficiency $\eta$ of the TO and EO strategies in figure 2(*c*), and the predator's energy consumption $E$ and travelling distance $L$ in figure 2(*d*). All the quantities exhibit a mirror symmetry about $\theta_0 = 45°$ when the prey is right in the north-eastern direction. In this particular configuration, the TO and EO predators adopt the same strategy – swimming straight towards the prey, as shown by figure 2(*e*). This symmetry can be anticipated because the forward $B_F$ or lateral $B_L$ mode alone allows the predator to approach the prey at $\theta_0 = 0°$ or $90°$, respectively. For the two modes sharing the same magnitude $B_F^{max} = B_L^{max} = 1$, the predatory scenario for the $90 - \theta_0$ orientation can be obtained by interchanging the time sequences of $B_F$ and $B_L$ for the $\theta_0$ counterpart. In addition, we see that the EO predators achieve an optimal efficiency of $\eta_{eo} \approx 0.15$ independent of $\theta_0$, which almost doubles that of the TO predator when $\theta_0 = 0°$ and $90°$. In contrast, the TO predator captures the prey slightly faster than the EO counterpart, reducing the capture time at most by around 0.2, which occurs when the prey is ahead of or beside the predator. Both the TO and EO predators catch their prey fastest that are initially in the north-eastern direction ($\theta_0 = 45°$), and take more time when they deviate from that orientation. Also, compared to the EO predator, the TO predator consumes more energy.

## 3.2. *The stresslet squirming mode facilitates predation*

Having observed how the predator exploits its squirming-induced disturbance flow to catch the prey faster, we then examine the influence of the stresslet mode $B_S$ known to vary the disturbance flow without affecting the swimming speed of an isolated squirmer (Blake 1971). According to our definition (with a sign difference compared to classical definitions), $B_S < 0$ corresponds to a puller microswimmer, e.g. the biflagellated algae *Chlamydomonas*; $B_S > 0$ indicates a pusher counterpart exemplified by most flagellated microorganisms. The puller attracts the fluid from its front and rear towards itself, while the pusher drives the flow oppositely. Compared to the baseline '$F + L$' predator using only the $B_F$ and $B_L$ modes, we show in figure 3 that introducing the stresslet mode $B_S$ can enhance significantly the all-round predatory performance under the EO policy. Figure 3(*a*) shows that the '$F + L + S$' (with $B_F$, $B_L$ and $B_S$ modes) predator captures the prey faster than the '$F + L$' (with $B_F$ and $B_L$ modes) competitor for all
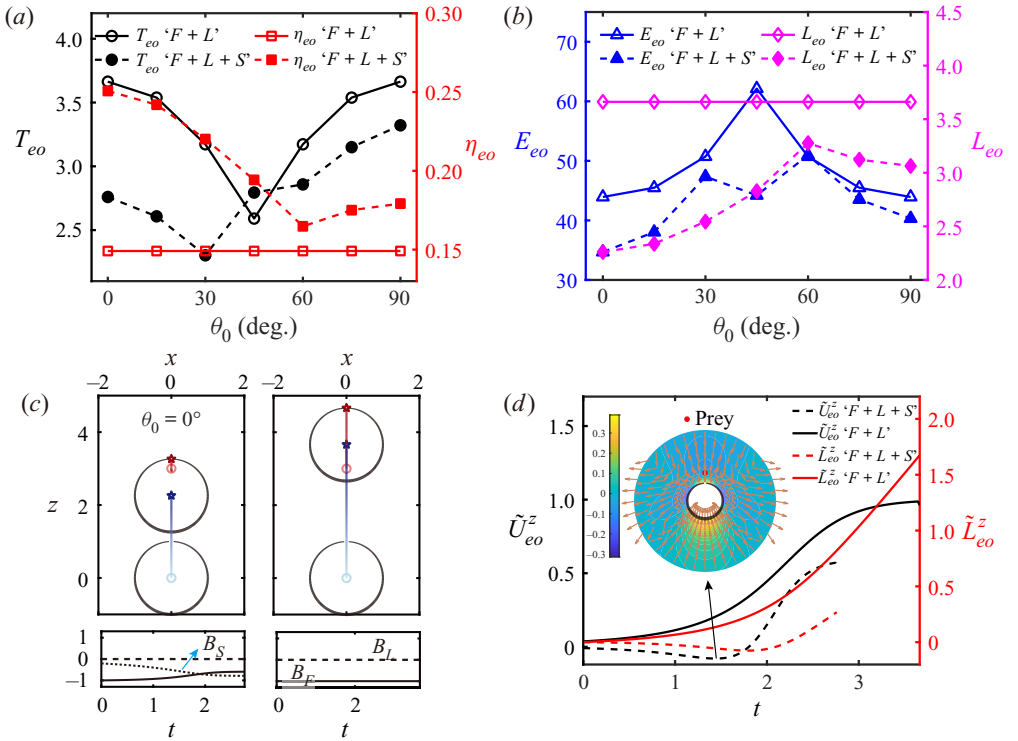
Figure 3. Adding a stresslet mode $B_S$ on top of the EO predator with $B_F$ and $B_L$ modes enhances its all-round performance. (*a*) Capturetime $T_{eo}$ and predatory efficiency $\eta_{eo}$ versus the initial bearing $\theta_0$ of the prey. (*b*) Similar to (*a*), but for the predator's energy consumption $E_{eo}$ and travelling distance $L_{eo}$. (*c*) Trajectories of the EO predator and prey with (left-hand image) and without (right-hand image) the $B_S$ mode; the bottom images show the time sequences of the squirming modes. (*d*) The prey's velocity $\tilde{U}_{eo}^z$ and accumulative travelling distance $\tilde{L}_{eo}^z$ in $e_z$, with and without the $B_S$ mode. The inset indicates the flow field when the $\tilde{U}_{eo}^z$ reaches the minimum.

the bearings of the prey $\theta_0$, except for when $\theta_0 = 45°$. Moreover, incorporating the stresslet enhances considerably the predatory efficiency $\eta_{eo}$ of the EO predators, with a maximum relative enhancement approaching 70 % when $\theta_0 = 0°$, as well as decreasing its energy consumption $E_{eo}$ and travelling distance $L_{eo}$ for all the prey's initial bearings (see figure 3*b*).

In particular, we examine in figure 3(*c*) the straight trajectories of EO predators with and without the stresslet mode, together with those of their respective prey. We observe that the prey has moved from the stresslet-equipped predator by a negligible distance compared to that chased by the stresslet-free one. The former predator with a negative $B_S$ mode (shown in the left bottom image of figure 3*c*) behaves as a puller swimmer, which sucks the prey ahead towards it significantly. The associated sucking disturbance flow generated by the predator is depicted in the inset of figure 3(*d*). This mechanism of stresslet-accelerated predatory process is also confirmed by the evident backward motion – negative $\tilde{U}_{eo}^z$ and $\tilde{L}_{eo}^z$ – of the prey shown in figure 3(*d*). In retrospect, it was found analogously that biflagellated organisms can enhance their feeding performance by adopting a puller-style locomotory gait (Dölger *et al.* 2017). As expected, when the prey is initially on the right
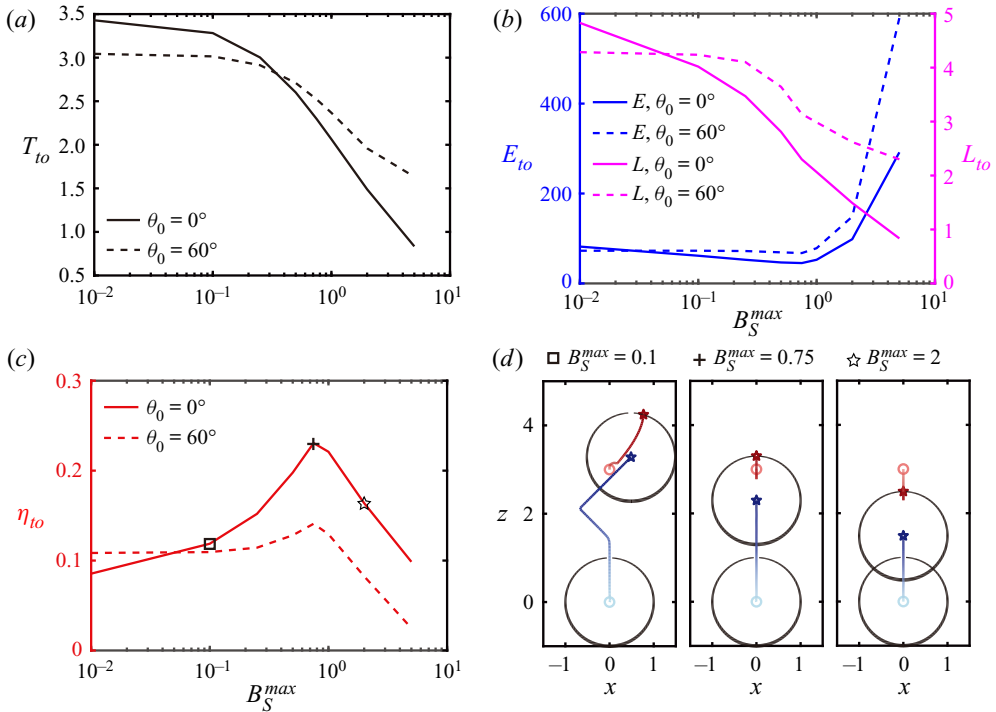
**944** A3-11

Figure 4. Varying the maximum magnitude $B_S^{max}$ of the stresslet mode for the TO predator that swims using the '$F + L + S$' combination of squirming modes. (*a*) Capture time $T_{to}$ versus $B_S^{max}$ for two initial orientations, $\theta_0 = 0°$ and $\theta_0 = 60°$, of the prey with respect to the predator. (*b*) Energy consumption $E_{to}$ and travelling distance $L_{to}$ versus $B_S^{max}$. (*c*) Similar to (*a*), but for the predating efficiency $\eta_{to}$. (*d*) Trajectories of the predator and prey for $B_S^{max} = 0.1$, 0.75 and 2 marked in (*c*).

side of the predator ($\theta_0 = 90°$), the latter would activate a positive $B_S$ mode; accordingly, the prey is attracted laterally towards to a pusher-style swimmer, which is not shown here.

Intuitively, we infer that a puller predator can exploit such a stresslet disturbance flow to accelerate capturing the prey. Hence the TO predator would naively turn on the full gear of $B_S$ mode for the fastest capture. This intuition is confirmed by figure 4(*a*), showing that $T_{to}$ decreases monotonically with increasing $B_S^{max}$ when $\theta_0 = 0°$ and 60°. On the other hand, the growing stresslet will produce higher power consumption of the predation due to the stronger viscous dissipation of the fluid. Figure 4(*b*) depicts that the energy consumption $E_{to}$ decreases weakly with $B_S^{max}$ when $B_S^{max} < 1$, but increases sharply with $B_S^{max} > 1$. The slightly negative relation between $E_{to}$ and $B_S^{max} < 1$ is due to: first, in this regime, the major power consumption is not from the stresslet flow but from the forward and lateral motions of squirmer; second, the decreasing time $T_{to}$ in this regime with $B_S^{max}$ tends to lower the total energy. When $B_S^{max}$ keeps growing from around 1, the stresslet-induced power becomes increasingly dominant, because the swimming power scales with these modes quadratically, while the forward ($B_F$) and lateral ($B_L$) modes are bounded to 1.

Next, we show in figure 4(*c*) that the predating efficiency $\eta_{to}$ depends non-monotonically on $B_S^{max}$, attaining the maximum when $B_S^{max}$ is approximately 0.7–0.8. This non-monotonic dependence can be explained by the variation of $T_{to}$ and $E_{to}$ according to $B_S^{max}$. We then examine in figure 4(*d*) three characteristic chasing scenarios for the $\theta_0 = 0°$ situation. In the case $B_S^{max} = 0.1$, the predator first approaches the prey straight forward. Then it takes

a two-fold zigzag path as a reminiscent of the TO predatory strategy in the absence of $B_S$ mode shown in figure 2($a$). The initial straight chasing reflects the predator's tendency to utilize the $B_S$-induced flow for sucking the prey. Increasing $B_S^{max}$ to 0.75 results in the optimal efficiency $\eta_{to} \approx 0.23$ that exceeds double the efficiency of the $B_S^{max} = 0.1$ and 5 predators. The suction flow of this $B_S^{max}$ level is strong enough to overcome the forward flow generated by the $B_F$ mode. Hence initially the prey moves backward towards the predator, and then it moves forward as their distance is decreased. When $B_S^{max}$ grows to 2, the stresslet-induced suction completely dominates the forward flow, hence enabling the prey to move continuously backward until being captured.

### 3.3. *Incorporating rotation or using only translations?*

In the above scenarios, the predator with a zero rotational mode $B_R = 0$ does not rotate, hence its orientation remains in the $\boldsymbol{e}_z$ direction. To capture a prey at an arbitrary orientation, the predator must activate both the forward and lateral translational modes. However, if allowed to rotate freely by adopting a non-zero $B_R$ mode, the predator needs only one translational mode. We then ask how such a combined rotational and translational mode compares with the combination of pure translational modes in the performance of TO predation. We show in figure 5($a$) the minimal capture time $T_{to}$, and in figure 5($b$) the corresponding predatory efficiency $\eta_{to}$ of the predator using three combinations of squirming modes: (1) forward plus rotational, '$F + R$'; (2) forward plus lateral, '$F + L$'; and (3) forward plus lateral plus rotational, '$F + L + R$'. The minimal capture time $T_{to}$ of these three combinations diminishes in order regardless of the prey's orientation $\theta_0$ with respect to the predator. Also, the '$F + L + R$' combination outperforms the other two in efficiency for most of the range of $\theta_0$. Moreover, for the '$F + R$' combination, $T_{to}$ (resp. $\eta_{to}$) increases (resp. decreases) with $\theta_0$ monotonically. This trend is in stark contrast to the symmetric (about $\theta_0 = 45°$) profiles of $T_{to}(\theta_0)$ and $\eta_{to}(\theta_0)$ for the '$F + L$' and '$F + L + R$' combinations.

We then analyse the detailed chasing dynamics for a better understanding. We first illustrate in figures 5($c$–$e$) how the '$F + R$' predator chases its prey oriented at $\theta_0 = 0°, 45°$ and $90°$. Intuitively, to reach an arbitrarily oriented prey, the predator using one rather than two translational modes has to rotate to align its swimming direction exactly towards the prey. For a special case when the prey with $\theta_0 = 0°$ is initially ahead of the predator, no rotational motion is required for the predator, as shown in figure 5($c$). When $\theta_0 = 45°$, the predator adopts a rotational mode $B_R = 1$ in full gear during $t \in [0, \tau_1]$, and then switches it off to swim straight towards the prey during $t > \tau_1$. It is revealed that for a non-zero $\theta_0$, the capture time comprises two parts: the first for rotational orientation and the second for straight swimming. The first part for orientation clearly increases with the initial angular difference $\theta_0$, which explains the monotonically increasing capture time $T_{to}$ with $\theta_0$. A less intuitive scenario occurs for $\theta_0 = 90°$, when the prey is initially on the predator's right side: first, the predator moves backward and rotates rightward simultaneously during $t < \tau_1$, with both modes in full gear; then it stops the backward translation, while maintaining the full right rudder until $t = \tau_2$, when it exactly faces the prey; finally, the predator swims straight forward to the prey. The initial backward movement of predator seems awkward, which tends to retard predating by lengthening the predator–prey distance at first glance. In fact, this seemingly awkward strategy embodies the wisdom of retreating in order to advance. Moving backward actually reduces the angle $\theta - \alpha$ (i.e. $\theta_0$ at $t = 0$ shown in figure 1$b$) between the predator–prey displacement
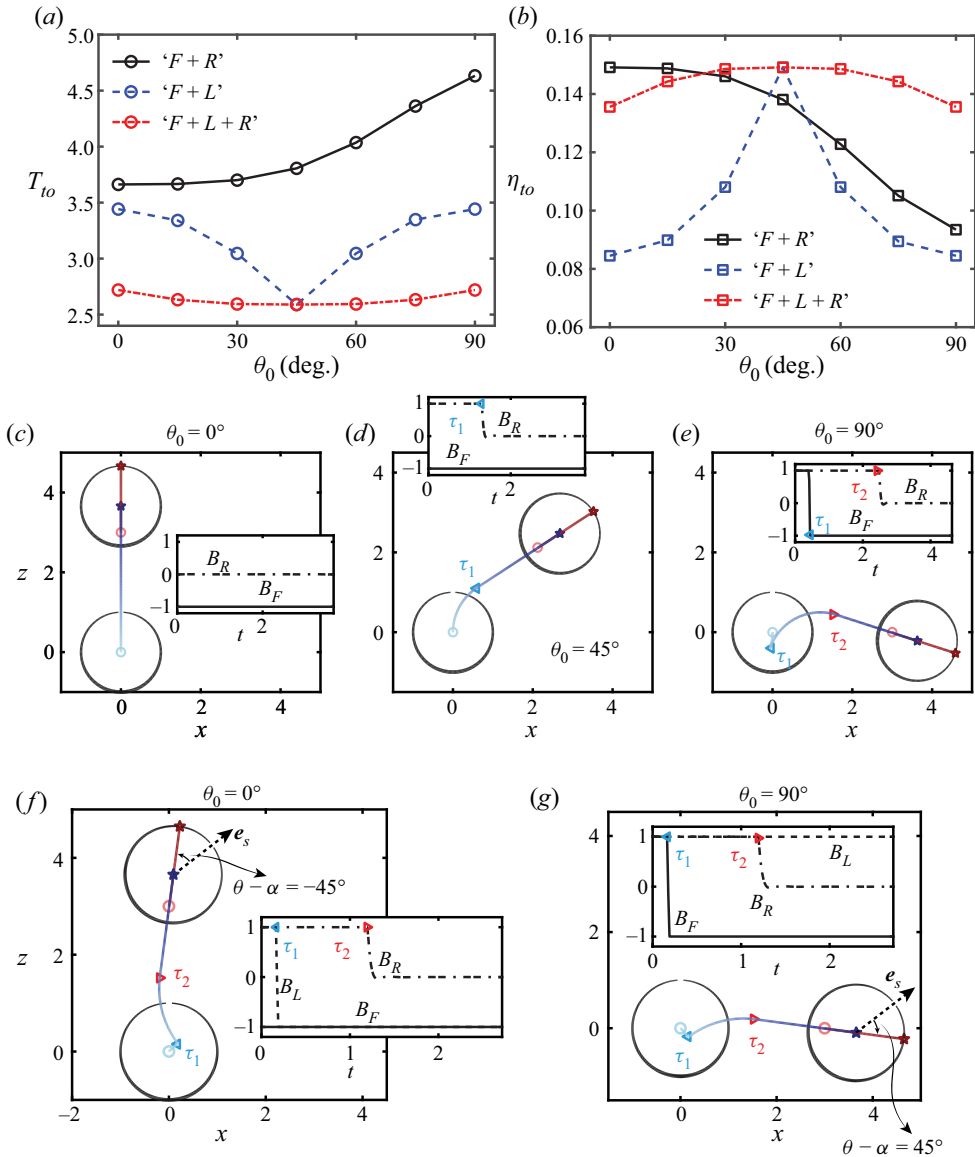
Figure 5. Effects of the rotational mode $B_R$ of the squirmer on its TO predatory performance. (*a*) Capture time $T_{to}$, and(*b*) predating efficiency $\eta_{to}$, are plotted versus the initial orientation $\theta_0$ of the prey with respect to the predator. Three combinations of squirming modes, '$F+R$', '$F+L$' and '$F+L+R$', are adopted. Chasing dynamics of the '$F+R$' mode are illustrated for (*c*) $\theta_0 = 0°$, (*d*) $\theta_0 = 45°$, and (*e*) $\theta_0 = 90°$; those of the '$F+L+R$' case are shown for (*f*) $\theta_0 = 0°$,and (*g*) $\theta_0 = 90°$. The insets plot the time evolutions of the modes.

and the predator's orientation, thus decreasing the time needed for orientation to achieve effectively a net time saving.

As discussed above, replacing the lateral mode by the rotational mode, i.e. shifting from the '$F+L$' to '$F+R$' combination, results in asymmetric distributions of $T_{to}(\theta_0)$ and $\eta_{to}(\theta_0)$ about $\theta_0 = 45°$. As reflected by the distinctive chasing dynamics for $\theta_0 = 0°$, $\theta_0 = 45°$ and 90°, this asymmetry is caused by the predator's necessity for a rotational

**944** A3-14

orientation to face the prey. Hence we postulate that the '$F + L + R$' squirming predator might exhibit similar asymmetric profiles of $T_{to}(\theta_0)$ and $\eta_{to}(\theta_0)$ owing to the rotational mode at play. In reality, this postulation is disproved by the symmetric profiles shown in figures 5(*a,b*), which can be elucidated by scrutinizing how the '$F + L + R$' predator chases its prey initially at $\theta_0 = 0°$ and $90°$ as depicted by figures 5(*f*) and 5(*g*), respectively. The set of trajectories of the predator and prey for $\theta_0 = 0°$ matches that for $\theta_0 = 90°$ in shape, and applying a $90°$ rotational transformation allows them to overlap each other. In contrast to the '$F + R$' predator rotating to exactly face the prey when $\theta - \alpha = 0°$, this '$F + L + R$' predator also orients itself but instead to face the prey sitting in its north-west ($\theta - \alpha = -45°$) or north-east ($\theta - \alpha = 45°$) direction before swimming straight towards the prey. This particular magnitude of $45°$ enables the predator to exploit its maximum translational speed, $\sqrt{(B_F^{max})^2 + (B_L^{max})^2} = \sqrt{2}$, to reach the prey in the post-rotation period $t > \tau_2$. This maximum translational speed exploited by the '$F + L + R$' predator leads to its faster predation compared to the '$F + R$' and '$F + L$' counterparts, as shown in figure 5(*a*). Indeed, the latter two can translate at a maximum velocity of 1 rather than $\sqrt{2}$. We comment that the difference between the initial bearing $\theta_0 = 0°$ and $90°$ explains the clockwise and counter-clockwise rotations of the predator, resulting in $\theta - \alpha = -45°$ and $45°$, respectively. This reasoning also justifies the $90°$ rotational mapping between the two sets of trajectories associated with the two bearings.

## 4. Results: RL for a point prey and a finite-size prey

### 4.1. *RL-based optimization in the case of a point prey*

To substantiate our study, we adopt an RL scheme to seek the optimal predatory strategies for a predator limited to the '$F + L$' squirming modes. Before studying a finite-size prey ($\chi > 0$), we first address the scenario of a point prey ($\chi = 0$), where the optimal solutions based on the optimal control approach (see figure 2) can be regarded as the globally optimal solutions for benchmarking. As shown in figure 6, both the minimal capture time $T_{to}$ and maximum efficiency $\eta_{eo}$ obtained by RL agree well with their counterparts by optimal control. Also as expected, a close inspection of figure 6 shows that RL performs slightly worse than the optimal control, which implies further that the latter has indeed provided the globally optimal solutions. Comparing further in figure 7 the learned trajectories of the TO predator and its prey to their counterparts based on optimal control, we observe that the RL-trained TO predator learns to execute a two-fold zigzag path identified by the optimal control approach. In particular, when $\theta_0 = 0°$, the trajectories obtained from the two approaches almost collapse on each other; the sharp turn in the trajectory and the associated steep jump in the lateral model (action) $B_L$ are captured quantitatively by RL, as shown in figures 7(*a,d*). However, as the initial bearing $\theta_0$ increases to $15°$ and $30°$, the RL solutions can reproduce the two-fold path only qualitatively, but fail to capture its sharp turn or the sudden jump of the swimming action (see figures 7(*b,c,e,f*)). In fact, the degrading performance of RL at a larger bearing angle $\theta_0 < 45°$ can be rationalized. For a two-fold path identified by optimal control, we define its two-foldedness $\tau/T$ as the time $\tau$ when the predator turns sharply scaled by the capture time. The foldedness decreases monotonically with $\theta_0 \in [0, 45]°$, becoming zero at $\theta_0 = 45°$ corresponding to a straight chasing path (see the inset of figure 6*b*). This trend implies that time saving gained by executing a two-fold path diminishes with increasing $\theta_0$; or, in other words, the extra time required by executing the straight path as a sub-optimal solution instead of the globally optimal version decreases with growing $\theta_0 < 45°$.

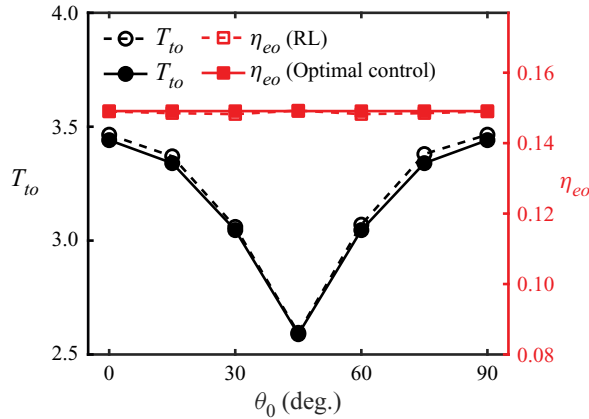Figure 6. Using RL to obtain the minimal capture time $T_{to}$ and maximum efficiency $\eta_{eo}$ of an '$F + L$' predator chasing a point prey. The results are compared against those based on the optimal control approach.

Hence at a sufficiently large $\theta_0$ featuring a negligible difference in the capture time between the sub-optimal and globally optimal strategies, it becomes challenging for the RL algorithm to pinpoint exactly the globally optimal one.

Returning to figure 7, strictly speaking, we will not regard our RL-trained strategies as globally optimal when $\theta_0 = 15°$ and $30°$. On the other hand, they indeed capture the essential features – two-fold path – of the globally optimal solutions. This promising observation thus motivates us to employ RL to optimize the predating strategy of a squirmer chasing a finite-size prey when the globally optimal solutions are not available, as we will present in § 4.2. We add two comments before proceeding. First, despite the RL solutions deviating from the globally optimal ones at increasing $\theta_0 < 45°$, when $\theta_0 = 45°$, featuring a globally optimal straight path depicted in figure 2(*e*), RL can again reproduce exactly this solution. Also, in perfect agreement with the optimal control approach, RL can identify the straight paths reaching the optimal efficiency regardless of the initial bearing of the prey. These optimal straight paths are not shown here. Second, we have realized the crucial role of $\Gamma$ (in (2.9)) that represents a positive reward upon the capture time. Without this reward, the RL approach results in straight chasing trajectories regardless of the initial orientation $\theta_0$ of the prey, evidently being trapped in locally optima.

### 4.2. *RL-based optimization in the case of a finite-size prey ($\chi > 0$)*

Here, we use RL to optimize the predatory strategies of an '$F + L$' squirmer for capturing a finite-size spherical prey of dimensionless radius $\chi > 0$. For the BIM-based RL environment, we use a cut-off distance $\epsilon = 5 \times 10^{-2}$ larger than the previous value $4 \times 10^{-3}$ due to the degrading accuracy of BIM as involving two sufficiently close surfaces. The finite-size effect of the prey does not change the typical two-foldedness of the TO chasing path, as exemplified by a $\chi = 0.25$ prey initially at $\theta = 0°$ and $30°$ (see figures 8(*a*) and 8(*b*), respectively). By examining further in figure 8(*c*) the $\chi$-dependent TO predatory trajectory for $\theta_0 = 15°$, we observe its decreasing two-foldedness in the case of a larger prey. For a sufficiently large prey of $\chi = 6$, the TO path becomes straight eventually. We here provide a phenomenological understanding of this change. As discussed in § 3.1, the two-fold path executed by the TO '$F + L$' predator enables
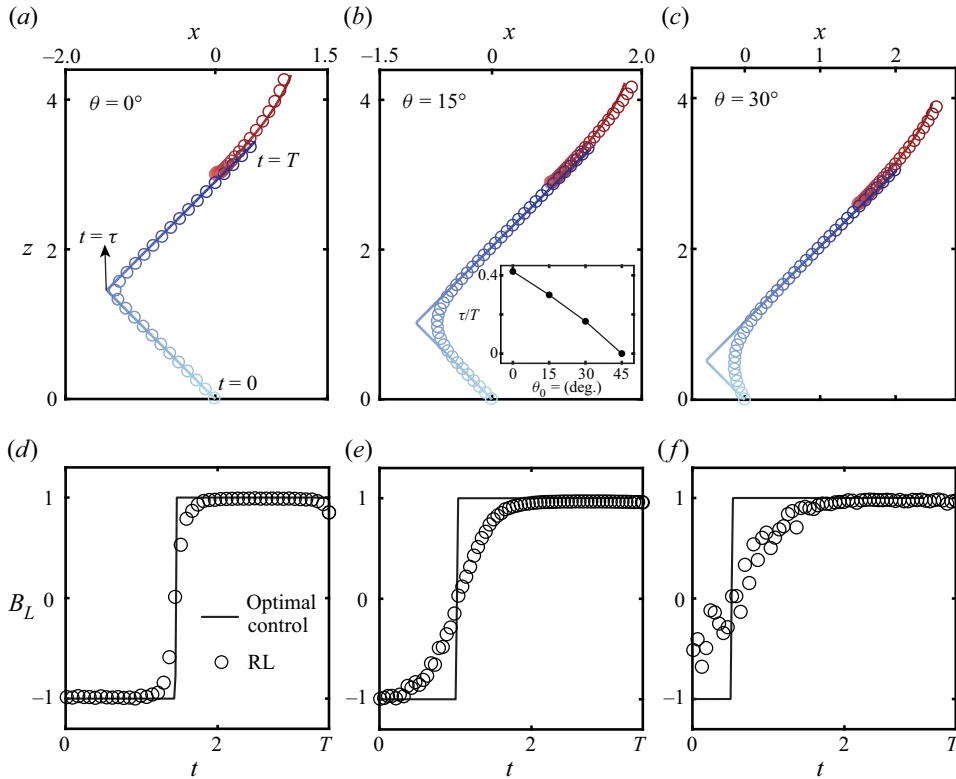
**944** A3-16

Figure 7. Trajectories of the TO predators (blue) and their point prey (red) with initial bearing (*a*) $\theta = 0°$, (*b*) $\theta = 15°$, and (*c*) $\theta = 30°$, based on optimal control (curves) and RL (symbols). The time evolutions of the lateral mode (action) $B_L$ of the predators are shown in (*d*), (*e*) and (*f*), respectively. The forward mode $B_F \cong -1$ as in figure 2(*a*) for both methods, and is thus not shown here. The inset of (*b*) shows $\tau/T$ versus $\theta_0$ of the optimal control solutions.

exploiting its propulsion-induced disturbance flow to adjust the advection of the prey. The effect, however, diminishes for a larger prey being more difficult to advect due to its increased hydrodynamic resistance coefficient. This trend will be responsible for the decreasing two-foldedness of the TO predatory path with increasing $\chi$. Furthermore, we depict in figure 8(*d*) the dependence of the minimal capture time $T_{to}$ on the prey size $\chi$. Despite the unchanged symmetric profile of $T_{to}(\theta_0)$ regardless of the varying $\chi$, the capture time $T_{to}$ declines monotonically with the prey size. This negative relation can be rationalized by realizing that the disturbance flow of the '$F + L$' squirmer, overall, expels the prey away from it, as evidenced by the prey's path (see figures 7 and 8*a*,*b*). Therefore, a larger expelled prey travels a shorter distance, reducing the capture time. We also note that the wide range of chosen $\chi$ has been motivated by the various realistic scenarios introduced in § 1.

## 5. Conclusion

In this work, we study, in the creeping flow regime, a swimming predator modelled by a spherical squirmer chasing a non-motile point or finite-size spherical prey advected by the disturbance flow generated by the former. Using optimal control for a point prey and RL for general situations, we optimize the predatory strategies of the squirmer that can
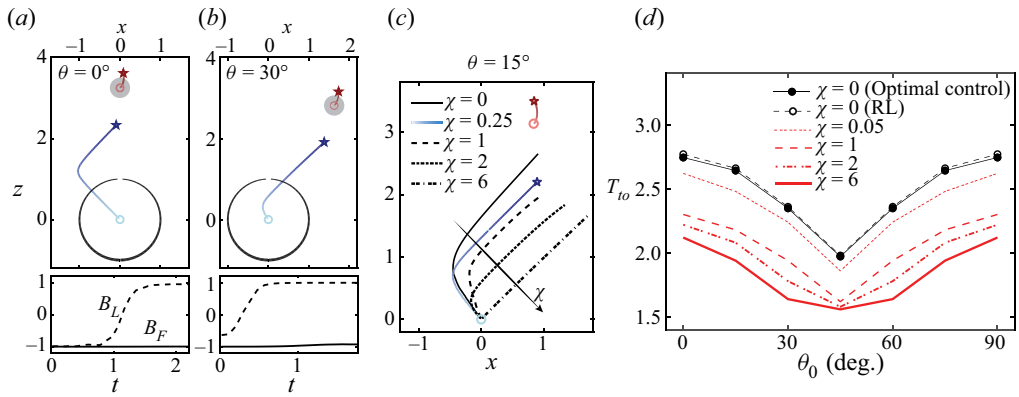
Figure 8. Trajectories of the RL-trained TO '$F + L$' predator and its finite-size spherical prey of radius $\chi = 0.25$ initially located at (*a*) $\theta = 0°$ and (*b*) $\theta = 30°$; the lateral $B_L$ and forward $B_F$ modes of the squirmer are depicted in the bottom images. The circles (resp. stars) denote the initial (resp. final) positions of the predator and prey. (*c*) Dependence of the TO trajectory on the prey size $\chi \in [0, 6]$ when $\theta_0 = 15°$. (*d*) Minimal capture time $T_{to}$ versus initial bearing $\theta_0$ and size $\chi$ of the prey; $\chi = 0$ corresponds to a point prey studied via both optimal control and RL, whereas the results for a finite-size prey ($\chi > 0$) are obtained by RL. For all cases here, the cut-off distance is $\varepsilon = 5 \times 10^{-2}$, and the initial surface-to-surface distance between predator and prey is $|d_0| = 2$ as in § 3.

translate forward ('$F$') and laterally ('$L$'), generating a stresslet ('$S$') flow or rotating ('$R$') in the fluid. We have identified the best time sequences of the squirming modes to achieve time-optimal (TO) or efficiency-optimal (EO) predating, in order to minimize the capture time or to maximize the predatory efficiency, respectively.

We first focus on a point prey. The EO '$F + L$' predator swims straight towards the prey regardless of its initial bearing with respect to the predator. In contrast, the TO predator follows an L-shaped route, hence travelling a longer distance than the EO predator. This chasing strategy can be understood by examining how the disturbance flow of the predator advects the prey: the EO predator generates a flow that persistently expels the prey away from it; the TO counterpart has been optimized to adapt its orientation with respect to the prey, such that its disturbance flow can be harnessed to advect the latter towards itself to some extent. This peculiar route may not be revealed easily intuitively. Also, we show that incorporating an additional stresslet mode of magnitude $B_S^{max} = 1$ allows the '$F + L + S$' EO predator to outperform considerably the '$F + L$' counterpart in every aspect of the predatory performance; generally, the former captures the prey faster, consumes less energy, travels a shorter distance, and gains a higher predatory efficiency (see figure 3). We recall that for an isolated squirmer, introducing the stresslet mode does not change its speed, but does increase the energy expenditure and decrease its efficiency (Blake 1971). For predation, the counterintuitive energy-saving and efficiency enhancement result from the predator's largely reduced capture time and travelling distance. This reduction is most pronounced when the prey is initially ahead of the predator, achieved by utilizing the stresslet flow to suck the prey towards itself. A similar scenario was revealed by Tam & Hosoi (2011) for a biflagellated swimmer that exploits its strokes-induced currents to achieve optimal nutrients uptake. Then we have examined how the maximum magnitude $B_S^{max}$ of the stresslet mode influences the performance of an '$F + L + S$' TO predator. Increasing the magnitude reduces the predator's capture time and travelling distance as expected; however, it increases significantly the consumed energy. The competition results

in the non-monotonic variation of the predatory efficiency versus $B_S^{max}$; accordingly, the TO predator attains the highest predatory efficiency at an optimal value $B_S^{max} \approx 0.7$–0.8. In addition, we have also investigated the potential role of rotational motion in the TO predation. Compared to a translating '$F + L$' predator, the '$F + R$' counterpart combining the forward translation and rotation spends more time catching the prey, while the '$F + L + R$' squirmer using two translational motions and rotation seizes the prey faster than the former two compeers. Unlike the '$F + L$' TO predator following an L-shaped path, the '$F + R$' compeer first rotates to face the prey exactly and then swims straight to it (see figure 5*d*). Thus the total capture time comprises two parts – one used for rotational reorientation, and the other for straight chasing. For a prey initially exactly on the right side requiring a considerable rotation, the TO predator has been optimized to adopt a non-intuitive strategy of retreating in order to advance: it first swims backwards, leaving rather than approaching the prey in appearance; this trick reduces effectively the angular difference between its orientation and the prey's bearing, and thus the corresponding time for rotation, leading to a net time saving. We comment that the activated rotational motion of the predator would fail to represent the actual situation due to the unphysical zero power consumption of this motion. Future improvement in modelling is needed to account for the reorientation of the predator.

Besides optimal control, we have used RL to seek the optimal strategies of an '$F + L$' squirmer chasing a point prey or a spherical one of radius $\chi > 0$. For the latter, a BIM solver is developed to emulate the RL environment. For a point prey, we have observed that our RL-based solutions reproduce perfectly the globally optimal EO strategies (featuring straight chasing paths) derived by optimal control, while for the TO counterpart, the former qualitatively (even quantitatively in some cases) agree with the globally optimal ones. In particular, RL can capture the non-intuitive two-foldedness of the TO path. Applying RL to a spherical prey of radius $\chi > 0$, we have identified that the two-foldedness of the TO path decreases with increasing $\chi$, and the TO path eventually becomes straight at a sufficiently large $\chi$. We have also shown that the minimal capture time decreases monotonically with $\chi$ because a larger prey is more difficult to be advected by the predator.

It is worth noting that despite the substantial amount of work applying RL to optimize the locomotory gaits or path planning of different swimmers, to the best of our knowledge, no individual work implies that the RL-trained swimming strategy represents or resembles the globally optimal one. It is indeed well-known that RL is easily trapped to local optima (Liepins & Vose 1991; Lehman & Stanley 2011; Sutton & Barto 2018). The only exception might be the very recent work Nasiri & Liebchen (2022), having used RL to find asymptotically optimal navigating strategies of a point swimmer, which replicate closely the globally optimal solutions identified previously by Daddi-Moussa-Ider, Löwen & Liebchen (2021). Together with Daddi-Moussa-Ider *et al.* (2021) and Nasiri & Liebchen (2022), our work indicates that RL-based optimization of swimming gaits or paths can be trapped to locally-optimal solutions as anticipated, but it can also identify the global optima by using a proper RL implementation.

We finally discuss the squirmer as the model predator. The model is adopted here for its mathematical simplicity and important features such as time-dependent strokes and the resulting disturbance flows, as well as its finite-size effect. This model does not capture the critical merits of certain predatory microflagellates in planktonic environments. The marine microflagellates are mostly sessile predators rather than free swimmers as studied here, and they intercept the prey or nutrient particles by using their flagella, filters or slender tentacles (Fenchel 1986) instead of chasing directly the prey or particles.

**Declaration of interests.** The authors declare no conflicts of interests.

**Author ORCIDs.**

Lailai Zhu https://orcid.org/0000-0002-3443-0709.

**Author contributions.** L.Z. designed the research. L.Z. and G.Z. built the numerical implementation. W.F. and G.Z. performed the research. All authors analysed the results and made the figures. L.Z. and G.Z. wrote the manuscript.

## REFERENCES

ACHIAM, J. 2018 Spinning up in deep reinforcement learning. https://spinningup.openai.com/en/latest/.

ALAGESHAN, J.K., VERMA, A.K., BEC, J. & PANDIT, R. 2020 Machine learning strategies for path-planning microswimmers in turbulent flows. *Phys. Rev.* E **101** (4), 043110.

ANDERSEN, A. & KIØRBOE, T. 2020 The effect of tethering on the clearance rate of suspension-feeding plankton. *Proc. Natl Acad. Sci. USA* **117** (48), 30101–30103.

BLAKE, J.R. 1971 A spherical envelope approach to ciliary propulsion. *J. Fluid Mech.* **46** (1), 199–208.

BORRA, F., BIFERALE, L., CENCINI, M. & CELANI, A. 2022 Reinforcement learning for pursuit and evasion of microswimmers at low Reynolds number. *Phys. Rev. Fluids* **7** (2), 023103.

CEYLAN, H., YASA, I.C., KILIC, U., HU, W. & SITTI, M. 2019 Translational prospects of untethered medical microrobots. *Prog. Biomed. Engng* **1** (1), 012002.

CLARK, A.J. & ADELBERG, E.A. 1962 Bacterial conjugation. *Annu. Rev. Microbiol.* **16** (1), 289–319.

COLABRESE, S., GUSTAVSSON, K., CELANI, A. & BIFERALE, L. 2017 Flow navigation by smart microswimmers via reinforcement learning. *Phys. Rev. Lett.* **118** (15), 158004.

DADDI-MOUSSA-IDER, A., LÖWEN, H. & LIEBCHEN, B. 2021 Hydrodynamics can determine the optimal route for microswimmer navigation. *Commun. Phys.* **4** (1), 15.

DASHIFF, A., JUNKA, R.A., LIBERA, M. & KADOURI, D.E. 2011 Predation of human pathogens by the predatory bacteria *Micavibrio aeruginosavorus* and *Bdellovibrio bacteriovorus*. *J. Appl. Microbiol.* **110** (2), 431–444.

DATT, C. & ELFRING, G.J. 2019 Active particles in viscosity gradients. *Phys. Rev. Lett.* **123** (15), 158006.

DATT, C., ZHU, L., ELFRING, G.J. & PAK, O.S. 2015 Squirming through shear-thinning fluids. *J. Fluid Mech.* **784**, R1.

DÖLGER, J., NIELSEN, L.T., KIØRBOE, T. & ANDERSEN, A. 2017 Swimming and feeding of mixotrophic biflagellates. *Sci. Rep.* **7** (1), 39892.

FENCHEL, T. 1980 Suspension feeding in ciliated protozoa: structure and function of feeding organelles. *Arch. Protistenkd.* **123** (3), 239–260.

FENCHEL, T. 1986 The ecology of heterotrophic microflagellates. In *Advances in Microbial Ecology* (ed. K.E. Nelson), pp. 57–97. Springer.

GAZZOLA, M., TCHIEU, A.A., ALEXEEV, D., DE BRAUER, A. & KOUMOUTSAKOS, P. 2016 Learning to school in the presence of hydrodynamic interactions. *J. Fluid Mech.* **789**, 726–749.

HANSEN, B., BJORNSEN, P.K. & HANSEN, P.J. 1994 The size ratio between planktonic predators and their prey. *Limnol. Oceanogr.* **39** (2), 395–403.

ISHIKAWA, T., BRUMLEY, D.R. & PEDLEY, T.J. 2021 Rheology of a concentrated suspension of spherical squirmers: monolayer in simple shear flow. *J. Fluid Mech.* **914**, A26.

ISHIMOTO, K. & GAFFNEY, E.A. 2013 Squirmer dynamics near a boundary. *Phys. Rev.* E **88** (6), 062702.

JABBARZADEH, M. & FU, H.C. 2018 Viscous constraints on microorganism approach and interaction. *J. Fluid Mech.* **851**, 715–738.

JAKOBSEN, H.H. 2001 Escape response of planktonic protists to fluid mechanical signals. *Mar. Ecol. Prog. Ser.* **214**, 67–78.

Jiao, Y., Ling, F., Heydari, S., Kanso, E., Heess, N. & Merel, J. 2021 Learning to swim in potential flow. *Phys. Rev. Fluids* **6** (5), 050505.

Kiørboe, T. 2003 Marine snow microbial communities: scaling of abundances with aggregate size. *Aquat. Microb. Ecol.* **33** (1), 67–75.

Kiørboe, T. 2016 Fluid dynamic constraints on resource acquisition in small pelagic organisms. *Eur. Phys. J.* **225** (4), 669–683.

Kiørboe, T., Andersen, A., Langlois, V.J., Jakobsen, H.H. & Bohr, T. 2009 Mechanisms and feasibility of prey capture in ambush-feeding zooplankton. *Proc. Natl Acad. Sci. USA* **106** (30), 12394–12399.

Kiørboe, T., Jiang, H., Gonçalves, R.J., Nielsen, L.T. & Wadhwa, N. 2014 Flow disturbances generated by feeding and swimming zooplankton. *Proc. Natl Acad. Sci. USA* **111** (32), 11738–11743.

Kuhnle, A., Schaarschmidt, M. & Fricke, K. 2017 Tensorforce: a tensorflow library for applied reinforcement learning. Web page. https://tensorforce.readthedocs.io/en/latest/.

Kurzthaler, C., Zhu, L., Pahlavan, A.A. & Stone, H.A. 2020 Particle motion nearby rough surfaces. *Phys. Rev. Fluids* **5** (8), 082101.

Lambert, R.A., Picano, F., Breugem, W. -P. & Brandt, L. 2013 Active suspensions in thin films: nutrient uptake and swimmer motion. *J. Fluid Mech.* **733**, 528–557.

Langlois, V.J., Andersen, A., Bohr, T., Visser, A.W. & Kiørboe, T. 2009 Significance of swimming and feeding currents for nutrient uptake in osmotrophic and interception-feeding flagellates. *Aquat. Microb. Ecol.* **54** (1), 35–44.

Lehman, J. & Stanley, K.O. 2011 Novelty search and the problem with objectives. In *Genetic Programming Theory and Practice IX* (ed. R. Riolo, E. Vladislavleva & J.H. Moore), pp. 37–56. Springer.

Li, G., Lauga, E. & Ardekani, A.M. 2021 Microswimming in viscoelastic fluids. *J. Non-Newtonian Fluid Mech.* **197**, 104655.

Liepins, G.E. & Vose, M.D. 1991 Deceptiveness and genetic algorithm dynamics. In *Foundations of Genetic Algorithms* (ed. G.J.E. Rawlins), vol. 1, pp. 36–50. Elsevier.

Lighthill, M.J. 1952 On the squirming motion of nearly spherical deformable bodies through liquids at very small Reynolds numbers. *Commun. Pure Appl. Maths* **5**, 109–118.

Lintuvuori, J.S., Würger, A. & Stratford, K. 2017 Hydrodynamics defines the stable swimming direction of spherical squirmers in a nematic liquid crystal. *Phys. Rev. Lett.* **119** (6), 068001.

Magar, V., Goto, T. & Pedley, T.J. 2003 Nutrient uptake by a self-propelled steady squirmer. *Q. J. Mech. Appl. Maths* **56** (1), 65–91.

Mathijssen, A.J.T.M., Jeanneret, R. & Polin, M. 2018 Universal entrainment mechanism controls contact times with motile cells. *Phys. Rev. Fluids* **3** (3), 033103.

Michelin, S. & Lauga, E. 2011 Optimal feeding is optimal swimming for all Péclet numbers. *Phys. Fluids* **23** (10), 101901.

Mirzakhanloo, M., Esmaeilzadeh, S. & Alam, M.-R. 2020 Active cloaking in Stokes flows via reinforcement learning. *J. Fluid Mech.* **903**, A34.

More, R.V. & Ardekani, A.M. 2020 Motion of an inertial squirmer in a density stratified fluid. *J. Fluid Mech.* **905**, A9.

Muiños-Landin, S., Fischer, A., Holubec, V. & Cichos, F. 2021 Reinforcement learning with artificial microswimmers. *Sci. Rob.* **6** (52), eabd9285.

Nasiri, M. & Liebchen, B. 2022 Reinforcement learning of optimal active particle navigation. arXiv:2202.00812.

Nelson, B.J., Kaliakatsos, I.K. & Abbott, J.J. 2010 Microrobots for minimally invasive medicine. *Annu. Rev. Biomed. Engng* **12**, 55–85.

Pak, O.S. & Lauga, E. 2014 Generalized squirming motion of a sphere. *J. Engng Maths* **88** (1), 1–28.

Pérez, J., Moraleda-Muñoz, A., Marcos-Torres, F.J. & Muñoz-Dorado, J. 2016 Bacterial predation: 75 years and counting! *Environ. Microbiol.* **18** (3), 766–779.

Pushkin, D.O., Shum, H. & Yeomans, J.M. 2013 Fluid transport by individual microswimmers. *J. Fluid Mech.* **726**, 5–25.

Qiu, J., Huang, W., Xu, C. & Zhao, L. 2020 Swimming strategy of settling elongated micro-swimmers by reinforcement learning. *Sci. China* **63** (8), 284711.

Qiu, J., Mousavi, N., Gustavsson, K., Xu, C., Mehlig, B. & Zhao, L. 2022 Navigation of micro-swimmers in steady flow: the importance of symmetries. *J. Fluid Mech.* **932**, A10.

Reigh, S.Y., Zhu, L., Gallaire, F. & Lauga, E. 2017 Swimming with a cage: low-Reynolds-number locomotion inside a droplet. *Soft Matt.* **13** (17), 3161–3173.

Rieck, M., Bittner, M., Grüter, B., Diepolder, J. & Piprek, P. 1999 Falcon.m user guide.

Schneider, E. & Stark, H. 2019 Optimal steering of a smart active particle. *Europhys. Lett.* **127** (6), 64003.

SCHULMAN, J., WOLSKI, F., DHARIWAL, P., RADFORD, A. & KLIMOV, O. 2017 Proximal policy optimization algorithms. arXiv:1707.06347.

SENGUPTA, A., KRUPPA, T. & LÖWEN, H. 2011 Chemotactic predator–prey dynamics. *Phys. Rev.* E **83** (3), 031914.

SŁOMKA, J., ALCOLOMBRI, U., SECCHI, E., STOCKER, R. & FERNANDEZ, V.I. 2020 Encounter rates between bacteria and small sinking particles. *New J. Phys.* **22** (4), 043016.

SPAGNOLIE, S.E. & LAUGA, E. 2012 Hydrodynamics of self-propulsion near a boundary: predictions and accuracy of far-field approximations. *J. Fluid Mech.* **700**, 105–147.

STRICKLER, J.R. 1998 Observing free-swimming copepods mating. *Phil. Trans. R. Soc. Lond.* B **353** (1369), 671–680.

SUTTON, R.S. & BARTO, A.G. 2018 *Reinforcement Learning: An Introduction*. MIT Press.

SVENSEN, C. & KIØRBOE, T. 2000 Remote prey detection in *Oithona similis*: hydromechanical versus chemical cues. *J. Plankton Res.* **22** (6), 1155–1166.

TAM, D. & HOSOI, A.E. 2011 Optimal feeding and swimming gaits of biflagellated organisms. *Proc. Natl Acad. Sci. USA* **108** (3), 1001–1006.

TSANG, A.C.H., TONG, P.W., NALLAN, S. & PAK, O.S. 2020 Self-learning how to swim at low Reynolds number. *Phys. Rev. Fluids* **5** (7), 074101.

VERMA, S., NOVATI, G. & KOUMOUTSAKOS, P. 2018 Efficient collective swimming by harnessing vortices through deep reinforcement learning. *Proc. Natl Acad. Sci. USA* **115** (23), 5849–5854.

VISSER, A. 2001 Hydromechanical signals in the plankton. *Mar. Ecol. Prog. Ser.* **222**, 1–24.

VISSER, A.W. & KIØRBOE, T. 2006 Plankton motility patterns and encounter rates. *Oecologia* **148** (3), 538–546.

WÄCHTER, A. & BIEGLER, L.T. 2006 On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math. Program.* **106** (1), 25–57.

ZHU, L. & GALLAIRE, F. 2017 Bifurcation dynamics of a particle-encapsulating droplet in shear flow. *Phys. Rev. Lett.* **119** (6), 064502.

ZHU, L., LAUGA, E. & BRANDT, L. 2013 Low-Reynolds-number swimming in a capillary tube. *J. Fluid Mech.* **726**, 285–311.