LICENTIATE THESIS

Smart microswimmers in complex flows

NAVID MOUSAVI

Department of Physics
University of Gothenburg
Göteborg, Sweden 2022

Smart microswimmers in complex flows
Navid Mousavi

ISBN 978-91-8009-831-1 (PRINT)
ISBN 978-91-8009-832-8 (PDF)

This thesis is electronically published, available at
https://hdl.handle.net/2077/71497
Department of Physics
University of Gothenburg
SE-412 96 Göteborg
Sweden
Telephone: +46 (0)31-786 00 00

Printed by Stema Specialtryck AB
Borås, Sweden 2022

ABSTRACT

Zooplankton are playing a pivotal role in balancing the life on Earth. By grazing on phytoplankton and then serving as food source to larger aquatic animals, they pave the way of redistributing the Sun's energy and form the second level of the aquatic food chain. They are also important for the global carbon cycle and, as a result, their effect upon the climate is another reason that highlights their importance.

Being small, they experience the flow as viscous. Nevertheless, they are able to each day migrate long distances efficiently. Their daily vertical migration is the largest natural migration of biomass on Earth, which is not well understood. In this thesis, we used a model to analyze the optimal navigation strategies for vertical migration of planktonic microswimmers in turbulent flows.

Passive strategies for vertical swimming, such as gyrotaxis, where the swimmer is bottom-heavy and hence obtains a tendency to point upwards, do not have as good performance in turbulent flows as they have in quiescent flows. We present here active mechanisms that a microswimmer, similar to a juvenile copepod, can exploit to significantly increase its vertical migration efficiency in turbulent flows. We find that the modeled swimmer utilizes different mechanisms in two and three spatial dimensions. In two dimensions, they mimic longer swimmers by actively reorienting. This results in an increase in the rate of upwelling region sampling, which leads to a significant increase in swimming speed against gravity. On the other hand, in three dimensions, it turns out that actively keeping the swimming direction aligned against gravity is more efficient. Both mechanisms are found to be robust to moderate perturbations of the flow and swimmer parameters, and they explain how swimmers that do not benefit from passive gyrotaxis can obtain notable vertical migration rates.

Keywords: microswimmer, turbulent flow, optimal navigation, diel vertical migration, reinforcement learning

List of papers

This thesis consists of an introductory text and the following two appended papers that can be found in Part III of the thesis:

Paper A

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

Paper B

Qiu, J., Mousavi, N., Zhao, L., & Gustavsson, K. Active gyrotactic stability of microswimmers using hydromechanical signals. *Phys. Rev. Fluids* **7**, 014311.

ACKNOWLEDGEMENTS

First and foremost I would like to express my gratitude to my supervisor Kristian Gustavsson, who gave me the great opportunity of working with him by accepting me as his student. He has always been patient with my mistakes which are infinite, and supportive in any aspect, during this weird time of the world, since I have started my PhD.

After that, I would like to thank my co-supervisor Giovanni Volpe and my examiner Bernhard Mehlig whose tireless enthusiasm in explaining physics is one of the best experiences I have ever had. I am also very grateful to our collaborators Jingran Qiu and Lihao Zhao at Tsinghua University for their contribution and hard work.

I also wish to thank my colleagues Anshuman and Ludvig for friendly discussions and particularly Ehsan for proof-reading my thesis and his friendship. I am also thankful to Saeed for the enjoyable talks on fluids, reinforcement learning and politics we had.

I express my thanks to my family, especially my father who tolerated many hardships and sacrificed a lot that I can study, and if it was not for him, things would have been very different in my life. I thank all friends who made and make life easier, and last but not least, I cannot thank Niloofar enough for her love and always present support. I am lucky to have her in my life.

CONTENTS

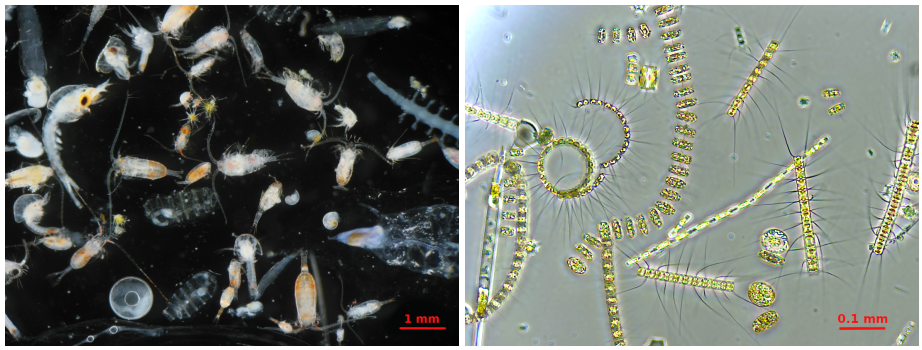
1 Introduction

Plankton! A name that we all have heard, but maybe not everyone knows how dependent our lives are on these creatures. Plankton are organisms, drifting around in water. The name is derived from the Greek word planktos, referring to wandering/drifting and was first used by German physiologist Victor Hensen (1835-1924) in 1887 [1]. Roughly 98% of the oceanic biomass consist of plankton [2], with a very diverse set of shapes, sizes, and behaviors. Their sizes range from viruses of less than 1 micron to siphonophores such as *Praya dubia* that can reach more than 50 meters when extending their fishing filaments [2].

Two of the main groups of planktonic swimmers are *phytoplankton* and *zooplankton* (Figure 1.1). Phytoplankton, are tiny unicellular creatures that use photosynthesis as their energy resource with oxygen as a residual [3]. In fact, we should be thankful to them for every other breath that we take in, since half of the oxygen on Earth is produced by these tiny plants. As they need to be exposed to the sunlight, phytoplankton live in the upper regions of the ocean. Zooplankton are on the other hand animals, meaning that they depend on other species as their food resource. The diversity of zooplankton is also very wide, being divided into 4 major groups based on the size, micro-zooplankton (2 to 200 μm), meso-zooplankton (200 μm to 2 mm), macro-zooplankton (2 to 20 mm) and mega-zooplankton ($>$ 20 mm) [4]. Many organisms live their whole life as a zooplankton, while others only spend a fraction of their life being a zooplankton, for example fish larva [2].

Tiny phytoplankton, are the primary producers in the aquatic food web. After those are the zooplankton, which graze on phytoplankton, and become the reason why larger animals, from small fishes to whales do not starve [2]. Plankton also play an important role in the carbon cycle, transporting carbon from the surface to the depths of ocean [5]. Their impact on climate change [6], as well as the consequences of climate change on their life [7, 8, 9], are very important open questions. A better understanding of this relationship can result in methods that can be used to mitigate the current climate change issue, for example efficient fertilisation of the ocean to stimulate plankton CO_2 uptake [10].

Every day when the sun goes down, zooplankton migrate from deep regions of the ocean towards the surface in order to feed on phytoplankton. At



(a) **(b)**

Figure 1.1: Samples of different (a) zooplankton and (b) phytoplankton species. Taken from Wikipedia.

sunrise they return to the darker depths, most likely to protect themselves from predators [11]. This is called the daily migration of zooplankton (Figure 1.2) and is referred to as the largest biomass migration on Earth [11]. This migration plays an important role in the aquatic food chain, energy transport, carbon cycle, and last but not least the survival of zooplankton.

Diel vertical migration has been known for over two centuries (see Ref. [12] for a review of its historical discoveries up to now). Yet, we still lack a complete understanding of how zooplankton swim and exploit the flow in order to efficiently migrate. In fact, being efficient in swimming should be very crucial for zooplankton, as their small size puts them in the viscous flow regime. Adding to that, the lack of power to swim against the flow currents for the same reason, makes the task of swimming towards a target harder, because the swimmer will be more dependent on the flow currents. Despite this, observations show that zooplankton can migrate up to hundreds of meters within a few hours [11], which for organisms with millimeter-sized bodies, means a migration even up to a million times their body length per day.

To understand their dynamics better, new models on the individual level that take into account the interactions of the plankter with its ambient flow, and with other individuals, are needed. Their small size and the long range of their migration, makes it hard to obtain experimental data. However, with the recent increasing power of computational facilities, it is possible to use computer simulations to study individual plankter in the turbulent ocean

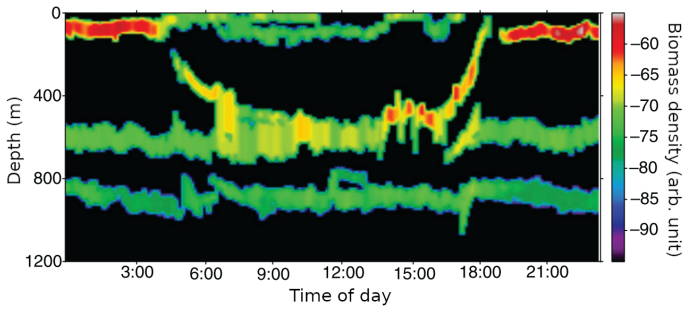


Figure 1.2: Time evolution of the biomass density in ocean. The density is smaller close to surface during the day compared to nights. Figure is taken from [11].

environment [13]. Using simulations to study and model the individual level behaviors of planktonic swimmers can also become helpful in better understanding of their crowd behavior. Previously, the planktonic swimmers were simulated using simple run and tumble processes and random walks with different characteristics, such as variable diffusivity [14], or as swimmers with constant speed, swimming in their instantaneous direction [15]. Although, the word plankton means drifter, many species can do more than simply drift with the flow. Observations show that they are able to contribute to their dynamics by actively rotating, jumping, regulating their buoyancy, changing their shape, etc. [16, 17]. It is possible to add such capabilities to models by setting a static policy based on educated guesses, for example the model of jumping copepods in Ref. [18] or the model of buoyancy regulating settling phytoplankton in Ref. [19]. However, a downside of this approach could be that it is heuristic and may therefore miss important mechanisms that exist in more general models and in nature. To account for this, machine learning [20], can be used to build more realistic models of plankton. Different machine learning approaches provide fascinating frameworks to improve our understanding of active microswimmers, such as plankton, in different aspects [21]. One of the suitable paradigms in machine learning, to deal with this problem, is reinforcement learning [22], which is an approach in between of supervised and unsupervised learning algorithms. Reinforcement learning provides the opportunity to train smart swimmers that learn to achieve a certain goal optimally, by interacting with their environment

and learning by trial and error.

The potential of applying reinforcement learning to the navigation problem of microswimmers in complex flows, is shown in the proof of principle studies in Refs. [23, 24]. The results in Refs. [23, 24] indicate that actively responding to flow signals allows for much more efficient strategies of vertical migration in complex flows.

In this thesis I study the optimal strategies that a microswimmer, such as plankton, given a degree of control, can exploit in order to efficiently perform vertical migration. For this purpose, we modeled an active microswimmer from a Lagrangian point of view and gave it the ability to interact with its background flow. Being able to measure some properties of its local flow and reorienting itself in response to the measured quantities, we developed optimal active reorientation strategies that enable the swimmer to increase its vertical migration drastically, when compared to a swimmer that only benefits from passive strategies of vertical migration such as gyrotaxis [3]. A gyrotactic swimmer is a swimmer that experiences a torque induced by gravitational force that aligns it parallel to gravity direction.

The thesis is structured in three parts. Part I, which consists of background information, starts with a discussion on how a planktonic microswimmer can be modeled from a physicist's point of view. Chapter 2, introduces the shape, dynamics, and sensing abilities of such swimmers. Then, in Chapter 3 a brief introduction to the flow models used as background flow in the simulations is presented. In Chapter 4, the reinforcement learning is discussed, which is the machine learning technique used to find optimal strategies of vertical migration for the modeled swimmer. Part II provides an introduction to the main results of the thesis projects, that resulted in two papers [Papers A, B]. Chapter 5 corresponds to the study of microswimmers in two-dimensional flows [Paper A], and Chapter 6 introduces the main results from our model in three spatial dimensions [Paper B]. The conclusion and outlook is given in Chapter 7, and finally Part III contains the published papers.

PART I
BACKGROUND

In this Part, the background information needed for better understanding the work done in the thesis is presented. Chapter 2, describes how and why one can model the dynamics of an individual plankter, such as a microswimmer in the low Reynolds number flow regime. The introduced model is later used in the course of the work done in the thesis.

To describe the living environment of the plankton, two flow models, Taylor-Green Vortex (TGV) flow and Gaussian random velocity (GRV) fields are used. A brief introduction to these flow models is given in Chapter 3.

Finally, in Chapter 4, the basics of reinforcement learning and in particular Q-learning are discussed. This is the framework of machine learning used here to train smart swimmers capable of navigating efficiently in complex environments, such as turbulent flow.

2 Planktonic microswimmer

Now that we know why the study of plankton is important (Chapter 1), we can try to model them from a physicist's perspective. First of all, it should be noted that for the purpose of generality, we do not consider a specific plankton species here. The motivation for this is that mechanisms found from studying a simple generic model can be relevant to many different species of plankton with different body shapes. However, the model we use in the thesis is highly inspired by copepods. Copepods are the most ubiquitous organisms among zooplankton, with more than 14,000 species. Their size ranges from 0.2 to 10 mm, but usually they are smaller than 6 mm, and can be found as free-living species, symbiotic, and even parasitic relationships, in fresh water, oceans, lakes, rivers, and even in a puddle [2, 25].

Many zooplankton species actively navigate to mate, to feed, to escape perceived threats, and in general to survive in the turbulent fluid environment they live in. But how do they navigate? The interesting character of copepods in view of this thesis is their ability to interpret hydromechanical signals and act upon these signals to navigate with their small control freedom [26, 27, 28, 29]. Some species can make huge jumps, 150 times larger than their body length, in order to escape from a predator when they feel a strong fluid strain [30, 31]. Some others, such as Limnocalanus macrurus, can continuously cruise with a small roughly constant swimming speed [16]. There are models based on diffusion that describes the movement of plankton [14, 32, 33]. Although this approach is helpful in understanding the observation, but cannot tell us anything about the mechanism of their swimming, or how they react to different environmental cues to navigate more efficiently on an individual scale. For that purpose, we need to study the individual plankter and the interaction it has with its environment.

The first step is to model the plankter as an active swimmer with small control over its dynamics. In what follows, a brief introduction to shapes of these creatures and how our model swimmer looks like is given in Section 2.1. The swimmer's dynamics is discussed in Section 2.2, and finally in Section 2.3, the way a plankter perceives its environment is briefly discussed.

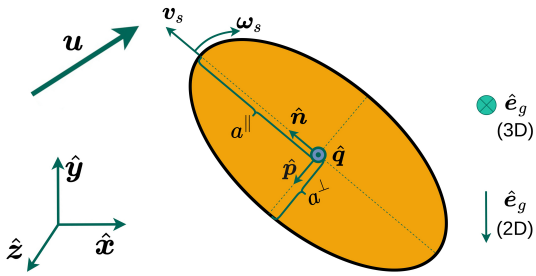


Figure 2.1: Cross section schematics of the swimmer's body modeled in the thesis.

2.1 Shape and mass distribution

Copepods exist with innumerous different body shapes and structures, but nearly all copepods have elongated bodies. A number of examples of this diversity are presented in Figure 2.2. Their antennae are important for sensing the surroundings and holds a high concentration of sensory receivers. The number of legs vary between four to five pairs in different species [34].

To approximate the shape of the copepods (Figure 2.2), we considered a swimmer with an axisymmetric spheroidal body. Since it is easier to illustrate the body properties in two dimensions, the cross section of the model swimmer is shown in Figure 2.1. The three-dimensional swimmer body, can be created by rotation of the ellipse in Figure 2.1, around the swimming direction $\hat{\boldsymbol{n}}$.

The parameter that differentiates a spheroid from a sphere is called the shape factor Λ, and is defined to be

$$\Lambda = \frac{\lambda^2 - 1}{\lambda^2 + 1}. \tag{2.1}$$

Here, λ is the aspect ratio of the body and is equivalent to the ratio between the semi-axis parallel with and perpendicular to the swimming direction, *i.e.* $\lambda = a^{\parallel}/a^{\perp}$ in Figure 2.1.

In our model, we used estimates on the parameters describing the copepod body based on the data for juvenile copepods. Juvenile copepods are small enough to enable us neglecting their influence on the flow. Otherwise, one needs to take into account the disturbances of the swimmer on

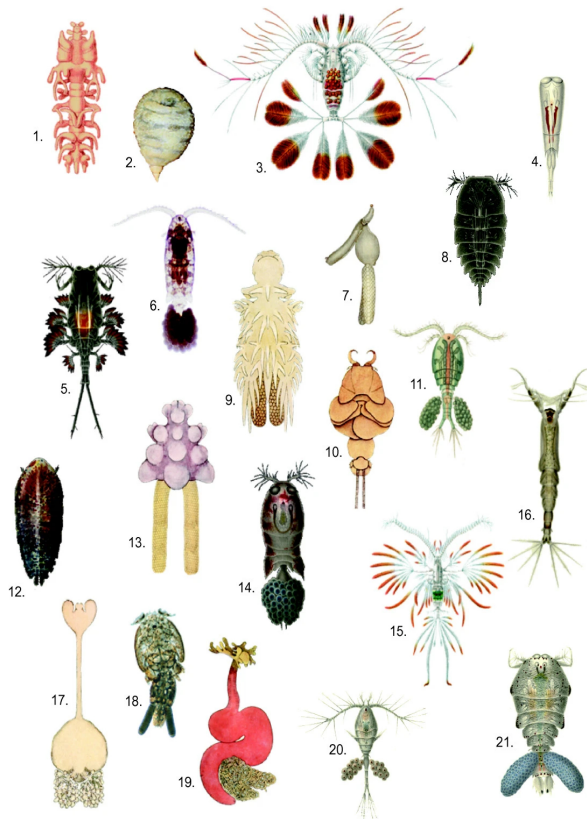


Figure 2.2: Illustration showing diversity of copepod forms. 1. Philichthys xiphiae 2. Sarcotaces sp. 3. Calocalanus pavo 4. Farranula rostrata 5. Copilia vitrea 6. Paracalanus parvus 7. Clavella adunca 8. Copilia quadrata 9. Chondracanthus zei 10. Phyllothyreus cornutus 11. Acanthocyclops vernalis 12. Sapphirina ovatolanceolata 13. Chondracanthus ornatus 14. Corycaeus obtusus 15. Euaugaptilus filigerus 16. Monstrilla longispinosa 17. Sphyrion lumpi 18. Caligus elongatus 19. Lernaeocera branchialis 20. Oithona nana 21. Sapphirina auronitens. Taken with permission from [35].

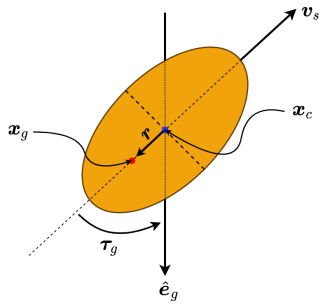


Figure 2.3: A bottom-heavy swimmer which experiences a gyrotaxis torque τ_g. This can simply happen if the center of mass x_g, does not coincide with geometrical center of the body x_c. Modified from [26].

the flow in the model. Juvenile copepods have body lengths in the range 0.1 to 0.5 mm [36, 37]. In Ref. [38] the ratio of body width to length during development of Centropages typicus is reported to stay fairly constant around 0.4 to 0.5, which is equivalent to having an aspect ratio λ between 2 to 2.5. The mass density ρ_p of Centropages finmarchius ranges from 1.0274 to 1.0452 g cm^{-3} [39]. Using the mass density $\rho_f = 1.025$ g cm^{-3} of sea water with 3.5% salinity at 20 °C [40], we chose the ratio ρ_p/ρ_f being equal to 1.017.

Another assumption here is that the body is rigid. Although this is not completely true, but it makes a tremendous simplification in modeling body-flow interactions and hydromechanical signal perceiving of the swimmer. Also, as stated in Ref. [41], this is not a very bad assumption since the copepods are relatively rigid compared to their surrounding flow.

Many plankton in nature are bottom-heavy, having structures in their body that aligns them in response to gravity. This can happen simply by having a non-uniform mass distribution. Figure 2.3 shows such a condition where the center of mass has an offset compared to the geometric center [26]. This gives rise to a net gravitational torque $\tau_g \propto \hat{e}_g \times r$, where r is the vector from the geometrical center x_c to the center of mass x_g, and gravity $\hat{e}_g$ points in the direction defined in Figure 2.1.

2.2 Dynamics

Copepods can swim with a swimming velocity of 0.33 to 3.76 mm s^{-1} [37] in the seawater with kinematic viscosity $\nu_s \simeq 1.0$ mm s^{-1} [42]. A juvenile copepod that has a body length of $a^{\parallel} \simeq 0.2$ mm [36, 37], and a swimming velocity $v_s \simeq 1.0$ mm s^{-1}, will have a particle Reynolds number $\mathrm{Re_p} = 2a^{\parallel} v_s/\nu$, smaller than unity. Therefore, the swimmer experiences its surroundings as a viscous environment. This puts the copepod in a world completely different from what we are experiencing in our daily life [43]. In addition, the Stokes number St, which is defined to be the ratio of the swimmer's relaxation time-scale τ_p, to the characteristic time of the flow τ_f, is small for the parameters of a juvenile copepod. Therefore, it is reasonable to assume that the swimmer experiences an overdamped dynamics. Due to the bottom-heaviness we considered here (Figure 2.3), the swimmer in our model experiences a gyrotactic torque from gravity as well. The other contribution from gravity is the settling velocity which is equal to [44]

$$\boldsymbol{v}_g = v_g^{\perp}\hat{\boldsymbol{e}}_g + (v_g^{\parallel} - v_g^{\perp})(\hat{\boldsymbol{e}}_g \cdot \hat{\boldsymbol{n}})\hat{\boldsymbol{n}}, \tag{2.2}$$

where $v_g^{\perp}$ and $v_g^{\parallel}$ are the settling speeds in quiescent flow for a spheroid settling with its symmetry axis perpendicular and parallel with the direction of gravity respectively [44],[Paper A]. Finally, advection by the flow should give a contribution to the swimmer's velocity.

Apart from the external forces discussed, the swimmer itself can also affect its dynamics. To account for that, we assume that the swimmer can cruise with a constant translational speed. In addition, we assume that it can exert an angular velocity to control its swimming direction to some extent. The swimming speed is estimated from experimental data being in a range 0.33-3.76 mm.s^{-1} [37].

Putting all of these assumptions together, we write the translational and rotational equations of motion for active swimmer in three spatial dimensions, with swimming direction $\hat{\boldsymbol{n}}$, in the following form [45, 46], [Paper B]

$$\dot{\boldsymbol{x}} = \boldsymbol{v}, \quad \dot{\hat{\boldsymbol{n}}} = \boldsymbol{\omega} \times \hat{\boldsymbol{n}}, \quad \dot{\hat{\boldsymbol{p}}} = \boldsymbol{\omega} \times \hat{\boldsymbol{p}}, \quad \hat{\boldsymbol{q}} = \hat{\boldsymbol{n}} \times \hat{\boldsymbol{p}}, \tag{2.3a}$$

where $\boldsymbol{v}$ and $\boldsymbol{\omega}$ are the translational and rotational velocities of the swimmer

and can be written as

$$\boldsymbol{v} = \boldsymbol{u}(\boldsymbol{x}, t) + v_s \hat{\boldsymbol{n}} + \boldsymbol{v}_g, \tag{2.3b}$$

$$\boldsymbol{\omega} = \mathbb{O}(\boldsymbol{x}, t) + \Lambda \hat{\boldsymbol{n}} \times \mathbb{S}(\boldsymbol{x}, t)\hat{\boldsymbol{n}} - \frac{1}{2B}\hat{\boldsymbol{n}} \times \hat{\boldsymbol{e}}_g + \boldsymbol{\omega}_s. \tag{2.3c}$$

Here, $\boldsymbol{u}(\boldsymbol{x}, t)$ is the fluid velocity at the swimmer's position at time t, v_s is the swimmer's propulsion speed that is considered to be a constant, and $\boldsymbol{v}_g$ is the settling velocity given by Eq. (2.2).
The first and the second terms in the RHS of Eq. (2.3c), describe the response of the swimmer to the flow vorticity and deformation respectively. These were first calculated by Jeffery in his work on hydrodynamic torque on a general spheroid in shear flow [47]. The first term does not depend on the shape and simply describes the torque exerted on the swimmer by fluid vorticity. The second term is the response of the swimmer to linear deformation of the flow and will be non-zero only for swimmers with non-spherical shapes. The shape parameter Λ accounts for the shape dependency in Eq. (2.3c), which is given by Eq. (2.1) for an axisymmetric swimmer. $\mathbb{S}$ and $\mathbb{O}$ are the symmetric and asymmetric parts of the flow gradient matrix $\mathbb{A}$, which is given for a three-dimensional flow $\boldsymbol{u} = (u_x\hat{\boldsymbol{x}} + u_y\hat{\boldsymbol{y}} + u_z\hat{\boldsymbol{z}})$, by

$$\mathbb{A} = \begin{bmatrix} \partial_x u_x & \partial_y u_x & \partial_z u_x \\ \partial_x u_y & \partial_y u_y & \partial_z u_y \\ \partial_x u_z & \partial_y u_z & \partial_z u_z \end{bmatrix}. \tag{2.4}$$

The symmetric part $\mathbb{S}$, is called strain rate tensor and describes deformation of the flow and asymmetric part $\mathbb{O}$, is called rotation and as the name suggests describes rotation of the flow. $\mathbb{S}$ and $\mathbb{O}$ are given by

$$\mathbb{S} = \frac{(\mathbb{A} + \mathbb{A}^T)}{2}, \tag{2.5a}$$

$$\mathbb{O} = \frac{(\mathbb{A} - \mathbb{A}^T)}{2}. \tag{2.5b}$$

Rotation $\mathbb{O}$, is related to the flow vorticity through the relation $\mathbb{O}\hat{\boldsymbol{n}} = \frac{1}{2}\boldsymbol{\omega}_f \times \hat{\boldsymbol{n}}$, where the flow vorticity is given by curl of the flow $\boldsymbol{\omega}_f = \nabla \times \boldsymbol{u}$.

The third term is the gyrotactic torque, resulting from the assumed bottom-heaviness (Section 2.1). The gyrotactic torque tends to align the swimming

direction $\hat{\boldsymbol{n}}$ against the direction of gravity $\hat{\boldsymbol{e}}_g$. B is the reorientation time-scale and determines the strength of the gyrotactic torque [45, 46, 48].

Finally, the last term in Eq. (2.3c), is a control parameter that we assume the swimmer can change. It is through this term that the swimmer can actively contribute to its rotational dynamics.

In our model we also ignore the jumping behavior of many species of copepods [36]. A Lagrangian model for jumping copepods is studied in Ref. [18], where the swimmer can perform jumps with very large speeds (up to 250 times of the flow characteristic velocity) or advected by the flow. Our swimmer is only capable of cruising with constant speed and the control, as mentioned above is only through the contribution $\boldsymbol{\omega}_s$ (Eq. (2.3c)).

The dynamics written in Eq. (2.3), can be reduced to a two-dimensional model. For a swimmer living in the, say $\boldsymbol{x}-\boldsymbol{y}$ plane, one can write $\hat{\boldsymbol{q}}=\hat{\boldsymbol{z}}$ and consequently $\dot{\hat{\boldsymbol{q}}}=0$. Without any loss of generality we assume the direction of gravity $\hat{\boldsymbol{e}}_g$ to be $-\hat{\boldsymbol{z}}$ in three spatial dimensions and $-\hat{\boldsymbol{y}}$ in two-dimensional model. Finally, it must be noted that in the two-dimensional dynamics, gyrotactic swimmers can get trapped in the limit cycles and stable fixed points of the dynamics, or show structurally unstable periodic orbits [15, 49]. However, by adding noise to the dynamics this trapping effect can be removed [49]. Therefore, to study the swimmer in two spatial dimensions, we added independent Gaussian noise with zero-mean to the dynamics in Eq. (2.3) in order to prevent such trapping behaviors [Paper A].

2.3 Sensing abilities

Many plankton species can perceive their environment and act upon different cues. They can feel if there is a predator in their surroundings and make escape jumps with speeds up to 150 body lengths per second as the result [31]. They can feel if there is a potential mate around them, and find the same species in an environment filled with different zooplankton species, by exploiting chemical cues [50, 51]. They can find food and capture a prey using feeding currents when they sense that one is close to them [52]. Many planktonic species have a single or a pair of eyes [53], that enables them to respond to light variations, for example in escape jumps of Acardia tonsa discussed above [54]. In conclusion, despite being tiny, plankton have sensory systems which are responsive to a wide range of information sources,

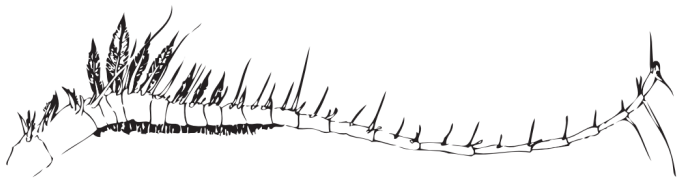


Figure 2.4: The antenna of the copeopod Labidocera madurae. Taken from [26]

allowing them to exploit hydromechanical, chemical, and electromagnetic signals, for better survival.

Sensitivity to the light and chemical cues are more familiar to us humans and therefore easier to understand, but how can plankton perceive the hydromechanical signals? Copepods have arrays of setae all over their body, especially on their most sensitive organ, the antennae (Figure 2.4). A detailed description of hydromechanical signal sensitivity in plankton is given in Ref. [55]. The hydromechanical sensory system is also discussed in Refs. [29, 30]. Here, I briefly introduce how different hydromechanical signals can be inferred using arrays of setae.

Before beginning with that, let me point out why the main source of information in plankton is believed to be perceived by their hydromechanical receptors. The visible spectrum of electromagnetic waves is not a particularly efficient messenger in water due to high extinction coefficient, opaqueness, and strong diffraction [56]. Therefore, unless the species has evolved a complex eye structure [57], which is not the case for copepods having simple photoreceptors [56], visual stimulus is limited. Also, the copepods response to light is slower than their response to hydromechanical signals. For example, the response time of Acartia tonsa to photonic stimulus ranges between 30 to more than 150 ms, while their response time to hydromechanical escape signals is roughly 4 ms [54].

Chemical signals, on the other hand, spread in the environment by diffusion and diffusion is inherently much slower compared to fluid disturbances in the flow regime that the tiny plankton live [28]. Diffusivity of biological molecules like oxygen are of the order 10^{-9} $\mathrm{m^2\,s^{-1}}$ [50]. Since, kinematic viscosity is in fact the diffusivity of fluid momentum and has the same dimension as diffusivity (L^2T^{-1}), we can compare the time scale of chemical

diffusion and hydromechanical disturbances. For the diffusion we have $T_D \sim L^2 D^{-1}$ and for the viscosity $T_\nu \sim L^2 \nu^{-1}$. Knowing that the ocean has a kinematic viscosity of the order $\nu \sim 10^{-6}\mathrm{m}^2\,\mathrm{s}^{-1}$ [42], one sees that the diffusion is 10^3 times slower compared to hydromechanical signals. This will be a matter of life and death, when the poor plankter needs to locate a predator [28]. However, the discussion above should not convey the message that the diffusion is useless. For the same reason of being slow, diffusive signals dissipate slower as well. This means that the signal will remain in the environment much longer in comparison to a hydromechanical disturbance, which dissipates very fast due to high damping rate in the fluid regime that plankton are living in. This could become useful for some tasks like tracking a mate exploiting pheromone diffusion.

Coming back to the array of setae, these organs power neurophysiological pulses in response to bending [26]. Having this in mind, a copepod cannot sense any hydromechanical signal if it is advected by a flow that is constant over its body, since this situation will not exert any force that bends the setae. Evidently, there must be a velocity difference between the swimmer's rigid body and the ambient flow for the setae to be deformed. Therefore, the array of setae are sensitive to velocity differences. The lowest velocity difference that is reported for plankton to sense is as small as $20\ \mu\mathrm{m\,s}^{-1}$ [58].

Velocity differences between the body of a swimmer and the ambient flow, originating from acceleration, rotation, and deformation of the flow have the potential to create different kinds of bending patterns on the setae arrays. Figure 2.5 shows a schematic of different possible patterns for different causes. This suggests the possibility of differentiating between distinct hydromechanical signals by the swimmer. In fact different setae on the plankton's body have different properties, and combination of various bending patterns of the setae arrays enables plankton to have a detailed three-dimensional perception of their environment [28, 59]. As an analogy, imagine you could feel someone is getting close to you by bending patterns of your body hair, and even the patterns were different if the person was your friend or was a thief. This is the case for a zooplankter, being able to distinguish between a prey and a predator based on the different hydromechanical signals they produce [60].

In order to model a swimmer in Refs. [Papers A, B], it is assumed that the swimmer is capable of making a distinction between various sources of velocity differences as depicted in Figure 2.5. The question that arises here

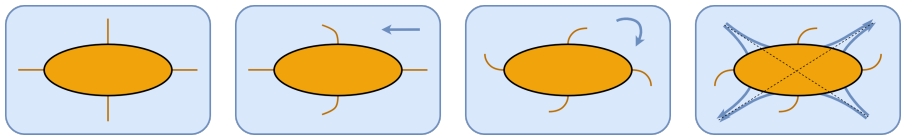


Figure 2.5: Different setal patterns in response to different flow gradients. (a) No velocity gradients, (b) acceleration, (c) rotation, and (d) deformation (compressing and stretching perpendicular to each other), of the flow. Modified from Ref. [26].

is, how strong should a disturbance be to become recognizable by the swimmer? There should be a threshold for a disturbance strength that induces a response from the swimmer. Therefore, we assign a threshold to each of the signals depicted in Figure 2.5, and throughout thesis the thresholds are denoted by adding * as a superscript to the signal. We know from the experiments in Refs. [30, 61], that the copepods make escape jumps in response to strain rates above $\sim 0.5\ \mathrm{s}^{-1}$. Therefore, we employ the same value as the threshold of being sensitive to strain rate $S^\star$, for our model swimmer as well.

If we denote the points on the swimmer's body with a vector from its geometrical center, such that the vector $\boldsymbol{r}$ signifies a certain point on the swimmer's body, recalling rigidity and no disturbances from the swimmer on the surrounding flow, the velocity difference between the swimmer's body and the flow at the point $\boldsymbol{r}$, arising from a strain rate $\mathbb{S}$, would be equivalent to $\mathbb{S}|\boldsymbol{r}|$ [30]. Accordingly, we choose the threshold for the velocity difference between swimmer's body and the surrounding flow to be $\Delta u^\star = a^\parallel S^\star$, evaluating the largest velocity difference between the flow and the swimmer, with $a^\parallel$ being the semi major-axis in our swimmer's spheroid body (Figure 2.1). Other points on the body will have smaller velocity differences and can be disregarded compared to the chosen value [30]. Similarly, for angular velocity differences we have $\Delta\Omega^\star = \Delta u^\star / a^\parallel = S^\star$.

With the chosen value of $a^\parallel$ in our model, the corresponding threshold for velocity difference will be $\Delta u^\star = 50\ \mu\mathrm{m\,s}^{-1}$. This is slightly higher than the smallest sensitivity of copepods being $20\ \mu\mathrm{m\,s}^{-1}$ [58].

3 Flow

Since the dynamics of the copepods happen in aquatic environments, modeling the underlying flow is an inseparable part of the study. In this Chapter I present an introduction without going into details to the two flow models that are used in the simulations of the swimmer.

Generally, to describe the the dynamics of the flow we adopt the Navier-Stokes equations [62]

$$\rho_f(\partial_t \boldsymbol{u} + \boldsymbol{u}\cdot\boldsymbol{\nabla}\boldsymbol{u}) = -\boldsymbol{\nabla}p + \mu\nabla^2\boldsymbol{u} + \boldsymbol{f}. \tag{3.1}$$

The Navier-Stokes equations are in fact the Newton's second law, written for an incompressible fluid ($\boldsymbol{\nabla}\cdot\boldsymbol{u} = 0$), which determines the evolution of the fluid velocity $\boldsymbol{u}(\boldsymbol{x}, t)$. The LHS in Eq. (3.1) describes the force due to inertial acceleration and the RHS accounts for the forces on a fluid parcel rising from pressure gradients $\boldsymbol{\nabla}p$, viscous forces $\mu\nabla^2\boldsymbol{u}$, and the external forces $\boldsymbol{f}$. The flow density and the dynamic viscosity are denoted respectively by ρ_f, and $\mu = \rho_f\,\nu$, which are assumed to remain constant.

General analytical solutions of Navier-Stokes equation are extremely hard to find, and numerical solutions are highly costly to compute. This is a result of the non-linear term $\boldsymbol{u}\cdot\boldsymbol{\nabla}\boldsymbol{u}$, which becomes more important for high Reynolds numbers $\mathrm{Re} = u_0\eta/\nu$, often referred to as inertial regime, where u_0 and η are respectively the characteristic speed- and length- scales of the flow. This term is responsible for turbulence and chaotic behavior of the flow.

In this thesis, instead of solving Navier-Stokes equation (Eq.(3.1)), we used two simpler models for turbulence, Taylor-Green vortex (TGV) flow and Gaussian random velocity (GRV) fields in order to model the underlying flow. Section 3.1 describes the TGV flow, and the GRV field basics and the settings used in our simulations are briefly discussed in Section 3.2. In Ref. [Paper B], our collaborators also used direct numerical solution (DNS) to model the background flow. These simulations are not covered in the thesis and reader is referred to Ref. [Paper B] for details.

3.1 Taylor-Green vortex flow

One of the simplest fluid models which is used to model turbulence is the Taylor-Green vortex (TGV) flow. The structure of the TGV flow in two spatial

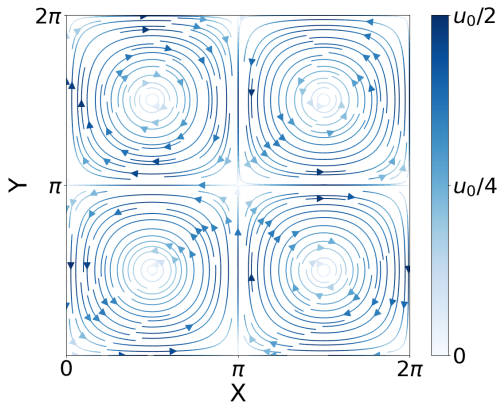


Figure 3.1: The structure of a steady TGV flow (Eq. (3.4)) for a box of side 2π. The parameters η, and u_0 are chosen to be unity here.

dimensions is shown in Figure 3.1 . The TGV flow was introduced to model one of the main characteristics of turbulence which is the breakup of larger eddies into smaller ones until they get small enough that the rate of their dissipation due to viscosity becomes larger compared to their production rate by the breakup process, in a viscous incompressible fluid. It has a definite length-scale and some of the properties of statistically uniform istortropic turbulence [63], for example it is invariant under $n\pi$ rotations.

The TGV flow is an analytical solution of Navier-Stokes equation (Eq. (3.1)) in the limit of zero external forces ($\boldsymbol{f} = 0$). Therefore, similar to other exact solutions [64], it is widely used as a tool to benchmarking numerical methods. A two-dimensional incompressible TGV flow $\boldsymbol{u}(x, y, t)$, consists of an array of periodic two-dimensional cells of eddies (Figure 3.1), and can be defined with the following equation

$$\boldsymbol{u}(x, y, t) = \boldsymbol{\nabla} \times \psi(x, y, t)\hat{\boldsymbol{z}}, \tag{3.2}$$

where $\psi(x, y, t)$ is the stream function and is equivalent to [63]

$$\psi(x, y, t) = \frac{u_0 \eta}{2} \cos(\frac{x}{\eta}) \cos(\frac{y}{\eta}) \exp[-\frac{2\nu t}{\eta^2}]. \tag{3.3}$$

Here, u_0 is the maximal flow speed, obtained on the cell boundaries, η is the length scale of the flow, and ν is the kinematic viscosity of the flow. In the

case of a steady TGV flow, the flow velocity components reduce to

$$u_x = \partial_y \psi(x,y) = \frac{u_0}{2}\cos(\frac{x}{\eta})\sin(\frac{y}{\eta}), \tag{3.4a}$$

$$u_y = -\partial_x \psi(x,y) = -\frac{u_0}{2}\sin(\frac{x}{\eta})\cos(\frac{y}{\eta}). \tag{3.4b}$$

Figure 3.1 represents the TGV flow structure for a box of side 2π, with $\eta = u_0 = 1$. The two-dimensional steady TGV flow is used in Ref. [Paper A] as a model of background flow in the simulations of the swimmer, with the structure depicted in Figure 3.2b.

3.2 Gaussian random velocity field

Another numerically efficient model for the small scales of turbulence suitable in the limit of plankton scales (small particles compared to the smallest turbulent length scale) is the Gaussian random velocity (GRV) field. Our implementation of GRV fields are statistically isotropic and homogeneous in space. The following formalism is taken from Refs. [65, 66].

GRV fields describing incompressible fluid velocity fields $\boldsymbol{u}(\boldsymbol{x},t)$, can be generated using Eq. (3.5) in two and three spatial dimensions.

$$\boldsymbol{u}(\boldsymbol{x},t) = \mathcal{N}_2 \nabla \times \psi(\boldsymbol{x},t)\hat{\boldsymbol{z}}, \tag{3.5a}$$

$$\boldsymbol{u}(\boldsymbol{x},t) = \mathcal{N}_3 \nabla \times \boldsymbol{A}(\boldsymbol{x},t). \tag{3.5b}$$

Here, $\mathcal{N}_2$ and $\mathcal{N}_3$ are normalization coefficients and are given by $\mathcal{N}_d = [d(d-1)]^{-1/2}$. The function $\psi(\boldsymbol{x},t)$ and components of the vector potential $\boldsymbol{A}(\boldsymbol{x},t)$ are independent Gaussian random functions $\phi(\boldsymbol{x},t)$, each with correlation function

$$\langle \phi(\boldsymbol{x},t)\phi(\boldsymbol{x}',t')\rangle \equiv u_0^2\eta^2 \exp\Big[-\frac{|\boldsymbol{x}'-\boldsymbol{x}|^2}{2\eta^2} - \frac{|t'-t|}{\tau}\Big]. \tag{3.6}$$

In Eq. (3.6), η and τ are respectively the flow's characteristic length and time correlation scales, and $u_0 = \langle \boldsymbol{u}^2\rangle^{1/2}$ is the flow speed scale.

To construct the scalar functions $\phi(\boldsymbol{x},t)$, in d-dimensions, we use a Fourier series in the form

$$\phi(\boldsymbol{x},t) = (2\pi)^{d/4}\frac{u_0\eta^{d/2+1}}{\ell^{d/2}}\sum_{\boldsymbol{k}} a_{\boldsymbol{k}}(t)\exp[i\boldsymbol{k}\cdot\boldsymbol{x} - k^2\eta^2/4] \tag{3.7}$$

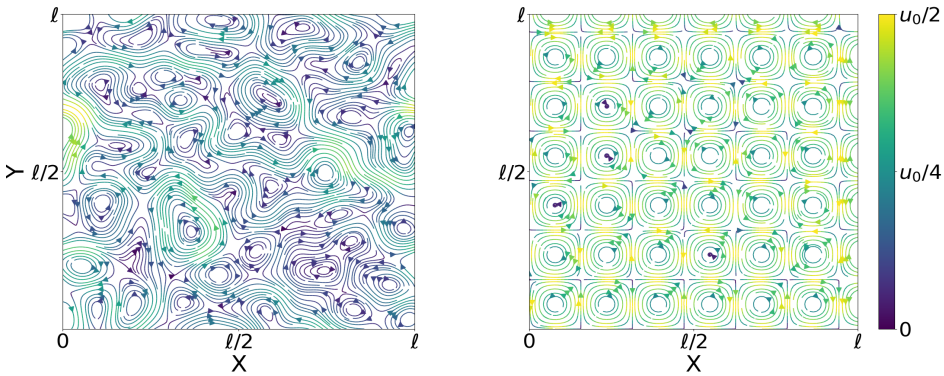


Figure 3.2: Comparison of the flow structures of the GRV field and TGV flow used in Ref. [Paper A]. The stream plot for (a) a GRV field realization, and (b) the TGV flow, with parameters corresponding to turbulent ocean's Kolmogorov scales [42].

The size of the system is denoted by ℓ, and the coefficients $a_{\boldsymbol{k}}(t)$ are complex time-dependent random Fourier coefficients with $\langle a_{\boldsymbol{k}}(t)\rangle = 0$ and $\langle a_{\boldsymbol{k}}(t)a_{\boldsymbol{k}}(t)^*\rangle = 1$, where the * means complex conjugation. The size of the system ℓ is assumed to be much larger in comparison with the spatial correlation length η. The evolution of $a_{\boldsymbol{k}}(t)$ is determined by

$$\dot{a}_{\boldsymbol{k}}(t) = -\frac{1}{\tau}a_{\boldsymbol{k}}(t) + b_{\boldsymbol{k}}(t). \tag{3.8}$$

Eq. (3.8) is an Ornstein-Uhlenbeck process [67] with $b_{\boldsymbol{k}}(t)$ being complex Gaussian random increments which satisfy $b_{-\boldsymbol{k}} = b^*_{\boldsymbol{k}}$, and have zero mean and covariance $\langle b_{\boldsymbol{k}}(t)b^*_{\boldsymbol{k}'}(t')\rangle = 2(\delta t/\tau)\delta_{\boldsymbol{k},\boldsymbol{k}'}\delta_{t,t'}$.

In many cases the GRV fields produce the same qualitative and quantitative results as obtained for the dynamics of particles in simulations of turbulence [68, 69] as well as our results in Ref. [Paper B]. We remark however, that the GRV fields are intended to model small scales in the dissipative range of homogeneous isotropic turbulence [65]. This means they cannot be applied to the inertial regime of the flow. In addition, the GRV field models a small fluid parcel that is swept by larger ones, and in this respect sweeping is taken into account. However, if large scales are important for the dynamics

(for example, if the swimmer swims very rapidly so that it decorrelates due to swimming rather than displacement from the flow), then the model fails because it does not model correlations larger than the smooth length scale η and the inertial range becomes important [65].

Another difference is the distribution of the velocity gradients at small scales (especially at high Re). Despite being smooth, the distribution is not Gaussian in turbulence [70], unlike what is assumed in the GRV model. A big advantage of the GRV model is that in the limit of rapid fluid fluctuations ($\tau \to 0$), it can be simplified using for example diffusion approximation which allows to find analytical solutions. These can result in important useful insights on physical mechanism and interpretations.

In Ref. [Paper A], in addition to the TGV background flow, we used a two-dimensional steady GRV filed. The parameters in both flows were chosen to approximately match the turbulent ocean with kinematic viscosity $\nu \sim 10^{-6}\ \mathrm{m^2\,s^{-1}}$, and the energy dissipation rate $\varepsilon = 1.6 \times 10^{-5}\ \mathrm{m^2\,s^{-3}}$ [42]. Therefore, using Kolmogorov scales

$$\eta = \Big(\frac{\nu^3}{\varepsilon}\Big)^{1/4}, \tag{3.9a}$$

$$u_\eta = (\nu\varepsilon)^{1/4}, \tag{3.9b}$$

with these values of ν and ε, the characteristic velocity is chosen to be $u_0 = 2\ \mathrm{mm\,s^{-1}}$, and the length scale is set to be $\eta = 0.5$ mm. Moreover, the system size $\ell = 20\eta$, was chosen large to reduce encounter of the same flow structures due to periodic boundary conditions. Figure 3.2 shows the stream structure in both the TGV and GRV flows with the parameters used in Ref. [Paper A].

The best agreement between the GRV model and the turbulence occurs at finite but large Ku numbers. $\mathrm{Ku} = \tau u_0/\eta$ is the ratio between Eulerian time correlation τ and the Lagrangian time scale of the spatial decorrelation due to advection, η/u_0. Therefore, when $\mathrm{Ku} \gg 1$, it means that the flow decorrelates faster spatially rather than temporally. In this limit, τ is proportional to the Kolmogorov time scale τ_K, and $\tau_K = \eta/(\sqrt{5}u_0)$ in the GRV model [65].

To model the background flow in Ref. [Paper B], similar GRV fields, but with three spatial dimensions were used.

4 Reinforcement learning

Almost every living species make decisions at various points during their lifespan. For example we decide whether to pursue a Ph.D. or not, as well as which pizza to have for dinner. We take such decisions based on our previous experiences and knowledge. If we refer to the process of acquiring the experiences that allow for better decision-making as *learning*, then we *learn* many things during our life time. A simple example could be that we learn that it is not a good idea to drink a hot tea just after it is poured. That is because we (or someone else) have tried it before and know how it would feel to be burnt by a hot beverage. Thus we prevent occurrence of that hurtful experience again by changing the timing of when we drink tea.

Unlike the decision on doing a Ph.D., in many cases the decisions are pre-programmed in the species DNA, and we just follow the instinct commands we receive in response to the environmental cues. For example, as an infant you knew how to express your hunger, and it is instincts that make you close your eyes to protect them, when faced with an extremely bright flash of light. These kinds of defensive behaviors are in fact responses we have learned through evolution [71], and no matter how automatic they are, they are still decisions made by our body.

The life is full of decision making situations and finding the best set of decisions is a vital problem that we deal with practically every moment. In mathematics these set of problems are normally formulated in terms of Markovian decision processes (MDPs) [72]. One of the approaches to tackle MDPs is reinforcement learning (RL). The idea of RL is to (similar to ourselves) learn an optimal policy from interaction with the surroundings and improving the actions based on the signals received from the environment, with respect to maximizing (minimizing) the total expected utility (cost).

An interesting approach to study the living active matter like plankton, is to model them using an RL framework [21], and study the possible optimal strategies that they might have in following certain goals, for example the diel vertical migration described in Chapter 1.

As discussed in Chapter 2, different species of plankton have acquired various strategies to survive and accomplish their goals in their life cycle. Using RL in modeling them can shed light on understanding the optimal mechanisms from a mechanistic point of view and result in a better picture

of their evolutionary processes in the first place. Moreover, the dynamics of an active organism does not only depend on the external forces, but is also a function of the actions that the organism takes at different stages of time. This is one of the hardships of modeling active matters, and RL can act as a tool to easier deal with such active dynamics, since one does not need to tune actions manually or define a heuristic policy like what is done in Refs. [18, 73, 74].

In a proof of concept study in Ref. [23], the potential of using RL in solving this family of problems is presented by modeling a gyrotactic microswimmer in TGV background flow (Section 3.1), with the goal of efficient vertical navigation, using Q-learning algorithm [75]. The swimmer in Ref. [23] can choose between up, down, right, and left directions as its preferred swimming direction, provided access to the global directional information. The optimal policy found (Figure2 in Ref. [23]) shows an interesting result, despite the goal being to maximize upward displacement, the swimmer takes the down action at various points of its dynamics. This might sound contradictory with the goal. The explanation is that maximization of a utility over a long time, does not necessarily mean taking actions that maximizes the short term reward[1].

In this Chapter RL is introduced with a focus on the settings used in the thesis work. The discussion is based on Refs. [20, 22, 75]. Section 4.1 gives a brief introduction to the concept of MDPs, and Section 4.2 discusses the Q-learning algorithm, which is used in the thesis work. Finally, Section 4.3 covers an introduction to applying RL framework to microswimmer problems.

4.1 Markovian decision processes

As mentioned earlier, the family of decision making problems are formulated classically in the framework of Markovian decision processes (MDPs). MDPs consist of a decision maker that is in continuous contact with its environment. The decision maker is called the *agent* and it can be in different *states* $s \in \mathscr{S}$ ($\mathscr{S}$ denotes the set of possible states of the problem). The agent can take

[1] As a side note, selecting behavioral patterns that only maximizes the immediate rewards as opposed to delayed consequences of the behavior is a deficit in self-control behavior in humans, and is counted as one of the causes for depression [76].

different *actions* $a \in \mathscr{A}$, ($\mathscr{A}$ is the set of possible actions accessible to the agent), at each state. Actions are the means for the agent to change its interaction with its environment. Taking actions results in responses from the environment in the form of *reward* signals $r \in \mathscr{R}$ ($\mathscr{R} \subset \mathbb{R}$ is the set of possible rewards that agent can receive). The goal in an MDP is to maximize the total reward, commonly denoted utility. Figure 4.1, shows a schematic of agent-environment interactions in an MDP problem.

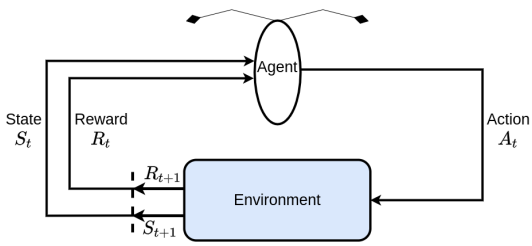


Figure 4.1: Schematic presentation of an MDP problem. The agent continuously interacts with its environment and refines the actions with respect to maximizing the total reward it receives from the environment. Modified from [22]

The dynamics of an MDP can be represented by the function p, which determines the probability of transition to state s' and receive reward r, given that the action a was taken in state s,

$$p(s', r|s, a) \equiv \Pr\{S_t = s', R_t = r|S_{t-1} = s, A_{t-1} = a\}. \tag{4.1}$$

For the case of a Markovian system, the probability of transition between state pairs by taking an action $a \in \mathscr{A}$, is constant, and does not depend on the history of the process. In addition, the following condition holds

$$\sum_{s' \in \mathscr{S}} \sum_{r \in \mathscr{R}} p(s', r|s, a) = 1 \quad \text{for all } s \in \mathscr{S}, a \in \mathscr{A}. \tag{4.2}$$

As mentioned, the agent estimates how good a state is, based on the reward signals it receives from the environment. The choice of the reward is to some extent arbitrary. However, since RL works with maximizing the reward to solve an MDP problem, it is the definition of the reward that sets the goal for the agent. Therefore, it is important to choose a reward that tells

the agent what we want it to achieve. Even though, we should try not to tell the agent how to achieve the goal through the reward. This could result in unwilling consequences, such as to optimize the decisions to reach a subgoal that can cause failure in accomplishing the ultimate goal. For example, if we give a reward to an agent learning to play chess for occupying the center, it might do that at the cost of failing to take out one of the opponents pieces, or even losing the game. In contrast, the reward should ideally reinforce optimizing the ultimate goal, which is winning in a chess game, without any assumptions of how to achieve winning state.

We call the sum of the total reward utility U. It is divided into two categories. One for the tasks that are called episodic and one for continuous tasks. When a task is episodic, it means that there exists a well defined final state and when the agent reaches that state, the episode is finished. In this case the utility at time t is

$$U_t \equiv R_{t+1} + R_{t+2} + R_{t+3} + \cdots + R_T = \sum_{k}^{T-(t+k)} R_{t+k+1}, \tag{4.3}$$

where T is the time of reaching to the final or terminal state, which is itself a stochastic variable and can be different in different realizations of the process. On the other hand, when there is no final state, one should use the discounted reward in order to prevent divergence of utility,

$$\begin{aligned} U_t &\equiv R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots = \sum_{k}^{\infty} \gamma^k R_{t+k+1} \qquad (4.4) \\ &= R_{t+1} + \gamma(R_{t+2} + \gamma R_{t+3} + \cdots) \\ &= R_{t+1} + \gamma U_{t+1}. \qquad (4.5) \end{aligned}$$

Here, γ is a discount factor, $0 \leq \gamma < 1$, and it is introduced to prevent divergence of utility in an infinite number of state changes. Moreover, it makes the rewards far in the future less important compared to the more immediate rewards by giving them smaller weight. The value of γ determines how far in the future the agent should look and optimize the actions with respect to.

Other important elements in the RL framework are the value function $V(s)$ and the policy π. It is more helpful to first define the policy π, and then we can use this to define the value function $V(s)$. The policy π, is a mapping between states and actions. Practically, it is the rule that tells the agent which

action it should take, at each state of the environment. Therefore, one can write the policy as a function $\pi(a|s)$, that gives the probability of taking action a given that state is s.

The value of a state $V(s)$ is the expected utility that the agent can get starting from a state s. The value $V(s)$ is a function of the policy $\pi(a|s)$ itself, because the expected utility depends on which policy is followed. The value of a state $V(s)$ subjected to a policy $\pi(a|s)$, can be written as

$$\begin{aligned} V_\pi(s) &\equiv \mathbb{E}_\pi[U_t|S_t = s] \quad , \quad \text{for all } s \in \mathscr{S}, \qquad (4.6)\\ &= \mathbb{E}_\pi[R_{t+1} + \gamma U_{t+1}|S_t = s] \\ &= \sum_a \pi(a|s) \sum_{s'} \sum_r p(s', r|s, a)\Big[r + \gamma \mathbb{E}_\pi[U_{t+1}|S_{t+1} = s']\Big] \\ &= \sum_a \pi(a|s) \sum_{s'} \sum_r p(s', r|s, a)\Big[r + \gamma V_\pi(s')\Big]. \qquad (4.7) \end{aligned}$$

Here, $\mathbb{E}_\pi$ means the expected value with respect to policy π. The recursive relation in Eq. (4.7) connects the value of a state s, to the value of its next state s'. This relation is known as the Bellman's equation who derived it in 1950's [77]. Bellman's equation can be used to break the MDP problems into smaller sub-problems that are easier to solve, and the solution to the main problem can be found recursively. This is called dynamic programming and makes the basis for many approaches in RL to solve the MDP problems by learning the value function $V_\pi(s)$ [22].

The question is more clear now, we are looking for a policy $\pi(a|s)$ that maximizes the utility U. This policy is called the optimal policy and is usually denoted in literature by $\pi^*(a|s)$. The method that we used to find the optimal policies for vertical navigation of microswimmers is called Q-learning, and the discussion in Section 4.2 explains the details of it.

4.2 Q-learning

Similar to what was said about the state value function $V(s)$ in Section 4.1, one can formulate everything for the state-action pairs instead of the states only. The value for a state-action pair is called the quality of that state-action pair or simply the Q-value in the literature and is denoted by $Q(s, a)$.

Q-learning is a method to find the optimal policy $\pi^*(a|s)$ by estimating the Q-values $Q(s, a)$, using temporal difference learning (TD). TD-learning,

tries to refine the agent's estimate of the Q-values of state-action pairs, by each state-transition that the RL agent experiences. One can start from a random or an arbitrary initialization of the Q-values estimate, and after enough iterations, the estimates can converge to the optimal Q-values [75]. It should be noted, although we use Q-values to explain the TD-learning here, it does not only work with Q-values, but can be applied to state value functions $V(s)$ as well.

There exist two main advantages of TD-learning compared to other RL methods like Monte-Carlo methods or dynamic programming approaches. One is the fact that TD-learning does not need a model of the environment to learn the optimal policy. The second is being online. This means that the agent is not forced to wait until it reaches to the terminal state in order to update its estimate for the state-action pairs quality $Q(s,a)$. This is very useful in training situations where no final state exists (continuous tasks) or the final state is reached very late.

To address the TD-learning formulation, the estimate of each state-action pair quality $Q(s,a)$ is updated after observing the quality of the new state action-pair $Q(s',a')$, and the corresponding reward r, using the value called TD-error δ_{TD} which is defined to be

$$\delta_{TD} = r + \gamma Q(s',a') - Q(s,a). \tag{4.8}$$

The factor $Q(s',a')$ is called the future expected reward based on the current estimate from the next state and action. The new Q-value after a transition $s \rightarrow s'$ by taking action a is $Q(s,a) \leftarrow Q(s,a) + \alpha\delta_{TD}$. Using Eq. (4.8) it can be rewritten as

$$Q(s,a) \leftarrow Q(s,a) + \alpha\Big[r + \gamma Q(s',a') - Q(s,a)\Big], \tag{4.9}$$

where α is the learning rate and determines how much weight is given to the new observation result, in comparison to the agent's previous estimation of the Q-value $Q(s,a)$. The update formula in Eq. (4.9) is called SARSA[2] update rule. This is an on-policy update, since it depends on the current policy of the agent to estimate the future expected reward term $Q(s',a')$.

Q-learning makes life even easier by an off-policy Q-value update. In contrast with SARSA, in Q-learning the expected future reward is calculated

[2]State - Action - Reward - State - Action sequence

simply by maximizing the Q-value of the new state s' with respect to the possible actions $\max_{a'} Q(s',a')$. Thus, we can write the update rule as

$$Q(s,a) \leftarrow Q(s,a) + \alpha\Big[r + \gamma \max_{a'} Q(s',a') - Q(s,a)\Big]. \tag{4.10}$$

The update rule in Eq. (4.10) is called the one-step Q-learning and is proved to converge to the optimal policy $\pi^*(s|a)$ if all the state-action pairs are visited for a large enough number of times, and the system dynamics are Markovian [75].

Following the greedy policy $\pi(a|s)$ which deterministically takes action $a = \arg\max_{a'}(Q(s,a'))$, during training has the danger of getting stuck in sub-optimal policies, which prevents exploring all of the state-action pairs. In order to ensure that such unfortunate events does not happen, one can add noise to the updating process, similar to the noise added to gradient descent learning. One way is to add a so-called exploration rate ϵ, where rather than following the greedy policy $\pi(a|s)$, the agent follows an ϵ-greedy policy $\pi_\epsilon(a|s)$ during training. This means that the agent takes random actions every now and then (by probability ϵ), and follows the greedy policy with probability $1-\epsilon$,

$$a = \begin{cases} \arg\max_{a'}(Q(s,a')) & \text{with Prob.}(1-\epsilon) \\ \text{random} & \text{with Prob.}(\epsilon) \end{cases}. \tag{4.11}$$

This is to explore the actions that are labeled as bad by the Q-values and to make sure that the labels are accurate. Also, to exploit what is learned so far by probability $1-\epsilon$, that results in faster convergence to the optimal policy. In other words, a successful training depends on the trade-off between exploration with probability ϵ and exploitation with probability $1-\epsilon$. The result of the training will be numerical qualities assigned to each state-action pair, where the optimal policy $\pi^*(s|a)$ is the greedy choice of the actions available to each state.

This was a general introduction to the Q-learning algorithm used in the work done in this thesis. In the Section 4.3 the specific settings to use Q-learning for the problem of an active swimmer is discussed.

4.3 Smart swimmer

In order to apply Q-learning to any problem one needs to define the Q-learning elements such as agent, environment, state, action, and reward function (Figure 4.1). For the case of training a single smart microswimmer, choosing the agent seems to be straightforward, as the only actor in the problem is the swimmer (Chapter 2).

The environment consists of the background flow (Chapter 3), with which the swimmer interacts dynamically. To employ the Q-learning algorithm, one needs to designate clearly defined states that can be occupied by, and are reachable to, the agent. In many cases the easiest way would be dividing the environment into a grid of favorable size [78, 79]. These kinds of problems are widely known as grid-world problems, where each state corresponds to a grid cell in the environment, and normally one grid cell is chosen to be the terminal state (Figure 4.3b). Another reasonable approach is to use the relevant environmental signals that the agent is able to measure, in order to define the states. For example, one can use the vorticity of the background flow as a state signal to train a swimmer with the goal of finding flow regions with the highest vorticities [80]. In many cases, it is helpful to combine a few different environmental signals to define the states. As an example, the swimmer in Ref. [23] can measure its local vorticity in addition to being able to distinguish between global up, down, right, and left directions in two spatial dimensions. It is also possible, to use other local information accessible to the swimmer, such as the local rate of strain, the swimmer's slip velocity, etc., which are potentially possible to be measured locally by a planktonic microswimmer in nature (Section 2.3).

The next important element is the reward function, which in fact imposes the goal of the problem and tells the swimmer what should be optimized. Recall that the reward should not tell the agent how it can reach the goal, instead it is safer to only convey what the goal is. This is to prevent optimizing for a wrong goal and to not impart our prejudices on what a good solution to the problem is (Section4.1). For example, many zooplankton species perform diel vertical migration as explained in Chapter 1. A simple reward function that can reinforce the upwards migration, is the vertical coordinate difference just after entering to and exiting from each state. For the case of a two-dimensional swimmer with the task of efficient navigation to increase its displacement in y-coordinate, one can use the reward function similar to

the reward in Ref. [23]

$$r_{n+1} = \frac{y(s_{n+1}) - y(s_n)}{\Delta t}, \tag{4.12}$$

where $y(s_n)$ and $y(s_{n+1})$ are the initial y-coordinate of the swimmer respectively in states s_n and s_{n+1}, and Δt is the time it takes the swimmer to transit between states $s_n \rightarrow s_{n+1}$. The simple reward in Eq. (4.12), on the one hand, reinforces the actions that increase the vertical migration with positive rewards, and on the other hand, punishes the actions that decrease it with a negative reward. However, planktonic swimmers in nature do not only follow one goal, but in most cases, they are expected to optimize a trade-off between several goals simultaneously. This can also be done using a reward function with different components for each goal. For example, planktonic swimmers try to avoid high strain regions of the flow to reduce the probability of being eaten by a predator [28, 30]. Therefore, to train a swimmer that is expected to avoid high strain regions and at the same time swim upwards, one can modify the reward in Eq. (4.12) to be in the form

$$r_{n+1} = \frac{[y(s_{n+1}) - y(s_n)] - [\mathbf{S}(s_{n+1}) - \mathbf{S}(s_n)]}{\Delta t}, \tag{4.13}$$

where $\mathbf{S}(s_n)$ is the initial measure of the flow strain rate at state s_n. When having a reward function with multiple components, it is important to make sure that the different terms responsible for different tasks, are of the same order of magnitude to make the competition between different optimization goals fair and controllable. Thus one might normalize the rewards, for example with their corresponding maximum values, to make them first of all, non-dimensional, and also having the same order.

For the grid-world problems, normally a negative reward is given for moving into any cell, except for transition to the target cell [78]. However, this is not an obligation. For example, in Ref. [79] the reward is the sum of the swimming time with negative sign to encourage shorter swimming times, and the relative improvement in distance with the goal after every state transition, to reinforce decreasing the distance with the target

$$r_{n+1} = -\Delta t + \frac{|\boldsymbol{x}_B - \boldsymbol{X}_n|}{V_s} - \frac{|\boldsymbol{x}_B - \boldsymbol{X}_{n+1}|}{V_s}. \tag{4.14}$$

Here, $\boldsymbol{x}_B$ is the target coordinate, $\boldsymbol{X}_n$ is the swimmer's position at state n, and V_s is the swimming speed of the swimmer.

The final components of an RL training are the actions. Active microswimmers in nature are able to affect their dynamics by performing different actions. For example, some species are capable of choosing a preferred direction of swimming to navigate towards a certain direction, being gravitactic, phototactic, or chemotactic [28, 81, 82]. The swimmer modeled in Refs. [23, 24] exploits actions inspired by these behaviors, allowing the swimmer to have access to global directional information and choose its preferred swimming direction at each state. This gives the swimmer, four possible actions in two spatial dimensions [23], and six actions in three dimensions [24].

Another possible action is to allow the swimmer to change its shape as it is the case for many micro-organisms in nature. For example, some planktonic swimmers can change their surface area to regulate the drag, or control their buoyancy by changing the concentration of certain ions in their body [17]. An example of a simplified model of such swimmers that can swim by shape transformation is studied in Refs. [73, 83]. There, the swimmer is modeled as a number of spheres connected by extensible rods that can swim by controlling the length of the connecting rods. A different option for action is to allow the swimmer to add an active swimming contribution to its angular dynamics ($\boldsymbol{\omega}_s$ in Eq. (2.3c)). A swimmer capable of tuning its angular velocity, can control its dynamics in competition with its surrounding flow [Papers A, B]. Many zooplankton species are able to navigate by steering while cruising [16], which implies that they should have the means to control their swimming velocity and direction to some extent.

As a simple example of applying Q-learning algorithm to a swimmer in a two-dimensional grid-world problem, here I have presented a swimmer in a discrete version of the Matheron-de Marsily (MdM) velocity field [84]. The MdM model was introduced to show that the transport in porous media is not always diffusive. A particle in a two-dimensional MdM model follows the dynamics

$$x_{t+1} = x_t + u(y(t)), \quad y_{t+1} = y_t + \xi(t), \tag{4.15}$$

where $u(y(t))$ is the horizontal velocity applied initially to each vertical layer of the grid (y-coordinate) and stays constant in time (Figure 4.2). In the continuous dynamics, $\xi(t)$ is a Gaussian random noise with zero mean and $\langle \xi(t)\xi(t')\rangle = \delta_{t,t'}$, implying that the motion in vertical coordinate is diffusive. However, here we assume that both $u(y(t))$ and $\xi(t)$ can only take values

$[-1, 1]$ uniformly random. In addition, we consider that at each time step, a passive particle can either be advected horizontally or diffuse vertically only one grid cell, with equal probabilities.

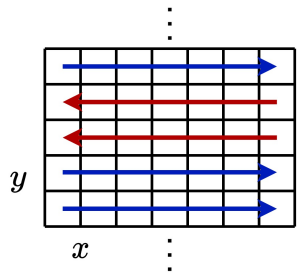


Figure 4.2: Schematic of Matheron-de Marsily (MdM) discrete velocity field. Each vertical layer has a uniformly random chosen advection direction $u(y)$.

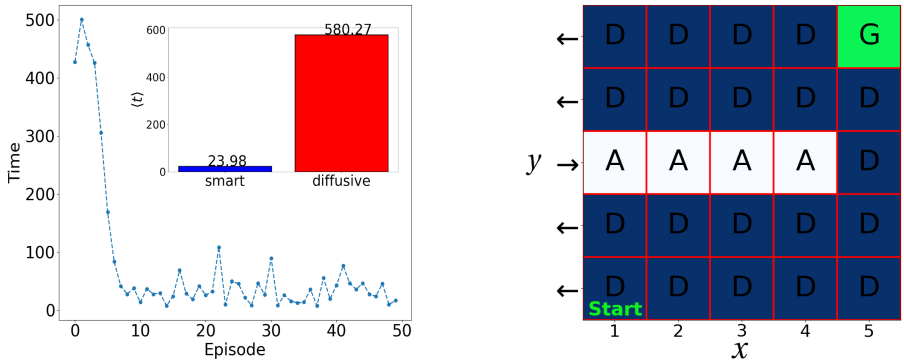
We place a point-like swimmer in a small 5×5 MdM velocity field grid with bounce-back boundary condition. The size of the grid is chosen small for the purpose of easier illustration of the Q-table. Each grid cell signifies a state, and to define the possible actions, the swimmer is allowed to choose at each state whether to diffuse D, or to be advected by the background velocity field A. The goal is to reach to the cell on the upper-right, starting from the lower-left cell in the grid (Figure 4.3)

The swimmer is then trained using one step Q-learning (Section 4.2), with reward defined as

$$r_t = \begin{cases} 0 & \text{if } s_{t+1} = G \\ -t & \text{otherwise} \end{cases}, \tag{4.16}$$

where t is the number of time steps, and G is the goal state. The trained swimmer finds the optimal strategy depicted in Figure 4.3b and gets to the goal state roughly 25 times faster compared to a passive particle that equally likely diffuse or is advected at each cell grid of the same velocity field (inset in Figure 4.3a).

This concludes our discussion on the background part. Hopefully, the reader now has a clear view on the shape properties of the microswimmer and its dynamics in our model (Chapter 2). Also, the background flow used in the thesis work (Chapter 3), and how RL functions and can be applied to microswimmers control problem (Chapter 4). In the next part, the main



(a) **(b)**

Figure 4.3: Training result for a swimmer in Matheron-de Marsily (MdM) velocity field. (a) shows the time it takes each training episode to reach the goal as training proceeds. The fluctuations after the agent learns the optimal path ($\sim$ episode 10) are due to ϵ-greedy update and the intrinsic stochasticity of the problem. The inset compares the average time it takes a passive particle (red) and the smart swimmer (blue) to reach the goal cell from the start position. The averages are taken over 5×10^4 realizations. (b) depicts the environment, as well as the optimal Q-table found by training. The arrows in the labels of the vertical y-axis indicate the direction of $u(y)$ for the used realization of MdM velocity field, which is kept constant during training. The optimal strategy will be different for different realizations of velocity field. Each training episode starts by placing the swimmer in starting position and finishes when the swimmer gets to the goal cell denoted by G. Each grid cell is considered to be a different state, where the swimmer can decide what action to take at any of them. The optimal actions at each state are shown by letters A and D, corresponding to advection by background flow and diffusion in y coordinate respectively.

findings of the thesis work are briefly introduced to facilitate understanding of the publications attached in Part III.

PART II
PRESENT WORK

This Part, is meant to act as an introduction to the work my collaborators and I have done during the past two years on using reinforcement learning to study the strategies for optimal vertical navigation of microswimmers in steady complex flows [Papers A, B].

Many zooplankton species perform a daily vertical migration (DVM), swimming towards regions close to the surface of the turbulent ocean to feed and returning back to darker safer regions by dusk and dawn. It is intriguing how efficient these creatures are in DVM, despite their tiny sizes, that makes swimming in water for them as hard as swimming in honey for us (Section 2.2). Therefore, they have developed effective swimming behaviors, that allows them to navigate efficiently in the ocean.

The work done in Refs. [Papers A, B], suggests that small zooplankton species can potentially exploit the hydromechanical signals they receive, to actively improve their vertical navigation performance, either swimming against or along gravity.

In what follows, Chapter 5 contains the introduction to the main findings published in Ref. [Paper A], where a smart microswimmer in a two-dimensional flow was studied. Then, the discussion in Chapter 6 addresses the results of extending our swimmer model to three spatial dimensions [Paper B]. Interestingly, different mechanisms in two and three dimensions are found for efficient vertical migration of microswimmers. Finally, this Part ends with Chapter 7 presenting the conclusions and outlook of the work.

5 Plankton navigation in the plane

The discussion in this Chapter addresses the main results in Ref. [Paper A]. There, we studied a highly idealized two-dimensional model for microswimmers in complex steady flows, and managed to find efficient strategies for upwards migration against gravity. The swimmer is assumed to follow the dynamics in Eq. (2.3) written for two dimensions (the $\boldsymbol{x}$-$\boldsymbol{y}$ plane with gravity pointing in the negative $\hat{\boldsymbol{y}}$ direction). As discussed in Chapter 2, other assumptions are to have a spheroidal shape (Figure 2.1) and to experience a gyrotactic torque due to bottom-heaviness (Figure 2.3). Section 2.1 discusses the shape parameters of the swimmer model (the specific parameter values are given in Table 1 in Ref. [Paper A]).

In our model, the swimmer has access only to local information, and can only interact with its environment by taking local actions that do not rely on having access to any global information such as global directional sense. This is in contrast with previous similar studies on smart microswimmers [23, 24, 78, 80, 85, 86, 87], where the swimmers have access to global information. Our aim is to investigate how local hydromechanical signals, in the the swimmer's frame of reference can be used for efficient navigation. The results on which local states and actions are chosen is discussed in Section 5.1.

To model the swimmer's environment, both steady Taylor-Green vortex (TGV) flow (Section 3.1) and frozen realizations of Gaussian random velocity (GRV) field (Section 3.2), which are both statistically isotropic turbulent flow models, were used. (see Chapter 3).

Now, imagine yourself as a microswimmer in an isotropic flow, and consider that you can only measure the velocity and rotation differences between yourself and your surrounding fluid (Section 2.3). The inherent symmetry and isotropy of the ambient flow, makes it impossible to find a clue telling you whether you are moving in a certain direction or not, while you do not have an external guide. In result, it becomes impossible to have a net displacement in a certain direction. We found that in order to have a successful training for navigation in a certain direction, the environment's symmetry must be broken by either actions, environmental signals, or the dynamics of the swimmer. The importance of breaking this symmetry for successful navigation is stressed out in Section 5.2.

Finally, Section 5.3, provides an overview of the optimal strategy we found and answers the question of how the simple strategy for active reorientation that was the training result, can in practice, help the swimmer to sample upwelling regions more often.

5.1 Loosing the map

Optimal navigation, while having access to the global information for a microswimmer in complex flows, being subjected to a chaotic environment [79], is essentially a hard problem to solve. However, having a sense of the global information, is like that the fortunate microswimmer has a GPS in its pocket, that precisely tells where it should head, if the goal is in a certain direction. Although many species can have access to such sorts of information, for example by being sensitive to light [53, 88], or being able to follow the invisible trails of chemical signals [89], or being bottom-heavy (Figure 2.3), there are also many real-world situations where this is not the case. For example, what happens when there is not enough light, or if it is hard for the basic eyes of copepods [56] to determine the direction of the light source. Even, if the microswimmer has a preferred direction, for example given by a gyrotactic torque, efficient navigation still remains a very complicated problem. One reason is that the swimmer does not have any knowledge of the time dependent structure of the flow between itself and its target. For example, only heading in one direction while being advected by the flow can make the vertical navigation inefficient, because the swimmer might get trapped in vortices, if it does not know how to navigate in such regions. Therefore, being able to exploit the hydromechanical signals, that zooplankton species are experts in perceiving (Section 2.3), seems to be necessary in order to help them efficiently navigate in the turbulent ocean.

We discussed the possible signals that a planktonic microswimmer, such as copepod, can obtain from its surrounding fluid using the arrays of setae on their body in Section 2.3. But the question is which of these local signals are feasible to be used by a microswimmer in two-dimensional space for the purpose of vertical navigation?

To model local signals, we attach an orthogonal reference frame to a swimmer in two spatial dimensions (Figure 2.1) by using the swimming direction $\hat{\boldsymbol{n}}$ and the vector normal to that $\hat{\boldsymbol{p}}$, where $\hat{\boldsymbol{n}}$ and $\hat{\boldsymbol{p}}$ form a right-handed orthog-

onal basis. Now, we can describe the local components of the flow signals, using this frame of reference. Starting with the strain rate of the flow $\mathbb{S}$, which is one of the hydromechanical signals that plankton is capable of measuring (Section 2.3), we can decompose it into two independent components in this frame

$$S_{nn} = \hat{\boldsymbol{n}} \cdot \mathbb{S} \cdot \hat{\boldsymbol{n}}, \tag{5.1a}$$

$$S_{np} = \hat{\boldsymbol{n}} \cdot \mathbb{S} \cdot \hat{\boldsymbol{p}}. \tag{5.1b}$$

In addition to the strain rate $\mathbb{S}$, we assume that the swimmer can measure its velocity differences, both translational $\Delta \boldsymbol{u}$ and angular $\Delta \boldsymbol{\Omega}$, with respect to its surrounding fluid (Section 2.3). These two can also be written in the local frame we defined. For the translational velocity difference we have

$$\Delta u_n = (\boldsymbol{u} - \boldsymbol{v}) \cdot \hat{\boldsymbol{n}}, \tag{5.2a}$$

$$\Delta u_p = (\boldsymbol{u} - \boldsymbol{v}) \cdot \hat{\boldsymbol{p}}, \tag{5.2b}$$

where $\boldsymbol{u}$ is the flow velocity at the swimmer's position and $\boldsymbol{v}$ is defined in Eq. (2.3b). Since the angular velocity in the $\hat{\boldsymbol{x}}$-$\hat{\boldsymbol{y}}$ plane is always along the $\hat{\boldsymbol{z}}$ axis, the angular velocity difference $\Delta \boldsymbol{\Omega} = \Delta\Omega \hat{\boldsymbol{z}}$, and can simply be written as

$$\Delta\Omega \hat{\boldsymbol{z}} = (\boldsymbol{\omega}_f / 2 - \boldsymbol{\omega}), \tag{5.3}$$

where $\boldsymbol{\omega}_f = \boldsymbol{\nabla} \times \boldsymbol{u}$ is the flow vorticity at the swimmer's position and $\boldsymbol{\omega}$ is defined in Eq. (2.3c).

To use the above signals in the training with tabular reinforcement learning (RL) introduced in Chapter 4, we need to discretize the signals into different levels, to define a state space $\mathscr{S}$. The approach we take here, is dividing each signal into three levels, using a constant positive threshold value, denoted by adding a * as a superscript to the signal's name. As an example, for a signal σ, we have a threshold $\sigma^\star > 0$. The resulting state space then contains three states, that can be expressed by

$$\mathscr{S} = \{\sigma^-, \sigma^0, \sigma^+\}, \tag{5.4}$$

where σ^- corresponds to $\sigma \in (-\infty, -\sigma^\star)$, σ^0 corresponds to $\sigma \in [-\sigma^\star, +\sigma^\star]$, and σ^+ corresponds to $\sigma \in (+\sigma^\star, +\infty)$.

Since, we have multiple environmental signals here, it is possible to build larger states spaces by combining different signals and the resulting size of

$\mathscr{S}$ is 3^n, where n is the number of signals used. It is important to notice that the state space $\mathscr{S}$ size increases exponentially with the number of signals. Large state spaces make it very inefficient or even impossible to use Q-learning, because one of the conditions for convergence is that the agent should sample all the possible states a large enough number of times.

In order to completely move into the local frame of the swimmer, in addition to the signals, the control parameter $\boldsymbol{\omega}_s$ (Eq. (2.3c)), *i.e.* in an RL framework speaking, the actions that the swimmer can take, should also be local. This means the action should not have any dependency on the laboratory frame, or provide any other global information.

For this reason, we define the action to be contributing to angular dynamics by adding an angular velocity which has a constant magnitude $\omega_s = |\boldsymbol{\omega}_s|$. The swimmer can decide on, whether it should be clockwise $\boldsymbol{\omega}_s = -\omega_s \hat{\boldsymbol{z}}$, anti-clockwise $\boldsymbol{\omega}_s = \omega_s \hat{\boldsymbol{z}}$, or shouldn't be added at all $\boldsymbol{\omega}_s = 0$. These produce an action space $\mathscr{A}$ of size 3, which provides enough degrees of freedom for the swimmer to control its orientation in competition with the background flow and gravity. At the same time, the size of $\mathscr{A}$ remains small, which makes the further analysis of the trained strategies easier.

5.2 Settling helps ascension

Using the hydromechanical signals and the actions defined in Section 5.1, the microswimmer only has access to local information. In its isotropic universe, only being able to measure the velocity differences, in practice means that it will not be able to differentiate between moving up or down, without access to an external guide. This can be a disaster in order to navigate in a preferred direction. For example, the swimmer will not be able to swim upwards due to existence of vertical-reflection symmetry.

For the training to be successful, this symmetry needs to be broken somehow. When the actions and the state signals are local, the only possibility is to break the symmetry through the dynamics (Eq. (2.3)). While it sounds contradictory, having a settling velocity term in the dynamics can become helpful for a swimmer that is trying its best to swim upwards. Symmetry-breaking terms, such as settling velocity or gyrotaxis (Figure 2.3) can break the vertical-reflection symmetry, and in result, enable the RL training to find strategies that increase the rate at which swimmer samples the upwelling

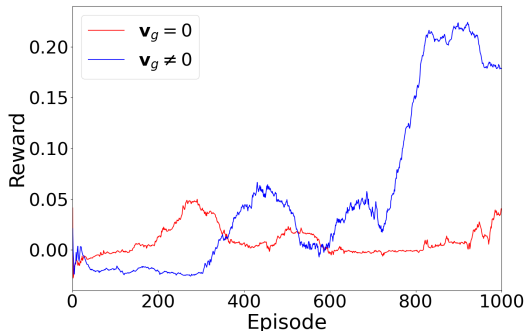


Figure 5.1: Accumulative reward per episode during training for swimmers with (blue) and without (red) settling velocity in their dynamics. There is no gyrotaxis ($B \to \infty$), and all other parameters are the same for both swimmers. The curves are smoothed by a moving average with a window of size 150 episodes.

regions of the flow.

Figure 5.1 compares the total accumulated reward during each training episode for two spheroidal swimmers ($\Lambda = 0.6$), with and without settling velocity. Both swimmers do not experience gyrotaxis ($B \to \infty$), and only have access to their local information. The goal is to find strategies for upward navigation. As the background, TGV flow (Section 3.1) is used. Figure 5.1, shows that, the swimmer with settling velocity, despite the fact that the settling velocity is pulling the swimmer down, achieves higher vertical migration compared to the swimmer without settling velocity in its dynamics. This is because the swimmer without settling does not learn how to swim upwards, and gains less reward, although it is not experiencing any downwards force. The reason is that the terms such as settling break the vertical-reflection symmetry in the problem, so the swimmer can have a sense of directions, still having access only to the local information.

It can also be shown that the same thing happens when the vertical-reflection symmetry is broken by other means, such as action or signals, as were the case in earlier studies [23, 24, 78, 80, 85, 86, 87], where the action or state signals are not local. Figure 2 in Ref. [Paper A], compares the normalized average vertical velocities for the training results with broken and unbroken vertical-reflection symmetry and different mechanisms of breaking this symmetry (Table 3 in Ref. [Paper A]).

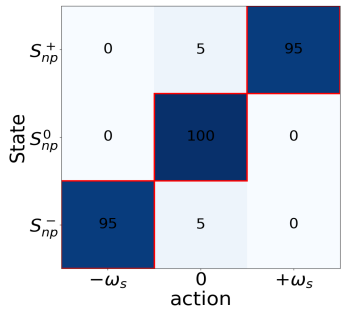


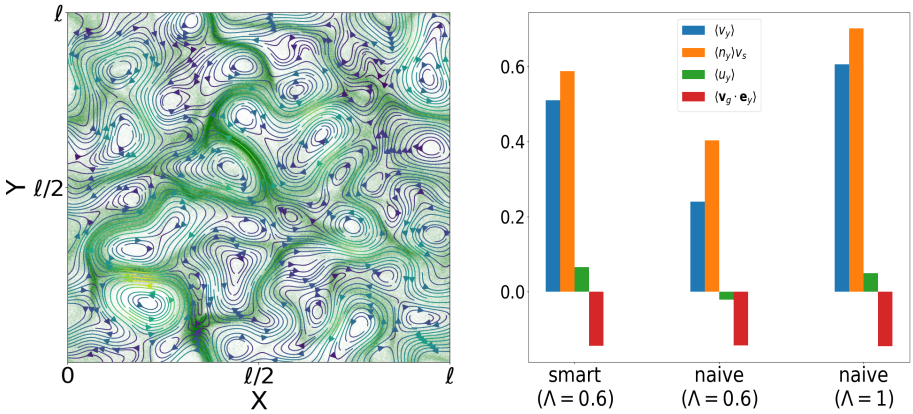
Figure 5.2: Q-table resulting from training with signal S_{np}. The numbers show the percentage of the trainings converged to each specific action.

Despite the fact, that we studied a highly idealized two-dimensional model, the importance of breaking this symmetry is a generic requirement to optimally navigate vertically upwards, when using local signals and actions. This means that breaking of such symmetries must also be necessary in real-case scenarios, where microswimmers such as copepods do not have access to any kind of global information in the vast isotropic turbulent ocean.

5.3 Becoming more slender

The result of training a swimmer with a reward function reinforcing positive vertical migration (Eq. (4.12)), using different sets of local signals (Section 5.1) is strategies that help the swimmer displace vertically upwards more efficiently compared to the same swimmer without active reorientation. The same strategies were found in training with both TGV flow (Section 3.1) and GRV fields (Section 3.2) as the background flows. This underlines the robustness of the found mechanism, which is independent of the flows used.

The strategy found using the signal defined by component of the strain rate $\mathbb{S}$ normal to the swimming direction $\hat{\boldsymbol{n}}$, which is denoted by S_{np} here (Eq. (5.1b)) is shown in Figure 5.2. Although, it is simple, the strategy is very efficient. Actively rotating, clockwise at state S_{np}^{-}, counter-clockwise at state S_{np}^{+}, and not rotating actively, whenever at state S_{np}^{0}, keeps the swimmer in left side of negative vortices, and the right side of positive vortices, as shown



(a) **(b)**

Figure 5.3: (a) Density plot of swimmers following the optimal strategy found (Figure 5.2).(b) Performance comparison for a smart swimmer in GRV field with two naïve swimmers in the same flow, one having exactly the same parameters ($\Lambda = 06$), and one being more slender ($\Lambda = 1$). The values are normalized by the swimmers swimming velocity v_s.

in Figure 5.3a. This simple mechanism increases the average vertical velocity of the swimmer by a factor of roughly 2, compared to a naïve strategy, which is swimming without any active reorientation *i.e.* $\boldsymbol{\omega}_s = 0$.

Training with other signals which are accessible to the swimmer (Section 5.1), shows that, the dominant signal is S_{np}. For a comparison between the different strategies we found for training with different state-spaces, the reader is referred to Figure 3 in Ref. [Paper A], which depicts that the signal S_{np} is more important for vertical navigation than the other signals.

A valid question that arises here, is to ask how does the strategy shown in Figure 5.2 help the swimmer in navigating upwards more efficiently? By another look at the dominant signal S_{np} in Eq. 5.1b, one can see it is possible to rewrite it on the form

$$S_{np} = \boldsymbol{n} \cdot \mathbb{S} \cdot \boldsymbol{p} = (\boldsymbol{n} \times \mathbb{S}\boldsymbol{n}) \cdot \hat{\boldsymbol{z}}. \tag{5.5}$$

This is similar to the Jeffery's term in the equation of motion (Eq. (2.3c)), which describes how the strain rate of the flow rotates a non-spherical swimmer ($\Lambda > 0$). If $S_{np} > 0$, the swimmer is rotated counter-clockwise and the

opposite happens when $S_{np} < 0$. This is exactly what the active navigation term is also adding to the dynamics (Figure 5.2). The meaning of this is that the trained swimmer actively mimics a more slender swimmer (increasing the effective Λ), but why should it do so? The importance of the shape parameter Λ in preferential sampling of upwelling regions in the flow, has been extensively studied in our group before [46]. It is shown that spherical gyrotactic swimmers preferentially sample the downwelling regions of the flow [45]. However, the results in Ref. [46] predict that a prolate gyrotactic swimmer (naïve strategy here), starts to sample upwelling regions when it become more slender than a critical length ($\Lambda > \Lambda_c$). This transition point (Λ_c) depends on the parameters of the flow and the swimmer. Therefore, actively mimicking a more slender swimmer can increase the rate of preferential sampling of upwelling regions.

Figure 5.3b compares the vertical migration performance of a smart swimmer following the optimal strategy shown in Figure 5.2 with $\Lambda = 0.6$, against the naïve swimmers with all the same parameters, except with different Λ values, swimming in GRV fields (Section 3.2). Figure 5.3b shows that the performance of the smart swimmer is almost two times better compared to a naïve swimmer with the same parameters and it is close to the performance that a naïve swimmer with roughly 2 times larger shape parameter Λ has. For the same comparison in TGV flow (Section 3.1), look at the Figure 5 in Ref. [Paper A].

In conclusion, following the simple active reorientation mechanism that we found here (Figure 5.2), increases the vertical migration of a prolate naïve swimmer with a factor of approximately 2. In addition, it provides the possibility of sampling upwelling regions of the flow, even for spherical swimmers, which was not possible before [45]. Figure 7 in Ref. [Paper A] shows that the mechanism is robust to small perturbations of the parameters, such as threshold values, and the translational and angular swimming velocities. However, under substantial changes to the parameters, the optimal behavior is also expected to become different. For example, if the translational swimming velocity is greatly increased, the swimmer can simply ignore the background flow and preferential sampling becomes weaker [46]. As a result, it is not easy to outperform the naïve strategy (passive gyrotactic swimmer) in this limit. However, the fact that we found the same mechanism in two different flow models, puts an emphasis on the robustness of the mechanism for the parameters used.

6 Plankton navigation in three dimensions

We continue with Chapter 6, that provides an introduction to the main findings reported in Ref. [Paper B]. There, we generalized our microswimmer model to three dimensions. Dealing with a three-dimensional environment is more challenging, since on the one hand, the number of signals that can affect the swimmer's navigation increases, and on the other hand, the dynamics in three spatial dimensions become more complex. Here as well, the swimmer is assumed to follow the dynamics written in Eq. (2.3), and has a spheroidal shape as described in Section 2.1. Gravity points in the negative z direction, $-\hat{\boldsymbol{z}}$. The goal is to find optimal navigation strategies for swimming upwards (against gravity). The swimmer is bottom-heavy (Figure 2.3) and settling is present.

The discussion in Section 6.1 is to point out the difference in available signals and actions to be exploited by the swimmer, in comparison to the two-dimensional model. In Section 6.2, the results of solving a simplified version of the problem in three dimensions is presented. The reason of simplification is that the number of possible strategies in three-dimensional model becomes of the order 10^{77}, which makes it extremely hard to converge to the optimal solution and analyze the results. Therefore, we constrain the swimmer to have only access to information in a certain plane. In Section 6.3, the results from training a swimmer with full information accessible in a three-dimensional model can be found. Having access to full information improves the performance of the swimmer in vertical migration, but finds the same mechanism as we found for simplified model (Section 6.2), suggesting robustness of the mechanism. Finally, we discuss the difference between the optimal strategies in two (Section 5.3) and three (Section 6.3) dimensions.

For the three-dimensional model we used DNS simulation and GRV fields (Section 3.2) as the background flow model. I did the simulations in GRV fields and the DNS simulations were carried out by our collaborators, Jingran Qiu and Lihao Zhang at Tsinghua university in Beijing. It is worth mentioning, that the optimal strategies found in both flows are in agreement, which shows the robustness of the active gyrotactic mechanism discussed in Sections 6.2 and 6.3, in different flow environments.

6.1 Sense and act

In this Section, the possible signals and their validity are briefly discussed. The arguments in Section 5.1, on the local signals and necessity of the symmetry breaking holds here as well. Therefore, without repeating, I just describe the available signals, and the actions that our microswimmer can take in three dimensions.

To describe the local frame of the swimmer, we add an axis normal to our previously defined $\hat{\boldsymbol{n}}$-$\hat{\boldsymbol{p}}$ plane (Section 5.1), and call it $\hat{\boldsymbol{q}}$, such that $\hat{\boldsymbol{n}}$-$\hat{\boldsymbol{p}}$-$\hat{\boldsymbol{q}}$ form a right handed basis (Figure 2.1). Projecting the potential hydromechanical signals accessible to the swimmer (Figure 2.5), namely the translational slip velocity $\Delta\boldsymbol{u}$, angular slip velocity $\Delta\boldsymbol{\Omega}$, and the strain rate tensor S_{ij} (Eq. (2.5a)), in the defined local frame, we examine the applicability of each signal. Following the dynamics described in Section 2.2, the local components of the translational velocity difference $\Delta\boldsymbol{u}$ are

$$\Delta u_n = (\boldsymbol{u}-\boldsymbol{v})\cdot\hat{\boldsymbol{n}} = v_{\parallel}^{(g)} n_z - v_s\,, \tag{6.1a}$$

$$\Delta u_p = (\boldsymbol{u}-\boldsymbol{v})\cdot\hat{\boldsymbol{p}} = v_{\perp}^{(g)} p_z, \tag{6.1b}$$

$$\Delta u_q = (\boldsymbol{u}-\boldsymbol{v})\cdot\hat{\boldsymbol{q}} = v_{\perp}^{(g)} q_z, \tag{6.1c}$$

and for the angular velocity difference $\Delta\boldsymbol{\Omega}$ one can write,

$$\Delta\Omega_n = (\boldsymbol{\omega}_f/2-\boldsymbol{\omega})\cdot\hat{\boldsymbol{n}} = -\omega_n^{(s)}, \tag{6.1d}$$

$$\Delta\Omega_p = (\boldsymbol{\omega}_f/2-\boldsymbol{\omega})\cdot\hat{\boldsymbol{p}} = \Lambda S_{nq} + \frac{1}{2B} q_z - \omega_p^{(s)}, \tag{6.1e}$$

$$\Delta\Omega_q = (\boldsymbol{\omega}_f/2-\boldsymbol{\omega})\cdot\hat{\boldsymbol{q}} = -\Lambda S_{np} + \frac{1}{2B} p_z - \omega_q^{(s)}, \tag{6.1f}$$

where $S_{ij} = \hat{\boldsymbol{i}}\cdot\mathbb{S}\cdot\hat{\boldsymbol{j}}$, with $\hat{\boldsymbol{i}},\hat{\boldsymbol{j}} \in \{\hat{\boldsymbol{n}},\hat{\boldsymbol{p}},\hat{\boldsymbol{q}}\}$, themselves form 9 candidate signals. Not all of the above signals are independent and meaningful to be used in the training. For example, signal Δu_n (Eq. (6.1a)) is much greater than the threshold value we used for sensing the slip velocity (Section 2.3), therefore, we ignore this signal to avoid having to add more thresholds. Also, Eq. (6.1d), does not provide any new information, since it is simply equivalent to the swimmer's control angular velocity projected on $\hat{\boldsymbol{n}}$. Eqs. (6.1e), and (6.1f), depend on p_z, q_z, and the local components of the $\mathbb{S}$, which are already given by Eqs. (6.1c), (6.1b), and by $\mathbb{S}$ components. This means Eqs. (6.1e) and (6.1f)

are not providing any independent information. Furthermore, no other component of the strain rate tensor $\mathbb{S}$, appears in the dynamics in Eq. (6.1), except S_{np} and S_{nq}, as a result other components should be less important in the dynamics and one can disregard them as well. Therefore, we neglect the signal $\Delta\boldsymbol{\Omega}$, and the components of $\mathbb{S}$ other than the two appearing in Eq. (6.1). At the end, we are left with 4 signals being p_z, q_z, S_{np}, and S_{nq} that we expect to be the most important independent signals for training.

Similar to Section 5.1, in order to define the training state space $\mathscr{S}$ we divide each signal into three levels (Eq. (5.4)). Having four independent signals, the largest state space $\mathscr{S}$ that we can build has 81 states.

To define the action space $\mathscr{A}$, we follow the same procedure as in the two-dimensional model (Section 5.1). In addition to the possibility of rotation with three different angular velocities we assumed in Section 5.1, taking into account that the swimmer in three spatial dimensions can rotate around axis $\hat{\boldsymbol{p}}$ as well, gives rise to a larger action space $\mathscr{A}$. Therefore, the swimmer's reorientation in three dimensions, can be expressed as a superposition of the rotations around two axes $\hat{\boldsymbol{p}}$ and $\hat{\boldsymbol{q}}$. Finally, we write the swimmer's contribution to its angular dynamics (Eq. (2.3c)) as

$$\boldsymbol{\omega}_s = \boldsymbol{\omega}_p + \boldsymbol{\omega}_q \,, \tag{6.2}$$

where $\boldsymbol{\omega}_p = \omega_s \hat{\boldsymbol{p}}$ is the rotation around $\hat{\boldsymbol{p}}$, and similarly, $\boldsymbol{\omega}_q = \omega_s \hat{\boldsymbol{q}}$ is the rotation around $\hat{\boldsymbol{q}}$. Giving the swimmer a three level control over ω_s similar to the two-dimensional model (Section 5.1), to control each component $\boldsymbol{\omega}_p$ and $\boldsymbol{\omega}_q$ independently, produces an action space $\mathscr{A}$ of size 9 in three-dimensional space.

6.2 Planar simplification

In general, any change in the swimming direction $\hat{\boldsymbol{n}}$, can be described as a combination of rotations around the local basis vectors $\hat{\boldsymbol{p}}$ and $\hat{\boldsymbol{q}}$ (Figure 2.1). Therefore, if the swimmer has a full control over its rotation around axes $\hat{\boldsymbol{p}}$ and $\hat{\boldsymbol{q}}$ by applying an angular velocity $\boldsymbol{\omega}_s$ (Eq. (6.2)), it should have perfect control of the steering. As described in Section 6.1, our swimmer, can only choose between nine possible discrete values to set its angular velocity $\boldsymbol{\omega}_s$. In addition, the magnitude of the control angular velocity $\boldsymbol{\omega}_s$ is of the same order as the flow vorticity, which makes the navigation even harder.

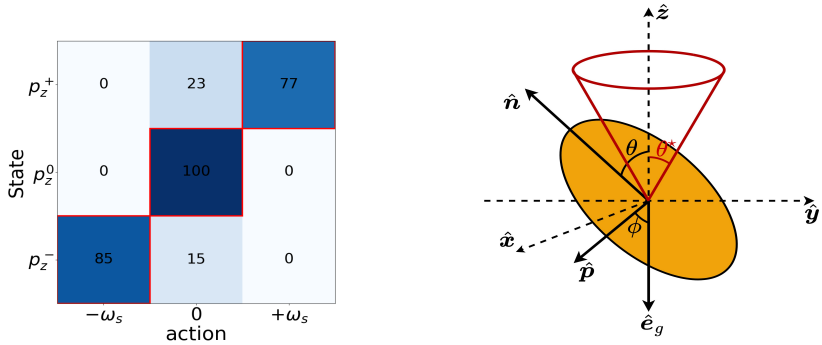


Figure 6.1: (a) The optimal strategy found from training in the planar model using signal p_z. The numbers show the percentage of trainings converged to the specific actions for each state. (b) Schematic explaining how the found mechanism aligns the swimming direction with the target direction.

From a reinforcement learning perspective, even with our idealized model, having only 9 actions and 4 independent signals introduced earlier, each with only 3 levels (Section 6.1), we will have a Q-table of dimension 81×9. This means that the number of possible strategies accessible is of the order 10^{77}. This large number makes it extremely hard for both the training to converge, and the interpretation of the resulting optimal strategy.

In order to prevent these complications, we simplify by reducing the size of the state space of the problem, considering first a planar model. In this model, the swimmer can only actively rotate around the axis $\hat{\boldsymbol{q}}$, and no active rotation around $\hat{\boldsymbol{p}}$ is allowed ($\boldsymbol{\omega}_p = 0$). This assumption, reduces the number of actions to 3. Moreover, we consider that the signals can only be measured in the $\hat{\boldsymbol{n}}$-$\hat{\boldsymbol{p}}$ plane, so the valid signals are S_{np} and p_z. This is actually similar to the two-dimensional model that we discussed in Chapter 5, with the difference that the $\hat{\boldsymbol{n}}$-$\hat{\boldsymbol{p}}$ plane itself, does not have a fixed orientation, since its direction changes as the swimmer rotates.

Comparing the training results for the two cases of either only the p_z signal, or using both S_{np} and p_z, shows that adding the signal S_{np} does not alter the performance very much. Therefore, the dominant signal is p_z, and of course by symmetry q_z, if we consider the planar model in the $\hat{\boldsymbol{n}}$-$\hat{\boldsymbol{q}}$ plane. The resulting strategy found for the planar model is shown in Figure 6.1a.

It is very efficient in increasing the vertical displacement compared to a naïve swimmer that has the same properties, but without the active rotation term in the dynamics ($\boldsymbol{\omega}_s = 0$). Looking at Figure 6.1a, and following the state notation defined in Eq. (5.4), one can express the resulting strategy as clockwise rotation around $\hat{\boldsymbol{q}}$ when the swimmer is in the state p_z^-, counter-clockwise rotation around $\hat{\boldsymbol{q}}$ when in state p_z^+, and no active rotation when the state is p_z^0.

To understand the mechanism of the optimal strategy (Figure 6.1a), notice that if we denote the angle between the swimming direction and the preferred direction ($\hat{\boldsymbol{z}}$) as $\theta = \cos^{-1}(\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{z}})$ (Figure 6.1b), the threshold $p_z^\star$ that we chose for p_z in order to define the three states (Eq. 5.4), corresponds to an equivalent angular threshold $\theta^\star$, which determines how much deviation from preferred direction is tolerable before the swimmer actively contributes to its angular dynamics (Eq. (2.3c)), following the optimal policy (Figure 6.1a). In addition, we can see that p_z is proportional to the cosine of the angle between axis $\hat{\boldsymbol{p}}$ and the gravity direction $\hat{\boldsymbol{e}}_g = -\hat{\boldsymbol{z}}$ (Figure 6.1b). Naming this angle $\phi = \cos^{-1}(\hat{\boldsymbol{p}} \cdot \hat{\boldsymbol{e}}_g)$, and following the optimal strategy found (Figure 6.1a), implies that when the angle $\phi < \pi/2 - \theta^\star$ (*i.e.* state p_z^-), the resulting rotation around $\hat{\boldsymbol{q}}$ will be clockwise, which tends to align the swimming direction with $\hat{\boldsymbol{z}}$. The same effect can be produced by counter-clockwise rotation when $\phi > \pi/2 + \theta^\star$ (*i.e.* state p_z^+). Finally, when $\phi \in [\pi/2 - \theta^\star, \pi/2 + \theta^\star]$, the active contribution is zero. Therefore, the mechanism found here gives an active alignment of the swimming direction towards the target direction.

6.3 Full three-dimensional model

Now that we know from the planar model results (Section 6.2), that the dominant signals are p_z and q_z, we can simply ignore the strain components S_{np} and S_{nq} in the beginning. Considering the full three-dimensional model, where the swimmer can actively rotate around both axes $\hat{\boldsymbol{p}}$ and $\hat{\boldsymbol{q}}$, by controlling the corresponding angular velocity components $\boldsymbol{\omega}_p$ and $\boldsymbol{\omega}_q$. Thus, the swimmer can have access to all of the 9 actions defined in Eq. (6.2).

Taking into account the right handedness of the local coordinate system defined, and the mechanism we found for the planar model (Section 6.2), one would expect to obtain the same behavior for the added rotation around axis $\hat{\boldsymbol{p}}$, but with a negative sign. Therefore, the optimal strategy should be

clockwise rotation around $\hat{\boldsymbol{p}}$ when state is q_z^+, counter-clockwise rotation around $\hat{\boldsymbol{p}}$ when state is q_z^-, and no active rotation when state is q_z^0, in addition to the rotations around $\hat{\boldsymbol{q}}$ as described in Section 6.2.

The training results with 9 states and 9 actions are in agreement with the expectation, as the resulting Q-table presented in Figure 2 of Ref. [Paper B], shows. The numbers in the Q-table show the percentage of the trainings that converged to a certain action in each state. Since the process is stochastic, each training can end up in sub-optimal strategies that are slightly different but have similar performance. This effect increases by enlarging the size of the state space, since the number of possible strategies increases drastically, and as a result larger number of strategies can produce the similar expected long term reward. In summary, the trend expected from the mechanism described above, is clearly visible in the Q-table presented in Figure 2 of Ref. [Paper B].

Combining all of the possible 4 signals and 9 actions (Section 6.1), we can perform trainings to search for the optimal strategies when the swimmer has access to the full possible information in our three-dimensional model. As mentioned before, the large size of the state space in this case makes it impossible to interpret the resulting Q-table properly. However, we can at least compare the performance of the strategy we find by training with full information, against a strategy that we propose by combining the result of our two-dimensional study (Section 5.2) and the mechanism found in Section 6.2. The dominant signals in three dimensions were found to be p_z and q_z, so we give them priority over S_{np} and S_{nq}. Therefore, we follow the mechanism described in Section 6.2, but add a refinement using S_{np} and S_{nq} signals when the swimmer is in the state with both p_z^0 and q_z^0. In that state the swimmer follows the mechanism found in two dimensions (Figure 5.2), *i.e.* rotating around $\hat{\boldsymbol{p}}$ counter-clockwise when the swimmer is in state S_{nq}^+ and do the opposite when it is in state S_{nq}^-. In addition, to rotate around $\hat{\boldsymbol{q}}$ with the same mechanism but based on the signal S_{np}. It turns out that the optimal strategy found by training with the full information (81 states) has a performance on the same order as the suggested strategy, and does not over-perform it.

It should be noted, that the same mechanism was found by training in both GRV fields (Section 3.2), and DNS simulation, which shows the robustness of the mechanism found in different flows. Also, the robustness of the

mechanisms were examined in different flow regimes and against perturbations in threshold values. Sensitivity of the performance of the above mechanism, with respect to the signal threshold values $p_z^\star$, was tested and the result suggests that as long as the threshold value is smaller than 0.5, the performance is on the same level (Figs. 3a and 3b in Ref. [Paper B]). In Figure 3c in Ref. [Paper B], the dependence of the average vertical velocity on the dimensionless angular velocity is shown, and one can see that the performance is reaching a plateau for $|\omega_s|\tau_\eta > 0.8$. This implies that the performance is independent of the swimming angular velocity for a large range of values. Moreover, Figure 4 in Ref. [Paper B] compares the performance of the active gyrotactic mechanism found for different number of signals with the naïve strategy *i.e.* passive gyrotactic swimmer performance, over a large range of different flow parameters. The comparison shows that the performance of the mechanism is very stable over a large parameter range of τ_η and u_0.

Another interesting result in the three-dimensional model, is that an efficient strategy for swimming downwards is obtained by mirroring the optimal strategy for upwards swimming. This means that, the swimmer accessing the same information and having the same set of actions can optimally swim down to deeper regions of the flow, just by mirroring its actions, *i.e.* simply switching the clockwise rotations with counter-clockwise and vice-versa in the optimal strategy shown in Figure 6.1a.

Finally, we contrast the optimal strategies we found in the two- and three-dimensional swimmer models. In two dimensions, the dominant signal is the strain component S_{np}, while in three dimensions it is the velocity difference (or equivalently p_z and q_z, Eq. (6.1)). The reason is that in the two-dimensional steady flows with the parameters used in Ref. [Paper A], a swimmer navigating using the optimal strategy based on p_z found in three dimensions often ends up in limit cycles and stable fixed points of the dynamics. The same trapping effect can be observed for swimmers without active navigation in shear flows [15, 49, 90]. However, this is not happening when the flow is three-dimensional or shows quick fluctuations because of the turbulence. It should be noted that we found the similar mechanism to three-dimensional case, in two spatial dimensions as well, when the swimmer was trained with signals $\{\Delta u_p, S_{np}\}$, but it was only the dominant effect when the signal $S_{np} \sim 0$ (Figure 3d in Ref. [Paper A]).

7 Conclusions and Outlook

Many species of zooplankton, such as copepods, are active microswimmers. They can measure hydromechanical signals, using the arrays of setae over their body and antennae (Section 2.3), and have the ability to refine their motion, for example by changing their orientation or jump, in order to achieve certain goals. One of the important behaviors that these creatures show, is their large-scale daily vertical migration [11]. Therefore, it is crucial for their survival to have efficient strategies, developed by evolution, to swim long distances upwards and downwards. Zooplankton exploit passive strategies that depend on the physical properties of the them for this purpose, for example, gyrotaxis due to the plankton being bottom-heavy. However, relying only on the passive strategies does not seem to be realistic, since passive strategies do not perform well in many situations. For example, in turbulent flows gyrotactic swimmers end up being trapped in certain flow regions [15, 49]. Moreover, the existence of passive strategies does not rule out the possibility of developing active strategies as well, since they after all have access to a wide range of hydromechanical signals and can control their navigation. This is what we tried to address in this thesis. We used reinforcement learning (RL) (Chapter 4), which provides a tremendous framework to deal with these types of problems. RL brings the possibility of efficiently finding optimal strategies in extremely huge spaces of possible strategies for control problems. Therefore, We modeled microswimmers with a small control over their angular velocity (Eq. 2.3c) in an RL framework, and searched for optimal active strategies for vertical navigation.

Unlike previous works on navigation of microswimmers, our swimmer is allowed to only exploit local information. Our results show the importance of symmetry breaking for successful navigation in an isotropic environment, such as the turbulent ocean (Section 5.2). As an example, in our quest for positive vertical migration, vertical reflection symmetry must be broken somehow, either by allowing the swimmer access to non-local information in its signals and actions [23, 24, 78, 80, 85, 86, 87], or through the dynamics by symmetry-breaking mechanisms, such as settling [Papers A, B].

Two different mechanisms for active vertical migration were found in models of active swimmers in two- and three-dimensional flows. The strategies differ in that the most relevant signals turn out to be different in the

two cases. The mechanism in three dimensions is to actively keep the alignment towards the target direction (Section 6.3). This mechanism fails in the two-dimensional case because it tends to cause the swimmer to end-up in limit cycles or stable fixed points of the dynamics. The same kind of trapping happens for gyrotactic passive swimmers [15, 49, 90]. The mechanism of the optimal strategy in two dimensions is instead to mimic more slender swimmers (Section 5.3), since this increases the sampling rate of upwelling regions [46].

In both cases the mechanism was found from trainings in different flows, suggesting that the mechanisms are robust to changes in the flows with the parameters used. Also, the strategies remain robust against small perturbations of the swimmer- and flow-parameters (Secs. 5.3, 6.3). Moreover, the mechanism found in three-dimensional flow also provides an efficient strategy for downwards swimming, if the swimmer switches the direction of active rotation in the optimal upwards navigation strategy (Figure 6.1a).

In summary, both mechanisms enhance the vertical migration drastically, and provide means to have positive vertical migration independent of the swimmer's shape, and passive gyrotactic ability. This suggests that micro-organisms, such as copepods, might have developed active navigation strategies, in addition to the passive strategies that they exploit, for efficient vertical migration which is an important part of their life.

Outlook

Many questions remain to be answered to understand the behavior of plankton in nature. Below I pose a number of important questions related to different aspects of the model I used in this thesis.

Our assumptions that the swimmer is a rigid spheroid without any deformation, or that it has a constant translational swimming speed are not the case in nature. In reality, swimmers can deform and use their deformation as a tool for more efficient swimming as has been studied on fish and larval fish in Refs. [91, 92]. In addition, many microswimmers in nature can change their swimming speed and copepods can perform jumps with huge momentary speeds [54, 61]. Therefore, our model is restricted to species that are only cruising *i.e.* not jumping, or to the time intervals during which a jumping microswimmer moves with constant speed. How does adding

shape deformation, and the possibility of taking actions such as stopping and jumping affect the performance in efficient vertical migration, are interesting questions to be answered.

Another consideration is that the swimmer's movement does not alter the surrounding flow in our model. Although, it is believed that microswimmers such as copepods can distinguish between their own and the external disturbances in the flow [60, 93], a more realistic model, taking into account the flow disturbances produced by the swimmer itself, might change the signals the swimmer measures, and in result change the optimal strategy.

An interesting question is how the strategy will change if there were different goals to be optimized? Here we only considered the vertical migration, but in nature plankton pursuit competing goals such as reducing energy consumption, or avoiding high strain regions in the flow for protection against predators. Copepods try to stay in regions of the flow with moderate strain rate to increase the accuracy in detecting the predator's flow disturbances [28, 30]. For example, in Ref. [18], a Lagrangian model of jumping copepods is given, where the swimmers can jump with large speeds in its instantaneous swimming direction, if the strain rate is larger than a threshold. It would be interesting to study whether the same threshold can be found by training a smart jumping swimmer, or how the result changes if the swimmer can actively choose its jumping direction. Moreover, in the mechanism that we found in two-dimensional flow, the swimmer exploits high strain regions of the flow to more efficiently swim upwards. It seems intriguing to study if it is possible to efficiently swim upwards and at the same time avoid high-strain regions. In addition, studying the optimal navigation for the purposes of increasing the encounter rate with other swimmers to mate or feed are other lines of research to pursue.

For the training environment, we used different models of turbulent flow, such as Taylor-Green vortex (TGV) flow, Gaussian random velocity (GRV) field, and direct numerical simulation (DNS), but all of them were frozen in time. A more realistic picture can be obtained by using time-dependent flows. This will increase the computational costs and make the training more complicated, but we are working on implementing time-dependent flows with a new method that decreases the computational costs, and makes it more efficient to perform the training in time-dependent flows. Using unsteady flows might change the optimal navigation strategies or the stability of found mechanisms. In addition, a swimmer which is trained in a time-

varying flow might be able to exploit temporal cues as well as spatial structure of the flow, and find more generic and robust navigation mechanisms that might be more realistic.

Taking into account the above considerations, will increase the size of the swimmer's observation space. Using tabular Q-learning makes it hard, if not impossible, to deal with large state spaces. Therefore, in order to solve more realistic models of the swimmer, we are aiming to use deep Q-learning [94], which replaces the Q-table with a Q-function using artificial neural networks [20]. Using Deep RL methods have other benefits as well. For example, there is no need to manually specify the threshold values, since they can be found by the swimmer (agent) itself. Also, because of using a function approximator instead of a lookup table, the agent will be able to deal with unseen states, which makes the training results more general and more robust. On the downside, understanding the mechanism is easier by interpreting a Q-table as small as possible, rather than interpreting the training results with full information.

After all of these technical improvements in modeling mentioned above have been carried out, an interesting suggestion is to conduct experiments that can show whether our found active mechanisms are valid in reality or not? Can a microorganism such as a copepod exploit the proposed mechanisms in this thesis as a complementary guide for its efficient swimming towards its goals? Also, in the future it may be possible to build artificial swimmers with similar properties modeled here. It would be very exciting to see if an artificial robotic swimmer can swim smartly in the unpredictable fluid environment following the mechanisms and strategies reported in this thesis.

Bibliography

[1] HENSEN, V 1887 *Ueber die Bestimmung des Plankton's oder des im Meere treibenden Materials an Pflanzen und Thieren.*

[2] SARDET, C 2015 *Plankton: wonders of the drifting world*. University of Chicago Press.

[3] GUASTO, J. S, RUSCONI, R & STOCKER, R 2012 Fluid mechanics of planktonic microorganisms. *Annual Review of Fluid Mechanics* **44**, 373–400.

[4] GIBBONS, M. J 1997 *An introduction to the Zooplankton of the Benguella current Region.*

[5] ARIZA, A, GARIJO, J, LANDEIRA, J, BORDES, F & HERNÁNDEZ-LEÓN, S 2015 Migrant biomass and respiratory carbon flux by zooplankton and micronekton in the subtropical northeast atlantic ocean (canary islands). *Progress in Oceanography* **134**, 330–342.

[6] WOODWARD, F. I, BARDGETT, R. D, RAVEN, J. A & HETHERINGTON, A. M 2009 Biological approaches to global environment change mitigation and remediation. *Current Biology* **19** (14), R615–R623.

[7] MOORE, M & FOLT, C 1993 Zooplankton body size and community structure: effects of thermal and toxicant stress. *Trends in ecology & evolution* **8** (5), 178–183.

[8] BRUN, P, STAMIESZKIN, K, VISSER, A. W, LICANDRO, P, PAYNE, M. R & KIØRBOE, T 2019 Climate change has altered zooplankton-fuelled carbon export in the north atlantic. *Nature ecology & evolution* **3** (3), 416–423.

[9] LJUNGSTRÖM, G, CLAIREAUX, M, FIKSEN, Ø & JØRGENSEN, C 2020 Body size adaptions under climate change: zooplankton community more important than temperature or food abundance in model of a zooplanktivorous fish. *Marine Ecology Progress Series* **636**, 1–18.

[10] LAMPITT, R. S, ACHTERBERG, E. P, ANDERSON, T. R, HUGHES, J, IGLESIAS-RODRIGUEZ, M, KELLY-GERREYN, B. A, LUCAS, M, POPOVA, E, SANDERS,

R, SHEPHERD, J *et al.* 2008 Ocean fertilization: a potential means of geoengineering? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **366** (1882), 3919–3945.

[11] BRIERLEY, A. S 2014 Diel vertical migration. *Current biology* **24** (22), R1074–R1076.

[12] BANDARA, K, VARPE, Ø, WIJEWARDENE, L, TVERBERG, V & EIANE, K 2021 Two hundred years of zooplankton vertical migration research. *Biological Reviews* **96** (4), 1547–1589.

[13] EVERETT, J. D, BAIRD, M. E, BUCHANAN, P, BULMAN, C, DAVIES, C, DOWNIE, R, GRIFFITHS, C, HENEGHAN, R, KLOSER, R. J, LAIOLO, L *et al.* 2017 Modeling what we sample and sampling what we model: challenges for zooplankton model assessment. *Frontiers in Marine Science* **4**, 77.

[14] VISSER, A. W 1997 Using random walk models to simulate the vertical distribution of particles in a turbulent water column. *Marine Ecology Progress Series* **158**, 275–281.

[15] DURHAM, W. M, KESSLER, J. O & STOCKER, R 2009 Disruption of vertical motility by shear triggers formation of thin phytoplankton layers. *science* **323** (5917), 1067–1070.

[16] JIANG, H, OSBORN, T. R & MENEVEAU, C 2002 The flow field around a freely swimming copepod in steady motion. part i: Theoretical analysis. *Journal of Plankton Research* **24** (3), 167–189.

[17] MOLLOY, P & COWLING, M 1999 Buoyancy mechanisms of marine organisms: lessons from nature. *Underwater Technology* **24** (2), 41–49.

[18] ARDESHIRI, H, BENKEDDAD, I, SCHMITT, F. G, SOUISSI, S, TOSCHI, F & CALZAVARINI, E 2016 Lagrangian model of copepod dynamics: Clustering by escape jumps in turbulence. *Physical Review E* **93** (4), 043117.

[19] BORGNINO, M, ARRIETA, J, BOFFETTA, G, DE LILLO, F & TUVAL, I 2019 Turbulence induces clustering and segregation of non-motile, buoyancy-regulating phytoplankton. *Journal of the Royal Society Interface* **16** (159), 20190324.

[20] MEHLIG, B 2021 *Machine Learning with Neural Networks: An Introduction for Scientists and Engineers*. Cambridge University Press.

[21] CICHOS, F, GUSTAVSSON, K, MEHLIG, B & VOLPE, G 2020 Machine learning for active matter. *Nature Machine Intelligence* **2** (2), 94–103.

[22] SUTTON, R. S & BARTO, A. G 2018 *Reinforcement learning: An introduction*. MIT press.

[23] COLABRESE, S, GUSTAVSSON, K, CELANI, A & BIFERALE, L 2017 Flow navigation by smart microswimmers via reinforcement learning. *Physical review letters* **118** (15), 158004.

[24] GUSTAVSSON, K, BIFERALE, L, CELANI, A & COLABRESE, S 2017 Finding efficient swimming strategies in a three-dimensional chaotic flow by reinforcement learning. *The European Physical Journal E* **40** (12), 1–6.

[25] LALLI, C & PARSONS, T 1997 *Biological oceanography: an introduction*. Elsevier.

[26] VISSER, A. W 2011 Small, wet and rational. *Dr. Techn. Dissertation. Technical University of Denmark, Copenhagen* .

[27] JIANG, H & KIØRBOE, T 2011 The fluid dynamics of swimming by jumping in copepods. *Journal of the Royal Society Interface* **8** (61), 1090–1103.

[28] KIØRBOE, T 2018 *A mechanistic approach to plankton ecology*. Princeton University Press.

[29] JIANG, H & OSBORN, T. R 2004 Hydrodynamics of copepods: a review. *Surveys in geophysics* **25** (3), 339–370.

[30] KIØRBOE, T & VISSER, A. W 1999 Predator and prey perception in copepods due to hydromechanical signals. *Marine Ecology Progress Series* **179**, 81–95.

[31] JAKOBSEN, H. H 2001 Escape response of planktonic protists to fluid mechanical signals. *Marine Ecology Progress Series* **214**, 67–78.

[32] SEURONT, L, SCHMITT, F. G, BREWER, M. C, STRICKLER, J. R & SOUISSI, S 2004 From random walk to multifractal random walk in zooplankton swimming behavior. *Zool. Stud* **43** (2), 498–510.

[33] VISSER, A. W 2008 Lagrangian modelling of plankton motion: From deceptively simple random walks to fokker–planck and back again. *Journal of Marine Systems* **70** (3-4), 287–299.

[34] SVETLICHNY, L, LARSEN, P. S & KIØRBOE, T 2020 Kinematic and dynamic scaling of copepod swimming. *Fluids* **5** (2), 68.

[35] BRON, J. E, FRISCH, D, GOETZE, E, JOHNSON, S. C, LEE, C. E & WYNGAARD, G. A 2011 Observing copepods through a genomic lens. *Frontiers in zoology* **8** (1), 1–15.

[36] TITELMAN, J 2001 Swimming and escape behavior of copepod nauplii: implications for predator-prey interactions among copepods. *Marine Ecology Progress Series* **213**, 203–213.

[37] TITELMAN, J & KIØRBOE, T 2003 Motility of copepod nauplii and implications for food encounter. *Marine Ecology Progress Series* **247**, 123–135.

[38] CARLOTTI, F, BONNET, D & HALSBAND-LENK, C 2007 Development and growth rates of centropages typicus. *Progress in Oceanography* **72** (2-3), 164–194.

[39] KNUTSEN, T, MELLE, W & CALISE, L 2001 Determining the mass density of marine copepods and their eggs with a critical focus on some of the previously used methods. *Journal of Plankton Research* **23** (8), 859–873.

[40] MILLERO, F. J, CHEN, C.-T, BRADSHAW, A & SCHLEICHER, K 1980 A new high pressure equation of state for seawater. *Deep Sea Research Part A. Oceanographic Research Papers* **27** (3-4), 255–264.

[41] HAURY, L. R, KENYON, D. E & BROOKS, J. R 1980 Experimental evaluation of the avoidance reaction of Calanus finmarchicus. *Journal of Plankton Research* **2** (3), 187–202.

[42] YAMAZAKI, H & SQUIRES, K. D 1996 Comparison of oceanic turbulence and copepod swimming. *Marine Ecology Progress Series* **144**, 299–301.

[43] PURCELL, E. M 1977 Life at low Reynolds number. *American Journal of Physics* **45** (1), 3–11.

[44] KIM, S & KARRILA, S. J 2013 *Microhydrodynamics: principles and selected applications*. Courier Corporation.

[45] DURHAM, W. M, CLIMENT, E, BARRY, M, DE LILLO, F, BOFFETTA, G, CENCINI, M & STOCKER, R 2013 Turbulence drives microscale patches of motile phytoplankton. *Nature communications* **4** (1), 1–7.

[46] GUSTAVSSON, K, BERGLUND, F, JONSSON, P & MEHLIG, B 2016 Preferential sampling and small-scale clustering of gyrotactic microswimmers in turbulence. *Physical review letters* **116** (10), 108104.

[47] JEFFERY, G. B 1922 The motion of ellipsoidal particles immersed in a viscous fluid. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character* **102** (715), 161–179.

[48] KESSLER, J. O 1985 Hydrodynamic focusing of motile algal cells. *Nature* **313** (5999), 218–220.

[49] SANTAMARIA, F, DE LILLO, F, CENCINI, M & BOFFETTA, G 2014 Gyrotactic trapping in laminar and turbulent kolmogorov flow. *Physics of Fluids* **26** (11), 111901.

[50] BAGØIEN, E & KIØRBOE, T 2005 Blind dating—mate finding in planktonic copepods. i. tracking the pheromone trail of centropages typicus. *Marine Ecology Progress Series* **300**, 105–115.

[51] TITELMAN, J, VARPE, Ø, ELIASSEN, S & FIKSEN, Ø 2007 Copepod mating: chance or choice? *Journal of Plankton Research* **29** (12), 1023–1030.

[52] FIELDS, D & YEN, J 1997 Implications of the feeding current structure of euchaeta rimana, a carnivorous pelagic copepod, on the spatial orientation of their prey. *Journal of plankton research* **19** (1), 79–95.

[53] PORTER, M. L, STECK, M, RONCALLI, V & LENZ, P. H 2017 Molecular characterization of copepod photoreception. *The Biological Bulletin* **233** (1), 96–110.

[54] BUSKEY, E. J & HARTLINE, D. K 2003 High-speed video analysis of the escape responses of the copepod acartia tonsa to shadows. *The Biological Bulletin* **204** (1), 28–37.

[55] VISSER, A. W 2001 Hydromechanical signals in the plankton. *Marine Ecology Progress Series* **222**, 1–24.

[56] STRICKLER, J. R & BAL, A. K 1973 Setae of the first antennae of the copepod cyclops scutifer (sars): their structure and importance. *Proceedings of the National Academy of Sciences* **70** (9), 2656–2659.

[57] LYTHGOE, J. N 1988 Light and vision in the aquatic environment. In *Sensory biology of aquatic animals*, pp. 57–82. Springer.

[58] YEN, J, LENZ, P. H, GASSIE, D. V & HARTLINE, D. K 1992 Mechanoreception in marine copepods: electrophysiological studies on the first antennae. *Journal of Plankton Research* **14** (4), 495–512.

[59] FIELDS, D. M, SHAEFFER, D & WEISSBURG, M. J 2002 Mechanical and neural responses from the mechanosensory hairs on the antennule of gaussia princeps. *Marine Ecology Progress Series* **227**, 173–186.

[60] HWANG, J.-S & STRICKLER, R 2001 Can copepods differentiate prey from predator hydromechanically? *ZOOLOGICAL STUDIES-TAIPEI-* **40** (1), 1–6.

[61] BUSKEY, E, LENZ, P & HARTLINE, D 2002 Escape behavior of planktonic copepods in response to hydrodynamic disturbances: high speed video analysis. *Marine Ecology Progress Series* **235**, 135–146.

[62] TRITTON, D. J 2012 *Physical fluid dynamics*. Springer Science & Business Media.

[63] TAYLOR, G. I & GREEN, A. E 1937 Mechanism of the production of small eddies from large ones. *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences* **158** (895), 499–521.

[64] WANG, C 1989 Exact solutions of the unsteady navier-stokes equations. *Applied Mechanics Reviews;(United States)* **42** (CONF-8901202-).

[65] GUSTAVSSON, K & MEHLIG, B 2016 Statistical models for spatial patterns of heavy particles in turbulence. *Advances in Physics* **65** (1), 1–57.

[66] GUSTAFSSON, K 2009 *Advective collisions in random flows*. Department of Physics; Institutionen för fysik.

[67] JACOBS, K 2010 *Stochastic processes for physicists: understanding noisy systems*. Cambridge University Press.

[68] GUSTAVSSON, K, JUCHA, J, NASO, A, LÉVÊQUE, E, PUMIR, A & MEHLIG, B 2017 Statistical model for the orientation of nonspherical particles settling in turbulence. *Physical review letters* **119** (25), 254501.

[69] BORGNINO, M, GUSTAVSSON, K, DE LILLO, F, BOFFETTA, G, CENCINI, M & MEHLIG, B 2019 Alignment of nonspherical active particles in chaotic flows. *Physical review letters* **123** (13), 138003.

[70] ISHIHARA, T, KANEDA, Y, YOKOKAWA, M, ITAKURA, K & UNO, A 2007 Small-scale statistics in high-resolution direct numerical simulation of turbulence: Reynolds number dependence of one-point velocity gradient statistics. *Journal of Fluid Mechanics* **592**, 335–366.

[71] WIEDENMAYER, C. P 2009 Plasticity of defensive behavior and fear in early development. *Neuroscience & Biobehavioral Reviews* **33** (3), 432–441.

[72] BELLMAN, R 1957 A Markovian Decision Process. *Indiana University Mathematics Journal* **6** (4), 679–684.

[73] NAJAFI, A & GOLESTANIAN, R 2004 Simple swimmer at low reynolds number: Three linked spheres. *Physical Review E* **69** (6), 062901.

[74] MICHALEC, F.-G, FOUXON, I, SOUISSI, S & HOLZNER, M 2017 Zooplankton can actively adjust their motility to turbulent flow. *Proceedings of the National Academy of Sciences* **114** (52), E11199–E11207.

[75] WATKINS, C. J & DAYAN, P 1992 Q-learning. *Machine learning* **8** (3), 279–292.

[76] REHM, L. P 1977 A self-control model of depression. *Behavior therapy* **8** (5), 787–804.

[77] BELLMAN, R 1957 A markovian decision process. *Journal of mathematics and mechanics* pp. 679–684.

[78] MUIÑOS-LANDIN, S, FISCHER, A, HOLUBEC, V & CICHOS, F 2021 Reinforcement learning with artificial microswimmers. *Science Robotics* **6** (52), eabd9285.

[79] BIFERALE, L, BONACCORSO, F, BUZZICOTTI, M, CLARK DI LEONI, P & GUSTAVSSON, K 2019 Zermelo's problem: Optimal point-to-point navigation in 2d turbulent flows using reinforcement learning. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **29** (10), 103138.

[80] COLABRESE, S, GUSTAVSSON, K, CELANI, A & BIFERALE, L 2018 Smart inertial particles. *Physical Review Fluids* **3** (8), 084301.

[81] KIM, H.-J, YAMADE, T, IWASAKI, K, MARCIAL, H. S & HAGIWARA, A 2019 Phototactic behavior of the marine harpacticoid copepod tigriopus japonicus related to developmental stages under various light conditions. *Journal of Experimental Marine Biology and Ecology* **518**, 151183.

[82] MAIBAM, C, FINK, P, ROMANO, G, BUIA, M. C, BUTERA, E & ZUPO, V 2015 Centropages typicus (crustacea, copepoda) reacts to volatile compounds produced by planktonic algae. *Marine Ecology* **36** (3), 819–834.

[83] TSANG, A. C. H, TONG, P. W, NALLAN, S & PAK, O. S 2020 Self-learning how to swim at low reynolds number. *Physical Review Fluids* **5** (7), 074101.

[84] MATHERON, G & DE MARSILY, G 1980 Is transport in porous media always diffusive? a counterexample. *Water resources research* **16** (5), 901–917.

[85] SCHNEIDER, E & STARK, H 2019 Optimal steering of a smart active particle. *EPL (Europhysics Letters)* **127** (6), 64003.

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

[87] GUNNARSON, P, MANDRALIS, I, NOVATI, G, KOUMOUTSAKOS, P & DABIRI, J. O 2021 Learning efficient navigation in vortical flow fields. *Nature communications* **12** (1), 1–7.

[88] COHEN, J. H & FORWARD JR, R. B 2002 Spectral sensitivity of vertically migrating marine copepods. *The Biological Bulletin* **203** (3), 307–314.

[89] WEISSBURG, M. J, DOALL, M & YEN, J 1998 Following the invisible trail: kinematic analysis of mate-tracking in the copepod temora longicornis. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **353** (1369), 701–712.

[90] CENCINI, M, BOFFETTA, G, BORGNINO, M & DE LILLO, F 2019 Gyrotactic phytoplankton in laminar and turbulent flows: a dynamical systems approach. *The European Physical Journal E* **42** (3), 1–15.

[91] VERMA, S, NOVATI, G & KOUMOUTSAKOS, P 2018 Efficient collective swimming by harnessing vortices through deep reinforcement learning. *Proceedings of the National Academy of Sciences* **115** (23), 5849–5854.

[92] MANDRALIS, I, WEBER, P, NOVATI, G & KOUMOUTSAKOS, P 2021 Learning swimming escape patterns for larval fish under energy constraints. *Physical Review Fluids* **6** (9), 093101.

[93] YEN, J & STRICKLER, J. R 1996 Advertisement and concealment in the plankton: what makes a copepod hydrodynamically conspicuous? *Invertebrate Biology* pp. 191–205.

[94] MNIH, V, KAVUKCUOGLU, K, SILVER, D, RUSU, A. A, VENESS, J, BELLEMARE, M. G, GRAVES, A, RIEDMILLER, M, FIDJELAND, A. K, OSTROVSKI, G *et al.* 2015 Human-level control through deep reinforcement learning. *nature* **518** (7540), 529–533.

Part III
Research papers

Paper A

QIU, J., MOUSAVI, N., GUSTAVSSON, K., XU, C., MEHLIG, B., & ZHAO, L. Navigation of micro-swimmers in steady flow: The importance of symmetries. *Journal of Fluid Mechanics* **932**, A10.

DOI: https://doi.org/10.1017/jfm.2021.978

Paper B

QIU, J., MOUSAVI, N., ZHAO, L., & GUSTAVSSON, K. Active gyrotactic stability of microswimmers using hydromechanical signals. *Phys. Rev. Fluids* **7**, 014311.

DOI:https://doi.org/10.1103/PhysRevFluids.7.014311