International School for Advanced Studies (SISSA)

Optimal Search Strategies in Physics and Biology

Thesis for the Ph. D. in Statistical Physics

Supervisor
Antonio CELANI

Candidate
Alberto PEZZOTTA

IX Cycle — Academic Year 2017/2018

Contents

Acknowledgements

I am deeply grateful to my supervisor, Antonio Celani, for all the support that he gave me throughout my PhD, for all the insightful thoughts and discussions, for his ability of transmitting the passion for research, for his infinite patience and for being a caring guide in life beyond science.

I sincerely thank Matteo, my colleague and friend with whom I have shared excitement, frustration and curiosity about science, as well as much of the good time during these four years of my PhD. Special thanks to Mihir who, together with Matteo, was the best colleague and office mate I could ever hope for.

I want to express my gratitude to all the other members of the *Quantitative Life Sciences* section at ICTP, current and past: Valerio, Erica, Claudio, Andrea, Anjan, Sophie, Edgar, Matteo, Asja, Ryan, Bruno, Clélia. They have made the workplace a nice and happy place to be, beyond being source of inspiring ideas and discussions.

Thanks also to all the current and past members in the *Statistical Physics* sector at SISSA.

Infinite thanks to all the friends who, near or far, made my time in Trieste a pleasant one, with a chat on the phone, in front of a beer, or on a trip in the mountains: Teo, Uopo, Albe, Richi, Gabri, Beppe, Luca, Carlo, Nico, Cri, Claudia.

Special thanks also to all the volunteers of the Italian courses with the *Centro delle Culture*. I thank Serena, in particular, because she has been for me an example of strength and solidarity that I will never forget.

I want to thank my parents, and all my family: without their love and invaluable support, it would have been impossible to undertake and get to the end of this path.

Finally, but above all, I thank Vizhe. Friend, companion of many adventures, and family. She was always supportive and caring, standing next to me all these years, bearing with all my flaws and giving me strength whenever I needed. Whatever achievement of mine, is hers too.

Abstract

In this thesis, navigation and search strategies are investigated from an optimal control and statistical physics perspective.

In Ch. 1, a multi-agent decision making problem, a cooperative search game, is treated in the framework of optimal control. It is shown that, surprisingly, phenomenological equations that describe chemotaxis –including perfect adaptation and fold-change response– emerge as the optimal solution, in a mean-field approximation, to this cooperative task. To our knowledge, such an equivalence was never noticed before, and it provides an interesting functional interpretation of chemotaxis. The content of this Chapter is available in Pezzotta et al. *Phys. Rev. E* **98**, 042401 (2018).

In Ch. 2, the dynamics of a statistical mechanical model (the Conformational Spread model) –that accurately reproduces the dynamics of the flagellar motor switch in *Escherichia coli* (E. coli)– is studied analytically by means of multi-scale techniques (decimation and averaging), providing a cooperative binding model which effectively describe the locked-state time distribution –ultimately determining the run-and-tumble behaviour of E. coli. Studies of the dynamics of this model were previously limited to numerical simulations, and analytical results were achieved only at equilibrium. This work has been published as a research article in Pezzotta, et al. *J. Stat. Mech.* 023402 (2017).

In Ch. 3, we formulate a collective navigation task as an optimal control problem, in which agents have an incentive to align their velocities. A multi-scale analysis (averaging

and homogenization) is used for studying the optimally controlled dynamics in the over-damped limit. The analytical solution of the effective equations at the steady state is given in particular instances of the problem (two agents on a torus).

In Ch. 4, it is shown that the conditioning of Markov processes to lie within a confined region of space can be regarded as an optimal search problem. As a case study, we analysed a jump process conditioned to stay within an infinitely long cylindrical domain, and to go from one end to the other. This example is inspired by the problem of sampling configurations of polymers confined in nanochannels. This work has been published in Adorisio, et al. *J. Stat. Phys* **170**, 79-100 (2017), where more details can be found pertaining to the physics of polymers.

Introduction and Outline

Traditionally, the discipline devoted to the study of animal behaviour is *ethology*, elevated to the rank of science in the XX century by eminent figures like the 1973 Nobel Prize winners Konrad Lorentz, Nikolaas Tinbergen and Karl von Frisch [1]. The main object of ethology is the observation of the behaviour of animals in their natural context. For some decades since its early days, the nature of ethology has been purely descriptive. The influence of *psychophysics* –a discipline born in XIX century with the seminal works by Ernst Weber and later on by Gustav Fechner [2]– began to gain prominence in ethology and contributed to a paradigm shift in the way behaviour is studied. Ethology then started to be shaped into a quantitative discipline, rooted in the physics of perception. The possibility of quantifying sensory stimuli allows to perform constrained experiments in order to pinpoint causal relationships between perception and action.

The works of Burrhus Skinner in the 1930s [3], preceded by those of Ivan Pavlov [4] and Edward Thorndike [5], paved the way towards a theory of behaviour in which memory and rewarding signals are involved in learning, predicting, and hence *acting*, under the influence of sensory stimuli. To this picture, algorithmic structure was given by, notably, the Rescorla–Wagner model, laying the basis for modern *reinforcement learning* theory [6], which nowadays constitutes a research field in its own right.

The central role occupied by physics and computer science is what makes ethology a modern subject. Physics has proved invaluable in uncovering the mechanisms subserving the biology of the cell [7, 8] and the growth of organisms [9]. It may then provide insightful tools in bridging perception and sensing to the actions that ultimately result in the observed behaviour. The predictive power of theories and models stemming from the most traditional fields of physics, together with the rapidly growing field of machine learning, may contribute to unravel the complexity of behaviour.

All the behaviours that involve orienting and moving in the environment can be ascribed to the the general context of *navigation*. Search processes are the behaviours that we will be discussing in this thesis, and that we look at from a physics perspective.

Connecting different levels of description

In 1976, Marr and Poggio [10] proposed a paradigmatic scheme in which three different levels of analysis contribute, in a complementary way, to the understanding of biological systems. Their thesis hinges upon examples coming from the study of the central nervous system, and vision in particular, but extends to all realms of the biological sciences, and therefore behaviour. This scheme is usually referred to as *Marr's levels of analysis*. In this thesis, we would like to fit navigation in all its complexity into this paradigm.

The first and highest level in the scheme of Marr is the *computation* level, dealing with the nature of the problem that the biological agent is trying to solve. In the study of cognition and behaviour, the abstract problem is referred to as the *task* or *goal* of the agent. For instance, bacteria need to climb gradients of nutrients or escape noxious substances. This task is essential for their survival. At the microscale, the ability to respond to chemical stimuli and move under their influence, is called *chemotaxis* [11, 12]. In most situations, organisms face complex environments, in which discerning what is the task that they need to perform is a challenging problem. In the general context of navigation, though, the ultimate goals are often clear: search for food, escape predators, find mates. What is seldom transparent, is the environmental and physical conditions that the agents experience, which are also to be taken into account at this level. Therefore, the computation level is the most difficult to address in a quantitative way, and arguably, because of this, the most neglected.

The second and intermediate level in Marr's scheme is the *algorithmic* level, which pertains to the possible strategies that solve an abstract computational problem. Here sits the behaviour itself, meant as the decision-making rules that result in the observed phenomena. For example, microorganisms integrate chemical signals in time and respond by controlling their motion in ways that differ from species to species [13]. However, one algorithm which is common to many species from bacteria, to insects and mammalian cells, is the fold-change response to sensory stimuli [14, 15] –which encompasses adaptation and Weber–Fechner law. This level, then, relates to the sequence of simple operations that an agent may perform in order to solve a complex task. In a general navigation task, the set of operations include perceiving sensory stimuli, processing them, and interpreting them to make decisions about whether to move and where.

The lowest level of analysis, dealing with the microscopic mechanisms with which the single operations are carried out, is the *implementation* level. Analysis of behaviour, here, requires a detailed knowledge of the system that performs the operations. The biochemical pathway that is responsible for the decision-making in most bacteria, for instance, is very well

known, from the sensing to the motor control [16]. It is obvious that the precise instructions in an algorithm are limited not only by the nature of the problem but also by the biophysical hardware that the organism has at its disposal. Some bacteria perform runs and tumbles, like *E. coli* [17, 18], while others alternate reversals and flicks, like *V. alginolyticus* [19]. Other organisms, like the amoeba *D. discoideum*, move by elongating and contracting their body in the direction of the gradient [20]. Figuring out how the microscopic hardware can operate in order to produce actions, can provide arguments for the biological feasibility of algorithms that solve abstract tasks.

In order to understand behaviour, it is necessary to answer questions at each level. However, there are logical and causal relations between the different levels, and investigation at one level might inspire questions and possibly suggest explanations at another. One way of proceeding –going from a higher to a lower level, or viceversa– may be found more convenient than the other. When studying behaviour, though, the complexity of the biochemical processes is often overwhelming and reconstructing the final goal of an agent from such microscopic details proves to be a daunting task. Already at a mechanistic level of description, it is often exceedingly difficult to predict emerging properties of the many microscopic parts. And even these mechanisms being comprehensively studied, it is often unclear how they fit into a decision-making process.

In some instances, then, it may be more sensible to proceed with a top-down approach, asking first what the nature of the task is. To do so, it is necessary to take into account the ecological scaffold that restricts the sort of sensing and ranges of action of an organism. The physics and the biochemical conditions experienced by an agent are key to understand its behaviour. What are the length and time scales probed by an agent, and how do the laws of physics look like at those scales? What is the sensory information available to an organism, and how does that affect its behaviour? Answering these questions makes it possible to understand what kind of tasks an organism needs to engage in. For instance, life at high and low Reynolds numbers definitely takes very different forms [21]: a tapered body offers the big cetaceans a great advantage in swimming long distances in the ocean, but is not necessarily useful to the microbes in our gut; vision is undoubtedly important for the colourful fish of the coral reefs, but not as much for the darker and smaller-eyed fish living in deep waters.

Searching for optimality

Focused primarily on search and navigation tasks, the main message of this thesis is that it may be possible to *explain* behaviour in terms of optimality. In this picture, the behaviour of an individual emerges as the optimization of an objective (or cost) function, subject to the likely physical limits that it experiences. Looking for optimality principles is particularly appealing from the point of view of evolution, as it can be argued that behaviour is shaped

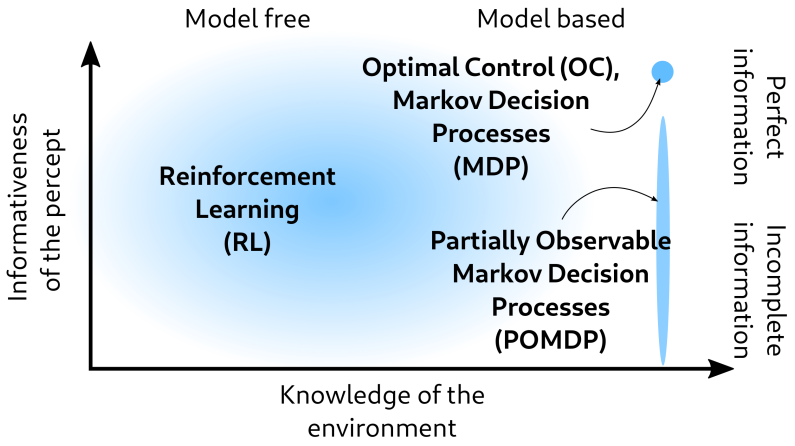


Fig. 1. **Decision-making theory and optimization.** Several optimization techniques can be used in order to find optimal strategies for decision-making tasks. They depend, loosely speaking, on the knowledge of the laws governing the environment and on the amount of information available to the agent about the state of the environment. Special cases are those in which a model of the dynamics is assumed to be known by the agent as well as the full configuration of the system. This is the realm of optimal control and of MDP, where optimal solutions are calculated from the prediction of future outcomes according to a known model of the dynamics. This is the limit dealt with in this thesis. An extension of MDP to situations in which the agent can access only partially the state of the environment is POMDP. All these approaches are model-based, in that they require the complete knowledge of the laws of Nature. Another framework is RL, which does not assume knowledge of either the state of the environment nor the model of its evolution.

the way it is by natural selection.

In this thesis, we formulate navigation and search tasks in precise mathematical terms as control problems by defining utility functions representing the goal of the agents, and then *derive* the strategies that maximize such utility. The approach which spans most of this thesis is *optimal control theory*, a framework which is prominent in the mathematical and engineering sciences [22–25]. Optimal control, as much as the well established theory of stochastic optimization and dynamic programming, which were pioneered by Richard Bellman, among others [26–29], constitutes the basis of other techniques to solve decision-making problems. One of them is *Reinforcement Learning* (RL) [6, 30], based on the paradigm of classical and operant conditioning [3–5]. The scope of RL is much broader and virtually encompasses all possible decision-making problems, from those pertaining to the regime of optimal control –where the agent is assumed to know everything about the state of the world– to those in which partial information about the environment is accessible to the agent (see Fig. 1). Beyond providing solutions to decision-making problems, RL –as it might be obvious from the name– offers the algorithms for the *learning* process. Not only, then, one might hope to draw conclusions about the optimality of the behaviours that one observes

among living things, but also about the computations required to learn from experience. A notable example is the dopamine circuit, which is seen to be the biological implementation of temporal difference learning algorithms [31]. Therefore, RL is appealing for a twofold reason: on one hand it may encompass, at an algorithmic level, the process of optimization carried out by evolution (in the spirit of the so-called evolutionary algorithms [32]); on the other hand, it might offer plausible arguments as to how naive biological agents become experienced in performing tasks essential for their survival.

An RL approach allows to connect all possible levels in Marr's scheme, and provides a new perspective from which to look at behaviour. It is the case of many animals, due to their short life span, to be required to perform some task at the first attempt in an efficient way. A compelling example are male moths that manage to locate females far away using the highly volatile pheromone signals without previous experiences [33]. The ability of behaving so efficiently is shaped by evolutionary and developmental processes; while the moth searches for its mate, then, it has to learn its position from sensory cues. Hence, learning occurs (in different forms) at various time-scales, pertaining to evolution, development, and the task itself. The ability to learn and flexibly adjust one's strategies to diverse situations is a common feature of all organisms. Understanding how learning takes place at different time-scales may give invaluable insights on behaviour itself.

Towards synthetic biology and artificial intelligence

The scheme of Marr's levels of analysis not only offers a paradigm shift in the study of biological systems, but also inspires possible directions towards the development of artificial intelligence systems and synthetic biology. In recent years, the field of machine learning and artificial intelligence has seen unprecedented progress, with contributions from scientists working in very diverse fields. Machine learning techniques have proved to be invaluable tools in the life sciences [34, 35], as well as in the understanding of fundamental problems in particle physics [36] and astrophysics [37]. Such techniques that directly connect with decision making problems are those in the reinforcement learning framework [6]. By incorporating the notion of learning from experience, not only are they particularly suited to understand behaviour and learning in biological systems, but also raise the possibility of designing intelligent machines [38–40].

One purpose of this thesis is to show that optimality, beyond potentially characterizing an organizing principle in biology, can also be of considerable importance as a *design principle* for artificial forms of intelligent systems.

Chemotaxis, an optimal search strategy

This thesis focuses primarily on what the optimum strategies are, rather than how they are learned from experience. However, even in the framework of optimal control it is possible to

address the question of how learning occurs.

In **Chapter 1**, we formulate a cooperative search game as an optimal control problem. The objective of this collective task is that every individual in the group reaches the target quickly, while paying a small price for energy expenditure and limiting the risk of collision with other members. For large collectives (where mean-field provides a reasonable approximation, both for physical [41] and for game-theoretical [42] arguments) the optimal solution turns out to be governed by the same equations that describe the motion of microorganisms directed by chemical signals. A precise dictionary can be compiled with entries from decision-making theory –such as costs for time, energy, collisions, optimal control, attitude towards risk– and their biophysical correlates in chemotaxis –individual motility, collective migration, logarithmic sensing, degradation and consumption rates. In this exact correspondence, chemicals are the media over which individuals share information about the environment. The mathematical structure of the problem, and its equivalence with chemotaxis, suggests biologically plausible ways to dynamically construct the optimal solution, i.e. to *learn* it. On one hand, the result of Chapter 1 may shed light on the *functional* role of chemotaxis in microorganisms, and on the other hand might inspire artificial intelligence algorithms for solving, e.g. traffic problems.

A look into the molecular mechanisms

In one part of this thesis, navigation is looked at also at the implementation level. The case study, which is presented in **Chapter 2**, is the mechanism at the basis of the chemotactic behaviour of *E. coli*. Prokaryotic chemotaxis is based on very similar molecular machinery across different species. A sensing apparatus, detecting molecules of attractants or repellents, triggers a relay signal which eventually controls the motor output. The motor can operate in two states, which are the rotation in the clockwise (CW) or counter-clockwise (CCW) fashion. A flagellum, extending outside the cell, is attached to the motor and produces motion similarly to a helix. The resulting type of motility (run and tumble [17], reverse and flick [19], etc.) depends on the interaction between the flagella, the fluid and the shape of the cell body of the bacterium. However, the motility ultimately depends on the frequency with which the motors change their rotational states, from CW to CCW or viceversa. This strategy results in a more persistent or more erratic motion that over long times leads to the achievement of the final goal: climbing gradients of food or descending gradients of repellents.

Bacterial chemotaxis is very well understood at the molecular level, and an analysis of these microscopic mechanisms may complete the quantitative explanation of the behaviours that we observe. A common feature of all the cellular processes is the presence of widely separated time scales. Conformational changes of proteins involved in signalling occur on μs – ms, folding of proteins can take up to few seconds while the transcription and translation processes may last for minutes [43]. This allows one to operate controlled approximations

in order to work out the effective dynamics at the time scales relevant for the behaviour. The framework of multi-scale analysis of Markov processes and dynamical systems is well established in the mathematical and physical literature [44]. Decimation, averaging and homogenization techniques lay on a solid mathematical ground [45] and find innumerable applications, from cell biology, to evolution, to the thermodynamics of mesoscopic systems [46].

Together with the theory of optimal control, the multi-scale analysis of Markov processes constitutes the main technical ingredient of this thesis. In Chapter 2 these techniques allow to map analytically the microscopic mechanisms into the *actions* –the average frequencies of the CW/CCW switch. In Chapter 3 the same analytical tools are used in the study of active particles which engage in a collective navigation task, formulated as an optimal control problem.

An optimality perspective on collective behaviour

While many phenomenological descriptions proved useful in accounting for many collective phenomena observed in Nature [47, 48], an explanation for their functional role is usually not clear. To distil optimality principles in the behaviour of large collectives, i.e. understand why a group provides benefits to the single individuals, is a fascinating issue to look at in the context of decision-making. It is apparent that the group offers protection from predators to its members [49]. Sometimes advantages are of physical nature [50]. In some cases it can be argued that the collective is more accurate in performing estimates, an example of the "wisdom of the crowds" which was noted by Sir Francis Galton [51]. Nicolas de Condorcet had also realized, based on simple probabilistic arguments, that the group enhances the ability of assessing the correctness of some statement, as the members of a jury do in Court [52]. These arguments suggest that some animals may prefer navigating in close contact in order to share information about the environment more efficiently, and hence take decisions more reliably [53].

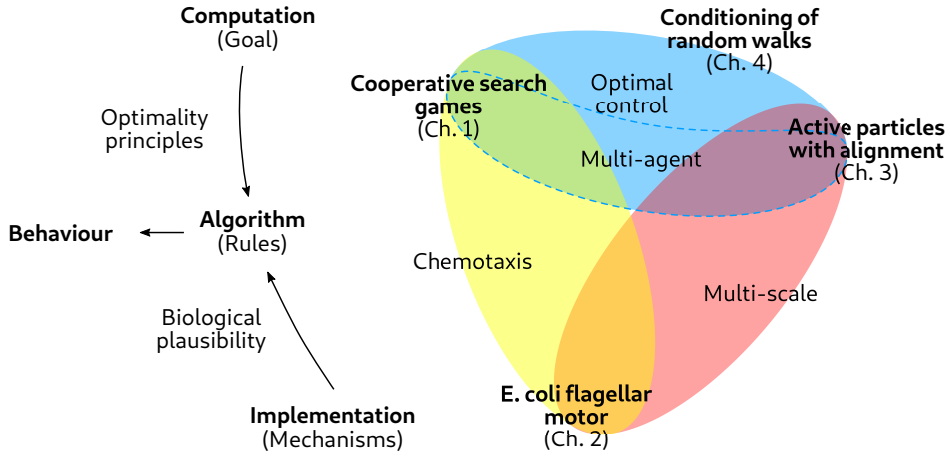
In **Chapter 3**, we look at collective behaviours from the perspective of optimality, with the same approach as in Chapter 1. We define a collective navigation task, in which the agents pay a cost associated to the amount of force that they apply against the drag and random forces of the environment as well as a cost for deviating from the direction of motion of their neighbours. We derive the exact equations for the optimal control, and the resulting dynamics is then analysed in the overdamped limit by means of averaging and homogenization techniques. The resulting driven-diffusive process features a non-trivial relation between the force and the noise term. The solution of the stationary state distribution for the simple scenario of two agents on a torus exhibits an aggregation effect, and gives analytical support to the findings from numerical approaches of similar models.

Conditioning as an optimal search

Optimal control is useful in the physics of out-of-equilibrium systems, from the thermodynamics of mesoscopic systems [54, 55] to quantum-state preparation [56]. In **Chapter 4** we show how an optimal control approach, such as the conditioning of Markov processes [57–59]. It is often a challenging problem to sample trajectories of Markov processes which have to satisfy some hard constraints, e.g. being confined in a bounded region of space. A straightforward approach would be to generate trajectories according to some transition probability and then retain only those satisfying the constraint. This strategy turns out to be very costly from the computational standpoint, as the constraint might select only rare trajectories [60]. Another approach, pioneered by Doob [61], is to introduce an auxiliary process which implements the constraints in the dynamics, so that it reproduces all and only those trajectories that would be selected by rejection.

In the last part of this thesis, Chapter 4, we regard the auxiliary (Doob-transformed) process as an optimally controlled search process. Our case study is the conditioning of a discrete-time process –which in free space performs jumps with an isotropic distribution– to lie within a bounded region of space (a cylinder), which provides a simple model for polymers under confinement [62, 63]. Under confinement, the walker biases its motion along the cylinder, avoiding jumps outside. We give an exact formula for this bias, which can be regarded as a specific form of Gibbs exponential reweighting. We also study numerically the geometric properties of the paths sampled with the optimally controlled transition probabilities, and also offer some analytical quantification.

The content and the logic of this thesis is summarized in the following scheme.



1

Cooperative Search Games

Individuals in a group often have to face complex situations which require concerted actions [64–66]. Among the various collective intelligence problems, here we focus our attention on cooperative navigation tasks, where all agents share the common goal of locating a target and reaching it in the most efficient way. For instance, a crowd may need to quickly escape from an enclosed space while averting stampedes. Similarly, birds in a flock or fish in a school try to reduce exposure to predators and avoid harmful collisions. In addition, individuals are also confronted with the limits posed by the energetic costs of locomotion. The very same kind of objectives and challenges lie at the heart of multi-agent autonomous robotics[67–70].

Intelligent agents should aim at acting optimally in these contexts. That is, they should cooperate in order to minimize some cost function that compounds the many objectives at play: short time for completing the task, small energy spent in the process, and reduced damage by collisions. What is the optimal strategy? How universal is it across environments and agents? How is information shared by agents? How is it translated into actions? Can the optimal behavior be reliably and quickly learned by agents facing unknown environments? Is the optimal strategy actually employed by living organisms?

In this Chapter we answer these questions by formulating the cooperative search game in terms of stochastic optimal control. We first discuss how optimal solutions can be mapped into quantum states of an interacting many-body system. Unfortunately, the exact solution of this quantum problem is very difficult even in simple geometries. However, in the limit

of very large collectives, a mean-field theory yields very simple and well-known effective equations.

Indeed, the mean-field equations for optimal cooperative search turn out to be identical to a long-known phenomenological model of chemotaxis, the celebrated Patlak–Keller–Segel model [71, 72] with Weber–Fechner logarithmic response (see e.g. [14] for a general discussion about fold-change detection). The chemical attractant can therefore be interpreted as the medium that agents use to share information about the location of the target and the density of individuals in the group. The biophysical processes by which the concentration is altered, namely production, consumption and degradation, correspond to the actions of writing information on the memory, erasing and forgetting, respectively. We show that there is a dictionary that maps concepts from decision-making theory – strategies, desirability, costs for control and for collisions, cost per time elapsed, attitude toward risk – into precise physico-chemical and biological correlates – concentration levels, diffusion coefficients, degradation and consumption rates, chemotactic coefficients (see Table 1.1 for the detailed analogy).

Decision making		**Chemotaxis**	
ζ	Desirability	s	Chemoattractant concentration
D	Uncontrolled dynamics	D	Random motility
u^*	Optimal control	$\chi \nabla \log s$	Chemotactic drift with logarithmic sensing
γ	Weight for the cost of control	$\chi = \dfrac{2D}{(1-2D\alpha\gamma)}$	Chemotactic coefficient
α	Risk sensitivity		
q	Time cost rate	$D_s = D/\epsilon$	Diffusion coefficient of chemoattractant
g	Collision cost rate	$k = q\,\dfrac{1-2D\alpha\gamma}{2D\gamma\epsilon}$	Degradation rate of chemoattractant
$1/\epsilon$	Learning rate	$\beta = g\,\dfrac{1-2D\alpha\gamma}{2D\gamma\epsilon}$	Consumption rate of chemoattractant per cell
Eq. (1.73)	Hamilton-Jacobi-Bellman equation	Eq. (1.75)	Patlak-Keller-Segel equations

Tab. 1.1. **Spoiler for the impatient reader: a dictionary between optimal cooperative search and chemotaxis.** The table illustrates the correspondence between quantities in mean-field optimal control and their counterparts in chemotaxis. See Sec. 1.3.

1.1 Optimal control equations

Let us consider a group of N agents in d dimensions following the stochastic dynamics [*]

$$\frac{d\bar{X}^t}{dt} = \bar{u}(\bar{X}^t) + \sqrt{2D}\,\bar{\eta}(t)\ , \tag{1.1}$$

where $\bar{\eta}$ is a Gaussian white noise, with $\langle \eta_{i,\alpha}(t)\rangle = 0$ and $\langle \eta_{i,\alpha}(t)\,\eta_{j,\beta}(t')\rangle = \delta_{\alpha,\beta}\delta_{i,j}\delta(t-t')$. The region of space where the agents move contains some absorbing states which we call *target*. The uncontrolled motility is D, which we choose to be constant. Notice that in this case there is no ambiguity about the regularization, and Ito or Stratonovich conventions are equivalent. The deterministic part in Eq. (1.1), $\bar{u}$, is the *control*, which is in the hands of the agents. In general, the control depends on the positions of all the agents.

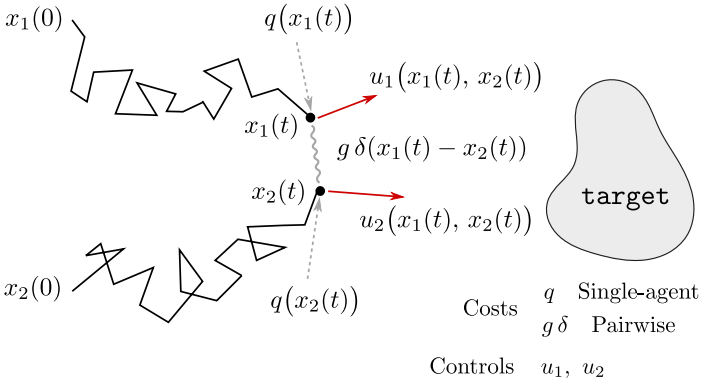


Fig. 1.1. **Scheme of the cooperative search game for $N = 2$ agents.** The agents are driven by the controls u_1 and u_2, which depend on both their positions x_1 and x_2. They pay individual costs associated with control and time expenditure, respectively with a rate $\gamma u^2/2$ and q (possibly position-dependent), and a pairwise cost associated with the interaction – collision – with a rate proportional to a Dirac-δ function. Optimal controls minimize the average cost (or an exponential average in the risk-sensitive case).

At every instant, agent i pays a cost per unit time

$$c_i = \underbrace{q(X_i)}_{\text{time}} + \underbrace{\frac{\gamma}{2}u_i^2}_{\text{energy}} + \underbrace{\frac{g}{2}\sum_{j\neq i}\delta(X_i - X_j)}_{\text{collisions}}\ . \tag{1.2}$$

The cost features three contributions. The first one is the penalty for the time spent before reaching the target. We denote the associated cost per unit time as q, which may in general

[*]Symbols expressed with a *bar* indicate N-tuples whose index corresponds to the label of the agent; e.g. $\bar{x} \equiv \{x_1 \dots x_N\}$. We will use the notation $x_{i,\alpha}$ to indicate the spatial component α of the i-th agent position, and similarly for any other d-dimensional variable related to each individual agent. As concerns the position variables, we denote the random variables with a capital letter labelled by a superscript indicating time, $\bar{X}^t$, while with the lower case we indicate the specific realization of the variable, $\bar{x}$.

depend on the spatial location to account for spatial inhomogeneities. The second one is the cost of control, that we take as $\gamma u^2/2$. This is reminiscent of the power dissipation due to motion in a viscous medium, but can also be interpreted as the Kullback-Leibler (KL) divergence [73] from a random strategy in the decision-making context [74]. Finally, the last term arises from collisions. This combination of factors embodies the trade-offs between different costs that lead to nontrivial solutions of the optimization problem: for instance, a fast search and a low collision risk cannot be achieved without a consequent expenditure in control cost. Every agent quits the game as soon as it reaches the target, and the cost that it pays does not increase thereafter.

The *goal* of the agents is to find the set of controls $\bar{u}$ for which the cost functional

$$C = \sum_{i=1}^{N} \int_0^{T_i} dt \left(\frac{\gamma}{2} u_i(\bar{X}^t, t)^2 + h_i(\bar{X}^t) \right) \tag{1.3}$$

is minimum, where

$$h_i(\bar{X}^t) = q(X_i^t) + \frac{g}{2} \sum_{j \neq i} \delta\big(X_i^t - X_j^t\big) \ ,$$

$q > 0$ and $g > 0$, $\gamma > 0$ and T_i is the exit time (first passage at the target) for the i-th agent. The upper extreme of integration in time indicates that an agent stops contributing to the total cost as soon as it reaches the target. The functional C is itself a random variable, and we shall clarify later what we mean by its minimization. In the next sections we shall derive the equations for optimal controllers that minimize different statistics of the cost functional (1.3), referred to as *risk-neutral* and *risk-sensitive* controls. The first case is simpler and will be described in detail in Sec. 1.1.1; the second, which is more general, will be worked out in Sec. 1.1.2.

It is worth commenting on the choice of the cost for control. As already mentioned, the quadratic form in the drift stems from the definition of the KL divergence from the uncontrolled path measure $\mathbb{P}_0$ to the controlled one $\mathbb{P}_u$. It is therefore the excess information (relative entropy) that one needs to specify in order to describe the statistics of the controlled paths given the knowledge of the statistics of the uncontrolled one. This is very interesting because of its information-theoretical origin, and would deserve attention by itself. Indeed, the KL cost is one of the several examples of entropy-based regularization schemes which are the subject matter of very recent literature on control theory, optimization and notably deep-learning [75, 76]. Among the possible entropy regularization schemes, the KL cost is particularly handy in that it allows to cast the optimal control equations into a linear form, as we will show in the following.

A comment about the optimization setup is also necessary. For the search game discussed here, we are interested in a terminal-state setup, in which the cost C is accumulated until all of the agents are absorbed at the target. In this setup there is no hard constraint on the

arrival times, so the upper extreme of integration in time in Eq. (1.3) extends to infinity. In this case, if the costs do not depend explicitly on time, the optimal control equations are time independent. Another case is the finite-horizon setup, in which the objective of the optimization is a functional of the cost accumulated over the fixed interval of time $[0, T]$, denoted C_0^T,

$$C_0^T = \sum_{i=1}^{N} \int_0^{\min\{T,T_i\}} dt \left(\frac{\gamma}{2} u_i(\bar{X}^t, t)^2 + h_i(\bar{X}^t) \right) , \tag{1.4}$$

and the agents are conditioned to reach the target within such time window. In this case the problem is generally time-dependent. One can recover the terminal-state setup in the limit $T \to \infty$ (i.e. no time constraint), so that we can write $C = \lim_{T\to\infty} C_0^T$. Hence, for the sake of generality, we shall derive the optimal control equations in the finite-horizon setup, and we will implicitly set $T \to \infty$ when specializing the results to the search game.

1.1.1 Risk-neutral case

In this section we describe the derivation of the optimal control equations for the risk-neutral case. A risk-neutral controller is defined as the one that minimizes the mean cost, $\langle C \rangle$, where C is the functional of the trajectories defined in Eq. (1.3):

$$\begin{aligned} \mathcal{F}[\bar{u}] &= \left\langle C_0^T \right\rangle \\ &= \int_0^T dt \int d^N x \sum_i \left(\frac{\gamma}{2} u_i(\bar{x}, t)^2 + h_i(\bar{x}) \right) P(\bar{x}, t) , \end{aligned} \tag{1.5}$$

where P is the N-agent probability density distribution, which satisfies the Fokker–Planck equation associated with Eq. (1.1). It is convenient to couch the minimization problem by including the dynamics as a constraint and minimize the auxiliary functional

$$\mathcal{L}[P,\, \bar{u},\, \Phi] = \mathcal{F} + \int_0^T dt \int d^N x\, \Phi(\bar{x}, t) \left[\partial_t P(\bar{x}, t) + \sum_i \nabla_i \cdot \Big(u_i(\bar{x}, t) P(\bar{x}, t) - D \nabla_i P(\bar{x}, t) \Big) \right] , \tag{1.6}$$

where $\Phi(\bar{x}, t)$ is a Lagrange multiplier. This is an application of the so-called Pontryagin minimum principle [22]. The condition of null variation of $\mathcal{L}$ with respect to Φ trivially yields the Fokker–Planck equation for the N-particle density P. The variation of $\mathcal{L}$ with respect to u_i, at the saddle point,

$$\left. \frac{\delta \mathcal{L}}{\delta u_i} \right|_* = \big(\gamma\, u_i^* - \nabla_i \Phi \big) P = 0 , \tag{1.7}$$

gives the optimal control

$$u_i^*(\bar{x}, t) = \gamma^{-1}\, \nabla_i \Phi(\bar{x}, t) . \tag{1.8}$$

Finally, the variation with respect to P gives

$$\left.\frac{\delta\mathcal{L}}{\delta P}\right|_* = -\partial_t\Phi + \sum_i \left(\frac{\gamma}{2}(u_i^*)^2 - u_i^*\cdot\nabla_i\Phi - D\nabla_i^2\Phi + h_i\right) = 0 \ , \tag{1.9}$$

which, together with Eq. (1.8), yields the *Hamilton–Jacobi–Bellman* (HJB) equation:

$$\partial_t\Phi + \frac{1}{2\gamma}\sum_i \left(\nabla_i\Phi\right)^2 + D\sum_i \nabla_i^2\Phi = \sum_i h_i \ . \tag{1.10}$$

The function $\Phi(\bar{x},t)$ has a very clear interpretation in the context of decision making. It is the optimal *value function* at time t and in the state $\bar{x}$, up to an additive constant: this is (minus) the expected cost-to-go under the optimal control $\bar{u}^*$ when the system is conditioned to be in state $\bar{x}$ at time t, namely

$$\Phi(\bar{x},t) \equiv -\left\langle C_t^T \middle| \bar{X}^t = \bar{x}\right\rangle_* = -\int_t^T dt' \int d^N x' \sum_i \left(\frac{\gamma}{2}u_i^*(\bar{x}',t')^2 + h_i(\bar{x}')\right) P(\bar{x}',t') \ , \tag{1.11}$$

where P satisfies the Fokker–Planck equation with control $\bar{u}^*$ and has, as initial condition,

$$P(\bar{x}',t) = \delta^N(\bar{x}' - \bar{x}) \ .$$

Indeed, it can be directly verified that the r.h.s. of Eq. (1.11) satisfies the saddle-point equation (1.9).

The HJB equation, Eq. (1.10), is non-linear. However, it is possible express it in terms of a function Z, related to Φ through the so-called Hopf–Cole transformation

$$\Phi = 2D\gamma \log Z \ , \tag{1.12}$$

for which it has a linear form

$$\partial_t Z + D\sum_i \nabla_i^2 Z = \frac{1}{2D\gamma}\sum_i h_i\, Z \ . \tag{1.13}$$

It ensues that the optimal control of the agent i is

$$u_i^* = 2D\,\nabla_i \log Z \ . \tag{1.14}$$

The function Z is known in decision-making theory as the *desirability function* and its interpretation in terms of the optimal average cost is straightforward from Eq. (1.11):

$$Z(\bar{x},t) = e^{-\langle C_t^T | \bar{X}^t = \bar{x}\rangle_*/2D\gamma} \ . \tag{1.15}$$

It is then clear that the optimal control biases the motion of the agents towards configurations with lower expected cost. Eq. (1.13) has to be supplemented by appropriate boundary

conditions, that are discussed in the next subsection. Notice that the optimal controls for the individual particles generally depends on the positions of all the N agents, because of the interaction term. It is interesting, and perhaps surprising, that Eq. (1.13) is equivalent to the imaginary time Schrödinger equation of a quantum-mechanical many-body system of identical particles [77, 78]. In the infinite-horizon limit that we are interested in for the cooperative search task defined at the beginning, i.e. a time-independent optimal control problem, Eq. (1.13) becomes a stationary, zero-energy, Schrödinger equation. To the best of our knowledge, an exact solution that satisfies the appropriate boundary conditions is not known for a generic N, even for simple geometries. Moreover, a numerical approach appears to be a daunting task already for three agents in a two-dimensional domain. Approximation schemes are therefore very valuable in order to proceed further.

Boundary conditions

Let us assume that all the agents are inside the domain except one, which we choose to be agent N without loss of generality, which sits at the target: $X_N^0 = x_N \in \texttt{target}$. Since the agent on the target is effectively out of the game, the average cost-to-go function is accounted for by the agents in the internal part of the domain. Hence the desirability is a function of those agents only [†]

$$Z^{(N)}(x_1 \dots x_{N-1}, x_N)\big|_{x_N \in \texttt{target}} = Z^{(N-1)}(x_1 \dots x_{N-1}) \ . \tag{1.16}$$

More generally, if the agents $1 \dots i$ (up to relabeling) are not yet at the target, while the others have already reached it, one has

$$Z^{(N)}(x_1 \dots x_i, x_{i+1} \dots x_N)\big|_{x_{i+1} \dots x_N \in \texttt{target}} = Z^{(i)}(x_1 \dots x_i) \ , \tag{1.17}$$

and for the problem with only one agent one has to impose

$$Z^{(1)}(x)|_{x \in \texttt{target}} = 1 \ . \tag{1.18}$$

If there are boundaries forbidden to the agents, one must set $Z = 0$ if any of the agents lies there (the agents receive an infinitely large penalty whenever they touch those boundaries). If reflecting boundaries are present, the Neumann boundary condition (null orthogonal gradient) is required with respect to the coordinates of the agents close to them:

$$\hat{n}(x_i) \cdot \nabla_i Z^{(N)}(x_1 \dots x_N)\big|_{x_i \in \texttt{ refl. boundary}} = 0 \ , \tag{1.19}$$

where $\hat{n}(x_i)$ is the unit vector normal to the boundary at the position of agent i.

[†]For the sole purpose of illustrating the boundary conditions defining the many-particle optimal control problem, we introduce the notation $Z^{(n)}$ to indicate the desirability function Z for the problem with n agents (when Z is a function of n spatial variables).

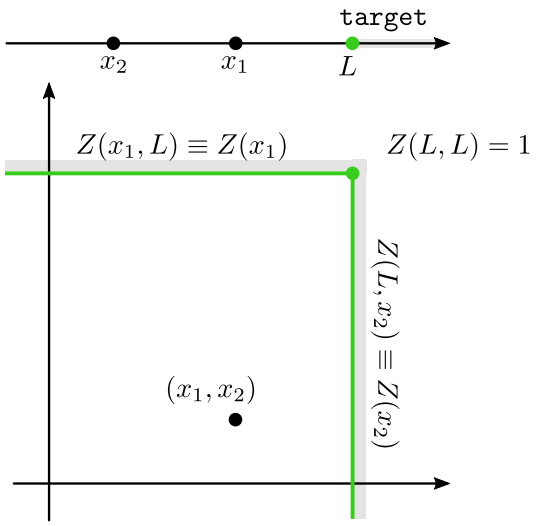


Fig. 1.2. **Boundary conditions for 2 agents in 1D.** Two agents on a line can be represented as a process in a plane. If the target is at position L, then the two-dimensional domain is limited by a square boundary (depicted in green). Such boundary is an absorbing state, and once reached, the two-dimensional process is then restricted to it. The task ends when the corner (green dot) is reached (both agents are at the target), where the desirability is 1. The boundary condition for Z on the semi-line $\{x_2 = L\}$, is given by the 1-agent desirability $Z(x_1)$; analogously, on $\{x_1 = L\}$ is $Z(x_2)$.

Eqs. (1.16) and (1.17) indicate that the boundary conditions are provided by the solution of the problem with fewer agents (the simple case of two particles in a one-dimensional domain is shown in Fig. 1.2). It ensues that in order to solve the cooperative search problem with N agents, one needs to solve a hierarchy of N equations. This recursive structure makes it seemingly impossible to solve analytically the optimal control equations even in simple geometries, and approximation schemes are most valuable. We will see in Sec. 1.2 and that a reasonable approximation that has deep roots in both statistical physics and game theory can yield equations which are amenable to analytical calculations and have interesting interpretations.

1.1.2 Risk-sensitive case

In a more general situation, it might be of interest to minimize a different statistics of the cost. For example, one may wish to limit the effect of improbable but very expensive events by means of a risk-averse strategy. In this section we present the derivation of the optimal control equation for a risk-sensitive controller, of which the risk-neutral one represents nothing but a special case.

A convenient way of incorporating the notion of risk in decision making is to introduce a parameter α which exponentially weighs the fluctuations of the cost [79, 80]. In this setting

the objective functional to be minimized is

$$\mathcal{F}_\alpha = \frac{1}{\alpha} \log \left\langle e^{\alpha C_0^T} \right\rangle = \frac{1}{\alpha} \log \mathcal{G}_\alpha \ , \tag{1.20}$$

where C_0^T is defined in Eq. (1.3). This choice ensures the invariance of the optimal control under a global offset of the costs. Moreover, as presented in [80], we will show that under this choice the optimal control equations in the risk-sensitive case are linearly solvable and are equivalent to the risk-neutral ones, up to a change in the parameters. In this section we only consider $\alpha > 0$, so the problem is equivalent to the minimization of the functional $\mathcal{G}_\alpha = \exp \alpha \mathcal{F}_\alpha$. It is easy to generalize the derivation for $\alpha < 0$.

The parameter α defines the *risk sensitivity* of the optimal control problem: one recognizes that in the limit $\alpha \to 0$, $\mathcal{F}_\alpha \to \mathcal{F} = \left\langle C_0^T \right\rangle$, and the control problem is exactly the risk-neutral one, discussed above; if $\alpha > 0$ the optimal solution is the one that reduces the most the fluctuations towards high values of the cost, and as a consequence corresponds to a *risk-averse* strategy; finally, $\alpha < 0$ corresponds to an optimization in which more weight is given to the values of C_0^T which are smaller than average, therefore leading to *risk-seeking* strategies. The limits $\alpha \to \infty$ and $\alpha \to -\infty$ are referred to as *min-max* and *min-min* optimization problems, respectively. We shall see that in the cooperative search example treated here, α is bound from above, i.e. no control exists which can reduce fluctuations of the cost more than a certain amount which is set by a critical value $\alpha_{\max}$.

We apply the Pontryagin principle to the minimization of the functional $\mathcal{G}_\alpha$. The definition of the cost functional C_0^t, Eq. (1.3), can be put in differential form as

$$dC_0^t = \sum_i \left(\frac{\gamma}{2} u_i(\bar{X}^t, t)^2 + h_i(\bar{X}^t) \right) dt \equiv c(\bar{X}^t)\, dt \ ,$$

and interpreted as an auxiliary process, whose evolution is determined by the dynamics of $\bar{X}$. The Fokker–Planck equation associated to this system of equations, which describes the evolution of the probability density function $p(\bar{x}, C, t)$ for the process $\{\bar{X}^t, C_0^t\}$, is

$$\partial_t P + \partial_C \big(C\, P \big) + \sum_i \nabla_i \big(u_i\, P - D \nabla_i P \big) = 0 \tag{1.21}$$

The functional $\mathcal{G}_\alpha = \exp \alpha \mathcal{F}_\alpha$, is the average of $\exp \alpha C_0^t$ over all trajectories which have arrived at $\bar{x}$ at time t. It can be expressed as a linear functional of P,

$$\mathcal{G}_\alpha = \int d^N x\, dC\, P(\bar{x}, C, T)\, e^{\alpha C} \equiv \int d^N x\, G_\alpha(\bar{x}, T) \ . \tag{1.22}$$

The function $G_\alpha(\bar{x}, t)$ is the density of agents in the configuration $\bar{x}$ at time t, reweighed by an exponential function of the cost paid along their trajectories up to time t. The exponential reweighing is similar in spirit to the multiplication by a Boltzmann factor. If one interprets

the risk-neutral cost functional $\mathcal{F}$ as an *energy* function, then one can think of the objective function $\mathcal{F}_\alpha$ as a *free-energy*, where α plays the role of an inverse temperature. By multiplying Eq. (1.21) by $\exp \alpha c$ and integrating over C, one recovers the forward Feynman–Kac equation for G_α:

$$\partial_t G_\alpha + \sum_i \nabla_i \cdot \Big(u_i G_\alpha - D \nabla_i G_\alpha \Big) = \alpha \, c \, G_\alpha \ . \tag{1.23}$$

The Pontryagin principle is then applied to the minimization of the functional $\mathcal{F}_\alpha$ subject to the constraint imposed by Eq. (1.23). This is equivalent to perform the unconstrained minimization of the Lagrange functional

$$\begin{aligned} \mathcal{L}_\alpha = \underbrace{\int d^N x \, G_\alpha(\bar{x}, T)}_{\mathcal{G}_\alpha} + \int_0^T dt \int d^N x \, \Psi_\alpha(\bar{x}, t) \Big[\partial_t G_\alpha \\ + \sum_i \nabla_i \cdot \Big(u_i G_\alpha - D \nabla_i G_\alpha \Big) - \alpha \sum_i \Big(\frac{\gamma}{2} \, u_i^2 + h_i \Big) G_\alpha \Big] \ . \end{aligned} \tag{1.24}$$

At the saddle point, variation with respect to the control u_i yields

$$\left. \frac{\delta \mathcal{L}_\alpha}{\delta u_i} \right|_* = -G_\alpha \Big(\nabla_i \Psi_\alpha + \alpha \gamma \, u_i^* \Psi_\alpha \Big) = 0 \ , \tag{1.25}$$

so the optimal control is

$$u_i^* = -\frac{1}{\gamma \alpha} \nabla_i \log \Psi_\alpha \ . \tag{1.26}$$

The variation with respect to G_α gives

$$\left. \frac{\partial \mathcal{L}_\alpha}{\delta G_\alpha} \right|_* = \delta(t - T) - \partial_t \Psi_\alpha - \sum_i u_i^* \cdot \nabla_i \Psi_\alpha - D \sum_i \nabla_i^2 \Psi_\alpha - \alpha \sum_i \Big(\frac{\gamma}{2} \, u_i^{*2} + h_i \Big) \Psi_\alpha = 0 \ , \tag{1.27}$$

which is the backward Feynman–Kac equation for the function Ψ_α, which can be interpreted as

$$\Psi_\alpha(\bar{x}, t) \equiv \big\langle e^{\alpha C_t^T} \big| \bar{X}^t = \bar{x} \big\rangle \ ; \tag{1.28}$$

the δ-function in time sets the condition at the final time, if a finite-horizon problem is considered; in the terminal-state setup, $T \to \infty$, and the final-time condition translates into a boundary condition at the terminal states. Note that $\Phi_\alpha \equiv -\alpha^{-1} \log \Psi_\alpha$, plays the role of the value, in that $u_i^* = \gamma^{-1} \nabla_i \Phi_\alpha$. Indeed, in the limit $\alpha \to 0$, Φ_α reduces to the (risk-neutral) value function Φ [‡]. We therefore identify Φ_α with the *risk-sensitive value function* [79, 81].

[‡] Ψ_α, as defined in Eq. (1.28), is the moment generating function for the statistics of the cost C_t^T conditioned to $\bar{X}^t = \bar{x}$. When $\alpha \to 0$,

$$\Phi_\alpha = -\alpha^{-1} \log \Psi_\alpha \to -\left. \frac{\partial \Psi_\alpha}{\partial \alpha} \right|_{\alpha = 0} = -\langle C_t^T | \bar{X}^t = \bar{x} \rangle \equiv \Phi \ .$$

The HJB equation for the risk-sensitive value function then is

$$\partial_t \Phi_\alpha = -\underbrace{\Big(\frac{1}{2\gamma} - D\alpha\Big)}_{\equiv 1/2\tilde{\gamma}} (\nabla_i \Phi_\alpha)^2 - D\sum_i \nabla_i^2 \Phi_\alpha + \sum_i h_i \ ; \tag{1.29}$$

this equation has the same form as Eq. (1.10), where the parameter γ is replaced by $\tilde{\gamma} \equiv \gamma/(1-2D\alpha\gamma)$. In the same way as for the risk-neutral case, a linear version of the HJB equation is obtained through the Hopf–Cole transformation

$$Z_\alpha = \exp\left(\Phi_\alpha / 2D\tilde{\gamma}\right) \tag{1.30}$$

to obtain

$$\partial_t Z_\alpha + D\sum_i \nabla_i^2 Z_\alpha = \frac{1}{2D\tilde{\gamma}} \sum_i h_i \, Z_\alpha \ . \tag{1.31}$$

The optimal control is hence obtained from Z_α as

$$u_i^* = \frac{2D\tilde{\gamma}}{\gamma} \nabla_i \log Z_\alpha \ . \tag{1.32}$$

It is straightforward that for $\alpha = 0$, the optimal control equations (1.31) and (1.32) exactly reduce to the risk-neutral ones, respectively (1.13) and (1.14). As for the boundary conditions for Eq. (1.31), the same considerations made for the risk-neutral case hold here: Eqs. (1.16) and (1.17) are valid also in this case, with Z being replaced by Z_α.

Exact solution for the 1D non-interacting case

In this section we show the exact analytic calculation for the non-interacting search in one dimension. The results also provide an approximation to the solution for the multi-dimensional case at large distances from the target.

A single agent is initially at x on the real axis and the target is at the origin; the generating function of the cost under the control u, $\tilde{G}_s(x) = \langle \exp(-sC_0^T) | X^0 = x \rangle$, satisfies the (stationary) Feynman–Kac equation

$$u\,\tilde{G}_s' + D\,\tilde{G}_s'' = s\Big(\frac{\gamma}{2}u^2 + q\Big)\,\tilde{G}_s \ , \tag{1.33}$$

with boundary conditions $\tilde{G}_s(0) = 1$ and $\tilde{G}_s(\pm\infty) = 0$ if $s > 0$ and $\tilde{G}_s(\pm\infty) = +\infty$ if $s < 0$. Assuming that the control is constant and pointing toward the origin [§], $u = -\mathrm{sign}(x)\,U$, with U positive, one can easily find that Eq. (1.33) is solved by

$$\tilde{G}_s = \exp\Big\{\beta|x|\big(1 - \sqrt{1+\Gamma s}\big)\Big\} \ , \tag{1.34}$$

[§] We know from the exact solution that the optimal control in one dimension is constant: this follows from the fact that the solution of the HJB equation $D\nabla^2 Z_\alpha = q/(2D\tilde{\gamma})\, Z_\alpha$ with the boundary conditions specified above is solved by $Z_\alpha = \exp\{-[q/(2D^2\tilde{\gamma})]^{1/2}|x|\}$, which produces $u^* = 2D\tilde{\gamma}/\gamma\, \nabla \log Z_\alpha = -\mathrm{sign}(x)(2q\tilde{\gamma})^{1/2}/\gamma$, whose amplitude is independent of the coordinate x.

where $\beta = U/(2D)$ and $\Gamma = 2D(\gamma + 2q/U^2)$. We can indeed check that the optimal control U^* obtained from the one-dimensional HJB equation is the one which minimizes $\mathcal{F}_\alpha|_{X^0=x} = \alpha^{-1}\log\tilde{G}_{-\alpha}$:

$$\frac{\partial}{\partial U}\mathcal{F}_\alpha|_{X^0=x} = \frac{\partial}{\partial U}\left\{\frac{U|x|}{2D\alpha}\left(1-\sqrt{1-2D\alpha\Big(\gamma+\frac{2q}{U^2}\Big)}\right)\right\} = 0\ , \tag{1.35}$$

solved by

$$U^* = \sqrt{\frac{2q}{\gamma(1-2D\alpha\gamma)}}\ . \tag{1.36}$$

The probability distribution of the cost C_0^T under the control U is found by applying the inverse Laplace transform to Eq. (1.34),

$$\begin{aligned} p(c|X^0=x) = \frac{1}{2\pi\mathrm{i}}\int_{-\mathrm{i}\infty}^{+\mathrm{i}\infty} ds\,\tilde{G}_s &= \frac{e^{\beta|x|-c/\Gamma}}{\beta^2x^2\Gamma}\,\frac{1}{2\pi\mathrm{i}}\int_{0^+-\mathrm{i}\infty}^{0^++\mathrm{i}\infty} dt\,e^{-\sqrt{t}}e^{t\,c/(\beta^2x^2\Gamma)} \\ &= \frac{\beta|x|\,\sqrt{\Gamma}\,e^{\beta|x|}}{2\sqrt{\pi}}\,c^{-3/2}\,e^{-(\beta^2x^2\Gamma)/(4c)-c/\Gamma}\ , \end{aligned} \tag{1.37}$$

To obtain the second equality the change of variable $t = (\beta x)^2(1+\Gamma s)$ has been made, while in the last equality one makes use of Eq. (31^8) from Ref. [82]. The important remark is that in Eq. (1.37) the right tail of the probability density of the cost has an exponential cutoff with rate

$$\frac{1}{\Gamma} = \frac{1}{2D(\gamma+2q/U^2)} < \frac{1}{2D\gamma} = \alpha_{\max}\ . \tag{1.38}$$

This result implies that, for any value of U, $\langle e^{\alpha C_0^T}\rangle$ diverges when $\alpha > \alpha_{\max}$. In particular, in the limit $\alpha \to \alpha_{\max}^-$, the functional $\langle e^{\alpha C_0^T}\rangle$ diverges also for controls arbitrarily close to the optimal one, for which $\Gamma^{-1} = \alpha_{\max}$.

Robustness of the optimal solution

The analytical solution in one dimension also allows to address the question about the robustness of the cost against perturbations in the control away from optimality. For controls U close to the optimal value U^*, the risk-sensitive cost $\mathcal{F}_\alpha$ can be approximated by a quadratic function,

$$\underbrace{\mathcal{F}_\alpha - \mathcal{F}_\alpha^*}_{\delta\mathcal{F}_\alpha} \simeq \frac{1}{2}\mathcal{F}_\alpha''(U^*)\,\big(\underbrace{U-U^*}_{\delta U}\big)^2\ .$$

In this approximation one can calculate the maximum tolerance on the control amplitude U, given an allowed level of suboptimality. Using the results from the previous subsection one obtains

$$\frac{\delta\mathcal{F}_\alpha}{\mathcal{F}_\alpha^*} = \frac{1}{2(1-2D\alpha\gamma)}\left(\frac{\delta U}{U^*}\right)^2 = \frac{\chi}{4D}\left(\frac{\delta U}{U^*}\right)^2 \tag{1.39}$$

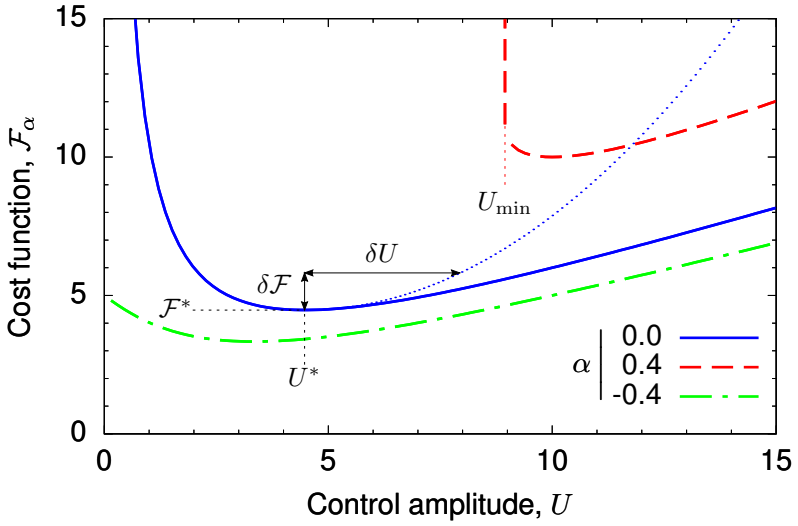


Fig. 1.3. **Robustness of the optimal solution.** The risk-sensitive cost function $\mathcal{F}_\alpha$ is plotted as a function of the control amplitude U, in a risk-neutral (solid blue), risk-averse (dashed red) and risk-seeking (dashed-dotted green) situation. The parabolic approximation around the minimum is shown for the risk-neutral case (dotted blue line). A deviation of the control from the optimum by a quantity δU corresponds to an increase in the cost $\delta\mathcal{F}$ (in the parabolic approximation). These two quantities are related by Eq. (1.39). Curves have been obtained with the same parameters as in Figs. 2 and 4 of the main text.

(see also Fig. 1.3). In the risk-neutral case, a relative error in the choice of U^* of 10% results in a small increase of 0.5% for the total cost incurred. Risk-averse strategies tend to be less tolerant to errors, while risk-seeking ones are more forgiving, in the sense specified by Eq. (1.39). We also remark that for $\alpha > 0$, the control amplitude is bound to be larger than a minimum value $U_{\min} = \sqrt{2D\alpha\gamma}\,U^*$, below which the risk-sensitive cost diverges.

1.1.3 Other forms for the cost of the control

The assumption that the cost for control is quadratic is particularly useful for two reasons. First, as we already remarked, it has a direct interpretation in terms of entropy and distance (Kullback–Leibler divergence) between ensembles of paths. Second, it has the property that the optimal control problem is linearly solvable (through the Hopf–Cole transformation the optimality equations can be cast into a linear form). Moreover, it has a physical interpretation, as the power dissipated while moving in a viscous medium. Obviously, this is not the most general form for the control cost which can be physically motivated. One can use a generic function of $|u|$. For instance, the control cost of the form $\eta\,|u|$ corresponds, in the low noise limit, to the minimization of the path length to the target.

Here we derive the optimal HJB equation for a target location problem with a single agent

and in the risk neutral case, where the control cost has the form an extra contribution which is linear in the control amplitude, $\eta\,|u|$. The minimization of the cost function constrained by the dynamics given in Eq. (1.1) is translated to the unconstrained minimization of the Lagrange functional

$$\mathcal{L}[u,p,\phi] = \int_0^\infty dt \int dx \left(\underbrace{\frac{\gamma}{2} u(x,t)^2 + \eta\,|u(x,t)|}_{\text{control}} + \underbrace{q(x)}_{\text{time}} \right) p(x,t) + \int_0^\infty dt \int dx\, \phi(x,t) \Big(\partial_t p + \nabla \cdot \big(u\,p\big) - D\nabla^2 p \Big)$$

The stationarity with respect to u yields the equation for the optimal control

$$\gamma\, u^* + \eta\, \frac{u^*}{\big|u^*\big|} = \nabla\phi \tag{1.40}$$

and with respect to p gives the optimality HJB equation

$$\partial_t \phi + D\nabla^2\phi + \frac{1}{2\gamma}\big|\nabla\phi\big|^2 - \eta\big|\nabla\phi\big| = q - \frac{\eta^2}{2\gamma}\ . \tag{1.41}$$

Through the Hopf–Cole transformation $\phi = 2D\gamma \log Z$, the HJB equation becomes

$$\partial_t Z + D\nabla^2 Z + \eta\big|\nabla Z\big| = \frac{1}{2D\gamma}\left(q - \frac{\eta^2}{2\gamma}\right) Z\ . \tag{1.42}$$

The dynamics of the desirability (chemoattractant) Z acquires a ballistic contribution, such that in addition to the diffusive motion it also propagates as a front.

This is a point that can be interesting to address in further developments of this work. However, throughout this Chapter, we restrict the study to the case of a quadratic cost function.

1.2 Mean-Field approximation

As previously noted, the Bellman equation (1.31), in the infinite-horizon case, becomes

$$-\,D\sum_i \nabla_i^2 Z_\alpha + \frac{1}{2D\tilde{\gamma}} \sum_i h_i\, Z_\alpha = 0\ , \tag{1.43}$$

equivalent to the stationary Schrödinger equation with zero energy for a quantum many-body problem of N identical particles with short-range interaction. Guided by this interpretation, and reckoning Eq. (1.43) to be impossible to solve with the boundary conditions discussed above, we adopt a mean-field approximation scheme. Such an approach is often successful in capturing the large-scale features of interacting systems [83]. We observe in passing that the

mean-field approach that we take here is exactly equivalent to the game-theoretical notion of cooperative mean-field games [42] which has been applied to crowd dynamics in a fast evacuation scenario [84].

Mean-field solutions are in general suboptimal, since a certain amount of information is discarded by the agents in the evaluation of the optimal action. However, if N is large and the system is diluted enough, a mean-field approximation for Eq. (1.13) is expected to yield an excellent approximation, as suggested by exact results in the closely related problem of a confined, repulsive Bose gas [41].

In the mean-field approximation one assumes that the controls are identical and independent, meaning

$$u_i^*(\bar{x}) \overset{\text{MF}}{=} u(x_i) \ . \tag{1.44}$$

While in principle the control of agent i depends on the entire configuration of the multi-agent system, the mean-field ansatz imposes that it actually depends only on the state of the agent i itself; the interchangeability of the agents, due to their identical nature, requires that the functional dependence of the control is irrespective of the label i. Consistently, the mean-field ansatz implies that the costs for individual particles are independent and identically distributed. Due to the formal similarity with the problem of finding the ground state of a gas of interacting bosons, one can draw an analogy between the optimal control equations in the mean-field approximation and the Gross–Pitaevski equations, as we are going to see.

We shall see in detail the derivation of the mean-field optimal control equations in the next two sections.

1.2.1 Risk-neutral case

As a consequence of the assumption of identical and independent controls, Eq. (1.44), the probability density function of the multi-agent system is factorizable into identical single-agent density functions ρ,

$$P(x_1 \dots x_N) = \prod_{i=1}^{N} \rho(x_i) \ , \tag{1.45}$$

provided that the initial positions of the N agents are independent. From the definition of the desirability function (as the exponential of minus the average cost-to-go) and as a straightforward consequence of Eq. (1.45), one finds that Z is factorized into single-agent desirability functions ζ

$$Z(x_1 \dots x_N) \overset{\text{MF}}{=} \prod_{i=1}^{N} \zeta(x_i) \ . \tag{1.46}$$

The single-particle density ρ then satisfies

$$\partial_t \rho + \nabla \cdot (u\, \rho) = D \nabla^2 \rho \ . \tag{1.47}$$

One can then replace the ansatz in Eqs. (1.44) and (1.45) in the cost functional $\mathcal{F}$ to obtain the cost-per-particle functional

$$\tilde{\mathcal{F}} = \int dt \int dx \left(\frac{\gamma}{2} u(x,t)^2 + q(x) + \frac{N-1}{2} \int dy\, v(x,y)\rho(y,t) \right) \rho(x,t) \ . \tag{1.48}$$

The mean-field optimal control equations follow by applying the Pontryagin principle to the functional $\tilde{\mathcal{F}}$ under the constraint (1.47), i.e. as the saddle point equations for the Lagrange functional

$$\tilde{\mathcal{L}} = \tilde{\mathcal{F}} + \int dt \int dx\, \phi(x,t) \Big(\partial_t \rho + \nabla \cdot (u\,\rho) - D\nabla^2 \rho \Big) \ , \tag{1.49}$$

where variations have to be calculated with respect to the single particle functions u, ρ (and ϕ, yielding the constraint). This leads to

$$\left. \frac{\delta \tilde{\mathcal{L}}}{\delta u(x,t)} \right|_* = \rho \left(\gamma u^* - \nabla \phi \right) = 0 \qquad \Rightarrow \qquad u^*(x,t) = \gamma^{-1} \nabla \phi(x,t) \ , \tag{1.50}$$

and

$$\begin{aligned} \left. \frac{\delta \tilde{\mathcal{L}}}{\delta \rho(x,t)} \right|_* &= \frac{\gamma}{2} u^{*2} + h_{\rm MF} - \partial_t \phi - u^* \cdot \nabla \phi - D\nabla^2 \phi \\ &= -\partial_t \phi - \frac{1}{2\gamma} (\nabla \phi)^2 - D\nabla^2 \phi + h_{\rm MF} = 0 \ , \end{aligned} \tag{1.51}$$

where $h_{\rm MF}$ is the mean-field cost

$$h_{\rm MF}(x,t) = q(x) + (N-1) \int dy\, v(x,y)\, \rho(y,t) \ . \tag{1.52}$$

When v is a contact interaction potential, $v(x,y) = g\,\delta(x-y)$, one has

$$h_{\rm MF} = q + g\,(N-1)\,\rho \ . \tag{1.53}$$

By applying the Hopf–Cole transformation $\phi = 2D\gamma \log \zeta$, the optimal control is

$$u^*(x) = 2D \nabla \log \zeta(x) \ , \tag{1.54}$$

where ζ is the mean-field desirability introduced in Eq. (1.46), satisfying the HJB equation

$$\partial_t \zeta - D\nabla^2 \zeta = \frac{1}{2D\gamma} h_{\rm MF}\, \zeta \ , \tag{1.55}$$

which for contact potential reads

$$\partial_t \zeta - D\nabla^2 \zeta = \frac{1}{2D\gamma} \Big(q + g(N-1)\rho \Big) \zeta \ . \tag{1.56}$$

Optimality of the cooperative solution

In a general multi-agent decision-making problem, optimality might require the agents to split into two or more sub-populations playing different strategies. The mean-field approximation introduced above fails to represent such scenarios, in that the control is assumed to be the same for all agents. However, we shall see in this section that the definition of the task imposes that the best mean-field solution is indeed the cooperative one, expressed in Eq. (1.44).

Let us assume that out of the total N agents, N_1 are of one species and $N_2 = N - N_1$ are of a second species, with different control. We can, for the sake of generality, introduce different collision costs depending on the species of the agents involved. Therefore, the cost rate for an agent of species α colliding with an agent of species β is $g_{\alpha\beta}/2$. The mean-field costs incurred by individual agents of species 1 and 2 are

$$C_1 = \bar{C}_1 + \frac{1}{2}\int dt\, dt \Big[g_{11}(N_1 - 1)\, \rho_1^2 + g_{12}(N - N_1)\rho_1\, \rho_2 \Big] \,, \tag{1.57a}$$

$$C_2 = \bar{C}_2 + \frac{1}{2}\int dt\, dt \Big[g_{21}\, N_1\, \rho_1\, \rho_2 + g_{22}(N - N_1 - 1)\rho_2^2 \Big] \,, \tag{1.57b}$$

where $\bar{C}_\alpha$ is the control and time cost for an agent of species α,

$$\bar{C}_\alpha = \int dt\, dx\, \rho_\alpha \left(\frac{\gamma_\alpha}{2} u_\alpha^2 + q_\alpha \right) . \tag{1.58}$$

The goal of each agent of species α is to maximize the cost C_α. We shall see that if the collision costs do not depend on the species involved, i.e. $g_{ij} = g$, the best partition of the system is a trivial one, i.e. either $N_1 = 0$ or $N_1 = N$.

One observes that C_1 and C_2 both have linear dependence on N_1:

$$\frac{\partial C_1}{\partial N_1} = \frac{g}{2}\int dt\, dt \Big[\rho_1^2 - \rho_1\, \rho_2 \Big] \,, \tag{1.59a}$$

$$\frac{\partial C_2}{\partial N_1} = \frac{g}{2}\int dt\, dt \Big[\rho_1\, \rho_2 - \rho_2^2 \Big] \,. \tag{1.59b}$$

If C_1 decreases with N_1 and C_2 increases with N_1 (i.e. decreases with N_2), or viceversa, then the two species should coexist. Assuming $\partial C_1/\partial N_1 < 0$ and $\partial C_2/\partial N_1 > 0$ would imply

$$0 > \frac{\partial C_1}{\partial N_1} - \frac{\partial C_2}{\partial N_1} = \frac{g}{2}\int dt\, dx \Big[\rho_1 - \rho_2 \Big]^2$$

which is not possible. The same conclusion holds for $\partial C_1/\partial N_1 > 0$ and $\partial C_2/\partial N_1 < 0$. Therefore, one must have both C_1 and C_2 decreasing (or increasing) functions of N_1, which makes it more desirable to have $N_1 \to N$ (or $N_1 \to 0$) for both species. This proves by *reductio ad absurdum* that the cooperative strategy is the best among mean-field solutions.

1.2.2 Risk-sensitive case

We now derive the mean-field equation for the risk-sensitive case, in particular the risk-averse one, $\alpha > 0$ (easily extended to the risk-seeking one, $\alpha < 0$). In this case, it is convenient to write the evolution of the joint stochastic process of particle positions and individual costs as the $2N$ coupled equations

$$\begin{cases} dX_i^t = u_i\big(X_1^t, \dots X_N^t\big)\, dt + \sqrt{2D} dW_i^t \ , \\ dC_i^t = \Big(\frac{\gamma}{2} u_i(X_1^t, \dots X_N^t, t)^2 + q(X_i^t) + \frac{1}{2}\sum_{j\neq i} v(X_i^t,\, X_i^t)\Big)\, dt \equiv c_i\big(X_1^t, \dots X_N^t, t\big)\, dt \ , \end{cases} \tag{1.60}$$

The multi-agent probability density function P, associated to Eqs. (1.60), is extended to include the individual cost variables in addition to the positions and in the mean-field ansatz $P(x_1, C_1 \dots x_N, C_N, t)$ is factorized

$$P(x_1, C_1 \dots x_N, C_N, t) \overset{\text{MF}}{=} \prod_i p(x_i, C_i, t) \ . \tag{1.61}$$

Notice that the marginal probability density for the individual position x, is the single-agent probability density function $\rho(x,t) = \int dC\, p(x, C, t)$, introduced above. It follows that the cost functional $\mathcal{G}_\alpha$ also factorizes:

$$\mathcal{G}_\alpha = \int dx_1\, dC_1 \dots dx_N\, dC_N\, P(x_1, C_1 \dots x_N, C_N, T)\, e^{\alpha \sum_i C_i} \overset{\text{MF}}{=} \Big(\int dx\, dC\, p(x, C, T) e^{\alpha C} \Big)^N . \tag{1.62}$$

The Fokker–Planck equation associated with Eqs. (1.60) is

$$\partial_t P + \sum_i \partial_{C_i}\big(c_i(\bar{x}, t)\, P\big) + \sum_i \nabla_i\big(u_i\, P - D\nabla_i P\big) = 0 \ , \tag{1.63}$$

and can be marginalized to the single-particle one by integrating over all particles but one:

$$\partial_t p + \nabla \cdot (up) + \partial_C \Big[\Big(\frac{\gamma}{2} u^2 + q + \frac{N-1}{2} \int dx'\, dC'\, v(x, x') p(x', C') \Big) p \Big] - D\nabla^2 p = 0 \ . \tag{1.64}$$

The mean-field optimal control equations are then derived (applying Pontryagin principle) as the saddle point equations of the functional

$$\begin{aligned} \tilde{\mathcal{L}}_\alpha[u, p, \chi] = & \int dx\, dC\, p(x, C, T) e^{\alpha C} \\ & + \int dx\, dC\, dt\, \chi_\alpha(x, C, t) \Big\{ \partial_t p(x, C, t) + \nabla \cdot (u(x, t) p(x, C, t)) - D\nabla^2 p(x, C, t) \\ & + \partial_C \Big[\Big(\frac{\gamma}{2} u(x, t)^2 + q(x) + \frac{N-1}{2} \int dx'\, dC'\, v(x, x') p(x', C', t) \Big) p(x, C, t) \Big] \Big\} \ . \end{aligned} \tag{1.65}$$

The function χ is the Lagrange multiplier enforcing the constraint expressed by Eq. (1.64). Stationarity with respect to the control yields

$$\left.\frac{\delta\tilde{\mathcal{L}}_\alpha}{\delta u(x,t)}\right|_* = -\int dC\, p(x,C,t)\Big[\nabla\chi_\alpha(x,C,t)+\gamma u(x,t)\partial_C\chi_\alpha(x,C,t)\Big] = 0\ , \tag{1.66}$$

and with respect to p gives

$$\left.\frac{\delta\tilde{\mathcal{L}}_\alpha}{\delta p(x,C,t)}\right|_* = e^{\alpha C}\,\delta(t-T) - \Big(\partial_t\chi_\alpha + u^*\cdot\nabla\chi_\alpha + D\nabla^2\chi_\alpha + \big(\gamma u^{*2}/2 + h_{\text{MF}}\big)\partial_C\chi_\alpha\Big) = 0\ . \tag{1.67}$$

The last one is the (backward) equation for the functional $\chi_\alpha(x,C,t) = \big\langle \exp\alpha C_0^T \big| X^t = x,\, C_0^t = C\big\rangle_*$, where h_{MF} is the mean-field cost Eq. (1.52).

$$h_{\text{MF}}(x,t) = q(x) + (N-1)\int dx' \underbrace{\int dC' p(x',C',t)}_{\rho(x',t)} v(x,x')\ . \tag{1.68}$$

The δ-function at the final time sets the condition $\chi_\alpha|_{t=T} = e^{\alpha C}$. Note that the function χ_α can be expressed as

$$\chi_\alpha(x,C,t) = \big\langle e^{\alpha(C_0^t + C_t^T)}\big| X^t = x,\, C_0^t = C\big\rangle_* = e^{\alpha C}\big\langle e^{\alpha C_t^T}\big| X^t = x\big\rangle_* \equiv e^{\alpha C}\psi_\alpha(x,t) \tag{1.69}$$

We therefore see that the optimal control can be written in terms of $\psi_\alpha(x,t)$

$$u^*(x,t) = \frac{1}{\gamma}\nabla\Big(-\frac{1}{\alpha}\log\psi_\alpha(x,t)\Big)\ , \tag{1.70}$$

i.e. as the gradient of the risk-sensitive (mean-field) value function $\phi_\alpha = -\alpha^{-1}\log\psi_\alpha$, which satisfies the HJB equation

$$\partial_t\phi_\alpha + \frac{1}{2\tilde{\gamma}}\big(\nabla\phi_\alpha\big)^2 + D\nabla^2\phi_\alpha = h_{\text{MF}}\ , \tag{1.71}$$

where we recall from the previous section that $\tilde{\gamma} = \gamma/(1-2D\alpha\gamma)$. The mean-field desirability $\zeta_\alpha = \exp(\phi_\alpha/2D\tilde{\gamma})$ then satisfies the linear HJB equation

$$\partial_t\zeta_\alpha + D\nabla^2\zeta_\alpha = \frac{1}{2D\tilde{\gamma}}h_{\text{MF}}\,\zeta_\alpha\ . \tag{1.72}$$

Perhaps surprisingly, also in the mean-field approximation, the HJB equation for the desirability in the risk-sensitive case exhibit the same form as the risk-neutral one.

1.3 The emergence of chemotaxis

In the last two sections we derived the optimal control equations for a cooperative search task and approached them in an approximation scheme which takes the inspiration from the theory of interacting quantum gases, as well as from game theory.

Our findings can be summarized as follows. In the terminal-state setup, which is relevant for the search task studied here, the mean-field desirability ζ solves the Hamilton–Jacobi–Bellman equation

$$D\nabla^2\zeta = \frac{1}{2D\tilde{\gamma}}\Big(q + g(N-1)\,\rho\Big)\,\zeta\ , \tag{1.73}$$

where $\tilde{\gamma} = \gamma/(1-2D\alpha\gamma)$, for a risk-sensitivity parameter α and contact interaction potential $v(x_i,\,x_j) = g\,\delta^d(x_i - x_j)$. The boundary conditions with which Eq. (1.73) is supplemented follow from the interpretation of ζ_α in terms of the optimal cost-to-go, and are $\zeta_\alpha = 1$ at the target, $\zeta_\alpha = 0$ at the forbidden boundaries as well as infinitely far from the target, and $\nabla_\perp\zeta_\alpha$ at the reflecting boundaries. The optimal control follows from the desirability as $u^* = 2D\tilde{\gamma}/\gamma\,\nabla\log\zeta_\alpha$, and the single-agent density ρ then evolves according to the Fokker–Planck equation

$$\partial_t\rho + \nabla\cdot\big(\rho\,u^*\big) = D\nabla^2\rho\ . \tag{1.74}$$

Remarkably, the set of equations (1.73) and (1.74) is identical to a limiting case of the well-known Patlak–Keller–Segel equations, which was first introduced to model microbial chemotaxis at the population level [71, 72]

$$\begin{aligned} &\partial_t n + \nabla\cdot(\chi n\nabla\log s) = D\nabla^2 n\,,\\ &D_s\nabla^2 s - ks - \beta n s = 0\ . \end{aligned} \tag{1.75}$$

Here n is the number density of microbes and s is the chemoattractant concentration. Comparing the Bellman equation Eq. (1.46) with the second row of Eq. (1.75) one sees that the desirability ζ is proportional to the chemoattractant concentration c, to which agents respond logarithmically – they sense only fold-changes in levels, in accord with the Weber–Fechner law [14]. The chemotactic coefficient, measuring the strength of the response to the relative gradients, is $\chi = 2D\tilde{\gamma}/\gamma$ in this case. The chemoattractant is kept at constant concentration at the target, is degraded with rate k proportional to $q/(2D\tilde{\gamma})$ and consumed at rate β proportional to $g/(2D\tilde{\gamma})$ per cell. We note in passing that perfect adaptation is an implicit feature of Eqs. (1.73) and (1.74), in that there is no chemokinesis, i.e. the random motility D does not depend on ζ.

Optimal cooperative search can then be realized by biophysical systems in which the target emits a diffusible chemical cue in the environment. Agents can behave optimally by responding to fold–changes of this signal and actively modifying it by consumption.

Learning to search optimally: scouts, beacons and recruitment

It would be useful to extend this setting to the relevant case when the location of the target is a priori unknown and the target does not spontaneously send out signals to facilitate the work of the agents. In other words, we seek a way to include the process of discovery of the target's location and the successive construction of the solution to Eq. (1.46). As we show below, this can be accomplished by adding a production term in the equation for concentration, which is the analog of the process by which information is written on memory.

Our solution to the learning problem goes as follows. Initially, the concentration is set to a constant everywhere in space. In the first part of the search process, agents wander at random and the concentration decays as the chemoattractant evaporates and is consumed. As a result, agents explore space away from their initial location. This is called the *scouting* process. When agents eventually reach the target, they start the production of chemoattractant on site, either releasing it themselves, e.g. in the form of a pheromone-like cue [85, 86], or inducing its production by the target, which may happen in practice by triggering specific gene expression [87] or by transforming it into attractive waste material. The net effect is that a beacon signal is emitted from the target, and it leads to the recruitment of all other agents towards it. A mathematically precise description of the process outlined above requires only the addition of two terms to the optimality equation (1.46)

$$\underbrace{\epsilon\partial_t\zeta}_{\text{relaxation}} \quad -D\nabla^2\zeta + \frac{1}{2D\gamma}\big(q + g(N-1)\rho\big)\,\zeta = \underbrace{f(t)1_{\texttt{target}}}_{\text{production}}\,, \tag{1.76}$$

where $f(t) = \int_0^t dt' \int_{\texttt{target}} ds \cdot J_\rho$ is the cumulated number of agents that have reached the target up to time t, and $J_\rho = (2D\nabla \log\zeta)\rho - D\nabla\rho$ is the spatial flux of agents. The indicator function $1_{\texttt{target}}$ specifies that production takes place only at the target. The relaxation term is interpreted as a delay in writing information on memory, i.e. a learning rate. When production and diffusion balance, the optimal solution, Eq. (1.46), is attained.

1.4 Applications

In this section we illustrate the learning dynamics at work in two examples of cooperative search games. In the first example, we discuss the effect of risk sensitivity for agents searching independently for a circular target. In this example the mean-field ansatz yields the exact solution and can be calculated analytically. In a second example we show the behaviour of the mean-field solution in a maze-escape task, and point out how the cooperative interactions make the random search faster.

1.4.1 A model for bacterial predation

The first example features a circular target in an infinite two-dimensional domain and can be thought of as a basic model for bacterial predation [87, 88]. One can think of the agents as microbes whose target is a host cell that they are trying to infect. In this example we do not include the costs for collision. Since the agents are independent at all times provided that they are at the start, the mean-field equations are exact. In this simple geometry, it is possible to find the analytical solution. This is a case study that can be useful in understanding the effect of the time cost as well as the risk sensitivity.

Exact analytical solution

If the target has radius R and we choose the origin of the coordinate system to be its center, the HJB equation in cylindrical coordinates reads

$$\frac{D}{r}\partial_r\Big(r\partial_r\zeta_\alpha\Big) - \frac{q}{2D\tilde{\gamma}}\,\zeta_\alpha = 0\ , \tag{1.77}$$

where the desirability ζ_α depends only on the radial coordinate, and, from previous sections, $\tilde{\gamma} = \gamma/(1-2D\alpha\gamma)$. Given the connection of the desirability with the expected cost-to-go

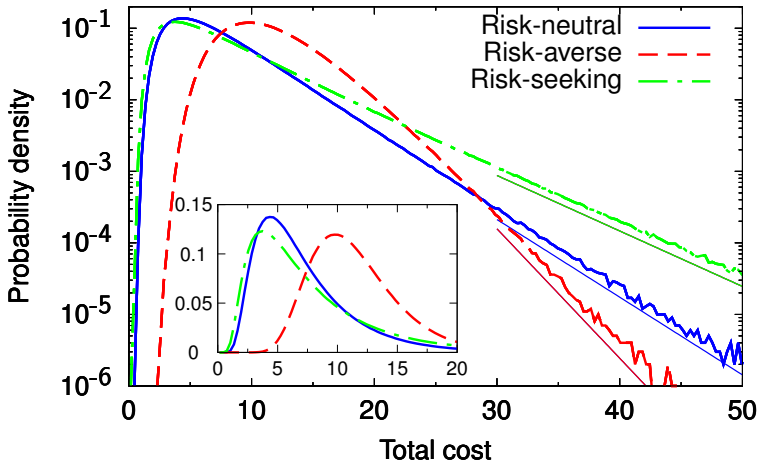


Fig. 1.4. **Risk-sensitive cooperative predation.** Probability density functions of the total cost per agent under the optimal risk-neutral control (solid blue), risk-averse ($\alpha = 0.4$, dashed red) and risk-seeking ($\alpha = -0.4$, dashed-dotted green). Other parameters are set as in Fig. 1.6. The same distributions are plotted on linear scale in the inset. Positive values of α (optimal risk-aversive behavior) lead to steeper exponential decay for large values of the cost while negative values (optimal risk-seeking behavior) enhance fluctuations with lower-than-average cost, at the price of increasing the frequency of higher-than-average fluctuations. The thin solid lines are the analytical estimates for the right tails, which can be calculated in the one-dimensional case (see Sec.1.1.2).

function, the boundary conditions (see Sec. 1.1.1) for Eq. (1.77) are $\zeta(r \to \infty) = 0$ and $\zeta(R) = 1$. The solution to this problem is

$$\zeta_\alpha(r) = \frac{K_0(r/\lambda)}{K_0(R/\lambda)} , \qquad \text{where} \quad \lambda = \sqrt{\frac{2D^2\tilde{\gamma}}{q}} , \tag{1.78}$$

which yields the optimal control

$$u^* = \frac{2D\tilde{\gamma}}{\gamma}\nabla \log \zeta_\alpha = -\frac{\sqrt{2\tilde{\gamma}q}}{\gamma}\frac{K_1(r/\lambda)}{K_0(r/\lambda)}\,\hat{e}_r , \tag{1.79}$$

where K_ν are the modified Bessel functions of second kind and $\hat{e}_r \equiv x/r$, i.e. the outward unit vector pointing to x from the origin.

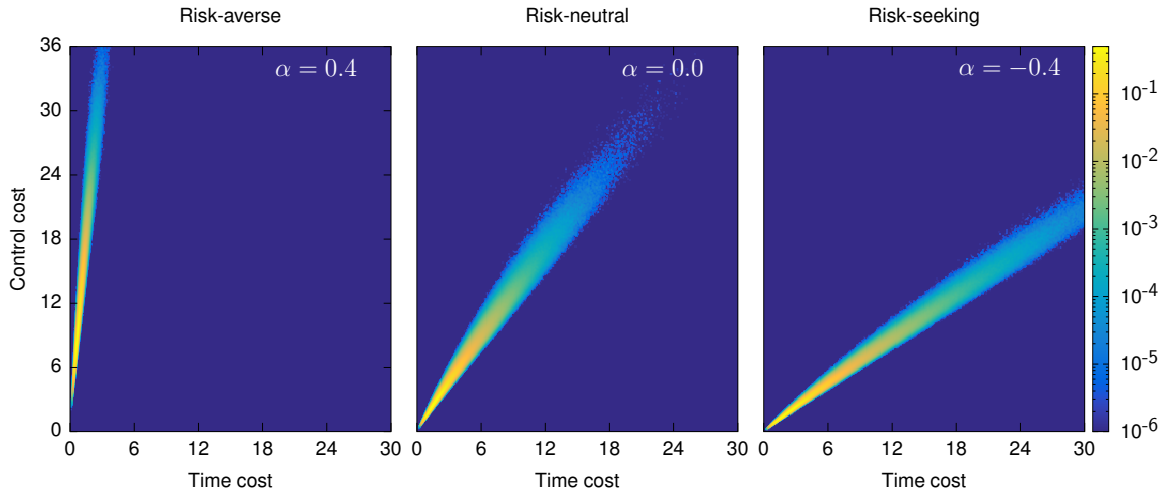


Fig. 1.5. **Risk-neutral vs Risk-sensitive.** Two-dimensional histograms of time-expenditure and control costs, from simulations of Eq. (1.1) with the mean-field control in Eq. (1.79), for the risk-averse, risk-neutral and risk-seeking situation. The costs for time-expenditure and control are positively and almost linearly correlated. The optimal control for the risk-neutral problem is such that control and time-expenditure costs are very similar. Instead, the risk-averse optimal controller tends to pay more on control (reducing possible dangerous fluctuations towards high values of the cost), whereas the risk-seeking controller allows for large time-expenditure cost while reducing the control.

In Figs. 1.4 and 1.5 we see the effect of risk-sensitivity on the distribution of the costs, obtained via numerical simulations of the exact optimal control, Eq. (1.79). In Fig. 1.4 we compare the distribution of the total cost (sum of time and control costs) for different values of the risk-sensitivity parameter α, other parameters being the same. We notice that increasing risk-aversion (growing α) leads to steeper exponential decrease in the right tail of the distribution. In Fig. 1.5, we look at the joint probability distribution of time and control cost,

showing that they are highly correlated. One can see that the risk-sensitivity parameter sets an imbalance between the control cost and the time-expenditure cost: risk-averse strategies are prone to pay much more price on control than on time-expenditure, while for risk-neutral strategies the difference is much less pronounced; risk-seeking controllers, instead, pay less cost in control, confident of being driven to the target by the noise.

Testing the learning dynamics

In Fig. 1.6 we show results from the simulation of a large number of agents under the controlled dynamics with drift from the numerical solution of Eq. (1.76), and compare them with the uncontrolled dynamics. From visual inspection, the gain in the number of agents that have reached the target is apparent. More quantitatively, the time average cost per agent as a function of time (Panels c and d of Fig. 1.6), which is proportional to the number of agents which have not reached the target at a given time, falls off exponentially for the controlled

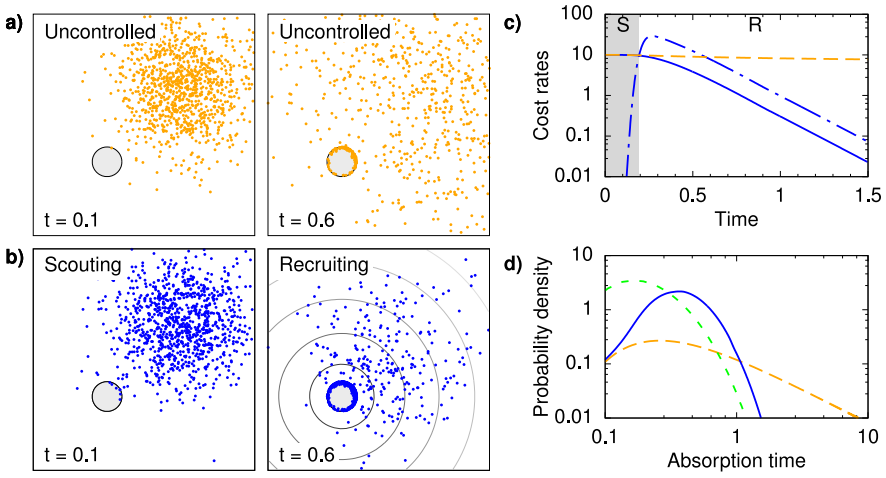


Fig. 1.6. **Optimal cooperative predation.** Comparison between the uncontrolled *a)* and controlled dynamics *b)*, in the non-interacting case ($g = 0$). The agents are initially localized in a small region of space and are required to reach the target (grey disk). They initially undergo unbiased diffusion during the scouting phase and when some reach the target, the recruiting phase begins. The chemical cue is emitted from the target and degraded at constant rate, resulting in a gradient (grey contour lines, in logarithmic scale) which elicits a drift toward the target in all other agents. In these simulations the parameters are: $\gamma = 1$, $q = 10$, $D = 1$, $g = 0$, $\epsilon = 0.1$. *c)*: Average cost rate for time (uncontrolled: dashed orange line; controlled: solid blue line) and for control (dash-dotted blue line) The scouting phase (S, shaded) and the recruiting (R) phase are dominated by time cost and by control cost, respectively. *d)*: Probability density function of the time to reach the target for controlled and uncontrolled agents (color code as in *c*), compared to the case with optimal solution (without learning, green dashed line). The distributions for the dynamics with are similar at small times, while at large times controlled agents display an exponential decay against a $-3/2$ power law for uncontrolled ones. The learning agents behave optimally after a delay due to the existence of a scouting phase.

case while it exhibits a very slow decay for uncontrolled diffusion.

1.4.2 Escaping Villa Pisani

The second example is crowd evacuation from a complicated domain. Agents, initially localized in the center of a maze, are required to find the exit with the minimal cost. The domain in which we performed this numerical experiment is a reproduction of the historical maze in the gardens of Villa Pisani (Stra, Italy). In this example, agents are introduced at the center of the maze at a constant injection rate. In Fig. 1.7 we see the emergence of the phases of scouting and recruitment, and eventually, we observe that the agents trace out the optimal path to the exit. Notice that, during the scouting phase, the density of agents propagates as

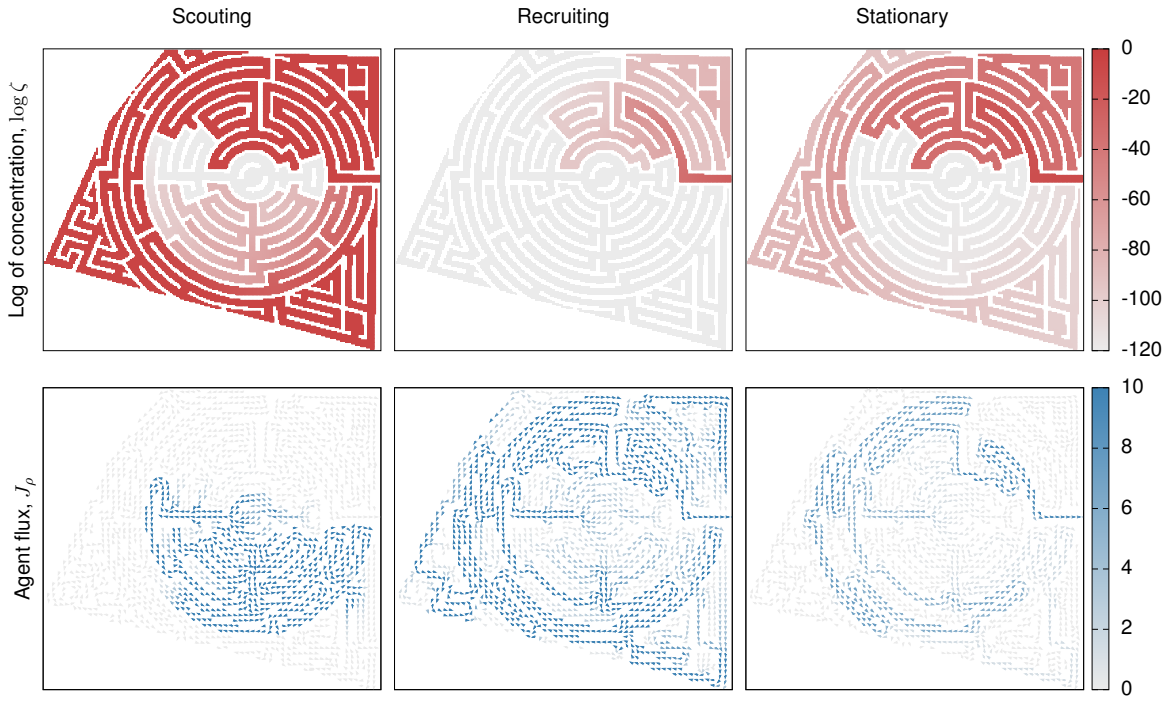


Fig. 1.7. **Optimal crowd evacuation.** Agents are injected at a constant rate at the center of the maze and have to find the exit (on the right side of the maze), as quickly as possible while minimizing collisions. The panels show a numerical simulation of Eqs. (1.54), (1.74) and (1.76). The desirability (=concentration, see Table 1.1) is shown in the top panels, while the flux of agents J_ρ is displayed in the bottom panels. During scouting (left column), the population consumes the chemical, leading to an outward-driven scouting process, faster than pure diffusion. Upon reaching the target, agents lay the beacon signal and recruit those who lag behind to the target (middle column). Eventually, since agents are continuously injected in this case, a stationary state is reached where agents track the optimal path from the center to the exit (right column). The parameters are $D = 1$, $\gamma = 1$, $\epsilon = 10^{-1}$, $q = 10$ and $g(N-1) = 100$.

a front with speed which is proportional to $\sqrt{N}$ (see SI, Sec. 2.1.2) so that the collective is much faster in finding the target than the single agent (which instead reaches it diffusively).

Effect of the collision cost: travelling wave solution of PKS in 1D

As noted in the main text, the optimal control equations in the mean-field approximation are equivalent to the Patlak–Keller–Segel (PKS) equations with logaritmic response. In this section we show that in the case where $q = 0$ (no time expenditure cost) it is possible to find a travelling wave solution to the PKS equations in one dimension [89]. We shall see that the combination $g(N-1)$ enters the definition of the travelling wave velocity.

In one dimension, the optimal control equations for $q = 0$ are

$$\begin{cases} D\partial_x^2\zeta - \dfrac{1}{2D\gamma}\big(q + g(N-1)\rho\big)\zeta = 0 \ , \\ \partial_t\rho + 2D\partial_x\big(\rho\,\partial_x\zeta/\zeta\big) - D\partial_x^2\rho = 0 \ . \end{cases} \tag{1.80}$$

We impose the boundary conditions $\zeta(-\infty) = 0$, $\zeta(+\infty) = 1$, $\rho(-\infty) = \rho_\infty$ and $\rho(+\infty) = 0$. Such boundary conditions correspond to the situation in which a constant supply of agents is provided very far on the left and the target is far on the right. The system admits a travelling wave solution, $\zeta(x,t) = Z(x-ct)$ and $\rho(x,t) = R(x-ct)$, with velocity $c > 0$. With this ansatz, Eqs. (1.80) write

$$\begin{cases} DZ'' - \dfrac{g(N-1)}{2D\gamma}RZ = 0 \ , \\ -cR' + 2D\big(R\,Z'/Z\big)' - DR'' = 0 \ , \end{cases} \tag{1.81}$$

where the symbol $'$ indicates the derivative with respect to the single variable $z = x - ct$. The second equation in the system (1.81) can be straightforwardly integrated once and gives

$$R' = (2Z'/Z - \kappa)\,R + \beta \ ,$$

where $\kappa = c/D$. If we impose the boundary conditions $R|_\infty = 0$ and $R'|_\infty = 0$, the integration constant β vanishes. One further integration gives

$$\log R = 2\log Z - \kappa z + \alpha \qquad \Rightarrow \qquad R(z) = A\,e^{-\kappa z}\,Z(z)^2 \ .$$

By replacing this solution into the first equation of the system (1.81), and by defining $Z(z) = e^{\kappa z/2}\chi(z)$

$$D\chi'' + c\chi' + \frac{1}{2D}\Big(\frac{c^2}{2} - \frac{g(N-1)}{\gamma}\,A\chi^2\Big)\chi = 0 \ , \tag{1.82}$$

which is the traveling-wave form of the Fisher–Kolmogorov–Petrovski–Pisounov (FKPP) equation with cubic nonlinearity,

$$\partial_t u = D\partial_x^2 u + r\,u(1-u^2) \ .$$

Notice that $R = A\,\chi^2$, and therefore the boundary conditions for χ follow from the ones for ρ. From the stability condition $c^2/2 - g(N-1)A\chi^2(-\infty)/\gamma = 0$, it follows that the speed of the wave front is

$$c = \sqrt{\frac{2\rho_\infty}{\gamma}\, g(N-1)}\ . \tag{1.83}$$

The number of agents therefore influences the speed at which the agent density profile propagates: more agents consume the chemoattractant more rapidly, hence giving rise to the steepest gradients of its concentration ζ, in turn yielding a stronger drift..

1.5 Discussion

We have shown how the problem of finding the optimal solution to a cooperative search game can be solved by agents which produce and consume a diffusible chemical cue, establishing a close correspondence between notions from decision-making theory and biophysical properties of chemotactic microorganisms. What are the implications of our results?

From the standpoint of search theory, our findings provide a solid theoretical rationale for the many solution inspired by chemotaxis, from computational [90–92], to biological [93, 94] and physico-chemical ones [95, 96]. At the practical level, we offer explicit expressions for the optimal choice of the parameters that appear in these biomimetic approaches, allowing to shortcut the painstaking procedure of parameter tuning. When the exact optimal values are difficult to realize in practice, as it may happen for real biological or physico-chemical systems, our analysis permits to evaluate the impact of the suboptimal choice on the performance of the search method. Conversely, from the viewpoint of chemotaxis, we remark that the dictionary in Table 1.1 can also be read in reverse, which allows to solve the inverse problem of retrieving the decision-making parameters from biophysical observations. Indeed, from the experimental knowledge of the effective microbial diffusivity D, the chemotactic coefficient χ, the chemoattractant diffusivity D_s, the degradation rate k and consumption rate β, one can invert the equations in Table 1.1 and extract the values of α, γ, q and g to within an overall, irrelevant multiplicative factor. For example, bacterial chemotaxis experiments (Fig. 6 in Ref. [97]) give $\chi/D \approx 12$, which translates into $2D\alpha\gamma \approx 5/6$. This value is very close to the upper limit for risk aversion, suggesting that bacteria try to minimize the impact of unfavorable fluctuations – a conclusion that has also been reached by other means [98].

Our results open many questions and directions for further research. Among them, extending the present analysis to cooperative Markov games [99] would allow to address problems of traffic control on arbitrary networks, for example, as well as many problems of cooperative resource allocation and transfer [100]. What will be the equivalent of chemical communication in these contexts? Another exciting direction to investigate is the realm of non-cooperative games and ecology: will optimal strategies and population dynamics be shaped by chemical signalling in this case as well?

2

From Conformational Spread to Allosteric and Cooperative Binding models of E. coli flagellar motor

In the previous Chapter we saw that chemotaxis emerges as the optimal solution of a collective decision making problem. Chemotaxis is then the *algorithm* that microbes are able to implement in order to navigate their environment, searching for food or escaping from noxious chemicals. In this Chapter we address the issue of *how* chemotaxis works at the molecular level, looking at it from a statistical mechanical perspective.

Escherichia coli (*E. coli*) is one of the model organisms for studying bacterial chemotaxis [18]. Thanks to its flagella, activated by bi-directional rotary motors, *E. coli* is able to move towards more favorable environments by optimally alternating *runs* and *tumbles*, which approximately consist of straight lines and random "turns", respectively.

The biochemical mechanisms underlying the chemotactic response of *E. coli* are well understood at the molecular level [16]. A sensing apparatus is devoted to detecting information about the environment, by measuring the concentration of chemicals (generally called, in this context, *chemoeffectors*). The arrangement and functioning of the receptors present on the *E. coli*'s cellular membrane have been extensively investigated also from the theoretical point

of view (see, e.g. [101, 102]). The information collected by the receptors is transduced to the flagellar motors through the "messenger molecule" CheY. The cytoplasmic concentration of its phosphorylated form CheY-P varies according to the activity of the membrane receptors. The CheY-P molecule then acts as a regulator of the activity of the flagella by binding to their motors. These are constituted by rings of Fli molecules, arranged in units called *protomers*. Motors are biased by the Fli occupancies to rotate counterclockwise (CCW) or clockwise (CW). When all the motors are in the CCW state, flagella form a bundle which propels the cell in a forward run; if at least one motor is in the CW state instead, the bundle splits apart and the cell tumbles.

Such a mechanism is an example of *allosteric* (or *indirect*) regulation, where the activity of protein complexes changes collectively upon independent binding of external molecules. The original model which encodes the concept of cooperativity in indirect regulation is the one proposed by Monod, Wyman and Changeux (MWC), commonly known as *concerted* model [103, 104].

Shortly after the paper by Monod, Wyman and Changeux, Eigen realized that the concerted model can be extended in order to offer a more graded interplay between the interactions within allosteric complexes and their binding affinities [105]. When the interactions are local, this generalized model takes the name of *conformational spread* model (see Sec. 2.1) and is nowadays understood in a statistical mechanical framework in the light of the ferromagnetic Ising model, to which it is formally equivalent [101, 106].

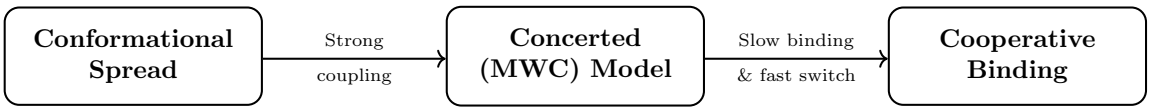
These allosteric models have found application in bacterial chemotaxis. In Ref. [107], the authors showed how the MWC model is able to reproduce the activity of the flagellar motor of *E. coli* as a function of the concentration of cytoplasmic CheY-P. [107]. In this paper, the authors recognized that the balance between the different CheY-P affinity in the two activity states and the size of the motor protein complex was essential in explaining the observed cooperative behaviour of the switch. The MWC model turned out to be particularly suitable for describing the flagellar switch of *E. coli*, in that it accounts for the correct degree of cooperativity with a proper choice of the parameters.

The conformational spread model has been applied to bacterial chemotaxis, both for the membrane receptors [101, 106] and for the flagellar rotary motors [108]. By means of a simulation of its associated Glauber dynamics [109], a numerical test of the conformational spread model against the experimental measurement of the rotation speed of the flagella has been performed [110]. Such analysis showed an excellent agreement between experiments and numerical simulations regarding several aspects of the dynamics, such as the switching time distribution at fixed values of the cytoplasmic CheY-P concentration and the sensitivity of the switch upon small variation of CheY-P. A more detailed numerical analysis of the model followed up [111], in which also other dynamical properties of the conformational spread model were quantified (like the locked-state behaviour, namely, the time spent by the motor in a rotational state between two consecutive switches) and a more precise estimation of the

parameters of the model which best fit the experimental results was given.

From the analytical point of view, one major obstacle to the study of the conformational spread model resides in the large number of states. The one-dimensional nature of the ring allows nonetheless for an exact calculation of its partition function at equilibrium via the transfer matrix method [112]. However, no analytical treatment of the non-equilibrium behaviour of the model has ever been attempted, to our knowledge.

In this Chapter we present an analytical derivation of the non-equilibrium properties of the conformational spread as a model of the flagellar switch. Our analysis hinges upon the presence of a hierarchy of widely separated time scales, as confirmed by experiments. Due to the strong interaction between the protomers, the coarsening of activity domains in the ring is much faster than the nucleation of a domain, i.e. the transitions away from the state of all active or all inactive protomers. This allows the treatment of the whole motor as an allosteric switch in two different activity states (CW and CCW), essentially described by the MWC model. The nucleation of a domain is in turn much more frequent than the binding/unbinding of a CheY-P molecule by one protomer, which makes it possible to operate a quasi-static approximation for the number of bound CheY-P and get a description of the slow binding dynamics, to which the activity is slaved. This separation of time scales allows us to reduce the complexity of the full conformational spread dynamics by progressively averaging the faster degrees of freedom and obtain, in the end, an effective cooperative model which captures the relevant features of the flagellar switch on the slowest time scales. The effective rates of the emergent "coarse-grained" cooperative binding model are expressed in terms of the rates of the original "microscopic" conformational spread model. In short, the rationale of our approach can be schematically summarized as follows:



The Chapter is structured as follows: in Sec. 2.1 we present the conformational spread model, outlining its equilibrium properties and introducing the dynamics (satisfying detailed balance) which is relevant for our study and is the object of our multiscale analysis; in Sec. 2.2 we show that, in our experimentally justified assumptions, it is possible to reduce the conformational spread to the concerted MWC model; a further time-scale separation is the subject matter of Sec. 2.3, resulting in a cooperative binding model (formally, a birth-and-death process with site-dependent rates) that is compared with experiments in Sec. 2.4.

2.1 Conformational Spread Model

The ring of proteins forming the motor of the *E. coli* flagella has been shown to be very well described by the conformational spread model [106, 108, 110]. This model consists in N identical units, or *protomers*, each of which can appear in two different states, active (A) or inactive (I): a protomer in the active state increases the probability of CW rotation of the motor, and of CCW rotation in the inactive state (see Fig. 2.1). Moreover, each protomer

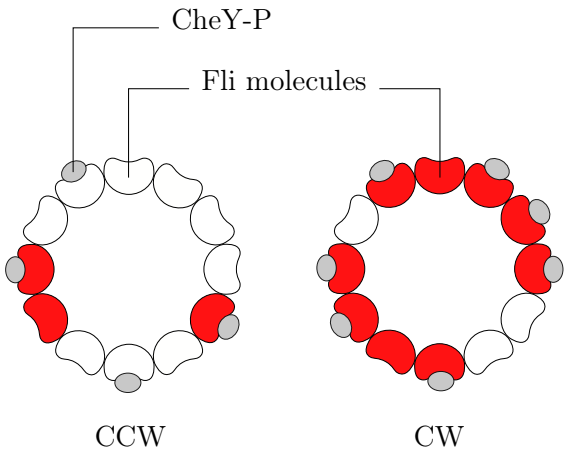


Fig. 2.1. **The flagellar motor.** The Fli molecules are depicted in white (inactive state, I) and red (active state, A), while the grey spots represent the CheY-P regulator. The motor rotates counterclockwise when most of the protomers are in the inactive state (left) and clockwise otherwise (right).

can also bind a *ligand*, corresponding to the CheY-P chemotactic regulator: we refer to the protomer as in the *bound* (B) state when a ligand is attached to it, or *unbound* (U) otherwise. Therefore, the single protomers can be in 4 different states, corresponding to all the possible *activity* and *binding* configurations.

The state diagram of a single protomer is depicted in Fig. 2.2: the A state is energetically more favorable than the I state when a ligand is bound and vice versa. This property ensures that this is a good model for allosteric regulation. Namely, a bias in the activity of the motor depends on the number of bound CheY-P molecules: at fixed high concentration of cytoplasmic CheY-P (denoted by c) the motor will most probably spin clockwise. The state of the full system is specified by the sequence $s = \{(\alpha_1, \ell_1), \dots (\alpha_N, \ell_N)\}$, where the subscripts label the N protomers, α indicates the activity state A or I, and ℓ stands for the binding state B ($\ell = 1$) or U ($\ell = 0$): hence, the number of possible configurations of the ring with N protomers is $(2 \times 2)^N$.

In addition, the protomers are coupled via a nearest neighbour interaction, which depends on their activity states *only*: in particular, the energy is lowered by a quantity J when the

neighbouring protomers are in the same activity state A or I. It turns out that the activity of the ring (fraction of active protomers) is more sensitive to small variations of concentration of ligands in the interacting case than in a system of N independent protomers. Therefore, the coupling is an essential ingredient which enhances the sensitivity of the whole complex.

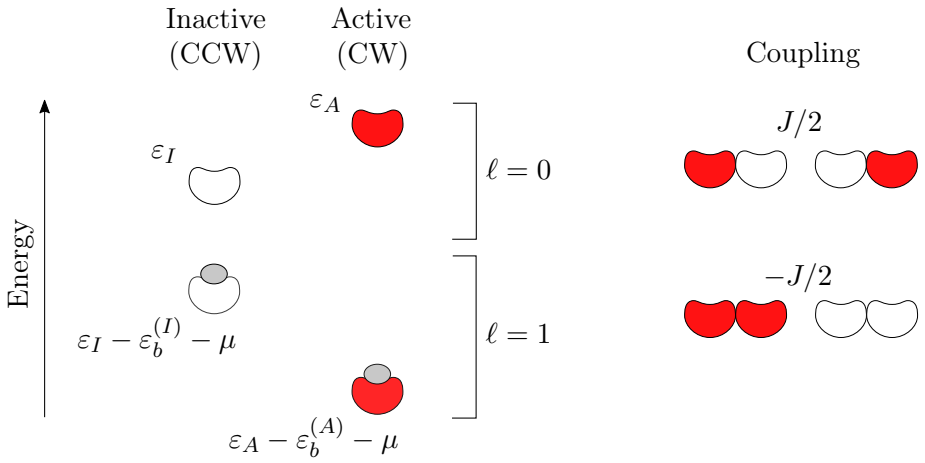


Fig. 2.2. **State diagram and couplings in the Conformational Spread Model.** On the left, the energy levels of the single protomer states: the active (CW) configuration is energetically favorable in the unbound case ($\ell = 0$), while the inactive (CCW) has lower energy when in the bound case ($\ell = 1$); the binding regulates the activity of the protomers. The notation and the general scheme has been borrowed from [104]. On the right, the coupling energy: the "ferromagnetic" coupling (independent of ℓ) accounts for the high sensitivity of the response of the ring upon binding.

The conformational spread model is very reminiscent of the Ising model. In fact, if one associates to each protomer a *spin* variable σ_i taking value $+1$ when the protomer is active ($\alpha_i = A$), or -1 when it is inactive ($\alpha_i = I$), one can represent the states of the system as $s = \{(\sigma_i, \ell_i)\}_{i=1}^N$ and the equilibrium properties of the model are determined by the Hamiltonian

$$H = -\frac{J}{2} \sum_{\langle i,j \rangle} \sigma_i \, \sigma_j - \sum_i h(\sigma_i, \, \ell_i) \ , \tag{2.1}$$

where J is a positive constant and h is the single-protomer contribution, reproducing the energy diagram in Fig. 2.2,

$$h(\sigma, \, \ell) = \frac{1}{2}\Big[\varepsilon_I - \varepsilon_A - (\varepsilon_b^{(I)} - \varepsilon_b^{(A)})\ell\Big]\sigma \\ - \frac{1}{2}\Big[\varepsilon_I + \varepsilon_A - (\varepsilon_b^A + \varepsilon_b^I + 2\mu)\ell\Big] \ .$$

The one in Eq. (2.1) is an Ising Hamiltonian with ferromagnetic coupling J, where h plays the role of an external local magnetic field, set by the occupation ℓ; in Eq. (2.2), μ is the

chemical potential, determined by the concentration of CheY-P, c, by

$$\mu = \mu_0 + \frac{1}{\beta} \ln \frac{c}{c_0} \ , \tag{2.2}$$

where μ_0 and c_0 are reference chemical potential and concentration, respectively. Hereafter, the notation σ and α will be used interchangeably, according to the situation. The partition function $Z = \sum \exp(-\beta H(s))$ (where $\beta = 1/k_B T$ is the inverse temperature and the sum is done over the 4^N possible states of the ring of protomers) has been calculated exactly via transfer matrix approach [112]. The analytic results found therein fit very well the experimental curves [16] of the ligand occupancy (average fraction of bound protomers) and the activity (fraction of protomers in the A state) as a function of the concentration of CheY-P.

If on one hand the equilibrium properties of the conformational spread model are exactly known, on the other hand a full-fledged analytic treatment of the stochastic dynamics of this model seems difficult. In the definition of the conformational spread model given above, there is no prescription about the dynamics. A natural choice which satisfies detailed balance is the Glauber-like [109] Markovian dynamics, used in numerical simulations of this model in Refs. [110, 111]. In such prescription, the process $\{S_t\}_t$ which accounts for the kinetics of the conformational spread model is governed by the master (Kolmogorov) equation

$$\frac{\partial}{\partial t} P(s,t) = \sum_{s'} \left[P(s',t)\, K(s' \to s) - P(s,t)\, K(s \to s') \right] \ , \tag{2.3}$$

where $P(s,t) = \mathrm{Prob}\{S_t = s\}$ and K are the rates defined as

$$\begin{aligned} K(s \to s') = \Bigg\{ & \frac{\omega_f}{1-\gamma} \Big(1 - \gamma\, \sigma_i \frac{\sigma_{i+1} + \sigma_{i-1}}{2} \Big) e^{\beta\, h(-\sigma_i, \ell_i)}\, \delta_{\sigma'_i, -\sigma_i}\, \delta_{\ell'_i, \ell_i} \\ & + \omega_s\, e^{\beta\, h(\sigma_i, 1-\ell_i)}\, \delta_{\sigma'_i, \sigma_i}\, \delta_{\ell'_i, 1-\ell_i} \Bigg\} \prod_{j \neq i} \delta_{\sigma'_j, \sigma_j}\, \delta_{\ell'_j, \ell_j} \ , \end{aligned} \tag{2.4}$$

where the product of Kronecker δ indicates that the rates K only involve one protomer at a time. Each term in Eq. (2.4) is obtained from detailed balance up to multiplicative factors ω_f and ω_s: these constants account for typical time scales of the flipping and binding process, respectively. The constant γ in the spin-flip contribution is set by the strength of the coupling, $\gamma = \tanh(\beta J)$. For a system made of a one protomer (or for a single protomer in absence of interaction, $\gamma = 0$), according to Eq. (2.4), we define the constants k_a and k_i as the rates for activation and inactivation with $\ell = 0$,

$$k_a = \omega_f\, e^{-\beta \varepsilon_A} \ , \qquad \text{and} \qquad k_i = \omega_f\, e^{-\beta \varepsilon_I} \ ; \tag{2.5}$$

their counterparts for $\ell = 1$ are

$$k_a \frac{c}{K_d^A} \qquad \text{and} \qquad k_i \frac{c}{K_d^I} \ . \tag{2.6}$$

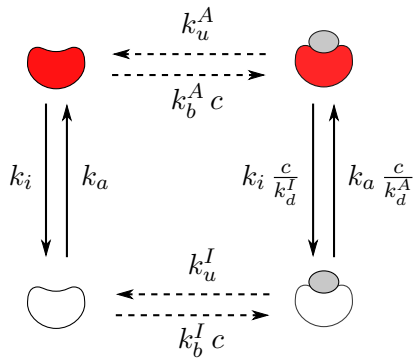


Fig. 2.3. **Single-protomer dynamics.** Contributions to the transition rates $K^{(i)}$ from the single-body term in the Hamiltonian (2.1): vertical arrows are labeled by the rates of activation/inactivation; horizontal arrows by the binding/unbinding ones.

The rates of binding and unbinding are respectively given by

$$c\, k_b^\alpha = \frac{c}{K_d^\alpha}\, \omega_s\, e^{-\beta\varepsilon_\alpha} \qquad \text{and} \qquad k_u^\alpha = \omega_s\, e^{-\beta\varepsilon_\alpha}\ , \tag{2.7}$$

when it is in the activity state α. The ratio between the rate constants k_u^α/k_b^α is the *dissociation constant* of the binding process, K_d^α:

$$K_d^\alpha = \frac{k_u^\alpha}{k_b^\alpha} = c_0\, e^{-\beta(\varepsilon_b^{(\alpha)}+\mu_0)}\ . \tag{2.8}$$

The dynamics of a single isolated protomer is depicted in Fig. 2.3. The ratios of the rate constants $k_{u,b}^\alpha$ and $k_{i,a}$ are determined by the equilibrium statistics, while their specific values affect the kinetics.

It is worth noticing that the binding/unbinding rates at one protomer only depend on the state of the protomer itself and no other protomer in the ring: this assumption of independent binding is typical of allosteric models.

Deriving an exact solution for the conditional probability $P(s,t|s_0,0)$ by directly attacking the Kolmogorov equation (2.3) is far from being an easy task. However, as experiments show [110], in the flagellar motor regulation mechanism of *E. coli* it is possible to identify a hierarchy of widely separated time scales. This opens up the possibility of operating a reduction of the set of states by gradually *integrating out/decimating* fast degrees of freedom, operating a *quasi-stationary* approximation: the time scale of the slow degrees of freedom is much longer than the time needed for the fast variables to relax to a stationary distribution; hence, the fast degrees of freedom enter the slow dynamics only through quantities averaged over such stationary distribution (conditioned to the state of the slow variables) [44–46]. The application of such techniques to the study of the allosteric regulation of the motor of *E. coli* will be the subject of the following sections. The approximation scheme is depicted in Fig. 2.4.

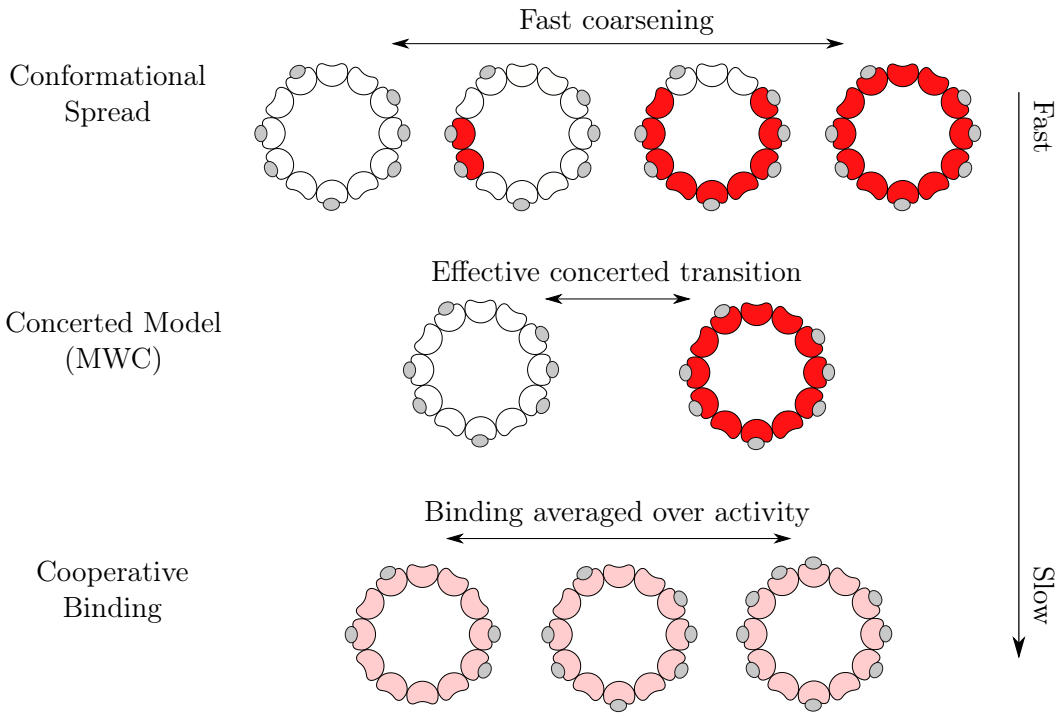


Fig. 2.4. **Time-scale separation in the Conformational Spread Model.** Graphic representation of the time-scale separation scheme. Short-lived transient states containing domain walls are decimated in a first time-scale separation, leading from the Conformational Spread to the MWC model, while the binding dynamics is kept frozen. Then, over the binding time scales the activity states are averaged out, resulting into a cooperative binding model.

2.2 From the Conformational Spread to the MWC model

In the present problem, the fastest degrees of freedom are associated with the spin-activity variables: the (concerted) conformational transition between CW and CCW state is much faster than the time scale for binding/unbinding of CheY-P, respectively occurring on typical times of $10^{-3}\,s$ and $10^{-1}\,s$. In the associated Glauber dynamics in Eq. (2.4), this can be encoded in the limit $\omega_f \gg \omega_b$.

Furthermore, it can be seen that the coarsening dynamics of the spin-activity variables occurs over time scales much shorter than the typical time interval between two successive nucleations of an activity domain, the latter setting the frequency of the switch from CW to CCW and vice versa, while the binding $\{\ell_i\}$ is fixed. This is due to the strong coupling between the neighbouring protomers, $\beta J \gg 1$, or equivalently, $\gamma \to 1$. In this limit, the transition rates away from the fully aligned configurations (all σ_i equal) are of order $\omega_f/(1-$

γ), while all other spin transitions are much slower, with typical rate $\omega_f \ll \omega_f/(1-\gamma)$.

The discussion of this latter time-scale separation is the subject matter of this section: it will be shown that the strong coupling limit amounts to considering the conformational spread model effectively equivalent to the Monod–Wyman–Changeux model, on the time scale of the switch. At the time scales typical of these fast processes, the binding state $\{\ell_i\}$ enters via a *quenched* external field term, playing a parametric role in determining the quasi-stationary distribution towards which the activity states relax. The slow binding dynamics will be discussed in the next section.

The role of the coupling

The ferromagnetic coupling in the conformational spread model is an essential ingredient which accounts for high sensitivity of the motor to the variation of concentration of CheY-P, due to the resulting cooperative response. The implementation of a large coupling J is suggested by the experimental determination of this high sensitivity, quantified by a Hill coefficient ~ 10. As pointed out in [111], though, the estimation of the Hill coefficient does not impose severe constraints on the parameters of the model, especially on J; in fact, the numerical simulations performed therein show that the sensitivity depends more strongly on the activation energy of the single protomer ($\varepsilon_{A,I}$) than on the cooperativity. However, combining the experimental knowledge of the Hill coefficient with the information about other quantities, such as the mean locked state time and the mean switch time, Ma et al. [111] were able to provide a very precise estimation of J, which is $\sim 4.5\, k_B T$. For such value of J the formation of domain walls is strongly disfavored. At equilibrium, in fact, the ratio between the probability of configurations with $2m$ domains and the probability of a coherent one can be estimated as (see Ref. [108])

$$\frac{P(2m)}{P(0)} \simeq \binom{N}{2m} \exp(-2m\beta J) \ , \tag{2.9}$$

where the binomial factor counts all possible ways of dividing N protomers into $2m$ domains; for $N = 30 \gg 1$, the limit $P(2m) \ll P(0)$ corresponds to

$$\beta J > \log\, N \sim 3.5 = \beta J_* \ , \tag{2.10}$$

satisfied by the estimate of J performed in [111]. The stationary equilibrium configuration, at fixed binding states $\{\ell_i\}$, is therefore concentrated only on the two states with all the protomers in the same state. From a dynamical point of view, this means that states with one or several domain walls are just short-lived transients between coherent states: as soon as a domain is nucleated inside a coherent configuration, it either immediately expands to invade the whole ring or is suddenly absorbed, typically much before another nucleation occurs.

Decimation of fast variables

To realize the fast "emptying" of configurations with several domain walls, it is necessary to analyse the structure of the transition rate matrix $K(s \to s')$, when the limits of the time-scale separation ($\omega_s \ll \omega_f \ll \omega_f/(1-\gamma)$) are concerned.

In the limit $\gamma \to 1$, in fact, the non-vanishing entries of the matrix K, at frozen binding $\{\ell_i\}$, are either of order $\omega_f/(1-\gamma)$, or of order ω_f: the latter rates (slow) are defined for transitions consisting in a nucleation of a domain, i.e. creation of pairs of domain walls, and are denoded by K_s; the former (fast) are defined for all other transitions, i.e. motion and destruction of domain walls, and are denoted by K_f. We can therefore write $K = K_f + K_s$, with

$$K_f(s \to s') = K(s \to s')\,(1 - \delta_{\sigma_1 \dots \sigma_N}) \sim \frac{\omega_f}{1-\gamma}\ , \tag{2.11}$$

and

$$K_s(s \to s') = K(s \to s')\,\delta_{\sigma_1 \dots \sigma_N} \sim \omega_f\ , \tag{2.12}$$

where K are defined in Eq. (2.4), and $\delta_{\sigma_1 \dots \sigma_N}$ indicates that the spin-activity variables in s have all the same value. One notices that the coherent configurations (all protomers active or inactive) are the only absorbing states of the fast process, since in such cases the entries of K_f vanish. The dynamics specified by K_f forbids the creation of pairs of domain walls and only allows translation or absorption of domain walls. As a result, the fast dynamics leads to one or the other coherent configuration with a typical rate $\sim \omega_f/(1-\gamma)$. As an explicative example, the case of $N = 4$ is depicted in Fig. 2.5. On a time scale set by $1/\omega_f$, the nucleation of an activity domain can occur. In the coherent activity configurations, the process involving the spin-activity variables has slow rates K_s. It is then possible to apply the standard techniques of time-scale separation [44–46], eliminating incoherent activity configurations from the dynamics at time scales comparable with $1/\omega_f$ or longer. The net effect of the fast coarsening dynamics is included in an effective way into rates, denoted by K_c, which provide the description of the dynamics at the nucleation time scale: a concerted transition between the two coherent configurations I (all protomers inactive, $\sigma_i = -1$) and A (all active, $\sigma_i = 1$), besides slow binding processes. In this model, the N-protomer complex can be in 2 different activity states, each of which present in 2^N binding configurations (2 for each protomer): therefore, the model contains 2×2^N states, and corresponds to the concerted allosteric model of Monod, Wyman and Changeux (MWC) [103, 104].

The structure of the state diagram of the MWC model with its rates K_c is depicted in Fig. 2.6. In App. A.1, the decimation procedure leading from the conformational spread to the MWC model, in the case of $N = 2$ has been worked out exactly. In general, the rate of a concerted switch from the activity state $\alpha = I$ (or A) to $\alpha' = A$ (or I) is

$$K_c(\alpha \to \alpha',\, \{\ell_i\}) = \sum_{j=1}^{N} K_s(\alpha \to \alpha^{(j)},\, \{\ell_i\})\, P_{abs}^{(j)}(\alpha')\ , \tag{2.13}$$

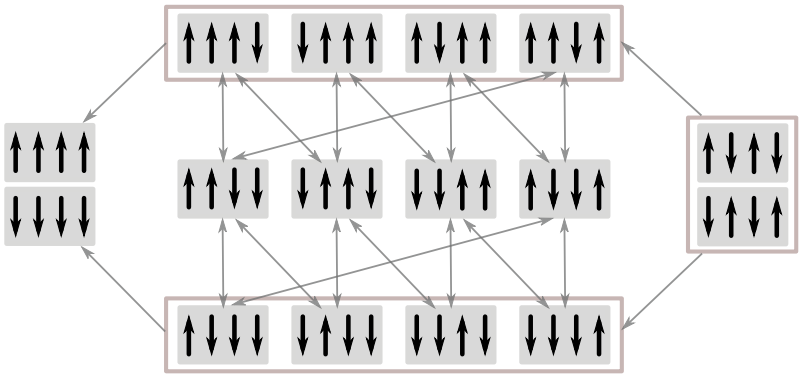


Fig. 2.5. **Fast coarsening dynamics.** Schematic representation of the fast rates of single-spin flipping K_f for a system of 4 protomers. Grey boxes correspond to the states; periodic boundary conditions are understood. Arrows are drawn between two states (or groups of states) for which K_f is non vanishing for some i. In particular: reversible transitions are allowed between states with equal number of domain walls; transitions to states with less domain walls are irreversible. Starting from any state, the dynamics leads to one of the coherent configurations in a time $\sim (1-\gamma)\,{\omega_f}^{-1}$ [see Eqs. (2.4) and (2.11)]; such states are the only two activity states in the MWC allosteric model.

where $\alpha^{(j)}$ denotes the state where all the spins but the j-th are in the state α, and $P_{abs}^{(j)}(\alpha')$ is the probability of absorption in the state α' conditioned to the initial state $\alpha^{(j)}$.

A direct analytic derivation of the rates K_c (or the probabilities $P_{abs}^{(j)}$) for a generic N-protomer ring can be extremely complicated. However, in the time-scale separation assumptions, the fast dynamics after the nucleation of an activity domain from a coherent state reaches one of its 2 absorbing states before another nucleation could possibly occur. This means that a calculation of the effective activity switching rates in the MWC model, does not require to include all the incoherent states, but only those with just two domain walls: the coarsening process can be seen as the expansion or contraction of the domain which has been nucleated. The nucleated domain can either expand until it invades the whole ring (complete switch), or be "absorbed" back (failed attempts).

Since the detailed balance is still respected by the rates in the decimated dynamics, all their pairwise ratios are fixed by the equilibrium distribution. Hence, since the equilibrium distribution of the MWC model is known from the Hamiltonian (2.1) (where the coupling part is just a constant term), it is sufficient to determine only *one* effective rate K_c exactly. In the case where $\ell_i = 0$ for all protomers, one is able to calculate the rate of switching from the I to the A state, by mapping the coarsening process into a simple birth and death process, the random variable being the size of the domain with active protomers (see App. A.1).

Regarding the binding process, the rate K_c is just the binding/unbinding contribution in the rates K, defined in Eq. (2.4).

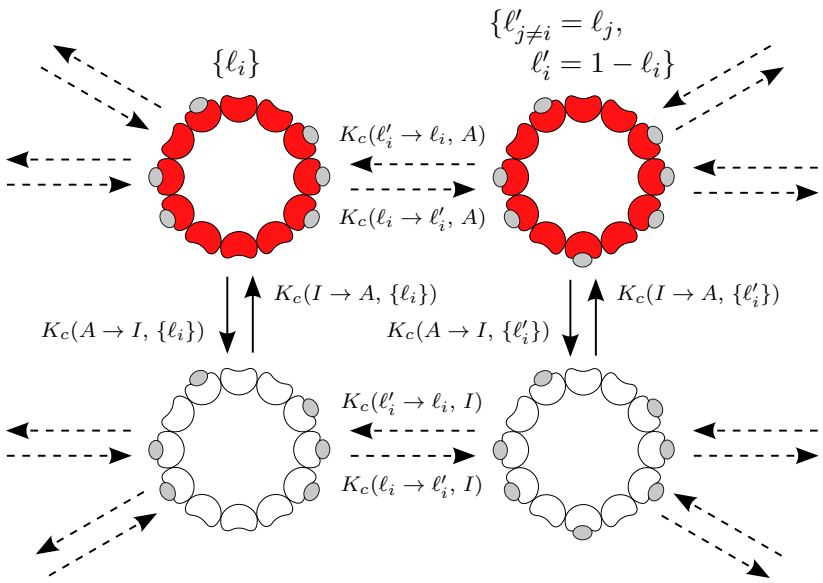


Fig. 2.6. **Dynamics of the MWC model.** The figure contains only a small portion of the model, corresponding to two possible binding states, $\{\ell_i\}$ and $\{\ell'_i\}$ (differing only by the occupation of the protomer at the bottom of the ring). The spin-activity variables of the ring of protomers are involved in a fast concerted transition (solid arrows), with rates K_c depending on the occupation $\{\ell_i\}$ in a highly non trivial way. The transitions between different binding configuration is slow (dashed arrows); only one of the possible binding/unbinding transitions is explicitly represented, while the others are symbolically indicated by unlabeled arrows.

Although the dynamics of the MWC model depends on the detailed binding configuration $\{\ell_i\}$ in a highly non-trivial way, the equilibrium distribution depends on the total occupancy $l = \sum \ell_i$ only,

$$P_{eq}(I,\, l) \;\; = \frac{\left(\frac{c}{K_d^I}\right)^l \binom{N}{l}}{\left(1+\frac{c}{K_d^I}\right)^N + L^{-1}\left(1+\frac{c}{K_d^A}\right)^N} \;, \tag{2.14}$$

$$P_{eq}(A,\, l) \;\; = \frac{L^{-1}\left(\frac{c}{K_d^A}\right)^l \binom{N}{l}}{\left(1+\frac{c}{K_d^I}\right)^N + L^{-1}\left(1+\frac{c}{K_d^A}\right)^N} \;, \tag{2.15}$$

where L is called *allosteric constant* of the N-protomer MWC molecule,

$$L = \left(\frac{k_i}{k_a}\right)^N = e^{\beta\,(\varepsilon_A - \varepsilon_I)\,N} \;. \tag{2.16}$$

There is an important comment to be made about the equilibrium distribution of the MWC model, in particular about the marginal probability for the active state, defined as the

activity of the MWC molecule,

$$P_{eq}(A) \quad = \sum_{l=0}^{N} P_{eq}(l,\, A) = \frac{1}{1+L\left(\frac{K_d^A}{K_d^I}\right)^N \left(\frac{c+K_d^I}{c+K_d^A}\right)^N} \ . \tag{2.17}$$

In our problem, this corresponds to the CW *bias* of the flagellar motor, which is a function of the CheY-P concentration c. In order for the MWC molecule to be a good allosteric switch, it needs to be almost certainly active for high enough concentration c and, vice versa, inactive when c is low:

$$P_{eq}(A) \sim \begin{cases} [1+L]^{-1} \to 0 \quad \text{for} \quad c \to 0 \\ \left[1 + L^{-1}\left(\dfrac{K_d^A}{K_d^I}\right)^N\right]^{-1} \to 1 \quad \text{for} \quad c \to \infty \end{cases} .$$

These limits impose the following constraints:

$$1 \ll L \ll \left(\frac{K_d^I}{K_d^A}\right)^N \ . \tag{2.18}$$

Since the single protomer has higher ligand affinity (smaller dissociation constant K_d) when

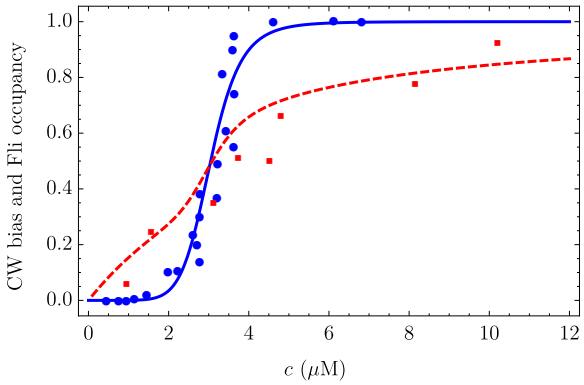


Fig. 2.7. **Activity and mean Fli occupancy at equilibrium.** Analytic results for $P_{eq}(A)$ (solid blue line) and mean Fli relative occupancy $\langle l \rangle / N$ (dashed red line) as a function of the CheY-P concentration, c. The dots are the experimental results presented in Ref.[18]. Here we chose the dissociation constants to be $K_d^A = 1.84\,\mu M$ and $K_d^I = 5.52\,\mu M$, respectively, while the allosteric constant has been set to be $L = 10^7$. The plot shows the effect of allostery: the activity response is much more sensitive than the binding to changes of concentration of CheY-P.

in the active state than in the inactive one, it is required that $K_d^A < K_d^I$. From this last relation one realizes that the number of protomers sets the *sensitivity* of the switch: since $K_d^I > K_d^A$, the larger N, the larger the r.h.s of the condition given by Eq. (2.18). Incidentally, depending on environmental stimuli *E. coli* is able to regulate the number of protomers of the flagellar motor [113–115].

2.3 From MWC to a cooperative binding model

As we already said at the beginning of Sec. 2.2, the binding is much slower than the switching dynamics. We can assume that on the time scale at which one of the protomers binds or releases a CheY-P (set by a typical time $\tau_b \sim 10^{-1}\, s$), the activity of the ring safely reaches the equilibrium configuration, conditioned to the (quasi-static) value of l:

$$P_{eq}(I|l) \;\; = \frac{P_{eq}(l,I)}{P_{eq}(l)} = \frac{P_{eq}(l,I)}{P_{eq}(l,I)+P_{eq}(l,A)} = \frac{1}{1+L^{-1}\left(\frac{K_d^I}{K_d^A}\right)^l} \;, \tag{2.19}$$

$$P_{eq}(A|l) \;\; = \frac{P_{eq}(l,A)}{P_{eq}(l)} = \frac{P_{eq}(l,A)}{P_{eq}(l,I)+P_{eq}(l,A)} = \frac{1}{1+L\left(\frac{K_d^A}{K_d^I}\right)^l} \;. \tag{2.20}$$

Then, on time scales comparable to (or larger than) τ_b, the relevant dynamics is essentially the slow binding/unbinding one, while the fast activation/inactivation dynamics is *averaged* over the equilibrium conditional probabilities in Eqs. (2.19) and (2.20), to give the effective rates $\bar{K}$ for the variable l:

$$\bar{K}(l \to l') = \sum_{\alpha \in \{I,A\}} P_{eq}(\alpha|l)\, K(l \to l',\, \alpha \to \alpha) \;. \tag{2.21}$$

This averaging procedure is guaranteed to give an effective dynamics of the slow variables which still enjoys the Markov property. The effective binding/unbinding rates of the whole allosteric complex are, in fact,

$$\begin{aligned} \bar{K}(l \to l+1) &= (N-l)\, c\, \bar{k}_b^{(l)} \equiv b_l \;, \\ \bar{K}(l \to l-1) &= l\, \bar{k}_u^{(l)} \equiv u_l \;, \end{aligned} \tag{2.22}$$

where

$$\bar{k}_{b,u}^{(l)} = \frac{k_{b,u}^A}{1 + L\left(\frac{K_d^A}{K_d^I}\right)^l} + \frac{k_{b,u}^I}{1 + L^{-1}\left(\frac{K_d^I}{K_d^A}\right)^l} \;, \tag{2.23}$$

depending only on the current value of l.

A comment about the range of validity of this result is in order: for the time-scale separation to hold, the rates $\bar{K}$ must be small enough to guarantee that the binding/unbinding process is still much slower than the activation/inactivation. In particular, this implies that the concentration of ligands in the environment c cannot be exceedingly large; then, in the time-scale separation approximation, we keep ourselves far from this regime.

The reduced system is also a Markov process, governed by the following master equation:

$$\partial_t P_t(l) = b_{l-1}\, P_t(l-1) + u_{l+1}\, P_t(l+1) - [b_l + u_l]\; P_t(l) \;. \tag{2.24}$$

The process hence obtained is a *birth-and-death process*, restricted on the set of integers between $l = 0$ and $l = N$. These extremes are reflecting boundary states. This dynamics eventually leads to the equilibrium state $P_{eq}(l)$, easily calculated by marginalizing the joint probability distribution $P_{eq}(\alpha, l)$, given in Eqs. (2.14) and (2.15):

$$P_{eq}(l) = P_{eq}(A, l) + P_{eq}(I, l) \ . \tag{2.25}$$

Albeit much reduced, this model still encodes a lot of information about the actual dynamics of the switch. Indeed, the flagellar motor switch is triggered by the number of ligands bound to the allosteric complex. In the next section we present some numerical analysis of the dynamical properties of the effective cooperative binding model obtained above.

2.4 Dynamics of the effective cooperative binding model

In this Section we analyze the case of a motor constituted by $N = 30$ Fli molecules. The allosteric constant L and the dissociation constants K_d^A and K_d^I have been chosen consistently with Ref. [18] and works cited therein: $L = 10^7$, $K_d^A = 1.84\ \mu M$ and $K_d^I = 5.52\ \mu M$; these values provide a qualitatively good fit of the activity as a function of the CheY-P concentration c (see Fig. 2.7). With this choice of the parameters, we can easily see that the bound in Eq. (2.18) is safely satisfied, so that the motor displays a switch behaviour, manifest in the response curve in Fig. 2.7. One also notices that the motor *operates* within a range of concentration c roughly between K_d^A and K_d^I. The maximum sensitivity is found around a value c_*, which correspond to a CheY-P concentration such that the CW (active) and the CCW (inactive) states occur with equal probabilities at equilibrium.

As already remarked above, the specific values of the rate constants $k_{b,u}^{\alpha}$ are irrelevant for the equilibrium properties of the model, but they determine the characteristic time scale for the motor switch. Out of these four constants, only two are actually independent, since we already defined their ratios $K_d^A = k_u^A/k_b^A$ and analogously $K_d^I = k_u^I/k_b^I$. Then, the dynamics of the cooperative binding model can be specified only the parameters k_b^I and k_b^A; the qualitative behaviour is determined only by their ratio, while their specific values gives information about the overall time scale (of the binding process). Here, we set $k_b^A = 2.8\, s^{-1}$ and $k_b^I = 5.0\, s^{-1}$, consistently with those recommended by Bai et al. [110].

As previously discussed, the cooperative binding model obtained so far must provide an accurate description of the statistics of slow observables, namely those which vary over time scales typical of the binding process or longer. From experimental results, it is clear that the mean–locked state time (i.e. the time in which the motor stays in a certain rotational state between two consecutive switches) is such an observable; we show, indeed, that the cooperative binding model captures very well its statistics.

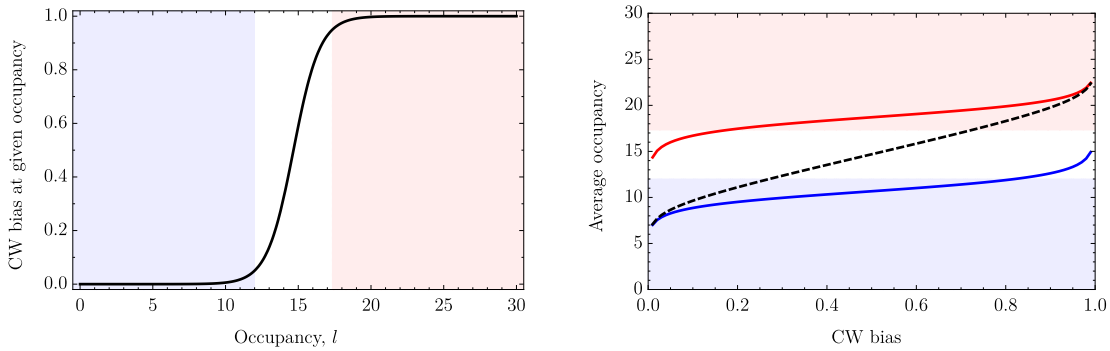


Fig. 2.8. **Motor switch ruled by the cooperative binding.** On the left, the probability of the CW state conditioned on the Fli occupancy; for values of l fixed in the shaded regions, the motor is in the CW or CCW state with 95% probability. On the right, average occupancy as a function of the CW bias: the dashed line corresponds to the unconditional average (see also Fig. 2.7) while the solid lines represent the averages conditioned to the CW state (red), $\bar{l}_A$, and CCW state (blue), $\bar{l}_I$. The values of $\bar{l}_A$ and $\bar{l}_I$ lie in the respective 95%-confidence intervals, with a CW bias between $\simeq 0.1$ and $\simeq 0.9$. The locked–state time can be interpreted as the first passage time between $\bar{l}_A$ and $\bar{l}_I$ in the cooperative binding model.

Let us denote by $\bar{l}_I$ and $\bar{l}_A$ the averages of the occupancy l conditioned, respectively, to the inactive state (CCW) and active state (CW). One can see that the probability of the CW state is very close to unity if the Fli occupancy is conditioned to $\bar{l}_A$, and almost vanishing when conditioned to $\bar{l}_I$ (Fig. 2.8). Therefore, since the fast activity variables are slaved to the slow binding ones, we can state that a good measure of the locked–state time is the first passage time between $\bar{l}_A$ and $\bar{l}_I$.

Let us then study the first arrival time at $\bar{l}$ from a generic state k. If we denote by f_k the probability density function of this time interval, its moment generating function

$$g_k(\lambda) = \int_0^\infty d\tau \, e^{-\lambda \tau} \, f_k(\tau) \ , \tag{2.26}$$

satisfies

$$\sum_l g_l(\lambda) \, (M_{l,k} - \lambda \, \delta_{l,k}) = -\delta_{k,\bar{l}} \ , \tag{2.27}$$

where $\mathbf{M}$ is the generator of the process in which absorbing conditions have been put at $\bar{l}$. From Eq. (2.27), we can derive the equation for the mean first passage time at $\bar{l}$, using $\langle \tau_k \rangle = g_k'(\lambda)|_{\lambda=0}$:

$$\sum_l \langle \tau_l \rangle \, M_{l,k} = -1 \ . \tag{2.28}$$

Exact results are obtained by inverting Eq. (2.28) and are shown in Fig. 2.9, with $k = \bar{l}_i$ and $\bar{l} = \bar{l}_a$, and vice versa, for several values of the bias $P_{eq}(A)$.

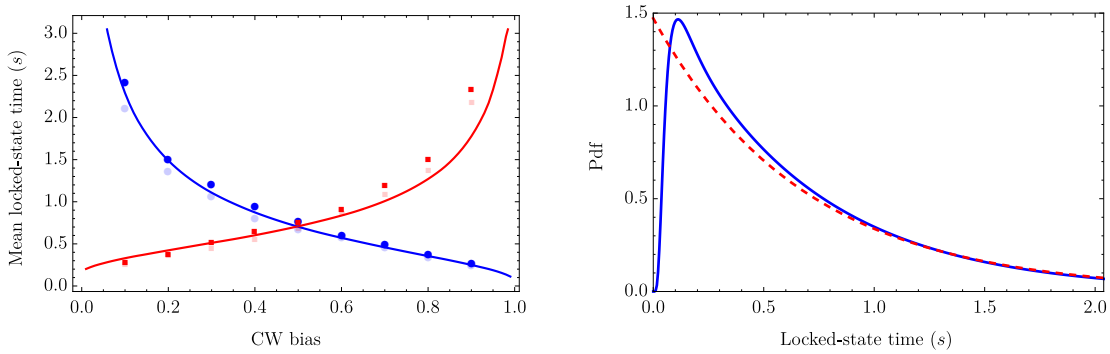


Fig. 2.9. **Statistics of the locked–state time.** On the left, average locked–state time in the CCW state (blue line and circles) and CW state (red line and squares): the points are experimental (lighter color) and numerical (darker color) results from [110]; the lines are the theoretical results from the cooperative binding model, estimated as the first passage time at $\bar{l}_a$ conditioned to $\bar{l}_i$ at $t = 0$ (mean CCW time, blue), and vice versa (mean CW time, red). On the right, probability distribution of the first passage time at $\bar{l}_a$ from $\bar{l}_i$ in the unbiased case: the solid blue line is the exact result found as the inverse Laplace transform of the generating function obtained by solving Eq. (2.27); the dashed red line is the exponential distribution with the same average. See for comparison the experimental and numerical results in Refs.[110, 111].

We also extract the probability density by solving Eq. (2.27) and numerically performing the inverse Laplace transform. The resulting distribution is very similar to the experimental and numerical results presented in [110] and [111], confirming that the effective cooperative binding model gives an excellent description of the motor kinetics.

2.5 Discussion

In this work we pursued an analytic approach to the description of the dynamics of the conformational spread model, a phenomenological model that reproduces the allosteric regulation of the flagellar motor in *E. coli*. Our analysis was based on the existence of a hierarchy of widely separated time scales in the biochemistry of the *E. coli*'s motor. Namely, over the time-scale of transitions between CW and CCW states in the Fli molecules (protomers, constituents of the flagellar motor) incoherent states are very short-lived, and only coherent states of activity are sufficiently long-lived. In such a limit we have reduced the conformational spread to the well known Monod–Wyman–Changeux model. For a motor with $N = 30$ protomers, this approximation amounts to reducing the number of states in the model from $4^N \sim 10^{18}$ to $2^{N+1} \sim 10^9$.

Moreover, the binding of CheY-P to the Fli molecules occurs much less frequently than the switch from the completely active to inactive state, allowing to average out the fast activity states under quasi-stationary Fli occupancy (number of CheY-P bound to the motor). This

allowed to reduce the number of states further and get a cooperative binding model containing only $N + 1 = 31$ states, the possible values of the overall occupancy. The resulting Markov process is a birth-and-death process which can be studied semi-analytically, with virtually no computational cost.

This effective model for the slow variables is able to capture the dynamics of observables varying on time scales of $10^{-1}\,s$ or longer. Two of such observables are the CW and CCW locked-state time, which correspond to the duration of tumbles and runs, respectively, with time scales typically of the order of seconds. We showed that our model reproduces the statistics of the locked state time and is in extremely good quantitative agreement with experimental measurements. This is particularly important, because it quantitatively connects the hardware mechanisms to the way *choices* are made by *E. coli* in order to navigate, i.e. to run or to tumble.

In perspective, our approach could be extended to include the even slower kinetics of motor remodeling. Indeed, it is known that over time scales much longer than the binding times (typically minutes), *E. coli* is also able to modify the flagellar motors by changing the number of Fli molecules, i.e. the protomers [113, 114]. This mechanism provides an adaptation layer at the output and restores the sensitivity of the motor when CheY-P concentration are kept off the dynamic range for a long time [115].

3

Active particles with alignment costs

Animals often engage in tasks that involve the cooperation of all the individuals in the group [64, 65] It is arguably the case that, for instance, some species of birds gain some advantage in navigating in flocks. The same must hold true for the schools of fish, or herds of sheep. Undoubtedly, the group offers protection from the attacks of predators [49]. Perhaps, for fish, the benefit is of hydrodynamic origin, as animals experience less drag forces by swimming in the wake of their companions [50, 116].

Collective phenomena akin to flocking are ubiquitous in the animal and microbial world, at both large and microscopic scales. Coherent motions are observed in bacterial suspensions [117] as well as in tissues during developmental processes [118].

The mechanisms used to coordinate and make decisions constitute a fascinating subject of investigation and have been at the center of many recent studies. Most prominent among theoretical approaches to flocking are agent-based models, a prototypical example of which is the celebrated Vicsek model [119]. The appealing feature of these models is the minimalistic nature of the dynamics and the interactions among the agents, which nevertheless result in non-trivial collective motions [120, 121]. However, despite the simple behavioural rules, it is generally impossible to predict the dynamics of agent-based models, and their investigation is limited to numerical approaches. Alternative approaches are inspired by the similarity of collective phenomena with the physics of gases and fluids. For example, hydrodynamic descriptions prove to be useful in capturing large-scale features of group dynamics [122, 123].

In some instances, analytical tools borrowed from the tradition of statistical physics and critical phenomena have helped to gain insights about the nature of the interactions and their roles in collective decision-making [124, 125].

In the context of animal behaviour, it is interesting to understand whether non-trivial collective dynamics can emerge as a consequence of an underlying optimality principle. Behavioural rules are then interpreted as the strategies that the individuals in a group can implement at an algorithmic level in order to solve an optimization task.

In this Chapter, in the same spirit as in Ch. 1, we formulate a collective task in the framework of optimal stochastic control. The goal of the agents is to navigate the environment while keeping their orientation close to that of their neighbours, and their strategy (the control) is included as a deterministic self-driving force. We derive the exact equations for the optimal control, and the resulting dynamics is then analysed in the overdamped limit. The resulting driven-diffusive process features a non-trivial relation between the force and the noise. While the case of an arbitrary number of agents still escapes, to our knowledge, a full-fledged analytical treatment, we were able to show that for two agents in a compact translational invariant domain (torus), this dynamics exhibits aggregation. This result corroborates numerical findings in the Vicsek model, and is worth investigating further.

3.1 Optimally controlled Langevin–Kramers dynamics

Let us study a system constituted of N agents moving in space with velocities evolving according to the (controlled) Ornstein–Uhlenbeck process*,

$$\begin{aligned} dX_i &= V_i\,dt\ , \\ m\,dV_i &= f_i(\bar{X},\bar{V})\,dt - \gamma V_i\,dt + \sqrt{2k_BT\gamma}\,dB_i\ . \end{aligned} \tag{3.1}$$

The variable X_i defines the position of agent i, moving with an instantaneous velocity V_i which changes in time under the effect of a deterministic force f_i (control), a drag force with friction coefficient γ and a stochastic force arising from the thermal fluctuations of the medium at temperature T. The constants k_B and m are, respectively, the Boltzmann constant and the mass of the agents. If one fixes a unit of length L_0 and a reference temperature T_0, Eq. (3.1) can be written in terms of dimensionless variables by mapping $X \mapsto L_0\,X$, $T \mapsto T_0\,T$, $V \mapsto \sqrt{k_BT_0/m}V$, $t \mapsto t_0\,t = (L_0/V_0)t$, $t \mapsto \epsilon^{-1}(k_BT/L_0)f$, $\gamma \mapsto \epsilon(m/t_0)\gamma$, hence obtaining

$$\begin{aligned} d\bar{X} &= \bar{V}\,dt \\ d\bar{V} &= \epsilon^{-1}\bar{f}(\bar{X},\bar{V})\,dt - \epsilon^{-1}\gamma\bar{V}\,dt + \epsilon^{-1/2}\sqrt{2D_v}\,d\bar{W} \end{aligned} \tag{3.2}$$

*The notation $\bar{\cdot}$ indicates Nd-tuples, where N is the number of agents and d is the dimension of the space they move into. For instance, $\bar{v} = (v_{1,1}\dots v_{1,d}, v_{2,1}\dots v_{N,1}\dots v_{N,d})$, regarded as an element of $\mathbb{R}^N \otimes \mathbb{R}^d$ (the first index is the label for the particle, and the second of the spatial component. If only one index is specified, e.g. v_i, that indicates the d-tuple pertaining to particle i.

where $D_v = T\gamma$ (in the new units) and $\tilde{W}$ is a standard Gaussian white noise. The quantity ϵ is a dimensionless parameter which sets the relative timescale between the dynamics of the (adimensional) positions X and velocities V. When $\epsilon \ll 1$, the dynamics of V is much faster than that of X: this is the *overdamped* limit of the Langevin–Kramers dynamics, corresponding to the case of large friction coefficient γ and/or small mass m in Eq. (3.1).

The cost per unit time (rate) paid by agent i is

$$c_i = \epsilon^{-1}\frac{\eta}{2} f_i(\bar{X},\bar{V})^2 + \frac{\epsilon^{-1}}{4}\sum_j k\big(X_i - X_j\big)\,\big|V_i - V_j\big|^2 + \epsilon\, q_i(\bar{X})\ . \tag{3.3}$$

The first term is the cost for control, originating as the Kullback–Leibler divergence from the uncontrolled to the controlled Langevin–Kramers path measures [73, 74]. The second term is a cost for the velocity variables, favouring alignment of their directions. The third and last term, is a generic cost for the spatial variables: it can account for any kind of positional cost (e.g. a confinement) and/or interaction "potential" (e.g. penalizing collisions). Notice that, in the limit $\epsilon \ll 1$, the alignment interaction is at the leading order in the cost, and that the function k, defined to be a positive, monotonically decreasing function of the distance between agents, varies over length-scales of order 1, and is approximately constant for arbitrarily small distances, as depicted in Fig. 3.1. The same dependence on the spatial variables is assumed for the positional cost q. This particular scaling form of the costs and of the variables upon taking the overdamped limit has been chosen so that the *fast* part of the control provides local alignment of the velocity variables, while a *slow* part enters the effective dynamics of the large-scale variables. We rigorously show this in the next sections.

The optimal control $\bar{f}^*$, is the one that minimizes the expectation value of the sum of the

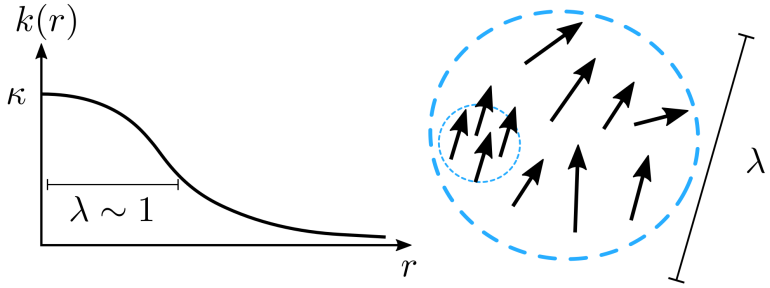


Fig. 3.1. **Alignment interaction**. The interaction kernel only depends on the distance between agents, and it possesses a typical length $\lambda \sim 1$. It is maximum at short distances, and it decreases monotonically for increasing distance. The optimally controlled dynamics is expected to have spatial domains of size λ in which velocities are positively correlated.

cumulated costs over time

$$\mathcal{F}_t = \Big\langle \sum_i \int_t^\infty dt\, c_i \Big\rangle = \int_t^\infty dt \int d^N x\, d^N v \left[\epsilon^{-1} \frac{\eta}{2} \bar{f}(\bar{x},\bar{v})^2 + \frac{\epsilon^{-1}}{4} \sum_{i,j} k\big(x_i - x_j\big) \left| v_i - v_j \right|^2 + \epsilon\, q(\bar{x}) \right] P(\bar{x},\bar{v},t) \ , \quad (3.4)$$

where P is the N-particle probability density function associated to the dynamics in Eq. (3.2), and where $q = \sum_i q_i$.

Optimal control equations

The constrained minimization of the cost functional (3.4) can be performed by invoking the Pontryagin minimum principle [22], and translated into the unconstrained minimization of

$$\mathcal{L}[\bar{f}, P, \Phi, \mu] = \mathcal{F}_t + \int_t^\infty dt \int d^N x\, d^N v\, \Phi(\bar{x},\bar{v},t) \Big[\partial_t P + \bar{v}^\top \nabla_{\bar{x}} P + \epsilon^{-1} \nabla_{\bar{v}}^\top \Big(\bar{f}\, P - \gamma\, \bar{v}\, P \Big) - \epsilon^{-1} D_v\, \nabla_{\bar{v}}^2 P \Big] + \mu \Big(1 - \int d^N x\, d^N v P \Big) \ . \quad (3.5)$$

The saddle point condition with respect to Φ and μ trivially yields the Fokker–Planck equation for P and its normalization while the ones with respect to $\bar{f}$ and the density function P give

$$\bar{f}^* = \frac{1}{\eta} \nabla_{\bar{v}} \Phi \quad (3.6)$$

and

$$\partial_t \Phi + \bar{v}^\top \nabla_{\bar{x}} \Phi + \epsilon^{-1} \left[D_v \nabla_{\bar{v}}^2 \Phi + \frac{1}{2\eta} \left| \nabla_{\bar{v}} \Phi \right|^2 - \gamma\, \bar{v}^\top \nabla_{\bar{v}} \Phi \right] = \epsilon^{-1} \frac{1}{8 D_v \eta} \sum_{i,j} k(x_i - x_j) |v_i - v_j|^2 + \epsilon\, q(\bar{x}) - \mu \ . \quad (3.7)$$

One can check that the function Φ is the value function, defined as (minus) the expected cost-to-go from a given initial condition in the position-velocity space:

$$\Phi(\bar{x},\bar{v},t) = -\Big\langle \sum_i \int_t^T dt'\, c_i \,\Big|\, \bar{X}_t = \bar{x},\, \bar{V}_t = \bar{v} \Big\rangle \ . \quad (3.8)$$

Through the Hopf–Cole transformation $\Phi = 2 D_v \eta \log Z$, the system reads

$$\bar{f}^* = 2 D_v \nabla_{\bar{v}} \log Z \quad (3.9)$$

with

$$\partial_t Z + \bar{v}^\top \nabla_{\bar{x}} Z + \epsilon^{-1}\Big[D_v \nabla^2_{\bar{v}} - \gamma\, \bar{v}^\top \nabla_{\bar{v}}\Big] Z \\ = \left(\epsilon^{-1}\frac{1}{8D_v\eta}\sum_{i,j} k(x_i - x_j)|v_i - v_j|^2 + \epsilon\frac{q(\bar{x})}{2D_v\eta} - \frac{\mu}{2D_v\eta}\right) Z\ . \quad (3.10)$$

The function Z is the so-called desirability. Notice that the optimal control $\bar{f}^*$ is a gradient in velocity space of a single function of all positions and velocities of the N agents.

3.2 Multi-scale analysis in the overdamped limit

When the friction is high, or alternatively, the mass of the agents is very small, inertia plays a very weak role compared to the stochastic forces due to thermal fluctuations. This situation is recovered, in the adimensional version of the Langevin–Kramers dynamics in Eq. (3.2), in the limit $\epsilon \ll 1$. In this case the dynamics of the velocities is much faster –by a factor ϵ^{-1}– than the dynamics of the spatial variables.

When a system is characterized by widely separated length and time scales, it is convenient to introduce the auxiliary variables that "live" at each characteristic scale [44]. If we fix a reference time scale $O(1)$, processes with a time resolution $O(\epsilon)$ is suitably described by the auxiliary variable $t_f = \epsilon^{-1}t$; similarly, processes varying on much longer time scales, e.g. $O(\epsilon^{-1})$, require the more natural time variable $t_s = \epsilon\, t$. Time derivatives are replaced, accordingly, as $\partial_t \mapsto \epsilon^{-1}\partial_{t_f} + \partial_t + \epsilon\partial_{t_s}$. The introduction of the auxiliary variables is just a clever mathematical trick to change the unit with which time is measured.

In the system discussed in this Chapter, the velocity variables have a well defined time scale which is $O(\epsilon)$, and the relevant time variable is t_f. Spatial variables, instead, live at order 1 or higher. However, we shall see that it is not necessary to explicitly introduce variables at length-scales $O(\epsilon^{-1})$ as only $\bar{x}$ is relevant in this case.

In this section we derive the effective equations of the slow dynamics. The rationale of this calculation is the following. First one finds the stationary state of the fast process, in the assumption that the other variables are slowly changing over this time scale and have a smooth behaviour at the associated typical length scales. Then the average dynamics of the slow variables $\bar{x}$ is computed by integrating out the fast variables $\bar{v}$. This is the so-called *averaging* procedure. It can be regarded as a generalization of the law of large numbers for stochastic processes, in which the value of the process at a given time is replaced by its ensemble average. The presence of a hierarchy of widely separate time scales allows one to replace the ensemble average by the time average over an interval $t \pm \tau$, with $\epsilon \ll \tau \ll 1$. Finally, we study the fluctuations of the averaged process, occurring at the time scales of order ϵ^{-1}. This step is referred to as *homogenization*. It is equivalent to the central limit

theorem for stochastic processes, accounting for the fluctuations of the mean around the average in the Gaussian approximation.

Here we deal with an optimization problem at the equilibrium state, i.e. in the average setup [74]. Therefore, the time derivatives will be omitted. By introducing the auxiliary variables, the (tilted) generator of the full process, L, is expressed as an expansion in the time-scale separation parameter ϵ, and the HJB equation for the desirability Z reads

$$\big(L+\tilde{\mu}\big)\,Z \equiv \big(\epsilon^{-1}M + L_0 + \epsilon\,L_1 + \tilde{\mu}\big)Z = 0\ , \tag{3.11}$$

where

$$M = D_v\nabla^2_{\bar{v}} - \gamma\,\bar{v}^{\top}\nabla_{\bar{v}} - \frac{1}{8D_v\eta}\sum_{i,j} k\big(x_i - x_j\big)|v_i - v_j|^2\ ,$$

$$L_0 = \bar{v}^{\top}\,\nabla_{\bar{x}}\ ,$$

$$L_1 = -\frac{1}{2D_v\eta}\,q(\bar{x})\ ,$$

which are all operators of order 1. The leading term in Eq. (3.11) is the one involving the fastest-varying variables, i.e. the velocities. The sub-leading terms, give a contribution of order ϵ and ϵ^2 with respect to the leading one. Therefore, it is sensible to look for a solution of Eq. (3.11) in the form of an expansion,

$$Z = Z^{(0)}(\bar{v},\,\bar{x}) + \epsilon\,Z^{(1)}(\bar{x}) + \epsilon^2\,Z^{(2)}(\bar{x})\ldots\ , \tag{3.12}$$

in which all the functions $Z^{(l)}$ are of order 1. The leading contribution is regarded as a function of the fastest variables $\bar{v}$ only, parametrically depending on the slower variables $\bar{x}$. Similarly, the sub-leading terms only depend on the variables changing on longer scales. An equivalent expansion is consistently assumed for the eigenvalue $\tilde{\mu}$,

$$\tilde{\mu} = \mu_f(\bar{x}) + \epsilon\,\mu^{(0)} + \epsilon^{(2)}\mu^{(1)} + \ldots\ . \tag{3.13}$$

The leading and subleading terms in the expansion depend on the variables "living" at slower timescales. Indeed, they correspond to the largest eigenvalues of the tilted generators at the corresponding orders, and are calculated conditioned on the slower variables. The dynamics at every order is *slaved* to the one occurring at slower timescales.

The ansatz (3.12) and (3.13) are then replaced in Eq. (3.11), which provides an equation to be solved term by term in the ϵ-expansion.

Order ϵ^{-1}, equilibration of the fast dynamics. The equation at the leading order is the Bellman equation for the optimal control of the Ornstein–Uhlenbeck process on the

velocity variables, in which the pairwise alignment cost depends parametrically on the relative distances. The particular case in which the alignment cost is independent on the positions of the agents is described in detail in App. A.2. If all the agents lie within a region of linear size $\ll 1$, hence much shorter than the scale over which the interaction kernel k varies substantially, then the velocities of the agents relax to *exactly* the stationary state derived in App. A.2. However, we assume that the density of agents is spread over large scales, and the interaction will discern among agents close to or far from each other.

The leading contribution to the desirability, $Z^{(0)}$, is the right eigenvector corresponding to the largest eigenvalue of the operator M,

$$M\, Z^{(0)} + \mu_f\, Z^{(0)} = 0 \ . \tag{3.14}$$

The fast generator M features a transport term in the velocity variables which makes it non-Hermitian. However, via the transformation $Z^{(0)} = \exp(\frac{\gamma}{4D_v}\bar{v}^\top \bar{v})\,\Psi$, Eq. (3.14) is cast into the smallest eigenvalue problem of a Hermitian operator H,[†]

$$H\,\Psi = \Big(\frac{\gamma}{2} N d + \mu_f\Big)\,\Psi \ , \qquad H = -D_v \nabla^2_{\bar{v}} + \frac{\gamma^2}{4D_v}\bar{v}^\top \big(Q(\bar{x}) \otimes \mathbb{I}_d\big)\bar{v} \ , \tag{3.15}$$

where $Q(\bar{x})$ is the $N \times N$ matrix with entries

$$Q_{i,j} = \begin{cases} 1 + \frac{N-1}{\gamma^2 \eta}\,\kappa & \text{if } i = j \ , \\ -\frac{1}{\gamma^2\eta}\,k(x_i - x_j) & \text{if } i \neq j \ . \end{cases}$$

in which κ is the value of the interaction kernel k at very short distances, $\kappa = \lim_{x\to 0} k(x)$. We recognize H to be the Hamiltonian operator for a (Nd)-dimensional harmonic oscillator in the velocity variables with potential defined by the quadratic form $Q(\bar{x}) \otimes \mathbb{I}_d$. Then, the HJB equation (3.15) is equivalent to the ground-state problem of this harmonic oscillator, given by the Gaussian function

$$Z^{(0)}(\bar{v},\, \bar{x}) = z(\bar{x})\, \exp\Big\{ -\frac{\gamma}{4D_v}\bar{v}^\top \big(\big(Q^{1/2}(\bar{x}) - \mathbb{I}_N\big) \otimes \mathbb{I}_d\big)\bar{v}\Big\} \ , \tag{3.16}$$

where z is an arbitrary function of the slow variables only. For simplicity of notation, we do not explicitly indicate the dependence of Q on the slow variables $\bar{x}$; similarly, we denote $k_{i,j} = k(x_i - x_j)$.

[†]If A is an $N \times N$ matrix and B a $d \times d$ one, the matrix expressed as the Kronecker product $A \otimes B$ acts on a vector $\bar{v}$ as

$$\big[(A \otimes B)\bar{v}\big]_{i,\alpha} \equiv \sum_{j,\beta} A_{i,j}\, B_{\alpha,\beta} v_{j,\beta} \ .$$

The derivation of the average process requires calculating the corresponding left eigenvector ℓ of the generator of the fast dynamics M,

$$M^\dagger\,\ell = D_v\nabla^2_{\bar v}\ell + \gamma\,\nabla^\top_{\bar v}\big(\ell\,\bar v\big) - \frac{1}{8D_v\eta}\sum_{i,j}k_{i,j}|v_i - v_j|^2\,\ell = -\mu_f\,\ell\ . \tag{3.17}$$

By expressing $\ell = \exp(-\frac{\gamma}{4D_v}\bar v^\top\bar v)\,\tilde\Psi$, one finds that $\tilde\Psi$ satisfies the same eigenvalue equation as Ψ. Therefore, the solution $\tilde\Psi$ is the same as Ψ and

$$\ell(\bar v,\,\bar x) \propto\ \exp\left\{-\frac{\gamma}{4D_v}\bar v^\top\big(\big(Q^{1/2}+\mathbb{I}_N\big)\otimes\mathbb{I}_d\big)\bar v\right\}\ . \tag{3.18}$$

The proportionality constant is conveniently chosen such that

$$\begin{aligned}\ell\,Z^{(0)} = \tilde\Psi\,\Psi\,z = w(\bar v|\bar x)\,z\\ = \frac{\big(\det Q^{1/2}\big)^{d/2}}{\big(2\pi D_v/\gamma\big)^{Nd/2}}\,\exp\left\{-\frac{1}{2D_v/\gamma}\sum_{i,j}\bar v^\top\left(Q^{1/2}\otimes\mathbb{I}_d\right)\bar v\right\}z(\bar x)\end{aligned} \tag{3.19}$$

i.e. the product of left and right eigenvectors equals the corresponding local (Maxwellian) equilibrium distribution for the velocity variables, $w(v)$, with $\int d\bar v\,w(\bar v) = 1$, times an arbitrary function of the slow variables only, z. This function is determined by the dynamics at higher orders, which we are going to see in the following.

Order ϵ^0, solution at the intermediate timescales. At next-to-leading order, the equation to solve for the correction $Z^{(1)}$ is

$$-\big(M+\mu_f\big)Z^{(1)} = \bar v^\top\nabla_{\bar x}Z^{(0)} + \mu^{(0)}Z^{(0)}\ , \tag{3.20}$$

where $Z^{(0)}$ is given from the calculation at the previous order. Formally, one needs to invert the operator $(M+\mu_f)$. However, this operator is not invertible, as we calculated $Z^{(0)}$ as its right null vector. In order to solve Eq. (3.20), one needs to specify a condition on the terms on the right hand side. This solvability condition, referred to as the *Fredholm alternative*, states that the solution must be looked for in the subspace orthogonal to the kernel of the operator on the left hand side, i.e. $Z^{(0)}$. Therefore, it is obtained by projecting Eq. (3.20) onto the left null vector of $(M+\mu_f)$, practically done by multiplying by ℓ and integrating over the velocity variables. From Eq. (3.19) one notices that this procedure is equivalent to averaging Eq. (3.20) over the local equilibrium distribution for $\bar v$, conditioned on the quenched variables $\bar x$, $w(\bar v|\bar x)$. Since w has vanishing average, the solvability condition reads $\mu^{(0)} = 0$, and $Z^{(1)}$ solves

$$-\big(M+\mu_f\big)Z^{(1)} = \bar v^\top\nabla_{\bar x}Z^{(0)}\ . \tag{3.21}$$

Notice that, for a generic operator f,

$$-\big(M+\mu_f\big)\,f\,Z^{(0)} = \Big[-D_v\nabla_{\bar v}^2 f + \gamma\,\bar v^\top\big(Q^{1/2}\otimes\mathbb{I}_d\big)\nabla_{\bar v} f\Big]Z^{(0)} \equiv \big(Lf\big)Z^{(0)}\ . \tag{3.22}$$

Hence, by writing $Z^{(1)} = f(\bar v)\,Z^{(0)}$, Eq. (3.21) is equivalent to

$$(Lf)z = \bar v^\top\nabla_{\bar x} z\ . \tag{3.23}$$

We recognize that the operator L is the generator of the Ornstein–Uhlenbeck process with a friction tensor $\Gamma = \gamma\,\big(Q^{1/2}\otimes\mathbb{I}_d\big)$, and its eigenfunctions are Hermite polynomials in the velocities along the eigenvectors of Γ. Linear functions are therefore eigenvectors of L, and the ansatz

$$f(\bar v) = \bar v^\top A\ ,$$

solves Eq. (3.23) for

$$A = \frac{1}{\gamma}\big(Q^{-1/2}\otimes\mathbb{I}_d\big)\,\nabla_{\bar x}\log z\ . \tag{3.24}$$

Then, at the next-to-leading order in the perturbation expansion, the solution of the HJB equation in the overdamped limit is

$$Z \simeq \Big(1+\frac{\epsilon}{\gamma}\,\bar v^\top\,\big(Q^{-1/2}\otimes\mathbb{I}_d\big)\,\nabla_{\bar x}\log z\Big)\,Z^{(0)}\ . \tag{3.25}$$

Notice that the solvability condition provides the effective dynamics of the variables $\bar x$, occurring on time scales of order 1. However, it turns out that such dynamics is trivial, i.e. no dynamics occurs on these time scales in this particular case. The effective dynamics takes place at longer time scales $O(\epsilon^{-1})$, and this requires to proceed to the analysis of the second order in the ϵ expansion of the HJB equation.

Order ϵ, effective slow dynamics. At this order, we need to solve

$$-\big(M+\tilde\mu\big)Z^{(2)} = \bar v^\top\nabla_{\bar x}Z^{(1)} + \frac{1}{2D_v\eta}q(\bar x)\,Z^{(0)} + \mu^{(1)}Z^{(0)}\ . \tag{3.26}$$

As it was described above, one needs to invoke the Fredholm alternative in order to solve this equation for the second order correction $Z^{(2)}$. Here, the solvability condition becomes the effective HJB equation for the desirability z,

$$\frac{1}{\gamma}\big\langle\bar v^\top\nabla_{\bar x}\big(\big(Q^{-1/2}\otimes\mathbb{I}_d\big)\,\nabla_{\bar x}z\big)\big\rangle - q(\bar x)\,z + \mu^{(1)}z = 0\ . \tag{3.27}$$

Written explicitly in components, the first term in Eq. (3.27) is

$$\frac{1}{\gamma}\sum_{i,\alpha}\big\langle v_{i,\alpha}\,\partial_{x_{i,\alpha}}\sum_{j,\beta}v_{j,\beta}\sum_{k,\sigma}\big(Q^{-1/2}_{j,k}\delta_{\beta,\sigma}\big)\,\partial_{x_{k,\sigma}}z\big\rangle = \frac{1}{\gamma}\sum_{i,j,k}\sum_{\alpha,\beta}\underbrace{\langle v_{i,\alpha}v_{j,\beta}\rangle}_{\frac{D_v}{\gamma}Q^{-1/2}_{j,i}\delta_{\beta,\alpha}}\ \partial_{x_{i,\alpha}}\,Q^{-1/2}_{j,k}\,\partial_{x_{k,\beta}}z$$

$$= \frac{D_v}{\gamma^2}\sum_{i,\alpha}\underbrace{\sum_j Q^{-1/2}_{i,j}\,\partial_{x_{j,\alpha}}}\ \underbrace{\sum_k Q^{-1/2}_{i,k}\,\partial_{x_{k,\alpha}}}\ z\ .$$

The effective HJB equation for the desirability z for the slow variables is then, in a more compact notation,

$$-\frac{D_v}{\gamma^2}\big[\big(Q^{-1/2}\otimes\mathbb{I}_d\big)\nabla_{\bar{x}}\big]^\top\big(Q^{-1/2}\otimes\mathbb{I}_d\big)\nabla_{\bar{x}}z+q(\bar{x})\,z=\mu^{(1)}z\ . \tag{3.28}$$

The optimal control $\bar{f}$ is

$$\begin{aligned}\bar{f}&=2D_v\,\nabla_{\bar{v}}\Big[\log Z^{(0)}+\log\Big(1+\epsilon\,\frac{Z^{(1)}}{Z^{(0)}}+\dots\Big)\Big]\\&\simeq-\gamma\,\big(\big(Q^{1/2}-\mathbb{I}_N\big)\otimes\mathbb{I}_d\big)\bar{v}+\epsilon\frac{2D_v}{\gamma}\,\big(Q^{-1/2}\otimes\mathbb{I}_d\big)\,\nabla_{\bar{x}}\log z\ .\end{aligned}$$

Then the dynamics of the optimally controlled system, at the first order approximation in ϵ for Z, is

$$\begin{cases}d\bar{X}=\bar{V}\,dt\ ,\\ d\bar{V}=-\epsilon^{-1}\,\gamma\,\big(Q^{1/2}\otimes\mathbb{I}_d\big)\bar{V}dt+\dfrac{2D_v}{\gamma}\,\big(Q^{-1/2}\otimes\mathbb{I}_d\big)\,\nabla_{\bar{x}}\log z\,dt+\sqrt{2D_v}\,d\bar{W}^t\ .\end{cases} \tag{3.29}$$

The probability density function p for this process satisfies the (backward) Kolmogorov (Fokker–Planck) equation

$$\begin{aligned}\partial_t p+\epsilon^{-1}\Big[D_v\nabla^2_{\bar{v}}p-\gamma\bar{v}^\top\big(Q^{1/2}\otimes\mathbb{I}_d\big)\nabla_{\bar{v}}p\Big]&\\+\bar{v}^\top\nabla_{\bar{x}}p+\frac{2D_v}{\gamma}\,(\nabla_{\bar{x}}\log z)\big(Q^{-1/2}\otimes\mathbb{I}_d\big)\,\nabla_{\bar{v}}p=0\ .&\end{aligned} \tag{3.30}$$

By applying the multi-scale analysis to this equation[‡], we see that the effective equation for the slow variables is

$$\begin{aligned}\partial_t p+2\,D\,\big[\big(Q^{-1}(\bar{x})\otimes\mathbb{I}_d\big)\nabla_{\bar{x}}\log z\big]^\top\nabla_{\bar{x}}p&\\+D\big[\big(Q^{-1/2}\otimes\mathbb{I}_d\big)\nabla_{\bar{x}}\big]^\top\big(Q^{-1/2}\otimes\mathbb{I}_d\big)\nabla_{\bar{x}}p=0\ ,&\end{aligned} \tag{3.31}$$

where $D=D_v/\gamma$ is a typical spatial diffusivity constant. Equation (3.31) is the backward Fokker–Planck equation associated to the drift diffusion process

$$d\bar{X}=2D\,\big(Q^{-1}(\bar{X})\otimes\mathbb{I}_d\big)\nabla_{\bar{x}}\log z\,dt+\sqrt{2D}\,\big(Q^{-1/2}\otimes\mathbb{I}_d\big)\circ d\bar{W}^t\ , \tag{3.32a}$$

[‡]The calculations are the same as for the HJB equation. An ansatz in the form $p=p^{(0)}+\epsilon p^{(1)}+\dots$ is used, as well as auxiliary variables $t_f=\epsilon^{-1}t$, $t_s=\epsilon t$, $\bar{x}_f=\epsilon^{-1}\bar{x}$ and $\bar{x}_s=\epsilon\,\bar{x}$ are introduced. $p^{(0)}$ is constant in the fast variables, and the corrensponding left eigenvector is the local equilibrium distribution for the Ornstein–Uhlenbeck process with friction tensor $\Gamma=\gamma\big(Q^{1/2}\otimes\mathbb{I}_d\big)$. The calculations at the lower orders are exactly the same.

or, in components

$$dX_{i,\alpha} = 2D \sum_j Q^{-1}_{i,j}(\bar{X}) \partial_{x_{j,\alpha}} \log z \, dt + \sqrt{2D} \sum_j Q^{-1/2}_{i,j}(\bar{X}) \circ dW^t_{j,\alpha} \ , \tag{3.32b}$$

which is understood in the sense of Stratonovich. Notice that the noise term in the backward Fokker–Planck equation can also be written as

$$D \sum_{i,\alpha} \sum_j Q^{-1/2}_{i,j} \, \partial_{x_{j,\alpha}} \sum_k Q^{-1/2}_{i,k} \, \partial_{x_{k,\alpha}} p \\ = \sum_\alpha \sum_{i,j} Q^{-1}_{i,j} \partial_{x_{i,\alpha}} \partial_{x_{j,\alpha}} p + \sum_{i,\alpha} \Big(\sum_{k,j} Q^{-1/2}_{k,j} \, \partial_{x_{j,\alpha}} Q^{-1/2}_{k,i} \Big) \, \partial_{x_{i,\alpha}} p \ ,$$

arising naturally from the Ito stochastic equation

$$dX_{i,\alpha} = 2D \sum_j Q^{-1}_{i,j}(\bar{X}) \partial_{x_{j,\alpha}} \log z \, dt \\ + D \Big(\sum_{k,j} Q^{-1/2}_{k,j} \, \partial_{x_{j,\alpha}} Q^{-1/2}_{k,i} \Big) dt + \sqrt{2D} \sum_j Q^{-1/2}_{i,j}(\bar{X}) \, dW^t_{j,\alpha} \ . \tag{3.33}$$

The Ito stochastic dynamics features a spurious drift term which is due to the inhomogeneity of the noise. In general, this spurious drift is very difficult to calculate, since it involves the explicit calculation of the derivative of $Q^{-1/2}$ for any configuration of the positions of the N agents. However, some explicit results can be worked out for $N = 2$.

2-agents dynamics

The spurious drift term has a simple expression when $N = 2$. Indeed, one can see by direct diagonalization of the matrix Q that its eigenvalues are

$$\lambda_1 \equiv \lambda_{\min}(\bar{x}) = 1 + \frac{1}{\gamma^2 \eta} \big(\kappa - k(|x_1 - x_2|) \big) \ , \\ \lambda_2 \equiv \lambda_{\max}(\bar{x}) = 1 + \frac{1}{\gamma^2 \eta} \big(\kappa + k(|x_1 - x_2|) \big) \ ,$$

with corresponding (normalized) eigenvectors

$$r^{(1)} = \Big(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \Big)^\top \ , \\ r^{(2)} = \Big(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \Big)^\top \ .$$

The orthogonal matrix $R = (r^{(1)}, r^{(2)})^\top$, such that $Q = R^\top \operatorname{diag}(\lambda_1, \lambda_2) \, R$, does not depend on the coordinates of the two agents, and the calculation of the derivative of $Q^{-1/2}$ boils

down to the calculation of the derivatives of the eigenvalues $\lambda_{1,2}$. The spurious drift can be calculated to be

$$\sum_{j,k} Q_{k,j}^{-1/2}\,\partial_{x_{j,\alpha}} Q_{k,i}^{-1/2} = \frac{1}{2}(-1)^i \partial_{x_{2,\alpha}} \lambda_2^{-1} = \frac{1}{2}\,\partial_{x_{i,\alpha}} \Big[1 + \frac{1}{\gamma^2\eta}\big(\kappa + k(|x_1 - x_2|)\big)\Big]^{-1} ,$$

obtained explicitly from the spectral decomposition of the matrix Q. The fact that $\partial_{x_{1,\alpha}}\lambda_l = -\partial_{x_{2,\alpha}}\lambda_l$, and that $\partial_{x_{j,\alpha}}\lambda_1 = -\partial_{x_{j,\alpha}}\lambda_2$ is used in the derivation of the formula for the spurious drift, and is simply a consequence of the symmetry properties of the alignment interaction kernel k. We note in passing that this drift term is a gradient of a scalar function of the position of the two agents.

For the 2-agent system we can provide closed formulas for the stationary distribution. In order to do this, we convert the backward Fokker–Planck equation, Eq. (3.31), into its forward counterpart, and seek for the stationary limit, $\partial_t p = 0$. The forward FP equation writes

$$\partial_t p + D \sum_{i,\alpha} \partial_{x_{i,\alpha}} \sum_j Q_{i,j}^{-1} \Big[2p\partial_{x_{j,\alpha}} \log z - \partial_{x_{j,\alpha}} p + \frac{1}{2} p \partial_{x_{j,\alpha}} \log \Big[1 + \frac{1}{\gamma^2\eta}\big(\kappa + k(|x_1 - x_2|)\big)\Big]\Big] = 0 , \quad (3.34)$$

where the identity

$$\sum_{k,l} Q_{j,k}^{1/2}\,\partial_{x_{l,\alpha}} Q_{l,k}^{-1/2} = -\frac{1}{2}\,\partial_{x_{j,\alpha}} \log \Big[1 + \frac{1}{\gamma^2\eta}\big(\kappa + k(|x_1 - x_2|)\big)\Big]$$

has been used, obtained in a similar way as for the spurious drift. The equilibrium solution is found by integrating once the stationary equation (3.34), and solving it explicitly:

$$p_{\mathrm{eq}}(x_1, x_2) \propto \lambda_{\max}^{1/2}\, z^2 = \Big[1 + \frac{1}{\gamma^2\eta}\big(\kappa + k(|x_1 - x_2|)\big)\Big]^{1/2} z^2 , \quad (3.35)$$

where z is the solution of the HJB equation (3.28). If the agents move on a torus (2-dimensional domain with periodic boundary conditions), and $q = 0$, i.e. no confinement or any inter-particle positional interaction is present, the solution of Eq. (3.28) is constant. It is interesting to notice that, if the kernel k is a positive and monotonically decreasing function of the distance, as we assumed to be the case, the equilibrium distribution is a decreasing function of the relative distance between the two agents. At very long time-scales, then, the request to minimize a cost favouring alignment of the velocities results, in the overdamped limit, in a net aggregation of the agents. This result corroborates the findings in numerical investigations of agent-based Vicsek-like models, where aggregation is found as an effect of the dynamics over long times. Here, we derive it analytically in the case of Langevin particles, as a consequence of an optimization procedure.

3.3 Discussion and perspective

In this Chapter we showed that a model of persistent random walk (the Langevin-Kramers dynamics) allows for aggregation when it optimizes a cost function with an alignment term. This result is obtained in the overdamped limit by studying the system of optimality equations (Hamilton–Jacobi–Bellman and Fokker–Planck) by means of averaging and homogenization techniques. The analytic solution of the overdamped optimally controlled dynamics is found only in the case of two agents on a torus.

In perspective, this preliminary result may be extended to more general scenarios, e.g. with spatial confinement and/or featuring more than two agents. The calculation of the stationary state distribution for a generic N seems a daunting task, because it involves the diagonalization of the $N \times N$ matrix Q, which contains all agent-agent distances. However, since the statistics of the matrix Q is determined by the spatial distribution of agents, it might be possible to exploit results from random matrix theory [126] to calculate, in a self-consistent way, the stationary state also in this case. Another possibility is that, with a suitable choice of the interaction kernel k, the solution of the Bellman equation and of the stationary Fokker–Planck equation might be found to have a formal connection with models in statistical physics, where exact results could shed light on some general principles [127–129]. Moreover, it is interesting to investigate the possible analogies with existing physical models of ferromagnetic liquids, e.g. spin fluids [130], and gain insights from these systems.

4

Exact and efficient sampling of constrained random walks

Generating constrained walks by means of stochastic techniques is generally computationally challenging [60]. This is readily illustrated by considering the inherent inefficiency of simple resampling strategies. In such approaches one could, in principle, generate by Monte Carlo or other schemes, a large ensemble of unrestricted walks and then reject *a posteriori* those violating the constraints. This naive strategy is bound to incur in a rejection rate that increases exponentially fast with the walk length.

A different approach was pioneered by Doob [131] for diffusive processes and recently revisited in the physics literature, see e.g. [57–59, 132]. In short, this method is based on the observation that any constrained random walk is exactly equivalent to an auxiliary unconstrained one, via a suitable reweighting of the transition probability. Clearly, the unconstrained version of the original process is typically much more amenable than the former to computational, and even analytical treatment.

In this Chapter, we show how it is possible to represent the unconstrained auxiliary process as an optimal search problem. We illustrate our approach with the example of a persistent walk confined inside a cylinder and forced to reach one of its ends. The searcher pays a cost proportional to the time spent outside of the cylinder. By finding the search strategy that minimizes this cost function –vanishing at optimality– one is assured that all

the trajectories generated lie within the cylinder. For this problem, the exact expression of the transition probability for the jumps can be obtained analytically. We also devise an exact sampling algorithm with very low rejection rate for these transition probabilities.

This method allows one to generate walks with arbitrarily large contour length and thus characterize different regimes (such as the transition from weak to strong confinement, i.e. different limits of the ratio between the persistence length and the radius of the cylinder) that are otherwise difficult to access when the length becomes asymptotically large. We show numerical simulations of these confined random walks performed by means of our algorithm and we characterize analytically and numerically some of the geometric properties of these walks relevant to the physics of polymers [62, 63, 133].

4.1 Optimal control representation of a confined walk

In this section we are going to formulate the problem of sampling a random walk confined in a bounded region of space as a optimal search process. We derive general equations for arbitrary geometries and will eventually specialize (and solve) them for the case of a discrete-time process with exponentially distributed jumps confined in an infinitely long cylinder. This example is a random walk which can be seen as a simple model of polymers with a given persistence length.

Let us consider a searcher whose motion is described by a discrete-time Markov process X_t with transition probabilities $p(x'|x)$, so that the probability of a path $\{x_1 \dots x_T\}$, conditioned on an initial state x_0, is written as

$$\mathbb{P}(x_1, \dots x_T|x_0) = p(x_1|x_0)\, p(x_2|x_1)\, \dots\, p(x_T|x_{T-1})\ . \tag{4.1}$$

The transition probabilities $p(x'|x)$ determine the time evolution of the probability vector ρ_t (i.e. the 1-point pdf of the variable X) through the (forward) Kolmogorov equation

$$\rho_t(x) = \sum_{x'} \rho_{t-1}(x') p(x|x')\ . \tag{4.2}$$

We refer to this as the *uncontrolled* process. This serves as a reference process, i.e. as the model describing the random searcher in absence of constraints.

Let us imagine that the reference process X_t lives in $\mathcal{S} \subseteq \mathbb{R}^d$, and that a portion of such a domain is a target $\mathcal{T}$, which is absorbing. Let us also assume that the random walker is allowed to occupy a portion of the total space, $\mathcal{R} \subset \mathcal{S}$, while trying to reach the target. Anywhere which is not inside $\mathcal{R}$ or $\mathcal{T}$ is forbidden to the searcher.

Imposing hard constraints on the reference process is equivalent –at the level of the ensemble of paths– to constructing an auxiliary process (or a controlled one) that satisfies such constraints *in probability*. It is possible to represent the auxiliary process as the one

that minimizes a cost functional of the trajectory. We shall see that when such a functional is local in time, i.e. is an additive Markov process, the controlled process can be expressed as a reweighting of the reference one.

We then define a Markov process $X_t^{(u)}$ with transition probabilities $u(x'|x)$, that we refer to as the *controlled* process, so that the probability of the same path writes, analogously to Eq. (4.1),

$$\mathbb{P}_u(x_1, \dots x_T|x_0) = u(x_1|x_0)\, u(x_2|x_1)\, \dots\, u(x_T|x_{T-1})\ . \tag{4.3}$$

The probability vector, under this process, evolves in time according to

$$\rho_t(x) = \sum_{x'} \rho_{t-1}(x') u(x|x')\ . \tag{4.4}$$

Notice that, although it is not included in the notation, the transition probabilities of the auxiliary process might also depend explicitly on time.

A way of incorporating the hard constraint of staying inside the region $\mathcal{R}$ while reaching the target $\mathcal{T}$ is to include the notion of running cost, which is paid by the searcher only when it falls outside $\mathcal{R}$ or $\mathcal{T}$,

$$q(x) = \begin{cases} 0 & \text{if } x \in \mathcal{R} \\ -r & \text{if } x \in \mathcal{T} \\ c & \text{otherwise}\ . \end{cases} \tag{4.5}$$

The positive quantity r is a terminal reward that the walker receives when it reaches the target. The searcher accumulates these costs along its trajectory, until it reaches the target, paying a total cost

$$C = \sum_{t=0}^{T-1} q(X_{t+1}) = c \times (\#\text{ of steps outside } \mathcal{R}) - r\ , \tag{4.6}$$

where T is the first passage time at $\mathcal{T}$. Notice that the terminal reward constitutes just a constant shift in the cost function, if we consider only trajectories that arrive at the target. We assume here that the initial condition is inside the domain $\mathcal{R}$. It is clear that a searcher that reaches $\mathcal{T}$ by paying a vanishing cost (minus the reward r) is one that necessarily satisfies the constraint. This searcher is characterized as the controlled process $X_t^{(u)}$ which minimizes the cost functional

$$\mathcal{C} = \langle C \rangle_u = \sum_{t=0}^{\infty} \sum_{x \in \mathcal{S}} q(x)\, p_{t+1}(x) = \sum_{t=0}^{\infty} \sum_{x,x' \in \mathcal{S}} q(x')\, u(x'|x)\, \rho_t(x)\ , \tag{4.7}$$

where $\langle \cdot \rangle_u$ denotes the average over the ensemble of paths with probability given by Eq. (4.3), so ρ_t here denotes the probability distribution at time t of the controlled process. We conventionally assign a cost for the *final* point of each transition. The calculation can be generalized

to a cost function of the form $\sum_t f(X_t, X_{t+1})$. Hereafter it is implied that the domain over which we perform the summation over values of the process is $\mathcal{S}$.

It is convenient, for a reason that we shall see later, to introduce a different cost functional that includes a regularization term which is proportional to the Kullback–Leibler divergence from the path measure of the reference (uncontrolled) process to the path measure of the auxiliary (controlled) process:

$$\begin{aligned}\mathcal{F} &= \mathcal{C} + \epsilon\, D_{\mathrm{KL}}\big(\mathbb{P}_u \big\| \mathbb{P}\big) \\ &= \mathcal{C} + \epsilon \sum_{t=1}^{\infty} \sum_{x} \rho_t(x)\, D_{\mathrm{KL}}\big(u(\cdot|x) \big\| p(\cdot|x)\big) \\ &= \sum_{t=1}^{\infty} \sum_{x,x'} \Bigg(\underbrace{q(x') + \epsilon\, \log \frac{u(x'|x)}{p(x'|x)}}_{c_t(x',x)} \Bigg)\, u(x'|x)\, \rho_t(x)\ , \end{aligned} \tag{4.8}$$

where ρ_t is constrained by the dynamics determined by the transition probabilities u. The last equality in Eq. (4.8) shows that the KL regularization can be incorporated into the running "cost" $c_t(x_{t+1}, x_t)$, as an additional on-line *fluctuating entropy* term, which, on average, penalizes search strategies u that are far from the reference one. The KL divergence is indeed a convex functional of u which vanishes only if u is equal to p everywhere. The parameter ϵ sets a trade-off between the control cost and the cost that represents the constraint, C. Then, it is intuitive that the hard constraint is realized in the limit $\epsilon \to 0$. However, this regularization is convenient in that it makes the optimal control problem globally convex and linearly solvable [74, 80, 134]. This problem then is a generalization of the diffusive case discussed in Chapter 1.

We can derive the optimal control equations by applying the Pontryagin minimum principle [22]. This is equivalent to the unconstrained minimization of the Lagrange functional

$$\begin{aligned}\mathcal{L}[u,\, p,\, \phi,\, \lambda] = \mathcal{F} + \sum_t \sum_{x'} \phi_t(x') \Big(p_{t+1}(x') - \sum_x u(x'|x)\, \rho_t(x) \Big) \\ + \sum_t \sum_x \lambda_t(x) \rho_t(x) \Big(\sum_{x'} u(x'|x) - 1 \Big)\ , \end{aligned} \tag{4.9}$$

where the arrival time at the target is not constrained. The two extra terms impose the constraints of the dynamics: null variations with respect to λ_t impose the normalization of the transition probability u, while the one with respect to ϕ_t gives the (forward) Kolmogorov equation for the probability ρ_t. In principle, the controlled transition probability u might

depend explicitly on t. The variations with respect to u and p give:

$$\left.\frac{\partial \mathcal{L}}{\partial u(x'|x)}\right|_* = \rho_t(x)\left[\epsilon\left(\log\frac{u(x'|x)}{p(x'|x)}+1\right)+q(x')+\lambda_t(x)-\phi_t(x')\right]=0\ , \tag{4.10}$$

$$\left.\frac{\partial \mathcal{L}}{\partial \rho_t(x)}\right|_* = \sum_{x'} u(x'|x)\left[\epsilon\,\log\frac{u(x'|x)}{p(x'|x)}+q(x')-\phi_t(x')+\phi_{t-1}(x)\right]=0\ . \tag{4.11}$$

Multiplying the first by $u(x'|x)$ and summing over x', and then comparing with the second equation, yields $\epsilon+\lambda(x)=\phi_{t-1}(x)$. This fixes the form of the *optimal controller* $u^*(x'|x)$ to be

$$u^*(x'|x) = p(x'|x)\,e^{-q(x')/\epsilon}\,e^{[\phi_t(x')-\phi_{t-1}(x)]/\epsilon}\ . \tag{4.12}$$

The normalization condition for $u^*(\cdot|x)$ is the Hamilton–Jacobi–Bellman equation

$$e^{\phi_{t-1}(x)/\epsilon} = \sum_{x'} p(x'|x)\,e^{-q(x')/\epsilon}\,e^{\phi_t(x')/\epsilon}\ . \tag{4.13}$$

If we define $Z_t(x)=\exp\phi_t(x)/\epsilon$, Eq. (4.13) is cast into a linear equation

$$Z_{t-1}(x) = \sum_{x'} p(x'|x)\,e^{-q(x')/\epsilon}\,Z_t(x')\ , \tag{4.14}$$

and

$$u^*(x'|x) = \frac{Z_{t+1}(x')}{Z_t(x)}\,e^{-q(x')/\epsilon}\,p(x'|x)\ . \tag{4.15}$$

These equations are generally time-dependent. When no constraint is imposed on the time of arrival at the target $\mathcal{T}$, as in the problem formulated here as well as for the search game of Ch. 1, the optimal strategy is time-independent, so Eqs. (4.14) and (4.15) can be written as

$$\begin{aligned} Z(x) &= \sum_{x'} p(x'|x)\,e^{-q(x')/\epsilon}\,Z(x')\ ,\\ u^*(x'|x) &= \frac{Z(x')}{Z(x)}\,e^{-q(x')/\epsilon}\,p(x'|x)\ . \end{aligned} \tag{4.16}$$

As we already noted, the hard constraint of being inside $\mathcal{R}$ until reaching the target is satisfied by setting $\epsilon\to 0$. In this limit, according to the definition of the cost in Eq. (4.5), the exponential factor in the linear HJB equation is

$$\lim_{\epsilon\to 0} e^{-q(x)/\epsilon} = \begin{cases} +\infty & \text{if } x\in\mathbb{T} \\ \mathbb{I}_{\mathcal{R}}(x) & \text{otherwise}\ , \end{cases}$$

where $\mathbb{I}_{\mathcal{R}}$ is the indicator function of the allowed region $\mathcal{R}$.

It is important to note that the choice of the reference process, i.e. the uncontrolled one with transition probabilities $p(x'|x)$, is completely arbitrary, and is solely determined on the basis of physical arguments (for instance, in the problem we deal with in this Chapter, we would like it to represent a polymer with given persistence length). Indeed, it only matters when ϵ is finite. In such cases the cost for control would be traded-off with the cost for being outside the allowed domain, and this would "soften" the constraint.

Other cost functions

The cost functional $\mathcal{C}$ as defined in Eq. (4.7), is only a convenient choice in this case. Other forms of the cost can be used for this conditioning problem, yielding the same optimal control equations. We have seen in Ch. 1 that the "risk-sensitive" cost functional $\mathcal{F}_\alpha = \alpha^{-1} \log\langle \exp \alpha(C + \text{"reg."})\rangle_u$, where "reg" stands for a regularizing cost for control, can also be solved linearly. As shown in [80, 134], for this more general form of the cost, the regularization which makes the optimal control equations linearly solvable is the Renyi divergence, that recovers the KL divergence in the limit $\alpha \to 0$,

$$D_{1+\alpha}\big(u(\cdot|x)\|p(\cdot|x)\big) = \frac{1}{\alpha}\log\sum_{x'} u(x'|x)^{1+\alpha}\, p(x'|x)^{-\alpha} \underset{\alpha\to 0}{\longrightarrow} D_{\mathrm{KL}}(u(\cdot|x)\|p(\cdot|x)) \ . \tag{4.17}$$

The difference between KL and Renyi divergence is apparent only in discrete-time or discrete-state processes, while in the diffusive limit the two converge to the same functional.

4.1.1 Conditioning as reweighting

Another way of constructing an auxiliary process that satisfies the hard constraint represented by the conditioning is the reweighting of the transition probabilities. Perhaps, this is more natural if one thinks of a naive rejection algorithm in which the full trajectory is discarded whenever a jump is taken outside the allowed region. This can be formally done by assigning vanishing probability to jumps ending outside $\mathcal{R}$.

The probability of a path $\{x_1 \dots x_T\}$ under the reweighted transition probability q can be written as

$$\mathbb{Q}(x_1, \dots x_T|x_0) = \frac{1}{Z_T(x_0)} \prod_{t=1}^{T} p(x_t|x_{t-1})\, \mathbb{I}_{\mathcal{R}}(x_t) \ , \tag{4.18}$$

where Z_T is a normalization factor. Notice that it satisfies a recursive relation, which has

the form of a (backward) Feynman-Kac equation

$$\begin{aligned} Z_T(x_0) &= \sum_{\{x_i\}_{i=1}^T} \prod_{t=1}^{T} p(x_t|x_{t-1})\, \mathbb{I}_{\mathcal{R}}(x_t) \\ &= \sum_{x_1} p(x_t|x_0)\, \mathbb{I}_{\mathcal{R}}(x_1) \sum_{\{x_i\}_{i=2}^T} \prod_{t=1}^{T} p(x_t|x_{t-1})\, \mathbb{I}_{\mathcal{R}}(x_t) \\ &= \sum_{x_1} p(x_t|x_0)\, \mathbb{I}_{\mathcal{R}}(x_1)\, Z_{T-1}(x_1) \ . \end{aligned} \tag{4.19}$$

In the limit $T \to \infty$, relevant to the terminal-state setup, Eq. (4.19) becomes the stationary equation (4.16) and the path probability $\mathbb{Q}$ can be written as the telescopic product

$$\mathbb{Q}(x_1, \dots x_T|x_0) = \prod_{t=1}^{T} p(x_t|x_{t-1})\, \mathbb{I}_{\mathcal{R}}(x_t) \frac{Z(x_t)}{Z(x_{t-1})} \ , \tag{4.20}$$

i.e. the probability of a path generated by the Markov process with transition probabilities

$$q(x'|x) = p(x'|x)\, \mathbb{I}_{\mathcal{R}}(x') \frac{Z(x')}{Z(x)} \ . \tag{4.21}$$

These are exactly the same as the optimal control transition probability u^*, so $\mathbb{Q} = \mathbb{P}_{u^*}$.

4.2 Constraining a jump process inside a cylindrical channel

In some cases Eq. (4.16) can be solved analytically, providing a very effective method to sample configurations that would otherwise be exceedingly rare in an unbiased sampling procedure. One such instance is when the confining region is a cylinder. In Fig. 4.1 we report a scheme of the conditioning problem and some trajectories for different confining regimes. We note here that a straightforward rejection scheme would be extremely costly in generating constrained trajectories under strong confinement, such as those in panels *c* and *d* in Fig. 4.1.

Besides being amenable to extensive characterization within the aforementioned framework, this system was chosen for its connection with polymer chains inside nano-channels.

4.2.1 Exact solution for exponentially distributed jumps

In this paragraph we show explicitly how to solve Eq. (4.16) (in the limit $\epsilon \to 0$) when the domain $\mathcal{R}$ is a cylindrical channel of radius R parallel to the z axis, with axial length $2H$, and the terminal domain to be at one of its two extremes along z:

$$\begin{aligned} \mathcal{R} &= \left\{ x, y, z \,\middle|\, \rho = \sqrt{x^2 + y^2} \leq R, |z| < H \right\} \ , \\ \mathcal{T} &= \left\{ x, y, z \,\middle|\, \rho = \sqrt{x^2 + y^2} \leq R, z > H \right\} \end{aligned} \tag{4.22}$$

(see Fig. 4.1). For this problem, the Hamilton–Jacobi–Bellman equation (4.16) reads

$$Z(\bar{x}) = \int_{\mathcal{R}} d^3x'\, p(\bar{x}'|\bar{x})\, Z(\bar{x}') \tag{4.23}$$

from which the transition probability for the auxiliary process writes

$$u^*(\bar{x}'|\bar{x}) = p(\bar{x}'|\bar{x})\, \mathbb{I}_{\mathcal{R}}(\bar{x}')\, \frac{Z(\bar{x}')}{Z(\bar{x})}\ . \tag{4.24}$$

The exact results can be obtained in the limit of an infinitely long cylinder, $H \to \infty$.

Let us consider the following reference (uncontrolled) process

$$p(x'\,|\,x) = \frac{m^2}{4\,\pi}\, \frac{e^{-m\,|\bar{x}'-\bar{x}|}}{4\pi\,|\bar{x}'-\bar{x}|}\,, \tag{4.25}$$

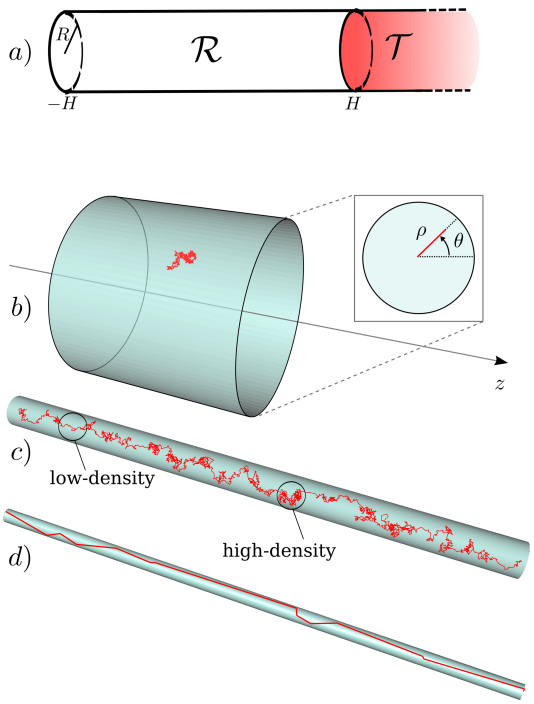


Fig. 4.1. **Scheme of the conditioning problem and qualitative features of the sample paths.** *a)* The domain in which walks are confined is a cylinder $\mathcal{R}$ with radius R and longitudinal extension $2H$. The target that the walks have to reach, denoted by $\mathcal{T}$, is the continuation of the cylinder on one of its sides. All the remaining part of the 3-dimensional space, is forbidden to the searcher. Below are depicted some sample trajectories in different confinement regimes: *b)* weak, *c)* intermediate, and *d)* strong. Different confinements yield different scaling regimes of geometric quantities, such as end-to-end distance, when looking at a sub-portion of a given number of jumps (see text). Also, density fluctuations, highlighted in panel *c)*, are interesting and can be quantified analytically.

where $\bar{x} \equiv (x, y, z)$ indicates the position vector in $\mathbb{R}^3$, and the notation $|\cdot|$ indicates the Euclidean norm. The parameter m controls the mean length of a jump,

$$\ell_f \equiv \langle\, |x_{t+1} - x_t| \,\rangle_p = 2/m\,. \tag{4.26}$$

We note in passing that the discrete-time process with transition probabilities given by Eq. (4.25) can be obtained by considering a three-dimensional diffusion process with coefficient D and sampling it at random time intervals distributed exponentially with mean $\tau = 1/(Dm^2)$,

$$p(\bar{x}'\,|\,\bar{x}) = \int_0^\infty \frac{dt}{\tau}\, e^{-\frac{t}{\tau}}\, \frac{e^{-\frac{(\bar{x}'-\bar{x})^2}{4Dt}}}{(2\pi\, D\, t)^{3/2}}\,. \tag{4.27}$$

Expressing the free propagator of the Brownian motion in Fourier space and integrating over time yields

$$p(\bar{x}'\,|\,\bar{x}) = \int \frac{d^3k}{(2\pi)^3}\, \frac{m^2}{k^2+m^2}\, e^{-i\,\bar{k}\cdot(\bar{x}'-\bar{x})}\,, \tag{4.28}$$

which is the Fourier representation of Eq. (4.25) (see Fig. 4.2).

As noted in the previous section, the particular form of the free transition probability p is not important in the conditioning scheme, as long as it contains the necessary ingredients of the physical system that one wants to describe. In our example, the only requirement of

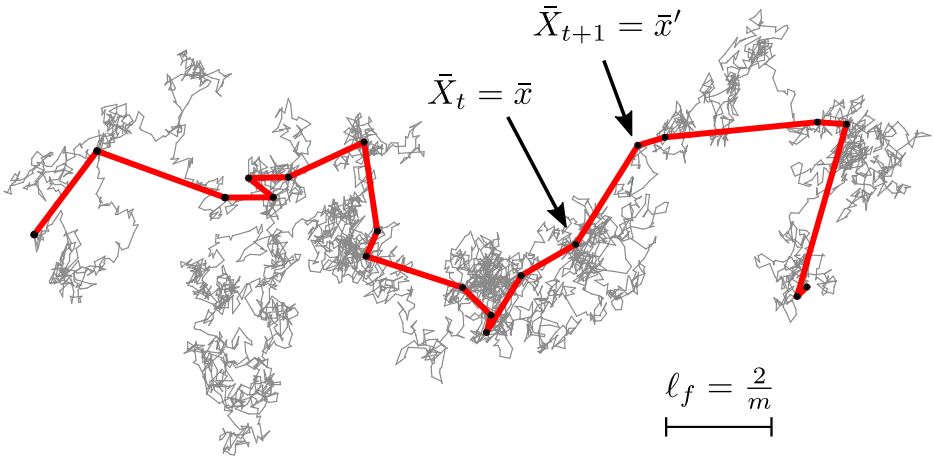


Fig. 4.2. **Exponential jumps as a coarse-grained Brownian walk.** The reference process with transition probabilities $p(\bar{x}'|\bar{x})$ (red line), with mean jump-length ℓ_f, is depicted in red. This process can be thought of as a coarse graining of the unbiased diffusion process with diffusivity constant D, represented in grey. Such coarse-graining is done by sampling a point on the diffusive trajectory every interval of time distributed exponentially with average $\tau = 1/(Dm^2) = \ell_f^2/(4D)$.

the free process is to have a characteristic length which can be attributed to the persistence of the polymer. The particular choice (4.25) for the transition probability $p(\bar{x}' \,|\, \bar{x})$ presents the advantage of satisfying

$$(\nabla^2 - m^2)\, p(\bar{x}' \,|\, \bar{x}) = -m^2\, \delta(\bar{x}' - \bar{x}) \;, \tag{4.29}$$

where δ is the Dirac delta function. Thus, applying the differential operator $\nabla^2 - m^2$ to Eq. (4.23) gives

$$(\nabla^2 - m^2)\, Z(\bar{x}) = -m^2 \int_{\mathcal{R}} d^3x'\, \delta(\bar{x}' - \bar{x})\, Z(\bar{x}') \;,$$

that leads to the system of equations

$$\begin{cases} \nabla^2\, Z(\bar{x}) = 0 & \text{for } \bar{x} \;\text{ in }\; \mathcal{R} \\ \left(\nabla^2 - m^2\right)\; Z(\bar{x}) = 0 & \text{elsewhere.} \end{cases} \tag{4.30}$$

In cylindrical coordinates the problem is separable in the longitudinal variable z and the polar variables (θ, ρ), and the general solution of Eq.(4.30) reads

$$Z(\bar{x}) = \begin{cases} A\, \exp(\lambda z)\, J_0\big(\lambda\rho\big) & \text{for } \bar{x} \;\text{ in }\; \mathcal{R} \\ B\, \exp(\lambda z)\, K_0\big(\sqrt{m^2 - \lambda^2}\,\rho\big) & \text{elsewhere,} \end{cases} \tag{4.31}$$

where A and B are real constants, and λ is the unique solution of the eigenvalue equation associated to Eq. (4.23),

$$\begin{aligned} &\lambda\, J_1(\lambda R)\, K_0\left(\sqrt{m^2 - \lambda^2}R\right) \\ &\qquad = \sqrt{m^2 - \lambda^2}\, J_0(\lambda R)\, K_1\left(\sqrt{m^2 - \lambda^2}R\right) \;. \end{aligned} \tag{4.32}$$

Here K_ν are the modified Bessel functions of the second kind. The function $\phi(\bar{x}) = \epsilon \log Z(\bar{x})$ is the optimal value function for the search problem, defined as (minus) the expected cost-to-go conditioned on the starting point $\bar{x}$. Therefore, when starting from the terminal region $\mathcal{T}$, the walker receives a positive reward and hence $Z \to \infty$ for $\epsilon \to 0$, which selects the growing exponential in the longitudinal variable for the solution of Eq. (4.30). Combining Eqs. (4.24) and (4.31) gives

$$u^*(\bar{x}'|\bar{x}) = p(\bar{x}'|\bar{x})\, e^{\lambda(z'-z)}\, \frac{J_0(\lambda\rho')}{J_0(\lambda\rho)}\, \Theta(R - \rho') \;, \tag{4.33}$$

where J_0 is the Bessel function of the first kind of order 0, and Θ is the Heaviside theta function, representing the indicator function of the infinitely long cylinder in polar coordinates. According to Eq. (4.33), λ^{-1} can be seen as a confinement-dependent length controlling the size distribution of the jumps in the positive z-direction: larger values of λ^{-1} will reflect in larger longitudinal jumps on average.

The knowledge of the transition probability (4.33) allows for a direct sampling of the constrained walk. With this method, the complexity of generating trajectories is independent of the strength of the confinement, and grows linearly with the number of jumps. This allows us to produce large samples of confined trajectories without rejections. This is especially useful in the strong confinement limit (see Fig. 4.1), when the channel diameter is much smaller than the jumps of the unconstrained process. In this case, virtually all free trajectories would violate the constraints. Figure 4.1 displays three realizations of the walk, when the reference jump length of the unconstrained walk, ℓ_f, increases with respect to the cylinder radius R, i.e. from weak to strong confinement.

4.2.2 Continuum limit

In the continuum limit $l_f/R \to 0$, the process becomes a controlled Brownian motion for which an analytical description is affordable. The effect of confining a Wiener process in the cylindrical channel $\mathcal{R}$ is subsumed by an additional drift term, $\bar{u}(\bar{x})$, which we refer to as the control. The Langevin equation for the walker thus reads

$$\frac{d\bar{x}}{dt} = \bar{u}(\bar{x}) + \sqrt{2D}\,\bar{\eta}_t \;, \tag{4.34}$$

where each component of $\bar{\eta}_t$ is an independent white noise. Following the same derivation as in Sec. 1.1.1 and 4.1, one can see that the optimal control is

$$\bar{u}^*(\bar{x}) = 2D\,\nabla \log Z(\bar{x}) \;, \tag{4.35}$$

where Z satisfies the linear version of the HJB equation (inside the cylinder $\mathcal{R}$), which in this case is the Laplace equation

$$\nabla^2 Z = 0 \;,$$

supplemented with Dirichlet boundary conditions everywhere but at the interface with the target $\mathcal{T}$, where it has a finite positive value. Indeed, this corresponds to the HJB equation for the optimal search process with no running cost, but terminal state reward. In the limit of an infinitely long cylinder $H/R \to \infty$, the solution of the HJB equation is

$$Z(\bar{x}) \equiv Z(\rho,\, z) \propto e^{\lambda z}\, J_0(\lambda\,\rho) \;, \tag{4.36}$$

where $\lambda = z_{0,1}/R$, with $z_{0,1}$ being the smallest zero of the Bessel function J_0. We then find that the drift $\bar{u}(\bar{x})$ is only a function of the radial coordinate and takes the form

$$\bar{u}(\rho) = 2D\lambda\,\hat{e}_z - 2D\lambda\,\frac{J_1(\lambda\,\rho)}{J_0(\lambda\,\rho)}\,\hat{e}_\rho \;, \tag{4.37}$$

where $\hat{e}_z$ and $\hat{e}_\rho$ are, respectively, the unit vector pointing longitudinally to the cylinder towards the target and the outward radial unit vector. A complete derivation of the optimal control, Eq. (4.37), is presented in App. A.3.

Density fluctuation along the cylinder . In the continuum limit it is also possible to calculate analytically the statistics of the local density along the cylindrical channel, calculated as the number of jumps counted within a longitudinal section of the cylinder, $[z_0, z_0 + \Delta]$. Due to translational invariance of the process along the cylinder, we can shift the origin of the coordinate system so that $z_0 = 0$. For a diffusive process, this is given by the residence time in such interval, defined as

$$\phi_\Delta = \int_0^\infty \mathrm{d}t\, \mathbb{I}_\Delta(z_t)\,, \tag{4.38}$$

where $\mathbb{I}_\Delta$ is the indicator function of $[0, \Delta]$. If $\Delta \gg R$, one expects that the counting statistics is approximately a Poisson distribution, with average given by the $\Delta/(2D\lambda)$, i.e. the length by the average "velocity", the fluctuations from the homogeneous configurations being unlikely. Conversely, if $\Delta \ll R$, the same average is expected, but deviations from the homogeneous situation are large. Note that in the continuum limit the drift in the z direction and the one in the radial direction are decoupled, so the statistics of ϕ_Δ is determined by the longitudinal drift only. In our estimates of ϕ_Δ, we are interested only in the case $z_0 < 0$, since we assume that the initial condition is infinitely far on the right, so that at $z = 0$ the process has equilibrated on the transverse direction. In App. A.3 it is shown that the generating function for ϕ_Δ is

$$G_\Delta(s,\, z_0 < 0) = \left\langle e^{-s\phi_\Delta} \middle| z_0,\, t = 0 \right\rangle = \frac{4\lambda\,\alpha\, e^{\Delta(\alpha+\lambda)}}{(\alpha+\lambda)^2\, e^{2\Delta\alpha} - (\lambda-\alpha)^2}\,, \tag{4.39}$$

where we denote $\alpha \equiv \alpha(s) = \sqrt{\lambda^2 + s/D}$. Note that G_Δ can be written in the scaling form $G_\Delta(s) = \tilde{g}(\Delta^2 s/D,\, \Delta\,\lambda)$, where $\tilde{g}$ is

$$\tilde{g}(u,\, v) = \frac{4v\sqrt{u+v^2}\, e^{\Delta\left(\sqrt{u+v^2}+v\right)}}{\left(\sqrt{u+v^2}+v\right)^2 e^{2\sqrt{u+v^2}} - \left(\lambda - \sqrt{u+v^2}\right)^2}\,.$$

In particular, the diffusion constant D can be absorbed in the scaling variable u. Hence it follows that the probability density of the residence time ϕ_Δ, denoted $F_\Delta(\phi_\Delta)$, is given by the inverse Laplace transform

$$F_\Delta(\phi_\Delta) = \frac{D}{\Delta^2}\, \tilde{f}\left(\frac{D\,\phi_\Delta}{\Delta^2},\, \Delta\lambda\right)\,, \tag{4.40}$$

where $\tilde{f}$ is the inverse Laplace transform of $\tilde{g}$ with respect to its first variable.

Numerical method

As the transition probability (4.33) is invariant under translation along the z-axis, the distribution of the walker positions in the transverse direction reaches a stationary state:

$$P_{\rm st}(\rho,\theta) = \frac{\rho\, J_0^2(\lambda\rho)}{\pi\, R^2\, [J_0^2(\lambda R) + J_1^2(\lambda R)]}\,, \tag{4.41}$$

that verifies, for all $\rho < R$,

$$\int_0^R d\rho' \int_0^{2\pi} \mathrm{d}\theta'\, u^*(\rho,\theta\,|\,\rho',\theta')\, P_{st}(\rho',\theta') = P_{st}(\rho,\theta)\,,$$

where

$$\begin{aligned}\int_0^{2\pi} d\theta' u^*(\rho,\theta\,|\,\rho',\theta') &= \int_{-\infty}^{+\infty} dz \int_0^{2\pi} \mathrm{d}\theta'\, u^*(\rho,\theta,z\,|\,\rho',\theta',0)\,,\\ &= m^2\, \frac{J_0(\lambda\rho)}{J_0(\lambda\rho')} \begin{cases} I_0(c_\lambda\rho')\, K_0(c_\lambda\rho) & \text{if } \rho > \rho' \\ I_0(c_\lambda\rho)\, K_0(c_\lambda\rho') & \text{else.}\end{cases}\end{aligned}$$

The first point of the path, with arbitrary z coordinate, is sampled from $P_{\rm st}$. The angular variable θ is trivially extracted from the uniform distribution in $[-\pi,\,\pi]$, while the sampling of ρ can be done via inverse transform sampling*. The cumulative distribution function for the marginal $P_{\rm st}(\rho)$ is

$$\int_0^r d\rho\, P_{\rm st}(\rho) = \frac{r\, J_1(\lambda r)}{R\, J_1(\lambda R)}$$

and the inversion can be done numerically with standard (Newton) root finding methods.

We then generate confined trajectories using sampling jumps according to the distribution in Eq. (4.33). It is convenient to introduce the instrumental probability distribution

$$f(\bar{x}'|\bar{x}) = \frac{m^2\, e^{-m\,|\bar{x}'-\bar{x}|}}{4\pi\,|\bar{x}'-\bar{x}|\, C(\bar{x})}\, e^{\lambda(z'-z)}\, \mathbb{I}_{\rho'<R}\,, \tag{4.42}$$

where $C(\bar{x}) \equiv C(\rho) = m^2\big(1 - c_\lambda R\, I_0(c_\lambda\rho)\, K_1(c_\lambda R)\big)/c_\lambda^2$, obviously related to the controlled transition probability by a multiplicative factor, $u(\bar{x}'|\bar{x}) = f(\bar{x}'|\bar{x})\, C(\rho) J_0(\lambda\rho')/J_0(\lambda\rho)$. It is possible to sample from the distribution in Eq. (4.42), by means of the inverse transform

*If u is a random number extracted uniformly in $[0,1)$, if f is a probability density function and F the corresponding cumulative distribution function, the integral equation

$$u = \int_{-\infty}^{y} dx\, f(x) = F(y)$$

has a unique solution $y = F^{-1}(u) \sim f(\cdot)$.

method, and then use the rejection in order to correct for the extra factor. For all $\bar{x}'$ and $\bar{x}$ in the cylinder, $u^*(\bar{x}'\,|\,\bar{x}) \le k\, f(\bar{x}'\,|\,\bar{x})$, where $k = C(\rho)/J_0(\lambda\rho)$, which sets the rejection threshold to $J_0(\lambda\rho') \in [0,1]$. This threshold decreases when the walk gets closer to the boundaries ($\rho' \to R$), where it is vanishingly small in the diffusion limit ($\lambda R \to z_{0,1}$). In practice, then, once the next proposed position $\bar{x}'$ is sampled from $f(\cdot\,|\,\bar{x})$, it is accepted with probability $J_0(\lambda\rho')$.

We here present an algorithm for sampling a three-dimensional variable $\bar{x}' \sim f(\cdot|\bar{x})$, which is entirely based on the inverse transform method. Using the change of variables

$$\begin{cases} \ell = |\bar{x}' - \bar{x}|\,, \\ \xi = \dfrac{z'-z}{\ell} \ \in [-1,1]\,, \\ \tan\varphi = \dfrac{y'-y}{x'-x}\,, \ \ \varphi \in [-\pi,\pi]\,, \end{cases} \tag{4.43}$$

we can rewrite the measure of the jump as

$$f(\bar{x}'|\bar{x})\, d^3x' = \mathbb{I}_{\ell<\ell^*}\, \frac{m^2\, e^{-m\,\ell+\lambda\,\ell\,\xi}}{4\pi\, C(\rho)}\ \ell\, \mathrm{d}\ell\ \mathrm{d}\xi\ \mathrm{d}\varphi\,, \tag{4.44}$$

where $\ell^*(\bar{x},\xi,\varphi)$ is the maximum length that a walker starting from $\bar{x}$ can travel within the cylinder in the direction given by (ξ,φ):

$$\ell^* = \frac{b(\bar{x},\varphi)}{f(\varphi\,|\bar{x})}$$

with $b(\bar{x},\varphi) = \sqrt{R^2 - \rho^2\sin^2(\varphi-\theta)} - \rho\cos(\varphi-\theta)$. It is convenient to write the probability density $f(\cdot|\bar{x})$ as

$$f(\ell,\,\xi,\,\varphi\,|\bar{x}) = f(\ell\,|\,\xi,\,\varphi,\,\bar{x})\, f(\xi\,|\,\varphi,\,\bar{x})\, f(\varphi\,|\,\bar{x})$$

where

$$f(\varphi\,|\,\bar{x}) = \int_0^{+\infty} d\ell\,\ell^2 \int_{-1}^{1} \mathrm{d}\xi\, f(\ell,\,\xi,\,\varphi\,|\bar{x})\,,$$

$$f(\xi\,|\,\varphi,\,\bar{x}) = \frac{1}{f(\varphi\,|\bar{x})} \int_0^{+\infty} d\ell\,\ell^2\, f(\ell,\,\xi,\,\varphi\,|\bar{x})\,,$$

$$f(\ell\,|\,\xi,\,\varphi,\,\bar{x}) = \frac{\ell^2\, f(\ell,\xi,\varphi\,|\bar{x})}{f(\xi\,|\,\varphi,\bar{x})\, f(\varphi\,|\bar{x})}\,.$$

Starting from $\bar{x}$, we thus sample first the angle φ from

$$f(\varphi|\bar{x}) = \frac{m^2}{2\pi\, c_\lambda^2\, C(\rho)}\big[1 - c_\lambda\, b(\bar{x},\varphi)\, K_1\big(c_\lambda b(\varphi)\big)\big]\,, \tag{4.45}$$

then, given the angle φ, we obtain ξ from the pdf

$$f(\xi\,|\,\varphi) = \frac{m^2}{4\pi\, C(\rho)\, f(\varphi)}\, \frac{1 - e^{-(m-\lambda\xi)\ell^*}[1 + (m-\lambda\xi)\ell^*]}{(m-\lambda\xi)^2}.$$

For both φ and ξ we invert the cumulative distribution numerically. Notice that having to proceed numerically does not pose any computational issue, since the distributions have compact support, $\varphi \in [-\pi, \pi]$ and $\xi \in [0, 1]$. Finally, ℓ is sampled from

$$f(\ell\,|\,\xi, \varphi) = \frac{(m - \lambda\,\xi)^2 \quad \ell\, e^{-(m-\lambda\xi)\ell}\, \mathbb{I}_{\ell \le \ell^*}}{1 - e^{-(m-\lambda\xi)\ell^*}\,[1 + (m-\lambda\xi)\ell^*]}\,, \tag{4.46}$$

which we recognize to be a truncated Gamma distribution. Its generating function can be inverted by means of the negative branch of the Lambert function, W_{-1}.

4.3 Monte Carlo simulations of polymer confinement

We simulated trajectories produced by the optimal control probabilities $u^*(x'|x)$, with the algorithm just described. Each trajectory, containing a given number of jumps N, can be thought of as the portion of a polymer that extends from one end to the other end of the infinite cylinder. We have generated 10^4 realizations of the random walk confined in an infinite cylinder, and considered subportions of $N = 10^4$ bonds, for different values of the average jump length of the free polymer, ranging from $\ell_f = 9 \cdot 10^{-5} R$ to $\ell_f = 3R$.

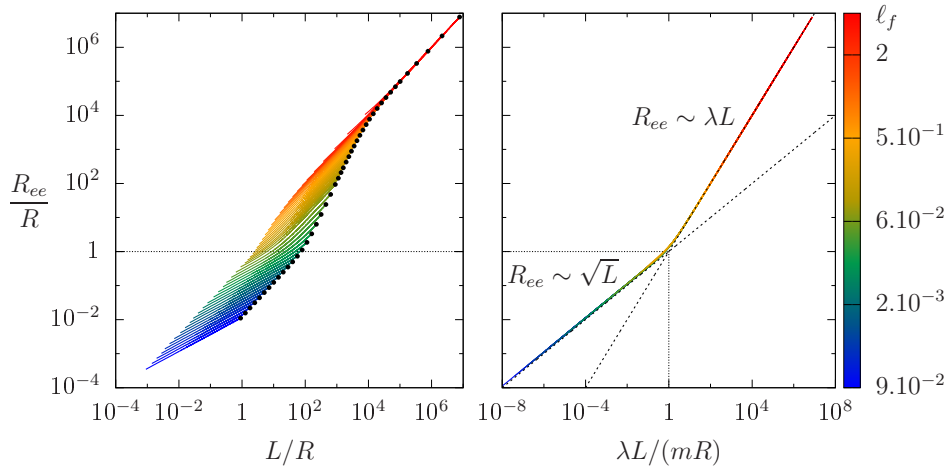


Fig. 4.3. **End-to-end distance versus contour length.** Each colored curve is an average over 10^4 realizations of the walk with a fixed value of the parameter ℓ_f at varying total length of the polymer ($N = 1$ to $N = 10^4$ jumps). The black dots represent the end-to-end distance while varying ℓ_f, at fixed number of jumps $N = 10^4$. On the right, the same curves are plotted against a properly rescaled contour length.

We investigated the behavior of several observables that have a straightforward interpretation in the language of polymer physics: the mean extension of the polymer along the z-axis, $L_z = \langle z_N - z \rangle$, the end-to-end distance $R_{ee} = \sqrt{\langle \|\bar{x}_N - \bar{x}\|^2 \rangle}$ and the contour length $L = \langle \sum_{i=0}^{N-1} \|\bar{x}_{t+1} - \bar{x}_t\| \rangle$. (see e.g. [135–137]). For details, see the published paper [138].

Three distinct regimes are observed, as seen Fig. 4.1. We talk about *weak* confinement, when the average bond length ℓ_c of the constrained walk is still comparable to the free one, and $R_{ee} \ll R$. In this regime the metric properties of the chain are only slightly perturbed with respect to the free case. *Intermediate* confinement is realized when the average bond length is still small relative to the cylinder diameter but the end-to-end distance is comparable or larger, $\ell_c \sim \ell_f \ll R \lesssim R_{ee}$. The walk is under *strong* confinement when the chain is affected even at the scale of individual bonds, $\ell_c \gg \ell_f \simeq R$. These situations can be identified respectively as the *bulk*, the *de Gennes*, and the *Odijk* regimes known from polymer physics [62, 63].

In Fig. 4.3 we show the scaling behaviour of the end-to-end distance versus the contour length. The blue and green curves correspond to the weak confinement case, the red ones are for the strong confinement while the yellow and orange ones (and the black circles) explore the intermediate confinement regime. The emergence of two different scaling regimes becomes evident upon rescaling the chain contour length, $L \to \lambda L/m$, which produces a nearly exact collapse of the data.

In the left panel of Fig. 4.4 we show the deviation of the average length ℓ_c from the

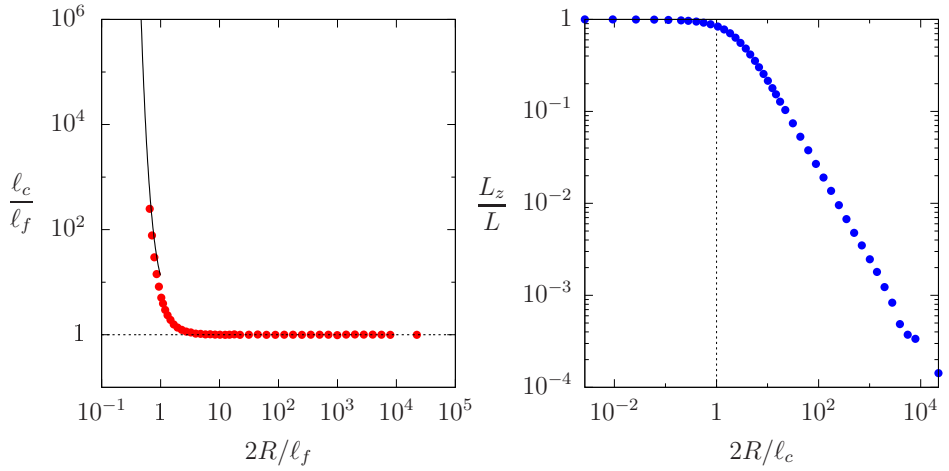


Fig. 4.4. **Jump length and longitudinal extension under confinement.** On the left, numerical values of the ratio beetwen the mean jump length ℓ_c of the confined walker and the free jump size as a function of the parameter $2R/\ell_f$. For $\ell_f \ll 2R$, ℓ_c is essentially equal to the free jump size ℓ_f (dashed line), whereas for $\ell_f \gtrsim 2R$ it increases exponentially fast. On the right, numerical values of the extension L_z/L as a function of the effective channel size $2R/\ell_c$. Each value of L_z and L is obtained over 10^4 realizations of the walk with fixed parameter ℓ_f.

average length ℓ_f in the free case as a function of the strength of the confinement $2R/\ell_f$. Under weak confinement $\ell_f \ll R$, ℓ_c is essentially equal to the mean length ℓ_f of the bonds for the free polymer, whereas under strong confinement $\ell_f \gtrsim R$, ℓ_c increases exponentially fast $\ell_c/R \sim 4(R/\ell_f)^3 \exp[(\ell_f/R)^2]$. This behavior signals the appearance of long stretches of nearly linear polymer configurations in the limit of strong confinement when the free average bond length ℓ_f exceeds the channel radius R. Another customary way to present results on the elongation statistics is given in the right panel of Figure 4.4. It displays L_z/L for different values of $2R/\ell_c$, i.e. upon varying the strength of confinement. We observe two main regimes, for $2R/\ell_c$ smaller or larger than 1.

Fluctuations in the density of jumps along the channel

Visual inspection of the confined paths (see Fig. 4.1) suggests that they are not homogeneously dense along the channel but rather feature an alternation of densely and sparsely occupied regions. To quantify this effect, we considered a measure of the variations of the local density of jumps along the channel. This is defined as the number of jumps that fall inside a cylindrical region of width Δ along the z-axis, corresponding to the variable ϕ_Δ introduced above. Fig. 4.5 clearly highlights the presence of local inhomogeneities in the distribution. In fact, the shape of the right tail suggests that in the confined process there is a higher probability to have regions with higher than average number of points. The location of the peak of the distribution also shows that less dense regions are more likely in the confined case. As detailed in Appendix A.3 the shape of these distributions is well captured by the diffusive optimally controlled process.

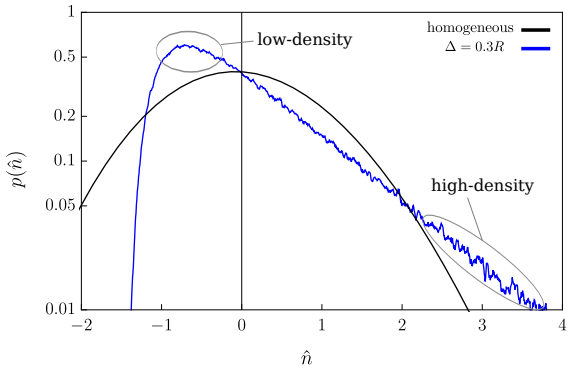


Fig. 4.5. **Inhomogeneity of the local density.** The probability density for $\hat{n} = (\hat{\phi}_\Delta - \langle\hat{\phi}_\Delta\rangle)/\sigma_\Delta$ in the homogeneous case (Poisson distribution, black curve) and for $\Delta = 0.3R$ (blue curve) in the case $\ell_f/R = 4.511 \cdot 10^{-2}$. The encircled regions highlight inhomogeneities in the density of jumps. The right tail indicates the presence of regions with higher density of points and the peak at negative values stems from the presence of regions with lower density of points. Plots for different values of Δ and comparison with analytical results are shown in App. A.3.

4.4 Conclusions and perspectives

In this Chapter, a general framework based on optimal stochastic control for the exact and efficient generation of constrained random walks has been proposed. The formalism can be applied to all Markov processes. In general, one has to solve numerically the linear equation (4.16) and use its solution to obtain the transition probability for the constrained process which can then be directly sampled by any suitable technique. Sometimes it is possible to obtain the exact transition probability and gain exceptional numerical efficiency.

Here, we have applied this method to a discrete-time process constrained inside a cylindrical region. Such a system was inspired by the classical problem of polymer chains confined inside nano-channels. For the minimalistic process considered here, we have shown that the proposed strategy based on optimal control offers an effective way of implementing confining constraints that would otherwise make the problem intractable with simple rejection-based sampling strategies.

A possible extension of this work is to look at self-avoiding paths, which provide more realistic models of polymers. This problem can be cast into an optimal control problem of the same kind as the one discussed in Ch. 1: each point on the path would be regarded as a walker that minimizes a cost function –a reasonable one could account for inter-particle distance. If exact or reasonably approximate solutions could be worked out, this approach would provide an efficient tool for simulations of macromolecules.

Appendices

A.1 Decimation of the fast coarsening dynamics in the Conformational Spread

In this appendix we show that the decimation of the short-living incoherent states in the Glauber-like dynamics of conformational spread model leads to the MWC model, in which the rates generally depend on the full binding state $\{\ell_i\}_{i=1}^N$ (kept frozen at this step). Nevertheless, the equilibrium properties of the resulting Markovian model only depend on the global variable $l = \sum \ell_i$.

For the sake of simplicity, we will describe in detail the case of $N = 2$. In the case of generic N, and in the limit of large coupling, we compute exactly the probability of completion of the switch after the nucleation of one domain, when all the protomers are unbound ($\ell_i = 0$): this provides information about the overall time scale of the concerted switching process; the switching rates in presence of a generic occupation can be then obtained from this by means of the detailed balance condition.

Coarsening with 2 protomers

At frozen binding, the dynamics of the activity variables is equivalent to the one of N spins with nearest-neighbour ferromagnetic interaction with strength J, subjected to a "magnetic field" given by Eq. (2.2):

$$\beta\, h(\sigma_i,\, \ell_i) = h_i\, \sigma - \lambda_i \ , \tag{A.1}$$

where

$$h_i = \frac{\beta}{2}\left[\varepsilon_I - \varepsilon_A - (\varepsilon_b^{(I)} - \varepsilon_b^{(A)})\ell_i\right] , \quad \lambda_i = \frac{\beta}{2}\left[\varepsilon_I + \varepsilon_A - (\varepsilon_b^{A} + \varepsilon_b^{I} + 2\mu)\ell_i\right] . \tag{A.2}$$

According to Eq. (2.4), and to the considerations given in Sec. 2.2, neglecting the binding dynamics, we can decompose the full matrix of the rates K, when $\gamma \to 1$, as the sum the fast contribution K_f,

$$K_f = \omega_f \frac{1+\gamma}{1-\gamma} \begin{pmatrix} 0 & e^{h_2-\lambda_2} & e^{h_1-\lambda_1} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & e^{-h_1-\lambda_1} & e^{-h_2-\lambda_2} & 0 \end{pmatrix} , \tag{A.3}$$

and a slow part K_s,

$$K_s = \omega_f \begin{pmatrix} 0 & 0 & 0 & 0 \\ e^{-h_2-\lambda_2} & 0 & 0 & e^{h_1-\lambda_1} \\ e^{-h_1-\lambda_1} & 0 & 0 & e^{h_2-\lambda_2} \\ 0 & 0 & 0 & 0 \end{pmatrix} , \tag{A.4}$$

where the row and the column index respectively correspond to the final and initial state, labelled as in Fig. 6. We notice that the fast dynamics has two absorbing states, which are the equilibrium configurations in the time-scale separation limit: these states are the coherent configurations, namely those with all the protomers in the same activity state (all spins aligned).

On the slow time scales (much longer than the coarsening process but much shorter than the binding), we can calculate effective rates of passing from one coherent state to the other, given by Eq. (2.13). These rates are limited by the rate of flipping one spin from the starting coherent configuration: this is the slow process, since such transition costs an energy $\sim J \gg \beta$. The rates are then affected by the probabilities that, once this flip has occurred, the process reaches the other coherent state and is not absorbed back in the starting coherent state: such probabilities are completely determined by the fast dynamics. The difficulty in calculating such probability stems from the fact that, for a generic number of protomers N, a huge number of paths contributes, with amplitudes strongly dependent on the binding configuration $\{\ell_i\}$. The general way of proceeding is presented in Refs.[44–46].

In the simple example where $N = 2$, there are only two paths which give contribution to the concerted transition: the flip of the first spin followed by the flip of the second one, or the flip of the second followed by the flip of the first one. Summing up the rates of these

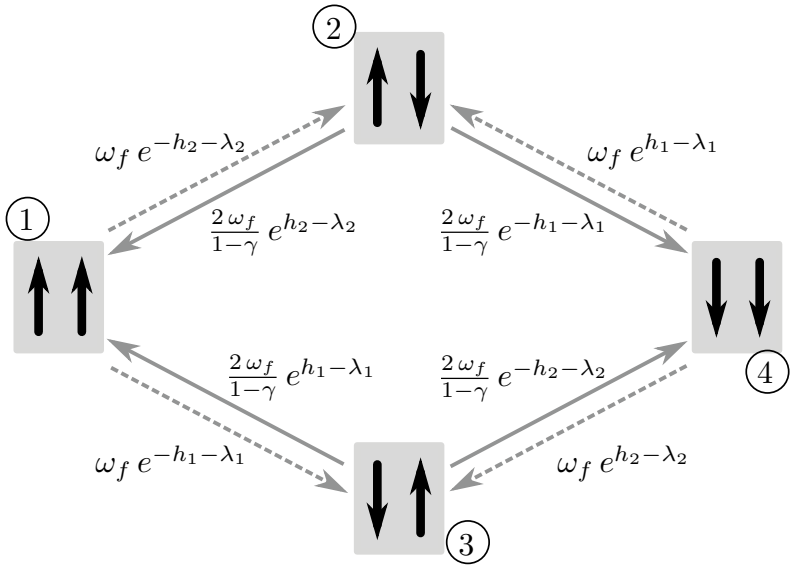


Fig. 6. **Time-scale separation in the dynamics of the spin-activity variable.** Solid and dashed arrows respectively indicate fast and slow transition rates. Numbers within circles are labelling of the states. Configurations with anti-parallel spins are quickly emptied, and coherent configurations are rarely escaped (they are absorbing state for the fast dynamics). In the case of $N = 2$ protomers reproduced here, there are two paths joining one coherent configuration and the other.

possible channels yields the effective rates of the concerted switch [†]:

$$K_c(I \to A) = K_s(\downarrow\downarrow \to \uparrow\downarrow)\, \frac{K_f(\uparrow\downarrow \to \uparrow\uparrow)}{K_f(\uparrow\downarrow \to \downarrow\downarrow) + K_f(\uparrow\downarrow \to \uparrow\uparrow)} \tag{A.5}$$

$$+\, K_s(\downarrow\downarrow \to \downarrow\uparrow)\, \frac{K_f(\downarrow\uparrow \to \uparrow\uparrow)}{K_f(\downarrow\uparrow \to \uparrow\uparrow) + K_f(\downarrow\uparrow \to \downarrow\downarrow)}\ , \tag{A.6}$$

and

$$K_c(A \to I) = K_s(\uparrow\uparrow \to \downarrow\uparrow)\, \frac{K_f(\downarrow\uparrow \to \downarrow\downarrow)}{K_f(\downarrow\uparrow \to \uparrow\uparrow) + K_f(\downarrow\uparrow \to \downarrow\downarrow)} \tag{A.7}$$

$$+\, K_s(\uparrow\uparrow \to \uparrow\downarrow)\, \frac{K_f(\uparrow\downarrow \to \downarrow\downarrow)}{K_f(\uparrow\downarrow \to \downarrow\downarrow) + K_f(\uparrow\downarrow \to \uparrow\uparrow)}\ . \tag{A.8}$$

By substituting Eqs. (2.11) and (2.12) into these expressions, we find

$$K_c(A \to I,\, \{\ell_i\}) = \omega_f \left\{ \frac{e^{-h_1-\lambda_1}}{1 + e^{h_1+h_2-\lambda_1+\lambda_2}} + \frac{e^{-h_2-\lambda_2}}{1 + e^{h_1+h_2+\lambda_1-\lambda_2}} \right\} \tag{A.9}$$

[†]The parametric dependence of the rates on the binding state $\{\ell_i\}$ is understood, but not made explicit in the notation.

and

$$K_c(I \to A,\, \{\ell_i\}) = \omega_f \left\{ \frac{e^{h_1-\lambda_1}}{1+e^{-(h_1+h_2)-\lambda_1+\lambda_2}} + \frac{e^{h_2-\lambda_2}}{1+e^{-(h_1+h_2)+\lambda_1-\lambda_2}} \right\} . \tag{A.10}$$

Even in this simple case with 2 protomers, the first time-scale separation yields effective rates which depend on the full binding state in a non trivial way, and not only on the sum $l = \sum \ell_i$.

However, being detailed balance preserved by the coarse-graining procedure, we have

$$\frac{K_c(A \to I,\, \{\ell_i\})}{K_c(I \to A,\, \{\ell_i\})} = \frac{P_{eq}(I|\{\ell_i\})}{P_{eq}(A|\{\ell_i\})} = e^{-2(h_1+h_2)} , \tag{A.11}$$

consistently with the general formula [‡]

$$P_{eq}(\sigma|\{\ell_i\}) = \frac{1}{Z(\{\ell_i\})}\, e^{\sigma \sum_i h_i} , \tag{A.12}$$

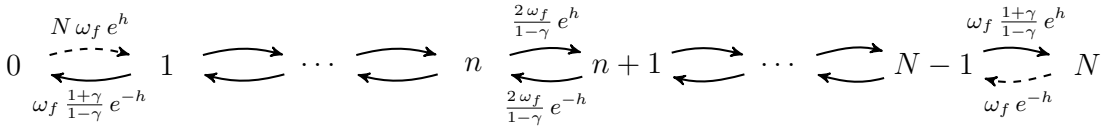
which gives the Boltzmann weights according to the Hamiltonian (2.1) restricted to the coherent configuration (in which case the coupling term is a constant contribution cancelled by the normalization Z). From Eq. (A.1), it is obvious that such Boltzmann weights only depend on the global occupancy $l = \sum \ell_i$.

N-protomer, completely unbound case

In the general case with N protomers, all unbound [§], we are able to map the coarsening process into a simple birth–and–death process which, in the limit $\gamma \to 1$, has site-independent rates. First of all, as we remarked in the main text, because of the time-scale separation, the fast coarsening process does not involve any state but the coherent (which are the long-living states) and the ones with only two domain walls. Then, the coarsening is simply a motion of the domain walls, namely an expansion or contraction of the domain which has been nucleated, until one of the coherent states is reached. The expansion/contraction of this domain can happen by a flip of a spin at its right or left border: if the protomer occupancy ℓ is the same for all sites, it is not important at which side of the domain the expansion/contraction occurs. Therefore, we can label the states visited by the coarsening process just by the number of protomers in the, *e.g.*, active state (spin up), denoted by n: far from the absorbing states $n = 0$ and $n = N$, the rate of increasing or decreasing the number of active protomers is twice the rate of moving a domain wall; the absorption rates from the state $n = 1$ and $n = N - 1$, are the rates of absorbing the two domain walls. The process is schematically represented here,

[‡] The spin variable σ in this expression corresponds to the spin-activity variable of the whole coherent system.

[§] In general, with all the protomers with the same occupancy.



where the (reduced) magnetic field h is the value of h_i given by Eq. (A.2) specified to $\ell_i = 0$, so

$$h = \frac{\beta}{2}(\varepsilon_I - \varepsilon_A) \ .$$

Such (negative) value is responsible for the downward alignment of the spin-activity variables, namely favouring the inactive state. The solid arrows represent the fast rates, while the dashed ones are the slow rates of nucleation of one domain.

In the strong-coupling limit, $\gamma \to 1$, and all the fast rates are asymptotically equal; the coarsening dynamics is then formally described by a birth-and-death process with site-independent rates, defined on the integer numbers between 0 and N, *i.e.* as asymmetric random walk with absorbing boundary conditions. The probability of being absorbed in the state N, starting from the state 1, is easily calculated to be

$$P_{abs}(N \,|\, 1) = e^{h\,(N-1)}\,\frac{\sinh h}{\sinh Nh} \ . \tag{A.13}$$

Once multiplied by the slow exit rate from 0 to 1, this gives the effective rate of switching from the inactive to the active state, in absence of ligands:

$$K_c(I \to A,\, \{\ell_i = 0\}) = N\,\omega_f\,e^{h\,N}\,\frac{\sinh h}{\sinh Nh} = N\,\omega_f\,L^{-1/2}\,\frac{\sinh h}{\sinh Nh} \ , \tag{A.14}$$

where L is the allosteric constant of the N-protomers MWC molecule, defined in the main text as $L = (k_i/k_a)^N = \exp N\beta(\varepsilon_A - \varepsilon_I)$. Similarly, for the opposite switch, one has

$$K_c(A \to I,\, \{\ell_i = 0\}) = N\,\omega_f\,e^{-h\,N}\,\frac{\sinh h}{\sinh Nh} = N\,\omega_f\,L^{1/2}\,\frac{\sinh h}{\sinh Nh} \ . \tag{A.15}$$

These rates are exact up to a correction of order $1-\gamma$, which is exponentially small in the coupling βJ (see Sec. 2.1). We notice that the ratio between these switching rates is the allosteric constant L, as expected from the detailed balance condition and from the Hamiltonian (2.1). The one in Eq. (A.14), or, alternatively, Eq. (A.15), is the overall frequency scale of the switching dynamics in the MWC model. All the other switching rates are found from these ones by applying the detailed balance condition.

A.2 Optimal control of the Ornstein-Uhlenbeck process

Let us consider N controlled Ornstein–Uhlenbeck processes

$$m\,dV_i = \big(f_i - \gamma\,V_i\big)dt + \sqrt{2k_BT\gamma}dW_i^t \ , \tag{A.16}$$

where m is the mass of the particle, f_i is the control (force) on the i-th particle's velocity, γ is the friction coefficient and T is the temperature (Einstein relation). In order to simplify the notation, we express all the quantities in units in which $m = 1$, and we write

$$dV_i = \big(f_i - \gamma\, V_i\big)dt + \sqrt{2D_v}dW_i^t \ , \tag{A.17}$$

where D_v is a "diffusion" coefficient for the velocity. Every agent pays a cost per unit time

$$c_i(\bar{v}) = \frac{\eta}{2}{f_i}^2 + \frac{\kappa}{4}\sum_j |v_j - v_i|^2 \ : \tag{A.18}$$

the first term is proportional to the (rate of) Kullback–Leibler divergence of the controlled path measure from the uncontrolled one, therefore measuring a cost of control in terms of relative entropy; the second term is a velocity-dependent anisotropic term, which favours alignment of all the velocities.

Optimal control equations

At any time t, the vector of controls $\bar{f}$ is chosen to minimize the cost functional

$$\begin{aligned}\mathcal{F}_t &= \Big\langle \sum_i \int_t^\infty dt\, c_i \Big\rangle = \int_t^\infty dt \int d^N v \left[\frac{\eta}{2}\bar{f}^2 - \frac{\kappa}{2}\sum_{i,j}|v_j - v_i|^2\right] P(\bar{v},t) \\ &= \int_t^\infty dt \int d^N v \left[\frac{\eta}{2}\bar{f}^\top \bar{f} + \frac{\kappa}{2}\bar{v}^\top\big(\big(N\mathbb{I}_N - \mathbb{U}\big)\otimes \mathbb{I}_d\big)\bar{v}\right] P(\bar{v},t) \ ,\end{aligned} \tag{A.19}$$

($\mathbb{U}$ is the $N \times N$ matrix with all entries equal to 1) over the stationary state distribution P of the process in Eq. (A.17), and therefore satisfying the (forward) Kolmogorov (Fokker–Planck) equation

$$\partial_t P + \nabla^\top\Big(P\big(\bar{f} - \gamma\,\bar{v}\big)\Big) = D_v\nabla^2 P \ . \tag{A.20}$$

This constrained minimization problem corresponds to the unconstrained optimization problem of the functional

$$\begin{aligned}\mathcal{L}[\bar{f}, P, \Phi, \mu] = \mathcal{F}_t + \int_t^\infty dt \int d^N v\, \Phi(\bar{v},t)\left[\partial_t P + \nabla^\top\Big(P\big(\bar{f} - \gamma\,\bar{v}\big)\Big) - D_v\nabla^2 P\right] \\ - \mu\left(1 - \int d^N v\, P\right)\end{aligned} \tag{A.21}$$

where Φ and μ are Lagrange multipliers enforcing the constraints, respectively, of the dynamics and the normalization of P. The optimality equations are found by seeking for stationary points of the functional $\mathcal{L}$. This is a formulation of the so-called *Pontryagin minimum principle*.

Null variation with respect to Φ trivially yields the dynamics of P, and the one with respect to μ gives the normalization condition. Stationarity with respect to $\bar{f}$ and P yield the expression for the optimal control,

$$\bar{f}^* = \frac{1}{\eta}\,\nabla\Phi \tag{A.22}$$

(in components $f^*_{i,\alpha} = \eta^{-1}\partial_{v_{i,\alpha}}\Phi$), and the HJB equation for Φ,

$$\cancel{\partial_t\Phi} + D_v\nabla\Phi + \frac{1}{2\eta}\big|\nabla\Phi\big|^2 - \gamma\bar{v}^\top\nabla\Phi - \frac{\kappa}{2}\bar{v}^\top\big(\big(N\mathbb{I}_N - \mathbb{U}\big)\otimes\mathbb{I}_d\big)\bar{v} + \mu = 0\ . \tag{A.23}$$

where the time derivative of Φ is neglected because we require optimality at the steady state (for $t \to -\infty$). The HJB equation can be cast into a linear form by means of the Hopf–Cole transformation, $\Phi = 2D_v\eta\log Z$:

$$-\,D_v\nabla^2 Z + \gamma\bar{v}^\top\nabla Z + \frac{\kappa}{4D_v\eta}\bar{v}^\top\big(\big(N\mathbb{I}_N - \mathbb{U}\big)\otimes\mathbb{I}_d\big)\bar{v}\,Z = \tilde{\mu}Z \tag{A.24}$$

where $\tilde{\mu} = \mu/2D_v\eta$, and then

$$\bar{f}^* = 2D_v\nabla\log Z\ . \tag{A.25}$$

By substituting $Z = \exp\big(\gamma\bar{v}^\top\bar{v}/4D_v\big)\,\Psi$, the HJB equation (A.24) reads

$$-\,D_v\nabla^2\Psi + \frac{\gamma}{4D_v}\bar{v}^\top\big(Q\otimes\mathbb{I}_d\big)\bar{v}\,\Psi = \Big(\frac{\gamma}{2}N\,d + \tilde{\mu}\Big)\,\Psi\ , \tag{A.26}$$

where

$$Q_{i,j} = \begin{cases} 1 + (N-1)\frac{\kappa}{\gamma^2\eta} & \text{if } i = j\ , \\ \\ -\frac{\kappa}{\gamma^2\eta} & \text{if } i \neq j\ . \end{cases}$$

Notice that in terms of the function Ψ the control is

$$\bar{f}^* = \gamma\,\bar{v} + 2D_v\nabla\log\Psi\ ,$$

hence the dynamics of $\bar{V}$ is

$$d\bar{V} = 2D_v\nabla\log\Psi\,dt + \sqrt{2D_v}dW_i^t$$

Exact solution for the stationary state

The optimization problem is reduced to finding the eigenvector corresponding to the smallest eigenvalue of the operator H, formally equivalent to the quantum-mechanical Hamiltonian of N d-dimensional coupled harmonic oscillators with frequencies determined by the matrix $Q\otimes\mathbb{I}_d$. The ground state of this quantum-mechanical problem is

$$\Psi = \exp\Big\{-\frac{\gamma}{4D_v}\bar{v}^\top\big(Q^{1/2}\otimes\mathbb{I}_d\big)\,\bar{v}\Big\}\ , \tag{A.27}$$

where $Q^{1/2}$ is the matrix such that $\left(Q^{1/2}\right)^2 = Q$. The corresponding eigenvalue is such that

$$\tilde{\mu} = \frac{\gamma}{2}\Big(\mathrm{tr}\big(Q^{1/2}\otimes \mathbb{I}_d\big) - N\,d\Big) = \frac{\gamma\,d}{2}\Big(\mathrm{tr}\,Q^{1/2} - N\Big)\ . \tag{A.28}$$

The matrix Q can be written as

$$Q = (1+Na)\mathbb{I}_N - a\mathbb{U} \qquad \text{with } a = \frac{\kappa}{\gamma^2\eta}\ ,$$

where $\mathbb{U}$ is the $N\times N$ matrix with all entries equal to 1. Then the eigenvectors of Q are the same as those of $\mathbb{U}$, and the eigenvalues λ are related to those of $\mathbb{U}$, λ_U, by

$$\lambda = 1 + a(N - \lambda_U)\ .$$

The matrix $\mathbb{U}$ is N times the projector onto the space spanned by the vector $r^{(1)} = (1,\,\ldots 1)^\top$. Therefore, its eigenvalues are N, with multiplicity 1 and corresponding eigenvector $r^{(1)}$, and 0 with multiplicity $N-1$ and corresponding eigenvectors lying in the space orthogonal to $r^{(1)}$:

$$\begin{aligned} &\lambda_{\min} = 1 \quad \to \quad r^{(1)} = (1/\sqrt{N},\,\ldots 1/\sqrt{N})^\top\ ,\\ &\lambda_{\max} = 1 + \frac{\kappa\,N}{\gamma^2\eta} \quad \to \quad r^{(2)},\ldots r^{(N)} \in \{r^{(1)}\}^\perp\ . \end{aligned} \tag{A.29}$$

If R is the $N\times N$ orthonormal matrix whose rows are the eigenvectors of Q, an arbitrary power $1/s$ of the matrix Q can be calculated as

$$Q^s = R^\top\cdot \mathrm{diag}\big(\lambda_{\min}^s, \underbrace{\lambda_{\max}^s,\ldots\lambda_{\max}^s}_{N-1}\big)\cdot R\ .$$

By means of the orthogonality relations

$$(R^\top R)_{i,j} = \sum_k r_i^{(k)} r_j^{(k)} = \delta_{i,j} \qquad \text{and} \qquad (RR^\top)_{i,j} = \sum_k r_k^{(i)} r_k^{(j)} = r^{(i)}\cdot r^{(j)} = \delta_{i,j}$$

and by using the explicit expression for $r^{(1)}$, one can show that

$$Q^s{}_{i,j} = \frac{1}{N}\big(\lambda_{\min}^s - \lambda_{\max}^s\big) + \delta_{i,j}\,\lambda_{\max}^s\ .$$

The components of the controls along the eigenvectors are

$$\gamma\,v - \begin{cases} \gamma\,v & \text{along the softest modes,}\\ \gamma\big[1+\frac{\kappa N}{\gamma^2\eta}\big]^{1/2} v & \text{otherwise,}\end{cases}$$

which lead to an anisotropic Ornstein–Uhlenbeck process with different frictions along different normal modes. Notice that when the particles are all moving in the same direction, they are uncontrolled.

The eigenvectors of the matrix $Q\otimes\mathbb{I}_d$ are $r^{(i)}\otimes\hat{e}^{(\alpha)}$ (where $\hat{e}^{(\alpha)}$ is the canonical basis vector in a d-dimensional space). One notices that the softest modes of the harmonic potential (in the direction of the eigenvectors corresponding to $\lambda_{\min}$) are characterized by all the particles oscillating coherently in the same direction. Therefore, large oscillations characterize these modes, while for the orthogonal ones they are much smaller.

The Fokker–Planck equation (A.20) becomes

$$\partial_t P + \nabla^\top \big(P(2D_v \nabla \log \Psi)\big) = D_v \nabla^2 P \tag{A.30}$$

from which one can see that the equilibrium distribution is

$$P_{\rm eq}(\bar{v}) \equiv w(\bar{v}) \propto \Psi^2(\bar{v}) = \exp\left\{ -\frac{\gamma}{2D_v} \bar{v}^\top \big(Q^{-1/2}\otimes\mathbb{I}_d\big)\bar{v} \right\} . \tag{A.31}$$

The covariance between the components α and β of the velocities of particles i and j, at equilibrium, is given by

$$\langle v_{i,\alpha} v_{j,\beta}\rangle_w = \frac{D_v}{\gamma}\, Q_{i,j}^{-1/2}\, \delta_{\alpha,\beta} \ , \tag{A.32}$$

where

$$Q_{i,j}^{-1/2} = \begin{cases} \frac{1}{N}\left[1 + \frac{N-1}{\sqrt{1+N\kappa/\gamma^2\eta}}\right] & \text{if } i = j \ , \\ \frac{1}{N}\left[1 - \frac{1}{\sqrt{1+N\kappa/\gamma^2\eta}}\right] & \text{otherwise} \ . \end{cases} \tag{A.33}$$

It is important to notice that, for large number of agents, the variance of the velocity of each agent is $\sim N^{-1/2}$, much larger than the covariance between the velocities of different agents, scaling like $\sim N^{-1}$.

A.3 Brownian searcher in an infinitely long cylinder

In this appendix we offer the exact solution for the optimal control of a diffusive searcher which is constrained to move in the cylinder $\mathcal{R}$ and reach the target $\mathcal{T}$. In the diffusive limit, it is also possible to study analytically the statistics of some observables. One which is of interest to quantify density fluctuations is the residence time in a longitudinal segment of the cylinder, for which we can derive the exact moment generating function.

Exact solution of the optimal control equations

Following the same calculations as in Sec. 4.1, in which the reference process is an unbiased random walk with diffusion constant D, and the controlled process features an additional drift term, $u(\bar{x})$,

$$\frac{d\bar{x}}{dt} = \bar{u}(\bar{x})\, dt + \sqrt{2D}\, \bar{\eta}(t)$$

The cylinder has a radius R and in the z direction it extends from $-H$ to H. We impose the following boundary conditions for the Laplace equation:

$$Z(R,\,\theta,\,z) = 0 = Z(\rho,\,\theta,\,-H)$$
$$Z(\rho,\,\theta,\,H) = 1$$

The equation is separable and, looking for a solution of the kind $Z(\rho,\,\theta,\,z) = P(\rho)\,\Theta(\theta)\,\zeta(z)$, it can be written as the following equivalent system of coupled ordinary differential equation:

$$\zeta''(z) = \lambda^2\,\zeta(z),$$
$$\Theta''(\theta) = -\mu^2\,\Theta(\theta),$$
$$\rho^2 P''(\rho) + \rho P'(\rho) + (\lambda^2\rho^2 - \mu^2)P(\rho) = 0,$$

where here λ and μ are real parameters. The solution to the equation for ζ which satisfies the Dirichlet boundary conditions on the left end of the cylinder is

$$\zeta(z) = const \times \sinh[\lambda(z+H)]$$

The equation for Θ satisfying the rotational invariance about the longitudinal axis of the cylinder selects the value $\mu = 0$ and is just a constant:

$$\Theta(\theta) = const$$

Finally, the solution for P is the regular Bessel function of first kind of order zero:

$$P(\rho) = const \times J_0(\lambda\,\rho)$$

the allowed values of λ are all and only those for which $P(R) = 0$, so $\lambda_n = z_{0,n}/R$, where we denote by $z_{0,n}$ the n-th zero of $J_0(x)$.
Therefore, the solution of the Laplace equation in the cylindrical geometry specified above is, dropping the θ dependence,

$$Z(\rho,\,z) = \sum_{n=1}^{\infty} c_n\,\sinh[z_{0,n}\,(z+H)/R]\,J_0(z_{0,n}\rho/R)$$

The vanishing conditions at $\rho = R$ and $z = -H$ is already implemented in the solution, while the boundary condition $Z|_{z=H} = 1$ fixes the coefficients c_n as the solution of

$$\sum_{n=1}^{\infty} \tilde{c}_n\,J_0(z_{0,n}x) \equiv \sum_{n=1}^{\infty} \tilde{c}_n\,J_{0,n}(x) = 1 \qquad \forall\, x = \frac{\rho}{R} \in [0,1)$$

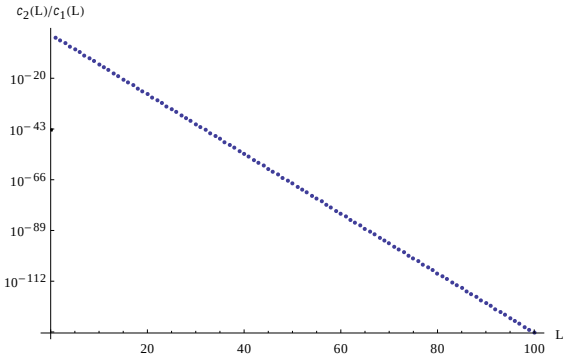


Fig. 7. The ratio between the coefficients of the first subleading term and the leading one against the length of the cylinder $L = 2H$, in logarithmic scale: the suppression of the subleading terms is exponential in L.

where $\tilde{c}_n = c_n \sinh[2\,a\,z_{0,n}/R]$.

The set $\{J_{0,n}(x)\}_{n=1}^{\infty}$ is a basis of the set of function in the interval $[0,1)$ and they are mutually orthogonal therein with respect to the measure $d\mu(x) = x\,dx$ [¶]:

$$\int_0^1 dx\, x\, J_{0,n}(x)\, J_{0,m}(x) = \frac{J_1(z_{0,n})^2}{2}\,\delta_{m,n}$$

The coefficients $\tilde{c}_n$ are therefore found to be the (properly normalized) inner products between the function $f(x) = 1$ and $J_{0,n}(x)$ within $[0,1)$:

$$\tilde{c}_n = \frac{2}{J_1(z_{0,n})^2}\int_0^1 dx\, x\, J_{0,n}(x) = \frac{2}{J_1(z_{0,n})\, z_{0,n}}$$

so that the full solution Z of the Laplace equation is

$$Z(\rho,\, z) = \sum_{n=1}^{\infty} \frac{2}{J_1(z_{0,n})\, z_{0,n}} \frac{\sinh[z_{0,n}\,(z+H)/R]}{\sinh[2H\, z_{0,n}\,/R]}\, J_0(z_{0,n}\rho/R)$$

In the limit $H/R \to \infty$ (infinite cylinder) with finite z, only the first term of the expansion can be retained:

$$Z(\rho,\, z) \propto \exp(z_{0,1}\, z/R)\, J_0(z_{0,1}\,\rho/R)$$

The drift in the effective Langevin dynamics of the conditioned Brownian motion is then

$$\mathbf{u}_*(\rho,\, z) = 2\,D\,\nabla \log Z(\rho,\, z) = 2D\lambda\,\mathbf{e}_z - 2D\lambda\,\frac{J_1(z_{0,1}\rho/R)}{J_0(z_{0,1}\rho/R)}\,\mathbf{e}_\rho,$$

where $\lambda = z_{0,1}/R$.

[¶]For $m \neq n$ one can use Gradshteyn–Ryzhik, 6.521, to check the orthogonality condition; for $m = n$ one can integrate twice by parts using $d(x\,J_1(x))/dx = x\,J_0(x)$.

Density fluctuations

To study the density of beads along the cylinder, we focus on the evolution of the driven Brownian walker (see Sec. 4.2.2) along the z-axis, described by the stochastic process:

$$\mathrm{d}z_t = 2D\lambda\,\mathrm{d}t + \sqrt{2D}\,\mathrm{d}W_t \tag{A.34}$$

where $\mathrm{d}W_t$ is the standard Wiener process. As described in Sec. 4.2.2 the first term is the drift along the z-axis due to confinement. Consider now the interval $[0,\,\Delta]$ along the z-axis, we define the residence time of the walker therein as

$$\phi_\Delta = \int_0^\infty \mathrm{d}t\,\mathbb{I}_\Delta(z_t)\,, \tag{A.35}$$

where $\mathbb{I}_\Delta$ is the characteristic function of $[0,\,\Delta]$, equal to 1 within the interval and 0 otherwise. In general, ϕ_Δ is a random variable, whose statistics depends on the initial conditions of the process. Its moment generating function is defined as

$$G_\Delta(s,\,z_0) = \left\langle e^{-s\phi_\Delta}\,\middle|\,z_0\right\rangle \tag{A.36}$$

and satisfies the stationary Feynman–Kac equation

$$2D\lambda\,\frac{\partial G_\Delta}{\partial z_0} + D\,\frac{\partial^2 G_\Delta}{\partial z_0^2} = s\,\mathbb{I}_\Delta(z_0)\,. \tag{A.37}$$

In Eq. (A.36), the average is taken with respect to the measure of the paths generated by the dynamics in Eq. (A.34). The drift in Eq. (A.34), that drives the process towards increasing values of z_t, fixes the boundary conditions of $G_\Delta(s,\,z_0)$:

$$\begin{cases} G_\Delta(s,\,z_0) \underset{z_0\to+\infty}{\longrightarrow} 1\,, & \text{as } \phi_\Delta \to 0 \\ G_\Delta(s,\,z_0) \underset{z_0\to-\infty}{\longrightarrow} \mathrm{const}(s)\,. \end{cases} \tag{A.38}$$

The general solution of Eq. (A.37) then reads

$$G_\Delta(s,\,z_0) = \begin{cases} A_l\,e^{-2\lambda\,z_0} + B_l & \text{for } z_0 < 0 \\ e^{-\lambda z_0}(A_+ e^{\alpha\,x_0} - A_- e^{-\alpha\,z_0}) & \text{for } z_0 \in [0,\Delta] \\ A_r\,e^{-2\lambda\,z_0} + B_r & \text{for } z_0 > \Delta \end{cases}$$

where $\alpha = \sqrt{\lambda^2 + s/D}$ and the A_i and B_i are constants with respect to z_0. The conditions of Eq. (A.38) then set

$$A_l = 0 \qquad \text{and} \qquad B_r = 1\,. \tag{A.39}$$

The four other constants are uniquely determined by imposing continuity and differentiability of G_Δ at $z_0 = 0$ and $z_0 = \Delta$. Note that, for $z_0 < 0$, $G_\Delta(s,\,z_0)$ doesn't depend on z_0, and thus, the statistics of ϕ_Δ are independent of the specific value of z_0.

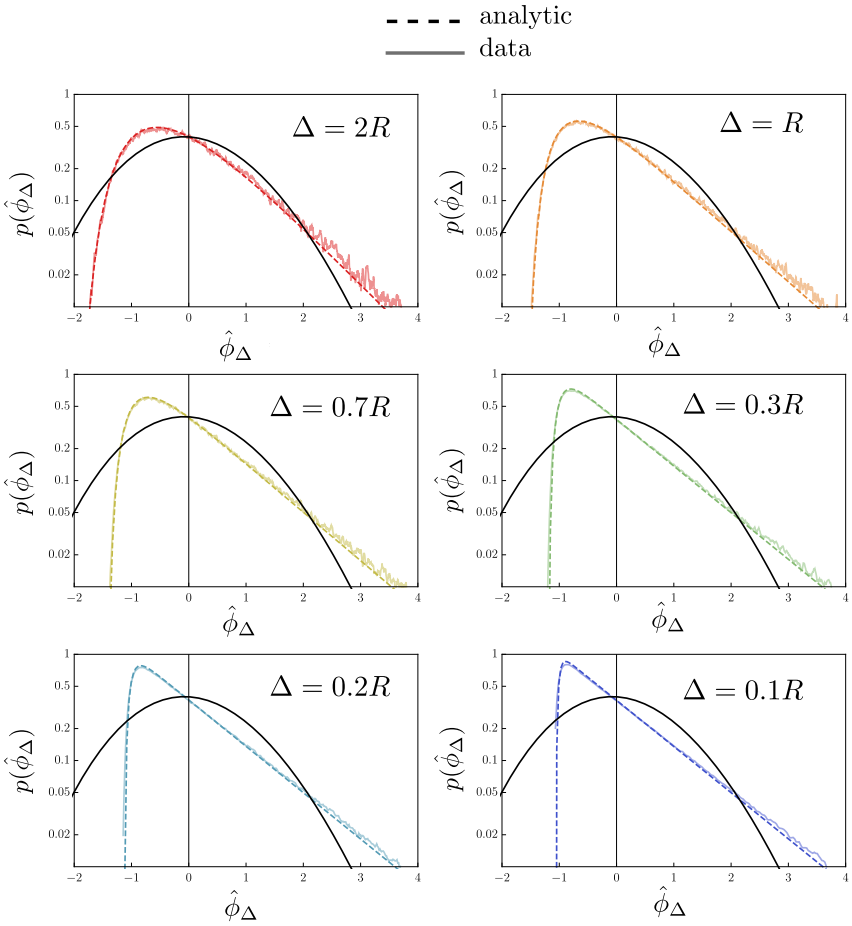


Fig. 8. **Density fluctuations along the channel, diffusive approximation.** Probability density functions of the rescaled residence time $\hat{\phi}_\Delta = (\phi_\Delta - \langle\phi_\Delta\rangle)/\sigma_{\phi_\Delta}$ for decreasing values of Δ (colored curves) in the case $\ell_f/R = 4.511 \cdot 10^{-2}$, compared with the homogeneous case (Poisson distribution with the same average, black curve). As discussed in the main text the peak and the tail of each curve highlight inhomogeneities in the system. The right tail indicates the presence of regions with higher density of points and the peak at negative values stems from the presence of regions with lower density of points.

Bibliography

1. Marler, P. & Griffin, D. R. The 1973 Nobel Prize for Physiology or Medicine. *Science* **182,** 464–466 (1973).

2. Fechner, G. T. *Elemente der psychophysik* (1860).

3. Skinner, B. F. *The behavior of organisms: An experimental analysis* (BF Skinner Foundation, 1990).

4. Pavlov, I. P. & Anrep, G. *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex* (Oxford University Press London, 1928).

5. Thorndike, E. L. Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements* **2** (1898).

6. Sutton, R. S. & Barto, A. G. *Reinforcement learning: an introduction.* (2018).

7. Alon, U. *An introduction to systems biology: design principles of biological circuits* (Chapman and Hall/CRC, 2006).

8. Phillips, R., Kondev, J., Theriot, J. & Garcia, H. *Physical biology of the cell* (Garland Science, 2012).

9. Thompson, D. W. *et al. On growth and form.* (Cambridge Univ. Press, 1942).

10. Marr, D. C. & Poggio, T. *From Understanding Computation to Understanding Neural Circuitry* 1976.

11. Berg, H. C. *E. coli in Motion* (Springer Science & Business Media, 2008).

12. Purcell, E. M. Life at low Reynolds number. *Am. J. Phys* **45,** 3–11 (1977).

13. Berg, H. C. & Purcell, E. M. Physics of chemoreception. *Biophys. J.* **20,** 193–219 (1977).

14. Adler, M., Szekely, P., Mayo, A. & Alon, U. Optimal regulatory circuit topologies for fold-change detection. *Cell systems* **4,** 171–181 (2017).

15. Adler, M. & Alon, U. Fold-change detection in biological systems. *Current Opinion in Systems Biology* **8,** 81–89 (2018).

16. Sourjik, V. & Berg, H. C. Binding of the Escherichia coli response regulator CheY to its target measured in vivo by fluorescence resonance energy transfer. *Proceedings of the National Academy of Sciences* **99,** 12669–12674 (2002).

17. Berg, H. C. *Random Walks in Biology* (Princeton University Press, 1993).

18. Berg, H. C. The Rotary Motor of Bacterial Flagella. *Annu. Rev. Biochem.* **72,** 19–54 (2003).

19. Stocker, R. Reverse and flick: Hybrid locomotion in bacteria. *Proceedings of the National Academy of Sciences* **108,** 2635–2636 (2011).

20. Devreotes, P. N. & Zigmond, S. H. Chemotaxis in eukaryotic cells: a focus on leukocytes and Dictyostelium. *Annual review of cell biology* **4,** 649–686 (1988).

21. Vogel, S. *Life in moving fluids: the physical biology of flow* (Princeton University Press, 1996).

22. Pontryagin, L. S. *Mathematical theory of optimal processes* (CRC Press, 1987).

23. Ogata, K. & Yang, Y. *Modern control engineering* (Prentice hall India, 2002).

24. Bertsekas, D. P., Bertsekas, D. P., Bertsekas, D. P. & Bertsekas, D. P. *Dynamic programming and optimal control* **3** (Athena scientific Belmont, MA, 2005).

25. Fleming, W. H. & Soner, H. M. *Controlled Markov processes and viscosity solutions* (Springer Science & Business Media, 2006).

26. Robbins, H. & Monro, S. A Stochastic Approximation Method. *Ann. Math. Statist.* **22,** 400–407 (1951).

27. Kiefer, J. & Wolfowitz, J. Stochastic Estimation of the Maximum of a Regression Function. *Ann. Math. Statist.* **23,** 462–466 (1952).

28. Bellman, R. On the theory of dynamic programming. *Proceedings of the National Academy of Sciences* **38,** 716–719 (1952).

29. Bellman, R. *Dynamic programming* (Courier Corporation, 2013).

30. Littman, M. L. Reinforcement learning improves behaviour from evaluative feedback. *Nature* **521,** 445–451 (2015).

31. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275,** 1593–1599 (1997).

32. Vikhar, P. A. *Evolutionary algorithms: A critical review and its future prospects* in *2016 Int. Conf. Glob. Trends Signal Process. Inf. Comput. Commun.* (2016), 261–265.

33. Celani, A., Villermaux, E. & Vergassola, M. Odor landscapes in turbulent environments. *Phys. Rev. X* **4,** 041015 (2014).

34. Esteva, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542,** 115 (2017).

35. Shen, D., Wu, G. & Suk, H.-I. Deep Learning in Medical Image Analysis. *Annual Review of Biomedical Engineering* **19,** 221–248 (2017).

36. Radovic, A. *et al.* Machine learning at the energy and intensity frontiers of particle physics. *Nature* **560,** 41–48 (2018).

37. Kremer, J., Stensbo-Smidt, K., Gieseke, F., Pedersen, K. S. & Igel, C. Big Universe, Big Data: Machine Learning and Image Analysis for Astronomy. *IEEE Intelligent Systems* **32,** 16–22 (2017).

38. Mnih, V. *et al.* Human-level control through deep reinforcement learning. *Nature* **518,** 529 (2015).

39. Silver, D. *et al.* Mastering the game of Go without human knowledge. *Nature* **550,** 354 (2017).

40. Reddy, G., Wong-Ng, J., Celani, A., Sejnowski, T. J. & Vergassola, M. Glider soaring via reinforcement learning in the field. *Nature*. ISSN: 1476-4687 (2018).

41. Lieb, E. H., Seiringer, R. & Yngvason, J. in *The Stability of Matter: From Atoms to Stars* 685–697 (Springer, 2001).

42. Lasry, J.-M. & Lions, P.-L. Mean field games. *Jap. J. Math.* **2,** 229–260 (2007).

43. Shamir, M., Bar-On, Y., Phillips, R. & Milo, R. SnapShot: Timescales in Cell Biology. *Cell* **164,** 1302–1302.e1 (2016).

44. E, W. *Principles of Multiscale Modeling* 1st ed. (Cambridge University Press, 2011).

45. Pavliotis, G. A. & Stuart, A. *Multiscale methods: averaging and homogenization* (Springer Science & Business Media, 2008).

46. Bo, S. & Celani, A. Multiple-scale stochastic processes: Decimation, averaging and beyond. *Physics Reports* **670,** 1–59 (2017).

47. Couzin, I. D., Krause, J., James, R., Ruxton, G. D. & Franks, N. R. Collective Memory and Spatial Sorting in Animal Groups. *Journal of Theoretical Biology* **218,** 1–11 (2002).

48. Bialek, W. *et al.* Statistical mechanics for natural flocks of birds. *Proceedings of the National Academy of Sciences* (2012).

49. Hamilton, W. D. Geometry for the selfish herd. *Journal of theoretical Biology* **31,** 295–311 (1971).

50. Gazzola, M., Tchieu, A. A., Alexeev, D., de Brauer, A. & Koumoutsakos, P. Learning to school in the presence of hydrodynamic interactions. *Journal of Fluid Mechanics* **789,** 726–749 (2016).

51. Galton, F. Vox Populi. *Nature* **75,** 450 (Mar. 1907).

52. De Condorcet, N. *et al. Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix* (Cambridge University Press, 2014).

53. Marshall, J. A. R., Brown, G. & Radford, A. N. Individual Confidence-Weighting and Group Decision-Making. *Trends in Ecology and Evolution* **32,** 636–645 (2017).

54. Aurell, E., Mej\'\ia-Monasterio, C. & Muratore-Ginanneschi, P. Optimal Protocols and Optimal Transport in Stochastic Thermodynamics. *Phys. Rev. Lett.* **106,** 250601 (2011).

55. Muratore-Ginanneschi, P., Mejía-Monasterio, C. & Peliti, L. Heat Release by Controlled Continuous-Time Markov Jump Processes. *J. Stat. Phys.* **150,** 181–203 (2013).

56. Caneva, T. *et al.* Optimal Control at the Quantum Speed Limit. *Phys. Rev. Lett.* **103,** 240501 (2009).

57. Chetrite, R. & Touchette, H. Nonequilibrium {M}arkov processes conditioned on large deviations. *Ann. Henri Poincaré* **16,** 2005–2057 (2015).

58. Chetrite, R. & Touchette, H. Variational and optimal control representations of conditioned and driven processes. *J. Stat. Mech. Theory Exp.* P12001 (2015).

59. Angeletti, F. & Touchette, H. Diffusions conditioned on occupation measures. *Journal of Mathematical Physics* **57** (2016).

60. Bucklew, J. *Introduction to rare event simulation* (Springer Science, 2004).

61. Doob, J. L. Conditional Brownian motion and the boundary limits of harmonic functions. *Bull. Soc. Math. France* **85,** 431–458 (1957).

62. Daoud, M. & De Gennes, P. Statistics of macromolecular solutions trapped in small pores. *Journal de Physique* **38,** 85–93 (1977).

63. Odijk, T. The statistics and dynamics of confined or entangled stiff polymers. *Macromolecules* **16,** 1340–1344 (1983).

64. Bonabeau, E., Theraulaz, G., Deneubourg, J.-L., Aron, S. & Camazine, S. Self-organization in social insects. *Tr. in Ecol. & Evol.* **12,** 188–193 (1997).

65. Bonabeau, E., Dorigo, M. & Theraulaz, G. *Swarm intelligence: from natural to artificial systems* **1** (Oxford, 1999).

66. Garnier, S., Gautrais, J. & Theraulaz, G. The biological principles of swarm intelligence. *Sw. Intel.* **1,** 3–31 (2007).

67. Panait, L. & Luke, S. Cooperative multi-agent learning: The state of the art. *Aut. Ag. and Mul.-Ag. Sys.* **11,** 387–434 (2005).

68. Virágh, C. *et al.* Flocking algorithm for autonomous flying robots. *Bioinsp. & Biomim.* **9,** 025012 (2014).

69. Gómez, V., Thijssen, S., Symington, A., Hailes, S. & Kappen, H. J. Real-time stochastic optimal control for multi-agent quadrotor swarms. *Robotics and Aut. Sys. arXiv* **1502,** 04548 (2015).

70. Ani Hsieh, M. e. a. in *Robotics Research: Volume 2* 387–402 (2018).

71. Patlak, C. S. Random walk with persistence and external bias. **15,** 311–338 (1953).

72. Keller, E. F. & Segel, L. A. Model for chemotaxis. **30,** 225–234 (1971).

73. Cover, T. M. & Thomas, J. A. *Elements of information theory* (John Wiley & Sons, 2012).

74. Todorov, E. Efficient computation of optimal actions. **106,** 11478–11483 (2009).

75. Goodfellow, I., Bengio, Y. & Courville, A. *Deep learning* (MIT press Cambridge, 2016).

76. Neu, G., Jonsson, A. & Gómez, V. A unified view of entropy-regularized Markov decision processes (2017).

77. Lieb, E. H. & Liniger, W. Exact analysis of an interacting Bose gas. I. The general solution and the ground state. *Phys. Rev.* **130,** 1605 (1963).

78. Lieb, E. H. Exact Analysis of an Interacting Bose Gas. II. The Excitation Spectrum. *Phys. Rev.* **130,** 1616–1624 (4 May 1963).

79. Howard, R. A. & Matheson, J. E. Risk-sensitive Markov decision processes. *Man. Sc.* **18,** 356–369 (1972).

80. Dvijotham, K. & Todorov, E. A unified theory of linearly solvable optimal control. *Artificial Intelligence (UAI),* 1 (2011).

81. Maddison, C. J. *et al.* Particle Value Functions. *arXiv preprint arXiv:1703.05820* (2017).

82. Gradshteyn, I. S. & Ryzhik, I. M. *Table of integrals, series, and products* (Academic, 2014).

83. Parisi, G. *Statistical field theory* (Addison-Wesley, 1988).

84. Burger, M., Di Francesco, M., Markowich, P. A. & Wolfram, M.-T. *On a mean field game optimal control approach modeling fast exit scenarios in human crowds* in *Dec. & Cont. (CDC), 2013 IEEE 52nd Ann. Conf.* (2013), 3128–3133.

85. Wyatt, T. D. *Pheromones and animal behaviour: communication by smell and taste* (Cambridge, 2003).

86. Hrolenok, B., Luke, S., Sullivan, K. & Vo, C. *Collaborative foraging using beacons* in *Proc. 9th Int. Conf. on Autonomous Agents and Multiagent Systems: Vol. 3* (2010), 1197–1204.

87. Humphreys, D., Davidson, A., Hume, P. J. & Koronakis, V. Salmonella Virulence Effector SopE and Host GEF ARNO Cooperate to Recruit and Activate WAVE to Trigger Bacterial Invasion. *Cell Host & Microbe* **11,** 129–139. ISSN: 1931-3128 (2012).

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

89. Wang, Z.-A. Mathematics of traveling waves in chemotaxis—review paper. *Discrete and Continuous Dynamical Systems Series B* **13,** 601–641 (2013).

90. Passino, K. M. Biomimicry of bacterial foraging for distributed optimization and control. *IEEE Contr. Syst.* **22,** 52–67 (2002).

91. Muller, S. D., Marchetto, J., Airaghi, S. & Kournoutsakos, P. Optimization based on bacterial chemotaxis. *IEEE Tr. Evol. Comp.* **6,** 16–29 (2002).

92. Reynolds, A. Maze-solving by chemotaxis. **81,** 062901 (2010).

93. Nakagaki, T., Yamada, H. & Tóth, Á. Intelligence: Maze-solving by an amoeboid organism. *Nature* **407,** 470–470 (2000).

94. Nakagaki, T. *et al.* Minimum-Risk Path Finding by an Adaptive Amoebal Network. *Phys. Rev. Lett.* **99,** 068104 (6 2007).

95. Lagzi, I., Soh, S., Wesson, P. J., Browne, K. P. & Grzybowski, B. A. Maze solving by chemotactic droplets. *J. Am. Chem. Soc.* **132,** 1198–1199 (2010).

96. Jin, C., Krüger, C. & Maass, C. C. Chemotaxis and autochemotaxis of self-propelling droplet swimmers. **114,** 5089–5094 (2017).

97. Kalinin, Y. V., Jiang, L., Tu, Y. & Wu, M. Logarithmic sensing in Escherichia coli bacterial chemotaxis. **96,** 2439–2448 (2009).

98. Celani, A. & Vergassola, M. Bacterial strategies for chemotaxis response. **107,** 1391–1396 (2010).

99. Littman, M. L. in *Mach. Learn. Proc. 1994* 157–163 (1994).

100. Hohzaki, R. Search games: Literature and survey. *J. Oper. Res. Soc. Japan* **59,** 1–34 (2016).

101. Shi, Y. & Duke, T. Cooperative model of bacterial sensing. *Physical Review E* **58,** 6399 (1998).

102. Shimizu, T. S. *et al.* Molecular model of a lattice of signalling proteins involved in bacterial chemotaxis. *Nat. Cell Biol.* **2,** 792 (2000).

103. Monod, J., Wyman, J. & Changeux, J.-P. On the nature of allosteric transitions: a plausible model. *Journal of molecular biology* **12,** 88–118 (1965).

104. Marzen, S., Garcia, H. G. & Phillips, R. Statistical mechanics of Monod–Wyman–Changeux (MWC) models. *Journal of molecular biology* **425,** 1433–1460 (2013).

105. Eigen, M. *Kinetics of reaction control and information transfer in enzymes and nucleic acids* in *Nobel Symp* **5** (1967), 333–369.

106. Duke, T. & Bray, D. Conformational spread in ring of proteins: a generic mechanism of allosteric switching. *BIOPHYS J* **80,** 248A–248A (2001).

107. Alon, U. *et al.* Response regulator output in bacterial chemotaxis. *The EMBO journal* **17,** 4238–4248 (1998).

108. Duke, T., Le Novere, N. & Bray, D. Conformational spread in a ring of proteins: a stochastic approach to allostery. *Journal of molecular biology* **308,** 541–553 (2001).

109. Krapivsky, P. L., Redner, S. & Ben-Naim, E. *A kinetic view of statistical physics* (Cambridge University Press, 2010).

110. Bai, F. *et al.* Conformational Spread as a Mechanism for Cooperativity in the Bacterial Flagellar Switch. *Science* **327,** 685–689 (2010).

111. Ma, Q., Nicolau Jr, D. V., Maini, P. K., Berry, R. M. & Bai, F. Conformational spread in the flagellar motor switch: a model study (2012).

112. Mochrie, S. G. J., Mack, A. H. & Regan, L. Allosteric conformational spread: Exact results using a simple transfer matrix method. *Phys. Rev. E* **82** (3 2010).

113. Delalez, N. J. *et al.* Signal-dependent turnover of the bacterial flagellar switch protein FliM. *Proc. Natl. Acad. Sci.* **107,** 11347 LP –11351 (2010).

114. Lele, P. P., Branch, R. W., Nathan, V. S. J. & Berg, H. C. Mechanism for adaptive remodeling of the bacterial flagellar switch. *Proc. Natl. Acad. Sci.* **109,** 20018–20022. ISSN: 0027-8424 (2012).

115. Yuan, J., Branch, R. W., Hosu, B. G. & Berg, H. C. Adaptation at the output of the chemotaxis signalling pathway. *Nature* **484,** 233 (Apr. 2012).

116. Weihs, D. Hydromechanics of Fish Schooling. *Nature* **241,** 290 (1973).

117. Chen, C., Liu, S., Shi, X.-q., Chaté, H. & Wu, Y. Weak synchronization and large-scale collective oscillation in dense bacterial suspensions. *Nature* **542,** 210 (2017).

118. Weijer, C. J. Collective cell migration in development. *J. Cell Science* **122,** 3215–3223 (2009).

119. Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I. & Shochet, O. Novel Type of Phase Transition in a System of Self-Driven Particles. *Phys. Rev. Lett.* **75,** 1226–1229 (1995).

120. Gautrais, J. *et al.* Deciphering Interactions in Moving Animal Groups. *PLOS Comput. Biol.* **8,** e1002678 (2012).

121. Calovi, D. S. *et al.* Swarming, schooling, milling: phase diagram of a data-driven fish school model. *New J. Phys.* **16,** 15026 (2014).

122. Toner, J., Tu, Y. & Ramaswamy, S. Hydrodynamics and phases of flocks. *Ann. Phys.* **318,** 170–244 (2005).

123. Marchetti, M. C. *et al.* Hydrodynamics of soft active matter. *Rev. Mod. Phys.* **85,** 1143–1189 (2013).

124. Cavagna, A. *et al.* Scale-free correlations in starling flocks. *Proc. Natl. Acad. Sci.* **107,** 11865–11870 (2010).

125. Attanasi, A. *et al.* Information transfer and behavioural inertia in starling flocks. *Nature Physics* **10,** 691 (2014).

126. Mehta, M. L. *Random matrices* (Elsevier, 2004).

127. Olshanetsky, M. A. & Perelomov, A. M. Classical integrable finite-dimensional systems related to Lie algebras. *Physics Reports* **71,** 313–400 (1981).

128. Olshanetsky, M. A. & Perelomov, A. M. Quantum integrable systems related to Lie algebras. *Phys. Rep.* **94,** 313–404 (1983).

129. Baxter, R. J. *Exactly solved models in statistical mechanics* (Elsevier, 2016).

130. Lomba, E., Weis, J.-J., Almarza, N. G., Bresme, F. & Stell, G. Phase transitions in a continuum model of the classical Heisenberg magnet: The ferromagnetic system. *Phys. Rev. E* **49,** 5169 (1994).

131. Doob, J. L. Conditional Brownian motion and the boundary limits of harmonic functions. *Bull. Soc. Math. Fr.* **85,** 431–458 (1957).

132. Majumdar, S. N. & Orland, H. Effective Langevin equations for constrained stochastic processes. *Journal of Statistical Mechanics: Theory and Experiment* **2015** (2015).

133. Tegenfeldt, J. O. *et al.* The dynamics of genomic-length DNA molecules in 100-nm channels. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 10979–10983 (2004).

134. Van den Broek, B., Wiegerinck, W. & Kappen, H. J. Risk Sensitive Path Integral Control. arXiv: 1203.3523 (2012).

135. Reisner, W. *et al.* Statics and Dynamics of Single DNA Molecules Confined in Nanochannels. *Phys. Rev. Lett.* **94** (19 2005).

136. Yang, Y., Burkhardt, T. W. & Gompper, G. Free energy and extension of a semiflexible polymer in cylindrical confining geometries. *Phys. Rev. E* **76** (1 2007).

137. Wang, Y., Tree, D. R. & Dorfman, K. D. Simulation of DNA Extension in Nanochannels. *Macromolecules* **44,** 6594–6604 (2011).

138. Adorisio, M., Pezzotta, A., de Mulatier, C., Micheletti, C. & Celani, A. Exact and Efficient Sampling of Conditioned Walks. *J. Stat. Phys.* **170,** 79–100 (2018).