Extreme value statistics of stochastic processes : from Brownian motion to active particles

Francesco Mori

université PARIS-SACLAY

Extreme value statistics of stochastic processes : from Brownian motion to active particles

Statistique d'extrêmes des processus : du mouvement brownien aux particules actives

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n° 564, Physique en Île-de-France (PIF)
Spécialité de doctorat : Physique
Graduate School : Physique, Référent : Faculté des sciences d'Orsay

Thèse préparée dans la unité de **LPTMS (Université Paris-Saclay, CNRS)**, sous la direction de **Satya N. MAJUMDAR, Directeur de Recherche CNRS**.

Thèse soutenue à Orsay, le 10 Juin 2022, par

Francesco MORI

Composition du jury

Cécile MONTHUS Directrice de Recherche CNRS, IPhT (Université Paris-Saclay)	Présidente
David S. DEAN Professeur, Université de Bordeaux	Rapporteur & Examinateur
Matteo MARSILI Research Scientist, Abdus Salam International Centre for Theoretical Physics	Rapporteur & Examinateur
Martin R. EVANS Professeur, University of Edinburgh	Examinateur
Joachim KRUG Professeur, University of Cologne	Examinateur
Grégory SCHEHR Directeur de Recherche CNRS, LPTHE (Sorbonne Université)	Examinateur
Raphaël VOITURIEZ Directeur de recherche CNRS, LPTMC (Sorbonne Université)	Examinateur
Satya N. MAJUMDAR Directeur de Recherche CNRS, LPTMS (Université Paris-Saclay)	Directeur de thése

THESE DE DOCTORAT
NNT : 2022UPASP043

Contents

Aknowledgements

I thank my supervisor Satya for sharing with me his great passion for science. He has been an invaluable guide and has become a good friend. He and Grégory, who has been a mentor more than a collaborator, have always been available and passionate in answering my questions. From them, I have learned how curiosity can be the main driver of research. They have been the best guides I could ever ask for and it has been an honor to be their student.

I warmly thank David Dean and Matteo Marsili for accepting to review this thesis and for their many useful comments. I am grateful to Cécile Monthus, Martin Evans, Joachim Krug, and Raphaël Voituriez for agreeing to be part of the jury of my thesis and for their very relevant suggestions.

I thank all the members of the Laboratoire de Physique Théorique et Modèles Statistiques (LPTMS), that has been a second home for the last three years. I am particularly grateful to the current director Alberto Rosso and the former one, Emmanuel Trizac. Their support went well beyond what was required by their duties. I thank Claudine and Karolina for their remarkable effectiveness and their help on many occasions. I am grateful to the students and postdocs of LPTMS (Andrea, Antonio, Fabian, Federico, Félix, Flavio, Gabriel, Giorgio, Hao, Ivan, Jules, Lara, Lenart, Li, Lorenzo, Louis, Luca, Lucas, Marco, Mathieu, Matteo, Mauro, Mert, Nadia, Nina, Pierre, Sap, Saverio, Sebastian, Valerio, Vanja, Vicio). I feel extremely lucky to have been part of this community. I would like to thank my friend Lorenzo for being a great office mate and for always laughing at my silly jokes. A special thanks also to Ana and Benjamin, from whom I learned that life is all about the people that surround you.

I thank the members of my *comité de suivi*, Alain Comtet and Leticia Cugliandolo, for their help and their many useful suggestions. I aknowledge the Laboratoire de Physique Théorique et Hautes Énergies for frequently hosting. I am grateful to my collaborators Marco Biroli, Giacomo Gradenigo, Bertrand Lacroix-A-Chez-Toine, Pierre Le Doussal, and Hendrik Schawe, from whom I have learned a lot. I thank Andriani and Léo for proof checking this thesis. I thank my long-time friends from Arezzo for their unconditional support and the "Italian community in Paris" (Alessandro, Camilla, Chiare, Gaia, Giulia, Giulio, Justine, Leonardo, Matteo, Saverio, Vito, ...) for all the fun we have had throughout the years. I thank my friend Checco for the great moments that we have shared.

I thank my parents from the bottom of my heart. They have always fueled my

curiosity with their love and have allowed me to choose my way in complete freedom. I thank my sister, who showed me how to deal with difficulties with stupid jokes and a good dose of courage. I am grateful to my grandmother Gina, who taught me the value of working hard without taking things too seriously. Finally, I thank Francesca for filling these years with love and joy. She has always pushed me to recognize my worth and overcome my limits. These achievements would have been impossible without her.

List of publications

1 *Time between the maximum and the minimum of a stochastic process.* F. Mori, S. N. Majumdar, and G. Schehr, **Phys. Rev. Lett. 123**, 200201 (2019).

2 *Distribution of the Time Between Maximum and Minimum of Random Walks.* F. Mori, S. N. Majumdar, and G. Schehr, Phys. Rev. E **101**, 052111 (2020).

3 *Distribution of the time of the maximum for stationary processes.* F. Mori, S. N. Majumdar, and G. Schehr, **Europhys. Lett. 135**, 30003 (2021).

4 *Universal survival probability for a d-dimensional run-and-tumble particle.* F. Mori, P. Le Doussal, S. N. Majumdar, and G. Schehr, **Phys. Rev. Lett. 124**, 090603 (2020).

5 *Universal properties of a run-and-tumble particle in arbitrary dimension.* F. Mori, P. Le Doussal, S. N. Majumdar, and G. Schehr, Phys. Rev. E **102**, 042133 (2020).

6 *Universal survival probability for a correlated random walk and applications to records.*, B. Lacroix-A-Chez-Toine and F. Mori, J. Phys. A: Math. Theor. **53**, 495002 (2020).

7 *Condensation transition in the late-time position of a Run-and-Tumble particle.*, F. Mori, P. Le Doussal, S. N. Majumdar, and G. Schehr, Phys. Rev. E **103**, 062134 (2021).

8 *First-order condensation transition in the position distribution of a run-and-tumble particle in one dimension*, F. Mori, G. Gradenigo, and S. N. Majumdar, J. Stat. Mech. 103208 (2021).

9 *Resetting in Stochastic Optimal Control*, B. De Bruyne and F. Mori, preprint arXiv:2112.11416 (2021).

10 *Mean perimeter and area of the convex hull of a planar Brownian motion in the presence of resetting*, S. N. Majumdar, F. Mori, H. Schawe, and G. Schehr, Phys. Rev. E **103**, 022135 (2021).

11 *Number of distinct sites visited by a resetting random walker*, M. Biroli, F. Mori, and S. N. Majumdar, J. Phys. A: Math. Theor. **55**, 244001 (2022).

List of acronyms

- BM Brownian motion
- BB Brownian bridge
- BE Brownian excursion
- CLT central limit theorem
- EVS extreme value statistics
- EW Edwards-Wilkinson
- FBC free boundary conditions
- i.i.d. independent and identically distributed
- LDT large deviation theory
- MCMC Markov chain Monte Carlo
- MFPT mean first-passage time
- PBC periodic boundary conditions
- PDF probability density function
- RTP run-and-tumble particle
- RW random walk
- KPZ Kardar-Parisi-Zhang

Résumé en français

L'incertitude est l'une des caractéristiques déterminantes de l'expérience humaine. À toutes les échelles humaines, des individus aux sociétés globales, beaucoup d'efforts sont consacrés à anticiper les événements futurs et choisir en conséquence la meilleure ligne de conduite. La plupart du temps, nos projections s'avèrent correctes, ce qui renforce la confiance dans nos capacités de prédiction. Cependant, bien que rarement, des événements complètement inattendus se produisent, changeant nos plans. Malgré leur faible probabilité, ces événements extrêmes peuvent jouer un rôle majeur dans nos vies. Pour comprendre l'impact dévastateur des événements extrêmes, il suffit de penser aux épisodes récents tels que les crises financières et les pandémies mondiales. D'autres exemples peuvent être trouvés dans différentes situations, y compris l'ingénierie et la physique. De plus, l'étude des phénomènes météorologiques extrêmes, par exemple les vagues de chaleur, les ouragans et les tsunamis, est devenue vitale en raison des changements climatiques [1–4]. Par conséquent, acquérir une meilleure compréhension de ce type d'événements est une tâche d'une importance fondamentale.

Les caractéristiques statistiques des événements extrêmes ont été systématiquement étudiées au cours du siècle dernier dans le domaine de la théorie des valeurs extrêmes (TVE) [5–7] (voir [8] pour une revue récente à ce sujet). Un modèle stochastique spécifique est typiquement considéré dans le but de calculer la distribution de probabilité des propriétés extrêmes. Par exemple, les quantités pertinentes sont l'ampleur des extrêmes (maxima ou minima) et le moment auquel ils se produisent. Fait intéressant, on obtient souvent des résultats généraux qui s'avèrent indépendants des détails spécifiques du modèle, un phénomène généralement appelé *universalité*. L'un des résultats les plus fondamentaux dans la TVE est le fait que la distribution du maximum de nombreuses variables aléatoires indépendantes et identiquement distribuées appartient toujours à l'une des trois classes d'universalité, indépendamment de la distribution spécifique des variables [7]. L'identification de faits généraux sur les événements extrêmes, valables pour divers systèmes et disciplines, est l'un des principaux objectifs de la TVE.

Au cours des dernières décennies, l'étude des extrêmes a trouvé des applications dans une pléthore de situations. Par exemple, la TVE a été appliquée au domaine de la biologie évolutive, où l'universalité du maximum a été utilisée pour estimer la distribution de l'augmentation de la valeur sélective après une mutation génétique [9]. Les outils de la TVE permettent également d'étudier des systèmes désordonnés [10–12], puisque leur physique à basse température est gouvernée par des niveaux d'énergie

proches de l'état fondamental, c.-à.-d. l'état d'énergie minimale.

Même si les résultats classiques de la TVE traitent de variables aléatoires non corrélées [7], plusieurs solutions exactes ont été obtenues pour le cas le plus difficile où des corrélations sont présentes. Par exemple, des variables extrêmes ont été étudiées pour des interfaces fluctuantes dans les classes d'universalité Edwards-Wilkinson ou Kardar-Parisi-Zhang [13–16]. La TVE a également été appliquée au domaine de la théorie des matrices aléatoires, où il est pertinent d'étudier la distribution de la plus grande valeur propre d'une matrice aléatoire [17–22]. De plus, les méthodes de la TVE ont été utilisées pour comprendre comment de nouveaux records sont établis, par exemple dans les domaines du sport, du climat ou de la finance [23].

L'un des exemples les plus célèbres d'universalité est le théorème central limite (TCL), qui stipule que lorsque de nombreuses variables indépendantes sont additionnées, leur distribution se rapproche de celle d'une variable aléatoire gaussienne [24]. Cependant, le TCL ne décrit que les événements typiques, c.-à-d. ceux proches du pic de la distribution et il ne peut pas être utilisé pour décrire les valeurs atypiques dans les queues de la distribution de probabilité. En effet, dans plusieurs cas, le TCL sous-estime la probabilité d'événements rares, ce qui peut entraîner des quantités qui sont des ordres de grandeur plus importantes que celles attendues. Par conséquent, utiliser le TCL pour guider les décisions peut potentiellement avoir des conséquences désastreuses. Le cadre mathématique qui décrit les queues de la distribution de probabilité en dehors de la plage de validité du TCL est connu sous le nom de théorie de grandes déviations (TGD) [25, 26]. Cette technique d'investigation d'événements inattendus est complémentaire à la TVE et permet de quantifier les probabilités d'événements rares. En particulier, toutes les informations importantes sur la queue de la distribution peuvent être encodées dans une seule fonction appelée *fonction de taux* ou *fonction de Cramér*.

Dans le contexte de la mécanique statistique hors équilibre, les grandes déviations ont été étudiées pour différents systèmes, y compris les matrices aléatoires [19–22]. De plus, la TGD a également été largement utilisée dans le contexte de la thermodynamique stochastique, où elle a fourni les principaux outils pour dériver les relations fondamentales entre les quantités thermodynamiques, telles que l'entropie et le travail, à l'échelle microscopique [27–29]. Fait intéressant, dans certains cas, la fonction de taux contient des informations qui vont au-delà de la quantification de la probabilité d'événements rares. En particulier, il a été observé que les singularités dans la fonction de taux sont généralement associées à une transition de phase dans le système [22,25]. Les transitions de phase sont souvent observées dans la nature, un exemple classique est la transition entre différents états de l'eau. Mathématiquement, les transitions de phase sont associées à une singularité dans l'énergie libre du système. En revanche, lorsque le système n'admet pas de description d'équilibre, la fonction de taux joue le rôle de l'énergie libre, signalant la présence de transitions de phase hors d'équilibre dans le régime de grandes déviations. Une transition de phase du troisième ordre (ce qui signifie qu'une discontinuité est présente dans la dérivée troisième de la fonction de taux) a été observée, par exemple, dans le régime de grandes déviations de la plus grande valeur propre d'une matrice aléatoire [22].

Dans cette thèse, nous étudions les propriétés extrémales d'une large classe de processus stochastiques, y compris le mouvement brownien et les "run-and-tumble particles" (RTP). Ce travail se pose trois objectifs principaux : (i) fournir de nouveaux résultats

exacts, pour améliorer notre compréhension des propriétés extrémales des processus stochastiques, (ii) identifier de nouvelles propriétés universelles, valables pour de nombreux systèmes et (iii) rechercher de nouveaux phénomènes physiques intéressants, par exemple, des transitions de phase dans des systèmes physiques hors d'équilibre. La recherche d'une solution exacte est motivée par le besoin de modèles simples et solubles pour guider notre compréhension dans des situations plus complexes. Notamment, les solutions exactes révèlent souvent de nouveaux aspects fondamentaux du système étudié, y compris des quantités universelles et des transitions de phase inattendues. La simplicité et la généralité des systèmes que nous allons considérer pourraient rendre nos résultats applicables dans une variété de contextes. Avant d'entrer dans le détail de nos résultats, nous présentons brièvement les principaux modèles stochastiques que nous allons considérer.

Le mouvement brownien tire son nom du botaniste écossais Robert Brown qui a décrit ses caractéristiques en étudiant la dynamique de petites particules en suspension dans l'eau. Le mouvement des particules observé par Brown était dominé par les collisions avec les molécules d'eau rapides et chaotiques, résultant en une *marche aléatoire*. En raison de la généralité de ce phénomène, les marches aléatoires apparaissent dans un large éventail de disciplines, de la finance à la physique. La première description mathématique de ce processus est attribuée à Louis Bachelier, qui, dans sa thèse de doctorat *La théorie de la spéculation* publiée en 1900 [30], a introduit un modèle stochastique, qui correspond précisément aux marches aléatoires, décrivant l'évolution des prix en finance. Quelques années plus tard, en 1905, Albert Einsein étudia le mouvement brownien dans sa théorie de la diffusion [31], introduisant ce modèle dans la communauté des physiciens.

Depuis lors, le mouvement brownien est sans doute devenu le processus stochastique le plus étudié, avec des applications en physique, en informatique et en finance. Dans le contexte de la TVE, les marches aléatoires font partie des rares modèles de variables corrélées pour lesquelles des résultats exacts peuvent être obtenus (les positions du processus à différents instants sont fortement corrélées). Par exemple, la distribution de probabilité du maximum global est connue pour un mouvement brownien à une dimension [24, 32–34], ainsi que pour plusieurs généralisations [35, 36]. De plus, la distribution de probabilité du moment auquel une marche aléatoire atteint son maximum global dans une fenêtre de temps fixe peut être calculé exactement [37, 38]. Cette quantité, le temps du maximum, a trouvé des applications dans plusieurs domaines différents, dont la finance [39] et la thermodynamique stochastique [40]. De plus, les statistiques complètes des records se sont avérées universelles pour une large classe de marches aléatoires [23, 41]. Dans cette thèse, nous considérerons différentes propriétés extrémales du mouvement brownien et de plusieurs généralisations, dont le mouvement brownien contraint [16, 42, 43], le mouvement brownien dans un potentiel de confinement et le mouvement brownien avec "resetting" [36].

La deuxième classe de processus stochastiques que nous allons considérer décrit des systèmes de *matière active*. Contrairement à leur homologue passif (par exemple le mouvement brownien), les systèmes actifs sont composés d'unités individuelles capables d'absorber localement l'énergie, la convertissant en travail thermodynamique. Des exemples de matière active peuvent être observés à plusieurs échelles dans la nature, des bactéries [44, 45] aux vols d'oiseaux [46, 47]. Ces dernières années, il y a eu un grand intérêt pour les propriétés statistiques de ces systèmes [48–50].

Afin de décrire le mouvement persistant des particules actives, plusieurs modèles ont été proposés. Par exemple, dans le processus actif d'Ornstein-Uhlenbeck [51], la vitesse des particules est choisie pour être un processus d'Ornstein-Uhlenbeck, dont la fonction d'autocorrélation décroît de manière exponentielle dans le temps. Un autre modèle est le mouvement brownien actif [49], dont l'angle d'orientation effectue un mouvement brownien. Dans cette thèse, nous nous concentrerons principalement sur le modèle RTP, qui décrit le mouvement persistant d'une classe de bactéries, dont l'*Escherichia Coli*. Dans sa formulation la plus simple, les particules se déplacent avec une série de « runs » (périodes de mouvement dans une direction fixe), séparées par des randomisations instantanées de la direction (« tumblings ») à des moments aléatoires. Malgré son apparente simplicité, le calcul exact des propriétés statistiques pertinentes de ce modèle est hautement non trivial. De plus, ce modèle simple présente plusieurs caractéristiques intéressantes dans le cas à plusieurs particules [49, 52]. En particulier, les particules autopropulsées ont tendance à ralentir à des densités élevées (généralement pour des raisons biochimiques), entraînant une séparation de phase entre une phase dense et une phase diluée, même en l'absence d'interactions attractives. Ce phénomène, appelé « motility-induced phase separation » (séparation de phase induite par la motilité), est un effet de non-équilibre sans équivalent en physique d'équilibre [52].

Malgré la popularité du modèle RTP, de nombreuses questions ouvertes restent à résoudre. En effet, contrairement aux processus passifs étudiés depuis longtemps dans le cadre de la TVE, la caractérisation systématique des propriétés extrémales des processus actifs n'a commencé que récemment. Compte tenu de la nature générale de la TVE et de l'omniprésence des processus actifs dans les systèmes vivants, il est fondamental d'améliorer notre compréhension des propriétés extrémales de la matière active. Il est significatif de noter que de nombreuses caractéristiques statistiques surprenantes du modèle RTP, par exemple l'état d'équilibre non-Bolzmann dans un potentiel [53–56], ont été observées même au niveau d'une seule particule. La description analytique du modèle RTP peut être grandement simplifiée dans une dimension, où plusieurs résultats ont été dérivés dans le cadre de la TVE, notamment les propriétés de persistance [54, 57, 58]. Obtenir des résultats exacts dans des systèmes multidimensionnels est un problème mathématique majeur car le processus décrivant la position de la particule n'est pas markovien [1]. L'identification de modèles non-markoviens pour lesquels les propriétés extrémales peuvent être calculées avec précision est en général assez difficile. Ainsi, l'étude du modèle RTP est également pertinente d'un point de vue théorique.

Panorama de la thèse

Cette thèse se concentre sur les statistiques des valeurs extrêmes et les grandes déviations des processus stochastiques.

Le chapitre 1 donne un aperçu des principaux résultats et techniques que nous emploierons dans le reste de la thèse. Dans un premier temps, nous présenterons les résultats classiques de la TVE dans le cas de variables aléatoires indépendantes et identiquement distribuées, en mettant l'accent sur trois observables principales : la valeur

[1] Un processus stochastique est markovien si la probabilité d'un événement ne dépend que du résultat du précédent événement (et non de l'historique complet du processus). Cette propriété simplifie grandement la description mathématique du système correspondant.

du maximum global de la séquence, l'instant auquel ce maximum se produit et la statistique de records. Ensuite, nous généraliserons ces résultats aux systèmes corrélés, à la fois en temps continu et en temps discret.

Dans le chapitre 2, nous nous intéresserons à la distribution des instants auxquels les extrema des processus stochastiques se produisent. Dans une première section, nous étudierons le temps entre un maximum global et le minimum global d'un processus stochastique unidimensionnel. Nous calculerons la distribution exacte dans le cas paradigmatique du mouvement brownien puis nous considérerons plusieurs généralisations. Dans une deuxième section, nous nous intéresserons à la distribution de l'instant auquel le maximum se produit dans des processus stochastiques stationnaires, à l'équilibre et hors équilibre.

Dans le chapitre 3, nous calculerons exactement la probabilité de survie du modèle RTP dans des dimensions arbitraires, montrant que cette quantité est complètement universelle, c.-à-d. indépendante de la dimension du système ainsi que des fluctuations de vitesse de la particule. De plus, nous allons également étendre ce résultat à d'autres observables (temps du maximum et records) et à différentes généralisations du modèle RTP.

Dans le chapitre 4, nous considérerons les propriétés de grandes déviations de la distribution de la position d'une seule RTP, pour une dimension arbitraire et différentes distributions de vitesse. Nous montrerons comment, sous certaines conditions, une transition de condensation peut être observée dans ce système : au-delà d'une valeur seuil du déplacement total de la particule, la trajectoire RTP est dominée par un seul long saut. Cette transition est signalée par une singularité dans la fonction de taux de la distribution de la position du RTP.

Enfin, dans le chapitre 5, nous fournirons d'abord une courte introduction à la théorie des processus stochastiques avec resetting. Ensuite, nous introduirons une nouvelle méthode, basée sur la théorie du contrôle optimal, qui permet de contrôler de manière optimale un système stochastique à travers des redémarrages.

Introduction

Uncertainty is one of the defining features of the human experience. At all human scales, from individuals to global societies, a great deal of effort is put to anticipate future events and accordingly select the best course of action. Most of the time, our guesses turn out to be correct, reinforcing the confidence in our prediction capabilities. However, rarely but surely, unexpected events occur, interfering with our plans. Despite their modest probability, these extreme events tend to play a major role in shaping our lives. To understand the devastating impact of extreme events it is sufficient to think of recent circumstances such as financial crises and global pandemics. Crucially, the study of extreme weather events, e.g., heat waves, hurricanes, and tsunamis, has become vital in the context of climate change [1–4]. Several other examples can be found in different contexts, including engineering and physics. Therefore, gaining a better understanding of this type of event is a task of clear practical importance.

The statistical features of extreme events have been systematically studied during the last century within the field of Extreme Value Statistics (EVS) [5–7] (for a recent review see [8]). In the typical setting, a specific stochastic model is considered with the goal of computing the probability distribution of its extremal properties. Relevant quantities can be the magnitude of the extremes (maxima or minima) and the time at which they occur. Interestingly, one often obtains general results which turn out to be independent of the specific details of the model, a phenomenon called *universality*. One of the fundamental results in EVS is the fact that the distribution of the maximum of many independent and identically distributed (i.i.d.) random variables always belongs to one of three universality classes, independently of the specific distribution of the variables [7]. Identifying general facts about extreme events, valid across systems and disciplines, is one of the goals of EVS.

Throughout the last decades, the study of extremes has found applications in a plethora of situations. For instance, EVS has been applied to the field of evolutionary biology [59–61], where the universality of the maximum has been used to estimate the distribution of the increment in fitness after a genetic mutation [9]. Tools from EVS can be used to study disordered systems [10–12], since their low-temperature physics is governed by energy levels close to the ground state, i.e., the state of minimal energy.

Even though the classical results of EVS deal with uncorrelated random variables [7], several exact solutions have been obtained for the more challenging case where correlations are present. For instance, extremal variables have been investigated for fluctuating interfaces in the Edwards-Wilkinson or the Kardar-Parisi-Zhang universality

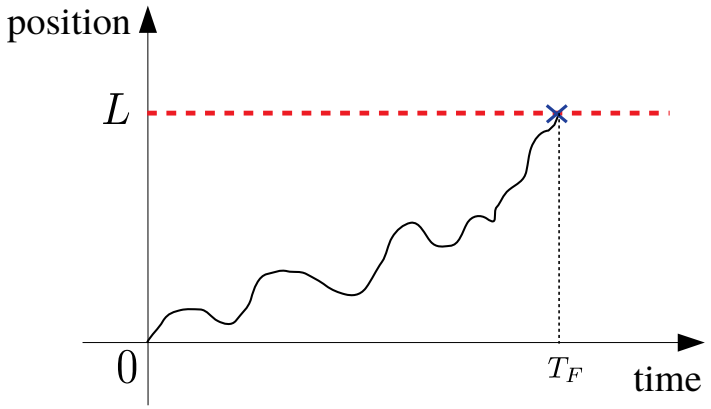

Figure 0.1: Schematic representation of a stochastic process in the presence of an absorbing boundary at position L (dashed red line). The process is absorbed (or "dies") when the boundary at L is reached for the first time. The first passage time is denoted by T_F.

classes [13–16]. EVS has also been applied to the field of random matrix theory, where it is relevant to study the distribution of the largest eigenvalue of a random matrix [17–22].

Moreover, ideas from EVS have been applied to understand how new records[2] are set. The statistical properties of records have been widely investigated both in the mathematical literature [62–65] and, more recently, in the context of physics [23, 41, 66, 67]. Questions related to records naturally emerge in many different disciplines, including climate science [4, 68], finance [69, 70], and theoretical evolution [59–61]. Typical observables related to records are the total number of records in a given sequence and the time between two successive record-breaking events. Interestingly, the distributions of these quantities were shown to be universal for a large class of RWs [23, 41, 67]. Note that the statistics of records are closely related to EVS. For instance, the current record value in a sequence corresponds to the global maximum of the sequence.

Interestingly, the field of EVS has also deep connections with *first-passage* properties, which have been widely investigated both in mathematics [38, 71, 72] and in physics [32, 73]. The first time T_F at which a stochastic process reaches a target state is a quantity of fundamental importance in several situations, from chemical reactions [74] to search processes [75, 76]. The connection with EVS becomes clear with the following simple example. Let us consider a one-dimensional stochastic process evolving in the presence of an absorbing barrier at position L (see Fig. 0.1). The process is absorbed as soon as the barrier is reached for the first time. We denote by $Q_L(t)$ the *survival probability*, i.e., the probability that the process is not absorbed up to time t. Then, denoting by M_t the maximal position reached by the process up to time t, it is clear that the survival probability is exactly the cumulative distribution of M_t, i.e., that $Q_L(t) = \text{Prob.}(M_t \leq L)$. Thus, computing the survival probability $Q_L(t)$ of the process is equivalent to investigating the distribution of the global maximum, which is a central quantity in EVS.

One of the most celebrated examples of universality is the central limit theorem

[2] In a sequence of values, an entry is said to be a record if it is larger than every previous entry.

(CLT), which states that when many independent variables are summed up their distribution approaches that of a Gaussian random variable (the famous "bell shape") [24]. However, the CLT only describes typical events, i.e., those close to the peak of the bell, and it cannot be used to describe the atypical values in the "tails" of the probability distribution. Indeed, in many cases, the CLT underestimates the probability of rare events, which can result in quantities that are orders of magnitude larger than the expected ones. Therefore using it to guide decisions can potentially lead to disastrous consequences. The mathematical framework that describes the tails of probability distributions outside the range of validity of the CLT is known as Large Deviation Theory (LDT) [25, 26]. This theory allows making precise statements about the probability of occurrence of rare events. In particular, all the relevant information about the tail of a distribution can be encoded in a single function known as *rate function* or *Cramér function*.

In the context of nonequilibrium statistical mechanics, large deviations have been investigated for different systems, including single-file diffusion [77], random matrices [19–22], and disordered systems [78–80]. Moreover, LDT has also been widely used in the context of stochastic thermodynamics, where it has provided the main tools to derive fundamental relations between thermodynamic quantities, e.g., entropy and work, at the microscopic scale [27–29]. Interestingly, in some cases, the rate function contains information that goes beyond the quantification of the probability of rare events. In particular, it has been observed that singularities in the rate function are usually associated with phase transitions in the underlying system [22, 25]. Phase transitions are widely observed in nature, a classical example being the transition between different states of water. At equilibrium, phase transitions are associated with a singularity in the free energy of the system. On the other hand, when the system does not admit an equilibrium description, the rate function plays the role of the free energy, signaling the presence of nonequilibrium phase transitions in the large deviation regime. As an example, dynamical phase transitions have been observed in the relaxation to the steady state of stochastic processes with resetting [81].

In this thesis, we investigate the extremal properties of a large class of stochastic processes, including Brownian motion (BM) and run-and-tumble particles (RTPs). This work has three main goals: (i) providing new exact results to improve our understanding of the extremes of stochastic processes, (ii) identifying new universal properties valid across systems, and (iii) searching for new relevant physical phenomena, e.g., phase transitions, in simple models of nonequilibrium physics. The search for exact solutions is motivated by the need for simple and solvable models to guide our understanding of more complex settings. Notably, exact solutions often uncover new fundamental aspects of the system under investigation, including universal quantities and unexpected phase transitions. The simplicity and generality of the systems that we will consider will hopefully make our results applicable in a variety of contexts. Before going into the details of our results, let us briefly introduce the main stochastic models that we will consider.

Brownian motion is named after the Scottish botanist Robert Brown who famously described its features by looking at the dynamics of small particles suspended in water. The motion of the particles observed by Brown was dominated by the collisions with the fast and chaotic water molecules, apparently resulting in a *random walk* (RW).

Due to the generality of this phenomenon, BM appears in a wide range of disciplines, from finance to physics. The first mathematical description of this process is attributed to Louis Bachelier, who in 1900 introduced BM as a stochastic model to describe the evolution of prices in finance [30]. A few years later, in 1905, Albert Einstein studied BM in his theory of diffusion [31], introducing this model to the physics community.

Since then BM has become arguably the most studied stochastic process, with applications to physics, computer science, and finance. In the context of EVS, RWs and BM [3] are among the few models of correlated random variables for which exact results can be obtained. Indeed, the positions of the process at different times are strongly correlated. For instance, the full probability distribution of the global maximum is known for a BM in one dimension [24,32–34], as well as for several generalizations [35,36] (for a comprehensive list of references see the review article [8]). Moreover, the distribution of the time at which an RW reaches its global maximum within a fixed time window can be exactly computed [37,38]. The time of the maximum has found applications in different fields, including finance [39] and stochastic thermodynamics [40]. In this thesis, we will consider different extremal properties of BM [82] and several generalizations, including constrained Brownian motion [16, 42, 43] and BM subject to a confining potential.

A variant of BM which has attracted a lot of attention in the last decade is BM with stochastic resetting (for a recent review, see [83]). Stochastic resetting describes situations in which a dynamical process is restarted at random times from some fixed configuration. Restarting processes can be observed in the different contexts, from computer science [84] to chemical reactions [85]. The simplest model of resetting, introduced in [36], consists of a Brownian particle that is reset to its starting position at a constant rate r. The restarting dynamics lead to several nontrivial features, including the convergence to a nonequilibrium steady state [36, 86, 87] and dynamical phase transitions [88–90]. Besides Brownian motion, resetting has been investigated for several other stochastic systems, from active particles [91–94] to the Ising model [95].

The second class of stochastic processes that we will consider describes systems of *active matter*. At variance with their passive counterpart, e.g., Brownian motion, which is driven by the random-like collisions with the surrounding environment, active systems are composed of individual units that are able to locally absorb energy, converting it into work (e.g., self-propelled motion). Examples of active matter can be observed in nature at different scales, from bacteria [44, 45] to flocks of birds [46, 47]. In recent years, there has been a surge of interest in the statistical properties of these systems [48–51]. At a mathematical level, the noise that drives BM is memoryless, while for active particles the noise has a short but finite memory.

Several models of active particles have been proposed to encode the memory effects in different ways. For instance, in the active Ornstein-Uhlenbeck process [51] the particle velocity is chosen to be an Ornstein-Uhlenbeck process, whose autocorrelation function decays exponentially in time. Another model is the active Brownian particle [49], whose orientation angle performs a BM. In this thesis, we will instead focus on the RTP model, which describes the persistent motion of a class of bacteria, including *Escherichia Coli*. In its simplest formulation, the particles move with a series of "runs" (periods of motion in a fixed direction), separated by instantaneous randomizations of the direction

[3] To be precise, the term random walk usually refers to a discrete-time process, while Brownian motion is defined in continuous time.

("tumblings") at random times. Despite its apparent simplicity, computing exactly the relevant statistical properties of this model is highly nontrivial. Moreover, this simple model displays several interesting features in the many-particle case [49, 52]. In particular, self-propelled particles tend to slow down at high densities (usually for biochemical reasons), leading to phase separation between a dense and a dilute phase, even in the absence of attractive interactions. This phenomenon, called *motility-induced phase separation*, is a nonequilibrium effect with no counterpart in equilibrium physics [52, 96]. Other nontrivial effects, such as clustering at the boundaries [49] and jamming [97, 98], have also been observed.

Despite the popularity of the RTP model, many interesting open questions remain to be addressed even at the single-particle level, i.e., without interactions. Indeed, at variance with passive processes that have been investigated in the context of EVS for a long time, the systematic characterization of extremal properties of active processes has only begun recently. Given the general nature of EVS and the ubiquity of active processes in living systems, it is paramount to improve our understanding of the extremal properties of active matter. Interestingly, many relevant statistical features of the RTP model, e.g., non-Bolzmann steady state in a potential [53–56], have been observed even at the single-particle level. The analytical description of the RTP model can be greatly simplified in one dimension, where several results have been derived within the framework of EVS. These include persistence properties [54, 57, 58] and the distribution of the time of the maximum [99]. Deriving exact results in higher dimensions is a major mathematical challenge because the process describing the position of the particle is non-Markovian [4]. Identifying non-Markovian models for which the extremal properties can be exactly computed is in general quite challenging. Thus, studying the RTP model is also relevant from a theoretical point of view.

Overview of the thesis

This thesis is focused on extreme value statistics and large deviations of stochastic processes.

Chapter 1 provides an overview of the main results and techniques that we will employ in the rest of the thesis. First, we will present the classical EVS results in the case of independent and identically distributed random variables, with a focus on three main quantities: the value of the global maximum of the sequence, the time at which this maximum occurs, and the statistics of records. Then, we will generalize these results to correlated systems, both in continuous and discrete time.

In Chapter 2, we will focus on the distribution of the times at which extrema of stochastic processes occur. In Section 2.1, we will investigate the time between the global maximum and the global minimum of a one-dimensional stochastic process. We will compute the exact distribution in the paradigmatic case of Brownian motion and then we will consider several generalizations. In Section 2.2, we will focus on the distribution of the time of the maximum of stationary stochastic processes, both at equilibrium and out-of-equilibrium.

[4] A stochastic process is Markovian if the probability of an event only depends on the outcome of the previous event (and not on the full history of the process). The Markovian property greatly simplifies the mathematical description of the corresponding system.

In Chapter 3, we will compute exactly the survival probability of the run-and-tumble particle model in arbitrary dimension, showing that this quantity is completely universal, i.e., independent of the dimension of the system as well as of the speed fluctuations of the particle. Moreover, we will also extend this result to other quantities (time of the maximum and record statistics) and different generalizations of the RTP model.

In Chapter 4, we will consider the large-deviation properties of the position distribution of a single RTP, for arbitrary dimension and different speed distributions. We will show how, under certain conditions, a condensation transition can be observed in this system: above a threshold value of the total displacement of the particle, the RTP trajectory is dominated by a single long running phase. This transition is signaled by a singularity in the rate function of the position distribution of the RTP and can be of arbitrary order.

Finally, in Chapter 5, we will first provide a short introduction to the theory of stochastic processes with resetting. Then, we will introduce a new framework, based on optimal control theory, that allows to optimally drive a stochastic system through restarts.

Main results

We provide here a summary of the main results of the thesis. In addition, we also present the results obtained in Articles 10 and 11, which are related to the geometrical properties of resetting Brownian motion and are not discussed in this thesis.

In this thesis, we investigate the extremal properties of three solvable stochastic processes: (i) Brownian motion (BM), (ii) resetting BM, and (iii) run-and-tumble particles (RTPs). These models are defined as follows.

- *Brownian motion.* We will consider a single overdamped Brownian particle, whose position $x(t)$ evolves according to the Langevin equation

$$\frac{dx(t)}{dt} = \sqrt{2D}\eta(t)\,, \tag{0.1}$$

  where $\eta(t)$ is Gaussian white noise with zero mean and correlator $\langle \eta(t)\eta(t')\rangle = \delta(t-t')$ and $D > 0$ is the diffusion constant. We will also consider the case in which the particle is subject to an external confining potential.

- *Resetting Brownian motion.* We will also consider the case of BM with resetting. In this case, the position $x(t)$ of the particle still follows Eq. (0.1) and in addition the particle is reset to some position x_0 with constant rate r. In a small time interval dt, this dynamics can be described as

$$x(t+dt) = \begin{cases} x(t) + \eta(t)dt & \text{with probability } 1 - r dt\,, \\ \\ x_0 & \text{with probability } r dt\,. \end{cases} \tag{0.2}$$

- *Run-and-tumble particle.* We will consider a single RTP, moving in a d-dimensional space. Initially, the particle chooses a random direction (isotropically in space) and a random speed $v_1 > 0$, drawn from some distribution $W(v)$, and starts moving in that direction with constant speed v_1. After a random time τ_1, exponentially distributed with rate γ, the particle tumbles, i.e., it randomizes its direction and chooses a new speed v_2, independently drawn from $W(v)$. After this "tumbling", which is assumed to be instantaneous, the particle "runs" for another time interval τ_2, again exponential with rate γ, then it tumbles again, and so on. This process continues up to time t. Note that by definition the distribution $W(v)$ has positive

support. We will also consider several generalizations of this model, including non-exponential running times and non-instantaneous tumblings.

Even though we have defined BM and resetting BM in one dimension ($d = 1$), these processes can be straightforwardly generalized to $d > 1$.

Time between the maximum and the minimum of stochastic processes

In Section 2.1 of Chapter 2, we consider the distribution $P(\tau = t_{\max} - t_{\min}|T)$ of the time difference τ between the maximum and the minimum of a one-dimensional stochastic process of duration T. This time difference has applications to finance, where one can model the evolution of the price of a stock as a random walk. In the case of Brownian motion with diffusion coefficient D, using a path-decomposition technique we derive the exact result

$$P(\tau|T) = \frac{1}{T} f_{\rm BM}\left(\frac{\tau}{T}\right) , \tag{0.3}$$

where

$$f_{\rm BM}(y) = \frac{1}{|y|} \sum_{n=1}^{\infty} (-1)^{n+1} \tanh^2\left(\frac{n\pi}{2}\sqrt{\frac{|y|}{1-|y|}}\right) . \tag{0.4}$$

We generalize this result to Brownian bridges, i.e., periodic Brownian motions of period T. Moreover, we apply our results to other stochastic processes, including discrete-time random walks and fluctuating interfaces.

From the analysis of the properties of the time τ between the maximum and the minimum, we also obtain a universal result for discrete-time random walks. We prove that the probability that an n-step random walk with continuous and symmetric jumps always remains above its starting position and reaches the global maximum at the final step is $1/(2n)$, independently of the jump distribution and for any $n \geq 1$.

These results have been published in the following articles:

1 *Time between the maximum and the minimum of a stochastic process.* F. Mori, S. N. Majumdar, and G. Schehr, Phys. Rev. Lett. **123**, 200201 (2019).
2 *Distribution of the Time Between Maximum and Minimum of Random Walks.* F. Mori, S. N. Majumdar, and G. Schehr, Phys. Rev. E **101**, 052111 (2020).

Time of the maximum of stationary processes

In Section 2.2 of Chapter 2, we consider a one-dimensional stationary process of duration T. We investigate the distribution $P(t_{\max}|T)$ of the time $t_{\max}$ at which the process reaches its global maximum before time T. We compute exactly the distribution $P(t_{\max}|T)$ for several processes, both at equilibrium (for instance, for the Ornstein-Uhlenbeck process) and out of equilibrium (e.g., for Brownian motion with resetting). Moreover, we find that for a large class of equilibrium processes, corresponding to an overdamped Brownian particle in a confining potential, the distribution of $t_{\max}$ becomes universal at late times.

In addition, we show that for any equilibrium process the probability density function $P(t_{\max}|T)$ is symmetric around its midpoint $t_{\max} = T/2$, i.e., $P(t_{\max}|T) = P(T - t_{\max})$. We also show that this is usually not the case if the process is nonequilibrium. Thus, this symmetry provides a simple method to detect nonequilibrium fluctuations in stationary time series.

These results have led to the following publication:

3 *Distribution of the time of the maximum for stationary processes*, F. Mori, S. N. Majumdar, and G. Schehr, Europhys. Lett. **135**, 30003 (2021).

Universal properties of a run-and-tumble particle

In Chapter 3, we consider a single run-and-tumble particle moving in d dimensions and with arbitrary speed distribution (after each tumbling event the speed of the particle is randomized, with PDF $W(v)$). We investigate the survival probability $S(t)$, i.e., the probability that the x component of the position of the particle does not change sign up to time t. We show that this probability $S(t)$ is completely universal, i.e., independent of d and $W(v)$, and is given by

$$S(t) = \frac{1}{2} e^{-\gamma t/2} \left[I_0\left(\frac{\gamma t}{2}\right) + I_1\left(\frac{\gamma t}{2}\right) \right] , \tag{0.5}$$

where γ is the tumbling rate of the particle and $I_0(z)$ and $I_1(z)$ are modified Bessel functions. Note that this result is exact for any t and not only for late times.

To derive this result, we develop a technique, based on a mapping from the RTP trajectory to a discrete-time random walk, that allows extending the universality to other observables (time of the maximum and records) and to more complicated RTP models (including a persistent random walk model in discrete time).

These results have been published in the following articles:

4 *Universal survival probability for a d-dimensional run-and-tumble particle.* F. Mori, P. Le Doussal, S. N. Majumdar, and G. Schehr, Phys. Rev. Lett. **124**, 090603 (2020).

5 *Universal properties of a run-and-tumble particle in arbitrary dimension.* F. Mori, P. Le Doussal, S. N. Majumdar, and G. Schehr, Phys. Rev. E **102**, 042133 (2020).

6 *Universal survival probability for a correlated random walk and applications to records.* B. Lacroix-A-Chez-Toine and F. Mori, J. Phys. A: Math. Theor. **53**, 495002 (2020).

Large deviations and condensation phase transitions in the RTP model

In Chapter 4, we investigate the distribution $P(\vec{R}, N)$ of the position $\vec{R}$ of a single RTP after N running phases. Considering a family of RTP models, parametrized by the dimension d of the system and the speed distribution $W(v)$ of the particle, we compute the rate function associated with the large deviations of the position of the particle.

Depending on the model, this large deviation regime corresponds to $||\vec{R}|| \sim \mathcal{O}(N)$ or $||\vec{R}|| \sim \mathcal{O}(N^\alpha)$, with $1/2 < \alpha < 1$. We observe that in several cases, this rate function becomes singular at a critical value of the particle displacement. This singularity turns out to be associated with a condensation phase transition. Indeed, upon varying the final position of the particle, a sharp transition occurs between a *fluid phase*, in which the different runs of the particle are roughly of the same order, and a *condensed phase*, in which a long run dominates the trajectory.

To characterize this transition, we identify an order parameter, i.e., the participation ratio (defined precisely in Chapter 4), which we compute exactly as a function of the total displacement of the particle. Moreover, we identify a simple criterion to determine whether or not an RTP model will display a condensation transition and we show that the order of the transition can be tuned by changing the parameters of the model. We verify our theoretical results with numerical simulations, performed with a constrained Markov chain Monte Carlo method, which allows us to sample rare events with probabilities smaller than 10^{-100}.

These results have led to the following articles:

7 *Condensation transition in the late-time position of a Run-and-Tumble particle.* F. Mori, P. Le Doussal, S. N. Majumdar, and G. Schehr, Phys. Rev. E **103**, 062134 (2021).

8 *First-order condensation transition in the position distribution of a run-and-tumble particle in one dimension.* F. Mori, G. Gradenigo, and S. N. Majumdar, J. Stat. Mech. 103208 (2021).

Optimal control with resetting

In Chapter 5, we consider the problem of optimally driving a dynamical system through restarts. We consider a generic resetting system, i.e., a dynamical (stochastic or deterministic) system that can be reset to a known state from time to time. Associating a cost to each resetting, we investigate the optimal restarting procedure to achieve a given goal.

We develop a general framework, analogous to the Hamilton-Jacobi-Bellman equation from optimal control theory [100, 101], to identify the optimal resetting policy in a wide range of situations. We also apply our technique to different control problems, both with an infinite and a finite time horizon.

These results have led to the following preprint article:

9 *Resetting in Stochastic Optimal Control.* B. De Bruyne and F. Mori, preprint arXiv:2112.11416 (2021).

Geometrical properties of random walks with resetting

Finally, we have also worked on the geometrical properties of Brownian motion under stochastic resetting. Although these works are not discussed in this thesis, we include here a short summary of the main results.

In a first work, we have investigated the spatial spread of a two-dimensional resetting BM with constant resetting rate r and diffusion constant D. To do so, we have computed different statistical properties of the convex hull of the process. In particular, we have computed the exact expression, at any time t, of the mean perimeter $\langle L(t)\rangle$ and the mean area $\langle A(t)\rangle$ of the process, showing that

$$\langle L(t)\rangle = 2\pi\sqrt{\frac{D}{r}} f_1(rt)\,, \tag{0.6}$$

and

$$\langle A(t)\rangle = 2\pi\frac{D}{r} f_2(rt)\,, \tag{0.7}$$

where the scaling function $f_1(z)$ and $f_2(z)$ are computed exactly. In particular, for late times, we show that the perimeter and the area of the convex hull grow extremely slowly as $\langle L(t)\rangle \sim \log(rt)$ and $\langle A(t)\rangle \sim \log^2(rt)$ for late times. Our results indicate that, as a consequence of resetting, the convex hull approaches a circular shape at late times.

These results have led to the following article:

10 *Mean perimeter and area of the convex hull of a planar Brownian motion in the presence of resetting.* S. N. Majumdar, F. Mori, H. Schawe, and G. Schehr, Phys. Rev. E **103**, 022135 (2021).

Moreover, in a second work, we have studied the number $V_p(n)$ of distinct sites visited by an n-step random walk on a d-dimensional hypercubic lattice with resetting probability p. We have computed the average number of distinct sites $\langle V_p(n)\rangle$ and we have derived analytically the late-time behavior of $\langle V_p(n)\rangle$ for any d , showing that it grows logarithmically slowly as $\sim [\log(n)]^d$. Moreover, in the case $d = 1$, we have computed the full probability distribution of $V_p(n)$ at late times.

These results have led to the following preprint article:

11 *Number of distinct sites visited by a resetting random walker.* M. Biroli, F. Mori, and S. N. Majumdar, preprint arXiv:2202.04906 (2022).

Chapter 1

A brief introduction to Extreme Value Statistics

In this chapter, we provide an introduction to some basic concepts in EVS. We will only consider a few results that will be useful in the coming Chapters. For a more complete pedagogical introduction to the topic, we recommend the lecture notes [34] and the review article [8]. We will first illustrate the classical results on the universal distributions of the extremes of i.i.d. random variables. We will then show how these facts can be generalized to the case of weakly-correlated random variables. Finally, we will consider the more challenging case of strongly correlated random variables and we will present different techniques to study extremal properties of random walks, both in discrete and continuous time.

Let us first define the observables that we will be interested in. We consider a sequence of T random variables $X_1, \dots, X_T$, where the variable X_i represents the configuration of the process at the discrete time i (see Fig. 1.1). We will be mainly interested in the statistical properties of three observables: the maximal value, the time of this maximum, and the records. The global maximum up to time T is simply defined as

$$M = \max_{1 \le i \le T} X_i \,. \tag{1.1}$$

Similarly, one can also define the minimum

$$m = \min_{1 \le i \le T} X_i \,. \tag{1.2}$$

The time $n_{\max}$ at which the process reaches the global maximum is

$$n_{\max} = \text{argmax}_{1 \le i \le T} X_i \,. \tag{1.3}$$

Similarly, the time of the minimum is defined as

$$n_{\min} = \text{argmin}_{1 \le i \le T} X_i \,. \tag{1.4}$$

Finally, records can be defined as follows. A new (positive) record is set at step i if X_i is larger than all previous entries of the sequence, i.e., if $X_i > X_j$ for all $j < i$ (see Fig. 1.6). Similarly, one can also define negative records, that occur at step i if $X_i < X_j$

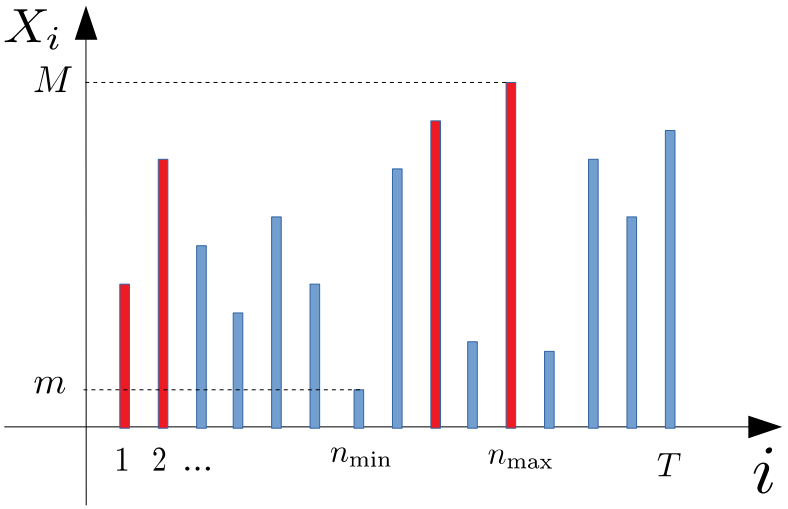

Figure 1.1: Schematic representation of a random sequence X_i as a function of the discrete time i, for $1 \leq i \leq T$. The global maximum M is reached at time $n_{\max}$, while the global minimum m occurs at time $n_{\min}$. Records are highlighted in red.

for all $j < i$. The properties of records are intimately related to extremes. Indeed, it is clear that the last record in the sequence corresponds to the global maximum. There are different relevant observables associated with records, including the number $N_R(T)$ of records up to time T and the times at which the records occur.

The definitions above can be easily generalized to a continuous-time process $X(\tau)$, where now τ is a continuous variable with $0 \leq \tau \leq T$. In particular, the maximum and the minimum are respectively defined as

$$M = \max_{0 \leq \tau \leq T} X(\tau)\,, \quad m = \min_{0 \leq \tau \leq T} X(\tau)\,. \tag{1.5}$$

Similarly the times of the maximum and the minimum read

$$t_{\max} = \mathrm{argmax}_{0 \leq \tau \leq T}\, X(\tau)\,, \quad t_{\min} = \mathrm{argmin}_{0 \leq \tau \leq T}\, X(\tau)\,. \tag{1.6}$$

Records are usually not considered for continuous-time processes since their number typically diverges in continuous time.

1.1 Independent and identically distributed random variables

We first consider the case of i.i.d. random variables, i.e., we assume that the variables $X_1, \ldots X_T$ are uncorrelated and drawn from the same probability density function (PDF) $p(X)$.

Distribution of the maximum

We are interested in the distribution of the global maximum M. As anticipated, this distribution becomes universal at late times [5–7]. This is a classical result that we

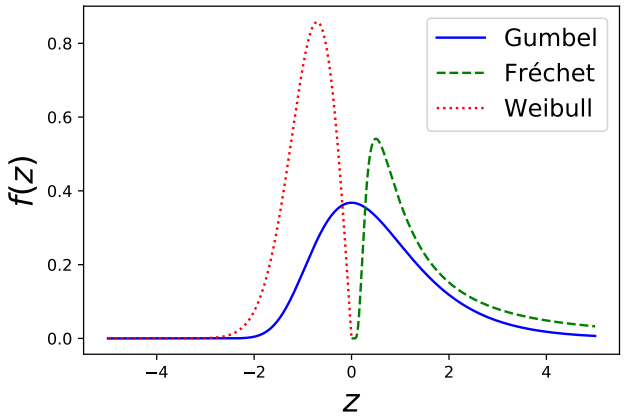

Figure 1.2: Universal scaling functions for the maximum of i.i.d. random variables. The Fréchet distribution shown in figure corresponds to $\mu = 1$, while the Weibull distribution has been obtained with $\alpha = 2$.

present without derivation, for a proof see [7]. Using the fact that the joint distribution of $X_1, \ldots X_T$ factorizes into the product of the marginal distribution, the cumulative distribution of M can be written as

$$S_T(x) = \left[\int_{-\infty}^{x} dy\ p(y)\right]^T = \left[1 - \int_{x}^{\infty} dy\ p(y)\right]^T , \tag{1.7}$$

where $S_T(x) = \text{Prob.}(M \leq x)$. By analyzing this expression in the limit of large T, it is possible to show that

$$S_T(x) \approx F\left(\frac{x - a_T}{b_T}\right) , \tag{1.8}$$

for large x. Here a_T and b_T are scaling factors and are not universal, i.e., they depend on the specific details of the distribution $p(X)$. In particular, a_T represents the typical value of the maximum and b_T describes the fluctuations around it. On the other hand, the scaling function $F(z)$ is completely universal and only depends on the tail behavior of $p(X)$. Interestingly, there are only three possible forms for $F(z)$. Before considering the specific cases, it is interesting to note that the result in Eq. (1.8) implies that, for large T, the maximum behaves as

$$M \approx a_T + b_T\ z\,, \tag{1.9}$$

where z is a random variable with cumulative distribution $F(z)$.

Unbounded distributions with a tail faster than power-law. We start by considering the case where the distribution $p(x)$ decays, for large x, faster than any power, i.e., $p(x) \ll x^{-\mu}$ for any $\mu > 0$. In particular, if $p(x) \sim e^{-x^\alpha}$ with $\alpha > 0$, one finds

$$F(z) = e^{-e^{-z}}\,, \tag{1.10}$$

with $a_T = (\log(T))^{1/\alpha}$ and $b_T = (\log(T))^{1/\alpha - 1}/\alpha$. The corresponding PDF reads

$$f(z) = F'(z) = e^{-z - e^{-z}}\,, \tag{1.11}$$

Table 1.1: Summary of the universality classes for maxima of i.i.d. variables.

Class	$p(x)$	a_T	b_T	$F(z)$
Gumbel	$\sim e^{-x^\alpha}$	$(\log(T))^{1/\alpha}$	$\sim (\log(T))^{1/\alpha-1}$	$e^{-e^{-z}}$
Fréchet	$\sim x^{-1-\mu}$	0	$\sim T^{1/\mu}$	$e^{-z^{-\mu}}\theta(z)$
Weibull	$\sim (c-x)^{\alpha-1}\theta(c-x)$	c	$\sim T^{-1/\alpha}$	$e^{-(-z)^\alpha}\theta(-z)+\theta(z)$

which is the celebrated *Gumbel* distribution [7]. In this case, the global maximum in Eq. (1.9) behaves deterministically at leading order with subleading random fluctuations.

Distributions with a power-law tail. In the case where $p(x) \sim x^{-1-\mu}$ for large x with $\mu > 0$, one gets

$$F(z) = e^{-z^{-\mu}}\theta(z) \tag{1.12}$$

where $\theta(z)$ is the Heaviside theta function, with $\theta(z) = 1$ for $z > 0$ and $\theta(z) = 0$ for $z < 0$. The scaling factors in this case behave as $a_T = 0$ and $b_T \sim T^{1/\mu}$ and the PDF is given by

$$f(z) = \frac{\mu}{z^{\mu+1}} e^{-z^{-\mu}}\theta(z)\,, \tag{1.13}$$

which is known as the *Fréchet* distribution. Thus, for power-law distributions, the late-time behavior of the global maximum is dominated by random fluctuations (see Eq. (1.9)).

Distributions with upper-bounded support. In the case where $p(x) \sim (c-x)^{\alpha-1}\theta(c-x)$ for $x \to c^-$ with $\alpha > 0$, one gets

$$F(z) = \begin{cases} e^{-(-z)^\alpha} & \text{for } z \le 0\,, \\ \\ 1 & \text{for } z > 0\,, \end{cases} \tag{1.14}$$

The scaling factors are $a_T = c$ and $b_T \sim T^{-1/\alpha}$. The PDF in this case is given by

$$f(z) = \alpha(-z)^{\alpha-1} e^{-(-z)^\alpha}\theta(-z)\,, \tag{1.15}$$

which is known as the *Weibull* distribution. Thus, in this case, the maximum will approach the value c for large T. The three universality classes are summarized in Table 1.1 and the corresponding universal scaling functions $f(z)$ are shown in Fig. 1.2.

Time of the maximum

Since the variables are i.i.d., the probability that one particular value is the global maximum is simply $1/T$. Thus, the distribution of the time $n_{\max}$ of the maximum is uniform in the interval $[1, T]$

$$P(n_{\max}|T) = \frac{1}{T}\,. \tag{1.16}$$

Note that here we are assuming that the marginal distribution $p(X)$ of the variables $X_1, \ldots, X_N$ is continuous. Similarly, one can show that the distribution of the time $n_{\min}$ of the minimum is also uniform, i.e., that

$$P(n_{\min}|T) = \frac{1}{T}\,. \tag{1.17}$$

Moreover, the joint distribution of $n_{\max}$ and $n_{\min}$ simply reads

$$P(n_{\max}, n_{\min}|T) = \frac{1}{T(T-1)}\,, \tag{1.18}$$

if $n_{\max} \neq n_{\min}$ and $P(n_{\max}, n_{\min}|T) = 0$ if $n_{\max} = n_{\min}$ (since the maximum and the minimum of the sequence cannot occur at the same time). From this, one can easily compute the distribution of quantities that depend on both $n_{\max}$ and $n_{\min}$. One of these is the time $\tau = n_{\min} - n_{\max}$ between the global maximum and the global minimum (with $-T < \tau < T$). One can easily show that

$$P(\tau|T) = \frac{T - |\tau|}{T(T-1)} \tag{1.19}$$

for $\tau \neq 0$ and $P(\tau|T) = 0$ for $\tau = 0$.

From the distribution in Eq. (1.18) one can also compute the covariance function of the variables $n_{\max}$ and $n_{\min}$. We recall that the covariance function is defined as

$$\mathrm{cov}(n_{\max}, n_{\min}) = \langle n_{\max} n_{\min}\rangle - \langle n_{\max}\rangle\langle n_{\min}\rangle\,. \tag{1.20}$$

Using the expression in Eq. (1.18), it is easy to show that

$$\mathrm{cov}(n_{\max}, n_{\min}) = -\frac{T+1}{12}\,. \tag{1.21}$$

Thus, the times $n_{\max}$ and $n_{\min}$ are negatively correlated. These anti-correlations simply arise from the fact that the times of maximum and minimum cannot occur at the same time, i.e., $n_{\max} \neq n_{\min}$.

The number of records

With a similar argument, one can easily compute the average number of records. Indeed, the total number of records up to step T can be written as

$$N_R(T) = \sum_{i=1}^{T} \sigma_i\,, \tag{1.22}$$

where

$$\sigma_i = \begin{cases} 1 & \text{if } X_i \text{ is a record}\,, \\ \\ 0 & \text{otherwise}\ . \end{cases} \tag{1.23}$$

Taking the average of Eq. (1.22), we get

$$\langle N_R(T)\rangle = \sum_{i=1}^{T} \mathrm{Prob.}(X_i \text{ is a record})\,, \tag{1.24}$$

where the symbol $\langle \ldots \rangle$ indicates the average over all realizations of the variables. Since the variables are i.i.d. one simply has

$$\mathrm{Prob.}(X_i \text{ is a record}) = \frac{1}{i}\,, \tag{1.25}$$

and thus

$$\langle N_R(T)\rangle = \sum_{i=1}^{T} \frac{1}{i} \approx \log(T) + \gamma\,, \tag{1.26}$$

where $\gamma = 0.5772\ldots$ is the Euler-Mascheroni constant.

Thus, the average number of records grows very slowly with the sequence length T. This is a consequence of the fact that it becomes harder and harder to set a new record for late times. Indeed, the rate of new records in Eq. (1.25) decays as $1/i$ with the discrete time i.

Interestingly, it is possible to compute the full distribution of the number $N_R(T)$ of records up to time T. In particular, for large T, one can show that this distribution approaches the Gaussian form [67]

$$\text{Prob.}(N_R(T) = N) \approx \frac{1}{\sqrt{2\pi\log(T)}} \exp\left[-\frac{(N-\log(T))^2}{2\log(T)}\right]. \tag{1.27}$$

Thus, for late times, the distribution of $N_R(T)$ concentrates around the mean value $\langle N_R(T)\rangle \approx \log(T)$ with fluctuations which scale as $\sqrt{\log(T)}$. Note that these properties are completely independent of the marginal distribution $p(X)$ (even for finite T).

1.2 Correlated random variables

Computing the extremal properties of correlated random variables is in general a challenging task. Nevertheless, there exist cases in which one can make progress. The first example that we will consider is that of weakly correlated random variables, where some of the results obtained in the case of i.i.d. variables turn out to be still applicable [8]. There are also few models with strong correlations for which the statistical properties of extremes can be computed exactly. In particular, we will focus on the case of random walks.

1.2.1 Weak correlations

Here we consider the case where the correlations decay exponentially as

$$\langle x(t)x(t')\rangle - \langle x(t)\rangle\langle x(t')\rangle \sim e^{-|t-t'|/\xi}\,, \tag{1.28}$$

where ξ is the correlation length. In this case, one can use the following "blocking" argument, valid for $T \gg 1/\xi$. We first regroup the variables in $N \approx T/\xi$ blocks of duration ξ (see Fig. 1.3). The first block contains the variables $x(t)$ for $0 < t < \xi$, the second block the variables $x(t)$ for $\xi < t < 2\xi$, and so on. We define M_i, with $1 \le i \le N$, as the maximum within block i. Then, the global maximum M can be written as

$$M = \max_{1\le i\le N} M_i\,. \tag{1.29}$$

Moreover, since the size of the blocks coincides with the correlation length, variables belonging to different blocks are roughly uncorrelated. As a consequence, the local maxima $M_1\,,\ldots\,,M_N$ can be considered as independent variables. Hence, one can apply

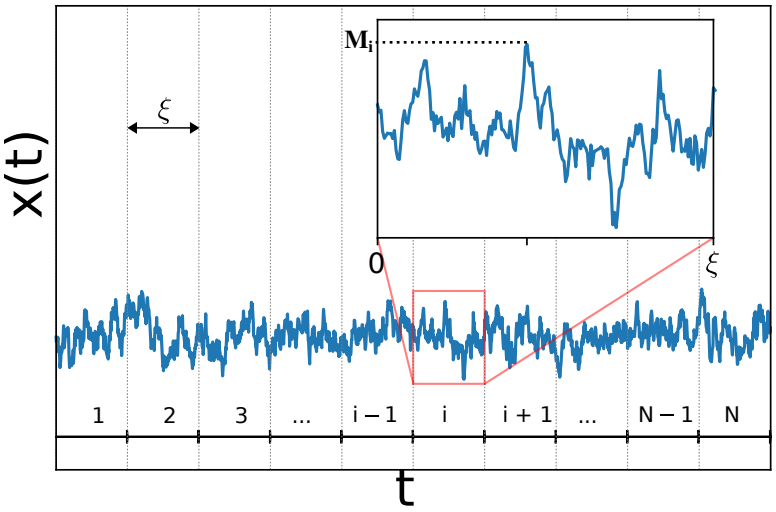

Figure 1.3: Typical realization of a weakly-correlated stochastic process $x(t)$ as a function of time t. The process is divided into N blocks of duration ξ, where ξ is the correlation time of the process. The maximal value within block i is denoted by M_i. Then the variables $M_1, \ldots, M_N$ can be considered independent and identically distributed.

the known results for i.i.d. variables and obtain the distribution of M. Note that the only missing piece of knowledge is the tail behavior of the distribution $p(m)$ of the local maxima m. However, the expression of $p(m)$ for large m can generally be guessed. For instance, in the case of stationary processes, one can assume that the tail behavior of $p(m)$ is the same as the one of the distribution of the variables X_i.

Using the same argument, one also finds that the distribution of the time of the maximum is approximately uniform

$$P(n_{\max}|T) \approx \frac{1}{T}\,. \tag{1.30}$$

As discussed in Chapter 2, this typically turns out to be correct only in the bulk of the distribution, i.e., where $n_{\max}, T \gg 1/\xi$ with $0 < n_{\max}/T < 1$. In the edge regimes, $n_{\max} \to 0$ and $n_{\max} \to T$, the result obtained with the blocking argument is often not correct. Finally, note also that the blocking argument above can be used to show that for weakly correlated random variables the average number of records grows logarithmically with T.

1.2.2 Strong correlations: Brownian motion

When strong correlations are present it is in general harder to obtain exact results. Here we consider the case of random walks, both in the discrete-time description and the continuous-time one, corresponding to BM. There are several other models of strongly correlated variables for which analytical progress is possible, including fluctuating interfaces [13–16] and random matrices [17–22].

Distribution of the maximum

We consider a one-dimensional BM $x(\tau)$ with $0 \leq \tau \leq T$, evolving in time according to the Langevin equation

$$\frac{dx}{dt} = \sqrt{2D}\eta(t)\,, \tag{1.31}$$

where $\eta(t)$ is Gaussian white noise with zero mean and correlator $\langle \eta(t)\eta(t')\rangle = \delta(t-t')$. The constant $D > 0$ is usually called *diffusion coefficient*. For simplicity we also assume $x(0) = 0$. The correlation function of BM reads

$$\langle x(t)x(t')\rangle = 2D\min(t,t')\,, \tag{1.32}$$

thus the process is strongly correlated.

For this model, the distribution $P(M|T)$ of the maximum can be computed as follows. We define $Q^L(x,t)$ as the probability that the process remains below position L up to time t, starting from position x. This quantity $Q^L(x,t)$ is usually called *survival probability*, since one asserts that the process is immediately "killed" when it arrives at position L. This then selects only the trajectories which stay below position L up to time t. The probability $Q^L(x,t)$ satisfies the backward Fokker-Planck equation

$$\partial_t Q^L(x,t) = D\partial_{xx} Q^L(x,t)\,. \tag{1.33}$$

The first boundary condition is

$$Q^L(L,t) = 0\,, \tag{1.34}$$

since if the process starts at L, it will immediately go above L. The second boundary condition simply reads

$$Q^L(-\infty,t) = 1\,, \tag{1.35}$$

since if the process starts from $-\infty$, it will never reach position L in finite time. Similarly, the initial condition is $Q^L(x,0) = 1$. Solving this differential equation, we find [32]

$$Q^L(x,t) = \mathrm{erf}\left(\frac{L-x}{\sqrt{4Dt}}\right)\theta(L-x)\,. \tag{1.36}$$

Interestingly, the survival probability is related to the distribution of the global maximum M. Indeed, it is clear that

$$Q^L(0,t) = \mathrm{Prob.}(M \leq L) = \mathrm{erf}\left(\frac{L}{\sqrt{4Dt}}\right)\theta(L)\,. \tag{1.37}$$

Thus, differentiating with respect to L, we find that the distribution of the global maximum reads

$$P(M|T) = \frac{1}{\sqrt{\pi DT}} e^{-M^2/(4Dt)}\theta(M)\,. \tag{1.38}$$

Note that the PDF of the maximum M has no support in $M < 0$ because the process is starting from $x = 0$ and thus the maximum cannot be negative. From this expression, it is easy to compute the average maximum

$$\langle M\rangle = \frac{2}{\sqrt{\pi}}\sqrt{DT}\,. \tag{1.39}$$

We next consider a discrete time random walk x_k of T steps. The position of the walker evolves according to

$$x_k = x_{x-1} + \eta_k \,, \tag{1.40}$$

where the steps $\eta_1 \,, \dots \,, \eta_T$ are i.i.d. random variables with distribution $f(\eta)$ and we assume $x_0 = 0$. For simplicity, we also assume that $f(\eta)$ is symmetric around $\eta = 0$, i.e., that $f(\eta) = f(-\eta)$.

When the variance of the steps

$$\sigma^2 = \int_{-\infty}^{\infty} d\eta \; f(\eta)\eta^2 \,, \tag{1.41}$$

is finite, the result in Eq. (1.39) is still valid for large T at leading order (with $\sigma^2 = 2D$). The corrections to this asymptotic results turns out to be rather nontrivial [33, 102]. In the case of Lévy flights, corresponding to $f(\eta) \sim 1/\eta^{\mu+1}$ with $1 < \mu < 2$, the step variance diverges and the average maximum grows instead as [33]

$$\langle M \rangle \sim T^{1/\mu} \,. \tag{1.42}$$

Time of the maximum

The distribution of the time of the maximum can be exactly computed both for BM and for RWs. The techniques used to compute this distribution will be fundamental to obtaining the main results of this thesis. For this reason, we will present the full derivation in the cases of both continuous and discrete time.

Continuous-time case - Path-decomposition method. We consider a one-dimensional BM $x(\tau)$, starting from $x(0) = x_0$ and evolving according to Eq. (1.31) up to time T. A typical realization of the process is shown in Fig. 1.4. In order to compute the distribution of the time $t_{\max}$ of the maximum, we will use a *path-decomposition* technique. Doing so, we first obtain the joint distribution $P(t_{\max}, M|T)$ of $t_{\max}$ and of the maximum $M = x(t_{\max})$. Then, integrating over M, we will obtain $P(t_{\max}|T)$. This approach was for instance used in Ref. [43] and can be described as follows. Using the Markov property of the process, we can write the joint probability of $t_{\max}$ and M as the product of the probabilities of two independent segments: (I) $[0, t_{\max}]$ and (II) $[t_{\max}, T]$ (see Fig. 1.4). In the first interval (I), the process starts from the origin and reaches the global maximum M at time $t_{\max}$. In the second interval (II), the BM starts from position M at time $t_{\max}$ and has to remain below this position M up to time T.

To compute the probability weight of the first interval, it is useful to define the constrained propagator $G^M(x, t|x_0)$ as the probability that the process goes from position x_0 to position x after time t, while always remaining below position M. This quantity can be computed using the method of images [32] and reads

$$G^M(x, t|x_0) = \frac{1}{\sqrt{4\pi D t}} \left[e^{(x-x_0)^2/(4Dt)} - e^{(2M-x-x_0)^2/(4Dt)} \right] . \tag{1.43}$$

Since in the first interval the particle goes from x_0 to M in a time $t_{\max}$ while remaining below position M (since M is the global maximum), one would naively guess that the probability weight of this interval is $G^M(M, t|x_0)$. However, using the expression in Eq. (1.43), we get

$$G^M(M, t|x_0) = 0 \,. \tag{1.44}$$

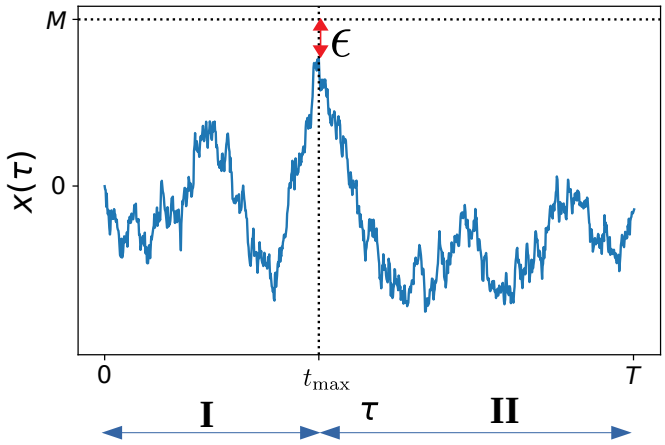

Figure 1.4: Brownian motion $x(\tau)$ during the time interval $[0, T]$. The value of the global maximum is $M - \epsilon$, with $\epsilon > 0$, and the time of the maximum is $t_{\max}$. The time interval $[0, T]$ is divided into the two subintervals $[0, t_{\max}]$ (I) and $[t_{\max}, T]$ (II).

This is a consequence of the continuous-time nature of the process. Indeed, it is possible to show that if the BM arrives at position M at time $t_{\max}$, it will go above position M infinitely many times in any time interval $[t_{\max} - \delta, t_{\max}]$ with $\delta > 0$. In other words, one cannot simultaneously impose that the BM remains below the level M (i.e., $x(\tau) < M$) and that it arrives at M at a specific time (i.e., $x(t_{\max}) = M$).

A way around this issue is to introduce a cutoff $\epsilon > 0$ and to impose that the process arrives at position $M - \epsilon$ (instead of M) at time $t_{\max}$ (see Fig. 1.4). We will compute the distribution $P(t_{\max}|T)$ with ϵ fixed and then we will take the limit $\epsilon \to 0$. Thus, the probability weight of the first interval ($[0, t_{\max}]$) reads

$$P_{\rm I} = G^M(M - \epsilon, t_{\max}|x_0)\,. \tag{1.45}$$

In the second interval ($[t_{\max}, T]$) the process starts from position $M - \epsilon$ at time $t_{\max}$ and remains below position M up to time T. This is precisely the survival probability defined above and hence the weight of the second interval reads

$$P_{\rm II} = Q^M(M - \epsilon, T - t_{\max})\,, \tag{1.46}$$

where the function $Q^L(x, t)$ is given in Eq. (1.36).

Using the Markov property of the process, we can write the joint distribution of M and $t_{\max}$ as

$$P(M, t_{\max}|T, \epsilon) = P_{\rm I}\; P_{\rm II} = \mathcal{N}(\epsilon) G^M(M - \epsilon, t_{\max}|x_0) Q^M(M - \epsilon, T - t_{\max})\,, \tag{1.47}$$

where $\mathcal{N}(\epsilon)$ is a normalization constant. Integrating this expression over $M > 0$, we obtain

$$P(t_{\max}|T, \epsilon) = \mathcal{N}(\epsilon) \int_{x_0}^{\infty} dM\; G^M(M - \epsilon, t_{\max}|x_0) Q^M(M - \epsilon, T - t_{\max})\,. \tag{1.48}$$

Finally, taking the limit of small ϵ, we find the following formula for the distribution of $t_{\max}$

$$P(t_{\max}|T) = \lim_{\epsilon \to 0} \left[\mathcal{N}(\epsilon) \int_{x_0}^{\infty} dM\; G^M(M - \epsilon, t_{\max}|x_0) Q^M(M - \epsilon, T - t_{\max}) \right]\,. \tag{1.49}$$

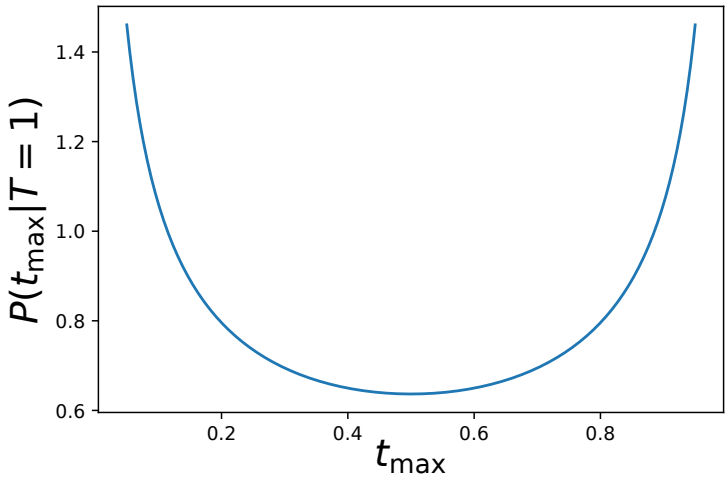

Figure 1.5: Probability density function $P(t_{\max}|T)$ of the time $t_{\max}$ of the maximum for $T = 1$. The blue curve corresponds to Lévy's Arcsine Law, given in Eq. (1.53).

Substituting the expressions for $G^M(M-\epsilon, t_{\max}|x_0)$ and $Q^M(M-\epsilon, T-t_{\max})$, respectively given in Eqs. (1.43) and (1.36), and expanding for small ϵ, we get

$$P(t_{\max}|T) = \lim_{\epsilon\to 0}\left[\mathcal{N}(\epsilon)\epsilon^2\right]\frac{1}{2\pi D\sqrt{t_{\max}^3(T-t_{\max})}}\int_0^\infty dM\ M e^{-M^2/(4Dt_{\max})} \quad (1.50)$$

$$= \lim_{\epsilon\to 0}\left[\mathcal{N}(\epsilon)\epsilon^2\right]\frac{1}{\pi\sqrt{t_{\max}(T-t_{\max})}}\,, \quad (1.51)$$

where we have set $x_0 = 0$. Finally, imposing that $P(t_{\max}|T)$ is correctly normalized to unity

$$\int_0^T dt_{\max}\ P(t_{\max}|T) = 1\,, \quad (1.52)$$

we find $\mathcal{N}(\epsilon) = \epsilon^{-2}$ and hence

$$\boxed{P(t_{\max}|T) = \frac{1}{\pi\sqrt{t_{\max}(T-t_{\max})}}}\,. \quad (1.53)$$

This result is the celebrated *Arcsine Law* of Lévy [1]. This distribution (see Fig. 1.5) is maximal for $t_{\max} \to 0$ and for $t_{\max} \to T$, meaning that the maximum is more likely to occur at the beginning or at the end of the time interval $[0, T]$.

Discrete-time case - The Sparre Andersen theorem

We consider a discrete-time random walk x_k moving on a line according to the evolution rule in Eq. (1.40). We assume that the steps are independent and distributed according to the PDF $f(\eta)$, which is continuous and symmetric around $\eta = 0$, i.e., $f(\eta) = f(-\eta)$ (this corresponds to a random walk without drift).

[1] The reason for this name is that the cumulative distribution of $t_{\max}$ is Prob.$(t_{\max} < t|T) = (2/\pi)\sin^{-1}(\sqrt{t/T})$.

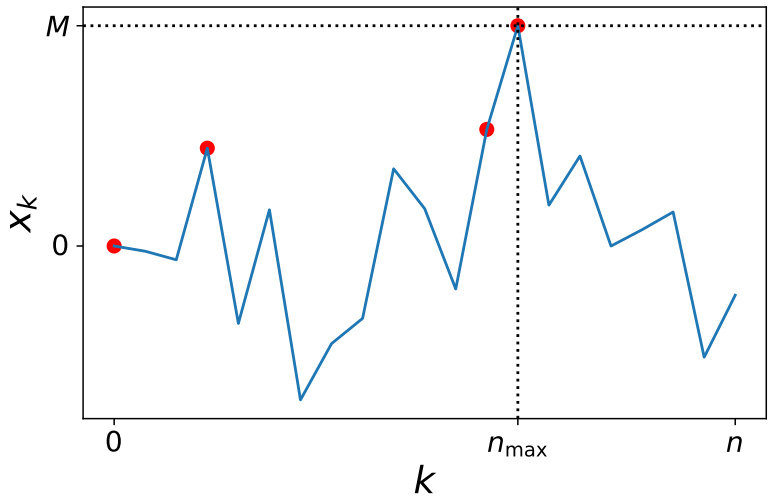

Figure 1.6: Typical trajectory of an n-step random walk x_k. The global maximum M is reached at step $n_{\max}$. Records are signaled by filled red circles.

It is useful to define the survival probability q_n as the probability that the position of the walker remains positive up to step n, i.e.,

$$q_n = \text{Prob.}(x_1 > 0\,, x_2 > 0\,, \ldots\,, x_n > 0 | x_0 = 0)\,. \tag{1.54}$$

Quite remarkably, the Sparre Andersen theorem [38] states that the survival probability is completely independent of the jump distribution and is given by the simple formula

$$\boxed{q_n = \binom{2n}{n} 2^{-2n}\,.} \tag{1.55}$$

Note that this result is completely universal for any n (and not only asymptotically for large n). It is also valid for fat-tailed distributions $f(\eta)$ for which the average value and/or the variance are divergent. Interestingly, the survival probability can be used as a building block to compute several other quantities, including the distribution of the time of the maximum and the records statistics.

In order to compute the distribution of the step $n_{\max}$ at which the global maximum is reached (see Fig. 1.6), we adopt a strategy similar to the path-integral method described above. We divide the interval $[0, n]$ into the two subintervals $[0, n_{\max}]$ (I) and $[n_{\max}, n]$ (II). The probability weight of the first interval is

$$P_{\rm I} = \text{Prob.}(x_0 < x_{n_{\max}}\,, x_1 < x_{n_{\max}}\,, \ldots\,, x_{n_{\max}-1} < x_{n_{\max}})\,. \tag{1.56}$$

We define the RW $y_k = x_{n_{\max}} - x_{n_{\max}-k}$ for $k = 0\,, \ldots\,, n_{\max}$. Then, the probability of the interval $[0, n_{\max}]$ can be rewritten in terms of y_k as

$$P_{\rm I} = \text{Prob.}(y_1 > 0\,, y_2 > 0\,, \ldots y_{n_{\max}} > 0 | y_0 = 0)\,. \tag{1.57}$$

Moreover, it is easy to show that the steps of the random walk y_k are distributed according to the PDF $f(\eta)$. Thus, recalling the definition of q_n in Eq. (1.54), we find

$$P_{\rm I} = q_{n_{\max}}\,. \tag{1.58}$$

With an analogous argument and using the transformation $y_k = x_{n_{\max}} - x_{n_{\max}+k}$, it is easy to show that

$$P_{\rm II} = q_{n-n_{\max}} \,. \tag{1.59}$$

Thus, we find that the probability distribution of the step $n_{\max}$ of the maximum is

$$P(n_{\max}|n) = q_{n_{\max}} q_{n-n_{\max}} \,. \tag{1.60}$$

Using the Sparre Andersen theorem in Eq. (1.55), we finally obtain

$$\boxed{P(n_{\max}|n) = \binom{2n_{\max}}{n_{\max}} \binom{2(n-n_{\max})}{n-n_{\max}} 2^{-2n}} \,. \tag{1.61}$$

Remarkably, as a consequence of the Sparre Andersen theorem, the distribution of the time of the maximum is completely universal, i.e., it does not depend on the particular shape of the jump distribution $f(\eta)$ (as long as $f(\eta)$ is continuous and symmetric).

Interestingly, taking the Brownian limit $n, n_{\max} \to \infty$ in Eq. (1.61) one recovers Arcsine law in Eq. (1.53). Note however that the result in Eq. (1.61) is also valid for fat-tailed distribution for which the first and second moments may not be defined and which do not converge to a BM in the limit of many steps.

The number of records

Using the Sparre-Andersen theorem, it is possible to show that the full record statistics of RW with continuous and symmetric steps is completely universal. In particular, the distribution of the number $N_R(n)$ of records up to step n reads [23]

$$\text{Prob.}(N_R(n) = N) = \binom{2n-N+1}{n} 2^{-2n+N-1} \,, \tag{1.62}$$

which is completely universal for any finite n. Interestingly, in the large-n limit, this distribution assumes the scaling form

$$\text{Prob.}(N_R(n) = N) \approx \frac{1}{\sqrt{n}} g\left(\frac{N}{\sqrt{n}}\right) \,, \tag{1.63}$$

where

$$g(z) = \frac{1}{\sqrt{\pi}} e^{-z^2/4} \theta(z) \,. \tag{1.64}$$

Here $\theta(z)$ is the Heaviside theta function. Thus, the distribution of the number of records converges for late times to a half-Gaussian distribution and is characterized by the single scale $\sqrt{n}$. Indeed, the average number of records reads

$$\langle N_R(n) \rangle = (2n+1) \binom{2n}{n} 2^{-2n} \,, \tag{1.65}$$

which for large n grows as

$$\langle N_R(n) \rangle \approx \frac{2}{\sqrt{\pi}} \sqrt{n} \,. \tag{1.66}$$

Thus, in the case of RW, the number of records grows faster than in the i.i.d. case, where the growth was logarithmic in time. Note that the statistics of records have been investigated for several generalizations of the RW model, including RWs with resetting [103, 104], RWs with a drift [105], and RWs in a random landscape [106].

In the next chapters, we will show how these classical results and techniques can be used to investigate more complicated systems.

Chapter 2

Time of the maximum of stochastic processes

In this chapter, we present our main results on the times at which the extrema (maxima and minima) of stochastic processes are reached. The time at which a time series of duration T attains its maximal value is a quantity of practical interest in several contexts, including physics [40,107], finance [39,108], and sports [109]. For instance, in finance, it is important to know when the price of a stock will reach its maximal or minimal value within a given time window (e.g., a trading day).

Since the classical result of Lévy, who computed the full distribution of $t_{\max}$ for a one-dimensional BM (see Eq. (1.53)), the distribution of the time of the maximum has been studied for a variety of stochastic processes. For instance, this quantity has been investigated for several generalizations of BM, including constrained BM [43, 82, 110], BM with drift [111, 112], fractional BM [113, 114], heterogeneous diffusion [115], and resetting BM [116]. Moreover, the distribution of $t_{\max}$ has also been computed for systems composed of many particles [117,118]. In addition, the time of the maximum has been investigated for a single run-and-tumble particle in one dimension [99]. Similarly, the distribution of the time $t_{\min}$ of the minimum has been studied. For instance, in the case of BM, $t_{\max}$ and $t_{\min}$ have the same distribution, given by the Arcsine law in Eq. (1.53).

This chapter is divided into two main sections. In the first one, we present the exact computation of the distribution of the time between the maximum and the minimum of BM. This first part corresponds to the Articles 1 and 2. In the second one, corresponding to Article 3, we investigate the distribution of $t_{\max}$ for a wide range of stationary processes.

2.1 Time between the maximum and the minimum

The statistical properties of the time $t_{\max}$ of the maximum and the time $t_{\min}$ of the minimum have been investigated for several different processes of fixed duration T. Despite this, even in the case of BM, the joint distribution $P(t_{\max}, t_{\min}|T)$ of $t_{\max}$ and $t_{\min}$ was only recently computed in the Articles 1 and 2. Indeed, as a consequence of the correlations between $t_{\max}$ and $t_{\min}$, the joint distribution does not factorize into the

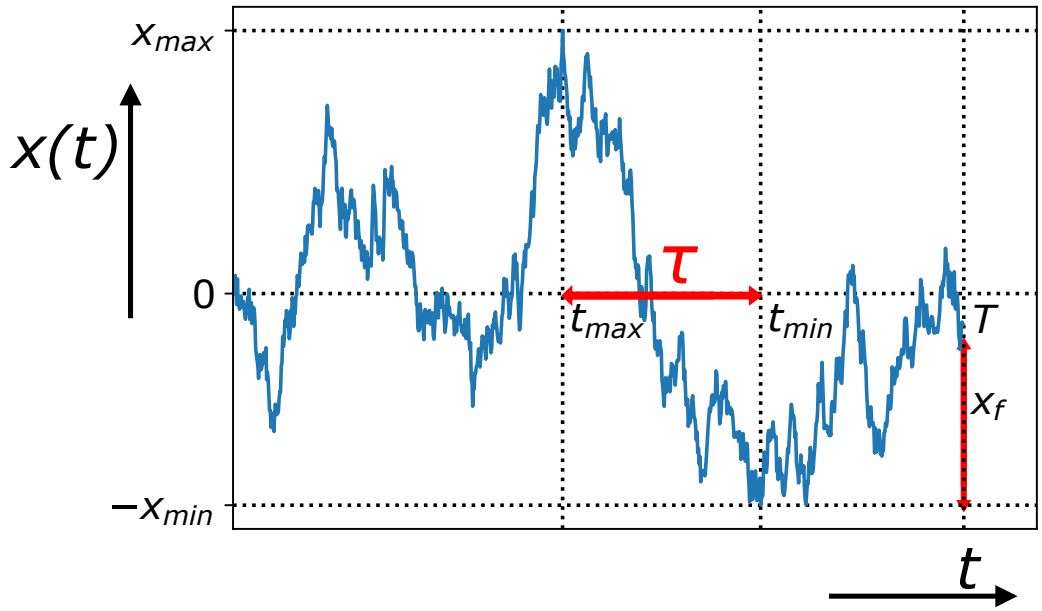

Figure 2.1: Typical trajectory of a BM $x(t)$ in the time interval $[0, T]$. The global maximum $x_{\max}$ is reached at time $t_{\max}$, while the global minimum $-x_{\min}$ is reached at time $t_{\min}$. The time between the maximum and the minimum is $\tau = t_{\max} - t_{\min}$. The final position of the process, measured from the global minimum is $x_f = x(T) + x_{\min}$.

product of the marginals $P(t_{\max}|T)$ and $P(t_{\min}|T)$. To understand the origin of these correlations, it is sufficient to think that if the global maximum occurs at time $t_{\max}$, it is very unlikely that the minimum will be reached immediately before or after. This introduces an effective repulsion between $t_{\max}$ and $t_{\min}$, which we will precisely quantify.

Besides encoding the correlations between $t_{\max}$ and $t_{\min}$, the joint probability distribution $P(t_{\max}, t_{\min}|T)$ allows us also to investigate other important observables. One of these is the time difference between $t_{\min}$ and $t_{\max}$: $\tau = t_{\min} - t_{\max}$. This observable takes values in $[-T, T]$ and has applications in finance. Indeed, one of the simplest models to describe the evolution of the price of a stock is to consider the exponential of a BM. Within this description, τ would represent the time between the occurrence of the maximal and minimal prices. As an example, in the case $t_{\max} < t_{\min}$ (as in Fig. 2.1), an agent would typically want to sell their assets at the time $t_{\max}$, when the price is the highest, and to buy the stocks back at time $t_{\min}$, when the price drops to its minimal value. Thus, in the context of finance, τ represents the optimal time between buying and selling a stock. Interestingly, quantities related to τ, e.g., the time between a local maximum and a local minimum, have been studied empirically for financial data [119]. Therefore, finding an exact expression for the probability density function $P(\tau|T)$ is a task of fundamental importance. Interestingly, before our Articles 1 and 2, the exact expression for $P(\tau|T)$ was not known even in the paradigmatic case of BM.

In this section, following the content of Articles 1 and 2, we will present the exact computation of the PDF $P(\tau|T)$ for BM and other stochastic processes. The main

technique that we will employ is the path-decomposition method presented in Chapter 1. We will first present the derivation of $P(\tau|T)$ in the case of BM. Then, we will consider generalizations to other stochastic processes. We will sometimes omit the details of the more technical parts, which can be found in Article 2.

2.1.1 Brownian motion

To compute the distribution of the time τ between the maximum and the minimum, we will use the path-decomposition technique described in Chapter 1. Doing so, we will first compute the joint PDF of the time $t_{\max}$ of the maximum, the time $t_{\min}$ of the minimum, the maximum $x_{\max}$, and the minimum $-x_{\min}$. Note that by definition $x_{\min} > 0$.

We consider a one-dimensional BM, whose position $x(t)$ evolves according to

$$\frac{dx}{dt} = \sqrt{2D}\eta(t)\,, \tag{2.1}$$

where $\eta(t)$ is Gaussian white noise with zero mean and correlator $\langle \eta(t)\eta(t')\rangle = \delta(t-t')$. Without loss of generality, we assume that the process is initially located at the origin: $x(0) = 0$ (note that τ does not change after a constant shift of the process).

We will first study the case $t_{\max} < t_{\min}$, where the global maximum occurs first. Using the Markov property of the process, we can write the joint grand PDF of $t_{\max}$, $t_{\min}$, $x_{\max}$, and $x_{\min}$ as the product of the probability weights $P_{\rm I}$, $P_{\rm II}$, and $P_{\rm III}$ of the three intervals (see Fig. 2.2): $[0, t_{\max}]$ (I), $[t_{\max}, t_{\min}]$ (II), and $[t_{\min}, T]$ (III). In each of the three intervals, the process has to remain inside the box $[-x_{\min}, x_{\max}]$. In the first time interval $[0, t_{\max}]$ the process starts from the origin and arrives at position $x_{\max}$ at time $t_{\max}$. In the second interval, it goes from the global maximum $x_{\max}$ to the global minimum $x_{\min}$ in time $t_{\min} - t_{\max}$. Finally, in the last interval the process starts from the global minimum $-x_{\min}$ and remains inside the interval $[-x_{\min}, x_{\max}]$ up to the final time T. As described in Chapter 1, due to the continuous-time nature of BM, one cannot simultaneously impose that the process remains inside the box $[-x_{\min}, x_{\max}]$ ($-x_{\min} < x(t) < x_{\max}$) while constraining it to arrive exactly at the boundary of the interval at some fixed time ($x(t_{\max}) = x_{\max}$ and $x(t_{\min}) = x_{\min}$). This is because if we impose for instance that $x(t_{\max}) = x_{\max}$, then the process will go above $x_{\max}$ infinitely many times in any time interval $[t_{\max} - \delta, t_{\max}]$, with $\delta > 0$. As a consequence, we introduce a small cutoff $\epsilon_1 > 0$ and we constrain the trajectory to arrive at $x_{\max} - \epsilon_1$ at time $t_{\max}$. Similarly, we constrain the process to reach position $-x_{\min} + \epsilon_2$, with $\epsilon_2 > 0$ and small, at time $t_{\min}$. To obtain the joint distribution of $t_{\max}$ and $t_{\min}$, we will consider the limit $\epsilon_1, \epsilon_2 \to 0$ at the end of the computation. For simplicity, we take $\epsilon_1 = \epsilon_2 = \epsilon$ and $D = 1/2$. We also denote by

$$M = x_{\max} + x_{\min}\,, \tag{2.2}$$

the size of the interval $[-x_{\min}, x_{\max}]$.

The main building block that we need is the constrained propagator $G_M(x, t|x_0, t_0)$, defined as the probability density that the process goes from position x_0 at time t_0 to position x at time t while always remaining inside the box $[0, M]$. An explicit expression for this propagator can be found by solving the diffusion equation

$$\partial_t G_M(x, t|x_0, t_0) = \frac{1}{2}\partial_x^2 G_M(x, t|x_0, t_0)\,, \tag{2.3}$$

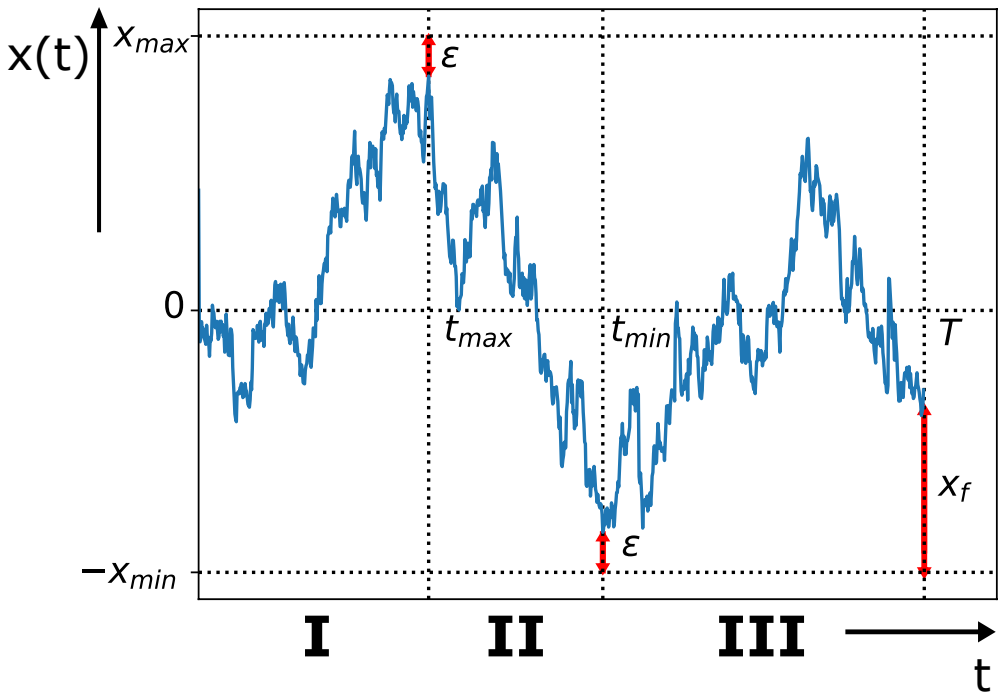

Figure 2.2: Path decomposition of the trajectory of a Brownian motion $x(t)$ for $0 \leq t \leq T$. In the first interval $[0, t_{\max}]$ (I), the process goes from the origin to position $x_{\max} - \epsilon$. In the second interval $[t_{\max}, t_{\min}]$ (II), the process goes from the global maximum $x_{\max} - \epsilon$ to the global minimum $-x_{\min} + \epsilon$. Finally, in the last interval $[t_{\min}, T]$, the process has to remain inside the box $[-x_{\min}, x_{\max}]$.

with absorbing boundary conditions $G_M(M, t|x_0, t_0) = G_M(0, t|x_0, t_0) = 0$. Solving this equation yields [32]

$$G_M(x, t|x_0, t_0) = \frac{2}{M} \sum_{n=1}^{\infty} \sin\left(\frac{n\pi x}{M}\right) \sin\left(\frac{n\pi x_0}{M}\right) e^{-\frac{n^2\pi^2}{2M^2}(t-t_0)} \,. \tag{2.4}$$

By shifting the origin in Fig. 2.2 to $-x_{\min}$, we can now use the propagator to compute the weights $P_{\rm I}$, $P_{\rm II}$, and $P_{\rm III}$. We start with the segment (I), by definition we have

$$P_{\rm I} \propto G_M(M - \epsilon, t_{\max}|x_{\min}, 0) \,, \tag{2.5}$$

where we recall that $M = x_{\max} + x_{\min}$. Substituting the expression in Eq. (2.4) and expanding to leading order in ϵ, we get

$$P_{\rm I} \propto -\frac{2\pi\epsilon}{M^2} \sum_{n_1=1}^{\infty} (-1)^{n_1} n_1 \sin\left(\frac{n_1 \pi\, x_{\min}}{M}\right) e^{-\frac{n_1^2\pi^2}{2M^2}\, t_{\max}} \,. \tag{2.6}$$

Similarly, in segment II, the probability weight reads

$$P_{\rm II} \propto G_M(\epsilon, t_{\min}|M - \epsilon, t_{\max}) \,. \tag{2.7}$$

Plugging again the expression in Eq. (2.4), and expanding for small ϵ, we obtain

$$P_{\rm II} \propto -\frac{2\pi^2\,\epsilon^2}{M^3}\sum_{n_2=1}^{\infty}(-1)^{n_2} n_2^2\, e^{-\frac{n_2^2\pi^2}{2M^2}(t_{\rm min}-t_{\rm max})}\,. \tag{2.8}$$

Finally, to compute the weight of the third time segment (III), we call x_f the final position of the motion, measured from the global minimum $-x_{\rm min}$ (see Fig. 2.2). Integrating over all possible values of x_f, we get

$$P_{\rm III} \propto \int_0^M G_M(x_{\rm f}, T|\epsilon, t_{\rm min})\, dx_{\rm f}\,, \tag{2.9}$$

Using Eq. (2.4) and performing the integral over x_f, we obtain

$$P_{\rm III} \propto \frac{2\epsilon}{M}\sum_{n_3=1}^{\infty}\left[1-(-1)^{n_3}\right] e^{-\frac{n_3^2\pi^2}{2M^2}(T-t_{\rm min})}\,, \tag{2.10}$$

where we have expanded to leading order for small ϵ.

Using the expressions for $P_{\rm I}$, $P_{\rm II}$, and $P_{\rm III}$, respectively given in Eqs. (2.6), (2.8), and (2.10), we can now write the grand joint PDF of $t_{\rm max}$, $t_{\rm min}$, $x_{\rm max}$, and $x_{\rm min}$ as

$$P(x_{\rm min}, x_{\rm max}, t_{\rm min}, t_{\rm max}|T) \propto P_{\rm I}P_{\rm II}P_{\rm III} \propto \frac{\epsilon^4}{M^6}\sum_{n_1=1}^{\infty}(-1)^{n_1} n_1 \sin\left(\frac{n_1\pi\, x_{\rm min}}{M}\right)$$
$$\times\quad e^{-\frac{n_1^2\pi^2}{2M^2}t_{\rm max}}\sum_{n_2=1}^{\infty}(-1)^{n_2} n_2^2\, e^{-\frac{n_2^2\pi^2}{2M^2}(t_{\rm min}-t_{\rm max})}\sum_{n_3=1}^{\infty}\left[1-(-1)^{n_3}\right] e^{-\frac{n_3^2\pi^2}{2M^2}(T-t_{\rm min})}\,, \tag{2.11}$$

which is valid for $t_{\rm min} > t_{\rm max}$. Note that this expression in Eq. (2.11) was previously derived in [82] using real-space renormalization group techniques. Integrating over $x_{\rm min} > 0$ and $x_{\rm max} > 0$, we obtain the joint distribution of $t_{\rm max}$ and $t_{\rm min}$

$$P(t_{\rm min}, t_{\rm max}|T) = \int_0^\infty dx_{\rm min}\int_0^\infty dx_{\rm max} P(x_{\rm min}, x_{\rm max}, t_{\rm min}, t_{\rm max}|T)\,. \tag{2.12}$$

Substituting the expression in Eq. (2.11) and performing the integrals over $x_{\rm max}$ and $x_{\rm min}$, we obtain

$$P_<(t_{\rm min}, t_{\rm max}|T) = A_< \sum_{n_1,n_2,n_3=1}^{\infty}\frac{(-1)^{n_2+1} n_2^2[1-(-1)^{n_1}][1-(-1)^{n_3}]}{\left[n_1^2 t_{\rm max} + n_2^2(t_{\rm min}-t_{\rm max}) + n_3^2(T-t_{\rm min})\right]^2}\,, \tag{2.13}$$

where the subscript $<$ indicates that the expression is valid in the case $t_{\rm max} < t_{\rm min}$ and $A_<$ is a normalization constant. To fix this constant $A_<$, we use the normalization condition

$$\int_0^T dt_{\rm max}\int_{t_{\rm max}}^T dt_{\rm min}\; P_<(t_{\rm min}, t_{\rm max}|T) = {\rm Prob.}(t_{\rm max} < t_{\rm min}) = \frac{1}{2}\,, \tag{2.14}$$

since the maximum will occur before the minimum for half of the trajectories by the $x \to -x$ symmetry of the process. Using the expression in Eq. (2.13), we obtain $A_< = 4/\pi^2$. Thus, we get

$$P_<(t_{\rm min}, t_{\rm max}|T) = \frac{4}{\pi^2}\sum_{n_1,n_2,n_3=1}^{\infty}\frac{(-1)^{n_2+1} n_2^2[1-(-1)^{n_1}][1-(-1)^{n_3}]}{\left[n_1^2 t_{\rm max} + n_2^2(t_{\rm min}-t_{\rm max}) + n_3^2(T-t_{\rm min})\right]^2}\,. \tag{2.15}$$

Repeating the computation in the case $t_{\min} < t_{\max}$ we find

$$P_{>}(t_{\min}, t_{\max}|T) = \frac{4}{\pi^2} \sum_{n_1,n_2,n_3=1}^{\infty} \frac{(-1)^{n_2+1} n_2^2 [1-(-1)^{n_1}][1-(-1)^{n_3}]}{\left[n_1^2 t_{\min} + n_2^2 (t_{\max}-t_{\min}) + n_3^2 (T-t_{\max})\right]^2} , \tag{2.16}$$

which is valid for $t_{\min} < t_{\max}$. We notice the nontrivial symmetry

$$P_{<}(t_{\min}, t_{\max}|T) = P_{>}(t_{\max}, t_{\min}|T) , \tag{2.17}$$

which can be understood as a consequence of the $x \to -x$ symmetry of the process. Overall, the joint distribution of $t_{\min}$ and $t_{\max}$ can be written as

$$\begin{aligned} P(t_{\min}, t_{\max}|T) = & \theta(t_{\max}-t_{\min}) \frac{4}{\pi^2} \sum_{n_1,n_2,n_3=1}^{\infty} \frac{(-1)^{n_2+1} n_2^2 [1-(-1)^{n_1}][1-(-1)^{n_3}]}{\left[n_1^2 t_{\min} + n_2^2 (t_{\max}-t_{\min}) + n_3^2 (T-t_{\max})\right]^2} \\ & + \theta(t_{\min}-t_{\max}) \frac{4}{\pi^2} \sum_{n_1,n_2,n_3=1}^{\infty} \frac{(-1)^{n_2+1} n_2^2 [1-(-1)^{n_1}][1-(-1)^{n_3}]}{\left[n_1^2 t_{\max} + n_2^2 (t_{\min}-t_{\max}) + n_3^2 (T-t_{\min})\right]^2} , \end{aligned} \tag{2.18}$$

where $\theta(z)$ is the Heaviside theta function.

We can now compute the PDF of the time $\tau = t_{\min} - t_{\max}$ between maximum and minimum. The PDF of τ can be obtained from the joint PDF of $t_{\min}$ and $t_{\max}$ as

$$P(\tau|T) = \int_0^T dt_{\max} \int_0^T dt_{\min} P(t_{\min}, t_{\max}|T) \delta(t_{\min} - t_{\max} - \tau) , \tag{2.19}$$

Substituting the expression for $P(t_{\min}, t_{\max}|T)$ in Eq. (2.18), we obtain, after few steps of algebra

$$\boxed{P(\tau|T) = \frac{1}{T} f_{\rm BM}\left(\frac{\tau}{T}\right) ,} \tag{2.20}$$

where

$$\boxed{f_{\rm BM}(y) = \frac{1}{|y|} \sum_{n=1}^{\infty} (-1)^{n+1} \tanh^2\left(\frac{n\pi}{2}\sqrt{\frac{|y|}{1-|y|}}\right) ,} \tag{2.21}$$

which is defined for $-1 \le y \le 1$. This exact result is shown in Fig. 2.3a) and is in excellent agreement with numerical simulations.

Asymptotic behaviors

As a consequence of the $x \to -x$ symmetry of BM, this scaling function $f_{\rm BM}(y)$ is symmetric around $y = 0$, i.e., $f_{\rm BM}(y) = f_{\rm BM}(-y)$. Moreover, it has asymptotic behaviors

$$f_{\rm BM}(y) \approx \begin{cases} \dfrac{1}{2} + \dfrac{1-|y|}{2} & \text{as} \quad y \to \pm 1 \\ \dfrac{8}{y^2} e^{-\pi/\sqrt{|y|}} - \dfrac{8}{|y|} e^{-\pi/\sqrt{|y|}} & \text{as} \quad y \to 0 . \end{cases} \tag{2.22}$$

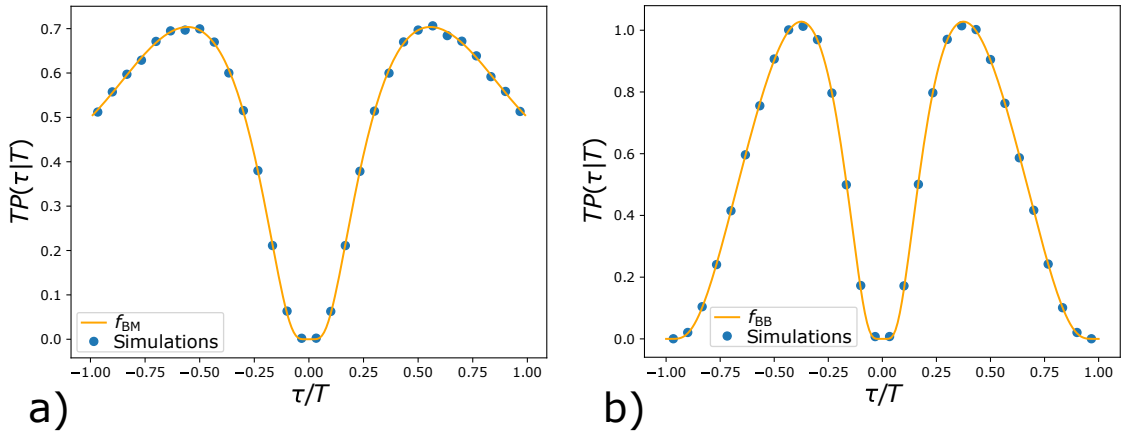

Figure 2.3: **a)** The scaled probability density function $T\,P(\tau|T)$ as a function of the scaled time τ/T between maximum and minimum for the BM. The solid orange line corresponds to the exact scaling function $f_{\rm BM}(y)$, given in Eq. (2.21), while the filled blue circles are the results of simulations. **b)** The same scaled probability density function for the Brownian bridge where the exact scaling function $f_{\rm BB}(y)$ is given in Eq. (2.44).

The first asymptotic behavior ($y \to \pm 1$) can be obtained directly from Eq. (2.21) by using the relation [1]

$$\sum_{n=1}^{\infty} (-1)^{n+1} = \frac{1}{2}\,. \tag{2.23}$$

On the other hand, the second asymptotic behavior (for $|y| \to 0$) can be obtained using the alternative representation

$$f_{\rm BM}(y) = \frac{2(1-|y|)}{|y|^2} \sum_{m=-\infty}^{\infty} \frac{2m+1}{\sinh\left((2m+1)\pi\sqrt{\frac{1-|y|}{|y|}}\right)}\,, \tag{2.24}$$

obtained using Poisson summation formula.

We observe that for $y \to 0$ (corresponding to $\tau \to 0$ and hence $t_{\rm max} \to t_{\rm min}$) the scaling function $f_{\rm BM}(y)$ has an essential singularity. As a consequence, it is very unlikely to observe a Brownian trajectory in which the maximum and the minimum are close in time. Indeed, we will show that $t_{\rm max}$ and $t_{\rm min}$ are anticorrelated.

On the other hand, for $y > 0$, the distribution of τ is non-monotonic. For small values of $y = \tau/T$ the probability is increasing as a function of y. The probability is maximal at some intermediate value $y = y^*$. This maximum y^* can be computed by solving the equation

$$\frac{df_{\rm BM}(y)}{dy} = 0\,. \tag{2.25}$$

We find $y^* = 0.5563\ldots$ where we have solved the equation numerically with Mathematica.

[1] This relation has to be interpreted as the limit $\lim_{\alpha\to -1} \sum_{n\geq 1} \alpha^{n+1}$.

Finally, for $\tau \to \pm T$, the PDF of τ goes to the limit value $1/(2T)$. In the coming sections, we will show that this is a very general result. The value $\tau = T$ can only be obtained if $t_{\min} = T$ and $t_{\max} = 0$. This corresponds to configurations in which the process always remains below its starting position and reaches its minimal value at the final time. We will show that even for discrete-time random walks with n steps, the probability of such an event is always $1/(2n)$, independently of the step distribution (even if this distribution is fat-tailed). This is a nice example of universality that holds for all times and not only asymptotically (à la Sparre Andersen).

Moments of τ

Due to the symmetry of $P(\tau|T)$ around $\tau = 0$, the odd moments of the distribution trivially vanish. Hence, we decide to study the moments of the absolute value $|\tau|$.

It turns out to be quite challenging to compute the moments of τ from the scaling function $f_{\rm BM}$. Luckily, we were able to find the following integral identity satisfied by $f_{\rm BM}(y)$

$$\int_0^1 dy \frac{f_{\rm BM}(y)}{1+uy} = \int_0^\infty dz \frac{1}{\sinh(z)} \tanh^2\left(\frac{z}{2\sqrt{1+u}}\right) . \tag{2.26}$$

The derivation of this relation is given in Article 2, where it appears quite naturally in the investigation of τ for discrete-time random walks. Interestingly, the moments of τ can be obtained by expanding both sides of this identity in powers of u. In particular, we find that the first four moments are

$$\begin{aligned}
\langle|\tau|\rangle &= \frac{4\log(2)-1}{3} T = (0.5908\ldots)\, T\,, \\
\langle\tau^2\rangle &= \frac{7\zeta(3)-2}{16} T^2 = (0.4009\ldots)\, T^2\,, \\
\langle|\tau|^3\rangle &= \frac{147\zeta(3)-34}{480} T^3 = (0.2972\ldots)\, T^3\,, \\
\langle\tau^4\rangle &= \frac{1701\zeta(3)-930\zeta(5)-182}{3840} T^4 \\
&= (0.2339\ldots)\, T^4\,,
\end{aligned} \tag{2.27}$$

where $\zeta(z)$ is the Riemann zeta function. These results are in perfect agreement with numerical simulation.

Covariance of $t_{\max}$ and $t_{\min}$

The covariance of $t_{\min}$ and $t_{\max}$ can be computed from the second moment of τ. The covariance function is defined as

$$\mathrm{cov}\,(t_{\min}, t_{\max}) = \langle t_{\min} t_{\max}\rangle - \langle t_{\min}\rangle\langle t_{\max}\rangle\ . \tag{2.28}$$

Computing this quantity from the joint PDF $P(t_{\min}, t_{\max}|T)$ in Eq. (2.13) turns out to be quite lengthy. We can instead apply the following idea. Using $\tau = t_{\min} - t_{\max}$, we have

$$\langle\tau^2\rangle = \langle t_{\min}^2\rangle + \langle t_{\max}^2\rangle - 2\langle t_{\min}\, t_{\max}\rangle\ . \tag{2.29}$$

And hence

$$\mathrm{cov}\,(t_{\min}, t_{\max}) = \frac{1}{2}\left(\langle t_{\min}^2\rangle + \langle t_{\max}^2\rangle - \langle \tau^2\rangle\right) - \langle t_{\min}\rangle\langle t_{\max}\rangle\ . \tag{2.30}$$

Thus, we just need to compute the first two moments of $t_{\min}$, $t_{\max}$. As anticipated, the marginal distributions of $t_{\min}$ and $t_{\max}$ are both given by the Arcsine law

$$P(t_{\min}|T) = \frac{1}{\pi\sqrt{t_{\min}(T-t_{\min})}}\ , \tag{2.31}$$

and

$$P(t_{\max}|T) = \frac{1}{\pi\sqrt{t_{\max}(T-t_{\max})}}\ . \tag{2.32}$$

From Eqs. (2.31) and (2.32), we obtain

$$\langle t_{\min}\rangle = \langle t_{\max}\rangle = \frac{T}{2}\ , \tag{2.33}$$

$$\langle t_{\min}^2\rangle = \langle t_{\max}^2\rangle = \frac{3}{8}T^2\ . \tag{2.34}$$

Plugging the results from Eqs. (2.27), (2.33), and (2.34) into Eq. (2.30), we finally find

$$\mathrm{cov}_{\mathrm{BM}}(t_{\min}, t_{\max}) = -\frac{7\zeta(3)-6}{32}\,T^2 = -(0.0754\ldots)\,T^2. \tag{2.35}$$

As expected, the random variables $t_{\min}$ and $t_{\max}$ are negatively correlated.

2.1.2 Brownian Bridge

In this section, we compute the distribution of $\tau = t_{\min} = t_{\max}$ in the case of a Brownian bridge, i.e., a periodic BM of fixed period T. The process starts from $x(0) = 0$ and is conditioned to go back to the starting position at the final time T (for a typical realization, see Fig. 2.4). The results derived for a BB will be directly applicable to the case of fluctuating interfaces, both Edwards-Wilkinson (EW) and Kardar-Parisi-Zhang (KPZ) types, with periodic boundary conditions (PBC). Moreover, constrained BM trajectories and in particular BBs have been applied to model the movement of animals while foraging for food [120–123]. Indeed, animals search for food starting from and going back to some fixed location, e.g., their nest.

In the case of BB, the PDF of the time $t_{\max}$ of the maximum (and by symmetry also the PDF of $t_{\min}$) is uniform in the interval $[0, T]$ [24]

$$P(t_{\max}|T) = \frac{1}{T}\,. \tag{2.36}$$

This is a consequence of the periodicity of the process. Nevertheless, we will show that the distribution of the time τ between the maximum and the minimum is instead highly nontrivial.

There are two possible ways of deriving the distribution of $P(\tau|T)$ for a BB. The first one is to employ a path-decomposition strategy analogous to the one used for a

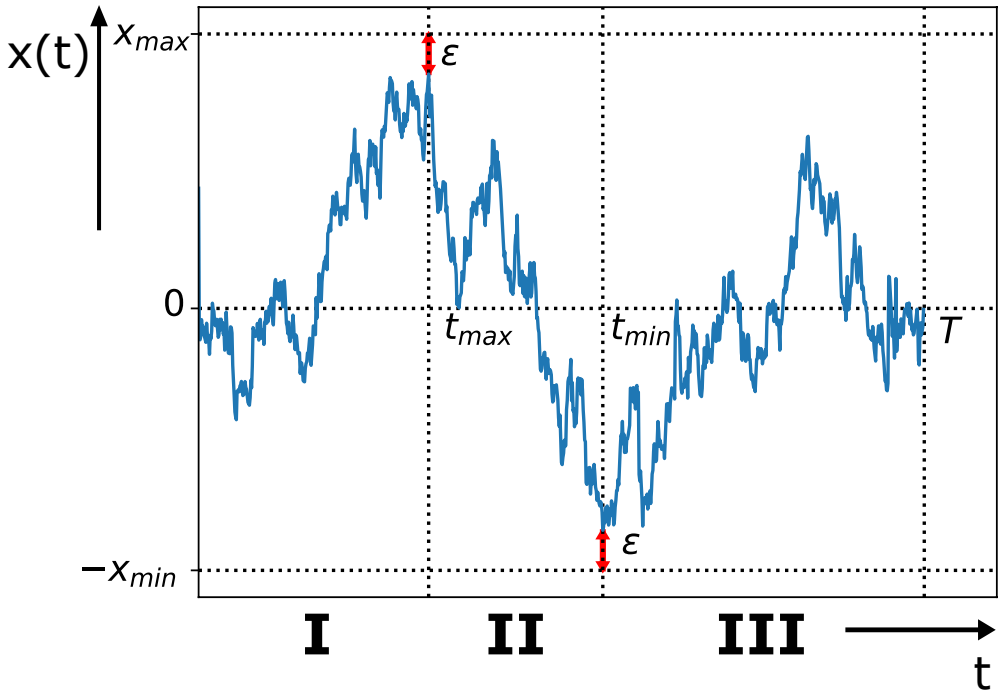

Figure 2.4: Path decomposition of the trajectory of a Brownian bridge $x(t)$ for $0 \leq t \leq T$. The process is constrained to go back to its starting position $x(0) = 0$ at the final time T. In the first interval $[0, t_{\max}]$ (I), the process goes from the origin to position $x_{\max} - \epsilon$. In the second interval $[t_{\max}, t_{\min}]$ (II), the process goes from the global maximum $x_{\max} - \epsilon$ to the global minimum $-x_{\min} + \epsilon$. Finally, in the last interval $[t_{\min}, T]$, the process has to go from the global minimum $x_{\min} + \epsilon$ to the origin. Moreover, the process has to always remain inside the box $[-x_{\min}, x_{\max}]$.

BM, using the same path-decomposition as in Fig. 2.4. The main difference with the BM computation is that one does not have to integrate over the final position x_f, which is instead fixed. We will not present the details of this computation, which can be found in Article 2.

The second way of computing the distribution $P(\tau|T)$ for a Brownian bridge is to use a mapping to a Brownian excursion (BE), which is a BB with the additional constraint that the process has to remain above its starting position during the whole time interval $[0, T]$. The mapping is known as Vervaat construction in the literature of probability theory [124, 125].

The Vervaat construction (see Fig. 2.5) can be used to map any BB trajectory to a BE trajectory. Moreover, Vervaat showed that the resulting BE is sampled with the correct statistical weight [124]. For our purposes, it is relevant to notice that the time τ between the maximum and the minimum of a BB gets mapped into the time $t^{\rm BE}_{\max}$ of the maximum (measured from the right end of the interval $[0, T]$) of the BE. We denote the PDF of $t^{\rm BE}_{\max}$ by

$$P_{\rm BE}(\tau|T) = \text{Prob.}\left(t^{\rm BE}_{\max} = \tau|T\right) . \tag{2.37}$$

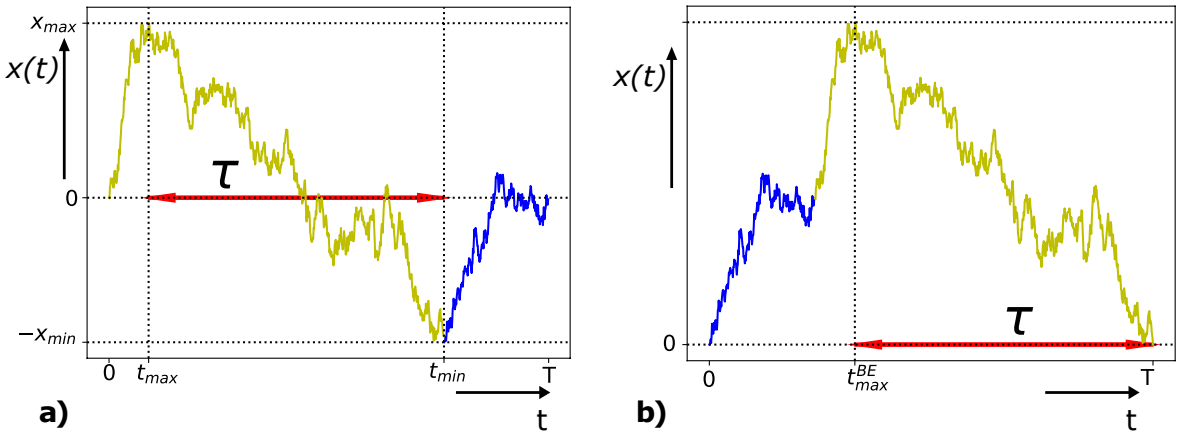

Figure 2.5: Mapping between a Brownian bridge (BB) of duration T (panel **a**) and a Brownian excursion (BE) (panel **b**). Starting with the BB in panel **a)**, we first identify the time $t_{\rm min}$ of the global minimum. Then we decompose the trajectory into two parts: the one in the interval $[0, t_{\rm min}]$ (yellow part) and the one in the interval $[t_{\rm min}, T]$ (blue part). We then shift the yellow part by the total time T, gluing it to the right of the blue part. Finally, we shift the origin of the x-axis to $-x_{\rm min}$. In this way, we obtain the BE shown in panel **b)**. Crucially, the time τ between the maximum and the minimum of the BB (red double arrow) gets mapped into the time of the maximum $t_{\rm max}^{\rm BE}$ (measured from the right end of the interval) of the Brownian excursion.

Note however that with this mapping different BB trajectories can be mapped into the same BE. For the mapping to be one-to-one, one has to also fix the value of the time $t_{\rm min}$ of the minimum of the BB. As a consequence, considering the case $\tau > 0$, we get the exact relation

$$P_{\rm BB}(t_{\rm min} - t_{\rm max}|t_{\rm min}, T) = P_{\rm BE}(t_{\rm min} - t_{\rm max}|T) \;, \tag{2.38}$$

where we used the subscript "BB" in the left-hand side to stress that we are considering the distribution of τ for a BB. The PDF $P_{\rm BE}(\tau|T)$ of the time of the maximum of a BE was computed exactly in [43] and reads

$$P_{\rm BE}(\tau|T) = 3\,T^{3/2} \sum_{m,n=1}^{\infty} \frac{(-1)^{m+n} m^2 n^2}{[m^2\tau + n^2(T-\tau)]^{5/2}} \;. \tag{2.39}$$

To exploit the mapping, we first rewrite the joint PDF of $t_{\rm max}$ and $t_{\rm min}$ for a BB as

$$P_{\rm BB}(t_{\rm max}, t_{\rm min}|T) \;=\; P_{\rm BB}(t_{\rm max} - t_{\rm min}|t_{\rm min}, T) P_{\rm BB}(t_{\rm min}|T) \,. \tag{2.40}$$

Using Eq. (2.38), we get

$$P_{\rm BB}(t_{\rm max}, t_{\rm min}|T) \;=\; P_{\rm BE}(t_{\rm max} - t_{\rm min}|T) P_{\rm BB}(t_{\rm min}|T) \,. \tag{2.41}$$

The time $t_{\rm min}$ of the minimum of a BB is uniformly distributed in the time interval $[0, T]$ [24]. Thus, substituting the expressions for $P_{\rm BE}(t_{\rm min} - t_{\rm max}|T)$, given in Eq. (2.39), into Eq. (2.41), we obtain

$$P_{\rm BB}(t_{\rm max}, t_{\rm min}|T) = 3\sqrt{T} \sum_{m,n=1}^{\infty} \frac{(-1)^{m+n} m^2 n^2}{[m^2(t_{\rm min} - t_{\rm max}) + n^2(T - t_{\rm min} + t_{\rm max})]^{5/2}} \,. \tag{2.42}$$

Finally, integrating over $t_{\max}$ and $t_{\min}$, with $\tau = t_{\min} - t_{\max}$ fixed, we find that the PDF of the time τ between the maximum and the minimum for a BB is given by

$$P(\tau|T) = \frac{1}{T} f_{\rm BB}(\frac{\tau}{T}) \tag{2.43}$$

where

$$f_{\rm BB}(y) = 3(1-|y|) \sum_{m,n=1}^{\infty} \frac{(-1)^{m+n} m^2 n^2}{\left[m^2 |y| + n^2 (1-|y|)\right]^{5/2}} \,. \tag{2.44}$$

We observe that the distribution $P(\tau|T)$ of τ is symmetric around $\tau = 0$. This is again a consequence of the $x \to -x$ symmetry of the process.

This exact distribution in Eq. (2.44) is shown in Fig. 2.3**b)** and is in perfect agreement with numerical simulations. To evaluate numerically the double sum in Eq. (2.44) we have used the PYTHON package MPMATH. The validity of this numerical scheme is confirmed by the excellent agremment with numerical simulations. We observe that, for $y > 0$, the scaling function $f_{\rm BB}(y)$ is non-monotonic as a function of y and that it attains its maximal value at $y = 0.3749\ldots$ which we compute by maximizing the function $f_{\rm BB}(y)$ with Mathematica. Note that to perform numerical simulations of BB trajectories, we used the sampling technique described in [125]. These methods have been recently generalized to discrete-time RWs and more complicated processes [123, 126–128].

Moreover, the asympototic behaviors of $f_{\rm BB}(y)$ can be obtained from those of $P_{\rm BE}(\tau|T)$, which were computed in [43]. We get

$$f_{\rm BB}(y) \approx \begin{cases} \frac{\sqrt{2}\,\pi^2}{(1-|y|)^{\frac{5}{4}}} e^{-\pi/\sqrt{1-|y|}} & \text{as} \quad y \to \pm 1 \\ \\ \frac{\sqrt{2}\pi^2}{|y|^{\frac{9}{4}}} e^{-\pi/\sqrt{|y|}} & \text{as} \quad y \to 0 \,. \end{cases} \tag{2.45}$$

Thus, in this case, as a consequence of the Bridge constraint, an essential singularity appears also for $\tau \to \pm 1$.

Moments of τ for BB

As in the case of BM, explicitly computing the moments of τ from the exact scaling function in Eq. (2.44) turns out to be rather cumbersome. Luckily, an integral identity as the one in Eq. (2.26) can be derived also for BB (see Appendix D in Article 2 for the derivation) and reads

$$\int_0^1 dy \, \frac{f_{\rm BB}(y)}{\sqrt{1+uy}} = \int_0^\infty dz \, \frac{\frac{z}{\sqrt{1+u}} \coth\left(\frac{z}{\sqrt{1+u}}\right) - 1}{\sinh(z) \sinh\left(\frac{z}{\sqrt{1+u}}\right)} \,. \tag{2.46}$$

Expanding in powers of u, we find

$$\begin{aligned} \langle|\tau|\rangle &= \frac{\pi^2-6}{9}\,T = (0.4299\ldots)\,T\,, \\ \langle\tau^2\rangle &= \frac{\pi^2-6}{18}\,T^2 = (0.2149\ldots)\,T^2\,, \\ \langle|\tau|^3\rangle &= \frac{375\pi^2-14\pi^4-1530}{6750}\,T^3 = (0.1196\ldots)\,T^3\,, \\ \langle\tau^4\rangle &= \frac{125\pi^2-7\pi^4-390}{2250}\,T^4 = (0.0719\ldots)\,T^4\,. \end{aligned} \tag{2.47}$$

These moments are in perfect agreement with numerical simulations.

Covariance of $t_{\min}$ and $t_{\max}$

As before, the covariance between $t_{\min}$ and $t_{\max}$ can be computed using the exact relation in Eq. (2.30). The first two moments of $t_{\max}$ and $t_{\min}$ can be simply computed using the fact that these random variables are uniformly distributed in $[0,T]$ and read

$$\langle t_{\min}\rangle = \langle t_{\max}\rangle = \frac{T}{2}\,, \tag{2.48}$$

$$\langle t_{\min}^2\rangle = \langle t_{\max}^2\rangle = \frac{1}{3}T^2\,. \tag{2.49}$$

Plugging these results and the expression for $\langle\tau^2\rangle$, given in Eq. (2.47), into Eq. (2.30) we get

$$\mathrm{cov}_{\mathrm{BB}}(t_{\min},t_{\max}) = -\frac{\pi^2-9}{36}T^2 = -(0.0241\ldots)T^2\,. \tag{2.50}$$

We observe that $t_{\max}$ and $t_{\min}$ are less negatively correlated in the BB case than in the BM case (compare with Eq. (2.35)).

2.1.3 Discrete-time random walks

We next consider the distribution of the time between the maximum and the minimum of a discrete-time RW. We consider an n-step random walker x_k moving on the real line according to the evolution rule

$$x_k = x_{k-1} + \eta_k\,, \tag{2.51}$$

where η_k are i.i.d. random variables from some symmetric distribution $p(\eta)$. We denote by $n_{\max}$ ($n_{\min}$) the discrete step at which the process reaches the global maximum (minimum). In the eventuality of degenerate extrema, we take $n_{\max}$ ($n_{\min}$) as the time at which the maximum (minimum) is reached for the first time. In this case, the time $\tau = n_{\max} - n_{\min}$ is an integer number, with $-n \le \tau \le n$.

In general, one can consider both continuous and discrete step distributions. In the continuous case, as discussed in detail in Chapter 1, the marginal distribution of $n_{\max}$ is completely universal (as a consequence of the Sparre Andersen theorem [38]) and is given for any n by the expression

$$\boxed{P(n_{\max}|n) = \binom{2n_{\max}}{n_{\max}}\binom{2(n-n_{\max})}{n-n_{\max}}2^{-2n}\,.} \tag{2.52}$$

By symmetry, $n_{\min}$ has the same distribution. We recall that this exact result is valid even for fat-tailed distributions, for which the mean value and/or the variance are not defined). Note however that this result is not valid in the case of discrete jump distributions (e.g., lattice walks), which we discuss separately below. Expanding the expression in Eq. (2.52) for large n, one obtains the Arcsine law in Eq. (2.32), valid for a continuous-time BM. It is thus natural to ask whether or not the distribution of τ converges to the one that we have computed for BM in the limit of large n.

In the case where the jump variance

$$\sigma^2 = \int_{-\infty}^{\infty} d\eta \; \eta^2 \; p(\eta) \tag{2.53}$$

is finite, the CLT guarantees that the random walk x_k converges to a BM at late times. As a consequence, it is natural to expect that the distribution of any observable of the trajectory x_k converges to its Brownian counterpart. From what stated above, we know this fact to be true in the case of the step $n_{\max}$ of the maximum. However, verifying directly the convergence of the distribution of τ for an RW to the exact result in Eq. (2.20) and (2.21), valid for Brownian motion, is in general rather complicated. We will verify analytically this convergence in two exactly solvable cases. Moreover, we will perform numerical simulations for several choices of $p(\eta)$. When the step variance σ^2 diverges, e.g., in the case where $p(\eta) \sim 1/|\eta|^{\mu+1}$ for large $|\eta|$ with $\mu < 2$, the CLT does not apply. Thus, it is relevant to ask whether the distribution of τ still converges to the Brownian result, as in the case of the distribution of $n_{\max}$.

In the case of discrete-time RW, a path-decomposition approach similar to the one described in Section 2.1.1 can be used. The only difference is that in this case no cutoff ϵ is required, since the process is discrete in time. In the case $n_{\max} < n_{\min}$, we decompose the interval $[0, n]$ into the three subintervals $[0, n_{\max}]$, $[n_{\max}, n_{\min}]$, and $[n_{\min}, n]$, with corresponding probabilities $P_{\rm I}$, $P_{\rm II}$, and $P_{\rm III}$. As before the probabilities can be expressed in terms of the constrained propagator $G_M(x, t|x_0, t_0)$, which is defined as the probability to go from position x_0 at step t_0 to position x at step t, while always remaining inside the interval $[0, M]$. Thus, denoting by $x_{\max}$ and $x_{\min}$ the maximum and the minimum of the process and by $M = x_{\max} + x_{\min}$ the span, we find

$$P_{\rm I} = G_M(x_{\max}, n_{\max}|0, 0)\,, \tag{2.54}$$

$$P_{\rm II} = G_M(x_{\min}, n_{\min}|x_{\max}, n_{\max})\,, \tag{2.55}$$

and

$$P_{\rm III} = \int_0^M dx_f \; G_M(x_f, n|x_{\min}, n_{\min})\,, \tag{2.56}$$

where the integral over the final position x_f has to be replaced with a sum if the steps are discrete. Thus, the grand joint distribution of $x_{\max}$, $x_{\min}$, $t_{\max}$, and $t_{\min}$ reads

$$\begin{aligned} P(x_{\max}, x_{\min}, n_{\max}, n_{\min}|n) &= P_{\rm I}P_{\rm II}P_{\rm III} = G_M(x_{\max}, n_{\max}|0, 0)G_M(x_{\min}, n_{\min}|x_{\max}, n_{\max}) \\ &\times \int_0^M dx_f \; G_M(x_f, n|x_{\min}, n_{\min})\,. \end{aligned} \tag{2.57}$$

From this exact relation, integrating over $x_{\max}$ and $x_{\min}$ and summing over $n_{\max}$ and $n_{\min}$ with $\tau = n_{\min} - n_{\max}$ fixed, one can obtain the probability distribution of τ.

The constrained propagator $G_M(x,t|x_0,t_0)$ satisfies the following recursion relation

$$G_M(x,t|x_0,t_0) = \int_0^M dx'\; G_M(x',t-1|x_0,t_0)p(\eta = x-x')\;, \tag{2.58}$$

for $t \geq 1$ and with the initial condition $G_M(x,t_0|x_0,t_0) = \delta(x-x_0)$. This equation can be understood as follows. At time $t-1$, the walker arrives at position x' with probability $G_M(x',t-1|x_0,t_0)$ and then it takes a jump to position x with probability $p(x-x')$. However, this simple equation cannot be solved in general for any jump distribution $p(\eta)$ [2].

Finite jump variance

In the case of finite jump variance, i.e., $\sigma^2 < \infty$, we expect the distribution of τ to converge to the BM result. Fist, we check this numerically for different jump distributions (see Fig. 2.6). Moreover, we verify it analytically with two exactly solvable models: RWs with a double-exponential jump distribution and lattice walks.

Random walks with double-exponential jumps: We first consider the case of the double-exponential jump distribution

$$p(\eta) = \frac{1}{2}e^{-|\eta|}\;. \tag{2.59}$$

This distribution has the special property that it satisfies the simple differential equation

$$p''(x) = p(x) - \delta(x)\;. \tag{2.60}$$

It is useful to consider the generating function in Eq. (2.58), yielding

$$\tilde{G}_M(x,s|x_0) = s\int_0^M dx'\; \tilde{G}_M(x',s|x_0)p(x-x') + sp(x)\;, \tag{2.61}$$

where we have defined

$$\tilde{G}_M(x,s|x_0) = \sum_{t=t_0}^{\infty} G_M(x,t|x_0,t_0)s^{t-t_0}\;. \tag{2.62}$$

Differentiating Eq. (2.61) twice with respect to x and using the relation in Eq. (2.60), we obtain

$$\partial_{xx}\tilde{G}_M(x,s|x_0) = (1-s)\tilde{G}_M(x,s|x_0) - s\delta(x-x_0)\;. \tag{2.63}$$

This is a much simpler differential equation, which can be exactly solved, yielding (see Article 2 for the details)

$$\tilde{G}_M(x,s|x_0) = \frac{1-\sqrt{1-s}}{1-\left(\frac{1-\sqrt{1-s}}{1+\sqrt{1-s}}\right)^2 e^{-2\sqrt{1-s}\,M}}\left[e^{-\sqrt{1-s}\,(x-x_0)} - \frac{1-\sqrt{1-s}}{1+\sqrt{1-s}}e^{-\sqrt{1-s}\,(2M-x+x_0)}\right]\;. \tag{2.64}$$

It is worth mentioning that the nice property in Eq. (2.60) allows to solve exactly many RW problems. Indeed, one can often easily write an integral recursion relation

[2] Note that in the limit $M \to \infty$, the recursion relation in Eq. (2.58) reduces to the Wiener-Hopf equation, which can be solved for any symmetric and continuous jump distribution $p(\eta)$ using Ivanov formula [129, 130].

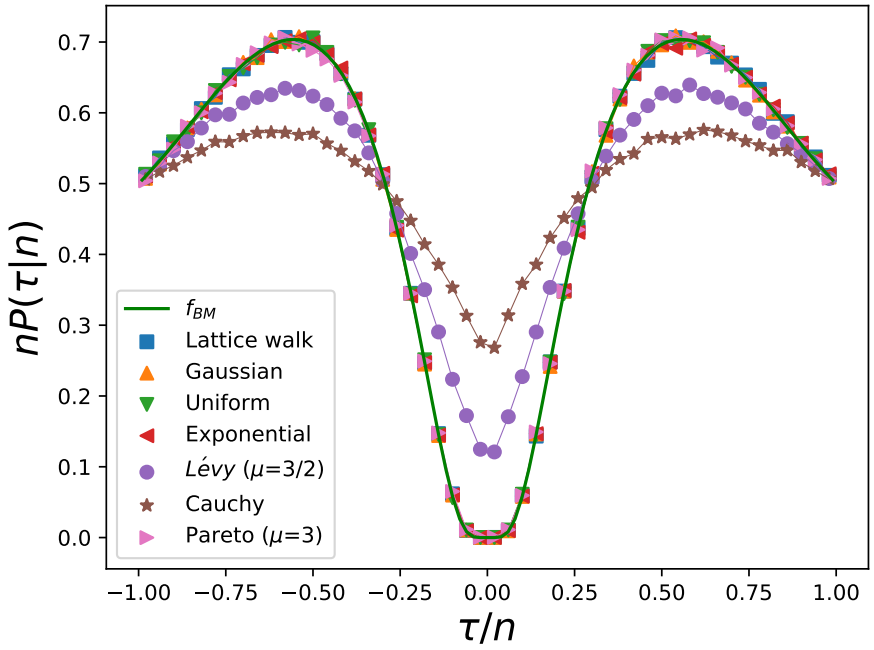

Figure 2.6: The scaled distribution $nP(\tau|n)$ as a function of τ/n for discrete-time random walks (RWs). The continuous green line corresponds to the exact scaling function $f_{\rm BM}(y)$, given in Eq. (2.21). The symbols correspond to numerical simulations performed with RWs with different jump distributions and $n = 10^5$ steps. When the variance of the jumps is finite (lattice walks, Gaussian, uniform, double-exponential and Pareto) the numerical curves collapse onto the scaling function $f_{\rm BM}(y)$, as a consequence of the Central Limit Theorem. When the jump variance is divergent, i.e. Lévy flights with index $\mu = 3/2$ and $\mu = 1$ (Cauchy distribution), the scaling function $f_\mu(y)$ depends continuously on μ. At the endpoints $\tau = \pm n$, we recover the universal result $P(\tau = n|n) = 1/(2n)$, valid even for fat tailed distributions.

similar to the one in Eq. (2.58) for the probability distribution of a given quantity. However, solving this kind of integral relation is in general highly nontrivial. In the case of the double exponential distribution, Eq. (2.60) can be used to obtain a differential equation, which is typically much easier to solve.

Using the relation in Eq. (2.57) together with the expression for the constrained propagator in Eq. (2.64), one can compute the joint distribution of $x_{\max}$, $x_{\min}$, $t_{\max}$, and $t_{\min}$. From that, one can obtain the probability distribution of τ. In particular, one obtains that in the limit of large-n

$$P(\tau|n) \approx \frac{1}{n} f_{\rm BM}\left(\frac{\tau}{n}\right) , \tag{2.65}$$

where $f_{\rm BM}(y)$ is the scaling function that we have derived in the case of BM, which is given in Eq. (2.21). The derivation of this result is technically quite complicated and is given in Article 2. Interestingly, from this computation one also obtains the integral

relation for $f_{\rm BM}(y)$ given in Eq. (2.26), which we have used to compute the moments of τ.

Lattice walks: We next consider the case of a one-dimensional random walk on a discrete lattice, corresponding to the jump distribution

$$p(\eta) = \frac{1}{2}\delta(\eta - 1) + \frac{1}{2}\delta(\eta + 1)\,. \tag{2.66}$$

In other words, at each timestep, the walker takes a unit step either to the right or to the left with the same probability. As a consequence, the constrained propagator $G_M(x, t|x_0, t_0)$ satisfies the recursion relation

$$G_M(x, t|x_0, t_0) = \frac{1}{2}G_M(x - 1, t - 1|x_0, t_0) + \frac{1}{2}G_M(x - 1, t - 1|x_0, t_0)\,. \tag{2.67}$$

The boundary conditions are

$$G_M(-1, t|x_0, t_0) = G_M(M + 1, t|x_0, t_0) = 0\,, \tag{2.68}$$

while the initial condition reads

$$G_M(x, t_0|x_0, t_0) = \delta_{x,x_0}\,. \tag{2.69}$$

This simple recursion relation can be solved by standard generating-function techniques and one obtains

$$\tilde{G}_M(x, s|x_0) = \frac{2}{s}\left(\frac{w(s)^{x+1}}{1 - w(s)^{2(M+2)}} + \frac{w(s)^{-(x+1)}}{1 - w(s)^{-2(M+2)}}\right)\,, \tag{2.70}$$

where $\tilde{G}_M(x, s|x_0)$ is the generating function of $G_M(x, t|x_0, t_0)$, defined in Eq. (2.62), and

$$w(s) = \frac{1}{s}\left(1 - \sqrt{1 - s^2}\right)\,. \tag{2.71}$$

Using this result for the constrained propagator and applying again the path-decomposition method described above, it is possible to show that also in this case

$$P(\tau|n) \approx \frac{1}{n} f_{\rm BM}\left(\frac{\tau}{n}\right)\,, \tag{2.72}$$

for large n.

Divergent jump variance

In the case where the variance σ^2 of the jumps is divergent, it is hard to make analytical progress. Thus, we performed extensive numerical simulations to investigate this case. In particular, we considered the case of Lévy walks, corresponding to the jump PDF

$$p(\eta) \sim \frac{1}{|\eta|^{\mu+1}}\,, \tag{2.73}$$

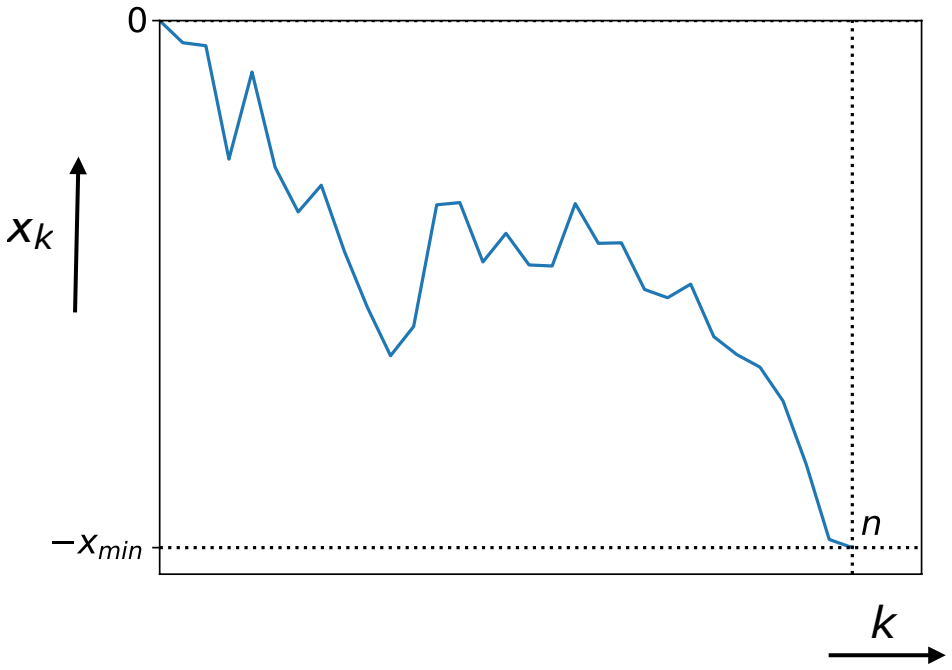

Figure 2.7: Typical realization of an n-step random walk x_k with $\tau = n$ as a function of the step k. The global maximum $x_{\max} = 0$ is reached at time zero ($t_{\max} = 0$), while the global minimum is reached at the final step ($t_{\min} = n$).

for large $|\eta|$, with $0 < \mu < 2$. Note that when $1 < \mu < 2$ the variance of η is divergent, while when $0 < \mu < 1$ both the mean and the variance are divergent.

We show numerically (see Fig. 2.6) that in the limit of large n the PDF of τ assumes the scaling form

$$P(\tau|n) \approx \frac{1}{n} f_\mu\left(\frac{\tau}{n}\right) , \tag{2.74}$$

where the scaling function $f_\mu(y)$ appears to be continuously dependent on the Lévy index μ. For $\mu = 2$, we have $f_2(y) = f_{\rm BM}(y)$. Thus, our result is somewhat less universal than the Arcsine law, which describes the asymptotic distribution of $n_{\max}$ in the limit of large number n of steps.

It is interesting to investigate the value $f_\mu(0)$ different values of $0 < \mu \leq 2$. This value $f_\mu(0)$ describes the probability that the maximum and the minimum occur very close in time. Indeed, from Eq. (2.74) we have, for large n,

$$P(\tau = 1|n) \approx \frac{f_\mu(0)}{n} . \tag{2.75}$$

For $\mu = 2$ the random walk converges for large times to Brownian motion and we find $f_2(0) = f_{\rm BM}(0) = 0$. This is in agreement with the fact that Brownian motion has a continuous path. Indeed, for the maximum and the minimum to occur at successive time-steps one needs the presence of very large jumps, which are not present for Brownian motion. These jumps, which make the trajectory discontinuous, can instead occur for

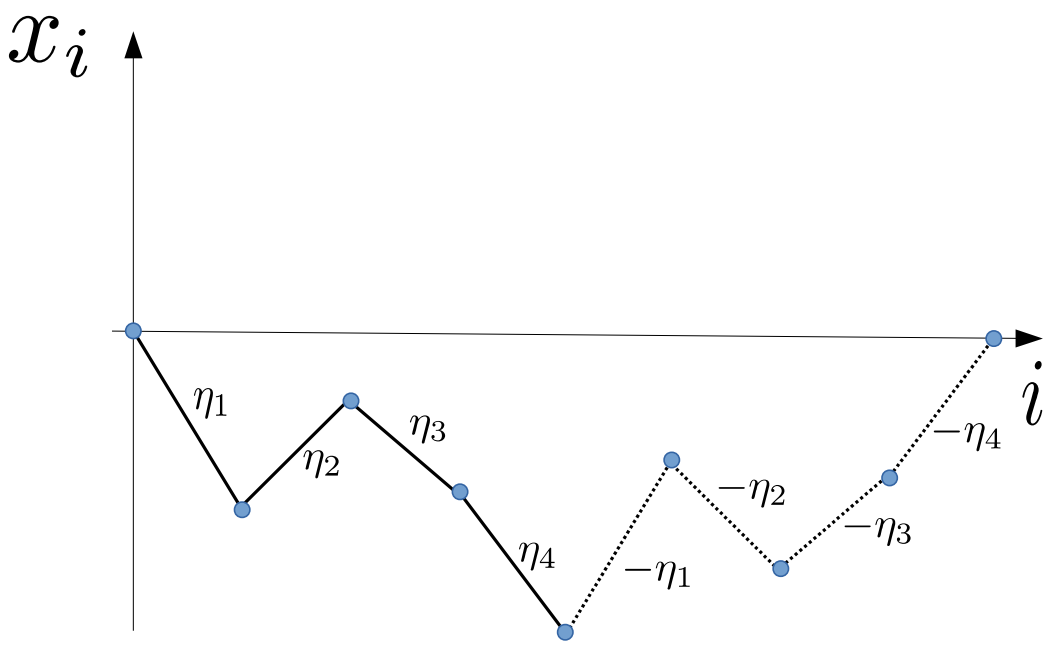

Figure 2.8: Mapping between an n-step random walk with $\tau = n$ (here $n = 4$) to a bridge configuration with $2n$ steps.

$\mu < 2$ and indeed we observe that $f_\mu(0) > 0$ for $\mu < 2$ (see Fig. 2.6). Moreover, we find that $f_\mu(0)$ increases for decreasing μ.

Interestingly, our numerical simulations suggest that, for any μ, $f_\mu(\pm 1) = 1/2$, indicating that the probabilities of the events "$\tau = n$" and "$\tau = -n$" are completely universal. Remarkably, this universality turns out to be exactly valid even when the number of steps is finite, as long as the distribution is continuous and symmetric. In other words, if $p(\eta)$ is continuous and $p(\eta) = p(-\eta)$, then

$$\boxed{P(\tau = n|n) = 1/(2n)}\,. \tag{2.76}$$

Note that the event $\tau = n$ can only occur if $t_{\max} = 0$ and $t_{\min} = n$. This corresponds to trajectories that always remain below their starting position and reach their global minimum at the final step (see Fig. 2.7). The universal result in Eq. (2.76) is shown in Fig. 2.9 and is in perfect agreement with numerical simulations. Note that in the case of lattice walks, which have a discrete jump distribution, the result in Eq. (2.76) applies only for large n.

We will next sketch the proof of this result[3] based on a combinatorial theorem by Spitzer [131]. We start by reproducing the statement of Theorem 2.1 in [131]:

Let $y = (y_1, \ldots, y_m)$ be a vector such that $y_1 + y_2 + \ldots + y_m = 0$, but no other partial sum of distinct components vanishes. Let $y_{k+m} = y_k$, and $y(k) = (y_k, y_{k+1}, \ldots, y_{k+m})$, for $k = 1, \ldots m$. Then, for each $r = 0, 1, \ldots, m-1$, exactly one of the cyclic permutations $y(k)$ of y is such that exactly r of its successive sums are positive.

[3] We warmly thank Mark Holmes for his suggestions on how to prove this result.

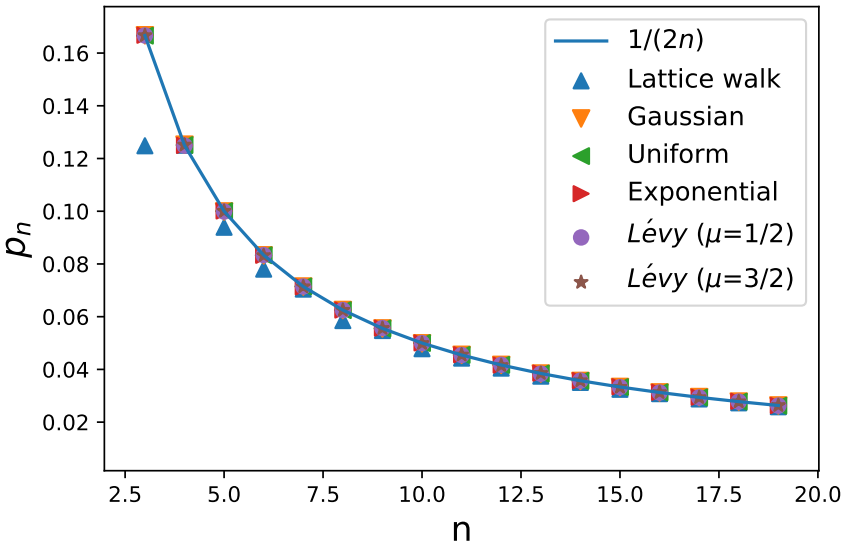

Figure 2.9: The probability $p_n = P(\tau = n|n)$ as a function of n for discrete-time random walks with n steps. Symbols correspond to the results of numerical simulations of RW with different jump distributions. The continuous blue line represents the universal exact result $p_n = 1/(2n)$, valid for continuous and symmetric jump distributions. Numerical results for different jump distributions (except lattice walks that are discrete in space) collapse onto the analytical curve for any n.

To apply this theorem to our problem, we consider an n-step RW

$$x_k = \sum_{i=1}^{k} \eta_i \,, \tag{2.77}$$

for $1 \leq k \leq n$. The increments $\eta_1 \,, \ldots \,, \eta_n$ are i.i.d. random variables with distribution $p(\eta)$, which we assume to be continuous and symmetric. Consider the vector of increments

$$y = (\eta_1 \,, \eta_2 \,, \ldots \,, \eta_{n-1} \,, \eta_n \,, -\eta_1 \,, -\eta_2 \,, \ldots \,, -\eta_{n-1} \,, -\eta_n) \tag{2.78}$$

of length $m = 2n$. This corresponds to mapping the random walk into a bridge configuration as shown in Fig. 2.8. Note that by definition

$$y_i = \begin{cases} \eta_i \,, & \text{for } 1 \leq i \leq n \,, \\ -\eta_{i-n} \,, & \text{for } n+1 \leq i \leq 2n \,. \end{cases} \tag{2.79}$$

By construction we have $y_1 + y_2 + \ldots + y_m = 0$. Moreover, since the steps are continuous, the vector satisfies almost surely the hypothesis that no other partial sum of distinct components vanishes. Thus, using the theorem above, exactly one of the $2n$ cyclic permutations of y is such that none of its successive sum are positive ($r = 0$). This means that there exists only one cyclic permutation $y(k)$ for which $\sum_{i=1}^{l} y(k)_i \leq 0$ for any l. Moreover, using the fact that the distribution of the steps is symmetric

(meaning that η and $-\eta$ have the same distribution) it is easy to show that the $2n$ cyclic permutations are equiprobable, implying

$$\text{Prob.}\left(\sum_{i=1}^{l} y_i \leq 0 \text{ for any } 1 \leq l \leq 2n\right) = \frac{1}{2n}\,. \tag{2.80}$$

We would now like to show that this event "$\sum_{i=1}^{l} y_i \leq 0$ for any $1 \leq l \leq 2n$" is equivalent to the event "$\tau = n$". Thus, we need to show that $\sum_{i=1}^{l} y_i \leq 0$ for any $1 \leq l \leq 2n$ if and only if i) the global maximum of the random walk x_k is reached at the first step ($t_{\max} = 0$) and ii) the global minimum is reached at the last step ($t_{\min} = n$).

Let us start with the first condition i). Since by definition $y_i = \eta_i$ for $i \leq n$ (see Eq. (2.79)), the condition $\sum_{i=1}^{l} y_i \leq 0$ can be rewritten as, for $1 \leq l \leq n$,

$$\sum_{i=1}^{l} \eta_i \leq 0\,. \tag{2.81}$$

This is equivalent to

$$x_l \leq 0 \tag{2.82}$$

for all $1 \leq l \leq n$. Meaning that the maximum is reached at the first step.

To obtain the second condition ii), i.e., that the global minimum of the random walk x_k is reached at the final step n, we consider the condition $\sum_{i=1}^{l} y_i \leq 0$ with $n+1 \leq l \leq 2n$. This condition can be written as

$$\sum_{i=1}^{n} y_i + \sum_{i=n+1}^{l} y_i \leq 0\,. \tag{2.83}$$

Using the fact that $y_i = \eta_i$ for $1 \leq i \leq n$ and that $y_i = -\eta_{i-n}$ for $n+1 \leq i \leq 2n$ (see Eq. (2.79)), we get

$$\sum_{i=1}^{n} \eta_i - \sum_{i=n+1}^{l} \eta_{i-n} \leq 0\,. \tag{2.84}$$

Shifting the index in the second sum, we find

$$\sum_{i=1}^{n} \eta_i - \sum_{i=1}^{l-n} \eta_i \leq 0\,. \tag{2.85}$$

This can be rewritten in terms of the random walk variables x_k as

$$x_n \leq x_{l-n}\,, \tag{2.86}$$

with $n+1 \leq l \leq 2n$. Defining $j = l - n$, this condition can be rewritten as

$$x_n \leq x_j\,, \tag{2.87}$$

with $1 \leq j \leq n$. Finally, this implies that the global minimum is reached at the final step n, which was condition ii).

Thus, we have shown that the events "$\sum_{i=1}^{l} y_i \leq 0$ for any $1 \leq l \leq 2n$" and "$\tau = n$" are actually the same event, since they correspond to the same conditions on the random

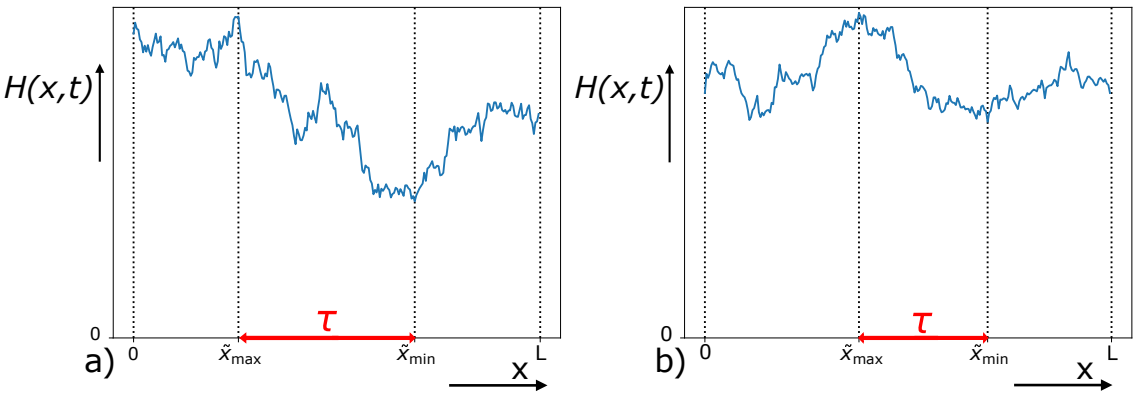

Figure 2.10: Kardar-Parisi-Zhang interfaces in $(1+1)$ dimensions, growing on a finite substrate of size L with free boundary conditions (**a**) and periodic boundary conditions (**b**). The interface profile $H(x,t)$ is displayed as a function of x and for t fixed. The distance between the position $\tilde{x}_{\max}$ of maximal height and the position $\tilde{x}_{\min}$ of minimal height is denoted by τ and indicated by red double arrows.

walk x_k. This means that they occur with the same probability. Thus, using the result in Eq. (2.80), we find that

$$P(\tau = n|n) = \text{Prob.}\left(\sum_{i=1}^{l} y_i \leq 0 \text{ for any } 1 \leq l \leq 2n\right) = \frac{1}{2n}\,. \tag{2.88}$$

This result can also be explicitly verified for any finite n in the case of RWs with double-exponential distribution (see Article 2). Note that this result was only conjectured in Article 2 and that this demonstration, to the best of our knowledge, was not published before.

2.1.4 Fluctuating interfaces

Our results of the previous sections turn out to be directly applicable to the study of extremal properties of Kardar-Parisi-Zhang (KPZ) and Edwards-Wilkinson (EW) interfaces. We consider a $(1+1)-$dimensional interface evolving over a substrate of finite size L. We denote by $H(x,t)$ the height of the interface at position x and at time t, where $0 \leq x \leq L$ (see Fig. 2.10 for a typical realization) [132–136]. The interface height evolves according to the celebrated KPZ equation [133]

$$\frac{\partial H(x,t)}{\partial t} = \frac{\partial^2 H(x,t)}{\partial x^2} + \lambda\left(\frac{\partial H(x,t)}{\partial x}\right)^2 + \eta\,(x,t)\,, \tag{2.89}$$

where $\lambda \geq 0$ and $\eta(x,t)$ is a Gaussian white noise with zero mean and correlator $\langle \eta\,(x,t)\,\eta\,(x',t')\rangle = 2\delta(x-x')\delta(t-t')$. Note that for $\lambda > 0$ the KPZ equation is non-linear. Setting $\lambda = 0$, one obtains the EW equation [132], which is instead linear. For the Eq. (2.89) one can either consider free boundary conditions (FBC), with the height values $H(0,t)$ and $H(L,t)$ at the ends of the substrate let free to evolve, or periodic boundary conditions (PBC), with the constraint $H(0,t) = H(L,t)$.

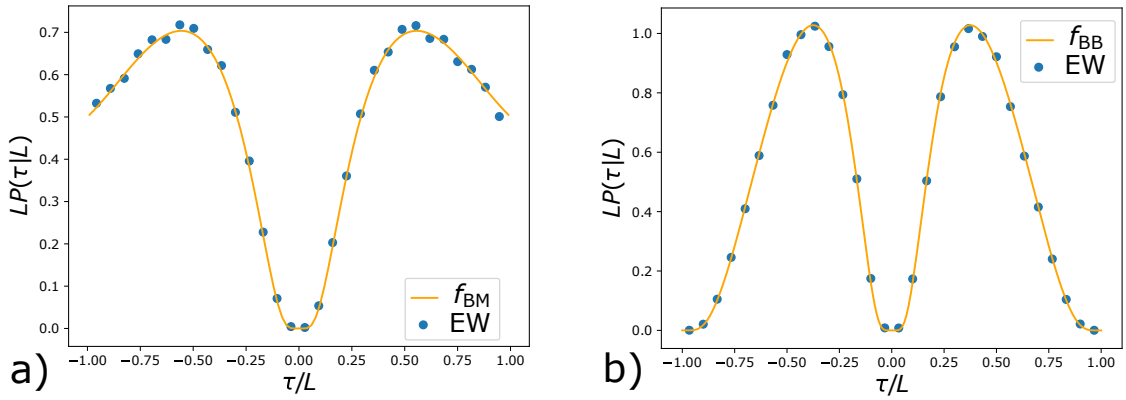

Figure 2.11: The scaled distribution $L\ P(\tau|L)$ for Edwards-Wilkinson interfaces as a function of the scaled distance τ/L between the points of maximal and minimal height with free boundary conditions (**a**) and periodic boundary conditions (**b**). The solid blue lines show the exact scaling functions $f_{\mathrm{BM}}(y)$ (in panel **a**), given in Eq. (2.21) and $f_{\mathrm{BB}}(y)$ (in panel **b**), given in Eq. (2.44). The filled blue dots represent the results of numerical simulations with $L = 512$.

We will be interested in investigating the extremal properties of the steady state of these systems. Note however that the space-averaged height (which is a random variable)

$$\overline{H(t)} = \frac{1}{L}\int_0^L H(x,t)\,dx\,, \tag{2.90}$$

never reaches a steady state. As a consequence, a more appropriate observable that does reach a steady state is the displacement from the space-averaged height, also known as relative height, which is defined as

$$h(x,t) = H(x,t) - \overline{H}(t)\,. \tag{2.91}$$

Note that, by construction, the space average of $h(x,t)$ is zero, i.e., that

$$\frac{1}{L}\int_0^L dx\ h(x,t) = 0\,. \tag{2.92}$$

It can be shown that the relative height reaches a steady state for $t \to \infty$.

To make a connection with the problems studied in the previous sections, we introduce the position $\tilde{x}_{\max}$ ($\tilde{x}_{\min}$) at which the steady-state relative height $h(x)$ is maximal (minimal):

$$\tilde{x}_{\max} = \mathrm{argmax}_{0\le x\le L}\left(h(x)\right)\,, \tag{2.93}$$

and

$$\tilde{x}_{\min} = \mathrm{argmin}_{0\le x\le L}\left(h(x)\right)\,. \tag{2.94}$$

These two quantities are analogous to the time $t_{\max}$ of the maximum and the time $t_{\min}$ of the minimum of a stochastic process. We also denote by $h_{\max} = h(\tilde{x}_{\max})$ and $h_{\min}(\tilde{x}_{\min})$ the values of the maximal and minimal relative height (see Fig. 2.10).

Edwards-Wilkinson case

We first study the linear case $\lambda = 0$, corresponding to EW interfaces. In the case of FBC, the steady-state relative height distribution reads [15, 16, 137]

$$P_{\rm st}\left(\{h\}\right) = \sqrt{2\pi} L^{3/2}\, e^{-\frac{1}{2}\int_0^L dx (\partial_x h)^2} \delta\left[\int_0^L h(x) dx\right] , \tag{2.95}$$

where the δ function imposes the zero-area constraint in Eq. (2.92). In the case of PBC, the steady state instead is given by

$$P_{\rm st}\left(\{h\}\right) = L\, e^{-\frac{1}{2}\int_0^L dx (\partial_x h)^2} \delta\left[\int_0^L h(x) dx\right] \delta\left(h(0) - h(L)\right) , \tag{2.96}$$

where the term $\delta\left(h(0) - h(L)\right)$ enforces the periodicity constraint.

Interestingly, for FBC (PBC), the relative height $h(x)$ behaves in the steady state as a BM (BB) with a global constrain on the area. To connect this problem to the discussion of the previous sections, we identify: space with time, i.e. $x \Leftrightarrow t$, the length L of the substrate with the total duration T of the process, i.e. $L \Leftrightarrow T$, and the stationary relative height $h(x)$ in the case of FBC (or PBC) with the position $x(t)$ of a BM (or BB), i.e. $h(x) \Leftrightarrow x(t)$.

In addition to the usual BM/BB dynamics, the process $x(t)$ obtained with this mapping has to also satisfy the zero-area constraint

$$\int_0^T dt\, x(t) \;=\; 0\,, \tag{2.97}$$

which derives from the condition in Eq. (2.92). This additional condition will in general affect the statistical properties of the process, e.g., the distribution of the maximal height $h_{\rm max}$ is known to differ from that of a BM [15, 16]. However, the zero-area constraint does not affect the joint distribution of $\tilde{x}_{\rm max}$ and $\tilde{x}_{\rm min}$. Indeed, this constraint corresponds to a vertical shift of the interface profile which does not affect the positions at which the extrema are attained. As a consequence, the steady-state distribution of the distance τ between the points of maximal and minimal height coincides with the distribution of the time τ between the maximum and the minimum of a BM or a BB (depending on the boundary conditions).

Thus, in the case of FBC, we obtain

$$P(\tau = \tilde{x}_{\rm min} - \tilde{x}_{\rm max}|L) = \frac{1}{L} f_{\rm BM}(\frac{\tau}{L})\,, \tag{2.98}$$

where the scaling function $f_{\rm BM}(y)$ is given in Eq. (2.21). On the other hand, for PBC, we get

$$P(\tau = \tilde{x}_{\rm min} - \tilde{x}_{\rm max}|L) = \frac{1}{L} f_{\rm BB}(\frac{\tau}{L})\,, \tag{2.99}$$

where $f_{\rm BB}(y)$ is given in Eq. (2.44). These results are shown in Fig. 2.11 and are in good agreement with numerical simulations (for the details of how the simulations were performed, see Article 2).

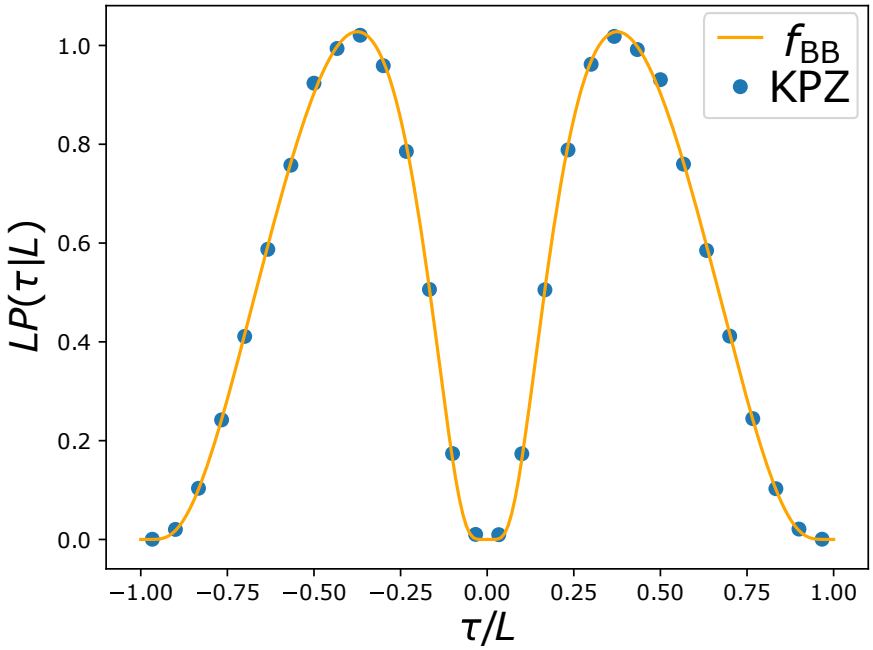

Figure 2.12: The scaled distribution $L\ P(\tau|L)$ for Kardar-Parisi-Zhang interfaces as a function of the scaled distance τ/L between the points of maximal and minimal height with periodic boundary conditions. The continuous orange line is the exact scaling function, given in Eq. (2.44), while the filled blue circles represent the results of numerical simulations with $L = 512$.

Kardar-Parisi-Zhang case

When $\lambda > 0$ in Eq. (2.89), corresponding to KPZ interfaces, the stationary measure in (2.95), valid for FBC, is only valid in the limit $L \to \infty$. Thus, the result for $P(\tau|L)$ in Eq. (2.98) is expected to hold for the KPZ equation only for large L. However, due to nonlinear term in Eq. (2.89), numerical integration of the KPZ equation is known to be challenging [138, 139]. On the other hand, the steady state in Eq. (2.96) for PBC is valid even for finite L. Using the discretization scheme proposed in [139], we were able to confirm our theoretical prediction in the case of PBC (see Fig. 2.12). However, for FBC such discretization scheme does not work and verifying numerically the result in Eq. (2.95) appears to be computationally challenging.

2.2 Time of the maximum of stationary processes

The time $t_{\max}$ of the maximum of a one-dimensional stochastic process $x(t)$, with $t \in [0, T]$, has been investigated for a wide range of models. However, to the best of our knowledge, before our Article 3, it was never studied in the case of stationary processes, i.e., processes that are invariant under a time shift. Stationary processes can be observed at different scales in nature, from Brownian engines [140] to climate sys-

tems [141]. They appear in two main categories: equilibrium and out-of-equilibrium. At equilibrium, all currents (of energy, matter, or even probability) vanish, while by definition nonequilibrium stationary processes display a nonzero current in the steady state. Equivalently, a stochastic process is said to be at equilibrium if it satisfies the detailed balance condition. Many classical results from statistical physics apply to equilibrium systems and thus their physical properties are generally well understood. In contrast, much less is known about nonequilibrium systems and only few exact results have recently become available [27, 28, 142–145].

We will be interested in the case where the autocorrelation function of the process decays exponentially fast in time

$$\langle x(t)x(t')\rangle - \langle x(t)\rangle\langle x(t')\rangle \sim e^{-|t-t'|/\xi}\,, \tag{2.100}$$

where ξ is the autocorrelation time. Then, depending on the observation time T, the system can be described either as strongly or weakly correlated. Indeed, when $T \ll \xi$, the random variables $x(t)$ at different times t are strongly correlated. On the other hand for $T \gg \xi$, as discussed in detail in Chapter 1, one can apply a "block renormalization" argument, which maps the process into a collection of i.i.d. random variables, where many exact results are known for the statistics of the extremes. In particular, one would expect the universal results on the distribution of the maximum M (valid for i.i.d. variables) to be valid for $T \gg \xi$.

It is therefore relevant to ask if one can identify models for which the extremal properties (e.g., the value M or the time $t_{\max}$ of the maximum) can be exactly computed for any $T > 0$. Indeed, an exact result valid for any T would interpolate between a strongly-correlated system (for small T) and a weakly correlated one (for large T). Interestingly, the distribution of the maximum M has been studied for different stationary processes, both at equilibrium, e.g., the Ornstein-Uhlenbeck process [8], and out-of-equilibrium, e.g. BM with stochastic resetting [36, 146]. In both cases, for $T \gg \xi$, it was explicitly shown that the distribution of M approaches the universal Gumbel form. However, it is not clear a priori whether or not this universality also extends to the distribution of the time $t_{\max}$ of the maximum.

At any time T, since by definition the stochastic process is invariant under a time translation, one could naively expect the distribution of $t_{\max}$ to be uniform in the interval $[0, T]$, i.e.,

$$P(t_{\max}|T) = \frac{1}{T}\,. \tag{2.101}$$

Interestingly, we show that, as a consequence of the correlations of the process, this is not the case in general.

Moreover, even though the uniform distribution in Eq. (2.101) does not exactly describe the distribution of $t_{\max}$, applying the block argument discussed in Chapter 1 one obtains, for $T \gg \xi$,

$$P(t_{\max}|T) \approx \frac{1}{T}\,. \tag{2.102}$$

As we will show, this result is only valid in the "bulk" of the distribution of $t_{\max}$, i.e., for $\xi \ll t_{\max} \ll (T - \xi)$. We will observe that the i.i.d. approximation breaks down in the "edge" regions where $t_{\max} \sim \xi$ (left edge) and $T - t_{\max} \sim \xi$ (right edge).

To be precise, in this section, we consider a one-dimensional process $x(t)$ with $t \in [0, T]$ and we assume that at time $t = 0$ the system starts from the steady state. This is equivalent to assuming that we prepare the system in some initial condition at time $t = -\infty$ and that we start to observe it at time $t = 0$. We denote by $P_{\rm st}(x)$ the stationary distribution of the process. The initial position $x_0 = x(0)$ is thus a random variable independently drawn from $P_{\rm st}(x)$.

In the following, we will present the main steps leading to our results. These results were recently announced in our letter 3. The precise details of the computations will be published in an article in preparation [147].

2.2.1 Equilibrium processes

We start with the case of equilibrium systems. The process that we consider is a one-dimensional BM with diffusion coefficient D subject to an external confining potential $V(x)$. We assume that this potential grows as $V(x) \approx \alpha |x|^p$ for large x, with $\alpha > 0$ and $p > 0$. The Langevin equation describing the evolution of the system reads

$$\frac{dx(t)}{dt} = -V'(x) + \eta(t)\,, \tag{2.103}$$

where $V'(x) = dV/dx$ and $\eta(t)$ is Gaussian white noise with zero mean and correlator $\langle \eta(t)\eta(t')\rangle = 2D\delta(t-t')$. For this system, the stationary equilibrium distribution is given by the Gibbs-Boltzmann measure

$$P_{\rm st}(x) = \frac{1}{Z} e^{-V(x)/D}\,, \tag{2.104}$$

where Z is a normalization constant.

To compute the distribution of $t_{\max}$, one can use the path-decomposition strategy presented in Chapter 1. Indeed, the result in Eq. (1.49) still applies in the case of stationary processes, with the only difference that one has to average over the initial position x_0, yielding

$$\begin{aligned} P(t_{\max}|T) &= \lim_{\epsilon \to 0} \Big[\mathcal{N}(\epsilon) \int_{-\infty}^{\infty} dx_0 P_{\rm st}(x_0) \\ &\times \int_{x_0}^{\infty} dM G^M(M-\epsilon, t_{\max}|x_0) Q^M(M-\epsilon, T-t_{\max}) \Big]\,, \end{aligned} \tag{2.105}$$

where $P_{\rm st}(x_0)$ is given in Eq. (2.104). We recall that $G^M(x,t|x_0)$ is the constrained propagator, defined as the probability density of going from position x_0 to position x in time t, while always remaining below position M. We also recall that the survival probability $Q^M(x,t)$ is defined as the probability of remaining below position M up to time t, having started from position x at the initial time.

For the Langevin equation in Eq. (2.103), the propagator satisfies the forward Fokker-Plank equation [24]

$$\partial_t G^M(x,t|x_0) = D\partial_x^2 G^M(x,t|x_0) + \partial_x \left[V'(x)\; G^M(x,t|x_0) \right]\,, \tag{2.106}$$

with initial condition

$$G^M(x,0|x_0) = \delta(x-x_0)\,. \tag{2.107}$$

The boundary conditions are

$$G^M(M,t|x_0) = 0\,, \tag{2.108}$$

corresponding to an absorbing wall at $x = M$, and

$$\lim_{x\to-\infty} G^M(x,t|x_0) = 0\,. \tag{2.109}$$

This condition in Eq. (2.109) follows from the observation that if the particle starts from position x_0 at the initial time, it cannot be infinitely far from this initial position at the finite time t.

On the other hand, the survival probability evolves according to the backward Fokker-Plank equation [24]

$$\partial_t Q^M(x,t) = \partial_x^2 Q^M(x,t) - V'(x)\partial_x Q^M(x,t)\,, \tag{2.110}$$

with initial condition

$$Q^M(x,0) = 1\,. \tag{2.111}$$

The boundary conditions are

$$Q^M(M,t) = 0\,, \tag{2.112}$$

meaning that a particle starting from the boundary is immediately absorbed and

$$\lim_{x\to-\infty} Q^M(x,t) = 1\,, \tag{2.113}$$

since a particle starting infinitely far away from the absorbing wall will never be absorbed in a finite time.

The Fokker-Planck equations (2.106) and (2.110) are usually easier to solve in Laplace space, i.e., taking a Laplace transform with respect to t. Thus, we rewrite the relation in Eq. (2.105) in Laplace space. It is useful to introduce the variables $t_1 = t_{\max}$, corresponding to the time of the maximum, and $t_2 = T - t_{\max}$, corresponding to the remaining time after the maximum. Considering the double Laplace transform of Eq. (2.105) with Laplace variables s_1 and s_2, corresponding to t_1 and t_2 respectively, we obtain

$$\begin{aligned} &\int_0^\infty dt_1\ e^{-s_1 t_1} \int_0^\infty dt_2\ e^{-s_2 t_2}\ P(t_{\max} = t_1 | T = t_1 + t_2) \qquad (2.114)\\ &= \lim_{\epsilon\to 0}\left[\mathcal{N}(\epsilon)\int_{-\infty}^\infty dx_0\ \int_{x_0}^\infty dM\ P_{\rm st}(x_0)\tilde{G}^M(M-\epsilon, s_1|x_0)\tilde{Q}^M(M-\epsilon, s_2)\right], \end{aligned}$$

where we have defined

$$\tilde{G}^M(x,s|x_0) = \int_0^\infty dt\ e^{-st} G^M(x,t|x_0)\,. \tag{2.115}$$

and

$$\tilde{Q}^M(x,s) = \int_0^\infty dt\ e^{-st} Q^M(x,t)\,. \tag{2.116}$$

Solving the differential equations (2.106) and (2.110), we will derive an exact expression for the distribution of $P(t_{\max}|T)$ for the potentials $V(x) = \alpha|x|$ (corresponding to $p = 1$) and $V(x) = \alpha x^2$ (corresponding to $p = 2$). Moreover, in the limit $T \gg 1$, we will improve the simple blocking argument presented in Chapter 1, showing that the full distribution $P(t_{\max}|T)$ becomes universal at late times for any $p > 0$.

The case $p = 1$

We start by investigating the case $p = 1$, corresponding to $V(x) = \alpha|x|$. The equilibrium distribution of the process reads

$$P_{\rm st}(x) = \frac{D}{2\alpha} e^{-\alpha|x|/D} \,. \tag{2.117}$$

The forward Fokker-Plank equation (2.106) can be exactly solved in this case and, at leading order in ϵ, we find

$$\tilde{G}^M(M-\epsilon, s|x_0) \approx \begin{cases} \frac{\epsilon}{D} e^{(\alpha-k)(M-x_0)/(2D)} & \text{if } x_0 < M < 0 \,, \\ \frac{\epsilon}{D} \frac{(k-\alpha)e^{kx_0/D}+\alpha}{(k-\alpha)e^{kM/D}+\alpha} e^{(-\alpha+k)(M-x_0)/(2D)} & \text{if } 0 < x_0 < M \,, \\ \frac{k\epsilon}{D} \frac{e^{(k-\alpha)x_0/(2D)} e^{(-k-\alpha)M/(2D)}}{k-\alpha+\alpha e^{-kM/D}} & \text{if } x_0 < 0 \text{ and } M > 0 \,, \end{cases} \tag{2.118}$$

where $k = \sqrt{\alpha^2 + 4sD}$. Similarly, solving the backward equation (2.110), we obtain

$$\tilde{Q}^M(M-\epsilon, s) \approx \begin{cases} \frac{\epsilon}{s} \frac{k-\alpha}{2D} \frac{(k+\alpha)e^{kM/D}-\alpha}{(k-\alpha)e^{kM/D}+\alpha} & \text{if } M > 0 \,, \\ \frac{\epsilon}{s} \frac{k-\alpha}{2D} & \text{if } M < 0 \,, \end{cases} \tag{2.119}$$

where we have expanded at leading order in ϵ.

Substituting the expressions for $P_{\rm st}(x)$, $\tilde{G}^M(M-\epsilon, s|x_0)$, and $\tilde{Q}^M(M-\epsilon, s)$, into Eq. (2.105), we get

$$\begin{aligned} &\int_0^\infty dt_1 \; e^{-s_1 t_1} \int_0^\infty dt_2 \; e^{-s_2 t_2} \; P(t_{\rm max} = t_1 | T = t_1 + t_2) = \lim_{\epsilon \to 0} \left[\mathcal{N}(\epsilon)\epsilon^2\right] \frac{k_2 - \alpha}{4D\alpha s_2} \\ \times \; &\left\{ \int_{-\infty}^0 dM \int_{-\infty}^M dx_0 \; e^{\alpha x_0/D} e^{(\alpha - k_1)(M-x_0)/(2D)} + \int_0^\infty dM \int_0^M dx_0 \; e^{-\alpha x_0/D} \right. \\ \times \; &\frac{(k_1-\alpha)e^{k_1 x_0/D} + \alpha}{(k_1-\alpha)e^{k_1 M/D} + \alpha} e^{(-\alpha+k_1)(M-x_0)/(2D)} \frac{(k_2+\alpha)e^{k_2 M/D} - \alpha}{(k_2-\alpha)e^{k_2 M/D} + \alpha} \\ + \; &\left. \int_0^\infty dM \int_{-\infty}^0 dx_0 \; e^{\alpha x_0/D} k_1 \frac{e^{(k_1-\alpha)x_0/(2D)} e^{(-k_1-\alpha)M/(2D)}}{k_1 - \alpha + \alpha e^{-k_1 M/D}} \frac{(k_2+\alpha)e^{k_2 M/D} - \alpha}{(k_2-\alpha)e^{k_2 M/D} + \alpha} \right\} , \end{aligned} \tag{2.120}$$

where we have defined $k_1 = \sqrt{\alpha^2 + 4Ds_1}$ and $k_2 = \sqrt{\alpha^2 + 4Ds_2}$.

The constant $\mathcal{N}(\epsilon)$ can be determined using the normalization condition of $P(t_{\rm max}|T)$ by setting $s_1 = s_2 = s$ on both sides of Eq. (2.120). Indeed, the left-hand side becomes

$$\begin{aligned} \int_0^\infty dt_1 \int_0^\infty dt_2 \; e^{-s(t_1+t_2)} \; P(t_{\rm max} = t_1 | T = t_1 + t_2) &= \int_0^\infty dT \; e^{-sT} \int_0^T dt_{\rm max} \; P(t_{\rm max}|T) \\ &= \int_0^\infty dT \; e^{-sT} = \frac{1}{s} \,. \end{aligned} \tag{2.121}$$

Setting $s_1 = s_2 = s$ and computing the integrals over x_0 and M, we find that the right-hand side of Eq. (2.120) becomes

$$\lim_{\epsilon \to 0} \left[\mathcal{N}(\epsilon)\epsilon^2\right] \frac{D}{\alpha^2 s} \,. \tag{2.122}$$

Thus, equating the left and the right-hand side, we find

$$\lim_{\epsilon\to 0}\left[\mathcal{N}(\epsilon)\epsilon^2\right] = \frac{\alpha^2}{D}\,. \tag{2.123}$$

Using this expression for the normalization constant and computing the integrals over x_0, we obtain, after few steps of algebra

$$\begin{aligned}&\int_0^\infty dt_1\; e^{-s_1 t_1}\int_0^\infty dt_2\; e^{-s_2 t_2}\; P(t_{\max}=t_1|T=t_1+t_2) \qquad (2.124)\\ =\;& \frac{2\alpha}{(k_1+\alpha)(k_2+\alpha)}\left[\frac{D}{\alpha}+\int_0^\infty e^{-\alpha M/D}\frac{(k_1+\alpha-\alpha e^{-k_1 M/D})(k_2+\alpha-\alpha e^{-k_2 M/D})}{(k_1-\alpha+\alpha e^{-k_1 M/D})(k_1-\alpha+\alpha e^{-k_2 M/D})}\right].\end{aligned}$$

Interestingly, the PDF $P(t_{\max}|T)$ can be rewritten in the scaling form

$$\boxed{P(t_{\max}|T) = \frac{\alpha^2}{4D}F_1\left(\frac{\alpha^2}{4D}t_{\max}, \frac{\alpha^2}{4D}(T-t_{\max})\right)}\,. \tag{2.125}$$

Plugging this expression into Eq. (2.125), we find that the double Laplace transform of $F_1(T_1,T_2)$ is given by

$$\begin{aligned}\tilde{F}_1(s_1,s_2) &= \frac{1}{2(1+\sqrt{1+s_1})(1+\sqrt{1+s_2})} \qquad (2.126)\\ \times\;& \left[1+\int_0^\infty dz\, e^{-z}\frac{\left(\sqrt{1+s_1}+1-e^{-\sqrt{1+s_1}z}\right)\left(\sqrt{1+s_2}+1-e^{-\sqrt{1+s_2}z}\right)}{\left(\sqrt{1+s_1}-1+e^{-\sqrt{1+s_1}z}\right)\left(\sqrt{1+s_2}-1+e^{-\sqrt{1+s_2}z}\right)}\right].\end{aligned}$$

where we have defined

$$\tilde{F}_1(s_1,s_2) = \int_0^\infty dT_1 e^{-s_1 T_1}\int_0^\infty dT_2 e^{-s_2 T_2}F(T_1,T_2)\,. \tag{2.127}$$

The variables $T_1 = \alpha^2 t_{\max}/(4D)$ and $T_2 = \alpha^2(T-t_{\max})/(4D)$ represent the rescaled versions of the time of the maximum $t_{\max}$ and of the time $T-t_{\max}$ after the maximum. We will use this notation in the coming sections as well.

Since the Laplace transform in Eq. (2.126) is invariant under exchange of s_1 and s_2, we find that $F_1(T_1,T_2) = F_1(T_2,T_1)$. This in turn implies that the PDF $P(t_{\max}|T)$ is symmetric around the midpoint $t_{\max} = T/2$, i.e., that $P(t_{\max}|T) = P(T-t_{\max}|T)$. This is a consequence of the time-reversal symmetry of equilibrium processes. From this observation it is easy to show that the first moment of $t_{\max}$ is given by

$$\langle t_{\max}\rangle = \frac{T}{2}\,. \tag{2.128}$$

The exact expression in Eq. (2.126) interpolates between the limit of strongly correlated variables at short times ($T \ll D/\alpha^2$) and the weakly-correlated regime at late times ($T \gg D/\alpha^2$). Indeed, for this process, the autocorrelation function decays over a typical timescale $\xi \sim D/\alpha^2$. Even if it is hard to exactly invert the double Laplace transform in Eq. (2.126) at all times T, this expression can be used to extract the asymptotic behavior in the two regimes.

In particular, the short-time limit $T \ll 4D/\alpha^2$ can be obtained by first expanding the expression in Eq. (2.126) for large s_1 and s_2, and then inverting the double Laplace transform, yielding

$$P(t_{\max}|T) \approx \frac{1}{\pi\sqrt{t_{\max}(T-t_{\max})}}\,, \tag{2.129}$$

corresponding to the Arcsine law, describing the distribution of the time of the maximum for free BM (see Eq. (1.53)). Thus, when observed during a short time window, the process locally behaves like a BM and the states of the process at different times are strongly correlated.

On the other hand, when we observe the system for a very long time $T \gg D/\alpha^2$, one expects the process to roughly behave like a set of i.i.d. random variables and the distribution of $t_{\max}$ to be uniform (see Eq. (2.102)). Interestingly, this argument correctly predicts the shape of the PDF $P(t_{\max}|T)$ only in the bulk of the distribution, i.e., for $\xi \ll t_{\max} \ll (T-\xi)$ (where the correlation time is $\xi \sim D/\alpha^2$ in this case). The bulk regime corresponds, in Laplace space, to the limit $s_1\,, s_2 \to 0$ with s_1/s_2 fixed (recall that the Laplace variables s_1 and s_2 are respectively conjugate to $t_{\max}$ and $T - t_{\max}$). Taking this limit on the right-hand side of Eq. (2.126) and inverting the double Laplace transform one indeed obtains

$$P(t_{\max}|T) \approx \frac{1}{T}\,. \tag{2.130}$$

However, this result in Eq. (2.130) is not valid in the edge regimes $t_{\max} \sim \xi$ (left edge) and $T - t_{\max} \sim \xi$ (right edge). We will just focus on the left edge, the right edge can be obtained using the symmetry $P(t_{\max}|T) = P(T - t_{\max})$. To study this edge behavior, we have to take the limit $s_2 \to 0$ (corresponding to $T \to \infty$) in Eq. (2.126), while keeping $s_1 \sim \mathcal{O}(1)$, i.e., $t_{\max} \sim \mathcal{O}(1)$. Taking this limit in Eq. (2.126) and then inverting the double Laplace transform, we obtain [147]

$$P(t_{\max}|T) \approx \frac{1}{T} G\left(\frac{\alpha^2}{4D} t_{\max}\right) \tag{2.131}$$

where

$$G(z) = \frac{1}{2}\left[1 + \mathrm{erf}(\sqrt{z}) + \frac{1}{\sqrt{\pi z}} e^{-z}\right]\,. \tag{2.132}$$

This function $G(z)$ has asymptotic behaviors

$$G(z) \approx \begin{cases} 1/(2\sqrt{\pi z}) & \text{for } z \to 0\,, \\ \\ 1 + e^{-z}/(4\sqrt{\pi} z^{3/2}) & \text{for } z \to \infty\,. \end{cases} \tag{2.133}$$

Thus, for $t_{\max} \ll D/\alpha^2$, we find that $P(t_{\max}|T)$ diverges as $1/\sqrt{t_{\max}}$. On the other hand, for $t_{\max} \gg D/\alpha^2$ we find $P(t_{\max}|T) \approx 1/T$, smoothly connecting to the bulk regime. Similarly, using the symmetry of $P(t_{\max}|T)$, we find that in the right-edge regime where $T - t_{\max} \sim D/\alpha^2$ and $T \gg D/\alpha^2$

$$P(t_{\max}|T) \approx \frac{1}{T} G\left(\frac{\alpha^2}{4D}(T - t_{\max})\right)\,. \tag{2.134}$$

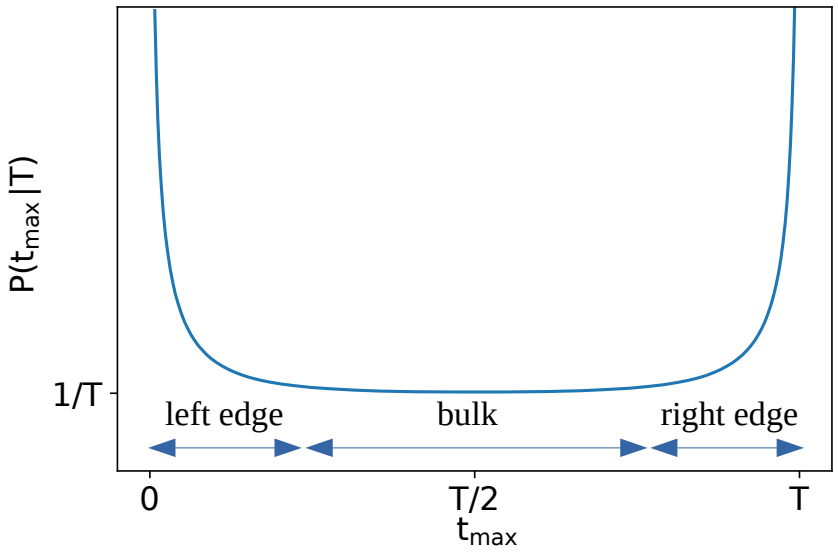

Figure 2.13: Schematic representation of the distribution $P(t_{\max}|T)$ at late times. In the "bulk" regime where $t_{\max}/T \sim \mathcal{O}(1)$ the distribution is flat, while it diverges for $t_{\max} \to 0$ and $t_{\max} \to T$, in the "edge" regime. The crossover between the two regimes is described by the function $G(z)$, given in Eq. (2.132).

To summarize, the function $G(z)$ describes the crossover between the edge and the bulk regime (see Fig. 2.13). Interestingly, we will show that this behavior of the PDF $P(t_{\max}|T)$ is very general.

The case $p = 2$ (the Ornstein-Uhlenbeck process)

We now investigate the case $p = 2$, i.e., the case of BM in a harmonic confining potential $V(x) = \alpha|x|^2$, corresponding to the Ornstein-Uhlenbeck process. In this case, the equilibrium state reads

$$P_{\rm st}(x_0) = \sqrt{\frac{\alpha}{\pi D}} \exp\left(-\frac{\alpha}{D} x_0^2\right) . \tag{2.135}$$

As before, in order to compute the distribution of $P(t_{\max}|T)$, we need to compute the constrained propagator $G^M(x,t|x_0)$ and the survival probability $Q^M(x,t|x_0)$. The procedure is similar to the one presented in the previous section.

Solving the forward Fokker-Planck equation (2.106) with $V(x) = \alpha x^2$, we obtain, at leading order in ϵ,

$$\tilde{G}^M(M-\epsilon, s|x_0, 0) \approx \frac{\epsilon}{D} e^{-(M^2-x_0^2)\alpha/(2D)} \frac{D_{-s/(2\alpha)}\left(-\sqrt{2\alpha/D}x_0\right)}{D_{-s/(2\alpha)}\left(-\sqrt{2\alpha/D}M\right)} , \tag{2.136}$$

where $D_p(z)$ is the parabolic cylinder function. Similarly, solving the backward Fokker-

Plank equation (2.110), we find

$$\tilde{Q}^M(M-\epsilon,s) \approx \frac{\epsilon}{s}\left[\frac{M2\alpha}{D} + \sqrt{\frac{2\alpha}{D}}\frac{D_{1-s/(2\alpha)}\left(-\sqrt{\frac{2\alpha}{D}}M\right)}{D_{-s/(2\alpha)}\left(-\sqrt{\frac{2\alpha}{D}}M\right)}\right], \tag{2.137}$$

where we have expanded to leading order in ϵ.

Plugging these expressions in Eqs. (2.135), (2.136), and (2.137), into the expression for the double Laplace transform of $P(t_{\max}|T)$ in Eq. (2.115), we find

$$\int_0^\infty dt_1 \int_0^\infty dt_2\, e^{-s_1 t_1 - s_2 t_2}\, P(t_{\max}=t_1|T=t_1+t_2) \tag{2.138}$$

$$= \frac{1}{\sqrt{8\pi\alpha}}\int_{-\infty}^{\infty} dz\, e^{-z^2/2}\frac{D_{-1-s_1/(2\alpha)}(-z)}{D_{-s_1/(2\alpha)}(-z)}\frac{D_{-1-s_2/(2\alpha)}(-z)}{D_{-s_2/(2\alpha)}(-z)}, \tag{2.139}$$

where, as before, we have determined the constant $\mathcal{N}(\epsilon)$ by using the normalization condition of $P(t_{\max}|T)$. Note that the distribution $P(t_{\max}|T)$ is symmetric around the midpoint $t_{\max} = T/2$, as a consequence of the time-reversal symmetry of equilibrium processes. This symmetry is confirmed by numerical simulations (see Fig. 2.17a). Using this symmetry it is easy to show that the first moment is simply given by

$$\langle t_{\max}\rangle = \frac{T}{2}. \tag{2.140}$$

We can rewrite the PDF of $t_{\max}$ in the following scaling form

$$\boxed{P(t_{\max}|T) = \alpha F_{\rm OU}(\alpha t_{\max}, \alpha(T-t_{\max}))\,.} \tag{2.141}$$

where the scaling function $F_{\rm OU}(T_1,T_2)$ satisfies

$$\int_0^\infty dT_1 \int_0^\infty dT_2\, e^{-s_1T_1-s_2T_2}F_{\rm OU}(T_1,T_2) = \int_{-\infty}^{\infty} dz\, \frac{e^{-z^2/2}}{\sqrt{8\pi}}\frac{D_{-1-s_1/2}(-z)}{D_{-s_1/2}(-z)}\frac{D_{-1-s_2/2}(-z)}{D_{-s_2/2}(-z)}. \tag{2.142}$$

Note that this result in Eqs. (2.141) and (2.142) was presented without derivation in [148]. Here we have presented an intuitive derivation of this result, based on a path-decomposition approach. Note that the asymptotic analysis of the distribution of $t_{\max}$ is, to the best of our knowledge, completely new.

As in the case $p = 1$, we can investigate the distribution of $t_{\max}$ in two opposite regimes: the regime of strong correlations ($T \ll \xi$) and the regime of weak correlations ($T \gg \xi$), where the correlation timescale is $\xi \sim 1/\alpha$.

To investigate the regime $T \ll \xi$, we take the limit $s_1\,, s_2 \to \infty$ in Eq. (2.142) and then we invert the double Laplace transform, yielding

$$P(t_{\max}|T) \approx \frac{1}{\pi\sqrt{t_{\max}(T-t_{\max})}}. \tag{2.143}$$

Thus, on short timescales, the process is strongly correlated and behaves like a BM.

On the other hand, in the opposite limit $T \gg \xi$, a detailed analysis is required. As in the case $p = 1$, three distinct regimes are present, depending on the value of $t_{\max}$:

the bulk regime for $\xi \ll t_{\max} \ll (T-\xi)$ and the two edge regimes for $t_{\max} \ll \xi$ and $t_{\max} \gg (T-\xi)$.

The bulk regime corresponds in Laplace space to the limit $s_1\,, s_2 \to 0$ with s_1/s_2 fixed. Taking this limit in Eq. (2.142) and then inverting the Laplace transform, we get

$$P(t_{\max}|T) \approx \frac{1}{T}\,. \tag{2.144}$$

Thus, in the bulk of the distribution $P(t_{\max}|T)$, we obtain the flat distribution that one would expect from i.i.d. random variables, as we did for the case $p=1$. The analysis of the edge regimes instead shows the signatures of the correlations of the process. Let us consider the left-edge regime, where $t_{\max} \ll 1/\alpha$, (the right-edge regime can be obtained by symmetry). A detailed analysis of the expression in Eq. (2.142) shows that the edge regime corresponds in this case to the limit $s_1 \to \infty$ and $s_2 \to 0$. Taking this limit in Eq. (2.142) and inverting the Laplace transform yields [147]

$$P(t_{\max}|T) \approx \frac{1}{T} G\left[(\log T)\alpha t_{\max}\right]\,, \tag{2.145}$$

where $G(z)$, given in Eq. (2.132), is the same function as in the case $p=1$. Interestingly, the shape of the edge regime is the same in the cases $p=1$ and $p=2$. The only difference between the two cases is the width of the edge: in the case $p=1$, the width of the edge regime is $\mathcal{O}(1)$ at late times (see Eq. (2.134)), while the edge shrinks as $\mathcal{O}(1/\log(T))$ for $p=2$ (see Eq. (2.145)).

Thus, once appropriately rescaled near the edges, the distribution of $t_{\max}$ turns out to be the same for $p=1$ and $p=2$ at late times. This universality is rather unexpected and led us to investigate whether or not this result can be generalized to any $p>0$. Interestingly, in the next section, we will show that this universality indeed applies to any potential with $p>0$.

Universality at late times

To show that the distribution of $t_{\max}$ becomes universal at late times for any $p>0$, we will use a "blocking" argument similar to the one presented in Chapter 1. This argument requires that the autocorrelation function of the process decays in time as

$$\langle x(t)x(t')\rangle \sim e^{-|t-t'|/\xi}\,. \tag{2.146}$$

This turns out to be true for $p \geq 1$ [149]. For $0<p<1$, we have verified numerically that the autocorrelation function has a stretched-exponential decay in time. For instance, for $p=1$, we verified numerically that (see Fig. 2.14)

$$\langle x(t)x(t')\rangle \sim e^{-\sqrt{|t-t'|/\xi}}\,, \tag{2.147}$$

for some timescale $\xi>0$. Thus, also for $0<p<1$ one has a typical timescale over which correlations decay and one can still apply the blocking argument[4].

[4]Note that in Article 3 we had shown that the universality was valid for $p \geq 1$. Here we show that this result can also be extended to $0<p<1$.

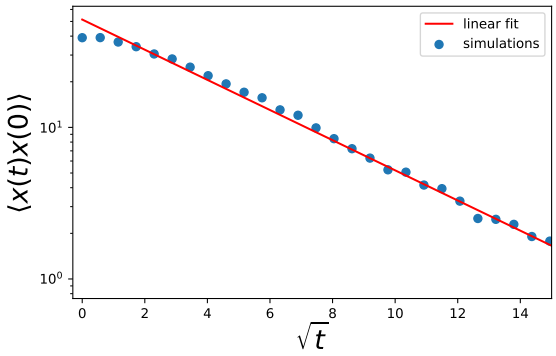

Figure 2.14: Semi-logarithmic plot of the correlation function $\langle x(t)x(0)\rangle$ as a function of $\sqrt{t}$ for Brownian motion with diffusion constant $D = 1$ in the potential $V(x) = \sqrt{|x|}$. The initial position $x(0)$ is drawn from the equilibrium state of the system. The continuous red line shows a stretched-exponential decay of the type $\langle x(t)x(0)\rangle \sim e^{-\sqrt{t/\xi}}$.

To proceed, we consider a long trajectory $x(t)$ of the system, with $0 < t < T$ and $T \gg \xi$. We divide the trajectory into $N = T/\xi$ blocks of size ξ. Since ξ is the correlation time, variables in different blocks are uncorrelated. Thus, denoting by m_i the maximum of the i-th block, we can approximate the local maxima $m_1, \dots, m_N$ to be i.i.d. variables. As a consequence, the probability that global maximum

$$M = \max_{1 \le i \le N} (m_i) \,, \tag{2.148}$$

is reached in a given block is just $1/N$, implying that the distribution of the time $t_{\max}$ of the maximum is flat, i.e.,

$$P(t_{\max}|T) \approx \frac{1}{T} \,. \tag{2.149}$$

As shown in the previous sections, this argument is only correct in the bulk of the distribution $P(t_{\max}|T)$, i.e., for $\xi \ll t_{\max} \ll (T - \xi)$. A more detailed analysis is required for the edges.

We first consider the left edge of the distribution $P(t_{\max}|T)$, for $t_{\max} \ll \xi$. This corresponds to the case in which the global maximum is reached in the first block. Thus, we condition on the event $t_{\max} < \xi$, which happens with probability $1/N = \xi/T$. If the maximum is reached within the first block, the position of the process $x(t)$ for $0 < t < \xi$ will be close to the value M of the maximum, since the process is strongly correlated within each box. Thus, we can linearize the Langevin equation (2.103) around the value M of the global maximum, yielding

$$\frac{dx(t)}{dt} = -V'(M) + \eta(t) \,. \tag{2.150}$$

Thus, the particle is subject to a constant drift $\mu = V'(x)$.

Moreover, using the results on the universality of the maximum of many i.i.d. variables, it is possible to show that at late times the maximum behaves deterministically

at leading order as (see for instance Article 3)

$$M \sim \left(\frac{D}{\alpha}\log(T)\right)^{1/p} . \tag{2.151}$$

In particular, this late-time behavior of M only depends on the large-x behavior of the potential $V(x)$, which we assume to be $V(x) \approx \alpha x^p$. Thus, for large T, the constant drift μ is given by

$$\mu = V'(M) \approx \alpha\, p \left(\frac{D}{\alpha}\log(T)\right)^{(p-1)/p} . \tag{2.152}$$

The distribution of the time at which a BM with drift reaches its global maximum has been exactly computed in [112] and reads

$$P(t_{\max}|\xi) = \frac{h_\mu(t_{\max})h_{-\mu}(\xi - t_{\max})}{\pi\sqrt{t_{\max}(\xi - t_{\max})}} , \tag{2.153}$$

where

$$h_\mu(\tau) = e^{-\mu^2\tau/(4D)} + \mu\sqrt{\frac{\pi\tau}{4D}}\,\mathrm{erfc}\left(-\mu\sqrt{\frac{\tau}{4D}}\right) . \tag{2.154}$$

Thus, for $0 \leq t_{\max} \leq \xi$ and $T \gg 1$, the distribution of $t_{\max}$ can be written as

$$P(t_{\max}|T) = P(t_{\max} < \xi|T)P(t_{\max}|\xi) \approx \frac{\xi}{T}\frac{h_\mu(t_{\max})h_{-\mu}(\xi - t_{\max})}{\pi\sqrt{t_{\max}(\xi - t_{\max})}} , \tag{2.155}$$

where the drift μ is given in Eq. (2.152). Note that, since we do not know the precise value of ξ, the result in Eq. (2.155) gives us the expression of $P(t_{\max}|t)$ up to a multiplicative constant. This prefactor can be computed by imposing that the edge expression in Eq. (2.155) matches with the bulk expression in Eq. (2.149), yielding

$$P(t_{\max}|T) \approx \frac{1}{T}G\left(\frac{t_{\max}}{\lambda(T)}\right) , \tag{2.156}$$

where $G(z)$ is given in Eq. (2.132) and the width of the edge region is

$$\lambda(T) = \frac{4D}{\alpha^2 p^2}\left(\frac{D}{\alpha}\log(T)\right)^{-2(p-1)/p} . \tag{2.157}$$

Remarkably, the shape of the edge region of $P(t_{\max}|T)$ is completely universal for any $p > 01$ and is described by the function $G(z)$. All the details about the potential $V(x)$ are contained in the width $\lambda(T)$ of the edge region. This quantity is increasing as $\log(T)^{2(1-p)/p}$ for $0 < p < 1$, it is of order one for $p = 1$, while it shrinks extremely slowly as $\log(T)^{-2(p-1)/p}$. Quite remarkably this simple argument reproduces the correct asymptotic behavior that we have obtained from the exact solutions in the cases $p = 1$ and $p = 2$ (see Eqs. (2.134) and (2.145)). We have verified the validity of the argument above by performing numerical simulations for different values of p (see Figs. 2.15 and 2.16). Note that the behavior in the right edge (i.e., for $t_{\max} \to T$) can be obtained from Eq. (2.156) using the symmetry $P(t_{\max}|T) = P(T - t_{\max}|T)$.

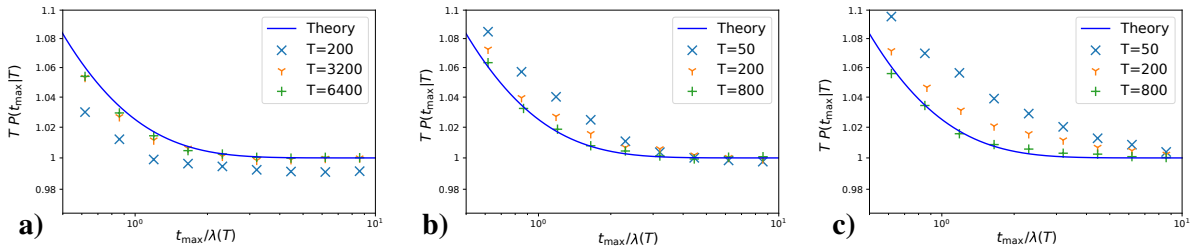

Figure 2.15: Left edge of the scaled distribution $TP(t_{\max}|T)$ as a function of the scaled time of the maximum $t_{\max}/\lambda(T)$ for $p = 1$ (panel **a**), $p = 2$ (panel **b**), and $p = 3$ (panel **c**). The continuous blue lines display the universal result in Eq. (2.156), which is the same for all values of p. The symbols are the results of numerical simulations of Brownian motion in the potential $V(x) = |x|^p$ with different total times T. We observe that for large T the numerical results approach the analytical prediction.

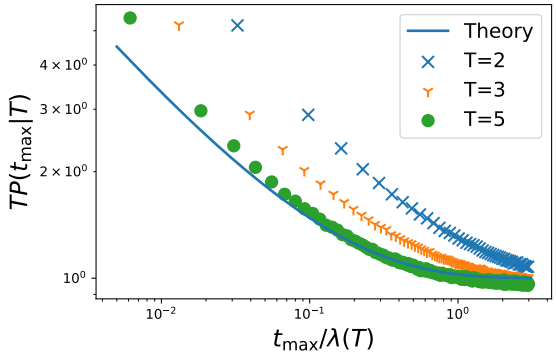

Figure 2.16: Left edge of the scaled distribution $TP(t_{\max}|T)$ as a function of the scaled time of the maximum $t_{\max}/\lambda(T)$ for $p = 1/2$. Note that here the width $\lambda(T) = (16D/\alpha^2)(D\log(T)/\alpha)^2$ is increasing with T. The continuous blue line corresponds to the universal result in Eq. (2.156). The symbols are the results of numerical simulations of Brownian motion in the potential $V(x) = 4x^2$ for $|x| < 1$ and $V(x) = 4\sqrt{|x|}$ for $|x| > 1$ with different total times T. We chose the quadratic part for small x to avoid the divergence in the first derivative $V'(x)$. We observe that already at $T = 5$ the numerical results are in excellent agreement the analytical prediction.

To summarize, we have shown that at late times and for any $p > 0$, the distribution of the time $t_{\max}$ of the maximum approaches the universal form

$$P(t_{\max}|T) \approx \begin{cases} \frac{1}{T} G\left(\frac{t_{\max}}{\lambda(T)}\right) & \text{for} \quad t_{\max} \lesssim \lambda(T) \\ \frac{1}{T} & \text{for} \quad \lambda(T) \ll t_{\max} \ll T - \lambda(T) \\ \frac{1}{T} G\left(\frac{T - t_{\max}}{\lambda(t)}\right) & \text{for} \quad t_{\max} \gtrsim T - \lambda(T)\,, \end{cases} \tag{2.158}$$

where the width $\lambda(T)$ of the edge regime is given in Eq. (2.157) and depends on the

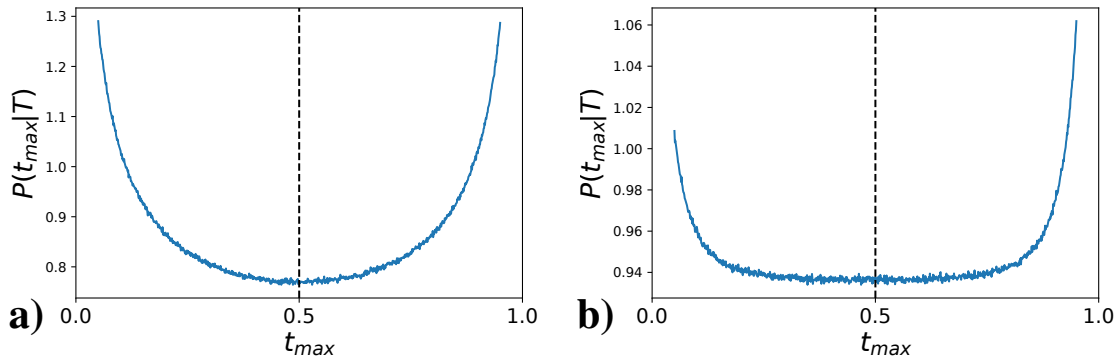

Figure 2.17: Empirical probability density function for the Ornstein-Uhlenbeck process with $\alpha = D = T = 1$ (panel **a**) and for resetting Brownian motion with $D = T = 1$ and $r = 10$ (panel **b**). In the equilibrium case in panel **a** the distribution is symmetric around the midpoint $t_{\max} = T/2$ (vertical dashed line), while this is not the case for the nonequilibrium process in panel **b**.

details of the model.

2.2.2 Out-of-equilibrium processes

We now turn our attention to the case of nonequilibrium stationary processes. We will consider two models for which the distribution $P(t_{\max}|T)$ of the time $t_{\max}$ of the maximum can be computed exactly. In this case, the distribution $P(t_{\max}|T)$ turns out to be asymmetric around the midpoint $t_{\max} = T/2$, i.e., $P(t_{\max}|T) \neq P(T - t_{\max}|T)$. This observation will lead us to formulate a simple criterion to detect nonequilibrium dynamics in stationary systems.

Resetting Brownian motion

In this section, we consider the case of a Brownian particle subject to stochastic resetting with constant rate r [36]. In a small time interval $[t, t + dt]$ the position of the particle evolves according to

$$x(t + dt) = \begin{cases} x(t) + \sqrt{2D}\eta(t)dt & \text{with probabiltiy } 1 - r dt\,, \\ \\ 0 & \text{with probabiltiy } r dt\,, \end{cases} \tag{2.159}$$

where $D > 0$ is the diffusion constant and $\eta(t)$ is Gaussian white noise. For a typical realization of the process, see Fig. 2.18. At late times, one can show that the system reaches the stationary state [36]

$$P_{\rm st}(x_0) = \frac{1}{2}\sqrt{\frac{r}{D}}\exp\left(-\sqrt{\frac{r}{D}}|x_0|\right)\,. \tag{2.160}$$

Interestingly, the resetting dynamics are completely irreversible, inducing a net probability current toward the origin and driving the system out of equilibrium.

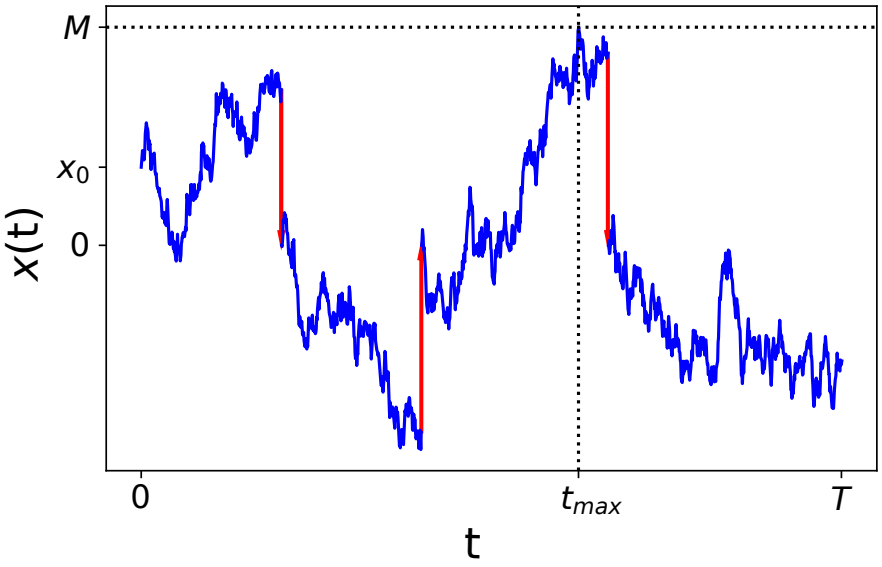

Figure 2.18: Typical trajectory $x(t)$ of a Brownian motion with stochastic resetting in the interval $[0, T]$. The red segments indicate the resetting events. The particle starts from position x_0, drawn from the steady state (2.160) and reaches the global maximum M at time $t_{\max}$.

We assume that at time $t = 0$ the initial position of the system x_0 is drawn from the nonequilibrium steady state in Eq. (2.160). Then, we let the system evolve according to the evolution equation (2.159) and we investigate the time $t_{\max}$ of the maximum. Note that the distribution of $t_{\max}$ in the case of BM with stochastic resetting was previously studied in [116] in the case where the starting position is fixed $x_0 = 0$. Here instead we let the starting position x_0 fluctuate according to the stationary distribution $P_{\rm st}(x_0)$.

To compute the distribution of $t_{\max}$, we will employ again a path-decomposition technique as in the previous sections. The relation in Eq. (2.105) can be written as

$$\begin{aligned} P(t_{\max}|T) &= \lim_{\epsilon\to 0}\left[\mathcal{N}(\epsilon)\int_{-\infty}^{\infty} dx_0 P_{\rm st}(x_0)\right. \\ &\times \left.\int_{x_0}^{\infty} dM\; G_r^M(M-\epsilon, t_{\max}|x_0) Q_r^M(M-\epsilon, T-t_{\max})\right], \end{aligned} \tag{2.161}$$

where the subscript "r" highlights the dependence of the constrained propagator and of the survival probability on the resetting rate.

We first compute the survival probability $Q_r^M(x,t)$, defined as the probability that the process remains below position M up to time t, having started from position $x < M$. In the case $M < 0$, no resetting event can occur (otherwise the system would jump above position M) and hence

$$Q_r^M(x,t) = e^{-rt} Q_0^M(x,t)\,, \tag{2.162}$$

where the term e^{-rt} is the probability that no resetting event occurs up to time t. The term $Q_0^M(x,t)$ is the survival probability of BM without resetting, which is given by [32]

$$Q_0^M(x,t) = \mathrm{erf}\left(\frac{M-x}{\sqrt{4Dt}}\right). \tag{2.163}$$

Taking a Laplace transform with respect to t and setting $x = M - \epsilon$, we find that

$$\tilde{Q}_r^M(M-\epsilon, s) \approx \frac{\epsilon}{\sqrt{D(s+r)}}\,, \tag{2.164}$$

at leading order in ϵ.

On the other hand, in the case $M > 0$, the survival probability satisfies the renewal equation [36]

$$Q_r^M(x,t) = e^{-rt} Q_0^M(x,t) + r\int_0^t d\tau\; e^{-r\tau} Q_0^M(x,\tau) Q_r^M(0, t-\tau)\,. \tag{2.165}$$

The first term on the right-hand side corresponds to the case where no resetting occurs, while the second term corresponds to the case where the first resetting event occurs at time τ. By taking a Laplace transform of both sides of Eq. (2.165) one can find the following exact expression for the survival probability

$$\tilde{Q}_r^M(x,s) = \frac{1 - e^{-\sqrt{(s+r)/D}(M-x)}}{s + re^{-\sqrt{(s+r)/D}M}},. \tag{2.166}$$

Setting $x = M - \epsilon$ and expanding to leading order in ϵ, we obtain

$$\tilde{Q}_r^M(M-\epsilon, s) \approx \frac{\epsilon}{\sqrt{D}} \frac{\sqrt{s+r}}{s + re^{-\sqrt{(s+r)/D}M}}\,. \tag{2.167}$$

We next focus on the constrained propagator $G_r^M(x,t|x_0)$, defined as the probability that the process goes from position x_0 to position x in time t, while always remaining below position M. In the case $M < 0$, as before, no resetting can occur and hence

$$G_r^M(x,t|x_0,0) = e^{-rt} G_0^M(x,t|x_0,0)\,, \tag{2.168}$$

The propagator without resetting $G_0^M(x,t|x_0,0)$ can be computed using the method of images [32] and reads

$$G_0^M(x,t|x_0,0) = \frac{1}{\sqrt{4\pi D t}} \left(e^{-(x-x_0)^2/(4Dt)} - e^{-(2M-x+x_0)^2/(4Dt)}\right)\,. \tag{2.169}$$

Taking a Laplace transform with respect to t and setting $x = M - \epsilon$, we get

$$\tilde{G}_0^M(M-\epsilon, s|x_0,0) \simeq \frac{\epsilon}{D} e^{-\sqrt{s/D}(M-x_0)}\,, \tag{2.170}$$

where we have expanded to leading order in ϵ. Considering the Laplace transform of Eq. (2.168) with respect to t and using Eq. (2.170), we get

$$\tilde{G}_r^M(M-\epsilon, s|x_0,0) \simeq \frac{\epsilon}{D} e^{-\sqrt{(s+r)/D}(M-x_0)}\,. \tag{2.171}$$

On the other hand, in the case $M > 0$, the constrained propagator satisfies the renewal equation [36]

$$G_r^M(x,t|x_0,0) = e^{-rt} G_0^M(x,t|x_0,0) + r\int_0^t d\tau\, e^{-r\tau} Q_r^M(x_0, t-\tau) G_0^M(x,\tau|0,0)\,. \tag{2.172}$$

The first term on the right-hand side corresponds to the case where no resetting occurs while the second term corresponds to the case where the last resetting event occurs at time $t-\tau$. The factor $Q_r^M(x_0, t-\tau)$ is the probability that the particle remains below position M up to time $t-\tau$. This relation can be solved by taking a Laplace transform with respect to t and by using the expressions in Eqs. (2.166) and (2.169), yielding

$$\tilde{G}_r(M-\epsilon, s|x_0, 0) \simeq \frac{\epsilon}{D} \frac{r + s\, e^{\sqrt{(s+r)/D} x_0}}{r + s\, e^{\sqrt{(s+r)/D} M}}\,, \tag{2.173}$$

where we have set $x = M - \epsilon$ and we have expanded to leading order in ϵ.

We now have all the ingredients required to compute the distribution of $t_{\max}$ using Eq. (2.161). Indeed, substituting the expressions in Eqs. (2.160), (2.164), (2.167), (2.171), and (2.173) into Eq. (2.161), we find that the distribution of $t_{\max}$ can be written in the scaling form

$$\boxed{P(t_{\max}|T) = r F_R(r t_{\max}, r(T - t_{\max}))\,,} \tag{2.174}$$

where

$$\int_0^\infty dT_1\; e^{-s_1 T_1} \int_0^\infty dT_2\; e^{-s_2 T_2}\; F_R(T_1, T_2) = \frac{1}{2} \frac{1}{(1+\sqrt{1+s_1})\sqrt{1+s_2}} \tag{2.175}$$
$$+ \quad \frac{1}{2} \frac{\sqrt{1+s_2}}{\sqrt{1+s_1} - 1} \int_0^\infty dz\; e^{-(1+\sqrt{1+s_1})z} \frac{e^{z\sqrt{1+s_1}} s_1 - \sqrt{1+s_1} + 1}{\left(s_1 + e^{-z\sqrt{1+s_1}}\right)\left(s_2 + e^{-z\sqrt{1+s_2}}\right)}\,.$$

Interestingly, we observe that the distribution of $t_{\max}$ is not symmetric, i.e., that $P(t_{\max}|T) \neq P(T - t_{\max}|T)$. This is a signature of the nonequilibrium nature of the process. This asymmetry is confirmed by numerical simulations (see Fig. 2.17b).

Using Eqs. (2.174) and (2.176), we can investigate the asymptotic behaviors of $P(t_{\max}|T)$ in the limits of small and large T. In particular, for $T \ll 1/r$ we obtain

$$P(t_{\max}|T) \approx \frac{1}{\pi \sqrt{t_{\max}(T - t_{\max})}}\,, \tag{2.176}$$

which corresponds to the distribution of $t_{\max}$ for BM without resetting (see Eq. (1.53)). Indeed, since $1/r$ is the typical time between two subsequent resetting events, for $T \ll 1/r$ typically no resetting event has occurred yet.

On the other hand, for $T \gg 1/r$ we obtain [147]

$$P(t_{\max}|T) \approx \begin{cases} \frac{1}{T} G(r t_{\max}) & \text{for } t_{\max} \ll 1/r \\ \\ \frac{1}{T} & \text{for } 1/r \ll t_{\max} \ll (T - 1/r) \\ \\ \frac{1}{T}\left[2G(rT - r t_{\max}) - 1\right] & \text{for } t_{\max} \ll 1/r\,, \end{cases} \tag{2.177}$$

where $G(z)$ is given in Eq. (2.132). Interestingly, the late-time shape of the distribution $P(t_{\max}|T)$ for BM with resetting is qualitatively similar to the universal result obtained for equilibrium processes (see Eq. (2.158)). Indeed, in the bulk regime $(1/r) \ll t_{\max} \ll (T - 1/r)$, the distribution is flat (as one would expect from a collection of i.i.d. random

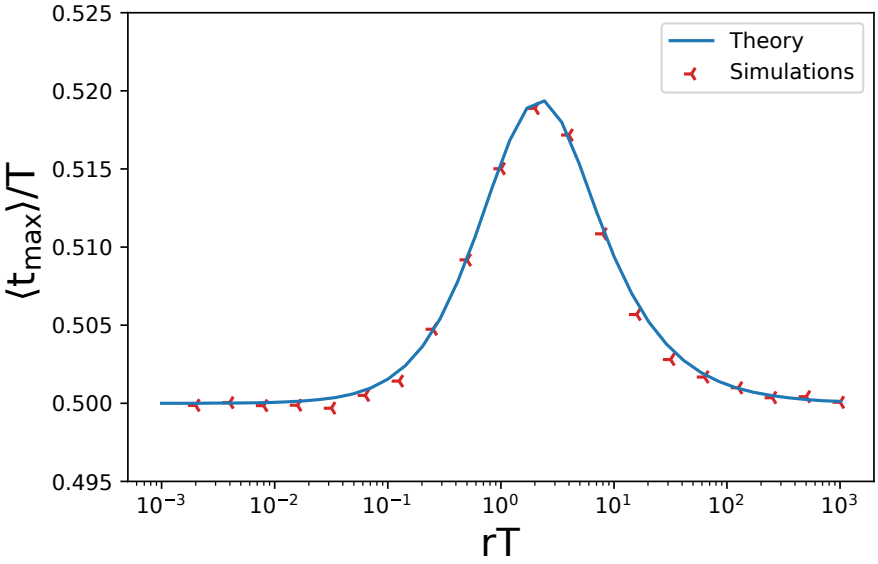

Figure 2.19: The scaled average $\langle t_{\max}\rangle/T$ as a function of the scaled total time rT for Brownian motion with stochastic resetting. The continuous blue line corresponds to the exact result in Eq. (2.178), while the symbols correspond to numerical simulations. Note that for any equilibrium process one expects $\langle t_{\max}\rangle = T/2$.

variables). On the other hand, in the edge regimes $t_{\max} \to 0$ and $t_{\max} \to T$, the distribution is described once again by the function $G(z)$. Note, however, that in this case the distribution is not symmetric.

Since in this case the distribution is asymmetric, i.e., $P(t_{\max}|T) \neq P(T - t_{\max}|T)$, the first moment of $t_{\max}$ is nontrivial. This quantity can be obtained from the expression in Eq. (2.176) by (i) differentiating with respect to s_1, (ii) setting $s_1 = s_2 = s$, and (III) inverting the Laplace transform (this last step turns out to be rather nontrivial in this case). This yields [147]

$$\langle t_{\max}(T)\rangle = r f(rT)\,. \tag{2.178}$$

where the scaling function $f(t)$ is given by

$$\begin{aligned} f(t) &= \frac{1}{96}\Big[-4t(2t^2+3t-18) + \frac{2}{\sqrt{\pi}}\sqrt{t}(3+16t+4t^2)e^{-t} \\ &+ (-3-30t+36t^2+8t^3)\,\mathrm{erf}(\sqrt{t})\Big] + \frac{1}{2}\left[e^{-t} - \frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2},t\right)\right] + \sum_{k=1}^{\infty} g_k(t)\,, \end{aligned} \tag{2.179}$$

where $\Gamma(a,t) = \int_t^\infty x^{a-1}e^{-x}$ is the upper incomplete Gamma function. The function $g_k(t)$ reads

$$g_k(t) = (-1)^k \frac{1}{2}(k+1)(k+2)\int_0^t d\tau\, h_k(t-\tau)\tau^{k+1}\left(\frac{1}{(k+1)!} + \frac{\tau}{(k+2)!}\right)\,, \tag{2.180}$$

where

$$h_k(t) = \frac{1}{k^2}\left\{-e^{-t+t/k^2}k(1-k)^2 + e^{-t}\frac{k\left[k(1+k)^3-2k^3t\right]}{\sqrt{\pi t}(1+k)^3}\right\} \tag{2.181}$$
$$+ \frac{1}{k^2}\left[\mathrm{erf}\left(\frac{\sqrt{t}}{k}\right)e^{-t+t/k^2}(1-k)^2\right]\frac{1}{(1+k)^4}e^{-t+t/(1+k)^2}\left[(1+k)^2(k^2-2)+2kt\right]$$
$$\times \left[1-\mathrm{erf}\left(\frac{\sqrt{t}}{(k+1)}\right)\right].$$

Evaluating the integral over τ and the sum over k above numerically, we obtain the analytical curve in Fig. 2.19, which is in good agreement with numerical simulations.

Run-and-tumble particle in a confining potential

In this section, we investigate the time $t_{\max}$ of the maximum for a run-and-tumble particle (RTP) moving in a confining potential $V(x) = \mu|x|$. The position x of the particle evolves according to

$$\frac{dx}{dt} = f(x) + v_0\,\sigma(t)\,, \tag{2.182}$$

where $v_0 > 0$ is the velocity of the particle, $f(x) = -V'(x)$ is the external force. The term $\sigma(t) = \pm 1$ is telegraphic noise, flipping its sign at a constant rate γ. Thus, for timescales of the order of $1/\gamma$ the motion of the particle is ballistic, breaking the detailed balance condition. Thus, this system is out of equilibrium. To understand this it is sufficient to observe that in a small time interval dt the system can go from the state $(x,+)$, i.e., from position x with positive direction, to the state $(x+(v_0+f(x))dt,+)$. However, the reversed transition is not possible, inducing a probability current in the phase space.

We will consider the case $v_0 > \mu$, for which the stationary distribution reads [54]

$$P_{\rm st}(x_0) = \frac{\gamma\ \mu}{v_0^2-\mu^2}\exp\left(-\frac{2\gamma\mu}{v_0^2-\mu^2}|x_0|\right). \tag{2.183}$$

It is useful to consider also the joint distribution $P^{\sigma}_{\rm st}(x_0)$ of the position x_0 and of the direction $\sigma = \pm$ of the particle in the steady state, which is given by [54]

$$P^{\pm}_{\rm st}(x_0) = \frac{1}{2}\left(1 \pm \frac{\mu}{v_0}\,\mathrm{sign}(x)\right)\frac{\gamma\ \mu}{v_0^2-\mu^2}\exp\left(-\frac{2\gamma\mu}{v_0^2-\mu^2}|x_0|\right). \tag{2.184}$$

We consider a single RTP starting from position x_0 and direction σ, jointly drawn from the stationary state in Eq. (2.184). We want to compute the distribution $P(t_{\max}|T)$ of the time $t_{\max}$ at which the maximum of the position is reached. To do this, we will use a path-decomposition strategy similar to the one used in previous sections. Note that as a consequence of the persistent motion of the particle, the events $t_{\max} = 0$ and $t_{\max} = T$ happen with a finite probability and have to be considered separately.

We start by considering the case $0 < t_{\max} < T$. In this case, the time of the maximum necessarily coincides with a tumbling event (the particle necessarily arrives at the global maximum with positive velocity and then it immediately tumbles). We

thus split the time interval $[0,T]$ into the three subintervals $[0,t_{\max}]$ (I), $[t_{\max},t_{\max}+dt]$ (II), where dt is assumed to be small, and $[t_{\max}+dt,T]$ (III). In the interval (I), the particle starts from position x_0 with direction σ, it stays below the maximal value M and reaches M for the first time at time $t_{\max}$. Thus, the probability weight of the first interval is $G_M^+(M,t_{\max}|x_0,\sigma)$, where the constrained propagator $G_M^{\pm}(x,t|x_0,\sigma)$ is defined as the probability that the particle reaches position x with direction $\pm$ at time t while always remaining below position M, having started from position x_0 with direction σ. In the small time interval (II), the particle has to tumble, i.e., to change its direction from positive to negative. This will happen with probability γdt. In the last interval $[t_{\max}+dt,T]$ the particle starts from position M with negative velocity and remains below position M up to time T, with probability weight $Q_M^-(M,T-t_{\max})$. Here, the survival probability $Q_M^{\sigma}(x,t)$ is defined as the probability that the particle remains below position M up to time t, starting from position x with direction σ. Since the joint process (x,σ) is Markov, the distribution of $t_{\max}$ can be written as the product of the three probability weights. Thus, integrating over both the initial position x_0 and the maximal value $M>x_0$ and summing over the initial direction σ, we get

$$P(t_{\max}|T) = A\,\gamma \sum_{\sigma=\pm}\int_{-\infty}^{\infty} dx_0\; P_{\rm st}^{\sigma}(x_0) \int_{x_0}^{\infty} dM\; G_M^+(M,t_{\max}|x_0,\sigma)\; Q_M^-(M,T-t_{\max})\,, \tag{2.185}$$

where A is a normalization constant. This expression is valid for $0<t_{\max}<T$. Note that no cutoff is required in this case.

Let us now focus on the events $t_{\max}=0$ and $t_{\max}=T$. In particular, $t_{\max}=0$ corresponds to trajectories where the particle starts from position x_0 with negative velocity and remains below its starting position x_0 up to time T. Thus, since the starting position and direction are drawn from the stationary distribution $P_{\rm st}^{\sigma}(x_0)$, we obtain

$$\text{Prob.}(t_{\max}=0|T) = \int_{-\infty}^{\infty} dx_0\; P_{\rm st}^-(x_0) Q_{x_0}^-(x_0,T)\,. \tag{2.186}$$

With a similar argument, we obtain

$$\text{Prob.}(t_{\max}=T|T) = \sum_{\sigma=\pm}\int_{-\infty}^{\infty} dx_0\; P_{\rm st}^{\sigma}(x_0)\; G_M^+(M,T|x_0,\sigma)\,. \tag{2.187}$$

Thus, using the Eqs. (2.185), (2.186), and (2.187), we find that for $0\le t_{\max}\le T$

$$\begin{aligned} P(t_{\max}|T) &= \delta(t_{\max})\int_{-\infty}^{\infty} dx_0\; P_{\rm st}^-(x_0)Q_{x_0}^-(x_0,T) \\ &+ A\,\gamma \sum_{\sigma=\pm}\int_{-\infty}^{\infty} dx_0\; P_{\rm st}^{\sigma}(x_0) \int_{x_0}^{\infty} dM\; G_M^+(M,t_{\max}|x_0,\sigma)\; Q_M^-(M,T-t_{\max}) \\ &+ \delta(t_{\max}-T)\sum_{\sigma=\pm}\int_{-\infty}^{\infty} dx_0\; P_{\rm st}^{\sigma}(x_0)\; G_M^+(M,T|x_0,\sigma)\,, \end{aligned} \tag{2.188}$$

where the constant A can be computed from the normalization condition

$$\int_0^T dt_{\max}\; P(t_{\max}|T) = 1\,. \tag{2.189}$$

We now need to compute the constrained propagator $G_M^{\pm}(x,t|x_0,\sigma)$ and the survival probability $Q_M^{\pm}(x,t)$. These quantities can be obtained by solving the Fokker-Planck equation associated with the process.

The propagator $G_M^{\pm}(x,t|x_0,\sigma)$ satisfies [54]

$$\begin{cases} \partial_t\, G_M^+(x,t|x_0,\sigma) &= -\partial_x\left[(-\mu\,\mathrm{sign}(x)+v_0)\,G_M^+(x,t|x_0,\sigma)\right]-\gamma\; G_M^+(x,t|x_0,\sigma) \\ &\quad +\gamma\; G_M^-(x,t|x_0,\sigma)\,, \\ \partial_t\, G_M^-(x,t|x_0,\sigma) &= -\partial_x\left[(-\mu\,\mathrm{sign}(x)-v_0)\,G_M^-(x,t|x_0,\sigma)\right]-\gamma\; G_M^-(x,t|x_0,\sigma) \\ &\quad +\gamma\; G_M^+(x,t|x_0,\sigma)\,, \end{cases} \tag{2.190}$$

with initial condition

$$G_M^{\pm}(x,t=0|x_0,\sigma)=\delta\,(x-x_0)\,\delta_{\sigma,\pm}\,, \tag{2.191}$$

and boundary conditions

$$\begin{cases} G_M^{\pm}(-\infty,t|x_0,\sigma)=0 \\ \\ G_M^{-}(M,t|x_0,\sigma)=0\,. \end{cases} \tag{2.192}$$

The boundary condition on the second line of Eq. (2.192) can be derived with the following argument. If a particle arrives at position M with a negative velocity, it must have visited the region $x>M$. However, we are constraining the particle to remain below position M and thus $G_M^-(M,t|x_0,\sigma)$ has to vanish. Note that $G_M^+(M,t|x_0,\sigma)$ remains instead unspecified.

This differential equation can be solved by taking a Laplace transform with respect to t and defining

$$\tilde{G}_M^{\pm}(x,s|x_0,\sigma)=\int_0^{\infty} dt\; e^{-st}G_M^{\pm}(x,t|x_0,\sigma)\,. \tag{2.193}$$

In particular, for $x=M$, we obtain [147]

$$\tilde{G}_M^+(M,s|x_0,+) \tag{2.194}$$

$$= \begin{cases} \dfrac{1}{v_0+\mu}e^{-(k-(s+\gamma)\mu)(M-x_0)/(v_0^2-\mu^2)} & \text{for } x_0<0\,,M<0\,, \\ \\ k\dfrac{e^{-(\mu(s+\gamma)+k)M/(v_0^2-\mu^2)}\;e^{(-\mu(s+\gamma)+k)x_0/(v_0^2-\mu^2)}}{v_0(k-\mu(\gamma+s))+\mu(v_0(\gamma+s)-k)e^{-2kM/(v_0^2-\mu^2)}} & \text{for } x_0<0\,,M>0\,, \\ \\ \dfrac{1}{v_0-\mu}\;\dfrac{(k-v_0(s+\gamma))\mu+e^{2kx_0/(v_0^2-\mu^2)}v_0((s+\gamma)\mu-k)}{(k-v_0(s+\gamma))\mu+e^{2kM/(v_0^2-\mu^2)}v_0((s+\gamma)\mu-k)} & \\ \times e^{(k-\mu(s+\gamma))(M-x_0)/(v_0^2-\mu^2)} & \text{for } x_0>0\,,M>0\,, \end{cases}$$

where

$$k=\sqrt{s^2v_0^2+2sv_0^2\gamma+\gamma^2\mu^2}\,. \tag{2.195}$$

Similarly, in the case $\sigma = -$, we obtain

$$\tilde{G}^{+}_{M}(M,s|x_0,-) \tag{2.196}$$
$$= \begin{cases} \frac{v_0(\gamma+s)-k}{\gamma(v_0^2-\mu^2)} e^{-(k-(s+\gamma)\mu)(M-x_0)/(v_0^2-\mu^2)} & \text{for } x_0<0\,, M<0\,, \\ \\ \frac{k(v_0(\gamma+s)-k)}{\gamma(v_0-\mu)} \frac{e^{-(\mu(s+\gamma)+k)M/(v_0^2-\mu^2)}\ e^{(-\mu(s+\gamma)+k)x_0/(v_0^2-\mu^2)}}{v_0(k-\mu(\gamma+s))+\mu(v_0(\gamma+s)-k)e^{-2kM/(v_0^2-\mu^2)}} & \text{for } x_0<0\,, M>0\,, \\ \\ \frac{v_0(s+\gamma)-k}{\gamma(v_0^2-\mu^2)} \frac{(k-v_0(s+\gamma))\mu+e^{2kx_0/(v_0^2-\mu^2)}v_0((s+\gamma)\mu-k)}{(k-v_0(s+\gamma))\mu+e^{2kM/(v_0^2-\mu^2)}v_0((s+\gamma)\mu-k)} & \\ \times e^{(k-\mu(s+\gamma))(M-x_0)/(v_0^2-\mu^2)} & \text{for } x_0>0\,, M>0\,, \end{cases}$$

Similarly, the survival probability satisfies the following backward Fokker-Planck equations [54]

$$\begin{cases} \partial_t\ Q^{+}_{M}(x,t) = (-\mu\,\mathrm{sign}(x)+v_0)\,\partial_x Q^{+}_{M}(x,t) + \gamma\ Q^{+}_{M}(x,t) - \gamma\ Q^{-}_{M}(x,t)\,, \\ \\ \partial_t\ Q^{-}_{M}(x,t) = (-\mu\,\mathrm{sign}(x)-v_0)\,\partial_x Q^{-}_{M}(x,t) + \gamma\ Q^{-}_{M}(x,t) - \gamma\ Q^{+}_{M}(x,t) \end{cases} \tag{2.197}$$

with initial condition

$$Q^{\pm}_{M}(x,t=0) = 1\,, \tag{2.198}$$

for any $x < M$. The boundary conditions in this case are given by

$$\begin{cases} Q^{\pm}_{M}(-\infty,t) = 1\,, \\ Q^{+}_{M}(M,t) = 0\,. \end{cases} \tag{2.199}$$

The second boundary condition encodes the fact that if the particle starts at M with a positive velocity, it will immediately go above M. Note that in this case the boundary condition for $Q^{-}_{M}(M,t)$ remains unspecified.

This coupled equations can be solved by considering the Laplace transform with respect to t and defining

$$\tilde{Q}^{\pm}_{M}(x,s) = \int_0^\infty dt\ e^{-st} Q^{\pm}_{M}(x,t)\,. \tag{2.200}$$

In particular, for $x = M$, we get

$$\tilde{Q}^{-}_{M}(M,s) = \begin{cases} \frac{1}{s}\frac{k+v_0 s-\gamma\mu}{k+v_0(s+\gamma)} & \text{for } M<0\,, \\ \\ \frac{1}{s}\frac{1}{k+v_0(s+\gamma)} & \\ \times\left[k+v_0 s+\mu\gamma-\frac{2k\gamma\mu(v_0-\mu)}{(v_0(s+\gamma)-k)\mu+v_0(k-(s+\gamma)\mu)e^{2kM/(v_0^2-\mu^2)}}\right] & \text{for } M>0\,. \end{cases} \tag{2.201}$$

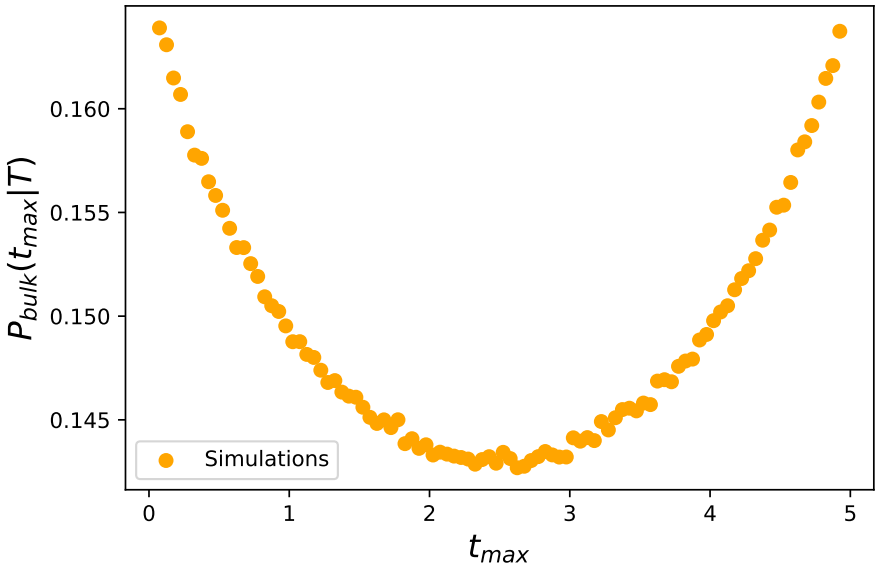

Figure 2.20: Probability density function $P_{\rm bulk}(t_{\rm max}|T)$ as a function of the time $t_{\rm max}$ of the maximum for a run-and-tumble particle in a confining potential $V(x) = |x|$, for $0 < t_{\rm max} < T$. Note that since the events "$t_{\rm max} = 0$" and "$t_{\rm max} = T$" occur with finite probability, the distribution is not normalized to unity for $0 < t_{\rm max} < T$. The curve is obtained by numerical simulations with $T = 5$ and $v_0 = 2$. The distribution $P_{\rm bulk}(t_{\rm max}|T)$ appears to be symmetric around the midpoint $t_{\rm max} = T/2$. We find numerically that $P_0(T) = {\rm Prob.}(t_{\rm max} = 0) \approx 0.087$ and $P_1(T) = {\rm Prob.}(t_{\rm max} = T) \approx 0.165$.

We now have all the ingredients to use the formula for $P(t_{\rm max}|T)$ in Eq. (2.188). To proceed, we first consider a double Laplace transform of Eq. (2.188) with respect to $t_1 = t_{\rm max}$ and $t_2 = T - t_{\rm max}$, yielding

$$\begin{aligned} &\int_0^\infty dt_1 \int_0^\infty dt_2 \, e^{-s_1 t_1 - s_2 t_2} P(t_{\rm max} = t_1 | T = t_1 + t_2) = \int_{-\infty}^\infty dx_0 \; P^-_{\rm st}(x_0) \tilde{Q}^-_{x_0}(x_0, s_2) \\ &+ \; A \, \gamma \sum_{\sigma = \pm} \int_{-\infty}^\infty dx_0 \; P^\sigma_{\rm st}(x_0) \int_{x_0}^\infty dM \; \tilde{G}^+_M(M, s_1 | x_0, \sigma) \; \tilde{Q}^-_M(M, s_2) \\ &+ \; \sum_{\sigma = \pm} \int_{-\infty}^\infty dx_0 \; P^\sigma_{\rm st}(x_0) \; \tilde{G}^+_M(M, s_1 | x_0, \sigma) \, . \end{aligned} \tag{2.202}$$

First of all, using the expressions above for $P^\sigma_{\rm st}$, $\tilde{G}^+_M$, and $\tilde{Q}^-_M$ and imposing normalization, we find that $A = 1$. It is useful to rewrite $P(t_{\rm max}|T)$ as

$$P(t_{\rm max}|T) = P_0(T) \delta(t_{\rm max}) + P_{\rm bulk}(t_{\rm max}|T) + P_1(T) \delta(t_{\rm max} - T) \, , \tag{2.203}$$

where

$$\begin{aligned} &\int_0^\infty dt_1 \int_0^\infty dt_2 \, e^{-s_1 t_1 - s_2 t_2} P_{\rm bulk}(t_{\rm max} = t_1 | T = t_1 + t_2) \\ &= \; \gamma \sum_{\sigma = \pm} \int_{-\infty}^\infty dx_0 \; P^\sigma_{\rm st}(x_0) \int_{x_0}^\infty dM \; \tilde{G}^+_M(M, s_1 | x_0, \sigma) \; \tilde{Q}^-_M(M, s_2) \, , \end{aligned} \tag{2.204}$$

$$\int_0^\infty dT\ e^{-sT} P_0(T) = \int_{-\infty}^\infty dx_0\ P_{\rm st}^-(x_0)\tilde{Q}_{x_0}^-(x_0,s)\,, \tag{2.205}$$

and

$$\int_0^\infty dT\ e^{-sT} P_1(T) = \sum_{\sigma=\pm}\int_{-\infty}^\infty dx_0\ P_{\rm st}^\sigma(x_0)\ \tilde{G}_M^+(M,s|x_0,\sigma)\,. \tag{2.206}$$

Interestingly, for $0 < t_{\rm max} < T$ one can check that $P_{\rm bulk}(t_{\rm max}|T) = P_{\rm bulk}(T - t_{\rm max}|T)$, i.e., that the central part of the distribution of $t_{\rm max}$ is symmetric around the midpoint $t_{\rm max} = T/2$. However, it is easy to show that the amplitudes $P_0(T)$ and $P_1(T)$ of the delta functions in $t_{\rm max} = 0$ and $t_{\rm max} = T$ are in general different. Thus, the full distribution $P(t_{\rm max}|T)$, for $0 \le t_{\rm max} \le T$ is not symmetric around $t_{\rm max} = T/2$. This is a consequence of the nonequilibrium nature of the process.

Criterion to detect nonequilibrium fluctuations in steady states

As we have shown in the previous sections, at equilibrium the PDF $P(t_{\rm max}|T)$ is usually symmetric around its midpoint $t_{\rm max} = T/2$, i.e., $P(t_{\rm max}|T) = P(T - t_{\rm max}|T)$. On the other hand, we have also observed that in general this is not the case for nonequilibrium processes. In this section, we will first show that the distribution of $t_{\rm max}$ is symmetric for any equilibrium process, as a consequence of the time-reversal symmetry. This will lead us to formulate a simple criterion to detect nonequilibrium fluctuations in stationary time series.

To keep the notation simple, we consider a discrete-time process x_i, with $1 \le i \le T$ (here i and T are integer numbers). It is immediate to generalize the derivation to processes in continuous time. Denoting by $P(\{x_i\})$ the probability of observing the trajectory $\{x_i\} = \{x_1, \dots, x_T\}$, the distribution of the time $t_{\rm max}$ of the maximum can be written as

$$P(t_{\rm max}|T) = \int_{-\infty}^\infty dx_1 \dots \int_{-\infty}^\infty dx_T\ \Theta_{t_{\rm max}}(\{x_i\}) P(\{x_i\})\ , \tag{2.207}$$

where

$$\Theta_k(\{x_i\}) = \prod_{i\neq k} \theta(x_k - x_i) \tag{2.208}$$

and $\theta(z)$ is the Heaviside step function, i.e., $\theta(z) = 1$ for $z > 0$ and $\theta(z) = 0$ otherwise. In other words, $\Theta_k(\{x_i\})$ is one if the maximum of the trajectory $\{x_i\}$ is reached at step k and zero otherwise. Thus, in Eq. (2.207), we sum over all trajectories for which the time of the maximum is $t_{\rm max}$. We next define $\{\bar{x}_i\} = \{x_{T-i}\}$ the time-reversed version of $\{x_i\}$. For an equilibrium process, it is immediate to show that, as a consequence of the detailed balance condition, $P(\{x_i\}) = P(\{\bar{x}_i\})$ (in other words, the process is symmetric under time reversal). Using this result in Eq. (2.207) and changing variables $x_i \to \bar{x}_i = x_{T-i}$, we obtain

$$P(t_{\rm max}|T) = \int_{-\infty}^\infty d\bar{x}_1 \dots \int_{-\infty}^\infty d\bar{x}_T\ \Theta_{t_{\rm max}}(\{\bar{x}_{T-k}\}) P(\{\bar{x}_k\})\ . \tag{2.209}$$

Using the relation $\Theta_{t_{\rm max}}(\{\bar{x}_{T-i}\}) = \Theta_{T-t_{\rm max}}(\{\bar{x}_i\})$, we find

$$P(t_{\rm max}|T) = \int_{-\infty}^\infty d\bar{x}_1 \dots \int_{-\infty}^\infty d\bar{x}_T\ \Theta_{T-t_{\rm max}}(\{\bar{x}_i\}) P(\{\bar{x}_i\})\ . \tag{2.210}$$

The expression on the right-hand side is precisely $P(T-t_{\max}|T)$ (compare with Eq. (2.207)), yielding $P(t_{\max}|T) = P(T-t_{\max}|T)$, which is thus a necessary, but not a sufficient, condition for a stationary process to be at equilibrium. As a consequence, if for a stationary process the distribution of $t_{\max}$ is not symmetric, then the process is necessarily out of equilibrium.

Interestingly, this observation can be used to detect nonequilibrium fluctuations in stationary time series. Imagine that you have access to a long time series $x(t)$ but you do not know the precise details of the dynamics. This trajectory could represent the position of a molecular motor or a diffusive particle. This situation is particularly relevant in light of the recent developments in single-particle tracking [150]. Then, a very natural question is whether or not the underlying system is at equilibrium. This kind of question is quite common in the context of living systems, where nonequilibrium dynamics signal active consumption of energy.

In recent years, many different techniques have been developed to both detect nonequilibrium fluctuations and estimate the associated dissipation (for a recent review, see [151]). One of the most popular methods is based on the verification of the fluctuation-dissipation theorem, which relates correlation and response functions at equilibrium [152–155]. In the presence of a violation of this theorem, one can immediately conclude that the process is out of equilibrium. For instance, this technique has been applied to show the nonequilibrium nature of the motion of red blood cells.

Note however that to verify the validity of the fluctuation-dissipation theorem one has to measure the response function. This requires perturbing the system, which is usually quite nontrivial in practice. More recently, several other methods have been developed based for instance on the detection of violations of the detailed balance condition [156–159] or on the analysis of waiting-time distributions [160, 161]. Some of these techniques also provide a quantitative measure of the nonequilibrium fluctuations of the system, usually as a bound on the entropy production [162–166].

Exploiting the symmetry of $P(t_{\max}|T)$ at equilibrium, we propose a new method to decide whether the stationary process $x(t)$ is nonequilibrium. First, (i) divide the time series into N blocks of duration T, (ii) for each block, compute the time $t^i_{\max}$ at which the maximum is reached (where the index i identifies the block, (iii) from these N values $t^1_{\max}, \dots t^N_{\max}$, build the empirical PDF $P(t_{\max}|T)$. If $P(t_{\max}|T)$ is not symmetric around $t_{\max} = T/2$, then the process is necessarily out of equilibrium. On the other hand, if $P(t_{\max}|T)$ is symmetric, our test is inconclusive (since there can be nonequilibrium processes for which $P(t_{\max}|T)$ is symmetric). Note that in the case of a multidimensional system, one can apply our criterion to any of its one-dimensional components. If $P(t_{\max}|T)$ is not symmetric in at least one of these components, then the full system is nonequilibrium.

Note that there are nonequilibrium processes for which the distribution of $t_{\max}$ is symmetric. A large class of examples is provided by Gaussian stationary processes. Indeed, it is possible to show that any one-dimensional Gaussian stationary process is invariant under time-reversal symmetry (this is a consequence of the fact that Gaussian processes only depend on the first two moments of the process distribution). As a particular example, we can consider a single one-dimensional active Ornstein-Uhlenbeck process (AOUP) in a harmonic potential $V(x) = \alpha x^2$ [51]. The position $x(t)$ of the

AOUP evolves according to the Langevin equation

$$\frac{dx(t)}{dt} = -\alpha x(t) + v(t) + \sqrt{2D}\xi(t)\,, \tag{2.211}$$

where $\xi(t)$ is a Gaussian white noise with zero mean and correlator $\langle \xi(t)\xi(t')\rangle = \delta(t-t')$ and $v(t)$ evolves as the Ornstein-Uhlenbeck process

$$\frac{dv(t)}{dt} = -\frac{v}{\tau_a} + \frac{\sqrt{2D_a}}{\tau_a}\zeta(t)\,, \tag{2.212}$$

where $D_a > 0$, $\tau_a > 0$, and $\zeta(t)$ is Gaussian white noise. We also assume that $\xi(t)$ and $\zeta(t)$ are uncorrelated. Note that for $\tau_a > 0$ the motion of the particle is persistent. Note also that even though $x(t)$ depends on the evolution of $v(t)$, there is no feedback from $v(t)$ to $x(t)$. One can show that this creates probability currents in the phase space (x, v) and thus the system is out of equilibrium. Interestingly, since the equations of motion for x and v are linear, the process is Gaussian. Thus, it follows that the distribution $P(t_{\max}|T)$ of the time $t_{\max}$ at which the position x is maximal is always symmetric around $t_{\max} = T/2$.

Article 1

Time between the maximum and the minimum of a stochastic process

F. Mori, S. N. Majumdar, and G. Schehr,
Phys. Rev. Lett. **123**, 200201 (2019).

☞ https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.123.200201
☞ https://arxiv.org/abs/1909.05594

Abstract

We present an exact solution for the probability density function $P(\tau = t_{\min} - t_{\max}|T)$ of the time-difference between the minimum and the maximum of a one-dimensional Brownian motion of duration T. We then generalise our results to a Brownian bridge, i.e. a periodic Brownian motion of period T. We demonstrate that these results can be directly applied to study the position-difference between the minimal and the maximal height of a fluctuating $(1+1)$-dimensional Kardar-Parisi-Zhang interface on a substrate of size L, in its stationary state. We show that the Brownian motion result is universal and, asymptotically, holds for any discrete-time random walk with a finite jump variance. We also compute this distribution numerically for Lévy flights and find that it differs from the Brownian motion result.

Article 2

Distribution of the Time Between Maximum and Minimum of Random Walks

F. Mori, S. N. Majumdar, and G. Schehr,
Phys. Rev. E **101**, 052111 (2020).

☞ https://journals.aps.org/pre/abstract/10.1103/PhysRevE.101.052111
☞ https://arxiv.org/abs/2002.12352

Abstract

We consider a one-dimensional Brownian motion of fixed duration T. Using a path-integral technique, we compute exactly the probability distribution of the difference $\tau = t_{\min} - t_{\max}$ between the time $t_{\min}$ of the global minimum and the time $t_{\max}$ of the global maximum. We extend this result to a Brownian bridge, i.e. a periodic Brownian motion of period T. In both cases, we compute analytically the first few moments of τ, as well as the covariance of $t_{\max}$ and $t_{\min}$, showing that these times are anti-correlated. We demonstrate that the distribution of τ for Brownian motion is valid for discrete-time random walks with n steps and with a finite jump variance, in the limit $n \to \infty$. In the case of Lévy flights, which have a divergent jump variance, we numerically verify that the distribution of τ differs from the Brownian case. For random walks with continuous and symmetric jumps we numerically verify that the probability of the event "$\tau = n$" is exactly $1/(2n)$ for any finite n, independently of the jump distribution. Our results can be also applied to describe the distance between the maximal and minimal height of $(1+1)$-dimensional stationary-state Kardar-Parisi-Zhang interfaces growing over a substrate of finite size L. Our findings are confirmed by numerical simulations. Some of these results have been announced in a recent Letter [Phys. Rev. Lett. **123**, 200201 (2019)].

Article 3

Distribution of the time of the maximum for stationary processes

F. Mori, S. N. Majumdar, and G. Schehr,
Europhys. Lett. **135**, 30003 (2021).

☞ https://iopscience.iop.org/article/10.1209/0295-5075/ac19ee
☞ https://arxiv.org/abs/2104.07346

Abstract

We consider a one-dimensional stationary stochastic process $x(\tau)$ of duration T. We study the probability density function (PDF) $P(t_{\rm m}|T)$ of the time $t_{\rm m}$ at which $x(\tau)$ reaches its global maximum. By using a path integral method, we compute $P(t_{\rm m}|T)$ for a number of equilibrium and nonequilibrium stationary processes, including the Ornstein-Uhlenbeck process, Brownian motion with stochastic resetting and a single confined run-and-tumble particle. For a large class of equilibrium stationary processes that correspond to diffusion in a confining potential, we show that the scaled distribution $P(t_{\rm m}|T)$, for large T, has a universal form (independent of the details of the potential). This universal distribution is uniform in the "bulk", i.e., for $0 \ll t_{\rm m} \ll T$ and has a nontrivial edge scaling behavior for $t_{\rm m} \to 0$ (and when $t_{\rm m} \to T$), that we compute exactly. Moreover, we show that for any equilibrium process the PDF $P(t_{\rm m}|T)$ is symmetric around $t_{\rm m} = T/2$, i.e., $P(t_{\rm m}|T) = P(T - t_{\rm m}|T)$. This symmetry provides a simple method to decide whether a given stationary time series $x(\tau)$ is at equilibrium or not.

Chapter 3

Universal properties of run-and-tumble particles

This and the next chapters are dedicated to the study of the statistical properties of run-and-tumble particles (RTPs). This model belongs to a larger class of nonequilibrium processes known as *active matter* [45, 48]. The common feature of active matter systems is that they are composed of individual elements, able to locally absorb energy and convert it into some form of work (for instance, intracellular cargo transport for kinesin motors). This differentiates active systems from their passive counterpart (e.g. Brownian motion), whose motion is governed by the thermal fluctuations of the environment. Active matter systems can be observed in several artificial and natural situations, e.g., vibrated granular matter [167], active gels [168], and bacteria [44, 45].

The RTP model was first known in the stochastic processes literature as persistent random walk [169–174]. In recent years, it has been applied to describe the persistent motion of a class of bacteria, including *E. coli* [44], which move along a fixed direction with constant velocity (they "run"), randomizing their direction (they "tumble") at random times. Interestingly, this simple model displays several nontrivial features, including clustering at the boundaries of a confining domain [49], non-Boltzmann distribution in the steady state [53–56], and jamming [97, 98]. Moreover, the RTP model has also been applied to investigate the dynamics of active particles in the presence of obstacles [175].

In this chapter, we consider a single RTP in d dimensions and with speed distribution $W(v)$. We assume that at the initial time the particle is at the origin and that it starts moving in a direction (chosen isotropically) with speed $v_1 > 0$, drawn from the distribution $W(v)$. After some random time τ_1, exponentially distributed with average $1/\gamma$, the particle tumbles, i.e., it chooses a new direction and a new speed v_2, independently drawn from $W(v)$, and starts moving in the new direction with the new velocity. This run-and-tumble motion continues up to the fixed time t. For a schematic trajectory of the process, see Fig. 3.1. The model is therefore parametrized by three quantities: (i) the tumbling rate γ, which sets the timescale of the RTP process, (ii) the dimensionality d of the system, and (iii) the speed distribution $W(v)$. We assume that the PDF $W(v)$ is normalized to unity for $v > 0$, i.e., $\int_0^\infty dv\ W(v) = 1$. Note that the standard RTP model with constant speed $v_0 > 0$ corresponds to the choice $W(v) = \delta(v - v_0)$. The

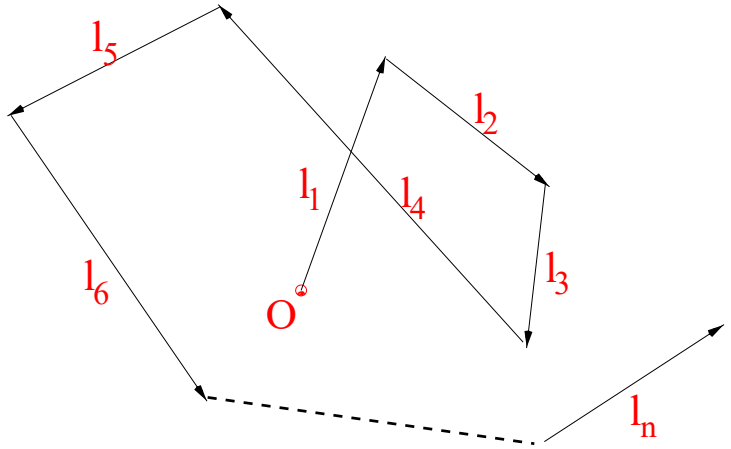

Figure 3.1: Typical trajectory of a run-and-tumble particle in $d = 2$. The particle is initially at the origin O and it starts moving in a random direction (chosen isotropically) for a distance $l_1 = v_1\tau_1$, where v_1 is a random speed, drawn from the probability distribution $W(v)$ and τ_1 is a random time, exponential with rate γ. After that, the particle tumbles, i.e., it chooses a new orientation and starts moving in that direction with velocity v_2, independently drawn form $W(v)$. After a time τ_2, exponential with rate γ, it tumbles again and so on. This process continues up to the total fixed time t. The number n of tumblings is also a random variable. Note that by definition $n \geq 1$, since the starting point is considered as a tumbling event.

goal of this chapter is to present our recent results on the extremal properties of this model. Interestingly, many of these features turn out to be completely universal, i.e., independent of d and $W(v)$.

3.1 Survival probability in arbitrary dimension

The first time t_f at which a process reaches a given state, also known as *first-passage time*, is a central quantity in several contexts, from chemistry, in which the first encounter time between two molecules plays a crucial role [74], to finance, in which agents use limit orders to buy an asset only below a given price. Due to the generality of first-passage problems, several new applications are often found. For a pedagogical introduction to first-passage processes see [32]. As explained in the introduction of this thesis, questions on first-passage times are intimately related to the statistics of the extremes of the process.

To characterize the statistical properties of the first-passage time t_f one is usually interested in computing its PDF, known as *first-passage probability*. Equivalently, one can also investigate its cumulative distribution $S(t) = \text{Prob.}(t_f > t)$, which is usually called *survival probability*. For a single RTP, the survival probability has been computed in one dimension and with constant velocity $v_0 > 0$, both with and without an external potential [57, 58, 172, 176, 177]. The main analytical simplification in the case $d = 1$ is that the particle velocity can only take two values $\pm v_0$. As a consequence, it is easy

to track the velocity degree of freedom and to write a Fokker-Plank equation for the survival probability. However, the survival probability is much harder to compute in $d > 1$ and only a few analytical results have been obtained in that case [178]. In this section, we investigate the first-passage properties of the RTP model for arbitrary d and $W(v)$.

First of all, let us define precisely the survival probability. We define $S(t)$ as the probability that the x-component of the particle remains positive up to time t. This indeed corresponds to the survival probability in the presence of an absorbing hyperplane passing through the origin and perpendicular to the x-axis. Note that in the case of a passive Brownian motion the different components of the process are completely independent. Thus, computing the survival probability of one component simply reduces to a one-dimensional problem. However, for the RTP process, the different components are correlated and thus the first-passage properties of the x component do not immediately reduce to the one-dimensional case. In the following, we present a mapping from the RTP trajectory in d dimensions and continuous time to a one-dimensional discrete-time random walk. This mapping allows us to compute several first-passage properties of the RTP model that cannot be derived with the standard approach, based on solving the Fokker-Planck equation associated with the process.

We denote by τ_i the duration of the i-th running phase, i.e., the time between the i-th and the $(i+1)$-th tumbling. For the first $(n-1)$ time intervals, these variables are simply exponentially distributed with average value $1/\gamma$, i.e.,

$$p(\tau) = \gamma e^{-\gamma \tau} \, . \tag{3.1}$$

However, the last running phase is yet to be completed (since we are fixing the total time to be t). As a consequence, the probability weight of the last time τ_n is

$$\int_{\tau_n}^{\infty} d\tau \; \gamma e^{-\gamma \tau} = e^{-\gamma \tau} \, . \tag{3.2}$$

Therefore, the joint probability distribution of the running times $\{\tau_i\} = \{\tau_1, \tau_2, \dots, \tau_n\}$ and of the number of tumblings n reads

$$P\left(\{\tau_i\}, \, n | t\right) = \left[\prod_{i=1}^{n-1} \gamma \, e^{-\gamma \, \tau_i} \right] \, e^{-\gamma \, \tau_n} \, \delta \left(\sum_{i=1}^{n} \tau_i - t \right) \, , \tag{3.3}$$

where the delta function constrains the total time to be t.

We next consider the d-dimensional vectors $\{\vec{l}_i\} = \{\vec{l}_1, \vec{l}_2, \dots \vec{l}_n\}$ describing the displacement of the particle during each running phase (see Fig. 3.1). The orientation of these vectors is random and isotropically distributed. Their norms $\{l_i\} = \{l_1, l_2, \dots, l_n\}$ are given by $l_i = v_i \, \tau_i$, where the random speeds v_i are independently drawn from $W(v)$. Thus, from Eq. (3.3), we find that the joint distribution of $\{l_i\}$ and the number of tumblings n reads

$$P\left(\{l_i\}, \, n | t\right) = \frac{1}{\gamma} \prod_{i=1}^{n} \int_0^{\infty} dv_i \, W(v_i) \frac{\gamma}{v_i} \, e^{-\gamma \, l_i / v_i} \, \delta \left(\sum_{i=1}^{n} \frac{l_i}{v_i} - t \right) \, . \tag{3.4}$$

From this expression, we now compute the distribution of the displacements $\{x_i\} = \{x_1 \, , \dots \, , x_n\}$ of the particle in the x direction. Note that the variable x_i is the first

entry of the random vector $\vec{l}_i$. If we fix the norm l_i of the vector $\vec{l}_i$, then one can show that the distribution of x_i is given by (see Article 4)

$$P(x_i|l_i) = \frac{1}{l_i} f_d\left(\frac{x_i}{l_i}\right) , \tag{3.5}$$

where

$$f_d(z) = \frac{\Gamma(d/2)}{\sqrt{\pi}\,\Gamma((d-1)/2)} (1-z^2)^{(d-3)/2}\,\theta(1-|z|)\,. \tag{3.6}$$

Here $\Gamma(y)$ is the Gamma function and $\theta(y)$ is the Heaviside step function: $\theta(y) = 1$ if $y \geq 0$ and $\theta(y) = 0$ if $y < 0$. Note that this function $f_d(z)$ is symmetric around $z = 0$, i.e., $f_d(z) = f_d(-z)$.

The joint distribution of the x-component displacements $\{x_i\}$ conditioned on the variables $\{l_i\}$ simply reads

$$P\left(\{x_i\}|\{l_i\}\right) = \prod_{i=1}^{n} \frac{1}{l_i} f_d\left(\frac{x_i}{l_i}\right) , \tag{3.7}$$

where we have used the independence of the variables $\{x_i\}$. Using Bayes theorem, we get

$$\begin{aligned} P\left(\{x_i\},\{l_i\},n|t\right) &= P\left(\{x_i\}|\{l_i\}\right) P\left(\{l_i\},n|t\right) \\ &= \frac{1}{\gamma} \prod_{i=1}^{n} \int_0^\infty dv_i\, W(v_i) \frac{1}{l_i} f_d\left(\frac{x_i}{l_i}\right) \frac{\gamma}{v_i}\, e^{-\gamma\, l_i/v_i} \delta\left(\sum_{i=1}^{n} \frac{l_i}{v_i} - t\right) , \end{aligned} \tag{3.8}$$

where we have used the results in Eqs. (3.4) and (3.7). Integrating over the variables $\{l_i\}$, we find that the joint distribution of $\{x_i\}$ and n reads

$$P\left(\{x_i\},n|t\right) = \frac{1}{\gamma} \prod_{i=1}^{n} \int_0^\infty dv_i\, W(v_i) \int_0^\infty dl_i\, \frac{1}{l_i} f_d\left(\frac{x_i}{l_i}\right) \frac{\gamma}{v_i}\, e^{-\gamma\, l_i/v_i}\, \delta\left(\sum_{i=1}^{n} \frac{l_i}{v_i} - t\right) . \tag{3.9}$$

This result describes the effective process obtained by projecting the RTP motion onto the x-axis. Note that the resulting process still explicitly depends on d.

The formula in Eq. (3.9) is not very handy due to the delta function, which constrains the sum of the x_i variables. To solve this problem and decouple the variables, it is convenient to take a Laplace transform with respect to t, yielding

$$\begin{aligned} \int_0^\infty dt\, e^{-st} P\left(\{x_i\},n|t\right) &= \frac{1}{\gamma} \prod_{i=1}^{n} \int_0^\infty dv_i\, W(v_i) \int_0^\infty dl_i\, \frac{1}{l_i} f_d\left(\frac{x_i}{l_i}\right) \frac{\gamma}{v_i}\, e^{-(\gamma+s) l_i/v_i} \\ &= \frac{1}{\gamma} \left(\frac{\gamma}{\gamma+s}\right)^n \prod_{i=1}^{n} \tilde{p}_s(x_i)\,, \end{aligned} \tag{3.10}$$

where we have defined

$$\tilde{p}_s(x) = \int_0^\infty dl\, \frac{1}{l} f_d\left(\frac{x}{l}\right) \int_0^\infty dv\, W(v)\, \frac{\gamma+s}{v} e^{(\gamma+s)l/v}\,. \tag{3.11}$$

In going from the first to the second line in Eq. (3.10), we have divided and multiplied by a constant factor $(\gamma+s)^n$. In this way, the function $\tilde{p}_s(x)$ is now correctly normalized

to unity. Moreover, since $\tilde{p}_s(x)$ is always positive, this quantity can be interpreted as a probability density function. Note that due to the symmetry $f_d(z) = f_d(-z)$, $\tilde{p}_s(x)$ is symmetric around $x = 0$. Thus, $\tilde{p}_s(x)$ is continuous and symmetric.

Inverting the Laplace transform in Eq. (3.10) formally, we finally obtain

$$P\left(\{x_i\}, n|t\right) = \int \frac{ds}{2\pi\, i} e^{s\,t}\, \frac{1}{\gamma} \left(\frac{\gamma}{\gamma + s}\right)^n \prod_{i=1}^{n} \tilde{p}_s(x_i)\,, \tag{3.12}$$

where the integral is over the Bromwich contour (imaginary axis in this case) in the complex s plane. Interestingly, this relation in Eq. (3.12) can be used to map the x-component projection of the RTP motion to a one-dimensional RW process.

The survival probability $S(t)$ of the RTP process can be expressed in terms of the variables $\{x_i\}$ as follows. It is useful to define the partial sums

$$X_k = x_1 + x_2 + \ldots + x_k \tag{3.13}$$

with $1 \leq k \leq n$. Then, since the RTP moves ballistically between tumblings, $S(t)$ is simply the probability that all the sums X_k, for $1 \leq k \leq n$ are positive. Thus, summing over the number n of tumblings, we get

$$S(t) = \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} dx_1 \ldots \int_{-\infty}^{\infty} dx_n\, \theta(X_1) \ldots \theta(X_n) P\left(\{x_i\}, n|t\right)\,, \tag{3.14}$$

where the theta functions constrain the partial sums to be positive. Substituting the expression for $P\left(\{x_i\}, n|t\right)$ given in Eq. (3.12) into Eq. (3.14), we get

$$\begin{aligned} S(t) &= \int \frac{ds\, e^{s\,t}}{2\pi\, i\gamma} \sum_{n=1}^{\infty} \left(\frac{\gamma}{\gamma + s}\right)^n \int_{-\infty}^{\infty} dx_1 \ldots \int_{-\infty}^{\infty} dx_n \prod_{i=1}^{n} \theta(X_i)\tilde{p}_s(x_i) \\ &= \int \frac{ds}{2\pi\, i}\, \frac{e^{s\,t}}{\gamma} \sum_{n=1}^{\infty} \left(\frac{\gamma}{\gamma + s}\right)^n q_n\,, \end{aligned} \tag{3.15}$$

where we have defined

$$q_n = \int_{-\infty}^{\infty} dx_1 \ldots \int_{-\infty}^{\infty} dx_n \prod_{i=1}^{n} \theta(X_i)\tilde{p}_s(x_i)\,. \tag{3.16}$$

Notably, this quantity q_n is precisely the survival probability of a discrete-time random walk of n steps. Indeed, the partial sum X_k describes the position of the walker at step k and the variable x_k represent the k-th step. Indeed, we have

$$X_k = X_{k-1} + x_k\,, \tag{3.17}$$

with $X_0 = 0$. The PDF $\tilde{p}_s(x_k)$ can be then interpreted as the jump distribution of the RW and the variables x_k are i.i.d. with distribution $\tilde{p}_s(x_k)$. As explained in Chapter 1, since $\tilde{p}_s(x)$ is continuous and symmetric, the Sparre Andersen theorem [38] guarantees that q_n is completely universal, i.e., independent of the specific form of $\tilde{p}_s(x)$, and given by

$$q_n = \binom{2n}{n} 2^{-2n}\,. \tag{3.18}$$

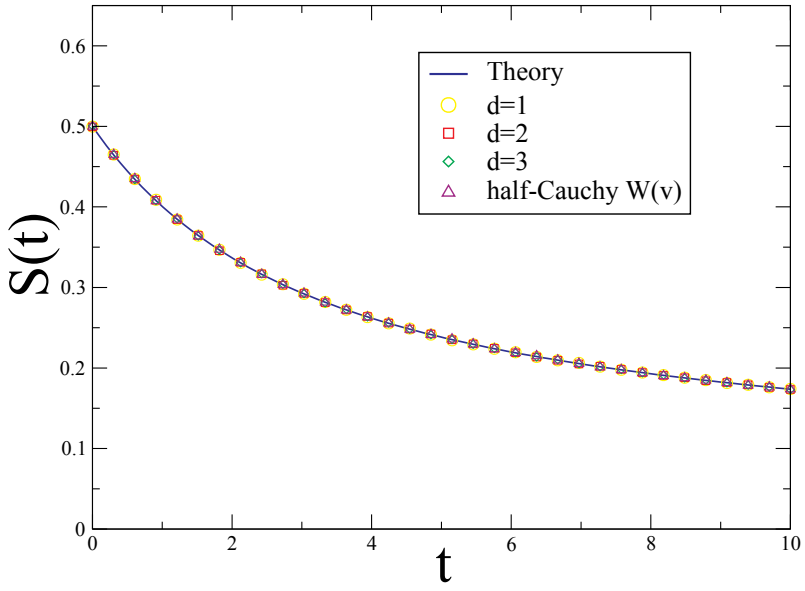

Figure 3.2: Survival probability $S(t)$ as a function of t (for $\gamma = 1$). The continuous line corresponds to the exact universal result in Eq. (3.21). The symbols correspond to numerical simulations performed for systems with different dimensions $d = 1, 2, 3$ and constant velocity and in $d = 2$ with $W(v) = 2/(\pi(1 + v^2))$ (half-Cauchy). The symbols obtained from simulations with different model parameters collapse into the same curve.

Crucially, this result is exact and valid for any n. The generating function of q_n thus reads

$$\sum_{n=0}^{\infty} q_n \, z^n = \sum_{n=0}^{\infty} \binom{2n}{n} \left(\frac{z}{4}\right)^n = \frac{1}{\sqrt{1-z}} \,. \tag{3.19}$$

Using this result (3.19) in Eq. (3.16) we get

$$S(t) = \int \frac{ds}{2\pi\, i} e^{s\,t} \, \frac{1}{\gamma} \left[\sqrt{\frac{\gamma + s}{s}} - 1\right] . \tag{3.20}$$

Remarkably, this result is universal: it does not depend on the dimension d of the system nor on the speed distribution $W(v)$. Indeed, the details of the model enter in the formula in Eq. (3.16) only through the PDF $\tilde{p}_s(x)$. Inverting the Laplace trasform, we get [179]

$$\boxed{S(t) = \frac{1}{2}\, e^{-\gamma t/2} \left(I_0\left(\frac{\gamma}{2}t\right) + I_1\left(\frac{\gamma}{2}t\right)\right) ,} \tag{3.21}$$

where $I_0(z)$ and $I_1(z)$ are modified Bessel functions. This universal result is exact and valid for any t (and not only asymptotically for late times). In the case $d = 1$, we recover, using a completely different method, the result obtained in previous works [58, 172, 176].

The function $S(t)$ is shown in Fig. 3.2 and it is in excellent agreement with numerical simulations performed in dimensions $d = 1, 2, 3$ with fixed velocity v_0, i.e., choosing $W(v) = \delta(v - v_0)$ and in $d = 2$ with speed distribution $W(v) = 2/(\pi(1 + v^2))$ (half-

Cauchy). From Eq. (3.21), we obtain the following asymptotic behaviors

$$S(t) \approx \begin{cases} \frac{1}{2} - \frac{\gamma}{8} t & \text{for } t \to 0\,, \\ \\ \frac{1}{\sqrt{\pi \gamma t}} - \frac{1}{4\sqrt{\pi \gamma^3 t^3}} & \text{for } t \to \infty\,. \end{cases} \tag{3.22}$$

In the short-time limit the survival probability converges to the limit value $1/2$ because, by symmetry, half of the trajectory will immediately go in the negative-x direction.

Note that this universal result is only valid if the tumbling events occur with a constant rate γ, i.e., if the distribution of the running time is exponential $p(\tau) = \gamma e^{-\gamma \tau}$. Nevertheless, one can show that if $p(\tau)$ has a well-defined first moment, then the survival probability still decays for late times as $S(t) \sim t^{-1/2}$. On the other hand, when $p(\tau) \sim \tau^{-\mu-1}$ for large τ with $0 < \mu < 1$, the first moment is not defined and one can show that $S(t) \sim t^{-\mu/2}$ for late times (see Article 5).

The main results of this section are: (i) the technique to map the continuous-time process to an effective discrete-time RW in Laplace space and (ii) the unexpected universality of the survival probability. The mapping that we have introduced is very general and can be used to investigate several different observables of the RTP model. The universal result of the survival probability can be generalized to several other observables and models, as discussed in detail in the following sections.

3.2 Generalizations to other observables

The result obtained in the previous section for the survival probability can be used as a building block to compute the distribution of other observables of the RTP trajectory. This derivation is somewhat similar to the one presented in Chapter 1 for discrete-time RWs, in which we have used the universal expression of the survival probability to show that the distribution of the time of the maximum and the record statistics are also universal.

3.2.1 Time of the maximum

We first consider the time $t_{\max}$ at which the x-component of the RTP reaches the global maximum before time t. The derivation of the distribution of this quantity is based on a path-decomposition technique, very similar to the one presented in Chapter 1. The central idea is to divide the time interval $[0, t]$ into the two subintervals $[0, t_{\max}]$ (I) and $[t_{\max}, t]$ (II). In the first time interval, the RTP starts from the origin and reaches the global maximum at time $t_{\max}$, while in the second interval the particle has to stay below the maximum up to time t. It is possible to show that the probability weights of both intervals can be expressed in terms of the survival probability $S(t)$. This is clear for the interval (II), since the particle in $[t_{\max}, t]$ has to remain below its starting position, which happens with probability $S(t - t_{\max})$. The derivation for the first interval is slightly less obvious, one has to consider the RTP trajectory in the time interval $[0, t_{\max}]$ after inverting the direction of time. Then, once again the particle has to remain below its starting position for a total time $t_{\max}$ and the probability weight of the first interval is $S(t_{\max})$. For a more precise formulation of this argument, see Article 5. Thus, we

obtain that the PDF of $t_{\max}$ is given by the product of the probability weights of two time intervals, yielding

$$P(t_{\max}|t) = \gamma S(t_{\max}) S(t - t_{\max})\,, \tag{3.23}$$

where the factor γ comes from the fact that the particle tumbles right after reaching the maximum (the probability of tumbling in a small time interval dt is γdt).

It is easy to check that this expression is not correctly normalized to unity over $t_{\max}$. This is because the construction above is only valid for $0 < t_{\max} < t$. It turns out that with finite probability the events "$t_{\max} = 0$" and "$t_{\max} = T$" can also occur (both with probability $S(t)$). Including also these events, we get

$$\boxed{P(t_{\max}|t) = S(t)\delta(t_{\max}) + \gamma S(t_{\max}) S(t - t_{\max}) + S(t)\delta(t - t_{\max})\,,} \tag{3.24}$$

where $S(t)$ is given in Eq. (3.21). One can check that this result is correctly normalized to unity. Quite remarkably the universality of the survival probability extends also to the distribution of the time $t_{\max}$ of the maximum. The result in Eq. (3.24) is valid for any t and is completely independent of the dimension d of the system and of the speed distribution $W(v)$. Note that in the special case $d = 1$ and $W(v) = \delta(v - v_0)$, the result in Eq. (3.24) can be derived by solving the Fokker-Plank equation associated to the system [99].

3.2.2 Record statistics

In general, it is quite nontrivial to compute the statistics of records for a sequence of correlated random variables [67]. Here, we show that the universality of the survival probability can be also generalized to the statistics of records. In particular, we consider lower records (upper records can be studied analogously). We consider a single RTP in d dimensions starting at the origin. The $x-$component X_k of the positions of the RTP at the end of the k-th running phase can be written as

$$X_k = x_1 + x_2 + \ldots + x_k\,. \tag{3.25}$$

Even though the process happens in continuous time, we can define records in the following way. We say that a lower record happens at step k if X_k is lower than all the previous values, i.e., $X_k < \min\{X_0 = 0, X_1, \cdots, X_{k-1}\}$. By convention, we consider $X_0 = 0$ as a lower record. Note that the final position X_n can also be a record.

We denote by $S_N(t)$ the probability that there are exactly N lower records up to time t. Clearly, when $N = 1$ this corresponds to the event that the position has never gone below 0 up to time t and thus $S_1(t) = S(t)$ (this is the reason why we have considered lower records). In this sense, $S_N(t)$ can be considered as a generalization of the survival probability $S(t)$.

Moreover, as explained in Chapter 1, for a discrete-time RW with continuous and symmetric jumps, the distribution of the number $N_R(n)$ of records up to step n is universal, i.e., independent of the jump distribution. Indeed, it can be shown that [23]

$$\text{Prob.}(N_R(n) = N) = \binom{2n - N + 1}{n} 2^{-2n+N-1}\,, \tag{3.26}$$

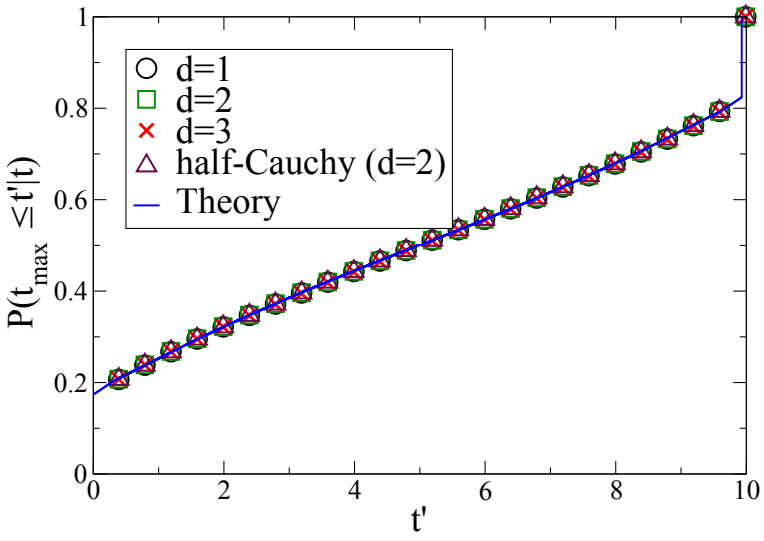

Figure 3.3: Cumulative distribution $P(t_{\max} \leq t'|t)$ of the time $t_{\max}$ of the maximum for the RTP model as a function of t', with $\gamma = 1$ and $t = 10$. The continuous blue line corresponds to the exact result in Eq. (3.24). The symbols correspond to numerical simulations with $d = 1, 2, 3$ ($W(v) = \delta(v-1)$) and $d = 2$ with $W(v) = 2/(\pi\,(1+v^2))$ (half-Cauchy). We observe that the numerical curves collapse on the corresponding analytical blue line.

with generating function

$$\sum_{n=N-1}^{\infty} \text{Prob.}(N_R(n) = N) z^n = \frac{(1-\sqrt{1-z})^{N-1}}{\sqrt{1-z}} \,. \tag{3.27}$$

Using the mapping in Eq. (3.12) between the RTP and a discrete-time RW, we can apply directly the result for the record statistics of RWs to our problem. Using the result in Eq. (3.27), we get

$$S_N(t) = \int \frac{ds}{2\pi\, i} e^{s\,t}\, \frac{1}{\gamma} \sqrt{\frac{\gamma+s}{s}} \left(1 - \sqrt{\frac{s}{\gamma+s}}\right)^{N-1} \,, \tag{3.28}$$

which is exact and valid for any N and t. For $N = 1$, as expected, we find $S_1(t) = S(t)$. Setting $N = 2$ and inverting the Laplace transform, we get

$$S_2(t) = \frac{1}{2} e^{-\gamma t/2} \left(I_0\left(\gamma t/2\right) + I_1\left(\gamma t/2\right)\right) \,. \tag{3.29}$$

Finally, from the result in Eq. (3.28), one can also obtain the average number $\langle N(t)\rangle$ of records up to time t, which reads

$$\boxed{\langle N(t)\rangle = \frac{e^{-\frac{\gamma t}{2}}}{2} \left((2\gamma t + 3) I_0\left(\frac{\gamma t}{2}\right) + (2\gamma t + 1) I_1\left(\frac{\gamma t}{2}\right)\right) .} \tag{3.30}$$

This quantity is shown in Fig. 3.4 and is in perfect agreement with numerical simulations.

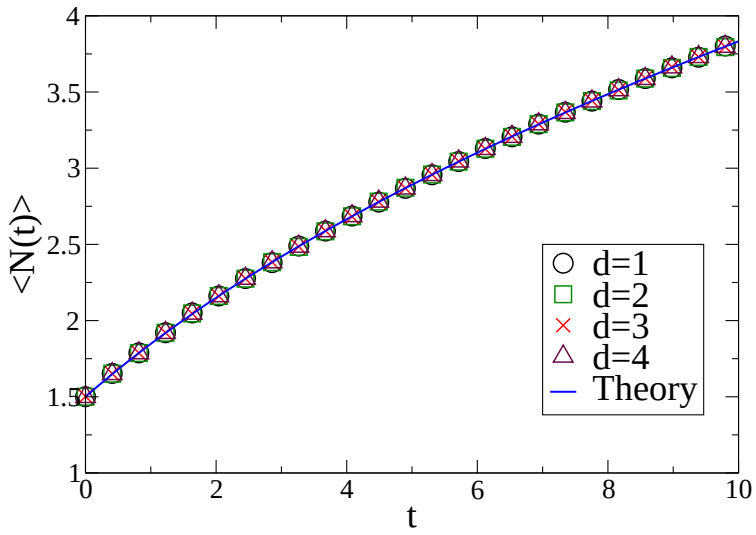

Figure 3.4: Average number of records $\langle N(t)\rangle$ vs the time t for the RTP model. The solid line is given by the exact formula (3.30) while the symbols represent numerical simulations in $d = 1, 2, 3, 4$ with $\gamma = 1$ and $v_0 = 1$.

3.3 Generalizations to other models

Interestingly, both the mapping to a discrete RW problem and the universality of the survival probability can be generalized to more complicated RTP models. The purpose of this section is to provide a concise presentation of these generalizations. For the details of the models and the computations, see Articles 5 and 6.

3.3.1 Run-and-tumble particles with instantaneous runs

The first model that we consider is a simplified version of the standard RTP model. We assume that the particle waits a random time during a tumbling and then jumps instantaneously to its new position. We refer to this model as the instantaneous-run (IR) model. To be precise, consider a particle that starts from the origin and evolves in d dimensions up to time t. In the beginning, the particle remains at the origin for a random time T_1, distributed according to the PDF $P_W(T)$, then it chooses a new direction uniformly at random and takes an instantaneous jump of length $v_1 T_1$ in that direction, where $v_1 \geq 0$ is drawn from $W(v)$. Then, it waits a random time T_2, drawn from the distribution $P_W(T)$, then it jumps, and so on. For an example of a typical trajectory, see Fig. 3.5**b**. Note that a discrete-time version of this model was recently investigated in [180].

We investigate the survival probability $S^{\mathrm{IR}}(t)$, defined as the probability that the x-component of the particle remains positive up to time t. Using a construction analogous to the one presented for the standard RTP model, we are able to show that the survival

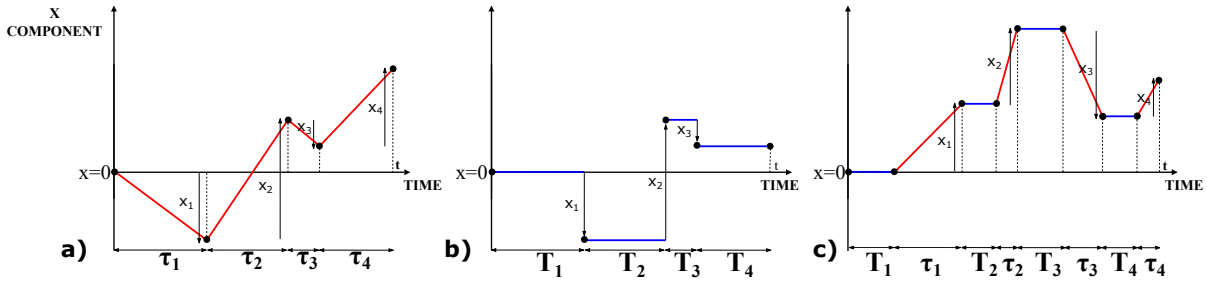

Figure 3.5: Projection of the RTP trajectory onto the x axis for the standard RTP model in panel **a**, the RTP model with instantaneous runs (IR model) in panel **b**, and the RTP model with non-instantaneous tumblings (mixed model) in panel **c**. In the standard RTP model (panel **a**), the particle tumbles instantaneously and the durations of the running phases are indicated by $\tau_1, \ldots, \tau_n$, where n is the number of tumblings (here $n = 4$). In the IR model (panel **b**), the particle jumps instantaneously after each waiting phase. The durations of the waiting phases are indicated by $T_1, \ldots, T_n$. In the mixed model (panel **c**), both the running and the waiting phases have finite duration.

probability for the IR model is given by

$$S^{\rm IR}(t) = \int \frac{ds}{2\pi i} e^{st} \frac{\sqrt{1-\tilde{P}_W(s)}}{s}\,, \tag{3.31}$$

where

$$\tilde{P}_W(s) = \int_0^\infty dt\; e^{-st} P_W(t)\,. \tag{3.32}$$

This result is again completely independent of the dimension d of the system and the speed distribution $W(v)$. Quite interestingly, the expression in Eq. (3.31) is valid for any waiting-time distribution $P_W(t)$ and not only for exponential times. Using the exact relation in Eq. (3.31) one can also infer the late-time behavior of the survival probability of the standard RTP model with non-exponential running times. Note that a very similar result was also obtained in [57] by combinatorial techniques.

In particular, in the case $P_W(t) = ae^{-at}$, we find $\tilde{P}_W(s) = a/(a+s)$ and thus [179]

$$S^{\rm IR}(t) = \int \frac{ds}{2\pi i} e^{st} \frac{1}{\sqrt{s(a+s)}}\,, = e^{-at/2} I_0\left(\frac{at}{2}\right)\,. \tag{3.33}$$

This exact result is shown in Fig. 3.6 and is in excellent agreement with numerical simulations. Using the survival probability as a building block, one can also compute the distribution of the time of the maximum and the record statistics, which are also universal.

3.3.2 Run-and-tumble particles with non-instantaneous tumblings

One of the main assumptions of the standard RTP model is that the tumbling events happen instantaneously. However, in many cases, this is not realistic. For instance,

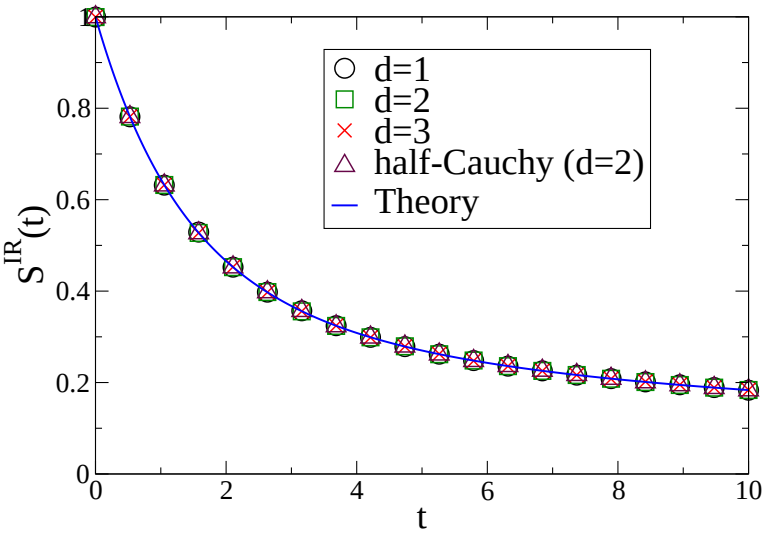

Figure 3.6: Survival probability $S^{\rm IR}(t)$ as a function of t for the IR model with $P_W(t) = e^{-t}$. The continuous line corresponds to the exact universal result in Eq. (3.33). The symbols correspond to numerical simulations performed for systems with different dimensions $d = 1, 2, 3$ and constant velocity and in $d = 2$ with $W(v) = 2/(\pi(1 + v^2))$ (half-Cauchy). The symbols obtained from simulations with different model parameters collapse into the same curve.

from experiments on the motion of *E. coli* bacteria, it was observed that the ratio of the typical tumbling time (i.e., the time required for the bacterium to change direction) to the typical running time is approximately 0.1 [44]. Thus, it is relevant to ask how a finite tumbling time would affect the properties of the model.

We assume that the particle is initially at the origin in d dimensions and chooses a direction uniformly at random. Orienting itself in the new direction requires a time T_1, which is random and distributed according to the PDF $P_W(T)$. Then, the particle starts to move in the chosen direction with velocity v_1, drawn from $W(v)$. After some random time τ_1, exponentially distributed with rate γ, the particle tumbles again, i.e., it does not move for a time T_2, independently drawn from the same PDF $P_W(T)$. Then, it starts moving in a new direction with a new velocity v_2, and so on up to time t. Since this model is a combination of the standard RTP model and the IR model, we call it "mixed model". For a typical trajectory of this process, see Fig. 3.5**c**.

The survival probability for this model can be computed for any waiting-time distribution $P_W(T)$ and is given by

$$S^{\rm Mixed}(t) = \int \frac{ds}{2\pi i} e^{st} \left(\frac{\left(1 - \tilde{P}_W(s)\right)}{s\, h(s)} + \frac{1}{\gamma} \left(\frac{1}{h(s)} - 1 \right) \right) , \tag{3.34}$$

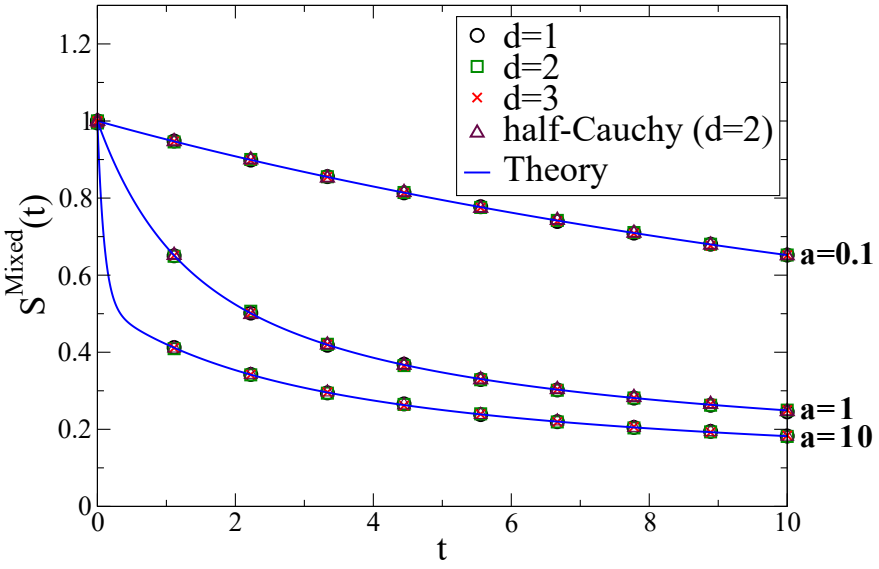

Figure 3.7: Survival probability $S^{\rm Mixed}(t)$ as a function of t for the mixed model with $P_W(t) = ae^{-at}$. The continuous line corresponds to the exact universal result in Eq. (3.36) for $\gamma = 1$ and $a = 0.1, 1, 10$. The symbols correspond to numerical simulations performed for systems with different dimensions $d = 1, 2, 3$ and constant velocity and in $d = 2$ with $W(v) = 2/(\pi(1 + v^2))$ (half-Cauchy). The symbols obtained from simulations with different model parameters collapse into the same curve.

where

$$h(s) = \sqrt{1 - \tilde{P}_W(s)\frac{\gamma}{\gamma + s}}\,. \tag{3.35}$$

We recall that $\tilde{P}_W(s)$ is the Laplace transform of the waiting-time distribution $P_W(T)$.

In particular, when the waiting times are exponentially distributed with rate a, i.e., $P_W(t) = a^{-at}$, we obtain

$$S^{\rm Mixed}(t) = \frac{\gamma}{4} e^{-\gamma t/2} \int_0^t dt'\, e^{-at'} \left(I_0\left(\frac{\gamma}{2}t'\right) + I_1\left(\frac{\gamma}{2}t'\right)\right) \tag{3.36}$$
$$\times \quad \left(I_0\left(\frac{\gamma}{2}(t-t')\right) + I_1\left(\frac{\gamma}{2}(t-t')\right)\right) + \frac{1}{2}\left(1 + e^{-at}\right) e^{-\gamma t/2} \left(I_0\left(\frac{\gamma}{2}t\right) + I_1\left(\frac{\gamma}{2}t\right)\right)\,,$$

Where the integral over t' can be computed numerically for given values of a and γ. In Fig. 3.7 we compare this theoretical result in Eq. (3.36) with numerical simulations for $\gamma = 1$ and different values of a, finding an excellent agreement. Interestingly, also for the mixed model, the survival probability can be used as a building block to compute the distribution of the time of the maximum and the record statistics.

3.3.3 Discrete-time persistent random walk

The derivation presented in the previous sections can be adapted also to discrete-time persistent RWs, which are defined as follows. We consider a one-dimensional RW

$$x_{n+1} = x_n + \sigma_n \eta_n\,, \tag{3.37}$$

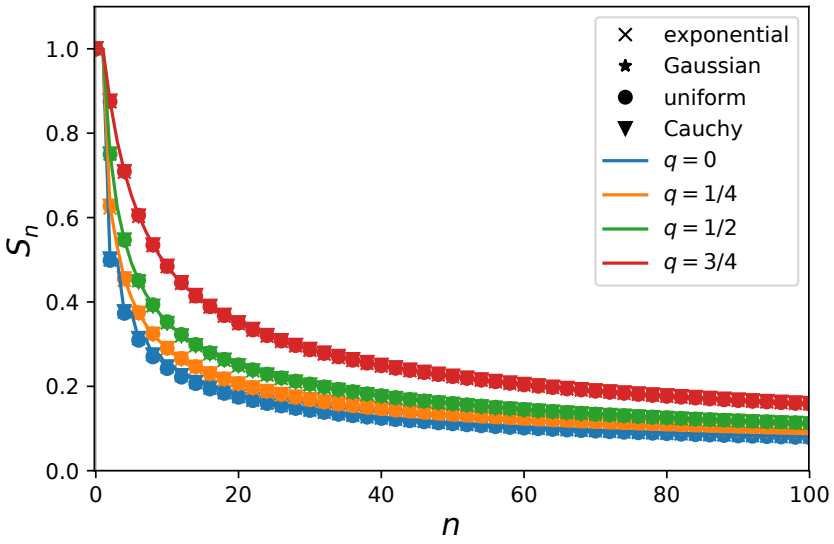

Figure 3.8: Survival probability S_n as a function of n for the persistent random walk with persistence probability q. The continuous lines correspond to the exact universal result in Eq. (3.39) for different values of q and different jump distributions $p(\eta)$. For a given value of q, the symbols obtained from simulations with different jump distributions $p(\eta)$ collapse into the same curve.

where $\sigma_n = \pm 1$ are binary variables evolving according to

$$\sigma_n = \begin{cases} \sigma_{n-1} & , \text{ with probability } q \, , \\ -\sigma_{n-1} & , \text{ with probability } 1-q \, . \end{cases} \tag{3.38}$$

The step $\eta_1 \, , \eta_2 \, , \ldots$ are positive i.i.d. random variables drawn from the continuous PDF $p(\eta)$, with $\eta > 0$. We assume that the RW starts at position $x = 0$ in state $\sigma_0 = +1$.

Note that the steps $\sigma_i \eta_i$ of the random walker are correlated for this model (one can easily show that the correlation decay exponentially in time). Interestingly, in the limit $n \to \infty$, $q \to 1$, with $n(1-q) = \gamma t$ fixed, one recovers the standard RTP model. On the other hand, for $q = 1/2$, one finds the standard RW model without persistence.

We want to investigate the survival probability S_n, i.e., the probability that the walker stays above $x = 0$ up to step n. It turns out that the mapping to the standard RW with symmetric jumps presented in Section 3.1 can be generalized to this model. Using this mapping, we show that the survival probability is completely universal, i.e., independent of $p(\eta)$, and is given by

$$S_n = \frac{2^{-2n}}{1-q} \binom{2n}{n} {}_2F_1 \left(-\frac{1}{2}, -n, \frac{1}{2} - n \, ; 2q-1 \right) \, , \tag{3.39}$$

where ${}_2F_1 \left(a, b, c \, ; x \right)$ is the hypergeometric function defined as

$${}_2F_1 \left[a, b, c \, ; x \right] = \sum_{n \geq 0} \frac{(a)_n (b)_n}{n! (c)_n} x^n \, , \tag{3.40}$$

and $(a)_n = \Gamma(a+n)/\Gamma(a)$ the rising factorial. For the details of the derivation, see Article 6. This exact result is shown in Fig. 3.8 and is in perfect agreement with numerical simulations, performed for different values of q and with different jump distributions. One can show that in the limits $q \to 1/2$ and $q \to 1$, one respectively obtains the known results for the survival probability of a RW and an RTP. Moreover, the survival probability can be used to show that the distribution of the time of the maximum and the statistics of records are completely universal for this model.

Article 4

Universal survival probability for a d-dimensional run-and-tumble particle

F. Mori, P. Le Doussal, S. N. Majumdar, and G. Schehr,
Phys. Rev. Lett. **124**, 090603 (2020).

☞ https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.124.090603
☞ https://arxiv.org/abs/2001.01492

Abstract

We consider an active run-and-tumble particle (RTP) in d dimensions and compute exactly the probability $S(t)$ that the x-component of the position of the RTP does not change sign up to time t. When the tumblings occur at a constant rate, we show that $S(t)$ is independent of d for any finite time t (and not just for large t), as a consequence of the celebrated Sparre Andersen theorem for discrete-time random walks in one dimension. Moreover, we show that this universal result holds for a much wider class of RTP models in which the speed v of the particle after each tumbling is random, drawn from an arbitrary probability distribution. We further demonstrate, as a consequence, the universality of the record statistics in the RTP problem.

Article 5

Universal properties of a run-and-tumble particle in arbitrary dimension

F. Mori, P. Le Doussal, S. N. Majumdar, and G. Schehr,
Phys. Rev. E **102**, 042133 (2020).

☞ https://journals.aps.org/pre/abstract/10.1103/PhysRevE.102.042133
☞ https://arxiv.org/abs/2006.06989

Abstract

We consider an active run-and-tumble particle (RTP) in d dimensions, starting from the origin and evolving over a time interval $[0, t]$. We examine three different models for the dynamics of the RTP: the standard RTP model with instantaneous tumblings, a variant with instantaneous runs and a general model in which both the tumblings and the runs are non-instantaneous. For each of these models, we use the Sparre Andersen theorem for discrete-time random walks to compute exactly the probability that the $x-$component does not change sign up to time t, showing that it does not depend on d. As a consequence of this result, we compute exactly other $x-$component properties, namely the distribution of the time of the maximum and the record statistics, showing that they are universal, i.e. they do not depend on d. Moreover, we show that these universal results hold also if the speed v of the particle after each tumbling is random, drawn from a generic probability distribution. Our findings are confirmed by numerical simulations. Some of these results have been announced in a recent Letter [Phys. Rev. Lett. **124**, 090603 (2020)].

Article 6

Universal survival probability for a correlated random walk and applications to records

B. Lacroix-A-Chez-Toine, and F. Mori,
J. Phys. A: Math. Theor. **53**, 495002 (2020).

☞ https://iopscience.iop.org/article/10.1088/1751-8121/abc129
☞ https://arxiv.org/abs/2007.10969

Abstract

We consider a model of space-continuous one-dimensional random walk with simple correlation between the steps: the probability that two consecutive steps have same sign is q with $0 \leq q \leq 1$. The parameter q allows thus to control the persistence of the random walk. We compute analytically the survival probability of a walk of n steps, showing that it is independent of the jump distribution for any finite n. This universality is a consequence of the Sparre-Andersen theorem for random walks with uncorrelated and symmetric steps. We then apply this result to derive the distribution of the step at which the random walk reaches its maximum and the record statistics of the walk, which show the same universality. In particular, we show that the distribution of the number of records for a walk of $n \gg 1$ steps is the same as for a random walk with $n_{\text{eff}}(q) = n/(2(1-q))$ uncorrelated and symmetrically distributed steps. We also show that in the regime where $n \to \infty$ and $q \to 1$ with $y = n(1-q)$, this model converges to the run-and-tumble particle, a persistent random walk often used to model the motion of bacteria. Our theoretical results are confirmed by numerical simulations.

Chapter 4

Large deviations and condensation

In this chapter, we investigate the position distribution of a single run-and-tumble particle (RTP) in d dimensions and with speed distribution $W(v)$. We will either observe an RTP trajectory for a total fixed time T (*fixed-T ensemble*) or for exactly N complete running phases (*fixed-N ensemble*). Note that the fixed-N ensemble is a discrete-time RW in d-dimensions. Even though for short times the two ensembles correspond to different stochastic processes (when N is fixed, T is free to fluctuate, and vice versa), one expects that the late-time behavior (large N or large T) is qualitatively similar. This equivalence was explicitly verified in Article 7 in the case $W(v) = \delta(v - v_0)$ and arbitrary d, but we expect it to be valid for arbitrary $W(v)$. Thus, we will mostly focus on the fixed-N case, which is easier to study analytically.

One of the most natural quantities that one can investigate for the RTP model is the position distribution $P(\vec{R}, T)$, i.e., the probability density of finding the particle at position $\vec{R}$ at time T[1]. For short times, this distribution will depend on the specific details of the model, i.e., the tumbling rate γ, the dimension d of the system, and the speed distribution $W(v)$. On the other hand, at late times, one expects the PDF $P(\vec{R}, T)$ to converge to a Gaussian distribution, as a consequence of the central limit theorem (CLT),

$$P(\vec{R}, T) \approx \frac{1}{\sqrt{4\pi DT}} e^{-R^2/(4DT)}\,, \tag{4.1}$$

where $R = ||\vec{R}||$ and the parameters of the model enter the distribution only through the diffusion constant D. Thus, for late times, the distribution of the position of an active particle converges to that of a passive Brownian motion.

It is thus natural to ask whether or not any signatures of the "activity" of the particle remain at late times. Interestingly, it turns out that the large-deviation tails of the distribution $P(\vec{R}, T)$ carry these signatures of the activity (note that the CLT is only valid around the peak of the distribution $P(\vec{R}, T)$). Thus, it is relevant to investigate the rare events where the RTP is very far from its starting position at time T. Typically, one expects the total displacement of the particle to grow as $R \sim \sqrt{T}$, while the large deviation regime corresponds to $R \sim T^\alpha$, where $\alpha > 1/2$ depends on the model (for most models, we will find $\alpha = 1$). In this regime, the distribution $P(\vec{R}, T)$ assumes the

[1]Here T stands for either the total elapsed time or the number N of running phases.

following large-deviation form

$$P(\vec{R},T) \sim \exp\left[-T^{2\alpha-1}\psi\left(\frac{R}{T^{\alpha}}\right)\right], \tag{4.2}$$

where the rate function $\psi(z)$ describes the probability of the large deviations. For small z, one has $\psi(z) \approx z^2/(4D)$, describing the universal Gaussian fluctuations in Eq. (4.1).

The position distribution of an RTP with constant velocity was first computed in [170] for $d = 2$. Later, in Ref. [181], this result was extended to arbitrary dimension d. However, these authors did not investigate the large deviations, which were first studied in [182], where the rate function $\psi(z)$ was computed for $W(v) = \delta(v - v_0)$ and arbitrary $d > 0$. In this case, one finds that the large deviation regime occurs for $R \sim T$, corresponding to $\alpha = 1$ in Eq. (4.2). Interestingly, the authors of [182] noticed that for $d > 5$ the rate function $\psi(z)$ has a singularity at some critical point $z = z_c$, corresponding to the critical value of the displacement $R_c = z_c N$. This singularity signals the presence of a dynamical phase transition. This turns out to be a *condensation* transition: for $R < R_c$ (*fluid phase*) the displacements of the particle during different runs contribute to the total displacement roughly by the same amount, while for $R > R_c$ the total displacement is dominated by a single long jump (*condensed phase*). Thus, the total displacement R plays the role of the control parameter for this transition. For a schematic representation of the transition, see Fig. 4.1. Moreover, in [183], a similar condensation transition was observed for a one-dimensional RTP with a half-Gaussian speed distribution $W(v) = \sqrt{2/\pi}\, e^{-v^2/2}\, \theta(v)$ (where $\theta(v)$ is the Heaviside step function) and in the presence of a constant drift.

These two examples suggest that condensation could be a general feature of the RTP model. However, in the canonical RTP model with fixed speed v_0, this condensation occurs only in $d > 5$, which is not accessible physically. In this chapter, we will first provide a general criterion for the condensation transition. Then, we will investigate specific models for the speed distribution $W(v)$ for which the transition can be observed in the physical dimensions $d = 1, 2, 3$. Note also that the large deviations of the position have been also investigated for several generalizations of the RTP model, including an RTP on a lattice [184, 185], an RTP model with space-dependent velocities [186] and an RTP model with generalized telegraphic noise [187].

Condensation transitions in physics are traditionally observed in momentum space, the most famous example being the Bose-Einstein condensation. This transition occurs for an ideal Bose gas in $d > 2$: below a critical temperature, a macroscopic number of particles condense in the single-particle ground state. Interestingly, condensation transitions have been observed in real space as well [188]. This typically arises in situations where a conserved quantity (e.g., the total mass) is shared between different sites or units. Above a critical value of this quantity, a condensate appears in real space, meaning that a single site absorbs a macroscopic fraction of the mass. Examples of real-space condensation can be observed in a variety of situations, including traffic models [189–191], models of mass transport [192–200], macroeconomic models [201, 202], discrete nonlinear Schrödinger equation [203], financial models [204, 205]. For example, in the case of traffic models, condensation describes the formation of traffic jams, while in wealth models, a condensate is an extremely rich individual. In the case of the RTP model, the condensate is a single long run that dominates the particle trajectory.

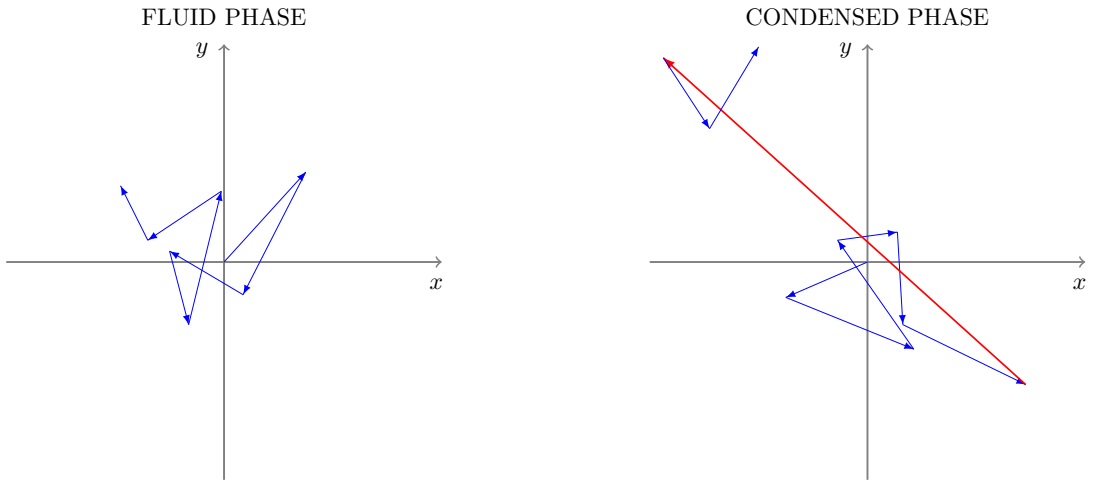

Figure 4.1: **Left panel:** Typical trajectory of a run-and-tumble particle (RTP) in two dimensions in the fluid phase. The different runs contribute to the displacement by roughly the same amount. **Right panel:** Typical trajectory of an RTP in the condensed phase. One single run (red arrow) dominates the trajectory.

The condensation transition that we investigate in this chapter has also applications in a broader context. Indeed, the position of the RTP can be simply written as the sum of many i.i.d. random variables. Let us consider for simplicity the fixed-N ensemble. First of all, since the direction is chosen uniformly in space after each run, the distribution of the position $\vec{R}$ is isotropic, i.e., $P(\vec{R}, N) = P(R, N)$[2]. Moreover, let X be the x-component of the position $\vec{R}$ and $P(X, N)$ its associated PDF. Then, it is easy to show that $P(\vec{R}, N)$ and $P(X, N)$ have the same rate function $\psi(z)$ that describes the large-deviation regime where $R \sim N^\alpha$ and $X \sim N^\alpha$ (see Article 7). Thus, to investigate the large deviations, it is sufficient to consider the distribution of X. We denote by x_i the displacement of the particle during the i-th running phase, such that

$$X = \sum_{i=1}^{N} x_i \,. \tag{4.3}$$

Then, the distribution of X can be written as

$$P(X, N) = \int_{-\infty}^{\infty} dx_1 \, \ldots \int_{-\infty}^{\infty} dx_N \, \prod_{i=1}^{N} p(x_i) \; \delta\left(\sum_{i=1}^{N} x_i - X \right) \,, \tag{4.4}$$

where $p(x)$ is the PDF of the variables $x_1 \,, \ldots \,, x_N$. Large deviations of the sum of many i.i.d. random variables have been widely investigated in the mathematical literature [24,206–208] and have recently been investigated in the case of correlated variables [209]. In the case of the RTP model, in Chapter 3 we have shown that

$$p(x) = \int_0^\infty dv \; \frac{1}{v} W(v) \int_0^\infty d\ell \; \frac{1}{\ell} f_d\left(\frac{x}{\ell} \right) \; \gamma e^{-\gamma \ell / v} \,, \tag{4.5}$$

[2]Note that for simplicity we will use the same notation for the fixed-N ($P(R, N)$) and the fixed-T ($P(R, T)$) ensembles.

where

$$f_d(z) = \frac{\Gamma\left(d/2\right)}{\sqrt{\pi}\Gamma\left((d-1)/2\right)}(1-z^2)^{(d-3)/2}\theta(1-|z|)\,, \tag{4.6}$$

and $\Gamma(y)$ is the Gamma function. Due to the isotropy of the RTP dynamics we observe that $p(x) = p(-x)$ and $P(X,N) = P(-X,N)$. Thus, it is sufficient to investigate the case $X > 0$.

Assuming that the variance of $p(x)$ is finite, the distribution of X approaches a Gaussian shape for $|X| \sim \sqrt{N}$. On the other hand, when $|X| \gg N$ we expect $P(X,N) \sim Np(X)$, corresponding to a single variable that completely dominates the sum. Then, for a given expression of $p(x)$, it is worth investigating whether between the two regimes there is a smooth crossover or a sharp transition (as in Fig. 4.2).

4.1 Criterion for condensation

In this section, we present a simple criterion, based on a grand-canonical argument, to determine whether or not condensation occurs for a given jump distribution $p(x)$ (see Article 7, for the derivation). Note however that this criterion does allow to determine the order of the transition, which we will discuss in the next sections. A similar criterion was previously derived in the context of mass-transport models [192,193]. However, the criterion in [192,193] is only valid in the case where the variables $x_1\,,\ldots\,,x_N$ are positive and it does not directly apply to our case. To proceed, it is useful to define the limit value

$$c = -\lim_{x\to\infty}\frac{\log(p(x))}{x}\,. \tag{4.7}$$

We consider three cases, depending on c.

The first case is $c = \infty$. This corresponds to distributions $p(x)$ that decay faster than any exponential, i.e.,

$$p(x) \ll e^{-bx}\,, \tag{4.8}$$

for $x \gg 1$ and for any $b > 0$. For instance, this is the case when $p(x)$ is Gaussian. In this case, no condensation transition occurs.

The case $0 < c < \infty$ corresponds to distributions that can be written as

$$p(x) = \tilde{p}(x)e^{-c|x|}\,, \tag{4.9}$$

where the function $\tilde{p}(x)$ decays slower than any exponential. In this case, a condensation transition occurs only if $\tilde{p}(x)$ decays faster than $1/x^2$, i.e., if

$$\tilde{p}(x) \ll 1/x^2\,, \tag{4.10}$$

for large x. For instance, if $p(x) = e^{-x}x^{-\beta}$, a condensation transition occurs only for $\beta > 2$.

Finally, in the case $c = 0$, we expect condensation to occur if

$$p(x) \ll 1/|x|^3\,, \tag{4.11}$$

for large x. This is for instance the case where $p(x)$ has a stretched-exponential tail $p(x) \sim e^{-x^\beta}$ with $0 < \beta < 1$.

Going back to the RTP model, let us provide a few examples of speed distributions $W(v)$ for which condensation occurs:

- $W(v) = \alpha(v_0 - v)^{\alpha-1}$ with $0 < v < v_0$. The constant $v_0 > 0$ represents the maximal speed that the particle can reach. Note that this family includes the standard RTP model where the speed is fixed. Indeed, in the limit $\alpha \to 0$, we get

$$W(v) = \delta(v - v_0). \tag{4.12}$$

  Moreover, other relevant distributions belong to this class. For instance, for $\alpha = 1$, one obtains the uniform distribution. Given α and d, one can show that for large $|x|$

$$p(x) \sim e^{-|x|} \frac{1}{|x|^\nu} \ , \tag{4.13}$$

  where

$$\nu = \frac{(d + 2\alpha - 1)}{2} \, . \tag{4.14}$$

  In this case we get $c = 1$ from Eq. (4.7) and, applying the criterion described above, we find that the transition is possible only for $\nu > 2$, i.e., for $d + 2\alpha > 5$. Thus, for $\alpha = 0$, i.e., for the standard RTP model, we find that condensation only occurs for $d > 5$, as already observed in [182].

- $W(v) = \sqrt{\frac{2}{\pi}} \, e^{-v^2/2}$ with $v > 0$. Considering $d = 1$, one can show that, for $|x| \gg 1$,

$$p(x) \sim |x|^{-1/3} e^{-3|x|^{2/3}/2} \, . \tag{4.15}$$

  In this case $p(x)$ decays slower than any exponential, thus $c = 0$. Moreover, $p(x)$ decays faster than $1/|x|^3$ and thus we expect a condensation transition to occur.

In the next sections, we will analyze in detail the two examples above.

4.2 Arbitrary dimension and random speed: second-order phase transition

In this section, we investigate the case where the particle moves in d dimensions with the speed distribution

$$W(v) = \alpha(v_0 - v)^{\alpha-1} \, , \tag{4.16}$$

where $\alpha > 0$ and $v_0 > 0$. For simplicity in the following we will take $\gamma = v_0 = 1$. Our goal is to compute the large-deviation function $\psi_{d,\alpha}(z)$, where we now use the subscripts α and d to stress the dependence on these parameters. The results of this section are presented in detail in Article 7.

4.2.1 Position distribution

To proceed, we recall the integral representation of the δ function

$$\delta(X) = \frac{1}{2\pi i} \int_\Gamma dq \ e^{-qX} \, , \tag{4.17}$$

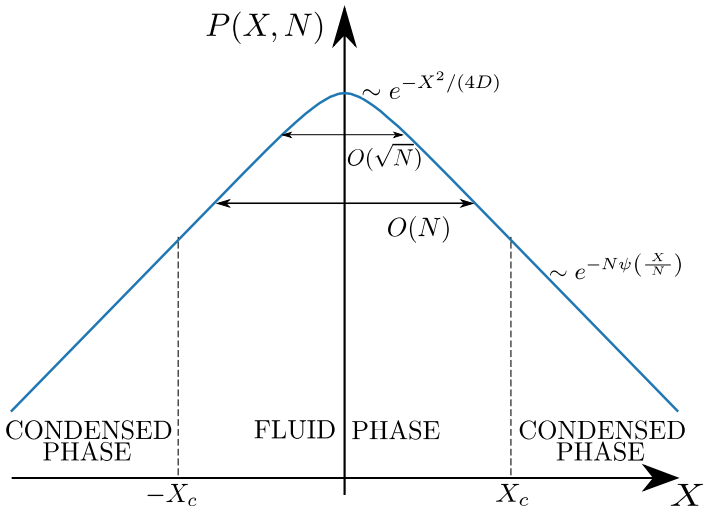

Figure 4.2: Schematic representation of the probability density function $P(X,N)$ of the RTP position in the presence of condensation. For $X \sim \sqrt{N}$, the fluctuations of X are Gaussian, as predicted by the central limit theorem. In the large deviation regime where $X \sim N$, the distribution assumes the scaling form $P(X,N) \sim e^{-N\psi(X/N)}$. The rate function $\psi(z)$ is singular at the critical value $z_c = X_c/N$. For $X < X_c$, the system is in the fluid phase and the different runs are roughly of the same order. For $X > X_c$, a condensate appears, i.e., a single run contributes to a macroscopic fraction of the total displacement.

where the integral runs over the imaginary-axis Bromwich contour Γ in the complex q plane. Substituting this representation in Eq. (4.4), we find

$$P(X,N) = \frac{1}{2\pi i}\int_\Gamma dq\; e^{qX}\,[\hat{p}(q)]^N\;, \tag{4.18}$$

where

$$\hat{p}(q) = \int_{-\infty}^{\infty} dx\; e^{-qx} p(x)\,. \tag{4.19}$$

Plugging $W(v) = \alpha\,(1-v)^{\alpha-1}$ in Eq. (4.5), we first evaluate $p(x)$ and then compute $\hat{p}(q)$ using Eq. (4.19), yielding

$$\hat{p}(q) = {}_4F_3\left(\frac{1}{2},\frac{1}{2},1,1;\frac{d}{2},\frac{1+\alpha}{2},\frac{2+\alpha}{2};q^2\right)\;, \tag{4.20}$$

where ${}_pF_q(\alpha_1\,,\ldots\alpha_p;\beta_1\ldots\beta_q;q)$ denotes the generalized hypergeometric function, defined as

$${}_pF_q(\alpha_1\,,\ldots\alpha_p;\beta_1\ldots\beta_q;z) = \sum_{n=0}^{\infty}\frac{(\alpha_1)_n\ldots(\alpha_p)_n}{(\beta_1)_n\ldots(\beta_q)_n}\frac{z^n}{n!}\,, \tag{4.21}$$

where $(a)_n = \Gamma(a+n)/\Gamma(a)$ is the rising factorial. Thus, we find

$$P(X,N) = \frac{1}{2\pi i}\int_\Gamma dq\;\exp\left[qX + NS_{d,\alpha}(q)\right]\;, \tag{4.22}$$

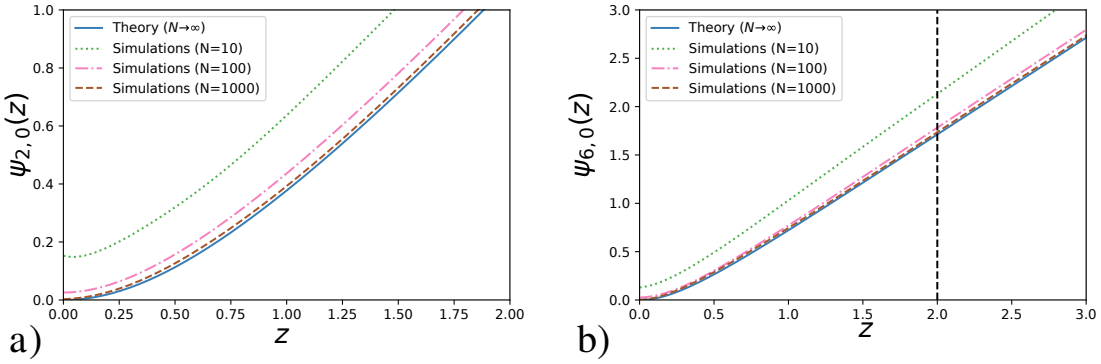

Figure 4.3: Rate function $\psi_{d,\alpha}(z)$ as a function of z for $d = 2$, $\alpha = 0$ (panel **a**) and $d = 6$, $\alpha = 0$ (panel **b**). The continuous blue lines correspond to the exact results in Eqs. (4.30) and (4.33). The dashed lines are obtained from importance sampling simulations for different system sizes N. In the case $d = 6$ (panel **b**), the rate function is singular at $z = 2$ (dashed vertical line), signaling a condensation transition. For the details of the numerical simulations, see Section 4.4.

where

$$S_{d,\alpha}(q) = \log\left[{}_4F_3\left(\frac{1}{2},\frac{1}{2},1,1;\frac{d}{2},\frac{1+\alpha}{2},\frac{2+\alpha}{2};q^2\right)\right] . \tag{4.23}$$

Note that this result is exact for any X and N. We are now interested in extracting the behavior of $P(X, N)$ in the limit of large N.

We first investigate the typical regime where $|X| \sim \sqrt{N}$. Rescaling $X = \sqrt{N}y$, where $y \sim \mathcal{O}(1)$ and changing variable $q \to q\sqrt{N}$ in Eq. (4.22) yields

$$P(X = y\sqrt{N}, N) = \frac{1}{2\pi i\sqrt{N}}\int_\Gamma dq\; e^{qy+NS_{d,\alpha}(q/\sqrt{N})} \approx \frac{1}{2\pi i\sqrt{N}}\int_\Gamma dq\; e^{qy+2q^2/(d(\alpha+1)(\alpha+2))} , \tag{4.24}$$

where we have used the small-argument expansion of $S_{d,\alpha}(q)$. Computing the Gaussian integral, we get

$$P(X, N) \approx \frac{1}{\sqrt{4\pi DN}}e^{-X^2/(4DN)} , \tag{4.25}$$

where $D = 2/(d(\alpha + 1)(\alpha + 2))$. Thus, in the typical regime where $|X| \sim \sqrt{N}$ the fluctuations of X are Gaussian, in agreement with the CLT.

We next investigate the large-deviation regime where $|X| \sim N$. We define the variable $z = X/N \sim \mathcal{O}(1)$. We will focus on the case $z > 0$. From (4.22), we get

$$P(X = zN, N) = \frac{1}{2\pi i}\int_\Gamma dq\;\exp\left[N\left(qz + S_{d,\alpha}(q)\right)\right] . \tag{4.26}$$

We first try to compute this integral by saddle-point approximation. The saddle-point condition $\frac{d}{dq}[qz + S_{d,\alpha}(q)] = 0$ gives the equation

$$z = -\frac{4q}{d(\alpha+1)(\alpha+2)}\frac{{}_4F_3\left(\frac{3}{2},\frac{3}{2},2,2;\frac{2+d}{2},\frac{3+\alpha}{2},\frac{4+\alpha}{2};q^2\right)}{{}_4F_3\left(\frac{1}{2},\frac{1}{2},1,1;\frac{d}{2},\frac{1+\alpha}{2},\frac{2+\alpha}{2};q^2\right)} , \tag{4.27}$$

where we have used the expression of $S_{d,\alpha}(q)$ in Eq. (4.23).

A detailed analysis of the function on the right-hand side shows that for $\nu < 2$, where $\nu = (d + 2\alpha - 1)/2$, the saddle-point equation admits a solution $q^*(z)$ for any z. This yields

$$P(X, N) \sim \exp\left[-N\psi_{d,\alpha}\left(\frac{X}{N}\right)\right], \tag{4.28}$$

where the rate function reads

$$\psi_{d,\alpha}(z) = -zq^*(z) - S_{d,\alpha}(q^*(z)). \tag{4.29}$$

Thus, for $\nu < 2$, $\psi_{d,\alpha}(z)$ does not display any singularities as a function of z, as predicted by the criterion presented in the previous section. For instance, for $d = 2$ and $\alpha = 0$, we find

$$\boxed{\psi_{2,0}(z) = \frac{1}{2}\left[\sqrt{1+4z^2} - 1 + \log\left(\frac{\sqrt{1+4z^2}-1}{2z^2}\right)\right].} \tag{4.30}$$

This function is shown in Fig. 4.3 and is in good agreement with numerical simulations, performed with importance-sampling techniques.

Interestingly, for $\nu > 2$, the saddle-point equation admits a solution only for $z < z_c$, where

$$\boxed{z_c = \frac{4}{d(\alpha+1)(\alpha+2)} \frac{{}_4F_3\left(\frac{3}{2}, \frac{3}{2}, 2, 2; \frac{2+d}{2}, \frac{3+\alpha}{2}, \frac{4+\alpha}{2}; 1\right)}{{}_4F_3\left(\frac{1}{2}, \frac{1}{2}, 1, 1; \frac{d}{2}, \frac{1+\alpha}{2}, \frac{2+\alpha}{2}; 1\right)}.} \tag{4.31}$$

Thus, for $z < z_c$, the integral is dominated by the solution $q^*(z)$ of the saddle point equation and thus the rate function $\psi_{d,\alpha}(z)$ is still given by Eq. (4.29). On the other hand, for $z > z_c$ the saddle-point equation does not admit a solution. Nevertheless, the integral is dominated by the value corresponding to $q^*(z_c) = -1$, yielding

$$\psi_{d,\alpha}(z) = z - S_{d,\alpha}(-1). \tag{4.32}$$

To summarize, we have obtained, for $\nu > 2$

$$\boxed{\psi_{d,\alpha}(z) = \begin{cases} -z\ q^*(z) - S_{d,\alpha}(q^*(z)) & \text{for } z < z_c \\ \\ z - S_{d,\alpha}(-1) & \text{for } z > z_c\,. \end{cases}} \tag{4.33}$$

Interestingly, in this case, the rate function displays a singularity at $z = z_c$. This signals the presence of the condensation transition. This exact result is shown in Fig. 4.3 for $d = 6$ and $\alpha = 0$ and is in good agreement with numerical simulations performed with a constrained Markov chain Monte Carlo technique (see Section 4.4 for the details).

Note that for simplicity we have presented this result for the fixed-N ensemble, i.e., in the case where the total number of running phases is fixed. One can compute exactly the rate function also for the fixed-T ensemble (see Article 7). Although the precise expression of the rate function is different for the fixed-T ensemble, the qualitative features (e.g., the presence of the phase transition) are the same.

It is also interesting to investigate the order of the phase transition. We say that the system undergoes a transition of order n if the n-th derivative of the rate function $\psi(z)$

is discontinuous at the critical point $z = z_c$. From the result in Eq. (4.33) it is possible to show that the order of the transition is

$$n = \begin{cases} \left\lceil \frac{\nu-1}{\nu-2} \right\rceil & \text{for } 2 < \nu < 3, \\ \\ 2 & \text{for } \nu > 3\,, \end{cases} \tag{4.34}$$

where $\lceil y \rceil$ indicates the smallest integer larger than or equal to y. Note that, for $2 < \nu < 3$, the order of the transition can be arbitrarily large and n diverges for $\nu \to 2$ ($n = \infty$ corresponds to a smooth rate function with no transition).

Investigating the rate function, we have identified the presence of a phase transition and we have computed its order. We have also anticipated that this is a condensation transition. We are now interested in characterizing precisely the mechanism of the transition.

4.2.2 Single-run marginal distribution

To understand how the phase transition takes place we need to investigate what a typical configuration of the system looks like in the two phases, for $z < z_c$ and $z > z_c$. To do so, we consider the probability distribution $p(x|X)$ of a single-run displacement x of the RTP in the x-direction, conditioned on the total displacement X. Note that since the N single-run displacements $x_1\,, \dots\,, x_N$ are i.i.d. random variables, we can choose x to be any of them, for instance $x = x_1$. In the fluid phase $X < X_c$ we expect that the different single-run displacements are roughly of the same order, and hence we expect $p(x|X)$ to be peaked around values $x \sim \mathcal{O}(1)$. On the other hand, for $X > X_c$, we have anticipated that one of these variables $x_1\,, \dots\,, x_N$ (randomly chosen) will absorb a finite fraction of the total displacement X. This would correspond to a "bump" in the large-x tail of $p(x|X)$ (see Fig. 4.4). The location of the bump is the typical size of the condensate, while its width describes the fluctuations of the condensate size. The area under the condensate bump is the probability that one particular running phase, say x_1, becomes the condensate. By symmetry, in the presence of a single condensate, we expect this area to be $1/N$.

The marginal PDF $P(x|X)$ can be simply related to the position distribution $P(X, N)$. Indeed, by definition, we have

$$p(x|X) = \frac{p(x) \int_{-\infty}^{\infty} dx_2 \, \dots \int_{-\infty}^{\infty} dx_N \, \left[\prod_{i=2}^{N} p(x_i)\right] \delta\left(X - x - \sum_{i=2}^{N} x_i\right)}{\int_{-\infty}^{\infty} dx_1 \, \dots \int_{-\infty}^{\infty} dx_N \, \left[\prod_{i=1}^{N} p(x_i)\right] \delta\left(X - \sum_{i=1}^{N} x_i\right)}\,. \tag{4.35}$$

Using the definition of $P(X, N)$ in Eq. (4.4), this expression can be rewritten as

$$p(x|X) = \frac{p(x) P(X - x, N-1)}{P(X, N)}\,. \tag{4.36}$$

Thus, we can use the results of the previous section for $P(X, N)$ and the large-x asymptotic behavior of $p(x)$, given in Eq. (4.13) to compute $p(x|X)$.

In particular, for $X < X_c$, we obtain

$$p(x|X) \sim x^{-\nu} e^{-x/\xi}\,, \tag{4.37}$$

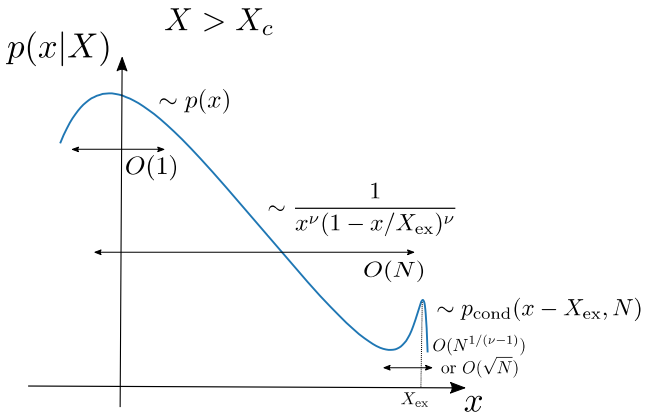

Figure 4.4: Schematic representation of the single-run marginal distribution $p(x|X)$ as a function of the single-run size x in the condensed phase $X > X_c$. A bump appears in the tail of the distribution for $x \sim \mathcal{O}(N)$.

where $\nu = (d + 2\alpha - 1)/2$ and

$$\xi = \frac{1}{1 + q^*(z)} \,. \tag{4.38}$$

We recall that $q^*(z)$ is the unique solution of the saddle-point equation (4.27). Thus, below the transition, the marginal PDF $p(x|X)$ decays exponentially over the typical distance ξ. As a consequence, the N running phases are typically of the same order, with $x \sim \xi$. As $z \to z_c$ from below, the solution $q^*(z)$ tends to the limit value -1 and hence the typical length ξ diverges.

For $X > X_c$, and for large x, we obtain

$$p(x|X) \sim \frac{1}{x^\nu (1 - x/X_{\rm ex})^\nu} \,, \tag{4.39}$$

where $X_{\rm ex} = X - X_c$ is the excess displacement above the critical value X_c. Thus, above the transition, the marginal distribution $p(x|X)$ has a power-law tail. Note however that the expression in Eq. (4.39) breaks down for $x \to X_{\rm ex}$. Indeed, at $x \approx X_{\rm ex}$ a bump appears in the tail of $p(x|X)$, with

$$p(x|X) \approx \frac{1}{N} p_{\rm cond}(x - X_{\rm ex}, N) \,, \tag{4.40}$$

where the function $p_{\rm cond}(x - X_{\rm ex}, N)$ is normalized to unity and describes the shape of the bump (i.e., the probability of the fluctuations of the condensate size around the value $X_{\rm ex}$). Moreover, the width of the bump vanishes relative to its location $X_{\rm ex}$ for large N. See Fig. 4.4 for a schematic representation of $p(x|X)$ in the condensed phase. Since $p_{\rm cond}(y, N)$ is normalized to unity, the area under the bump is precisely $1/N$, indicating the presence of a single condensate.

The precise expression of $p_{\rm cond}(y, N)$ depends on ν. For $\nu > 3$, the fluctuations of the condensate size scale as $\sqrt{N}$ and $p_{\rm cond}(y, N)$ is Gaussian around its peak

$$p_{\rm cond}(y, N) \approx \frac{1}{\sqrt{4 a_{d,\alpha} \pi N}} e^{-y^2/(4 a_{d,\alpha} N)} \,, \tag{4.41}$$

where $a_{d,\alpha}$ is a positive constant. On the other hand, for $2 < \nu < 3$, the condensate size has anomalous fluctuations of order $N^{1/(\nu-1)}$ and

$$p_{\rm cons}(y,N) \approx \frac{1}{N^{1/(\nu-1)}} V_\nu\left(\frac{y}{N^{1/(\nu-1)}}\right) , \tag{4.42}$$

where

$$V_\nu(y) = \frac{1}{\pi}\int_0^\infty dr\; e^{b_{d,\alpha}\sin(\pi\nu/2) r^{\nu-1}} \cos\left[b_{d,\alpha}\cos(\pi\nu/2) r^{\nu-1} + yr\right] , \tag{4.43}$$

and $b_{d,\alpha}$ is a positive constant. This function $V_\nu(y)$ has asymptotic behaviors

$$V_\nu(y) \sim \begin{cases} |y|^{-\nu} & \text{for } y \to -\infty \\ \\ y^{(3-\nu)/(2(\nu-2))} e^{-c\; y^{(\nu-1)/(\nu-2)}} & \text{for } y \to \infty\,, \end{cases} \tag{4.44}$$

where c is a positive constant. One can check that $V_\nu(y)$ is correctly normalized to unity.

Thus, for $\nu > 3$ and $z > z_c$, the system is in a *normal condensate phase*, where the condensate fluctuations are Gaussian and scale as $\sqrt{N}$. On the other-hand for $2 < \nu < 3$ and $z > z_c$, the system is in an *anomalous condensate phase*, where the condensate fluctuations scale as $N^{1/(\nu-1)}$ and are described by the function $V_\nu(y)$. These different phases are shown in Fig. 4.5 in the (α, z) phase space for $d = 3$ (left panel) and $d = 7$ (right panel). The presence of two different condensate phases (normal and anomalous) was already observed in the context of mass-transport models [193].

Finally, we identify the order parameter that characterizes the transition: the *participation ratio* $Y_2(z)$, defined as

$$Y_2(z) = \left\langle \frac{\sum_{i=1}^N x_i^2}{\left(\sum_{i=1}^N x_i\right)^2} \right\rangle_z , \tag{4.45}$$

where the symbol $\langle \ldots \rangle_z$ indicates the average over trajectories of the system with $z = X/N$ fixed. In the fluid phase, the variables x_i are of order one and hence the numerator is of order N. Since the denominator is $X^2 \sim \mathcal{O}(N^2)$, the participation ratio vanishes as $1/N$. Thus, for $z < z_c$, we obtain (in the thermodynamic limit)

$$Y_2(z) = 0\,. \tag{4.46}$$

On the other hand, for $z > z_c$, the numerator is dominated by the condensate, corresponding to a single-run displacement of length $X_{\rm ex} = X - X_c \sim \mathcal{O}(N)$. Thus, in the condensed phase, the numerator and the denominator are of the same order and we obtain

$$Y_2(z) = \frac{(z-z_c)^2}{z^2}\,, \tag{4.47}$$

valid for $z > z_c$. The participation ratio is thus zero in the fluid phase and non-zero in the condensed phase. Note that in this case $Y_2(z)$ is a continuous function of z.

To summarize, for the speed distribution in Eq. (4.16) a condensation transition occurs for $\nu > 2$ at a critical value z_c of the control parameter $z = X/N$. For $z < z_c$, the system is in the fluid phase and all runs are of order one. For $z > z_c$, a single running phase absorbs the whole excess displacement $X_{\rm ex} = X - X_c \sim \mathcal{O}(N)$ with Gaussian

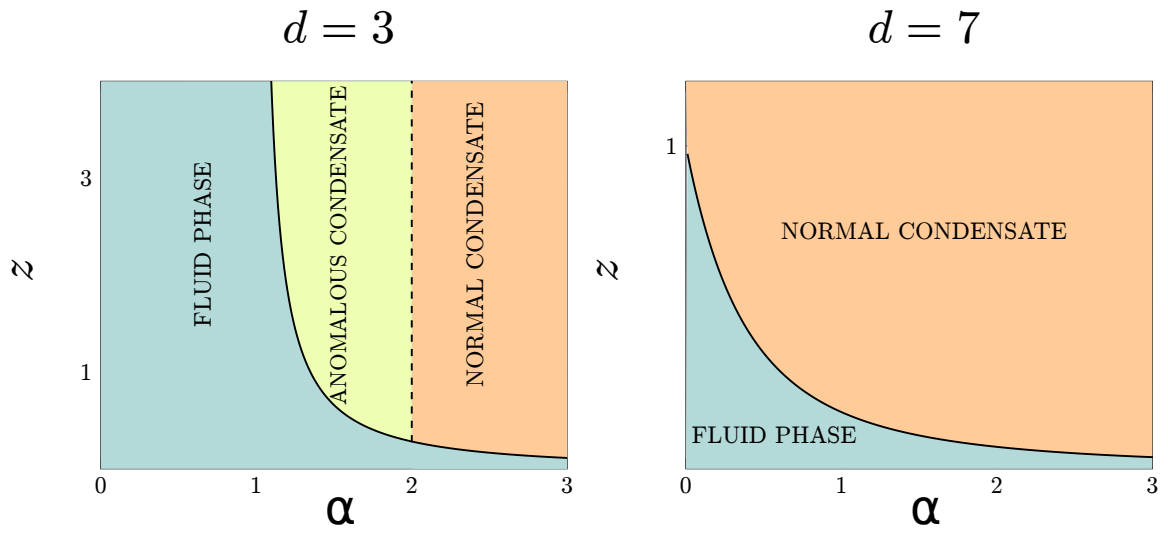

Figure 4.5: Phase diagram in the (α, z) plane for $d = 3$ (left panel) and $d = 7$ (right panel). For $d = 3$ there are three possible phases. In the fluid phase the different running phases contribute to the total displacement by roughly the same amount. In the condensate phase, a condensate appears and absorbs a macroscopic fraction of the total displacement X. The fluctuations of the condensate can either be of order $\sqrt{N}$ and Gaussian (normal condensate phase) or of order $N^{1/(\nu-1)}$ and described by the function $V_\nu(y)$ in Eq. (4.43) (anomalous condensate phase).

fluctuations of order $\sqrt{N}$ for $\nu > 3$ and anomalous fluctuations of order $N^{1/(\nu-1)}$ for $2 < \nu < 3$. Interestingly, in the anomalous condensate phase, the order n of the transition depends continuously on ν (with $n \geq 3$), while in the normal condensate phase we find $n = 2$. We also identify the order parameter $Y_2(z)$ associated with the transition. In the next section, we will investigate an RTP model for which the condensation transition is of order one.

4.3 One dimension and Gaussian speed: first-order phase transition

In the previous section, we have shown that for a large class of RTP models, a second (or higher) order transition appears in the large deviation regime of the position of the particle, for $X \sim \mathcal{O}(N)$, where N is the number of running phases. In this section, we present a model for which a first-order transition appears for $X \sim \mathcal{O}(N^{3/4})$. The results presented in this section are derived in detail in Article 8.

We consider a single RTP in one dimension with constant tumbling rate γ. We assume that the speed of the particle is independently drawn, after each tumbling event, from the half-Gaussian distribution

$$W(v) = \sqrt{\frac{2}{\pi\sigma^2}} \exp\left[-v^2/(2\sigma^2)\right] \theta(v)\,. \tag{4.48}$$

To simplify the notation, we consider $\gamma = \sigma^2 = 1$. Then, using the expression in Eq. (4.5), we find that the distribution of the displacement of the particle during a single run is

$$p(x) = \frac{1}{\sqrt{2\pi}} \int_0^\infty d\tau \; \frac{1}{\tau} e^{-\tau - x^2/(2\tau^2)} \,, \tag{4.49}$$

for $x > 0$. From this expression, we obtain the first two moments of the displacement x, which are $\langle x \rangle = 0$ and $\langle x^2 \rangle = 2$. Note that since the distribution $p(x)$ is an even function of x, also the PDF $P(X, N)$ of the final position of the particle will be symmetric around $X = 0$. Thus, we will just consider the case $X > 0$, as the complementary case can be obtained by symmetry. Moreover, in the limit of large x, the integral in Eq. (4.49) can be evaluated by saddle-point approximation and one obtains

$$p(x) \approx \frac{1}{\sqrt{3}|x|^{1/3}} e^{-3|x|^{2/3}/2} \,, \tag{4.50}$$

valid for large $|x|$. Thus, the jump distribution $p(x)$ has a stretched-exponential (also known as semiexponential) tail for large $|x|$, with exponent $2/3$. This satisfies the criterion for condensation presented in Section 4.1 and thus we expect a condensation transition to occur.

Computing the distribution of the position X of the RTP consists in finding the distribution of the sum of N i.i.d. stretched-exponential random variables. Interestingly, the general case where $p(x) \sim e^{-a|x|^b}$, with $a > 0$ and $0 < b < 1$ was first studied by Nagaev [206, 207], who found that the distribution of X admits a nontrivial large-deviation regime for $X \sim \mathcal{O}(N^{1/(2-\beta)})$ (where $a = 3/2$ and $\beta = 2/3$ in our case). Moreover, in a recent mathematical paper, the rate function associated with this large-deviation regime was derived [208]. In this section, we present an alternative derivation of this result. Moreover, we analyze in detail the mechanism of the phase transition, by studying the marginal probability of a single variable and identifying the order parameter associated with the transition. These aspects were not investigated in [208].

To identify the scale at which the transition occurs, let us consider different regimes, depending on the final position X of the particle. In the typical regime $X \sim \mathcal{O}(\sqrt{N})$, by applying the CLT, we have

$$P(X, N) \sim e^{-X^2/(4N)} \,. \tag{4.51}$$

On the other hand, in the large-deviation regime where the position scales linearly with N, we find that the total displacement is completely dominated by a single run. Correspondingly, we get

$$P(X, N) \approx N p(X) \,, \tag{4.52}$$

where $p(X)$ is the single-displacement distribution, given in Eq. (4.49), and the factor N comes from the fact that any of the N i.i.d. displacements can be the dominant one. Using the asymptotic expression of $p(X)$, given in Eq. (4.50), we obtain

$$P(X, N) \sim e^{-(3/2)X^{2/3}} \,. \tag{4.53}$$

In order to identify the correct scale at which the transition occurs, we match the Gaussian distribution in Eq. (4.50) with the stretched-exponential tail in Eq. (4.53)

$$e^{-X^2/(4N)} \sim e^{-(3/2)X^{2/3}} \,, \tag{4.54}$$

yielding $X \sim \mathcal{O}(N^{3/4})$.

4.3.1 Position distribution

Let us recall the relation in Eq. (4.18), which reads in this case

$$P(X,N) = \frac{1}{2\pi i}\int_\Gamma dq\; e^{qX + N\log(\hat{p}(q))}\,, \tag{4.55}$$

where, using the expression for $p(x)$ in Eq. (4.49),

$$\hat{p}(q) = \sqrt{\pi}\frac{e^{-1/(2q^2)}}{\sqrt{-2q^2}}\;\mathrm{erfc}\left[\frac{1}{\sqrt{-2q^2}}\right]. \tag{4.56}$$

It is useful to give the asymptotic behaviors of this function $\hat{p}(q)$ in the limit of small q, which read

$$\hat{p}(q) \approx \begin{cases} 1+q^2 & \text{if } \mathrm{Im}(q) < 0\,, \\ \\ 1+q^2+\frac{\sqrt{2\pi}}{\sqrt{-q^2}}e^{-1/(2q^2)} & \text{if } \mathrm{Im}(q) > 0\,. \end{cases} \tag{4.57}$$

The additional non-analytic term for $\mathrm{Im}(q) > 0$ turns out to play a central role in the computation of the rate function.

By analyzing this expression in Eq. (4.55) for $X \sim \mathcal{O}(\sqrt{N})$ (typical regime) and $X \sim \mathcal{O}(N)$ (extreme large deviation regime), one obtains the two asymptotic results in Eqs. (4.51) and (4.52) (see Article 8 for the details). Here, we would like to investigate the intermediate matching regime, which we call *anomalous large-deviation regime*, corresponding to $X \sim \mathcal{O}(N^{3/4})$.

To proceed, it is useful to define the variable $z = X/N^{3/4} > 0$, which we assume to be of order one. This variable z is the control parameter of the transition. Eq. (4.55) can then be written as

$$P(X = zN^{3/4}, N) = \frac{1}{2\pi i}\int_\Gamma dq\; e^{qzN^{3/4} + N\log[\hat{p}(q)]}\,, \tag{4.58}$$

where the contour of integration Γ is shown in Fig. 4.6. By using the small-q expansion in Eq. (4.57), we obtain

$$P(X = zN^{3/4}, N) \simeq \frac{1}{2\pi i}\int_{\Gamma^-} dq\; e^{qzN^{3/4}+Nq^2} + \frac{1}{2\pi i}\int_{\Gamma^+} dq\; e^{qzN^{3/4}+Nq^2+N\sqrt{2\pi}e^{-1/(2q^2)}/\sqrt{-q^2}}\,, \tag{4.59}$$

where we have split the contour Γ into the two parts Γ^+ (for $\mathrm{Im}(q) > 0$) and Γ^- (for $\mathrm{Im}(q) < 0$). We then use the expansion

$$\exp\left[N\sqrt{2\pi}\;\frac{e^{-1/(2q^2)}}{\sqrt{-q^2}}\right] \;\simeq\; 1 + N\sqrt{2\pi}\;\frac{e^{-1/(2q^2)}}{\sqrt{-q^2}}\,, \tag{4.60}$$

to rewrite the PDF $P(X,N)$ as the sum of a Gaussian term $P_G(X,N)$ and an "anomalous" term $P_A(X,N)$

$$P(X,N) \simeq P_G(X,N) + P_A(X,N)\,. \tag{4.61}$$

The Gaussian term reads

$$P_G(X,N) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} dq\; e^{qzN^{3/4}+Nq^2} = \frac{1}{2\sqrt{\pi N}}\;e^{-\sqrt{N}z^2/4}\,, \tag{4.62}$$

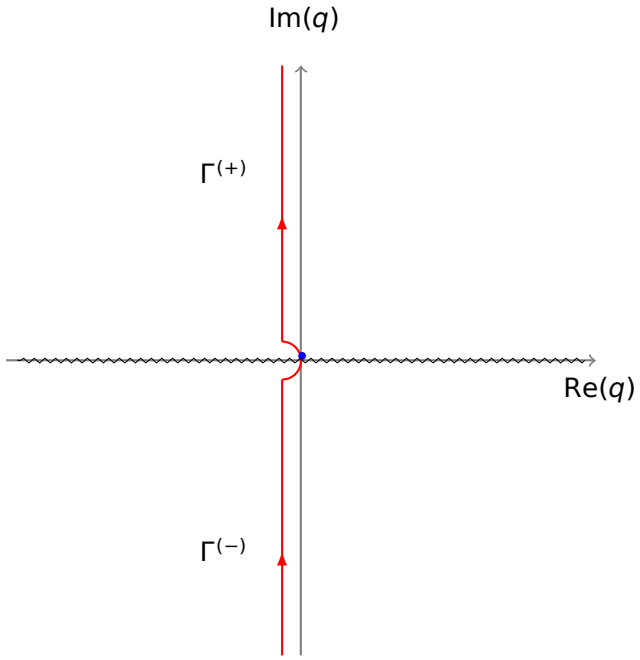

Figure 4.6: Analytic structure of $\hat{p}(q)$, given in Eq. (4.56), in the complex-q plane. The function $\hat{p}(q)$ has two branch cuts (wiggled lines), in the intervals $]-\infty, 0[$ and $]0, \infty[$ in the real axis. The red oriented line represents the Bromwich contour Γ that we use to compute the integral in Eq. (4.55).

while the anomalous term is given by

$$P_A(X, N) = \frac{N}{i} \int_{\Gamma^+} dq \, \frac{1}{\sqrt{-q^2}} \, e^{qzN^{3/4} + Nq^2 - 1/(2q^2)} . \tag{4.63}$$

To compute the complex integral in Eq. (4.63), we perform the change of variable $q \to s = qN^{1/4}$, yielding

$$P_A(X, N) = \frac{N}{i} \int_{\Gamma^+} ds \, \frac{1}{\sqrt{-s^2}} e^{\sqrt{N} G_z(s)} , \tag{4.64}$$

where

$$G_z(s) = zs + s^2 - \frac{1}{2s^2} . \tag{4.65}$$

Interestingly, the integral in Eq. (4.64) can be computed via saddle point if $z > z_\ell = 4(2/3)^{3/4} \approx 2.95115$. Indeed, the saddle point equation

$$G'_z(s) = z + 2s + \frac{1}{s^3} = 0 \tag{4.66}$$

admits a real solution s^* only for $z > z_\ell$. Hence, for $z > z_\ell$, we obtain

$$P_A(X = zN^{3/4}, N) \sim e^{-\sqrt{N}\chi(z)} , \tag{4.67}$$

where

$$\chi(z) = G_z(s^*(z)) = zs^*(z) + [s^*(z)]^2 - \frac{1}{2\,[s^*(z)]^2} . \tag{4.68}$$

Here $s^*(z)$ is the largest negative root of the saddle point equation in (4.66). An expression for $s^*(z)$ can be obtained by simply solving the cubic equation, yielding

$$s^*(z) = -\frac{z}{8} - \frac{1}{8}\sqrt{\frac{64(2/3)^{1/3}}{a(z)} + 4(2/3)^{1/3}a(z) + z^2} \tag{4.69}$$
$$+ \frac{\sqrt{2}}{8}\sqrt{z^2 + z^3\left(\frac{64(2/3)^{1/3}}{a(z)} + 4(2/3)^{1/3}a(z) + z^2\right)^{-1/2} - \left(\frac{2}{3}\right)^{1/3}\left(2a(z) + \frac{32}{a(z)}\right)},$$

where

$$a(z) = \left(9z^2 + \sqrt{81z^4 - 6144}\right)^{1/3}. \tag{4.70}$$

Thus, we find that the distribution of X can be written as the sum of the following two terms

$$P(X = zN^{3/4}, N) \sim e^{-\sqrt{N}z^2/4} + e^{-\sqrt{N}\chi(z)}. \tag{4.71}$$

The competition between these two terms in the large-N regime is at the origin of the first-order condensation transition of this model. Indeed, only the term with the smallest exponent dominates for large N, yielding

$$P(X = zN^{3/4}, N) \sim e^{-\sqrt{N}F(z)}, \tag{4.72}$$

where

$$F(z) = \min\left[\frac{z^2}{4}, \chi(z)\right]. \tag{4.73}$$

We recall that we have an explicit expression for $\chi(z)$ (given in Eq. (4.68)) only for $z > z_\ell$. By analyzing this function $\chi(z)$, we find that $\chi(z) < z^2/4$ only for $z > z_c = 2^{7/4} \approx 3.36358$ (note that $z_c > z_\ell$). Thus, we know the expression of $\chi(z)$ in the relevant region $z > z_c$. As a consequence, we obtain

$$\boxed{F(z) = \begin{cases} z^2/4 & \text{for } z < z_c, \\ \\ \chi(z) & \text{for } z > z_c. \end{cases}} \tag{4.74}$$

This rate function $F(z)$ is shown in Fig. 4.7 and is in good agreement with numerical simulations. Interestingly, the PDF $P(X, N)$ remains Gaussian in the large deviation regime, up to $X_c = z_c N^{3/4}$, well outside the range of validity of the CLT. Moreover, the rate function $F(z)$ has a first-order singularity at $z = z_c$, meaning that its derivative $F'(z)$ is discontinuous at the critical point.

The expression of $P(X, N)$ in Eq. (4.71) provides a nice interpretation of the mechanism of the transition. In analogy with first-order transition in classical thermodynamics, the transition arises from the competition between two different phases: the fluid phase (described by the Gaussian weight) and the condensed phase (associated with the anomalous term). Each phase is also associated with a rate function ($z^2/4$ for the fluid phase and $\chi(z)$ for the condensed phase), which plays the role of the free energy for nonequilibrium systems. At a given value of the control parameter z, the system will be in the phase with a lower rate function. Thus, the critical point z_c is by definition the value for which the two rate functions are equal.

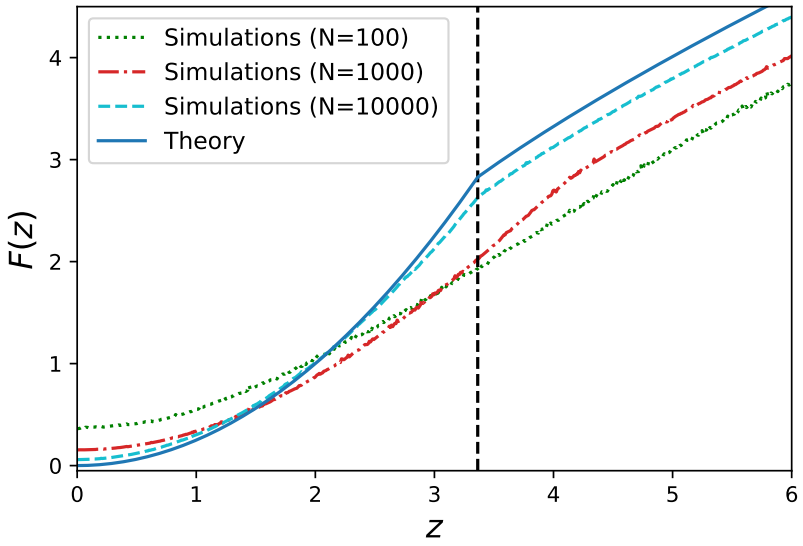

Figure 4.7: Rate function $F(z) \approx -\log[P(X,N)]/\sqrt{N}$ of the position distribution as a function of the rescaled position $z = X/N^{3/4}$. The continuous blue line indicates the exact result in Eq. (4.74), which has a first-order singularity for $z = z_c = 2^{7/4}$ (vertical dashed line). The dashed curves indicate the results of numerical simulations at different system sizes.

The asymptotic behaviors of the function $\chi(z)$ are given by

$$\chi(z) \approx \begin{cases} \sqrt{6} & \text{for } z \to z_\ell \\ \\ \frac{3}{2} z^{2/3} - z^{-2/3} & \text{for } z \to \infty \end{cases} \tag{4.75}$$

Thus, in the limit $z \to \infty$, we find

$$P(X = zN^{3/4}, N) \sim e^{-\sqrt{N}(3/2)z^{2/3}}, \tag{4.76}$$

smoothly connecting to the expression for $P(X,N)$ in the regime $X \sim \mathcal{O}(N)$ (see Eq. (4.52)).

Since we have observed that the rate function $F(z)$ has a first-order singularity at $z = z_c$, it is relevant to ask what is the precise mechanism of the transition. Moreover, it is worth investigating whether or not the transition is associated with a discontinuity of the order parameter, in analogy with thermodynamic first-order transitions.

4.3.2 Single-run marginal distribution

As done in the previous section, it is relevant to investigate the marginal PDF $p(x|X)$ of the displacement of the RTP during a single run, conditioned on the total displacement X. The relation between $p(x|X)$ and $P(X,N)$ in Eq. (4.36) remains valid in this case, yielding

$$p(x|X) = \frac{p(x)P(X-x, N-1)}{P(X,N)}. \tag{4.77}$$

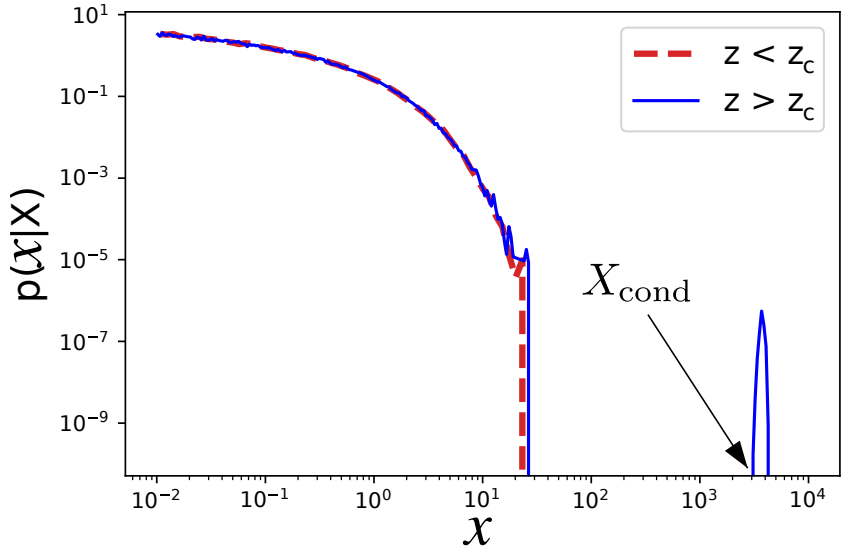

Figure 4.8: Single-run marginal distribution $p(x|X)$ as a function of x, obtained from numerical simulations of RTP trajectories with $N = 10^4$ running phases and for different values of $z = X/N^{3/4}$. When $z < z_c$ (dashed red line) the distribution is concentrated around values of x of order one (fluid phase). For $z > z_c$ (continuous blue line) a bump appears in the tail of the distribution at $X_{\rm cond} \sim \mathcal{O}(N^{3/4})$, indicating the presence of a condensate (condensed phase). For the details on how the numerical simulations have been performed, see Section 4.4.

Note that since the total displacement X is fixed and the N displacements are i.i.d., the mean value of $p(x|X)$ is simply given by

$$\int_{-\infty}^{\infty} dx\; x\; p(x|X) = \frac{X}{N}\,. \tag{4.78}$$

Thus, in the intermediate regime $X \sim \mathcal{O}(N^{3/4})$, this first moment vanishes for large N as $N^{-1/4}$.

As in the previous section, we expect that below the transition, in the fluid phase, the marginal distribution $p(x|X)$ is centered around values of x of order one. This corresponds to the fact that the different running phases are roughly of the same order of magnitude. On the other hand, for $X > X_c = z_c N^{3/4}$, we expect a condensate, i.e., a single displacement of length $X_{\rm cond}$ of order $N^{3/4}$ to emerge. This would appear in the tail of the distribution $p(x|X)$ as a "bump" located at $x \sim X_{\rm cond}$. The results of numerical simulations, shown in Fig. 4.8, confirm our expectation. We next investigate analytically the behavior of the marginal distribution $p(x|X)$ in the large-deviation regime where $X \sim \mathcal{O}(N^{3/4})$.

We will investigate two different regimes, depending on the value of x. If $x \sim \mathcal{O}(1)$, it is easy to show from Eq. (4.77) that, at leading order in N,

$$p(x|X) \approx p(x)\,, \tag{4.79}$$

where $p(x)$ is the unconstrained marginal distribution, given in Eq. (4.49). This result is valid both in the fluid and in the condensed phases. This can be observed in Fig. 4.8, where the curves for $z < z_c$ and $z > z_c$ collapse into the same curve when $x \sim \mathcal{O}(1)$.

Since the behavior of $p(x|X)$ is always the same when $x \sim \mathcal{O}(1)$, one needs to investigate the large-x tail to unveil the phase transition. Since we expect a condensate to appear for $z > z_c$ with typical size $X_{\rm cond} \sim O(N^{3/4})$, we consider the large-deviation tail of $p(x|X)$ where $x \sim \mathcal{O}(N^{3/4})$. By substituting the expressions for $p(x)$ and $P(X, N)$, respectively given in Eqs. (4.50) and (4.72), into the expression for $p(x|X)$ in Eq. (4.77) we are able to rewrite $p(x|X)$ in the large deviation form

$$p(x|X) \sim \exp\left[-\sqrt{N}\psi_z\left(\frac{x}{N^{3/4}}\right)\right], \tag{4.80}$$

where $z = X/N^{3/4}$ and

$$\psi_z(y) = \frac{3}{2}y^{2/3} + F(z-y) - F(z)\,. \tag{4.81}$$

We recall that $F(z)$ is the rate function of $P(X, N)$ given in Eq. (4.73). Note that the variable $y = x/N^{3/4}$ is the rescaled single-run displacement, while the variable $z = X/N^{3/4}$ is the rescaled total displacement.

We thus need to investigate the behavior of the rate function $\psi_z(y)$ as a function of y and for different values of z. Indeed, if $\psi_z(y) > 0$ for some value of y, the corresponding single-run displacement $x = yN^{3/4}$ will be observed with exponentially small probability, i.e., with probability which decays as $e^{-c\sqrt{N}}$ with $c > 0$. Instead, if for some value $y > 0$ we find $\psi_z(y) = 0$, the corresponding displacement can be observed in a typical RTP configuration as a condensate.

The behavior of $\psi_z(y)$ is shown in detail in Fig. 4.9. Below the transition, i.e., for $z < z_c$, we observe that $\psi_z(y) > 0$ for all y and thus no condensate can appear. On the other hand for $z > z_c$ there exist a unique point $y^*(z) > 0$ for which $\psi_z(y^*(z)) = 0$. Correspondingly, for $z > z_c$ a bump appears at $x \approx y^*(z)N^{3/4}$, indicating the presence of a condensate of size

$$X_{\rm cond} = y^*(z)N^{3/4}\,. \tag{4.82}$$

One can show that the variable $y^*(z)$ satisfies

$$\psi_z'(y^*) = (y^*)^{-1/3} - F'(z-y^*) = 0\,, \tag{4.83}$$

where $F'(z) = dF(z)/dz$ and $F(z)$ is given in Eq. (4.73). This equation can be solved analytically, for instance by using Mathematica. We find

$$y^*(z) = z\left[\frac{z^{4/3}}{8} + \frac{1}{8}\left(\frac{64\,(2/3)^{1/3}}{a(z)} + 4\,(2/3)^{2/3}\,a(z)\;z^{4/3} + z^{8/3}\right)^{1/2}\right. \tag{4.84}$$
$$\left. + \frac{1}{2}\left(\frac{4\,(2/3)^{1/3}}{a(z)} + \frac{a(z)z^{4/3}}{2^{4/3}3^{2/3}} - \frac{z^{8/3}}{8} - \frac{z^4}{8\sqrt{\frac{64(2/3)^{1/3}}{a(z)} + 4\left(\frac{2}{3}\right)^{2/3}a(z)\;z^{4/3} + z^{8/3}}}\right)^{1/2}\right]^{-3},$$

where $a(z)$ is given in Eq. (4.70). From this expression we find $y*(z_c) = z_c/2$. Thus, at the critical value $z = z_c$ the size of the condensate jumps to the value $X_{\rm cond} = X/2$,

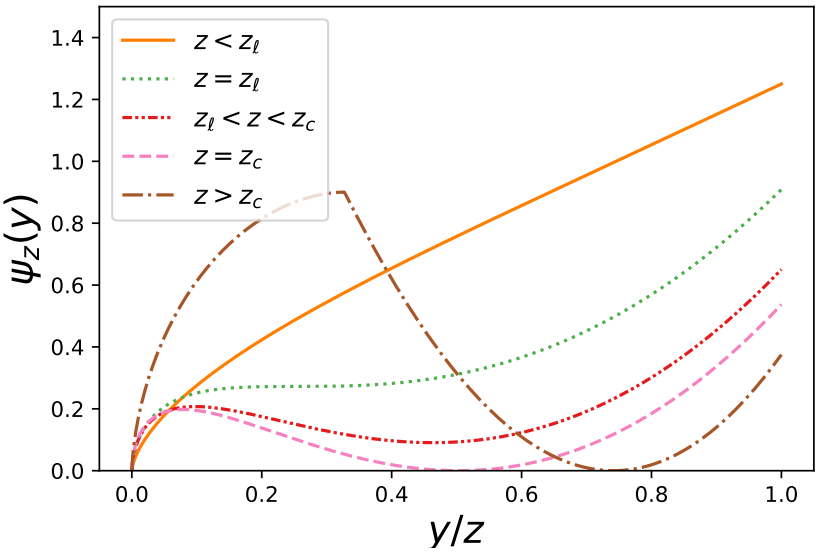

Figure 4.9: The rate function $\psi_z(y)$ as a function of y/z, for different values of z. The different curves correspond to the exact result in Eq. (4.81). For $z < z_\ell$, $\psi_z(y)$ is positive and monotonic for $y > 0$. At $z = z_\ell$ a minimum appears at some value $y^* > 0$, with corresponding rate function value $\psi_z(y^*) > 0$. Increasing z further, the value $\psi_z(y^*)$ decreases until at the critical point $z = z_c$, one finds $\psi_z(y^*) = 0$. For $z > z_c$, there always exists a unique value $y^* > 0$ for which $\psi_z(y^*) = 0$.

meaning that just above the transition the condensate is responsible for half of the total displacement X. This is quite different from what was observed in the previous section, where $X_{\rm cond} = X - X_c$ (and thus $X_{\rm cond} = 0$ at the critical point).

It is also interesting to observe that a local minimum of the rate function appears for $z > z_\ell$ at $y = y^*(z)$, where $y^*(z)$ is given in Eq. (4.84). However, for $z_\ell < z < z_c$, one has $\psi_z(y^*(z)) > 0$ and thus no condensate is observed. We can interpret this result as follows. For $z < z_\ell$ the only stable phase is the fluid one. For $z_\ell < z < z_c$, both the fluid and the condensate phases are in principle possible, but the condensate phase is metastable since its associated rate function $\psi_z(y^*(z))$ is positive. Increasing z, the value of the rate function $\psi_z(y^*(z))$ decreases, until at $z = z_c$ one has $\psi_z(y^*(z)) = 0$ and the whole system jumps to the condensate phase. This mechanism of the phase transition resembles that of thermodynamic first-order transitions, with the rate function $\psi_z(y)$ playing the role of the free energy.

Finally, we also investigate the behavior of the participation ratio $Y_2(z)$, introduced in the previous section, which is one of the possible order parameters for this transition[3]. This is defined as

$$Y_2(z) = \left\langle \frac{\sum_{i=1}^N x_i^2}{\left(\sum_{i=1}^N x_i\right)^2} \right\rangle_z . \tag{4.85}$$

In the fluid phase, the numerator in Eq. (4.85) scales as N, since $x_i^2 \sim \mathcal{O}(1)$ for all i,

[3]Another equivalent order parameter is the condensate fraction $m_c = X_{\rm cond}/X$, defined as the fraction of the total displacement X which is carried by the condensate.

while the denominator is $X^2 \sim \mathcal{O}(N^{3/2})$. Thus, for $z < z_c$, we expect the participation ratio to vanish as $1/\sqrt{N}$. On the other hand, in the condensed phase, the numerator is dominated by the condensate $X_{\rm cond} = y^*(z)N^{3/4}$ and thus

$$Y_2(z) \approx \frac{X_{\rm cond}^2}{X^2} = \left(\frac{y^*(z)}{z}\right)^2 . \tag{4.86}$$

To summarize, in the thermodynamic limit, we have

$$\boxed{Y_2(z) = \begin{cases} 0 & \text{for } z < z_c\,, \\ \\ \left(y^*(z)/z\right)^2 & \text{for } z > z_c\,, \end{cases}} \tag{4.87}$$

where $y^*(z)$ is the unique solution of Eq. (4.83). This exact result is shown in Fig. 4.10 and is in good agreement with numerical simulations. Interestingly, we observe that the order parameter $Y_2(z)$ is discontinuous at the critical value $z = z_c$. One can also show that, in the region $z > z_c$, the participation ratio has asymptotic behaviors

$$Y_2(z) \simeq \begin{cases} 1/4 + 2^{-7/4}(z - z_c) & \text{for } z \to z_c^+, \\ \\ 1 - 4z^{-4/3} & \text{for } z \to \infty\,. \end{cases} \tag{4.88}$$

Just above the transition, the participation ratio approaches the limit value $1/4$, corresponding to configurations in which the condensate contains half of the total displacement X (i.e., $X_{\rm cond} \approx X/2$). On the other hand, for large z, $Y_2(z)$ tends to the limit value one, meaning that the condensate absorbs the whole displacement X (i.e., $X_{\rm cond} \approx X$).

To summarize, we have shown that, for a simple one-dimensional RTP model with Gaussian speed, a first-order condensation transition can be observed in the large deviations of the position of the particle. This transition is signaled by a first-order discontinuity in the rate function of the position. We have characterized the mechanism of the transition and we have identified the corresponding order parameter, showing that it is discontinuous at the critical point. The results of these sections apply to the general problem of computing the distribution of the sum of several i.i.d. stretched exponential random variables.

4.4 Sampling rare events

In this last section, we present the details of the numerical simulations that we have performed to check the analytical results of the previous sections. To study the large-deviation regime, we employ a constrained Markov chain Monte Carlo (MCMC) algorithm, similar to the ones used in [183, 210, 211].

An RTP configuration $\mathcal{C}$ of N steps is described by the $3N$ random variables $\mathcal{C} = \{(\tau_i, v_i, u_i)\} = \{(\tau_1, v_1, u_1), \dots (\tau_N, v_N, u_N)\}$, where τ_i is the time of the i-th running phase, v_i is the running speed, and u_i is the x-component of the d-dimensional unit

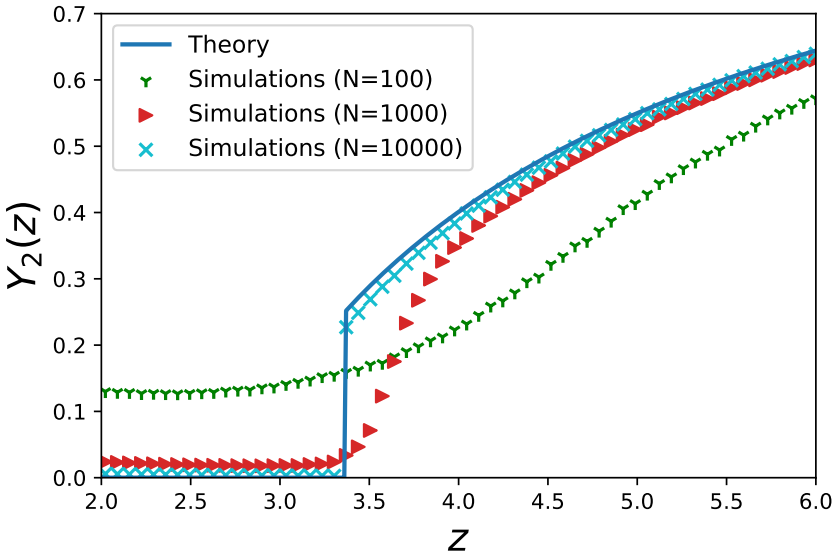

Figure 4.10: The participation ratio $Y_2(z)$ as a function of $z = X/N^{3/4}$. The continuous blue line corresponds to the exact result in Eq. (4.87), while the symbols are obtained from numerical simulations at different values of the system size N.

vector that represents the direction of the RTP (with $-1 \leq u_i \leq 1$). The probability weight associated with each configuration is

$$P(\mathcal{C}) = \prod_{i=1}^{N} p(\tau_i) W(v_i) f_d(u_i)\,, \tag{4.89}$$

where $p(\tau) = \gamma e^{-\gamma\tau}$ is the distribution of the running times, $W(v)$ is the speed distribution, and $f_d(u)$ is given in Eq. (4.6). The position X of the particle for an N-steps configuration $\mathcal{C}$ is given by

$$X(\mathcal{C}) = \sum_{i=1}^{N} \tau_i v_i u_i\,. \tag{4.90}$$

Thus, by standard sampling techniques one can directly sample the N couples $\{(\tau_i, v_i, u_i)\}$ and from this obtain a realization of the position X. Repeating this procedure several times, one can build a histogram that approximates the PDF $P(X, N)$ of X. However, with this *direct-sampling* approach one is usually only able to probe the typical regime of the PDF $P(X, N)$, i.e., one can only estimate this distribution for $X \sim \mathcal{O}(\sqrt{N})$. For instance, considering 10^6 samples, with this method one can only observe events that occur with probability of order 10^{-6} or higher. However, if one is interested in the rare events corresponding to the large-deviation tails of the distribution (for instance for $X \sim \mathcal{O}(N)$) this naive approach cannot be used anymore as it would be extremely expensive. For instance, if one wants to sample events that occur with probabilities of order 10^{-100}, a more sophisticated approach is required.

For this reason, we employ a constrained MCMC algorithm, which allows us to sample rare configurations with atypically large displacements X. We start by implementing an MCMC dynamics in the space of configurations $\mathcal{C} = \{(\tau_i, v_i, u_i)\}$, us-

ing the Metropolis-Hastings algorithm. Starting from any initial configuration, we choose at random one of the N running phases, say the i-th, and we propose a move $(\tau_i, v_i, u_i) \to (\tau_i^{\text{new}}, v_i^{\text{new}}, u_i^{\text{new}})$, where

$$\tau_i^{\text{new}} = \tau_i + \delta\tau_i\,, \quad v_i^{\text{new}} = v_i + \delta v_i\,, \quad u_i^{\text{new}} = u_i + \delta u_i\,. \tag{4.91}$$

Here the variables $\delta\tau_i$, δv_i, δu_i are random and uniformly distributed in the intervals $(-a, a)$, $(-b, b)$ and $(-c, c)$, respectively. The positive constants a, b and c are parameters of the algorithm. The proposed new configuration is thus

$$\mathcal{C}^{\text{new}} = \{(v_1, \tau_1, u_1)\,, \dots\,, (v_{i-1}, \tau_{i-1}, u_{i-1})\,, (v_i{}^{\text{new}}, \tau_i{}^{\text{new}}, u_i^{\text{new}})\,, (v_{i+1}, \tau_{i+1}, u_{i+1})\,, \dots\} \tag{4.92}$$

The move is accepted with probability

$$p_{\text{acc}} = \min\left[1, \frac{p(\tau_i^{\text{new}})W(v_i^{\text{new}})f_d(u_i^{\text{new}})}{p(\tau^i)W(v^i)f_d(u_i)}\right]\,, \tag{4.93}$$

and it is rejected otherwise. Initially, we let the system evolve for 10^7 sweeps, i.e., $10^7 N$ moves. In this way the MCMC reaches its equilibrium state. Then, we measure the position X of the RTP every 10^2 sweeps, to avoid correlations between samples. Using these samples, we build a histogram that approximates the PDF $P(X, N)$. Up to this point, the MCMC algorithm is completely equivalent to the direct-sampling strategy and it only allows to probe typical trajectories. To estimate numerically the large-deviations tails of $P(X, N)$, we need to bias the MCMC dynamics towards large values of X.

We start by choosing some large value X^*, depending on which regime we would like to probe. For instance, for the model discussed in Section 4.2, the large-deviation regime that we have discussed is found at a scale $X \sim \mathcal{O}(N)$ and thus one would choose $X^* \sim \mathcal{O}(N)$. On the other hand, for the one-dimensional RTP with Gaussian speed discussed in Section 4.3, we take $X^* \sim \mathcal{O}(N^{3/4})$. In the following, we will discuss the latter case. We initialize the MCMC from some initial condition with $X > X^*$. Then, we evolve the system according to the MCMC dynamics described above, adding the hard constraint $X > X^*$, meaning that any attempted update corresponding to $X < X^*$ is always rejected.

The histogram that we obtain from this biased algorithm will approximate the PDF $P(X, N|X > X^*)$, conditioned on the event $X > X^*$. This quantity is then simply related to the PDF $P(X, N)$ by

$$P(X, N|X > X^*) = \frac{P(X, N)}{P(X > X^*)}\,. \tag{4.94}$$

Taking the natural logarithm of both sides, we obtain

$$\log\left[P(X, N|X > X^*)\right] = \log\left[P(X, N)\right] - \log\left[P(X > X^*)\right]\,. \tag{4.95}$$

Diving both sides by $\sqrt{N}$ and recalling the definition of the rate function (for large N)

$$F\left(\frac{X}{N^{3/4}}\right) = -\frac{\log\left[P(X, N)\right]}{\sqrt{N}}\,, \tag{4.96}$$

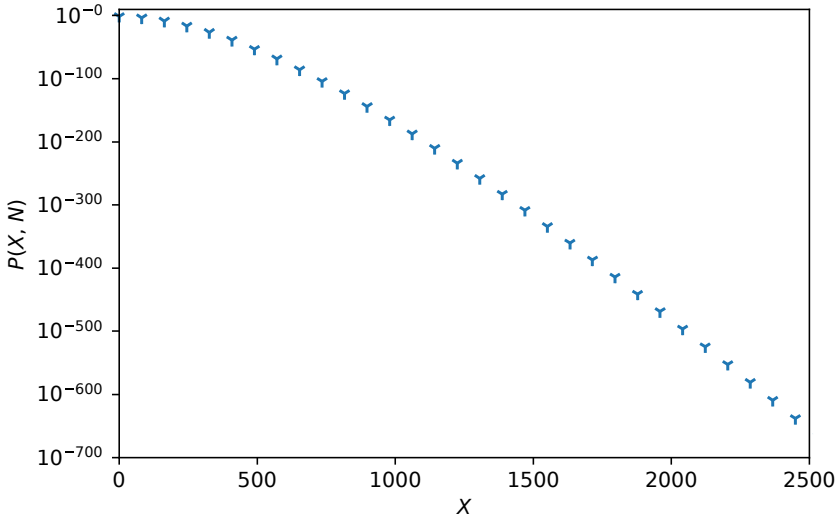

Figure 4.11: Numerical curve of the probability density function $P(X, N)$ as a function of X, for $d = 2$, speed distribution $W(v)$ in Eq. (4.16) with $\alpha = 2$, and $N = 1000$, obtained with a constrained Markov chain Monte Carlo algorithm.

we find

$$F\left(\frac{X}{N^{3/4}}\right) = -\frac{\log\left[P(X, N|X > X^*)\right]}{\sqrt{N}} + C_{X^*}\,. \tag{4.97}$$

Note that C_{X^*} is constant with respect to X.

Thus, to estimate numerically $F(z)$ we need to compute the value of C_{X^*}. This can be achieved by the following strategy. First, we perform an unbiased simulation, which will allow us to estimate $F(z)$ in a small interval around the origin. Then, we choose a value of X^* such that $z^* = X^*/N^{3/4}$ falls within the range of values for which $F(z)$ is known. The biased simulation will give us an estimate of $F(z)$ in a small region with $z > z^*$, up to the constant C_{X^*}. Since the two estimates of $F(z)$, the one obtained without the constraint and the one with the constraint, overlap for some values of z, one can compute the constant C_{X^*} by matching the two curves. This allows us to know $F(z)$ in a slightly larger interval. Then, we continue by performing a new MCMC simulation with a larger value of X^* and so on until $F(z)$ is known for the desired range of values. Note that to speed up the algorithm the procedure above can be parallelized by choosing a fine enough grid of values X^* in order to ensure the overlap between the different histograms. For instance, to obtain the numerical curves in Fig. 4.7 we have used 90 equispaced values of X^*. Using this procedure, one can probe events with very small probabilities, even smaller than 10^{-100}.

Interestingly, with the technique described above one can also obtain the marginal PDF $p(x|X)$ (see Fig. 4.8). Indeed, from the MCMC dynamics, one has access to the values of the single-run displacements $x_1, \dots x_N$. By measuring a randomly chosen displacement at every step one can build a histogram that will approximate the PDF $p(x|X > X^*)$. However, it turns out that when X^* is large, the system will typically stay in a small region to the right of X^*. In other words, even if X is free to fluctuate

during the simulation, it will typically remain close to X^*. Thus, one can approximate

$$p(x|X > X^*) \simeq p(x|X^*)\,. \tag{4.98}$$

Alternatively, one can avoid this approximation by performing an MCMC dynamics at fixed X (e.g., by proposing moves that conserve the total displacement X). However, the convergence of the algorithm turns out to be slower in this case. With the same technique one can also estimate the participation $Y_2(z)$ as functions of z (see Fig. 4.10). The results of our numerical simulations are in good agreement with the theory.

Article 7

Condensation transition in the late-time position of a Run-and-Tumble particle

F. Mori, P. Le Doussal, S. N. Majumdar, and G. Schehr,
Phys. Rev. E **103**, 062134 (2021).

☞ https://journals.aps.org/pre/abstract/10.1103/PhysRevE.103.062134
☞ https://arxiv.org/abs/2103.04637

Abstract

We study the position distribution $P(\vec{R}, N)$ of a run-and-tumble particle (RTP) in arbitrary dimension d, after N runs. We assume that the constant speed $v > 0$ of the particle during each running phase is independently drawn from a probability distribution $W(v)$ and that the direction of the particle is chosen isotropically after each tumbling. The position distribution is clearly isotropic, $P(\vec{R}, N) \to P(R, N)$ where $R = |\vec{R}|$. We show that, under certain conditions on d and $W(v)$ and for large N, a condensation transition occurs at some critical value of $R = R_c \sim O(N)$ located in the large deviation regime of $P(R, N)$. For $R < R_c$ (subcritical fluid phase), all runs are roughly of the same size in a typical trajectory. In contrast, an RTP trajectory with $R > R_c$ is typically dominated by a 'condensate', i.e., a large single run that subsumes a finite fraction of the total displacement (supercritical condensed phase). Focusing on the family of speed distributions $W(v) = \alpha(1 - v/v_0)^{\alpha-1}/v_0$, parametrized by $\alpha > 0$, we show that, for large N, $P(R, N) \sim \exp\left[-N\psi_{d,\alpha}(R/N)\right]$ and we compute exactly the rate function $\psi_{d,\alpha}(z)$ for any d and α. We show that the transition manifests itself as a singularity of this rate function at $R = R_c$ and that its order depends continuously on d and α. We also compute the distribution of the condensate size for $R > R_c$. Finally, we study the model when the total duration T of the RTP, instead of the total number of runs, is fixed. Our analytical predictions are confirmed by numerical simulations, performed using a constrained Markov chain Monte Carlo technique, with precision $\sim 10^{-100}$.

Article 8

First-order condensation transition in the position distribution of a run-and-tumble particle in one dimension

F. Mori, G. Gradenigo, and S. N. Majumdar,
J. Stat. Mech. 103208 (2021).

☞ https://iopscience.iop.org/article/10.1088/1742-5468/ac2899/meta
☞ https://arxiv.org/abs/2107.00338

Abstract

We consider a single run-and-tumble particle (RTP) moving in one dimension. We assume that the velocity of the particle is drawn independently at each tumbling from a zero-mean Gaussian distribution and that the run times are exponentially distributed. We investigate the probability distribution $P(X, N)$ of the position X of the particle after N runs, with $N \gg 1$. We show that in the regime $X \sim N^{3/4}$ the distribution $P(X, N)$ has a large deviation form with a rate function characterized by a discontinuous derivative at the critical value $X = X_c > 0$. The same is true for $X = -X_c$ due to the symmetry of $P(X, N)$. We show that this singularity corresponds to a first-order condensation transition: for $X > X_c$ a single large jump dominates the RTP trajectory. We consider the participation ratio of the single-run displacements as the order parameter of the system, showing that this quantity is discontinuous at $X = X_c$. Our results are supported by numerical simulations performed with a constrained Markov chain Monte Carlo algorithm.

Chapter 5

Optimal control with stochastic resetting

In this last chapter, we present some recent developments in the application of stochastic resetting to the context of optimal control theory. Stochastic processes with random restarts have recently become a very active area of research (for a recent review see [83]). This class of stochastic processes includes systems that are restarted at random times from a fixed configuration.

One of the simplest and most investigated models of resetting is that of a Brownian particle in one dimension with constant resetting rate r, which was first introduced in [36]. In a small time interval dt, the position of the particle evolves according to the stochastic rule

$$x(t+dt) = \begin{cases} x(t) + \sqrt{2D}\eta(t)dt & \text{with probability } 1 - r dt\,, \\ \\ 0 & \text{with probability } r dt\,, \end{cases} \tag{5.1}$$

where $\eta(t)$ is Gaussian white noise. This simple stochastic rule leads to several interesting features.

First of all, for $r > 0$, at late times the system reaches the steady state [83]

$$p_{\rm st}(x) = \frac{1}{2}\sqrt{\frac{r}{D}}\exp\left(-\sqrt{\frac{r}{D}}|x|\right)\,. \tag{5.2}$$

Moreover, since the resetting dynamics induce a net probability flux towards the origin, the system is out of equilibrium. As a consequence, resetting is often used as a toy model for nonequilibrium stationary processes (see for instance Chapter 2). Besides the case of one-dimensional diffusive particles, stochastic resetting has been investigated for a wide range of processes, including random walks [88, 89], active particles [91–93], fluctuating interfaces [81, 212, 213], Ising model [95], and predator prey models [214].

Another relevant aspect of stochastic resetting is related to search processes [215]. As an example, let us consider a Brownian particle in one dimension with position $x(t)$, initially at the origin $x(0) = 0$. We denote by T_F the first time at which the process reaches a target located at $x = L$. This *first-passage time* T_F has been studied for a wide

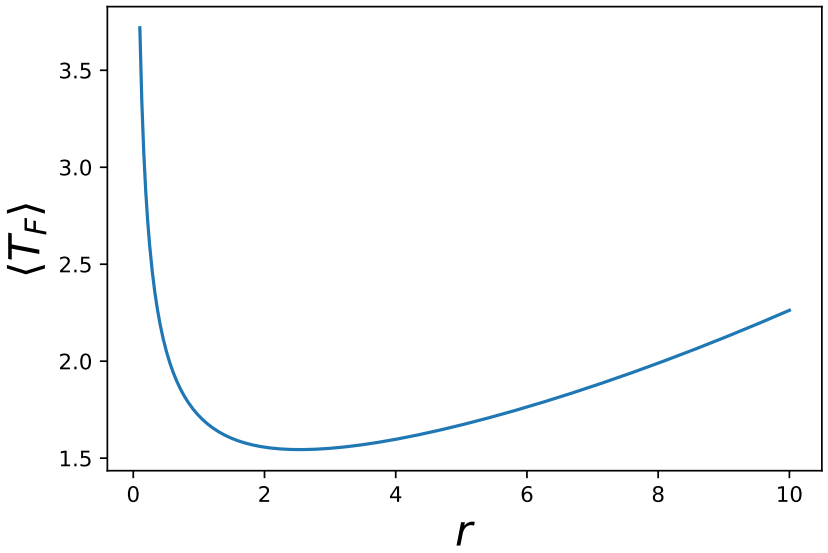

Figure 5.1: Mean first-passage time $\langle T_F \rangle$ to a target at position $x = L$ as a function of the resetting rate r for Brownian motion. The curve corresponds to the exact result in Eq. (5.3) with $L = D = 1$.

range of processes, with several applications [32, 75]. In the case of BM, the average value of T_F, known as *mean first-passage time* (MFPT) is infinite, meaning that on average it will take an infinite amount of time to find the target[1]. On the other hand, in the presence of resetting with constant rate r, it can be shown that [36]

$$\langle T_F \rangle = \frac{1}{r}\left[\exp\left(\sqrt{\frac{r}{D}}L\right) - 1\right] . \tag{5.3}$$

Thus, once we turn on the resetting rate r the MFPT becomes finite. Interestingly, as a function of r, the MFPT turns out to be non-monotonic and shows a global minimum for $r = r^*$, where $r^* = (2.53964\ldots)\, D/L^2$ (see Fig. 5.1). Thus, one can optimally choose the resetting rate r^* to minimize the MFPT to a given target.

Starting from this observation, it is natural to ask whether one could find a more general framework that allows driving stochastic systems through restarts. In particular, considering the more complicated case where the resetting rate depends on the state of the system and on time, i.e., $r = r(x, t)$, can we find the optimal resetting rate $r^*(x, t)$, as a function of x and t, to achieve a given goal? This question is best posed within the framework of optimal control theory. In this chapter, we will first introduce some classical ideas in the context of optimal control. Then, we will introduce a new framework that allows finding the optimal resetting policy in a wide range of situations.

[1] Note however that the probability of eventually finding the target is one.

5.1 Optimal control theory and the Hamilton-Jacobi-Bellman equation

Since the seminal works of Pontryagin [216] and Bellman [101], optimal control theory has found applications in a plethora of contexts, including artificial intelligence [217] and finance [218]. In the typical setting, one considers a dynamical (stochastic or deterministic) system, whose state $\vec{x}(t)$, expressed as a d-dimensional vector, evolves as a function of time t. In many situations, the dynamical equation that determines the evolution of the state of the system can be written as

$$\frac{d\vec{x}(t)}{dt} = \vec{f}(x,t) + \sqrt{2D}\vec{\eta}(t) + \vec{u}(x,t)\,, \tag{5.4}$$

where $\vec{f}(x,t)$ is the deterministic component of the dynamics of the system, $\vec{\eta}(t)$ is Gaussian white noise with zero mean and correlator $\langle \eta_i(t)\eta_j(t')\rangle = \delta_{i,j}\delta(t-t')$, and $D>0$ is the diffusion constant. The extra term $\vec{u}(x,t)$ represents the external control, that we can tune to achieve a given goal.

The goal is expressed in terms of the *reward function* $R(\vec{x},t)$ that quantifies the reward received if the system is in state $\vec{x}$ at time t. Moreover, one has to take into account the operating costs associated with controlling the system. For simplicity, we assume a quadratic cost in $\vec{u}(x,t)$. Thus, considering the evolution of the system in the time interval $[t,t_f]$, the payoff can be expresses as the difference between the reward and the cost

$$\int_t^{t_f} d\tau\; R(\vec{x}(\tau),\tau) - \int_t^{t_f} d\tau\; \frac{1}{2}\vec{u}^{\,2}(\vec{x}(\tau)\,\tau)\,. \tag{5.5}$$

The final time t_f is often called *time horizon*. Note that the payoff in Eq. (5.5) depends on the full trajectory of the system and is therefore a random variable. For this reason, it is useful to consider the expected payoff

$$\mathcal{F}_{\vec{x}_0,t}\left[\vec{u}\right] = \left\langle \int_t^{t_f} d\tau\; R(\vec{x}(\tau),\tau) - \int_t^{t_f} d\tau\; \frac{1}{2}\vec{u}^{\,2}(\vec{x},(\tau)\,\tau)\right\rangle_{\vec{x}_0}\,, \tag{5.6}$$

where the average $\langle\ldots\rangle_{\vec{x}_0}$ is taken over all stochastic trajectories starting from $\vec{x}_0$ at time t, i.e., with $\vec{x}(t)=\vec{x}_0$. Note that the expected payoff in Eq. (5.6) is a functional of the control policy $\vec{u}(\vec{x},\tau)$, for $\tau\in[t,t_f]$.

Then, two questions naturally arise: (i) What is the optimal control strategy

$$\vec{u}^{\,*}(\vec{x},\tau) = \mathrm{argmax}_{\vec{u}}\,\mathcal{F}_{\vec{x}_0,t}\left[u\right] \tag{5.7}$$

that maximizes the expected payoff? Note that one has to determine the optimal control as a function of $\vec{x}\in\mathbb{R}^d$ and $\tau\in[t,t_f]$. (ii) What is the *optimal payoff*

$$J(\vec{x}_0,t) = \max_{\vec{u}}\mathcal{F}_{\vec{x}_0,t}\left[\vec{u}\right] = \mathcal{F}_{\vec{x}_0,t}\left[\vec{u}^{\,*}\right]\,, \tag{5.8}$$

i.e., the payoff associated with the optimal cost? Note that, by definition, $J(\vec{x},t)$ is the expected payoff starting from position $\vec{x}_0$ at time t and optimally controlling the system up to time t_f. Answering these questions requires solving a quite complicated functional optimization problem.

Remarkably, Bellman and his collaborators introduced a general framework to solve this class of problems, known as dynamic programming. The main idea is to break down the optimization problem into several simpler subproblems that one can solve iteratively. This approach leads to the celebrated Hamilton-Jacobi-Bellman equation, which allows computing both $J(\vec{x}_0,t)$ and $\vec{u}^{\,*}(\vec{x},t)$ by solving a differential equation. We now sketch a simple derivation of this equation.

To proceed, it is useful to split the integrals in Eq. (5.6) as

$$\begin{aligned}\mathcal{F}_{\vec{x}_0,t}\left[\vec{u}\right] &= dtR(\vec{x}_0,t) - dt\frac{1}{2}\vec{u}^{\,2}(\vec{x}_0,t) \\ &+ \int d\vec{x}_1 P(\vec{x}_1|\vec{x}_0)\left\langle \int_{t+dt}^{t_f} d\tau\; R(\vec{x}(\tau),\tau) - \int_{t+dt}^{t_f} d\tau\; \frac{1}{2}\vec{u}^{\,2}(\vec{x}(\tau)\,\tau)\right\rangle_{\vec{x}_1},\end{aligned} \tag{5.9}$$

where dt is a small time increment and we have averaged over all possible values of $\vec{x}_1 = \vec{x}(t+dt)$, multiplying by its conditioned PDF $P(\vec{x}_1|\vec{x}_0)$. This distribution $P(\vec{x}_1|\vec{x}_0)$ depends on $\vec{x}_0$ and on the details of the dynamics. We recall that the symbol $\langle\ldots\rangle_{\vec{x}_1}$ indicates the average over all trajectories starting from the fixed state $\vec{x}_1$ at time $t+dt$. We now perform the maximization over $\vec{u}(\vec{x},\tau)$ (for $t \le \tau \le t_f$) on both sides, yielding

$$\begin{aligned}J(\vec{x}_0,t) &= \max_{\vec{u}}\left[dtR(\vec{x}_0,t) - dt\frac{1}{2}\vec{u}^{\,2}(\vec{x}_0,t)\right. \\ &+ \left.\int d\vec{x}_1 P(\vec{x}_1|\vec{x}_0)\left\langle \int_{t+dt}^{t_f} d\tau\; R(\vec{x}(\tau),\tau) - \int_{t+dt}^{t_f} d\tau\; \frac{1}{2}\vec{u}^{\,2}(\vec{x}(\tau)\,\tau)\right\rangle_{\vec{x}_1}\right].\end{aligned} \tag{5.10}$$

We perform the optimization over $\vec{u}(\vec{x},\tau)$ in two steps, first we maximize over $\vec{u}(\vec{x},\tau)$ for $t+dt \le \tau \le t_f$, which gives

$$J(\vec{x}_0,t) = \max_{\vec{u}(\vec{x}_0,t)}\left[dtR(\vec{x}_0,t) - dt\frac{1}{2}\vec{u}^{\,2}(\vec{x}_0,t) + \langle J(\vec{x}_1,t+dt)\rangle\right], \tag{5.11}$$

where we have used the definition of $J(\vec{x},t)$ in Eq. (5.8) and $\langle\ldots\rangle$ now indicates averaging over the random variable $\vec{x}_1$. Note that one only needs to maximize over the variable $\vec{u}(\vec{x}_0,t)$ (with $\vec{x}_0$ and t fixed).

Using the equations of motion in (5.4), we next express the variable $\vec{x}_1$ as

$$\vec{x}_1 = \vec{x}(t+dt) = \vec{x}_0 + \vec{f}(\vec{x}_0,t)dt + \sqrt{2Ddt}\vec{\eta}(t) + \vec{u}(\vec{x}_0,t)dt\,, \tag{5.12}$$

yielding

$$\begin{aligned}J(\vec{x}_0,t) &= \max_{\vec{u}(\vec{x}_0,t)}\left[dtR(\vec{x}_0,t) - dt\frac{1}{2}\vec{u}^{\,2}(\vec{x}_0,t)\right. \\ &+ \left.\langle J(\vec{x}_0 + \vec{f}(\vec{x}_0,t)dt + \sqrt{2D}dt\vec{\eta}(t) + \vec{u}(\vec{x}_0,t)dt, t+dt)\rangle\right],\end{aligned} \tag{5.13}$$

where the average $\langle\ldots\rangle$ is now performed over the Gaussian white noise $\vec{\eta}(t)$. Expanding the right-hand side to leading order in dt, we get

$$\begin{aligned}0 &= \max_{\vec{u}(\vec{x}_0,t)}\left[dtR(\vec{x}_0,t) - dt\frac{1}{2}\vec{u}^{\,2}(\vec{x}_0,t) + \left[\vec{f}(\vec{x}_0,t) + \vec{u}(\vec{x}_0,t)\right]\cdot\nabla_{\vec{x}_0}J(\vec{x}_0,t)dt\right. \\ &+ \left. D\Delta_{\vec{x}_0}J(\vec{x}_0,t)dt + \partial_t J(\vec{x}_0,t)dt\right].\end{aligned} \tag{5.14}$$

Renaming $\vec{x}_0 = \vec{x}$ to ease the notation, we find that optimal control is given by

$$\boxed{\vec{u}^{\,*}(\vec{x},t) = \nabla_{\vec{x}} J(\vec{x},t)\,.} \tag{5.15}$$

Plugging this expression back into Eq. (5.14), we obtain the differential equation

$$\boxed{-\partial_t J(\vec{x},t) = R(\vec{x},t) + \frac{1}{2}\left[\nabla_{\vec{x}} J(\vec{x},t)\right]^2 + \vec{f}(\vec{x},t)\cdot\nabla_{\vec{x}} J(\vec{x},t) + D\Delta_{\vec{x}} J(\vec{x},t)\,,} \tag{5.16}$$

which is the celebrated stochastic Hamilton-Jacobi-Bellman equation [100].

The HJB equation must be solved backward in time, starting from the final condition $J(\vec{x},t_f) = 0$. Once the full solution $J(\vec{x},t)$ for $t \in [0,t_f]$ is found, one can compute the optimal control by using Eq. (5.15). Note that the optimal control $\vec{u}^{\,*}$ always drives the system in the direction in which the expected payoff $J(\vec{x},t)$ increases the most. The quadratic term in $\nabla_{\vec{x}} J$ makes the equation nonlinear and, as a consequence, one has often to rely on numerical simulations to solve it.

Although analytical solutions of the HJB equation are usually hard to obtain, there exist few exactly solvable cases. For instance, let us consider the case $d = 1$, with $f(x,t) = 0$ and $R(x,t) = -(\alpha/2)(x - x_f)^2\delta(t - t_f)$. In this case, since the reward function contains a δ function centered at $t = t_f$, the reward will only depend on the final state $x(t_f)$ of the system. In particular, the reward is higher the closer $x(t_f)$ is to the target state x_f. In this case it is possible to show that the optimal control is given by

$$u^*(x,t) = -\alpha\frac{x - x_f}{1 + \alpha(t_f - t)}\,. \tag{5.17}$$

This optimal control continuously pushes the system in the direction of the target location x_f. Interestingly, in the limit $\alpha \to \infty$, one obtains the effective force to generate a Brownian bridge [125].

Stochastic optimal control has been used in a variety of systems, including pandemic management [219], supply-chain planning [220], and fluctuating interfaces [221]. The formalism described above requires the control to be local and continuous in time. However, a wide range of systems cannot be controlled by a sequence of infinitesimal local changes but require global changes. This can happen either because implementing continuous changes would be too complicated or because there are structural limitations that do not allow continuous control. A recent example of such a situation is the COVID-19 crisis, during which the main policies chosen by governments were global measures such as lockdowns.

Other examples can be found in the context of search processes, both in computer science [84] and in search-and-rescue missions [222]. A commonly observed strategy in search processes is to combine phases of local motion during which the searcher is actively looking for the target and phases of global motion, during which the target cannot be detected. These *intermittent search strategies* have been widely investigated and are typically quite efficient [75, 223]. However, the HJB equation (5.16) cannot be directly applied to investigate such global policies. One of the simplest global policies that one can implement is to restart the system from some known state.

5.2 Resetting optimal control

In order to model restarting policies we replace the control drift $\vec{u}(\vec{x},t)$ in Eq. (5.4) with a space-dependent resetting rate $r(\vec{x},t)$. In a small time interval dt, the system evolves according to

$$\vec{x}(t+dt) = \begin{cases} \vec{x}_{\rm res} & \text{with probability } r(\vec{x},t)dt\,, \\ \\ \vec{x}(t) + \vec{f}(\vec{x},t)dt + \sqrt{2D}dt\vec{\eta}(t) & \text{with probability } 1-r(\vec{x},t)dt\,. \end{cases} \tag{5.18}$$

In other words, with probability $r(\vec{x},t)dt$ the system is reset to the fixed resetting location $\vec{x}_{\rm res}$. With the complementary probability $1-r(\vec{x},t)dt$, the system evolves freely.

We would like to find the optimal resetting policy $r^*(\vec{x},t)$ that maximizes an expected payoff. As before, we consider a generic reward function $R(\vec{x},t)$. Moreover, we introduce a cost $c(\vec{x},t)$ associated with each resetting event. Thus, the expected payoff can be written as

$$\mathcal{F}_{\vec{x}_0,t}\left[r\right] = \left\langle \int_t^{t_f} d\tau\ R(\vec{x}(\tau),\tau) - \int_t^{t_f} d\tau\ c(\vec{x},(\tau)\,\tau) r(\vec{x},(\tau)\,\tau) \right\rangle_{\vec{x}_0}\,, \tag{5.19}$$

where we recall that the average is performed over all trajectories $\vec{x}(\tau)$ (for $t \le \tau \le t_f$) with $\vec{x}(t) = \vec{x}_0$. The optimal payoff and the optimal policy are respectively defined as

$$J(\vec{x}_0,t) = \max_r \mathcal{F}_{\vec{x}_0,t}\left[r\right]\,, \tag{5.20}$$

and

$$r^*(\vec{x},\tau) = \text{argmax}_r\, \mathcal{F}_{\vec{x}_0,t}\left[r\right]\,, \tag{5.21}$$

where the maximization is performed over all resetting functions $r(\vec{x},\tau)$. Note that also in this case the expected payoff $\mathcal{F}_{\vec{x}_0,t}\left[r\right]$ is a functional of the resetting strategy $r(\vec{x},t)$, which is a function of $\vec{x}$ and t. Thus, this is again a quite complicated functional optimization problem. Using ideas from dynamic programming, as done in the previous section, we derive a framework that allows solving this optimization problem in a very general setting.

We start by rewriting the expected payoff in Eq. (5.19) as

$$\begin{aligned} \mathcal{F}_{\vec{x}_0,t}\left[r\right] &= dtR(\vec{x}_0,t) - dtc(\vec{x}_0,t)r(\vec{x}_0,t) \\ &+ \int d\vec{x}_1 P(\vec{x}_1|\vec{x}_0) \left\langle \int_{t+dt}^{t_f} d\tau\ R(\vec{x}(\tau),\tau) - \int_{t+dt}^{t_f} d\tau\ c(\vec{x}_0,t)r(\vec{x}_0,t) \right\rangle_{\vec{x}_1}\,, \end{aligned} \tag{5.22}$$

where, as before, $\vec{x}_1 = \vec{x}(t+dt)$ and $P(\vec{x}_1|\vec{x}_0)$ indicates the distribution of $\vec{x}_1$ conditioned on the previous state $\vec{x}_0$. We next maximize both sides over $r(\vec{x},\tau)$, for $t \le \tau \le t_f$, yielding

$$\begin{aligned} J(\vec{x}_0,t) &= \max_r \Big[dtR(\vec{x}_0,t) - dtc(\vec{x}_0,t)r(\vec{x}_0,t) \\ &+ \int d\vec{x}_1 P(\vec{x}_1|\vec{x}_0) \left\langle \int_{t+dt}^{t_f} d\tau\ R(\vec{x}(\tau),\tau) - \int_{t+dt}^{t_f} d\tau\ c(\vec{x}_0,t)r(\vec{x}_0,t) \right\rangle_{\vec{x}_1} \Big]\,, \end{aligned} \tag{5.23}$$

Maximizing over $r(\vec{x},\tau)$ for $t+dt \leq \tau \leq t_f$, we get

$$J(\vec{x}_0,t) = \max_{r(\vec{x}_0,t)} \left[dtR(\vec{x}_0,t) - dtc(\vec{x}_0,t)r(\vec{x}_0,t) + \langle J(\vec{x}_1,t+dt)\rangle \right], \tag{5.24}$$

where the average $\langle \ldots \rangle$ is now taken over the variable $\vec{x}_1 = \vec{x}(t+dt)$. From Eq. (5.18) we have

$$\vec{x}_1 = \begin{cases} \vec{x}_{\rm res} & \text{with probability } r(\vec{x}_0,t)dt\,, \\ \\ \vec{x}_0 + \vec{f}(\vec{x}_0,t)dt + \sqrt{2D}dt\vec{\eta}(t) & \text{with probability } 1-r(\vec{x}_0,t)dt\,. \end{cases} \tag{5.25}$$

Thus, we obtain

$$\begin{aligned} J(\vec{x}_0,t) &= \max_{r(\vec{x}_0,t)} \Big[dtR(\vec{x}_0,t) - dtc(\vec{x}_0,t)r(\vec{x}_0,t) + r(\vec{x}_0,t)dtJ(\vec{x}_{\rm res},t+dt) \\ &+ (1-r(\vec{x}_0,t)dt) \left\langle J(\vec{x}_0 + \vec{f}(\vec{x}_0,t)dt + \sqrt{2D}dt\vec{\eta}(t), t+dt) \right\rangle \Big], \end{aligned} \tag{5.26}$$

where the average is now performed over the Gaussian noise $\vec{\eta}(t)$. Expanding to leading order in dt, we find

$$\begin{aligned} 0 &= \max_{r(\vec{x}_0,t)} \left[r(\vec{x}_0,t)dt \left[J(\vec{x}_{\rm res},t+dt) - J(\vec{x}_0,t) - c(\vec{x}_0,t) \right] \right] \\ &+ dtR(\vec{x}_0,t) + \vec{f}(\vec{x}_0,t)\cdot\nabla_{\vec{x}_0} J(\vec{x}_0,t)dt + D\Delta_{\vec{x}_0} J(\vec{x}_0,t)dt + \partial_t J(\vec{x}_0,t)dt\,, \end{aligned} \tag{5.27}$$

Note that since $r(\vec{x}_0,t)dt$ is a probability one has to perform the maximization with the constraint $0 \leq r(\vec{x}_0,t) \leq 1/dt$. Thus, since the function that we are maximizing is linear, we find that the optimal resetting policy is

$$r^*(\vec{x}_0,t)dt = \begin{cases} 0 & \text{if } J(\vec{x}_0,t) \geq J(\vec{x}_{\rm res},t) + c(\vec{x}_0,t)\,, \\ \\ 1 & \text{if } J(\vec{x}_0,t) < J(\vec{x}_{\rm res},t) + c(\vec{x}_0,t)\,. \end{cases} \tag{5.28}$$

Renaming $\vec{x}_0 \to \vec{x}$, this policy can be rewritten as

$$\boxed{r^*(\vec{x},t)dt = \begin{cases} 0 & \text{if } \vec{x} \in \Omega(t)\,, \\ \\ 1 & \text{if } \vec{x} \notin \Omega(t)\,, \end{cases}} \tag{5.29}$$

where we have defined the domain

$$\boxed{\Omega(t) = \{\vec{x} : J(\vec{x},t) \geq J(\vec{x}_{\rm res},t) - c(\vec{x},t)\}\,.} \tag{5.30}$$

Remarkably, we find that the optimal resetting policy r^* is *bang-bang* [224], meaning that the control switches between two possible states. The meaning of Eq. (5.29) is quite clear. The optimal strategy is to reset if the expected payoff at the resetting location $\vec{x}_{\rm res}$ is larger than the one at the current location $\vec{x}$, taking into account the cost $c(\vec{x},t)$ of resetting.

Moreover, from Eq. (5.27) we find that the optimal payoff $J(\vec{x},t)$ satisfies the differential equation

$$\boxed{-\partial_t J(\vec{x}_0,t) = \vec{f}(\vec{x}_0,t)\cdot\nabla_{\vec{x}_0}J(\vec{x}_0,t) + D\Delta_{\vec{x}_0}J(\vec{x}_0,t)dt + R(\vec{x},t)\,,} \tag{5.31}$$

with final condition $J(\vec{x},t_f)=0$. Note that the domain of definition of the differential equation (5.31) is the domain $\Omega(t)$ itself. One has to therefore solve the equation backwards in time, with a time-dependent domain of definition and with Neumann boundary conditions in space

$$\boxed{\nabla_{\vec{x}}J(\vec{x},t)\cdot\vec{n}(\vec{x}) = 0\,,} \tag{5.32}$$

where $\vec{n}(\vec{x})$ is the normal unit vector to the boundary. For the derivation of the boundary condition, see Article 9. Outside of the domain $\Omega(t)$, the solution is given by $J(\vec{x},t) = J(\vec{x}_{\rm res},t) - c(\vec{x},t)$.

We would like to stress the fact that the differential equation in Eq. (5.31) is of a special kind, since the domain of definition, given in Eq. (5.30), depends on the solution $J(\vec{x},t)$ of the differential equation. This kind of differential equation belongs to a class known as *moving boundary problems* or *Stefan problems* [225]. These problems often arise in the description of heat transfer between different phases of matter. A simple example is the motion of the interface between ice and water on a freezing lake. To obtain the temperature profile as a function of the distance from the lake surface, one must solve the heat equation with an interface that moves according to the temperature gradient. The interface typically moves with time and has to be treated as an additional variable, which must be computed jointly with the solution of the differential equation. In analogy with the freezing-lake example, in our case, the optimal payoff $J(\vec{x},t)$ corresponds to the temperature profile, while the boundary of the domain $\Omega(t)$ plays the role of the ice-water interface[2]. In general, the domain $\Omega(t)$ and the optimal payoff $J(\vec{x},t)$ have to be obtained numerically from Eqs. (5.30) and (5.31).

The framework derived above turns out to be very general and provides the optimal restart policy for a wide range of dynamical systems and several control problems. In the rest of this section, we present two cases in which one can make analytical progress.

5.2.1 Dirac delta final reward

One of the simplest control problems that one can consider is that of driving a one-dimensional system towards a desired location x_f at the final time t_f. This can be expressed with the δ function reward

$$R(x,t) = \alpha\delta(x-x_f)\delta(t-t_f)\,. \tag{5.33}$$

Thus, the reward is α if the system is at position x_f at the final time t_f and it is zero otherwise [3]. For simplicity, we also consider a constant resetting cost $c(x,t)=c>0$ and we set the deterministic component of the dynamics to zero, i.e., we assume $f(x,t)=0$. Thus, the process is a Brownian motion with restarts.

[2] Note however that the two problems have different boundary conditions at the interface.

[3] In some sense, this is like a *rendez-vous*, where one has to be at the right place at the right time.

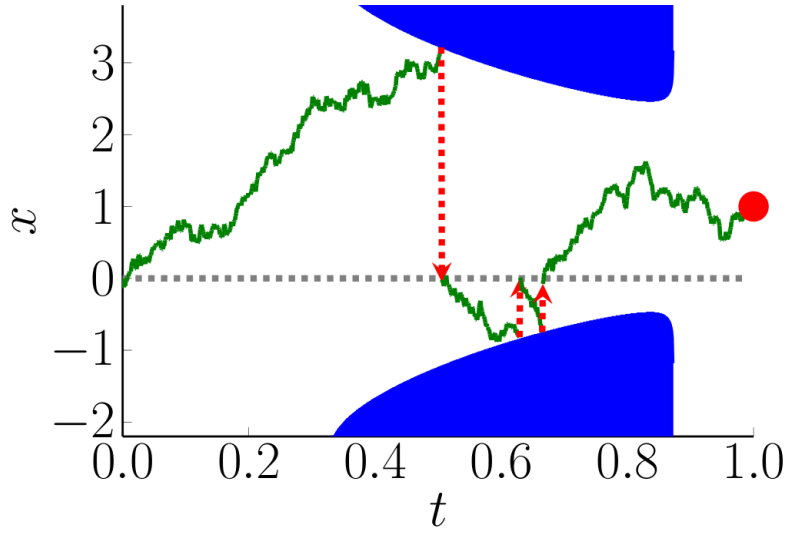

Figure 5.2: Optimal restarting strategy for the reward $R(x,t) = \alpha\delta(x-1)\delta(t-1)$, with $\alpha = 10$ and with unit resetting cost $c = 1$. The goal is to drive the process to the target location $x_f = 1$ (red dot) at the final time $t_f = 1$. Getting to the target location is associated with a constant reward α while a cost c is paid for each resetting event (red dashed arrows). The optimal strategy is to let the system free to evolve (continuous green line) when it is in $\Omega(t)$ (white region) and to reset it to $x_{\rm res} = 0$ (horizontal dashed line) upon touching the blue region, which is the complementary of $\Omega(t)$. The optimal domain $\Omega(t)$ is obtained from the numerical integration of Eqs. (5.30) and (5.31), with $D = 1$.

Before presenting the solution to this control problem, it is useful to consider two limiting cases. In the case where $\alpha < c$, even one resetting event would cost more than the potential reward for reaching the target location. Thus, in such a case, the optimal strategy is to never reset, corresponding to the domain $\Omega(t) = \mathbb{R}$. In the opposite situation where $\alpha \gg c$, the cost of resetting can be neglected. Thus, it is convenient to reset as soon as the distance $|x - x_f|$ of the current location x from the target x_f is larger than the distance of the resetting location $x_{\rm res}$ from the target. Thus, assuming $x_f > x_{\rm res}$, one expects $\Omega(t) \approx [x_{\rm res}, 2x_f - x_{\rm res}]$. However, when $\alpha/c \sim 1$ it is quite nontrivial to identify the optimal resetting protocol.

From Eqs. (5.30) and (5.31), we derive the optimal resetting strategy for any α and c. For $\alpha < \alpha_c$, where $\alpha_c = x_f c\sqrt{2\pi e} \approx 4.13273\, x_f c$, we find that the optimal strategy is to never reset, corresponding to $\Omega(t) = \mathbb{R}$ for all $0 \le t \le t_f$. The result is more complex for $\alpha > \alpha_c$, where resetting is only favorable in a specific time window, for $t < t_f$. In this case, the domain $\Omega(t)$ is quite nontrivial and has to be obtained by numerical integration (see Fig. 5.2).

To describe the case $\alpha > \alpha_c$, it is convenient to introduce the backward time $\tau = t_f - t$ and the backward payoff, defined as $I(x,\tau) = J(x, t_f - \tau)$. Using Eq. (5.31), it is easy to show that $I(x,\tau)$ satisfies the diffusion equation

$$\partial_\tau I(x,\tau) = D\partial_{xx} I(x,\tau)\,, \tag{5.34}$$

with initial condition $I(x, \tau = 0) = \alpha\delta(x - x_f)$. Assuming that no boundary appears for small τ, i.e., that $\Omega(t) = \mathbb{R}$ (to be verified a posteriori), the optimal payoff function is the Gaussian weight

$$I(x, \tau) = \alpha \frac{1}{\sqrt{4\pi D t}} e^{-(x-x_f)^2/(4\pi D\tau)} . \tag{5.35}$$

A boundary only appears when the condition $I(x, \tau) < I(0, \tau) + c$ is verified for the first time for some value of x. Using Eq. (5.35), this condition can be rewritten, for $|x - x_f| > x_f$, as

$$\alpha > \frac{c\sqrt{4\pi D\tau}}{e^{-x_f^2/(4D\tau)} - e^{-(x-x_f)^2/(4D\tau)}} . \tag{5.36}$$

Minimizing the right-hand side of Eq. (5.36) over x and τ, we obtain

$$\alpha > c x_f \sqrt{2e\pi} \tag{5.37}$$

Thus, for $\alpha < \alpha_c = c x_f \sqrt{2e\pi}$, no boundary appears and the cost function is given by Eq. (5.35) for any τ. On the other hand, for $\alpha > \alpha_c$, two boundaries appear at time τ^*, which is the smallest solution of the transcendental equation

$$\alpha = c\sqrt{4\pi D\tau}\; e^{x_f^2/(4D\tau)} . \tag{5.38}$$

This solution τ^* is shown in Fig. 5.3 as a function of the reward intensity α. Thus, for $\tau < \tau^*$ (i.e., for $t > t_f - \tau^*$), the cost function is given by Eq. (5.35), while it is hard to determine it analytically for $\tau > \tau^*$. We obtain numerically the boundary for $\tau > \tau^*$ (see Fig. 5.2). Note that at $\tau = \tau^*$ the condition in Eq. (5.36) is only verified for $x \to \pm\infty$, meaning that the two boundaries start from infinity at the critical time. The asymptotic behaviors of τ^* as a function of α are given by

$$\tau^* = \begin{cases} x_f^2/(2D) , & \text{for } \alpha \to \alpha_c , \\ [x_f^2/(4D)]/\log(\alpha) , & \text{for } \alpha \to \infty . \end{cases} \tag{5.39}$$

We observe that the optimal boundary in Fig. (5.2) evolves non-monotonically over time. This can be understood as follows. For very early times, it is not yet convenient to reset because after a resetting event the system would drift away from the target, requiring a second resetting. On the other hand, at late times, i.e., at $t \sim t_f$, it is not useful to reset either, since the probability to reach the target from the resetting location would be too small.

5.2.2 Infinite time horizon with discounted reward

Interestingly, one can also obtain a time-independent version of the problem in Eq. (5.31) by considering an infinite time horizon $t_f \to \infty$ and by assuming the discounted reward and cost

$$R(\vec{x}, t) = \mathcal{R}(\vec{x}) e^{-\beta t} , \tag{5.40}$$

and

$$c(\vec{x}, t) = \mathcal{C}(\vec{x}) e^{-\beta t} , \tag{5.41}$$

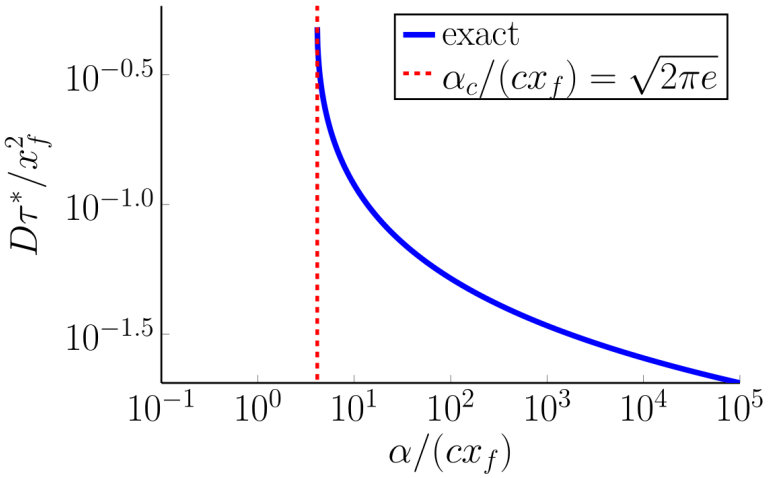

Figure 5.3: Scaled time $D\tau^*/x_f^2$ at which the barrier first appears as a function of $\alpha/(cx_f)$. The continuous blue line corresponds to the smallest positive solution in Eq. (5.38), while the dashed red line corresponds to the critical value $\alpha_c = \sqrt{2\pi e}cx_f$. For $\alpha < \alpha_c$, no barrier is present.

where $\beta > 0$ is the discount factor. Moreover, we also assume that the deterministic term does not change in time $\vec{f}(\vec{x},t) = \vec{f}(\vec{x})$. These type of discounted payoffs are quite common in the control theory literature [226] and describe situations in which the strength of the reward decays over a typical timescale $1/\beta$. This is for instance relevant in finance, where the discount rate β is used to model the fact that future benefits are considered less valuable than present benefits[4].

Using the ansatz $J(\vec{x},t) = \mathcal{J}(\vec{x})e^{-\beta t}$, Eq. (5.31) takes the time-independent form

$$\beta\mathcal{J}(\vec{x}) = D\Delta_{\vec{x}}\mathcal{J}(\vec{x}) + \vec{f}(\vec{x})\cdot\nabla_{\vec{x}}\mathcal{J}(\vec{x}) + \mathcal{R}(\vec{x})\,, \tag{5.42}$$

where the domain of definition $\Omega = \{\vec{x}:\ \mathcal{J}(\vec{x}) \geq \mathcal{J}(\vec{x}_{\rm res}) - \mathcal{C}(\vec{x})\}$ is also independent of time.

This differential equation can be analytically solved in the case of a one-dimensional system, with a quadratic reward $\mathcal{R}(x) = -\alpha x^2$, no external force $f(x) = 0$. For simplicity we also set $x_{\rm res} = 0$. Note that the system has to remain close to the resetting state to maximize the reward. Thus, the resetting strategy is quite clear: one should reset when the system goes too far from the resetting location, corresponding to $\Omega = [-u,u]$, where u depends on the problem parameters.

This statement can be made more precise by solving the differential equation (5.42). Setting $\beta = D = 1$, we obtain the exact expression

$$\mathcal{J}(x) = \alpha\left[-2 - x^2 + 2u(v)\coth(x)\right]\,, \tag{5.43}$$

for $x \in \Omega$. The variable $u(v)$ is the boundary of the symmetric domain Ω, i.e.,

$$\Omega = [-u(v), u(v)]\,. \tag{5.44}$$

[4]As an old Latin proverb says "Ad praesens ova cras pullis sunt meliora", which can be translated as "Better an egg today than a chick tomorrow".

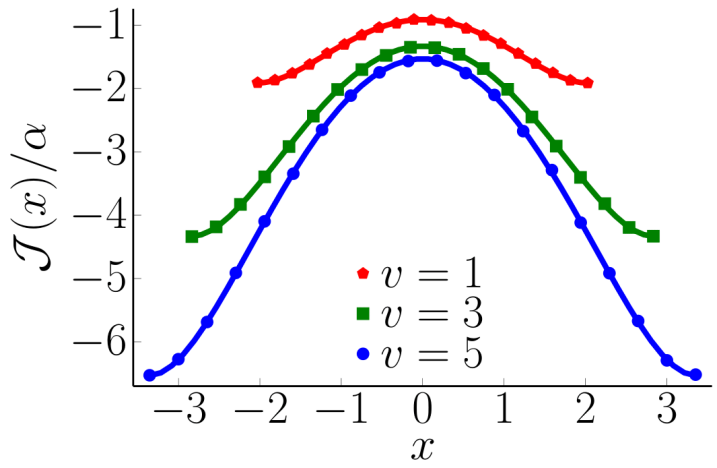

Figure 5.4: The optimal discounted payoff $\mathcal{J}(x)$ for a one-dimensional Brownian motion as a function of x for the discounted reward $\mathcal{R}(x) = -\alpha x^2$ and cost $\mathcal{C}(x) = c$. The continuous curves correspond to the exact result in Eq. (5.43) for different values of the cost-to-reward ratio $v = c/\alpha$. The symbols correspond to numerical simulations performed with $\beta = D = 1$.

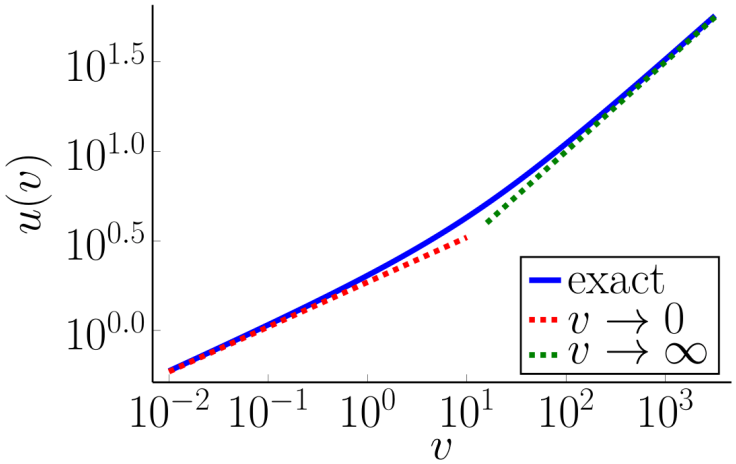

Figure 5.5: The optimal boundary $u(v)$ as a function of the cost-to-reward ration c/α (obtained for $\beta = D = 1$). The continuous blue line indicates the exact expression, obtained by numerically solving Eq. (5.45). The dashed lines indicate the asymptotic expressions for small and large v, which are given in the text.

This boundary $u(v)$ is a function of the cost-to-reward ratio $v = c/\alpha$ and is the unique positive solution of the transcendental equation

$$v - u^2(v) + 2u(v) \tanh\left(\frac{u(v)}{2}\right) = 0\,. \tag{5.45}$$

The exact result for the optimal payoff $\mathcal{J}(x)$ in Eq. (5.43) is shown in Fig. 5.4 and is in excellent agreement with numerical simulations.

In particular, when $v \ll 1$ (meaning that the cost is very small), one can position the boundary very close to the resetting location, allowing the system to remain close to the optimal location $x = 0$. Indeed, from Eq. (5.45), we obtain that for small v the optimal boundary goes as $u(v) \approx \sqrt{2}(3v)^{1/4}$. In the opposite case $v \gg 1$, resetting is expensive and one has to play a more conservative strategy by putting the boundary far from the resetting location. From Eq. (5.45), we find that for large v the optimal boundary grows as $u(v) \approx \sqrt{v} + 1$. The exact optimal boundary $u(v)$ is shown as a function of v in Eq. (5.3). Interestingly, the results obtained in this special case can also be recovered within the framework of first-passage resetting [227, 228] (see Article 9).

To conclude, we have derived a framework, analogous to the Hamilton-Jacobi-Bellman equation that provides the optimal restarting strategy in a wide range of control problems. Even if we have focused on two main illustrative examples, the technique is quite general and can be used, for instance, for systems in higher dimensions. It can be also generalized to other (non-linear) resetting costs and to discrete-time dynamics (see Article 9).

Article 9

Resetting in Stochastic Optimal Control

B. De Bruyne and F. Mori,
preprint arXiv:2112.11416 (2021).

☞ https://arxiv.org/abs/2112.11416

Abstract

"When in a difficult situation, it is sometimes better to give up and start all over again". While this empirical truth has been regularly observed in a wide range of circumstances, quantifying the effectiveness of such a heuristic strategy remains an open challenge. In this report, we combine the notions of optimal control and stochastic resetting to address this problem. The emerging analytical framework allows not only to measure the performance of a given restarting policy but also to obtain the optimal strategy for a wide class of dynamical systems. We apply our technique to a system with a final reward and show that the reward value must be larger than a critical threshold for resetting to be effective. Our approach, analogous to the celebrated Hamilton-Jacobi-Bellman paradigm, provides the basis for the investigation of realistic restarting strategies across disciplines.

Article 10

Mean perimeter and area of the convex hull of a planar Brownian motion in the presence of resetting

S. N. Majumdar, F. Mori, H. Schawe, and G. Schehr,
Phys. Rev. E **103**, 022135 (2021).

☞ https://journals.aps.org/pre/abstract/10.1103/PhysRevE.103.022135
☞ https://arxiv.org/abs/2011.06668

Abstract

We compute exactly the mean perimeter and the mean area of the convex hull of a two-dimensional isotropic Brownian motion of duration t and diffusion constant D, in the presence of resetting to the origin at a constant rate r. We show that for any t, the mean perimeter is given by $\langle L(t)\rangle = 2\pi\sqrt{\frac{D}{r}}\, f_1(rt)$ and the mean area is given by $\langle A(t)\rangle = 2\pi\frac{D}{r}\, f_2(rt)$ where the scaling functions $f_1(z)$ and $f_2(z)$ are computed explicitly. For large $t \gg 1/r$, the mean perimeter grows extremely slowly as $\langle L(t)\rangle \propto \ln(rt)$ with time. Likewise, the mean area also grows slowly as $\langle A(t)\rangle \propto \ln^2(rt)$ for $t \gg 1/r$. Our exact results indicate that the convex hull, in the presence of resetting, approaches a circular shape at late times, due to the isotropy of the Brownian motion. Numerical simulations are in perfect agreement with our analytical predictions.

Article 11

Number of distinct sites visited by a resetting random walker

M. Biroli, F. Mori, and S. N. Majumdar,
J. Phys. A Math. Theor. **55**, 244001 (2022).
☞ https://arxiv.org/abs/2202.04906
☞ https://iopscience.iop.org/article/10.1088/1751-8121/ac6b69/meta

Abstract

We investigate the number $V_p(n)$ of distinct sites visited by an n-step resetting random walker on a d-dimensional hypercubic lattice with resetting probability p. In the case $p = 0$, we recover the well-known result that the average number of distinct sites grows for large n as $\langle V_0(n)\rangle \sim n^{d/2}$ for $d < 2$ and as $\langle V_0(n)\rangle \sim n$ for $d > 2$. For $p > 0$, we show that $\langle V_p(n)\rangle$ grows extremely slowly as $\sim [\log(n)]^d$. We observe that the recurrence-transience transition at $d = 2$ for standard random walks (without resetting) disappears in the presence of resetting. In the limit $p \to 0$, we compute the exact crossover scaling function between the two regimes. In the one-dimensional case, we derive analytically the full distribution of $V_p(n)$ in the limit of large n. Moreover, for a one-dimensional random walker, we introduce a new observable, which we call *imbalance*, that measures how much the visited region is symmetric around the starting position. We analytically compute the full distribution of the imbalance both for $p = 0$ and for $p > 0$. Our theoretical results are verified by extensive numerical simulations.

Conclusion

Extreme events play a central role in several fields, from finance to climate science. Their statistical features can be systematically studied with tools from probability theory only in the case when the entries of a given time series are either uncorrelated or weakly correlated. However, for strongly correlated variables, examples of which are abundant in nature (such as a simple Brownian motion), there is no standard tool or method available. In the absence of a general theory, it is necessary to study simple solvable models to gain insights onto the extreme statistics of strongly correlated variables. In this thesis, we have achieved significant progress in that direction by computing several extreme observables in three different types of stochastic processes: (i) Brownian motion, (ii) the run-and-tumble particle model, and (iii) resetting Brownian motion. These are among the few exactly solvable systems with strong correlations. To study the extremal properties of these models, we have considered a wide range of observables, including the time of the maximum and the record statistics.

The common feature of the different works presented in this thesis is that of considering models for which the extremal properties can be studied exactly, i.e., without any approximations. These exact solutions allowed us to uncover several unexpected phenomena, including universal properties and nonequilibrium phase transitions. Moreover, since the models that we have considered are very general, we believe that our results could be useful in the description of a variety of systems across disciplines.

In Section 2.1 of Chapter 2, we have presented the exact computation of the distribution of time $\tau = t_{\min} - t_{\max}$ between the maximum and the minimum of a one-dimensional Brownian motion of fixed duration T. We have shown that this distribution assumes the scaling form $P(\tau|T) = (1/T) f_{\rm BM}(\tau/T)$, where we have computed the scaling function $f_{\rm BM}(y)$ exactly. We have generalized this result to other processes, including Brownian bridges, discrete-time random walks, and fluctuating interfaces.

Moreover, we have considered discrete-time random walks of n-step with continuous and symmetric jump distribution. We have studied the probability $p_n = P(\tau = n|n)$ that the time τ between the maximum and the minimum is exactly n. This corresponds to configurations in which the maximum is reached at the first step $t_{\max} = 0$ and the minimum is attained at the final step $t_{\min} = n$. We have shown that $p_n = 1/(2n)$ for any finite n, independently of the jump distribution, even if the distribution is fat-tailed. This result is a consequence of a combinatorial theorem by Spitzer [131].

For future works, it would be interesting to study the distribution of τ in the case of Lévy flights. Indeed, the scaling form obtained for BM is not valid for Lévy flights with

Lévy index μ, with $0 < \mu < 2$. From numerical simulations, we expect that in the limit of large n, where n is the number of steps, the distribution of τ can still be written in the scaling form

$$P(\tau|n) \approx \frac{1}{n} f_\mu\left(\frac{\tau}{n}\right) , \tag{5.46}$$

where the scaling function $f_\mu(y)$ depends explicitly on the index μ. It would be relevant to derive this scaling function $f_\mu(y)$ for general μ (note that $f_\mu(1) = 1/2$ as a consequence of our result on the probability of the event "$\tau = n$"). Moreover, it would be also relevant to study the distribution of τ in the case of Brownian motion with drift. This could be useful in the context of finance, where the price of a stock can be modeled as a drifted Brownian motion. Another interesting generalization of this result would be to compute the distribution of τ for a single run-and-tumble particle.

In Section 2.2 of Chapter 2, we have investigated the distribution $P(t_{\max}|T)$ of the time $t_{\max}$ of the maximum for stationary processes of fixed duration T. Using a path-decomposition method, we have found the exact expression for $P(t_{\max}|T)$ for several processes, both at equilibrium (including the Ornstein-Uhlenbeck process) and out-of-equilibrium (including Brownian motion with resetting). Moreover, in the case of an overdamped Brownian particle in a confining potential, we have shown that, at late times, the probability density function $P(t_{\max}|T)$, properly rescaled, becomes universal, i.e., independent of the details of the potential. Furthermore, we have shown that for any equilibrium process, the distribution $P(t_{\max}|T)$ is always symmetric around its midpoint $t_{\max} = T/2$, i.e., $P(t_{\max}|T) = P(T - t_{\max}|T)$. This property is a consequence of the time-reversal symmetry of the process and provides a simple test to detect nonequilibrium fluctuations in steady states. For future works, it would be relevant to test this criterion in real nonequilibrium systems.

In Chapter 3, we have investigated the first-passage properties of the RTP process in d dimensions. In a passive Brownian motion in d dimensions, the different components of the process are uncorrelated, hence the first-passage probability of a single component in d-dimensions reduces trivially to a one-dimensional problem. However, for the RTP process, the different components are strongly correlated and hence the first-passage probability of the x-component of an RTP does not immediately reduce to a one-dimensional problem, unlike in the Brownian case. One of the major achievements of this thesis was to discover a rather nontrivial mapping between the RTP in continuous time and a one-dimensional random walk problem in discrete time. This mapping helped us to derive several universal first-passage properties that are almost impossible to derive from the standard Fokker-Planck approach. This method allows one to directly apply numerous results from the theory of discrete-time random walks to investigate the statistical properties of the RTP model. In particular, using this technique, we have shown that several quantities, i.e., the survival probability, the distribution of the time of the maximum, and the record statistics, are completely universal, i.e., independent of d and the speed distribution of the particle. This is a consequence of the Sparre Andersen theorem [38], valid for discrete-time random walks. We have also shown that this universality also extends to several generalizations of the RTP model, including a model with non-instantaneous tumblings. Moreover, we have applied this method to compute the survival probability and the record statistics of a discrete-time persistent random walk.

Interestingly, in a recent work [229], the mapping to a random walk has been used to compute the survival probability of an RTP in the presence of a constant drift. For future works, it would be relevant to investigate whether the universality also extends to other observables of the x component of the RTP process. For instance, we expect the distribution of the occupation time of the x-component, i.e., the time spent by the x component on the positive side, to be also universal. This distribution was computed for $d = 1$ in [99] by a generalization of the Feynman-Kac method. We have verified this universality for $d \geq 1$ with numerical simulation but we have not been able so far to employ our mapping in this case. Thus, showing analytically the universality of this observable is an interesting open problem.

In Chapter 4, we have investigated the distribution of the position of a single RTP moving in a d-dimensional space with arbitrary speed distribution $W(v)$. We have considered the position X of the particle after N running phases. We have shown that the typical events are described by the central limit theorem: the probability density function $P(X, N)$ is approximately Gaussian for $X \sim \sqrt{N}$. However, the rare events where $X \gg \sqrt{N}$ are not described by the central limit theorem.

To estimate the probability of these events, we have computed exactly the rate function associated with the large-deviation regime $X \sim N$ (or $X \sim N^{3/4}$ in one case). We have considered several variations of the RTP model, for different d and $W(v)$. Interestingly, we have shown that, for several choices of the parameters of the model, the rate function becomes singular at some critical value X_c of the displacement of the particle. We characterize this phenomenon, showing that it consists in a condensation transition. Indeed, by increasing the total displacement X the system goes from a fluid phase (for $X < X_c$) in which the displacements associated with the different running phases are roughly of the same order, to a condensed phase, in which a single running phase dominates the trajectory. We have found that the order of the transition can be arbitrarily tuned upon changing the model parameters. In particular, we have shown that for an RTP moving in a line ($d = 1$) and with Gaussian speed fluctuations, the system undergoes a first-order transition. We have identified an order parameter associated with the transition, the participation ratio. We have shown that, in the case of a first-order transition, this order parameter is discontinuous as a function of the total displacement X, which plays the role of the control parameter.

Although we have mostly focused on RTP models in the fixed-N ensemble, i.e., fixing the number N of completed running phases, we expect our results to be qualitatively valid also in the fixed-T ensemble, i.e., when the total time T is fixed. It would be relevant to investigate explicitly the large deviation in the fixed-T ensemble, in particular in the case where we expect to observe a first-order transition (which has not been considered before).

Finally, in Chapter 5 we have combined optimal control theory and stochastic resetting into a new framework that allows finding the optimal restarting policy to drive a complex system. This technique is very general and works for a wide class of dynamical stochastic systems and several different control problems. As a result, we have obtained a moving boundary problem for the optimal payoff function, which we have solved in two illustrative examples.

There are several possible extensions of this work. First of all, it would be relevant to apply this technique to more complex and realistic systems, composed of several

interacting degrees of freedom. Moreover, it would be interesting to study cost functions which are first-passage functionals (meaning that the process stops when a given target is found for the first time). This could have applications in the context of search processes.

Bibliography

[1] Sidney Redner and Mark R Petersen. Role of global warming on the statistics of record-breaking temperatures. *Physical Review E*, 74(6):061114, 2006.

[2] Stefan Rahmstorf and Dim Coumou. Increase of extreme events in a warming world. *Proceedings of the National Academy of Sciences*, 108(44):17905–17909, 2011.

[3] Bo Christiansen. Changes in temperature records and extremes: Are they statistically significant? *Journal of Climate*, 26(20):7863–7875, 2013.

[4] Gregor Wergen, Andreas Hense, and Joachim Krug. Record occurrence and record values in daily and monthly temperatures. *Climate dynamics*, 42(5):1275–1289, 2014.

[5] Ronald Aylmer Fisher and Leonard Henry Caleb Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical proceedings of the Cambridge philosophical society*, volume 24, pages 180–190. Cambridge University Press, 1928.

[6] Boris Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, pages 423–453, 1943.

[7] Emil Julius Gumbel. *Statistics of extremes*. Columbia university press, 1958.

[8] Satya N Majumdar, Arnab Pal, and Grégory Schehr. Extreme value statistics of correlated random variables: a pedagogical review. *Physics Reports*, 840:1–32, 2020.

[9] H Allen Orr. The distribution of fitness effects among beneficial mutations. *Genetics*, 163(4):1519–1526, 2003.

[10] Bernard Derrida. Random-energy model: An exactly solvable model of disordered systems. *Physical Review B*, 24(5):2613, 1981.

[11] Jean-Philippe Bouchaud and Marc Mézard. Universality classes for extreme-value statistics. *Journal of Physics A: Mathematical and General*, 30(23):7997, 1997.

[12] Giulio Biroli, Jean-Philippe Bouchaud, and Marc Potters. Extreme value problems in random matrix theory and other disordered systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(07):P07019, 2007.

[13] Subhadip Raychaudhuri, Michael Cranston, Corry Przybyla, and Yonathan Shapir. Maximal height scaling of kinetically growing surfaces. *Physical review letters*, 87(13):136101, 2001.

[14] Géza Györgyi, Peter Holdsworth, Baptiste Portelli, and Zoltan Rácz. Statistics of extremal intensities for Gaussian interfaces. *Physical Review E*, 68(5):056116, 2003.

[15] Satya N Majumdar and Alain Comtet. Exact maximal height distribution of fluctuating interfaces. *Physical review letters*, 92(22):225501, 2004.

[16] Satya N Majumdar and Alain Comtet. Airy distribution function: from the area under a Brownian excursion to the maximal height of fluctuating interfaces. *Journal of Statistical Physics*, 119(3):777–826, 2005.

[17] Craig A Tracy and Harold Widom. Level-spacing distributions and the Airy kernel. *Communications in Mathematical Physics*, 159(1):151–174, 1994.

[18] Craig A Tracy and Harold Widom. On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics*, 177(3):727–754, 1996.

[19] David S Dean and Satya N Majumdar. Large deviations of extreme eigenvalues of random matrices. *Physical review letters*, 97(16):160201, 2006.

[20] David S Dean and Satya N Majumdar. Extreme value statistics of eigenvalues of Gaussian random matrices. *Physical Review E*, 77(4):041108, 2008.

[21] Satya N Majumdar and Massimo Vergassola. Large deviations of the maximum eigenvalue for Wishart and Gaussian random matrices. *Physical review letters*, 102(6):060601, 2009.

[22] Satya N Majumdar and Grégory Schehr. Top eigenvalue of a random matrix: large deviations and third order phase transition. *Journal of Statistical Mechanics: Theory and Experiment*, 2014(1):P01012, 2014.

[23] Satya N Majumdar and Robert M Ziff. Universal record statistics of random walks and Lévy flights. *Physical review letters*, 101(5):050601, 2008.

[24] William Feller. An introduction to probability theory and its applications. *1957*.

[25] Hugo Touchette. The large deviation approach to statistical mechanics. *Physics Reports*, 478(1-3):1–69, 2009.

[26] Satya N Majumdar and Gregory Schehr. Large deviations. *ICTS newsletter 2017, volume 3, issue 2*, 2017.

[27] Udo Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on progress in physics*, 75(12):126001, 2012.

[28] Andre C Barato and Udo Seifert. Thermodynamic uncertainty relation for biomolecular processes. *Physical review letters*, 114(15):158101, 2015.

[29] Jordan M Horowitz and Todd R Gingrich. Proof of the finite-time thermodynamic uncertainty relation for steady-state currents. *Physical Review E*, 96(2):020103, 2017.

[30] Louis Bachelier. Théorie de la spéculation. In *Annales scientifiques de l'École normale supérieure*, volume 17, pages 21–86, 1900.

[31] Albert Einstein. On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat (English translation, 1956). *Investigations on the theory of the Brownian movement*, 1905.

[32] Sidney Redner. *A guide to first-passage processes*. Cambridge university press, 2001.

[33] Alain Comtet and Satya N Majumdar. Precise asymptotics for a random walker's maximum. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(06):P06013, 2005.

[34] Satya N Majumdar. Universal first-passage properties of discrete-time random walks and Lévy flights on a line: Statistics of the global maximum and records. *Physica A: Statistical Mechanics and its Applications*, 389(20):4299–4316, 2010.

[35] Philippe Mounaix, Satya N Majumdar, and Grégory Schehr. Asymptotics for the expected maximum of random walks and Lévy flights with a constant drift. *Journal of Statistical Mechanics: Theory and Experiment*, 2018(8):083201, 2018.

[36] Martin R Evans and Satya N Majumdar. Diffusion with stochastic resetting. *Physical review letters*, 106(16):160601, 2011.

[37] Paul Lévy. Sur certains processus stochastiques homogènes. *Compositio mathematica*, 7:283–339, 1940.

[38] Erik Sparre Andersen. On the fluctuations of sums of random variables. *Mathematica Scandinavica*, pages 263–285, 1953.

[39] Charles Dale and Rosemarie Workman. The arc sine law and the treasury bill futures market. *Financial analysts journal*, 36(6):71–74, 1980.

[40] Andre C Barato, Édgar Roldán, Ignacio A Martínez, and Simone Pigolotti. Arcsine laws in stochastic thermodynamics. *Physical review letters*, 121(9):090601, 2018.

[41] Gregor Wergen, Satya N Majumdar, and Grégory Schehr. Record statistics for multiple random walks. *Physical Review E*, 86(1):011119, 2012.

[42] Svante Janson. Brownian excursion area, Wright's constants in graph enumeration, and other brownian areas. *Probability Surveys*, 4:80–145, 2007.

[43] Satya N Majumdar, Julien Randon-Furling, Michael J Kearney, and Marc Yor. On the time to reach maximum for a variety of constrained Brownian motions. *Journal of Physics A: Mathematical and Theoretical*, 41(36):365005, 2008.

[44] Howard C Berg. *E. coli in Motion*. Springer, 2004.

[45] Michael E Cates. Diffusive transport without detailed balance in motile bacteria: does microbiology need statistical physics? *Reports on Progress in Physics*, 75(4):042601, 2012.

[46] Tamás Vicsek, András Czirók, Eshel Ben-Jacob, Inon Cohen, and Ofer Shochet. Novel type of phase transition in a system of self-driven particles. *Physical review letters*, 75(6):1226, 1995.

[47] Sriram Ramaswamy. The mechanics and statistics of active matter. *Annu. Rev. Condens. Matter Phys.*, 1(1):323–345, 2010.

[48] M Cristina Marchetti, Jean-François Joanny, Sriram Ramaswamy, Tanniemola B Liverpool, Jacques Prost, Madan Rao, and R Aditi Simha. Hydrodynamics of soft active matter. *Reviews of modern physics*, 85(3):1143, 2013.

[49] Clemens Bechinger, Roberto Di Leonardo, Hartmut Löwen, Charles Reichhardt, Giorgio Volpe, and Giovanni Volpe. Active particles in complex and crowded environments. *Reviews of Modern Physics*, 88(4):045006, 2016.

[50] Sriram Ramaswamy. Active matter. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(5):054002, 2017.

[51] Étienne Fodor, Cesare Nardini, Michael E Cates, Julien Tailleur, Paolo Visco, and Frédéric Van Wijland. How far from equilibrium is active matter? *Physical review letters*, 117(3):038103, 2016.

[52] Michael E Cates and Julien Tailleur. Motility-induced phase separation. *Annu. Rev. Condens. Matter Phys.*, 6(1):219–244, 2015.

[53] Francisco J Sevilla, Alejandro V Arzola, and Enrique Puga Cital. Stationary superstatistics distributions of trapped run-and-tumble particles. *Physical Review E*, 99(1):012145, 2019.

[54] Abhishek Dhar, Anupam Kundu, Satya N Majumdar, Sanjib Sabhapandit, and Grégory Schehr. Run-and-tumble particle in one-dimensional confining potentials: steady-state, relaxation, and first-passage properties. *Physical Review E*, 99(3):032132, 2019.

[55] Emil Mallmin, Richard A Blythe, and Martin R Evans. Exact spectral solution of two interacting run-and-tumble particles on a ring lattice. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(1):013204, 2019.

[56] Urna Basu, Satya N Majumdar, Alberto Rosso, Sanjib Sabhapandit, and Grégory Schehr. Exact stationary state of a run-and-tumble particle with three internal states in a harmonic trap. *Journal of Physics A: Mathematical and Theoretical*, 53(9):09LT01, 2020.

[57] Roberto Artuso, Giampaolo Cristadoro, Mirko Degli Esposti, and Georgie Knight. Sparre-Andersen theorem with spatiotemporal correlations. *Physical Review E*, 89(5):052111, 2014.

[58] Kanaya Malakar, V Jemseena, Anupam Kundu, K Vijay Kumar, Sanjib Sabhapandit, Satya N Majumdar, S Redner, and Abhishek Dhar. Steady state, relaxation and first-passage properties of a run-and-tumble particle in one-dimension. *Journal of Statistical Mechanics: Theory and Experiment*, 2018(4):043215, 2018.

[59] Paolo Sibani, Michael Brandt, and Preben Alstrøm. Evolution and extinction dynamics in rugged fitness landscapes. *International Journal of Modern Physics B*, 12(04):361–391, 1998.

[60] Joachim Krug and Kavita Jain. Breaking records in the evolutionary race. *Physica A: Statistical Mechanics and its Applications*, 358(1):1–9, 2005.

[61] Jasper Franke, Alexander Klözer, J Arjan GM de Visser, and Joachim Krug. Evolutionary accessibility of mutational pathways. *PLoS computational biology*, 7(8):e1002134, 2011.

[62] KN Chandler. The distribution and frequency of record values. *Journal of the Royal Statistical Society: Series B (Methodological)*, 14(2):220–228, 1952.

[63] Valery B Nevzorov. *Records: mathematical theory*. American Mathematical Soc., 2001.

[64] Sidney I Resnick. *Extreme values, regular variation, and point processes*, volume 4. Springer Science & Business Media, 2008.

[65] Barry C Arnold, Narayanaswamy Balakrishnan, and Haikady N Nagaraja. *Records*, volume 768. John Wiley & Sons, 2011.

[66] Joachim Krug. Records in a changing world. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(07):P07001, 2007.

[67] Claude Godreche, Satya N Majumdar, and Gregory Schehr. Record statistics of a strongly correlated time series: random walks and Lévy flights. *Journal of Physics A: Mathematical and Theoretical*, 50(33):333001, 2017.

[68] Gregor Wergen and Joachim Krug. Record-breaking temperatures reveal a warming climate. *EPL (Europhysics Letters)*, 92(3):30008, 2010.

[69] Gregor Wergen, Miro Bogner, and Joachim Krug. Record statistics for biased random walks, with an application to financial data. *Physical Review E*, 83(5):051109, 2011.

[70] Behlool Sabir and MS Santhanam. Record statistics of financial time series and geometric random walks. *Physical Review E*, 90(3):032126, 2014.

[71] J McFadden. The axis-crossing intervals of random functions–ii. *IRE Transactions on Information Theory*, 4(1):14–24, 1958.

[72] Frank Aurzada and Thomas Simon. Persistence probabilities and exponents. In *Lévy matters V*, pages 183–224. Springer, 2015.

[73] Alan J Bray, Satya N Majumdar, and Grégory Schehr. Persistence and first-passage properties in nonequilibrium systems. *Advances in Physics*, 62(3):225–361, 2013.

[74] Peter Hänggi, Peter Talkner, and Michal Borkovec. Reaction-rate theory: fifty years after Kramers. *Reviews of modern physics*, 62(2):251, 1990.

[75] Olivier Bénichou, Claude Loverdo, Michel Moreau, and Raphael Voituriez. Intermittent search strategies. *Reviews of Modern Physics*, 83(1):81, 2011.

[76] Vincent Tejedor, Raphael Voituriez, and Olivier Bénichou. Optimizing persistent random searches. *Physical review letters*, 108(8):088103, 2012.

[77] Bernard Derrida. Microscopic versus macroscopic approaches to non-equilibrium systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(01):P01030, 2011.

[78] Cécile Monthus and Thomas Garel. Matching between typical fluctuations and large deviations in disordered systems: application to the statistics of the ground state energy in the SK spin-glass model. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(02):P02023, 2010.

[79] Cécile Monthus. Revisiting classical and quantum disordered systems from the unifying perspective of large deviations. *The European Physical Journal B*, 92(7):1–24, 2019.

[80] Cécile Monthus. Large deviations for the density and current in non-equilibrium-steady-states on disordered rings. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(2):023206, 2019.

[81] Satya N Majumdar, Sanjib Sabhapandit, and Grégory Schehr. Dynamical transition in the temporal relaxation of stochastic processes under resetting. *Physical Review E*, 91(5):052131, 2015.

[82] Grégory Schehr and Pierre Le Doussal. Extreme value statistics from the real space renormalization group: Brownian motion, Bessel processes and continuous time random walks. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(01):P01009, 2010.

[83] Martin R Evans, Satya N Majumdar, and Grégory Schehr. Stochastic resetting and applications. *Journal of Physics A: Mathematical and Theoretical*, 53(19):193001, 2020.

[84] Andrea Montanari and Riccardo Zecchina. Optimizing searches via rare events. *Physical review letters*, 88(17):178701, 2002.

[85] Shlomi Reuveni, Michael Urbakh, and Joseph Klafter. Role of substrate unbinding in Michaelis–Menten enzymatic reactions. *Proceedings of the National Academy of Sciences*, 111(12):4391–4396, 2014.

[86] Vicenç Méndez and Daniel Campos. Characterization of stationary states in random walks with stochastic resetting. *Physical Review E*, 93(2):022106, 2016.

[87] Stephan Eule and Jakob J Metzger. Non-equilibrium steady states of stochastic processes with intermittent resetting. *New Journal of Physics*, 18(3):033006, 2016.

[88] Lukasz Kusmierz, Satya N Majumdar, Sanjib Sabhapandit, and Grégory Schehr. First order transition for the optimal search time of Lévy flights with resetting. *Physical review letters*, 113(22):220602, 2014.

[89] Łukasz Kuśmierz and Ewa Gudowska-Nowak. Optimal first-arrival times in Lévy flights with resetting. *Physical Review E*, 92(5):052127, 2015.

[90] Daniel Campos and Vicenç Méndez. Phase transitions in optimal search times: How random walkers should combine resetting and flight scales. *Physical Review E*, 92(6):062115, 2015.

[91] Martin R Evans and Satya N Majumdar. Run and tumble particle under resetting: a renewal approach. *Journal of Physics A: Mathematical and Theoretical*, 51(47):475003, 2018.

[92] Jaume Masoliver. Telegraphic processes with stochastic resetting. *Physical Review E*, 99(1):012121, 2019.

[93] Vijay Kumar, Onkar Sadekar, and Urna Basu. Active Brownian motion in two dimensions under stochastic resetting. *Physical Review E*, 102(5):052129, 2020.

[94] Ion Santra, Urna Basu, and Sanjib Sabhapandit. Run-and-tumble particles in two dimensions under stochastic resetting conditions. *Journal of Statistical Mechanics: Theory and Experiment*, 2020(11):113206, 2020.

[95] Matteo Magoni, Satya N Majumdar, and Grégory Schehr. Ising model with stochastic resetting. *Physical Review Research*, 2(3):033182, 2020.

[96] Julien Tailleur and Michael E Cates. Statistical mechanics of interacting run-and-tumble bacteria. *Physical review letters*, 100(21):218103, 2008.

[97] Alexander B Slowman, Martin R Evans, and Richard A Blythe. Jamming and attraction of interacting run-and-tumble random walkers. *Physical review letters*, 116(21):218101, 2016.

[98] Matthew J Metson, Martin R Evans, and Richard A Blythe. Jamming of multiple persistent random walkers in arbitrary spatial dimension. *Journal of Statistical Mechanics: Theory and Experiment*, 2020(10):103207, 2020.

[99] Prashant Singh and Anupam Kundu. Generalised 'Arcsine' laws for run-and-tumble particle in one dimension. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(8):083205, 2019.

[100] Richard Bellman. Dynamic programming and a new formalism in the calculus of variations. *Proceedings of the National Academy of Sciences of the United States of America*, 40(4):231, 1954.

[101] Richard Bellman. Dynamic programming. *Science*, 153(3731):34–37, 1966.

[102] Benjamin De Bruyne, Satya N Majumdar, and Gregory Schehr. Expected maximum of bridge random walks & Lévy flights. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(8):083215, 2021.

[103] Satya N Majumdar, Philippe Mounaix, Sanjib Sabhapandit, and Gregory Schehr. Record statistics for random walks and Lévy flights with resetting. *Journal of Physics A: Mathematical and Theoretical*, 55(3):034002, 2021.

[104] Claude Godrèche and Jean-Marc Luck. Maximum and records of random walks with stochastic resetting. *arXiv preprint arXiv:2203.01102*, 2022.

[105] Satya N Majumdar, Grégory Schehr, and Gregor Wergen. Record statistics and persistence for a random walk with a drift. *Journal of Physics A: Mathematical and Theoretical*, 45(35):355002, 2012.

[106] Pierre Le Doussal and Kay Jörg Wiese. Driven particle in a random landscape: Disorder correlator, avalanche distribution, and extreme value statistics of records. *Physical Review E*, 79(5):051105, 2009.

[107] Satya N Majumdar, Alberto Rosso, and Andrea Zoia. Hitting probability for anomalous diffusion processes. *Physical review letters*, 104(2):020602, 2010.

[108] Jamil Baz and George Chacko. *Financial derivatives: pricing, applications, and mathematics*. Cambridge University Press, 2004.

[109] Aaron Clauset, M Kogan, and Sidney Redner. Safe leads and lead changes in competitive team sports. *Physical Review E*, 91(6):062815, 2015.

[110] Julien Randon-Furling and Satya N Majumdar. Distribution of the time at which the deviation of a Brownian motion is maximum before its first-passage time. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(10):P10008, 2007.

[111] Emannuel Buffet. On the time of the maximum of Brownian motion with drift. *Journal of Applied Mathematics and Stochastic Analysis*, 16(3):201–207, 2003.

[112] Satya N Majumdar and Jean-Philippe Bouchaud. Optimal time to sell a stock in the Black–Scholes model: comment on 'Thou Shalt Buy and Hold', by A. Shiryaev, Z. Xu and XY Zhou. *Quantitative Finance*, 8(8):753–760, 2008.

[113] Mathieu Delorme and Kay Jörg Wiese. Extreme-value statistics of fractional Brownian motion bridges. *Physical Review E*, 94(5):052105, 2016.

[114] Tridib Sadhu, Mathieu Delorme, and Kay Jörg Wiese. Generalized arcsine laws for fractional Brownian motion. *Physical review letters*, 120(4):040603, 2018.

[115] Prashant Singh. Extreme value statistics and arcsine laws for heterogeneous diffusion processes. *Physical Review E*, 105(2):024113, 2022.

[116] Prashant Singh and Arnab Pal. Extremal statistics for stochastic resetting systems. *Physical Review E*, 103(5):052119, 2021.

[117] Satya N Majumdar, Alain Comtet, and Julien Randon-Furling. Random convex hulls and extreme value statistics. *Journal of Statistical Physics*, 138(6):955–1009, 2010.

[118] Joachim Rambeau and Grégory Schehr. Distribution of the time at which N vicious walkers reach their maximal height. *Physical Review E*, 83(6):061146, 2011.

[119] Yongjie Zou and Honggang Li. Time spans between price maxima and price minima in stock markets. *Physica A: Statistical Mechanics and its Applications*, 395:303–309, 2014.

[120] Dennis D Murphy and Barry R Noon. Integrating scientific methods with habitat conservation planning: reserve design for northern spotted owls. *Ecological Applications*, 2(1):3–17, 1992.

[121] L Giuggioli, G Abramson, VM Kenkre, G Suzán, E Marcé, and TL Yates. Diffusion and home range parameters from rodent population measurements in Panama. *Bulletin of Mathematical Biology*, 67(5):1135–1149, 2005.

[122] Arnab Pal, Łukasz Kuśmierz, and Shlomi Reuveni. Search with home returns provides advantage under high uncertainty. *Physical Review Research*, 2(4):043174, 2020.

[123] Benjamin De Bruyne, Satya N Majumdar, and Gregory Schehr. Optimal resetting Brownian bridges. *arXiv preprint arXiv:2201.01994*, 2022.

[124] Wim Vervaat. A relation between Brownian bridge and Brownian excursion. *The Annals of Probability*, pages 143–149, 1979.

[125] Satya N Majumdar and Henri Orland. Effective Langevin equations for constrained stochastic processes. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(6):P06039, 2015.

[126] Benjamin De Bruyne, Satya N Majumdar, and Grégory Schehr. Generating discrete-time constrained random walks and Lévy flights. *Physical Review E*, 104(2):024117, 2021.

[127] Benjamin De Bruyne, Satya N Majumdar, and Grégory Schehr. Generating constrained run-and-tumble trajectories. *Journal of Physics A: Mathematical and Theoretical*, 54(38):385004, 2021.

[128] Benjamin De Bruyne, Satya N Majumdar, Henri Orland, and Gregory Schehr. Generating stochastic trajectories with global dynamical constraints. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(12):123204, 2021.

[129] VV Ivanov. Resolvent method: exact solutions of half-space transport problems by elementary means. *Astronomy and Astrophysics*, 286:328–337, 1994.

[130] Satya N Majumdar, Philippe Mounaix, and Grégory Schehr. Survival probability of random walks and Lévy flights on a semi-infinite line. *Journal of Physics A: Mathematical and Theoretical*, 50(46):465002, 2017.

[131] Frank Spitzer. A combinatorial lemma and its application to probability theory. *Transactions of the American Mathematical Society*, 82(2):323–339, 1956.

[132] Samuel Frederick Edwards and DR Wilkinson. The surface statistics of a granular aggregate. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 381(1780):17–31, 1982.

[133] Mehran Kardar, Giorgio Parisi, and Yi-Cheng Zhang. Dynamic scaling of growing interfaces. *Physical Review Letters*, 56(9):889, 1986.

[134] Timothy Halpin-Healy and Yi-Cheng Zhang. Kinetic roughening phenomena, stochastic growth, directed polymers and all that. Aspects of multidisciplinary statistical mechanics. *Physics reports*, 254(4-6):215–414, 1995.

[135] Joachim Krug. Origins of scale invariance in growth processes. *Advances in Physics*, 46(2):139–282, 1997.

[136] Gernot Akemann, Oriol Bohigas, Alexei Borodin, Jean-Philippe Bouchaud, Alain Comtet, Bertrand Eynard, Alice Guionnet, Jonathan P.. Keating, Aris L Moustakas, Leonid Petrov, et al. *Stochastic Processes and Random Matrices: Lecture Notes of the Les Houches Summer School*. Oxford University Press, 2017.

[137] Grégory Schehr and Satya N Majumdar. Universal asymptotic statistics of maximal relative height in one-dimensional solid-on-solid models. *Physical Review E*, 73(5):056103, 2006.

[138] Chi-Hang Lam and FG Shin. Anomaly in numerical integrations of the Kardar-Parisi-Zhang equation. *Physical Review E*, 57(6):6506, 1998.

[139] Chi-Hang Lam and Franklin G Shin. Improved discretization of the Kardar-Parisi-Zhang equation. *Physical Review E*, 58(5):5592, 1998.

[140] R Dean Astumian. Thermodynamics and kinetics of a Brownian motor. *science*, 276(5314):917–922, 1997.

[141] Jeffrey B Weiss, Baylor Fox-Kemper, Dibyendu Mandal, Arin D Nelson, and Royce KP Zia. Nonequilibrium oscillations, probability angular momentum, and the climate system. *Journal of Statistical Physics*, 179(5):1010–1027, 2020.

[142] Chris Jarzynski. Nonequilibrium work theorem for a system strongly coupled to a thermal environment. *Journal of Statistical Mechanics: Theory and Experiment*, 2004(09):P09005, 2004.

[143] Jorge Kurchan. Fluctuation theorem for stochastic dynamics. *Journal of Physics A: Mathematical and General*, 31(16):3719, 1998.

[144] Gavin E Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E*, 60(3):2721, 1999.

[145] Udo Seifert. Entropy production along a stochastic trajectory and an integral fluctuation theorem. *Physical review letters*, 95(4):040602, 2005.

[146] Satya N Majumdar, Francesco Mori, Hendrik Schawe, and Grégory Schehr. Mean perimeter and area of the convex hull of a planar Brownian motion in the presence of resetting. *Physical Review E*, 103(2):022135, 2021.

[147] Francesco Mori, Satya N. Majumdar, and Gregory Schehr. Time to reach the maximum for a stationary stochastic processes. *in preparation.*

[148] Andrei N Borodin and Paavo Salminen. *Handbook of Brownian motion-facts and formulae*. Springer Science & Business Media, 2015.

[149] Sanjib Sabhapandit and Satya N Majumdar. Freezing transition in the barrier crossing rate of a diffusing particle. *Physical Review Letters*, 125(20):200601, 2020.

[150] Hao Shen, Lawrence J Tauzin, Rashad Baiyasi, Wenxiao Wang, Nicholas Moringo, Bo Shuang, and Christy F Landes. Single particle tracking: from theory to biophysical applications. *Chemical reviews*, 117(11):7331–7376, 2017.

[151] Federico S Gnesotto, Federica Mura, Jannes Gladrow, and Chase P Broedersz. Broken detailed balance and non-equilibrium dynamics in living systems: a review. *Reports on Progress in Physics*, 81(6):066601, 2018.

[152] Leticia F Cugliandolo, David S Dean, and Jorge Kurchan. Fluctuation-dissipation theorems and entropy production in relaxational systems. *Physical review letters*, 79(12):2168, 1997.

[153] P Martin, AJ Hudspeth, and F Jülicher. Comparison of a hair bundle's spontaneous oscillations with its response to mechanical stimulation reveals the underlying active process. *Proceedings of the National Academy of Sciences*, 98(25):14380–14385, 2001.

[154] Daisuke Mizuno, Catherine Tardin, Christoph F Schmidt, and Frederik C MacKintosh. Nonequilibrium mechanics of active cytoskeletal networks. *Science*, 315(5810):370–373, 2007.

[155] Hervé Turlier, Dmitry A Fedosov, Basile Audoly, Thorsten Auth, Nir S Gov, Cécile Sykes, J-F Joanny, Gerhard Gompper, and Timo Betz. Equilibrium physics breakdown reveals the active nature of red blood cell flickering. *Nature physics*, 12(5):513–519, 2016.

[156] Royce KP Zia and Beate Schmittmann. Probability currents as principal characteristics in the statistical mechanics of non-equilibrium steady states. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(07):P07012, 2007.

[157] Christopher Battle, Chase P Broedersz, Nikta Fakhri, Veikko F Geyer, Jonathon Howard, Christoph F Schmidt, and Fred C MacKintosh. Broken detailed balance at mesoscopic scales in active biological systems. *Science*, 352(6285):604–607, 2016.

[158] Federica Mura, Grzegorz Gradziuk, and Chase P Broedersz. Nonequilibrium scaling behavior in driven soft biological assemblies. *Physical review letters*, 121(3):038002, 2018.

[159] Juan Pablo Gonzalez, John C Neu, and Stephen W Teitsworth. Experimental metrics for detection of detailed balance violation. *Physical Review E*, 99(2):022143, 2019.

[160] Yuhai Tu. The nonequilibrium mechanism for ultrasensitivity in a biological switch: Sensing by Maxwell's demons. *Proceedings of the National Academy of Sciences*, 105(33):11737–11741, 2008.

[161] Dominic J Skinner and Jörn Dunkel. Estimating entropy production from waiting time distributions. *Physical review letters*, 127(19):198101, 2021.

[162] Junang Li, Jordan M Horowitz, Todd R Gingrich, and Nikta Fakhri. Quantifying dissipation using fluctuating currents. *Nature communications*, 10(1):1–9, 2019.

[163] Sreekanth K Manikandan, Deepak Gupta, and Supriya Krishnamurthy. Inferring entropy production from short experiments. *Physical review letters*, 124(12):120603, 2020.

[164] Sreekanth K Manikandan, Subhrokoli Ghosh, Avijit Kundu, Biswajit Das, Vipin Agrawal, Dhrubaditya Mitra, Ayan Banerjee, and Supriya Krishnamurthy. Quantitative analysis of non-equilibrium systems from short-time experimental data. *Communications Physics*, 4(1):1–10, 2021.

[165] Édgar Roldán, Jérémie Barral, Pascal Martin, Juan MR Parrondo, and Frank Jülicher. Quantifying entropy production in active fluctuations of the hair-cell bundle from time irreversibility and uncertainty relations. *New Journal of Physics*, 23(8):083013, 2021.

[166] Shun Otsubo, Sreekanth K Manikandan, Takahiro Sagawa, and Supriya Krishnamurthy. Estimating time-dependent entropy production from non-equilibrium trajectories. *Communications Physics*, 5(1):1–10, 2022.

[167] Lee Walsh, Caleb G Wagner, Sarah Schlossberg, Christopher Olson, Aparna Baskaran, and Narayanan Menon. Noise and diffusion of a vibrated self-propelled granular particle. *Soft matter*, 13(47):8964–8968, 2017.

[168] Nitzan Razin, Raphael Voituriez, and Nir S Gov. Signatures of motor susceptibility to forces in the dynamics of a tracer particle in an active gel. *Physical Review E*, 99(2):022419, 2019.

[169] Mark Kac. A stochastic model related to the telegrapher's equation. *The Rocky Mountain Journal of Mathematics*, 4(3):497–509, 1974.

[170] Wolfgang Stadje. The exact probability distribution of a two-dimensional random walk. *Journal of statistical physics*, 46(1):207–216, 1987.

[171] Enzo Orsingher. Probability law, flow function, maximum distribution of wave-governed random motions and their connections with kirchoff's laws. *Stochastic Processes and their Applications*, 34(1):49–66, 1990.

[172] George H Weiss. Some applications of persistent random walks and the telegrapher's equation. *Physica A: Statistical Mechanics and its Applications*, 311(3-4):381–410, 2002.

[173] Peter Hänggi and Peter Jung. Colored noise in dynamical systems. *Advances in chemical physics*, 89:239–326, 2007.

[174] Ryszard Kutner and Jaume Masoliver. The continuous time random walk, still trendy: fifty-year history, state of art and outlook. *The European Physical Journal B*, 90(3):1–13, 2017.

[175] Thibault Bertrand, Yongfeng Zhao, Olivier Bénichou, Julien Tailleur, and Raphaël Voituriez. Optimized diffusion of run-and-tumble particles in crowded environments. *Physical Review Letters*, 120(19):198103, 2018.

[176] Pierre Le Doussal, Satya N Majumdar, and Grégory Schehr. Noncrossing run-and-tumble particles on a line. *Physical Review E*, 100(1):012113, 2019.

[177] Thibaut Demaerel and Christian Maes. Active processes in one dimension. *Physical Review E*, 97(3):032604, 2018.

[178] Jean-François Rupprecht, Olivier Bénichou, and Raphael Voituriez. Optimal search strategies of run-and-tumble walks. *Physical Review E*, 94(1):012117, 2016.

[179] Joel L Schiff. *The Laplace transform: theory and applications*. Springer Science & Business Media, 1999.

[180] Satya N Majumdar, Philippe Mounaix, and Gregory Schehr. Universal record statistics for random walks and Lévy flights with a nonzero staying probability. *Journal of Physics A: Mathematical and Theoretical*, 54(31):315002, 2021.

[181] K Martens, L Angelani, R Di Leonardo, and L Bocquet. Probability distributions for the run-and-tumble bacterial dynamics: An analogy to the Lorentz model. *The European Physical Journal E*, 35(9):1–6, 2012.

[182] Karel Proesmans, Raul Toral, and Christian Van den Broeck. Phase transitions in persistent and run-and-tumble walks. *Physica A: Statistical Mechanics and its Applications*, 552:121934, 2020.

[183] Giacomo Gradenigo and Satya N Majumdar. A first-order dynamical transition in the displacement distribution of a driven run-and-tumble particle. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(5):053206, 2019.

[184] Bart van Gisbergen and Frank Redig. Central limit theorem and large deviations for run and tumble particles on a lattice. *arXiv preprint arXiv:1910.03350*, 2019.

[185] Stephy Jose, Dipanjan Mandal, Mustansir Barma, and Kabir Ramola. Active random walks in one and two dimensions. *arXiv preprint arXiv:2202.02995*, 2022.

[186] Cécile Monthus. Large deviations at various levels for run-and-tumble processes with space-dependent velocities and space-dependent switching rates. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(8):083212, 2021.

[187] David S Dean, Satya N Majumdar, and Hendrik Schawe. Position distribution in a generalized run-and-tumble process. *Physical Review E*, 103(1):012130, 2021.

[188] Martin R Evans and Tom Hanney. Nonequilibrium statistical mechanics of the zero-range process and related models. *Journal of Physics A: Mathematical and General*, 38(19):R195, 2005.

[189] Joachim Krug and Pablo A Ferrari. Phase transitions in driven diffusive systems with random rates. *Journal of Physics A: Mathematical and General*, 29(18):L465, 1996.

[190] Martin R Evans. Bose-Einstein condensation in disordered exclusion models and relation to traffic flow. *EPL (Europhysics Letters)*, 36(1):13, 1996.

[191] OJ O'loan, Martin R Evans, and Michael E Cates. Jamming transition in a homogeneous one-dimensional system: The bus route model. *Physical Review E*, 58(2):1404, 1998.

[192] Satya N Majumdar, Martin R Evans, and Royce KP Zia. Nature of the condensate in mass transport models. *Physical review letters*, 94(18):180601, 2005.

[193] MR Evans, Satya N Majumdar, and Royce KP Zia. Canonical analysis of condensation in factorised steady states. *Journal of Statistical Physics*, 123(2):357–390, 2006.

[194] Martin R Evans, Tom Hanney, and Satya N Majumdar. Interaction-driven real-space condensation. *Physical review letters*, 97(1):010602, 2006.

[195] Martin R Evans and Satya N Majumdar. Condensation and extreme value statistics. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(05):P05004, 2008.

[196] Martin R Evans, Satya N Majumdar, Ignacio Pagonabarraga, and Emmanuel Trizac. Condensation transition in polydisperse hard rods. *The Journal of chemical physics*, 132(1):014102, 2010.

[197] Juraj Szavits-Nossan, Martin R Evans, and Satya N Majumdar. Constraint-driven condensation in large fluctuations of linear statistics. *Physical review letters*, 112(2):020602, 2014.

[198] Juraj Szavits-Nossan, Martin R Evans, and Satya N Majumdar. Condensation transition in joint large deviations of linear statistics. *Journal of Physics A: Mathematical and Theoretical*, 47(45):455004, 2014.

[199] Juraj Szavits-Nossan, Martin R Evans, and Satya N Majumdar. Conditioned random walks and interaction-driven condensation. *Journal of Physics A: Mathematical and Theoretical*, 50(2):024005, 2016.

[200] Giacomo Gradenigo and Eric Bertin. Participation ratio for constraint-driven condensation with superextensive mass. *Entropy*, 19(10):517, 2017.

[201] Jean-Philippe Bouchaud and Marc Mézard. Wealth condensation in a simple model of economy. *Physica A: Statistical Mechanics and its Applications*, 282(3-4):536–545, 2000.

[202] Zdzislaw Burda, D Johnston, Jerzy Jurkiewicz, M Kamiński, Maciej A Nowak, Gabor Papp, and Ismail Zahed. Wealth condensation in pareto macroeconomies. *Physical Review E*, 65(2):026102, 2002.

[203] Giacomo Gradenigo, Stefano Iubini, Roberto Livi, and Satya N Majumdar. Condensation transition and ensemble inequivalence in the discrete nonlinear Schrödinger equation. *The European Physical Journal E*, 44(3):1–6, 2021.

[204] Mario Filiasi, Elia Zarinelli, Erik Vesselli, and Matteo Marsili. Condensation phenomena in fat-tailed distributions: a characterization by means of an order parameter. *arXiv preprint arXiv:1309.7795*, 2013.

[205] Mario Filiasi, Giacomo Livan, Matteo Marsili, Maria Peressi, Erik Vesselli, and Elia Zarinelli. On the concentration of large deviations for fat tailed distributions, with application to financial data. *Journal of Statistical Mechanics: Theory and Experiment*, 2014(9):P09030, 2014.

[206] Aleksandr V Nagaev. Integral limit theorems taking large deviations into account when Cramér's condition does not hold. i. *Theory of Probability & Its Applications*, 14(1):51–64, 1969.

[207] Aleksandr V Nagaev. Integral limit theorems taking large deviations into account when Cramér's condition does not hold. ii. *Theory of Probability & Its Applications*, 14(2):193–208, 1969.

[208] Fabien Brosset, Thierry Klein, Agnès Lagnoux, and Pierre Petit. Probabilistic proofs of large deviation results for sums of semiexponential random variables and explicit rate function at the transition. *arXiv preprint arXiv:2007.08164*, 2020.

[209] Marc Höll and Eli Barkai. Big jump principle for heavy-tailed random walks with correlated increments. *The European Physical Journal B*, 94(10):1–17, 2021.

[210] Celine Nadal, Satya N Majumdar, and Massimo Vergassola. Phase transitions in the distribution of bipartite entanglement of a random pure state. *Physical review letters*, 104(11):110501, 2010.

[211] Celine Nadal, Satya N Majumdar, and Massimo Vergassola. Statistical distribution of quantum entanglement for a random bipartite state. *Journal of Statistical Physics*, 142(2):403–438, 2011.

[212] Shamik Gupta, Satya N Majumdar, and Grégory Schehr. Fluctuating interfaces subject to stochastic resetting. *Physical review letters*, 112(22):220601, 2014.

[213] Shamik Gupta and Apoorva Nagar. Resetting of fluctuating interfaces at power-law times. *Journal of Physics A: Mathematical and Theoretical*, 49(44):445001, 2016.

[214] Martin R Evans, Satya N Majumdar, and Grégory Schehr. An exactly solvable predator prey model with resetting. *arXiv preprint arXiv:2202.06138*, 2022.

[215] Georgia R Calvert and Martin R Evans. Searching for clusters of targets under stochastic resetting. *The European Physical Journal B*, 94(11):1–9, 2021.

[216] Lev Semenovich Pontryagin. *Mathematical theory of optimal processes*. CRC press, 1987.

[217] Stuart Russell and Peter Norvig. Artificial intelligence: a modern approach. 2002.

[218] Huyên Pham. *Continuous-time stochastic control and optimization with financial applications*, volume 61. Springer Science & Business Media, 2009.

[219] Joseph Samuel and Supurna Sinha. Optimal control in pandemics. *Physical Review E*, 103(1):L010301, 2021.

[220] Alexandre Dolgui, Dmitry Ivanov, Suresh P Sethi, and Boris Sokolov. Scheduling in production, supply chain and Industry 4.0 systems by optimal control: fundamentals, state-of-the-art and applications. *International Journal of Production Research*, 57(2):411–432, 2019.

[221] Konstantin Khanin, Sergei Nechaev, Gleb Oshanin, Andrei Sobolevski, and Oleg Vasilyev. Ballistic deposition patterns beneath a growing Kardar-Parisi-Zhang interface. *Physical Review E*, 82(6):061107, 2010.

[222] Michael F Shlesinger. Search research. *Nature*, 443(7109):281–282, 2006.

[223] O Bénichou, C Loverdo, M Moreau, and R Voituriez. Two-dimensional intermittent search processes: An alternative to Lévy flight strategies. *Physical Review E*, 74(2):020102, 2006.

[224] Robert F Stengel. *Optimal control and estimation*. Courier Corporation, 1994.

[225] John Crank. *Free and moving boundary problems*. Oxford University Press, USA, 1984.

[226] Wendell H Fleming and Halil Mete Soner. *Controlled Markov processes and viscosity solutions*, volume 25. Springer Science & Business Media, 2006.

[227] B De Bruyne, J Randon-Furling, and S Redner. Optimization in first-passage resetting. *Physical Review Letters*, 125(5):050602, 2020.

[228] B De Bruyne, J Randon-Furling, and S Redner. Optimization and growth in first-passage resetting. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(1):013203, 2021.

[229] Benjamin De Bruyne, Satya N Majumdar, and Grégory Schehr. Survival probability of a run-and-tumble particle in the presence of a drift. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(4):043211, 2021.

université PARIS-SACLAY | ÉCOLE DOCTORALE Physique en Île-de-France (PIF)

Titre : Statistique d'extrêmes des processus : du mouvement brownien aux particules actives
Mots clés : Statistiques d'extrêmes, mouvement brownien, physique hors équilibre, particules actives

Résumé : Bien que rares, les événements extrêmes peuvent jouer un rôle majeur dans un large éventail de situations, de la finance au climat. Dans cette thèse, nous étudions les propriétés extrêmes de plusieurs processus stochastiques, dont le mouvement brownien (MB), les particules actives et le MB avec réinitialisation.

Dans la première partie, nous étudions les instants auxquels les extrema des processus stochastiques unidimensionnels se produisent. En particulier, dans le cas d'un MB de durée fixée, nous calculons la distribution de probabilité du temps entre le maximum global et le minimum global. De plus, nous dérivons la distribution du temps du maximum pour une classe de processus stochastiques stationnaires, à la fois à l'équilibre et hors d'équilibre. Cette analyse conduit à la formulation d'un critère simple pour détecter des fluctuations hors d'équilibre dans les états stationnaires.

Dans la deuxième partie, nous nous concentrons sur le modèle de particules dit « run-and-tumble particle » (RTP). Nous calculons exactement la probabilité de survie pour une RTP dans un espace à d dimensions, montrant que cette quantité est complètement universelle, c'est-à-dire indépendante de d et des fluctuations de vitesse de la particule. Nous étendons cette universalité à d'autres observables et à certains modèles RTP généralisés. De plus, nous étudions également les grandes déviations de la position d'une RTP. Nous montrons que, sous certaines conditions, une transition de condensation est observée dans le régime de grandes déviations où la particule est éloignée de sa position de départ.

Enfin, nous introduisons une nouvelle technique, analogue à l'équation de Hamilton-Jacobi-Bellman, pour contrôler de manière optimale un système dynamique à travers des redémarrages.

Title : Extreme value statistics of stochastic processes : from Brownian motion to active particles
Keywords : Extreme value statistics, Brownian motion, nonequilibrium physics, active particles

Abstract : Rare extreme events tend to play a major role in a wide range of contexts, from finance to climate. Hence, understanding their statistical properties is a relevant task, which opens the way to many applications. In this thesis, we investigate the extremal properties of several stochastic processes, including Brownian motion (BM), active particles, and BM with resetting.

In the first part, we investigate the times at which extrema of one-dimensional stochastic processes occur. In particular, in the case of a BM of fixed duration, we compute the probability distribution of the time between the global maximum and the global minimum. Moreover, we derive the distribution of the time of the maximum for stationary stochastic processes, both at equilibrium and out-of-equilibrium. This analysis leads to the formulation of a simple criterion to detect nonequilibrium fluctuations in steady states.

In the second part, we focus on the run-and-tumble particle (RTP) model. We compute exactly the survival probability for a single RTP in d dimensions, showing that this quantity is completely universal, i.e., independent of d and the speed fluctuations of the particle. We extend this universality to other observables (time of the maximum and records) and generalized RTP models. Moreover, we also investigate the position distribution of a single RTP at late times. We show that, under certain conditions, a condensation transition can be observed in the large-deviation regime where the particle is far from its starting position.

Finally, we introduce a new technique, analog to the Hamilton-Jacobi-Bellman equation, to optimally control a dynamical system through stochastic resetting.