POLITECNICO
DI TORINO

Dipartimento
di Scienze Matematiche
G.L. Lagrange

DIPARTIMENTO
DI MATEMATICA
GIUSEPPE PEANO
UNIVERSITÀ DI TORINO

Politecnico di Torino
Università degli Studi di Torino

Dottorato in Matematica Pura e Applicata – XXXI ciclo

Tesi di Dottorato

Large deviations for rare realizations of dynamical systems

From extreme water waves to nonequilibrium transients
of "small systems": A general approach

Giovanni Dematteis

Supervisor
Prof. Lamberto Rondoni
Dipartimento di Scienze Matematiche
Politecnico di Torino

Co-supervisor
Prof. Eric Vanden-Eijnden
Courant Institute of Mathematical Sciences
New York University

Anno Accademico 2018-2019

A mamma e papà, che con il loro
esempio mi hanno insegnato la passione
al dare il meglio di me

Abstract

A central problem in uncertainty quantification is how to characterize the impact that our incomplete knowledge about models has on the predictions we make from them. This question naturally lends itself to a probabilistic formulation, by making the unknown model parameters random with given statistics. In the following this approach is used in concert with tools from *large deviation theory* (LDT) and optimal control to estimate the probability that some observables in a dynamical system go above a large threshold after some time, given the prior statistical information about the system's parameters and/or its initial conditions. It is established under which conditions the extreme events occur in a predictable way, as the minimizer of the LDT action functional, i.e. the *instanton*.

In the first physical application, the appearance of rogue waves in a long-crested deep sea is investigated. First, the leading order equations are derived for the wave statistics in the framework of wave turbulence (WT), showing that the theory cannot go beyond Gaussianity, although it remains the main tool to understand the energetic transfers. It is shown how by applying our LDT method one can use the incomplete information contained in the spectrum (with the Gaussian statistics of WT) as prior and supplement this information with the governing nonlinear dynamics to reliably estimate the probability distribution of the sea surface elevation far in the tail at later times. Our results indicate that rogue waves occur when the system hits unlikely pockets of wave configurations that trigger large disturbances of the surface height. The rogue wave precursors in these pockets are wave patterns of regular height but with a very specific shape that is identified explicitly, thereby potentially allowing for early detection. Finally, the first experimental evidence of hydrodynamic instantons is presented using data collected in a long wave flume, elevating the instanton description to the role of a unifying theory of extreme water waves.

Other applications of the method are illustrated: To the nonlinear Schrödinger equation with random initial conditions, relevant to fiber optics and integrable turbulence, and to a rod with random elasticity pulled by a time-dependent force. The latter represents an interesting nonequilibrium statistical mechanics setup with a strongly out-of-equilibrium transient (absence of local thermodynamic equilibrium) and a small number of degrees of freedom (small system), showing how the LDT method can be exploited to solve optimal-protocol problems.

Acknowledgements

I am grateful to Lamberto Rondoni, a supportive mentor in science and in life. I will never stop learning from his unconventional and deep way to look at problems, especially in the realm of thermodynamics. Moreover, it is indeed thanks to his selfless foresight that I was able to live out the great variety of academic and human experiences that made my PhD.

I am thankful to Eric Vanden-Eijnden. Without knowing much about myself, he gave me the opportunity to join his group at Courant, patiently supervised and directed my work for over two years teaching me a huge amount of things about large deviations and applied math, always challenging me with the highest work standards. Among many other things, I am in debt for the hard work that Eric put in shaping my writing under the motto of "be precise and concise"—at times a real pain for both of us, but which did help me a lot in the end.

A huge thanks goes to Tobias Grafke, great collaborator and friend. Even though not officially appearing as one of my supervisors, he is in the facts one of them. Apart from learning from him—and his proverbial patience and method—most of the things that I know about numerical optimization, I have profited immensely from all our conversations and discussions.

Among other people and institutions that I want to acknowledge:

Miguel Onorato, whose unique all-round approach to physics has helped us reach the important experimental results presented in this dissertation, not to mention that under his inspiring mentoring I am currently engaged in my first postdoc experience;

Hugo Touchette and Angelo Vulpiani, who kindly accepted to referee the thesis and sent detailed and useful comments that helped improve the presentation;

Compagnia di San Paolo, for providing the funds for my PhD scholarship;

Dipartimento di Scienze Matematiche (DISMA) of Politecnico di Torino, my home institution in the last three years;

The human resources staff of the Courant Institute of Mathematical Sciences (NYU), for all the help received with the Visa paperwork during my entire stay in 2016 and 2017.

Finally, I acknowledge the support of my whole great family, including Nonna Pia and Nonno Piero (we miss enormously celebrating with him on any good occasion), and of my closest friends. I also want to thank all the friends in NYC, who have been my family when away from home—in particular Tully, Roshard and Steve, whose sincere friendship and continuous support have meant to my life much more than I am able to express by words.

Contents

List of Figures

...ho promesso che avrei fatto le cose per bene, perché tutto quello che faccio, lo faccio a nome Suo [Gino Bartali]

Cross-fertilization between fields is a powerful engine of progress in science [Jim Yorke]

Chapter 1

Introduction

1.1 Extreme events in everyday life

In many aspects of the natural phenomena and of human life, events that are rare but extreme play a significant role. Often, these are really what matters the most. In order to realize this, it is sufficient to think of devastating situations we are familiar with, such as a boat hit by a rogue wave [123, 144], the large number of deaths caused by a heat wave [164], a stock market collapse [176], or the damage inflicted by failure events in combustion systems [166]. Many more examples can be done: The overheating of a component in an electric circuit, the adaptive ability of a bacterial colony based on very unlikely mutations that turn out crucially favorable, etc.

In an age where climate change is inexorably and unpredictably impacting our lives, we are becoming more and more accustomed to an Alpine climate with winters without snow (like the current season at the beginning of 2019 in Northwest of Italy) and with very hot summers. Over the past fourty years, some combination of these factors has literally melted down most of those majestic glaciers that had inspired the fantasies of Goethe and Dumas. To be convinced of this fact, it is enough to look at the pictures that people—among which the author's grandfather—used to take of glaciers in the '80s in places where nowadays one can see nothing but rocks and grass. Those glaciers had been there for millennia, since the last ice age! How could that happen so fast, considering that the average global temperature has increased slightly more than half of a Celsius degree since 1970? The answer is that climate is nonlinear and complex, a tiny average change can trigger domino effects whose extent and magnitude are hard to predict, and "the consequence of global warming is not the change in the averages but the overall increase of extreme events" [132]. Examples of climatic extreme events include heat waves [168, 11], droughts [120, 132], hurricanes and storms [179, 42].

Regardless of the field of application being fluid dynamics, climate, technology, finance or biology, all of the events mentioned above have a fundamental feature in

common: each of them is catastrophic. Despite the event being rare, an irreversible damage is inflicted on the system at hand. Usually, as intuition would suggest, an increasing *extremeness* goes along with an increasing *rareness*. At least on a general level, this principle is an implication of the *central limit theorem* (CLT), regulating the probability of deviations of the empirical mean from its expected value. Though, the mentioned phenomena make it apparent that their frequency is surely not sufficiently low to justify neglecting them. On the contrary, their abrupt and brutal character and the tragic impact motivate a systematic treatment, in order to understand better their likelihood and the underlying mechanisms. Indeed, concerning the *risk assessment*, the ultimate goal is to make such catastrophic occurrences predictable and, as a direct implication, to be able to minimize their irreversible consequences. A lot of effort has been put in this direction in recent years [170, 133, 166, 161, 194, 61], in the field of uncertainty quantification.

Indeed, we should say that not always the extreme events have a negative connotation. Looking at the problem from a different perspective, for instance one might be interested in achieving a difficult task by acting on a system. In this setting, being able to predict—and therefore reproduce—extreme realizations may represent a problem of great relevance as well, closely related to the aims of *optimal control* [189, 8].

1.2 Are the extreme events predictable?

Predicting the pathway to the extreme events in general is a hard problem, especially if the challenge is tackled naively. The straightforward approach consists of sampling a great amount of realizations out of a given statistical prior—the precise setup we have in mind will be defined in Chapter 2—, selecting the events that fulfill some extremeness requirement, and finally processing the signal obtained via a suitable averaging in order to extract useful information. Such blind procedure is complicated and can be highly inefficient.

Fortunately, another quite intuitive observation comes to help here: when an extreme rare realization of a random variable occurs, there are much fewer ways by which it can occur compared to more likely, typical realizations. As an example, let us think of the roll of two dice. If we wish to obtain 7 for the sum of the two dice, we have 6 possibilities out of 36, while to obtain 12 we have only 1. This simple example suggests the following idea: in order for a sample mean of a random variable (here the mean of two independent realizations, i.e. the two dice) to reach a very unlikely value, typically very large, all of the degrees of freedom must "conspire" together in a specific way. More simply, all of the extreme events look alike and this yields crucial implications for their predictability. Mathematically, this idea is effectively formulated in terms of a *saddle-point approximation* (*Laplace's method*), in a region where the least unlikely scenario is so much more likely than all of the others, that it dominates the statistics. The approximation works well if the number N of samples

over which the empirical mean is computed becomes very large. Then, all of the occurrencies of an extreme mean event are well approximated by the most likely realization of that mean event. This concept has far-reaching consequences and, not surprisingly, it has been widely exploited by physicists and rigorously formalized by mathematicians during the last century. In particular, we are referring to the concepts of *instanton* and of *large deviations*, which we are going to review in the following.

One of the main reasons why these *large deviations* are so important to have an entire theory named after them, is the fact that the CLT is restricted by its assumptions to the description of fluctuations comparable to the standard deviation of the random variable of interest—at least, in general, when the probability distribution of the latter is not itself a Gaussian. Therefore, a different theory is needed for fluctuations significantly larger than the standard deviation. On the other hand, as we have already suggested, the saddle-point approximations (mentioned above) represent powerful tools by which the computations can be done in practice.

Hereafter, we shall restrict ourselves to physics for the sake of the argument, and consider generic systems with a certain given temporal evolution law, which mathematically we think of as *dynamical systems*. It is known that the result of a measurement on a physical system can be reduced to an *ensemble average*, i.e. a sample mean over independent realizations, at equilibrium and in the thermodynamic limit. This is the well known *ergodic hypothesis* dating back to Boltzmann, and thermodynamic equilibrium is the setup where it is well justified [108, 26, 84, 35]. In fact, there is an entire domain of mathematical physics devoted to such problem, named *ergodic theory*. Notice that even the novel theories of fluctuations based on Markov processes such as *stochastic thermodynamics* [172, 68] are derived at least under the assumption of local thermodynamic equilibrium for the many degrees of freedom of the heat bath. Out of the context of thermodynamic equilibrium, instead, it can be misleading to assign physical meaning to ensemble averages [119]. Let us consider a single toss of a coin, where we measure $s = +1$ if the outcome is "head" and $s = -1$ if the outcome is "tail"—the two states of the system. Knowing that the expected value of the measurement is 0 does not help much, as it is a very different value from the possible outcomes of the event s: we have the right to say that here the sample mean is not a physical observable depending on the state of the system at a given moment. We should not forget that the sample mean of N independent tosses, converging to 0 as $N \to \infty$ by the law of large numbers, in this case represents the mean of N independent experiments. It is fair to study such quantity, but keeping in mind that it is not describing the single-system behavior, but a so-called *ensemble* behavior. On the other hand, when having a system of N coins and studying the behavior of the observable $N^{-1}\sum_{i=1}^{N} s_i$, the sample mean now is not just an ensemble property but a good physical observable well defined on the system. We hope that these simple examples helped illustrating the limits of a sample-mean approach that one should be aware of. For a clear presentation of

these subtle (and often overlooked) concepts, we refer the reader to Lebowitz [119].

One of the ways to go beyond sample means without giving up on the predictable character of the rare events, is to rescale the random variable by a small parameter ϵ, for instance interpreting it as a small random noise. If this is done, it is no longer necessary to consider N independent realizations of the system, but *Laplace's method* will work for an observable defined on a single realization, ϵ^{-1} playing the role of large parameter for the *saddle point* in place of N. Thus, we are now restricting to situations where such a small-noise limit is meaningful, and with the advantage that all of the single extreme realizations look like the most likely one. And as we have pointed out, in general there *is* a conceptual difference between results on the single system's behavior (*à la Boltzmann*) and the mean behavior of an ensemble of identical systems independent of each other (*à la Gibbs*) [119, 83].

A further step can be taken at this point, and this is precisely what the present dissertation aims at. Such step consists of considering the large size itself of the physical observable as the large parameter justifying the saddle-point approximation. Conceptually, this can often be thought of as the rescaling of a small-noise problem, where the noise is small and the observable of typical size, to an equivalent problem where the randomness parameters do not need to be small but one is interested in very large values of the observable. Sometimes, as we will show explicitly via a suitable rescaling, there is a true equivalence between the two approaches. In other situations, though, the relationship between the random parameters (or noise) and the observable is so complicated that it is not possible (at least, in practice) to view the "extreme size" problem as a "small noise" problem. It is for such cases that a method with the focus on the "extremeness" of the events becomes really important, as we will demonstrate in the following. Therefore, the saddle-point approximation will work only for the extreme events, the size z playing itself the role of large parameter. Moreover, also in this setup the results will hold for single extreme realizations which are not just ensemble quantities. On the other hand, being the large parameter the event size itself and not an external parameter, one will have to pay the price that the obtained asymptotic result does not apply in the range of the typical events.

Intentionally, so far we have kept the discussion on the level of mere ideas. These will be made precise in the following. In particular, the approach to the extreme events mentioned in the previous paragraph is the core of the dissertation. In order to go deeper into the problem, in the next two sections we will review how all of the ideas above have been developed in the context of physics, around the concept of *instanton*, and in the context of mathematics, originating that branch of probability known as *large deviation theory*.

1.2.1 Instanton theory

Rare and extreme realizations of stochastic systems are known to be described by *instantons* [165, 59, 171], saddle-point configurations of the action of the associated stochastic field theory. Developed originally in the context of quantum chromodynamics [171] starting from the late '70s, later on instantons have been used successfully as theoretical tool to predict limiting behavior of extreme events in different fields, e.g. for chemical reactions [59] or for turbulence in fluids [76, 90, 91]. At the core of the theory lies the realization that the evolution of any stochastic system, be it quantum or classical, reduces to a well-defined (semi-classical) limit in the presence of a smallness parameter. Concretely, the simultaneous evaluation of all possible realizations of the system subject to a given constraint results in a (classical or path-) integral, where the integrand contains an action functional $S(\psi)$. The dominating realization can then be obtained to leading order by approximating the integral by a *saddle-point approximation*, $\delta S(\psi)/\delta\psi = 0$. The critical point ψ^* of the action functional associated with a stochastic system is called the *instanton*, and it represents the maximum likelihood realization of the stochastic system subject to the given constraints.

One of the simplest prototypes of instanton computation is the quantum double well potential, with two stable fixed points separated by a potential barrier. In a classical deterministic system, a particle with total mechanical energy lower than the potential energy barrier would have no way to cross the barrier and would remain trapped indefinitely in one of the wells. On the other hand, a quantum particle with energy smaller than the energy barrier does have a chance to cross the barrier, via the well know quantum tunnelling effect. One simple way to see why this happens is, in the WKB approximation, to compute the transition probability from one well minimum ($x = a$) to the other ($x = b$). Using Feynman's path integrals, the expression of the transition probability from a to b writes

$$\begin{aligned} \mathbb{P}(x_f = b|x_0 = a) &= \int d[x(\tau)] e^{-\frac{S[x(\tau)]}{\hbar}} , \\ S[x(\tau)] &= \int_{\tau_a}^{\tau_b} \left(\frac{1}{2} \left(\frac{dx}{d\tau} \right)^2 + V(x) \right) d\tau , \end{aligned} \tag{1.1}$$

where V is the potential and $\tau = it$ is just a Wick rotation of the time variable t. If we consider a potential $V(x) = \frac{1}{4}(x^2 - 1)^2$, the action minimization via $\delta S[x(\tau)]/\delta x = 0$ yields

$$x_{inst}(\tau) = \tanh\left(\frac{1}{\sqrt{2}}(\tau - \tau_0) \right) . \tag{1.2}$$

Around the arbitrary jumping time τ_0, such solution transitions quickly from one classical minimum $a = -1$ to the other $b = +1$, tunnelling the barrier by "climbing" uphill along the time-reversed classical downhill solution. The rapid transient

character of the solution is self-explanatory for the origin of the name *instanton*. Moreover, in (1.1) we see that as $\hbar \to 0$ (essence of the WKB approximation) the saddle-point approximation

$$\mathbb{P}(x_f = b | x_0 = a) \propto e^{-\frac{S[x_{inst}(\tau)]}{\hbar}} \tag{1.3}$$

becomes more and more accurate: The instanton solution dominates the transition probability. Instanton calculus basically developed around this idea, focusing on the minimum-action trajectory and the bundle of trajectories that fluctuate close to the instanton trajectory.

For stochastic differential equations (SDEs), the concept is analogous. For instance, the double-well problem above can be formulated as a stochastic problem concerning a Brownian particle in a heat bath that determines the strength of the noise. If the noise is small compared to the size of the potential barrier, the transition probability from a to b can be cast exactly in the form of Eq. (1.1), where $\hbar$ is replaced by the noise strength $\epsilon \propto T$ (proportional to the temperature T). For SDEs, these ideas have found a rigorous justification in the Freidlin-Wentzell theory of large deviations [80] (see next section).

1.2.2 Large deviation theory

The *theory of large deviations* concerns the exponential-order decay of the probability tails of random variables, going beyond the results of the *central-limit theorem* (CLT) that is restricted to fluctuations of the order of the standard deviation. Historically, the first results are found in the '30s by Scandinavians studying risk analysis in the insurance market [67], and then by Cramér [47] who formalized his theorem for sums of independent identically distributed (*i.i.d.*) random variables. Sanov's theorem [169], instead, was derived for distributions of *i.i.d.* random variables. It is in the late '60s, then, that a rigorous formalization of large deviation theory (LDT) started, from the work of Varadhan [192], Gärtner [86] and Ellis [66], generalizing the results to Markov chains. The results on small random perturbations of dynamical systems are mainly due to Freidlin and Wentzell [80]. We mention some references covering most of the aspects of LDT [52, 50, 175, 65, 53, 79]. Nowadays, LDT has countless applications to the most diverse fields, and it is considered the underlying language of statistical mechanics [187]—notice that many of the ideas in large deviations were anticipated by Boltzmann about 140 years ago! [19] (see [64]).

Central to LDT is the *large deviation principle* (LDP). For the empirical mean of *i.i.d.* random variables $X_N = \frac{1}{N}\sum_{j=1}^{N} Y_j$, we wish to study the probability $P(a) \equiv \mathbb{P}(X_N > a)$. Such probability might be hard to obtain in general. If an exponential approximation holds of the form

$$P(a) \asymp e^{-NI(a)}, \tag{1.4}$$

—where the symbol "$\asymp$" indicates logaritmic equivalence in the limit, more precisely: $\lim_{N\to\infty} \frac{1}{N} \log P(a) = \lim_{N\to\infty} -I(a)$—we say that we are in the presence of an LDP. Different forms of LDP can be found, like

$$P(a) \asymp e^{-\frac{1}{\epsilon} I(a)} \tag{1.5}$$

for $P(a) \equiv \mathbb{P}(X > a)$, where $I(a)$ can be a function or a functional (called quasi-potential) as in the Freidlin-Wentzell theory for Brownian motion [80], and ϵ is the small strength of the noise—and $\epsilon \to 0$ the limit of the asymptotic equivalence. Yet, we may have

$$P(a) \asymp e^{-TI(a)}, \tag{1.6}$$

with $P(a) \equiv \mathbb{P}(X_T > a)$, X_T being a time-averaged dynamical observable of a Markov process in the $T \to \infty$ limit [188].

In each of the above cases, the LDP is not useful unless we have a way to access the *rate function* $I(a)$. The latter can be found straightforwardly by using asymptotic expansions like Stirling approximation, but such path is not always possible to follow. Alternatively, there is a major result in LDT to this end called *Gärtner-Ellis theorem* [86, 66], which expresses the rate function $I(a)$ as the *Legendre-Fenchel transform* of the *scaled cumulant generating function* $S(\lambda)$,

$$I(a) = \sup_{\lambda}(\lambda a - S(\lambda)), \qquad S(\lambda) \equiv \lim_{N\to\infty} N^{-1} \log \mathbb{E}\left(e^{N\lambda Y_j}\right) \tag{1.7}$$

referring to the LDP (1.4), where $S(\lambda)$ can be simply computed from the statistics of the Y_j's. The result (1.7) of Gärtner-Ellis is of great relevance both to establish when an LDP holds and to practically compute the rate function. It is easy to see how substantially it comes from concentration of an integral in the point of maximum of the integrand, i.e. a saddle-point approximation, from which the maximization of the Legendre-Fenchel transform comes about—see section III.C of [187]. We see therefore how LDT is the framework capturing the common intuitions about the extreme events that were discussed in section 1.1, actually also representing the underlying mathematics of instantons. The concept of an optimal point dominating the statistics, the *dominating point* [143] or *instanton* hereafter, is really intimately connected with the LDP. We will expand on this concept in chapter 2.

At this point, we can draw some general considerations that will be useful in the following to relate our results to the context of the existing theory:

- For the LDP, in any of its forms (1.4), (1.5) or (1.6), we need a large parameter with a specific meaning. In the applications, this parameter for instance can be the number of components of the system as in thermodynamics [187], the reciprocal of the noise strength for random perturbations of dynamical systems [92], or the time over which a time-averaged observable is computed in Markov processes [188]. The large parameter limit is the asymptotic regime in which the LDP equivalence is intended.

- There are two independent scales in LDT: one is the scale a_N by which we need to divide the observable to make it intensive, i.e. $X_N = {a_N}^{-1} \sum_{j=1}^{N} Y_j$; the other is the so-called *speed of the LDP* b_N, $P(a) \asymp e^{-b_N I(a)}$. The two scales are independent a-priori. The fact that in the example above we have $a_N = b_N = N$ is a coincidence for *i.i.d.* random variables, but it is not necessary in general.

- The meaning of the LDP (1.4) is that the dominant behavior of the tail probability $P(a)$ is a decaying exponential in N (or b_N to be more general), with $I(a)$ (at fixed a) a positive constant. Thus, with a simple passage we can invert the relationship and extract such constant: $I(a) = b_N^{-1} \log P(a) + o(b_N^{-1})$. Taking the limit, the result is a positive constant if the tail is actually an exponential in b_N, it is vanishing if the tail is sub-exponential and it is infinite if it is super-exponential. Only the first case yields a meaningful LDP.

- The Legendre-Fenchel transform is ubiquitous in LDT, and it provides the main framework for the optimization problems from which the rate function is computed.

- The rate function enjoys a number of important properties, some of which we limit ourselves to mention—they can be found in any good review, e.g. [187]: $I(a)$ is generally convex, it is positive and it is zero at its minimum. The point of minimum is where the probability concentrates as the limit is taken, and this is the way the CLT is included within the more general LDT.

- LDT extends the estimate of the CLT to extreme fluctuations well beyond the order of the standard deviation, where the CLT is supposed to hold—of course, the extension of the estimate is in the weaker sense of logaritmic equivalence proper of LDT.

1.2.3 A guiding example: multiscale modelling of the ocean surface

In order to make this introductive discussion more concrete, we bring the example of the ocean waves. This will motivate the need for a large deviation method that can be applied to fluctuations of dynamical systems where none of the scaling regimes in the previous section are meaningful. On the scale of the naval vessels, the fluctuations of the ocean surface are the waves that we all are familiar with. It is important to quantify how they can reach an extreme height, representing a serious danger for navigation. Such large fluctuations of the water surface are single events localized in space, so that an $N \to \infty$ limit is not meaningful. Moreover, they appear as fast transients, and a $T \to \infty$ scenario cannot capture such occurrences. Lastly, by considering a small energy of the surrounding wave field, in a

similar fashion to the Freidlin-Wentzell $\epsilon \to 0$, we would "kill" all of the nonlinear hydrodynamic effects that may play a significant role in the energy localization phenomena. Therefore, if we wish to use LDT, there is no large parameter here but the fluctuation size itself.

In the remainder of the section, we review the state of the art of ocean-wave modelling, by tracing an interesting parallel with nonequilibrium thermodynamics. This will help understanding the overlap of different levels of description at the different scales, from the thousands of kilometers of the atmospheric forcing down to the few meters of width of an oil rig.

The macroscopic scale

We start from the macroscopic scale, where the description is analogous to the continuum description of fluids where some local macroscopic fields (mass and energy densities, momentum, temperature) flow and evolve with given deterministic laws (PDE's such as the Navier Stokes equations). For the ocean surface, such description involves the evolution of the *wave action* (playing a similar role to the mass density), *momentum* and *energy* fields, whose integrals are the conserved quantities of the system[1]. The mass is proportional to H_s^2, where H_s is the significant wave height quantifying the *roughness* of the sea. A dimensional estimate of the energy is given by $\omega_0 H_s^2$. Fig. 1.1 is the output of the ECMWF global model, representing the macroscopic state of the Atlantic Ocean at a given moment. The flag at a specific point indicates $H_s = 6.8$ m and a spectral peak with period $T_0 = 10$ s, i.e. $\omega_0 = 2\pi/T_0 \simeq 0.63$ s^{-1} at that point, that is, an equivalent representation of the macroscopic fields. The color map, instead, represents the spatial distribution of $H_s = H_s(x, y)$, appearing as a smooth field. The forcing by the winds is where the ocean surface takes most of its energy from, on the large scales of hundreds of kilometers of the atmospheric perturbations. The counterpart of the hydrodynamic equations (PDE's) for the space-time evolution of mass and energy are not known. Nonetheless, the evolution of these quantities is implemented in the global models of weather forecast such as ECMWF or Wavewatch, where the ocean surface is coupled with the atmosphere dynamics and with the oceanic currents. This is done by using the more detailed description of wave turbulence (WT) involving a mesoscopic transport equation analogous to the Boltzmann equation.

The mesoscopic scale

Behind the macroscopic representation in Fig. 1.1 there is an assumption of local thermodynamic equilibrium (LTE) for the mesoscale level, in analogy to the LTE assumption that allows to define a continuous density to describe a gas that

[1]For simplicity, here we restrict the discussion to action and energy

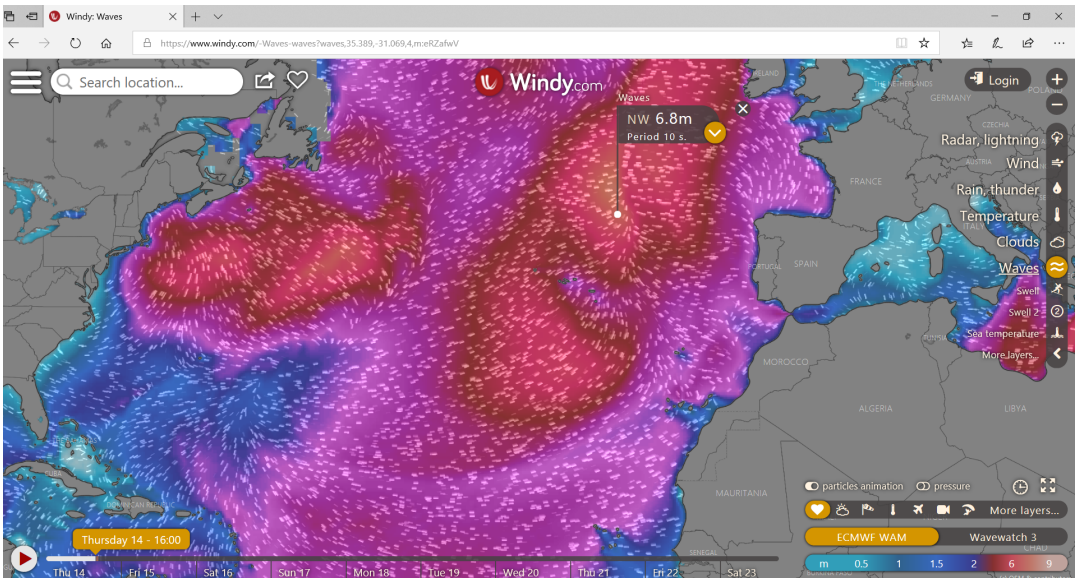

Figure 1.1. Global-model prediction of the state of the ocean, from the website *windy.com*.

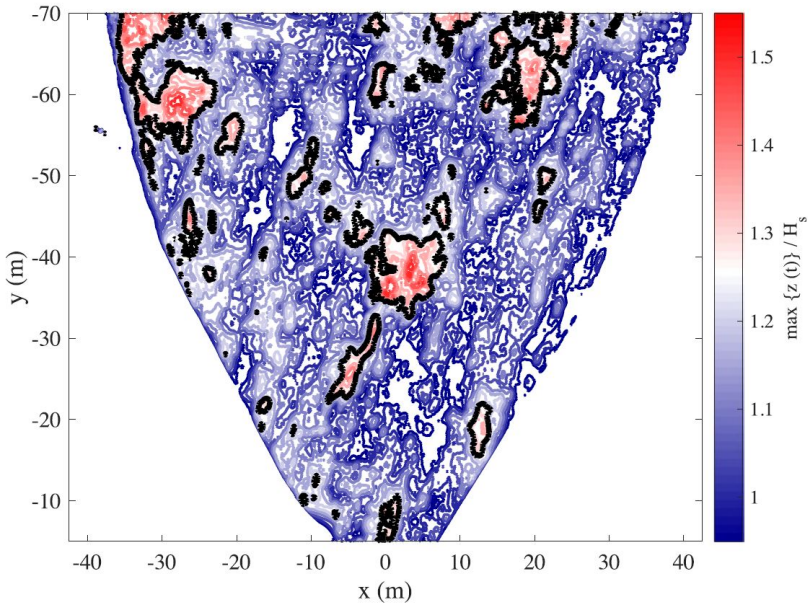

Figure 1.2. Pattern of surface elevation at the platform *Aqua Alta* in the Adriatic Sea.

is microscopically discontinuous. Without such assumption, it would not be possible to define a macroscopic deterministic density at all. The resolution of the ECMWF WAM model is 13 km, meaning that in Fig. 1.1 anything happening at smaller scales is coarse grained and averaged out resulting in a collective information representing the 13 km box. At the mesoscale level, within each box there is a stochastic "bath" of waves that interact nonlinearly exchanging energy on the nonlinear timescales. Such interactions are captured by the collision integral of the wave kinetic equation (WKE), in analogy to how the Boltzmann equation describes the molecular interactions in statistical mechanics. The quantity evolving with the WKE (playing the part of the single-particle density) is the spectrum $n(k)$, which is the wave-action density per frequency mode. The spectral energy density $e(k)$ relates to the wave-action density as $e(k) = w(k)n(k)$. If we suppose that the atmospheric forcing is constant on a box, the WKE reaches a nonequilibrium steady state called the Kolmogorov-Zakharov (KZ) spectrum—an $H-$theorem predicts the convergence.[2] The KZ spectrum is a (turbulent) nonequilibrium state since there is a net energy flow from the large scales of the forcing towards the small scales of dissipation, where the latter is mostly due to wave breaking. In absence of forcing, a true equilibrium state is attained, the Rayleigh-Jeans state—and in such case the $H-$theorem is analogous to the Boltzmann $H-$theorem predicting relaxation to the Maxwell-Boltzmann distribution. Thus, the spectrum is an averaged quantity over the whole box, describing the distribution of energy among the modes, analogously to how the Maxwell-Boltzmann distribution describes the velocity distribution of the molecular constituents at equilibrium in a large box. Clearly, the mesoscopic box must be very large (so as the observation time) in order for the law of large numbers to make the empirical spectrum converge to a deterministic limit. The WKE is the main object of the WT theory, and the resonant interactions within the collision integral (4-wave resonances for surface gravity waves) determine the spectral transfers which are included in the computations of the global models. For a good continuum description as in Fig. 1.1, it is fundamental that the macro and meso scales be separated: The mesoscopic box is large enough that the WT formalism holds thanks to the law of large numbers, but also small enough that the atmospheric forcing and the statistics of the wave field can be considered homogeneous over the box. To give an idea of such scale separation, under each anticyclone forcing the Atlantic in Fig. 1.1, of diameter on the order of a thousand of kilometers, there are hundreds of mesoscale boxes of side 13 km!

Now, we said that the mesoscopic box is characterized by random waves. But where does the randomness come from, if the underlying "microscopic" equations (describing the wave surface) are deterministic? There are at least two main sources

[2]The flux (turbulence cascade) is among the Fourier modes. Instead, the LTE assumption that was mentioned is spatial, assuming that the statistics is homogeneous inside the box and that the net fluxes at the boundaries of the box are small compared to the fluctuations inside.

of stochasticity: one is due to our poor knowledge of the initial conditions, therefore modelling the uncertainty coming from "ignorance" of the true state—*epistemic uncertainty*, in the language of philosophy of science; the other is due to the dynamical properties of dispersion and to the internal and external interactions of a large number of degrees of freedom, making the system *chaotic* or at least with good *mixing* properties—called *ontic uncertainty* by philosophers of science. This justifies an *ergodic hypothesis* analogous to Boltzmann's, but rather in the practical weaker sense of Khinchin [108, 35], which allows for a full use of ensemble averages. Such passages will be explained further in the WT derivation of section 4.2, when the *kinetic limit* (large box and weak nonlinearity) is taken. Whether the probability can be interpreted in an epistemic or ontic sense, or both, does not make a difference in the math, and for our concerns also to any practical purpose other than an interpretation of results. In the end, once the *bridge law* of ergodicity is accepted, the epistemic aspect of probability can be reasonably considered to become ontic as well, in the same way that Gibbs ensemble description is equivalent to Boltzmann's macrostate at thermodynamic equilibrium. Indeed, the idealized scale separation we are talking about are not even close to the "astronomic separation" of the molecular and the macroscopic scales of thermodynamics, of the order of Avogadro's number. Still, if one is to treat the macroscopic scale as a continuum, it is important to have clear in mind what type of assumptions are needed to that end. The "thumb rule" for LTE, i.e. a proper mesoscopic-box size allowing for the definition of a deterministic density, is that the fluctuations at the boundary of the box be negligible compared to the averages of what is inside the box. If this is fulfilled, then the finite size effects will be negligible, and the macroscopic density will be independent of the actual size and shape of the box itself.

The "microscopic" (human) scale

Finally, we come to the scale of the actual waves encountered on the ocean surface. For waves like in Fig. 1.1, where the flag is pointing we have a carrier period of 10 s. For waves in deep water, the dispersion relation is $\omega(k) = \sqrt{gk}$, where g is the gravity acceleration. Using this information and the definition of the wave length $\lambda = 2\pi/k$, we have that the carrier wave has a wave length of about 150 m. This means that within a mesoscopic box there are hundreds of waves. As required, the mesoscopic box is large compared to the dimensions of the single waves: a second sharp scale separation is present between the mesoscale and the human scale. In Fig. 1.2 we show the wave patterns on the small scale, recorded in the surroundings of a platform in the Adriatic Sea. On such scales, the governing equations are deterministic, and if one wishes to have any predictive power against the formation of extreme wave fronts, it is necessary to deal with the deterministic equations of motion. This is true especially if one wants to quantify single fluctuations and not

just average behaviors, i.e. if one wants to switch from a CLT to a LDT framework able to account reliably for the extreme excursions. Whatever the method employed, it will need to account for the uncertainty coming from the stochastic bath of the mesoscale, as well as the deterministic faster dynamics governing at the small scale. We add that in principle the fluctuations can also be studied inferring information from the mesoscopic stochastic theory to obtain predictions on the small scale fluctuations. This procedure was followed in the past leading to a theory named *quasi-determinism*. Forcingly, by following this path the nonlinear dynamical effects are neglected and the CLT Gaussianity of WT is enforced onto the small-scale fluctuations.

Most of the work that follows is oriented to devise an LDT method allowing us to bridge the WT statistical prior and the governing deterministic dynamics, and then to apply it to the problem of the rogue waves at sea. Actually, the results are quite striking: as shown in Chapter 4 the method will turn out successfull at reproducing the extreme waves in numerical experiments, proving that indeed a LDT regime is attained, and subsequently the results will be confirmed for real extreme water waves in a wave flume experiment.

1.3 Outline

Theoretical framework

The rest of the manuscript is organized as follows:

In chapter 2 the theoretical aspects of the method are worked out, starting from an introductive paragraph 2.1. The main result is the large deviation principle presented in the form of a theorem in section 2.2.1. It is established under which conditions the extreme events occur in a predictable way, as the minimizer of the LDT action functional.

In section 2.2.2 a simple illustration for the particular case of a Gaussian measure and a linear observable is used to familiarize with the method. It will be helpful to discuss how our setup relates to LDT. In this simple case a rescaling by a small parameter ϵ shows that our approach reduces to *Gärtner-Ellis theorem*—at least, in the tails—in particular cases where the speed of the LDP is known. In general, though, our main result is an unusual LDP where the speed is unknown and implicitly contained in the rate function.

Section 2.2.3 is devoted to argue why our result is indeed in the realm of LDT, and how it is oriented to the study of the extreme events in dynamical systems. Importantly, as will be clear in the Applications Part, our method is designed to investigate extreme events for observables depending on complicated, high-dimensional, nonlinear dynamics which would be difficult or even impossible to study by means of other conventional LDT tools. For clarity, we present another example in section 2.2.4, from which we gain an instructive insight on the conceptual difference

from Gärtner-Ellis.

In chapter 3 it is shown how the minimization can be numerically performed in an efficient way using tools from optimal control. Two alternative routes can be followed to compute the gradient of the LDT action, either via the direct method (section 3.1.1) involving the Jacobian matrix (sensitivity of the field with respect to the random parameters), or via the adjoint method (section 3.1.2). As discussed in section 3.2, the gradient computation is necessary at each step of an outer loop performing a gradient descent in the LDT action landscape.

The findings of chapter 2 are illustrated on various physical applications in the Applications Part (chapters 4 through 6).

Applications

By itself, chapter 4 represents an all-round study of the extreme events in the context of surface gravity waves.

Section 4.1 contains a brief review of the derivation of the Zakharov equation in the limit of small nonlinearity, making the connection with the 4-wave turbulence theory, and the subsequent derivation of the nonlinear Schrödinger equation (NLSE) in the narrow-band limit—both of which will be used in the following sections.

Starting from the setup of *wave turbulence* (WT), in section 4.2 we present an original derivation of the leading-order equation for the multimode PDF for 4-wave resonant systems—the class in which surface gravity waves belong. The technique employed is based on the Feynman-Wyld diagrams. With the help of numerics, it is explored how intermittent phenomena and non-Gaussianity could possibly arise in the WT framework. We find quite convincing evidence that deviations from Gaussianity cannot be retrieved within the mean-field formalism of WT, even when forcing and dissipation are present. Indeed, WT remains the effective theory describing the energetic transfers within the wave field, determining the Kolmogorov-Zakharov stationary spectrum.

In section 4.3, we start from the Gaussian statistical prior of WT, with a spectrum given by a parametrization often used by oceanographers—but which could in principle be the stationary spectrum of WT—and we use such prior for the LDT method, using the modified nonlinear Schrödinger equation (MNLS) which governs the dynamics on the fast timescale. Then, with a setup in analogy with Bayesian inference, the appearance of rogue waves in the modelled 1D deep sea is investigated. The rogue wave precursors are wave patterns of regular height but with a very specific shape that is identified explicitly, thereby potentially allowing for early detection.

In section 4.4 we present an *experimental study* performed in a water tank, where analogous results are confirmed for real-world extreme waves on the water surface.

This work represents the *first experimental evidence of hydrodynamic instantons*. Moreover, our theory yields a *unified picture of two existing theories* of extreme waves: in the linear limit, the quasi-determinism of Lindgren-Boccotti, and in the highly nonlinear limit, the gradient-catastrophe of Tovbis-Bertola. In the two limits, the instantons reduce to the existing predictions of the limiting theories, and they describe accurately the extreme events in the in-between regimes.

In chapter 5, the LDT method is applied to the nonlinear Schrödinger equation (NLS) with random initial conditions for nonlinear laser waves in fiber optics, in the context of what is known as *integrable turbulence*. We study the problem of the onset of rogue waves out of a bath of random waves taken as initial condition for NLS, and we discuss possible implications that our results may have.

The example of a rod with random elasticity pulled by a time-dependent force is treated in chapter 6. The LDP can be used here to infer the probability of atypically large extensions of the rod. Such example serves as a nontrivial application to nonequilibrium statistical mechanics, and it exemplifies a nice problem of search of the optimal protocol. Moreover, in this chapter the numerical optimization is carried out with the adjoint method.

Closing remarks are found in chapter 7, trying to clarify some concepts in light of the applications presented, and suggesting possible future directions both from the theoretical and the applicative perspectives.

Part I

Theoretical Framework

Chapter 2

Large deviation method for extreme events

Part of the material in this chapter was published in

G. Dematteis, T. Grafke, and E. Vanden-Eijnden. "Extreme event quantification in dynamical systems with random components". https://arxiv.org/abs/1808.10764, 2018 (to appear in "SIAM/ASA Journal of Uncertainty Quantification").

2.1 Probabilistic setup

The governing equations we use to model complex phenomena are often approximate. For example, we may not know exactly the initial and/or boundary conditions necessary to integrate these equations. Other parameters entering these equations can also be uncertain, either because we are not sure of the model itself or because these parameters may vary from situations to situations in a way that is difficult to predict in detail. The question then becomes whether we can quantify how our imperfect knowledge of the system's parameters impact its behavior. This question lends itself naturally to a probabilistic formulation. Consider for example the case of a dynamical system whose state at time t can be specified by some $u(t)$ which can be a vector or a field and satisfies

$$\partial_t u = b(u, \vartheta), \qquad u(t=0) = u_0(\vartheta). \tag{2.1}$$

Here $b(u, \vartheta)$ is a given vector field and ϑ denotes the set of parameters we are uncertain of. Assuming that these parameters take value in some set Ω, which can again be finite or infinite dimensional, it is then natural to equip Ω with a probability measure μ to quantify our uncertainty. This makes ϑ random, and therefore the solution to (2.1) becomes a stochastic process. Denoting it by $u(\cdot, \vartheta)$,

we can ask questions about the statistics of this process. For example, if $f(u)$ is a scalar valued observable, we can define

$$P_T(z) \equiv \mathbb{P}\left(f(u(T,\vartheta)) \geq z\right), \qquad z \in \mathbb{R}, \tag{2.2}$$

where $\mathbb{P}$ denotes the probability over μ and $T > 0$ is some observation time. The probability (2.2) is useful e.g. in the context of certification problem where, given $z \in \mathbb{R}$ and $\epsilon > 0$ (typically z large and ϵ small), we wish to verify that $P_T(z) \leq \epsilon$. Other quantities of interest include

$$\mathbb{P}\left(\int_0^T f(u(t,\vartheta))dt \geq z\right), \qquad \mathbb{P}\left(\sup_{0\leq t\leq T} f(u(t,\vartheta)) \geq z\right), \qquad \text{etc.} \tag{2.3}$$

The numerical estimation of (2.2) or (2.3) can be performed by Monte Carlo sampling methods: generate N independent realizations of ϑ, for each evaluate $f(u(T,\vartheta))$ via integration of (2.1), and compute the fraction of these realizations for which $f(u(T,\vartheta)) \geq z$. As $N \to \infty$, this fraction will converge to $P_T(z)$. This direct approach is not effective when $P_T(z)$ is small, however, since the relative error of the estimator just described is $\sqrt{(1-P_T(z))/(NP_T(z))} \sim 1/\sqrt{NP_T(z)}$. This means that in order to get an estimate accurate to order $\delta \ll 1$, we need to use $N = O\left(\delta^{-2}P_T^{-1}(z)\right)$ samples, which can become prohibitively expensive as $P_T(z)$ gets smaller. This is problematic since it excludes from consideration events that are rare but may nonetheless have dramatic consequences. Similar issues arise if we replace (2.1) by some time independent equation like

$$0 = b(u,\vartheta), \tag{2.4}$$

where $b(\cdot,\vartheta)$ is some function of u and possibly its derivatives and (2.4) is supplemented with boundary conditions that may also depend on the random parameter ϑ. The solution to (2.4) defines a complicated map $u(\vartheta)$, and given a scalar valued observable $f(u)$, the estimation of

$$\mathbb{P}\left(f(u(\vartheta)) \geq z\right), \qquad z \in \mathbb{R} \tag{2.5}$$

will again be challenging when this probability is small, i.e. when the event $f(u(\vartheta)) \geq z$ is rare.

In these situations alternative methods such as those proposed e.g. in [88, 106, 27, 87, 181, 191, 77, 164] must be used to estimate (2.2), (2.3), or (2.5). The approach we introduce in this chapter builds on earlier results found in [49] and uses large deviation theory (LDT) [50, 192] as a tool: we show that, if in (2.2) $P_T(z) \to 0$ as $z \to \infty$, then under some additional assumptions we have

$$P_T(z) \asymp \exp\left(-\min_{\theta\in\Omega(z)} I(\theta)\right) \qquad \text{where} \qquad \Omega(z) = \{\theta : f(u(T,\theta)) \geq z\} \subseteq \Omega. \tag{2.6}$$

Here $\asymp$ indicates that the ratio of logarithms of both sides tends to 1 as $z \to \infty$ and we defined

$$I(\theta) = \max_{\eta} \left(\langle \eta, \theta \rangle - S(\eta) \right), \tag{2.7}$$

where $\langle \cdot, \cdot \rangle$ is a suitable inner product on Ω and $S(\eta)$ is the cumulant generating function of ϑ:

$$S(\eta) = \log \mathbb{E} e^{\langle \eta, \vartheta \rangle} = \log \int_{\Omega} e^{\langle \eta, \theta \rangle} d\mu(\theta) \,. \tag{2.8}$$

We will also show that the minimizer of $I(\theta)$ in $\Omega(z)$, i.e.

$$\theta^\star(z) = \operatorname*{argmin}_{\theta \in \Omega(z)} I(\theta) \,, \tag{2.9}$$

is the point of maximum likelihood in $\Omega(z)$. The most likely way the event $\{f(u(T, \vartheta)) \geq z\}$ occurs is when $\vartheta = \theta^\star(z)$. Similar estimates hold for (2.3) and (2.5) upon straightforward redefinition of the set $\Omega(z)$ upon which the optimization is performed.

Establishing the large deviation principle (LDP) in (2.6) is one of the objectives of this chapter. As we will see in Sec. 2.2.1, this can be done by proving that $\theta^\star(z)$ is a dominating point in $\Omega(z)$, building on results derived e.g. in [20, 143, 23, 101] that provide us with a framework to justify the saddle-point approximations often used in physics [105, 82]. Eq. (2.6) is a somewhat unusual LDP however because there is no small (or large) parameter associated to the random variable ϑ: rather we play with the variable z being large. More precisely, instead of scaling ϑ so that events with a finite z become rare, we keep ϑ as is and look at rare events that occur in the tail of the distribution when $z \gg 1$. As a result, the standard approach developed in [20, 143, 101] must be adapted. Of course, both viewpoints are equivalent up to some appropriate rescaling of the variables ϑ and z, but this rescaling involves the so-called speed of the LDP, which is unknown to us *a priori*. The formulation we adopt can be viewed as a way to estimate this speed.

When (2.6) holds, we can reduce the evaluation of $P_T(z)$ to the minimization problem in (2.9), and a second objective here is to design numerical tools to perform this minimization. As we will see in Sec. 3, this can be done by adapting techniques used in optimal control [189, 21].

2.2 Large deviation principle

2.2.1 Formulation of the theorem

Here we establish (2.6), using background material that can be e.g. found in [20, 143, 101]. For simplicity, we will restrict ourselves to situations where ϑ is finite dimensional, i.e. we assume that $\vartheta \in \Omega \subseteq \mathbb{R}^M$ with $M \in \mathbb{N}$. In this case we can also assume that the inner product $\langle \cdot, \cdot \rangle$ appearing in (2.7) and (2.8) is the standard

Euclidean inner product on $\mathbb{R}^M$. Under appropriate assumptions, the results below will hold also in the infinite-dimensional set-up, when ϑ is a random field, but the arguments to establish them will require generalization (see e.g. [62, 112] for results in infinite dimension). To treat the problems in (2.2), (2.3) and (2.5) on the same footing we also define the map $F : \Omega \to \mathbb{R}$ via

$$\begin{aligned} F(\theta) &= f(u(T,\theta)), && \text{for (2.2)} \\ F(\theta) &= \int_0^T f(u(T,\theta))\, dt, \quad \text{or} \quad F(\theta) = \sup_{0\le t\le T} f(u(T,\theta)), && \text{for (2.3)} \\ F(\theta) &= f(u(\theta)) && \text{for (2.5)} \end{aligned} \tag{2.10}$$

so that we can recast these probabilities into

$$P(z) = \mathbb{P}(F(\vartheta) \ge z) = \mu(\Omega(z)) \qquad \text{where} \qquad \Omega(z) = \{\theta \,:\, F(\theta) \ge z\}\,. \tag{2.11}$$

To proceed, we start by making two assumptions:

Assumption 1 *The map F is continuously differentiable, and such that $|\nabla F(\theta)| \ge K > 0$ for all $\theta \in \Omega$.*

Assumption 2 *The measure μ is such that (this is (2.8))*

$$S(\eta) = \log \mathbb{E} e^{\langle \eta, \vartheta\rangle} = \log \int_\Omega e^{\langle \eta,\theta\rangle} d\mu(\theta) \tag{2.12}$$

exists for all $\eta \in \mathbb{R}^M$ and defines a differentiable function $S : \mathbb{R}^M \to \mathbb{R}$.

Ultimately, Assumption 1 is about the specifics of the governing equation in (2.1) or (2.4) and the observable f: since the field u is typically a complicated function of ϑ, establishing the conditions under which this assumption holds will have to be done on a case-by-case basis. Note that it guarantees that the set $\Omega(z)$ is simply connected with a boundary that is C^1 for all $z \in \mathbb{R}$, with inward pointing unit normal at $\theta(z) \in \partial\Omega(z)$ given by $\hat{n}(z) = \nabla F(\theta(z))/|\nabla F(\theta(z))|$. We could relax the constraint $|\nabla F(\theta)| > 0$, and allow e.g. for the sets $\Omega(z)$ to have several connected components (the number of which could depend on z), but this requires to modify the argument below. Assumption 2 allows us to introduce the tilted measure

$$d\mu_\eta(\theta) = \frac{e^{\langle \eta,\theta\rangle} d\mu(\theta)}{\int_\Omega e^{\langle \eta,\theta\rangle} d\mu(\theta)} = e^{\langle \eta,\theta\rangle - S(\eta)} d\mu(\theta)\,. \tag{2.13}$$

It is easy to see that the mean of μ_η is shifted compared to that of μ. A simple calculation shows that

$$\int_\Omega \theta\, d\mu_\eta(\theta) = \nabla S(\eta)\,, \tag{2.14}$$

and this will allow us to pick η such that the mean of μ_η is precisely at the point minimizing $I(\theta)$ in $\Omega(z)$. Note that

$$\Omega(z+\delta) \subseteq \Omega(z) \qquad \forall z \in \mathbb{R}, \delta \geq 0\,, \tag{2.15}$$

and to establish (2.6) we will find conditions such that (i) $\mu(\Omega(z))$ decreases fast with z and (ii) this probability is dominated by a small region around a single point on $\partial\Omega(z)$. This will require us to make additional assumptions on the geometry of $\Omega(z)$ that we discuss next in connection with properties of the rate function $I(\theta)$ defined in (2.7).

Letting

$$\theta^\star(z) = \operatorname*{argmin}_{\theta \in \Omega(z)} I(\theta)\,, \tag{2.16}$$

we first make:

Assumption 3 *There exists a finite z_0 such that, $\forall z \geq z_0$, $\theta^\star : [z_0, \infty) \to \Omega$ is continuously differentiable and $I(\theta^\star(\cdot))$ is strictly increasing with z with*

$$I(\theta^\star(z)) \to \infty \quad \textit{and} \quad |\nabla I(\theta^\star(z))| \geq K > 0 \quad \textit{as} \quad z \to \infty. \tag{2.17}$$

This assumption implies that $\theta^\star(z) \in \partial\Omega(z)$ for $z > z_0$, i.e. we can replace (2.16) with

$$\theta^\star(z) = \operatorname*{argmin}_{\theta \in \partial\Omega(z)} I(\theta)\,. \tag{2.18}$$

The Euler Lagrange equation for (2.18) is

$$\nabla I(\theta^\star(z)) = \lambda \nabla F(\theta^\star(z)) \tag{2.19}$$

for some Lagrange multiplier λ. Since by definition both S and I are convex functions, by the involution property of the Legendre transform we have

$$S(\eta) = \max_\theta \left(\langle \eta, \theta \rangle - I(\theta)\right)\,, \tag{2.20}$$

and this maximum is achieved at the solution of

$$\eta = \nabla I(\theta) \tag{2.21}$$

in θ. Therefore if we define $\eta^\star(z)$ via

$$\eta^\star(z) = \nabla I(\theta^\star(z)) \tag{2.22}$$

the mean of $\mu_{\eta^\star(z)}$ is $\theta^\star(z)$. From (2.20) this also implies that

$$\langle \eta^\star(z), \theta^\star(z) \rangle - S(\eta^\star(z)) = I(\theta^\star(z))\,, \tag{2.23}$$

which gives the following exact representation formula for $\mu(\Omega(z))$

$$\begin{aligned}\mu(\Omega(z)) &= \int_{\Omega(z)} e^{S(\eta^\star(z))-\langle\eta^\star(z),\theta\rangle} d\mu_{\eta^\star(z)}(\theta) \\ &= e^{-I(\theta^\star(z))} \int_{\Omega(z)} e^{-\langle\eta^\star(z),(\theta-\theta^\star(z))\rangle} d\mu_{\eta^\star(z)}(\theta)\,.\end{aligned} \tag{2.24}$$

To proceed further we need to make some assumptions about $\Omega(z)$. First:

Assumption 4 *For all $z \geq z_0$, the set $\Omega(z)$ is contained in the half-space whose boundary is tangent to $\Omega(z)$ at $\theta = \theta^\star(z)$, i.e.*

$$\Omega(z) \subseteq \mathcal{H}(z) = \{\theta \ : \langle \hat{n}^\star(z), \theta - \theta^\star(z)\rangle \geq 0\}\,, \tag{2.25}$$

where $\hat{n}^\star(z) = \nabla F(\theta^\star(z))/|\nabla F(\theta^\star(z))|$ denotes the inward pointing unit normal to $\partial\Omega(z)$ at $\theta^\star(z)$.

In the terminology of Ney [143], it means that $\theta^\star(z)$ is a dominating point in $\Omega(z)$. If we combine (2.19) and (2.22) we deduce that

$$\frac{\eta^\star(z)}{|\eta^\star(z)|} = \frac{\nabla F(\theta^\star(z))}{|\nabla F(\theta^\star(z))|} = \hat{n}^\star(z) \tag{2.26}$$

and as a result we can use Fubini's theorem to express (2.24) as

$$\mu(\Omega(z)) = e^{-I(\theta^\star(z))} \int_0^\infty e^{-|\eta^\star(z)|s} |\eta^\star(z)| G(z,s)\, ds\,. \tag{2.27}$$

Here we defined

$$G(z,s) = \mu_{\eta^\star(z)}\left(\Omega(z) \setminus \mathcal{H}(z,s)\right), \tag{2.28}$$

with

$$\mathcal{H}(z,s) = \{\theta \ : \ \langle \hat{n}^\star(z), (\theta - \theta^\star(z) - \hat{n}^\star(z)s)\rangle \geq 0\}\ . \tag{2.29}$$

Note that in (2.27) the lower limit of the integral is at $s = 0$ by Assumption 4. Since by definition we have

$$\forall s > 0 \ : \ G(z,s) \in (0,1), \qquad \forall s, s' > 0,\ s' > s \ : \ G(z,s') > G(z,s), \qquad \lim_{s\to 0^+} G(z,s) = 0\,, \tag{2.30}$$

from (2.27) we obtain the upper bound

$$\mu(\Omega(z)) \leq e^{-I(\theta^\star(z))} \int_0^\infty e^{-|\eta^\star(z)|s} |\eta^\star(z)| ds = e^{-I(\theta^\star(z))}\,, \tag{2.31}$$

which implies

$$\frac{\log \mu(\Omega(z))}{I(\theta^\star(z))} \leq -1\,. \tag{2.32}$$

To get a matching lower bound notice that for all $s_1 > 0$ we have

$$\begin{aligned}\mu(\Omega(z)) &\geq e^{-I(\theta^\star(z))} \int_0^{s_1} e^{-|\eta^\star(z)|s} |\eta^\star(z)| G(z,s) ds \\ &\geq e^{-I(\theta^\star(z))} G(z,s_1) \left(1 - e^{-|\eta^\star(z)|s_1}\right) \\ &\geq e^{-I(\theta^\star(z))} G(z,s_1) \frac{|\eta^\star(z)|s_1}{1+|\eta^\star(z)|s_1}\,.\end{aligned} \tag{2.33}$$

Therefore if we make:

Assumption 5 *There exists $s_1 > 0$ such that*

$$\lim_{z\to\infty} \frac{\log G(z,s_1)}{I(\theta^\star(z))} = 0, \tag{2.34}$$

for this s_1 we have (using also Assumption 3 that guarantees that $|\eta^\star(z)| \geq K > 0$)

$$\begin{aligned}\frac{\log \mu(\Omega(z))}{I(\theta^\star(z))} &\geq -1 + \frac{\log G(z,s_1) + \log\left(|\eta^\star(z)|s_1\right) - \log\left(1+|\eta^\star(z)|s_1\right)}{I(\theta^\star(z))} \\ &= -1 + \frac{\log G(z,s_1) - \log\left(1+|\eta^\star(z)|^{-1}s_1^{-1}\right)}{I(\theta^\star(z))} \\ &\to -1 \qquad \text{as} \quad z\to\infty\,.\end{aligned} \tag{2.35}$$

Combining (2.32) and (2.35) we finally deduce

Theorem 1 (Large deviation principle) *Under Assumptions 1–5, the following result holds:*

$$\lim_{z\to\infty} \frac{\log P(z)}{I(\theta^\star(z))} = \lim_{z\to\infty} \frac{\log \mu(\Omega(z))}{I(\theta^\star(z))} = -1. \tag{2.36}$$

Note that (2.36) is just a rephrasing of (2.6).

It is useful to comment on the assumptions on $\Omega(z)$ that lead to Theorem 1. Assumption 3 states that the event $\{F(\vartheta) \geq z\}$ becomes rare as $z \to \infty$, which is clearly necessary for an LDP to apply. Assumption 4 guarantees that all regions in $\Omega(z)$ remain much more unlikely than $\theta^\star(z)$: this assumption can be relaxed, but at the price of having to analyze more carefully how $I(\theta)$ behaves on $\partial\Omega(z)$ and exclude that regions with lower likelihood near this boundary accumulate and eventually dominate the probability. Finally, Assumption 5 is about the shape of the set $\Omega(z)$ near $\theta^\star(z)$. Since the mean of $\mu_{\eta^\star(z)}$ is $\theta^\star(z)$, we know that this measure must have mass in a region around $\theta^\star(z)$ but we need to make sure that this region has sufficient overlap with $\Omega(z)$. For example, if for each $z \geq z_0$ we can insert in $\Omega(z)$ a set that contains $\theta^\star(z)$ on its boundary and is such that its volume remains finite as $z \to \infty$, Assumption 5 will automatically hold. On the other hand, this assumption

could fail for example if $\Omega(z)$ becomes increasingly thin. More discussion about this kind of geometric assumptions can be found e.g. in [100, 112].

It is also interesting to note that (2.27) offers a way to derive asymptotic expansions for $\mu(\Omega(z))$ more refined than (2.36) if we assume that: (i) $|\eta^\star(z)|$ grows with z, i.e. we supplement (2.17) with

$$|\eta^\star(z)| = |\nabla I(\theta^\star(z))| \to \infty \qquad \text{as} \quad z \to \infty\,; \tag{2.37}$$

and (ii) $G(z,s)$ has a specific behavior near $s=0$ as $z\to\infty$. For example, suppose that there is a $C>0$ such that for all $u \geq 0$

$$G(z, |\eta^\star(z)|^{-1}u) \sim C|\eta^\star(z)|^{-\alpha}u^\alpha \qquad \text{with} \quad \alpha > 0 \quad \text{as} \quad z, |\eta^\star(z)| \to \infty\,, \tag{2.38}$$

where $f(z) \sim g(z)$ indicates that $\lim_{z\to\infty} f(z)/g(z) = 1$. Then we have

$$\begin{aligned} P(z) = \mu(\Omega(z)) &= e^{-I(\theta^\star(z))} \int_0^\infty e^{-u} G\left(z, |\eta^\star(z)|^{-1}u\right) du \\ &\sim e^{-I(\theta^\star(z))} C|\eta^\star(z)|^{-\alpha} \int_0^\infty e^{-u}u^\alpha du \\ &= C\Gamma(\alpha+1)|\eta^\star(z)|^{-\alpha} e^{-I(\theta^\star(z))}\,. \end{aligned} \tag{2.39}$$

It is interesting to note that both (2.27) and (2.39) are consistent with $\vartheta|_{\Omega(z)}$ (outcome of the event conditioned on $F(\vartheta) \geq z$) having fluctuations of order $O(|\eta^\star(z)|^{-1})$ away from $\theta^\star(z)$ in the direction parallel to $\eta^\star(z)$. Perpendicular to $\eta^\star(z)$ the fluctuations remain of order $O(1)$ even as $z\to\infty$, but integrating in these perpendicular directions only gives a sub-exponential correction to $\mu(\Omega(z))$. This correction depends on the geometry of the hypersurface $\partial\Omega(z)$ (in particular on its curvature) near $\theta^\star(z)$. This is what is accounted for in (2.39), and this picture will be confirmed in the numerical examples below.

2.2.2 Illustration: Gaussian measure with linear observable

Let us illustrate the LDT optimization in the simple case of a Gaussian random variable ϑ with mean 0 and covariance Id, taking values $\theta \in \mathbb{R}^N$. If we consider a linear observable

$$F(\theta) = \langle b, \theta\rangle, \qquad b \in \mathbb{R}^N\,, \tag{2.40}$$

we have

$$\mathbb{P}(\langle b, \vartheta\rangle \geq z) = (2\pi)^{-N/2} \int_{\langle b,\theta\rangle \geq z} \exp\left(-\tfrac{1}{2}|\theta|^2\right) d\theta\,, \tag{2.41}$$

and a direct calculation shows that

$$\mathbb{P}(\langle b, \theta\rangle \geq z) = \tfrac{1}{2}\operatorname{erfc}\left(\frac{z}{\sqrt{2}|b|}\right) \sim (2\pi)^{-1/2}|b|z^{-1}\exp\left(-\tfrac{1}{2}|b|^{-2}z^2\right) \quad \text{as} \quad z\to\infty\,. \tag{2.42}$$

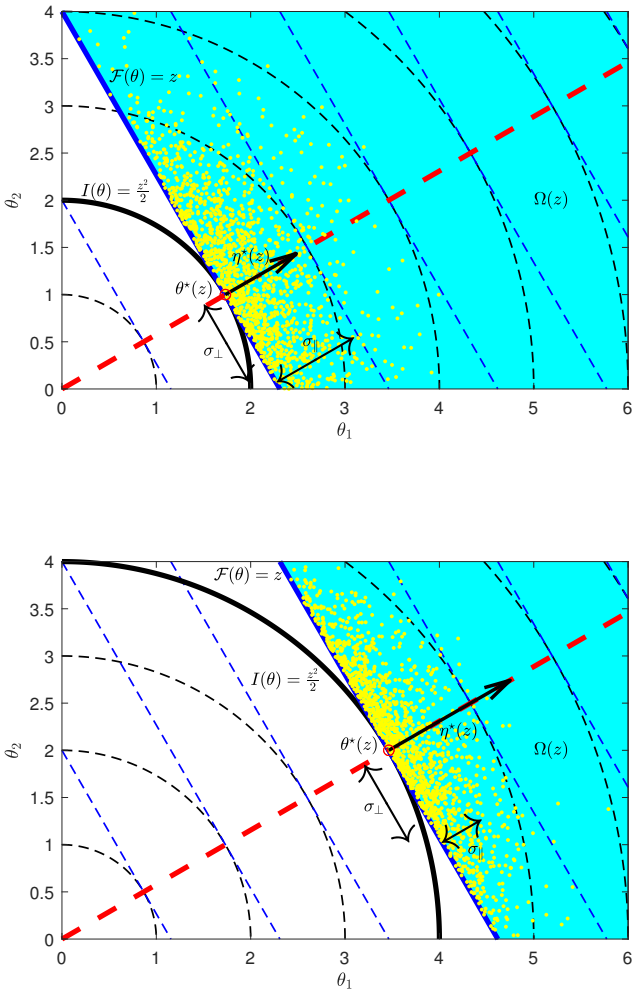

Figure 2.1. Representation of the concentration of the extreme realizations in the support of the parameters ϑ, in the case of $N = 2$, for the linear observable in (2.41) with $b = (\frac{\sqrt{3}}{2}, \frac{1}{2})$, and $I(\theta)$ from (2.43). Both figures represent about 2000 points sampled from the conditional event $F(\vartheta) \geq z$, and in the bottom figure the event size is twice the event size of the figure on top. We notice that as the size doubles, the standard deviation of the fluctuations in the parallel direction shrinks of a factor 2, i.e. the ratio of the lengths of $\eta^\star(z)$, according to the prediction below (2.39). It is in the direction of $\eta^\star(z)$ that the extreme events collapse onto the most likely realization $\theta^\star(z)$. Instead, in the perpendicular directions, tangent to the iso-surface of probability, the fluctuations remain of order $O(1)$. Such degeneration does not contribute to the exponential leading order of the probability, but just to a sub-exponential prefactor estimated in (2.39) via geometric arguments. The saddle-point approximation works as the events become extreme because the projection of the fluctuations $\vartheta - \theta^\star(z)$ in the direction of $\eta^\star(z)$ tends to zero.

Let us check that the LDP derived above is consistent with this result. Here

$$I(\theta) = \tfrac{1}{2}|\theta|^2, \qquad S(\eta) = \tfrac{1}{2}|\eta|^2. \tag{2.43}$$

If we minimize $I(\theta)$ subject to $\langle \theta, b\rangle \geq z$, we deduce

$$\theta^\star(z) = z|b|^{-2}b \qquad \text{and} \qquad I(\theta^\star(z)) = \tfrac{1}{2}|b|^{-2}z^2. \tag{2.44}$$

Comparing this result with (2.42) we see that it is consistent with the prediction in (2.36).

We can also test what the theory can say beyond the log-asymptotic estimate. Here, the planar condition corresponding to $\Omega(z) = \mathcal{H}(z)$ is exactly fulfilled by linearity of $F(\theta) = \langle b, \theta\rangle$. We need to estimate $G(z, |\eta^\star(z)|^{-1})$ as $z \to \infty$. From (2.22) and (2.23) we have that

$$\eta^\star(z) = \nabla I(\theta^\star(z)) = \theta^\star(z) = z|b|^{-2}b, \qquad S(\eta^\star(z)) = \tfrac{1}{2}z^2|b|^{-2}\,, \tag{2.45}$$

and the tilted measure (2.13) at $\eta = \eta^\star(z)$ reads

$$d\mu_{\eta^\star(z)}(\theta) = (2\pi)^{-N/2}\exp\left(-\tfrac{1}{2}|\theta|^2 + zb|b|^{-2}\langle b, \theta\rangle - \tfrac{1}{2}z^2|b|^{-2}\right)d\theta\,. \tag{2.46}$$

Using (2.45), we obtain

$$\begin{aligned} G(z,s) &= \int_{z\leq\langle b,\theta\rangle\leq z+s} d\mu_{\eta^\star(z)}(\theta) \\ &= (2\pi)^{-1/2}\int_0^s \exp\left(-\tfrac{1}{2}u^2\right)du \\ &= \tfrac{1}{2}\,\mathrm{erf}\left(\tfrac{1}{2}\sqrt{2}s\right)\,. \end{aligned} \tag{2.47}$$

As a result

$$G(z, |\eta^\star(z)|^{-1}s) \sim (2\pi)^{-1/2}|\eta^\star(z)|^{-1}s = (2\pi)^{-1/2}|b|sz^{-1} \qquad \text{as} \quad z\to\infty. \tag{2.48}$$

Comparing with (2.38), we see that here $C = (2\pi)^{-1/2}$ and $\alpha = 1$. Therefore (2.39) agrees with (2.42) as expected.

2.2.3 Relationship with Gärtner-Ellis

The example above can be interpreted as a usual LDP by introducing a rescaling by a small parameter ϵ. The meaning is either that the variance of the random parameters is very small, $\vartheta \to \epsilon\vartheta$ (which is conceptually close to the Freidlin-Wentzell approach), or equivalently that the size of the events we look at is very big $z \to \epsilon^{-1}z$. By simply introducing the latter substitution, we can interpret (2.42) as

$$P\left(\tfrac{z}{\epsilon}\right) \asymp \exp\left(-\epsilon^{-2}\tfrac{z^2}{2|b|^2}\right) \quad \text{as} \quad \epsilon\to 0\,, \tag{2.49}$$

where the limit is now taken for ϵ. That is, ϵ^{-2} is the so-called *speed of the LDP*, i.e. the correct scale to use for the limit commonly defining the (properly scaled) rate function:

$$I(z) \equiv -\lim_{\epsilon \to 0} \epsilon^2 \log P(\tfrac{z}{\epsilon}) = \tfrac{z^2}{2|b|^2} \,. \tag{2.50}$$

Now, we realize that instead of (2.36) we could have used Gärtner-Ellis theorem, obtaining the equivalent result (2.49) for the probability tails from the Legendre-Fenchel transform of the scaled cumulant generating function. On top of this, (2.49) also gives us information on the CLT concentration, by knowledge of the global minimum of the rate function at $z = 0$, whereas our asymptotic result (2.36) is restricted to large values of z (the probability tail). Thus, it is important to point out that whenever the speed of the LDP is known, (2.36) reduces to the tail part of some usual form of LDP. This is not completely a surprise, as the proof of (2.36) can be viewed as a parallel to the proof of Gärtner-Ellis theorem, just adapted in a different limit—particularly, we have seen that our assumptions are in analogy to those of [143], ensuring a *dominating-point* regime. On the other hand, often it is not possible to establish *a-priori* the LDP speed. We are going to see in the following that this is the case in many applications of interest. Therefore, the importance of the LDP (2.49) is not of establishing a result with a very hard or original proof, but rather of giving an emphasis on the magnitude of the rare events, independently of other large or small parameters. In the *Applications Part* the reader will find compelling evidence of the usefulness of such an approach, which paves the way for new kinds of application of the LDT techniques. Moreover, in (2.36) the LDP speed is implicitely hidden within the rate function $I(\theta^\star(z))$, but computation of $I(\theta^\star(z))$ can offer a way to establish such speed *a-posteriori*—see section 2.2.5.

2.2.4 Application to the *Example IV.7* of Ref. [187]

In order to understand better in which way our setting relates to other results in large deviations, let us spend some more time on another test example, that is, *Example IV.7* of reference [187]. The following random variable is considered

$$U_n = Y + \frac{1}{n}\sum_{i=1}^{n} X_i \,, \tag{2.51}$$

where Y takes the two values ± 1 with probability $\frac{1}{2}$ each, and the X_i's are i.i.d. normal random variables with zero mean and unit variance. Let us briefly review the main LDT results about such system. Here, the scaled cumulant generating function defined as

$$S_{sc}(\lambda) \equiv \lim_{n \to \infty} \tfrac{1}{n} \log \left[\mathbb{E}\left(e^{n\lambda U_n} \right) \right] \tag{2.52}$$

takes the form

$$S_{sc}(\lambda) = \frac{1}{2}\lambda^2 + |\lambda| \,. \tag{2.53}$$

Being $S_{sc}(\lambda)$ non-differentiable at $\lambda = 0$, the implication of Gärtner-Ellis, that the rate function $I_{sc}(z)$ can be found via Legendre-Fenchel transform of $S_{sc}(\lambda)$, does not hold here. One can see that by definition

$$I_{sc}(z) \equiv -\lim_{n\to\infty} \tfrac{1}{n} \log\left[\mathbb{P}\left(S_n \in ds\right)\right] = \begin{cases} I_-(z) & \text{if } z < 0 \\ I_+(z) & \text{if } z \geq 0 \end{cases} \tag{2.54}$$

where $I_\pm(z) = (z \mp 1)^2/2$, which is a non-convex function. Calculating the Legendre-Fenchel transform of $S_{sc}(\lambda)$, $I^{**}_{sc}(z)$, one obtains instead the convex envelope of $I_{sc}(z)$, which is identically zero for $-1 < z < 1$—due to non-differentiablility of S_n in $\lambda = 0$.

Now, let us see how (2.36) applies to this example. In our framework, we call $\theta_0 = \pm 1$ the binary variable spanning the support of Y and $\theta_i, i = 1, ...n$, the real variables spanning the support of X_i's. Using the definition (2.8), the cumulant generating function writes

$$S(\eta) = \tfrac{1}{2} \sum_{i=1}^{n} \eta_i^2 + \log\cosh \eta_0 \,, \tag{2.55}$$

which is everywhere differentiable in $\mathbb{R}^{n+1}$. Its $n+1-$dimensional Legendre-Fenchel transform (2.7), performing the maximization in η and after some computations involving inverse hyperbolic functions, yields

$$I(\theta) = \tfrac{1}{2} \sum_{i=1}^{n} \theta_i^2 + \sum_{s=\pm 1} \tfrac{1+s\theta_0}{2} \log(1 + s\theta_0) = \tfrac{1}{2} \sum_{i=1}^{n} \theta_i^2 + \log 2 \,, \tag{2.56}$$

where the last equivalence is true since θ_0 can only take the two values ± 1. The rate function will be given by the constrained minimization

$$I(\theta^\star(z)) = \min_{\{\theta : U_n = z\}} I(\theta) \,, \tag{2.57}$$

which in practise we can perform with a Lagrange multiplier enforcing the constraint and transforming it into the unconstrained minimization

$$I(\theta^\star(z)) = \min_{\theta} [I(\theta) - \lambda U_n(\theta)] \,. \tag{2.58}$$

Imposing that the gradient be zero, we find the minimizer $\theta_j^\star(\lambda) = \lambda/n$, function of λ, and then the constraint $U_n(\theta^\star(\lambda)) = z$ provides the parametrization $\lambda(z) = n(z - \theta_0)$. Substituting into $I(\theta)$, we obtain

$$I(\theta^\star(z)) = \min_{\theta_0 = \pm 1} \tfrac{n}{2} \left(z - \theta_0\right)^2 + \log(2) \,, \tag{2.59}$$

which is a non-convex rate function. Not only: if we divide by n, we notice that this is nothing but the correct rate function (2.54). On the right branch, as $z \to \infty$ we have

$$I(\theta^\star(z)) = \tfrac{n}{2} z^2 - nz + o(z) \,. \tag{2.60}$$

Such limit is conceptually different from the $n \to \infty$ limit of (2.54): There, the focus is on how concentration takes place as the numer n of samples increases. Here, given a fixed number of samples n, we ask how the probability tail behaves for the rare events. In the limit $z \to \infty$, since here we know the rate function explicitely, we can express the LDP (2.36) with an auxiliary scaling parameter ϵ (speed of the LDP) as

$$P(z/\epsilon) \asymp e^{\epsilon^{-2} I_{sc}(\theta^\star(z))}, \quad I_{sc}(\theta^\star(z)) = \lim_{\epsilon \to 0} \epsilon^2 \log(P(z/\epsilon)) = \tfrac{n}{2} z^2 . \tag{2.61}$$

This is a rigorous definition, although (2.60) has more detailed information on how the asymptotic regime is attained. We conclude by emphasizing the important difference between defining the cumulant generating function either on the projected space of the observable as (2.52) or in the full space of the parameters as (2.55). When performing the constrained minimization after taking the Legendre-Fenchel transform, the second procedure allowed us to obtain the correct non-convex rate function, whereas in the first procedure the Legendre-Fenchel transform of (2.52) is forcingly convex and some information was missed.

2.2.5 Estimate of the scales of the LDP

If we use ϵ as an auxiliary small parameter, we have that (2.36) writes

$$P\left(\tfrac{z}{\epsilon}\right) \asymp \exp\left(I\left(\theta^\star\left(\tfrac{z}{\epsilon}\right)\right)\right), \quad \text{as} \;\; \epsilon \to 0 . \tag{2.62}$$

We wish to see this as an LDP with a given speed in ϵ, which we need to infer somehow by knowledge of $I(\theta^\star(z))$. For simplicity, suppose that $I(\theta^\star(z)) = g(z)$, where $g(z)$ is a power law whose exponent is unknown. Hence, we can rewrite (2.62) as

$$P\left(\tfrac{z}{\epsilon}\right) \asymp \exp\left(g(\epsilon^{-1}) g(z)\right) , \tag{2.63}$$

in the form of a conventional LDP with rate function $g(z)$ and speed $g(\epsilon^{-1})$. This is a simple illustration of how one can try to unveil the hidden LDP structure of (2.36) a-posteriori.

As can be noticed from the proof in section 2.2.1, the main result (2.36) builds on the dominating-point property of the extreme realizations. This means that the approximation by the most likely realization improves as the fluctuations become smaller. After the rescaling by ϵ, the values taken by the observable are of the order of unity, $z = O(1)$—it is ϵ^{-1} that now tends to infinity. Thus, it is the fluctuations normalized by the signal that need to be small. We recall that the fluctuations in the direction of $\eta^\star(z) \equiv \nabla I(\theta^\star(z))$—the direction perpendicular to the iso-surface of probability, i.e. the direction that matters for the probability, as the others just contribute to a subexponential prefactor—are predicted to be of order $O(|\eta^\star(z)|)$. Heuristically, we can thus define a parameter

$$\delta_z(\epsilon) = \frac{1}{|\eta^\star(\frac{z}{\epsilon})|} \frac{\epsilon}{z} , \tag{2.64}$$

quantifying the relative magnitude of the fluctuations at event-size z as a function of the parameter ϵ: $\delta_z(\epsilon)$ can be seen as a *concentration parameter*. The meaning of it should be clear: when $\delta_z(\epsilon)$ approaches zero, the saddle-point approximation giving (2.36) works well as the events of size $\epsilon^{-1}z$ concentrate around the most likely one, with larger fluctuations in directions that do not contribute to the the leading exponential order of the probability. This parameter can be used to keep under control the speed at which the LDT regime is attained, that is, at which point of the tail the regime of the LDP (2.36) sets in.

For instance, in the Gaussian case above, after introducing the scaling in ϵ as in (2.49), we have $\eta^\star(\frac{z}{\epsilon}) = \frac{1}{\epsilon} zb/|b|^2$, using (2.45). Thus, the value of the concentration parameter is $\delta_z(\epsilon) = \epsilon^2 b/z^2$, which gives an idea of the rate at which the extreme events concentrate dependent of the scaling and of the event size.

A similar estimate holds for the LDP (2.61), for which $\eta^\star(\frac{z}{\epsilon}) = n(\frac{z}{\epsilon} - 1)$ so that $\delta_z(\epsilon) \sim \frac{\epsilon^2}{nz^2}$.

We will see in chapter 6 a case where the quantity $\eta^\star(z) \to \sqrt{M}\alpha$, a constant, as $z \to \infty$. There, for the concentration parameter we have $\delta_z(\epsilon) \sim \frac{\epsilon}{\sqrt{M}\alpha z}$, still tending to zero as $\epsilon \to 0$ but at a slower rate than in the previous cases. In a way, one can say that in the latter case the concentration of measure—and the subsequent saddle-point approximation—does a "worse" job and one needs to go to relatively larger events for the approximation to work properly.

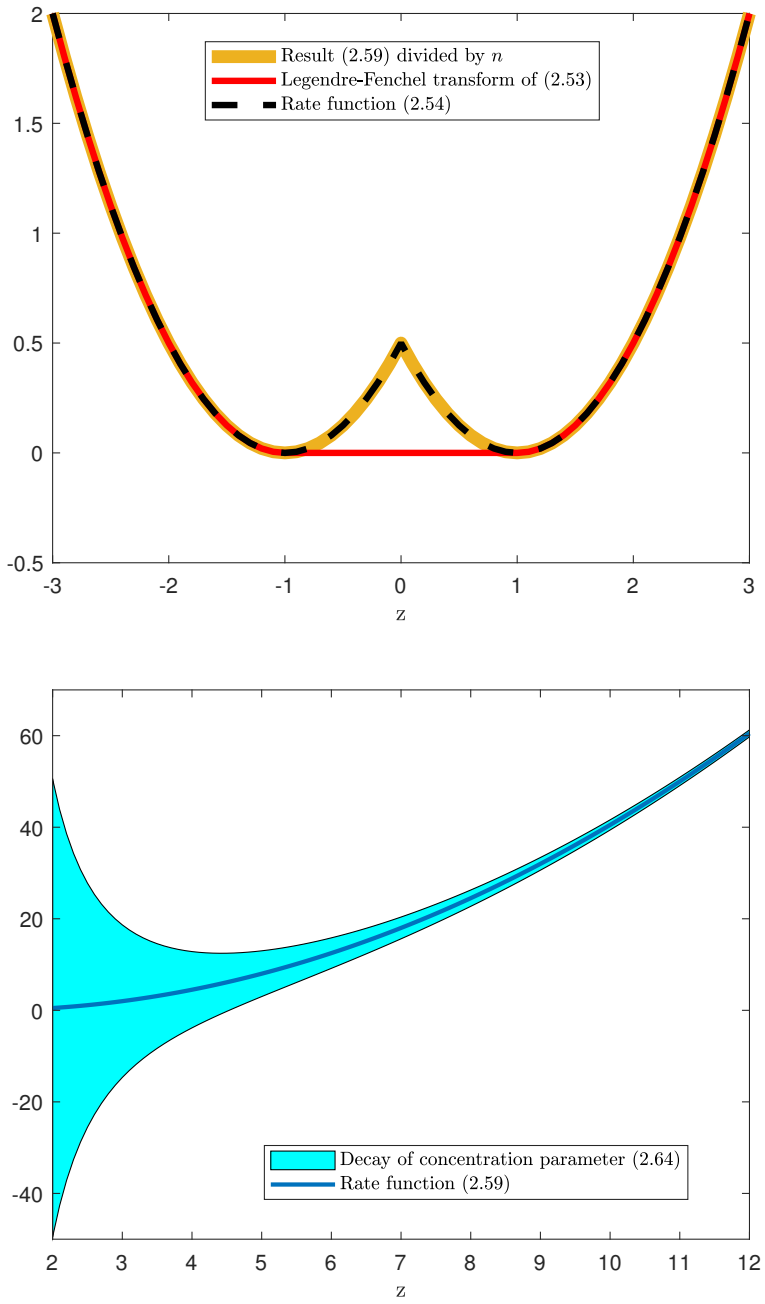

Figure 2.2. Top: The rate functions in section 2.2.4. Bottom: A cartoon of the concentration parameter behavior for the example in section 2.2.4, capturing how fast the LDT regime, based on a saddle point approximation, sets in. Therefore, the narrower the shaded area (stretched by an arbitrary factor in the y direction, so in arbitrary units) around the rate function, the better the rate-function approximation to the probability: all tail events concentrate onto the most likely ones.

Chapter 3

Numerical optimization

The material in this chapter was published in

G. Dematteis, T. Grafke, and E. Vanden-Eijnden. "Extreme event quantification in dynamical systems with random components". https://arxiv.org/abs/1808.10764, 2018 (to appear in "SIAM/ASA Journal of Uncertainty Quantification").

Here we review how to numerically perform the minimization in (2.6) and thereby estimate $P(z)$ – the method can be straightforwardly generalized to consider also the minimization associated with the calculation of (2.3) or (2.5). We impose the constraint $f(u(T)) \geq z$ by adding a Lagrange multiplier term to (2.6), so that the minimization can be rephrased in Hamiltonian formalism by [189, 21]:

$$E(u, \theta) = I(\theta) - \lambda f(u(T))) \,, \tag{3.1}$$

where $u(T)$ should itself be viewed as a function of θ obtained by solving (2.1) with $\vartheta = \theta$, that is

$$\partial_t u = b(u, \theta), \qquad u(t = 0) = u_0(\theta). \tag{3.2}$$

The minimization of (3.1) with $u(T)$ obtained from (3.2) can be performed via steepest descent with adaptive step (line search). This requires to compute the gradient of E with respect to θ, which can be achieved in two ways: by the direct and the adjoint methods [21, 160]. These steps are described next.

3.1 Gradient Calculation

3.1.1 Direct method

The gradient of the cost function with respect to the control reads:

$$\nabla_\theta E(u(T, \theta), \theta) = \partial_\theta E + (\partial_\theta u(T, \theta))^\top \, \partial_u E = \nabla_\theta I - \lambda \, J^\top(T, \theta) \, \partial_u f(u(T, \theta)) \,, \tag{3.3}$$

where $J = \partial_\theta u$ is the Jacobian—componentwise $J_{i,j} = \partial u_i / \partial \theta_j$. An evolution equation for J can be obtained by differentiating (3.2) with respect to θ:

$$\partial_t J = \partial_u b \, J + \partial_\theta b, \qquad J(0) = \nabla_\theta \, u_0. \tag{3.4}$$

Summing up, given the current state of the control, θ^n, we calculate the gradient of the objective function via:

1. *Field estimation:* Obtain the current field u^n by solving

$$\partial_t u^n = b(u^n, \theta^n), \qquad u^n(0) = u_0(\theta^n)\,. \tag{3.5}$$

2. *Jacobian estimation:* Obtain the Jacobian J^n by solving

$$\partial_t J^n = \partial_u b(u^n, \theta^n)\, J^n + \partial_\theta b(u^n, \theta^n), \qquad J^n(0) = \nabla_\theta \, u_0(\theta^n). \tag{3.6}$$

3. *Gradient calculation:* Compute the gradient $(\nabla_\theta E)^n$ via

$$(\nabla_\theta E)^n = \nabla_\theta I(\theta^n) - \lambda \, (J^n(T))^\top \, \partial_u f(u^n(T)). \tag{3.7}$$

3.1.2 Adjoint method

Let us introduce the adjoint field $p(t)$ solution of

$$\partial_t p = -(\partial_u b)^\top p, \qquad p(T, \theta) = \lambda \partial_u f(u(T, \theta)). \tag{3.8}$$

Using this equation as well as the transpose of (3.4) we deduce

$$\begin{aligned} \partial_t (J^\top p) &= \partial_t J^\top p + J^\top \partial_t p \\ &= J^\top (\partial_u b)^\top p + (\partial_\theta b)^\top p - J^\top (\partial_u b)^\top p = (\partial_\theta b)^\top p\,. \end{aligned} \tag{3.9}$$

As a result

$$\int_0^T (\partial_\theta b)^\top p dt = J^\top(T) p(T) - J^\top(0) p(0) = \lambda J^\top(T, \theta) \partial_u f(u(T, \theta)) - (\nabla_\theta u_0)^\top p(0, \theta). \tag{3.10}$$

This expression offers a way to write the gradient of the objective function in (3.3) as

$$\nabla_\theta E = \nabla_\theta I - (\nabla_\theta u_0)^\top p(0, \theta) - \int_0^T (\partial_\theta b)^\top p \, dt\,. \tag{3.11}$$

Using this expression instead of (3.3) is computationally advantageous because it avoid the calculation of the Jacobian J – note in particular that the adjoint field p has the same dimensions as u, independent of the dimensions of the space Ω. The price to pay is the field u must be computed and stored separately since (3.8) for p must be solved backward in time. Summarizing, the gradient of the objective function is now calculated via:

1. *Field estimation:* Obtain the current field u^n by solving

$$\partial_t u^n = b(u^n, \theta^n), \qquad u^n(0) = u_0(\theta^n). \tag{3.12}$$

2. *Adjoint field estimation:* Obtain the adjoint field p^n by solving

$$\partial_t p^n = -(\partial_u b(u^n, \theta^n))^\top p^n, \qquad p^n(T) = \lambda \partial_u f(u^n(T)). \tag{3.13}$$

3. *Gradient calculation:* Compute the gradient $(\nabla_\theta E)^n$ via

$$(\nabla_\theta E)^n = \nabla_\theta I(\theta^n) - (\nabla_\theta u_0(\theta^n))^\top p^n(0) - \int_0^T (\partial_\theta b(u^n(t), \theta^n))^\top p^n(t)\, dt\,. \tag{3.14}$$

Note that equations (3.12) for u and (3.13) for p are adjoint in both space and time. As a result the numerical simulation of these equations has to be done with care, as the integration scheme used for one equation needs to be the adjoint of the other. This is preferably done by using schemes that are self-adjoint. For recent literature on the topic we refer the reader to [197, 95, 193].

3.2 Descent with pre-conditioning of the gradient

Once we have calculated the gradient of the objective function at θ^n, we can make a downhill step in the cost function landscape using

(4) *Descent step with pre-conditioning:*

$$\theta^{n+1} = \theta^n - \alpha^n B^n (\nabla_\theta E)^n\,, \tag{3.15}$$

where B^n is a pre-conditioning $M \times M$ matrix (recall that $\theta \in \Omega \subseteq \mathbb{R}^M$), and $\alpha^n > 0$ is the step size that is tuned optimally at each iteration via line search: this can be done using classical merit functions as discussed in [198].

The estimate of the matrix B^n deserves some further comments. Ideally, B^n should be the inverse of the Hessian of the objective function $E(\theta^n)$, but this Hessian is typically difficult to calculate. Therefore, a simpler solution is to use the Hessian of the prior $I(\theta^n)$, which in the case of a Gaussian measure is simply the inverse covariance matrix C^{-1} (which is independent of θ. Since this estimate coincides with the Hessian of $E(\theta^n)$ only when $\lambda = 0$, it will deteriorate when λ increases and the pre-conditioning may become inefficient. If that is the case, it may be useful to switch to "quasi-Newton" methods such as the BFGS algorithm, or the Limited-Memory BFGS algorithm when M is very large (> 100). In the applications treated in this thesis, the naive pre-conditioning depending only on the prior $I(\theta)$ turned out to be sufficient to perform the optimization efficiently.

Since we are typically interested in calculating (2.6) for a range of values of z, instead of fixing z and trying to determine the corresponding Lagrange multiplier λ in (3.3), it is easier to vary λ and determine *a posteriori* which value of z this leads to. Indeed this offers a parametric representation of $\theta^\star(z)$ via

$$\theta^\star(z(\lambda)) = \tilde{\theta}^\star(\lambda), \qquad z(\lambda) = f(u(T, \tilde{\theta}^\star(\lambda))\,, \tag{3.16}$$

where $\tilde{\theta}^\star(\lambda)$ is the minimizer of $E(\theta)$ at λ fixed. We can then also calculate $I(\theta^\star(z(\lambda))) = I(\tilde{\theta}^\star(\lambda))$ and estimate $P(z(\lambda)) \asymp \exp(-I(\tilde{\theta}^\star(\lambda)))$.

Part II

Applications

Chapter 4

Surface gravity waves: from theory to experiment

Introductory paragraph

The modelling of the sea surface gravity waves is at once very interesting and highly challenging. Waves are essentially energetic perturbations of a medium that propagate in space and time. In this case the medium is the surface of water at the interface with the atmosphere, and all the waves of wavelength larger than about ten centimeters propagate due to gravity, following the dispersion relation $\omega(k) = \sqrt{gk}$ in the case of deep water. In order to characterize the phenomenon from a physicist perspective, we need to understand where the energy of the waves comes from, how it is transferred through different scales, how it happens to focus and form large waves, etc. Possibly, all of this should be done quantitatively by deriving and studying the equations that govern the different aspects.

The main theory to understand the energy transfers in the oceanic wave field is a statistical-mechanical framework called Wave Turbulence. When wind blows quite strongly on the sea surface, the stress forces the very small scale perturbations (gravity-capillary waves) that tend to grow due to resonance and linear instability effects (Miles-Phillips mechanism). Such effect represents the primary source of wave energy at high wave numbers (small wave lengths). Now, the gravity waves turn out to communicate energy by interacting in resonant quartets of waves of different wave number. These resonances are described by the so called Kinetic Equation, which in the case of surface gravity waves comes from the Zakharov equation based on a Hamiltonian with fourth-order nonlinearity, which implies 4-wave interactions. Such interactions transfer energy from the small scales toward the large scales, until the large waves reach a steepness large enough to break (main energy-dissipation process), or the wind eventually stops blowing, etc. The possible phenomenology is indeed very complex and rich and here we are rather interested in

giving an idea of the phenomena at play. The theory of Wave Turbulence describes this and similar out-of-equilibrium wave systems in a statistical way, quantifying the evolution of the energy spectral density, i.e. how the energy of the wave field is distributed and exchanged among the different scales. In section 4.1 we review the derivation of the Zakharov and Nonlinear Schrödinger equations, which are used in the following. In section 4.2 we perform an original derivation of the equations governing the statistics of 4-wave systems (including the Kinetic Equation).

At this point, we are interested in investigating how extreme random waves can arise on the sea surface. The high interest in this question is partly motivated by the fact that such occurrences (sometimes referred to as rogue waves) often appear with higher frequency than expected in a sea of random Gaussian waves. The derivation of section 4.2 leads to an H-theorem implying that the stationary state of Wave Turbulence is indeed a Gaussian state (maximum entropy state). Therefore, we are convinced that in order for a description to go beyond Gaussianity one needs to introduce different strategies. In section 4.3, we apply the LDT method introduced in chapters 2 and 3 to a Gaussian initial state and we look at how a governing nonlinear dynamics on short time scales is able to drive the probability tail of the surface elevation away from Gaussianity. We show that the instantons of the problem, that we can compute explicitly, dominate the statistics of the extreme waves. In section 4.4, finally, we provide experimental support to our finding, exploring different regimes. Our results indicate that the instanton describes the typical extreme realizations for any level of nonlinearity, from the quasi-linear case, reducing to earlier results known in the literature as "quasi-determinism", to the highly-nonlinear case, tending locally to the Peregrine soliton—i.e. an exact solution of the nonlinear Schrödinger equation that plays an important role in the so-called semi-classical (highly nonlinear) regime.

Part of the material in this chapter was published in

S. Chibbaro, G. Dematteis, C. Josserand, and L. Rondoni. "Wave-turbulence theory of four-wave nonlinear interactions". Physical Review E, 96(2):021101, 2017.

S. Chibbaro, G. Dematteis, and L. Rondoni. "4-wave dynamics in kinetic wave turbulence". Physica D: Nonlinear Phenomena, 362:24–59, 2018.

G. Dematteis, T. Grafke, and E. Vanden-Eijnden. "Rogue waves and large deviations in deep sea". Proceedings of the National Academy of Sciences, page 201710670, 2018.

G. Dematteis, T. Grafke, M. Onorato, and E. Vanden-Eijnden. "Experimental evidence of hydrodynamic instantons: The universal route to rogue waves". Under revision, 2018.

4.1 NLS-type equations for surface gravity waves

First, we review how from "first principles",[1] i.e. from the *Euler equations* for the water surface, one can derive the 4-wave Hamiltonian (A.10), on which the statistical description of the waves, named *wave turbulence*, is based. For the water waves such Hamiltonian is the one of the *Zakharov equation*. The derivation involves the *weak-nonlinearity* limit. In fact, upon that Hamiltonian dominated by 4-wave resonances, as for the water waves, we will build the WT theory of 4-wave systems in section 4.2.

Second, we show how one can easily derive the NLS equation from the Zakharov equation, by introducing a Taylor expansion of the dispersion relation up to second order, in the *narrow-band* limit. For this review section, we follow the book [151], and all of the calculations are done for the case of one horizontal dimension for simplicity.

[1] At any level of description, one needs to postulate some laws on which to operate accepting them as first principles. Indeed, Euler equations are obtained from Navier-Stokes' equations with the assumptions of inviscid, incompressible and irrotational fluid. Navier-Stokes can be considered as Newton's laws for a fluid in the continuum description. But Navier-Stokes' equations can be reasonably derived from the Boltzmann transport equation, by introducing a series of nontrivial conceptual ingredients taking us from a mesoscopic stochastic atomistic description to a macroscopic deterministic continuum description. The Boltzmann equation, in turn, was written by Boltzmann somehow "heuristically", and it is not merely a reduction from the microsopic Newton's laws for the fundamental constituents (molecules) to the mesoscopic level: in the passage, further fundamental assumptions are needed such as the *Stosszahlansatz* (bridging the microscopic and the macroscopic descriptions), and emergent phenomena appear, characteristic of the new point of view and extraneous to the lower level. The view that we find more convincing, therefore, is that there is not really a theory which is more fundamental than the others. Each level (or scale) of description typically has an effective theory which has some overlaps with the effective theories at the levels above and the ones at the levels below, but there exists no such thing as a legitimate ranking of fundamental-ness of theories—chemistry cannot be substituted by physics, although many of the first principles of chemistry may be justified by quantum mechanics. Theory reduction is not that simple. The computations of quantum mechanics become operationally almost unfeasible already for the two electrons of the Helium atom, the second element of the periodic table, and chemistry is very much needed if we wish to understand something in the realm of chemical reactions. The theories inform each other in a cooperative way in the overlapping regions, but at any level in general we find one theory (or a few) which is the effective "fundamental" theory at that level of description. If one is interested to these topics, a read to [36] is highly recommended.

4.1.1 From Euler equations to the Zakharov equation expanding in small nonlinearity

Under the assumptions of inviscid, incompressible and irrotational fluid, the equations for the water surface, known as the *Euler equations*, read

$$\begin{cases} \partial_{xx}\phi + \partial_{zz}\phi = 0\,, & -h < z < \eta(x,t) \\ \partial_t\phi + \partial_x\phi\,\partial_x\eta = \partial_z\phi & z = \eta(x,t) \\ \partial_t\phi + \dfrac{1}{2}\left[(\partial_x\phi)^2 + (\partial_z\phi)^2\right] + g\eta - \gamma w = 0\,, & z = \eta(x,t) \\ \partial_z\phi = 0\,, & z = -h \end{cases} \tag{4.1}$$

g is the gravity acceleration, γ is the ratio between the coefficient of surface tension and the fluid density and w is the surface curvature $w = \frac{\partial_x\eta}{\sqrt{1+(\partial_x\eta)^2}}$. Moreover, because of the irrotational assumption we have that $v = \nabla\phi$, where $v(x,t)$ is the velocity field and $\phi(x,t)$ is the velocity potential, while $\eta(x,t)$ is the surface elevation. Introducing the velocity potential at the free surface $\zeta(x,t) = \phi(x, z = \eta(x,t), t)$, in [201] it was shown that η and ζ are canonically conjugated via the Hamiltonian given by the sum of the kinetic energy and the potential energy. The normal variable is introduced as follows in Fourier space (here with the dimensions of an action density, which leads to simpler coefficients in the following),

$$\hat{u}_k(t) = \sqrt{\frac{g}{2\omega_k}}\,\hat{\eta}_k(t) + i\sqrt{\frac{\omega_k}{2g}}\,\hat{\zeta}_k(t)\,. \tag{4.2}$$

where the relation $\omega_k = \sqrt{(g|k| + \gamma|k|^3)\tanh(|k|h)}$ is used for the dispersion relation in arbitrary depth h. Limited to the current section, the index k can indicate either a continuous Fourier variable or a discrete one in the case of a finite physical space with periodic boundary conditions—therefore, the sign of integral will indicate a true integration in the k variable or just a summation, respectively. Expanding in small nonlinearity (i.e. in powers of $\hat{u}_k$) up to third order, from (4.1) and (4.2) we obtain the following equation

$$\begin{aligned} i\partial_t\hat{u}_1 = {} & \omega_1\hat{u}_1 + \int V^{(1)}_{1,2,3}\hat{u}_2\hat{u}_3\delta(k_1 - k_2 - k_3)dk_{23} + 2\int V^{(2)}_{1,2,3}\hat{u}^*_2\hat{u}_3\delta(k_1 + k_2 - k_3)dk_{23} \\ & + \int V^{(3)}_{1,2,3}\hat{u}^*_2\hat{u}^*_3\delta(k_1 + k_2 + k_3)dk_{23} + \int T^{(2)}_{1,2,3,4}\hat{u}^*_2\hat{u}_3\hat{u}_4\delta(k_1 + k_2 - k_3 - k_4)dk_{234} \\ & + \int T^{(1)}_{1,2,3,4}\hat{u}_2\hat{u}_3\hat{u}_4\delta(k_1 - k_2 - k_3 - k_4)dk_{234} + \int T^{(1)}_{1,2,3,4}\hat{u}^*_2\hat{u}^*_3\hat{u}_4\delta(k_1 + k_2 + k_3 - k_4)dk_{234} \\ & + \int T^{(4)}_{1,2,3,4}\hat{u}^*_2\hat{u}^*_3\hat{u}^*_4\delta(k_1 + k_2 + k_3 + k_4)dk_{234}\,. \end{aligned} \tag{4.3}$$

We use the notation: $a_j = a_{k_j}$, $V^{(1)}_{1,2,3} = V^{(1)}_{k_1,k_2,k_3}$, $dk_{123} = dk_1dk_2dk_3$, etc., the the simbols $\delta(\cdot)$ are Dirac's deltas. The equation is valid both for gravity and capillary

waves. However, assuming that the waves are long leads to the dispersion relation of surface gravity waves, $\omega_k = \sqrt{g|k|\tanh(|k|h)}$. For the energy transfers to be active in the equation above, the resonance conditions have to be fulfilled, that is, the following pairs of equations need to be fulfilled at the same time

$$\begin{aligned} k_1 - k_2 - k_3 = 0, \quad \omega_1 - \omega_2 - \omega_3 = 0 \\ k_1 + k_2 - k_3 = 0, \quad \omega_1 + \omega_2 - \omega_3 = 0 \\ k_1 + k_2 + k_3 = 0, \quad \omega_1 + \omega_2 + \omega_3 = 0 \end{aligned} \tag{4.4}$$

for the 3-wave terms. For the capillary waves, 3-wave interactions are resonant, and dominate the spectral energy transfers. For gravity waves, instead, 3-wave resonances are never satisfied—being the dispersion relation concave, see [137]—, and the 4-wave resonances are satisfied only for the interactions of 2 waves going into 2 waves, that correspond to the resonance condition

$$k_1 + k_2 - k_3 - k_4 = 0, \quad \omega_1 + \omega_2 - \omega_3 - \omega_4 = 0\,. \tag{4.5}$$

The integrals that do not satisfy the resonance conditions can be removed by a canonical transformation into a new variable a_k:

$$\begin{aligned} \hat{u}_1 = \hat{a}_1 &+ \int A^{(1)}_{1,2,3}\hat{a}_2\hat{a}_3\delta(k_1 - k_2 - k_3)dk_{23} + 2\int A^{(2)}_{1,2,3}\hat{a}^*_2\hat{a}_3\delta(k_1 + k_2 - k_3)dk_{23} \\ &+ \int A^{(3)}_{1,2,3}\hat{a}^*_2\hat{a}^*_3\delta(k_1 + k_2 + k_3)dk_{23} + \int B^{(2)}_{1,2,3,4}\hat{a}^*_2\hat{a}_3\hat{a}_4\delta(k_1 + k_2 - k_3 - k_4)dk_{234} \\ &+ \int B^{(1)}_{1,2,3,4}\hat{a}_2\hat{a}_3\hat{a}_4\delta(k_1 - k_2 - k_3 - k_4)dk_{234} + \int B^{(1)}_{1,2,3,4}\hat{a}^*_2\hat{a}^*_3\hat{a}_4\delta(k_1 + k_2 + k_3 - k_4)dk_{234} \\ &+ \int B^{(4)}_{1,2,3,4}\hat{a}^*_2\hat{a}^*_3\hat{a}^*_4\delta(k_1 + k_2 + k_3 + k_4)dk_{234}\,. \end{aligned} \tag{4.6}$$

The choice of the coefficients $A_{1,2,3}$ and $B_{1,2,3,4}$ is done in order to eliminate the quadratic and the cubic non-resonant terms in (4.3). For the new variable the evolution equation, named *Zakharov equation*, takes the simple form

$$i\partial_t\hat{a}_1 = \omega_1\hat{a}_1 + \int \tilde{T}^{(2)}_{1,2,3,4}\hat{a}^*_2\hat{a}_3\hat{a}_4\delta(k_1 + k_2 - k_3 - k_4)dk_{234}\,, \tag{4.7}$$

where $\tilde{T}^{(2)}_{1,2,3,4}$ is the effective interaction coefficient that depends on $T^{(1)}_{1,2,3,4}$ and on the terms $A_{1,2,3}$. It is easy to see that the Zakharov equation comes from Hamilton's equation for the canonical conjugate variables a_k and a^*_k, from the Hamiltonian

$$H = \int \omega_1|\hat{a}_1|^2dk_1 + \int \tilde{T}^{(2)}_{1,2,3,4}\hat{a}^*_1\hat{a}^*_2\hat{a}_3\hat{a}_4\delta(k_1 + k_2 - k_3 - k_4)dk_{1234}\,, \tag{4.8}$$

which is exactly in the form of (A.10), starting point for the WT approach of a system with 4-wave resonances.

4.1.2 Derivation of NLS in the narrow-band limit

The Zakharov equation (4.7) is not only the basis for the statistical description of WT, on which for instance the modern ocean forecast models are based, but it represents also a good starting point for the derivation of envelope equations like the NLS equation or higher-order variants. From (4.7), it is sufficient to take the narrow-band approximation, expanding the dispersion relation to second order around k_0,

$$\omega_k = \omega_0 + \left.\frac{d\omega_k}{dk}\right|_{k_0} (k-k_0) + \frac{1}{2}\left.\frac{d^2\omega_k}{dk^2}\right|_{k_0} (k-k_0)^2 + o((k-k_0)^3)\,, \tag{4.9}$$

with $\omega_0 = \omega_{k_0}$, and the interaction coefficient to zero-th order,

$$T^{(2)}_{k_1,k_2,k_3,k_4} = T^{(2)}_{k_0,k_0,k_0,k_0} \simeq k_0^3\,. \tag{4.10}$$

Furthermore, we use a new variable $\hat{b}_k = \hat{a}_k e^{-i\omega_0 t}$ removing a phase, and we Fourier transform to finally obtain

$$i\,(\partial_t b + c_g\,\partial_x b) = \beta\,\partial_{xx} b + \alpha\,|b|^2 b\,, \tag{4.11}$$

with $c_g = \left.\frac{d\omega_k}{dk}\right|_{k_0} = \frac{\omega_0}{2k_0}$ the group velocity, $\beta = \frac{1}{2}\left.\frac{d^2\omega_k}{dk^2}\right|_{k_0} = \frac{\omega_0}{8k_0^2}$ the dispersion coefficient, and $\alpha = k_0^3$ the coefficient of nonlinearity.

In equation (4.32) in the next section, the Dysthe equation, one can recognize the four terms of (4.8). The other terms come from an higher order expansion in both nonlinearity and narrow band [60].

4.1.3 NLS equation with spatial evolution

Due to the narrow-band and small-nonlinearity limits, the last two terms of the NLS equation (4.11), dispersion and nonlinearity respectively, are of higher order compared to the other terms. Thus, at leading order one can use this information to express the space derivative in terms of the time derivative,

$$\partial_x b = -\frac{1}{c_g}\partial_t b\,. \tag{4.12}$$

Differentiating the latter one more time in x and using the fact that the derivatives commute, we obtain

$$\partial_{xx} b = -\frac{1}{c_g}\partial_t\partial_x b = \frac{1}{c_g^2}\partial_{tt} b\,. \tag{4.13}$$

Replacing (4.12) and (4.12) in (4.11), we obtain a version of the NLS equation with evolution in space

$$i\left(\partial_x b + \frac{1}{c_g}\,\partial_t b\right) = \frac{\beta}{c_g^3}\,\partial_{tt} b + \frac{\alpha}{c_g}\,|b|^2 b\,. \tag{4.14}$$

4.2 Wave turbulence: a powerful "mean field" theory

4.2.1 Wave turbulence theory of 4-wave resonant systems

Dispersive waves are ubiquitous in nature, and their nonlinear interactions make them intriguing and challenging [196, 15]. Wave Turbulence is the theory that describes the statistical properties of large numbers of incoherent interacting waves, with tools such as the *wave kinetic equation* analytically derived in the late sixties. This equation describes the evolution of the wave spectrum in time, when homogeneity and weak nonlinearity are assumed [206, 137, 142]. It has been applied to numerous phenomena, including ocean waves [110, 73], capillary waves [163, 70] Alfvén waves [85], optical waves [157] and solid oscillations [54, 135, 22, 131, 98, 99]. It is the analogue of the Boltzmann equation for classical particles and it allows the Rayleigh-Jeans equilibrium state as well as non-equilibrium solutions, in terms of Kolmogorov-Zakharov (KZ) spectra [200].

To characterise the invariant measure of the dynamics, that is to find the complete statistical description concerning all quantities of interest, an important step has been taken by Sagdeev and Zaslavski [209], who obtained the Brout-Prigogine equation for the probability density function (pdf) of wave turbulence [24]. More recently, this statistical framework has been nicely revisited using the diagrammatic technique [137] and performing analytical calculations, in the 3-wave case [39, 38, 69]. Interestingly, many experimental and theoretical results have shown that deviations from wave-turbulence predictions can be found for rare events, *e.g. intermittency* [129, 73, 72, 125, 139, 75]. This seems to be the case when a more general theoretical framework [141, 17, 41, 127, 102] is required, because the nonlinearities are not small [34, 31].

4.2.2 Derivation of the $M-$mode PDF equation and its implications for the stationary state

Following the results that we obtained in [32, 33], in this section the complete wave-turbulence theory is developed for a fully general 4-wave system, whose hamiltonian is expressed by the following canonical expression[2] (the details of the derivation, based on the Feynman-Wyld diagrammatic averaging technique, can be found in

[2]We have seen in 4.1 that the Hamiltonian of the Zakharov equation is precisely in this form, with $\mathcal{H}_k^\sigma = T^{(2)}_{1,2,3,4}$, non-vanishing only when σ has two "plus" and two "minus" signs, i.e. interactions of type 2 waves $\rightarrow$ 2 waves.

Appendix A):

$$H = \frac{1}{2}\sum_{1} \omega_1 A_1^{\sigma_1} A_1^{-\sigma_1} + \epsilon \sum_{1234} \mathcal{H}^{\sigma}_{\underline{\mathbf{k}}}\, A_1^{\sigma_1} A_2^{\sigma_2} A_3^{\sigma_3} A_4^{\sigma_4}\, \delta_{\underline{\sigma}\cdot\underline{\mathbf{k}},\mathbf{0}}\ . \tag{4.15}$$

Here, ω_1 is the normal frequency of wave 1, that nonlinearly interacts with waves 2,3,4 with coupling constant $\mathcal{H}^{\sigma}_{\underline{\mathbf{k}}}$, $\sum_i \doteq \sum_{\sigma_i=\pm 1}\sum_{\mathbf{k}_i\in\Lambda_L^*}$, $\Lambda_L^* = \frac{2\pi}{L}\mathbb{Z}_M^d$. $A^{\sigma}_{\mathbf{k}} = \frac{1}{\sqrt{2}}(P_{\mathbf{k}} + i\sigma Q_{\mathbf{k}})$ are the canonical variables of the wave-field, whose real and imaginary parts are the coordinates and momenta. $\sigma = \pm 1$ represents the "spin" of a wave, so that $A^{+}_{\mathbf{k}} \doteq A_{\mathbf{k}}$, $A^{-}_{\mathbf{k}} \doteq A^{*}_{\mathbf{k}}$ (* is complex conjugation).

Given the Hamiltonian (A.10), we concisely derive the equations of motion in terms of canonical normal variables; the details are given in Ref. [33]. First, recall that the action-angle variables (amplitudes and phases) for the linear dynamics are defined by $J_{\mathbf{k}} = |A^{\sigma}_{\mathbf{k}}|^2$ and $\varphi_{\mathbf{k}} = \sigma \arg(A^{\sigma}_{\mathbf{k}})$, so that $A^{\sigma}_{\mathbf{k}} = \sqrt{J_{\mathbf{k}}}\psi^{\sigma}_{\mathbf{k}}$, where $\psi_{\mathbf{k}} = \exp(i\varphi_{\mathbf{k}})$. Then, the Liouville measure μ preserved by the Hamiltonian flow reads: $d\mu = \prod_{\mathbf{k}} dQ_{\mathbf{k}}dP_{\mathbf{k}} = \prod_{\mathbf{k}} \frac{1}{i}dA^{+}_{\mathbf{k}}dA^{-}_{\mathbf{k}} = \prod_{\mathbf{k}} \frac{1}{i}da^{+}_{\mathbf{k}}da^{-}_{\mathbf{k}} = \prod_{\mathbf{k}} dJ_{\mathbf{k}}d\varphi_{\mathbf{k}}$. $A^{\sigma}_{\mathbf{k}}$ and $a^{\sigma}_{\mathbf{k}}$ are linked by the rotation in the complex plane: $A^{\sigma}_{\mathbf{k}} = a^{\sigma}_{\mathbf{k}} e^{i\sigma\omega_{\mathbf{k}}t}$. The equations of motion with 4-wave interactions can thus be expressed by ($\sigma = +1$ when it is omitted):

$$\begin{aligned}\frac{\partial a_1}{\partial t} \;=\; & \epsilon \sum_{234} \mathcal{L}^{+\sigma_2\sigma_3\sigma_4}_{1234} a_2^{\sigma_2} a_3^{\sigma_3} a_4^{\sigma_4} \\ & \times \exp\left[i\left(-\omega_1 + \sigma_2\omega_2 + \sigma_3\omega_3 + \sigma_4\omega_4\right)t\right] \\ & \times \delta_{-\mathbf{k}_1+\sigma_2\mathbf{k}_2+\sigma_3\mathbf{k}_3+\sigma_4\mathbf{k}_4,\mathbf{0}}\end{aligned} \tag{4.16}$$

For a system with N modes in a box of size L, the complete statistical description of the field is given by the *generating function*, defined by:

$$\mathcal{Z}_L[\lambda,\mu,T] \doteq \left\langle \exp\left(\sum_{\mathbf{k}\in\Lambda_L^*} \lambda_{\mathbf{k}} J_{\mathbf{k}}(T)\right) \prod_{\mathbf{k}\in\Lambda_L^*} \psi_{\mathbf{k}}^{\mu_{\mathbf{k}}}(T) \right\rangle\ , \tag{4.17}$$

where $\lambda_{\mathbf{k}} \in \mathbb{R}$, $\mu_{\mathbf{k}} \in \mathbb{Z}$, $\forall \mathbf{k} \in \Lambda_L^*$.

Assuming that the canonical wavefield enjoys the *Random Phase* (RP) property at the initial time, we have averaged over phases using the *Feynman-Wyld diagrams* [137]. Further, taking the large-box limit, we have normalized the amplitudes in such a way that the wave spectrum remains finite. This step is crucial for the evaluation of the different diagrams [69]. Then, taking the large-box limit, followed by the *small nonlinearity* limit, and introducing the nonlinear time $\tau = \epsilon^2 T$, we have formally obtained the following closed equation for the generating function

(the *characteristic functional*):

$$\begin{aligned}
&\frac{d\mathcal{Z}[\lambda,\mu,\tau]}{d\tau} = -192\pi\delta_{\mu,0} \\
&\times\sum_{\underline{\sigma}}\int d^dk_1d^dk_2d^dk_3d^dk_4\lambda\left(\mathbf{k}_1\right)\left|\mathcal{H}_{1234}^{-\sigma_2\sigma_3\sigma_4}\right|^2\delta(\tilde{\omega}_{234}^1) \\
&\times\delta_{234}^1\left(\frac{\delta^3\mathcal{Z}}{\delta\lambda(\mathbf{k}_2)\delta\lambda(\mathbf{k}_3)\delta\lambda(\mathbf{k}_4)} - \sigma_2\frac{\delta^3\mathcal{Z}}{\delta\lambda(\mathbf{k}_1)\delta\lambda(\mathbf{k}_3)\delta\lambda(\mathbf{k}_4)} + \right. \\
&\left. -\sigma_3\frac{\delta^3\mathcal{Z}}{\delta\lambda(\mathbf{k}_1)\delta\lambda(\mathbf{k}_2)\delta\lambda(\mathbf{k}_4)} - \sigma_4\frac{\delta^3\mathcal{Z}}{\delta\lambda(\mathbf{k}_1)\delta\lambda(\mathbf{k}_2)\delta\lambda(\mathbf{k}_3)}\right),
\end{aligned} \tag{4.18}$$

which constitutes the main ingredient of the present section. The frequency in $\delta(\tilde{\omega}_{234}^1)$ has been renormalised [137] as $\tilde{\omega}_{\mathbf{k}} \doteq \omega_{\mathbf{k}} + \Omega_{\mathbf{k}}$, taking into account the self-interactions possible in 4-wave systems, that do not contribute to the nonlinear interactions but shift the linear frequency.

The characteristic functional constitutes the most detailed description of the phenomenon [134], for which the following holds: (i) the RP property of the initial field is preserved in time, implying the validity of eq.(A.139) for $\tau > 0$; (ii) eq.(A.139) has a solution preserving in time the stricter *Random Phase and Amplitude* (RPA) property of an initial wavefield, *i.e.* the possible factorization of $\mathcal{Z}[\lambda,\mu,0]$; (iii) differentiating with respect to the $\lambda_{\mathbf{k}}$'s, the *spectral hierarchy* for the moments, analogous to the BBGKY hierarchy in Kinetic Theory, is obtained. Then, RPA allows us to close the hierarchy, leading to the wave spectrum equation, the *kinetic equation*.

As the characteristic functional gives too detailed information, in relevant situations we have derived the equation for the *characteristic function* $\mathcal{Z}^{(M)}$, that concerns a number M of modes, and enjoys the same properties of $\mathcal{Z}[\lambda,\mu,\tau]$ [33].

Then, under the RPA hypothesis, we derived a closed fully general equation for the *1-mode pdf* that reads [33]:

$$\frac{\partial P}{\partial\tau} = -\frac{\partial F}{\partial s} = \frac{\partial}{\partial s}\left[s\left(\eta_{\mathbf{k}}\frac{\partial P}{\partial s} + \gamma_{\mathbf{k}}P\right)\right], \tag{4.19}$$

$$\begin{aligned}
\eta_{\mathbf{k}} &\doteq 192\pi\sum_{\underline{\sigma}}\int d^d\mathbf{k}_2d^d\mathbf{k}_3d^d\mathbf{k}_4\delta_{234}^{\mathbf{k}}\delta\left(\tilde{\omega}_{234}^{\mathbf{k}}\right)\left|\mathcal{H}_{\mathbf{k}234}^{-\sigma_2\sigma_3\sigma_4}\right|^2 \\
&\qquad\times n(\mathbf{k}_2)n(\mathbf{k}_3)n(\mathbf{k}_4) \geq 0, \\
\gamma_{\mathbf{k}} &\doteq 192\pi\sum_{\underline{\sigma}}\int d^d\mathbf{k}_2d^d\mathbf{k}_3d^d\mathbf{k}_4\delta_{234}^{\mathbf{k}}\delta\left(\tilde{\omega}_{234}^{\mathbf{k}}\right)\left|\mathcal{H}_{\mathbf{k}234}^{-\sigma_2\sigma_3\sigma_4}\right|^2 \\
&\qquad\times\left[\sigma_2n(\mathbf{k}_3)n(\mathbf{k}_4) + \sigma_3n(\mathbf{k}_2)n(\mathbf{k}_3) + \sigma_4n(\mathbf{k}_2)n(\mathbf{k}_3)\right]
\end{aligned} \tag{4.20}$$

The conservation equation for P explicitly expresses F, the flux of the 1-mode probability in the amplitude space. This is a nonlinear Markov evolution equation

in the sense of McKean. As a matter of fact, the solutions must satisfy a set of self-consistency conditions: $n(\mathbf{k},\tau) = \int ds\ sP(s,\tau;\mathbf{k})$, where $n(\mathbf{k},\tau)$ is the spectrum, that also appears in the formulas for the coefficients (A.126). The derivation of the standard kinetic equation from equation (A.142) is straightforward. Let us assume that the wave turbulence picture is valid for $s \in (0, s_{nl})$, where the upper bound of the interval can also be $+\infty$ (a fact that will be discussed later). Using (A.142), the definition of the wave spectrum $n(\mathbf{k}) = \int_0^{s_{nl}} sP(s)ds$ and integrating by parts, we obtain

$$\frac{\partial n}{\partial \tau} = \eta_{\mathbf{k}} - \gamma_{\mathbf{k}} n - s_{nl}(F(s_{nl}) + \eta_{\mathbf{k}} P(s_{nl}). \tag{4.21}$$

The last term is a null term that has to vanish in order for the equation to be satisfied in general, giving a boundary condition in the amplitude space at $s = s_{nl}$. What we are left with is nothing but the kinetic equation. To make it clear for a concrete example of a 4−wave resonant system where not only $2 \to 2$ wave interactions are present, we derive the kinetic equation for the vibrating plates [54]. Writing (4.21) in the 2−dimensional case, we obtain

$$\begin{aligned}\frac{\partial n}{\partial \tau} = 192\pi \sum_{\underline{\sigma}} \int d^2\mathbf{k}_1 d^2\mathbf{k}_2 d^2\mathbf{k}_3 \delta^{(2)\mathbf{k}}_{123} \delta(\tilde{\omega}^{\mathbf{k}}_{123}) |\mathcal{H}^{-\sigma_2\sigma_3\sigma_4}_{\mathbf{k}\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3}|^2 \\ \times n_{\mathbf{k}} n_1 n_2 n_3 \cdot \left(\frac{1}{n_{\mathbf{k}}} + \frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2} + \frac{\sigma_3}{n_3}\right),\end{aligned} \tag{4.22}$$

which is the same equation as in [54, 56]: the quantity $J_{-\mathbf{k}\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3}$ in [54] corresponds to $4i\mathcal{H}^{-\sigma_2\sigma_3\sigma_4}_{\mathbf{k}\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3}$ because of the way their coefficients relate to the Hamiltonian coefficients. Therefore, a factor 16 appears making the two equations identical. The equation for the pdf can be written also as the following set of stochastic differential equations

$$ds_{\mathbf{k}} = (\eta_{\mathbf{k}} - \gamma_{\mathbf{k}} s_{\mathbf{k}})d\tau + \sqrt{2\eta_{\mathbf{k}} s_{\mathbf{k}}} dW_{\mathbf{k}}, \tag{4.23}$$

interpreted in the Ito sense and with self-consistent determination of $n(\mathbf{k},\tau)$. An important solution of (A.142) is the distribution

$$Q(s,\tau;\mathbf{k}) = \frac{1}{n(\mathbf{k},\tau)} e^{-s/n(\mathbf{k},\tau)}\ . \tag{4.24}$$

In absence of forcing and dissipation, an H-theorem and the *law of large-numbers* for the empirical spectrum imply that the solution relaxes to Q, for typical initial wavefields [69, 33]. It strictly describes thermodynamic equilibrium only when n is stationary, but our results show (see Fig.4.1) that P tends to the asymptotic state Q before n has reached its stationary state. This justifies that Q be called distribution of equilibrium despite its formal dependence on time. Furthermore, the results in Fig.4.2 suggest that relaxation to equilibrium also extends to forced and damped systems.

The general stationary solution to eq.(A.142) reads [39, 137]

$$P(s) = Ce^{-s/\nu} - \frac{F_*}{\eta_{\mathbf{k}}}\,\mathrm{Ei}\left(\frac{s}{\nu}\right)e^{-s/\nu} \tag{4.25}$$

where $\mathrm{Ei}(x)$ is the integral exponential function $\mathrm{Ei}(x) = -\int_{-x}^{\infty}\frac{\mathrm{e}^{-t}}{t}\,\mathrm{d}t$. Eq.(4.25) is obtained enforcing a constant probability flux in amplitude space: $F(s) = -s\left(\eta_{\mathbf{k}}\frac{\partial P}{\partial s} + \gamma_{\mathbf{k}}P\right) \equiv F_*$. For the positivity of $P(s)$ for $s \gg \nu$, F_* must be negative, corresponding to a probability flux from the large to the small amplitudes. This must be physically motivated by the existence of strong nonlinear interactions (e.g. breaking of wave crests) which feed probability into the weak, near-Gaussian background. In this picture, this happens at $s = s_{nl}$ and due to the strong nonlinear effects $P(s)$ decays very quickly for $s > s_{nl}$. Thus, the cut-off amplitude s_{nl} and the stationary flux F_* are two aspects of the same phenomenon, connected to each other through the boundary condition that comes out of (4.21) in a natural way:

$$P(s_{nl}) = -F_*/\eta_{\mathbf{k}}. \tag{4.26}$$

This is consistent with the fact that if the weak-turbulence assumption holds over the whole amplitude space, $s \in (0, \infty)$, the normalization of probability implies $F_* = 0$, and the equilibrium exponential distribution is recovered, as expected in absence of strong nonlinear effects that would affect the dynamics. So, clearly the picture with cut-off is meant to describe systems where forcing and damping are present at some wave numbers, which are necessary to sustain the strong nonlinear phenomena. Then, the corrective term in (4.25) represents the increased probability in the tail of the distribution due to such nonlinear phenomena ($\mathrm{Ei}(x) \propto \frac{1}{x}$ for $x \gg 1$).

Before numerically verifying this scenario, some remarks are in order. At variance with previous studies [39, 69], we do not need a probability sink to allow the solution, because we have $F(s) = F_*$ for $s \in (0, s_{nl})$ (similarly as in [137]). Integrating (A.142) from 0 to s_{nl}, $\frac{\partial}{\partial t}\int_0^{s_{nl}} dsP(s) = F(s = s_{nl}) - F(s = 0) = 0$, it is seen that the normalization of the probability in the system is preserved. This appears natural when considering the logarithmic variable $\sigma = \ln(s)$, whose probability density $\Pi(\sigma)$ satisfies

$$\partial_t \Pi = \partial_\sigma F, \tag{4.27}$$

with the same F of Eq. (A.142). Imposing $F(s = 0) = F_*$, as in the rest of the interval, just means that there is a probability flux from $\sigma_{nl} = \ln(s_{nl})$ toward $\sigma = -\infty$, with probability transferred to infinitesimally small amplitudes. In the stationary state, using (4.26) and normalizing the probability yields:

$$C = \frac{1}{\nu}\left(1 + \frac{\Gamma + \ln\frac{s_{nl}}{\nu} - e^{-\frac{s_{nl}}{\nu}}\,\mathrm{Ei}\left(\frac{s_{nl}}{\nu}\right)}{e^{\frac{s_{nl}}{\nu}} - \mathrm{Ei}\left(\frac{s_{nl}}{\nu}\right)}\right)^{-1}, \tag{4.28}$$

where $\Gamma \simeq 0.5772$ is the Euler-Mascheroni constant, and $P(s) = \frac{1}{\nu} e^{-s/\nu}$, in the $s_{nl} \to \infty$ limit. As s_{nl} becomes finite, the complete solution has to be chosen (with $F_* < 0$) and this contribution brings a correction to the asymptotic solution. In conclusion, given the cut-off value s_{nl}, which enters as a parameter of the model, and the spectrum $\nu = \eta/\gamma$ in the equilibrium limit, the two free constants in (4.25) are fixed and a unique general solution with cut-off is obtained.

4.2.3 Numerical verification of convergence to Gaussianity

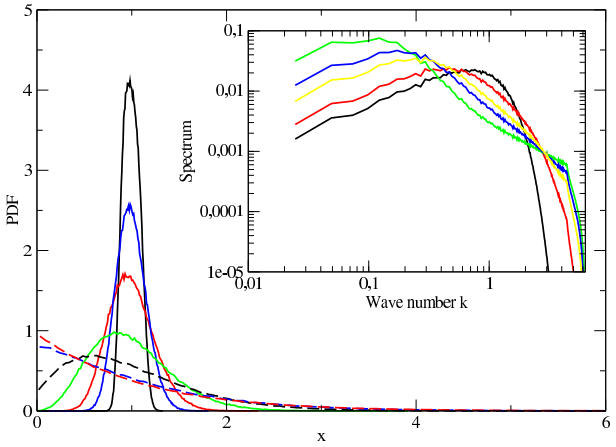

Figure 4.1. Normalized pdf of the modes $|\Psi_{\mathbf{k}}(\tau)|^2$ for $|\mathbf{k}| = 2$ as a function of the normalized quantity $x = |\Psi_k(\tau)|^2/n(k,\tau)$ where $n(k,\tau)$ is the mean value of $|\Psi_k(\tau)|^2$. The numerical simulation of the 2D NLSE is performed over a domain of size 256×256 using a regular square grid of mesh size $dx = 0.5$ so that 512×512 modes are simulated. The statistics and mean values are obtained both by an ensemble average over 128 realizations of the numerical simulation of the NLS equation starting at $\tau = 0$ with a Gaussian Fourier mode distribution with random phases, and using the isotropy of the fields allowing angular mean. The pdf are shown for $\tau = 0.01$, 0.03, 0.05, 0.1, 0.2, 0.5 and 1 time units respectively from top to bottom. The short time pdf are concentrated around the mean value while they converge at large time to the expected e^{-x} law (corresponding to the dashed red line, pdf for $\tau = 10$) and no more variations of the pdf are observed for $\tau > 10$. The inset shows the spectrum $n(k,\tau)$ for the times $\tau = 0.1$, 10, 30, 50 and 110, from bottom to top respectively looking at low k. The equipartition of energy spectrum $n(k,\tau) \propto 1/k^2$ is still not reached for the latest time shown here.

In order to validate these analytical predictions, we performed numerical simulations for two prototype equations of 4-wave turbulence. The first is the Nonlinear Schrödinger equation (NLSE) in two dimensions, modeling for instance the propagation of electromagnetic fields in optic fibers [58]:

$$i\partial_t \Psi = -\frac{1}{2}\Delta\Psi + |\Psi|^2\Psi, \tag{4.29}$$

where $\Delta = \partial_x^2 + \partial_y^2$ is the Laplacian operator and Ψ is a field taking complex values. The second is the Föppl Von-Karman equation in two space dimensions for the vibrations of elastic plates [115], which in dimensionless form reads:

$$\frac{\partial^2\zeta}{\partial t^2} = -\frac{1}{4}\Delta^2\zeta + \{\zeta,\chi\}; \tag{4.30}$$

$$\Delta^2\chi = -\frac{1}{2}\{\zeta,\zeta\}. \tag{4.31}$$

χ is the Airy stress function imposing the compatibility condition for the displacement field and the Poisson bracket $\{\cdot,\cdot\}$ is defined by $\{f,g\} \equiv f_{xx}g_{yy} + f_{yy}g_{xx} - 2f_{xy}g_{xy}$, so that $\{\zeta,\zeta\}$ is the Gaussian curvature.

The reason for investigating these two models is that they exhibit an important difference in the 4-wave interactions: while the NLSE only allows a two waves-two waves collision kernel, because of an additional conservation law, the FVK equation allows 1 wave-3 waves collisions as well. Both equations are solved in a periodic square domain using similar numerical schemes involving a pseudo-spectral method (see for instance [54] for details on the numerical methods). We first investigate the evolution of the fields starting with a Guassian distribution (consisting for NLSE of $|\psi(\mathbf{k},0)|^2 \propto e^{-k^2/k_0^2}$ with a random phase): the initial pdf of the amplitudes is given by $P(x) = \delta(x-1)$ for each mode, where $x = s/n(0)$ is the normalized amplitude. The evolution of the one mode pdf is shown in Fig.4.1 together with the time evolution of the density spectrum (inset). We can see that $P(x)$ converges rapidly to the exponential solution given by eq.(A.145), in agreement with the theory. Interestingly, the dynamics of the spectrum is different. The spectrum converges towards the equilibrium solution given by the Rayleigh-Jeans spectrum [206], but the characteristic time is much larger: the pdf has reached equilibrium when the spectrum is still far from it. That validates the theory and in particular it supports the RPA approximation, which appears to be verified from whatever initial conditions after extremely short times. The same dynamics was also observed for the elastic plate (not shown here). This evidence confirms the results already obtained for a general 3-waves system [182]. Then, we study the non-equilibrium wave turbulence energy cascade for the elastic plate dynamics obtained by injecting energy at large scale through a random noise in Fourier space at small k and a dissipation dominant at small scale. The balance between these two contributions leads to a stationary regime with a wave turbulence spectrum following roughly $|\zeta_k|^2 \sim k^{-4}$ at low

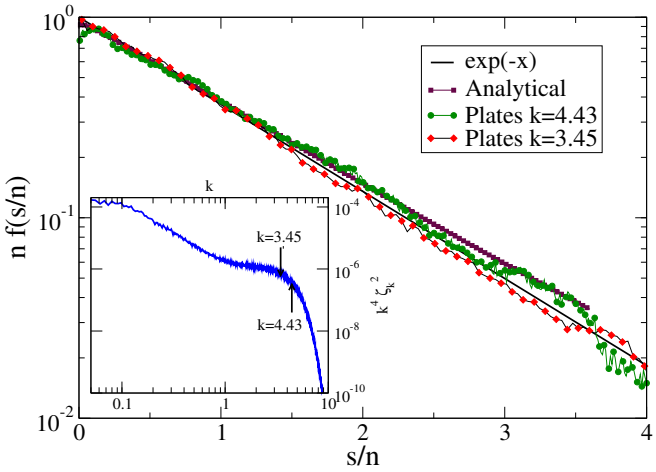

Figure 4.2. Normalized pdf of the Fourier modes $|\zeta_k|^2$ as a function of the rescaled parameter s/n for two different wave numbers $k = 3.45$ and $k = 4.43$, in a linear-log plot. The statistical average is made using angle average due to the isotropy of the system and time average, because of the statistically stationary regime reached in time. Here $dx = 0.25$ and the the square plate is $L \times L = 1024 \times 1024$, meaning that 4096×4096 modes are simulated. The pdf are reasonably well fitted by the equilibrium law e^{-x} although for $k = 4.43$ the generalized function (4.25) with the cut-off $s_{nl} = 3.6\, n(k)$ is a much better fit. The inset shows the compensated spectrum $k^4|\zeta_k|^2$ that exhibits a complex inertial regime, with a k^{-2} slope at large scale ($k \lesssim 1$) indicating intermittent behavior, and the expected weak turbulence spectrum $|\zeta_k|^2 \propto k^{-4}$ at smaller scales ($1 < k < 5$), where are located the two modes shown here [34]. The other modes pdf's show, outside of the forcing region ($k < 0.05$), the exponential Rayleigh distribution.

forcing (up to a logarithmic correction [54]) that corresponds to a constant flux of energy from the large to the small scales. It is thus tempting to compare the pdf of the Fourier modes of this dynamics with that of the Hamiltonian dynamics studied above, for which the theory has been derived. Indeed, no theoretical predictions can be easily made in such configuration, because the forcing-dissipation terms break the Hamiltonian structure. Moreover, while a distribution close to the one of the equilibrium situation could be expected at low forcing, intermittency at high forcing is supposed to heavily influence the pdf of the Fourier mode, similarly to what has been observed for the high moments of the structure function in real space [34]. Surprisingly, Fig. 4.2 shows that the pdf's are very close to the Rayleigh distribution predicted for the Hamitonian dynamics, in the absence of flux ($F_* = 0$) even at high forcing where the spectrum exhibits a k^{-6} slope at small k. However, a closer analysis shows a slight deviation from this distribution for modes at small scales, just before the dissipative range, where the pdf is better fitted by the generalized distribution (4.25) with $F_* \neq 0$. Similar results have also been observed for

the NLSE with no noticeable non-zero F_*. The weak value of F_* obtained for our systems suggests that while clear signature of intermittency is detected in physical space via structure functions [34], it is difficult to find anomalous scaling looking at the 1-mode spectral pdf.

4.2.4 Main results from wave turbulence

To conclude the section, we sum up the main results, that will be useful in the following section.

- Due to their fast linear rotation and to periodicity of their domain, the phases decorrelate extremely fast implying an effective *mixing* that physically justifies the RP (random-phase) assumption. An implication is that the RP assumption is likely bound to fail at low k's for systems with a dispersion relation $\omega(k)$ passing through the origin, the linear dynamics becoming very slow.

- The RP assumption is sufficient to derive a closed equation for the $M-$mode statistics, in the kinetic limit—large box and weak nonlinearity. This is in close analogy with the BBGKY hierarchy for the Boltzmann equation.

- The $M-$mode equation preserves a factorized form of the initial conditions. In the RPA assumption (uniform phases, uncorrelated amplitudes), the RPA property of the field is preserved in time and therefore one can derive the $1-$mode PDF equation.

- An $H-$theorem predicts convergence to Gaussianity of the $1-$mode PDF. From numerics, this is confirmed to happen very fast, even in the nonequilibrium case of a forced and damped dynamics.

- From the $1-$mode PDF equation we can easily derive the kinetic equation for the spectrum, which is consistent with the one typically found in WT and admits the Kolmogorov-Zakharov (KZ) spectra as its nonequilibrium steady states.

- Convergence to the KZ solutions is slow, with the nonlinear time scales of the system. Supported by numerics, we find that there is a clear scale separation between convergence to Gaussianity and convergence to the KZ spectrum. Indeed, this is consistent with the fact that the kinetic equation can be derived straigthforwardly assuming *quasi-Gaussianity* and RPA.

4.3 Large deviations and rogue waves

4.3.1 Rogue waves and nonlinear focusing effects

Rogue waves, long considered a figment of sailor's imagination, are now recognized to be a real, and serious, threat for boats and naval structures [136, 195]. Oceanographers define them as deep water waves whose crest-to-trough height H exceeds twice the significant wave height H_s, which itself is four times the standard deviation of the ocean surface elevation. Rogue waves appear suddenly and unpredictably, and can lead to water walls with vertical size on the order of 20–30 m [97, 144], with enormous destructive power. Although rare, they tend to occur more frequently than predicted by linear Gaussian theory [150, 138]. While the mechanisms underlying their appearance remain under debate [6, 5, 149], one plausible scenario has emerged over the years: it involves the phenomenon of modulational instability [13, 202], a nonlinear amplification mechanism by which many weakly interacting waves of regular size can create a much larger one. Such an instability arises in the context of the focusing nonlinear Schrödinger (NLS) equation [202, 114, 156, 7, 154, 207, 152] or its higher order variants [60, 178, 190, 46, 93], which are known to be good models for the evolution of a unidirectional, narrow-banded surface wave field in a deep sea. Support for the description of rogue waves through such envelope equations recently came from experiments in water tanks [145, 29, 30, 89], where Dysthe's MNLS equation in one spatial dimension [60, 178] was shown to accurately describe the mechanism creating coherent structures which soak up energy from its surroundings. While these experiments and other theoretical works [124, 43] give grounds for the use of MNLS to describe rogue waves, they have not addressed the question of their likelihood of appearance. Some progress in this direction has been recently made in [44], where a reduced model based on MNLS was used to estimate the probability of a given amplitude within a certain time, and thereby compute the tail of the surface height distribution. These calculations were done using an ansatz for the solutions of MNLS, effectively making the problem two-dimensional. The purpose of this chapter is to remove this approximation, and study the problem in its full generality. Specifically, we consider the MNLS with random initial data drawn from a Gaussian distribution [137]. The spectrum of this field is chosen to have a width comparable to that of the JONSWAP spectrum [96, 146] obtained from observations in the North Sea. We calculate the probability of occurrence of a large amplitude solution of MNLS out of these random initial data, and thereby also estimate the tail of the surface height distribution. These calculations are performed within the framework of large deviations theory (LDT), which predicts the most likely way by which large disturbances arise and therefore also explains the mechanism of rogue wave creation. Our results are validated by comparison with brute-force Monte-Carlo simulations, which indicate that rogue waves in MNLS are indeed within the realm

of LDT. Our approach therefore gives an efficient way to assess the probability of large waves and their mechanism of creation.

4.3.2 LDT method for MNLS with random initial conditions

Our starting point will be the MNLS equation for the evolution of the complex envelope $u(t,x)$ of the sea surface in deep water [60], in terms of which the surface elevation reads $\eta(t,x) = \Re\Big(u(t,x)e^{i(k_0x-\omega_0t)}\Big)$ (here k_0 denotes the carrier wave number, $\omega_0 = \sqrt{gk_0}$, and g is the gravitational acceleration). Measuring u and x in units of k_0^{-1} and t in ω_0^{-1} we can write MNLS in non-dimensional form as

$$\begin{aligned} &\partial_t u + \tfrac{1}{2}\partial_x u + \tfrac{i}{8}\partial_x^2 u - \tfrac{1}{16}\partial_x^3 u + \tfrac{i}{2}|u|^2 u \\ &+ \tfrac{3}{2}|u|^2\partial_x u + \tfrac{1}{4}u^2\partial_x\bar{u} - \tfrac{i}{2}\,|\partial_x|\,|u|^2 = 0, \quad x\in[0,L], \end{aligned} \tag{4.32}$$

where the bar denotes complex conjugation. We will consider Eq. 4.32 with random initial condition $u_0(x) \equiv u(0,x)$, constructed via their Fourier representation,

$$u_0(x) = \sum_{n\in\mathbb{Z}} e^{ik_nx}(2\hat{C}_n)^{1/2}\theta_n, \quad \hat{C}_n = Ae^{-k_n^2/(2\Delta^2)}, \tag{4.33}$$

where $k_n = 2\pi n/L$, θ_n are complex Gaussian variables with mean zero and covariance $\mathbb{E}\theta_n\bar{\theta}_m = \delta_{m,n}$, $\mathbb{E}\theta_n\theta_m = \mathbb{E}\bar{\theta}_n\bar{\theta}_m = 0$. This guarantees that $u_0(x)$ is a Gaussian field with mean zero and $\mathbb{E}(u_0(x)\bar{u}_0(x')) = 2\sum_{n\in\mathbb{Z}} e^{ik_n(x-x')}\hat{C}_n$.To make contact with the observational data, the amplitude A and the width Δ in Eq. 4.33 are picked so that $\hat{C}_n$ has the same height and area as the JONSWAP spectrum [96, 146].

Because the initial data for Eq. 4.32 are random, so is the solution at time $t > 0$, and our aim is to compute

$$P_T(z) \equiv \mathbb{P}\Big(F(u(T)) \geq z\Big), \tag{4.34}$$

where $\mathbb{P}$ denotes probability over the initial data and F is a scalar functional depending on u at time $T > 0$. Even though our method is applicable to more general observables, here we will focus on

$$F(u(T)) = \max_{x\in[0,L]} |u(T,x)|. \tag{4.35}$$

A brute force approach to calculate Eq. 4.34 is Monte-Carlo sampling: Generate random initial conditions $u_0(x)$ by picking random θ_n's in Eq. 4.33, evolve each of these $u_0(x)$ deterministically via Eq. 4.32 up to time $t = T$ to get $u(T,x)$, and count the proportion that fulfill $F(u(T)) \geq z$. While this method is simple, and will be used below as benchmark, it looses efficiency when z is large, which is precisely the regime of interest for the tails of the distribution of $F(u(T))$.

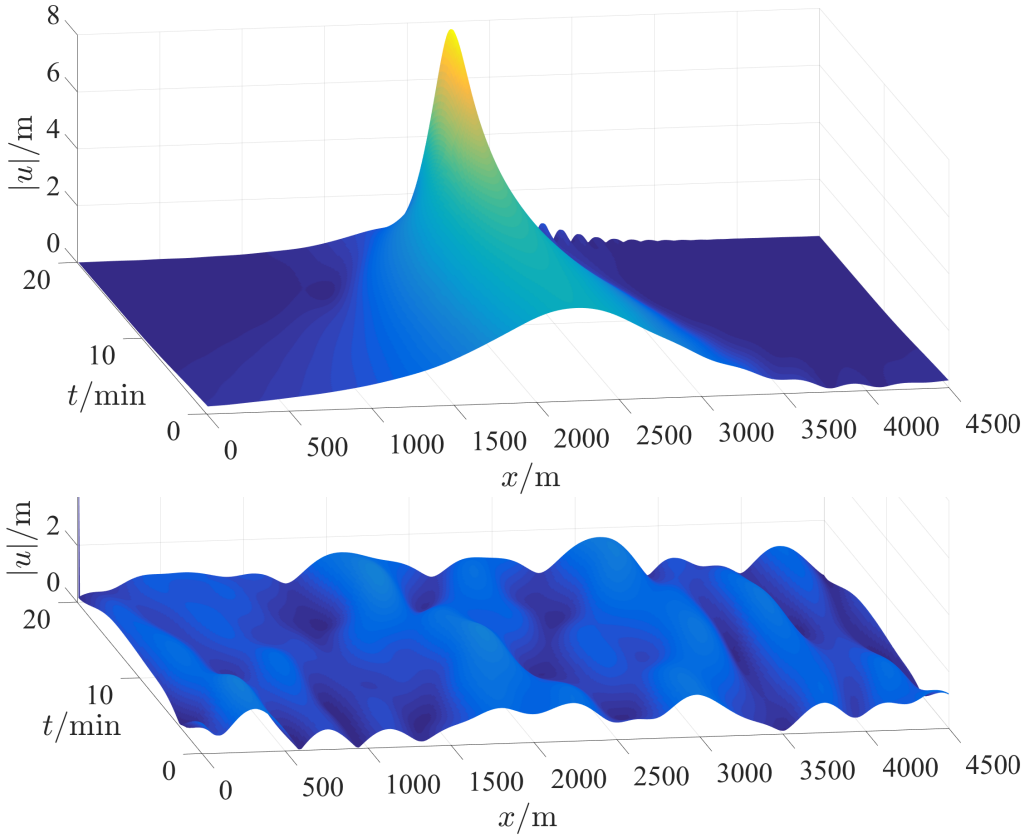

Figure 4.3. *Top:* Time evolution of $|u(t,x)|$ from an initial condition optimized for $\max_x |u(T,x)| \geq 8$ m at $T = 20$ min. *Bottom:* Same for a typical Gaussian random initial condition.

4.3.3 The LDT approach

In that regime, a more efficient approach is to rely on results from LDT which assert that Eq. 4.34 can be estimated by identifying the most likely initial condition that is consistent with $F(u(T)) \geq z$. To see how this result comes about, recall that the probability density of u_0 is formally proportional to $\exp(-\frac{1}{2}\|u_0\|_C^2)$, where $\|u_0\|_C^2$ is given by

$$\|u_0\|_C^2 = \sum_{n\in\mathbb{Z}} \frac{|\hat{a}_n|^2}{\hat{C}_n}, \quad \hat{a}_n = \frac{1}{L}\int_0^L e^{-ik_n x} u_0(x)dx\,. \tag{4.36}$$

To calculate Eq. 4.34 we should integrate this density over the set $\Omega(z) = \{u_0 : F(u(T,u_0)) \geq z\}$, which is hard to do in practice. Instead we can estimate the integral by Laplace's method. As shown in *Material and Methods*, this is justified for large z, when the probability of the set $\Omega(z)$ is dominated by a single $u_0(x)$ that contributes most to the integral and can be identified via the constrained

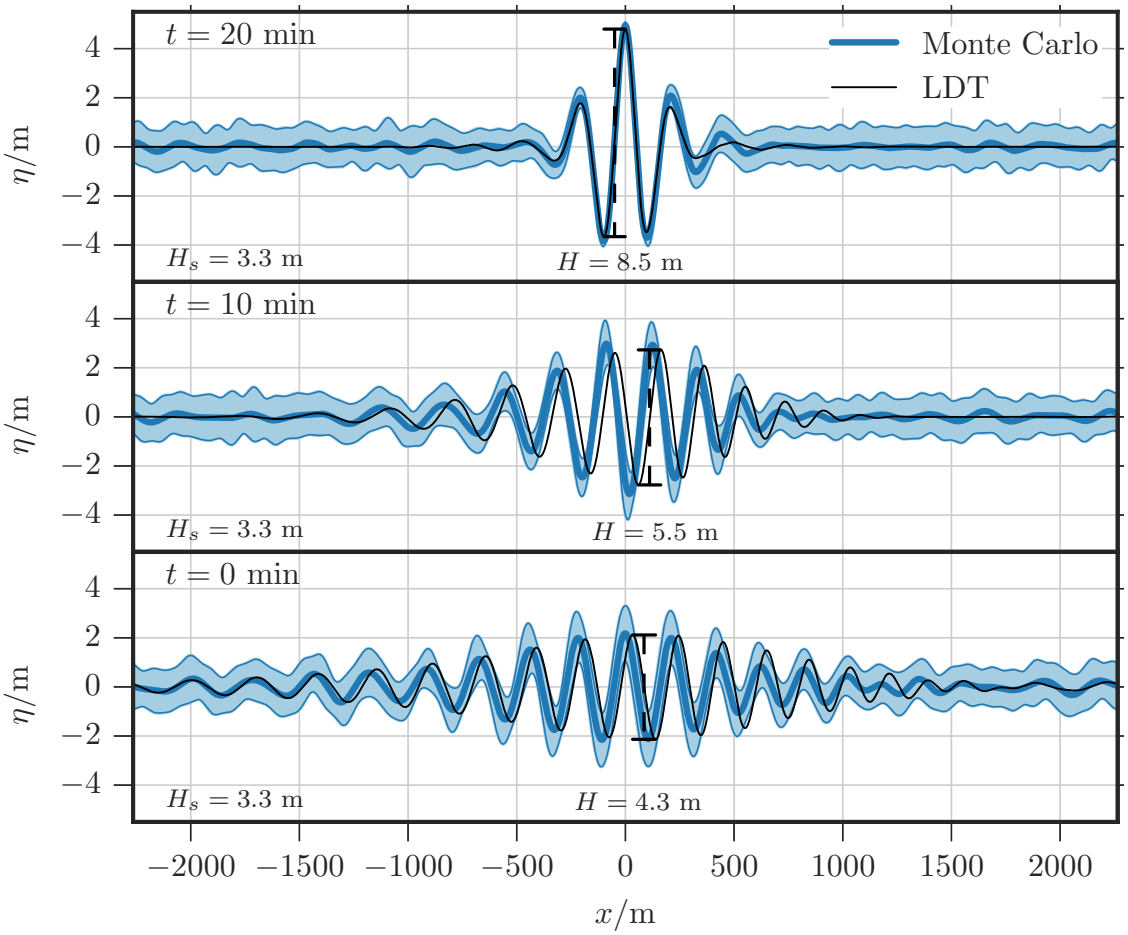

Figure 4.4. Comparison between the average realization reaching $\max_x |u(T,x)| \geq 4.8$ m at $T = 20$ min (dark blue) and one standard deviation around this mean (light blue), with the solution reaching the same amplitude starting from the maximum likelihood initial condition (black) for $t = 0$,10,20 min.

minimization problem

$$\tfrac{1}{2} \min_{u_0 \in \Omega(z)} \|u_0\|_C^2 \equiv I_T(z)\,, \tag{4.37}$$

which then yields the following LDT estimate for Eq. 4.34

$$P_T(z) \asymp \exp\left(-I_T(z)\right)\,. \tag{4.38}$$

Here $\asymp$ means that the ratio of the logarithms of both sides tends to 1 as $z \to \infty$. As discussed in *Material and Methods*, a multiplication prefactor can be added to (4.38) but it does not affect significantly the tail of $P_T(z)$ on a logarithmic scale.

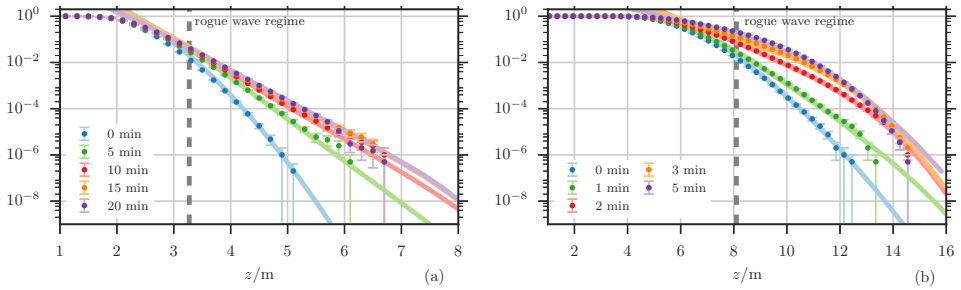

Figure 4.5. Probabilities $P_T(z)$ of $\max_x |u(T,x)|$ for different times T for Set 1 (a) and Set 2 (b). The probabilities estimated by Monte-Carlo sampling with 10^6 realizations (dots) are compared with those predicted by LDT (lines). These probabilities agree over about 5 orders of magnitude in probability, though LDT allows for the calculation of the tail where Monte-Carlo becomes prohibitively costly. The error bars in the Monte-Carlo results represent the statistical error of 2 standard deviations (95% confidence interval) for the Bernoulli distribution with parameter $P_T(z)$.

In practice, the constraint $F(u(T,u_0)) \geq z$ can be imposed by adding a Lagrange multiplier term to Eq. 4.37, and it is easier to use this multiplier as control parameter and simply see *a posteriori* what value of z it implies. That is to say, perform for various values of λ the minimization

$$\min_{u_0} \left(\tfrac{1}{2}\|u_0\|_C^2 - \lambda F(u(T,u_0)) \right) \equiv S_T(\lambda)\,, \tag{4.39}$$

over all u_0 of the form in Eq. 4.33 (no constraint), then observe that this implies the parametric representation

$$I_T(z(\lambda)) = \tfrac{1}{2}\|u_0^\star(\lambda)\|_C^2, \quad z(\lambda) = F(u(T,u_0^\star(\lambda)))\,. \tag{4.40}$$

where $u_0^\star(\lambda)$ denotes the minimizer obtained in Eq. 4.39. It is easy to see from Eqs. 4.37 and 4.39 that $S_T(\lambda)$ is the Legendre transform of $I_T(z)$ since:

$$S_T(\lambda) = \sup_{z\in\mathbb{R}}(\lambda z - I_T(z)) = \sup_{z\in\mathbb{R}}(\lambda z - \tfrac{1}{2}\inf_{u_0\in\Omega(z)}\|u_0\|_C^2), \tag{4.41}$$

4.3.4 Numerical results

We considered two sets of parameters. In Set 1 we took $A = 5.4\cdot 10^{-5}k_0^{-2}$ and $\Delta = 0.19k_0$. Converting back into dimensional units using $k_0^{-1} = 36$ m consistent with the JONSWAP spectrum [96, 146], this implies a significant wave height $H_s = 4\sqrt{C(0)} = 3.3$ m classified as a *rough sea* [153]. It also yields a Benjamin-Feir index $\mathrm{BFI} = 2\sqrt{2C(0)}/\Delta = 0.34$, [146, 104], meaning that the modulational instability of

a typical initial condition is of medium intensity. In Set 2 we took $A = 3.4 \cdot 10^{-4} k_0^{-2}$ and $\Delta = 0.19 k_0$, for which $H_s = 8.2$ m is that of a *high sea* and the BFI is 0.85, meaning that the modulational instability of a typical initial condition is stronger.

Fig. 4.3 (top) shows the time evolution of $|u(t,x)|$ starting from an initial condition from Set 1 optimized so that $\max_x |u(T,x)| = 8$ m at $T = 20$ min. For comparison, Fig. 4.3 (bottom) shows $|u(t,x)|$ for a typical initial condition drawn from its Gaussian distribution. To illustrate what is special about the initial conditions identified by our optimization procedure, in Fig. 4.4 we show snapshots of the surface elevation $\eta(t,x)$ at three different times, $t = 0$,10,20 min (black lines), using the constraint that $\max_x |u(T,x)| \geq 4.8$ m at $T = 20$ min. Additionally, we averaged all Monte-Carlo samples achieving $\max_x |u(t,x)| \geq 4.8$ m, translated to the origin. Snapshots of this mean configuration are shown in Fig. 4.4 (blue lines). They agree well with those of the optimized solution (black lines). The one standard deviation spread around the mean Monte-Carlo realization (light blue) is reasonably small, especially around the rogue wave at final time. This indicates that the event $\max_x |u(T,x)| \geq 4.8$ m is indeed realized with probability close to 1 by starting from the most likely initial condition consistent with this event, as predicted by LDT. The usefulness of LDT is confirmed in Figs. 4.5 (a,b) depicting the probabilities of $\max_x |u(T,x)|$ for both Sets 1 and 2 calculated via LDT optimization (lines), compared to Monte-Carlo sampling (dots). As can be seen, the agreement is remarkable, especially in the tail corresponding to the rogue wave regime. As expected, the Monte-Carlo sampling becomes inaccurate in the tail, since there the probabilities are dominated by unlikely events. The LDT calculation, in contrast, remains efficient and accurate far in the tail.

The probabilities plotted in Fig. 4.5 (a,b) show several remarkable features. First, they indicate that, as T gets larger, their tails fatten significantly. For example, in Set 1 $P_{T=20\,\mathrm{min}}(6\,\mathrm{m}) \approx 10^{-5}$, which is 5 orders of magnitude larger than initially, $P_{t=0\,\mathrm{min}}(6\,\mathrm{m}) \approx 10^{-10}$. Secondly, the probabilities converge to a limiting density for large T. This occurs after some decorrelation time $\tau_c \approx 10$ min in Set 1 and $\tau_c \approx 3$ min in Set 2. Similarly, the LDT results converge. In fact, this convergence can be observed at the level of the trajectories generated from the optimal $u_0^\star$. As Fig. 4.6 shows, reading these trajectories backward from $t = T$, their end portions coincide, regardless on whether $T = 20$ min, $T = 15$ min, or $T = 10$ min. The implications of these observations, in particular on the mechanism of creation of rogue waves and their probability of appearance within a time window, will be discussed in *Interpretation* below.

Before doing so, let us discuss the scalability of our results to larger domain sizes, referring the reader to the Appendix B for more details. As shown above, the optimization procedure based on large deviation theory predicts that the most likely way a rogue wave will occur in the domain is via the onset of a single large peak in $|u(t,x)|$. In the set-up considered before, this prediction is confirmed by the brute-force simulations using Monte-Carlo sampling. It is clear, however, that

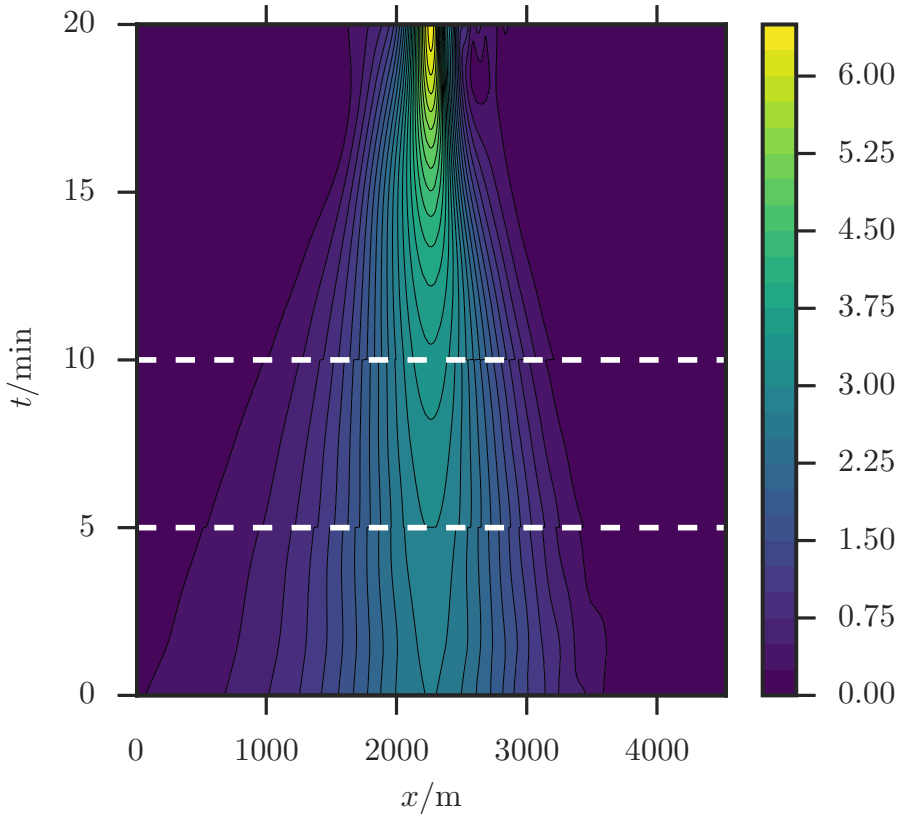

Figure 4.6. Contourplot of the optimal trajectories from LDT for $T = 10$, 15, and 20 min in Set 1. The trajectories, superposed to match at $t = T$, coincide, which is consistent with the convergence of the probabilities $P_T(z)$ for large T.

for increased domain size, e.g. by taking a domain size of NL with $N \gg 1$, it will become increasingly likely to observe multiple peaks, for the simple reason that large waves can occur independently at multiple sufficiently separated locations. In these large domains, the large deviation predictions remain valid if we look at the maximum of $|u(t,x)|$ in observation windows that are not too large (that is, about the size of the domain L considered above). However, they deteriorate if we consider this maximum in the entire domain of size NL, in the sense that the value $\mathbb{P}\left(\max_{x\in[0,NL]} |u(t,x)| \geq z\right)$ predicted by large deviation theory matches that from Monte-Carlo sampling at values of z that are pushed further away in the tails. This is an entropic effect, which is easy to correct for: events in different subwindows must be considered independent, and their probabilities superposed. That is, if we denote by

$$P_T^N(z) = \mathbb{P}\Big(\max_{x\in[0,NL]} |u(T,x)| \geq z\Big), \tag{4.42}$$

it can be related to $P_T(z) = \mathbb{P}(\max_{x\in[0,L]} |u(T,x)| \geq z)$ via

$$P_T^N(z) = 1 - (1 - P_T(z))^N . \tag{4.43}$$

This formula is derived in Appendix B and shown to accurately explain the numerical results. In effect, this provides us with a method to scale up our results to arbitrary large observation windows.

Details on LDT and Laplace's method

For the reader's convenience, here we recall some standard large deviations results that rely on the evaluation of Gaussian integrals by Laplace's method and are at the core of the method we propose. All of the following is justified rigorously by the results of chapter 2. It is convenient to rephrase the problem abstractly and consider the estimation of

$$P(z) = \mathbb{P}(\phi(\theta) > z) , \tag{4.44}$$

where $\theta \in \mathbb{R}^D$ are Gaussian random variables with mean zero and covariance Id, and $\phi : \mathbb{R}^D \to \mathbb{R}$ is some real valued function – as long as we truncate the sum in Eq. 4.33 to a finite number of modes, $|n| \leq M$, the problem treated in this chapter can be cast in this way, with θ playing the role of $C^{-1/2}u_0$ and $\phi(\cdot)$ that of $F(u(T, C^{-1/2}\cdot))$. The probability $P(z)$ in Eq. 4.44 is given by

$$P(z) = (2\pi)^{-D/2} \int_{\Omega(z)} e^{-\frac{1}{2}|\theta|^2} d\theta , \tag{4.45}$$

where $\Omega(z) = \{\theta : \phi(\theta) > z\}$. The interesting case is when this set does not contain the origin, $0 \notin \Omega(z)$, which we will assume is true when $z > 0$. We also make two additional assumptions:

1. The point on the boundary $\partial\Omega(z)$ that is closest to the origin is isolated: Denoting this point as

$$\theta^\star(z) = \operatorname*{argmin}_{\theta\in\partial\Omega(z)} |\theta|^2 , \tag{4.46}$$

we assume that

$$\begin{aligned} &\tfrac{1}{2}|\theta^\star(z)|^2 \text{ is strictly increasing with } z \geq 0 ; \\ &\lim_{z\to\infty} \tfrac{1}{2}|\theta^\star(z)|^2 = \infty . \end{aligned} \tag{4.47}$$

2. The connected piece of $\partial\Omega(z)$ that contains $\theta^\star(z)$ is smooth with a curvature that is bounded by a constant independent of z.

The point $\theta^\star(z)$ satisfies the Euler-Lagrange equation for Eq. 4.46, with the constraint incorporated via a Lagrange multiplier term:

$$\theta^\star(z) = \lambda \nabla\phi(\theta^\star(z)) \tag{4.48}$$

for some Lagrange multiplier λ. This implies that

$$\frac{\theta^\star(z)}{|\theta^\star(z)|} = \frac{\nabla\phi(\theta^\star(z))}{|\nabla\phi(\theta^\star(z))|} = \hat{n}(z)\,. \tag{4.49}$$

where $\hat{n}(z)$ denotes the inward pointing unit vector normal to $\partial\Omega(z)$ at $\theta^\star(z)$. If we move inside the set $\Omega(z)$ from $\theta^\star(z)$ in the direction of $\hat{n}(z)$, the norm $|\theta|^2$ increases under the assumptions in Eq. 4.47. Indeed, setting $\theta = \theta^\star(z) + \hat{n}(z)u$ with $u \geq 0$, we have

$$\begin{aligned} |\theta|^2 &= |\theta^\star(z)|^2 + 2\langle \hat{n}(z), \theta^\star(z)\rangle u + u^2 \\ &= |\theta^\star(z)|^2 + 2|\theta^\star(z)|z + z^2\,, \end{aligned} \tag{4.50}$$

where we used Eq. 4.49. In fact, if we were to perform the integral in that direction, the natural variable of integration would be to rescale $u \to u/|\theta^\star(z)|$. In particular, if we were to replace $\Omega(z)$ by the half space $P(z) = \{\theta \,|\, \hat{n}(z)\cdot(\theta - \theta^*(z)) > 0\}$, it would be easy to estimate the integral in Eq. 4.45 by introducing a local coordinate system around $\theta^*(z)$, whose first coordinate is in the direction of $\hat{n}(z)$. Indeed this would give:

$$\begin{aligned} &(2\pi)^{-D/2}\int_{P(z)} e^{-\frac{1}{2}|\theta|^2} d\theta \\ &= (2\pi)^{-D/2}\int_0^\infty e^{-\frac{1}{2}|\theta^\star(z)|^2 - |\theta^\star(z)|u - \frac{1}{2}u^2} du \int_{\mathbb{R}^{N-1}} e^{-\frac{1}{2}|\eta|^2} d\eta \\ &= (2\pi)^{-1/2} e^{-\frac{1}{2}|\theta^\star(z)|^2}\int_0^\infty e^{-|\theta^\star(z)|u - \frac{1}{2}u^2} du \\ &= (2\pi)^{-1/2}|\theta^\star(z)|^{-1} e^{-\frac{1}{2}|\theta^\star(z)|^2}\int_0^\infty e^{-v - \frac{1}{2}|\theta^\star(z)|^{-2}v^2} dv \\ &\sim (2\pi)^{-1/2}|\theta^\star(z)|^{-1} e^{-\frac{1}{2}|\theta^\star(z)|^2} \qquad \text{as } z \to \infty. \end{aligned} \tag{4.51}$$

The last approximation goes beyond a large deviations estimate (i.e. it includes the prefactor), and it implies

$$\lim_{z\to\infty} |\theta^\star(z)|^{-2} \log\left((2\pi)^{-D/2}\int_{P(z)} e^{-\frac{1}{2}|\theta|^2} d\theta\right) = -\frac{1}{2}\,. \tag{4.52}$$

This log-asymptotic estimate is often written as

$$\int_{P(z)} e^{-\frac{1}{2}|\theta|^2} d\theta \asymp e^{-\frac{1}{2}|\theta^\star(z)|^2} \qquad \text{as } z \to \infty. \tag{4.53}$$

Interestingly, while the asymptotic estimate in Eq. 4.51 does not necessarily apply to the original integral in Eq. 4.45 (that is, the prefactor may take different forms depending on the shape of $\partial\Omega(z)$ near $\theta^\star(z)$), the rougher log-asymptotic estimate in Eq. 4.53 does as long as the the boundary $\partial\Omega(z)$ is smooth, with a curvature that is bounded by a constant independent of z. This is because because the

contribution (positive or negative) to the integral over the region between $\Omega(z)$ and $P(z)$ is subdominant in that case, in the sense that the log of its amplitude is dominated by $|\theta^\star(z)|$. This is the essence of the large deviations result that we apply in this thesis.

Numerical details

To perform the calculations, we solved Eq. 4.32 with $L = 40\pi$ and periodic boundary conditions, and checked that this domain is large enough to make the effect of these boundary conditions negligible (see Appendix B). The spatial domain was discretized using 2^{12} equidistant gridpoints, which is enough to resolve the solution of Eq. 4.32. To evolve the field $u(t, x)$ in time we used a pseudo-spectral second order exponential time-differencing (ETD2RK) method [45, 107].

When performing the Monte-Carlo simulations, we used 10^6 realizations of the random initial data constructed by truncating the sum in Eq. 4.33 over the $M = 23$ modes with $-11 \leq n \leq 11$, i.e $-3\Delta \leq k_n \leq 3\Delta$: these modes carry most of the variance, and we checked that adding more modes to the initial condition did not affect the results in any significant way (see Appendix B).

Details on the optimization procedure

As explained above, the large deviation rate function $I_T(z)$ in Eq. 4.37 can be evaluated by solving the dual optimization problem in Eq. 4.39, which we rewrite as $S_T(\lambda) = \inf_{u_0} E(u_0, \lambda)$, where we defined the cost function

$$E(u_0, \lambda) \equiv \tfrac{1}{2}\|u_0\|_C^2 - \lambda F(u(T, u_0))\,. \tag{4.54}$$

We performed this minimization using steepest descent with adaptive step (line-search) and preconditioning of the gradient [21]. This involves evaluating the (functional) gradient of $E_T(u_0, \lambda)$ with respect to u_0. Using the chain rule, this gradient can be expressed as (using compact vectorial notation)

$$\frac{\delta E}{\delta u_0} = C^{-1}u_0 - \lambda J^T(T, u_0)\frac{\delta F}{\delta u} \tag{4.55}$$

where $J(t, u_0) = \delta u(t, u_0)/\delta u_0$ is the Jacobian of the transformation $u_0 \to u(t, u_0)$. Collecting all terms on the right-hand-side of the MNLS Eq. 4.32 into $b(u)$, this equation can be written as

$$\partial_t u = b(u), \quad u(t = 0) = u_0\,, \tag{4.56}$$

and it is easy to see that in this notation $J(t, u_0)$ satisfies

$$\partial_t J = \frac{\delta b}{\delta u} J, \qquad J(t = 0) = \mathrm{Id}. \tag{4.57}$$

Consistent with what was done in the Monte-Carlo sampling, to get the results presented above we truncated the initial data u_0 over $M = 23$ modes using the representation

$$u_0(x) = \sum_{n=-11}^{11} e^{ik_n x}\hat{a}_n, \quad k_n = 2\pi n/L. \tag{4.58}$$

This means that minimization of Eq. 4.54 was performed in the $2M - 1 = 45$ dimensional space spanned by the modes $\hat{a}_n$, accounting for invariance by an overall phase shift – to check convergence we also repeated this calculation using larger values of M and found no noticeable difference in the results (see Appendix B).

In practice, the evaluation of the gradient in Eq. 4.55 was performed by integrating both $u(t)$ and $J(t)$ up to time $t = T$. Eq. 4.57 was integrated using the same pseudo-spectral method as for Eq. 4.32 on the same grid. To perform the steepest descent step, we then preconditioned the gradient through scalar multiplication by the step-independent, diagonal metric with the components of the spectrum $\hat{C}_n$ as diagonal elements.

4.3.5 Interpretation: combining the long-timescale WT predictions and the fast nonlinear effects

The convergence of $P_T(z)$ towards a limiting function $P(z)$ has important consequences for the significance and interpretation of our method and its results. Notice first that this convergence can be explained if we assume that the probability distribution of the solutions to Eq. 4.32 with Gaussian initial data converges to an invariant measure. In this case, for large T, the Monte-Carlo simulations will sample the value of $\max_x |u|$ on this invariant measure, and the optimization procedure based on LDT will do the same. The timescale τ_c over which convergence occurs depends on how far this invariant measure is from the initial Gaussian measure of $u_0(x)$. Interestingly the values we observe for τ_c are in rough agreement with the timescales predicted by the theory of high-power pulse propagation [16, 185]. As recalled in the Appendix B, this approach predicts that the timescale of apparition of a focusing solution starting from a large initial pulse of maximal amplitude U_i and length-scale L_i is $\tau_c = \sqrt{T_{\rm nl}T_{\rm lin}}$, where $T_{\rm nl} = \left(\frac{1}{2}\omega_0 k_0^2 U_i^2\right)^{-1}$ is the nonlinear timescale for modulational instability and $T_{\rm lin} = 8\omega_0^{-1}k_0^2L_i^2$ is the linear timescale associated to group dispersion. Setting $U_i = H_s$ (the size at the onset of rogue waves) and $L_i = \sqrt{2\pi}\Delta^{-1}$ (the correlation length of the initial field) gives $\tau_c \simeq 18$ min for Set 1 and $\tau_c \simeq 8$ min for Set 2, consistent with the convergence times of $P_T(z)$. This observation has implications in terms of the mechanism of apparition of rogue waves, in particular their connection to the so-called Peregrine soliton, that has been invoked as prototype mechanism for rogue waves creation [156, 3, 174, 4, 150, 186], in particular for water waves [29, 30, 28], plasmas [10] and fiber optics [109, 180, 185]. This connection is discussed in Appendix B.

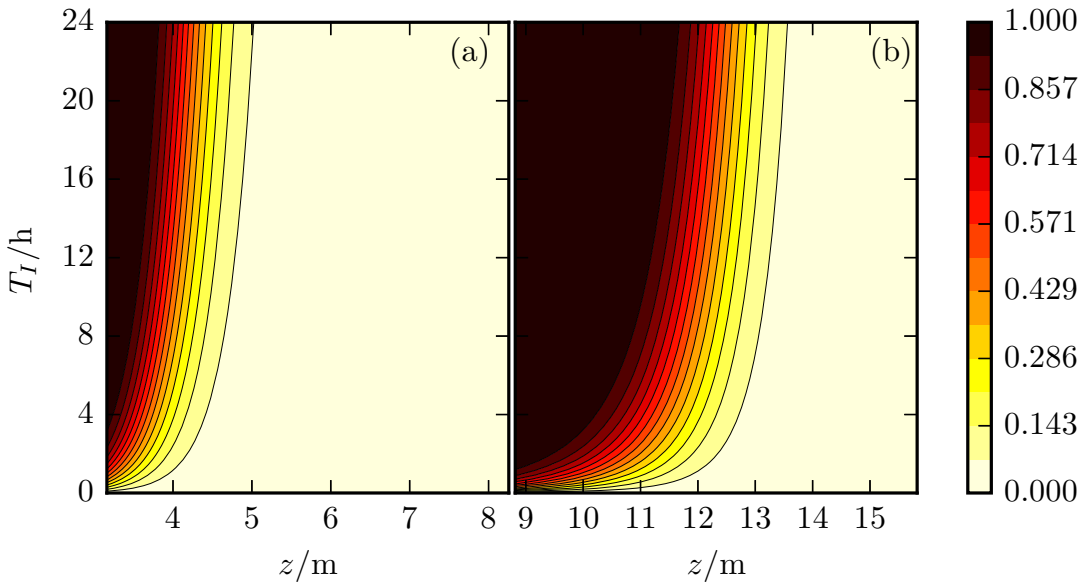

Figure 4.7. Contourplots of the probability to observe a wave whose amplitude exceeds z in the time window $[0, T_I]$ for Sets 1 (a) and 2 (b).

Our findings also indicate that, even though the assumption that $u_0(x)$ is Gaussian is incorrect in the tail (that is, $P_{T=0}(z)$ is not equal to the limiting $P(z)$ in the tail), it contains the right seeds to estimate $P(z)$ via $P_T(z)$ if $T \gtrsim \tau_c$[3]. Altogether this is consistent with the scenario put forward by Sapsis and collaborators in [133, 77] to explain how extreme events arise in intermittent dynamical systems and calculate their probability: they occur when the system hits small instability pockets which trigger a large transient excursion. In this scenario, as long as the initial probability distribution in these pockets is accurate, the dynamics will permit precise estimation of the distribution tail. In some sense, the distribution of the initial condition plays a role of the prior distribution in Bayesian inference[4], and the posterior can be effectively sampled by adding the additional information from the dynamics over short periods of time during which instabilities can occur. In [133], this picture was made predictive by using a two-dimensional ansatz for the initial condition $u_0(x)$ to avoid having to perform sampling in high-dimension over the original $u_0(x)$. What our results show is that this approximation can be

[3]This convergence occurs on the timescale τ_c which is much smaller than the mixing time for the solutions of Eq. 4.32, i.e. the time it would take from a given initial condition, rather than an ensemble thereof, to sample the invariant measure.

[4]Note in particular that the Gaussian field in Eq. 4.33 is the random field that maximizes entropy given the constraint on its covariance $C(x)$.

avoided altogether by using LDT to perform the calculations directly with the full Gaussian initial condition in Eq. 4.33.

Interestingly, we can use the results above to calculate the probability of occurrence of rogue waves in a given time window. More precisely, the probability $p(z, T_I)$ that a rogue wave of amplitude larger than z be observed in the domain $[0, L]$ during $[0, T_I]$ (i.e. that $\max_{t\in[0,T_I]} \max_{x\in[0,L]} |u(t,x)| \geq z$) can be estimated in terms of $P(z)$ and τ_c as

$$p \equiv \mathbb{P}\Big(\max_{t\in[0,T_I]} \max_{x\in[0,L]} |u(t,x)| \geq z \Big) \sim 1 - (1 - P(z))^{T_I/\tau_c}\,, \tag{4.59}$$

where we used the fact that rogue waves can be considered independent on timescales larger than τ_c and assumed $T_I \gg \tau_c$. The function p is plotted in Fig. 4.7 as a function of z and T_I. For example for Set 1, Eq. 6.7 indicates a 50% chance to observe a rogue wave of height $z = 4$ m (that is, about 8 m from crest-to-trough) after 11 hours (using $\tau_c = 10$ min and $P(z = 4\text{ m}) = 1.1 \cdot 10^{-2}$); if we wait 30 hours, the chance goes up to 85%. Similarly, for Set 2 the chance to observe a wave of 11 m height is about 50% after 3 hours and about 85% after 8 hours ($\tau_c = 3$ min and $P(z = 11\text{ m}) = 1.2 \cdot 10^{-2}$).

4.4 Experimental evidence of instantons

In the following section we apply the same method to an experiment in a real water tank. Conceptually, the setting is in analogy with the previous section, though the evolution is in space rather than time, and for semplicity we use the NLS equation (4.14) without the higher order terms of the Dysthe equation.

4.4.1 Experimental setup and analysis

The experimental data were recorded in the 270m long wave flume at Marintek (Norway) [147, 148], represented in Fig. 4.8a. On one end of the tank a plane-wave generator is able to perturb the water surface with a predefined random signal. The perturbations create long-crested wave trains that propagate along the tank toward the opposite end, where they eventually break on a smooth beach that suppresses reflections. The water surface $\eta(x, t)$ is measured by probes placed at different distances from the wave maker (x-coordinate). The signal at the wave maker $\eta(x = 0, t)$ is prepared according to a stationary random-phase statistics with amplitudes given by the JONSWAP spectrum [96], designed to replicate energy spectra of deep water waves in the ocean. The statistics of JONSWAP depend on two parameters H_s and γ, quantifying the significant wave height and the enhancement factor, respectively. Experimental data are collected for three different regimes: *quasi linear* ($\gamma = 1$, $H_s = 0.11$ m), *intermediate* ($\gamma = 3.3$, $H_s = 0.13$ m), and *highly nonlinear* ($\gamma = 6$, $H_s = 0.15$ m).

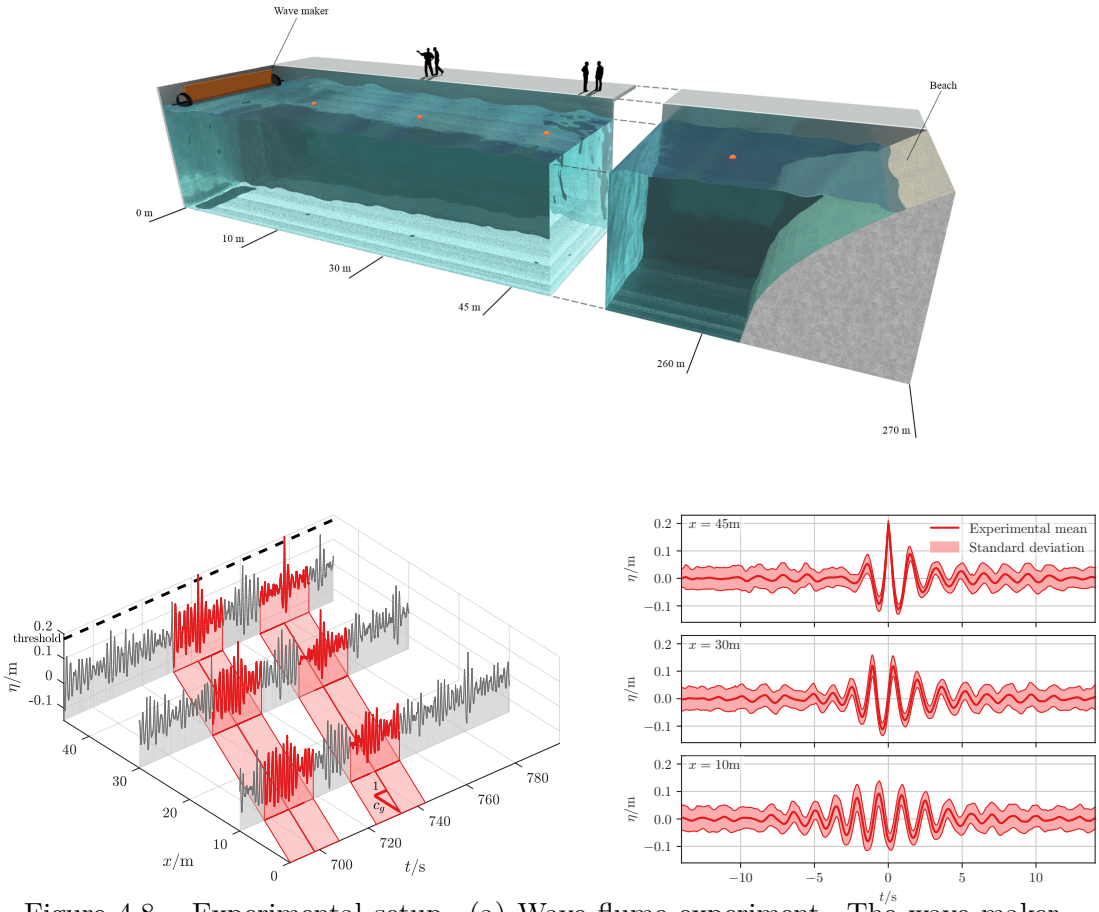

Figure 4.8. Experimental setup. (a) Wave flume experiment. The wave maker generates a random wave field with stationary Gaussian statistics. The planar wave fronts propagate along the water tank, where the surface elevation η is measured by vertical probes. (b) Extreme wave event selection. At $x = 45$ m, we pick a temporal maximum of the experimental data series of η which exceeds a large threshold z. We track the event evolution at probes located earlier in the channel, within an observation time window centered at the maximum and following the wave packet with group velocity c_g; we repeat this for the whole time series building a collection of extreme events and their evolution. (c) Mean extreme event. The thick line shows the mean extreme event at different points along the channel, the shaded area a 1 standard deviation range around it. The *noise to signal* ratio is small in the focusing region, leading naturally to the question: Is there a common pathway by which most of the rogue waves arise?

Due to the *deep-water*, *small-steepness* and *narrow-band* properties, the evolution of the system is accurately described by the one-dimensional nonlinear Schrödinger equation [202, 13, 146, 6, 150, 29]. The NLS equation describes the change of the

complex envelope ψ, that relates to the surface elevation via the Stokes' series truncated at second order. The so-called Benjamin-Feir index (BFI, ratio of the nonlinear term over the dispersive term) quantifies the degree of nonlinearity present in the system. Its value for the three chosen regimes, as visible from Table 4.1, justifies their names. In order to characterize the dynamics leading to extreme

Set	γ	H_s (m)	BFI	L_{lin} (m)	L_{Per} (m)
quasi linear	1	.11	.56	8.9	32
intermediate	3.3	.13	1.50	46	61
highly nonlinear	6	.15	2.11	69	65

Table 4.1. The three experimental regimes with different strength of nonlinearity quantified by the nondimensional parameter BFI, ratio of nonlinearity over dispersion. γ and H_s characterize the JONSWAP spectrum enforced by the wave maker.

events of the water surface, we adopt the following procedure: At a fixed location $x = L$ along the channel, we select small observation windows around all temporal maxima of η that exceed a threshold z. Knowing the group velocity $c_g = \omega_0/(2k_0)$, we can track the wave packet backward in space and look at its shape at earlier points in the channel. This allows us to build a collection of extreme events and their precursors. In Fig. 4.8b, this procedure is pictured for two extreme events at $x = 45$ m and their precursors at $x = 30$ m and $x = 10$ m. Now, we can do statistics of the conditional event by computing its average and standard deviation at the different positions along the channel, obtaining a result like in Fig. 4.8c for the highly-nonlinear case.

Further details

The surface elevation η is measured simultaneously by 19 probes placed at different locations along the axes at the center of the tank, recording data with a rate of 5 measurements per second. At each of two different positions ($x = 75$ m and $x = 160$ m) two extra probes closer to the sides are used to check that the wave fronts remain planar. The signal at the wave maker $\eta(x = 0, t) \equiv \eta_0(t)$ is prepared according to the stationary random-phase statistics with fixed spectral amplitudes,

$$\eta_0(t) = \sum_{j=1}^{N} \sqrt{2\hat{C}(\omega_j)\delta\omega}\cos(2\pi\omega_j t + \phi). \tag{4.60}$$

The phase ϕ is uniformly distributed in [0,2π], $\delta\omega = \frac{2\pi}{T}$, $\omega_j = j\delta\omega$, T is the time-series length. The spectrum $\hat{C}(\omega)$ is given by the JONSWAP parametrization [96]

$$\hat{C}(\omega) = \frac{\alpha g^2}{(2\pi)^4\omega^5}\exp\left(-\frac{5}{4}\left(\frac{\omega_0}{\omega}\right)^4\right)\gamma^{\exp\left(-\frac{(\omega-\omega_0)^2}{2\sigma^2\omega_0^2}\right)}, \tag{4.61}$$

with g the gravity acceleration, $\omega_0 = 4.19\ \mathrm{s}^{-1}$ the carrier frequency (spectral peak), $\sigma = 0.007$ if $\omega \leq \tilde{\omega}_0$, and $\sigma = 0.09$ if $\omega > \tilde{\omega}_0$. The value of ω_0 yields a carrier wave number $k_0 = \omega_0^2/g = 1.79\ \mathrm{m}^{-1}$, where the dispersion relation of surface gravity waves in deep water is used. We use data from 5 time series for each set, each of which is 25 min long.

Because of the asymmetry of the JONSWAP spectrum, the effective group velocity to track the wave packets along their evolution is slightly different from the nominal one: in the quasi linear case the correction is by a factor .90, in the intermediate and in the highly nonlinear cases by a factor .88. The choice of the threshold z is meant to select extreme events with a similar probability for all sets: the values of $z = H_s$ for the quasi linear set, $z = 1.1\,H_s$ for the intermediate set and $z = 1.2\,H_s$ for the highly-nonlinear set lead to 78, 99 and 88 conditional maxima of the surface elevation such that $\eta(x = 45\ \mathrm{m}, t) \geq z$, respectively.

The NLS equation with space evolution reads

$$\partial_x \psi + 2\partial_t \psi + i\partial_t^2 \psi + i|\psi|^2 \psi = 0\,. \tag{4.62}$$

This is the effective equation describing the evolution of the surface complex envelope $|u|$ when the conditions of *deep-water*, *small-steepness* and *narrow-spectrum* are at once satisfied. The first condition requires $k_0 h \gg 1$, where h is the water depth. The wave tank is everywhere 5 to 10 meters deep, meaning $k_0 h \geq 8.95 \gg 1$. For the second condition, in the most energetic set we have $k_0 H_s \simeq .26 \ll 1$, where H_s is an estimate of the height of the tallest waves. The third condition requires $\Delta\omega/\omega_0 \ll 1$: for increasing nonlinearity, our three sets have a value of $\Delta\omega/\omega_0$ of .25, .11 and .09, i.e. all of them have narrow-banded spectra. The surface elevation relates to the complex envelope via the Stokes series truncated at second order (being third and higher orders negligible)

$$\eta = |\psi|\cos(\theta) + \tfrac{1}{2} k_0 |\psi|^2 \cos(2\theta) + O(k_0^2 |\psi|^3)\,, \tag{4.63}$$

where $\theta = k_0 x - \omega_0 t + \varphi$ and φ is the phase of ψ. Considering the third order term of the series would lead to differences always smaller than 2% that are not relevant to our results.

4.4.2 LDT method

In order to theoretically predict rare events, we make use of *instanton theory*. Developed originally in the context of quantum chromodynamics [171], at its core lies the realization that the evolution of any stochastic system, be it quantum and classical, reduces to a well-defined (semi-classical) limit in the presence of a smallness parameter. Concretely, the simultaneous evaluation of all possible realizations of the system subject to a given constraint results in an integral (classical or path-integral), where the integrand contains an action functional $S(\psi)$. The dominating

realization can then be obtained to leading order by approximating the integral by a *saddle point approximation*, $\delta S(\psi)/\delta\psi = 0$. The critical point ψ^* of the action functional associated with a stochastic system is called the *instanton*, and it represents the maximum likelihood realization of the stochastic system subject to the given constraints.

In our case, the evaluation is taken over all realizations ψ_0 of the random energy input at the wave generator, weighted by their respective probability under the JONSWAP spectrum. We consider only those initial data that result in an evolution that will exceed the defined threshold further down the channel, subject to its evolution via the NLS equation. More precisely, we are interested trajectories of the elevation $\eta(t, x)$ that fulfill $\eta(0, L) \geq z$, i.e. trajectories that exceed the threshold z at spatial position L, where we use temporal invariance to designate $t = 0$ to be the point in time of the extreme event. We denote the set of initial conditions that fulfill this constraint with $\Lambda(z)$. It will turn out that we can use z^{-1} as our smallness parameter, justifying the instanton approximation. Denoting by $P_L(z) \equiv \mathbb{P}(\eta(0, L) \geq z)$, we then have

$$P_L(z) \sim \exp(-\min_{\psi_0 \in \Lambda(z)} S(\psi_0)) \,, \tag{4.64}$$

where $S(\psi_0)$ is the action associated with the random initial condition. The derivation of the action, and the validity of the instanton approximation, can be made precise within *large deviation theory* (LDT), as discussed in chapter 2. At this stage, it is clear from equation (4.64) that the stochastic sampling problem is replaced by a deterministic optimization problem, which we solve numerically. The solution of this optimization problem, i.e. the associated minimizer, corresponds to the instanton trajectory, which in the following we will compare to trajectories obtained from the experiment.

4.4.3 Agreement of experimental extreme waves and instantons

In Fig. 4.9 we show the comparison between the evolution of the average extreme event of the experiment and the instanton evolution, both constrained at $x = 45$ m. In all cases the instanton tracks the dynamics of the mean extreme event very closely during the whole evolution. Moreover, in the focusing region the standard deviation around the mean is small, especially toward the end of the evolution. Observing this is itself a statement that indeed all of the extreme excursions leading to $\eta(x,0) \geq z$ resemble the instanton plus small random fluctuations. The instanton approximation shows excellent agreement not only across different degrees of nonlinearity (and therefore substantially different physical mechanisms), but also captures the behavior of precursors earlier along the channel.

In Fig. 4.10 the envelope evolution of a single realization of the conditional event is shown at multiple points and is compared to the instanton evolution, in

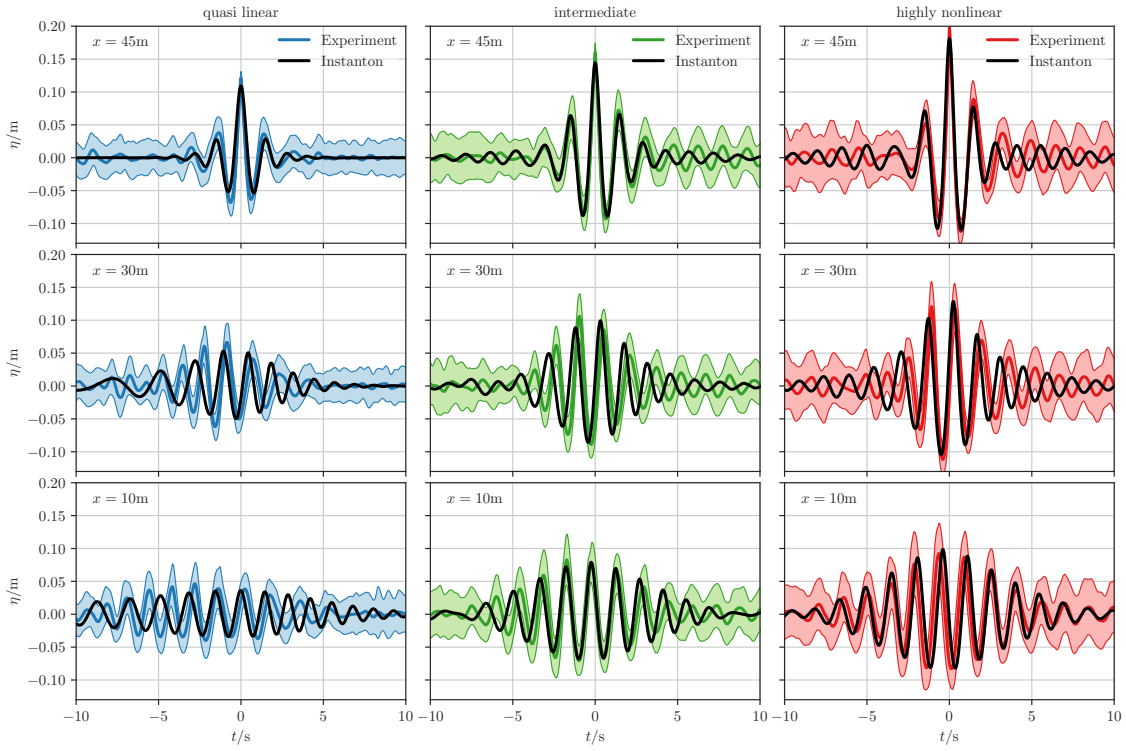

Figure 4.9. Validity of the instanton. The black lines display the instanton evolution along the channel, comparing it to the mean and standard deviation of the experimental extreme events in color, for different regimes of nonlinearity. The instanton prediction agrees with the experimental mean across all regimes, and captures the whole evolution along the channel. This confirms that the typical extreme event is well represented by the instanton, and the typical extreme events collapses onto this most likely one only with small fuctuations around it.

the highly-nonlinear case. Apart from random fluctuations, one can notice that in the focusing region the experimental sample shares with the instanton the same overall structure, which needs to be there in order for it to reach an extreme size.

4.4.4 A unifying theory of rogue waves

We want to particularly stress that previous theory described either the fully linear case or the fully nonlinear case. In the linear case, i.e. a Gaussian stationary field, the shape of an envelope time series with a large local maximum in $t = 0$ is expected to be given by the covariance of the wave field, i.e. the inverse Fourier transform of the spectrum. This is a well established result in probability [121]. In the oceanographic context, the result was rediscovered in the ′90s [18] and subsequently tested for some real quasi-Gaussian wave records in the ocean [183], also accounting for second-order Stokes' corrections [78]. We refer to such result as to the linear

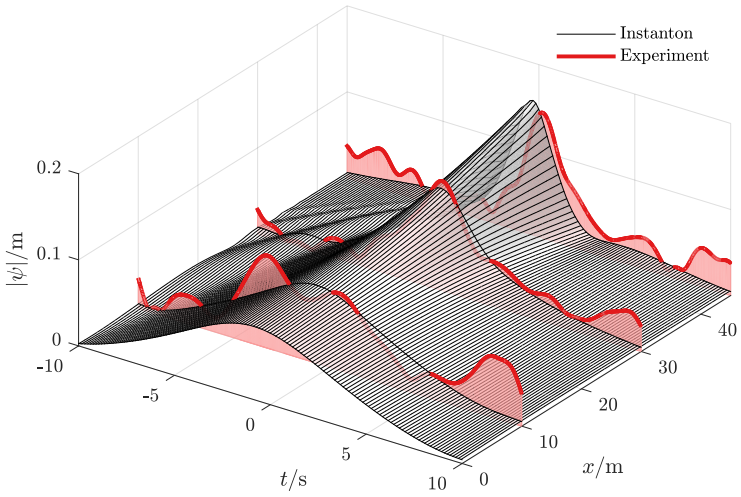

Figure 4.10. Agreement of the instanton with single extreme events. The evolution of a single realization of an extreme wave (depicted in red) is reasonably approximated by the instanton evolution (black/white surface), here for a sample of the highly nonlinear dataset. In order to capture the focusing pattern in an essential way, the envelope $|\psi|$ is plotted instead of the surface elevation η to remove carrier-frequency oscillations.

theory for extreme events.

The framework in which the mentioned result belongs and represents one of the central aspects is often called theory of *quasi-determinism*, valid for Gaussian fields – which strictly speaking requires a linear dynamics. A core result of the theory is the prediction that conditioning the surface elevation to have a maximum of size z at $t = 0$ (at a fixed location in space), the expected shape of the envelope is given by $z\,C(t)/C(0)$, where $C(t)$ is the covariance of the wave field, i.e. the inverse Fourier transform of the spectrum $\hat{C}(\omega_0 + \Omega)$.

In our case, the linear prediction is justified if the nonlinear focusing effects are small so that the statistics stay close to Gaussian along the tank, as in the quasi linear set. Then, conditioning on a temporal maximum of $\eta(L,0)$ at $x = L$, we can compute the history of the wave packet by evolving NLS backward in space. In Fig. 4.11a this linear prediction is plotted in comparison with the envelope of the mean conditional extreme event for the quasi linear set. A good agreement is observed at all spatial points considered. Moreover, the theoretical instanton found through the optimization procedure reduces perfectly to the linear prediction, proving that such result is included in the instanton theory and represents its limiting linear case.

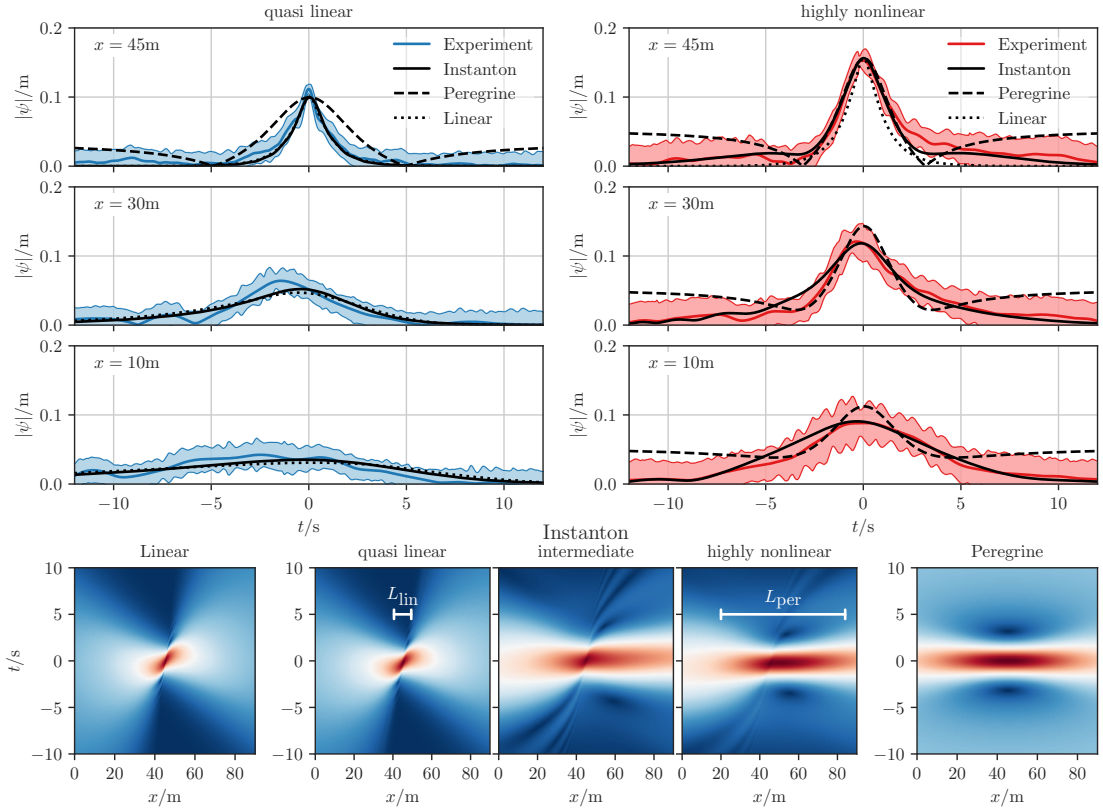

Figure 4.11. Comparison of the instanton to the limiting theories. (a) The quasi linear instanton converges to the linear prediction, correctly reproducing the mean extreme event of the experiment. The results clearly differ from the Peregrine soliton predicted by semi-classical theory. (b) The highly nonlinear instanton evolution, closely following the experimental mean extreme event, converges locally to a Peregrine around its space-time maximum, as predicted by the semi-classical theory, and reproduced by the instanton. The linear prediction instead fails particularly around the maximum. (c) The contour plots show agreement with the two limiting theories and recover the respective dominant length scales. In the linear limit, dominated by dispersion, the rogue waves arise and decay very rapidly. On the contrary, in the semi-classical limit, where nonlinearity dominates, the Peregrine-like structure of the extreme event is persistent, with a very slow dacay. The rogue waves in intermediate regimes, as exemplified by the central panel, display both linear and nonlinear features.

For the opposite regime, a recently formulated theorem [16] based on the regularization at the gradient catastrophe point states that in the zero-dispersion (semi-classical) regime of the NLSE any single localized pulse on a vanishing background leads locally to the emergence of a Peregrine soliton. By scale invariance of the NLSE, such a regime can be attained whenever a peak is large and focused enough

that the nonlinear term dominates over dispersion. In fiber optics [180], emerging Peregrine-like structures have been observed out of a random background. For the highly nonlinear case, in Fig. 4.11b we compare the instanton and the Peregrine reaching the same maximal height z at $x = 45$ m, finding that in the focusing region the two converge to the same shape, which is also closely followed by the envelope of the experimental mean event. Looking at the event precursor at earlier x, instead, we notice that the experimental mean stays close to the instanton, gradually diverging from the Peregrine. Thus, it appears that the instanton captures the mechanism underlying the rogue wave events also when nonlinearity rules over dispersion, tending locally to the Peregrine solution around the maximal focusing point, consistently with the regularization of the gradient catastrophe [16].

A useful quantification of the effective mechanisms consists in looking at the length scales at play. The linear length of dispersion is given by $L_{\text{lin}} = \omega_0^2/(k_0 \Delta\omega^2)$, while the characteristic length associated with the Peregrine soliton is $L_{\text{Per}} = \sqrt{L_{\text{lin}} L_{\text{nl}}}$, where $L_{\text{nl}} = 8/k_0^3 H_s^2$ is the nonlinear length of modulational instability. Looking at Fig. 4.11c, these length scales are clearly visible in space-time contours of the amplitude. In the linear and quasi linear regimes, the wave packet has a characteristic length around $L_{\text{lin}} \simeq 9$ m. Thus, we can state that linear superposition dominates and the expected mechanism leading to the extreme event is the linear dispersion of a coherent wave packet. The quasi linear instanton evolution is almost indistinguishable from the linear approximation. On the other hand, the extent of the structures in the highly nonlinear agrees with the length $L_{\text{Per}} \simeq 65$ m. The dynamics of the highly nonlinear instanton clearly converges to the Peregrine dynamics in proximity of the space-time maximal focusing, including the characteristic isolated "dips" of the amplitude around the extreme event. This picture highlights a sharp qualitative difference between the rapid evanescent linear rogue waves and the much more persistent nonlinear ones. Quite strikingly, the instanton is able to interpolate between those two limiting regimes, as evidenced by the intermediate instanton in Fig. 4.11c, which displays features of both the linear theory and the Peregrine soliton. The instanton therefore predicts extreme surface waves universally.

Chapter 5

Nonlinear waves in fiber optics

Introductive paragraph

In this chapter, we treat an application of the LDT method devised in chapters 2 and 3 to the problem of laser beams propagating within a Kerr optical fiber. The beams are randomly generated at the fiber input and evolve nonlinearly via the nonlinear Schrödinger equation with evolution in space. The situation is completely similar to the 1D evolution of random waves in a water flume analyzed in the previous chapter, as recently pointed out in [63]. Although in the following we carry out a purely numerical study, experiments of this kind are rapidly becoming more and more common. As expected, the mathematical setup is analogous to the one in chapter 4, but with different parameters. We compute the instantons of the problem and compare them to the "optical rogue waves". In the parameter range of optics, where no wave breaking occurs, one is able to explore regimes with very strong nonlinearity. The observed collapse of the extreme events onto the instanton is truly remarkable and the LDT probability estimate in the tail is accurate across the whole rogue-wave regime. We believe that these results will raise interest in the optics community and pave the way to new approaches such as for instance the theoretical design of optimal conditions to create highly focused energy concentration by exploiting nonlinearity.

Part of the material in this chapter was published in

G. Dematteis, T. Grafke, and E. Vanden-Eijnden. "Extreme event quantification in dynamical systems with random components". https://arxiv.org/abs/1808.10764, 2018.

5.1 The 1D NLS equation in optical turbulence

The nonlinear Schrödinger equation (NLSE) in one dimension arises in a variety of different contexts such as surface gravity waves [202, 150], nonlinear fiber optics [4], plasmas [10] and Bose-Einstein condensates [94, 159]. Here we will focus on applications of NLSE in nonlinear optics, a domain that has seen exciting experimental developments in recent years [109, 180, 185]. Specifically, we study the problem of the onset of rogue waves out of a bath of random waves taken as initial condition for NLSE, which is a key question in *integrable turbulence* [203, 167, 2, 44, 77].

In non-dimensional units, the 1D NLSE for the envelope of a light beam propagating in an optical fiber reads

$$\partial_\xi \Psi = i\frac{1}{2}\Psi_{\tau\tau} + i|\Psi|^2\Psi\,, \quad \tau \in \Gamma, \tag{5.1}$$

where $\Gamma = [0, T]$, with periodic boundary conditions $\Psi(\xi,0) = \Psi(\xi, T)$, and a suitable initial condition $\Psi(0, \tau) = \Psi_0(\tau)$, at the input end of the fiber $\xi = 0$. The non-dimensional distance ξ, time τ, and envelope Ψ are related to the respective physical quantities x, t, and ψ via characteristic constants that depend on the specifics of the optical fiber: $x = \mathcal{L}_0\xi$, $t = \mathcal{T}_0\tau$ and $\psi = \sqrt{\mathcal{P}_0}\Psi$. For instance, if we pick $\mathcal{T}_0 = 5$ ps, $\mathcal{L}_0 = 0.5$ km, $\mathcal{P}_0 = 0.5$ mW, the NLSE (5.1) models an optical fiber with dispersion $|\beta_2| = \mathcal{T}_0^2/\mathcal{L}_0 = 50$ ps^{2}km^{-1} and nonlinearity $\gamma = 1/(\mathcal{L}_0\mathcal{P}_0) = 4$ km^{-1}mW^{-1}.

Let us denote by $\{\hat\Psi_n\}_{n\in\mathbb{Z}}$ the Fourier component of $\{\Psi(\tau)\}_{\tau\in[0,T]}$, i.e.

$$\hat\Psi_n = \frac{1}{T}\int_0^T e^{-i\omega_n\tau}\Psi(\tau)d\tau, \qquad \Psi(\tau) = \sum_{n=-\infty}^{+\infty} e^{i\omega_n\tau}\hat\Psi_n\,, \tag{5.2}$$

where $\omega_n = 2\pi n/T$ and $n \in \mathbb{Z}$. Equation (5.1) is derived under the *quasi-monochromatic* assumption, meaning that the spectrum $\hat C_n$ defined as

$$\hat C_n = \frac{1}{T}\int_0^T e^{-i\omega_n\tau}C(\tau)d\tau\,, \qquad C(\tau-\tau') = \mathbb{E}(\Psi_0(\tau)\bar\Psi_0(\tau'))\,, \tag{5.3}$$

must be narrow – here and below the bar denoting complex conjugation. We will consider a Gaussian spectrum with

$$\hat C_n = \mathcal{A}e^{-\omega_n^2/(2\Delta)} \qquad \mathcal{A} > 0, \qquad \Delta > 0, \qquad -M \le n \le M, \quad M > 0, \tag{5.4}$$

and $\hat C_n = 0$ for $|n| > M$. Assuming that the initial $\Psi(0,\tau)$ is a Gaussian field with mean zero and covariance $C(\tau-\tau')$, this implies the representation

$$\Psi(0,\tau,\vartheta) = \sum_{n=-M}^{M} e^{i\omega_n\tau}\hat C_n^{1/2}\vartheta_n, \tag{5.5}$$

where ϑ_n are complex Gaussian variables with mean zero and covariance $\mathbb{E}\vartheta_n\bar\vartheta_m = \delta_{m,n}$, $\mathbb{E}\vartheta_n\vartheta_m = \mathbb{E}\bar\vartheta_n\bar\vartheta_m = 0$. Note that the spectral amplitude is related to the optical power $P(\xi,\tau) = |\psi(\xi,\tau)|^2$ (statistically homogeneous in τ) via $\mathcal{A} = \mathbb{E}(P)/\sum_n e^{-\omega_n^2/(2\Delta)}$.

The initial statistical state of the system is thus completely determined given the two parameters Δ and $\mathbb{E}(P)$, and the average power $\mathbb{E}(P)$ is relevant to optical experiments – it also enjoys the property of being invariant under the NLSE evolution in the variable ξ, i.e. it can be measured at the input or at the output of the optical fiber, equivalently.

In the set-up above, we will investigate extreme fluctuations of the optical power at the output of the optical fiber ($\xi = L$). Recalling that $|\Psi(L,\tau)| = \sqrt{P(L,\tau)}$, this amounts to looking at the statistics of

$$f(\Psi(\vartheta)) = \max_{\tau\in\Gamma} |\Psi(L,\tau,\vartheta)|\,, \qquad L > 0. \tag{5.6}$$

Analyzing this observable using the framework developed in Secs. 2.2.1 and 3 amounts to minimizing the cost function (this is (3.1))

$$E(\Psi,\theta) = I(\theta) - \lambda f(\Psi) \qquad \text{with} \qquad I(\theta) = \tfrac{1}{2}\sum_{n=-M}^{M} |\theta_n|^2\,. \tag{5.7}$$

This minimization must be performed on the $2\times(2M+1)$-dimensional space $\Omega \subseteq \mathbb{C}^{2M+1}$ of the initial conditions. The gradient of the cost function (5.7) is given by

$$\nabla_\theta E(\Psi(\theta),\theta) = \nabla_\theta I(\theta) + \Re(J(L,\tau_*))^T \frac{\Re(\Psi(L,\tau_*))}{|\Psi(L,\tau_*)|} + \Im(J(L,\tau_*))^T \frac{\Im(\Psi(L,\tau_*))}{|\Psi(L,\tau_*)|}\,, \tag{5.8}$$

where $\Psi(L,\tau_*) \equiv \max_{\tau\in\Gamma}|\Psi(L,\tau)|$. The field Ψ is evolved with (5.1) and the initial condition depends on the point $\theta\in\Omega$ through the mapping $\Psi(0,\theta)$ defined in (5.5), with the difference that here θ is no longer random. The matrix J (also complex) evolves according to

$$\partial_\xi J(\xi,\tau) = \int_0^L d\xi' \left(\frac{\delta b(\Psi(\xi))}{\delta\Psi(\xi')}\, J(\xi',\tau) + \frac{\delta b(\Psi(\xi))}{\delta\bar\Psi(\xi')}\, \bar J(\xi',\tau) \right), \tag{5.9}$$

where $b(\Psi(\xi)))$ is a shorthand for the right hand side of (5.1): explicitly

$$\int_0^L d\xi' \,\frac{\delta b(\xi)}{\delta\Psi(\xi')}\, J(\xi') = \left(\frac{i}{2}\partial_{\tau\tau} + 2i|\Psi(\xi)|^2\right) J(\xi)\,, \tag{5.10}$$

$$\int_0^L d\xi' \,\frac{\delta b(\xi)}{\delta\bar\Psi(\xi')}\, \bar J(\xi') = i\Big(\Psi(\xi)\Big)^2 J(\xi)\,. \tag{5.11}$$

The initial condition for (5.9) is

$$J(\xi=0,\theta) = \nabla_\theta\Psi(0,\theta)\,. \tag{5.12}$$

Before turning to the results, let us explain how the numerical simulations were performed. Equations (5.1) and (5.9) were evolved from $\xi = 0$ to $\xi = L$ (up to

$L = 0.2$) using the pseudo-spectral second order Runge-Kutta exponential-time-differencing method (ETDRK2) [45, 107] with step $d\xi = 5 \times 10^{-4}$ on a periodic box $[0, T]$ discretized by 2^{12} equidistant grid points. The size $T = 30$ is found large enough for the boundary conditions to not affect the statistics on the spatio-temporal scales considered. Each Monte Carlo simulation involves 10^6 realizations of the random initial data constructed via (5.5), with $M = 45$. Adding more modes to the initial condition does not affect the results in any significant way. The minimization was performed in the space Ω (with high dimension $2 \times (2M + 1) = 182$). This step was carried out via steepest descent with adaptive step (line search) and preconditioning of the gradient, using the covariance of the initial condition as metric, as explained in Sec. 3.

5.2 Instantons in optical fibers and the gradient catastrophe

For generality, we present the results for the normalized field $A(\xi, \tau) = \Psi(\xi, \tau)/\sqrt{\mathbb{E}(P)}$ using non-dimensional units. One can easily obtain the physical dimensions by applying the straightforward transformations given below equation (5.1). Four sets of parameters have been chosen to explore different regimes: In Set 1, we take $\Delta = \pi$, $\mathbb{E}(P) = 5/4$; in Set 2, $\Delta = \pi/2$, $\mathbb{E}(P) = 5/4$; in Set 3, $\Delta = 3\pi/2$, $\mathbb{E}(P) = 5/4$; and in Set 4, $\Delta = \pi$, $\mathbb{E}(P) = 5/9$.

There have been recent claims, supported by both numerical and experimental evidence [180, 185], about the universality of the Peregrine Soliton (PS) as a pathway to optical rogue waves out of a random background. For this reason, we carried out a comparison between the instantons and the PS. In Fig. 5.1, the path of occurrence of two extreme events is shown for Set 1, selected among the events in the random sampling with maximum power amplification $|A|^2 = P/\mathbb{E}(P)$ exceeding a value of 40. The instanton and the PS reaching the same power amplification are also plotted.

In Fig. 5.2 the probability $P(z) = \mathbb{P}(\max_\tau |A(L, \tau)| \geq z)$ is shown for various values of L, showing good agreement between the results from MC sampling and those from LDT optimization. A rough estimate for the onset threshold of optical rogue waves is $|A|_{RW} = 4\sqrt{2/\pi}\mathbb{E}(|A|) \simeq 2.8$ [63], independently of the set considered because of the use of the normalized variable A. As can be seen, the focusing NLSE increases the probability of large excursions of $|A(L, \tau)|$ compared to its initial Gaussian value with expectation $\mathbb{E}(|A(L = 0, \tau)|) = \sqrt{\pi/4}$. This happens gradually as the distance L separating the input from the output increases. The tail fattening can be interpreted quantitatively in terms of the typical lengths of the coherent structures of NLSE. Defining the linear length as $L_{\text{lin}} = 2/\Delta^2$ and the nonlinear length as $L_{\text{nlin}} = 1/\mathbb{E}(P)$, the typical length of emergence of a coherent structure starting from a small hump is $L_c = \frac{1}{2}\sqrt{L_{\text{lin}} L_{\text{nlin}}}$. This gives $L_c = 0.2$ for

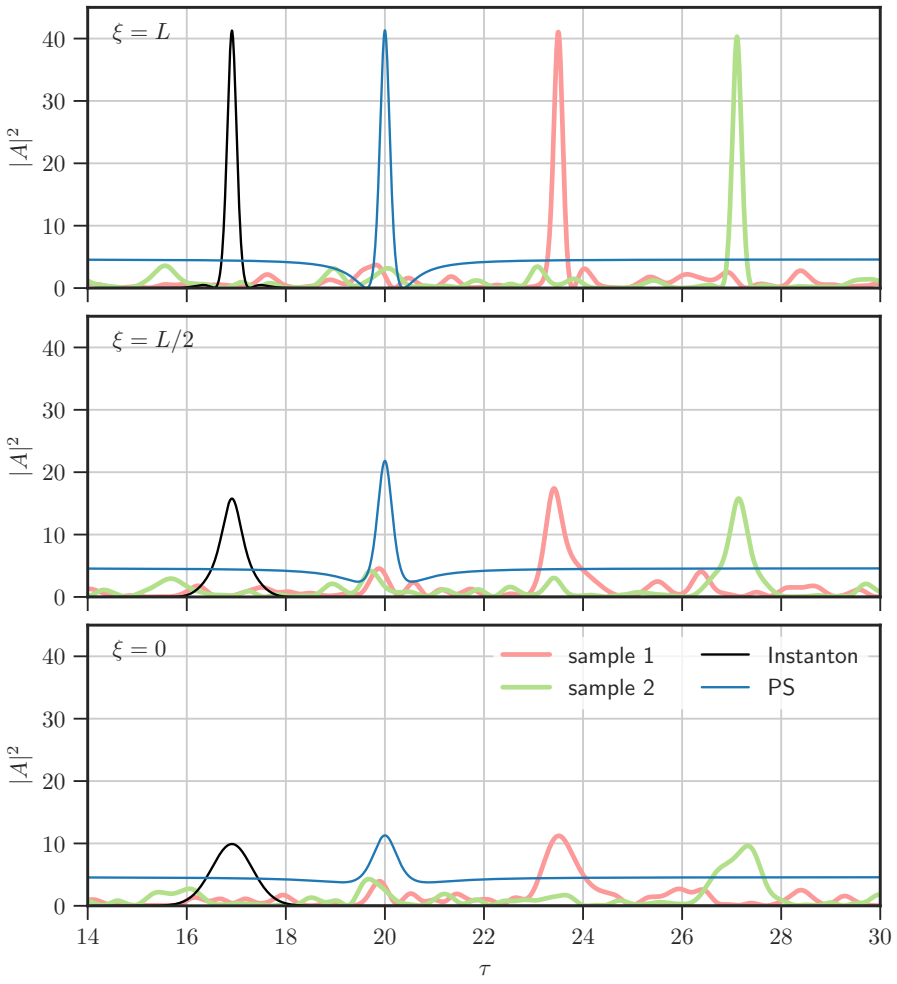

Figure 5.1. Set 1: The paths of occurrence of two extreme events plotted are compared with the instanton and the Peregrine solution reaching the same maximum power at $\xi = L$. Shown is the quantity $|A(\xi, \tau)|^2$, i.e. the power in units of average power, at three different locations ($L = 0.2$). The solution are shifted away from one another for clarity, exploiting homogeneity in τ.

Set 1, in good agreement with the width of the spatial transient over which the fast tail fattening takes place.

The asymptotic agreement of the probabilities shown in Fig. 5.2 is a numerical evidence that the focusing NLSE (5.1) with random initial data (5.5) satisfies an LDP. Additional support for the LDP is found in Fig. 5.3, where we compare the instanton with the sampling mean. Looking at the signal to noise ratio, one sees that the events reaching a certain extreme amplification are all very similar. According to the results in Sec. 2.2.1, these events are expected to have typical fluctuations in the direction perpendicular to the instanton in the space Ω: notice how away

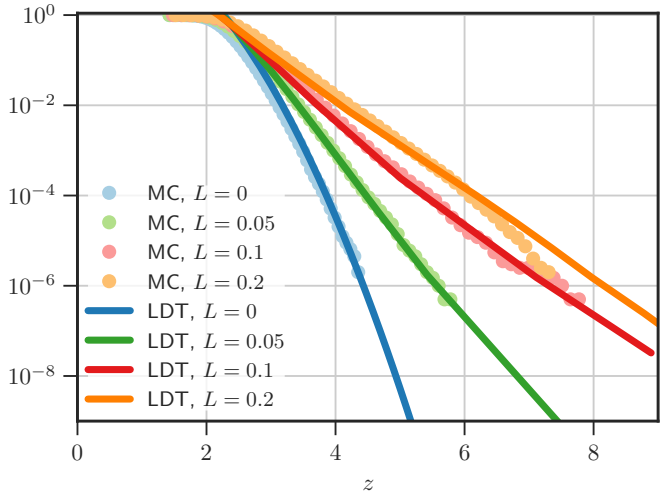

Figure 5.2. Set 1: Comparison between the probability distributions of $\max_\tau |A(L,\tau)|$ in the periodic time window $[0,T]$ obtained by MC with 10^6 samples, and their corresponding LDT estimates computed using the optimization method. The plot captures the tail fattening due to the NLSE dynamics, as the output point is taken at increasing distance L from the input. The rogue-wave threshold is $|A|_{RW} \simeq 2.8$. The characteristic length of emergence of the coherent structures is $L_c = 0.2$, compatible with the observed tail fattening.

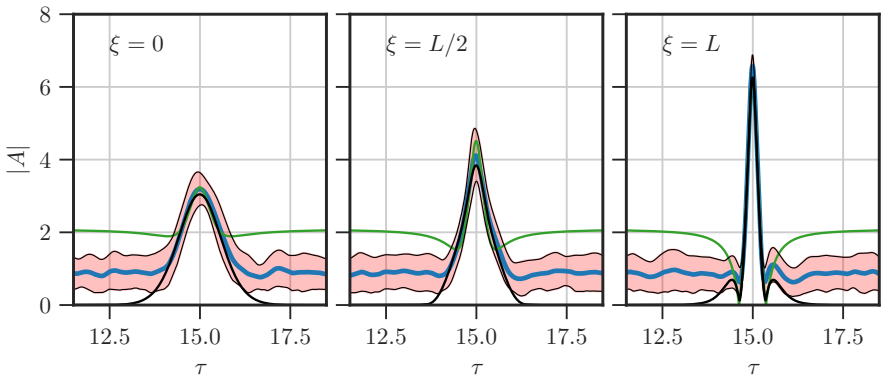

Figure 5.3. Set 1: Results of the conditioning on the sampling for $\max_{\tau\in\Gamma} |A(L,\tau)| \geq z = 6.25$, with $L = 0.2$. Shown is the average of the conditional event (blue line), surrounded by the 1 std range (red area). The instanton (black line) is the optimal event reaching maximal intensity $A = z$ at the output point $\xi = L$. The PS is also represented (green line), normalized to have intensity z at the point of maximal space-time focusing. From left to right, the panels are at $\xi = 0$, $\xi = L/2$ and $\xi = L$.

from the focusing region (determined by the direction perpendicular to the instanton because there the instanton is vanishing) the observable $|A|$ fluctuates with standard deviation $\sqrt{\mathbb{E}(P)/2}\sqrt{(4-\pi)/2}/\sqrt{\mathbb{E}(P)} \simeq 0.57$ around the expected value $\sqrt{\pi/2}\sqrt{\mathbb{E}(P)/2}/\sqrt{\mathbb{E}(P)} \simeq 0.89$, exactly as expected for typical events. Instead, the extreme size of the event is due to the component parallel to the instanton in Ω, with small fluctuations in this direction: As a matter of fact, in the focusing region (determined by the component parallel to the instanton) the signal to noise ratio becomes very big, meaning that, as z increases, the extreme rogue waves with $\max_\tau |A(\tau, L)| \geq z$ become closer to the instanton reaching $\max_\tau |A(\tau, L)| = z$.

Interestingly, from the knowledge of the LDT tails for a particular configuration of the parameters Δ and $\mathbb{E}(P)$ we can derive the LDT tails for any combination of Δ and $\mathbb{E}(P)$, using only analytical transformations. This is possible thanks to two properties: First, the scale invariance of the NLSE; second, the way the parameter $\mathbb{E}(P)$ appears in the cost function (5.7). Indeed the term $I(\theta)$ is independent of $\mathbb{E}(P)$, and from (5.5) the term $f(\Psi(\theta))$ can be seen as a function of $\sqrt{\mathbb{E}(P)}\theta$.

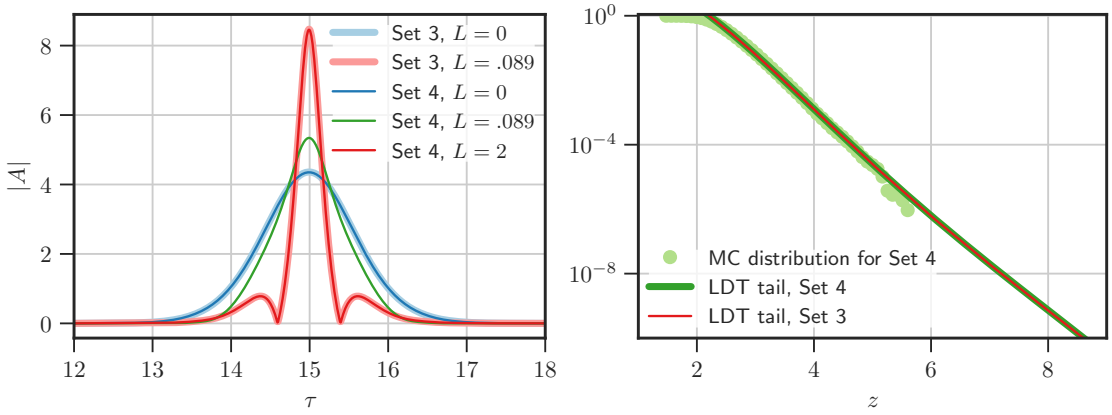

Figure 5.4. The two panels show how knowledge of the LDT tail at the output point L for a given Δ and $\mathbb{E}(P)$ allows us to recover the LDT tail for an arbitrary Δ', with the properly rescaled mean power $\mathbb{E}(P)'$, space L', and time τ'. Left: Instanton reaching $\max_{\tau \in \Gamma} |A(L, \tau)| = 8.5$, for sets 3 and 4. Right: $\mathbb{P}(\max_\tau |A(L, \tau|) \geq z$ from MC sampling and tail estimate for Sets 3 and 4, at an equivalent rescaled output point L. Note that not only the probability tail is the same for the two sets, but also the entire distribution, as the scale invariance establishes a complete equivalence between two sets having the same value of the ratio $\sqrt{\mathbb{E}(P)}/\Delta$.

- Starting from the second property, we have that given a fixed spectral width Δ and a mean power $\mathbb{E}(P)$, giving the cost function (5.7) $E(\theta, \lambda)$, the cost function $E'(\theta, \lambda)$ associated to a new mean power $\mathbb{E}(P)'$ (but same spectral width) can be

written as

$$E'(\theta,\lambda)=\frac{\mathbb{E}(P)}{\mathbb{E}(P)'}E(\theta',\lambda'),\quad \theta'=\theta\sqrt{\frac{\mathbb{E}(P)'}{\mathbb{E}(P)}},\quad \lambda'=\lambda\frac{\mathbb{E}(P)'}{\mathbb{E}(P)}\,. \tag{5.13}$$

Since λ' is nothing but a rescaling of λ, and they are both arbitrary variables, E and E' represent actually the same landscape, just differing by a positive factor and a rescaling of the variables. This implies that if we know an instanton $\theta^\star(z)$ and its associated probability $P(z)$ for the mean power $\mathbb{E}(P)$, we also know that for mean power $\mathbb{E}(P)'$ the same event will have instanton $\theta'^\star(z)=\theta^\star(z)\sqrt{\mathbb{E}(P)'/\mathbb{E}(P)}$ with associated probability

$$P'(z)=P(z)^{\frac{\mathbb{E}(P)}{\mathbb{E}(P)'}}\,. \tag{5.14}$$

Thus, keeping Δ fixed, the LDT tails for a given $\mathbb{E}(P)$ are sufficient to generate the LDT tails for any mean power $\mathbb{E}(P)'$, using (5.14).

- Using the scale invariance of the NLSE, it is possible to make a similar argument to extend the LDT tails to arbitrary Δ. Knowing that initial conditions with the same ratio $\sqrt{\mathbb{E}(P)}/\Delta$ are scale invariant for the NLSE, one can pick an arbitrary spectral width Δ'. This gives a new mean power $\mathbb{E}(P)'=\mathbb{E}(P)(\Delta'/\Delta)^2$, and allows us to compute the new length $L'=(\Delta/\Delta')^2L$ and time coordinate $\tau'=(\Delta/\Delta')\tau$. Thus, a bijection is established between the two parameter sets, where each pair is characterized by the same non-dimensional instanton and same probability. Hence, knowing the LDT tails at different L for one value of the spectral width, one is able to obtain the whole spatial transient of the LDT tails for an arbitrary spectral width. In Fig. 5.5 the invariance of the non-dimensional instanton and of the LDT tail is shown for Sets 3 and 4, which yield the same dynamics once the appropriate rescaling is performed.

Figs. 5.1 and 5.3 confirm that the high-power pulses arising spontaneously from a random background tend to the shape of the PS around its maximum space-time concentration [185]. Interpreting this in light of the gradient-catastrophe regularization [16], it is clear that such characteristic shape of the extreme power amplifications is independent of the solitonic content of the field, although it is shared with the local behavior of an exact solitonic solution. The random extreme realizations quickly diverge from the PS away from the maximum, however. In contrast, the instantons characterize all the essential dynamics of the extreme events in integrable turbulence. They give an approximation of the extreme excursions that is much more accurate than the PS, as can be observed in Fig. 5.3, and their shape adapts to the size of the event. In addition, unlike the PS, they come with probabilistic information and allows the estimation of the distribution tail, as seen in Fig. 5.2, with mathematical justification in the LDT result (2.36). Furthermore, the instantons depend on the statistical state of the random background, as shown in Fig. 5.5,

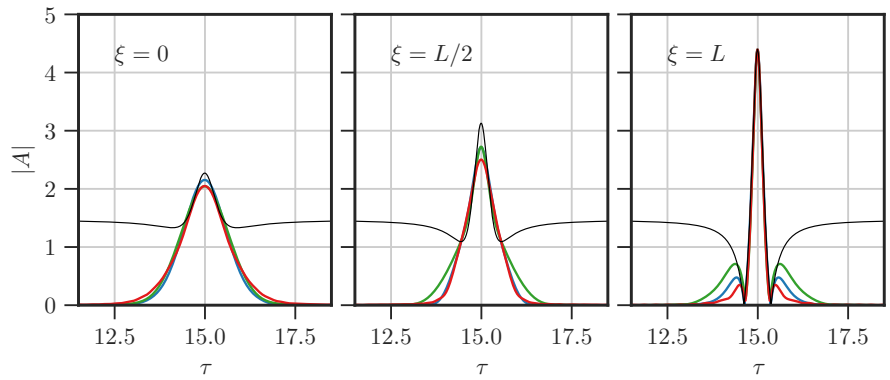

Figure 5.5. Snapshots at increasing spatial coordinate from left to right ($\xi = 0$, $\xi = 0.1$, $\xi = 0.2$) of instantons reaching the same peak intensity, for the three sets of parameters with different spectral width: Set 1 ($\Delta = \pi$) in red; Set 2 ($\Delta = \pi/2$) in green; and Set 3 ($\Delta = 3\pi/2$) in blue. The PS reaching the same final height (at the point of maximal focusing) is also plotted in black. For all the profiles, striking agreement is observed around the point of maximal focusing in space-time, while significant differences are observed away from that point.

while the PS is always the same. Because of these properties and their connection with the gradient catastrophe (which is their generating mechanism), the instantons can be important objects for further investigations in integrable turbulence. In this context, recent results [49] suggest that the formation of extreme coherent structures may not necessarily be linked to integrability, but may pertain to a more general class of systems with instabilities (e.g. due to non-resonant interactions) leading to spatio-temporal concentration phenomena.

Chapter 6

Nonequilibrium forcing of an inhomogeneous elastic rod

Introductory paragraph

In this section we study a model for a one-dimensional rod with random elasticity structure, under the action of a prescribed external mechanical forcing (i.e. pulling at one end). Harmonic interactions are used for instance for coarse-grained models of DNA stretching [25, 40, 118] to account for the elastic response from extension – although our minimal model neglects the elastic response from changes in the polymer configuration. In general, the discrete version of the model can be viewed as a simple *molecular mechanics* model for a short sector of a stiff polymer with harmonic interaction between nearest neighbors (Hooke's law) [1]. This being said about possible physical or biological applications, we want to emphasize that the present model is picked as a simple illustrative example case where a random structure (here the elasticity coefficients) and a given dynamics (here Hamiltonian) are known, and we do not claim the extent of our results to go beyond the intended illustrative purpose. In a strongly out-of-equilibrium regime enforced by an external time-dependent forcing, we characterize the dynamical response of the system, equipped with a random inhomogeneous elasticity that represents our uncertainty on the system's state. By LDT optimization we locate the optimal configurations leading to extreme responses and we show that such realizations dominate the statistics asymptotically.

In the case of forcing increasing linearly in time, we are able to derive analytical results. This is an important test case to validate the numerical method. Moreover, the linear protocol corresponds to pulling at constant speed, which is a common scenario in the experiments [122]. Then, our numerical method is exploited to study efficiently the extreme events with an arbitrary nonlinear forcing, and to yield results about the optimality of the protocol, a problem that has risen recent

interest for small nonequilibrium molecular systems [9, 48].

Part of the material in this chapter was published in

G. Dematteis, T. Grafke, and E. Vanden-Eijnden. "Extreme event quantification in dynamical systems with random components". https://arxiv.org/abs/1808.10764, 2018 (to appear in "SIAM/ASA Journal of Uncertainty Quantification").

6.1 Time-dependent Hamiltonian system with randomness

6.1.1 Continuous model with random structure

Consider a one-dimensional elastic rod of length 1 that is being pulled at one end with a time-dependent force and whose energy is specified in terms of its displacement field $u : [0,1] \to \mathbb{R}$ via

$$V(u,t) = \frac{1}{2}\int_0^1 \mathcal{D}(x)\,|\partial_x u|^2\,dx - r(t)u(1)\,, \tag{6.1}$$

where the first term is the total internal energy of the rod and the second term is the external energy (negative of the work potential); $\mathcal{D}(x) > 0$ is the elasticity coefficient, assumed to be spatially dependent, and $r(t)$ is a prescribed external forcing protocol acting on the right end of the rod – the specific form of $r(t)$ will be introduced later. The dynamics of the rod is governed by the Euler-Lagrange equation associated with (6.1):

$$\partial_t^2 u = \partial_x(\mathcal{D}(x)\partial_x u) \quad x \in (0,1)\,, \tag{6.2}$$

with initial conditions to be prescribed later and boundary conditions

$$u(t,0) = 0\,, \quad \mathcal{D}(1)\partial_x u(t,1) = r(t)\,, \quad \forall t \geq 0\,. \tag{6.3}$$

In order to introduce uncertainty in the model we make the elasticity random, i.e. we take $\mathcal{D}(x) \equiv \mathcal{D}(x,\vartheta)$. Here we will assume that $\mathcal{D}(x,\vartheta)$ is piecewise constant over blocks of size $1/M$ for some $M \in \mathbb{N}$, with independent values in each block. Specifically, we take:

$$\mathcal{D}(x,\vartheta) = \sum_{k=1}^{M} \varphi_k(x) g(\vartheta_k)\,, \tag{6.4}$$

where the functions $\{\varphi_k\}_{k=1}^M$ are given by

$$\varphi_k(x) = \begin{cases} 1 & \text{if} \quad M^{-1}(k-1) \leq x < M^{-1}k \\ 0 & \text{otherwise} \end{cases}; \tag{6.5}$$

g is a given function; and $\{\vartheta_k\}_{k=1}^M$ are i.i.d. random variables. Below we will consider two cases:

Case 1. Here we assume that $g : (0,\infty) \to (0,\infty)$ with

$$g(y) = y^{-1} \tag{6.6}$$

and we take the variable $\{\vartheta_k\}_{k=1}^M$ to be exponentially distributed, i.e.

$$\mathbb{P}(\vartheta_k \geq \theta_k) = e^{-\alpha\theta_k}, \qquad \theta_k \geq 0, \quad \alpha > 0\,. \tag{6.7}$$

This choice implies that

$$S(\eta) = \log \mathbb{E} e^{\langle \eta,\vartheta\rangle} = -\sum_{k=1}^{M} \log(1-\alpha^{-1}\eta_k), \qquad \eta_k < \alpha \quad \forall k = 1,\dots,M, \tag{6.8}$$

so that

$$I(\theta) = \sum_{k=1}^{M} (\alpha\theta_k - 1 - \log\theta_k) \qquad \theta_k > 0 \quad \forall k = 1,\dots,M\,. \tag{6.9}$$

Case 2. Here we assume that $g : \mathbb{R} \to (0,\infty)$ with

$$g(y) = \tfrac{1}{2}y + \sqrt{\tfrac{1}{4}y^2 + 1}, \tag{6.10}$$

and we take the variable $\{\vartheta_k\}_{k=1}^M$ to be normally distributed with variance $\sigma^2 > 0$, i.e.

$$\vartheta_k = \mathcal{N}(0,\sigma^2) \tag{6.11}$$

This choice implies that

$$S(\eta) = \tfrac{1}{2}\sum_{k=1}^{N} \sigma^2\eta_k^2, \qquad I(\theta) = \tfrac{1}{2}\sum_{k=1}^{N} \sigma^{-2}\theta_k^2\,. \tag{6.12}$$

Given this random input, our aim is to investigate the statistics of the displacement of the right end of the rod at time T: this amount to considering the observable $f(u(T)) = u(T,1)$, and studying the behavior of

$$P(z) = \mathbb{P}(u(T,1,\vartheta) \geq z) \qquad \text{for} \quad z \gg 1. \tag{6.13}$$

Below we will analyze the behavior of this quantity in two cases, when the forcing $r(t)$ in (6.1) is linear in t and when it is not – the first situation is amenable to analytical treatment whereas the second is not in general. Note that in both situations, the behavior of $P(z)$ for large z will depend on how fast $g(u)$ decays to zero: due to the shape of g this will depend on the right tail of the distribution of ϑ_k in Case 1 and on its left tail in Case 2.

6.1.2 Discrete model

To perform the numerics, we need to consider a spatially discretized version of the model above. We do so by introducing the discrete energy

$$V(u,t) = \frac{1}{2}\sum_{j=0}^{N-1} \mathcal{D}_{j+1}(\vartheta)\frac{(u_{j+1}-u_j)^2}{\Delta x} - r(t)\,u_N\,, \tag{6.14}$$

in which $u_j = u(j\Delta x)$, $\mathcal{D}_j = \mathcal{D}(j\Delta x)$, $\Delta x = 1/N$. Alternatively, (6.14) can be thought of as the energy for a system of $N+1$ beads u_j connected by N springs with random spring constants $\mathcal{D}_j(\vartheta)$. The dynamics obeys the system of ODEs

$$\partial_t^2 u_j = \frac{\mathcal{D}_{j+1}}{\Delta x^2}(u_{j+1}-u_j) - \frac{\mathcal{D}_j}{\Delta x^2}(u_j - u_{j-1})\,, \qquad j = 1,...,N-1\,, \tag{6.15}$$

with fixed boundary condition $u_0 = 0$ at the left end and dynamic boundary condition

$$\partial_t^2 u_N = -\frac{\mathcal{D}_N}{\Delta x^2}(u_N - u_{N-1}) + \frac{r(t)}{\Delta x} \tag{6.16}$$

at the right end. We will pick $N = PM$ for some $P \in \mathbb{N}$, so that by our choice for $\mathcal{D}(x,\vartheta)$ in (6.4) we have

$$\mathcal{D}_j(\vartheta) = g(\vartheta_k) \qquad \text{for } \lceil j/P \rceil = k,\quad j = 1,\dots,N,\quad k = 1\dots,M. \tag{6.17}$$

Since we focus on the statistics of the observable $f(u(T)) = u_N(T) = u(T,1)$ that measures the displacement at time T of the right end point with respect to its initial position, the cost function is

$$E(u,\theta) = I(\theta) - \lambda\, u_N(T)\,, \tag{6.18}$$

to optimize on the parameters $\{\theta_k\}_{k=1}^M$.

6.2 Analytical optimization for a linear forcing

Assume that $r(t) = at$ for some $a > 0$ and as initial conditions for (6.2) take

$$u(0,x) = 0\,, \quad \partial_t u(0,x) = a\int_0^x \frac{dx'}{\mathcal{D}(x',\theta)}\,, \quad \forall x \in [0,1]\,. \tag{6.19}$$

The solution to (6.2) equipped with the boundary conditions in (6.3) is

$$u(t,x,\vartheta) = at\int_0^x \frac{dx'}{\mathcal{D}(x',\vartheta)}\,. \tag{6.20}$$

Let us consider the implications of this formula in *Case 1*, which is suitable to derive analytical results. Eq. (6.20) implies that

$$u(T,1,\vartheta) = aT \int_0^1 \frac{dx'}{\mathcal{D}(x',\vartheta)} = \frac{aT}{M} \sum_{k=1}^{M} \vartheta_k \,, \tag{6.21}$$

where we used the specific form of $\mathcal{D}(x,\vartheta)$ given in (6.4) with g given in (6.6). Note that since the discrete equivalent to the initial conditions (6.56) is

$$u_j(0) = 0, \qquad \partial_t u_j(0) = \frac{a}{M} \sum_{k=1}^{j} \theta_k \,, \tag{6.22}$$

the result (6.21) also holds for the discretized model, i.e. we have

$$u_N(T,\vartheta) = \frac{aT}{M} \sum_{k=1}^{M} \vartheta_k \,. \tag{6.23}$$

From (6.7), this implies that $u(T,1,\vartheta) = u_N(T,\vartheta)$ follows a gamma distribution with shape parameter M and rate parameter $\alpha M(aT)^{-1}$:

$$\begin{aligned} P(z) &= \int_z^\infty \frac{(\alpha M(aT)^{-1})^M y^{M-1}}{(M-1)!} e^{-\alpha M(aT)^{-1} y} dy \\ &= \frac{1}{(M-1)!} \Gamma\left(M, \alpha M(aT)^{-1} z\right) \,, \end{aligned} \tag{6.24}$$

where $\Gamma(\cdot,\cdot)$ is the upper incomplete Gamma function. When $z \gg 1$ with M fixed, (6.24) gives

$$P(z) \sim \frac{(\alpha M(aT)^{-1} z)^{M-1}}{(M-1)!} e^{-\alpha M(aT)^{-1} z} \,, \tag{6.25}$$

meaning that

$$\log P(z) \sim -\alpha M(aT)^{-1} z + (M-1)\log(\alpha M(aT)^{-1} z) - \log(M-1)!. \tag{6.26}$$

In this last expression the second and third terms at the right hand side are sub-dominant over the first, $\alpha M(aT)^{-1} z$, and disappear in the limit as $z \to \infty$. It is useful to keep this terms for comparison with the result (2.36) in Theorem 1 and the result (2.39), which we do next.

If we solve

$$\min I(\theta) = \min \sum_{k=1}^{M} (\alpha\theta_k - 1 - \log \alpha\theta_k) \qquad \text{subject to} \qquad u(T,1,\theta) = \frac{aT}{M} \sum_{k=1}^{M} \theta_k = z \,, \tag{6.27}$$

we get

$$\theta_k^\star(z) = (aT)^{-1} z \qquad \text{for} \quad k = 1, \ldots, M \,. \tag{6.28}$$

As a result

$$I(\theta^\star(z)) = M\left(\alpha(aT)^{-1}z - 1 - \log(\alpha(aT)^{-1}z)\right), \tag{6.29}$$

which from (6.26) is consistent with $\log P(z) \sim -I(\theta^\star(z))$ as $z \to \infty$, as predicted by (2.36). Note also that here

$$\eta_k^\star(z) = \partial_{\theta_k} I(\theta^\star(z)) = \alpha - aTz^{-1} \qquad \text{for} \quad k = 1, \dots, M\,. \tag{6.30}$$

Since this implies that $|\eta^\star(z)| \to \sqrt{M}\alpha$ as $z \to \infty$, this means that the condition in (2.37) is not satisfied here.

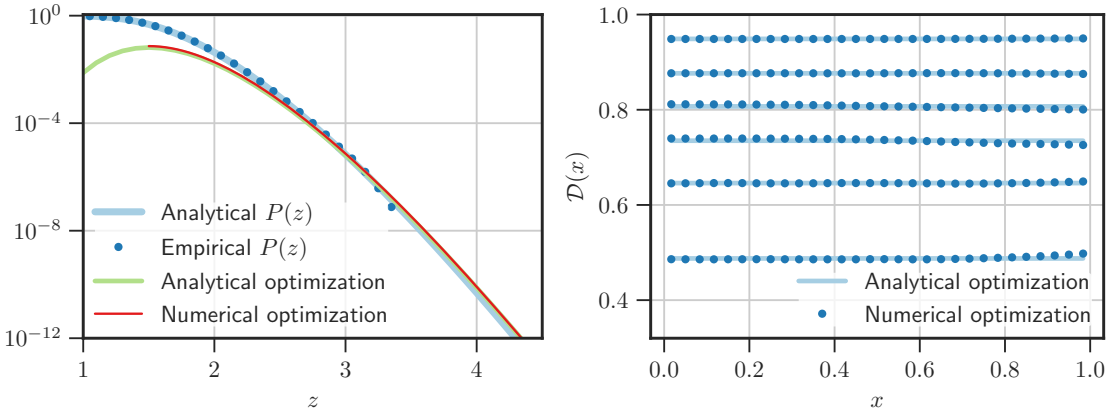

Figure 6.1. Linear forcing with $a = 0.1$, final time $T = 15$, initial conditions (6.22), and the statistical prior of *Case 1*. The numerics are performed with $M = N = 30$. Left panel: Comparison between the exact expression for $P(z)$ in (6.24), the empirical MC estimate with 2×10^7 samples, the analytical LDT estimate (6.29), and the LDT estimate obtained via numerical optimization. Right panel: Comparison between the analytical (6.28) and the numerical instantons, for $z = 1.58, 1.71, 1.85, 2.04, 2.32, 3.08$ from top to bottom.

In Fig. 6.1 we compare the asymptotic estimate (6.29) with the exact expression (6.24). We also check that the numerical optimization is consistent with the analytical one, which is important to validate the numerical code described below.

6.2.1 Relationship with Gärtner-Ellis

The result (6.29) can be interpreted in two different ways. In the sense that we intend it for, M is nothing but a constant included in the rate function, which has been fixed—say, the number of beads of the polymer— and does not have to be a large number. In this scenario, it is really the large size z of the events that we look at, and such largeness goes along with concentration of the events around the most likely realization. Knowing the rate function explicitly, we can equivalently

formulate the LDP as $P(z/\epsilon) \asymp \exp[\epsilon^{-1} I(\theta^\star(z/\epsilon))]$, as $\epsilon \to 0$, where now z is finite, and ϵ^{-2} is the speed of the LDP.

On the other hand, if one wishes to interpret (6.25) looking at the same observable but as the number of beads $M \to \infty$, then it is immediate to recognize (6.25) as an LDP for a sum of M i.i.d. random variables. In such case, the expression (6.25) divided by M is already the properly scaled rate function that one obtains from Gärtner-Ellis theorem, and the speed of that LDP is M, indeed.

6.3 Adjoint method optimization

We will minimize (6.18) using the adjoint method to compute the gradient. As shown below, the adjoint equations read

$$\partial_t^2 p_j = \frac{\mathcal{D}_{j+1}}{\Delta x^2}(p_{j+1} - p_j) - \frac{\mathcal{D}_j}{\Delta x^2}(p_j - p_{j-1})\,, \qquad j = 1, ..., N-1\,, \tag{6.31}$$

with conditions at the boundaries given by

$$p_0(t) = 0\,, \quad \partial_t^2 p_N = -\frac{\mathcal{D}_N}{\Delta x^2}(p_N - p_{N-1})\,, \tag{6.32}$$

and final conditions

$$p_j(T) = 0, \quad \partial_t p_j(T) = \lambda \delta_{j,N}\,. \tag{6.33}$$

The gradient of the cost function can be expressed as

$$\nabla_\theta E(u(\theta), \theta) = \nabla I(\theta) - G^\top \nabla \mathcal{D}(\theta)\,, \tag{6.34}$$

where $\nabla \mathcal{D}(\theta)$ is the $N \times M$ tensor with entries $\partial \mathcal{D}_j(\theta)/\partial \theta_k$, $j = 1, \dots, N$, $k = 1, \dots, M$, and G is a vector with entries

$$G_j = \int_0^T \frac{u_j - u_{j-1}}{\Delta x} \frac{\mu_j - \mu_{j-1}}{\Delta x}\, dt\,, \qquad j = 1, \dots, N\,. \tag{6.35}$$

6.3.1 Derivation of (6.31) and (6.34)

Using the convention that $\mathcal{D}_{N+1} = 0$, the evolution equation (6.15) can be rewritten as a system of first order ODEs,

$$\begin{cases} \partial_t u_j = v_j \\ \partial_t v_j = \dfrac{\mathcal{D}_{j+1}}{\Delta x^2}(u_{j+1} - u_j) - \dfrac{\mathcal{D}_j}{\Delta x^2}(u_j - u_{j-1}) + \delta_{j,N} \dfrac{r(t)}{\Delta x} \end{cases}, \quad j = 1, ..., N \tag{6.36}$$

with fixed boundary condition in the origin,

$$u_0(t) = 0\,, \tag{6.37}$$

and initial conditions

$$u_j(0) = 0, \quad v_j(0) = 0\,. \tag{6.38}$$

To make the notation compact, we will use:

$$X = \begin{pmatrix} u \\ v \end{pmatrix}, \quad Y = \begin{pmatrix} q \\ p \end{pmatrix}, \tag{6.39}$$

column vectors in $\mathbb{R}^{2N}$. Then, (6.36) can be written as

$$\partial_t X = b(X, \theta)\,, \tag{6.40}$$

where $b(X, \theta)$ is the $2N$-dimensional vector with the components of the RHS of (6.36). Note that (6.40) is in the general form (3.2) (linear system of ODEs), and this is helpful to make direct contact with the formulas (3.8) and (3.11), and thereby compute the gradient of the cost function (6.18) as

$$\nabla_\theta E = \nabla_\theta I(\theta) - \int_0^T (\partial_\theta b)^\top Y \, dt\,, \tag{6.41}$$

with Y the adjoint field to X. Let us start by deriving the adjoint equation. One can easily check that the linearization of the operator $b(X, \theta)$ for small variations of X reads

$$\begin{aligned} \partial_X b(\theta) &= \begin{pmatrix} 0 & \mathrm{Id} \\ B(\theta) & 0 \end{pmatrix}, \\ \text{with} \quad B_{jk} &= \frac{\mathcal{D}_{j+1}}{\Delta x^2}(\delta_{j+1,k} - \delta_{j,k}) - \frac{\mathcal{D}_j}{\Delta x^2}(\delta_{j,k} - \delta_{j-1,k}). \end{aligned} \tag{6.42}$$

Id is the $N \times N$ identity matrix and we recall that $\mathcal{D}_j = \mathcal{D}(\theta_j)$, by (6.10). It is the adjoint operator $(\partial_X b)^\top$ that we need to compute, defined implicitly by the identity

$$\left\langle (\partial_X b)^\top Y, X' \right\rangle_{\mathbb{R}^{2N}} = \langle Y, \partial_X b\, X' \rangle_{\mathbb{R}^{2N}}\,, \tag{6.43}$$

where $\langle \cdot, \cdot \rangle_{\mathbb{R}^{2N}}$ denotes the standard scalar product in $\mathbb{R}^{2N}$. Using (6.43) we obtain,

$$\begin{aligned} \langle Y, \partial_X b\, X' \rangle_{\mathbb{R}^{2N}} &= \sum_{j=1}^N \left(q_j v_j' + p_j \left(\frac{\mathcal{D}_{j+1}}{\Delta x^2}(u_{j+1}' - u_j') - \frac{\mathcal{D}_j}{\Delta x^2}(u_j' - u_{j-1}') \right) \right) \\ &= \sum_{j=1}^N \left(q_j v_j' + \left(\frac{\mathcal{D}_{j+1}}{\Delta x^2}(p_{j+1} - p_j) - \frac{\mathcal{D}_j}{\Delta x^2}(p_j - p_{j-1}) \right) u_j' \right)\,, \end{aligned} \tag{6.44}$$

where in the last passage we just reorganized the indices in the sum in an equivalent way, provided that we assume the boundary condition

$$p_0(t) = 0\,. \tag{6.45}$$

Comparing the last line of (6.44) with the LHS of (6.43), we deduce that

$$(\partial_X b)^\top = \begin{pmatrix} 0 & B(\theta) \\ \mathrm{Id} & 0 \end{pmatrix} \tag{6.46}$$

which is the transpose of the RHS of (6.42) ($B(\theta)$ is symmetric), as we should expect. Though, starting from the identity (6.43) is the rigorous way to obtain the adjoint operator, making the proper boundary conditions arise naturally. Plugging the result (6.46) into (3.8), we finally obtain the adjoint equation

$$\begin{cases} \partial_t q_j = \dfrac{\mathcal{D}_{j+1}}{\Delta x^2}(p_{j+1} - p_j) - \dfrac{\mathcal{D}_j}{\Delta x^2}(p_j - p_{j-1}) \\ \partial_t p_j = q_j \end{cases}, \quad j = 1, ..., N\,, \tag{6.47}$$

with boundary condition (6.45). To obtain the correct conditions at final time, it is sufficient to observe that the final conditions of (3.8) now read

$$q_j(T) = \lambda \partial_{u_j} f(u(T)) = \lambda \delta_{j,N}\,, \qquad p_j(T) = 0. \tag{6.48}$$

Let us now compute $(\partial_\theta b)^\top$, again starting from the definition of the adjoint operator:

$$\left\langle (\partial_\theta b)^\top Y, w \right\rangle_{\mathbb{R}^N} = \langle Y, \partial_\theta b\, w \rangle_{\mathbb{R}^{2N}}\,, \tag{6.49}$$

where $w \in \mathbb{R}^N$ and

$$\begin{aligned} (\partial_\theta b) &= \begin{pmatrix} 0 \\ \nabla_\theta B(\theta) \end{pmatrix} \quad \text{(two } N \times N \text{ blocks)} \\ (\nabla_\theta B)_{jk} &= \frac{\mathcal{D}'(\theta_{j+1})}{\Delta x^2}(u_{j+1} - u_j)\delta_{j+1,k} - \frac{\mathcal{D}'(\theta_j)}{\Delta x^2}(u_j - u_{j-1})\delta_{j,k}. \end{aligned} \tag{6.50}$$

With the convention that $\mathcal{D}'(\theta_{N+1} = 0)$, a straightforward calculation yields

$$\begin{aligned} \langle Y, \partial_\theta b\, w \rangle_{\mathbb{R}^{2N}} &= \sum_{j=1}^{N} p_j \left(\frac{\mathcal{D}'(\theta_{j+1})}{\Delta x^2}(u_{j+1} - u_j) w_{j+1} - \frac{\mathcal{D}'(\theta_j)}{\Delta x^2}(u_j - u_{j-1}) w_j \right) \\ &= \sum_{j=1}^{N} \left(\frac{\mathcal{D}'(\theta_j)}{\Delta x^2}(u_j - u_{j-1})(p_j - p_{j-1}) \right) w_j\,, \end{aligned} \tag{6.51}$$

from which, comparing with the LHS of (6.49), we observe that

$$((\partial_\theta b)^T Y)_j = \mathcal{D}'(\theta_j) \frac{u_j - u_{j-1}}{\Delta x} \frac{p_j - p_{j-1}}{\Delta x}\,. \tag{6.52}$$

Now, integrating in time according to (6.41),

$$\int_0^T ((\partial_\theta b)^\top Y)_j dt = \mathcal{D}'(\theta_j) \int_0^T \frac{u_j - u_{j-1}}{\Delta x} \frac{p_j - p_{j-1}}{\Delta x} dt\,, \tag{6.53}$$

leads to (6.34).

6.4 Nonlinear forcing

6.4.1 Nonlinear protocols and work normalization

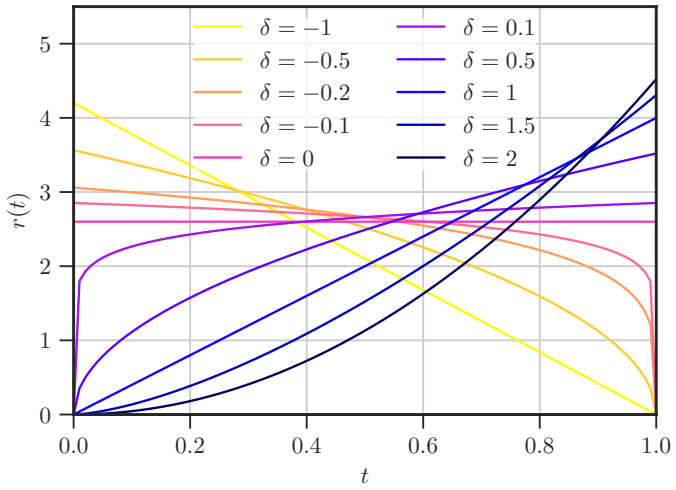

Figure 6.2. The forcing protocols $r_\delta(t)$ in (6.54), which are decreasing functions of t when $\delta < 0$ and increasing functions when $\delta > 0$.

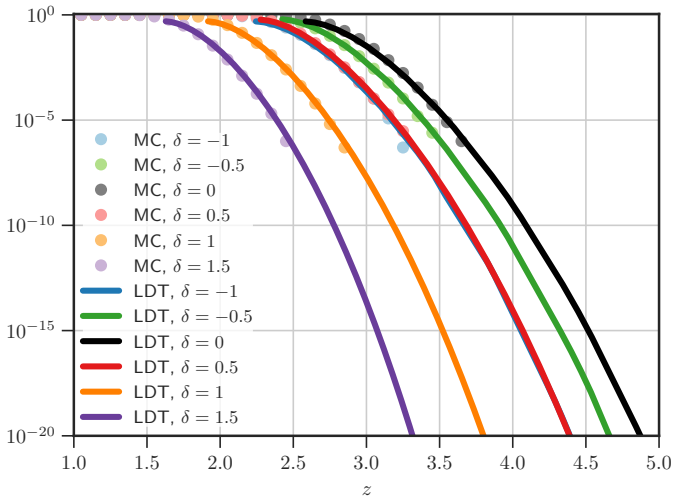

Figure 6.3. Comparison between the empirical distributions $P(z)$ obtained via MC sampling and their LDT estimate. The sampling works down to events whose probability is about the inverse of the MC sampling size, while the LDT optimization allows us to extend the tails to much smaller probabilities.

Next we consider nonlinear forcing protocols of the type

$$r(t) = at^\beta \qquad \text{and} \qquad r(t) = a(T-t)^\beta \qquad \text{both with} \quad a, \beta > 0\,. \tag{6.54}$$

Letting $s = +1$ if $r(t) = at^\beta$ and $s = -1$ if $r(t) = a(T-t)^\beta$, we will use $r_\delta(t)$ with $\delta = s\beta$ as shorthand to describe the family of forcing protocols, which is continuous in the parameter δ: notice that when $\beta \to 0$ both increasing and decreasing protocols tend to the constant protocol, $\delta = 0$. The different forcings are shown in Fig. 6.2.

As initial conditions for (6.2) we take

$$u(0,x) = 0\,, \quad \partial_t u(0,x) = 0\,, \quad \forall x \in [0,1]\,. \tag{6.55}$$

At discrete level these initial conditions read

$$u_j(0) = 0, \qquad \partial_t u_j(0) = 0\,. \tag{6.56}$$

In this section we restrict ourselves to *Case 2* and we use $M = N = 30$ and final time $T = 1$. Observing that the mean elasticity $\mathbb{E}\,(\mathcal{D}(x)) = 1$ (as for *Case 1*), the average velocity of propagation of the waves along the bar is also 1. Thus, 1 is the average time that a signal takes to propagate from the right end to the left end. This means that taking $T = 1$ we are considering a short transient strongly out of equilibrium, where the random structure will contribute in a non-homogeneous way.

To integrate (6.15) and (6.16) numerically, we use a velocity-Verlet integrator, which is of second order, symplectic, and time reversible, with a time step of 10^{-3}. The optimization is performed as described in Sec. 3, using (6.34) and (6.35).

We remind the reader that the meaning of $r(t)$ in (6.1) is that of external force applied to the right end of the rod. The work of the external forcing labelled by δ on the system is therefore defined by

$$W(\theta,\delta) = \int_0^T r_\delta(t)\dot{u}(t,1,\theta)dt\,, \tag{6.57}$$

with dependence on the elasticity structure (θ) and the forcing (δ). Considering that in the LDT regime the optimal realizations (instantons) dominate the statistics of the extreme events (this is going to be checked below), we define the work restricted to the optimal events by

$$W_o(z,\delta) = W(\theta^\star(z),\delta)\,, \tag{6.58}$$

where we recall that $\theta^\star(z)$ is the solution of the minimization

$$\min_\theta I(\theta) \qquad \text{subject to} \qquad u(T,1,\theta) \geq z\,. \tag{6.59}$$

We normalize the factor a of each protocol (thus becoming an $a(\delta)$) so that the curves $W_o(z)|_\delta$ lay on top of each other independent of δ (up to small deviations) in the range of interest of the extreme events. This means that under this normalization any given amount of work done on the system leads the instanton configurations to the same final extension, for all protocols. This will turn out useful to compare the different protocols from an optimality viewpoint.

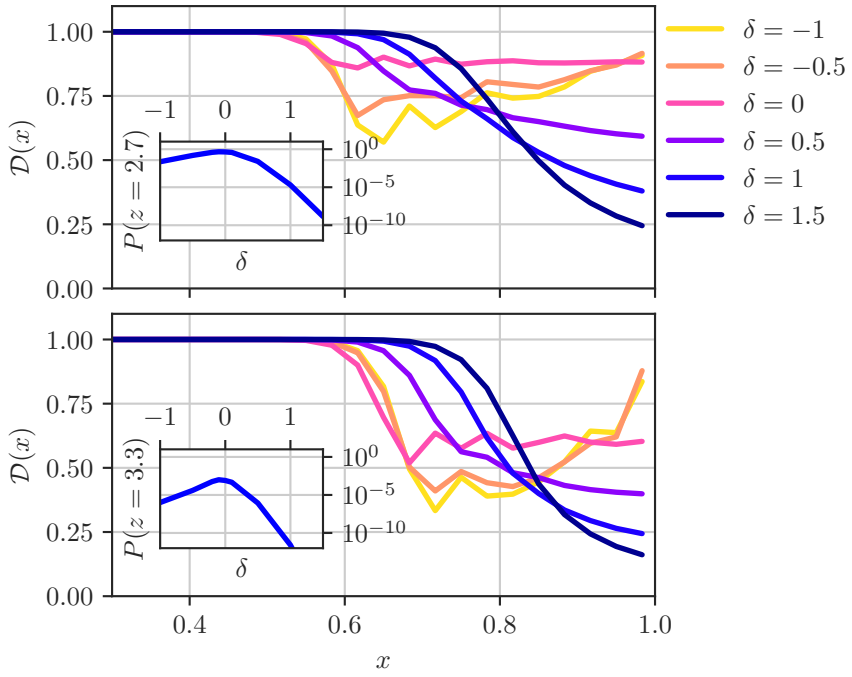

Figure 6.4. Top panel: Elasticity structure of the instantons for $z \geq 2.7$, for the different protocols labeled by δ. Inset: the probability $P(z = 2.7)$ as a function of the forcing protocol. Bottom panel: Same as in the top panel, but for $z \geq 3.3$. Inset: the probability $P(z = 3.3)$ as a function of δ in the forcing protocol.

6.4.2 Instantons and optimal protocol

Let us now describe our results. In Fig. 6.3 the LDT estimates of $P(z)$ are compared to the empirical estimates obtained via MC with 2×10^6 samples, showing good agreement. Next we look at the specific elasticity structure of the optimizers, $\mathcal{D}(x, \theta^\star(z))$. These are shown in Fig. 6.4. As can be seen, the region that is relevant for having an extreme extension $u(T = 1{,}1)$ occupies only the right half of the space domain, independent of the protocol. This makes sense since on average the signal takes a time 1 to cross the whole domain: For a point x_0 to influence $u(T = 1{,}1)$ the signal needs to have time to propagate to $x = 1$. As a result, the points on the left side will not have the possibility to influence the dynamics at all, and the optimal state of $\mathcal{D}(x, \theta)$ is determined by mere minimization of $I(\theta)$ with no dynamical constraint. In contrast, on the right side of the domain, $\mathcal{D}(x, \theta)$ must take low values to allow for large values of $u(T = 1{,}1)$—since these low values are unlikely, this also accounts for the drop in probability observed in Fig. 6.3. Fig. 6.4 also indicates that $\mathcal{D}(x, \theta^\star(z))$ depends on the forcing protocol. This dependency can again be interpreted intuitively by realizing that the region that impacts $u(T{,}1)$

the most will be the one that is reached by a strong signal (i.e. the propagation front of the most intense part of the forcing) and is able to send a strong feedback back to the right end at final time—this feedback is what is accounted for by the backward evolution of the adjoint equation in the optimization. So, the earlier the most intense part of the forcing takes place, the further from the right end a low elasticity peak appears. This explains why going towards negative δ the low-elasticity peak moves to the left in Fig. 6.4, and the constant forcing ($\delta = 0$) is the one where the low elasticity contribution is the most uniformly distributed.

Note that in this framework it is possible to compare how likely the protocols are to produce extreme realizations of a given size, as shown in the insets in Fig. 6.4. In this sense, the constant protocol appears to be the optimal one. This is consistent with the fact that $\delta = 0$ is the highest curve in Fig. 6.3. Furthermore, not only we gain quantitative information on which protocol is the most suitable to sample extreme events, but also we have access to the peculiar shape common to all the extreme realizations. For a given statistical prior, such specific shape depends only on the protocol (δ) and on the event size (z), and it determines the "instantons" of the problem, like the ones depicted in Fig. 6.4.

6.4.3 Instantons "filtered" out of the random sampling

To further clarify the role of the instantons and why they dominate the dynamics and the statistics of the extreme events, it is useful to "filter" the conditional events such that $u(T,1) \geq z$ in the following way: First, we fix a size z and generate via MC a large set of ϑ such that $u(T,1,\vartheta) \geq z$. Second, we average over such conditional set to obtain the mean conditional event and its fluctuations around the mean, which is generally very close to the instanton $\theta^\star(z)$. Third, we decompose the fluctuations $\vartheta - \theta^\star(z)$ into the components parallel and perpendicular to $\eta^\star(z)$, i.e. the normal to the hypersurface $\Omega(z)$. This procedure is then repeated for various z.

In Fig. 6.5 we show the outcome of this analysis for the protocol with $\delta = 1.5$ and for two different values of z – analogous results hold for the other kinds of forcing as well. As can be seen the average event $u(T,1) \geq z$ lies on top of the instanton $\theta^\star(z)$, with fluctuations independent of the size of the event and also of the position along the rod (upper panels). The decomposition shows that the components perpendicular to $\eta^\star(z)$ are independent of the size of the event, and basically independent of the dynamics too. Their mean and standard deviation are the mean and the standard deviation of the unconstrained random variables ϑ (central panels). In contrast, the parallel fluctuations are small and tend to zero as z increases (bottom panels). The scaling of the fluctuations is analyzed in more detail in Fig. 6.6, which shows that they are $O(1)$ in the direction perpendicular to $\eta^\star(z)$ and $O(|\eta^\star(z)|^{-1})$ in the direction parallel to it, consistent with the theoretical predictions of section 2.2.1.

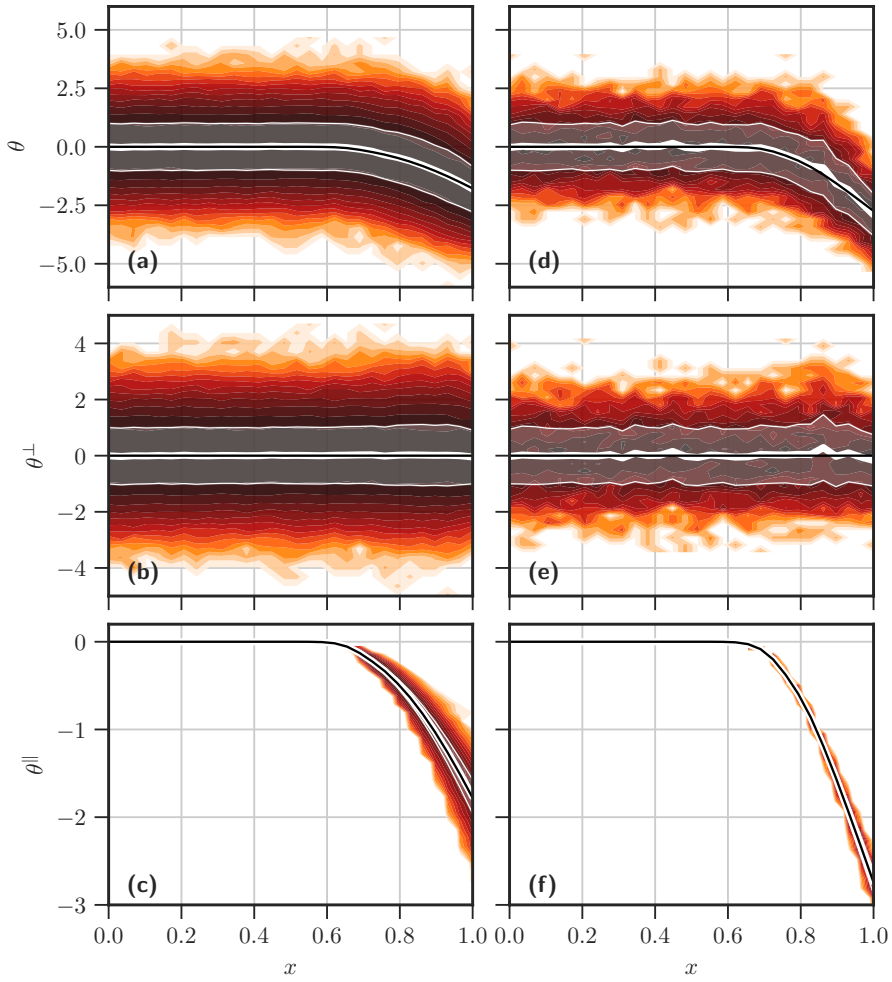

Figure 6.5. Comparison between the instanton $\theta^\star(z)$ (black solid line) and the Monte Carlo sampling on the distribution of ϑ, conditioned on $u_N(1) = u(1,1) \geq z$ (color map with intensity proportional to the empirical probability density; thick white line = mean; thin white lines = 1 standard deviation range around the mean). Left panels: $z = 2.10$, right panels: $z = 2.40$. The top panels show the full data: the instanton agrees with the mean, but the variance does not substantially change going to more extreme events. The two central panels show the fluctuations perpendicular to $\eta^\star(z)$, confirming that their amplitude is independent of the size of the event (left and right panels have the same variance) and homogeneous in space. The bottom panels show the fluctuations in the direction parallel to $\eta^\star(z)$, indicating that their amplitude decreases as z increases, as predicted by the theory in Sec. 2.2.1.

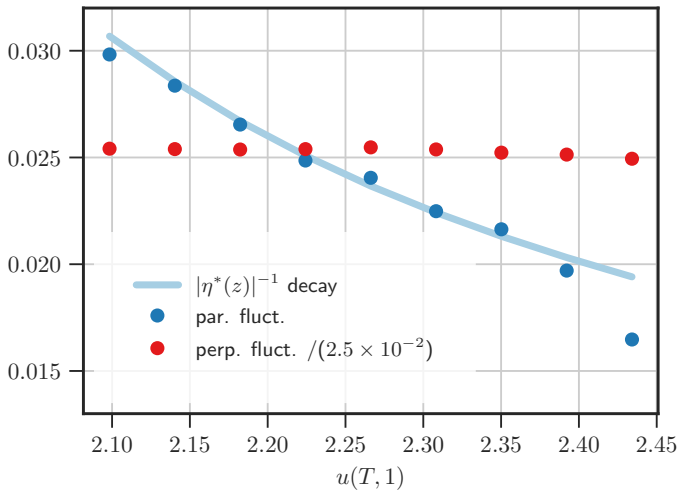

Figure 6.6. Increasing z, the fluctuations in the direction perpendicular to $\eta^\star(z)$ stay constant, whereas in the parallel direction they scale as $O(|\eta^\star(z)|^{-1})$. Both behaviors are predicted analytically and here confirmed numerically.

Chapter 7

Concluding remarks

After having analyzed thoroughly some relevant physical applications of the LDP (2.36), it is worth commenting further on some aspects that hopefully will clarify the state of the art and possibly indicate future directions.

1. The philosophy of the proof of the theorem (2.36) follows the Gärtner-Ellis theorem: the upper bound makes use of the Tchebycheff inequality and the lower bound needs differentiability of the cumulant generating function, implying convexity of $I(\theta)$ by definition (2.7). Indeed, since we consider the limit $z \to \infty$ rather than recalling the random variable itself, the proof needs to be adapted. More precisely, such reformulation is carried out in analogy with the dominating point results of Ney [143]. The example in section 2.2.4 shows that (2.36) amounts to taking the Legendre-Fenchel transform of the differentiable cumulant generating function (with domain the $N-$dimensional support of the random parameters) and then computing the rate function from it via a constrained minimization. This procedure leads to the correct rate function, whereas the Gärtner-Ellis theorem does not apply here (as explained in section $IV.D$ of [187]) being the scaled cumulant generating function (defined differently, the domain being the support of the scalar observable) non-differentiable.

2. From a mathematical viewpoint, what we do is not a very hard derivation either, since the five assumptions that are made are most of what is needed for the result (2.7), i.e. they are very general conditions ensuring a dominating-point situation. Furthermore, our mathematical assumptions are not immediate to test on a given dynamics and a choice of an observable, although they are related to them. What would be useful, for the applications, is to know what classes of dynamics and observables do satisfy the LDP (2.7), and which ones do not. For example, for dynamics characterized by fast transients due to instabilities, in the sense of [133, 77], a proper observable must be instantaneous or involve a time average on a short time

window. There, a long-time average would not surely satisfy a dominating-point LDP, since there would be a degenerate situation with instabilities at different points of the time window, all contributing to the same value of the observable.

In a quite tautological statement, the answer is that (2.7) holds whenever the statistical prior $I(\theta)$, the dynamics of the field $u(t)$ and the observable $f(u)$ are combined in such a way that the onset of a dominating-point regime is observed, when the size of f is conditioned to be large.

3. The result (2.7) is unusual as an LDP, as the speed of the LDP is implicitly contained within the non-scaled rate function. We refer the reader to the book [12], where similar ideas are developped in the formalism of the *movable saddle point*. In the examples in the sections 2.2.2, 2.2.4 and 6.2, we have shown how our result is equivalent to the usual LDP, via a rescaling by a small parameter ϵ. On the other hand, these remain very simple cases, while the numerical tools of our LDT method are specifically designed for cases where the mapping between the random parameters and the observable is extremely complicated, possibly involving a nonlinear dynamics in a very high-dimensional setting. For such cases, we have shown through all of the applications in chapters 4, 5 and 6 that the LDP (2.7) is directly applicable. Moreover, the prediction on the decay of the fluctuations in the direction parallel to $\eta^\star(z)$ ($O(|\eta^\star(z)|^{-1})$ is a useful quantification of the speed of concentration around the instanton realization, that is, a quantification of how fast the asymptotic behavior (2.7) is attained as $z \to \infty$. For the applications, this test can be applied a-posteriori to check whether the LDP dominating-point concentration is satisfied, as exemplified in Fig. 5.4. For instance, in the case of the elastic rod this was verified in detail in Figs. 6.5 and 6.6. For the rogue waves, the analysis is more complicated, although in Figs. 4.4, 4.9 and 5.3 it is clear that the variance of the conditional extreme event becomes small in the focusing region (somehow "parallel" to the instanton) while it stays $O(1)$ where the instanton is vanishing ("perpendicular" to the instanton), where parallel and perpendicular are now interpreted in the sense of the L^2 scalar product.

Appendix A

Derivation of the leading order equations of wave turbulence

The material in this chapter was published in

S. Chibbaro, G. Dematteis, and L. Rondoni. "4-wave dynamics in kinetic wave turbulence". Physica D: Nonlinear Phenomena, 362:24–59, 2018.

A.1 Introduction

Wave Turbulence (WT) theory concerns the dynamics of dispersive waves that interact nonlinearly over a wide range of scales [137]. In general the nonlinear interaction can be considered small, allowing a perturbative analysis and then an asymptotic closure for statistical observables [142]. For this reason, sometimes one then talks about Weak Wave Turbulence (WWT). Until recently, most of the attention was given to the energy spectrum, which is governed by a kinetic equation. Wave turbulence also provides exact solutions of the kinetic equation, which are related to equipartition, Rayleigh-Jeans solution, or stationary cascade, Kolmogorov-Zakharov solutions [206]. Many physical phenomena are studied within this general framework, for instance gravity [111, 74, 146, 138], capillary or Alfvèn waves [204, 163, 71, 85], non-linear optics [158] and elastic plates [54, 22, 135]. Furthermore, applications of WT to non dispersive systems such as the acoustic waves [208, 128] exist, even though the necessary statistical closure is subtler in such cases [140, 173].

In the last years, many experiments and numerical simulations were performed to verify the predictions of WT. The picture is relatively clear in the case of the capillary waves on a fluid surface (water, ethanol, liquid hydrogen or liquid helium):

both experiments and numerical simulations confirm the Kolmogorov-Zakharov spectrum predicted by WT in this case. For other cases, *e.g.* surface gravity waves or waves in vibrating elastic plates, the picture is more complicated: both numerics and experiments showed deviations from theoretical predictions, and the presence of intermittency [81, 57, 75, 130]. This was unexpected, since WT appears as a mean-field theory, based on an initial "quasi-gaussianity", previously believed to prevent sensible deviations from gaussianity.

An important step forward in this context has been the development of a more efficient formalism for non-gaussian wavefields [37, 38, 103, 137]. In particular, these works pointed out that probability density functions (PDF) are the relevant statistical objects to be analysed, reviving the interest in the study of PDFs in WT, that dates back to the works of Peierls, Brout, Prigogine, Zawlaski and Sagdeev [155, 24, 209]. These authors had considered waves in anharmonic crystals, which constitute a special case of 3-wave systems. In the recent developments a diagrammatic approach was proposed [137], based on Zakharov's pioneering work [134, 205], to analytically investigate PDF equations. Importantly, this has also clarified the role of the different assumptions needed for the statistical closure. In particular, the 3-wave resonant systems has been studied in details and a Peierls equation for the N-particles PDF has been proposed [37, 38, 137].

Nevertheless, the Peierls equation does not guarantee the strict preservation of the independence of phases and amplitudes, even though it can be argued that the property of *random phases and amplitudes* (RPA) is preserved in a weaker form [39, 137]. Starting from these premises, it has been shown that a proper normalization of the wave amplitudes is necessary for 3-wave resonant systems, in order to obtain a finite spectrum in the infinite-box limit, that leads to an amplitude density, dependent on the continuous variable $\mathbf{k}$ [69]. In particular, the original amplitudes must be normalized by a factor scaling as $1/V$, where V is the volume of the box. Adopting such a point of view, the Peierls equation for the multimode PDFs is not the leading-order asymptotic equation of the continuum limit of weakly interacting, incoherent waves. In Ref. [69], then, new multimode equations were derived, that importantly have the factorized exponential solutions excluded by the Peierls equation. This is equivalent to the preservation of the RPA property. In turn, the preservation of exponential solutions implies a law of large numbers (LLN) for the empirical spectrum at times $\tau > 0$, which is analogous to the propagation of chaos of the BBGKY hierarchy in the kinetic theory of gases. This LLN implies that the empirical spectrum satisfies the wave-kinetic closure equations for nearly every initial realization of random phases and amplitudes, without necessity of averaging. Just as the Boltzmann hierarchy has factorized solutions for factorized initial conditions, so does the kinetic wave hierarchy for all multi-point spectral correlation functions. An H-theorem corresponding to positive entropy variation holds as well. On the other hand, using these multimode equations, Ref. [69] shows that the 1-mode PDF equations is not altered by the different normalization, if the

modes initially enjoy the RPA property.

The 4-wave case has not yet been dealt with, although a formal analogy has been used to propose a possible extension of the 3-wave result to the 4-wave case [39]. Therefore, the present derivation is devoted to the case of 4-wave interactions, which is of particular interest. As a matter of fact, most of the known violations of gaussianity arise in gravity waves and in vibrating elastic plates, which are 4-wave resonant systems. Following the same diagrammatic approach of Ref. [137], and using the normalization proposed in Ref. [69], we first explicitly derive the continuos multimode equations, and then we obtain the equation for the M-mode PDF equation. These equations are different from the Peierls equations obtained by the formal analogy of Ref. [39]; they constitute instead a direct extension of the 3-wave case treated in Ref. [69]. The relation between the Peierls and our equations is thus discussed, showing the limit in which they coincide. Our framework also sheds some light on the issue of WT intermittency, as demonstrated by [32], in which the equations obtained here are confirmed by numerical simulations of two 4-wave resonant Hamiltonian systems.

The derivation is organized as follows. First, we describe our model and notation, which are consistent with previous works [137, 69]. Section A.2 discusses the probabilistic properties of RPA fields. The main results are reported in sections A.3 and A.4, where the multimode equations are derived and discussed. In Section A.3 the spectral generating functional and correlation functions are considered, while Section A.4 concerns the PDF generating function and the multipoint PDFs. Section A.5 summarizes our results.

A.1.1 Model and notation

Similarly to [69], we consider a complex wavefield $u(\mathbf{x}, t)$ in a d-dimensional periodic cube with side L. This field is a linear combination of the canonical coordinates and momenta. It is assumed that there is a maximum wavenumber $k_{\max}$, to avoid ultraviolet divergences. This can be achieved by a lattice regularization with spacing $a = L/M$, for some large integer M, so that $k_{\max} = \pi/a$. The location variable x then ranges over the physical space

$$\Lambda_L = a\mathbb{Z}_M^d, \tag{A.1}$$

with the usual notation $\mathbb{Z}_M$ for the field of integers, modulo M. This space has volume $V = L^d$. The dual space of wavenumbers is

$$\Lambda_L^* = \frac{2\pi}{L}\mathbb{Z}_M^d \tag{A.2}$$

with $k_{\min} = 2\pi/L$. The total number of modes is $N = M^d$, so that $V = Na^d$. The following index notation will be used:

$$u^\sigma(\mathbf{x}) = \begin{cases} u(\mathbf{x}) & \sigma = +1 \\ u^*(\mathbf{x}) & \sigma = -1 \end{cases} \tag{A.3}$$

for u and its complex-conjugate u^*. Likewise, we adopt the convention for (discrete) Fourier transform:

$$A^\sigma(\mathbf{k}) = \frac{1}{N} \sum_{\mathbf{x}\in\Lambda_L} u^\sigma(\mathbf{x},t) \exp(-i\sigma\mathbf{k}\cdot\mathbf{x}) \tag{A.4}$$

so that $A^+(\mathbf{k})$ and $A^-(\mathbf{k})$ are complex conjugates. This quantity converges to the continuous Fourier transform $\frac{1}{L^d}\int_{[0,L]^d} d^dx\ u^\sigma(\mathbf{x},t)\exp(-i\sigma\mathbf{k}\cdot\mathbf{x})$ in the limit $a\to 0$. The discrete inverse transform is

$$u^\sigma(\mathbf{x}) = \sum_{\mathbf{k}\in\Lambda_L^*} A^\sigma(\mathbf{k}) \exp(i\sigma\mathbf{k}\cdot\mathbf{x}). \tag{A.5}$$

The dynamics is assumed to be canonical Hamiltonian, with a 4^{th} power term in the Hamiltonian density (energy per volume) describing 4-wave interactions. As in [206] and with lattice regularization, we write:

$$H = H_0 + \delta H\ , \quad H_0 = \frac{1}{2}\sum_{\mathbf{k}\in\Lambda^*} \omega_{\mathbf{k}} |A_{\mathbf{k}}|^2 \tag{A.6}$$

Taking the most general Hamiltonian with any kind of 4-wave interactions, [206], one can write $\delta\mathcal{H}$ in the symmetrized compact form:

$$\delta H = \epsilon \sum_{1234} \mathcal{H}_{1234}^{\sigma_1\sigma_2\sigma_3\sigma_4}\ A_1^{\sigma_1} A_2^{\sigma_2} A_3^{\sigma_3} A_4^{\sigma_4}\ \delta_{1234} \tag{A.7}$$

with the coefficients satisfying the general relations:

$$(\mathcal{H}_{1234}^{\sigma_1\sigma_2\sigma_3\sigma_4})^* = \mathcal{H}_{1234}^{-\sigma_1-\sigma_2-\sigma_3-\sigma_4}\ , \quad \mathcal{H}_{1234}^{\sigma_1\sigma_2\sigma_3\sigma_4} = \mathcal{H}_{\Pi(1234)}^{\Pi(\sigma_1\sigma_2\sigma_3\sigma_4)}. \tag{A.8}$$

$\Pi \in S^4$ represents any permutation of the four elements. Introducing further notation:

$$\begin{aligned} &\underline{\sigma} \doteq (\sigma_1,\sigma_2,\sigma_3,\sigma_4)\ , \qquad \underline{\mathbf{k}} \doteq (\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4)\ , \qquad \delta_{\underline{\sigma}\cdot\underline{\mathbf{k}},\mathbf{0}} = \delta_{\sigma_1\mathbf{k}_1+\sigma_2\mathbf{k}_2+\sigma_3\mathbf{k}_3+\sigma_4\mathbf{k}_4,\mathbf{0}} \\ &\omega_1 \doteq \omega(\mathbf{k}_1)\ , \qquad A_1 \doteq A(\mathbf{k}_1)\ , \qquad \sum_1 \doteq \sum_{\sigma_1=\pm1}\sum_{\mathbf{k}_1\in\Lambda^*} \end{aligned} \tag{A.9}$$

the Hamiltonian can be written as:

$$H = \frac{1}{2}\sum_1 \omega_1 A_1^{\sigma_1} A_1^{-\sigma_1} + \epsilon \sum_{1234} \mathcal{H}_{\underline{\mathbf{k}}}^{\underline{\sigma}}\ A_1^{\sigma_1} A_2^{\sigma_2} A_3^{\sigma_3} A_4^{\sigma_4}\ \delta_{\underline{\sigma}\cdot\underline{\mathbf{k}},\mathbf{0}} \tag{A.10}$$

which leads to

$$\frac{\partial A_{\mathbf{k}}^\sigma}{\partial t} = i\sigma\omega_{\mathbf{k}} A_{\mathbf{k}}^\sigma + \epsilon \sum_{234} \mathcal{L}_{\mathbf{k}234}^{\sigma\sigma_2\sigma_3\sigma_4} A_2^{\sigma_2} A_3^{\sigma_3} A_4^{\sigma_4} \delta_{-\mathbf{k}+\sigma_2\mathbf{k}_2+\sigma_3\mathbf{k}_3+\sigma_4\mathbf{k}_4,\mathbf{0}} \tag{A.11}$$

where

$$\mathcal{L}^{\sigma\sigma_2\sigma_3\sigma_4}_{\mathbf{k}234} \doteq 4i\sigma\mathcal{H}^{(-\sigma)\sigma_2\sigma_3\sigma_4}_{\mathbf{k}234}. \tag{A.12}$$

Changing $\mathbf{k} \to \mathbf{k}_1$ and introducing the interaction representation[1] $A^\sigma_{\mathbf{k}} = a^\sigma_{\mathbf{k}} e^{i\sigma\omega_{\mathbf{k}}t}$, one obtains[2]:

$$\begin{aligned}\frac{\partial a_1}{\partial t} &= \epsilon \sum_{234} \mathcal{L}^{+\sigma_2\sigma_3\sigma_4}_{1234} a_2^{\sigma_2} a_3^{\sigma_3} a_4^{\sigma_4} \\ &\quad \times \exp\left[i\left(-\omega_1 + \sigma_2\omega_2 + \sigma_3\omega_3 + \sigma_4\omega_4\right)t\right] \delta_{-\mathbf{k}_1+\sigma_2\mathbf{k}_2+\sigma_3\mathbf{k}_3+\sigma_4\mathbf{k}_4,\mathbf{0}}\end{aligned} \tag{A.13}$$

With notation [69]:

$$\begin{aligned}\mathcal{L}_{1234} &\doteq \mathcal{L}^{+\sigma_2\sigma_3\sigma_4}_{1234}\ , \qquad \omega^1_{234} \doteq -\sigma_1\omega_1 + \sigma_2\omega_2 + \sigma_3\omega_3 + \sigma_4\omega_4 \\ \delta^1_{234} &\doteq \delta_{-\sigma_1\mathbf{k}_1+\sigma_2\mathbf{k}_2+\sigma_3\mathbf{k}_3+\sigma_4\mathbf{k}_4,\mathbf{0}}\end{aligned} \tag{A.14}$$

the dynamical equation of motion with 4-wave interactions now reads:

$$\dot{a}_1 = \epsilon \sum_{234} \mathcal{L}_{1234} a_2^{\sigma_2} a_3^{\sigma_3} a_4^{\sigma_4} \exp\left(i\omega^1_{234}t\right) \delta^1_{234} \tag{A.15}$$

A.2 Fields with random phases and amplitudes

In derivations of wave kinetic equations, it is often assumed that initial fields have Fourier coefficients with random statistically independent phases and amplitudes (RPA). This property is expected to be preserved in time, in some suitable sense, in the wave-kinetic limit.

Let N complex-valued random variables $a_{\mathbf{k}}$, $\mathbf{k} \in \Lambda^*_L$ be the Fourier coefficients of a random field:

$$u_L(\mathbf{x}) = \sum_{\mathbf{k}\in\Lambda^*_L} a_{\mathbf{k}} \exp(i\mathbf{k}\cdot\mathbf{x}). \tag{A.16}$$

Here $a_{\mathbf{k}}$ corresponds to $a^+_{\mathbf{k}}$, i.e. $A^+_{\mathbf{k}}$ in the previous section (no distinction need be made between the two at time $t = 0$). It will be crucial in the following to work with normalized variables

$$\widetilde{a}_{\mathbf{k}} = \left(\frac{L}{2\pi}\right)^{d/2} a_{\mathbf{k}} \tag{A.17}$$

which are assumed to remain finite in the large-box limit $L \to \infty$. This normalization is sufficient for the spectrum of the random field to be well defined in that

[1] Such a representation eliminates the fast linear oscillations, giving a variable $a^\sigma_{\mathbf{k}}$ that does not oscillate on fast scales.

[2] In our derivation, for simplicity and no loss of generality, we consider $\sigma = +1$. Trivially, the equations with $\sigma = -1$ are redundant, because obtained by complex conjugation of the ones with $\sigma = +1$. From now a_1 stands for a^+_1.

limit, as first pointed out in [69]. It is convenient to write the complex variables in polar coordinates (action-angle variables, or amplitudes and phases)

$$a_{\mathbf{k}} = \sqrt{J_{\mathbf{k}}} e^{i\varphi_{\mathbf{k}}} = \sqrt{J_{\mathbf{k}}}\psi_{\mathbf{k}} \tag{A.18}$$

with normalized action defined by

$$\tilde{J}_{\mathbf{k}} = \left(\frac{L}{2\pi}\right)^d J_{\mathbf{k}}. \tag{A.19}$$

We denote by $s_{\mathbf{k}}$ and $\xi_{\mathbf{k}}$ for possible values of the random variables $\tilde{J}_{\mathbf{k}} \in \mathbb{R}^+$ and $\psi_{\mathbf{k}} = e^{i\varphi_{\mathbf{k}}} \in S^1$.

$$d\mu(s,\xi) = \prod_{\mathbf{k}\in\Lambda_L^*} ds_{\mathbf{k}} \frac{|d\xi_{\mathbf{k}}|}{2\pi} \tag{A.20}$$

suitably normalized. The N-mode joint probability density function $\mathcal{P}^{(N)}(s,\xi)$ is defined with respect to the Liouville measure, such that the average of the random variable $f_{\tilde{J}_\psi}(s,\xi)$ is given by

$$\langle f_{\tilde{J}_\psi}\rangle = \int d\mu(s,\xi)\, \mathcal{P}^{(N)}(s,\xi) f(s,\xi) \tag{A.21}$$

where the integral is over (s,ξ) in the product space $(\mathbb{R}^+)^N \times (S^1)^N$.

The field $u_L(\mathbf{x})$ is called a *random-phase field* (RP) if for all $\mathbf{k} \in \Lambda_L^*$ the $\psi_{\mathbf{k}} = e^{i\varphi_{\mathbf{k}}}$ are independent and identically distributed (i.i.d.) random variables, uniformly distributed over the unit circle S^1 in the complex plane [137]. For the joint PDF, this is equivalent to:

$$\mathcal{P}^{(N)}(s,\xi) = \mathcal{P}^{(N)}(s) \tag{A.22}$$

Note that an RP $u_L(\mathbf{x})$ is a homogeneous random field on Λ_L, statistically invariant under space-translations by the finite group $a\mathbb{Z}_M^d$. In the limit $L \to \infty$ the field $u_L(\mathbf{x})$ defined with appropriately chosen $\tilde{J}_{\mathbf{k},L}$ will converge to a homogeneous random field $u(\mathbf{x})$ invariant under translations by $a\mathbb{Z}^d$. The standard definition of the spectrum $n(\mathbf{k}) = \lim_{L\to\infty}(L/2\pi)^d\langle|a_{\mathbf{k},L}|^2\rangle$ implies that one must choose

$$\lim_{L\to\infty}\langle \tilde{J}_{\mathbf{k}_L,L}\rangle = n(\mathbf{k}), \tag{A.23}$$

for $\mathbf{k} \in \Lambda^* = [-k_{\max}, +k_{\max}]^d$, where $\mathbf{k}_L = \frac{\mathbf{k}L}{2\pi} (\mathrm{mod}\, M) \cdot \frac{2\pi}{L} \in \Lambda_L^*$ converges to $\mathbf{k}$ as $L = aM \to \infty$ (for fixed a). So, $u_L(\mathbf{x})$ converges in distribution to a homogeneous field $u(\mathbf{x})$ with spectrum $n(\mathbf{k})$.

Let $u_L(\mathbf{x})$ be a *random-phase and amplitude field* (RPA) if $u_L(\mathbf{x})$ is RP and if also $\tilde{J}_{\mathbf{k}}$ are mutually independent random variables for all $\mathbf{k} \in \Lambda_L^*$. This is equivalent to the factorization of the N-mode PDF into a product of 1-mode PDFs:

$$\mathcal{P}^{(N)}(s) = \prod_{\mathbf{k}\in\Lambda_L^*} P(s_{\mathbf{k}};\mathbf{k}). \tag{A.24}$$

All homogeneous Gaussian random fields are RPA. Conversely, for any sequence of RPA fields satisfying condition (A.23) the spatial field $u_L(\mathbf{x})$ converges in distribution to the homogeneous Gaussian field with mean zero and spectrum $n(\mathbf{k})$ as $L \to \infty$ [113]. Here we note only that

$$u_L(\mathbf{x}) = \left(\frac{2\pi}{L}\right)^{d/2} \sum_{\mathbf{k}\in\Lambda_L^*} \sqrt{\tilde{J}_{\mathbf{k},L}}\, \exp(i\mathbf{k}\cdot\mathbf{x} + i\varphi_{\mathbf{k}}) \tag{A.25}$$

is a sum of N independent variables scaled by $1/\sqrt{N}$. It is important to emphasize that the Fourier coefficients $\tilde{a}_{\mathbf{k},L}$ can remain far from Gaussian in this limit. In physical space also there are non-vanishing cumulants for large but finite L.

Let us define the characteristic functional, containing information about the statistical distribution of amplitudes and phases:

$$\mathcal{Z}_L(\lambda,\mu) = \left\langle \exp\left[\int d\mathbf{k}(i\lambda_{\mathbf{k}} J_{\mathbf{k}} + i\mu_{\mathbf{k}}\varphi_{\mathbf{k}})\right]\right\rangle \tag{A.26}$$

A most important result for RPA fields is that the *empirical spectrum*

$$\widehat{n}_L(\mathbf{k}) = \left(\frac{2\pi}{L}\right)^{d} \sum_{\mathbf{k}_1\in\Lambda_L^*} \tilde{J}_{\mathbf{k}_1,L}\delta^d(\mathbf{k}-\mathbf{k}_1), \quad \mathbf{k}\in\Lambda^* \tag{A.27}$$

converges under the condition (A.23) to the deterministic spectrum $n(\mathbf{k})$ with probability going to 1 in the limit $L \to \infty$ (weak LLN). One can show that $\int d^dk\ \lambda(\mathbf{k})\ \widehat{n}_L(\mathbf{k})$ converges in probability to $\int d^dk\ \lambda(\mathbf{k})n(\mathbf{k})$ for every bounded, continuous λ. This is sufficient to infer that the amplitude characteristic function defined in (A.26) satisfies

$$\lim_{L\to\infty} \mathcal{Z}_L(\lambda) = \exp\left(i\int d^dk\ \lambda(\mathbf{k})n(\mathbf{k})\right) \tag{A.28}$$

with $n(\mathbf{k})$ the deterministic spectrum. The LLN means that for RPA fields the empirical spectrum $\widehat{n}_L(\mathbf{k})$ coincides with $n(\mathbf{k})$ at large L for almost every realization of the random phases and amplitudes.

Notice that for the above result one does not actually need the full independence assumption in RPA, but it suffices that

$$\lim_{L\to\infty} [\mathcal{N}_L^{(2)}(\mathbf{k}_1,\mathbf{k}_2) - \mathcal{N}_L^{(1)}(\mathbf{k}_1)\mathcal{N}_L^{(1)}(\mathbf{k}_2)] = 0, \tag{A.29}$$

where the M-th order correlation functions are defined as

$$\mathcal{N}_L^{(M)}(\mathbf{k}_1,...,\mathbf{k}_M) = \langle \widehat{n}_L(\mathbf{k}_1)\cdots\widehat{n}_L(\mathbf{k}_M)\rangle. \tag{A.30}$$

Property (A.29) is analogous to Boltzmann's *Stosszahlansatz* for his kinetic equation. Under this assumption, the M-th order correlations that exist will factorize in the large-box limit [116, 117]:

$$\lim_{L\to\infty} \mathcal{N}_L^{(M)}(\mathbf{k}_1,...,\mathbf{k}_M) = \prod_{m=1}^{M} n(\mathbf{k}_m). \tag{A.31}$$

Our results indicate that properties RP and (A.23,A.29) for the initial wave field, suffice for the wave kinetic equation and for the LLN for the empirical spectrum to hold at positive times.

RPA fields whose Fourier amplitudes possess the full independence property satisfy the even stronger LLN for the *empirical 1-mode PDF*

$$\widehat{P}_L(s;\mathbf{k}) = \left(\frac{2\pi}{L}\right)^d \sum_{\mathbf{k}_1 \in \Lambda_L^*} \delta(s - \tilde{J}_{\mathbf{k}_1})\delta^d(\mathbf{k}-\mathbf{k}_1). \tag{A.32}$$

Assume that the limiting random variables $\tilde{J}_{\mathbf{k}} = \lim_{L\to\infty} \tilde{J}_{\mathbf{k}_L,L}$ of an RPA field exist and have PDFs $P(s;\mathbf{k})$ which are continuous in $\mathbf{k}$. Then, the random functions $\widehat{P}_L(s;\mathbf{k})$ converge to $P(s;\mathbf{k})$ with probability approaching 1 as $L\to\infty$. This implies the previous LLN for the spectrum, since $\widehat{n}_L(\mathbf{k}) = \int_0^\infty ds\, s\widehat{P}_L(s;\mathbf{k})$ and $n(\mathbf{k}) = \int_0^\infty ds\, sP(s;\mathbf{k})$. Although the "empirical PDF" defined in (A.32) is mathematically very convenient, it is not a PDF for finite L. It is therefore more intuitive to use an alternative definition

$$\widehat{P}_L(s;\Delta) = \frac{1}{N_L(\Delta)} \sum_{\mathbf{k}\in\Lambda_L^*\cap\Delta} \delta(s-\tilde{J}_{\mathbf{k}}), \tag{A.33}$$

for any open set $\Delta \subset \Lambda^*$ and with $N_L(\Delta)$ the number of elements in $\Lambda_L^* \cap \Delta$. This quantity is nearly the same as $\frac{1}{|\Delta|}\int_\Delta d^dk\ \widehat{P}_L(s;\mathbf{k})$ for large L but it has the advantage that it defines a probability measure in s for each fixed Δ and L. Definition (A.33) also has a simple intuitive meaning, since it represents the instantaneous distribution of amplitudes of the large number of Fourier modes that reside in the set Δ for large box-size L. Under the same assumptions as above, it follows with probability going to 1 that

$$\lim_{L\to\infty} \widehat{P}_L(s;\Delta) = \frac{1}{|\Delta|}\int_\Delta d^dk\ P(s;\mathbf{k}) \equiv P(s;\Delta). \tag{A.34}$$

Strict independence is not necessary for this to hold; factorization of *multimode PDFs* for $\mathbf{k}_1,...,\mathbf{k}_M \in \Lambda^*$ is required:

$$\mathcal{P}_L^{(M)}(s_1,...,s_M;\mathbf{k}_1,...,\mathbf{k}_M) = \langle \delta(s_1-\tilde{J}_{\mathbf{k}_{1,L},L})\cdots\delta(s_M-\tilde{J}_{\mathbf{k}_{M,L},L})\rangle. \tag{A.35}$$

The factorization property for all pairs of distinct $\mathbf{k}_1,\mathbf{k}_2 \in \Lambda^*$

$$\lim_{L\to\infty}[\mathcal{P}_L^{(2)}(s_1,s_2;\mathbf{k}_1,\mathbf{k}_2) - \mathcal{P}_L^{(1)}(s_1;\mathbf{k}_1)\mathcal{P}_L^{(1)}(s_2;\mathbf{k}_2)] = 0 \tag{A.36}$$

suffices for the LLN for the empirical PDF and also the factorization of the multimode PDFs

$$\lim_{L\to\infty} \mathcal{P}_L^{(M)}(s_1,...,s_M;\mathbf{k}_1,...,\mathbf{k}_M) = \prod_{m=1}^M P(s_m;\mathbf{k}_m) \tag{A.37}$$

for all integers $M>2$ and distinct $\mathbf{k}_1,...,\mathbf{k}_M \in \Lambda^*$. The asymptotic independence is considerably weaker than RPA, permitting statistical dependence between Fourier modes at finite L. In the following, we show that properties (A.31), (A.37) are preserved by the limiting kinetic hierarchies of WT.

A.3 Multimode hierarchy equations

In this section we formally derive the multimode kinetic equations for the 4-wave dynamics of our system. Our analysis differs from those of previous works [69, 39, 38] mainly because of the nonlinear frequency shift, and because of the details of the $L \to \infty$ and $\epsilon \to 0$ limits.

The action-angle variables (amplitudes and phases) for linear dynamics are defined as $J_{\mathbf{k}} = |A^\sigma_{\mathbf{k}}|^2$ and $\varphi_{\mathbf{k}} = \sigma \arg(A^\sigma_{\mathbf{k}})$, so that $A^\sigma_{\mathbf{k}} = \sqrt{J_{\mathbf{k}}}\psi^\sigma_{\mathbf{k}}$, where $\psi_{\mathbf{k}} = \exp(i\varphi_{\mathbf{k}})$. Then, the Liouville measure μ conserved by the Hamiltonian flow can be written as

$$d\mu = \prod_{\mathbf{k}} dQ_{\mathbf{k}} dP_{\mathbf{k}} = \prod_{\mathbf{k}} \frac{1}{i} dA^+_{\mathbf{k}} dA^-_{\mathbf{k}} = \prod_{\mathbf{k}} \frac{1}{i} da^+_{\mathbf{k}} da^-_{\mathbf{k}} = \prod_{\mathbf{k}} dJ_{\mathbf{k}} d\varphi_{\mathbf{k}} \tag{A.38}$$

The canonical momenta and coordinates are given by real and imaginary parts of $A^\sigma_{\mathbf{k}} = \frac{1}{\sqrt{2}}(P_{\mathbf{k}} + i\sigma Q_{\mathbf{k}})$, and $A^\sigma_{\mathbf{k}}$ and $a^\sigma_{\mathbf{k}}$ are linked by the simple rotation in the complex plane used to obtain (A.13). Consistently with the general definition (A.26), the generating function of amplitudes and phases for finite box-size L is:

$$\mathcal{Z}_L[\lambda, \mu, T] \doteq \left\langle \exp\left(\sum_{\mathbf{k}\in\Lambda^*_L} \lambda_{\mathbf{k}} J_{\mathbf{k}}(T) \right) \prod_{\mathbf{k}\in\Lambda^*_L} \psi^{\mu_{\mathbf{k}}}_{\mathbf{k}}(T) \right\rangle \tag{A.39}$$

$$\lambda_{\mathbf{k}} \in \mathbb{R}, \quad \mu_{\mathbf{k}} \in \mathbb{Z} \qquad \forall \mathbf{k} \in \Lambda^*_L$$

A.3.1 Power Series Expansion in the Dynamical Equation

The frequency shift

Let us perturbatively expand the solution of Eq.(A.15) in ϵ at finite L. As explained in [69] and [137], we consider an intermediate time between the "linear time", that is the wave period, and the "nonlinear time" that represents the time scale of evolution of the wave amplitude statistics. To consider the long-time behavior of the wave field expanding in ϵ the solution of the dynamical equation, we need to renormalize the frequency [39, 137]. The equation for the order zero in ϵ has a constant solution:

$$a^{(0)}_1(T) = a_1(0), \tag{A.40}$$

Thus, the terms like $\sum_{234} \mathcal{L}_{1234} a^{(0)}_2 a^{(0)}_3 a_4 \exp(i\omega^1_{234} t)\, \delta^1_{234}$, for $\mathbf{k}_2 = \mathbf{k}_3$, $\sigma_2 = -\sigma_3$ and $\mathbf{k}_4 = \mathbf{k}_1$, play the role of linear terms in a_1, responsible for fast oscillations. We want to remove all terms of this kind, using an interaction representation and a frequency renormalization [39]:

$$\sum_{234}^{**} \doteq \sum_{\sigma_2\sigma_3\sigma_4} \sum_{\mathbf{k}_2\mathbf{k}_3\mathbf{k}_4} \delta_{\sigma_2,\sigma_1}\delta_{\sigma_3,-\sigma_4}\delta_{\mathbf{k}_2,\mathbf{k}_1}\delta_{\mathbf{k}_3,\mathbf{k}_4} + (2 \leftrightarrow 3) + (2 \leftrightarrow 4) \tag{A.41}$$

$$\sum_{234}^{*} \doteq \left(\sum_{\sigma_2\sigma_3\sigma_4} \sum_{\mathbf{k}_2\mathbf{k}_3\mathbf{k}_4} - \sum_{234}^{**} \right) \tag{A.42}$$

Recalling Eq.(A.15), we can write:

$$\begin{aligned}\dot{a}_1 &= \epsilon\left(\sum_{234}^{*}+\sum_{234}^{**}\right)\mathcal{L}_{1234}^{+\sigma_2\sigma_3\sigma_4}a_2^{\sigma_2}a_3^{\sigma_3}a_4^{\sigma_4}\exp\left(i\omega_{234}^{1}t\right)\delta_{234}^{1}\\ &= \epsilon\sum_{234}^{*}\mathcal{L}_{1234}^{+\sigma_2\sigma_3\sigma_4}a_2^{\sigma_2}a_3^{\sigma_3}a_4^{\sigma_4}\exp\left(i\omega_{234}^{1}t\right)\delta_{234}^{1}+i\,\Omega_1\,a_1+\epsilon^2\mathcal{D}_1a_1 \qquad \text{(A.43)}\end{aligned}$$

where

$$i\Omega_1 \doteq \epsilon\sum_{\sigma_2=\pm1}\sum_{\mathbf{k}_2}\mathcal{L}_{1122}^{++\sigma_2-\sigma_2}\left|a_2^{(0)}\right|^2+(2\leftrightarrow3)+(2\leftrightarrow4) \qquad \text{(A.44)}$$

and $\mathcal{D}_1 = O(1)$. Introducing a new interaction representation with

$$b_{\mathbf{k}} = a_{\mathbf{k}}e^{-i\Omega_{\mathbf{k}}t} \qquad \text{(A.45)}$$

Eq.(A.43) becomes:

$$\dot{b}_1 = \epsilon\sum_{234}^{*}\mathcal{L}_{1234}^{+\sigma_2\sigma_3\sigma_4}\,b_2^{\sigma_2}b_3^{\sigma_3}b_4^{\sigma_4}\,e^{i\tilde{\omega}_{234}^{1}t}\,\delta_{234}^{1}+\epsilon^2\mathcal{D}_1b_1 \qquad \text{(A.46)}$$

where the renormalized frequency with a shift is given by [39, 137]:

$$\tilde{\omega}_{\mathbf{k}} \doteq \omega_{\mathbf{k}}+\Omega_{\mathbf{k}} \qquad \text{(A.47)}$$

2^{nd} order equations

Considering an intermediate time between the linear and the nonlinear time ($\frac{2\pi}{\tilde{\omega}_{\mathbf{k}}} << T << \frac{2\pi}{\epsilon^2\tilde{\omega}_{\mathbf{k}}}$), the solution of Eq.(A.46) to second order in ϵ is:

$$b_{\mathbf{k}}(T) = b_{\mathbf{k}}^{(0)}(T)+\epsilon b_{\mathbf{k}}^{(1)}(T)+\epsilon^2b_{\mathbf{k}}^{(2)}(T)+O(\epsilon^3) \qquad \text{(A.48)}$$

which implies

$$b_1^{(0)}(T) = b_1(0) \qquad \text{(A.49)}$$

$$b_1^{(1)}(T) = \sum_{234}^{*}\mathcal{L}_{1234}b_2^{(0)}b_3^{(0)}b_4^{(0)}\Delta_T(\tilde{\omega}_{234}^{1})\delta_{234}^{1} \qquad \text{(A.50)}$$

$$\begin{aligned}b_1^{(2)}(T) = &\sum_{234567}^{*}\mathcal{L}_{1234}\mathcal{L}_{4567}\,b_2^{(0)}b_3^{(0)}b_5^{(0)}b_6^{(0)}b_7^{(0)}E_T\left(\tilde{\omega}_{23567}^{1},\tilde{\omega}_{234}^{1}\right)\delta_{234}^{1}\delta_{567}^{4}\\ &+(4\leftrightarrow3)+(4\leftrightarrow2)+\int_0^T\mathcal{D}_1b_1^{(0)}dt \qquad \text{(A.51)}\end{aligned}$$

where

$$\Delta_T(x) \doteq \int_0^T\exp(ixt)dt\ , \quad E_T(x,y) \doteq \int_0^T\Delta_t(x-y)\exp(iyt)dt \qquad \text{(A.52)}$$

and

$$\sum_{234567}^{**} \doteq \sum_{\sigma_2\sigma_3\ldots\sigma_7}\sum_{\mathbf{k}_2\mathbf{k}_3\ldots\mathbf{k}_7}\delta_{\sigma_2,\sigma_1}\delta_{\sigma_3,-\sigma_4}\delta_{\sigma_4,\sigma_5}\delta_{\sigma_6,-\sigma_7} \times\delta_{\mathbf{k}_2,\mathbf{k}_1}\delta_{\mathbf{k}_3,\mathbf{k}_4}\delta_{\mathbf{k}_4,\mathbf{k}_5}\delta_{\mathbf{k}_6,\mathbf{k}_7} + (2\leftrightarrow 3) + (2\leftrightarrow 4) +(5\leftrightarrow 6) + (5\leftrightarrow 7) + (2\leftrightarrow 3{,}5\leftrightarrow 6) + (2\leftrightarrow 4{,}5\leftrightarrow 6) +(2\leftrightarrow 3{,}5\leftrightarrow 7) + (2\leftrightarrow 4{,}5\leftrightarrow 7) \tag{A.53}$$

$$\sum_{234567}^{*} \doteq \left(\sum_{\sigma_2\sigma_3\ldots\sigma_7}\sum_{\mathbf{k}_2\mathbf{k}_3\ldots\mathbf{k}_7} - \sum_{234567}^{**}\right) \tag{A.54}$$

$$\mathcal{D}_1 \doteq \sum_{\sigma_2=\pm 1}\sum_{\mathbf{k}_2}\mathcal{L}_{1122}^{++\sigma_2-\sigma_2}\left(b_2^{(0)}b_2^{(1)*} + b_2^{(1)}b_2^{(0)*}\right) + (2\leftrightarrow 3) + (2\leftrightarrow 4) \tag{A.55}$$

A.3.2 Phase averaging: Feynman-Wyld Diagrams

In this section, we carry out the phase averaging using diagrammatic techniques, which are in essence those used in the 3-wave case in Ref. [38]. However, here we describe them in details, for completeness and also because we have introduced spin terms, σ_i, absent in [38].

An expansion like (A.48) for the original normal variables $A_{\mathbf{k}}$ may be written as

$$A_{\mathbf{k}}(T) = A_{\mathbf{k}}^{(0)}(T) + \epsilon A_{\mathbf{k}}^{(1)}(T) + \epsilon^2 A_{\mathbf{k}}^{(2)}(T) + O(\epsilon^3) \tag{A.56}$$

where:

$$b_{\mathbf{k}}^{(i)} = A_{\mathbf{k}}^{(i)}e^{-i\tilde{\omega}t}, \quad i = 0{,}1{,}2 \tag{A.57}$$

and a similar expansion Equation (A.56) leads to:

$$J_{\mathbf{k}}(T) = |A_{\mathbf{k}}(T)|^2 = |b_{\mathbf{k}}(T)|^2 \doteq J_{\mathbf{k}}^{(0)} + \epsilon J_{\mathbf{k}}^{(1)} + \epsilon^2 J_{\mathbf{k}}^{(2)} + O(\epsilon^3) \tag{A.58}$$

Definition (A.39) shows that $\mathcal{Z}_L$ satisfies the simmetry:

$$\mathcal{Z}_L\left[\lambda,\mu,T\right] = \mathcal{Z}_L^*\left[\lambda,-\mu,T\right] \tag{A.59}$$

Therefore, writing

$$\mathcal{Z}_L\left[\lambda,\mu,T\right] = \chi_L\left\{\lambda,\mu,T\right\} + \chi_L^*\left\{\lambda,-\mu,T\right\} \tag{A.60}$$

one eventually gets:

$$\chi_L\left\{\lambda,\mu,T\right\} = \chi_L\left\{\lambda,\mu{,}0\right\} + \left\langle \prod_{\mathbf{k}\in\Lambda_L^*} e^{\lambda_{\mathbf{k}}J_{\mathbf{k}}^{(0)}}\left[\epsilon\mathcal{J}_1 + \epsilon^2\left(\mathcal{J}_2+\mathcal{J}_3+\mathcal{J}_4+\mathcal{J}_5\right)\right]\right\rangle^J \tag{A.61}$$

where [38]:

$$\mathcal{J}_1 \doteq \left\langle \prod_{\mathbf{k}} \psi_{\mathbf{k}}^{(0)\mu_{\mathbf{k}}} \sum_1 \left(\lambda_1 + \frac{\mu_1}{2J_1^{(0)}}\right) b_1^{(1)} b_1^{(0)*} \right\rangle_\psi \tag{A.62}$$

$$\mathcal{J}_2 \doteq \frac{1}{2}\left\langle \prod_{\mathbf{k}} \psi_{\mathbf{k}}^{(0)\mu_{\mathbf{k}}} \sum_1 \left(\lambda_1 + \lambda_1^2 J_1^{(0)} - \frac{\mu_1^2}{4J_1^{(0)}}\right) |b_1^{(1)}|^2 \right\rangle_\psi \tag{A.63}$$

$$\mathcal{J}_3 \doteq \left\langle \prod_{\mathbf{k}} \psi_{\mathbf{k}}^{(0)\mu_{\mathbf{k}}} \sum_1 \left(\lambda_1 + \frac{\mu_1}{2J_1^{(0)}}\right) b_1^{(2)} b_1^{(0)*} \right\rangle_\psi \tag{A.64}$$

$$\mathcal{J}_4 \doteq \left\langle \prod_{\mathbf{k}} \psi_{\mathbf{k}}^{(0)\mu_{\mathbf{k}}} \sum_1 \left(\frac{1}{2}\lambda_1^2 + \frac{\mu_1}{4J_1^{(0)2}}(\frac{\mu_1}{2} - 1) + \frac{\lambda_1\mu_1}{2J_1^{(0)}}\right) (b_1^{(1)} b_1^{(0)*})^2 \right\rangle_\psi \tag{A.65}$$

$$\begin{aligned}\mathcal{J}_5 \doteq \frac{1}{2}\Big\langle \prod_{\mathbf{k}} \psi_{\mathbf{k}}^{(0)\mu_{\mathbf{k}}} \sum_{1\neq 2} \Big(&\lambda_1\lambda_2 (b_1^{(1)} b_1^{(0)*} + b_1^{(1)*} b_1^{(0)}) b_2^{(1)} b_2^{(0)*} \\ &+ (\lambda_1 + \frac{\mu_1}{4J_1^{(0)}}) \frac{\mu_2}{J_2^{(0)}} (b_2^{(1)} b_2^{(0)*} - b_2^{(1)*} b_2^{(0)}) b_1^{(1)} b_1^{(0)*} \Big) \Big\rangle_\psi\end{aligned} \tag{A.66}$$

The averages over phases and amplitudes have been separated. Furthermore,

$$\chi_L\{\lambda, \mu, 0\} \doteq \left\langle \prod_{\mathbf{k}} \exp\left[\lambda_{\mathbf{k}} J_{\mathbf{k}}^{(0)}\right] \right\rangle_J \tag{A.67}$$

Rules for *phase-averaging*

The terms in the perturbative solution of the equation of motion can be represented by Wyld diagram expansions [39, 69, 199, 205]. The main rules for such diagrams and for averages over phases follow.

Rule 1 **How to build the basic diagrams**
The various contributions are represented by tree diagrams illustrated in Figs. 1-3, for the zeroth-, first- and second-order terms, that we call "basic diagrams".

- Lines: a solid line labeled by an integer j represents factor $b_j^{(0)}$; dashed line indicates the absence of such a factor. An arrow added to a solid line, pointing away from j, indicates $\sigma_j = +1$ (source); if the arrow points toward j, it corresponds to $\sigma_j = -1$ (sink).
- Vertices: the vertex labelled 1234 represents $\mathcal{L}_{1234}^{\sigma_1,\sigma_2,\sigma_3,\sigma_4} e^{\tilde{\omega}_{234}^1 t} \delta_{234}^1$ with $\sigma_1 = +1$ when the arrow points out of the vertex and $\sigma_1 = -1$ when the arrow points into the vertex. The times at each vertex are ordered causally, with the latest times at the root of the tree, labelled by 1. Integrating from time 0 to T, one gets the various contributions to the perturbative solution.

Figure A.1. $b_1^{(0)+}$ and $b_1^{(0)-}$

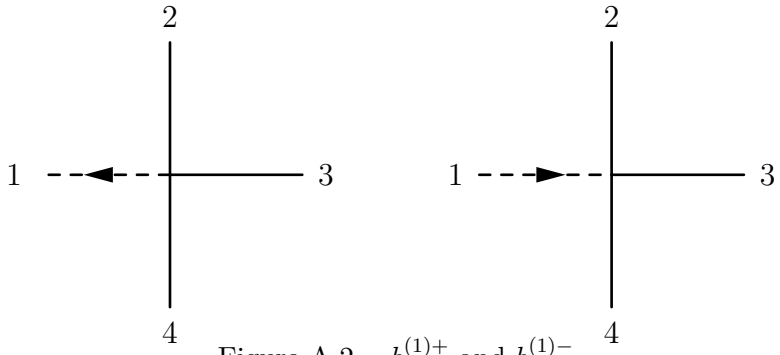

Figure A.2. $b_1^{(1)+}$ and $b_1^{(1)-}$

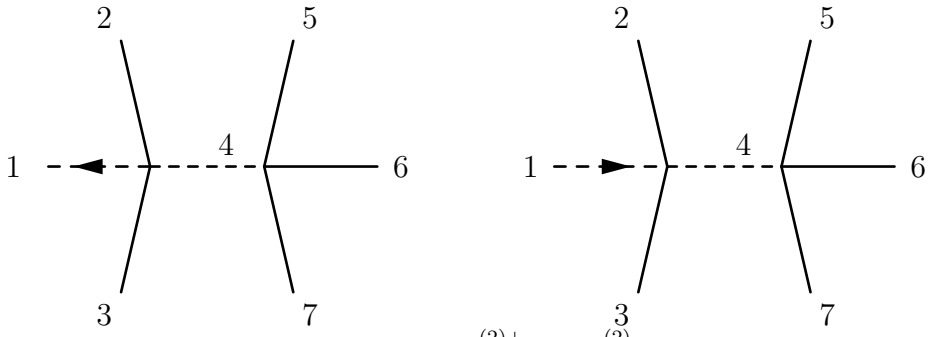

Figure A.3. $b_1^{(2)+}$ and $b_1^{(2)-}$

For completeness, observe that:

$$b_1^{(1)}(T) = \sum_{234}^{*} \mathcal{L}_{1234} b_2^{(0)} b_3^{(0)} b_4^{(0)} \Delta_T(\tilde{\omega}_{234}^1)\delta_{234}^1 \tag{A.68}$$

$$b_1^{(2)}(T) = \sum_{234567}^{*} \mathcal{L}_{1234}\mathcal{L}_{4567}\, b_2^{(0)} b_3^{(0)} b_5^{(0)} b_6^{(0)} b_7^{(0)} \int_0^T \Delta_t(\tilde{\omega}_{567}^4) \exp\left(i\tilde{\omega}_{234}^1 t\right) dt \delta_{234}^1 \delta_{567}^4$$
$$+(4 \longleftrightarrow 3) + (4 \longleftrightarrow 2) + \int_0^T \mathcal{D}_1 b_1^{(0)} dt \tag{A.69}$$

Rule 2 **How to combine basic diagrams before phase-averaging**
Before averaging over phases, the various contributions (A.62)-(A.66) can be represented by diagrams (see next section), combining the tree diagrams in Figs. A.1-A.3. The combination of two basic diagrams graphically represents the product of the two analytical terms to which the diagrams are associated, and this is performed by joining the trees with the same "root" indices, over which there must be a sum. Each of the integer labels indicates an index to be summed over independently of the others, except for the constraints imposed by Kronecker deltas at the vertices.

From now, we omit superscripts, as they are (0).

Rule 3 **Phase-averaging: diagrams closed by internal or external couplings**
The only contributions that survive the average over phases have phases summing to zero before averaging. Then, each $b^{(0)}$ either pairs with another $b^{(0)}$ so that their phases sum to zero or belong to a set of $b^{(0)}$'s that pair with a $\psi_{\mathbf{k}}^{(0)\mu_{\mathbf{k}}}$ making the sum of their phases vanish. The first is an internal coupling, represented by a solid line connecting the paired indices ij that contribute a factor $\delta_{\sigma_i+\sigma_j,0}\delta_{\mathbf{k}_i,\mathbf{k}_j}$ after phase averaging. The second is an external coupling, represented by joining all solid lines with indices $i_1, i_2, ..., i_p$ to a blob $\bullet$ labeled a, that represents the phase $\psi_{\mathbf{k}_a}^{(0)\mu_{\mathbf{k}_a}}$ which contributes a factor $\delta_{\sigma_{i_1}+\cdots\sigma_{i_p}+\mu_a,0}\prod_{j=1}^{p}\delta_{\mathbf{k}_j,\mathbf{k}_a}$ after phase averaging. We will say that the blob (Kronecker delta) makes the wavenumber $\mathbf{k}_j$ pinned to the value $\mathbf{k}_a$. Conventionally, we omit the letters labeling the blobs: factors such as $\delta_{\mathbf{k}_j,\mathbf{k}_a}\delta_{\mu_a+\sigma_j,0}$ are denoted by $\delta_{\mu_j+\sigma_j,0}$, meaning that $\mathbf{k}_j$ is constrained to the value $\mathbf{k}_a$ because of external coupling [38].

Call bridge the line connecting two vertices, labeled with just one number: e.g. the line labeled with 1 in presence of the factor $\mathcal{L}_{1234}^{+\sigma_2\sigma_3\sigma_4}\mathcal{L}_{1567}^{-\sigma_5\sigma_6\sigma_7}$. We distinguish between *in-internal coupling*, with two lines starting from the same vertex closed together, and *cross-internal* coupling, when two lines starting from two different vertices are bound. Let the number of degrees of freedom (or number of free wavenumbers) be the number of summations over all N modes, cf. Appendix B.

Lemma *Let us assume the initial wavefield is an RP field. Consider the phase average*

$$\left\langle \prod_l \psi_l^{\mu_l} \psi_1 \cdots \psi_p \psi_{p+1}^* \cdots \psi_q^* \right\rangle$$

and all the possible associated diagrams giving non-null contributions. Then, the degrees of freedom of each closed diagram are equal to the total number of internal couplings in the diagram, no matter if "in-" or "cross-internal" couplings. No degrees of freedom must be counted for a bridge.

This implies a new rule for the phase-averaging method.

Rule 4 **Distinguishing leading order graphs**
The terms with a larger number of internal couplings are greater in order, so the leading contributions come from the terms with the maximum number of internal couplings. Therefore, we can subdivide the diagrams in four different types: type 0 diagrams with three free wavenumbers; type I diagrams two; type II diagrams with one; type III with no free wavenumbers. The leading contributions are then given by type 0 or type I diagrams and, in some cases, by type II diagrams.

The symbol $\sum^*$ expresses the fact that the combinations of $\mathbf{k}_i$'s and σ_i's giving linear terms inside the interaction term are separated. Then, the interaction representation (A.45) allows us to remove such linear terms from the interaction, implying:

Rule 5 **Diagram "eliminated" by frequency renormalization**
Definition (A.42) implies that the Kronecker delta's inside (A.41) vanish in $\sum^*_{234}$ for any allowed configuration. Definition (A.54) implies that the delta's inside (A.53) also vanish in the term $\sum^*_{2\ldots7}$. Thus, a diagram for $b_1^{(1)}$ implying the arguements of the delta's inside (A.41) to be simultaneously equal to zero is not contributing. The same holds for a graph for $b_1^{(2)}$ whose particular state requires null arguements for the delta's inside (A.53).

Contributions $\mathcal{J}_1 - \mathcal{J}_5$

The graph associated to $\mathcal{J}_1$ before phase-averaging is represented in Fig. A.4, and analitically expressed by:

$$\begin{aligned}\mathcal{J}_1 \;=\; \Bigg\langle \prod_{\mathbf{k}} \psi_{\mathbf{k}}^{\mu_{\mathbf{k}}} \sum_{1234}^{*} \left(\lambda_1 + \frac{\mu_1}{2J_1}\right) \mathcal{L}_{1234}^{+\sigma_2\sigma_3\sigma_4} a_1^{-} a_2^{\sigma_2} a_3^{\sigma_3} a_4^{\sigma_4} \\ \times \Delta_T\left(-\omega_1 + \sigma_2\omega_2 + \sigma_3\omega_3 + \sigma_4\omega_4\right) \delta_{\mathbf{k}_1,\sigma_2\mathbf{k}_2+\sigma_3\mathbf{k}_3+\sigma_4\mathbf{k}_4} \Bigg\rangle_{\psi}\end{aligned} \tag{A.70}$$

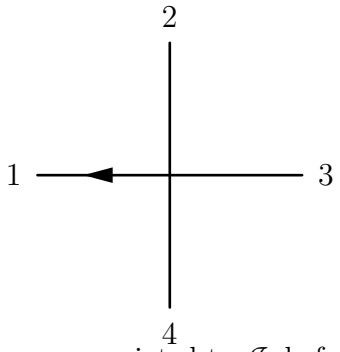

Figure A.4. Diagram associated to $\mathcal{J}_1$ before phase-averaging

Substituting the action-angle variables, we have:

$$\mathcal{J}_1 = \sum_{1234}^{*}\left(\lambda_1+\frac{\mu_1}{2J_1}\right)\sqrt{J_1J_2J_3J_4}\mathcal{L}_{1234}^{+\sigma_2\sigma_3\sigma_4}\left\langle\psi_1^{-1}\psi_2^{\sigma_2}\psi_3^{\sigma_3}\psi_4^{\sigma_4}\prod_{\mathbf{k}}\psi_{\mathbf{k}}^{\mu_{\mathbf{k}}}\right\rangle_{\psi} \times\Delta_T\left(-\omega_1+\sigma_2\omega_2+\sigma_3\omega_3+\sigma_4\omega_4\right)\delta_{\mathbf{k}_1,\sigma_2\mathbf{k}_2+\sigma_3\mathbf{k}_3+\sigma_4\mathbf{k}_4} \tag{A.71}$$

where only the term within angular brackets depends on phases. This term can be thought of as the sum of the contributions of all the possible closures (*Rule 3*) of the diagram in Fig. A.4.

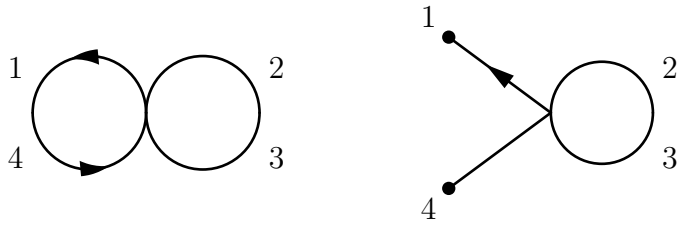

Figure A.5. Diagram 1 (type I, vanishing) and diagram 2 (type II)

1. The contribution associated with diagram 1 in Fig. A.5 may be directly written as

$$\sum_{\underline{\sigma}}\sum_{\underline{\mathbf{k}}}^{*}\left(\lambda_1+\frac{\mu_1}{2J_1}\right)\sqrt{J_1J_2J_3J_4}\mathcal{L}_{1234}^{+\sigma_2\sigma_2\sigma_4}\prod_m\delta_{\mu_m,0}\quad\Delta_T\left(0\right)\delta_{\mathbf{k}_1,\mathbf{k}_4}\delta_{\mathbf{k}_2,\mathbf{k}_3} \tag{A.72}$$

$$\underline{\sigma}=(1,\sigma_2,-\sigma_2,1)$$

The two Kronecker delta's of the internal couplings make the vertex delta redundant. Applying *Rule 5*, one sees that this kind of graph is missing in the interaction. The physics of this diagram has already been included in the frequency renormalization and thus it must not been considered here. This implies that this is not a leading order term in $\mathcal{J}_1$.

2. For diagram 2 in Fig. A.5, one has the following contribution to $\mathcal{J}_1$:

$$\sum_{\underline{\sigma}}\sum_{\underline{\mathbf{k}}}{}'\left(\lambda_1+\frac{\mu_1}{2J_1}\right)\sqrt{J_1J_2J_3J_4}\mathcal{L}_{1234}^{+\sigma_2\sigma_2\sigma_4}\delta_{\mu_1,1}\delta_{\mu_4,1}\delta_{\sigma_2,-\sigma_3} \times\prod_{m\neq1,-1}\delta_{\mu_m,0}\quad\Delta_T\left(-\omega_1-\omega_{-1}\right)\delta_{\mathbf{k}_2,\mathbf{k}_3}\delta_{\mathbf{k}_1,-\mathbf{k}_4} \tag{A.73}$$

Here $\underline{\sigma}=(1,\sigma_2,-\sigma_2,-1)$, because the internal coupling between 2 and 3 needs $\sigma_2=-\sigma_3$ for the phase of $\mathbf{k}_2$ to vanish. Then, $\mathbf{k}_1=\sigma_4\mathbf{k}_4=-\mathbf{k}_4$, as $\sigma_4=-1$.

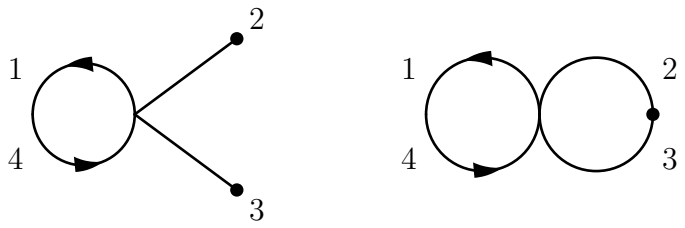

Figure A.6. Diagram 3 (type II) and diagram 4 (type II, vanishing)

3. For diagram 3 in Fig. A.6, the contribution to $\mathcal{J}_1$ reads:

$$\sum_{\underline{\sigma}} {\sum_{\underline{\mathbf{k}}}}' \left(\lambda_1 + \frac{\mu_1}{2J_1}\right) \sqrt{J_1 J_2 J_3 J_4}\, \mathcal{L}_{1234}^{+\sigma_2\sigma_2\sigma_4} \delta_{\mu_2,-\sigma_2} \delta_{\mu_3,-\sigma_2} \times \prod_{m\neq 2,3} \delta_{\mu_m,0} \quad \Delta_T\left(\sigma_2\left(\omega_2 + \omega_{-2}\right)\right) \delta_{\mathbf{k}_2,-\mathbf{k}_3} \delta_{\mathbf{k}_1,\mathbf{k}_4} \tag{A.74}$$

   and $\underline{\sigma} = (1, \sigma_2, \sigma_2, 1)$.

4. For diagram 4, the last Kronecker delta in (A.71), which represents momentum conservation at the vertex, implies $\mathbf{k}_2 = \mathbf{k}_3 = 0$. So this diagram does not represent an effective interaction. As a matter of fact, for spatially homogeneous WT fields there must be no coupling with the zero mode $\mathbf{k} = \mathbf{0}$ because such coupling would violate momentum conservation, cf. [137, 38]. If one of the arguments of $\mathcal{L}_{1234}$ vanishes, the matrix element is zero. That is to say that for any spatially homogeneous WT system $\mathcal{L}_{1234}$ is identically zero if one of $\mathbf{k}_1$, $\mathbf{k}_2$, $\mathbf{k}_3$ or $\mathbf{k}_4$ is zero. The situation is analogous for graphs obtained by permutations of the indices.

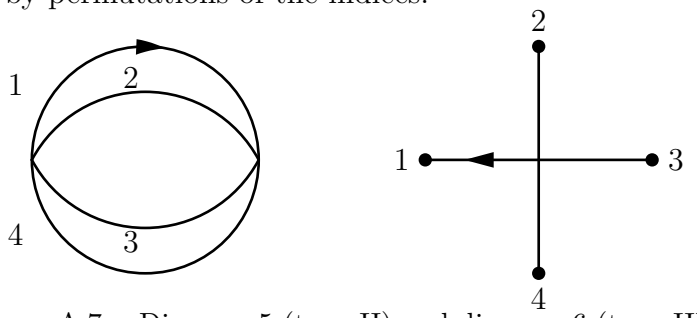

Figure A.7. Diagram 5 (type II) and diagram 6 (type III)

5. Diagram 5 in Fig. 4 contributes as

$$\sum_{\underline{\sigma}} \sum_{\substack{\underline{\mathbf{k}}, \\ \mathbf{k}_1=\mathbf{k}_2=\mathbf{k}_3=\mathbf{k}_4}} \left(\lambda_1 + \frac{\mu_1}{2J_1}\right) \sqrt{J_1 J_2 J_3 J_4}\, \mathcal{L}_{1234}^{+\sigma_2\sigma_2\sigma_4} \delta_{\sigma_2+\sigma_3+\sigma_4,1} \prod_m \delta_{\mu_m,0}\, \Delta_T(0) \tag{A.75}$$
$$\underline{\sigma} = (1, \sigma_2, \sigma_3, \sigma_4)$$

6. All other diagrams are type III (like *e.g.* diagram 6 in Fig. 4) and give subleading contributions.

Normalization of amplitudes: Let us introduce the change of variables

$$J_{\mathbf{k}} = \left(\frac{2\pi}{L}\right)^d \tilde{J}_{\mathbf{k}}, \qquad \lambda_{\mathbf{k}} = i\lambda(\mathbf{k}) \tag{A.76}$$

This substitution implies that the characteristic function be expressed by:

$$\mathcal{Z}_L[\lambda,\mu] \doteq \left\langle \exp\left(i \sum_{\mathbf{k}\in\Lambda_L^*} \left(\frac{2\pi}{L}\right)^d \lambda(\mathbf{k}) \tilde{J}_{\mathbf{k}} \right) \prod_{\mathbf{k}\in\Lambda_L^*} \psi_{\mathbf{k}}^{\mu_{\mathbf{k}}} \right\rangle \tag{A.77}$$

where $\lambda(\mathbf{k})$ is a smooth test function and $\mu_{\mathbf{k}}$ are integers. Here we keep the time dependence implicit, for sake of notation. This characteristic function, after the transformation of the sum to an integral thanks to the large-L limit, becomes a characteristic functional [69]. The change of variables (A.76) is the key to a finite, well defined expression, in the thermodynamic limit.

Main contributions to $\mathcal{J}_1$: Diagram 2 is the only type II diagram with mode $\mathbf{k}_1$ pinned to an external blob, so that $\mu_1 \neq 0$. This graph contributes to $O(1)$, as it is order $O(L^d)$ (one free wavenumber, that is an unconstrained sum) multiplyed by order $O(L^{-d})$ (term proportional to μ_1, see (A.62)). A factor 3 appears to account for the possible permutations of the indices. There is no other leading order term. The other graphs contribute to order $O(L^{-d})$ and vanish in the $L \to \infty$ limit. Summarizing, $\mathcal{J}_1$ may be written as:

$$\begin{aligned} \mathcal{J}_1 \;=\; & \frac{3}{2}\left(\frac{2\pi}{L}\right)^d \sum_{(1)} \left[\sqrt{\frac{\tilde{J}_2\tilde{J}_3\tilde{J}_4}{\tilde{J}_1}}\, \mathcal{L}_{1234}^{+\sigma_2\sigma_3\sigma_4} \delta_{\mu_1,1}\delta_{\mu_{-1},1} \prod_{m\neq\pm 1} \delta_{\mu_m,0} \Delta_T\left(-(\omega_1+\omega_{-1})\right) \right] \\ & + O\left(L^{-d}\right), \qquad \text{where} \qquad \sum_{(1)} \doteq \sum_{\substack{\sigma_1=-\sigma_4=1,\\ \sigma_2=-\sigma_3}} \sum_{\mathbf{k}_2} \end{aligned} \tag{A.78}$$

The contributions from the terms $\mathcal{J}_2, ..., \mathcal{J}_5$, are computed in the appendix of [33].

Contribution of $\mathcal{J}_2$

$$\begin{aligned} \mathcal{J}_2 \;=\; & \frac{i}{2}\left(\frac{2\pi}{L}\right)^{3d} \delta_{\mu,0} \Big\{ 9 \sum\nolimits_{(2)} \Big[\lambda(\mathbf{k}_1)\, \mathcal{L}_{1234}^{+\sigma_2\sigma_3\sigma_4} \mathcal{L}_{1567}^{-\sigma_5\sigma_6\sigma_7} \tilde{J}_2\tilde{J}_4\tilde{J}_5 |\Delta_T\left(\tilde{\omega}_1+\tilde{\omega}_{-1}\right)|^2 \Big] \\ & + 6 \sum\nolimits_{(3)} \Big[\lambda(\mathbf{k}_1)\, |\mathcal{L}_{1234}^{+\sigma_2\sigma_3\sigma_4}|^2 \tilde{J}_2\tilde{J}_3\tilde{J}_4 |\Delta_T\left(\tilde{\omega}_{234}^1\right)|^2 \Big] \Big\} + O\left(L^{-1}\right) \end{aligned} \tag{A.79}$$

$$\sum\nolimits_{(2)} \doteq \sum_{\underline{\sigma}} \sum_{\underline{\mathbf{k}}}{}' \delta_{\mathbf{k}_4,-\mathbf{k}_1} \delta_{\mathbf{k}_4,\mathbf{k}_7} \delta_{\mathbf{k}_2,\mathbf{k}_3} \delta_{\mathbf{k}_5,\mathbf{k}_6}, \;\; \underline{\sigma} = (1,\sigma_2,-\sigma_2,-1,\sigma_5,-\sigma_5,1) \tag{A.80}$$

$$\sum\nolimits_{(3)} \doteq \sum_{\underline{\sigma}} \sum_{\underline{\mathbf{k}}}{}' \delta_{234}^1 \delta_{\mathbf{k}_2,\mathbf{k}_5} \delta_{\mathbf{k}_4,\mathbf{k}_7} \delta_{\mathbf{k}_3,\mathbf{k}_6}, \;\; \underline{\sigma} = (1,\sigma_2,\sigma_3,\sigma_4,-\sigma_2,-\sigma_3,-\sigma_4) \tag{A.81}$$

Contribution of $\mathcal{J}_3$

$$
\begin{aligned}
\mathcal{J}_3 \;=\; & 18i\left(\frac{2\pi}{L}\right)^{3d}\delta_{\mu,0} \\
& \times\Big\{\sum\nolimits_{(4)}\Big[\lambda\left(\mathbf{k}_1\right)\mathcal{L}_{1234}^{+\sigma_2\sigma_3\sigma_4}\mathcal{L}_{4567}^{\sigma_4\sigma_5\sigma_6\sigma_7}\tilde{J}_1\tilde{J}_3\tilde{J}_5E_T\left(0,\sigma_3\left(\omega_3+\omega_{-3}\right)\right)\Big] \\
& \quad+\frac{1}{2}\sum\nolimits_{(5)}\Big[\lambda\left(\mathbf{k}_1\right)\mathcal{L}_{1234}^{+\sigma_2\sigma_3\sigma_4}\mathcal{L}_{4567}^{\sigma_4\sigma_5\sigma_6\sigma_7}\tilde{J}_1\tilde{J}_3\tilde{J}_6E_T\left(0,-\left(\tilde{\omega}_1+\tilde{\omega}_{-1}\right)\right)\Big] \\
& \quad+\sum\nolimits_{(6)}\Big[\lambda\left(\mathbf{k}_1\right)\mathcal{L}_{1234}^{+\sigma_2\sigma_3\sigma_4}\mathcal{L}_{4567}^{\sigma_4\sigma_5\sigma_6\sigma_7}\tilde{J}_1\tilde{J}_2\tilde{J}_3E_T\left(0,\tilde{\omega}_{234}^1\right)\Big]\Big\} \\
& +9\left(\frac{2\pi}{L}\right)^{2d}\Big\{\sum\nolimits_{(7)}\Big[\mathcal{L}_{1234}^{+\sigma_2\sigma_3\sigma_4}\mathcal{L}_{4567}^{\sigma_4\sigma_5\sigma_6\sigma_7}\sqrt{\frac{\tilde{J}_{-1}}{\tilde{J}_1}}\tilde{J}_3\tilde{J}_5\prod_{m\neq1,2}\delta_{\mu_m,0} \\
& \qquad\times E_T\Big(-\left(\tilde{\omega}_1+\tilde{\omega}_{-1}\right),-\left(\tilde{\omega}_1+\tilde{\omega}_{-1}\right)+\sigma_3\tilde{\omega}_3+\sigma_4\tilde{\omega}_4\Big)\Big] \\
& \quad+\frac{1}{2}\sum\nolimits_{(8)}\Big[\mathcal{L}_{1234}^{+\sigma_2\sigma_3\sigma_4}\mathcal{L}_{4567}^{\sigma_4\sigma_5\sigma_6\sigma_7}\sqrt{\frac{\tilde{J}_7}{\tilde{J}_1}}\tilde{J}_3\tilde{J}_5\prod_{m\neq1,7}\delta_{\mu_m,0} \\
& \qquad\times E_T\Big(-\left(\tilde{\omega}_1+\sigma_4\tilde{\omega}_4\right),-\left(\tilde{\omega}_1+\sigma_4\tilde{\omega}_4\right)\Big)\Big] \\
& \quad+\sum\nolimits_{(9)}\Big[\mathcal{L}_{1234}^{+\sigma_2\sigma_3\sigma_4}\mathcal{L}_{4567}^{\sigma_4\sigma_5\sigma_6\sigma_7}\sqrt{\frac{\tilde{J}_{-1}}{\tilde{J}_1}}\tilde{J}_3\tilde{J}_5\prod_{m\neq1,6}\delta_{\mu_m,0} \\
& \qquad\times E_T\Big(-\left(\tilde{\omega}_1+\tilde{\omega}_{-1}\right),\tilde{\omega}_{234}^1\Big)\Big]\Big\}+O\left(L^{-1}\right)
\end{aligned}
\tag{A.82}
$$

where

$$
\begin{aligned}
\sum\nolimits_{(4)} &\doteq \sum_{\underline{\sigma},\underline{\mathbf{k}}}{}'\delta_{\mathbf{k}_4,-\mathbf{k}_3}\delta_{\mathbf{k}_1,\mathbf{k}_2}\delta_{\mathbf{k}_3,\mathbf{k}_7}\delta_{\mathbf{k}_5,\mathbf{k}_6},\ \underline{\sigma}=(1,1,\sigma_3,\sigma_3,\sigma_5,-\sigma_5,-\sigma_3) \\
\sum\nolimits_{(5)} &\doteq \sum_{\underline{\sigma},\underline{\mathbf{k}}}{}'\delta_{\mathbf{k}_1,-\mathbf{k}_4}\delta_{\mathbf{k}_1,\mathbf{k}_5}\delta_{\mathbf{k}_3,\mathbf{k}_2}\delta_{\mathbf{k}_6,\mathbf{k}_7},\ \underline{\sigma}=(1,\sigma_2,-\sigma_2,-1,1,\sigma_6,-\sigma_6) \\
\sum\nolimits_{(6)} &\doteq \sum_{\underline{\sigma},\underline{\mathbf{k}}}{}'\delta_{234}^1\delta_{\mathbf{k}_1,\mathbf{k}_6}\delta_{\mathbf{k}_3,\mathbf{k}_7}\delta_{\mathbf{k}_5,\mathbf{k}_2},\ \underline{\sigma}=(1,\sigma_2,\sigma_3,\sigma_4,-\sigma_2,1,-\sigma_3) \\
\sum\nolimits_{(7)} &\doteq \sum_{\underline{\sigma},\underline{\mathbf{k}}}{}'\delta_{-\sigma_4\mathbf{k}_4,\sigma_3\mathbf{k}_3}\delta_{\mathbf{k}_1,-\mathbf{k}_2}\delta_{\mathbf{k}_3,\mathbf{k}_7}\delta_{\mathbf{k}_5,\mathbf{k}_6},\ \underline{\sigma}=(1,-1,\sigma_3,\sigma_4,\sigma_5,-\sigma_5,-\sigma_3) \\
\sum\nolimits_{(8)} &\doteq \sum_{\underline{\sigma},\underline{\mathbf{k}}}{}'\delta_{\mathbf{k}_1,\sigma_4\mathbf{k}_4}\delta_{\mathbf{k}_1,\sigma_7\mathbf{k}_7}\delta_{\mathbf{k}_3,\mathbf{k}_2}\delta_{\mathbf{k}_5,\mathbf{k}_6},\ \underline{\sigma}=(1,\sigma_2,-\sigma_2,\sigma_4,\sigma_5,-\sigma_5,\sigma_7) \\
\sum\nolimits_{(9)} &\doteq \sum_{\underline{\sigma},\underline{\mathbf{k}}}{}'\delta_{234}^1\delta_{\mathbf{k}_1,-\mathbf{k}_6}\delta_{\mathbf{k}_3,\mathbf{k}_7}\delta_{\mathbf{k}_5,\mathbf{k}_2},\ \underline{\sigma}=(1,\sigma_2,\sigma_3,\sigma_4,-\sigma_2,-1,-\sigma_3)
\end{aligned}
\tag{A.83}
$$

Contribution of $\mathcal{J}_4$

One finds that $\mathcal{J}_4=O\left(L^{-d}\right)$, so it represents a subleading contribution.

Contribution of $\mathcal{J}_5$

$$\begin{aligned}\mathcal{J}_5 &= 3\left(\frac{2\pi}{L}\right)^{2d}\sum_{(10)}\frac{\mu_1\mu_2}{4J_1J_2}\mathcal{L}_{1345}^{+\sigma_3\sigma_4\sigma_5}\mathcal{L}_{2678}^{+\sigma_6\sigma_7\sigma_8}\sqrt{J_1J_2J_3J_4J_5J_6J_7J_8}\\ &\quad\times\delta_{\mu_1,1}\delta_{\mu_2,1}\prod_{m\neq1,2}\delta_{\mu_m,0}\,\Delta_T\left(\tilde{\omega}_{345}^{1}\right)\Delta_T\left(\tilde{\omega}^{(-1)345}\right)\\ &+9\left(\frac{2\pi}{L}\right)^{2d}\sum_{(11)}\frac{\mu_1\mu_2}{4J_1J_2}\mathcal{L}_{1345}^{+\sigma_3\sigma_4\sigma_5}\mathcal{L}_{2678}^{+\sigma_6\sigma_7\sigma_8}\sqrt{J_1J_2J_3J_4J_5J_6J_7J_8}\\ &\quad\times\delta_{\mu_1,2}\delta_{\mu_2,2}\prod_{m\neq1,2}\delta_{\mu_m,0}\,\Delta_T\left(\tilde{\omega}_{(-1)44}^{1}\right)\Delta_T\left(\tilde{\omega}^{(-1)177}\right)+O\left(L^{-d}\right)\end{aligned}\tag{A.84}$$

$$\begin{aligned}\sum_{(10)} &\doteq \sum_{\underline{\sigma}}\sum_{\underline{\mathbf{k}}}{}'\delta_{\mathbf{k}_1,-\mathbf{k}_2}\delta_{\mathbf{k}_3,\mathbf{k}_6}\delta_{\mathbf{k}_4,\mathbf{k}_7}\delta_{\mathbf{k}_5,\mathbf{k}_8}\delta_{345}^{1},\ \underline{\sigma}=(1,1,\sigma_3,\sigma_4,\sigma_5,-\sigma_3,-\sigma_4,-\sigma_5)\\ \sum_{(11)} &\doteq \sum_{\underline{\sigma}}\sum_{\underline{\mathbf{k}}}{}'\delta_{\mathbf{k}_1,-\mathbf{k}_2}\delta_{\mathbf{k}_1,-\mathbf{k}_3}\delta_{\mathbf{k}_1,\mathbf{k}_6}\delta_{\mathbf{k}_4,\mathbf{k}_5}\delta_{\mathbf{k}_7,\mathbf{k}_8},\ \underline{\sigma}=(1,1,-1,\sigma_4,-\sigma_4,-1,\sigma_7,-\sigma_7)\end{aligned}\tag{A.85}$$

A.3.3 Dynamical Multi-Mode Equation

In this section we turn Eq.(A.61) into a dynamical equation for the characteristic functional Z taking the $L\longrightarrow\infty$ and $\epsilon\longrightarrow 0$ limits. The two limits do not commute: the large-box limit must be taken first, the weak-nonlinearity limit after. The physical meaning of this operation is that there is a vast number of quasi-resonances (introduced by the large box limit, which leads to a continuous $\mathbf{k}$-space sending $\Lambda_L^*\longrightarrow\Lambda^*$), each of which is as important as the exact resonances [137].

Large-box limit

Let us introduce the large-L asymptotics standard substitutions, and

$$\left(\frac{2\pi}{L}\right)^{d}\sum_{\mathbf{k}}\Longrightarrow\int d^dk\ ,\quad \left(\frac{L}{2\pi}\right)^{d}\delta_{\mathbf{k},\mathbf{k}'}\Longrightarrow\delta^d(\mathbf{k}-\mathbf{k}')\tag{A.86}$$

Recalling Eq.(A.60), using (A.78), (A.79), (A.82) and (A.84), and neglecting $O\left(L^{-d}\right)$ corrections, we can eventually write:

$$
\begin{aligned}
&\Big\langle \exp\Big[i\int d^dk\lambda(\mathbf{k})\tilde{J}_{\mathbf{k}}\Big]\Big\{6i\epsilon\Big(\sum_{\mathbf{k}_1}\delta_{\mu_1,1}\delta_{\mu_{-1},1}\prod_{m\neq\pm1}\delta_{\mu_m,0}\Big)\\
&\qquad\times\sum_{\sigma_2}\int d^dk_2\tilde{J}_2\sqrt{\frac{\tilde{J}_{-1}}{\tilde{J}_1}}\mathcal{H}^{-\sigma_2(-\sigma_2)-}_{1224}\Delta_T\Big(-\left(\tilde{\omega}_1+\tilde{\omega}_{-1}\right)\Big)+\\
&+8i\epsilon^2\delta_{\mu,0}\Big[9\sum_{\sigma_2,\sigma_5}\int d^dk_1d^dk_2d^dk_5\lambda\left(\mathbf{k}_1\right)\mathcal{H}^{-\sigma_2(-\sigma_2)-}_{122(-1)}\\
&\qquad\qquad\times\mathcal{H}^{+\sigma_5(-\sigma_5)+}_{155(-1)}\tilde{J}_{-1}\tilde{J}_2\tilde{J}_5|\Delta_T\left(\tilde{\omega}_1+\tilde{\omega}_{-1}\right)|^2\\
&\quad+6\sum_{\sigma_2,\sigma_3,\sigma_4}\int d^dk_1d^dk_2d^dk_3d^dk_4\lambda\left(\mathbf{k}_1\right)|\mathcal{H}^{-\sigma_2\sigma_3\sigma_4}_{1234}|^2\tilde{J}_2\tilde{J}_3\tilde{J}_4|\Delta_T\left(\tilde{\omega}^1_{234}\right)|^2\delta^1_{234}\Big]+\\
&+288i\epsilon^2\delta_{\mu,0}\Big[\sum_{\sigma_4,\sigma_5}\int d^dk_1d^dk_4d^dk_5\left(-\sigma_4\right)\lambda\left(\mathbf{k}_1\right)\mathcal{H}^{-+\sigma_4\sigma_4}_{11(-4)4}\\
&\qquad\qquad\times\mathcal{H}^{(-\sigma_4)\sigma_5(-\sigma_5)(-\sigma_4)}_{455(-4)}\tilde{J}_1\tilde{J}_4\tilde{J}_5E_T\left(0,\sigma_4\left(\omega_4+\omega_{-4}\right)\right)\\
&\qquad+\frac{1}{2}\sum_{\sigma_2,\sigma_6}\int d^dk_1d^dk_2d^dk_6\lambda\left(\mathbf{k}_1\right)\mathcal{H}^{-\sigma_2(-\sigma_2)-}_{122(-1)}\\
&\qquad\qquad\times\mathcal{H}^{++\sigma_6-(\sigma_6)}_{(-1)166}\tilde{J}_1\tilde{J}_2\tilde{J}_6E_T\left(0,-\left(\tilde{\omega}_1+\tilde{\omega}_{-1}\right)\right)\\
&\quad+\sum_{\sigma_2,\sigma_3,\sigma_4}\int d^dk_1d^dk_2d^dk_3d^dk_4\left(-\sigma_4\right)\lambda\left(\mathbf{k}_1\right)|\mathcal{H}^{-\sigma_2\sigma_3\sigma_4}_{1234}|^2\tilde{J}_1\tilde{J}_2\tilde{J}_3E_T\left(0,\tilde{\omega}^1_{234}\right)\delta^1_{234}\Big]+\\
&+144\epsilon^2\Big[\Big(\sum_{\mathbf{k}_1}\delta_{\mu_1,1}\delta_{\mu_{-1},1}\prod_{m\neq\pm1}\delta_{\mu_m,0}\Big)\sum_{\sigma_3,\sigma_4,\sigma_5}\int d^dk_3d^dk_4d^dk_5\\
&\qquad\times\left(-\sigma_4\right)\mathcal{H}^{--\sigma_3\sigma_4}_{1(-1)34}\mathcal{L}^{(-\sigma_4)\sigma_5(-\sigma_3)(-\sigma_5)}_{4535}\sqrt{\frac{\tilde{J}_{-1}}{\tilde{J}_1}}\tilde{J}_3\tilde{J}_5\,\delta\left(\sigma_3\mathbf{k}_3+\sigma_4\mathbf{k}_4\right)\\
&\qquad\times E_T\Big(-\left(\tilde{\omega}_1+\tilde{\omega}_{-1}\right),-\left(\tilde{\omega}_1+\tilde{\omega}_{-1}\right)+\sigma_3\tilde{\omega}_3+\sigma_4\tilde{\omega}_4\Big)\\
&\quad+\frac{1}{2}\sum_{\sigma_2,\sigma_4,\sigma_5,\sigma_7}\Big(\sum_{\mathbf{k}_1}\delta_{\mu_1,1}\sum_{\mathbf{k}_7}\delta_{\mu_7,-\sigma_7}\prod_{m\neq1,7}\delta_{\mu_m,0}\Big)\\
&\qquad\times\int d^dk_2d^dk_4d^dk_5d^dk_7\left(-\sigma_4\right)\mathcal{H}^{-\sigma_2(-\sigma_2)\sigma_4}_{1224}\mathcal{H}^{(-\sigma_4)\sigma_5(-\sigma_5)\sigma_7}_{4557}\sqrt{\frac{\tilde{J}_7}{\tilde{J}_1}}\tilde{J}_3\tilde{J}_5\\
&\qquad\times E_T\Big(-\left(\tilde{\omega}_1+\sigma_4\tilde{\omega}_4\right),-\left(\tilde{\omega}_1+\sigma_4\tilde{\omega}_4\right)\Big)\delta(\mathbf{k}_1-\sigma_4\mathbf{k}_4)\delta(\mathbf{k}_1-\sigma_7\mathbf{k}_7)\\
&\quad+\Big(\sum_{\mathbf{k}_1}\delta_{\mu_1,1}\delta_{\mu_{-1},1}\prod_{m\neq\pm1}\delta_{\mu_m,0}\Big)\sum_{\sigma_2,\sigma_3,\sigma_4}\int d^dk_2d^dk_3d^dk_4\\
&\qquad\times\left(-\sigma_4\right)\mathcal{H}^{-\sigma_2\sigma_3\sigma_4}_{1234}\mathcal{H}^{(-\sigma_4)(-\sigma_2)-(-\sigma_3)}_{42(-1)3}\sqrt{\frac{\tilde{J}_{-1}}{\tilde{J}_1}}\tilde{J}_2\tilde{J}_3E_T\Big(-\left(\tilde{\omega}_1+\tilde{\omega}_{-1}\right),\tilde{\omega}^1_{234}\Big)\delta^1_{234}\Big]+
\end{aligned}
$$

$$
\begin{aligned}
&-12\epsilon^2\left(\sum_{\mathbf{k}_1}\delta_{\mu_1,1}\delta_{\mu_{-1},1}\prod_{m\neq\pm1}\delta_{\mu_m,0}\right)\sum_{\sigma_3,\sigma_4,\sigma_5}\\
&\times\int d^dk_3d^dk_4d^dk_5\mathcal{H}^{-\sigma_3\sigma_4\sigma_5}_{1345}\mathcal{H}^{-(-\sigma_3)(-\sigma_4)(-\sigma_5)}_{(-1)345}\frac{\tilde{J}_3\tilde{J}_4\tilde{J}_5}{\sqrt{\tilde{J}_1\tilde{J}_{-1}}}\Delta_T\left(\tilde{\omega}^1_{345}\right)\Delta_T\left(\tilde{\omega}^{(-1)345}\right)\delta^1_{345}\\
&-36\epsilon^2\left(\sum_{\mathbf{k}_1}\delta_{\mu_1,2}\delta_{\mu_{-1},2}\prod_{m\neq\pm1}\delta_{\mu_m,0}\right)\sum_{\sigma_3,\sigma_4}\int d^dk_3d^dk_4\mathcal{H}^{--\sigma_3(-\sigma_3)}_{1(-1)33}\mathcal{H}^{--\sigma_4(-\sigma_4)}_{(-1)144}\tilde{J}_3\tilde{J}_4\\
&\qquad\times\tilde{J}_3\tilde{J}_4\Delta_T\left(\tilde{\omega}^1_{(-1)33}\right)\Delta_T\left(\tilde{\omega}^{(-1)144}\right)\Big\}\Big\rangle_J
\end{aligned}
\tag{A.87}
$$

Weak-nonlinearity limit

Recall that in section A.3.1, we took $\frac{2\pi}{\tilde{\omega}_{\mathbf{k}}}\ll T\ll\frac{2\pi}{\epsilon^2\tilde{\omega}_{\mathbf{k}}}$, with T between the wave period and the nonlinear time. We can now take $T\sim\frac{2\pi}{\epsilon\tilde{\omega}_{\mathbf{k}}}$, so that $\lim_{\epsilon\to0}T=\infty$. Then, in (A.87) we must take the $T\to\infty$ limit, consistently with the large-T asymptotics of Δ_T and E_T [69, 14]:

$$
\begin{aligned}
&\Delta_T(x)\sim\widetilde{\Delta}(x)=\pi\delta(x)+iP\left(\frac{1}{x}\right),\ E_T(x;y)\sim\Delta_T(x)\Delta_T(y)\sim\widetilde{\Delta}(x)\widetilde{\Delta}(y),\\
&|\Delta_T(x)|^2\sim2\pi T\delta(x)+2P\left(\frac{1}{x}\right)\frac{\partial}{\partial x},\ E_T(x;0)\sim\widetilde{\Delta}(x)\left(T-i\frac{\partial}{\partial x}\right),
\end{aligned}
\tag{A.88}
$$

Some considerations are in order.

1. in Eq.(A.87), only the terms containing $|\Delta_T(x)|^2$, $E_T(x;0)$ or $E_T(0;y)$ give secular contributions (proportional to T); the non-secular contributions are irrelevant in the $T\to\infty$ ($\epsilon\to0$) limit. Thus, only the terms with $\delta_{\mu,0}$ survive the weak-nonlinearity assumption, while those with $\delta_{\mu_1,1}$ etc. are subleading.

2. The μ-dependent part of $\mathcal{Z}$ is constrained to be 1 by $\delta_{\mu,0}$. Then, using (A.76), switching to $i\lambda(\mathbf{k})$ and taking the large-box limit leads to the functional derivative
$$
\left(\frac{L}{2\pi}\right)^d\frac{\partial}{\partial\lambda_{\mathbf{k}}}\Longrightarrow-i\frac{\delta}{\delta\lambda(\mathbf{k})}
\tag{A.89}
$$

3. Replace $(\mathcal{Z}[T]-\mathcal{Z}[0])/T$ with the time derivative $\dot{\mathcal{Z}}$. This can be done ([137],pg. 81) because time T is small compared to the characteristic time of averaged quantities such as $\mathcal{Z}$ (nonlinear time). Indeed, the istantaneous time derivative can be of same order or even greater than the rate of change described by our substitution, but such rapid changes are oscillatory and they drop out.

4. We introduce a new time variable $\tau \doteq \epsilon^2 T$.

5. Renaming indices, we split the integral with 4 wavenumbers into identical contributions

$$\begin{aligned}
\frac{d\mathcal{Z}[\lambda,\mu,\tau]}{d\tau} &= -192\pi\delta_{\mu,0}\sum_{\underline{\sigma}=(1,\sigma_2,\sigma_3,\sigma_4)}\int d^dk_1d^dk_2d^dk_3d^dk_4\lambda\left(\mathbf{k}_1\right)|\mathcal{H}^{-\sigma_2\sigma_3\sigma_4}_{1234}|^2 \\
&\quad \delta\left(\tilde{\omega}^1_{234}\right)\delta^1_{234}\Bigg(\frac{\delta^3\mathcal{Z}}{\delta\lambda(\mathbf{k}_2)\delta\lambda(\mathbf{k}_3)\delta\lambda(\mathbf{k}_4)}-\sigma_2\frac{\delta^3\mathcal{Z}}{\delta\lambda(\mathbf{k}_1)\delta\lambda(\mathbf{k}_3)\delta\lambda(\mathbf{k}_4)} \\
&\quad -\sigma_3\frac{\delta^3\mathcal{Z}}{\delta\lambda(\mathbf{k}_1)\delta\lambda(\mathbf{k}_2)\delta\lambda(\mathbf{k}_4)}-\sigma_4\frac{\delta^3\mathcal{Z}}{\delta\lambda(\mathbf{k}_1)\delta\lambda(\mathbf{k}_2)\delta\lambda(\mathbf{k}_3)}\Bigg) \\
&\quad -288\pi\delta_{\mu,0}\sum_{\underline{\sigma}=(1,\sigma_2,\sigma_3)}\int d^dk_1d^dk_2d^dk_3\lambda\left(\mathbf{k}_1\right) \\
&\quad \times\Bigg[\mathcal{H}^{--\sigma_2(-\sigma_2)}_{1(-1)22}\mathcal{H}^{++\sigma_3(-\sigma_3)}_{1(-1)33}\delta\left(\tilde{\omega}_1+\tilde{\omega}_{-1}\right) \\
&\quad \sum_{\sigma=\pm1}\frac{\delta^3\mathcal{Z}}{\delta\lambda(\sigma\mathbf{k}_1)\delta\lambda(\mathbf{k}_2)\delta\lambda(\mathbf{k}_3)}-2\sigma_2\mathcal{H}^{\sigma_2\sigma_2+-}_{2(-2)11} \\
&\quad \mathcal{H}^{(-\sigma_2)(-\sigma_2)\sigma_3(-\sigma_3)}_{2(-2)33}\delta\left(\tilde{\omega}_2+\tilde{\omega}_{-2}\right)\frac{\delta^3\mathcal{Z}}{\delta\lambda(\mathbf{k}_1)\delta\lambda(\mathbf{k}_2)\delta\lambda(\mathbf{k}_3)}\Bigg]
\end{aligned} \tag{A.90}$$

Resonance condition

Recall definition $\tilde{\omega}_{\mathbf{k}} \doteq \omega_{\mathbf{k}}+\Omega_{\mathbf{k}}$. The definition of $\Omega_{\mathbf{k}}$ (A.44) and the thermodynamic limit imply:

$$\Omega_1 \overset{L\to\infty}{\longrightarrow} 24\epsilon\int_{\Lambda^*}d^dk_2\mathcal{H}^{+-+-}_{1122}\tilde{J}^{(0)}_2\ , \qquad \Omega_{-1}=24\epsilon\int_{\Lambda^*}d^dk_2\mathcal{H}^{+-+-}_{(-1)(-1)22}\tilde{J}^{(0)}_2 \tag{A.91}$$

Note: each component of $\mathbf{k}_2$, defined in the dual space Λ^*, ranges in the interval $[-k_{max},k_{max}]$ and space isotropy implies that our system is symmetric under the $\mathbf{k}\to-\mathbf{k}$ transformation. Therefore:

$$\Omega_{-1}=24\epsilon\int_{(-\Lambda^*)}d^d(-k_2)\mathcal{H}^{+-+-}_{11(-2)(-2)}\tilde{J}^{(0)}_{-2} \tag{A.92}$$

where $(-\Lambda^*)$ means that we are integrating over each component of $\mathbf{k}_2$ from $+k_{max}$ to $-k_{max}$ and not from $-k_{max}$ to $+k_{max}$ as it would be for Λ^*. However, the integration over $-\mathbf{k}_2\in(-\Lambda^*)$ is equivalent to the integration over the variable $\mathbf{k}_3\in\Lambda^*$, and this leads to:

$$\Omega_{-1}=24\epsilon\int_{\Lambda^*}d^d(k_3)\mathcal{H}^{+-+-}_{1133}\tilde{J}^{(0)}_3\equiv\Omega_1 \tag{A.93}$$

Space isotropy also implies that $\omega_{\mathbf{k}} = \omega_{-\mathbf{k}}$ ($\omega_{\mathbf{k}}$ is positive for all $\mathbf{k} \in \Lambda^*$) and then:

$$\tilde{\omega}_{\mathbf{k}} + \tilde{\omega}_{-\mathbf{k}} = \omega_{\mathbf{k}} + \Omega_{\mathbf{k}} + \omega_{-\mathbf{k}} + \Omega_{-\mathbf{k}} = 2(\omega_{\mathbf{k}} + \Omega_{\mathbf{k}}) \tag{A.94}$$

Thus, the condition to fulfill for the resonance in the second term (last four lines) of equation (A.90) reads:

$$\omega_{\mathbf{k}} + \Omega_{\mathbf{k}} = 0 \tag{A.95}$$

Also, $\tilde{J}_2^{(0)}$ is positive, whereas the sign of $\mathcal{H}_{1122}^{+-+-}$ implies that it is impossible to generalize without looking at the specific problem we want to describe.

If we take as a paradigmatic example a relatively simple, 4-wave resonant system, namely the *Nonlinear Klein Gordon* system, we easily notice that the Hamiltonian coefficients are strictly positive, see also [56]. A more accurate analysis is needed in other cases, such as the deep water gravity waves, whose effective coefficients have been derived in [206]. If the Hamiltonian coefficients are positive, then $\Omega_{\mathbf{k}}$ is positive too.

Actually, many of the physical systems one usually considers have positive Hamiltonian coefficients. Furthermore, this last condition is even not necessary to satisfy our weaker condition $\Omega_{\mathbf{k}} \geq 0$, $\forall \mathbf{k} \in \Lambda^*$. The reason to rely on such a condition is justified by the fact that those systems enjoy the property:

$$\tilde{\omega}_{\mathbf{k}} = \omega_{\mathbf{k}} + \Omega_{\mathbf{k}} \geq 0, \forall \mathbf{k} \in \Lambda^*. \tag{A.96}$$

Thus, condition (A.95) is never fulfilled, implying that the arguments of the two Dirac delta's $\delta(\tilde{\omega}_1 + \tilde{\omega}_{-1})$ and $\delta(\tilde{\omega}_2 + \tilde{\omega}_{-2})$ in equation (A.90) cannot vanish for any value of $\mathbf{k}$ except from $\mathbf{k} = 0$, but in that case the Hamiltonian coefficients are identically null. As a consequence, for "positive renormalized frequency" systems (*i.e.* satisfying (A.96)) the last four lines of equation (A.90) give zero identically and the dynamical multi-mode equation reduces to the really compact form (A.97) given below. Let us also note that the frequency $\Omega_{\mathbf{k}}$, Eq.(A.44), contains a factor ϵ and that the sum in (A.44) is expected to converge if the energy of the system is finite, then $\Omega_{\mathbf{k}}$ is of order $O(\epsilon)$. Therefore, $\Omega_{\mathbf{k}} \ll \omega_{\mathbf{k}}$. As a matter of fact, even for a system where $\Omega_{\mathbf{k}}$ can be negative, $\Omega_{\mathbf{k}}$ does not nullify the frequency $\omega_{\mathbf{k}}$. Thus, the relevant equation for 4-wave resonant systems is:

$$\begin{aligned}\frac{d\mathcal{Z}[\lambda,\mu,\tau]}{d\tau} &= -192\pi\delta_{\mu,0} \sum_{\underline{\sigma}=(1,\sigma_2,\sigma_3,\sigma_4)} \int d^dk_1 d^dk_2 d^dk_3 d^dk_4 \lambda(\mathbf{k}_1) \left|\mathcal{H}_{1234}^{-\sigma_2\sigma_3\sigma_4}\right|^2 \\ &\quad \delta\left(\tilde{\omega}_{234}^1\right) \delta_{234}^1 \left(\frac{\delta^3\mathcal{Z}}{\delta\lambda(\mathbf{k}_2)\delta\lambda(\mathbf{k}_3)\delta\lambda(\mathbf{k}_4)} - \sigma_2 \frac{\delta^3\mathcal{Z}}{\delta\lambda(\mathbf{k}_1)\delta\lambda(\mathbf{k}_3)\delta\lambda(\mathbf{k}_4)} \right. \\ &\quad \left. -\sigma_3 \frac{\delta^3\mathcal{Z}}{\delta\lambda(\mathbf{k}_1)\delta\lambda(\mathbf{k}_2)\delta\lambda(\mathbf{k}_4)} - \sigma_4 \frac{\delta^3\mathcal{Z}}{\delta\lambda(\mathbf{k}_1)\delta\lambda(\mathbf{k}_2)\delta\lambda(\mathbf{k}_3)} \right)\end{aligned} \tag{A.97}$$

which is a natural generalization to the 4-wave case of Eq.(94) in [69]. It is worth emphasising that this equation has been obtained with the RP assumption but not with the RPA.

A.3.4 Derivation of the spectral hierarchy

We may now consider the characteristic functional of amplitudes only:[3]

$$\mathcal{Z}_L[\lambda, T] \doteq \left\langle \exp\left(\sum_{\mathbf{k}\in\Lambda_L^*} \lambda_{\mathbf{k}} J_{\mathbf{k}}(T) \right) \right\rangle \tag{A.98}$$

In analogy to [69], from (A.97) we derive a hierarchy of evolution equations for the M-mode spectral correlation functions defined in (A.30), in the kinetic limit:

$$\mathcal{N}^{(M)}(\mathbf{k}_1, ..., \mathbf{k}_M, \tau) = \lim_{\epsilon\to 0} \lim_{L\to\infty} \mathcal{N}^{(M)}_{L,\epsilon}(\mathbf{k}_1, ..., \mathbf{k}_M, \epsilon^{-2}\tau). \tag{A.99}$$

The hierarchy is easy to derive knowing the relation

$$\mathcal{N}^{(M)}(\mathbf{k}_1, ..., \mathbf{k}_M, \tau) = (-i)^M \frac{\delta^M \mathcal{Z}[\lambda, \tau]}{\delta\lambda(\mathbf{k}_1)\cdots\delta\lambda(\mathbf{k}_M)}\bigg|_{\lambda=0}. \tag{A.100}$$

By taking M functional derivatives of (A.97) and setting $\lambda \equiv 0$, one obtains:

$$\begin{aligned}
\dot{\mathcal{N}}^{(M)}(\mathbf{k}_1, ..., \mathbf{k}_M, \tau) &= 192\pi \sum_{j=1}^{M} \sum_{\underline{\sigma}} \int d^d\overline{k}_2 d^d\overline{k}_3 d^d\overline{k}_4 \delta\left(\tilde{\omega}^1_{234}\right) \delta^1_{234} |H^{\underline{\sigma}}_{\underline{\mathbf{k}}_j}|^2 \\
&\Big[\mathcal{N}^{(M+2)}(\mathbf{k}_1, ..., \mathbf{k}_{j-1}, \mathbf{k}_{j+1}, ..., \mathbf{k}_M, \overline{\mathbf{k}}_2, \overline{\mathbf{k}}_3, \overline{\mathbf{k}}_4, \tau) - \sigma_2 \mathcal{N}^{(M+2)}(\mathbf{k}_1, ..., \mathbf{k}_M, \overline{\mathbf{k}}_3, \overline{\mathbf{k}}_4, \tau) \\
&- \sigma_3 \mathcal{N}^{(M+2)}(\mathbf{k}_1, ..., \mathbf{k}_M, \overline{\mathbf{k}}_2, \overline{\mathbf{k}}_4, \tau) - \sigma_4 \mathcal{N}^{(M+2)}(\mathbf{k}_1, ..., \mathbf{k}_M, \overline{\mathbf{k}}_2, \overline{\mathbf{k}}_3, \tau)\Big]
\end{aligned} \tag{A.101}$$

We shall refer to this set of equations as to the *spectral hierarchy* of kinetic wave turbulence. It is exactly analogous to the hierarchy derived by Lanford from the BBGKY hierarchy in the low-density limit [116, 117]. If the spectral correlation functions satisfy bounds on their growth for large orders M that allow them to uniquely characterize the distribution of the empirical spectrum, then the spectral hierarchy (A.101) is not only a consequence of the equation (A.97) but is in fact equivalent to that equation. If the initial functional $\mathcal{Z}[\lambda,0]$ is of exponential form (A.28), as follows for an initial RP field with uncorrelated amplitudes, an exact solution of (A.97) is given by:

$$\mathcal{Z}[\lambda, \tau] = \exp\left(i \int d^d\mathbf{k}\, \lambda(\mathbf{k}) n(\mathbf{k}, \tau) \right), \tag{A.102}$$

where $n(\mathbf{k}, \tau)$ satisfies the standard wave kinetic equation with initial condition $n(\mathbf{k},0) = n(\mathbf{k})$. Equivalently, factorized Mth-order correlation functions (A.31) as initial data, entail a factorized solution of the spectral hierarchy (A.101):

$$\mathcal{N}^{(M)}(\mathbf{k}_1, ..., \mathbf{k}_M, \tau) = \prod_{m=1}^{M} n(\mathbf{k}_m, \tau). \tag{A.103}$$

[3] Then, we can consider equation (A.97) without the $\delta_{\mu,0}$ term.

Under suitable conditions [69] this is analogous to the propagation of chaos of Boltzmann's *Stosszahlansatz* [116, 117]. The results above have an important implication. As follows from our discussion in section A.2, the conditions (A.102) or (A.103) imply a law of large numbers for the empirical spectrum at positive times. That is, with probability going to 1 in the kinetic limit (first $L \to \infty$, then $\epsilon \to 0$), it follows that

$$\widehat{n}_L(\mathbf{k}, \epsilon^{-2}\tau) \simeq n(\mathbf{k}, \tau), \quad \tau > 0 \tag{A.104}$$

where $n(\mathbf{k}, \tau)$ is the solution of the wave kinetic equation. An interesting implication for laboratory and numerical experiments is that the wave kinetic equations hold for *typical* initial amplitudes and phases chosen from an RPA ensemble. Some technical comments are in order. As explained in section 2, the very definition of our generating function (A.26) entails that, in the thermodynamic limit, the solution has the form (A.28) if the initial field is RPA. It is important to remark that this is an exact solution of Eq.(A.97), which has been derived asymptotycally under the sole RP assumption. Therefore, the result is not trivial, besides constituting a consistency check.

A.4 Derivation of the PDF equation

With respect to Section A.3, we now consider a second possible limit involving only a fixed number of modes $\mathbf{k}_m$, $m = 1,2,...,M$, as the total number $N \to \infty$. As before, one must keep $\tilde{J}_{\mathbf{k}} = O(1)$ for all modes. We thus define the joint characteristic function:

$$\mathcal{Z}_L^{(M)}(\lambda_1, ..., \lambda_M, \mu_1, ..., T; \mathbf{k}_1, ...) \doteq \left\langle \exp\left[i \sum_{m=1}^{M} \lambda_m \tilde{J}_{\mathbf{k}_M}(T) \right] \prod_{m=1}^{M} \psi_{\mathbf{k}_m}^{\mu_m}(T) \right\rangle \tag{A.105}$$

This is the characteristic function (110) of [69], which corresponds to the generating function (68) of [38] but with $\lambda_{\mathbf{k}_m} = i\left(\frac{L}{2\pi}\right)^d \lambda_m$, $\quad J_{\mathbf{k}_m} = \left(\frac{2\pi}{L}\right)^d \tilde{J}_{\mathbf{k}_m}$, $\quad m = 1, ..., M$, and for all the other modes $\lambda_{\mathbf{k}} = 0$. It also corresponds to the generating functional (5.15) of [137], with same λ_m and $\tilde{J}_m$, but with an imaginary unit in the exponent, and with a finite number M of nonzero arguments.

The reason why λ_m is finite and $\lambda_{\mathbf{k}_m}$ is not is that the exponent of (A.105) contains finitely many terms $\lambda_m \tilde{J}_{\mathbf{k}_m}$, each of which is finite. Then, as $\tilde{J}_{\mathbf{k}_m}$ must be finite as $L \to \infty$, the same holds for λ_m.

We use the symbol $\mathcal{Z}_L^{(M)}(\lambda, \mu, T)$ when there is no possibility of confusion, and we use the perturbation expansion in ϵ giving (A.60) for the generating functions ($\forall M$), with the definitions (A.61) of $\chi_L(\lambda, \mu, T)$ and (A.62) - (A.66) of the $\mathcal{J}$'s. As λ_m is finite, different relations hold for the orders of the prefactors in the $\mathcal{J}$'s. In particular, for $\mathcal{J}_1$, $\mathcal{J}_2$, $\mathcal{J}_3$ we have:

$$\lambda_{\mathbf{k}_1} + \frac{\mu_{\mathbf{k}_1}}{2\tilde{J}_{\mathbf{k}_1}}, \quad \lambda_{\mathbf{k}_1} + \lambda_{\mathbf{k}_1}^2 \tilde{J}_{\mathbf{k}_1} - \frac{\mu_{\mathbf{k}_1}^2}{4\tilde{J}_{\mathbf{k}_1}} = O\left(L^d\right) \tag{A.106}$$

and for $\mathcal{J}_4$, $\mathcal{J}_5$ we have:

$$\frac{1}{2}\lambda_{\mathbf{k}_1}^2+\frac{\mu_{\mathbf{k}_1}}{4\tilde{J}_{\mathbf{k}_1}^2}\left(\frac{\mu_{\mathbf{k}_1}}{2}-1\right)+\frac{\lambda_{\mathbf{k}_1}\mu_{\mathbf{k}_1}}{2\tilde{J}_{\mathbf{k}_1}},\quad \lambda_{\mathbf{k}_1}\lambda_{\mathbf{k}_2},\quad \left(\lambda_{\mathbf{k}_1}+\frac{\mu_{\mathbf{k}_1}}{4\tilde{J}_{\mathbf{k}_1}}\right)\frac{\mu_{\mathbf{k}_2}}{\tilde{J}_{\mathbf{k}_2}}=O\left(L^{2d}\right) \quad \text{(A.107)}$$

To calculate the leading order contributions, one must note that some wavenumbers are discrete and take only M values (mode 1 for $\mathcal{J}_1$ - $\mathcal{J}_4$, modes 1 and 2 for $\mathcal{J}_5$), whereas all the others are continuous in the infinite-box limit. This is important to distinguish $O(L^d)$ from $O(M)$ terms.

A.4.1 Derivation of the PDF hierarchy

Collecting the contributions of the $\mathcal{J}_1-\mathcal{J}_5$ terms enumerated in the Appendix of [33], we can neglect nonsecular terms. Furthermore, $\tilde{\omega}_{\mathbf{k}}+\tilde{\omega}_{-\mathbf{k}}=0$ is never fulfilled, so we ignore the non-resonant terms with a $\delta(\tilde{\omega}_{\mathbf{k}}+\tilde{\omega}_{-\mathbf{k}})$ contribution. The two remaining contributions contain $\delta_{\mu,0}$, hence we can write:

$$\delta_{\mu,0}\langle\tilde{J}_j e^{\sum_m i\lambda_m\tilde{J}_m}\rangle_J=-i\delta_{\mu,0}\frac{\partial}{\partial\lambda_j}\mathcal{Z}^{(M)} \quad \text{(A.108)}$$

for wavenumber $\mathbf{k}_j, j=1,...,M$. Similarly, for mode $\mathbf{k}\neq\mathbf{k}_m, \forall m=0,...,M$ we have:

$$\delta_{\mu,0}i\langle\tilde{J}_{\mathbf{k}}e^{\sum_m i\lambda_m\tilde{J}_m}\rangle_J=\delta_{\mu,0}\left.\frac{\partial}{\partial\lambda_{\mathbf{k}}}\mathcal{Z}^{(M+1)}\right|_{\lambda_{\mathbf{k}}=0} \quad \text{(A.109)}$$

Subsequently, we consider an intermediate time between the linear time and the nonlinear time, cf. Section A.3.3, $T\sim\frac{1}{\epsilon}$, and we take the $\epsilon\to 0$ limit. Because $\mathcal{Z}^{(M)}(\lambda,\mu,T)=\chi^{(M)}(\lambda,\mu,T)+\chi^{(M)*}(-\lambda,-\mu,T)$, while $\chi^{(M)}(\lambda,\mu,T)=\chi^{(M)*}(-\lambda,-\mu,T)$ and $\mathcal{Z}^{(M)}(\lambda,\mu,T)=2\chi^{(M)}(\lambda,\mu,T)$, we get:

$$\frac{\mathcal{Z}^{(M)}(T)-\mathcal{Z}^{(M)}(0)}{\epsilon^2 T}\sim\frac{\partial\mathcal{Z}^{(M)}}{\epsilon^2\partial T}(\lambda,\mu,T)=\frac{\partial\mathcal{Z}^{(M)}}{\partial\tau}(\lambda,\mu,\tau) \quad \text{(A.110)}$$

where $\tau=\epsilon^2 T$ is the nonlinear time. This leads to:[4]

$$\begin{aligned}\frac{\partial\mathcal{Z}^{(M)}}{\partial\tau}(\lambda,\mu,\tau)&=-192\pi\delta_{\mu,0}\\&\times\sum_{j=1}^{M}\sum_{(1,\sigma_2,\sigma_3,\sigma_4)}\int d^d\overline{k}_2 d^d\overline{k}_3 d^d\overline{k}_4\delta^j_{234}\delta\left(\tilde{\omega}^j_{234}\right)\left|\mathcal{H}^{-\sigma_2\sigma_3\sigma_4}_{j234}\right|^2\\&\times\left\{\left(\lambda_j+\lambda_j^2\frac{\partial}{\partial\lambda_j}\right)\left.\frac{\partial^3\mathcal{Z}^{(M+3)}}{\partial\overline{\lambda}_2\partial\overline{\lambda}_3\partial\overline{\lambda}_4}\right|_{\overline{\lambda}_2=\overline{\lambda}_3=\overline{\lambda}_4=0}-\sigma_2\lambda_j\left.\frac{\partial^3\mathcal{Z}^{(M+2)}}{\partial\lambda_j\partial\overline{\lambda}_3\partial\overline{\lambda}_4}\right|_{\overline{\lambda}_3=\overline{\lambda}_4=0}\right.\\&\quad\left.-\sigma_3\lambda_j\left.\frac{\partial^3\mathcal{Z}^{(M+2)}}{\partial\lambda_j\partial\overline{\lambda}_2\partial\overline{\lambda}_4}\right|_{\overline{\lambda}_2=\overline{\lambda}_4=0}-\sigma_4\lambda_j\left.\frac{\partial^3\mathcal{Z}^{(M+2)}}{\partial\lambda_j\partial\overline{\lambda}_2\partial\overline{\lambda}_3}\right|_{\overline{\lambda}_2=\overline{\lambda}_3=0}\right\}\end{aligned} \quad \text{(A.111)}$$

[4]The continuous quantities are identified by a bar, and symmetrization is made in the three continuous modes $\overline{\mathbf{k}}_2$, $\overline{\mathbf{k}}_3$, $\overline{\mathbf{k}}_4$.

for $M = 1,2,3,....$ An important fact is that $\delta_{\mu,0}$ implies that the RP property of the initial wavefield is preserved in time. By Fourier transformation in the λ variables, one can obtain an equivalent hierarchy of equations for the joint PDFs $\mathcal{P}^{(M)}(s_1,...,s_M,\tau;\mathbf{k}_1,...,\mathbf{k}_M)$, which appears more practical to implement boundary conditions on the amplitudes.

A.4.2 The M-mode PDF equations

From the definition of the joint characteristic function of amplitudes, one has:

$$\begin{aligned}\mathcal{Z}^{(M)}(\lambda_1,...,\lambda_M) &= \left\langle e^{\sum_m i\lambda_m s_m}\right\rangle_J \\ &= \int ds_1...ds_M e^{\sum_m i\lambda_m s_m}\mathcal{P}^{(M)}(s_1,...,s_M)\end{aligned} \tag{A.112}$$

$\mathcal{P}^{(M)}$ is the Fourier transform of $\mathcal{Z}^{(M)}$, so that:

$$\mathcal{P}^{(M)}(s_1,...,s_M) = \frac{1}{2\pi}\int d\lambda_1...d\lambda_M e^{-\sum_m i\lambda_m s_m}\mathcal{Z}^{(M)}(\lambda_1,...,\lambda_M) \tag{A.113}$$

A straightforward Fourier transformation yields the following continuity equation:

$$\dot{\mathcal{P}}^{(M)} + \sum_{m=1}^{M}\frac{\partial}{\partial s_m}\mathcal{F}_m^{(M)} = 0, \tag{A.114}$$

$$\begin{aligned}\mathcal{F}_m^{(M)} = &-192\pi s_m \sum_{\underline{\sigma}=(1,\sigma_2,\sigma_3,\sigma_4)}\int d^d\overline{k}_2 d^d\overline{k}_3 d^d\overline{k}_4 \delta^m_{234}\delta\left(\tilde{\omega}^m_{234}\right)\left|\mathcal{H}^{-\sigma_2\sigma_3\sigma_4}_{m234}\right|^2 \\ &\times\Bigg[\int d\overline{s}_2 d\overline{s}_3 d\overline{s}_4 \overline{s}_2\overline{s}_3\overline{s}_4\frac{\partial\mathcal{P}^{(M+3)}}{\partial s_m} + \sigma_2\int d\overline{s}_3 d\overline{s}_4\overline{s}_3\overline{s}_4\mathcal{P}^{(M+2)} \\ &+ \sigma_3\int d\overline{s}_2 d\overline{s}_4\overline{s}_2\overline{s}_4\mathcal{P}^{(M+2)} + \sigma_4\int d\overline{s}_2 d\overline{s}_3\overline{s}_2\overline{s}_3\mathcal{P}^{(M+2)}\Bigg]\end{aligned} \tag{A.115}$$

This is not a closed equation for $\mathcal{P}^{(M)}$, as it contains $\mathcal{P}^{(M+2)}$ and $\mathcal{P}^{(M+3)}$, for $M = 1,2,3,...$

A.4.3 Relation with Peierls equation

As recalled above, a similar diagrammatic calculation for the 3-wave case was performed in Ref. [38]. Starting from the same defintion of generating function adopted here, the authors derived the canonical Peierls equation, in their version of the thermodynamic limit [38, 137]. Later, it was shown that certain terms contributing to the Peierls equation are actually negligible, if the variables are normalized so that the characteristic functional remains finite in the thermodynamic limit [69]. Consequently, an equation that at the leading order differs from the Peierls equation

was derived in Ref. [69].

Because the PDF, rather than the generating function, is the object of physical interest, in this subsection we investigate the relation between the two asymptotic equations for the PDF, and we show that under two assumptions the Peierls equation reduces to the other leading order equation. We compare our PDF equation (A.114,A.115), that follows from the leading-order equation (A.111), with the 4-wave Peierls PDF equation, Eq.(6.120) of Ref. [137], that has been derived taking the Laplace transform of the generating function equation, obtained in the thermodynamic limit, in formal analogy with the 3-wave case. Such a PDF equation, that takes the form:

$$\begin{aligned}\dot{\mathcal{P}} = \quad & \pi\epsilon^4 \int |W^{ij}_{nm}|^2 \delta(\omega^{ij}) \delta^{ij}_{nm} \left[\tfrac{\delta}{\delta s_j} + \tfrac{\delta}{\delta s_l} - \tfrac{\delta}{\delta s_m} - \tfrac{\delta}{\delta s_n}\right] \\ & \times \left(s_j s_l s_m s_n \left[\tfrac{\delta}{\delta s_j} + \tfrac{\delta}{\delta s_l} - \tfrac{\delta}{\delta s_m} - \tfrac{\delta}{\delta s_n}\right] \mathcal{P}\right) d\mathbf{k}_j d\mathbf{k}_l d\mathbf{k}_m d\mathbf{k}_n \end{aligned} \tag{A.116}$$

is meant to describe the behaviour of an infinite set of modes. Unlike our case, there are no spins in Eq.(A.116), but this is irrelevant for the following discussion. Equation (A.116) can also be written as a continuity equation, which reads:

$$\dot{\mathcal{P}} + \int \frac{\partial}{\partial s_j} \mathcal{F}_j d\mathbf{k}_j = 0, \tag{A.117}$$

$$\mathcal{F}_j = -4\pi\epsilon^4 s_j \int |W^{lj}_{nm}|^2 \delta(\omega^{lj}) \delta^{lj}_{nm} s_j s_l s_m s_n \left[\frac{\delta}{\delta s_j} + \frac{\delta}{\delta s_l} - \frac{\delta}{\delta s_m} - \frac{\delta}{\delta s_n}\right] \mathcal{P}\, d\mathbf{k}_l d\mathbf{k}_m d\mathbf{k}_n \,. \tag{A.118}$$

To compare with our M-mode equation, let us assume that the Peierls equation holds with same form also in the case of large but finite N, so that we can write:

$$\dot{\mathcal{P}}^{(N)} + \sum_{j=1}^{N} \frac{\partial}{\partial s_j} \mathcal{F}^{(N)}_j = 0, \tag{A.119}$$

$$\mathcal{F}^{(N)}_j = -4\pi\epsilon^4 s_j \left(\frac{2\pi}{L}\right)^{3d} \sum_{l,m,n=1}^{N} |W^{lj}_{nm}|^2 \delta(\omega^{lj}) \delta^{lj}_{nm} s_j s_l s_m s_n \left[\frac{\delta}{\delta s_j} + \frac{\delta}{\delta s_l} - \frac{\delta}{\delta s_m} - \frac{\delta}{\delta s_n}\right] \mathcal{P}^{(N)} \,. \tag{A.120}$$

This is tantamount to commute the thermodynamic limit and the $T \sim 1/\epsilon \to \infty$ limit with the Laplace transform, if the $N \to \infty$ limit can be taken without further specifications.[5] Now, one can integrate out $N - M$ variables, as in the standard BBGKY procedure, to obtain

$$\dot{\mathcal{P}}^{(M)} + \sum_{j=1}^{M} \frac{\partial}{\partial s_j} \mathcal{F}^{(M)}_j = 0, \tag{A.121}$$

[5]Strictly speaking, here the derivative is only a finite difference.

with the flux given by

$$\mathcal{F}_j^{(M)} = -4\pi\epsilon^4\left(\frac{2\pi}{L}\right)^{3d}\times \left\{\sum_{l,m,n=M+1}^{N}|W_{nm}^{lj}|^2\delta(\omega^{lj})\delta_{nm}^{lj}\left(\int s_l s_m s_n\frac{\partial\mathcal{P}^{M+3}}{\partial s_j}ds_l ds_m ds_n - \int s_m s_n\mathcal{P}^{M+2}ds_m ds_n + 2\int s_m s_l\mathcal{P}^{M+2}ds_m ds_l\right)\right.$$

$$+\sum_{m,n=1}^{M}\sum_{l=M+1}^{N}|W_{nm}^{lj}|^2\delta(\omega^{lj})\delta_{nm}^{lj}\left[-s_j s_m s_n\left(\mathcal{P}^M+\frac{\partial}{\partial s_j}\int ds_l s_l\mathcal{P}^{M+1}\right)\right]$$

$$\left.+\sum_{l=1}^{M}\sum_{m,n=M+1}^{N}|W_{nm}^{lj}|^2\delta(\omega^{lj})\delta_{nm}^{lj}2\left(\int s_l s_m s_n\frac{\partial\mathcal{P}^{M+2}}{\partial s_j}ds_m ds_n+\int s_l s_m\mathcal{P}^{M+1}ds_m\right)\right\}. \tag{A.122}$$

Analogously to the analysis of Ref. [69] for the 3-wave case, we note that this flux contains more terms than the leading order (A.114)-(A.115). Nonetheless, taking $N \gg M$, and assuming that all the terms are individually of the same size, one may obtain the leading order only from the sum having all three indices l, m, n in $[N-M, N]$. The remaining terms can be neglected simply because they constitute a negligibly small set compared to the others. Under this assumption, the last two lines of Eq.(A.123) can be discarded, and the flux can be written as:

$$\mathcal{F}_j^{(M)} = -4\pi\epsilon^4\left(\frac{2\pi}{L}\right)^{3d}\sum_{l,m,n=1}^{N}|W_{nm}^{lj}|^2\delta(\omega^{lj})\delta_{nm}^{lj}\left\{\left(\int s_l s_m s_n\frac{\partial\mathcal{P}^{M+3}}{\partial s_j}ds_l ds_m ds_n - \int s_m s_n\mathcal{P}^{M+2}ds_m ds_n + 2\int s_m s_l\mathcal{P}^{M+2}ds_m ds_l\right)\right., \tag{A.123}$$

Then, taking the thermodynamic limit ($N, L \to \infty$) leads to our equation (A.115).

In summary, the procedure based on the Peierls equation leads to our same results, provided: (i) the thermodynamic limit is not singular; (ii) the wave modes in the first M modes can be neglected, compared to all the others.

A.4.4 The 1-mode PDF equation

It is interesting to note that factorized initial conditions (which is equivalent to RPA property at the initial time) imply factorized solutions for Eq.(A.115), $\forall\tau \geq 0$:

$$\mathcal{Z}^{(M)}(\lambda_1, ..., \lambda_M, \tau; \mathbf{k}_1, ..., \mathbf{k}_M) = \prod_{m=1}^{M}\mathcal{Z}(\lambda_m, \tau; \mathbf{k}_m), \quad \tau \geq 0 \tag{A.124}$$

with each $\mathcal{Z}(\lambda_{\mathbf{k}},\tau;\mathbf{k})$ satisfying

$$\frac{\partial\mathcal{Z}(\lambda_{\mathbf{k}},\tau;\mathbf{k})}{\partial\tau}=i\eta_{\mathbf{k}}\lambda_{\mathbf{k}}\Big(1+\lambda_{\mathbf{k}}\frac{\partial}{\partial\lambda_{\mathbf{k}}}\Big)\mathcal{Z}(\lambda_{\mathbf{k}},\tau;\mathbf{k})-\gamma_{\mathbf{k}}\lambda_{\mathbf{k}}\frac{\partial\mathcal{Z}}{\partial\lambda_{\mathbf{k}}}(\lambda_{\mathbf{k}},\tau;\mathbf{k}) \tag{A.125}$$

where

$$\eta_{\mathbf{k}}\doteq 192\pi\sum_{\underline{\sigma}}\int d^d\mathbf{k}_2d^d\mathbf{k}_3d^d\mathbf{k}_4\delta^{\mathbf{k}}_{234}\delta\left(\tilde{\omega}^{\mathbf{k}}_{234}\right)\left|\mathcal{H}^{-\sigma_2\sigma_3\sigma_4}_{\mathbf{k}234}\right|^2 n(\mathbf{k}_2)n(\mathbf{k}_3)n(\mathbf{k}_4)\geq 0, \tag{A.126}$$

$$\begin{aligned}\gamma_{\mathbf{k}}\doteq 192\pi\sum_{\underline{\sigma}}\int d^d\mathbf{k}_2d^d\mathbf{k}_3d^d\mathbf{k}_4\delta^{\mathbf{k}}_{234}\delta\left(\tilde{\omega}^{\mathbf{k}}_{234}\right)\left|\mathcal{H}^{-\sigma_2\sigma_3\sigma_4}_{\mathbf{k}234}\right|^2\\ \times\Big[\sigma_2n(\mathbf{k}_3)n(\mathbf{k}_4)+\sigma_3n(\mathbf{k}_2)n(\mathbf{k}_3)+\sigma_4n(\mathbf{k}_2)n(\mathbf{k}_3)\Big]\end{aligned} \tag{A.127}$$

For the PDF hierarchy an analogous result holds. Substituting a factorized solution into Eq.(A.114) we get the equation for the 1-mode PDF:

$$\mathcal{P}^{(M)}(s_1,...,s_M,\tau;\mathbf{k}_1,...,\mathbf{k}_M)=\prod_{m=1}^{M}P(s_m,\tau;\mathbf{k}_m)\doteq\prod_{m=1}^{M}P_m,\quad\tau\geq 0 \tag{A.128}$$

Equation (A.114) transforms into:

$$\begin{aligned}\prod_{m\neq1}P_m\frac{\partial P_1}{\partial\tau}+\prod_{m\neq2}P_m\frac{\partial P_2}{\partial\tau}...=\frac{\partial}{\partial s_1}\Bigg\{192\pi s_1\sum_{\underline{\sigma}}\int d^dk_2d^dk_3d^dk_4\delta^1_{234}\delta\left(\tilde{\omega}^1_{234}\right)\\ \times\left|\mathcal{H}^{-\sigma_2\sigma_3\sigma_4}_{1234}\right|^2\prod_{m\neq1}P_m\Bigg[\frac{\partial\mathcal{P}_1}{\partial s_1}\int d\bar{s}_2\bar{s}_2P_2\int d\bar{s}_3\bar{s}_3P_3\int d\bar{s}_4\bar{s}_4P_4\\ +\Big(\sigma_2P_1\int d\bar{s}_3\bar{s}_3P_3\int d\bar{s}_4\bar{s}_4P_4+(2\leftrightarrow3)+(2\leftrightarrow4)\Big)\Bigg]+...\Bigg\}\end{aligned} \tag{A.129}$$

Recall that $\int ds_is_iP(s_i,\tau;\mathbf{k}_i)=n(\mathbf{k}_i,\tau)$, because of the definition of the wave spectrum. Equation (A.129) is made of M independent parts, each of which can be written in the continuity equation form:

$$\frac{\partial P}{\partial\tau}=\frac{\partial}{\partial s}\Big[s\Big(\eta_{\mathbf{k}}\frac{\partial P}{\partial s}+\gamma_{\mathbf{k}}P\Big)\Big] \tag{A.130}$$

where $\eta_{\mathbf{k}}$ and $\gamma_{\mathbf{k}}$ are the same defined in (A.126) and (A.127). These are nonlinear Markov evolution equations in the sense of McKean, since the solutions satisfy the set of self-consistency conditions:

$$n(\mathbf{k},\tau)=\int ds\;sP(s,\tau;\mathbf{k}) \tag{A.131}$$

where $n(\mathbf{k},\tau)$ is the same spectrum that appears in the formulas for the coefficients (A.126) and (A.127).

These equations are the exact solutions of a model of "self-consistent Langevin equations". Here, the model equations take the form of the stochastic differential equations

$$ds_{\mathbf{k}} = (\eta_{\mathbf{k}} - \gamma_{\mathbf{k}} s_{\mathbf{k}})d\tau + \sqrt{2\eta_{\mathbf{k}} s_{\mathbf{k}}} dW_{\mathbf{k}}, \tag{A.132}$$

interpreted in the Ito sense, with self-consistency determination of $n(\mathbf{k}, \tau)$ via (A.131). This generalizes the 3-wave case of Ref. [69] where, $P(s, \tau; \mathbf{k})$ relaxes to

$$Q(s, \tau; \mathbf{k}) = \frac{1}{n(\mathbf{k}, \tau)} \exp(-s/n(\mathbf{k}, \tau)), \tag{A.133}$$

which corresponds to a Gaussian distribution of the canonical variable $\tilde{b} = \left(\frac{2\pi}{L}\right)^{d/2} b$. For any solution $n(\mathbf{k}, \tau)$ of the wave kinetic equation, $Q(s, \tau; \mathbf{k})$ solves the 1-mode PDF equation (A.130). Also, the relaxation of a general solution P to Q is indicated by an H-theorem for the *relative entropy*

$$H(P|Q) = \int ds\ P(s) \ln\left(\frac{P(s)}{Q(s)}\right) = \int ds\ P(s) \ln P(s) + \ln n + 1. \tag{A.134}$$

This is a convex function of P, non-negative, and vanishing only for $P = Q$. Taking the time-derivative using (A.130), it is straightforward to derive

$$\frac{d}{d\tau} H(P(\tau)|Q(\tau)) = -\eta \int ds\ \frac{s|\partial_s P(s, \tau)|^2}{P(s, \tau)} + \frac{\eta}{n(\tau)}, \tag{A.135}$$

where

$$\int -s\partial_s P(s, \tau)\, ds = \int P(s, \tau)\, ds = 1 \tag{A.136}$$

is used to cancel terms involving γ. The self-consistency condition $n(\tau) = \int s\, P(s, \tau)\, ds$ implies

$$\frac{d}{d\tau} H(P(\tau)|Q(\tau)) = -\eta \left(\int ds\ \frac{s|\partial_s P(s, \tau)|^2}{P(s, \tau)} - \frac{1}{\int s\, P(s, \tau)\, ds} \right) \leq 0. \tag{A.137}$$

The inequality follows from the Cauchy-Schwartz inequality applied to (A.136)

$$1 = \int \sqrt{sP} \cdot \sqrt{\frac{s}{P}}(-\partial_s P)\, ds \leq \sqrt{\int sP(s)\, ds \cdot \int \frac{s|\partial_s P|^2}{P}\, ds}. \tag{A.138}$$

Equality holds and relative entropy production vanishes if and only if $\sqrt{sP} = c\sqrt{\frac{s}{P}}(-\partial_s P)$, or $P = -c\partial_s P$ for some c. The solution of this latter equation gives $P = Q$ with $n = c$. Then, $P(\tau)$ relaxes to $Q(\tau)$ as τ increases, assuming that kinetic theory holds over the entire amplitudes range $s \in (0, \infty)$.

A.5 Conclusions

1. We have worked within the framework of WWT. We have considered a *Hamiltonian system* in d dimensions, with a quartic small perturbation implying *4-wave interactions*. From Hamilton equations, we have derived the equations of motion expressed in canonical normal variables.

2. To reach a closure for the problem, we have assumed that the canonical wavefield has the RP property at the initial time, allowing a statistical description of the field through its modes. We have averaged over phases using a method based on the *Feynman-Wyld diagrams*.

3. For the large-box limit, we have normalized the amplitudes to keep the wave spectrum finite, which is crucial for a correct evaluation of the contributions of the different diagrams [69]. The result differs significantly from the previous approach of Ref. [137], but it has been shown that the approach of Ref. [137] is equivalent to ours, under two technical assumptions.

4. We have formally taken the large-box (*thermodynamic*) limit, followed by the small nonlinearity limit, obtaining the following closed equation:

$$\begin{aligned}\frac{d\mathcal{Z}[\lambda,\mu,\tau]}{d\tau} &= -192\pi\delta_{\mu,0}\sum_{\underline{\sigma}}\int d^dk_1d^dk_2d^dk_3d^dk_4\lambda\left(\mathbf{k}_1\right)|\mathcal{H}_{1234}^{-\sigma_2\sigma_3\sigma_4}|^2\delta(\tilde{\omega}_{234}^1)\\ &\times\delta_{234}^1\left(\frac{\delta^3\mathcal{Z}}{\delta\lambda(\mathbf{k}_2)\delta\lambda(\mathbf{k}_3)\delta\lambda(\mathbf{k}_4)}-\sigma_2\frac{\delta^3\mathcal{Z}}{\delta\lambda(\mathbf{k}_1)\delta\lambda(\mathbf{k}_3)\delta\lambda(\mathbf{k}_4)}+\right.\\ &\left.-\sigma_3\frac{\delta^3\mathcal{Z}}{\delta\lambda(\mathbf{k}_1)\delta\lambda(\mathbf{k}_2)\delta\lambda(\mathbf{k}_4)}-\sigma_4\frac{\delta^3\mathcal{Z}}{\delta\lambda(\mathbf{k}_1)\delta\lambda(\mathbf{k}_2)\delta\lambda(\mathbf{k}_3)}\right)\end{aligned} \tag{A.139}$$

   where τ is the nonlinear time. Note that:

   - Due to the $\delta_{\mu,0}$ factor, the RP property of the initial field is preserved as time goes on. This fact is crucial as it ensures the validity of the equation itself at $\tau > 0$.
   - The stricter initial RPA property for the wavefield, and thus a factorized form for $\mathcal{Z}[\lambda,\mu,0]$, entails a solution preserving the factorized form in time.
   - Differentiating the characteristic functional in the variables $\lambda_{\mathbf{k}}$'s, one obtains the *spectral hierarchy*, which is analogous to the BBGKY hierarchy of Kinetic Theory. Assuming RPA for the initial field, the hierarchy is closed obtaining the kinetic wave equation for the spectrum. This connects our work, that gives for the first time the general derivation for the 4-wave case, with the existing literature.

5. We have defined a different *characteristic function*, for a *finite number of modes* M, and derived a hierarchy of equations for its time evolution for any value of M, which reads:

$$\begin{aligned}
&\frac{\partial \mathcal{Z}^{(M)}}{\partial \tau}(\lambda,\mu,\tau) = -192\pi\delta_{\mu,0} \\
&\times \sum_{j=1}^{M} \sum_{(1,\sigma_2,\sigma_3,\sigma_4)} \int d^d\overline{k}_2 d^d\overline{k}_3 d^d\overline{k}_4 \delta^j_{234}\delta\left(\tilde{\omega}^j_{234}\right)\left|\mathcal{H}^{-\sigma_2\sigma_3\sigma_4}_{j234}\right|^2 \\
&\times\Bigg\{\left(\lambda_j+\lambda_j^2\frac{\partial}{\partial\lambda_j}\right)\frac{\partial^3\mathcal{Z}^{(M+3)}}{\partial\overline{\lambda}_2\partial\overline{\lambda}_3\partial\overline{\lambda}_4}\Bigg|_{\overline{\lambda}_2=\overline{\lambda}_3=\overline{\lambda}_4=0} - \sigma_2\lambda_j\frac{\partial^3\mathcal{Z}^{(M+2)}}{\partial\lambda_j\partial\overline{\lambda}_3\partial\overline{\lambda}_4}\Bigg|_{\overline{\lambda}_3=\overline{\lambda}_4=0} \\
&\quad - \sigma_3\lambda_j\frac{\partial^3\mathcal{Z}^{(M+2)}}{\partial\lambda_j\partial\overline{\lambda}_2\partial\overline{\lambda}_4}\Bigg|_{\overline{\lambda}_2=\overline{\lambda}_4=0} - \sigma_4\lambda_j\frac{\partial^3\mathcal{Z}^{(M+2)}}{\partial\lambda_j\partial\overline{\lambda}_2\partial\overline{\lambda}_3}\Bigg|_{\overline{\lambda}_2=\overline{\lambda}_3=0}\Bigg\}
\end{aligned} \tag{A.140}$$

6. By taking the Fourier transform of equation (A.140), we have derived a hierarchy for the *M-mode joint PDFs*, which can be written in *continuity equation* form:

$$\dot{\mathcal{P}}^{(M)} + \sum_{m=1}^{M}\frac{\partial}{\partial s_m}\mathcal{F}^{(M)}_m = 0, \tag{A.141}$$

$$\begin{aligned}
\mathcal{F}^{(M)}_m =& -192\pi s_m \sum_{\underline{\sigma}=(1,\sigma_2,\sigma_3,\sigma_4)} \int d^d\overline{k}_2 d^d\overline{k}_3 d^d\overline{k}_4 \delta^m_{234}\delta\left(\tilde{\omega}^m_{234}\right)\left|\mathcal{H}^{-\sigma_2\sigma_3\sigma_4}_{m234}\right|^2 \\
&\times\Bigg[\int d\overline{s}_2 d\overline{s}_3 d\overline{s}_4 \overline{s}_2\overline{s}_3\overline{s}_4\frac{\partial\mathcal{P}^{(M+3)}}{\partial s_m} + \sigma_2\int d\overline{s}_3 d\overline{s}_4\overline{s}_3\overline{s}_4\mathcal{P}^{(M+2)} \\
&\quad + \sigma_3\int d\overline{s}_2 d\overline{s}_4\overline{s}_2\overline{s}_4\mathcal{P}^{(M+2)} + \sigma_4\int d\overline{s}_2 d\overline{s}_3\overline{s}_2\overline{s}_3\mathcal{P}^{(M+2)}\Bigg]
\end{aligned}$$

where $\mathcal{F}^{(M)}_m$ is the flux for one of the M modes. As in the case of Eq.(A.139) we have:

- RP property for the wavefield at $\tau = 0$ remains fullfilled for the field at $\tau > 0$. So, equation (A.140) is valid for any nonlinear time $\tau \geq 0$.
- An initial RPA field remains RPA under the evolution of Eq.(A.141).
- Under RPA, the hierarchy (A.141) can be closed to yield the the 1-mode PDF equation:

$$\frac{\partial P}{\partial\tau} = \frac{\partial}{\partial s}\left[s\left(\eta_{\mathbf{k}}\frac{\partial P}{\partial s} + \gamma_{\mathbf{k}}P\right)\right], \tag{A.142}$$

$$\eta_{\mathbf{k}} \doteq 192\pi \sum_{\underline{\sigma}} \int d^d\mathbf{k}_2 d^d\mathbf{k}_3 d^d\mathbf{k}_4 \delta^{\mathbf{k}}_{234} \delta\left(\tilde{\omega}^{\mathbf{k}}_{234}\right) \left|\mathcal{H}^{-\sigma_2\sigma_3\sigma_4}_{\mathbf{k}234}\right|^2 \times n(\mathbf{k}_2)n(\mathbf{k}_3)n(\mathbf{k}_4) \geq 0, \tag{A.143}$$

$$\gamma_{\mathbf{k}} \doteq 192\pi \sum_{\underline{\sigma}} \int d^d\mathbf{k}_2 d^d\mathbf{k}_3 d^d\mathbf{k}_4 \delta^{\mathbf{k}}_{234} \delta\left(\tilde{\omega}^{\mathbf{k}}_{234}\right) \left|\mathcal{H}^{-\sigma_2\sigma_3\sigma_4}_{\mathbf{k}234}\right|^2 \times \Big[\sigma_2 n(\mathbf{k}_3)n(\mathbf{k}_4) + \sigma_3 n(\mathbf{k}_2)n(\mathbf{k}_3) + \sigma_4 n(\mathbf{k}_2)n(\mathbf{k}_3)\Big] \tag{A.144}$$

that can be efficiently treated numerically. The spectrum $n(\mathbf{k})$ in $\eta_{\mathbf{k}}$ and $\gamma_{\mathbf{k}}$ can be determined using the kinetic equation.

- Our Eq.(A.142) is more general than Eq.(6.51) of [137], as it contains all interactions, not only the "2 waves → 2 waves" interactions.[6]
- An important solution to (A.145) is represented by the Rayleigh distribution:

$$Q(s,\tau;\mathbf{k}) = \frac{1}{n(\mathbf{k},\tau)} e^{-s/n(\mathbf{k},\tau)} \tag{A.145}$$

corresponding to equilibrium. In absence of forcing and damping, P tends to the Rayleigh form (A.145) for any typical initial condition. This was tested numerically in [32].

7. In the most general case, Eqs.(A.139,A.140,A.141) would have some supplementary terms (see equation (A.90)). However, as argued in Section A.3.3, we think they are irrelevant for the known physical systems of wave turbulence, since the resonant condition is never fulfilled.

8. For any system where the leading nonlinear phenomena are N-wave resonances, our results suggest the conjecture that the coefficient preceding the right-hand side in Eq.(A.139) equals $12i^{2-N}A_N$, where A_N is a number and $A_3 = 3, A_4 = 16$. Integration over the N wavenumber variables, on which also the Hamiltonian coefficients and the two delta's depend yields:

$$\left(\frac{\delta^{N-1}\mathcal{Z}}{\delta\lambda(\mathbf{k}_2)\delta\lambda(\mathbf{k}_3)...\delta\lambda(\mathbf{k}_N)} - \sum_{i=2}^{N} \sigma_i \frac{\delta^{N-1}\mathcal{Z}}{\prod_{j\neq i}\delta\lambda(\mathbf{k}_i)}\right) \tag{A.146}$$

9. We conclude noting that our derivation of the wave kinetics is not mathematically rigorous, as is common in the specialized literature. In particular, analogously to Ref. [69] for 3-wave systems, we have not shown that $O(\epsilon^3)$ terms are negligible in the perturbation expansion (A.48). The kinetic limit

[6] Remarkably, see the system of vibrating elastic plates treated in [55].

consists indeed of a delicate combination of large box and small nonlinearity limits [177], whereas rigorous proofs based on asymptotic methods are problematic and presently limited to particular systems, see e.g. Ref. [126]. At the same time, treating a discrete system in a finite volume and successively taking a suitable large system limit makes physical sense: the various quantities are well defined, classical mechanics issues are naturally cast in a discrete formalism, and splitting schemes can be mathematically justified in a variety of circumstances, including kinetic equations [184, 51, 162]. Furthermore, this approach allows us to identify and test the properties of the leading order equations of the 4-wave dynamics. As a matter of fact, the recent work [32] has demonstrated the agreement of part of our results with the kinetic equation in [55], that had been derived through asymptotic methods [56]. Reference [32] also shows the agreement of the predictions of the PDF equation with direct numerical simulations of relevant 4-wave systems. This further vindicates the approach developed.

Appendix B

Supplemental material to Chapter 4

The material in this chapter was published as *Supporting Information* to

G. Dematteis, T. Grafke, and E. Vanden-Eijnden. "Rogue waves and large deviations in deep sea". Proceedings of the National Academy of Sciences, page 201710670, 2018.

B.1 Construction of the initial condition and dynamical consistency check

Our procedure requires to specify the statistics of the (complex) envelope at initial time, $u_0(x)$, whereas the experimental spectrum is for the surface elevation $\eta(x)$ which is related to $u_0(x)$ as

$$\eta(x) = \Re\Big(u_0(x)e^{ik_0x}\Big), \tag{B.1}$$

To construct the intial $u_0(x)$, we introduce the auxiliary variable $\zeta(x)$, [150]

$$\zeta(x) = \Im\Big(u_0(x)e^{ik_0x}\Big), \tag{B.2}$$

which we treat as a field independent of $\eta(x)$, with the same statistics. It is easy to see from Eqs. (B.1) and (B.2) that the envelope $u_0(x)$ can then be expressed as

$$u_0(x) = \Big(\eta(x) + i\zeta(x)\Big)e^{-ik_0x}\,. \tag{B.3}$$

Assuming that both $\eta(x)$ and $\zeta(x)$ are independent Gaussian fields with covariance $\mathbb{E}(\eta(x)\eta(x')) = \mathbb{E}(\zeta(x)\zeta(x')) = C_\eta(x-x')$, the envelope $u_0(x)$ is also Gaussian, with covariance $C_u(x-x') = \mathbb{E}(u(x)\bar{u}(x'))$ given by

$$C_u(x-x') = 2C_\eta(x-x')e^{-ik_0(x-x')}\,. \tag{B.4}$$

This relation implies that

$$\hat{C}_u(k) = 2\hat{C}_\eta(k + k_0)\,. \tag{B.5}$$

where we defined

$$\hat{C}_u(k) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-ikx} C_u(x)dx, \qquad \hat{C}_\eta(k) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-ikx} C_\eta(x)dx. \tag{B.6}$$

Recalling that k_0 is defined as the wave vector at which the spectrum of $\eta(x)$ should be centered, if we take a Gaussian ansatz for this spectrum, we should pick

$$\hat{C}_\eta(k) = \hat{C}_\eta(0)e^{-|k-k_0|^2/(2\Delta^2)}\,. \tag{B.7}$$

As a result,

$$\hat{C}_u(k) = 2\hat{C}_\eta(0)e^{-k^2/(2\Delta^2)}\,. \tag{B.8}$$

The spectrum for $u_0(x)$ used in the following is a discretized version of the one above, with $A = (2\pi/L)^2\hat{C}_\eta(0)$.

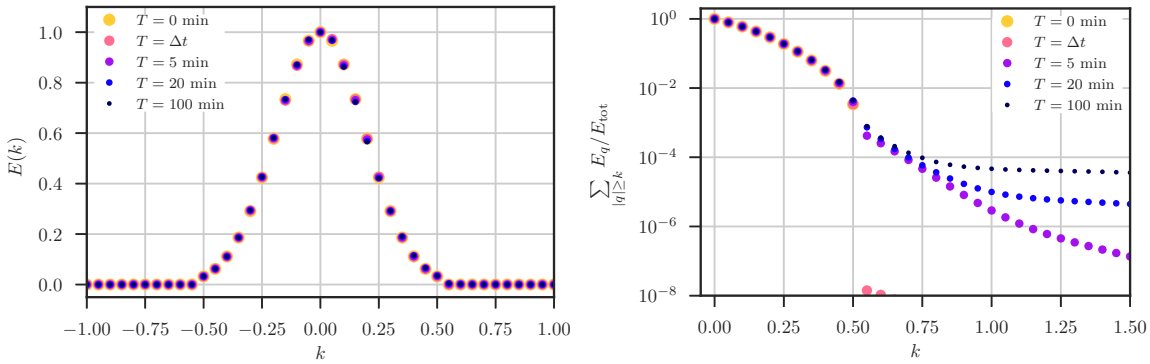

Figure B.1. Evolution of the spectrum of $u(t, x)$. The left panel shows that this spectrum stays essentialy constant in time over 100 min, which justifies our choice of prior: indeed, from the viewpoint of this prior, the time evolution of $u(t, x)$ leads to no changes. Of course, some features of $u(t, x)$ change, as apparent from the evolution of other observables such as $P_t(z) = \mathbb{P}(\max_x |u(t, x)|)$: These changes can be detected in the spectrum, but they require us to look at much finer energy scales, as shown on the right panel where we plot the energy contained in modes above $k > 0$ as k increases.

The results reported in main text require us to evolve the field $u(t, x)$ from its initial condition $u_0(x)$. As explained in main text, through this evolution, the probabities $P_t(z) = \mathbb{P}(\max_x |u(t, x)|)$ change with time t until they converge to some limit value. It is interesting to ask how much this evolution changes the prior information we used to construct the initial $u_0(x)$: that is, it is interesting to look at the spectrum of $u(t, x)$ and see how much it differs from that of $u_0(x)$. The results

of this calculation are shown in Fig. B.1, and they indicate that the spectrum stays essentialy constant in time over 100 min. This justifies our choice of prior: indeed, from the viewpoint of this prior, the time evolution of $u(t,x)$ leads to no significant changes. Of course, some features of $u(t,x)$ change, as apparent from the evolution of other observables such as $P_t(z) = \mathbb{P}(\max_x |u(t,x)|)$. Detecting the trace of these changes in the spectrum requires one to look at much finer energy scales: this can be seen on the right panel in Fig. B.1 where we plot the energy contained in modes above $k > 0$ for increasing values of k.

B.2 Influence of the size of the domain and of the observation window

In this section, we investigate the influence of the size of the domain and/or that of the observation window on our results. To this end, we conduct experiments in domains of size $L = L_0 = 40\pi$ (the domain size used in main text, which is $L_0 \approx 4.53 \times 10^3$ m in dimensional units), and compare to $L = 2L_0$, $L = 4L_0$, and $L = 8L_0$. The base domain size L_0 was chosen to be as small as possible for computational efficiency, but still large enough that the influence of the periodic boundary conditions be negligible (as checked below). Consequently, *the results below can be interpreted either by thinking of $L \geq L_0$ as the actual domain size, or as the size of the observation window in an even larger domain (including one that could be infinite).* We also stress that our results are numerically converged and consistent in terms of numerical resolution, in the sense that we doubled both the number of gridpoints in the domain and the number of modes in the initial data each time we doubled the domain size. In particular, we used 2^{12} gridpoints and $M = 23$ initial modes ($-11 \leq n \leq 11$) in the domain of size L, 2^{13} gridpoints and $M = 47$ initial modes ($-23 \leq n \leq 23$) in the domain of size $2L$, etc.

We begin by checking that the domain of size $L_0 = 40\pi$ is already large enough to render negligible the effect of the boundary conditions. To this end, let us consider a different observable than the one in main text, namely the probability that $|u(t,x)|$ be above a certain threshold at a given location x_0 in the domain,

$$P_0^L(t,z) = \mathbb{P}(|u(t,x_0)| > z), \qquad x_0 \in [0,L]. \tag{B.9}$$

By translational invariance, $P_0^L(t,z)$ is independent of x_0. As $L \to \infty$ this probability converges to a limiting value, $P_0^L(t,z) \to P_0(t,z)$, which makes it useful to consider here. As can be seen from Fig. B.2, convergence is already achieved for $L = L_0$, $P_0^{L_0}(t,z) \approx P_0(t,z)$. The results shown in Fig. B.2 are for $t = 15$ min, when the probability has converged to that on the invariant measure already. A similar conclusion can be made at intermediate times: Fig. B.3 shows that doubling the domain size makes no significant difference, i.e. $P_0^{2L_0}(t,z) \approx P_0^{L_0}(t,z)$, both in the results from Monte-Carlo sampling and in those from our large deviation approach.

The same invariance is also observed in the trajectories obtained by optimization in the large deviation approach, see Fig. B.4. Note that these results are not surprising since L_0 is already much larger than the correlation length of the initial field, $L_0 \simeq 10L_c$ – this is in fact why this value of L_0 was chosen to begin with.

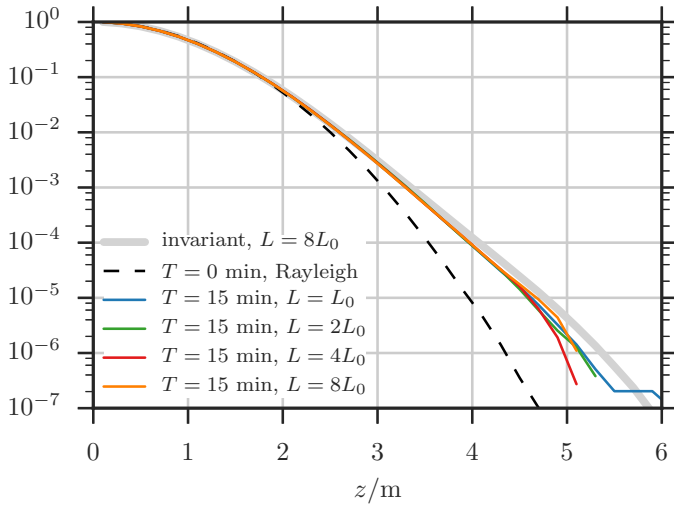

Figure B.2. Numerical verification of the invariance $P_0^L(z) = \lim_{t\to\infty} P_0^L(t,z)$ for $L \geq L_0$. The limiting value $P_0(z)$ (gray curve) was calculated by propagating 1500 samples up to time of 3000 min in the largest domain with $L = 8L_0$). Note that this also shows that $P_0^{L_0}(t,z)$ in the MC sampling has essentially converged to the invariant $P_0(z)$ after only 15 min.

Coming back to the quantity investigated in main text, let us denote

$$P_{\max}^L(t,z) = \mathbb{P}\left(\max_{x\in[0,L]} |u(t,x)| > z\right). \tag{B.10}$$

Unlike $P_0^L(t,z)$, the probability $P_{\max}^L(t,z)$ does depend on L – the larger L, the higher $P_{\max}^L(t,z)$. We can actually estimate this growth explicitly. To see how, consider a domain of size NL that can be partitioned into $N \geq 1$ sub-domains of size L, each large enough to be roughly statistically independent of the others. Then we have

$$1 - P_{\max}^{NL}(t,z) = \left(1 - P_{\max}^L(t,z)\right)^N, \qquad N \geq 1 \tag{B.11}$$

since in order for the maximum of $|u|$ to be less than z in the larger domain of size NL, it must be less than z in each of the (roughly independent) sub-domains of size L. Eq. (B.11) is the fundamental equation used in extreme value statistics. We confirmed its applicability for $L = L_0 = 40\pi$ in our system via direct estimation of

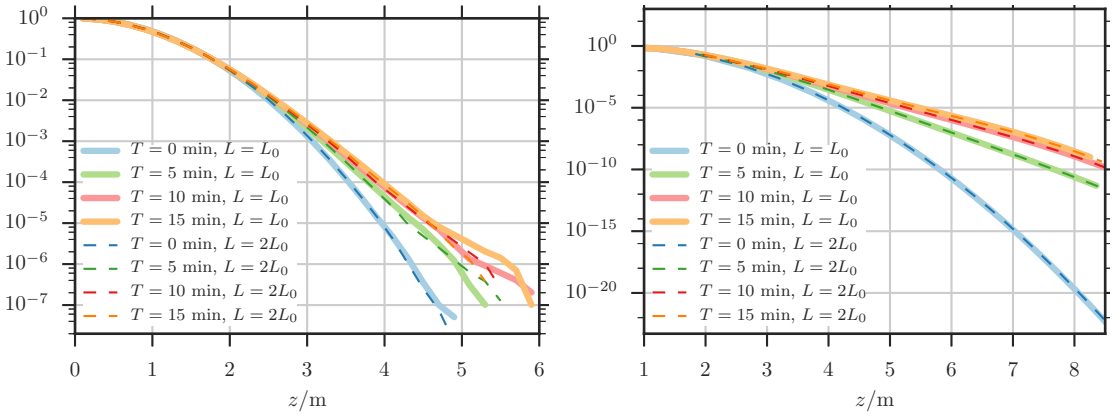

Figure B.3. Left: $P_0^L(t,z) = \mathbb{P}(|u(t,x_0)| > z)$ at a fixed location x_0 and different times t in domains of size $L = L_0$ and $L = 2L_0$ obtained by Monte Carlo (MC) sampling. Right: Same, obtained by optimization using our large deviation approach and a larger range of values for z (such large values cannot easily be reached by MC). As can be seen, the PDFs essentially lay on top of each other for the two different domains, confirming that the domain size L_0 is large enough for the periodic boundary conditions to not affect the results.

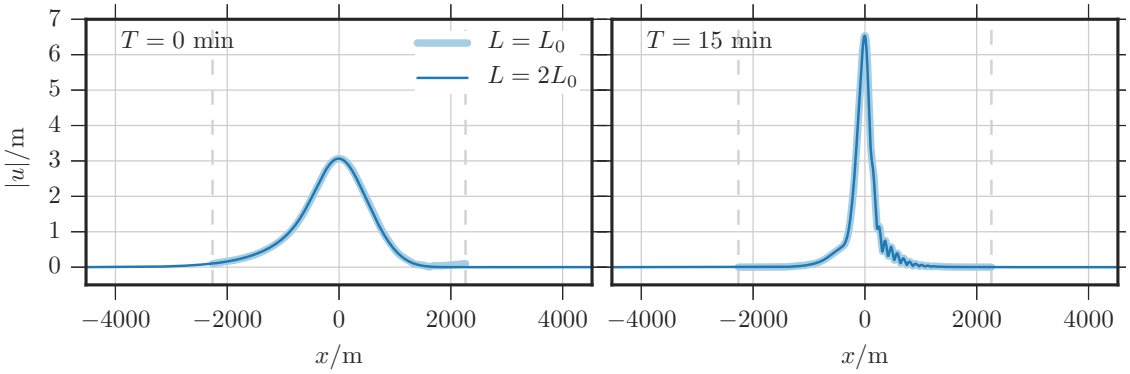

Figure B.4. Optimal trajectories calculated in the domains of size $L = L_0$ (thick line) and $L = 2L_0$ (thin line). As can be seen, the periodicity of the domain does not affect significantly the shape of the instanton inside this domain.

$P_{\max}^{NL_0}(t,z)$ for $N = 1,2,4,8$ by Monte Carlo sampling. These results are reported in Fig. B.5.

Since $L_0 = 40\pi$ is already large enough for (B.11) to hold, we can rewrite this equation as

$$1 - P_{\max}^{L}(t,z) = \left(1 - P_{\max}^{L_0}(t,z)\right)^{L/L_0}, \qquad L \geq L_0 \tag{B.12}$$

Note that this equation implies that, at fixed z, $P_{\max}^{L}(t,z)$ increases with L since

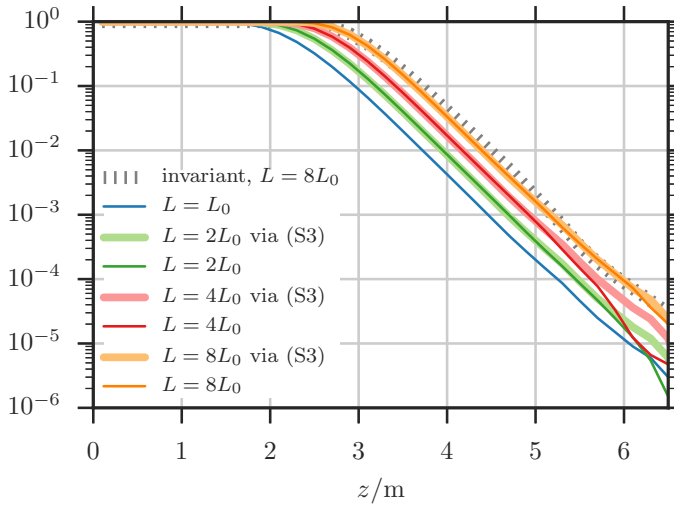

Figure B.5. Numerical verification of (B.11) for $L = L_0 = 40\pi$. These results confirm that adjacent boxes of size L_0 can be considered statistically independent. The probability $P^L_{\max}(z) = \lim_{t\to\infty} P^L_{\max}(t, z)$ for $L = 8L_0$, is also shown, indicating that this quantity can be estimated accurately from $P^{L_0}_{\max}$ at 15 min using (B.11).

$1-P^{L_0}_{\max}(t,z) < 1$ and therefore $1-P^{L}_{\max}(t,z) = \left(1 - P^{L_0}_{\max}(t,z)\right)^{L/L_0} \leq 1-P^{L_0}_{\max}(t,z)$ for $L \geq L_0$. Intuitively, this increase in $P^L_{\max}(t,z)$ stems from the fact that multiple large values of $|u|$ are expected to arise simultaneously in different sub-domains since they are statistically independent – this is usually referred to as an entropic effect, and it can be seen in the typical realizations from the Monte-Carlo sampling shown in Fig. B.6 for $L = L_0$ and $L = 8L_0$. Of course this effect is properly accounted for by Eq. (B.11). Indeed, realizations like those shown in Fig. B.6 are those from which the probabilities shown in Fig. B.5 were calculated.

It is also important to stress that this entropic effect cannot be accounted for directly by our large deviation approach. The solution obtained by optimization becomes independent of L for L large enough (which is the case already for $L = L_0$). This implies that, without correction, the results of the large deviation approach will deteriorate with increasing L. Eq. (B.12) shows that this issue can be easily fixed, however: indeed this formula indicates how the large deviation results at $L = L_0$ (i.e. in a domain that is large enough to not be influenced by the boundary condition, but small enough that the entropic effects remain negligible) can be extended to larger L.

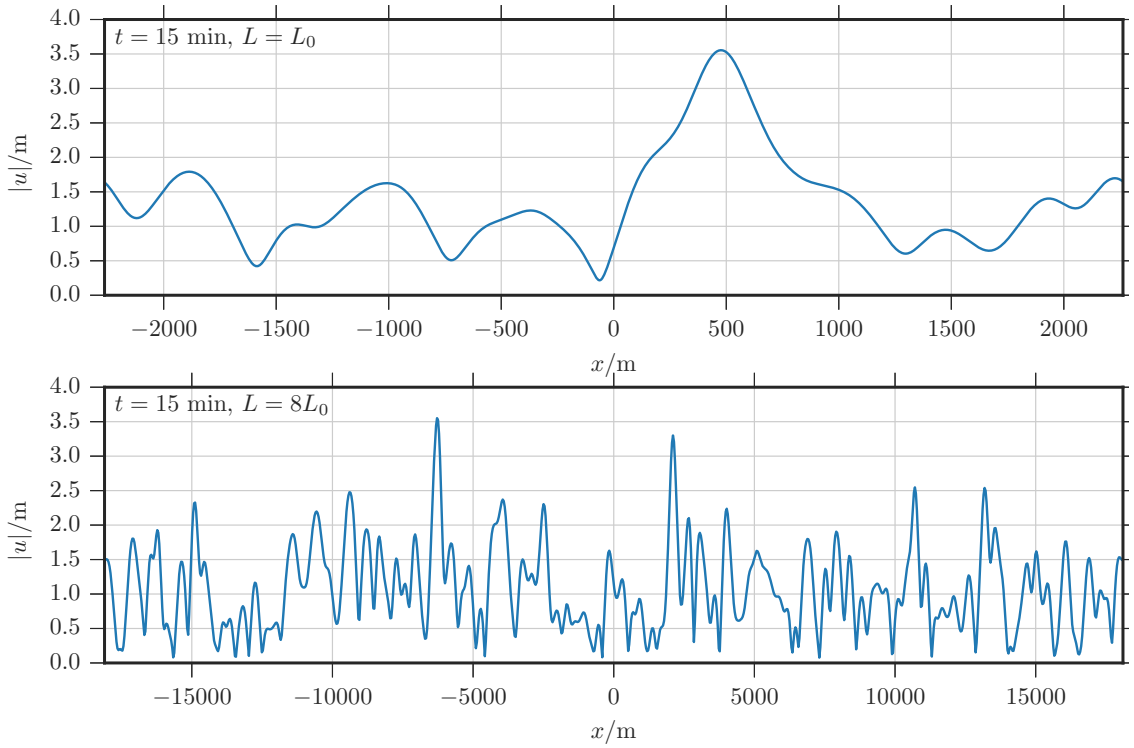

Figure B.6. Typical realizations from the Monte-Carlo sampling such that $\max_x |u(t,x)| \geq 3.5$ m at $t = 15$ min in the domains of size $L = L_0$ (top) and $L = 8L_0$ (bottom). As can be seen, as the domain size increases, it becomes increasingly likely to observe more than one large value of $u(t,x)|$ in the domain.

B.3 The case of the nonlinear Schrödinger equation (NLS) and the role of the Peregrine soliton (PS)

For completeness, we redid all of our calculations in the context of the standard Nonlinear Schrödinger (NLS) equation instead of the MNLS equation. Using the same non-dimensional variables as in MNLS, NLS reads

$$\partial_t u + \frac{i}{8}\partial_{xx} u + \frac{i}{2} u|u|^2 = 0. \tag{B.13}$$

Fig. B.7 shows the distributions for the spatial maximum of the envelope $|u|$ at different times calculated by both direct Monte-Carlo sampling and minimization using our large deviation approach, using the same random initial conditions as in MNLS. As can be seen, here too the approach based on large deviation theory does

an excellent job at capturing these PDFs.

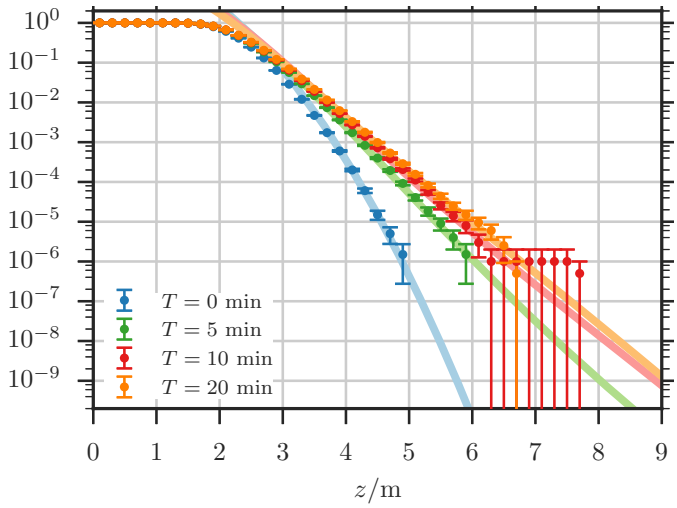

Figure B.7. $\mathbb{P}(\max_x |u(t,x)| \geq z)$ for $u(t,x)$ solution of NLS at different times calculated by Monte Carlo sampling using 10^6 realizations and compared to the results obtained via optimization in our large deviation approach.

The advantage of using NLS is that it permits us to assess the relevance of the Peregrine soliton (PS), which is an exact solution of NLS (though not of MNLS) that has been invoked as prototype mechanism for rogue waves creation [156, 3, 174, 4, 150, 186] – recent experimental results in the context of water waves [29, 30, 28], plasmas [10] and fiber optics [109, 180, 185] have lent support to this hypothesis. The PS reads

$$u(t,x) = U_i e^{-it/T_{\mathrm{nl}}}\left(\frac{4(1-2it/T_{\mathrm{nl}})}{1+4\big(t/T_{\mathrm{nl}}\big)^2+4\big(x/L_{\mathrm{nl}}\big)^2}-1\right),\ T_{\mathrm{nl}}=\frac{2}{U_i^2},\ L_{\mathrm{nl}}=\tfrac{1}{4}\sqrt{T_{\mathrm{nl}}}=\frac{\sqrt{2}}{4U_i}, \tag{B.14}$$

where $U_i > 0$ is a free parameter. It can be checked that this solution reaches its maximal amplitude $|u(0,0)| = 3U_i$ at $(t,x) = (0,0)$ and decays both forward and backward in time to $\lim_{t\to\pm\infty} |u(t,x)| = U_i$.

To compare the PS to our results, we translated t in (B.14) to make the time at which this solution reaches its maximal amplitude coincide with the time at which a prescribed value of the wave elevation is observed in either our minimization procedure or in the MC sampling. By adjusting U_i so that the maximal amplitude of the PS also coincides with this prescribed value of the amplitude, we can then verify how well the PS reproduces our instanton as well as the mean and variance of the solutions observed in the MC sampling. These results are reported in Fig. B.8. As can be seen, the PS captures the shape of the instanton at final time (i.e. when

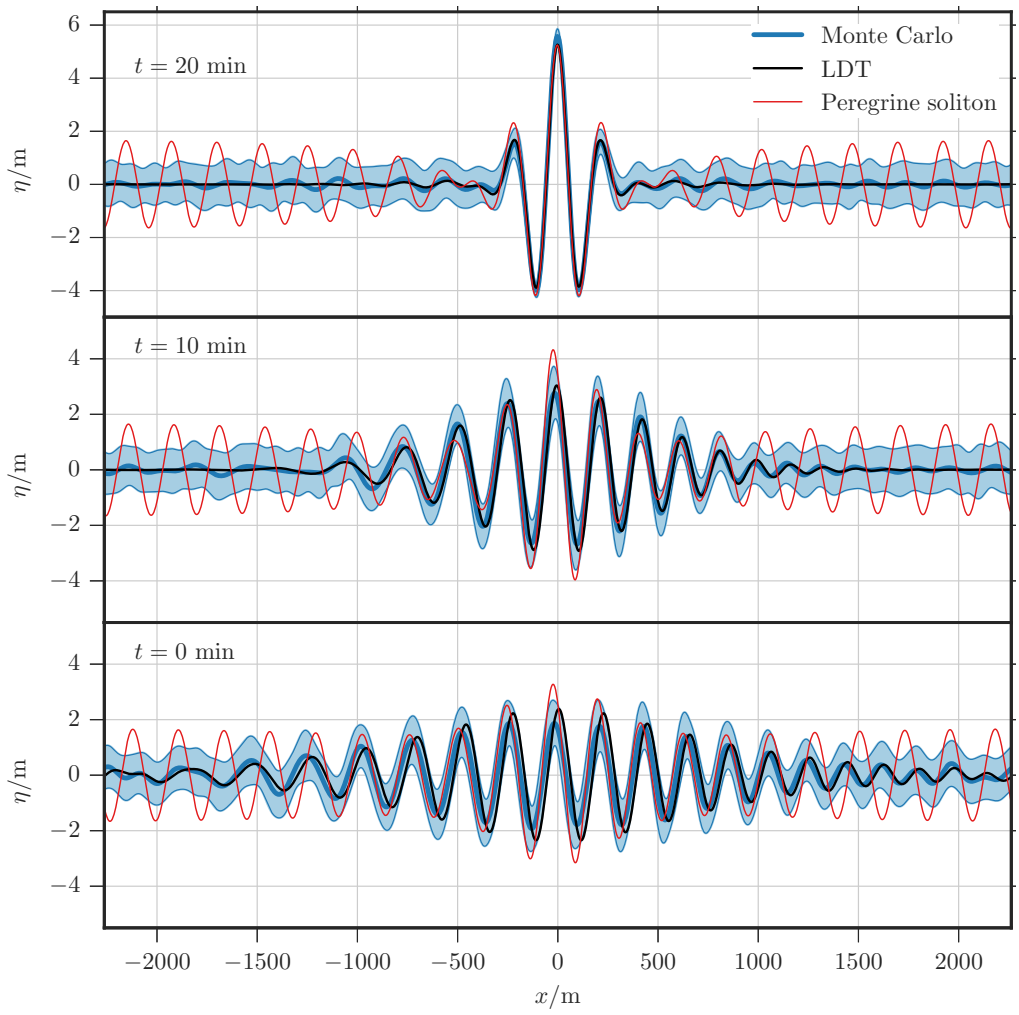

Figure B.8. Comparison of the optimized solution, the mean and standard deviation of the Monte-Carlo realizations, and the PS reaching the same maximal surface elevation at $T = 20$ min. From bottom to top the figures are at 0, 10 and 20 minutes respectively, and these results are for NLS.

the rogue wave occurs) reasonably well, at least near the location $x = 0$ where the maximum amplitude is observed (focusing region). The PS also does a reasonably good job at tracking the evolution of the solution that led to this extreme event. In particular, the focusing time scale of the optimized solutions (which we interpret to also describe the convergence time of the a-priori distribution to the invariant distribution) is in rough agreement with the effective focusing time scale of the PS starting from a pulse of size L_i [180, 185]. This time scale is given by $\tau_c = \sqrt{T_{\rm nl} T_{\rm lin}}$, where the nonlinear time $T_{\rm nl}$ is defined in (B.14) and the linear time $T_{\rm lin} = 8L_i^2$ is

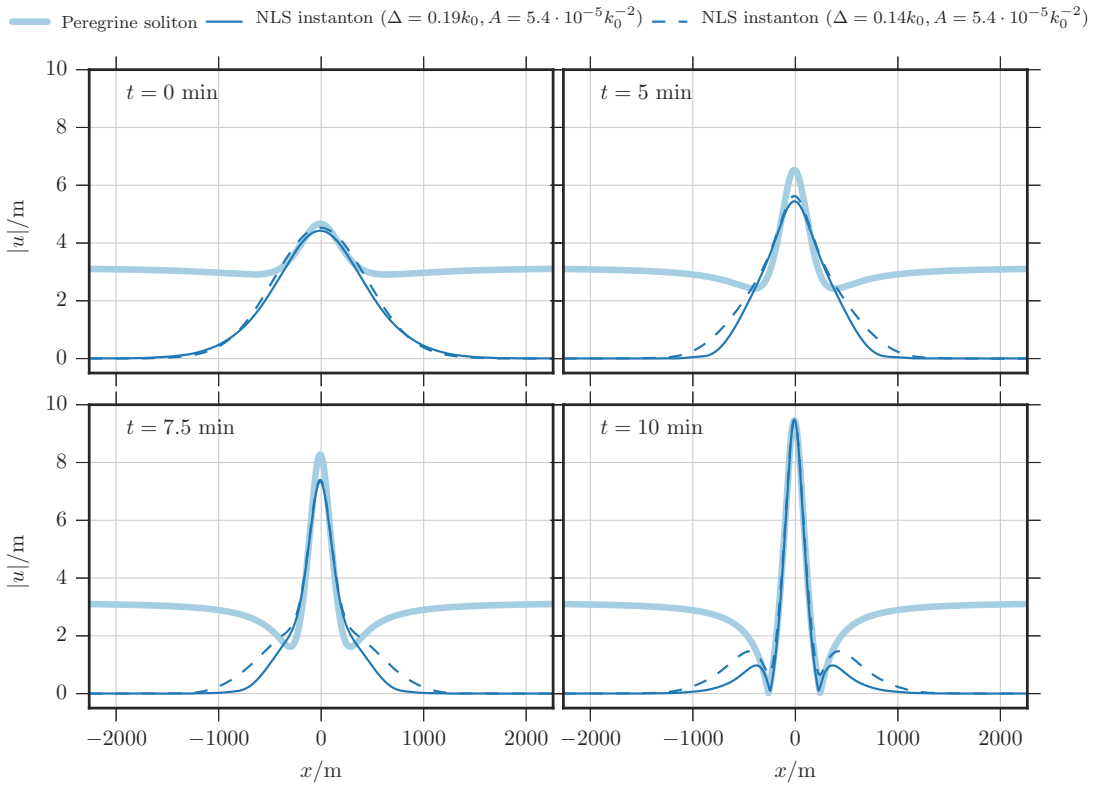

Figure B.9. Comparison between the optimized solution for a very extreme surface elevation and the PS reaching the same final height (after $T = 10$ minutes). Comparison with realization from the Monte-Carlo sampling is impossible due to the extreme rareness of such event on the ensemble of the initial conditions. The evolution is shown at times 0 min, 5 min, 7.5 min, and 10 min, respectively. These results are for NLS.

that associated with group velocity dispersion of the initial pulse – in dimensional units, these are $T_{\mathrm{nl}} = \left(\frac{1}{2}\omega_0 k_0^2 U_i^2\right)^{-1}$ and $T_{\mathrm{lin}} = 8\omega_0^{-1}k_0^2 L_i^2$.

The relative agreement both in shape and timescale between the optimized solution and the PS suggests that the main physical phenomenon responsible for the focusing in the NLS equation is the gradient catastrophe [16], which fosters a very unique evolution pathway as the point of maximum focusing is approached in space-time. Still, it should be stressed that the discrepancies between the PS and the actual solution we observe become more and more pronounced backward in time. These differences can also be observed in Fig. B.9 where we plot the amplitude of u for a more extreme event that is too rare to be observed by MC sampling. In this figure, we show the optimized solutions obtained for two different spectral widths Δ, whose shapes are slightly different from one another: clearly, these differences

cannot be captured by the PS since this solution is completely specified by the final amplitude, which is the same for both sets.

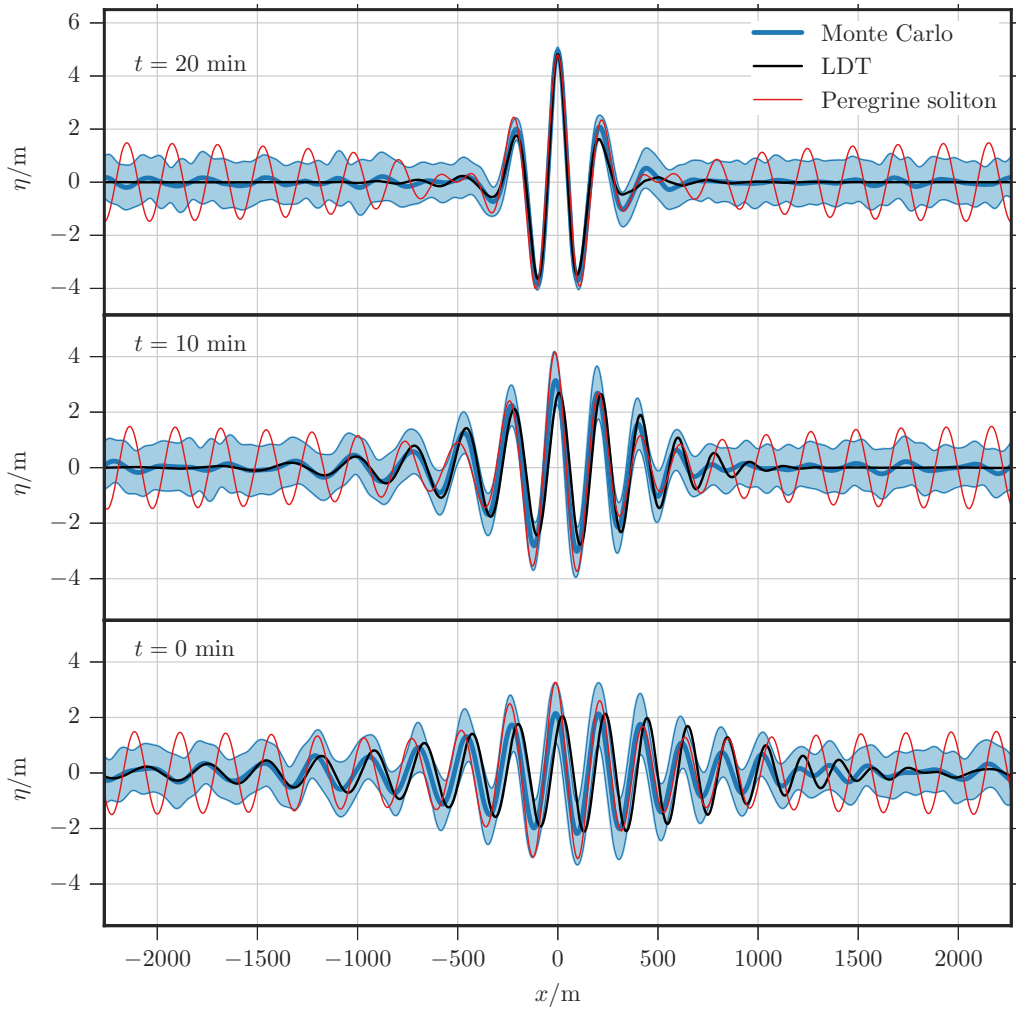

Figure B.10. Same as in Fig. B.8 for MNLS.

For completeness we also compared the PS with the solutions obtained in the context of MNLS. These results are reported on Fig. B.10 and show similar type of agreement, in particular in term of the shape of the rogue wave near its maximum and the time scale of its emergence. Note the discrepancies between the PS and our solutions is even more pronounced in this case, which is to be expected since PS is an exact solution of NLS, but not of MNLS.

To summarize, while the PS can explain some features of the rogue waves, in particular their shape as well as the focusing time scale over which these waves evolve from a large initial pulse, it does not capture the details of the formation of

these waves — indeed there is no reason why it should, since different sets of random initial conditions lead to waves with different shapes (and whose amplitudes have different statistics) and this information is not seen by the PS. In particular, the instanton solution for the initial data chosen here depends on two parameters, the significant wave height H_s and the BFI, while the PS only allows a single parameter U_i. Additionally and more importantly, the PS does not allow the estimation of the probability of observing rogue waves of given amplitude since this solution *per se* is devoid of a probabilistic framework.

Bibliography

[1] S. A. Adcock and J. A. McCammon. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chemical reviews*, 106(5):1589–1615, 2006.

[2] D. S. Agafontsev and V. E. Zakharov. Integrable turbulence and formation of rogue waves. *Nonlinearity*, 28(8):2791, 2015.

[3] N. Akhmediev, A. Ankiewicz, and M. Taki. Waves that appear from nowhere and disappear without a trace. *Physics Letters A*, 373(6):675–678, 2009.

[4] N. Akhmediev, J. M. Dudley, D. R. Solli, and S. K. Turitsyn. Recent progress in investigating optical rogue waves. *Journal of Optics*, 15(6):060201, 2013.

[5] N. Akhmediev and E. Pelinovsky. Editorial–introductory remarks on "discussion & debate: Rogue waves–towards a unifying concept?". *Eur. Phys. J. Special Topics*, 185(1):1–4, 2010.

[6] N. Akhmediev, J. M. Soto-Crespo, and A. Ankiewicz. Extreme waves that appear from nowhere: on the nature of rogue waves. *Physics Letters A*, 373(25):2137–2145, 2009.

[7] N. N. Akhmediev, V. M. Eleonskii, and N. E. Kulagin. Exact first-order solutions of the nonlinear Schrödinger equation. *Theoretical and Mathematical Physics*, 72(2):809–818, 1987.

[8] M. Athans and P. L. Falb. *Optimal control: an introduction to the theory and its applications*. Courier Corporation, 2013.

[9] E. Aurell, C. Mejía-Monasterio, and P. Muratore-Ginanneschi. Optimal protocols and optimal transport in stochastic thermodynamics. *Physical Review Letters*, 106(25):250601, 2011.

[10] H. Bailung, S. K. Sharma, and Y. Nakamura. Observation of Peregrine solitons in a multicomponent plasma with negative ions. *Physical Review Letters*, 107(25):255005, 2011.

[11] D. Barriopedro, E. M. Fischer, J. Luterbacher, R. M. Trigo, and R. García-Herrera. The hot summer of 2010: redrawing the temperature record map of europe. *Science*, 332(6026):220–224, 2011.

[12] C. M. Bender and S. A. Orszag. *Advanced mathematical methods for scientists and engineers I: Asymptotic methods and perturbation theory*. Springer Science & Business Media, 2013.

[13] T. B. Benjamin and J. E. Feir. The disintegration of wave trains on deep water Part 1. Theory. *Journal of Fluid Mechanics*, 27(03):417–430, Feb. 1967.

[14] D. J. Benney and A. C. Newell. Random wave closures. *Stud. Appl. Math.*, 48:29–53, 1969.

[15] M. Berry. Making waves in physics. *Nature*, 403(6765):21–21, 2000.

[16] M. Bertola and A. Tovbis. Universality for the focusing nonlinear Schrödinger equation at the gradient catastrophe point: rational breathers and poles of the tritronquée solution to Painlevé I. *Communications on Pure and Applied Mathematics*, 66(5):678–752, 2013.

[17] L. Biven, S. Nazarenko, and A. Newell. Breakdown of wave turbulence and the onset of intermittency. *Physics Letters A*, 280(1):28–32, 2001.

[18] P. Boccotti. *Wave mechanics for ocean engineering*, volume 64. Elsevier, 2000.

[19] L. Boltzmann. *Über die Beziehung zwischen dem zweiten Hauptsatze des mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung, respective den Sätzen über das Wärmegleichgewicht*. Kk Hof-und Staatsdruckerei, 1877.

[20] A. A. Borovkov and B. A. Rogozin. On the multi-dimensional central limit theorem. *Theory of Probability & its Applications*, 10(1):55–62, 1965.

[21] A. Borzì and V. Schulz. *Computational optimization of systems governed by partial differential equations*. Computational Science & Engineering. SIAM, 2011.

[22] A. Boudaoud, O. Cadot, B. Odille, and C. Touzé. Observation of wave turbulence in vibrating plates. *Physical Review Letters*, 100(23):234504, 2008.

[23] M. Broniatowski and A. Fuchs. Tauberian theorems, Chernoff inequality, and the tail behavior of finite convolutions of distribution functions. *Advances in Mathematics*, 116(1):12–33, 1995.

[24] R. Brout and I. Prigogine. Statistical mechanics of irreversible processes part viii: general theory of weakly coupled systems. *Physica*, 22(6):621–636, 1956.

[25] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith. Entropic elasticity of lambda-phage DNA. *Science*, 265(5178):1599–1599, 1994.

[26] C. Cercignani. The boltzmann equation. In *The Boltzmann Equation and Its Applications*, pages 40–103. Springer, 1988.

[27] F. Cérou and A. Guyader. Adaptive multilevel splitting for rare event analysis. *Stochastic Analysis and Applications*, 25(2):417–443, 2007.

[28] A. Chabchoub. Tracking breather dynamics in irregular sea state conditions. *Physical Review Letters*, 117(14):144103, 2016.

[29] A. Chabchoub, N. Hoffmann, and N. Akhmediev. Rogue wave observation in a water wave tank. *Physical Review Letters*, 106(20):204502, 2011.

[30] A. Chabchoub, N. Hoffmann, M. Onorato, and N. Akhmediev. Super rogue waves: observation of a higher-order breather in water waves. *Physical Review X*, 2(1):011015, 2012.

[31] S. Chibbaro, F. De Lillo, and M. Onorato. Weak versus strong wave turbulence in the majda-mclaughlin-tabak model. *Physical Review Fluids*, 2(5):052603, 2017.

[32] S. Chibbaro, G. Dematteis, C. Josserand, and L. Rondoni. Wave-turbulence theory of four-wave nonlinear interactions. *Physical Review E*, 96(2):021101, 2017.

[33] S. Chibbaro, G. Dematteis, and L. Rondoni. 4-wave dynamics in kinetic wave turbulence. *Physica D: Nonlinear Phenomena*, 362:24–59, 2018.

[34] S. Chibbaro and C. Josserand. Elastic wave turbulence and intermittency. *Physical Review E*, 94:011101, Jul 2016.

[35] S. Chibbaro, L. Rondoni, and A. Vulpiani. On the foundations of statistical mechanics: ergodicity, many degrees of freedom and inference. *Communications in Theoretical Physics*, 62(4):469, 2014.

[36] S. Chibbaro, L. Rondoni, and A. Vulpiani. Reductionism, emergence and levels of reality. *Springer, Berlin. by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on*, 3(20):17, 2014.

[37] Y. Choi, Y. V. Lvov, and S. Nazarenko. Probability densities and preservation of randomness in wave turbulence. *Physics Letters A*, 332(3):230–238, 2004.

[38] Y. Choi, Y. V. Lvov, and S. Nazarenko. Joint statistics of amplitudes and phases in wave turbulence. *Physica D: Nonlinear Phenomena*, 201(1):121–149, 2005.

[39] Y. Choi, Y. V. Lvov, S. Nazarenko, and B. Pokorni. Anomalous probability of large amplitudes in wave turbulence. *Physics Letters A*, 339(3):361–369, 2005.

[40] P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J.-L. Viovy, D. Chatenay, and F. Caron. DNA: an extensible molecule. *Science*, 271(5250):792–794, 1996.

[41] C. Connaughton, S. Nazarenko, and A. Newell. Dimensional analysis and weak turbulence. *Physica D: Nonlinear Phenomena*, 184(1):86–97, 2003.

[42] D. Coumou and S. Rahmstorf. A decade of weather extremes. *Nature climate change*, 2(7):491, 2012.

[43] W. Cousins and T. P. Sapsis. Unsteady evolution of localized unidirectional deep-water wave groups. *Physical Review E*, 91(6):063204, 2015.

[44] W. Cousins and T. P. Sapsis. Reduced-order precursors of rare events in unidirectional nonlinear water waves. *Journal of Fluid Mechanics*, 790:368–388, 2016.

[45] S. M. Cox and P. C. Matthews. Exponential time differencing for stiff systems. *Journal of Computational Physics*, 176(2):430–455, 2002.

[46] W. Craig, P. Guyenne, and C. Sulem. A Hamiltonian approach to nonlinear modulation of surface water waves. *Wave Motion*, 47(8):552–563, 2010.

[47] H. Cramér. Sur un nouveau théoreme-limite de la théorie des probabilités. *Actual. Sci. Ind.*, 736:5–23, 1938.

[48] S. J. Davie, O. G. Jepps, L. Rondoni, J. C. Reid, and D. J. Searles. Applicability of optimal protocols and the Jarzynski equality. *Physica Scripta*, 89(4):048002, 2014.

[49] G. Dematteis, T. Grafke, and E. Vanden-Eijnden. Rogue waves and large deviations in deep sea. *Proceedings of the National Academy of Sciences*, 115(5):855–860, 2018.

[50] A. Dembo and O. Zeitouni. *Large deviations techniques and applications. Corrected reprint of the second edition.* Stochastic Modelling and Applied Probability, 38. Springer-Verlag, Berlin, 2010.

[51] L. Desvillettes and S. Mischler. About the splitting algorithm for boltzmann and bgk equations. *Mathematical Models and Methods in Applied Sciences*, 6(08):1079–1101, 1996.

[52] J.-D. Deuschel and D. W. Stroock. *Large deviations*, volume 342. American Mathematical Soc., 2001.

[53] P. Dupuis and R. S. Ellis. *A weak convergence approach to the theory of large deviations*, volume 902. John Wiley & Sons, 2011.

[54] G. Düring, C. Josserand, and S. Rica. Weak turbulence for a vibrating plate: Can one hear a kolmogorov spectrum? *Physical Review Letters*, 97(2):025503, 2006.

[55] G. Düring, C. Josserand, and S. Rica. Weak turbulence for a vibrating plate: can one hear a kolmogorov spectrum? *Physical Review Letters*, 97.2:025503, 2006.

[56] G. Düring, C. Josserand, and S. Rica. Wave turbulence theory of elastic plates. *Physica D: Nonlinear Phenomena*, 87, 2017.

[57] G. Düring, A. Picozzi, and S. Rica. Breakdown of weak-turbulence and nonlinear wave condensation. *Physica D: Nonlinear Phenomena*, 238(16):1524–1549, 2009.

[58] S. Dyachenko, A. Newell, A. Pushkarev, and V. Zakharov. Optical turbulence: weak turbulence, condensates and collapsing filaments in the nonlinear schrödinger equation. *Physica D: Nonlinear Phenomena*, 57:96–160, 1992.

[59] M. I. Dykman, E. Mori, J. Ross, and P. Hunt. Large fluctuations and optimal paths in chemical kinetics. *The Journal of Chemical Physics*, 100(8):5735–5750, 1994.

[60] K. B. Dysthe. Note on a Modification to the Nonlinear Schrödinger Equation for Application to Deep Water Waves. *Proc. R. Soc. Lond. A*, 369(1736):105–114, Dec. 1979.

[61] L. Ebener, G. Margazoglou, J. Friedrich, L. Biferale, and R. Grauer. Instanton based importance sampling for rare events in stochastic pdes. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(6):063102, 2019.

[62] U. Einmahl and J. Kuelbs. Dominating points and large deviations for random vectors. *Probab. Theory Related Fields*, 105(4):529–543, Dec. 1996.

[63] R. El Koussaifi, A. Tikan, A. Toffoli, S. Randoux, P. Suret, and M. Onorato.

Spontaneous emergence of rogue waves in partially coherent waves: a quantitative experimental comparison between hydrodynamics and optics. *Physical Review E*, 97(1):012208, 2018.

[64] R. S. Ellis. The theory of large deviations: from boltzmann's 1877 calculation to equilibrium macrostates in 2d turbulence. *Physica D: Nonlinear Phenomena*, 133(1-4):106–136, 1999.

[65] R. S. Ellis. *Entropy, large deviations, and statistical mechanics*. Springer, 2007.

[66] R. S. Ellis et al. Large deviations for a general class of random vectors. *The Annals of Probability*, 12(1):1–12, 1984.

[67] F. Escher. On the probability function in the collective theory of risk. *Skand. Aktuarie Tidskr.*, 15:175–195, 1932.

[68] M. Esposito. Stochastic thermodynamics under coarse graining. *Physical Review E*, 85(4):041125, 2012.

[69] G. L. Eyink and Y.-K. Shi. Kinetic wave turbulence. *Physica D: Nonlinear Phenomena*, 241(18):1487–1511, 2012.

[70] C. Falcon, E. Falcon, U. Bortolozzo, and S. Fauve. Capillary wave turbulence on a spherical fluid surface in low gravity. *EPL (Europhysics Letters)*, 86(1):14002, 2009.

[71] C. Falcon, E. Falcon, U. Bortolozzo, and S. Fauve. Capillary wave turbulence on a spherical fluid surface in low gravity. *Europhysics Letters*, 86:14002, 2009.

[72] E. Falcon, S. Aumaître, C. Falcón, C. Laroche, and S. Fauve. Fluctuations of energy flux in wave turbulence. *Physical Review Letters*, 100(6):064503, 2008.

[73] E. Falcon, S. Fauve, and C. Laroche. Observation of intermittency in wave turbulence. *Physical Review Letters*, 98(15):154501, 2007.

[74] E. Falcon, C. Laroche, and S. Fauve. Observation of gravity-capillary wave turbulence. *Physical Review Letters*, 98(9):094503, 2007.

[75] E. Falcon, S. Roux, and C. Laroche. On the origin of intermittency in wave turbulence. *EPL (Europhysics Letters)*, 90(3):34005, 2010.

[76] G. Falkovich, I. Kolokolov, V. Lebedev, and A. Migdal. Instantons and intermittency. *Physical Review E*, 54(5):4896, 1996.

[77] M. Farazmand and T. P. Sapsis. A variational approach to probing extreme events in turbulent dynamical systems. *Science Advances*, 3(9):e1701533, 2017.

[78] F. Fedele and F. Arena. Weakly nonlinear statistics of high random waves. *Physics of fluids*, 17(2):026601, 2005.

[79] J. Feng and T. G. Kurtz. *Large deviations for stochastic processes*. Number 131. American Mathematical Soc., 2006.

[80] M. I. Freidlin and A. D. Wentzell. Random perturbations. In *Random Perturbations of Dynamical Systems*, pages 15–43. Springer, 1998.

[81] U. Frisch. *Turbulence: the legacy of AN Kolmogorov*. Cambridge university press, 1995.

[82] U. Frisch and D. Sornette. Extreme deviations and applications. *Journal de Physique I*, 7(9):1155–1171, 1997.

[83] G. Gallavotti. *Statistical mechanics: A short treatise*. Springer Science & Business Media, 2013.

[84] G. Gallavotti, W. L. Reiter, and J. Yngvason. *Boltzmann's legacy*, volume 5. European Mathematical Society, 2008.

[85] S. Galtier, S. Nazarenko, A. C. Newell, and A. Pouquet. A weak turbulence theory for incompressible magnetohydrodynamics. *Journal of Plasma Physics*, 63(05):447–488, 2000.

[86] J. Gärtner. On large deviations from the invariant measure. *Theory of Probability & Its Applications*, 22(1):24–39, 1977.

[87] C. Giardina, J. Kurchan, V. Lecomte, and J. Tailleur. Simulating rare events in dynamical processes. *Journal of Statistical Physics*, 145(4):787–811, 2011.

[88] P. Glasserman, P. Heidelberger, P. Shahabuddin, and T. Zajic. Multilevel splitting for estimating rare event probabilities. *Operations Research*, 47(4):585–600, 1999.

[89] A. Goullet and W. Choi. A numerical and experimental study on the nonlinear evolution of long-crested irregular waves. *Physics of Fluids*, 23(1):016601, 2011.

[90] T. Grafke, R. Grauer, and T. Schäfer. The instanton method and its numerical implementation in fluid mechanics. *Journal of Physics A: Mathematical and Theoretical*, 48(33):333001, 2015.

[91] T. Grafke, R. Grauer, T. Schäfer, and E. Vanden-Eijnden. Relevance of instantons in Burgers turbulence. *EPL (Europhysics Letters)*, 109(3):34003, 2015.

[92] T. Grafke and E. Vanden-Eijnden. Numerical computation of rare events via large deviation theory. *arXiv preprint arXiv:1812.00681*, 2018.

[93] O. Gramstad and K. Trulsen. Hamiltonian form of the modified nonlinear Schrödinger equation for gravity waves on arbitrary depth. *Journal of Fluid Mechanics*, 670:404–426, 2011.

[94] E. P. Gross. Structure of a quantized vortex in boson systems. *Nuovo Cimento*, 20(454), 1961.

[95] W. W. Hager. Runge-Kutta methods in optimal control and the transformed adjoint system. *Numerische Mathematik*, 87(2):247–282, 2000.

[96] K. Hasselmann, T. P. Barnett, E. Bouws, H. Carlson, D. E. Cartwright, K. Enke, J. A. Ewing, H. Gienapp, D. E. Hasselmann, P. Kruseman, et al. Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (jonswap). Technical report, Deutches Hydrographisches Institut, 1973.

[97] S. Haver. A possible freak wave event measured at the Draupner jacket january 1 1995. *Rogue waves 2004*, pages 1–8, 2004.

[98] T. Humbert, O. Cadot, G. Düring, C. Josserand, S. Rica, and C. Touzé. Wave turbulence in vibrating plates: The effect of damping. *Europhysics Letters*, 102(3):30002, 2013.

[99] T. Humbert, C. Josserand, C. Touzé, and O. Cadot. Phenomenological model for predicting stationary and non-stationary spectra of wave turbulence in vibrating plates. *Physica D: Nonlinear Phenomena*, 316:34–42, 2016.

[100] M. Iltis. Sharp asymptotics of large deviations in $\mathbb{R}^d$. *Journal of Theoretical Probability*, 8(3):501–522, 1995.

[101] M. Iltis. Sharp asymptotics of large deviations for general state-space Markov-additive chains in R-d. *Statistics & Probability Letters*, 47(4):365–380, 2000.

[102] P. Jakobsen and A. C. Newell. Invariant measures and entropy production in wave turbulence. *Journal of Statistical Mechanics: Theory and Experiment*, 2004(10):L10002, 2004.

[103] P. Jakobsen and A. C. Newell. Invariant measures and entropy production in wave turbulence. *J. Stat. Mech.-Theory E.*, 2004:L10002, 2004.

[104] P. A. E. M. Janssen. Nonlinear Four-Wave Interactions and Freak Waves. *Journal of Physical Oceanography*, 33(4):863–884, Apr. 2003.

[105] J. L. Jensen. *Saddlepoint approximations*. Oxford Statistical Science Series, 16. Oxford University Press, New York, 1995.

[106] S. Juneja and P. Shahabuddin. Rare-event simulation techniques: an introduction and recent advances. *Handbooks in operations research and management science*, 13:291–350, 2006.

[107] A.-K. Kassam and L. N. Trefethen. Fourth-order time-stepping for stiff PDEs. *SIAM Journal of Scientific Computing*, 26(4):1214–1233, 2005.

[108] A. I. Khinchin. *Mathematical foundations of statistical mechanics*. Courier Corporation, 1949.

[109] B. Kibler, J. Fatome, C. Finot, G. Millot, F. Dias, G. Genty, N. Akhmediev, and J. M. Dudley. The Peregrine soliton in nonlinear fibre optics. *Nature Physics*, 6(10):790, 2010.

[110] G. Komen, L. Cavaleri, M. Donelan, K. Hasselmann, H. Hasselmann, and P. Janssen. *Dynamics and modeling of ocean waves*. Cambridge University Press, Cambridge, 1994.

[111] G. J. Komen, L. Cavaleri, M. Donelan, K. Hasselmann, S. Hasselmann, and P. A. E. M. Janssen. *Dynamics and modelling of ocean waves*. Cambridge University Press, 1996.

[112] J. Kuelbs. Large deviation probabilities and dominating points for open convex sets: nonlogarithmic behavior. *Annals of probability*, pages 1259–1279, 2000.

[113] O. Kurbanmuradov. Convergence of numerical models for the Gaussian fields. *Russ. J. Numer. Anal. M.*, 10:311–323, 1995.

[114] E. A. Kuznetsov. Solitons in a parametrically unstable plasma. *Akademiia Nauk SSSR Doklady*, 236:575–577, Sept. 1977.

[115] L. Landau and E. Lifshitz. *Theory of Elasticity*. Pergamon Press, New York, 1959.

[116] O. E. Lanford. Time evolution of large classical systems. In J. Moser, editor, *Dynamical Systems, Theory and Applications*, volume 38 of *Lecture Notes in Physics*, pages 1–111. Springer Berlin / Heidelberg, 1975.

[117] O. E. Lanford. On the derivation of the Boltzmann equation. *Asterisque*, 40:117–137, 1976.

[118] F. Lankaš, J. Šponer, P. Hobza, and J. Langowski. Sequence-dependent elastic properties of DNA. *Journal of Molecular Biology*, 299(3):695–709, 2000.

[119] J. L. Lebowitz. Boltzmann's entropy and time's arrow. *Physics today*, 46:32–32, 1993.

[120] S. L. Lewis, P. M. Brando, O. L. Phillips, G. M. van der Heijden, and D. Nepstad. The 2010 Amazon drought. *Science*, 331(6017):554–554, 2011.

[121] G. Lindgren. Local maxima of Gaussian fields. *Arkiv för Matematik*, 10(1-2):195–218, 1972.

[122] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science*, 296(5574):1832–1835, 2002.

[123] P. C. Liu. A chronology of freauqe wave encounters. *Geofizika*, 24(1):57–70, 2007.

[124] E. Lo and C. C. Mei. A numerical study of water-wave modulation based on a higher-order nonlinear Schrödinger equation. *Journal of Fluid Mechanics*, 150:395–416, 1985.

[125] S. Lukaschuk, S. Nazarenko, S. McLelland, and P. Denissenko. Gravity wave turbulence in wave tanks: space and time statistics. *Physical Review Letters*, 103(4):044501, 2009.

[126] J. Lukkarinen and H. Spohn. Weakly nonlinear schrödinger equation with random initial data. *Inventiones mathematicae*, 183(1):79–188, 2011.

[127] Y. V. Lvov and S. Nazarenko. Noisy spectra, long correlations, and intermittency in wave turbulence. *Physical Review E*, 69:066608, 2004.

[128] V. S. L'vov, Y. L'vov, A. C. Newell, and V. Zakharov. Statistical description of acoustic turbulence. *Physical Review E*, 56(1):390, 1997.

[129] A. Majda, D. McLaughlin, and E. Tabak. A one-dimensional model for dispersive wave turbulence. *Journal of Nonlinear Science*, 7(1):9–44, 1997.

[130] R. Meyrand, K. Kiyani, and S. Galtier. Weak magnetohydrodynamic turbulence and intermittency. *Journal of Fluid Mechanics*, 770:R1, 2015.

[131] B. Miquel, A. Alexakis, C. Josserand, and N. Mordant. Transition from wave turbulence to dynamical crumpling in vibrated elastic plates. *Physical Review Letters*, 111(5):054302, 2013.

[132] A. K. Mishra and V. P. Singh. A review of drought concepts. *Journal of hydrology*, 391(1-2):202–216, 2010.

[133] M. A. Mohamad, W. Cousins, and T. P. Sapsis. A probabilistic decomposition-synthesis method for the quantification of rare events due to internal instabilities. *Journal of Computational Physics*, 322:288–308, 2016.

[134] A. S. Monin and A. M. Yaglom. *Statistical fluid mechanics, volume II: Mechanics of turbulence*, volume 2. Dover, 2013.

[135] N. Mordant. Are there waves in elastic wave turbulence? *Physical Review Letters*, 100(23):234505, 2008.

[136] P. Müller, C. Garrett, and A. Osborne. Rogue waves. *Oceanography*, 18(3):66, 2005.

[137] S. Nazarenko. *Wave Turbulence*, volume 825 of *Lecture Notes in Physics*. Springer, 2011.

[138] S. Nazarenko and S. Lukaschuk. Wave turbulence on water surface. *Annual Review of Condensed Matter Physics*, 7:61–88, 2016.

[139] S. Nazarenko, S. Lukaschuk, S. McLelland, and P. Denissenko. Statistics of surface gravity wave turbulence in the space and time domains. *Journal of Fluid Mechanics*, 642:395–420, 2010.

[140] A. C. Newell and P. J. Aucoin. Semidispersive wave systems. *Journal of Fluid Mechanics*, 49(03):593–609, 1971.

[141] A. C. Newell, S. Nazarenko, and L. Biven. Wave turbulence and intermittency. *Physica D: Nonlinear Phenomena*, 152:520–550, 2001.

[142] A. C. Newell and B. Rumpf. Wave turbulence. *Annual Review of Fluid Mechanics*, 43:59–78, 2011.

[143] P. Ney. Dominating points and the asymptotics of large deviations for random walk on $\mathbb{R}^d$. *The Annals of Probability*, 11(1):158–167, 1983.

[144] I. Nikolkina and I. Didenkulova. Rogue waves in 2006–2010. *Nat. Hazards Earth Syst. Sci.*, 11(11):2913–2924, 2011.

[145] M. Onorato, A. Osborne, M. Serio, L. Cavaleri, C. Brandini, and C. Stansberg. Observation of strongly non-Gaussian statistics for random sea surface gravity waves in wave flume experiments. *Physical Review E*, 70(6):067302, 2004.

[146] M. Onorato, A. R. Osborne, M. Serio, and S. Bertone. Freak waves in random oceanic sea states. *Physical Review Letters*, 86(25):5831, 2001.

[147] M. Onorato, A. R. Osborne, M. Serio, and L. Cavaleri. Modulational instability and non-gaussian statistics in experimental random water-wave trains. *Physics of Fluids*, 17(7):078101, 2005.

[148] M. Onorato, A. R. Osborne, M. Serio, L. Cavaleri, C. Brandini, and C. T. Stansberg. Extreme waves, modulational instability and second order theory: wave flume experiments on irregular waves. *European Journal of Mechanics-B/Fluids*, 25(5):586–601, 2006.

[149] M. Onorato, D. Proment, G. El, S. Randoux, and P. Suret. On the origin of

heavy-tail statistics in equations of the nonlinear Schrödinger type. *Physics Letters A*, 380(39):3173–3177, 2016.

[150] M. Onorato, S. Residori, U. Bortolozzo, A. Montina, and F. Arecchi. Rogue waves and their generating mechanisms in different physical contexts. *Physics Reports*, 528(2):47–89, 2013.

[151] M. Onorato, S. Resitori, and F. Baronio. *Rogue and shock waves in nonlinear dispersive media*, volume 926. Springer, 2016.

[152] M. Onorato, T. Waseda, A. Toffoli, L. Cavaleri, O. Gramstad, P. Janssen, T. Kinoshita, J. Monbaliu, N. Mori, A. R. Osborne, et al. Statistical properties of directional ocean waves: the role of the modulational instability in the formation of extreme events. *Physical Review Letters*, 102(11):114502, 2009.

[153] W. M. Organization, editor. *Manual on Codes - International Codes, Volume I.1, Annex II to the WMO Technical Regulations: part A- Alphanumeric Codes*. WMO- No. 306. Secretariat of the World Meteorological Organization, 2016.

[154] A. R. Osborne, M. Onorato, and M. Serio. The nonlinear dynamics of rogue waves and holes in deep-water gravity wave trains. *Physics Letters A*, 275(5):386–393, 2000.

[155] R. Peierls. Zur Kinetischen Theorie der Wärmeleitung in Kristallen. *Annalen der Physik*, 395(8):1055–1101, 1929.

[156] D. H. Peregrine. Water waves, nonlinear Schrödinger equations and their solutions. *The ANZIAM Journal*, 25(01):16–43, July 1983.

[157] A. Picozzi, J. Garnier, T. Hansson, P. Suret, S. Randoux, G. Millot, and D. Christodoulides. Optical wave turbulence: Towards a unified nonequilibrium thermodynamic formulation of statistical nonlinear optics. *Physics Reports*, 542(1):1–132, 2014.

[158] A. Picozzi, J. Garnier, T. Hansson, P. Suret, S. Randoux, G. Millot, and D. N. Christodoulides. Optical wave turbulence: Towards a unified nonequilibrium thermodynamic formulation of statistical nonlinear optics. *Physics Reports*, 542:1–132, 2014.

[159] L. P. Pitaevsky. Vortex lines in an imperfect Bose gas. *Soviet Physics - Journal of Experimental and Theoretical Physics*, 13(451), 1961.

[160] R.-E. Plessix. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503, 2006.

[161] D. A. Plotkin, R. J. Webber, M. E. O'Neill, J. Weare, and D. S. Abbot. Maximizing simulated tropical cyclone intensity with action minimization. *Journal of Advances in Modeling Earth Systems*, 11(4):863–891, 2019.

[162] L. Preziosi and L. Rondoni. Conservative energy discretization of boltzmann collision operator. *Quarterly of Applied Mathematics*, 57(4):699–721, 1999.

[163] A. N. Pushkarev and V. E. Zakharov. Turbulence of capillary waves. *Physical Review Letters*, 76(18):3320, 1996.

[164] F. Ragone, J. Wouters, and F. Bouchet. Computation of extreme heat waves in climate models using a large deviation algorithm. *Proceedings of the National Academy of Sciences*, 115(1):24–29, 2018.

[165] R. Rajaraman. Solitons and instantons. 1982.

[166] V. Raman and M. Hassanaly. Emerging trends in numerical simulations of combustion systems. *Proceedings of the Combustion Institute*, 37(2):2073–2089, 2019.

[167] S. Randoux, P. Walczak, M. Onorato, and P. Suret. Intermittency in integrable turbulence. *Physical Review Letters*, 113(11):113902, 2014.

[168] J.-M. Robine, S. L. K. Cheung, S. Le Roy, H. Van Oyen, C. Griffiths, J.-P. Michel, and F. R. Herrmann. Death toll exceeded 70,000 in Europe during the summer of 2003. *Comptes rendus biologies*, 331(2):171–178, 2008.

[169] I. N. Sanov. On the probability of large deviations of random magnitudes (in Russian). *Mat. Sb.*, 42 (84):11–44, 1957.

[170] T. P. Sapsis and A. J. Majda. Statistically accurate low-order models for uncertainty quantification in turbulent dynamical systems. *Proceedings of the National Academy of Sciences*, 110(34):13705–13710, 2013.

[171] T. Schäfer and E. V. Shuryak. Instantons in QCD. *Reviews of Modern Physics*, 70(2):323, 1998.

[172] U. Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics*, 75(12):126001, 2012.

[173] Y.-K. Shi and G. L. Eyink. Resonance van hove singularities in wave kinetics. *Physica D: Nonlinear Phenomena*, 332:55–72, 2016.

[174] V. I. Shrira and V. V. Geogjaev. What makes the Peregrine soliton so special as a prototype of freak waves? *Journal of Engineering Mathematics*, 67(1):11–22, 2010.

[175] A. Shwartz and A. Weiss. *Large deviations for performance analysis: queues, communication and computing*, volume 5. CRC Press, 1995.

[176] D. Sornette. *Why stock markets crash: critical events in complex financial systems*, volume 49. Princeton University Press, 2017.

[177] H. Spohn. The phonon Boltzmann equation, properties and link to weakly anharmonic lattice dynamics. *Journal of Statistical Physics*, 124(2-4):1041–1104, 2006.

[178] M. Stiassnie. Note on the modified nonlinear Schrödinger equation for deep water waves. *Wave motion*, 6(4):431–433, 1984.

[179] T. Stocker, D. Qin, G. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley. Ipcc, 2013: Climate change 2013: The physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change, 1535 pp, 2013.

[180] P. Suret, R. El Koussaifi, A. Tikan, C. Evain, S. Randoux, C. Szwaj, and S. Bielawski. Single-shot observation of optical rogue waves in integrable turbulence using time microscopy. *Nature Communications*, 7, 2016.

[181] J. Tailleur and J. Kurchan. Probing rare physical trajectories with Lyapunov weighted dynamics. *Nature Physics*, 3(3):203, 2007.

[182] M. Tanaka and N. Yokoyama. Numerical verification of the random-phase-and-amplitude formalism of weak turbulence. *Physical Review E*, 87(6):062922, 2013.

[183] M. A. Tayfun and F. Fedele. Expected shape of extreme waves in storm seas. In *ASME 2007 26th International Conference on Offshore Mechanics and Arctic Engineering*, pages 53–60. American Society of Mechanical Engineers, 2007.

[184] R. Temam. Sur la stabilité et la convergence de la méthode des pas fractionnaires. *Annali di Matematica pura ed applicata*, 79(1):191–379, 1968.

[185] A. Tikan, C. Billet, G. El, A. Tovbis, M. Bertola, T. Sylvestre, F. Gustave, S. Randoux, G. Genty, P. Suret, et al. Universality of the Peregrine soliton in the focusing dynamics of the cubic nonlinear Schrödinger equation. *Physical Review Letters*, 119(3):033901, 2017.

[186] S. Toenger, T. Godin, C. Billet, F. Dias, M. Erkintalo, G. Genty, and J. M. Dudley. Emergent rogue wave structures and statistics in spontaneous modulation instability. *Scientific reports*, 5, 2015.

[187] H. Touchette. The large deviation approach to statistical mechanics. *Physics Reports*, 478(1-3):1–69, 2009.

[188] H. Touchette. Introduction to dynamical large deviations of Markov processes. *Physica A: Statistical Mechanics and its Applications*, 504:5–19, 2018.

[189] F. Tröltzsch. *Optimal control of partial differential equations*. Graduate Studies in Mathematics, 112. American Mathematical Soc., 2010.

[190] K. Trulsen and K. B. Dysthe. A modified nonlinear Schrödinger equation for broader bandwidth gravity waves on deep water. *Wave motion*, 24(3):281–289, 1996.

[191] E. Vanden-Eijnden and J. Weare. Rare Event Simulation of Small Noise Diffusions. *Communications on Pure and Applied Mathematics*, 65(12):1770–1803, Sept. 2012.

[192] S. R. S. Varadhan. *Large deviations*. Courant Lecture Notes, 27. American Mathematical Soc., 2016.

[193] A. Walther. Automatic differentiation of explicit Runge-Kutta methods for optimal control. *Computational Optimization and Applications*, 36(1):83–108, 2007.

[194] R. J. Webber, D. A. Plotkin, M. E. O'Neill, D. S. Abbot, and J. Weare. Practical rare event sampling for extreme mesoscale weather. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(5):053109, 2019.

[195] B. S. White and B. Fornberg. On the chance of freak waves at sea. *Journal of Fluid Mechanics*, 355:113–138, 1998.

[196] G. B. Whitham. *Linear and nonlinear waves*, volume 42. John Wiley & Sons, 2011.

[197] L. C. Wilcox, G. Stadler, T. Bui-Thanh, and O. Ghattas. Discretely exact derivatives for hyperbolic PDE-constrained optimization problems discretized by the discontinuous Galerkin method. *Journal of Scientific Computing*, 63(1):138–162, 2015.

[198] S. Wright and J. Nocedal. *Numerical optimization*. Springer Series in Operations Research. Springer, 2000.

[199] H. W. Wyld. *Annals of Physics*, 14:143, 1961.

[200] V. Zakharov and N. Filonenko. Energy spectrum for stochastic oscillations of the surface of a liquid. In *Soviet Physics Doklady*, volume 11, page 881, 1967.

[201] V. E. Zakharov. Stability of periodic waves of finite amplitude on the surface of a deep fluid. *Journal of Applied Mechanics and Technical Physics*, 9(2):190–194, 1968.

[202] V. E. Zakharov. Stability of periodic waves of finite amplitude on the surface of a deep fluid. *Journal of Applied Mechanics and Technical Physics*, 9(2):190–194, 1968.

[203] V. E. Zakharov. Turbulence in integrable systems. *Studies in Applied Mathematics*, 122(3):219–234, 2009.

[204] V. E. Zakharov and N. N. Filonenko. *Zh. Prikl. Mekh. I Tekn. Fiz.*, 5:62, 1967.

[205] V. E. Zakharov and V. S. L'vov. Statistical description of nonlinear wave fields. *Radiophysics and Quantum Electronics*, 18(10):1084–1097, 1975.

[206] V. E. Zakharov, V. S. L'vov, and G. Falkovich. *Kolmogorov spectra of turbulence I. Wave turbulence, Springer series in nonlinear dynamics*. Springer Berlin, 1992.

[207] V. E. Zakharov and L. A. Ostrovsky. Modulation instability: The beginning. *Physica D: Nonlinear Phenomena*, 238(5):540–548, Mar. 2009.

[208] V. E. Zakharov and R. Z. Sagdeev. Spectrum of acoustic turbulence. In *Soviet Physics Doklady*, volume 15, page 439, 1970.

[209] G. Zaslavskii and R. Sagdeev. Limits of statistical description of a nonlinear wave field. *Soviet Journal of Experimental and Theoretical Physics*, 25:718, 1967.