NORTHWESTERN UNIVERSITY

Rare Events in a Mode-Locked Fiber Laser

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Applied Mathematics

By

Nathan Len Sanford

EVANSTON, ILLINOIS

December 2020

ABSTRACT

Rare Events in a Mode-Locked Fiber Laser

Nathan Len Sanford

Optical fibers utilize nonlinear effects to help transmit soliton or near soliton pulses in a variety of contexts including optical communication systems and fiber lasers. Fiber lasers produce ultra-short pulses, down to a few femtoseconds in duration, via a process called mode-locking where modes of the optical cavity are synchronized via a variety of physical mechanisms to produce stable solitons. Active mode-locking, where an optical device such as a modulator facilitates the synchronization, is a scheme popular for the level of control it gives to the mode-locking process. Even with such control, however, pulses in an actively mode-locked laser can occasionally slip relative to the timing signal leading to fluctuations in the pulse repetition rate.

Monte Carlo simulations can be used to capture such errors, but are ineffective when the events are rare as the number of simulations required to determine their probability is too computationally expensive. In modern nonlinear optical fibers, error rates are typically below 10^{-10} so traditional methods are not within the realm of possibility. Importance sampling modifies Monte Carlo simulations to make them feasible by

more efficiently sampling from regions in phase space that give rise to rare events via the introduction of biasing distributions. The identification of important regions is usually accomplished via the solution of an optimization problem. In nonlinear optics, the formulation of such problems is frequently guided by soliton perturbation theory. Here, quantification of position slip error rates in a model of a soliton-based mode-locked laser are obtained using importance-sampled Monte Carlo simulations where the physical effects of the active mode-locking are incorporated as small perturbations. Position slips are studied in two distinct cases: an overdamped regime where they are primarily direct, and an underdamped regime where they typically involve oscillations.

The qualitative distinction of direct versus oscillatory biasing paths between these regimes motivates the use of different importance sampling techniques. Quantification of the slip rate is shown to be more straightforward in the overdamped regime. Dynamic importance sampling, where error pathways are dynamically recomputed mid-simulation to maximize error yield, is found to be necessary to accurately and efficiently capture error rates. Dynamic importance sampling is shown to be more difficult to implement in the underdamped regime due to the existence of multiple routes by which position slip errors can occur. This motivates the development of a new importance sampling method blending dynamic importance sampling and multiple importance sampling, where multiple biasing distributions are used simultaneously. This algorithm is first applied to a toy model with multiple paths, a two-dimensional random walk past a transverse wall obstacle, and then is applied to the underdamped laser model. In both cases, this dynamic multiple importance sampling method is found to be far superior to using either dynamic importance sampling or multiple importance sampling in isolation.

Acknowledgements

I would like to thank and acknowledge the following people. Bill Kath, for advising my thesis and giving me invaluable advice and assistance. My family for always being there to support me, and in particular my sister Emily for also choosing to get a PhD (despite seeing firsthand how difficult it can be). My friends from my cohort and the department including, but not limited to, Albert Berahas, Brennan Sprinkle, Eric Autry, Nishant Nangia, and Yuxin Chen. And last, I would like to thank my partner Lauren Cox for believing in me even when I didn't believe in myself.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

The achievements in the field of optical fiber communication systems are many and varied. From facilitating the highest precision research in astrophysics through the use of optical clocks [**13**], to carrying the bulk of internet traffic between continents [**10, 79**], the world would be a drastically different place without the invention of fiber-optics. A significant insight that has accompanied these achievements has been the discovery of optical solitons, high amplitude solitary waves generated by combining nonlinear and dispersive effects. Maintaining a precise balance between these two effects, chromatic dispersion and Kerr nonlinearity, allows for the stable generation and transmission of high energy pulses. First proposed as optical carriers nearly a half-century ago, and experimentally demonstrated soon thereafter, these pulses operate as solitary waves and have proved effective in overcoming dispersion-related signal corruption [**49, 70**].

The study of nonlinear waves and solitons dates back to the early nineteenth century when Scottish engineer John Scott Russell noted a water wave propagating down the Union Canal in Edinburgh that did not change in character as it propagated, calling it a "wave of translation" [**28**]. This translational behavior was at odds with the understanding of hydrodynamics at the time, and it wasn't until the late nineteenth century when theories and equations such as the Korteweg-de Vries equation (KdV) were developed that explained such solitary wave behavior [**63**]. The field was further revolutionized in

the late 1960s with the discovery of the inverse scattering transform for solving KdV analytically and explicitly, from which solitons were found to act akin to particles [**39, 96**]. This analytical solution method was soon extended to other nonlinear partial differential equations with a similar structure, creating a class of so-called "integrable systems." The nonlinear Schrödinger equation was investigated in this context as it was also a nonlinear PDE arising in water waves: KdV arises in the shallow-water approximation while NLS arises in the deep-water asymptotic approximation to the problem of surface water waves [**55**]. In particular, the works of Zakharov and Ablowitz et. al. were instrumental in extending the integrability framework to NLS [**2, 98, 99**]. They obtained the NLS soliton from the inverse scattering transform, establishing its integrability in a rather robust fashion. At the same time, solitons were found to develop from a sequence of waves by means of a mechanism called "modulation instability" whereby periodic wave forms can focus into steeper pulses. This corroborated a multitude of findings in the preceding decade of the modulation instability occurring in a variety of physical systems [**9, 97**]. It was in this context and with this firm theoretical footing that solitons were proposed as potential carriers in optical systems [**49, 50**].

The primary factor limiting soliton use in practical systems was power loss from energy absorption in the propagation medium. While solitons were readily obtainable in laboratory settings throughout the 1980s, power loss kept them from being broadly applicable and useful in real-world applications since power loss leads to pulse broadening [**5**]. The development of fiber amplifiers offered a solution to this problem. In particular, Erbium-doped fiber amplifiers (EDFAs) came to attract attention as they operate in one

of the low-loss wavelength regimes of fused silica [**23**]. Fiber amplifiers, broadly speaking, work by injecting energy into the propagation medium which counteracts loss. The method of amplification can be either distributed (or acting throughout the medium) or lumped (acting at discrete points). Lumped amplification has been used in a variety of contexts as it requires less energy and can effectively combat loss if amplifiers are spaced at intervals much less than the characteristic length on which losses occur. With this technique a number of practical soliton-based schemes were developed in the 1990s, including dissipative solitons and dispersion-managed solitons among others [**6, 48**]. While lumped amplification has obvious benefits, it comes with the inherent drawback that it injects noise into the system in addition to combating loss. This noise, called amplified spontaneous emission (ASE) noise, is a pernicious phenomenon that can cause system failures.

In particular, additive ASE noise causes perturbations to soliton pulses during transmission that can lead to communication errors. To see how this occurs, consider a transmission scheme where solitons are propagated along the fiber in a sequence of specified bit-slots. The presence of a soliton in a bit-slot signals a "1" bit and no soliton is a "0." Such a transmission scheme is called "on-off keying" though other schemes are sometimes used [**5**]. Repeated injection of ASE noise, which is modeled as an instantaneous addition of Gaussian white noise at each amplifier, can perturb the pulse sufficiently so that a 1 is misdetected as a 0 or vice versa. The pulse perturbations through which this most commonly occurs are called parameter jitter; the pulse maintains its soliton character, but changes primarily through changes in its parameters [**52**]. The NLS soliton is characterized by four parameters; amplitude, frequency, position, and phase; with all values

admissible for each due to the integrable structure of the NLS equation. Amplifier noise can then cause these values to change without resistance, leading to large pulse deformations. Therefore, a pulse can undergo drastic reductions in amplitude, for instance, which would cause a 1-to-0 error. Such errors are typically rare in occurrence, so rare event theory is needed to understand their dynamics.

Rare events have been studied in a variety of contexts including financial markets, ocean waves, social systems, and climate disasters [**19, 22, 42, 62, 77**]. Two of the most important questions to address when studying rare events are assessing their prevalence or probability and understanding the dynamics that lead to their occurrence. As such, a number of theoretical and analytical approaches have been developed to address these questions [**38**]. However, computational studies play an important role in the study of rare events as well, as the nonlinear process of pulse propagation causes the Gaussian white noise input of the additive ASE noise to lead to distinctly non-Gaussian output statistics of pulse behavior. However, the use of Monte Carlo simulations, one of the most widespread techniques in the field of random simulation, is inherently limited in the case of studying rare events [**75**]. By definition a rare event has a low probability, and a large number of Monte Carlo trials and concomitant computational outlay is required to capture its probability reliably. In the case of optical communication systems, error rates have long been engineered to be extremely low; lower than one error per trillion bits [**73**]. Nevertheless, the increased demand for ever increasing transmission bandwidths and stable pulse repetition has motivated the computational study of errors to understand how they occur, so that they can be reduced as much as possible. The extremely low rate

of errors in already existing technologies provides a significant modeling and simulation challenge in this pursuit.

Adapting Monte Carlo simulations to rare event circumstances has been an ongoing field of study for decades. One such scheme is importance sampling where a *biasing distribution* is introduced that gives rise to the rare events of interest much more frequently. Samples are drawn from the biasing distribution, and then trials are weighted so that corrected probabilities of interest may be calculated. The selection of biasing distributions is a significant avenue of ongoing inquiry [**76**]. The chief criterion to keep in mind when investigating possible biasing distributions is that the variance of the Monte Carlo simulations should be reduced, and so importance sampling and similar techniques are referred to as variance reduction methods. In the systems considered in this thesis, it requires identifying the paths through state space which lead to errors most frequently. System reductions using soliton perturbation theory make it possible to approximate the infinite-dimensional function space of solitons in a finite-dimensional parameter space, making this objective attainable [**73, 83**]. Optimization methods are then implemented in the parameter space to find the likeliest error paths which then guide simulation of the full NLS. This methodology has been applied to studies of errors in a multitude of soliton-based systems, and will be taken here to study an actively mode-locked laser model which presents novel dynamic features and challenges.

Fiber ring lasers are a nonlinear optical system that is important in a wide variety of contexts. They have been utilized as pulse sources in optical communication systems, but have also had significant roles in scientific inquiry such as in optical frequency metrology and optical clocks for extremely precise and accurate measurements in spectroscopy

and other pursuits [**89**]. The mechanism by which such lasers generate sufficiently short pulses, as short as on a femtosecond time scale, is called mode-locking and has been studied for over fifty years [**21, 51**]. Mode-locking refers to the frequency domain process underlying pulse generation in such lasers, where stable lines at specified modes in the frequency domain underlie the soliton dynamics observed in the temporal domain. Such a mechanism relies on the ring-like structure of such a laser, and requires that the pulse be unchanged after propagating around the ring, which imposes the specified frequencies at which lasing can occur [**20**]. From a dynamical perpective, the characteristic balance of dispersion and nonlinearity present in a mode-locked soliton-system is augmented with a number of small perturbations achieved by adding physical elements to the fiber cavity. These perturbations help resist noise-induced parameter jitter, and thus help reduce the frequency of errors occurring.

Physically, the perturbative elements used in a mode-locked fiber laser are typically applied either throughout propagation or at a single point in the ring with the EDFA amplifier being a canonical example. The perturbative elements present in a mode-locked laser broadly fall under two regimes, active mode-locking and passive mode-locking. The distinction between the schemes relies on the physical element causing mode-locking, where either an added element causes the locking in active schemes, or the pulse itself causes the mode-locking in passive schemes [**20**]. In both cases, intensity-dependent perturbations (i.e. those that depend on the modulus of the optical field), also called intensity discrimination perturbations, are usually incorporated via the inclusion of absorbers, polarizers, and other gain elements [**65**]. Such elements help to fix the pulse amplitude at

a specific value. Other perturbations targeting non-amplitude-based jitter are also incorporated, such as with a frequency-dependent filtering element to combat perturbations to the pulse frequency. Of particular interest are active feedback mechanisms that combat pulse walk-off from the bit-slot center, also called timing jitter. This thesis examines a particular active mode-locking mechanism where an active phase modulator is used to explicitly resist timing jitter [**71**].

This thesis examines the dynamics of this actively mode-locked fiber laser model, as it produces dynamically interesting timing-jitter errors. The dynamical description of such errors, as well as the modifications to existing importance sampling algorithms to account for this novel dynamics, are the focus of this thesis. To that end, the thesis is organized in the following order. Chapter 2 provides an introduction to the physics of nonlinear optical systems, their numerical simulation, and soliton perturbation theory. Chapter 3 explains the basics of importance sampling using a random walk as an illustrative example. Augmentations of importance sampling, such as multiple importance sampling and dynamic importance sampling, are also discussed. Chapter 4 introduces an importance sampling algorithm that blends elements of multiple and dynamic importance sampling that will be used in the investigation of our actively mode-locked laser. We discuss why this algorithm is only advantageous to use in very specific circumstances, e.g., when there are multiple paths through phase space to the same rare event, and provide a toy random walk model where there are two paths past an obstacle. Lastly, Chapter 5 discusses the laser model, and shows how the multiple-path phenomenon arises, and how it may be dealt with in importance sampling simulations.

CHAPTER 2

Mathematics Governing Nonlinear Optical Systems

2.1. The Nonlinear Schrödinger Equation

The nonlinear Schrödinger equation (NLS) is a partial differential equation arising in many areas of scientific inquiry including Bose-Einstein Condensates and nonlinear water waves [**45, 98**]. However, its applications in nonlinear optics is the focus of this thesis. In this context, the nonlinear Schrödinger equation describes the propagation of short pulses through a fiber medium with significant dispersive and nonlinear effects.

2.1.1. Comments on its derivation

Consider the propagation of an optical field through an optical fiber medium. This process is governed by Maxwell's equations, which can be written as a wave equation involving the electric field $\mathbf{E}$ and the medium's response to the electric field $\mathbf{P}$ (called the induced electric polarization),

$$\nabla^2\mathbf{E} - \frac{1}{c^2}\frac{\partial^2\mathbf{E}}{\partial\tau^2} = \mu_0\frac{\partial^2\mathbf{P}}{\partial\tau^2}, \tag{2.1}$$

where τ, c, and μ_0 are time, the speed of light in a vacuum, and the vacuum's magnetic permeability. The derivation of NLS from Eq. (2.1) will not be done in full here, but its details can be found in [**5**]. Instead, we will simply state the assumptions necessary for deriving NLS, along with a few explanatory comments.

Here, we assume that the main nonlinear effects present in the fiber contributing to the polarization $\mathbf{P}$ are due to the third-order nonlinearity of the material, typically arising through nonlinear refraction. Further, we assume that the polarization can be decomposed into linear and nonlinear components so that $\mathbf{P} = \mathbf{P}_L + \mathbf{P}_{NL}$, with $\mathbf{P}_{NL}$ serving as a small perturbation to $\mathbf{P}_L$. With these two assumptions in place, the polarization can be Taylor expanded in the electric field with a linear and cubic term as the first two terms. A quadratic term is not present as the material, typically fused silica for fiber, is assumed to be centro-symmetric, causing its second-order susceptibility to be zero.

Another needed assumption about the fiber medium is that it is a single-mode fiber, meaning that it admits one propagating mode at typical wavelengths. Such fibers, which typically have a core radius below 2 μm and are single-mode for wavelengths $\lambda > 1.2$ μm, are used widely. Additionally, we assume that the single-mode optical field maintains its polarization along the fiber length, which is true for so-called "polarization-maintaining" fibers and is an approximation that holds for a variety of other fibers as well. With this approximation, we can decompose the vector dynamics of Eq. (2.1) into scalar dynamics of propagation along a single polarization direction.

Lastly, we assume that the optical field is "quasi-monochromatic" meaning that it has a spectrum with a single peak at a frequency ω_0 and a very narrow spectral width $\Delta\omega$ around this peak with $\Delta\omega \ll \omega_0$. In practice pulses as short as 0.1 ps obey this condition; typically communication systems use longer pulses, however. With these assumptions in hand, we utilize a slowly-varying envelope approximation by taking an electric field solution to Eq. (2.1) that has three parts: the fiber mode describing the transverse form of the field, a slowly-varying envelope U that depends on propagation distance and

either time or frequency, and a plane wave capturing the rapidly varying phase of the field. Letting z be the fiber propagation distance variable, the slowly-varying-envelope-approximation implies $\frac{\partial^2 U}{\partial z^2} \ll \frac{dU}{dz}$ as U is slowly varying so the second spatial derivative of the envelope is neglected.

This approximation results in

$$i\frac{dU}{dz} - \frac{\beta''}{2}\frac{\partial^2 U}{\partial t^2} + \gamma|U|^2 U + i\alpha U = 0, \tag{2.2}$$

where β'' is the group-velocity dispersion (GVD) coefficient, γ is the nonlinear coefficient, α is the linear loss coefficient (with $\alpha > 0$), and $t = \tau - z/v_g$ is the retarded time (where v_g is the group-velocity coefficient). The dispersion of the field has been dealt with approximately by taking a Taylor expansion of the plane wave's wavenumber β around the central frequency ω_0:

$$\beta(\omega) \approx \beta_0 + \beta_1(\omega - \omega_0) + \frac{1}{2}\beta_2(\omega - \omega_0)^2 \tag{2.3}$$

with higher-order terms neglected under the quasi-monochromatic assumption. Then, we can identify

$$v_g = \frac{1}{\beta_1} \quad \text{and}$$

$$\beta'' = \beta_2$$

which helps motivate the name of the GVD coefficient in particular as $\beta_2 = \beta''(\omega_0)$. Thus, Eq. (2.2) is called the "lossy" nonlinear Schrödinger equation in dimensional form.

2.1.2. Amplifiers and noise

A few more comments on the units used to describe the system are in order before we proceed to the nondimensional form. In dimensional form, U is normalized so that $|U|^2$ represents the optical power (in W), the retarded time t has units of ps, and z is in km. The dispersion constants v_g and β'' have units of km/ps and ps^2/km respectively. The GVD constant β'' is assumed to be strictly negative, putting the system in what is known as the anomalous dispersion regime, for reasons explained in the next section. β'' will be replaced by its absolute value in what follows. The loss coefficient has units of 1/km while the nonlinear coefficient γ has units of 1/(W-km) and is given by

$$\gamma = \frac{2\pi n_2}{\lambda A_{\text{eff}}}. \tag{2.4}$$

In equation (2.4) n_2 is the Kerr coefficient, λ is the light wavelength, and A_{eff} is the effective area of the fiber mode. The linear loss term is in units of 1/km, and can cause significant pulse attenuation

Erbium-doped fiber amplifiers (EDFAs) are placed periodically throughout the transmission length to counteract the linear loss in the system. First realized experimentally in 1989, such amplifiers are in common usage, though other types of amplification such as Raman amplification are still sometimes used [**4**]. These devices instantaneously produce an amount of power gain G as the signal passes through them that counteracts the accumulated loss between amplifiers. Unfortunately, amplified spontaneous emission (ASE) noise is also added as part of this process. Over time this noise builds up and can distort the signal to the point where accurate transmission is impeded.

ASE noise is approximated by additive zero-mean Gaussian white noise. While Gaussian white noise has a nonphysical infinite bandwidth, it suitably approximates ASE noise as its bandwidth is much larger than that of the signal [**44**]. The variance of ASE noise is controlled by its power spectral density,

$$S(\omega_0) = \hbar\omega_0\eta_{\text{sp}}(G-1),$$

where $\hbar$ is Planck's constant and η_{sp} is the spontaneous emission factor (with $\hbar\omega_0$ being the energy per photon of light at ω_0 the carrier frequency) [**44**]. Therefore, the noise injected by an amplifier is modeled with

$$\langle N(t)\rangle = 0 \quad \text{and} \tag{2.5a}$$

$$\langle N(t)N^*(t')\rangle = \hbar\omega_0\eta_{\text{sp}}(G-1)\delta(t-t') \tag{2.5b}$$

where half of the noise is added to each quadrature of the complex signal U.

We mentioned previously that the power gain parameter G from the EDFAs is chosen to counterbalance the linear loss term $i\alpha U$ in Eq. (2.2). To see how this parameter is tuned, we take the ansatz $U(z,t) = v(z)U_0(z,t)$ in the lossy NLS equation and require that $v(z)$ satisfies

$$\frac{dv}{dz} = -\alpha v \quad (\text{so } v(z) = v_0\exp(-\alpha z))$$

so v captures the linear loss in the system. The power gain coefficient measures the ratio of the signal power directly after to before the amplifier so a power gain of

$$G = \frac{(v(z_a))^2}{(v(0))^2} = \exp(2\alpha z_a) \tag{2.6}$$

is needed to compensate for loss over a fiber length of z_a km (taken to be the distance between amplifiers). The jump condition for v given by passing through an amplifier is then

$$v_+ = \sqrt{G}\, v_- + N(t) \tag{2.7}$$

where v_+ and v_- signify the solution directly after and before amplification. Therefore, the lossy NLS equation Eq. (2.2) with this ansatz and form for $v(z)$ gives (with $U_0 \to U$)

$$i\frac{dU}{dz} + \frac{|\beta''|}{2}\frac{\partial^2 U}{\partial t^2} + v(z)^2\gamma|U|^2U = 0 \tag{2.8}$$

with an associated amplifier jump condition of

$$v_0 U_+ = v_0 U_- + N(t). \tag{2.9}$$

Equation 2.8 can then be simplified by replacing $v(z)$ with its average value, chosen to be unity, in view of the fact that the attenuation length scale is typically much smaller than the length scales on which the dispersive and nonlinear effects take place [**60**]. This choice then implies

$$v_0^2 = \frac{2\alpha z_a}{(1-\exp(-2\alpha z_a))} = \frac{G\ln G}{G-1}. \tag{2.10}$$

The preceding averaging process leads to the following equation

$$i\frac{dU}{dz} + \frac{|\beta''|}{2}\frac{\partial^2 U}{\partial t^2} + \gamma|U|^2U = N(t,z), \tag{2.11}$$

with the periodically-added ASE noise modeled explicitly through the noise process N which takes the form

$$N(t,z) = \sum_{k=1}^{N_a} n_k(t)\delta(z - kz_a), \tag{2.12a}$$

$$\langle n_k(t)\rangle = 0, \tag{2.12b}$$

$$\langle n_k(t) n_l^*(t')\rangle = \frac{(G-1)^2}{G\ln G}\hbar\omega_0\eta_{\mathrm{sp}}\delta_{kl}\delta(t-t') \tag{2.12c}$$

where N_a is the total number of amplifiers, again z_a is the amplifier spacing, and the noise added at each amplifier is delta-correlated Gaussian white noise. It should be noted that the noise power Eq. (2.12c) differs from that given in Eq. (2.5) due to dividing through by v_0 because of Eq. (2.9).

2.1.3. Nondimensionalization

The dimensional form of Eq. (2.11) is often exchanged for a nondimensional form for the purpose of mathematical analysis and simulation. We will follow the nondimensionalization used in [**73**], though other nondimensionalizations have been used [**54**]. The first step in the nondimensionalization is to introduce two characteristic length scales associated with the dispersive (L_D) and nonlinear (L_{NL}) effects experienced by a characteristic pulse of width T_0 ps and power P_0 W with

$$L_D = \frac{T_0^2}{|\beta''|}, \text{ and} \tag{2.13a}$$

$$L_{NL} = \frac{1}{\gamma P_0}. \tag{2.13b}$$

Note that the scaling constants T_0 and P_0 are simply reference values and a pulse form has not been assumed at this point. The pulse dynamics we are interested in exploring take place in the regime where the nonlinear length and dispersion length are of the same order of magnitude, so we equate $L_D = L_{NL}$ and find that

$$P_0 = \frac{|\beta''|}{\gamma T_0^2}. \tag{2.14}$$

Now, we can introduce dimensionless coordinates via

$$\tilde{t} = \frac{t}{T_0}, \quad \tilde{z} = \frac{z}{L_D} = \frac{|\beta''|z}{T_0^2}, \quad \text{and} \quad u = \frac{U}{\sqrt{P_0}} = \frac{\sqrt{\gamma}\, T_0 U}{\sqrt{|\beta''|}} \tag{2.15}$$

to nondimensionalize the equation. With these dimensionless coordinates Eq. (2.11) becomes

$$i\frac{du}{dz} + \frac{1}{2}\frac{\partial^2 u}{\partial t^2} + |u|^2 u = N(t,z), \tag{2.16}$$

after dropping the tildes from $\tilde{z}$ and $\tilde{t}$. Note that the noise process is still Gaussian white noise with zero mean, but it has been scaled by the above dimensionless coordinates as well. The noise standard deviation is scaled by $(P_0 T_0)^{-1/2}$ due to the introduction of the factors in Eq. (2.15), and consequently the noise process' variance is scaled by $1/(P_0 T_0)$ so

$$N(t,z) = \sum_{k=1}^{N_a} n_k(t)\delta(z - kz_a) \quad \text{with} \tag{2.17a}$$

$$\langle n_k(t) n_l^*(t')\rangle = \frac{(G-1)^2}{G\ln G}\frac{\hbar\omega_0 \eta_{\mathrm{sp}} \gamma T_0}{|\beta''|}\delta_{kl}\delta(t-t') \equiv \sigma^2. \tag{2.17b}$$

The nondimensionalized noise variance, or power, shall be given by σ^2 for convenience.

2.2. Solitons and the Structure of the Nonlinear Schrödinger Equation

We will now consider the dimensionless NLS equation, Eq. (2.16), without noise (setting $N = 0$), as it admits a solution that will form the basis of this thesis: the soliton. In the anomalous dispersion regime, a precise balance between nonlinearity and dispersion has been found to give rise to a type of solitary wave that propagates without changing its form [**47**]. This solitary wave, called a soliton, has been studied extensively and is related to the remarkable mathematical structure of the NLS equation. The inverse scattering method, an analytical solution method for nonlinear partial differential equations whose details are outside the scope of this thesis, can been used to arrive at the soliton solution by Ablowitz et al. and Zakharov and Shabat [**2, 99**]. The solvability of NLS by the inverse scattering transform makes it what is called an integrable system. The NLS soliton has been found to be robust to the cycle of linear loss and amplification found in fibers, and indeed has been demonstrated experimentally [**49, 70**].

The single soliton solution is given by the form

$$u_s(z,t) = u_0(z,t)e^{i\Theta} \quad \text{where} \tag{2.18a}$$

$$u_0(z,t) = E\ \mathrm{sech}(E(t - T(z))), \quad \text{and} \tag{2.18b}$$

$$\Theta = \Omega(t - T(z)) + \Phi(z). \tag{2.18c}$$

The parameters of the soliton are as follows: E is the soliton amplitude (or energy), Ω is the frequency mismatch from the carrier frequency ω_0, $T(z)$ is the position of the center of the pulse, and $\Phi(z)$ is the phase of the pulse at its center. For convenience, Θ is the

shorthand for the entire argument of the complex exponential. E and Ω are constants, and $T(z) = \Omega z + T_0$ and $\Phi(z) = (E^2 + \Omega^2)z/2 + \phi$ evolve with the propagation from initial constants T_0, and ϕ. It is sometimes convenient to separate the hyperbolic secant from the complex exponential, as is done with u_0 in Eq. (2.18), for the purpose of analysis in soliton perturbation theory as we will see in the next section.

Before proceeding to an explanation of soliton perturbation theory, a few comments on the connections between the soliton form and the mathematical structure of NLS are warranted. By Noether's theorem, a symmetry or invariance in a PDE is always associated with a conservation law, or conserved quantity, of the system [**40, 94**]. The details of this result need not be reiterated here, except to note that conservation laws are technically associated with invariances of a variational principle of a PDE such as a Lagrangian or Hamiltonian.

The connections between the parameters of the soliton and conserved quantities and their associated invariances are laid out in Table 2.1. As the table elucidates, the soliton amplitude and frequency are readily defined by conservation laws whereas position and phase rely on modified conservation laws. A "modified" or "generalized" conservation law can be thought of as the analog of a generalized eigenvector. It is a conservation law that relies on another conservation law. The absolute phase cannot be defined via a conservation law, but changes in the phase can be defined using

$$\frac{i}{2}\int_{-\infty}^{\infty}\left(u\frac{\partial u^*}{\partial z} - u^*\frac{\partial u}{\partial z}\right)dt - \frac{d}{dz}\int_{-\infty}^{\infty}\frac{i}{2}t\left(u\frac{\partial u^*}{\partial t} - u^*\frac{\partial u}{\partial t}\right)dt = -\frac{3}{2}\int_{-\infty}^{\infty}\left(\left|\frac{\partial u}{\partial t}\right|^2 - |u|^4\right)dt \tag{2.19}$$

Parameter	Definition	Conservation Law	Associated Invariance
E	$E = \frac{1}{2}\int_{-\infty}^{\infty} \lvert u\rvert^2\, dt$	Mass (Photon number)	Phase invariance $u \to e^{i\nu}u$
Ω	$\Omega = \frac{1}{2E}\int_{-\infty}^{\infty} \mathrm{Im}(u^*\frac{\partial u}{\partial t})\, dt$	Momentum	Time translation $t \to t - t_0$
T	$T = \frac{1}{2E}\int_{-\infty}^{\infty} t\lvert u\rvert^2\, dt$	Modified	Galilean invariance $t \to t + \epsilon z$ $u \to e^{i\epsilon t - i\epsilon^2 t/2}u$
Φ	N/A (Changes only)	Modified	Scale invariance $u(z,t) \to au(\frac{z}{a}, \frac{t}{a^2})$

Table 2.1. The parameters of the soliton and associations with conservation laws and invariances of NLS. Note that only the amplitude E and frequency Ω of the soliton are associated with true conserved quantities of NLS. The other two quantities are associated with "modified" or "generalized" conservation laws, which can be seen as being related to the fact that their evolution is coupled to other soliton parameters. Lastly, the phase cannot be defined by a conservation law, but changes in phase are associated with Eq. (2.19).

which was noted in [**58**]. Various definitions are used to track the phase quantitatively (see [**81**]), but tracking the phase will not play a major role in this thesis so they will not be included here.

The verification of conservation laws is a calculation that will play a role in soliton perturbation theory. Such calculations also help to underline the distinction between a true and modified conservation law. Therefore, we will verify the conservation laws for amplitude and position. Consider the soliton amplitude, E, which is defined via

$$E = \frac{1}{2}\int_{-\infty}^{\infty} |u|^2\, dt.$$

E being a conservation law means that $\frac{dE}{dz} = 0$. Therefore, taking the derivative gives

$$\frac{dE}{dz} = \frac{d}{dz}\Big(\frac{1}{2}\int_{-\infty}^{\infty} |u|^2\,dt\Big), \tag{2.20a}$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} \frac{\partial}{\partial z}\big(|u|^2\big)\,dt, \tag{2.20b}$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} \frac{\partial u}{\partial z}u^* + \frac{\partial u^*}{\partial z}u\,dt, \tag{2.20c}$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} \big(\frac{i}{2}\frac{\partial^2 u}{\partial t^2} + i|u|^2u\big)u^* - \big(\frac{i}{2}\frac{\partial^2 u^*}{\partial t^2} + i|u|^2u^*\big)u\,dt, \tag{2.20d}$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} \frac{i}{2}\big(\frac{\partial^2 u}{\partial t^2}u^* - u\frac{\partial^2 u^*}{\partial t^2}\big) + i\big(|u|^4 - |u|^4\big)\,dt, \tag{2.20e}$$

$$= \frac{i}{4}\Big\{\big(\frac{\partial u}{\partial t}u^* - u\frac{\partial u^*}{\partial t}\big)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty}\big(\frac{\partial u}{\partial t}\frac{\partial u^*}{\partial t} - \frac{\partial u}{\partial t}\frac{\partial u^*}{\partial t}\big)\,dt\Big\} = 0, \tag{2.20f}$$

where we have used Eq. (2.16) and its conjugate in Eq. (2.20d), and integrated by parts in Eq. (2.20f). The decay conditions on u at $\pm\infty$ then give the desired result of 0, indicating that the pulse amplitude does not change under propagation. Now, consider the soliton position T defined by

$$T = \frac{1}{2E}\int_{-\infty}^{\infty} t|u|^2\,dt.$$

Taking the derivative with respect to z then gives

$$\frac{dT}{dz} = \frac{d}{dz}\Big(\frac{1}{2E}\int_{-\infty}^{\infty} t|u|^2\,dt\Big), \tag{2.21a}$$

$$= \frac{1}{2E}\int_{-\infty}^{\infty} t\frac{\partial}{\partial z}\big(|u|^2\big)\,dt, \tag{2.21b}$$

$$= \frac{1}{2E}\int_{-\infty}^{\infty} t\Big[\frac{i}{2}\big(\frac{\partial^2 u}{\partial t^2}u^* - u\frac{\partial^2 u^*}{\partial t^2}\big) + i\big(|u|^4 - |u|^4\big)\Big]\,dt, \tag{2.21c}$$

$$= \frac{i}{4E}\int_{-\infty}^{\infty} tu^*\frac{\partial^2 u}{\partial t^2} - tu\frac{\partial^2 u^*}{\partial t^2}\,dt, \tag{2.21d}$$

$$= \frac{i}{4E}\Big\{\big(tu^*\frac{\partial u}{\partial t} - tu\frac{\partial u^*}{\partial t}\big)\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty}\big(u + t\frac{\partial u}{\partial t}\big)\frac{\partial u^*}{\partial t} - \big(u^* + t\frac{\partial u^*}{\partial t}\big)\frac{\partial u}{\partial t}\,dt\Big\}, \tag{2.21e}$$

$$= \frac{i}{4E}\int_{-\infty}^{\infty} u\frac{\partial u^*}{\partial t} - u^*\frac{\partial u}{\partial t} + t\Big(\big|\frac{\partial u}{\partial t}\big|^2 - \big|\frac{\partial u}{\partial t}\big|^2\Big)\,dt, \tag{2.21f}$$

$$= \frac{i}{4E}\int_{-\infty}^{\infty} 2i\,\mathrm{Im}\big(u\frac{\partial u^*}{\partial t}\big)\,dt = \frac{1}{2E}\int_{-\infty}^{\infty}\mathrm{Im}\big(u^*\frac{\partial u}{\partial t}\big)\,dt = \Omega, \tag{2.21g}$$

which, while not 0, is another conservation law (that for Ω) which is the requirement for a modified conservation law. Additionally, note that the boundary terms in Eq. (2.21e) going to 0 does not simply require that $u \to 0$ at $\pm\infty$ due to the factor of t, but that the product of u and its derivative decay superlinearly. The boundary conditions placed on u are typically stronger than this, so the requirement is easily satisfied [**3**]. Also note that the result of Eq. (2.21) coincides with the parametrization in Eq. (2.18), but that the conservation laws and modified conservation laws discussed here do not depend on a particular functional form for u: it just needs to be a solution to NLS.

2.3. Soliton Perturbation Theory

Soliton perturbation theory (SPT) allows us to study effects of small perturbations to the NLS soliton system laid out in this chapter. This theory can be approached in a number of ways including using the inverse scattering transform or considering direct perturbations to the soliton [**47, 61**]. However, here we will present the theory as arising from a perturbative term in the NLS, which affects the soliton pulse's dynamics under propagation. Therefore, consider a rewritten version of Eq. (2.16) with a perturbative term f:

$$\frac{\partial u}{\partial z} = \frac{i}{2}\frac{\partial^2 u}{\partial t^2} + i|u|^2 u + f. \tag{2.22}$$

We consider this perturbation to be small in magnitude, and it may or may not include the noise process $N(t, z)$. This section will investigate the effects of this general term on the propagation of a soliton in two main ways. First, we will discuss what the conservation laws can say about these effects and then we will introduce a linearized theory that will be able to provide an extremely useful approximation for simulation purposes.

2.3.1. SPT using conservation laws

Under this perturbation, the conservation laws of the previous section provide not unchanging quantities but rather evolution equations. The derivation of these evolution equations is not particularly surprising: when the right-hand-side of NLS is substituted into the integrals in Table 2.1 extra terms arise from f. For instance, if we examine the law associated with the amplitude at Eq. (2.20d) we instead arrive at

$$\begin{aligned} \frac{dE}{dz} &= \frac{1}{2}\int_{-\infty}^{\infty} uf^* + fu^*\,dt, \\ &= \mathrm{Re}\int_{-\infty}^{\infty} u^* f\,dt \end{aligned} \tag{2.23}$$

as all other terms integrate out or cancel. Similarly, we can derive evolution equations for the frequency

$$\frac{d\Omega}{dz} = \frac{1}{E}\mathrm{Re}\int_{-\infty}^{\infty}(i\frac{\partial u}{\partial t} - u^*)f\,dt \tag{2.24}$$

and position

$$\frac{dT}{dz} = \Omega + \frac{1}{E}\mathrm{Re}\int_{-\infty}^{\infty}(t-T)u^* f\,dt. \tag{2.25}$$

Note that in these derivations E and Ω are functions of z and not constants. Lastly, using Eq. (2.19) we can also derive an evolution equation for the phase

$$\frac{d\Phi}{dz} = \frac{1}{2}(E^2 + \Omega^2) + \frac{1}{E}\int_{-\infty}^{\infty} \Big[\big(t - T\big)\big(2\Omega u^* - i\frac{\partial u}{\partial t}\big) - iu^*\Big] f \, dt \tag{2.26}$$

though this will not be used in the rest of the thesis.

There is a bit of ambiguity in the precise meaning of Eqs. (2.23-2.26). These equations involved no approximations and therefore capture the fully nonlinear behavior of Eq. (2.22). However, we have no guarantees that the solution u will continue to be a soliton and the parametric meanings of E, Ω, T, and Φ may quickly lose relevance. For instance, consider when $f = f_n(t)\delta(z - z_n)$, a perturbation similar to the discretely added ASE noise. We can plug this into Eqs. (2.23-2.26) and additionally integrate around z_n to get jump conditions:

$$\Delta E_n = \mathrm{Re}\int_{-\infty}^{\infty} u^* f_n \, dt, \tag{2.27a}$$

$$\Delta \Omega_n = \frac{1}{E}\mathrm{Re}\int_{-\infty}^{\infty} (i\frac{\partial u}{\partial t} - u^*) f_n \, dt, \tag{2.27b}$$

$$\Delta T_n = \frac{1}{E}\mathrm{Re}\int_{-\infty}^{\infty} (t - T) u^* f_n \, dt, \tag{2.27c}$$

$$\Delta \Phi_n = \frac{1}{E}\mathrm{Re}\int_{-\infty}^{\infty} \Big[\big(t - T\big)\big(2\Omega u^* - i\frac{\partial u}{\partial t}\big) - iu^*\Big] f_n \, dt, \tag{2.27d}$$

in the soliton parameters. These equations are very closely coupled to the soliton solution form and it is unclear if they will continue to be useful after many rounds of repeated amplification. Luckily though, the noise typically has a low noise power in physical systems and the soliton pulse retains its approximate shape under amplification with

stochastic fluctuations on top of it. Using a linear theory can help to explain what deeper changes to the soliton can occur in these circumstances.

2.3.2. Linearized theory

Let's consider the perturbation in Eq. (2.22) to be $\mathcal{O}(\epsilon)$ (i.e. write $f = \epsilon F$) where ϵ is a small number. Now, we additionally consider a perturbation v to the soliton solution that is in phase: let $u(z,t) = u_s(z,t) + \epsilon v(z,t)e^{i\Theta}$ with u_s and Θ from Eq. (2.18). This approach was taken in [**73**]. Upon substitution of this ansatz, the $\mathcal{O}(1)$ terms drop out and the $\mathcal{O}(\epsilon)$ terms are

$$\frac{dv}{dz} + \Omega\frac{\partial v}{\partial t} = \frac{i}{2}\frac{\partial^2 v}{\partial t^2} - \frac{i}{2}E^2 v + 2i|u_0|^2 v + iu_0^2 v^* \tag{2.28}$$

where, again, $u_0 = E\ \mathrm{sech}(E(t - T(z)))$. We have neglected the effect of the perturbation ϵF for the moment, but this term also shows up at this order and will be addressed later. The left-hand-side ensures that the perturbation propagates with the soliton. As Eq. (2.21) demonstrated, the frequency Ω is coupled to the position T of the pulse. Therefore, we transform into a moving coordinate system $z \equiv z + \Omega t$, and we define the right hand side as the linearized NLS operator so that we have

$$\frac{\partial v}{\partial z} = \mathcal{L}v \quad \text{where} \tag{2.29a}$$

$$\mathcal{L}v = \frac{i}{2}\frac{\partial^2 v}{\partial t^2} - \frac{i}{2}E^2 v + 2i|u_0|^2 v + iu_0^2 v^*. \tag{2.29b}$$

This operator governs the linear space of infinitesimal changes to the soliton. We then equip this space with the inner product

$$\langle f, g\rangle = \mathrm{Re}\int_{-\infty}^{\infty} f^* g\, dt \tag{2.30}$$

whose form is suggested by the conservation laws (see Eq. (2.23) for example). The linear operator is not self-adjoint in this normed space. The adjoint linearized operator is given by

$$\mathcal{L}^{\mathrm{adj}}\underline{v} = -\frac{i}{2}\frac{\partial^2 \underline{v}}{\partial t^2} + \frac{i}{2}E^2\underline{v} - 2i|u_0|^2\underline{v} + iu_0^2\underline{v}^*. \tag{2.31}$$

In particular, note that $\mathcal{L}^{\mathrm{adj}}v = i\mathcal{L}(iv)$ and $i\mathcal{L}^{\mathrm{adj}}(iv) = \mathcal{L}v$. Also, the adjoint of the left hand side of Eq. (2.28) is $\frac{d\underline{v}}{dz} - \Omega\frac{\partial \underline{v}}{\partial t}$ since the inner product is with respect to t. This motivates moving into a moving time frame $\frac{\partial}{\partial z_1} = \frac{\partial}{\partial z} + \Omega\frac{\partial}{\partial t}$ in Eq. (2.28) for the linear space and then in the adjoint equation so that

$$\frac{\partial \underline{v}}{\partial z_1} = \mathcal{L}^{\mathrm{adj}}\underline{v}$$

with z_1 running backwards (equivalent to taking the adjoint originally and then letting z run backwards). For this reason, $\mathcal{L}^{\mathrm{adj}}$ is also called the backwards linear operator.

The generalized nullspace of the linearized operator Eq. (2.29b) represents the tangent space for small changes in the soliton parameters. A biorthonormal basis for this space consists of eight modes, a forward and backwards mode associated with each soliton parameter. The functional forms of these modes can be found in Table 2.2. In this basis, there are two true modes and two generalized modes for the forward operator Eq. (2.29b)

Mode	Type	Action
$v_E = \frac{1}{E}\frac{\partial}{\partial t}((t-T)u_0)$	Generalized forward	$\mathcal{L}v_E = Ev_\Phi$
$v_\Omega = i(t-T)u_0$	Generalized forward	$\mathcal{L}v_\Omega = v_T$
$v_T = -\frac{\partial u_0}{\partial t}$	Forward	$\mathcal{L}v_T = 0$
$v_\Phi = iu_0$	Forward	$\mathcal{L}v_\Phi = 0$
$\underline{v}_E = -iv_\Phi = u_0$	Backward	$\mathcal{L}^{\text{adj}}\underline{v}_E = 0$
$\underline{v}_\Omega = iv_T/E = -\frac{i}{E}\frac{\partial u_0}{\partial t}$	Backward	$\mathcal{L}^{\text{adj}}\underline{v}_\Omega = 0$
$\underline{v}_T = -iv_\Omega/E = \frac{1}{E}(t-T)u_0$	Generalized backward	$\mathcal{L}^{\text{adj}}\underline{v}_T = \underline{v}_\Omega$
$\underline{v}_\Phi = iv_E = \frac{i}{E}\frac{\partial}{\partial t}((t-T)u_0)$	Generalized backward	$\mathcal{L}^{\text{adj}}\underline{v}_\Phi = E\underline{v}_E$

Table 2.2. The true and generalized modes of the forward and backwards (adjoint) linearized NLS operator. The modes can be found by IST theory or generated by considering invariances. The amplitude, frequency, position, and phase modes are associated with invariances with respect to scaling, Galilean transformations, time translation, and phase rotations respectively.

and its adjoint Eq. (2.31). A true forward mode for a soliton parameter corresponds to a generalized backwards mode and vice versa. To be clear, the biorthonormality works as follows

$$\langle v_k, v_m\rangle = 0 \quad \text{for } k \neq m, \tag{2.32a}$$

$$\langle \underline{v}_k, \underline{v}_m\rangle = 0 \quad \text{for } k \neq m, \tag{2.32b}$$

$$\langle v_k, \underline{v}_m\rangle = 0 \quad \text{for } k \neq m, \tag{2.32c}$$

$$\langle v_k, \underline{v}_k\rangle = 1 \quad \text{with } k, m \in \{E, \Omega, T, \Phi\}. \tag{2.32d}$$

These modes correspond to the "preferred" directions for changes in the soliton parameters, or the directions in function space that produce the largest parameters changes for fixed-size perturbations, as we will see shortly. The forward modes correspond to the directions of "clean" changes in soliton parameters or changes to the soliton that only affect a particular parameter. This can be seen from the fact that the forward modes show up in the derivatives of the soliton taken with respect to its parameters. In this procedure we leave the other parameters constant (i.e. we don't let $T = T(\Omega)$ when taking the derivative with respect to Ω). These derivatives are

$$\frac{\partial u_s}{\partial E} = v_E e^{i\Theta}, \tag{2.33a}$$

$$\frac{\partial u_s}{\partial \Omega} = v_\Omega e^{i\Theta}, \tag{2.33b}$$

$$\frac{\partial u_s}{\partial T} = (v_T - \Omega v_\Phi) e^{i\Theta}, \tag{2.33c}$$

$$\frac{\partial u_s}{\partial \Phi} = v_\Phi e^{i\Theta}. \tag{2.33d}$$

However, the backwards modes correspond with the most likely directions for additive noise to cause a particular soliton parameter to change [**73**].

The somewhat counterintuitive result that the adjoint soliton modes are associated with more-likely soliton perturbation directions comes about by considering the perturbation F as causing the constant soliton parameters to become slowly varying. Therefore, we employ a multiscale method with two time scales, z and $\zeta = \epsilon z$. Then $\frac{\partial}{\partial z} = \frac{\partial}{\partial z} + \epsilon \frac{\partial}{\partial \zeta}$ and we assume E, Ω, T, and Φ in u_s are dependent on the slow time scale ζ. Then, at

$\mathcal{O}(\epsilon)$ we get

$$\frac{\partial v}{\partial z} + \Omega \frac{\partial v}{\partial t} - \mathcal{L}v = Fe^{-i\Theta} - e^{-i\Theta}\Big(\sum_{Q\in\{E,\Omega,T,\Phi\}} \frac{\partial u_s}{\partial Q}\frac{\partial Q}{\partial \zeta}\Big). \tag{2.34}$$

The left hand side consists of the terms from Eq. (2.28) while the right hand side is the effect of the perturbation to NLS. It is clear that the right-hand-side contains secular terms to be removed by enforcing orthogonality with the linear modes (because of the Fredholm Alternative Theorem). Via these solvability conditions and the biorthogonality of the linear and adjoint linear modes, we are left with [**73**]

$$\frac{dE}{dz} = \epsilon \mathrm{Re} \int_{-\infty}^{\infty} \underline{v}_E^* e^{-i\Theta} F \, dt, \tag{2.35a}$$

$$\frac{d\Omega}{dz} = \epsilon \mathrm{Re} \int_{-\infty}^{\infty} \underline{v}_\Omega^* e^{-i\Theta} F \, dt, \tag{2.35b}$$

$$\frac{dT}{dz} = \Omega + \epsilon \mathrm{Re} \int_{-\infty}^{\infty} \underline{v}_T^* e^{-i\Theta} F \, dt, \tag{2.35c}$$

$$\frac{d\Phi}{dz} = \frac{1}{2}(E^2 + \Omega^2) + \epsilon \mathrm{Re} \int_{-\infty}^{\infty} (\underline{v}_\Phi^* + \Omega \underline{v}_T^*) e^{-i\Theta} F \, dt. \tag{2.35d}$$

If the perturbation is in the form of additive noise, then the most likely direction for a parameter change is that which maximizes the probability of the noise subject to the relevant constraint in Eq. (2.35). As the noise is zero mean, a perturbation minimizing the L^2 norm of the noise is equivalent to maximizing the probability. In this sense, the notions of the "most-likely" and "preferred" directions are linked. To see this, let the noise perturbation be a single instantiation of white noise $f_0(t)$ discretely added in z at $z = z_0$, to match the noise added at a single amplifier in Eq. (2.16), so $f(z,t) = f_0(t)\delta(z - z_0)$. Therefore, in order to find the most probable noise configuration that achieves a prescribed

or desired parameter change of size ΔY in the generic soliton parameter Y we have

$$\min_{f_0} \int_{-\infty}^{\infty} |f_0(t)|^2 \, dt \quad \text{subject to} \tag{2.36a}$$

$$\text{Re} \int_{-\infty}^{\infty} \underline{v}_Y^* e^{-i\Theta} f_0 \, dt = \Delta Y, \tag{2.36b}$$

where the side constraint's form comes from considering the solvability equations Eq. (2.35) and integrating across z_0 to get a jump condition. To be clear, the soliton mode $\underline{v}_Y$ and soliton total phase Θ are computed at z_0 the instant before the noise is added. The solution to this problem is computed in Appendix A, but takes the form

$$f_0(t) = \Delta Y \frac{\underline{v}_Y}{||\underline{v}_Y||^2} e^{i\Theta}. \tag{2.37}$$

This solution is arrived at via using standard calculus of variations methods [**40**]. It confirms that the adjoint modes are the preferred directions for noise-induced parameter perturbations.

2.3.3. Amplifier-induced changes

The pulses propagated in noisy simulations obviously do not obey the strict functional form of solitons after the addition of noise. However, they are usually approximately solitons, with random components added. Using the linearized theory of Section 2.3.2, the leading order change in a soliton after amplification can be given by decomposing the noise-induced change into the forward soliton modes (see Table 2.2). At the first amplifier the leading-order change in the soliton is

$$\Delta u_1(t) = v_E \Delta E + v_\Omega \Delta\Omega + v_T \Delta T + v_\Phi \Delta\Phi. \tag{2.38}$$

The scalars $\Delta E, \Delta\Omega, \Delta T$, and $\Delta\Phi$ are given by considering Eq. (2.35) with the discretely-added noise $f_1(t)$. They are easy to obtain from those continuous evolution equations, but explicitly these constants are

$$\Delta E = \mathrm{Re}\int_{-\infty}^{\infty} \underline{v}_E^* e^{-i\Theta} f_1(t)\, dt, \tag{2.39a}$$

$$\Delta\Omega = \mathrm{Re}\int_{-\infty}^{\infty} \underline{v}_\Omega^* e^{-i\Theta} f_1(t)\, dt, \tag{2.39b}$$

$$\Delta T = \mathrm{Re}\int_{-\infty}^{\infty} \underline{v}_T^* e^{-i\Theta} f_1(t)\, dt, \tag{2.39c}$$

$$\Delta\Phi = \mathrm{Re}\int_{-\infty}^{\infty} (\underline{v}_\Phi^* + \Omega\underline{v}_T^*) e^{-i\Theta} f_1(t)\, dt. \tag{2.39d}$$

To clarify this notation, if we have a clean soliton u_s at z_{1-} (right before amplification), then right after amplification (at z_{1+}) we have

$$\begin{aligned} u(z_{1+},t) &= u_s(z_{1-},t;E,\Omega,T,\Phi) + \Delta u_1(t) e^{i\Theta(t,z_{1-})} + R(t,z_{1+}), \\ &= u_s(z_{1+},t;E+\Delta E,\Omega+\Delta\Omega,T+\Delta T,\Phi+\Delta\Phi) + R(t,z_{1+}) \end{aligned} \tag{2.40}$$

where R is dispersive radiation also injected by the amplifier that does not change the soliton but instead propagates on top of it. This dispersive radiation grows under repeated amplification, but the portion injected at each amplifier is typically smaller than the cumulative soliton changes.

2.4. Numerical Methods and Considerations

There are a variety of numerical methods that can be used to solve NLS and similar lightwave systems. In this section we will discuss the main algorithm used to numerically

simulate the partial differential equations in this thesis: the split-step Fourier method. The numerical properties of the split-step method will also be discussed. Then, we will talk about simulating the properties and effects of ASE noise on a pulse as it propagates. We will discuss how to numerically deal with the amplifiers and also how to track soliton parameters under noisy propagation.

2.4.1. Split-step Fourier method

The split-step Fourier method, the principal numerical method of this thesis, is based on *operator splitting* [**8**]. This concept comes from considering a partial differential equation of the form

$$\frac{\partial u}{\partial z} = g_1(u, \frac{\partial u}{\partial t}, \frac{\partial^2 u}{\partial t^2}, \ldots) + g_2(u, \frac{\partial u}{\partial t}, \frac{\partial^2 u}{\partial t^2}, \ldots) \tag{2.41}$$

where g_1 and g_2 are two general functions of u and its temporal derivatives. Operator splitting is applied to this equation by considering two simpler equations

$$\frac{\partial u}{\partial z} = g_1(u, \frac{\partial u}{\partial t}, \frac{\partial^2 u}{\partial t^2}, \ldots) \quad \text{and} \tag{2.42a}$$

$$\frac{\partial u}{\partial z} = g_2(u, \frac{\partial u}{\partial t}, \frac{\partial^2 u}{\partial t^2}, \ldots). \tag{2.42b}$$

The principal benefit of operator splitting is that Eqs. (2.42a) and (2.42b) can be solved by different methods. As we will see shortly, NLS admits to different treatments of its dispersive and nonlinear terms. Once solution methods are chosen for each equation in Eq. (2.42), the overall solution method is to evolve in z while taking alternating steps

with the solvers for g_1 and g_2. However, operator splitting also comes with an associated cost of an error that is $\mathcal{O}(\Delta z^2)$ per step.

This error can be seen by replacing the functions Eq. (2.41) with differential operators G_1 and G_2 so that

$$\frac{\partial u}{\partial z} = (G_1 + G_2)u. \tag{2.43}$$

With this notation, the formally exact solution to Eq. (2.41) is

$$u(z,t) = \exp((G_1 + G_2)z)u_0(t) \tag{2.44}$$

where $u_0(t)$ is the initial condition. However, the solution implemented with operator splitting is

$$u(z,t) = \exp(G_1 z)\exp(G_2 z)u_0(t) \tag{2.45}$$

which is exact if G_1 and G_2 commute, but this is not true in general. Taylor expanding the exact and implemented solutions in z gives that the leading order error term is in fact proportional to the commutator of G_1 and G_2 or $[G_1, G_2] = G_1G_2 - G_2G_1$ which shows up at second-order. That is, for small Δz we have

(2.46)

$$\begin{aligned} e^{(G_1+G_2)\Delta z} - e^{G_1\Delta z}e^{G_2\Delta z} &= \frac{1}{2}\Big[(G_1+G_2)^2\Delta z^2 - (G_1^2 + G_2^2 + 2G_1G_2)\Delta z^2\Big] + H.O.T., \\ &= -\frac{1}{2}\big(G_1G_2 - G_2G_1\big)\Delta z^2 + H.O.T. \end{aligned}$$

which demonstrates that split-step methods are only first-order accurate [**5**].

Luckily, this accuracy can be improved by an order by considering a slight modification to the sequence in which the steps are taken in the implementation of operator splitting. This modification, called symmetrized operator splitting, is to alternate the order of the steps taken, i.e. $\exp(G_1\Delta z)\exp(G_2\Delta z)\exp(G_2\Delta z)\exp(G_1\Delta z)\ldots\exp(G_1\Delta z)\exp(G_2\Delta z)$. Practically, this is implemented by taking single steps of the form

$$u(z_0+\Delta z,t)=e^{G_1\frac{\Delta z}{2}}e^{G_2\Delta z}e^{G_1\frac{\Delta z}{2}}u_0(t). \tag{2.47}$$

When Taylor expanded, this gives an error of

$$\begin{aligned} e^{(G_1+G_2)\Delta z}-e^{G_1\frac{\Delta z}{2}}e^{G_2\Delta z}e^{G_1\frac{\Delta z}{2}} = \frac{1}{24}\big(&G_1^2G_2-2G_1G_2G_1-2G_1G_2^2+\\ &G_2G_1^2+4G_2G_1G_2-2G_2^2G_1\big)\Delta z^3+H.O.T. \end{aligned} \tag{2.48}$$

Therefore, symmetrized operator splitting is second-order accurate.

To apply operator splitting to NLS, first note that the NLS operator can be written

$$\begin{aligned} \frac{\partial u}{\partial z}&=(\tilde{D}+\tilde{N})u \quad \text{where}\\ \tilde{D}u=i\frac{\partial^2 u}{\partial t^2} \quad &\text{and} \quad \tilde{N}u=i|u|^2u. \end{aligned} \tag{2.49}$$

This motivates treating the dispersion and nonlinearity of NLS separately like

$$\frac{\partial u}{\partial z}=\tilde{D}u, \quad \text{and} \tag{2.50a}$$

$$\frac{\partial u}{\partial z}=\tilde{N}u. \tag{2.50b}$$

Note that this formalism allows for treating a more general equation than just the NLS equation. For example, if Eq. (2.22) involves the NLS and some small continuous perturbations, these can be incorporated into the numerical split-step scheme by adding any linear terms to $\tilde{D}$ and all other terms to $\tilde{N}$. Luckily, the perturbations considered in this thesis only include temporal derivatives of u in linear terms, otherwise the scheme presented here becomes more complicated and it may be advantageous to use a scheme other than operator splitting.

We solve the dispersive problem, Eq. (2.50a), using a pseudospectral method. Transforming it into Fourier space gives an ordinary differential equation (ODE) for each Fourier component

$$\frac{d\hat{u}(z,w)}{dz}e^{iwt} = (iw)^2\hat{u}(z,w)\,e^{iwt}. \tag{2.51}$$

That is, all temporal derivatives become scalar multiples in Fourier space, so the only work left to be done is to solve a simple ODE. Therefore, the dispersive solve is formally

$$\begin{aligned} u(z+\Delta z,t) &= \mathcal{F}_w^{-1}\Big[\int_z^{z+\Delta z}\exp(\mathcal{F}_w[\tilde{D}]z')\,dz'\,\mathcal{F}_w[u(z,t)]\Big], \\ &= \mathcal{F}_w^{-1}\Big[\exp(\mathcal{F}_w[\tilde{D}]\Delta z)\,\mathcal{F}_w[u(z,t)]\Big]. \end{aligned} \tag{2.52}$$

This takes advantage of the fact that generally $\tilde{D}$ is a linear differential operator so the only work being done is the Fourier transforms. Numerically, this involves taking the Fast Fourier Transform (FFT) with N Fourier modes, which is known to be fast and accurate.

The nonlinear problem can be treated in a number of ways. Some methods rely on the fact that the pulse energy is constant under propagation as it is a conserved quantity.

That is,

$$|u|^2 = |u_0|^2 = E$$

which was discussed in Section 2.3. Therefore, one can use an integrating factor scheme via

$$u(z+\Delta z,t) = \exp\big(\int_z^{z+\Delta z} \tilde{N}\, dz'\big) u(z,t), \tag{2.53}$$
$$= \exp(\tilde{N}\Delta z) u(z,t)$$

where $\tilde{N}$ is evaluated at z and is then treated as a constant in the integration. However, in the presence of other terms (arising via perturbations in Eq. (2.22)), the nonlinearity operator is not a constant multiplier and this integrating factor scheme can be improved upon. One easy improvement is to use an explicit trapezoidal method, which preserves the second-order accuracy of the scheme [**5**]. This method approximates the integral in Eq. (2.53) via

$$\int_z^{z+\Delta z} \tilde{N}(z')\, dz' \approx \frac{\Delta z}{2}\big[\tilde{N}(z) + \tilde{N}(z+\Delta z)\big] \tag{2.54}$$

and involves taking an initial predictor step to ascertain $\tilde{N}(z+\Delta z)$. We take a step of forward Euler as the predictor so in $\tilde{N}(z+\Delta z)$ any dependence on u is given by $\tilde{u} = u(z) + \Delta z \tilde{N}(z)$. This method is alternatively known as improved Euler or Heun's method and is well known in many contexts [**84**].

There are many other spectral methods for solving NLS, many of which have faster convergence rates. However, the split-step method is sufficient for our purposes, as the main simulations in this thesis will be Monte Carlo simulations where NLS is solved

many times. The primary advantage of step splitting is its speed, and its second-order convergence is satisfactory in this context. The symmetrized split step Fourier method is implemented by taking half steps of the dispersive ($\tilde{D}$ stands in for G_1 in Equation 2.47). The convergence rates of three variants of the split-step method discussed here, along with two other methods, are shown in Fig. 2.1. The other two methods included are the Crank-Nicholson Adams-Bashforth second-order scheme, an example of a semi-implicit spectral method, and exponential time differencing with a fourth order Runge Kutta stepper which uses an integrating factor spectral method [**78**]. Note that the most accurate split-step scheme is the symmetrized scheme that treats the nonlinearity as a constant, which is the exact solution of the nonlinearity in the NLS without perturbations. We should not expect it to be more accurate than the trapezoidal method for more general systems.

2.4.2. Implementing amplifier noise

A few practical considerations for implementing the ASE noise model from Eq. (2.17) should be considered. In a discretized version of the noise, its values at each point in the time domain are a complex random variable drawn from a zero-mean Gaussian distribution, that is, a Gaussian in the real and imaginary quadratures. The noise power σ^2 in Eq. (2.17b) is matched in a discrete approximation by drawing each Gaussian random variable and then scaling the variance with

$$\sigma_a^2 = \frac{\sigma^2}{2\,dt} \tag{2.55}$$

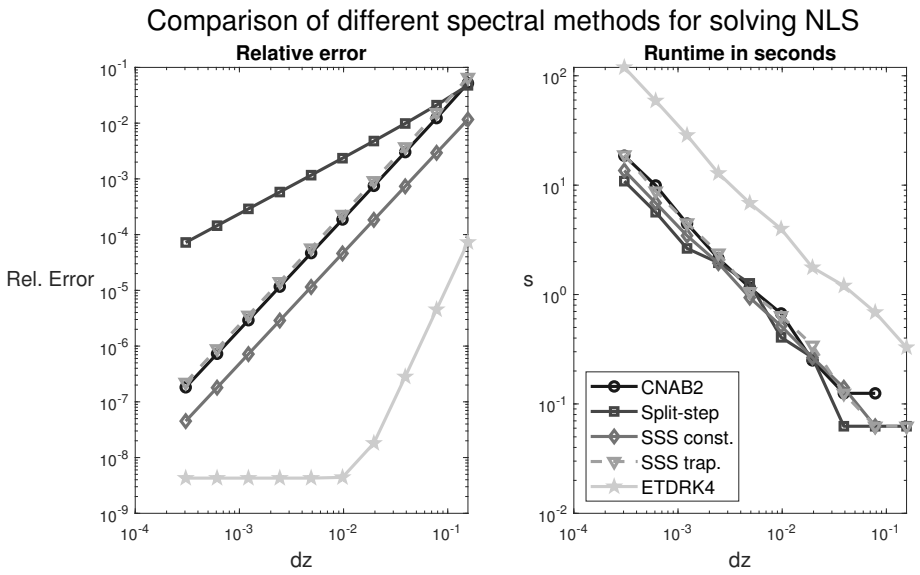

Figure 2.1. A comparative study of various numerical methods for solving unperturbed NLS. Five spectral methods were used with 256 Fourier modes to simulate a soliton with initial parameters $E = 1, \Omega = 0, T_0 = 0$, and $\phi = 0$ in a computational window 40 units wide. The relative L^2 error between the numerical solution and the exact solution at $z = 8$ was reported for different step sizes dz in the left panel, with the computation time shown in the right panel. The split-step method is first order accurate. The symmetrized split-step schemes treating the nonlinearity as a constant (SSS const.) and using the trapezoidal rule (SSS trap.) along with a semi-implicit scheme called Crank-Nicholson Adams-Bashforth second-order (CNAB2) are all second-order accurate. Lastly, exponential time differencing with a fourth order Runge Kutta stepper (ETDRK4) is a fourth-order accurate scheme (though it is significantly more computationally intensive). Note that ETDRK4 hits the "error floor," or the point where the numerical error is dominated by the spectral approximation, at $dz \approx 0.01$ so decreasing the step size past that point without using more Fourier modes provides no benefit. ETDRK4 is roughly an order of magnitude slower to execute than all the other methods for each step size.

where dt is the spatial step size in the discretization [**73**]. Standard Gaussian random number generators (such as `randn` in MATLAB) are sufficient to generate a zero mean Gaussian R.V. and the variance scaling is accomplished by multiplying by σ_a.

Another consideration is the noise bandwidth. The noise bandwidth in Eq. (2.17) is infinite, but this obviously cannot be replicated by a numerical discretization. Luckily, this noise model is itself an approximation of the ASE noise produced by an EDFA, which has a very large but finite bandwidth. This bandwidth is often on the order of 2500GHz which is much larger than typical pulse bandwidths of 50GHz [**5**]. Therefore, we should chiefly be concerned that the finite bandwidth of the numerical noise is much larger than the pulse used in simulation, as is the case in physical systems. The numerical noise bandwidth is proportional to the number of Fourier modes used, while the pulse bandwidth is fixed, so this condition can always be met if a fine-enough discretization is used. In practice, a minimum of 64 or 128 Fourier modes are sufficient to meet this requirement in most circumstances.

Now, we wish to track how the build up of noise from repeated amplification affects pulses in simulation. Therefore, we should not choose a numerical method with dissipative error as this would artificially dampen the noise and, potentially, lead to spurious conclusions about noise effects. Luckily, the split-step spectral method's error is not dissipative, so it is suitable to use in noisy simulation. Additionally, we should be concerned that our method does not succumb to aliasing errors, or the spurious transfer of energy from high wavenumber modes to low wavenumber modes caused by the Fourier discretization. As the white noise is infinite in noise bandwidth, such an error would be particularly pernicious, and could even lead to instability in the simulation, such as through spectral blocking [**78**]. Again, the split-step method is resilient to aliasing as it does not treat the cubic nonlinearity in the NLS equation with a Fourier method. An example of numerical pulse propagation with amplifier noise is provided in Fig. 2.2. In this figure, the back-

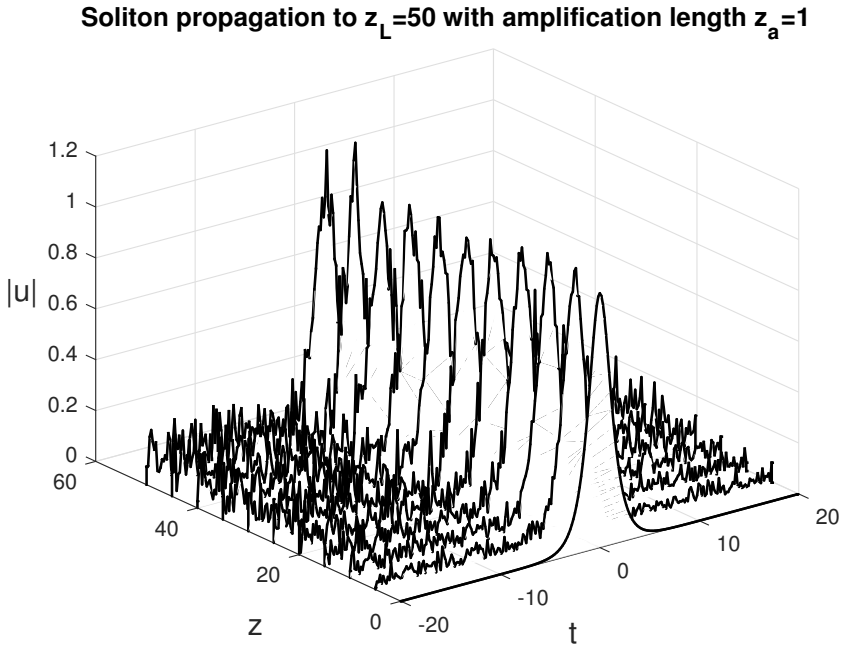

Figure 2.2. A soliton with $E_0 = 1, \Omega_0 = T_0 = \Phi_0 = 0$ was propagated to $z = 50$ with $N = 256$ Fourier modes and a step size of $dz = 0.05$ using the symmetrized split-step Fourier method and the trapezoidal rule. The solution's modulus is shown every 5 units and the build up of the noise, which has noise power $\sigma^2 = 5 \times 10^{-5}$, is striking. The underlying soliton is still easy to discern, though.

ground noise builds, which is indicative of the fact that the numerical method does not dissipate it away.

2.4.3. Tracking soliton parameters in noisy simulation

Assessing the state of the soliton, and how its parameters are affected by the ASE noise requires separating the soliton from the noise: a process called soliton extraction. Extracting the soliton from dispersive radiation from the additive noise is not a trivial task. There has been extensive study on discriminating soliton pulses from additive noise, and

many have opted for pre-processing the pulse by passing it through a series of filters (both in and outside of frequency space). Two representative methods are found in [**67, 73**]. In all of these methods the moments (the first three rows of Table 2.1 and Equation 2.19) are used to compute the parameters which then determine the soliton u_s. Another method involves utilizing the inverse scattering transform [**93**]. This method is very accurate, but is quite time consuming.

The method used here is to use a generic curve-fitting algorithm included in MATLAB called `lsqnonlin`. This nonlinear optimization routine relies on a trust region approach to minimize the user-supplied nonlinear objective function (for us the soliton parametrization from Equation 2.18) on the given noisy data in a least squares sense [**17**]. Implementation of it into our soliton extraction scenario requires giving the objective function's parametrization as well as initial guesses for each parameter (the parameter values at the previous amplifier). In addition to this approach's ease of use, it was found to be effective and efficient in implementation. Figure 2.3 shows the results of this process when applied to the soliton propagation from Fig. 2.2. The curve-fitting routine was applied at each amplifier in order to track the evolution of the parameters with noise. Only the parameters were saved, but their values completely determine the soliton. The random walk in each parameter is shown in the right panel and the comparison of the fit with the noisy pulse is shown in the left panel, along with another fit obtained using the moments method, which is obviously less accurate.

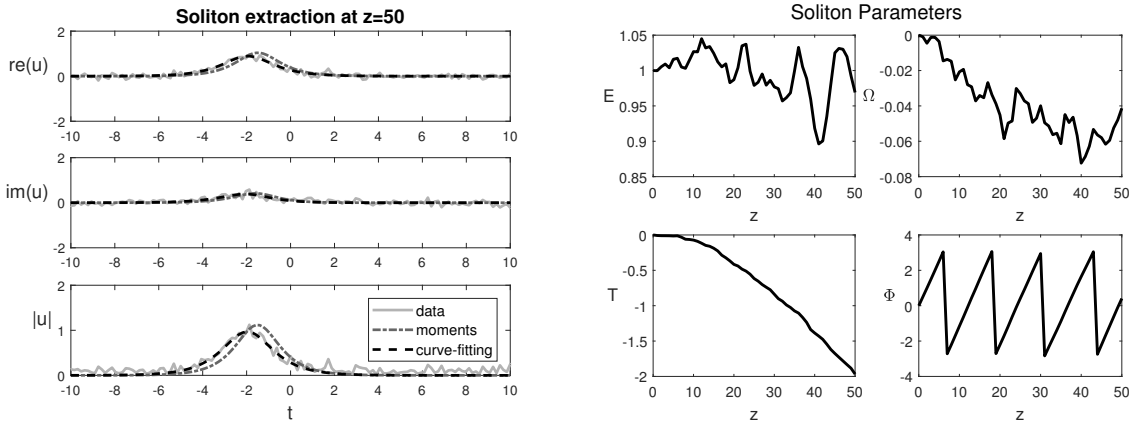

Figure 2.3. This figure shows the extraction process when applied to the soliton propagation from Fig. 2.2. The curve-fitting routine was applied at each amplifier in order to track the evolution of the parameters; each parameter's random walk is shown in the right panel. The comparison of the overall soliton fit with the noisy pulse is shown in the left panel. Another soliton extraction performed by simply using the moments (the first three rows of Table 2.1 and Equation 2.19) is shown for comparison. The curve-fitting does a better job than the moments of capturing the leftward drift in the soliton's position.

CHAPTER 3

Importance Sampling Methods

The goal of this thesis is to describe rare events in lightwave systems. Therefore, we will discuss the numerical techniques used to approximate the probabilities of such events occurring. The main technique is Monte Carlo (MC) sampling, a ubiquitous technique for simulating random events. As the events of interest are rare, importance sampling (ISMC or IS for brevity) is used to achieve accurate estimations of small probabilities with remarkable efficiency. Modifications of importance sampling are then discussed: among them multiple importance sampling (MIS) and dynamic importance sampling (DIS). This sets up Chapter 4, where the intersection of these two types of importance sampling are discussed at length, by means of a toy random walk problem where MIS and DIS must be used in tandem. This theory also presages the work of Chapter 5 where these techniques are all applied to lightwave system where mode-locking mechanisms in a laser model complicate the use of IS. Therefore, we end this chapter with a discussion of some of the basic aspects of performing importance-sampled Monte Carlo simulations in a soliton-based lightwave system.

3.1. Monte Carlo Integration

In its simplest form, Monte Carlo integration allows one to numerically approximate an integral by drawing a large number of random samples from a probability distribution. For our purposes, consider an integral that is the expectation of some real-valued function

f depending on a random variable (or vector of random variables)

$$I = E[f(\mathbf{X})]$$

where X is a vector of n real random variables with each component independently and identically distributed (i.i.d.) by the probability density function (P.D.F.) $p(\mathbf{x})$ so that

$$I = E[f(\mathbf{X})] = \int_{\mathbb{R}^n} f(\mathbf{x})p(\mathbf{x})\, d\mathbf{x}. \tag{3.1}$$

Monte Carlo integration works to approximate I by randomly sampling from p. For instance, after drawing N samples of n-dimensional real random variables distributed according to $p(\mathbf{x})$, our approximation to I is

$$\hat{I}_N = \frac{1}{N}\sum_{l=1}^{N} f(\mathbf{X}_l). \tag{3.2}$$

Now, the law of large numbers says that $\hat{I}_N$ approaches the true value of I as $N \to \infty$. However, the convergence rate is problematically slow. This is due to the fact that the variance of the estimator is

$$\text{var}[\hat{I}_N] = \frac{1}{N}\text{var}[f] \tag{3.3}$$

since the random samples are independent. The standard deviation of $\hat{I}_N$ is then $\mathcal{O}(N^{-1/2})$.

The slow decay in the standard deviation of the estimate makes it extremely inefficient to achieve small errors with MC. Chebyshev's inequality [**86**] allows us to get a bound on

the likelihood of an error of particular size occurring in our estimate

$$\Pr\left[(\hat{I} - I)^2 \geq \frac{\text{var}[f]}{N\delta}\right] \leq \delta. \tag{3.4}$$

This, in turn, tells us that if we want to have 99% confidence (i.e. $\delta = 0.01$) that an estimate has an error of less than 10^{-5} we have to draw at least

$$N \geq \frac{\text{var}[f]}{(10^{-5})^2(0.01)} = \text{var}[f] \times 10^{12}.$$

This means that we require an infeasible number of samples in order to accurately estimate small quantities (which require even smaller errors). An even more pernicious problem with using MC for estimating small probabilities comes from the fact that typically $\text{var}[f] \sim I$. This implies, using Eq. (3.3), that the convergence rate of the MC estimator for small probabilities is $\mathcal{O}\big((NI)^{-1/2}\big)$. Therefore, one must pick $N \gg 1/I$ which is infeasible when I is small as is typically the case in this thesis. Estimating small quantities is of interest to us because we seek to ascertain error rates in lightwave systems, which are typically small.

Additionally, we will typically consider the function $f(x)$ from Eq. (3.1) to be an indicator function on another function g: it takes the value 1 if $g(x)$ is in a particular range and is 0 otherwise. Therefore, I gives the probability of g being in the range and $\hat{I}$ estimates this probability. Partitioning a relevant sample space for g up into small ranges (called bins) then allows us to discretely estimate the probability density function over this space. It is from this perspective that we can then see that MC simulations can do a good job estimating P.D.F.'s in bulk but have a hard time converging in their tails (where probabilities are low).

3.2. Importance Sampling

The slow convergence in Monte Carlo simulations for small probabilities can be caused by a mismatch between the function f and the distribution p in Eq. (3.1). Obviously, the integral I will be dominated by regions where fp is approximately maximal, and in many cases p is not maximal in the same region. In other words, the main contributions to the integral come from the tails of the probability distribution p. When this happens, the "important" regions for I are undersampled and the estimate $\hat{I}$ converges very slowly. An augmentation to MC simulation called importance sampling (IS) corrects for this by introducing an artificial probability distribution, called a biasing distribution, into the simulation from which samples are drawn [**14**]. The goal of this technique is to replace p with a distribution that heavily weights regions of probability space where events of interest occur most often.

The insight which makes IS possible is that Eq. (3.1) can be rewritten as

$$I = \int_{\mathbb{R}^n} f(\mathbf{x}) \frac{p(\mathbf{x})}{p^*(\mathbf{x})} p^*(\mathbf{x})\, d\mathbf{x}. \tag{3.5}$$

The function p^* is the biasing distribution and will be treated as given for this discussion; the choice of biasing distribution is an application specific issue and investigating rationales for choosing biasing distributions is a rich field of inquiry in many contexts. With a given biasing distribution we perform MC simulations by drawing samples from p^* and then computing

$$\hat{I}_N = \frac{1}{N} \sum_{l=1}^{N} f(\mathbf{X}_l) \frac{p(\mathbf{X}_l)}{p^*(\mathbf{X}_l)}. \tag{3.6}$$

In an ISMC simulation, the sample results are now weighted by the likelihood ratio, or

$$L(\mathbf{X}_l) = \frac{p(\mathbf{X}_l)}{p^*(\mathbf{X}_l)}, \tag{3.7}$$

which corrects for the fact that samples were drawn by the biasing distribution and gives results equivalent to having used the original distribution in MC simulation.

The success of importance sampling is principally judged by two factors: that the resulting integral or probability computed is accurate, and that the convergence is suitably fast. The second criterion can be addressed by examining the variance of the IS integrand

$$\mathrm{var}^*\Big[f(\mathbf{X})\frac{p(\mathbf{X})}{p^*(\mathbf{X})}\Big] = \int_{\mathbb{R}^n} \big(f(\mathbf{x})\frac{p(\mathbf{x})}{p^*(\mathbf{x})} - I\big)^2 p^*(\mathbf{x})\, d\mathbf{x} \tag{3.8}$$

where it can be seen that the variance will be zero if p^* is chosen so that $p^*(\mathbf{x}) = (f(\mathbf{x})p(\mathbf{x}))/I$. However, this result assumes that the quantity of interest, I, is known prior to simulation. Additionally, note that Eq. (3.8) can be written

$$\mathrm{var}^*\Big[f(\mathbf{X})\frac{p(\mathbf{X})}{p^*(\mathbf{X})}\Big] = \int_{\mathbb{R}^n} f^2(\mathbf{x})\frac{p(\mathbf{x})}{p^*(\mathbf{x})}p(\mathbf{x})\, d\mathbf{x} - I^2$$

which allows for comparing the variances of the original integral and the importance-sampled version. Explicitly, this comparison is

$$\mathrm{var}[f(\mathbf{X})] - \mathrm{var}^*\Big[f(\mathbf{X})\frac{p(\mathbf{X})}{p^*(\mathbf{X})}\Big] = \int_{\mathbb{R}^n} f^2(\mathbf{x})\Big(1 - \frac{p(\mathbf{x})}{p^*(\mathbf{x})}\Big)p(\mathbf{x})\, d\mathbf{x} \tag{3.9}$$

which shows that the variance is reduced by choosing $p^*(\mathbf{x}) > p(\mathbf{x})$ when $f^2(\mathbf{x})p(\mathbf{x})$ is small and $p^*(\mathbf{x}) < p(\mathbf{x})$ when $f^2(\mathbf{x})p(\mathbf{x})$ is large. This too is not useful as a selection method as it requires knowledge of $f(\mathbf{x})$ that is not available in many cases, such as when

$f(\mathbf{x})$ can only be evaluated implicitly or numerically as is the case in tracking soliton parameter jitter. Additionally, potential biasing distributions may not be able to satisfy both inequalities simultaneously. Lastly, in some instances many biasing distributions may satisfy these criteria, so more work must be done to determine which is the best choice. In practice the choice of biasing distribution is guided by the principle that regions of state space that lead to events of interest should be sampled as much as possible. The inexact nature of importance sampling has been recognized since the technique was first introduced [**56**]. Luckily there are properties of the systems and noise of interest in this thesis that help to guide the choice of biasing distribution.

3.2.1. Parametric biasing and a Gaussian example

In this thesis, the probability distributions inherent in the rare-event problems allow parametric biasing, where biasing distributions are chosen by adjusting parameters in the original distribution. In all cases considered here these distributions are Gaussian and so their P.D.F.s take the form

$$p_G(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{(x-\mu)^2}{2\sigma^2}\Big). \tag{3.10}$$

In one dimension there are two inherent parameters: μ for the mean and σ for the standard deviation. There are then two natural options for parametric biasing: adjusting the mean or changing the standard deviation. These methods are also known as "mean-shifting" and "variance scaling" respectively. Mean-shifting is nearly always preferable to variance scaling which can be seen by considering a simple example from [**59**].

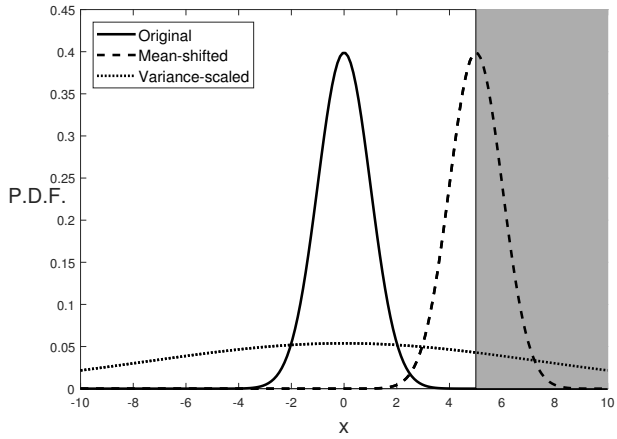

Figure 3.1. The standard Gaussian distribution and characteristic mean-shifted and variance-scaled biasing distributions are shown. The mean-shifted biasing distribution covers the region of interest ($x \geq 5$) more efficiently than the variance-scaled one.

Consider the case of attempting to compute $P = \Pr[X] \geq 5$ with X drawn from the Gaussian distribution. The exact solution to this problem is

$$P = \int_5^\infty p_G(x; 0, 1)\, dx = \frac{1}{2}\text{erfc}(5/\sqrt{2}) \tag{3.11}$$

where erfc is the complementary error function. However, if we were trying to compute P using MC simulations, we'd be limited by the fact that the unbiased Gaussian distribution samples this region very rarely. Parametrically biasing the Gaussian distribution allows us to shift probability mass from the region around 0 to this region for ISMC simulation. Mean shifting translates the distribution whereas variance scaling broadens the distribution, as seen in Fig. 3.1. From this we can develop a graphical intuition that mean-shifting is preferable to variance-scaling as mean-shifting allows for probability mass to be shifted into the region of interest, whereas variance scaling shifts mass into the region but also shifts mass away from the region (i.e. on the left side of the distribution).

Furthermore, we can compare the efficacy of mean-shifting and variance-scaling for this simple problem by checking the variance of the associated estimators. For mean-shifting with $\sigma = 1$ but μ arbitrary, the variance is

$$\text{var}^*\Big[f(X)\frac{p(X)}{p^*(X)}\Big] = \frac{1}{2}e^{\mu^2}\text{erfc}\big((5+\mu)/\sqrt{2}\big) - P^2 \tag{3.12}$$

with P from Eq. (3.11). This variance can then be examined for the optimal mean shift in the following way: replace the complementary error function (erfc) with the first term in its asymptotic expansion for a large argument

$$\text{erfc}(x) \sim \frac{1}{\sqrt{\pi}\,x}e^{-x^2},$$

take a logarithmic derivative of the resulting exponential factor (which dominates the expression), and neglect a few small terms while equating the rest with zero. The resulting optimal mean shift is then $\mu = 5$. With this mean shift, we can then see that the variance is $\mathcal{O}(P^2)$ which means that the convergence in the MC simulations will be good.

By contrast, consider variance-scaling by leaving $\mu = 0$ and leaving σ arbitrary. In this case the variance is

$$\text{var}^*\Big[f(X)\frac{p(X)}{p^*(X)}\Big] = \frac{\sigma^2}{2\sqrt{2\sigma^2-1}}\text{erfc}\Big(5\sqrt{1-\frac{1}{2\sigma^2}}\Big) - P^2. \tag{3.13}$$

The variance is dominated by the prefactor, so minimizing it by taking logarithmic derivatives and equating with zero results in $\sigma = 1$. As the optimal variance scaling is the same as the unbiased case, variance scaling is of no benefit. Further, the variance is $\mathcal{O}(P)$ so the convergence is quite slow. This is an illustration of what is known as the "dimensionality

problem" of variance scaling [**59**]. Though this is just a one dimensional illustration, the problem can become much worse in higher dimensional sample spaces where regions of interest can be very localized; mean-shifting is the preferable option in such circumstances.

3.2.2. Another example: a 1D Gaussian random walk

Consider a random variable given by a sum of N i.i.d. Gaussian random variables with mean zero and variance one. Let's choose to interpret this as a random walk with a Gaussian distributed step. Therefore,

$$Z_N = \sum_{j=1}^{N} X_j \tag{3.14}$$

gives the position of the walker on the real line after N steps. The probability distribution for Z_N is known *a priori* as it is simply a Gaussian with mean zero and variance N. However, recreating this P.D.F. numerically using ISMC is useful as it demonstrates a few things that will be of use later in the thesis: computing P.D.F.'s via binning, computing likelihood ratios of processes, and optimizing biasing distributions.

Let $\vec{X} = \{X_1, \ldots, X_N\}^T$ be the series of steps that produce Z_N. Then, analytically, we can compute the probability of Z_N landing in a particular range using an indicator function $\mathbf{I}$ that takes the values 0 or 1 based on logical criteria such as $a \leq Z_N$. Then the probability density in this range can be trivially computed using

$$P(a \leq Z_N) = \int_{-\infty}^{\infty} \mathbf{I}(z)p(z)\,dz = \int_{\mathbb{R}^N} \mathbf{I}(\sum_{n=1}^{N} x_n)p(\vec{x})\,d\vec{x}$$

where $\vec{X}$ is distributed by the multivariate Gaussian distribution

$$p(\vec{x};\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\Big(-\frac{||\vec{x}-\mu||^2}{2\sigma^2}\Big) \tag{3.15}$$

for i.i.d. Gaussians, so it is parametrized by a single value each for mean and variance.

The important regions for $\mathbf{I}$ in N-dimensional space may be hard to capture analytically, so we can approach this problem using ISMC. This means we choose instead a biasing distribution from which to draw a biased noise sample $\vec{X}^b$. We bias parametrically via mean-shifting, but since there are multiple steps it is unclear if we need to pick different shifts for each step. To pick the optimal series of mean-shifts, we seek to maximize the probability of a noise sample $\vec{X}$ meeting our criterion $\sum_{j=1}^{N} X_j = a$. Equivalent to maximizing the probability of a vector of zero-mean Gaussians is minimizing its L_2 norm (squared for smoothness). Therefore, we seek

$$\min_{\vec{X}\in\mathbb{R}^N} ||\vec{X}||^2 \tag{3.16a}$$

$$\text{subject to} \quad \sum_{j=1}^{N} X_j = a. \tag{3.16b}$$

This is a simple Lagrange multiplier problem with the solution $X_j = a$ for all j which gives the optimal mean-shift. Therefore, we pick the same mean-shift for all steps and are ready to begin sampling.

In order to assemble the samples into the ensemble, we need to compute their likelihood ratios. Luckily, since the steps are i.i.d., the probability of a particular sample $\vec{X}^b$ is simply the product of the probability of each step X_j^b. Therefore, the likelihood ratio can

be written as the likelihood ratio of each step

$$L(\vec{X}^b) = \frac{p(\vec{X}^b)}{p^*(\vec{X}^b)} = \prod_{j=1}^{N} \frac{p(X_j^b)}{p^*(X_j^b)} \tag{3.17}$$

which is a general result for the random walks under consideration in this thesis. In the case of this example, the probability distributions for a single step are $p(x) = p_G(x; 0, 1)$ and $p^*(x) = p_G(x; a/N, 1)$ respectively. Results for $N = 10$ comparing regular MC simulations and ISMC simulations are shown in Fig. 3.2. Fifty thousand samples were used for each simulation so the Monte Carlo simulation is only able to capture the P.D.F. of Z_{10} down to probabilities of roughly 10^{-4} or $-10 < Z < 10$. The ISMC simulation, which targets $Z_{10} = 10$ explicitly, is able to accurately capture probabilities down below 10^{-10} in the right tail of the distribution.

The accuracy and convergence of Monte Carlo simulations can be assessed using the coefficient of variation. The coefficient of variation (C.V.) is defined as the sample standard deviation divided by the sample mean of the simulation, or

$$C.V. = \frac{\hat{\sigma}}{\hat{I}}. \tag{3.18}$$

In the context of simulating a P.D.F., the C.V. is defined as the intra-bin standard deviation divided by the mean-simulated P.D.F.. It produces a curve rather than a single value, and can help to give a metric for the convergence of simulation throughout the state space of interest that is independent of the magnitude of probabilities throughout the region. As a general rule of thumb, a low C.V. shows good convergence and a high C.V. shows poor convergence. An example of a C.V. is shown in the lower panel of Fig.

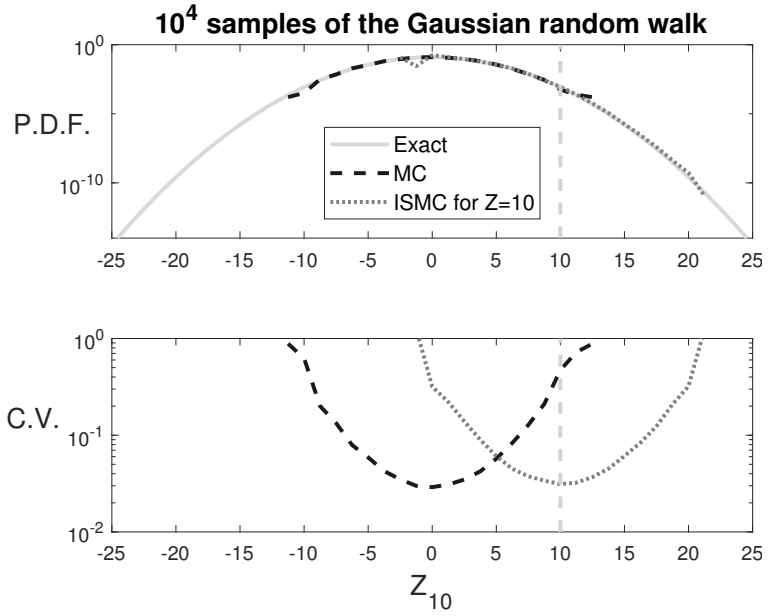

Figure 3.2. Exact and numerical approximations of the P.D.F. of the Gaussian random walk from Eq. (3.14) with $N = 10$ steps and 10,000 samples taken for each numerical approximation. The top panel shows three P.D.F.'s: the exact solution ($p_G(x; 0, N)$), one obtained by MC simulation, and one obtained using ISMC simulation targeting the right tail of the distribution. The bottom panel shows the coefficient of variation (C.V.) of the two MC simulations, which visually demonstrates their convergence properties. The vertical gray dashed lines indicate that the ISMC simulation targeted $Z = 10$.

3.2. The C.V. for the regular MC simulation (thick-dashed) shows that the convergence is best in the bulk of the distribution, and gets progressively worse as we get down into the tails. This makes sense: the vast majority of samples fall around $Z = 0$ and relatively few fall in the $|Z| > 10$ range. However, for the ISMC simulation (dotted) the C.V. is best for $Z = 10$ and gets progressively worse as the distance increases from that point. This is because the simulation was biasing toward $Z = 10$ so the majority of samples ended up near there, with a progressive decline in the amount of samples realized in regions increasingly far from that point.

3.3. Multiple Importance Sampling

Many variants of importance sampling have been developed that use multiple biasing distributions, for a variety of purposes and types of simulations [**15, 33, 74, 75**]. Here, we will discuss multiple importance sampling (MISMC or MIS), which uses multiple biasing distributions in the course of a single batch simulation. The use of multiple biasings can serve a variety of purposes, but for now we will discuss its application to simulating large regions of state space simultaneously. Figure 3.2 demonstrates that importance sampling can easily capture the tail of a P.D.F. by shifting samples so that they fall in that region. However, it has trouble capturing the P.D.F. over a wide range. By introducing multiple biasing distributions that shift samples to various regions of state space, an entire P.D.F. can be captured by a single simulation. This sort of simulation is shown in Fig. 3.3 where multiply-importance-sampled Monte Carlo simulations (MISMC) has been performed on the Gaussian random walk.

The mathematical basis for introducing multiple biasing distributions is through associating a weight with each distribution. If there are J biasing distributions where M_j is the number of samples drawn for the j^{th} biasing target then the way to compute the probability density function is

$$\hat{P} = \sum_{j=1}^{J} \hat{P}_j = \sum_{j=1}^{J} \frac{1}{M_j} \sum_{m=1}^{M_j} w_j(\vec{x}_{j,m}) I(\vec{x}_{j,m}) L_j(\vec{x}_{j,m}) \tag{3.19}$$

where $I(\vec{x}_{j,m})$, $w_j(\vec{x}_{j,m})$, and $L_j(\vec{x}_{j,m})$ are the indicator function, weight, and likelihood ratio associated with the m^{th} sample drawn from the j^{th} distribution. Note that there are J likelihood ratios computed with each sample: one for each biasing. Additionally,

the overall estimator $\hat{P}$ can be decomposed into individual estimators arising from each biasing $\hat{P}_j$ which can help to assess the performance of each biasing distribution in the overall ensemble.

The choice of the weights can be sample dependent, and in order to ensure an unbiased estimator, we need $\sum_{j=1}^{J} w_j(\vec{x}) = 1$ for all $\vec{x}$. This obviously leaves a large range of choices for the weights, but the primary one used in this thesis is the balance heuristic

$$w_j(\vec{x}_{j,m}) = \frac{M_j p_j^*(\vec{x}_{j,m})}{\sum_{j'=1}^{J} M_{j'} p_{j'}^*(\vec{x}_{j,m})} \tag{3.20}$$

introduced by Veach and Guibas and refined by Veach in his thesis [**91, 92**]. An intuition for the balance heuristic can be given by substituting Eq. (3.20) into Eq. (3.19) which gives

$$\begin{aligned} \hat{P} &= \sum_{j=1}^{J} \frac{1}{M_j} \sum_{m=1}^{M_j} \Big(\frac{M_j p_j^*(\vec{x}_{j,m})}{\sum_{j'=1}^{J} M_{j'} p_{j'}^*(\vec{x}_{j,m})} \Big) I(\vec{x}_{j,m}) \frac{p(\vec{x}_{j,m})}{p_j^*(\vec{x}_{j,m})}, \\ &= \sum_{j=1}^{J} \sum_{m=1}^{M_j} \frac{I(\vec{x}_{j,m}) p(\vec{x}_{j,m})}{\sum_{j'=1}^{J} M_{j'} p_{j'}^*(\vec{x}_{j,m})}. \end{aligned} \tag{3.21}$$

The denominator in this fraction can be thought of as the "combined sample density" or the probability of the sample averaged across all biasing distributions [**91**]. Therefore, the balance heuristic can be thought of as weighting samples by the probability of the j^{th} distribution divided by the probability averaged over all biasing distributions. It can be shown that no other weighting function performs significantly better in terms of reducing the estimator variance, in that the variance reduction achieved by all weights cannot be more than a small amount better than the balance heuristic [**91**]. That is, there are

specific cases in which different weights can be slightly better, but overall the balance heuristic is always at least close to optimal.

The variance associated with the MIS estimator can similarly be decomposed into contributions from each biasing distribution

$$\hat{\sigma}^2_{\hat{P}} = \sum_{j=1}^{J} \frac{1}{M_j(M_j-1)} \hat{S}_j \quad \text{where} \tag{3.22a}$$

$$\hat{S}_j = \sum_{m=1}^{M_j} \left(w_j(\vec{x}_{j,m}) I(\vec{x}_{j,m}) L_j(\vec{x}_{j,m}) - \hat{P}_j \right)^2. \tag{3.22b}$$

The forms for the MIS probability and variance estimators using the balance heuristic can be slightly simplified by noting that

$$w_j(\vec{x}_{j,m}) L_j(\vec{x}_{j,m}) = \frac{1}{\sum_{j'=1}^{J} M_{j'} \frac{1}{L_{j'}(\vec{x}_{j,m})}}. \tag{3.23}$$

Therefore, the inverse likelihood ratio for each biasing is often tracked with each sample, rather than the likelihood ratios. As before, the C.V. can be computed as $\hat{\sigma}_{\hat{P}}/\hat{P}$ and can be used to assess MIS simulation performance. Take Fig. 3.3 as an example, where MIS has been applied to the 1D Gaussian random walk of Eq. (3.14). Three biasing distributions were included in the simulation, one to cover each of the left tail, middle, and right tail of the distribution. The benefits and potential drawbacks of MIS are shown by comparing the MISMC estimate (dotted curve in the figure) against an MC simulation (thick dashed curve) with the same number of overall samples. The MISMC estimate accurately captures a much wider range of probability space than the MC simulation and has good convergence for approximately $|Z_{10}| < 15$ as indicated by the C.V. The estimate has a slightly worse C.V. for $|Z_5| < 15$ as less samples were drawn from the

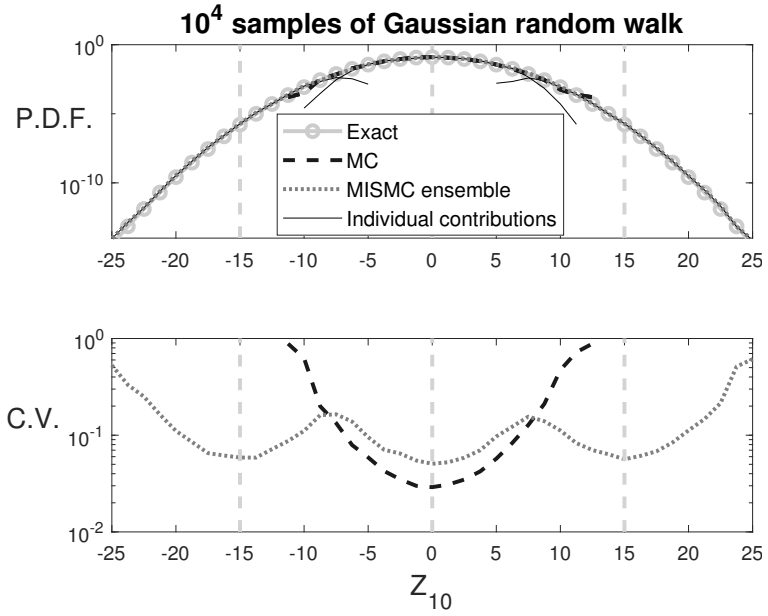

Figure 3.3. Exact and numerical approximations of the P.D.F. of the Gaussian random walk from Eq. (3.14) with $N = 10$ steps and 10,000 samples taken for each numerical approximation. The top panel shows three P.D.F.'s: the exact solution, one obtained by MC simulation, and one obtained using MISMC simulation using the balance heuristic with three biasing distributions for the left tail, bulk, and right tail of the distribution (the targets are given with vertical dashed lines). The contributions of each biasing distribution to the overall estimate are also depicted with thin black lines. The bottom panel shows the C.V. of the two MC simulations, which visually demonstrates their convergence properties. The vertical gray dashed lines indicate the MISMC biasing targets.

middle biasing of the MISMC than for the simple MC. Lastly, note that the MISMC C.V. undulates slightly, peaking in between the biasing targets at ± 5. This is because relatively fewer samples landed in these regions.

3.4. Dynamic Importance Sampling

Dynamic or targeted importance sampling (DISMC or DIS) is another modification to importance sampling that has been used in a variety of contexts [**12, 29, 30, 36**]. In

the context of computing a P.D.F. that results from a random process, it can be thought of as remedying an opposite problem to MIS: where MIS seeks to resolve large regions of phase space in a single batch simulation, DIS seeks to resolve highly localized regions of probability space that are hard to reach by means of conventional IS. The key concept in dynamic IS is that the biasing distribution is adjusted adaptively, mid-simulation, to ensure that a high proportion of samples cluster tightly around the target.

In application, this requires solving a series of optimization problems that are similar in character to Eq. (3.16), for each sample. The operative difference is that the adaptive adjustments to the biasing distribution slightly complicate the likelihood ratio for the sample from Eq. (3.17). Now, the biasing noise realization at each step may depend on prior steps so the product formulation for the overall likelihood ratio of the sample is no longer valid. However, it can be reformulated in terms of conditional probabilities so that the joint distribution of the entire sample is still factored into a product [**11**]. Mathematically this only relies on using the law of conditional probabilities (namely that $p(AB) = p(A|B)p(B)$) recursively so that the joint probability of the noise in a particular sample is written

$$
\begin{aligned}
p(\mathbf{z}_1, \ldots, \mathbf{z}_N) = & \, p(\mathbf{z}_N | \mathbf{z}_{N-1}, \mathbf{z}_{N-2}, \ldots, \mathbf{z}_2, \mathbf{z}_1) \\
& \times p(\mathbf{z}_{N-1} | \mathbf{z}_{N-2}, \ldots, \mathbf{z}_2, \mathbf{z}_1) \cdots p(\mathbf{z}_2 | \mathbf{z}_1) p(\mathbf{z}_1).
\end{aligned} \tag{3.24}
$$

This formalism based on conditional distributions then allows us to compute the likelihood ratios in a manner very similar to before. Considering a concrete example of dynamic importance sampling will make this more clear.

Constrained to parametric mean-shift biasing of Gaussian random noise, the implementation of dynamic importance sampling is not too complicated. When the biasing distribution is adjusted, the optimization problem is re-solved using a different set of initial conditions, to give a new set of mean-shifts to be used throughout the rest of the simulation of that sample. The likelihood ratio is computed the same as before, using a product of the biasings computed at each step. Practically, the only difficulties in using dynamic IS comes from repeatedly solving the optimization problem for the biasing, and choosing the frequency with which the biasing path will be recalculated.

Consider the 1D Gaussian random walk Eq. (3.14). This is not a problem that requires DISMC as all regions of probability space are easy to reach by regular IS, but it is useful for demonstration purposes. Recalculating the biasing path is trivial here. If we want to calculate the mean-shift for the n^{th} step when trying to reach a target c when there are N total steps we simply compute

$$\mu_n = \frac{c - \sum_{k=1}^{n-1} \mathbf{z}_k}{N - (n-1)}. \tag{3.25}$$

Now, we can choose to perform this recalculation of the mean-shift as often as we like. Recalculating more often will lead to a sample spread more tightly clustered around the intended target, as is illustrated in Fig. 3.4. Choosing a recalculation schedule is often a matter of balancing achieving a desired convergence against covering a wanted region of the probability distribution (and also factoring in the computational work of recalculation in harder examples). In Fig. 3.4, two different recalculation schedules are employed: recalculating the path once halfway through simulation and recalculating every step. As stated previously, the samples more tightly cluster around the target as recalculation is

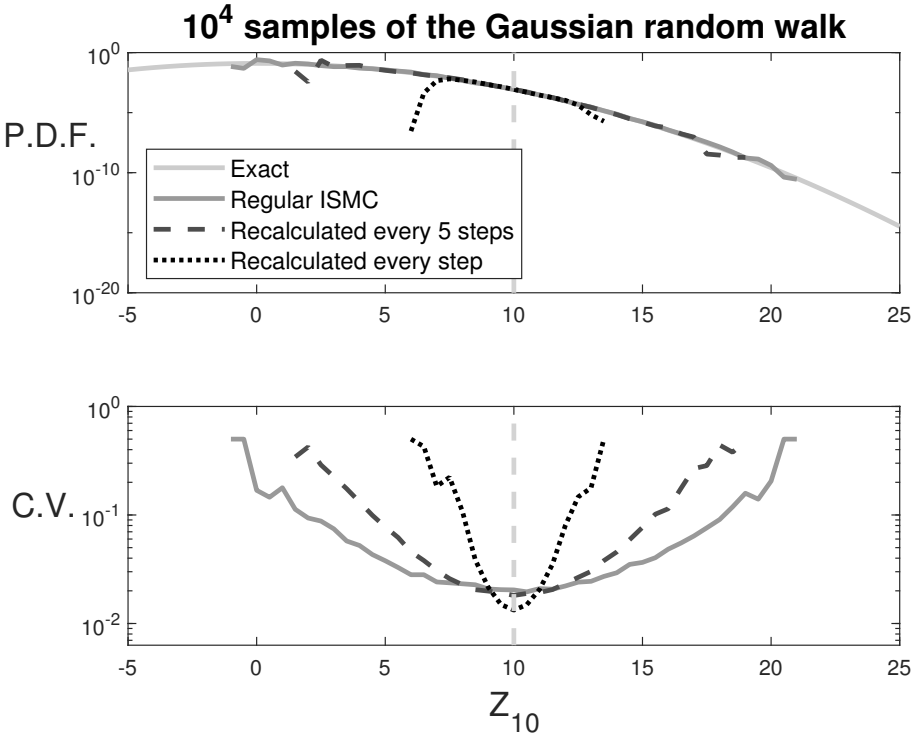

Figure 3.4. Exact and numerical approximations of the P.D.F. of the Gaussian random walk from Eq. (3.14) with $N = 10$ steps and 10,000 samples taken for each numerical approximation. The top panel shows four P.D.F.'s: the exact solution, one obtained by ISMC simulation, and two obtained using DISMC simulation with varying amounts of path recalculation. The bottom panel shows the C.V. of the three ISMC simulations, which visually demonstrates their convergence properties. The vertical gray dashed line indicates the ISMC biasing target, common to all simulations.

increased. Additionally, the convergence in the C.V. improves due to this clustering, though there is typically a saturation point beyond which additional recalculation leads to little improved convergence. Here, the convergence improvement relative to regular IS is small as regular IS does a reasonable job of guiding samples to the target. Additionally, it should be noted that other types of recalculation schedules than recalculating at regular

intervals may be employed, such as recalculating whenever the sample wanders outside of a prescribed distance from the anticipated trajectory by the biasing.

3.5. Importance Sampling for a Soliton System

A few comments on implementing importance sampling for soliton-based models like Eq. (2.16) are warranted in order to proceed to Chapter 5. Recall that at each amplifier ASE noise is added which can cause parameter jitter in the soliton. Such jitter can be unwanted in many physical contexts as it may cause a lightwave communication system to fail to transmit a signal correctly. Simple Monte Carlo simulation to assess large deviations in the parameters can be ineffective due to the low probability with which they occur. This provides a perfect opportunity to use importance sampling, and there has been extensive use of this technique in a variety of soliton systems [**11, 27, 68, 72, 82, 88**]. In particular, the review paper by Moore et. al. [**73**] is a concise and thorough introduction to this type of simulation. In this section, we will discuss a few of the implementation issues involved with using importance sampling, taking the example of simulating amplitude jitter in Eq. (2.16) with no deterministic perturbations as an introductory guiding problem.

In a discretized version of Eq. (2.16), both the soliton and the noise are characterized by complex numbers at N temporal discretization points. Therefore, the dimension of the complex noise vector $\vec{x}$ is $2N \times 1$. Each element in $\vec{x}$ here is a real-valued random number with the first N entries of $\vec{x}$ taken to be the real part of the noise and the last N the imaginary part. Due to the ASE noise process, these random numbers are i.i.d. Gaussians with mean zero. For an overall noisepower of σ^2, the variance of each Gaussian used in the discretization must be $\sigma_a^2 = \sigma^2/(2\Delta t)$. Therefore, the probability distribution

for such an overall noise vector is a multivariate normal distribution characterized by a mean vector $\vec{\mu}$ and a scalar variance

$$p_{\vec{x}}(\vec{x};\vec{\mu},\sigma_a^2) = \frac{1}{(2\sigma_a^2)^N}\exp\Big[-\frac{(\vec{x}-\vec{\mu})^T(\vec{x}-\vec{\mu})}{(2\sigma_a^2)}\Big] \tag{3.26}$$

(presented this way as we will mean-shift the noise for biasing and each point in the domain is noise shifted by a different amount). Consider a generic $2N\times 1$ biasing vector $\vec{b}$ ordered with real parts before imaginary parts as for $\vec{x}$. The biasing density function for a *biased* noise realization $\vec{x}^* = \vec{x}+\vec{b}$ is then simply $p^*_{\vec{x}}(\vec{x}^*) = p_{\vec{x}}(\vec{x}^*;\vec{b},\sigma_a^2) = p_{\vec{x}}(\vec{x};0,\sigma_a^2)$. Recall that the likelihood ratio, Eq. (3.7), is generically given by $L = p(\vec{x}^*)/p^*(\vec{x}^*)$ so explicitly we have

$$L(\vec{x}^*) = \frac{p_{\vec{x}}(\vec{x}^*;0,\sigma_a^2)}{p_{\vec{x}}(\vec{x}^*;\vec{b},\sigma_a^2)} = \frac{\exp[-(\vec{x}^*)^T(\vec{x}^*)/(2\sigma_a^2)]}{\exp[-(\vec{x}^*-\vec{b})^T(\vec{x}^*-\vec{b})/(2\sigma_a^2)]}. \tag{3.27}$$

Everything up to this point has been slightly simplified as it has only dealt with what happens at one amplifier, but of course in practice there are many amplifiers in a simulation. As before, the overall likelihood ratio can be written as the likelihood ratio at each individual amplifier. Therefore, the overall likelihood ratio for a sample with N_a amplifiers can be written compactly by assembling the noise vectors into a $2N\times N_a$ matrix like $X^* = (\vec{x}^*_1,\ldots,\vec{x}^*_{N_a})$, then

$$L(X^*) = \prod_{n=1}^{N_a} L(\vec{x}^*_n). \tag{3.28}$$

This notation is further complicated by multiple importance sampling. To compute Eq. (3.23) we only need to compute the likelihood ratios for each biasing distribution.

However, the biased noise realization and the biasing distribution do not generally share the same mean shifts. In other words, for MIS one needs to compute likelihood ratios for biasing distributions that differ from the distribution used to create the biased noise. Since the only difference between biasing distributions is a different set of mean-shifts, this is not too difficult to implement. For simplicity of notation, say there are two biasing distributions with two different biasing vectors at the n^{th} amplifier, $\vec{b}_{n,1}$ and $\vec{b}_{n,2}$. If the first distribution is used to bias the sample, then the biased noise is $\vec{x}_n^* = \vec{x}_n + \vec{b}_{n,1}$. The likelihood ratio for the first distribution is then unchanged from Eq. (3.27). However, the likelihood ratio for the second biasing distribution is

$$L_2(\vec{x}_n^*) = \frac{p_{\vec{x}}(\vec{x}_n^*; 0, \sigma_a^2)}{p_{\vec{x}}(\vec{x}_n^*; \vec{b}_{n,2}, \sigma_a^2)} = \frac{p_{\vec{x}}(\vec{x}_n + \vec{b}_{n,1}; 0, \sigma_a^2)}{p_{\vec{x}}(\vec{x}_n + \vec{b}_{n,1} - \vec{b}_{n,2}; 0, \sigma_a^2)}. \tag{3.29}$$

3.5.1. An example: biasing amplitude changes

Consider biasing for amplitude jitter in a soliton system where there are no continuous perturbations. The work in Section 2.3.2 and Appendix A explained that the biasing vector for a soliton parameter is proportional to the adjoint mode for that soliton parameter. Therefore, to bias the amplitude we add a term that is proportional to the adjoint amplitude mode $\underline{v}_E$ at each amplifier. The full term for the biasing at the n^{th} amplifier is

$$\vec{b}_n = \Delta E_n \frac{\underline{v}_E(z_n, t)}{||\underline{v}_E(z_n, t)||^2} \exp\left[i\Theta(z_n, t)\right] \tag{3.30}$$

which leads to an expected jump in the soliton amplitude of ΔE_n. Note in particular that the biasing has to be added in phase with the soliton at $z = z_n$, the location of

the n^{th} amplifier. The soliton phase Θ and adjoint linear mode are computed using the parametrization of the soliton extracted at $z = z_n$ prior to amplification.

Suppose we want to simulate the probability that a soliton undergoes an amplitude deviation of ΔE_{tot} after propagation through N_a total amplifiers. The most probable way for this deviation to occur is that which minimizes the cumulative L^2-norm of the noise. We can determine this most-probable noise configuration by minimizing the norm of the amplitude biasing function from Eq. (3.30). This significantly simplifies matters as it re-casts the infinite-dimensional problem of minimizing the noise into a finite-dimensional one where only the discrete amplitude changes, ΔE_n, need to be determined. These amplitude changes are also known as "biasing coefficients" and the minimization problem to determine them here takes the form

$$\min_{\Delta E_n} \sum_{n=1}^{N_a} \frac{\Delta E_n^2}{2E_n} \quad \text{subject to} \quad \sum_{n=1}^{N_a} \Delta E_n = \Delta E_{\text{tot}}. \tag{3.31}$$

This constrained optimization problem can be solved using Lagrange multipliers. Additionally, if the solution is expanded for a small total amplitude change (or ΔE_{tot} small) then the leading-order solution is $\Delta E_n = \frac{\Delta E_{\text{tot}}}{N_a}$. Note that higher order corrections can be found, but they are ignored here [**73**].

With the biasing coefficients found, importance sampling simulations can be performed. With a desired amplitude change, the soliton is simulated and at each amplifier the noise is biased by adding the mean-shift vector from Eq. (3.30) to the noise. Drawing multiple samples in this fashion allows us to approximate statistics for amplitude jitter, or the change in amplitude induced by ASE noise. Furthermore, picking multiple amplitude

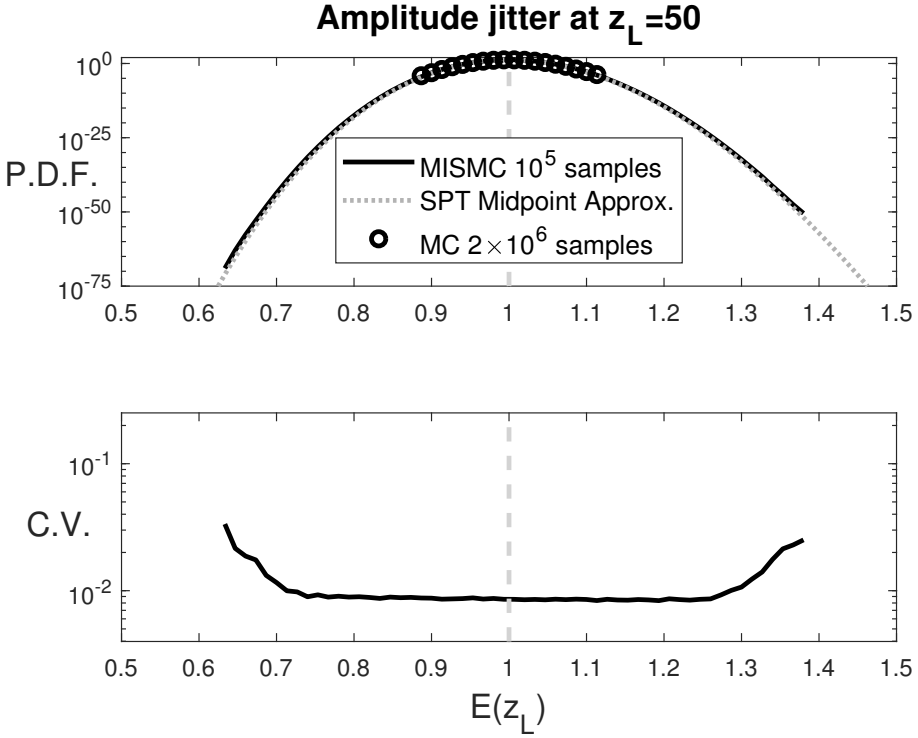

Figure 3.5. Amplitude jitter for NLS solitons simulated out to a dimensionless distance of $z_L = 50$ with amplifier spacing $z_a = 1$ and a dimensionless noisepower of $\sigma^2 = 10^{-5}$. The MISMC simulations used a series of amplitude targets between $E = 0.7$ and $E = 1.3$ with 10^5 total samples and agree with a midpoint model of amplitude jitter from SPT (derived in appendix D of [**73**]). Monte Carlo simulations with 2×10^6 samples also agree with the theory, but only down to density levels of 10^{-6} whereas the MISMC simulations with much fewer samples have good agreement down to density levels below 10^{-50}. The C.V. for the MISMC simulations in the bottom panel shows that the numerical convergence properties of the simulation throughout the region are good.

changes (multiple values of ΔE_{tot}) enables the simulation of the entire P.D.F. of amplitude jitter. Such a simulation, of a soliton with initial amplitude $E_0 = 1$, is performed in Fig. 3.5, which shows good agreement with the theory from [**73**].

CHAPTER 4

Dynamic Multiple Importance Sampling

This thesis is particularly concerned with problems where combining dynamic importance sampling and multiple importance sampling is warranted. Such a problem is considered in the laser model discussed in Chapter 5, for instance. This may seem like a counterintuitive proposition, as DIS and MIS have been presented thus far as solving opposite problems in Chapter 3. DIS targets a localized region of probability space in a probability distribution, whereas MIS allows for resolving a large region of probability space. In fact, MIS was originally devised for light rendering in computer graphics, where the optimal distribution varied from pixel to pixel, and very large numbers of pixels needed to be rendered quickly [**92**]. However, there is another use for MIS: when there are multiple biasing paths to a single region of probability space. It is well known that when there are multiple paths to a rare event of interest, all must be included in an ISMC simulation in order to get reasonable convergence. Rare events with multiple paths have been noted in toy problems [**43**], nucleation models [**32**], and nonlinear optics [**83**]. This is not to say that MIS is the only possible ISMC method that addresses this issue, as other variants of IS such as mixture IS also involve sampling from multiple distributions, but including information from all relevant biasing paths is needed in such methods, just as it is in multiple importance sampling [**15, 33, 75**]. However, the problem of combining DIS with MIS has been little studied, especially in an optical setting. In this chapter, we

will motivate the use of dynamic multiple importance sampling (DMIS), or using multiple importance sampling where individual biasing components involve dynamic path recalculation, with a simple example and discuss issues with its implementation.

4.1. A Simple Example with Multiple Paths

This section examines a problem with multiple paths. Consider a 2D discrete random walk similar to Eq. (3.14) where the steps are drawn from Gaussian distributions:

$$x_N = x_0 + \sum_{i=1}^{N} \Delta W_i^{(1)}; \qquad y_N = y_0 + \sum_{i=1}^{N} \Delta W_i^{(2)}. \tag{4.1}$$

The variances of each step distribution ($\Delta W_i^{(1)}$ and $\Delta W_i^{(2)}$) are drawn from the standard Gaussian distribution with mean zero and variance one ($p_G(x, 0, 1)$). We are concerned with the probability that the walker is at a particular x-position after N steps. Since the steps consist of uncorrelated Gaussians in the x- and y-directions, the y-dynamics are obviously irrelevant for the x-position. However, this random walk takes place in the presence of a reflecting barrier wall that is oriented vertically at a position $x = x_B$ that brings the y-steps into play.

This barrier extends from $y = -h$ to $y = h$ and is reflecting. This means that a walker will be unable to access x-regions beyond x_B unless it passes x_B with $|y| > h$. Therefore, we should expect that this wall should distort the Gaussian distribution that would be expected in the absence of the wall by shifting probability density from beyond the wall to before the wall. This is simply due to the fact that the reflecting wall makes it more difficult for the random walk to reach the region past the wall, so the probability there will

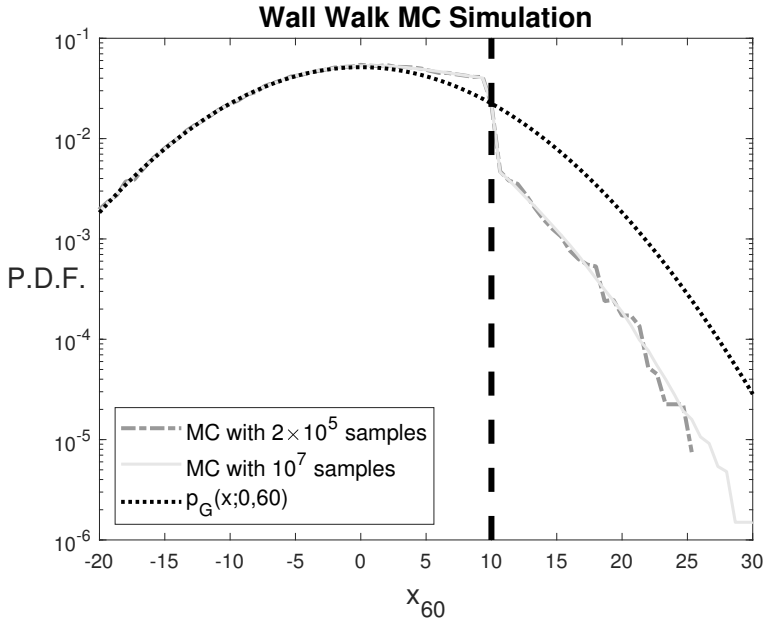

Figure 4.1. Two Monte Carlo simulations (one with 2×10^5 and one with 10^7 samples) of the random walk showing the P.D.F. of final walker x-position are shown along with a Gaussian distribution with variance 60. The right side of the distribution shows the effect of the wall, which is at $x = 10$. Samples are reflected back and excess probability density is observed for $0 < x_{60} < 10$, whereas probability density is lower for $x > 10$ relative to the Gaussian. Probabilities past the wall are much lower than probabilities before the wall and the jaggedness for large x-values reflects that a dearth of samples reach that region by simple chance.

be smaller. This is exactly what we see in Fig. 4.1, where we have chosen $h = 10, x_B = 10$, and $N = 60$ (which will be taken for all other simulations unless explicitly stated). The exact Gaussian distribution in the absence of the wall $p_G(x; 0, 60)$ is shown in the black-dashed curve, and we see that the numerical approximation obtained by MC simulation has more probability density for locations before the wall, lower probabilities for those past the wall, and close to a jump-discontinuity at the location of the wall.

Applying importance sampling to the wall walk model requires computing the optimal path for a walker to reach a location past the wall x_F at step N. The optimal path to the

target $x_N = x_F$ (with $x_F > x_B$) can be derived using Lagrange multipliers in a manner similar to the 1D walk. Since the noise is i.i.d. Gaussian noise, we just have to pick a series of mean-shifts (Δx_i for $\Delta W_i^{(1)}$ and Δy_i for $\Delta W_i^{(2)}$) that lead the walker to its target. As was the case in Eq. (3.16), the most probable series of mean shifts come from minimizing the L^2 norm of the total noise or

$$\min_{\substack{\Delta x_i, \Delta y_i, \\ i=1,\ldots,N}} \Big(\sum_{i=1}^{N} (\Delta x_i)^2 + (\Delta y_i)^2 \Big) \tag{4.2}$$

subject to $x_N = x_F$ and that the walker passes around the wall. Intuitively, we expect that there should be two paths to any x-value beyond the wall: one going above the wall and one going under. If $|y_0| \geq h$, then the optimal path should be that $\Delta x_i = (x_F - x_0)/N$ and $\Delta y_i = 0$ for all i. However, if $|y_0| < h$ then we expect that these paths should touch the top and bottom of the wall as this will minimize the length of the path; each sample will be started at $(0, 0)$ in the plane but we keep the starting location arbitrary for when we need to solve the optimization problem in dynamic IS. In this case we need to do more work to derive the optimal path(s).

First, we will assume that it takes $N_1 < N$ steps to reach the intermediary wall. That is, $x_{N_1} = x_B$ and we will split the minimization problem Eq. (4.2) up into a nested

	$i \leq N_1$	$i > N_1$
Δx_i	$\dfrac{x_B - x_0}{N_1}$	$\dfrac{x_F - x_B}{N - N_1}$
Δy_i	$\dfrac{\pm h - y_0}{N_1}$	0

Table 4.1. The mean-shifts for the wall walk model are presented. Note that there are two possible paths, as indicated by the plus/minus in the y-shifts.

minimization problem. Therefore we seek to minimize

$$\min_{N_1}\left(\min_{\substack{\Delta x_i,\Delta y_i,\\ i=1,\ldots,N}}\left(\sum_{i=1}^{N_1}\{(\Delta x_i)^2+(\Delta y_i)^2\}+\sum_{i=1}^{N-N_1}\{(\Delta x_{N_1+i})^2+(\Delta y_{N_1+i})^2\}\right)\right) \tag{4.3}$$

$$\text{subject to} \qquad x_N = x_0 + \sum_{i=1}^{N}\Delta x_i = x_F,$$

$$x_{N_1} = x_0 + \sum_{i=1}^{N_1}\Delta x_i = x_B,$$

$$\text{and } |y_{N_1}| = |y_0 + \sum_{i=1}^{N_1}\Delta y_i| \geq h.$$

The solution of the interior optimization problem is fairly straightforward. For simplicity, we will replace the third constraint in Eq. (4.3) with an equality for now. A rigorous treatment of the full inequality condition is given in Appendix C. As such we have a minimization problem with three side-constraints. To solve it, we give a Lagrange multiplier to each of the three constraints and take derivatives with respect to $\Delta x_i, \Delta y_i$, and the three Lagrange multipliers. Equating each of these derivatives to zero leads to the solutions shown in Table 4.1. The plus/minus in the expression for Δy_i refers to whether

the path goes above or below the wall, and can be derived by assuming either $\pm h$ in the third constraint.

In any case, this solution reduces the exterior problem in Eq. (4.3) to

$$\min_{N_1} \frac{(x_B - x_0)^2 + (\pm h - y_0)^2}{N_1} + \frac{(x_F - x_B)^2}{(N - N_1)}. \tag{4.4}$$

Now, temporarily treating the integer quantity N_1 as a continuous variable allows us to differentiate and equate with zero again. Note that this is an approximation and that, in fact, assuming that the walker goes through x_B exactly at one of its discrete steps is also an approximation. Let $A = (x_B - x_0)^2 + (\pm h - y_0)^2$ and $B = (x_F - x_B)^2$ so

$$\frac{d}{dN_1}\Big(A/N_1 + B/(N - N_1)\Big) = \frac{B}{(N - N_1)^2} - \frac{A}{N_1^2} = 0.$$

Solving this equation leads to

$$N_1 = N\left(\frac{A \pm \sqrt{AB}}{A - B}\right) \text{ or } N_1 = \frac{N}{2} \text{ if } A = B \tag{4.5}$$

from which we take the negative root in order to ensure that $N_1 < N$. Additionally, this value must be rounded to the nearest integer.

Visual representations of the biasing paths are shown in Fig. 4.2. Four paths are shown; two paths that originate at $(0, 0)$, and two paths that originate at a location above the wall (these will be relevant for DIS). From each location, there is an optimal path that goes over the wall, and an optimal path that goes under the wall. Due to the symmetry of the wall, we expect that the trajectories from the origin are equally likely to give samples hitting the target of $x_F = 20$. However, for the other two paths, we

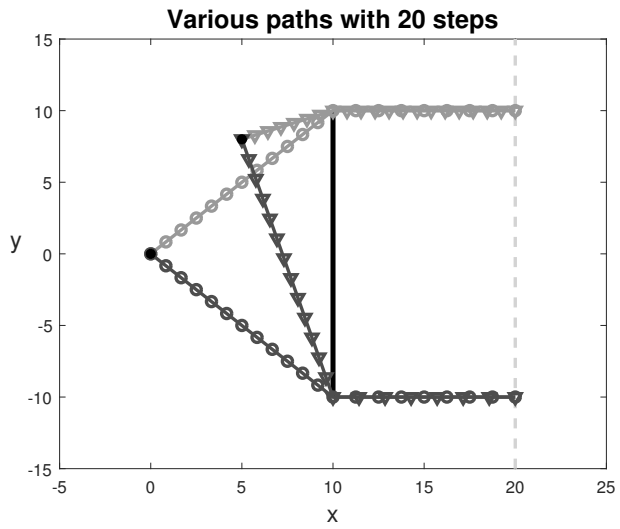

Figure 4.2. Four computed biasing paths from two different starting locations (x_0, y_0), with two passing above the wall and two passing below the wall that arrive at a final position of $x_F = 20$. Discrete markers are shown on the paths to show the size of the steps the walker is expected to take along each path. From this, one can see that the number of steps taken past the wall is dependent upon the distance traversed to get around the wall. The optimal path passing over the wall (not shown) when $y_0 > h$ is just a straight horizontal path to the target as the wall is irrelevant to the most probable path.

expect that the over wall path is more likely than the under wall path. This notion of the "likeliness" of a path will be developed with more rigor later, but for now the intuition that the over wall path is better should be plain from the fact that the walker needs to travel a much smaller distance to pass directly over the wall than to travel to the bottom of the wall and over. Of course, intermediate steps may push the walker down below the mid-line of the wall, and then the below-wall path would be the most probable path.

Now that we have the biasing paths we can perform importance-sampled Monte Carlo simulations for position probabilities past the wall. First, we demonstrate that the two equally likely paths (by symmetry) are required for the ISMC simulations to converge.

Figure 4.3 shows the results of an ISMC simulation, an MISMC simulation, and the MC results from Fig. 4.1 for comparison. The ISMC simulation only uses one of the two paths (the path going over the wall) while the MISMC simulation uses both paths, with half the samples using the over path and half using the under path with the balance heuristic weighting the contribution of each. Note that the ISMC results are extremely poor: the C.V. is as high as the MC simulation using the same number of samples while the P.D.F. is very jagged. In contrast, the MISMC simulations produce a P.D.F. that visually agrees with the MC simulations using 10^7 samples and achieves a similar rate of convergence near $x = 20$ despite using one fiftieth of the samples as those MC simulations. Therefore, it is clear that both paths must be included in importance sampling simulations for this problem. This simulation demonstrates that IS can only be applied to this problem by considering both paths. Next we will try to develop an importance sampling algorithm that utilizes dynamic importance sampling to try to improve upon this convergence while taking into account the requirement that both paths must be used.

4.2. The Dynamic Multiple Importance Sampling Algorithm

Dynamic multiple importance sampling is a rather straightforward combination of DIS and MIS as presented in Sections 3.3 and 3.4. Multiple biasing distributions (or paths) are used in a simulation where the distributions are adaptively recalculated during the drawing of samples. The intuition for dynamic multiple importance sampling relies on a simple rule: recalculating a biasing path should not depend on whether that path is actively biasing the current sample.

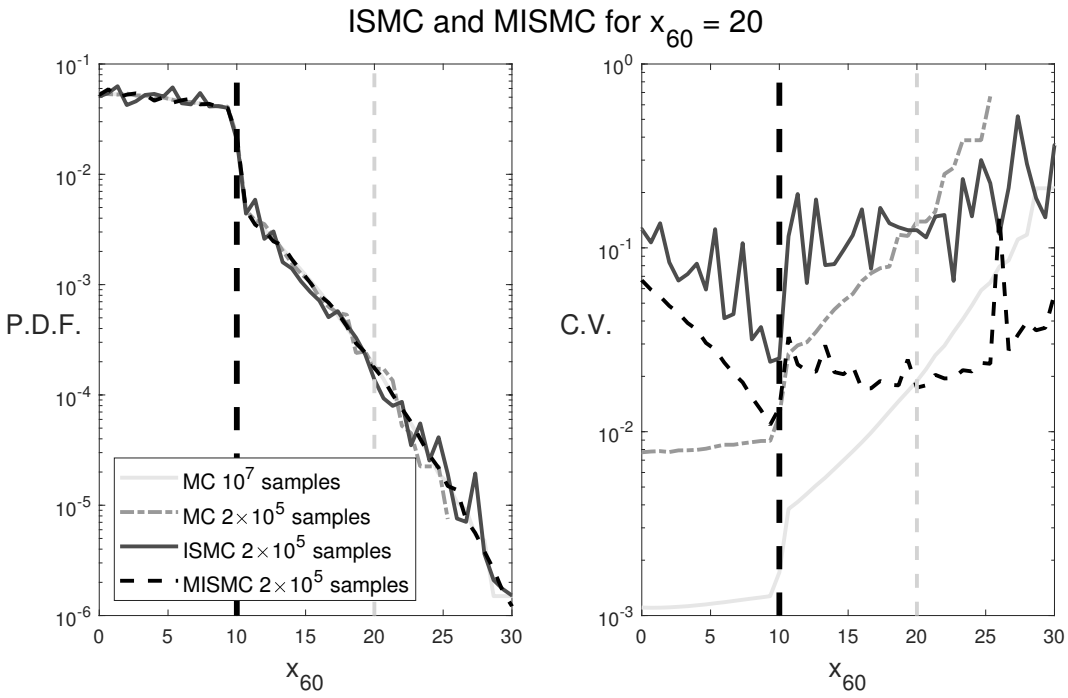

Figure 4.3. Importance sampling simulations of the final walker position, with the same parameters as in Fig. 4.1. The MISMC simulation, using both over- and under-wall paths, converges to the same probability distribution as the MC simulation while the ISMC simulation, using just the over-wall path, exhibits very poor convergence. Also note that the MISMC simulation's convergence is far superior to the convergence of the MC simulations with the same number of samples near the target of $x_{60} = 20$.

To make this concrete, consider the Gaussian random walk from Eq. (3.14). Say we have two probability targets, c_1 and c_2, and we want to use DMIS with path recalculation every step. In this case we must recalculate the paths to both targets for each biasing after every step of each sample. Consider a sample being biased towards c_1, at its n^{th} step. To be precise, before the n^{th} step, the steps $X_1, \ldots, X_{n-1}$ have been drawn according to dynamic importance sampling for the target c_1. When drawing the n^{th} step we must calculate the mean shift for biasing towards c_1, based on the state of the random walk at step $n-1$ (i.e. $Z_{n-1} = \sum_{i=1}^{n-1} X_i$). We also must do a similar calculation for c_2,

although this will have no role in the biasing of the current sample and will only be used in calculating a likelihood ratio. If we call these mean shifts $\mu_{n,1}$ and $\mu_{n,2}$ for the n^{th} step to targets 1 and 2 respectively, they are given simply by

$$\mu_{n,1} = \frac{c_1 - Z_{n-1}}{N-(n-1)} \quad \text{and} \tag{4.6}$$
$$\mu_{n,2} = \frac{c_2 - Z_{n-1}}{N-(n-1)}.$$

The n^{th} step is then drawn from a Gaussian with mean $\mu_{n,1}$, and $\mu_{n,2}$ is saved for the calculation of the likelihood ratio for the second biasing. The likelihood ratios are then combined according to the balance heuristic (Equation 3.20). When this is performed appropriately, the results appear as shown in the top panel of Fig. 4.4, with each biasing distribution's contribution exhibiting a combination of the features present in MIS and DIS. They balance appropriately in regions of overlap and sharply drop off away from their targets.

The bottom panel in Fig. 4.4 shows an incorrect attempt at DMIS where samples are treated inconsistently. Samples that are biased towards $Z = 10$ do not recalculate the path towards $Z = 15$ and vice versa. It is elucidating to consider this incorrect method and why it fails. Imagine that we are biasing a sample towards c_1 and calculate $\mu_{n,1}$ according to Eq. (4.6) (i.e. using the current state of the random walk), but only use the initial state of the random walk for $\mu_{n,2}$ (or $\mu_{n,2} = c_2/N$ as the initial state is $Z_0 = 0$) for all n. In this way, we are only recalculating the path that is "actively" being used in that sample, and are not recalculating the other "inactive path." In this scenario, all samples biased towards c_1 are treated in this manner, and all samples biased towards c_2 have

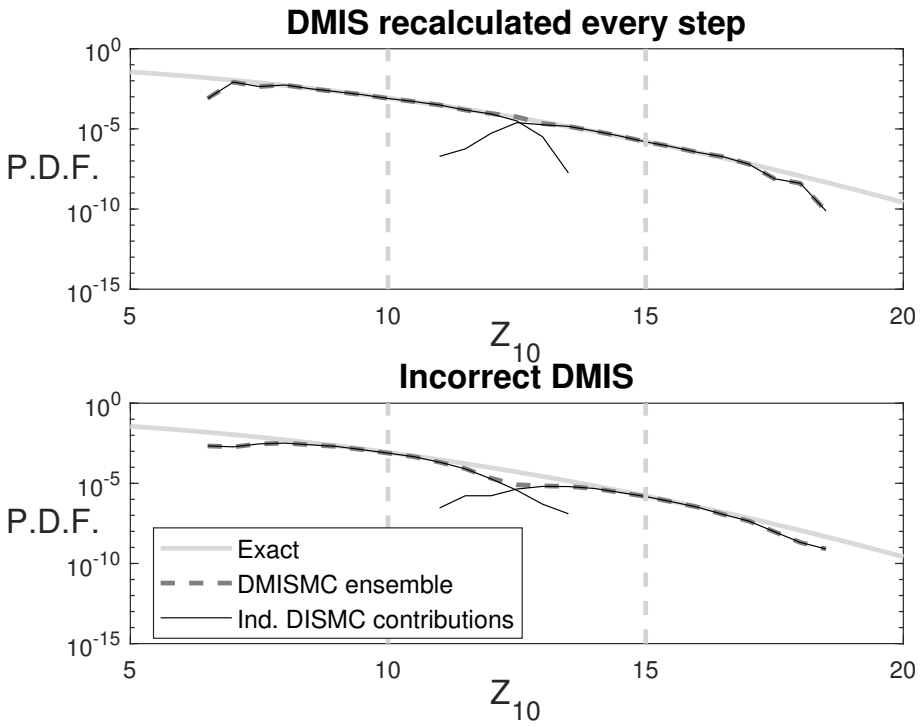

Figure 4.4. Dynamic multiple importance sampling implemented on the 1D Gaussian random walk. Two probability targets at $Z = 10$ and $Z = 15$ were used with all samples recalculated after each step of the random walk. In the top panel, DMIS is implemented correctly, with the path to each target recalculated after every step. In the bottom panel, DMIS is implemented incorrectly, with only the path to the current target recalculated after every step. The dip in the DMIS ensemble estimate between the two targets at roughly $Z = 12.5$ is caused by the inconsistent implementation.

$\mu_{n,2}$ recalculated with $\mu_{n,1}$ left unrecalculated. Samples that fall in between the biasing targets of $Z = 10$ and $Z = 15$ get down-weighted as the likelihood ratio overestimates the likelihood that the sample was biased toward the inactive target. This helps to underline that recalculation must be consistent across distributions. It is not surprising that this simulation method produces incorrect results, as it sometimes treats samples as if they were drawn from conditional distributions (i.e. using DIS: see Equation 3.24) and sometimes not.

Another way to illustrate this point is to consider an example where different recalculation criteria are used for each distribution. These recalculation criteria for each distribution must be consistent across all samples, regardless of which distribution is actively driving the sample. For instance, in Fig. 4.5, the Gaussian random walk is simulated with two distributions: one distribution targets $Z = 10$ and recalculates the path every five steps and the other targets $Z = 15$ with the path recalculated every step. Regardless of whether a sample is being generated from the distribution targeting position 10 or 15, the biasing path to $Z = 15$ is recalculated after every step, and the path to $Z = 10$ is recalculated after every fifth step and the likelihood ratios computed accordingly. The results of such a scheme are fairly predictable; the spread of samples is wider around the $Z = 10$ target and the convergence is slightly better near the $Z = 15$ target due to tighter clustering of samples from that biasing distribution. The consistency of recalculation makes the overall estimate match the exact answer for the P.D.F. in the region between the two targets.

4.3. Large Deviation Theory

Before discussing the DMIS algorithm in a multiple-path context a bit of background in large deviation theory is needed. This theory provides a more rigorous way to consider exit problems than those presented thus far in this thesis. I.e., instead of minimizing the noise on the grounds that the minimum is most probable due to the noise being zero-mean, this theory provides an asymptotic basis for considering noise-induced phenomena. To ground the large deviation theory, consider a generic stochastic differential equation

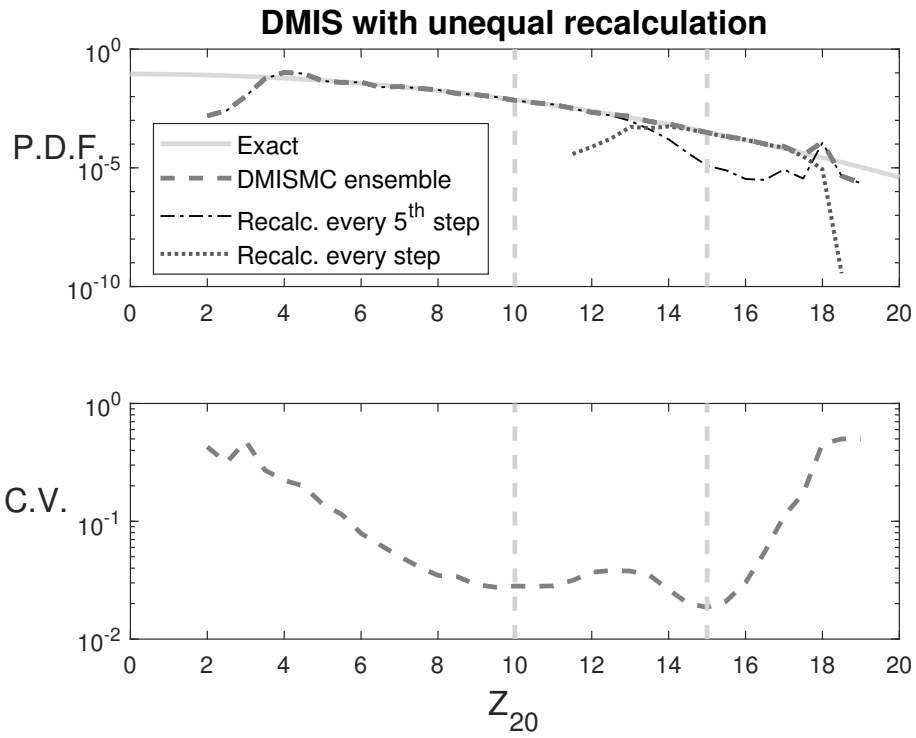

Figure 4.5. The 1D Gaussian random walk from Eq. (3.14) with 20 steps is simulated with two biasing distributions: one targeting $Z_{20} = 10$ that recalculates the path every 5 steps and one targeting $Z_{20} = 15$ with path recalculation every step. The individual contributions of each distribution to the DMIS estimate are shown with thinner curves. These contributions as well as the general pattern of the C.V. in the bottom panel show that samples are more tightly clustered, and convergence is improved, for the distribution with more recalculations.

(SDE) of the form

$$dX = b(X,t)dt + \varepsilon\, dW. \tag{4.7}$$

The SDE describes the evolution of a random variable X that is subject to both deterministic and stochastic dynamics. The first term on the right-hand-side captures the deterministic dynamics and the second term describes the stochastic dynamics where dW

is a noise process and ε is the small constant noise power. This noise process is continuous, and can be taken to be a Wiener process for our purposes. In the low noise limit ($\varepsilon \to 0$) the asymptotics governing the dynamics of Eq. (4.7) have been worked out in extensive detail by Freidlin and Wentzell [**38**].

When considering an exit problem (i.e from a domain $\mathcal{D}$), the path that dominates all paths that exit the domain is the minimizer of the Freidlin-Wentzell (F-W) least action functional

$$S_L = \min_{\phi_L \notin \mathcal{D}} \frac{1}{2} \int_0^L \left|\left| \frac{d\phi}{dt} - b(\phi, t) \right|\right|^2 dt. \tag{4.8}$$

This path-based functional can be thought of as minimizing the norm of the noise process as solving Eq. (4.7) for dW leaves $dW = (dX - b(X,t)dt)/\varepsilon$. In the literature, many methods consider time-minimization as well

$$S_\infty = \min_L S_L, \tag{4.9}$$

giving rise to the so-called "quasi-potential" which provides the basis for calculating exit paths in a variety of contexts [**90**]. However, these methods are outside the scope of this thesis, and do not provide information on how the optimal exit path changes as the propagation distance varies, which will prove interesting in Chapter 5.

Now, this theory is asymptotic in that it applies in the low-noise limit. In this limit, Wentzell and Freidlin have developed a beautiful and comprehensive theory of the least-action functional and its properties. For instance, they have proven it to be unique and applicable to exit problems in the infinitesimal noise limit [**38**]. In sampling problems, however, the noise power is not infinitesimal, and thus the appropriateness of low-noise

asymptotics is a matter of debate. There are finite-noise tools that have been developed in specialized conditions, such as the Onsager-Machlup functional [**31**]. This method amounts to the F-W (Freidlin-Wentzell) functional with the addition of another term proportional to the noise power. In small noise situations, the F-W functional has proven sufficient to capture rare-event paths when importance sampling is used to calculate the actual probabilities [**73**].Therefore, the additional computational work of the extra term is not necessary for the applications considered here.

In addition to the finite-noise question, the presence of multiple paths leads to another issue with the F-W theory. To understand this issue, it first must be noted that the Freidlin-Wentzell action also provides *a priori* exit probability estimates for an exit path X_t computed from the theory via the asymptotic relation

$$\log[P\{X_t\}] \sim -\varepsilon^{-2} S_L. \tag{4.10}$$

This logarithmic estimate provides a framework for assessing situations in which there are multiple exit paths (i.e. local minima in the minimization problem in Equation 4.8). For example, in the wall walk model, we know that there are two exit paths from the origin (over and under the wall) that should be equally likely based on symmetry. Computing the action of these two paths shows, in fact, that the paths are equally likely as they have the same asymptotic probabilities, but large deviation theory can also assess when we are examining exit paths from locations other than the origin (e.g., the other two paths depicted in Fig. 4.2), which is an inherent issue in implementing dynamic importance sampling in that problem.

If we are perturbed from a starting point with $y = 0$ (i.e. the halfway point between the ends of the wall), the equality of the actions associated with the different paths is broken, but the action estimate from Eq. (4.8) can still be used to estimate their relative likelihoods. For instance, Fig. 4.6 shows contours of constant relative probability, computed using Eq. (4.10), that the walker's exit path (path to $x = 20$) goes above the wall, with the number of steps left in the random walk held constant. The perfectly flat line from the halfway point of the wall reflects this symmetry, for instance. It is important to note that in simulation, we should expect locations further from the starting point of the random walk to have taken more steps to get there, and thus have less steps left to get to their exit destination. Due to the asymptotic nature of the action estimate, however, we cannot draw precise distinctions between locations that have relative probabilities with the same order of magnitude, so this figure is meant purely for illustration purposes.

4.4. Dynamic Multiple Importance Sampling with Multiple Paths

Now, with the requisite large deviation theory having been discussed, we can consider the class of problems for which dynamic multiple importance sampling is useful by returning to the two dimensional Gaussian random walk past a wall obstacle (from Section 4.1). In this problem, we naturally have to contend with two paths: over- and under-wall paths. The results of Section 4.3 demonstrate that proper convergence in ISMC simulations is only achievable when both paths are used. In order for convergence to be improved from this MIS baseline, DMIS must be implemented for tighter clustering of samples around the intended exit target of $x_{60} = 20$. In this sense, this is a more proper application of the DMIS algorithm than its use in the 1D random walks of Section 4.2.

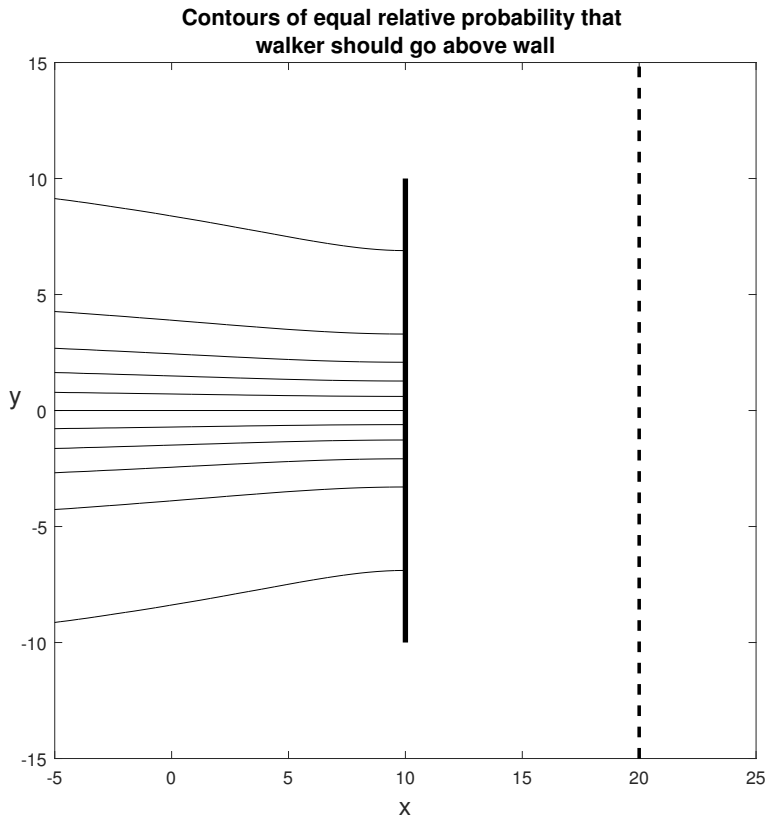

Figure 4.6. Contours of constant relative probability that the walker should go up over the wall when recalculating the path from that point. The contours, from bottom to top, represent $1\%, 10\%, 20\%, \ldots, 80\%, 90\%$, and 99% relative probability. The number of steps left was kept constant at $N = 60$ for all recalculations.

However, the multi-path nature of this problem brings up a number of further complications that must be discussed. For instance, from the geometry of the paths shown in Fig. 4.2 it is clear that as soon as a path crosses the position of the wall, the number of different paths to exit decreases to one path. Really, it is more proper to say that once a path crosses the vertical ends of the wall (i.e. $y = \pm 10$) the number of paths is one, just a rightward path to $x = 20$. This presages a concern that will be more relevant in Chapter 5 that paths sometimes disappear, and this will be discussed in more detail in Section

4.4.1 where more complicated dynamics lead to paths disappearing in far less predictable ways than those discussed here in this toy model. The opposite issue, where new paths arise mid-simulation, will also be mentioned in passing in Section 4.4.2, though this issue is far more complicated and is largely the provenance of future work outside the scope of this thesis.

Before getting to such complicating factors, let us apply DMIS to our 2D random walk in two different ways, and compare them to the methods mentioned previously: MC, MIS, and DIS. Results of such an application are shown in Fig. 4.7. The first DMIS method (called DMISMC in the figure and DMIS-both here) recalculates both the over-wall and under-wall paths every three steps for all samples (half the samples are "actively" biased over the wall and half under). In other words, it uses dynamic recalculation on both paths. The second DMIS method (denoted DMISMC over-wall in the figure and DMIS-over here) just recalculates the over-wall path every three steps for all samples. This second method still biases half of the samples above and below the wall, but only the over-wall path is recalculated for all samples in an asymmetric manner. This is similar to the simulations from Fig. 4.5 where the biasing distributions have different recalculation criteria, but these criteria are applied consistently across all samples. The comparison with the previous methods shows that the two DMIS methods provide convergent results, but DMIS-over has a slower convergence rate than DMIS-both. DMIS-over provides convergence similar to MIS. This is unsurprising as the distribution using the below-wall path, which is not recalculated, is exactly the same as the one used in the MIS simulations. The convergence benefits from recalculation of the over-wall path are negated by its inclusion. The DMIS-both method, by contrast, has a lower C.V. indicating improved convergence over MIS.

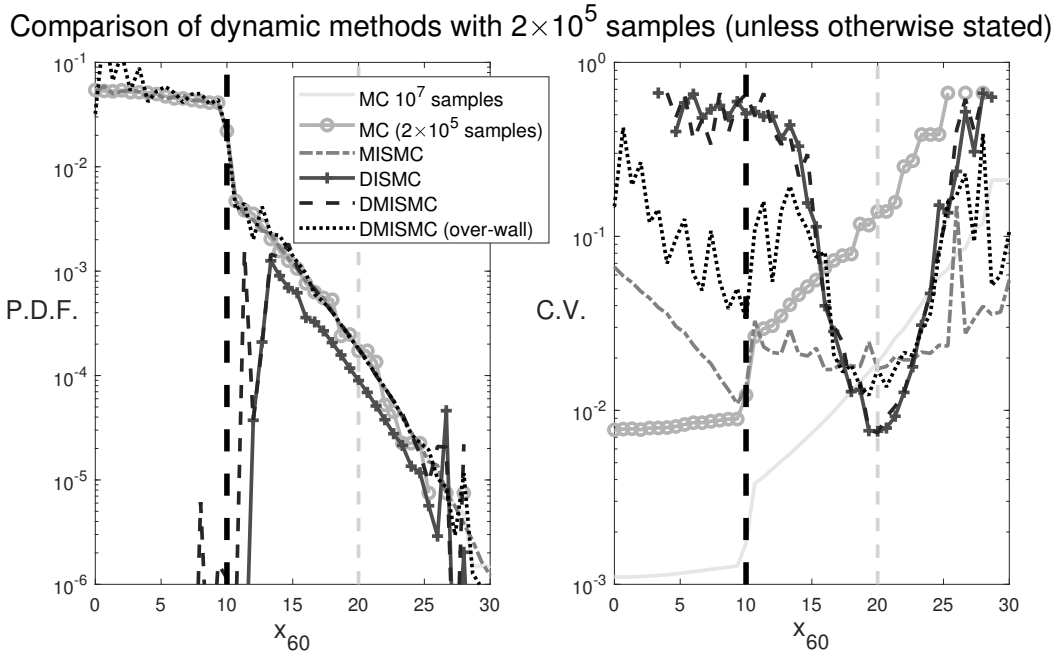

Figure 4.7. A comparison of different varieties of dynamic importance sampling for the wall walk model to the MISMC simulation from Fig. 4.3. Three methods using dynamic recalculation are compared: using only the above wall path (DISMC), using both paths with only the over-wall path recalculated (DMISMC over-wall), and using both paths with both paths recalculated (DMISMC). All recalculations occurred every 3 steps of the random walk. The DISMC converges to half of the correct probability, while the DMISMC over-wall simulation has convergence comparable to the MISMC simulation with no recalculation. Only the DMISMC simulation with both paths recalculated has improved convergence and correct probability estimation.

Additionally, DIS (using only the over-wall path without MIS) produces a C.V. that is roughly equivalent to DMIS, but it converges to the wrong probability: roughly half of the correct probability throughout the region of targeting. This is unsurprising as the results of Fig. 4.3 indicated that both over-wall and under-wall paths were necessary.

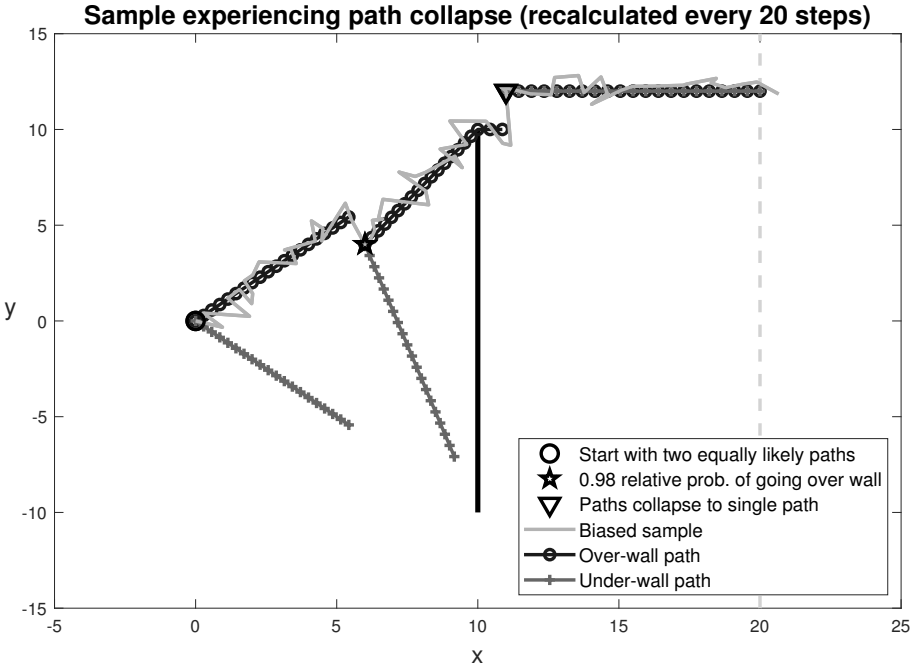

Figure 4.8. A sample of the random walk past a wall with sixty steps simulated with DMIS used to bias the sample over the wall. Both the over wall path and the under wall paths are recalculated every twenty steps. By the first recalculation, the relative probability of the walker going over the wall according to Eq. (4.10), is 98%. At the second recalculation the walker has passed the wall and the only path available is straight to the target of $x_{60} = 20$.

4.4.1. Path disappearance/collapse

A salient feature of this problem is that both possible paths collapse to the same path once the walker passes the wall's penumbra. An illustration of this situation is depicted in Fig. 4.8 where a sample is biased to go over the wall, with each path recalculated twice (every twenty steps in a sixty-step walk). By the second recalculation, the walker has passed the wall and both paths are the same. It is important to note that the situations in which the number of paths decreases is not as easily distinguished in many cases as it is in this 2D random walk. The multiple-path phenomenon can be conceptualized as multiple local

minima in a global landscape of paths that have similarly low action. However, in more complicated examples than this simple random walk the path landscape can be expected to be substantially more complicated. In fact, the path landscape for the laser model described in Section 5.2 will be found to be quite complex. It will be much more difficult to partition state space *a priori* into regions based on the number of different paths, as we can do here. This question is still poorly understood, as both the notion of similarly low action paths and the asymptotic nature of the underlying theory are inexact.

To demonstrate the ambiguity still present in the theory for this phenomenon, simulations were performed where the relative probability of each path was tracked at each recalculation point, and when it fell below a prescribed tolerance it was handled in two different fashions. In one method, the unlikely path was overwritten with the dominant path, and in the second the unlikely path was kept as in previous simulations. The first method treats the paths as if the unlikelier path has collapsed into the dominant path. Results of these simulations are shown in Fig. 4.9, where two different tolerances were used with recalculation performed at every step in the path. In the lower tolerance (relative probability of 0.01) simulations, both methods perform comparably well, and are better than the baseline convergence of MIS. However, in the higher tolerance simulations the quasi-path-collapse method performs much more poorly than the default DMIS method, as the spike in the C.V. in the bottom right panel near $x_{60} = 20$ indicates (even a single spike near the target can indicate poor convergence in an IS simulation as we expect samples landing near the target to have similar likelihoods if the biasing is determined properly [**38**]). This suggests that missing an unlikely path, and assuming that it has collapsed into the dominant path, is not problematic up to a point. Missing paths that

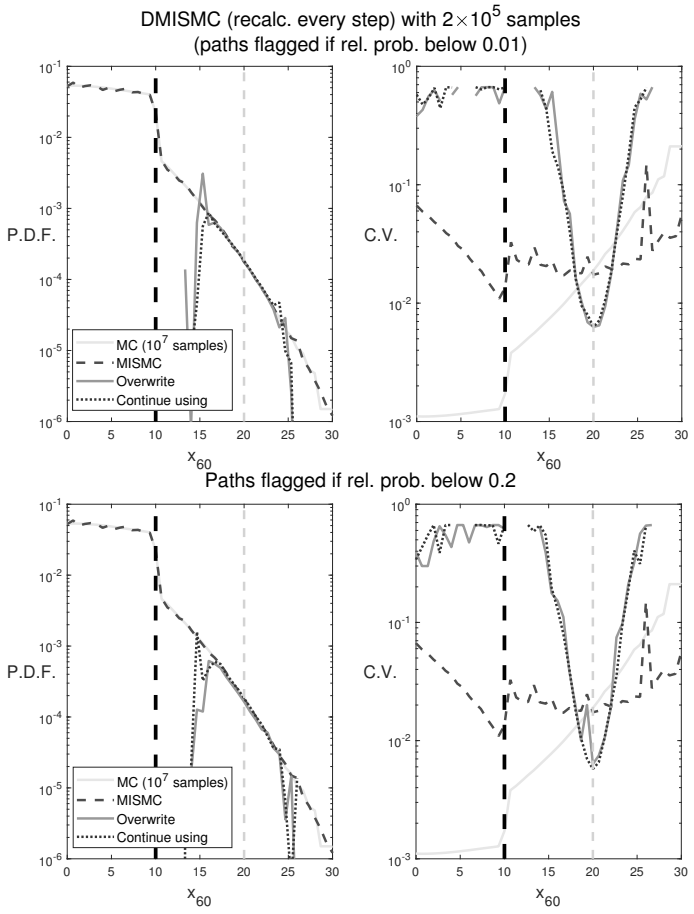

Figure 4.9. A comparison of DMIS methods (with every step recalculation) for the wall-walk problem where unlikely paths, measured using Eq. (4.10), are handled in two different ways and are compared with MISMC (dashed) and MC (light gray) simulations for accuracy and convergence. When the relative probability of a path is below a certain tolerance the dominant path overwrites that path (gray), or we continue to use that path as in Fig. 4.7 (black-dotted). Two different tolerances are presented, a tolerance of 0.01 in the top panels where both methods produce similar results that are far better than MISMC, and a tolerance of 0.2 in the bottom panels where overwriting produces a spike in the C.V. For the low tolerance simulations, roughly 16 steps of the random walk are below the tolerance on average, whereas 26 are below the higher tolerance, which gives an indication of the different convergence properties between the two simulations.

have a high enough likelihood can limit simulation convergence, however. These results are simply illustrative, and it is likely that in future systems with multipath phenomena the importance of tracking path collapses for simulation convergence will vary greatly.

4.4.2. Further considerations

In this section we will briefly mention qualitative behaviors of multi-path systems that we have not encountered in this simple random walk, but that may pose additional problems if they occur sufficiently frequently in future systems. In the previous section, we discussed many qualitative issues with paths disappearing during simulation, but it may be possible for paths to appear during simulation. One way for this to happen is illustrated in Fig. 4.10. In this hypothetical scenario, the walker rapidly moves to a location where both paths collapse to a single path, then back into a location with multiple different paths. We can loosely describe this as a path "reoccurring." Such a situation did not happen in any of the simulations of the random walk, and consideration of the large rapid displacements necessary for such a phenomenon to occur indicates why it is quite unlikely, but this does not mean that it is impossible for similar phenomena to occur in other systems. Additionally, in this 2D random walk it is trivial to discover when the second path reoccurs as it is simply a matter of the walker's y-position, but this may not be the case in other systems.

An even more challenging version of this phenomenon cannot be conceived of in the random walk wall obstacle as currently constructed, but can be thought of as a path appearing that is not possible at the outset of the dynamics. We can conceive of this in the wall-walk setup by adjusting the starting position of the walker, however. Consider if

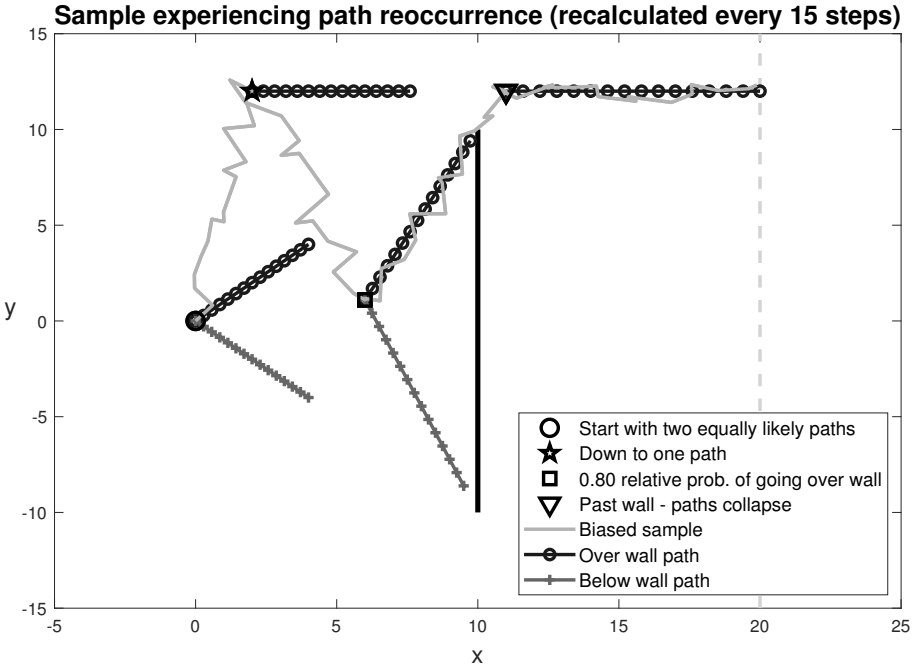

Figure 4.10. A sample of the random walk past a wall with 60 steps simulated with DMIS used to bias the sample over the wall, with the paths recalculated every 15 steps. In this hypothetical scenario, due to rapid vertical movement early on, the walker reaches a location at the first recalculation where there is only one path. However, due to rapid vertical movement thereafter, the walker has two viable paths. This example trajectory demonstrates path reoccurrence, although did not happen naturally in previous simulations.

the initial position of the walker was moved upwards, say from $y = 0$ to $y = 10$. Then, the path over the wall would have a far lower action than the under-wall path and it would not make sense to include under-wall paths in an MIS simulation. However, it may be the case that enough samples would stray downward to the point where a non-negligible number of samples would have a viable path below the wall. In this hypothetical scenario, it may be necessary to incorporate information about below-wall paths in the overall simulation. A qualitatively similar scenario is encountered in the laser model and is discussed in Section 5.4. In that case it is also a largely hypothetical phenomenon, but in systems where path

appearance happens sufficiently frequently (which may not need to be very frequently) it could pose a significant problem. The most challenging aspect of path appearance is that it might require acquiring a global sense of the path landscape in potentially a high number of dimensions. The landscape of the laser model is explored in the next chapter, for instance, at a high computational cost using continuation methods. It is also obviously the case that the path landscape can be arbitrarily complicated in future systems adding to the overall difficulty.

Lastly, we consider some algorithmic augmentations and alternatives that may improve on the performance of DMIS in certain circumstances. One hypothetical augmenation might be variable likelihood tracking. By this, I mean truncating (or potentially inserting) likelihood ratios into samples that are shorter than the overall simulation time, based on a path that only exists for part of the sample's trajectory. This is potentially possible as the likelihood ratio for a sample is the product of the likelihood ratio at each step, so multiple paths need not have the same number of steps to be combined via the balance heuristic. This provides a somewhat straightforward alternative for path collapse and reappearance situations, as paths can be truncated when they disappear, and concatenated to previous paths when they reappear. However, this does not ameliorate the issue of path appearance as the number of relevant paths may be different across samples, and again there would need to be *a priori* knowledge about the number of possible paths as they may not emerge during simulation. There are a number of importance sampling strategies that take a more empirical approach to path discovery by discovering paths during MC simulation and adjusting or creating biasing distributions based on results [**16, 53, 75**]. The limitation with this approach is that samples of interest (i.e. samples that both reach the target

condition but also pass through a region of multiple paths) would likely be far too rare to be discovered without further knowledge. The researcher might have to see paths take qualitatively different paths, for instance, which would involve saving all trajectories in simulations that may require billions of samples to get even two hits and involve untenably expensive computational outlay.

An alternative to MIS that has been studied in the literature is mixture importance sampling where multiple proposal densities are weighted and selected in a single simulation, rather than in multiple batch simulations of each distribution [**75**]. The key is that these distributions still must be weighted in some way, with the balance heuristic being just one of a class of possible "deterministic weights" that are devised before simulations begin. Similar to the discussion in the previous paragraph, there are methods that take an emergent approach to distributional weights, and are broadly described as adaptive schemes [**18**]. Both approaches would struggle to deal with the path appearance phenomenon, as it changes the number of distributions required, rather than just the weighting between distributions.

We have demonstrated that mixing multiple importance sampling with dynamic importance sampling has utility when applied to a problem with multiple routes to a rare event of interest. The combination of the two methods performs better than either of the two used in isolation. The application of dynamic multiple importance sampling also highlights important features of the dynamics necessary for achieving a large displacement past the wall, and foreshadows issues of path collapse and appearance that occur in the laser model and may occur in other systems. To now see how DMIS may be useful in

more complicated scenarios and explore the details further, we will analyze a model of an actively mode-locked laser in the next chapter of this thesis.

CHAPTER 5

A Mode-Locked Laser Model

The previous chapters in this thesis have provided the groundwork for the study of a mode-locked laser model. Mode-locked lasers are a technology with a rich history of application and study [**51, 65, 66**]. In recent years the breadth of applications employing such systems has grown dramatically, and simultaneously the performance of the underlying lasers has greatly improved. Optical communications is an oft-cited application [**41**], but other uses include optical frequency metrology [**21**], optical clocks [**13, 24, 89, 95**], spectroscopy [**87**], generation of high harmonics [**57**], measurement of fundamental constants [**35**], and optical storage rings [**46**].

Here, ISMC simulations using biasing distributions obtained using soliton perturbation theory (see Section 2.3) are used to describe "position slip" errors and determine the rates at which these slips occur in a mode-locked laser model involving an active feedback mode-locking mechanism [**26, 71**]. The active feedback mechanism can lead to two fundamentally different mode-locking behaviors: a non-oscillatory overdamped regime, and an oscillatory underdamped locking regime. Position slip errors, or errors of pulse position slippage relative to the mode-locking, are found to occur in both cases, but with qualitatively and quantitatively different behaviors. We will show that the importance sampling methods needed to capture error rates in the two cases are somewhat different, and that more sophistication is needed in the underdamped regime to deal with the more complicated paths arising due to the oscillations. In particular, we will apply the dynamic

multiple importance sampling method introduced in Chapter 4 to quantify error rates in this regime when multiple oscillatory paths to errors exist. We believe that this problem illustrates a general issue that may occur in systems where multiple routes to rare events may be present.

5.1. Model Formulation

We model an actively mode-locked fiber laser as an optical cavity oscillator including an amplifier, filter, polarization rotator (and polarizer), and phase modulator [**26, 71**]. This system is represented schematically in Fig. 5.1. Propagation through the optical fiber is described by the dimensionless nonlinear Schrödinger equation for the single polarization optical field envelope u [**73**]

$$\frac{\partial u}{\partial z} - \frac{i}{2}\frac{\partial^2 u}{\partial t^2} - i|u|^2 u = F(z,t), \tag{5.1}$$

with the right-hand-side representing the perturbative effects of the added physical elements used to provide the active mode-locking and stabilize the optical pulses. This governing equation is essentially the same as Eq. (2.22) with the perturbation specified to the physical elements of the active mode-locking.

The mode-locking elements are (linear) gain from the amplifier and polarization rotator, filtering, nonlinear gain (or loss) from the polarization rotator, and phase modulation [**65**]. A simple version of the filtering due to frequency dependent gain (or loss) can be described by the perturbing term $a\,\partial^2 u/\partial t^2$, where a is the filtering strength. A simplified

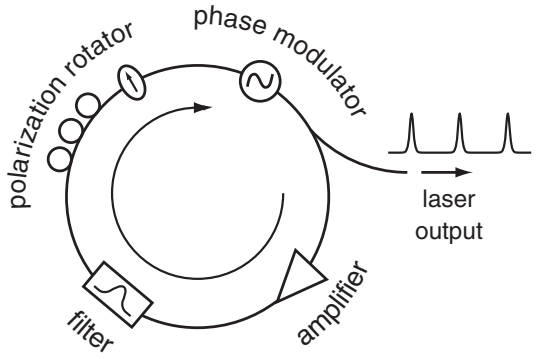

Figure 5.1. A schematic of a fiber ring laser where an optical pulse passes through an amplifier, filter, polarization rotator, and phase modulator on each pass around the ring.

model of excess linear and nonlinear gain can be described by terms of the form [**65**]

$$c_1 u + c_2 |u|^2 u + c_3 |u|^4 u.$$

Typically c_1 will be negative, so really this term models a small excess linear loss. In addition, a simplified model of the active phase modulation is a term of the form $ib\cos(\omega t)u$ [**71**], where b is the modulation strength and the external modulation frequency is ω. In what follows we will assume $\omega = 2\pi/25$, i.e., a 25 ps modulation period. It should be noted that, in practice, all of the above effects are discrete or lumped, i.e., the perturbations are each applied to the pulse once per pass through the loop. Since the net change to the pulse per pass is small, it is permissible to replace the lumped effects with their averaged continuous versions as above [**60**].

In addition, the linear gain provided by the amplifier is also accompanied by amplified spontaneous emission noise as described in Section 2.1.2. This additional perturbation is modeled as zero-mean, delta-correlated Gaussian white noise that is added to the pulse

as it passes through the amplifier on every round trip through the fiber ring [**73**], and takes the form

$$\sum_{n=0}^{N} f_n(t)\delta(z-n), \text{ where} \tag{5.2a}$$

$$\langle f_i(t)\rangle = 0, \text{ and} \tag{5.2b}$$

$$\langle f_i(t) f_j^*(t')\rangle = \sigma^2 \delta(t-t')\delta_{ij}\,. \tag{5.2c}$$

For simplicity we have assumed that the noise is added once per dispersion length; if the fiber ring is shorter than this, we can combine the noise from multiple passes into a single larger perturbation once per dispersion length, as long as the total remains small. The noise variance is then $\sigma^2 = [(G-1)^2 \eta_{sp} \hbar\omega_0 T_0 \gamma]/[G \ln G |\beta|] = 2.667 \times 10^{-5}$ [**73**], where we have used $G = 40$ (or, 16 dB) as the amplifier gain needed to compensate loss due to the components in the loop, $\eta_{sp} = 2.0$ is the amplified spontaneous emission (excess noise) factor, $\hbar$ is Planck's constant, $\omega_0 = 1.22 \times 10^{15}$ 1/s is the carrier frequency, $T_0 = 1$ ps, and $\beta = 0.2$ ps^2/km. Thus, the full perturbative term F on the right-hand-side of Eq. (5.1) is

$$F(z,t) = a\frac{\partial^2 u}{\partial t^2} + ib\cos(\omega t)u + c_1 u + c_2|u|^2 u + c_3|u|^4 u + \sum_{n=0}^{N} f_n(t)\delta(z-n). \tag{5.3}$$

When $F \equiv 0$ the NLS admits the well-known soliton solution of Eq. (2.18).

5.1.1. Soliton perturbation theory and representative parameter regimes

We first consider Eq. (5.1) with the deterministic perturbations in Eq. (5.3) (i.e., without the noise). Assuming the initial solution is a soliton and the perturbative terms are small in magnitude, then the leading-order effect is to cause the soliton parameters to change

slowly. Using soliton perturbation theory [**3, 47**], the evolution equations for the soliton parameters are given in Eqs. (2.23-2.25) (ignoring the phase as this model is phase-insensitive). Using Eq. (2.18) for u and to evaluate the integrals to leading order, we obtain

$$\frac{dE}{dz} = (2c_1 - 2a\Omega^2)E + \left(\frac{4}{3}c_2 - \frac{2}{3}a\right)E^3 + \frac{16}{15}c_3E^5, \tag{5.4}$$

$$\frac{d\Omega}{dz} = -\frac{4}{3}aE^2\Omega - \frac{\omega^2 b\pi}{2E}\mathrm{csch}(\frac{\pi\omega}{2E})\sin(\omega T), \quad \text{and} \tag{5.5}$$

$$\frac{dT}{dz} = \Omega. \tag{5.6}$$

First of all, it is seen that for $c_1 < 0$ (excess loss), $E = 0$ is stable. Strictly speaking, soliton perturbation theory is not applicable in this limit, but a linear stability analysis of the full NLS equation shows that $u = 0$ is stable due to the linear excess damping (the c_1 term) and filtering. For $c_2 > 0$ and $c_3 < 0$ there will be a positive stable solution of Eq. (5.4) for E when $\Omega = 0$, which we shall denote E_s. Because the filtering coefficient a is typically small, the pulse energy will be more or less independent of Ω if it is not too large. Since there are two stable pulse energies ($E = 0$ and $E = E_s$) it is possible for added noise to induce transitions between the two, leading to either pulse dropouts or the spontaneous creation of new pulses. For the parameters we will use, however, the probability of either event will be very small [**26**].

For $E = E_s$, the dynamics of Ω and T can be underdamped or overdamped [**26, 71**]. In both cases there are stable steady states at $(E, \Omega, T) = (E_s, 0, 2n\pi/\omega)$ and unstable saddles at $(E, \Omega, T) = (E_s, 0, (2n+1)\pi/\omega)$ where $n \in \mathbb{Z}$, but the structure of the phase space is different in the two cases. Phase planes for Ω and T with E constant are

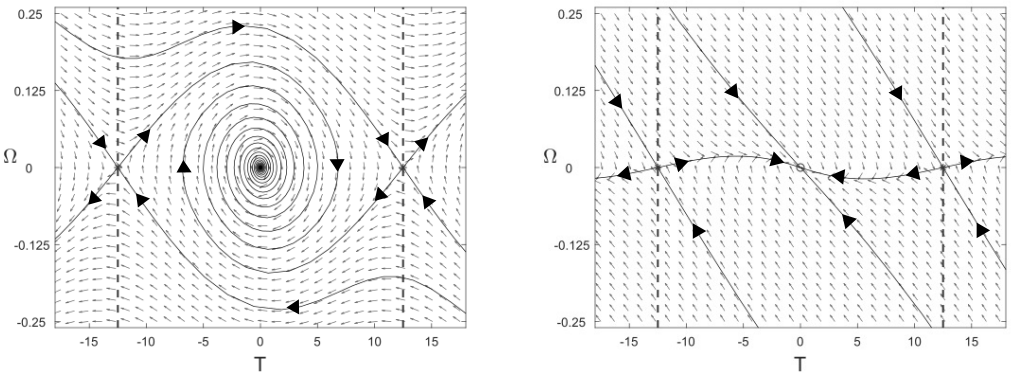

Figure 5.2. The continuous dynamics of the soliton ODE system Eqs. (5.5) and (5.6) are shown in (T, Ω) phase space. Left: The underdamped state with a stable spiral fixed point at $(0, 0)$ and saddles $(\pm\pi/\omega, 0)$ (at bit-slot boundaries, depicted with dashed vertical lines, as defined by the modulation frequency). For a pulse to move out of its bit-slot, it must escape the basin of attraction of a stable spiral and cross the stable manifold of the saddle point. Right: The overdamped state where the stable fixed point at $(0, 0)$ is now a nodal sink rather than a spiral. The parameters for the filtering, modulation, nonlinear gain strengths, and steady-state amplitude E_s in each are the representative values listed in the text.

shown in Fig. 5.2. In the underdamped case, decay to the non-zero equilibrium state is oscillatory, while in the overdamped case a soliton displaced from equilibrium will be restored smoothly to the steady state. The relative strengths of filtering and modulation determine the type of dynamics, with greater filtering leading to overdamping [**71**]. We will use two representative parameter sets. For the underdamped regime, we will assume $a = 0.002, b = 0.01, c_1 = -0.01, c_2 = 0.034, c_3 = -0.02$ with a stable amplitude of $E_s = 1.177$. For the overdamped regime we assume $a = 0.015, b = 0.002, c_1 = -0.01, c_2 = 0.04, c_3 = -0.02$ with a stable amplitude of $E_s = 1.150$.

One effect of the noise is that it adds jitter to the pulse parameters, broadening the laser's linewidth [**1**]. A more dynamically interesting event is for a noise-induced *position*

slip error to occur [**71**]. Here the noise induces the pulse position to transition from one stable equilibrium to another. Since the goal of mode-locking is to keep a pulse synchronized to the active mode-locking signal, noise-induced sliding of a pulse's position is undesirable as it leads to a random drift of the mode-locked laser's repetition rate. The rate at which these position slips occur, of course, is directly related to the rate of this random drift. Because the parameters being considered are such that pulse energy dropouts are very improbable relative to position slips, in what follows we will assume the amplitude to be fixed at $E = E_s$ and only concern ourselves with position slips. We monitored the pulse energy in simulations to verify that this assumption holds.

5.2. Biasing for Position Slip Errors

When considered as the sole perturbation, the effect of the ASE noise is to cause random walks in the soliton parameters as discussed in Section 2.4. Since any values of these parameters produce perfectly valid solutions of the NLS equation [**73**], it is unable to resist such random changes. The random walks that result in the absence of the deterministic perturbations described in Section 5.1 eventually lead to large deviations in the soliton parameters even if the noise added at a single amplifier is small. The purpose of the mode-locking terms, of course, is to limit the growth of perturbations in the pulse parameters. Even in this case, however, there is a small but non-zero probability for the pulse to experience a large deviation.

To address the question of precisely how likely a position slip is to occur in the presence of mode-locking, we pose a stochastic exit problem. Specifically, we wish to determine the most likely way for a pulse starting at equilibrium ($\Omega = 0$, $T = 0$) to escape the effective

potential well imposed by the mode-locking and exit a bit-slot by undergoing a position drift of $\pi/\omega = 12.5$. (By symmetry, an exit to $-\pi/\omega = -12.5$ is equally probable.) To answer this question, we first will derive optimized exit paths using the approximating system of ODEs for soliton parameter evolution, Eqs. (5.4)-(5.6).

We also use the soliton perturbation theory of Sections 2.3.2 and 2.3.3 to describe the noise process' effects that lead to position slips via its effects on the frequency and position of the soliton. The relevant results from those sections are summarized as follows. Taking F in Eq. (5.1) to be the Gaussian white noise terms (ignoring the deterministic perturbations for the moment, and neglecting changes in E), we find at leading-order that the noise-induced parameter jumps are

$$\Delta\Omega_n = \mathrm{Re}\int_{-\infty}^{\infty} \underline{v}_\Omega^* e^{-i\Theta} f_n \, dt \quad \text{and} \tag{5.7}$$

$$\Delta T_n = \mathrm{Re}\int_{-\infty}^{\infty} \underline{v}_T^* e^{-i\Theta} f_n \, dt. \tag{5.8}$$

with the relevant adjoint soliton modes given by

$$\underline{v}_\Omega = -\frac{i}{E}\frac{\partial u_0}{\partial t} \quad \text{and} \tag{5.9a}$$

$$\underline{v}_T = \frac{1}{E}(t-T)u_0. \tag{5.9b}$$

When the noise f_n is given by Eq. (5.2), $\Delta\Omega_n$ and ΔT_n are zero-mean Gaussian random variables with variances [**73**]

$$\mathbf{E}[\Delta\Omega_n^2] = \sigma_\Omega^2 = \sigma^2 E/3 \quad \text{and} \tag{5.10}$$

$$\mathbf{E}[\Delta T_n^2] = \sigma_T^2 = \pi^2\sigma^2/(12E^3)\,. \tag{5.11}$$

5.2.1. Derivation of Optimal Exit Paths

Equations (5.5) and (5.6) with added random jumps given by Eqs. (5.7) and (5.8) form an approximate version of the soliton dynamics in the full model, Eq. (5.1) with Eq. (5.3). This system is, letting $A = -4aE^2/3$ and $B = -\omega^2 b\pi\mathrm{csch}(\pi\omega/2E)/(2E)$, explicitly given by

$$\frac{d\Omega}{dz} = A\Omega + B\sin(\omega T) + \sum_{n=0}^{N}\Delta\Omega_n\delta(z-n) \quad \text{and} \tag{5.12a}$$

$$\frac{dT}{dz} = \Omega + \sum_{n=0}^{N}\Delta T_n\delta(z-n)\,. \tag{5.12b}$$

In particular, these equations approximately describe the dynamics of a position slip error, and allow us to derive exit paths for such an event. The most probable series of parameter kicks $\Delta\Omega_n$ and ΔT_n are those which have the lowest combined weighted L_2 norm, as the underlying noise is Gaussian. Let the functions $\eta_\Omega(z)$ and $\eta_T(z)$ be continuous functions that stand-in for the discrete parameter kicks (this continuum approximation is justified as the length scale on which the deterministic dynamics work in the system is much longer than the amplifier spacing, here $\Delta z = 1$). Then the constrained optimization problem

needed to be solved for an optimal exit path is to minimize [**72, 73, 83**]

$$\min_{\eta_\Omega,\eta_T} S = \min_{\eta_\Omega,\eta_T} \frac{2}{\sigma^2} \int_0^{z_L} C\eta_\Omega^2 + D\eta_T^2 \, dz \tag{5.13}$$

subject to

$$\frac{d\Omega}{dz} = A\Omega + B\sin(\omega T) + \eta_\Omega, \quad \text{and} \tag{5.14a}$$

$$\frac{dT}{dz} = \Omega + \eta_T \tag{5.14b}$$

where $C = 1/\sigma_\Omega^2 = 3/(E\sigma^2)$ and $D = 1/\sigma_T^2 = 12E^3/(\pi^2\sigma^2)$. We want the final soliton position to be $T(z_L) = \hat{T}$ (the one that gives a position slip), so we have three boundary conditions: $T(0) = T_0$, $\Omega(0) = \Omega_0$, and $T(z_L) = \hat{T}$. Here the idea is to maximize the probability associated with a particular set of deviations. This is equivalent to finding the Freidlin-Wentzell least action path [**38**] (as elucidated in Section 5.3) and related methods [**7, 37, 80**]. Following standard variational calculus methods for optimization with differential side constraints, we construct a constrained functional using Lagrange multipliers [**40**]

(5.15)
$$\int_0^{z_L} \left\{ C\eta_\Omega^2 + D\eta_T^2 + \lambda_1(z)\left[\frac{d\Omega}{dz} - A\Omega - B\sin(\omega T) - \eta_\Omega\right] + \lambda_2(z)\left[\frac{dT}{dz} - \Omega - \eta_T\right] \right\} dz.$$

Taking variations with respect to the unknown functions $\eta_\Omega, \eta_T, \Omega$, and T, we find

$$\delta\eta_\Omega \to 2C\eta_\Omega - \lambda_1 = 0, \tag{5.16a}$$

$$\delta\eta_T \to 2D\eta_T - \lambda_2 = 0, \tag{5.16b}$$

$$\delta\Omega \to -A\lambda_1 - \lambda_2 - \frac{d\lambda_1}{dz} = 0, \quad \text{and} \tag{5.16c}$$

$$\delta T \to -B\omega\lambda_1 \cos(\omega T) - \frac{d\lambda_2}{dz} = 0\,. \tag{5.16d}$$

We use Eqs. (5.16a) and (5.16b) to specify the Lagrange multipliers $\lambda_1 = 2C\eta_\Omega$ and $\lambda_2 = 2D\eta_T$ giving

$$\frac{d\eta_\Omega}{dz} = -A\eta_\Omega - \frac{D}{C}\eta_T \quad \text{and} \tag{5.17a}$$

$$\frac{d\eta_T}{dz} = -\frac{B\omega C}{D}\eta_\Omega \cos(\omega T)\,, \tag{5.17b}$$

which along with Eqs. (5.14a) and (5.14b) form a boundary value problem with four boundary conditions: $T(0) = T_0$, $\Omega(0) = \Omega_0$, $T(z_L) = \hat{T}$, and $\eta_\Omega(z_L) = 0$. The last condition is a natural boundary condition arising from the free boundary for $\Omega(z_L)$ [**40**].

5.2.2. Exit Path Behaviors

The solution of the boundary value problem given by Eqs. (5.14) and (5.17) with $T_0 = 0$, $\Omega_0 = 0$, and $\hat{T} = 12.5$ gives an optimal exit path after propagation through $\lfloor z_L \rfloor$ amplifiers. This boundary value problem is solved numerically with MATLAB's BVP4C package. Representative numerical solutions are shown in Fig. 5.3 and 5.4 for the overdamped and underdamped regimes, respectively. The exit paths for the overdamped regime have

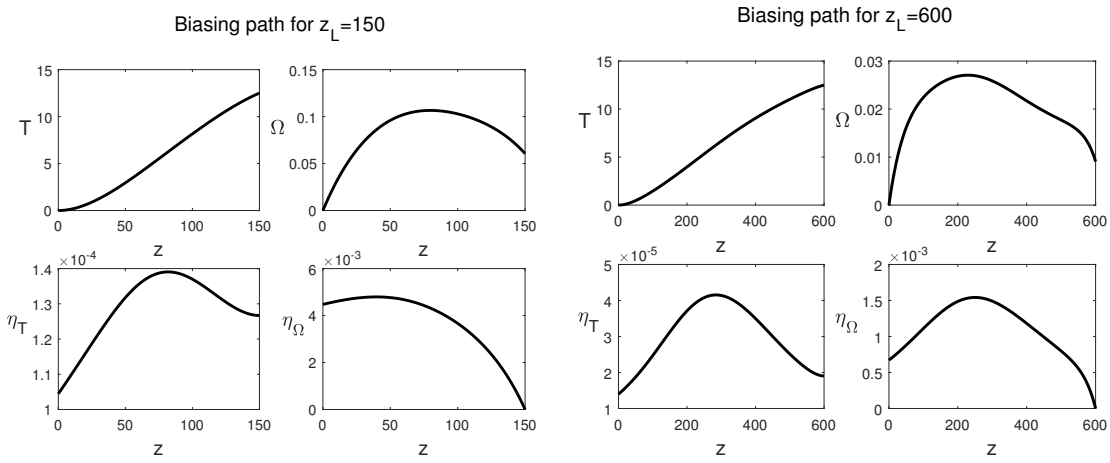

Figure 5.3. Two representative solutions for the exit path in the overdamped parameter regime showing four-part solutions with different propagation distances, z_L. The lower panels show the biasing path coefficients and the the upper panels show the expected soliton parameter trajectories under biasing. The solutions were computed numerically using the MATLAB boundary value problem solver BVP4C.

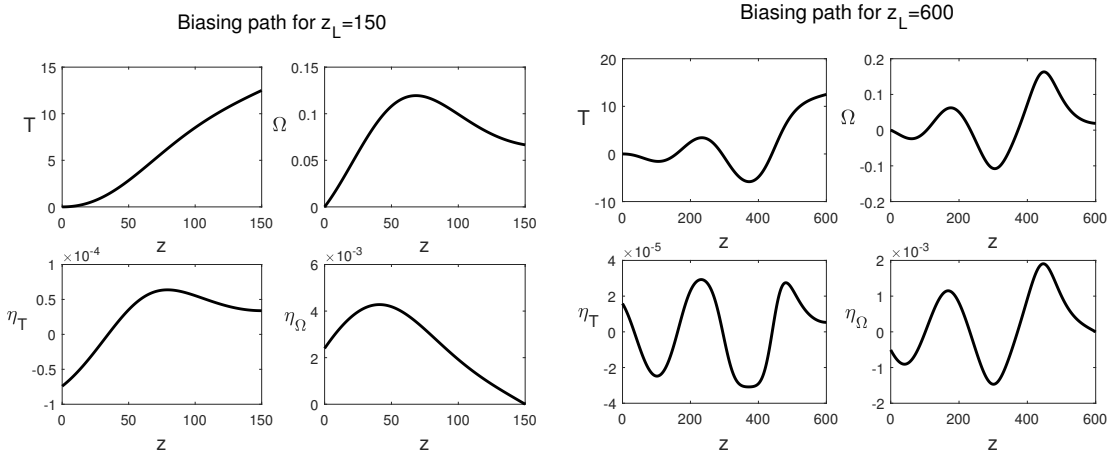

Figure 5.4. Two representative solutions for the exit path in the underdamped regime showing four-part solutions with different propagation distances, z_L. The lower panels show the biasing path coefficients for the frequency and position modes and the upper panels show the expected frequency and position trajectories under biasing. The solutions were computed numerically using the MATLAB boundary value problem solver BVP4C. For a long distance, such as the solution shown on the right, the biasing path is oscillatory, but for a short distance (left) it is not.

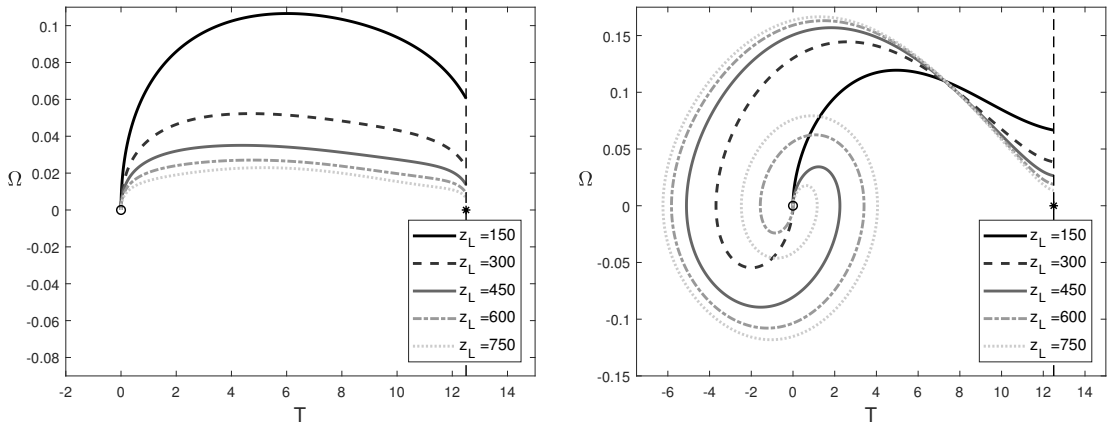

Figure 5.5. Computed biasing path trajectories visualized in (T, Ω) phase space for different system lengths in the overdamped and underdamped regimes. Left: the overdamped regime. All paths monotonically approach exit and do not change qualitatively as the system length increases. For longer propagation distances, less frequency driving is necessary to force an exit. Right: the underdamped regime. As z_L increases the biasing paths undergo more and more oscillations before exiting due to the oscillatory dynamics near the critical point.

a similar character for all system lengths as the optimal way to resist the overdamped mode-locking is to monotonically push the pulse toward its exit. However, the exit paths in the underdamped regime exhibit significant qualitative differences as the system length is increased. For a sufficiently long system length, the optimal path takes advantage of the oscillatory dynamics inherent in the mode-locking and wraps around the spiral before exiting, with the number of oscillations (or "loops") dependent on the overall propagation distance. Figure 5.5 shows the qualitative behavior changes in the exit paths in both regimes as the system length is varied.

It is best to understand the bifurcating oscillatory paths in the underdamped regime as different modes of exit. As the system length increases, an infinite cascade of biasing paths occur which are each local minima in the least action problem of Eqs. (5.14) and

(5.17). Continuing the paths in the right panel of Fig. 5.5 shows that multiple paths of exit from the fixed point with different numbers of oscillations, or loops, coexist at all distances past $z_L = 206$. Figure 5.6 shows a bifurcation diagram continuing these paths in the propagation distance, z_L, with the path action Eq. (5.13) as the dependent variable. Additionally, the bifurcation diagram demonstrates that there are multiple system lengths z_L for which there are two types of paths that are roughly equal in action (e.g., see the region $205 < z_L < 220$).

5.3. Numerical Simulations

To quantify the probabilities of position slips happening in this system, Monte Carlo simulations were performed. The full NLS equation, Eq. (5.1), was integrated numerically with different noise realizations many times, and statistics were computed based on the final results. Each sample in the Monte Carlo simulations was solved numerically using the split-step Fourier method [**85**] with 256 Fourier modes, a computational window in time t of width 80, and a propagation stepsize of $dz = 0.05$. The initial condition was a soliton with the stable amplitude for the parameter regime being simulated, as given in Section 5.1.1.

Position slip probabilities due to the mode-locking can be quite small so the Monte Carlo simulations were augmented with importance sampling. In these simulations, the underlying probability distribution in the problem (Gaussian white noise) is replaced with a biasing distribution from which samples are drawn. Then, the correct probabilities are computed using the likelihood ratio, which is the ratio between the original and the biased probabilities used to generate that sample [**73**]. Here, the Gaussian biasing

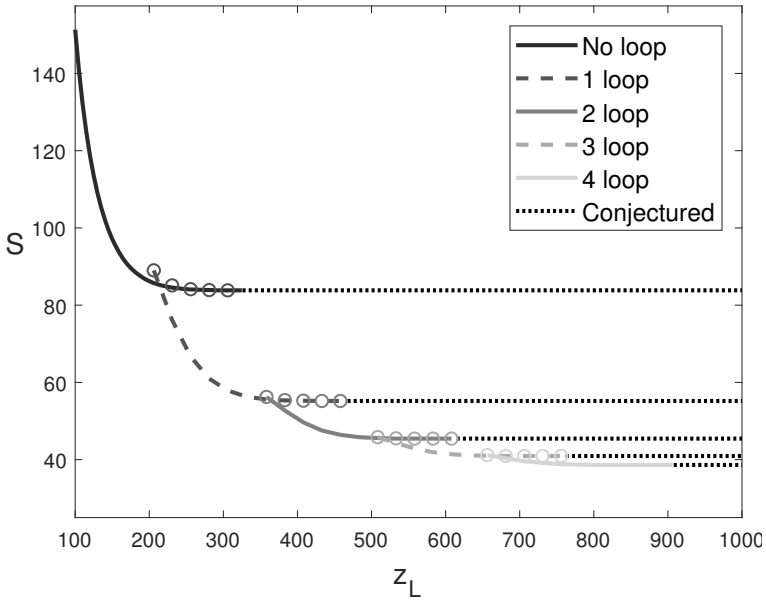

Figure 5.6. Bifurcation diagram showing biasing paths in the underdamped regime with different numbers of oscillations or loops numerically continued by varying z_L and measured in terms of their action according to Eq. (5.13). The oscillatory biasing paths are conjectured to arise through a bifucation at infinity and each has a critical propagation distance at which it disappears (visualized here as folds in the bifurcation diagram). Each branch above its fold is visualized with circles to show that its action closely matches that of the branch with one less oscillation. As z_L increases, the difference in L^2-norm between above-fold solutions and solutions with one less oscillation also goes to zero, which provides the basis for conjecturing that these solution families continue to infinity. The black dot-dashed lines depict such conjectured extensions of computed solution families.

distributions utilized means obtained from solutions to the system Eqs. (5.14) and (5.17) using coefficients, $\underline{v}_\Omega$ and $\underline{v}_T$ respectively, as described previously [**73**].

Recall from Section 2.3.2 that the Freidlin-Wentzell least action minimizer for a prescribed soliton change after passage through one amplifier is in the direction of the adjoint modes of the linearized NLS equation [**73**]. In this case, the relevant modes that should be biased at each amplifier are the position and frequency adjoint modes, the latter being

included since frequency perturbations become position perturbations upon propagation. Explicitly, the mean shift at the n^{th} amplifier due to the biasing is

$$f_n(t) = \left(\eta_\Omega(z_n)\frac{\underline{v}_\Omega}{||\underline{v}_\Omega||^2} + \eta_T(z_n)\frac{\underline{v}_T}{||\underline{v}_T||^2}\right)e^{i\Theta} \tag{5.18}$$

where η_Ω and η_T are the biasing coefficients that give the magnitude of the shift in the direction of each mode (where the modes and total phase of the soliton are computed at the amplifier). These biasing coefficients are the same as the functions used in Eqs. (5.14a) and (5.14b). With this formulation, the functional minimized in the optimization problem Eq. (5.13) to maximize the probability of an exit path is the L_2 norm of Eq. (5.18) (note that there is no cross term since the modes are orthogonal with respect to the inner product associated with the linearized NLS operator) [**73**]. The normalization constants are chosen so that the mean shift at each amplifier Eq. (5.18) agrees with Eqs. (5.14a) and (5.14b) after evaluating the jump conditions Eqs. (5.7) and (5.8). The constants C and D in (5.13) are then seen to be related to the adjoint mode norms,

$$\frac{\sigma^2}{2}C = ||\underline{v}_\Omega||^{-2} = \frac{3}{2E} \quad \text{and} \tag{5.19a}$$

$$\frac{\sigma^2}{2}D = ||\underline{v}_T||^{-2} = \frac{6E^3}{\pi^2} \tag{5.19b}$$

and are solely dependent on $E(z) \approx E_s$, and are therefore treated here as constants. (An overall multiplicative factor does not affect the minimization, of course.)

Taking the biasing coefficients resulting from the solutions of Eqs. (5.14) and (5.17) in Section 5.2.1 and incorporating them into Eq. (5.18) allows us to bias the noise to optimally produce a desired position slip. The simulation process to create a biased

sample is to solve the deterministic part of Eq. (5.1) until an amplifier is encountered. The split-step solver is then stopped, the mean shift Eq. (5.18) is computed using the precomputed biasing coefficients (extracted by fitting a soliton through the noisy pulse) and the soliton's current parameters (for construction of the adjoint modes and total phase). After adding the biased noise using this mean shift, the split-step solver is then restarted and the process is repeated until the simulation is terminated. Thousands of samples are typically generated and the final pulse position is binned to estimate its probability distribution. Additionally, in order to form full probability density functions of pulse position that resolve positions throughout a bit-slot and into neighboring bit-slots, multiple biasings were used. That is, to cover larger regions, biasing paths with different position targets $\hat{T} = T(z_L)$ were included in multiply importance-sampled Monte Carlo simulations and then combined using the balance heuristic [**91**]. Thus, distributions producing low numbers of samples in specific regions are down-weighted when others producing larger numbers of samples are available.

5.3.1. Importance Sampling in the Overdamped Regime

We performed ISMC to assess pulse position probabilities in the laser model using the overdamped parameters given in Section 5.1.1. Probability density functions of the soliton's position after propagation through varying values of z_L amplifiers are included in Fig. 5.7. The bit-slot boundaries are indicated with dashed vertical lines, and we see that the probability of a large position shift is greatly increased as the system length grows. Included in these figures is the *coefficient of variation*, or the measure of the intra-bin standard deviation divided by the bin's overall probability, which provides one way of

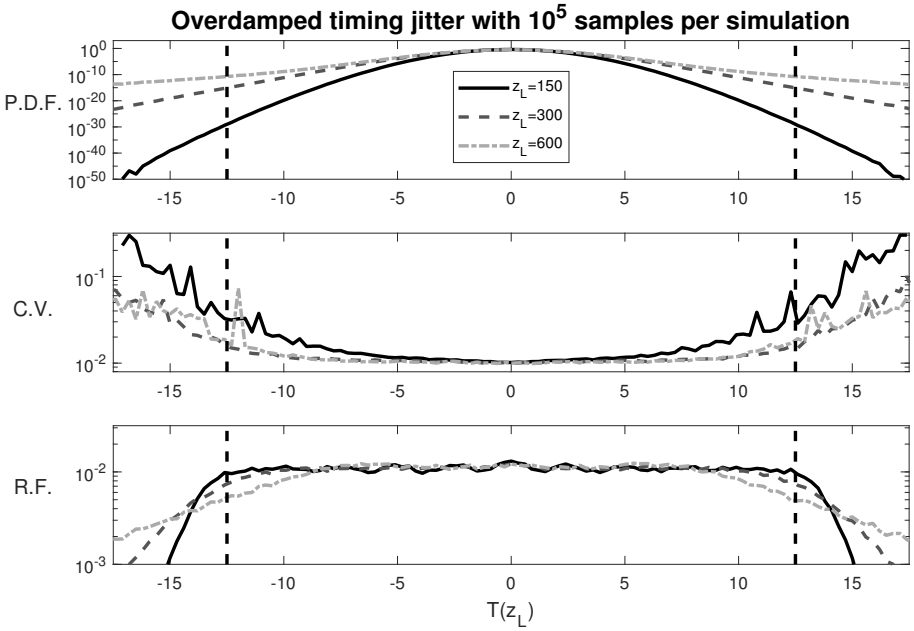

Figure 5.7. ISMC simulations of soliton position in the overdamped regime for 3 system lengths. Top panel: the probability density function (P.D.F.) of soliton position T after propagation through z_L amplifiers. Eleven position targets were included throughout the domain (evenly spaced between $T(z_L) = -12.5$ and $T(z_L) = 12.5$) with 100,000 total samples. Middle panel: the coefficient of variation (C.V.) indicates simulation convergence: a lower and smoother the C.V. indicates better convergence. We see that convergence is generally good for moderate propagation distances and worse for very short or long distances. Bottom panel: the raw relative frequency histogram (R.F.) of sample location for each simulation. Less samples reach the bit-slot boundary for the longer-distance simulations.

assessing convergence of the simulations, even when probabilities are small. In all simulations the samples converge well for regions at or near equlibrium, but not necessarily in the tails. This is disadvantageous as the tails are precisely the regions where position slips occur. For very short distances and very long distances, the simulations do not converge well in the tails (as in the left and right panels of Fig. 5.7). However, for intermediate distances, the simulations converge well. For short distances, the probabilities

that we are attempting to capture are extremely low, so it is natural that convergence is slower, but for long distances the poor convergence requires additional explanation since the probabilities are larger.

The poor convergence for long system lengths can be understood by examining the bottom panel of Fig. 5.7. The relative frequency histogram indicates the spread of samples in the standard ISMC simulations, which were conducted with eleven position targets spread evenly throughout the bit-slot and at the left and right boundaries. The samples are evenly spread throughout the interior of the bit-slot for all distances, but fewer samples are clustered around the boundaries as the system length increases. For $z_L = 600$ too many simulations either get ahead of or behind the biasing, and end up either exiting early and getting pushed to the center of the neighboring bit-slot, or get pushed back into the current bit-slot (a situation which is shown in Fig. 5.8).

This issue can be ameliorated by using dynamic importance sampling [**29**]. That is, we recompute biasing paths mid-simulation while a sample is being drawn as discussed in Section 3.4. A demonstration of this scheme applied to laser simulations is shown in Fig. 5.8. Eqs. (5.14) and (5.17) are re-solved during the simulation by changing the left-hand boundary conditions to be the soliton's current frequency and position, as shown in the bottom left panel of Fig. 5.8. The biasing path expects the pulse to be at the position of the dark grey pulse, but due to accumulated noise the pulse is actually that indicated in light grey. Therefore, we correct the biasing path by solving the biasing BVP using the frequency and position of the soliton shown with dark grey dashes, which is the best-fit soliton to the noisy pulse, and use the resulting biasing path throughout the rest of the simulation, or until the path is dynamically recalculated again. The criterion used

to decide when to recalculate the pulse can be varied; one can recalculate the path after passage through a fixed number of amplifiers or when the deviation between the expected and computed positions is larger than some prescribed tolerance, for example. Here, we choose to re-target periodically with a specific number of amplifiers between recalculations in order to precisely control how many times the BVP needs to be solved, since computing its solution is expensive relative to the computational costs of the Fourier split-step solver. The convergence benefits of dynamic ISMC typically come from more tightly clustering samples around a desired target [**30, 36**]. Here, dynamically re-calculating the optimal path improves the number of samples that finish in the vicinity of the bit-slot boundary. This can be seen for $z_L = 600$ simulations in Fig. 5.9; using dynamic ISMC eliminates the fluctuations in the C.V. near $T(z_L) = \pm 12.5$.

Overall, performing ISMC in the overdamped region is mostly straightforward as the biasing paths are typically similar in character and solving (and re-solving) the biasing BVP is simple. Accumulated deviations from biasing paths cause poorer convergence for exit probabilities at longer distances, but this can be counteracted by the use of dynamic ISMC. Short simulation distances have very low exit probabilities, and require a large number of samples to be resolved irrespective of the amount of recalculation. Dynamic ISMC typically provides notable convergence benefits with a relatively low amount of recalculation. Performing many path recalculations provides diminishing returns in convergence, however, and greatly increases the computational cost. Performing simulations to track the probability of exit directly as a function of system length z_L confirms these characteristics. These simulations are important as they directly address the formulation of the position-slip as an exit problem, and are shown in Fig. 5.10, which go out

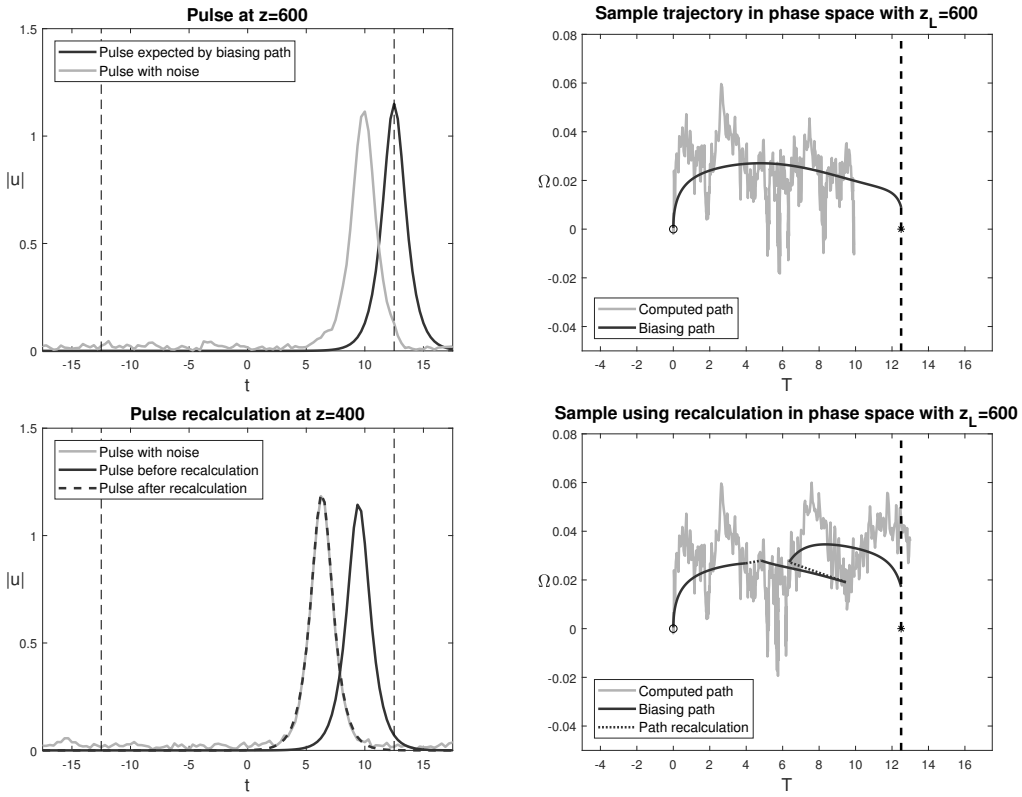

Figure 5.8. A single sample trajectory with $z_L = 600$ showing the effects of dynamic path recalculation on pulse position. Top panels: a pulse and its trajectory computed using ISMC without path recalculation. The pulse lags behind the optimal sampling path (light grey in the right-top panel) and approaches the saddle from a direction which causes it to be forced back to the left, away from the exit target of $\hat{T} = 12.5$. Bottom panels: the same noise realization as in the top panels is used in a sample drawn with recalculation of the path every 200 amplifiers. Bottom left: the biasing path recalculation at $z = 400$ showing the computed pulse, the pulse originally expected by the biasing path, and the pulse used for biasing after the path is recalculated. Bottom right: the sample trajectory in phase space again showing the parameter random walk and the expected biasing path, which is now piecewise-smooth due to the periodic recalculation. The pulse now exits the specified interval and ends up much closer to the intended position target.

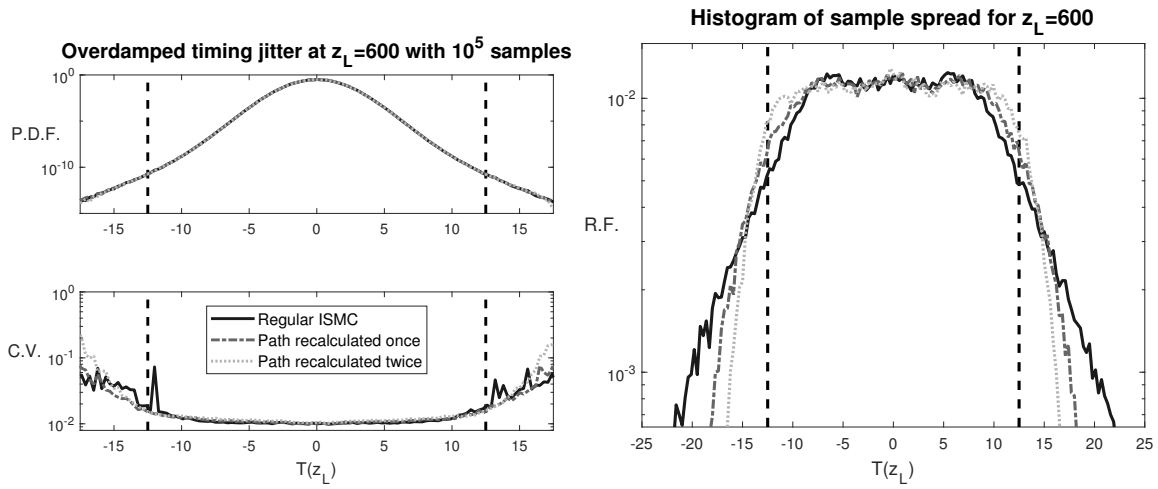

Figure 5.9. ISMC simulations for the overdamped case showing different amounts of dynamic recalculation, and their effects on convergence, for $z_L = 600$ simulations. In addition to the black curve, which has no recalculation, curves showing simulations where the path was recalculated once (at $z = 300$) or twice (at $z = 200$ and $z = 400$) are shown. Dynamic ISMC concentrates more samples in the region of the bit-slot boundary and improves the convergence there. The improvement in convergence between recalculating the path twice versus once is small.

to $z_L = 600$ with 1,000,000 total samples. The figure shows the overall exit probability, i.e., the probability that $T \geq \hat{T}$ vs. z_L. For an accurate exit probability estimate, the P.D.F. for $T(z_L)$ must be computed accurately at and just beyond the exit point. The non-dynamic ISMC simulations (black-solid) produce smooth estimates for intermediate distances, and experience variance fluctuations for longer distances. Dynamic ISMC with one path recalculation greatly improves the convergence, eliminating the fluctuations and generally lowering the C.V. with the same number of samples. Recalculating the path twice slightly lowers the C.V. level relative to one recalculation, but recalculating further does not provide any noticeable additional benefit.

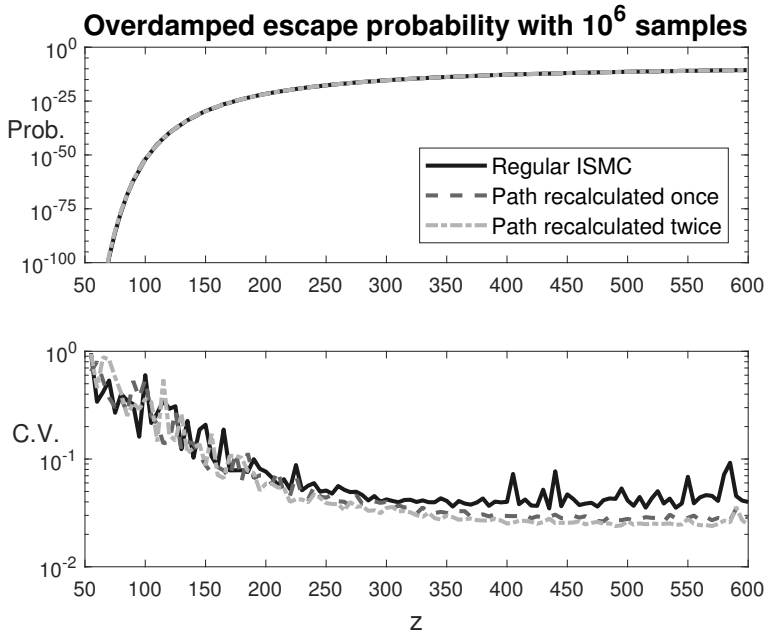

Figure 5.10. ISMC simulations of soliton exit simulations in the overdamped case, with varying amounts of dynamic recalculation. Top: the probability density function of pulse exit probability as a function of distance. Bottom: the C.V. shows improved convergence for the dynamic simulations compared with the regular ISMC, with additional but saturating benefit as more recalculation is used.

5.3.2. The Underdamped Regime

ISMC simulation in the underdamped regime is more difficult due to the oscillatory nature of the biasing paths. The numerical solution of the biasing problem Eqs. (5.14) and (5.17) involves making an initial guess, and then iterating to find a solution. While this works well in the overdamped case, in the underdamped regime a generic initial guess (for example, a constant or linear function) and use of the MATLAB boundary value solver BVP4C does not always converge to a proper solution at every distance. When optimal biasing solutions can be found, simulations for the pulse position P.D.F. can be completed; examples are shown in Fig. 5.11. As before, the general trend is that ISMC

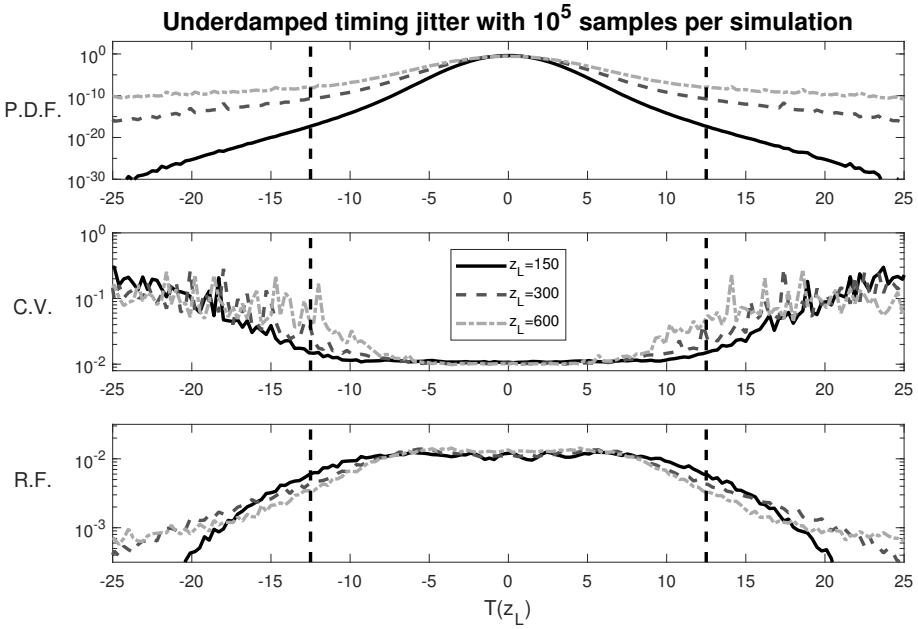

Figure 5.11. ISMC simulations of soliton position in the underdamped regime at 3 system lengths. Top panel: the probability density function of soliton position T after propagation through z_L amplifiers. Eleven position targets were included throughout the domain (evenly spaced between $T(z_L) = -12.5$ and $T(z_L) = 12.5$) with 100,000 total samples. Middle panel: the coefficient of variation provides an estimate of simulation convergence. Bottom panel: the raw relative frequency histogram of sample location for each simulation.

simulations converge better for shorter than for longer distances. The distances with good convergence are much shorter than before, however; by $z_L = 300$, the C.V. near the bit-slot boundaries exhibits significant fluctuations and we see that progressively fewer samples fall in the proper range as system length increases, as shown in the bottom panel of Fig. 5.11.

The reasons behind this behavior can be explored by considering Figs. 5.6 and 5.12. In particular, the convergence of the exit probability simulations is good up until $z_L \approx 200$, but past that point the convergence degrades. Furthermore, multiple locally optimal paths

with similar actions appear soon after this, at $z_L \approx 206$. Multiple importance sampling and other similar variants [**75**] predict that all events with similar probabilities must be included in ISMC simulations, and omitting one or more similarly likely paths has been shown to lead to convergence problems for the results [**83**]. This issue is observed here if one performs ISMC simulations for the position P.D.F. at $z_L = 210$, since this distance has three distinct paths with similar action values, as shown in Fig. 5.13. Simulations were performed to find the P.D.F. using the least action path, the two paths with the smaller actions, and all three paths. One of the results was obtained using only the path with the smallest action (the largest probability) and it is seen that this can substantially mis-estimate the P.D.F. near the target. The simulation with one path converged to a significantly lower estimate for the probabilities, while the simulations with two and three paths converged to the same values. None of the three simulations converged quickly (as measured by the C.V.), however, with the single path simulations having the worst convergence.

Applying dynamic ISMC in the underdamped case is also more difficult than before, again due to problems solving the biasing BVP. Performing dynamic simulations for the full P.D.F. (i.e., with multiple position targets) for even intermediate distances does not succeed as the boundary value solver fails to converge to a usable path sufficiently frequently. The problem is not so severe for simulations of only the region around the exit point (i.e., a single position target) and can be shown to be effective in some cases (e.g., Fig. 5.14), but the boundary value solver failure rate is an impediment for relatively long distances (such as $z_L = 600$) and there are many distances where dynamic ISMC appears to not improve convergence.

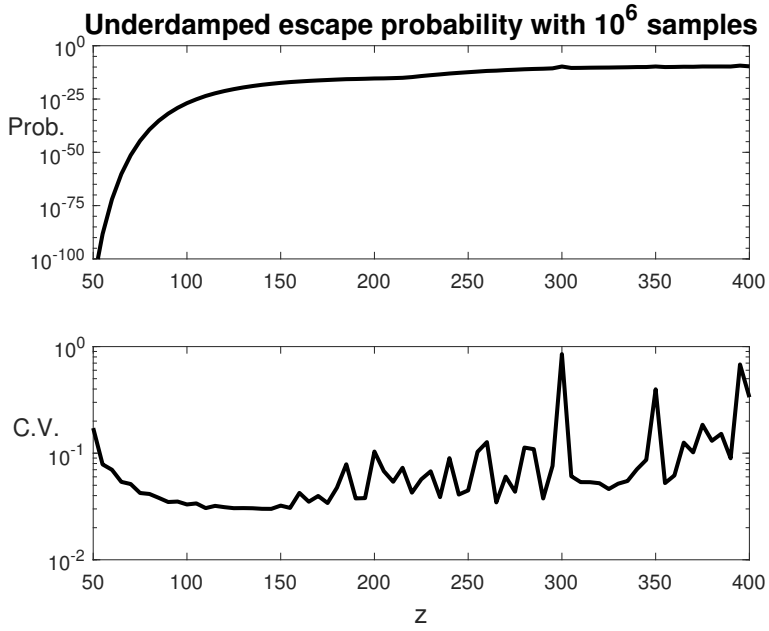

Figure 5.12. ISMC simulations of soliton exit simulations in the underdamped regime. Top: the probability density function of pulse exit probability as a function of distance. Bottom: the coefficient of variation shows good convergence for only the shortest system lengths.

In part, the issues with dynamic ISMC at longer distances is due to the multiple-path phenomenon, and such considerations will be discussed in Section 5.4. However, not all simulation distances past the onset of the oscillatory paths require the use of multiple paths in the simulation. Examination of Fig. 5.6 indicates that between roughly $250 < z_L < 350$ the path with one oscillation dominates the exit probability, for instance. Dynamic ISMC is still difficult to implement at longer distances, however, and the fraction of failed samples due to nonconvergence in the BVP solver is a persistent issue, particularly for $z_L > 300$. In a nonconvergent sample, the BVP solver, using the remaining portion of the previously computed biasing path as an initial guess, fails to converge to a prescribed tolerance. The convergence rate of the BVP solver can be considerably

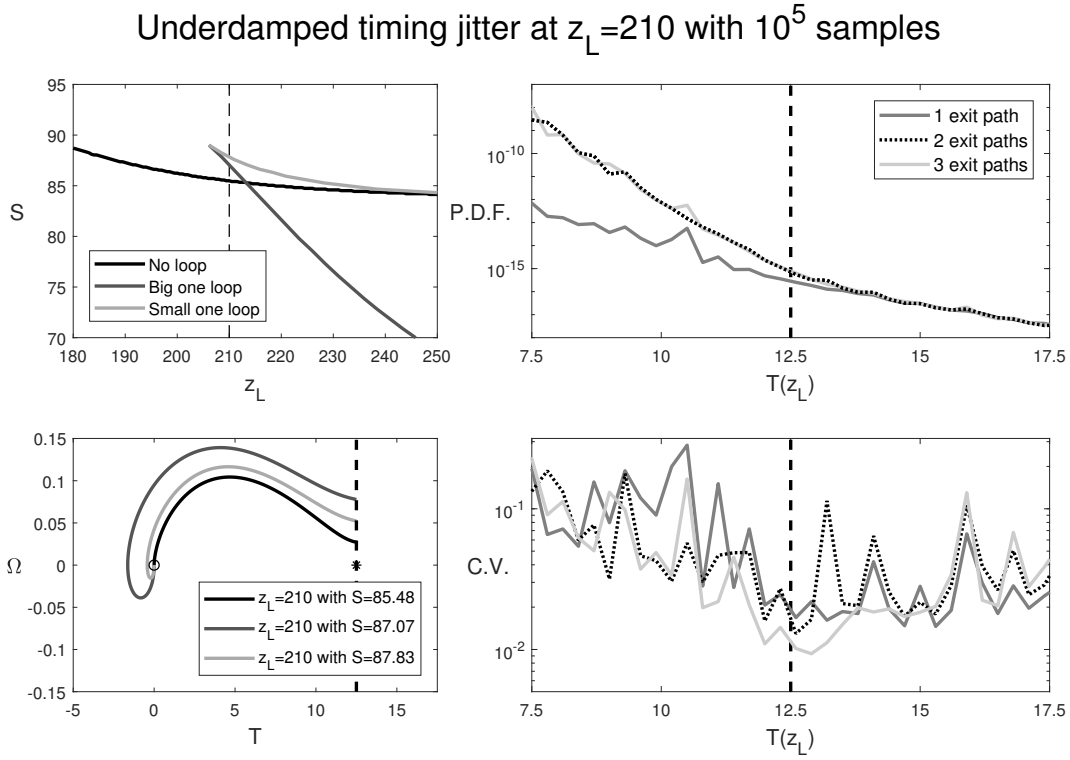

Figure 5.13. MISMC in the first region of multiple paths. Left top: a zoomed version of the bifurcation diagram. Left bottom: the three paths and their actions. Right: MISMC results using the three paths. The dark grey solid curve uses only the global minimum curve, the dotted curve uses the two paths with least action, and the light grey solid curve uses all three. The P.D.F. of pulse position shows that using multiple paths provides a substantial correction to the probability in the tail, though the convergence in the C.V. is not very good regardless of how many paths are used.

improved, however, by updating the optimal path using a series of intermediate steps between the previously-computed path and the current location in (T, Ω) phase space if the path recalculation initially fails (implementing a simple homotopy algorithm in this case). While this approach is computationally expensive, the rate at which samples experience a failed recalculation can be cut by an order of magnitude, as shown in Fig. 5.15. The lowered failure rate makes dynamic ISMC effective in this longer propagation regime.

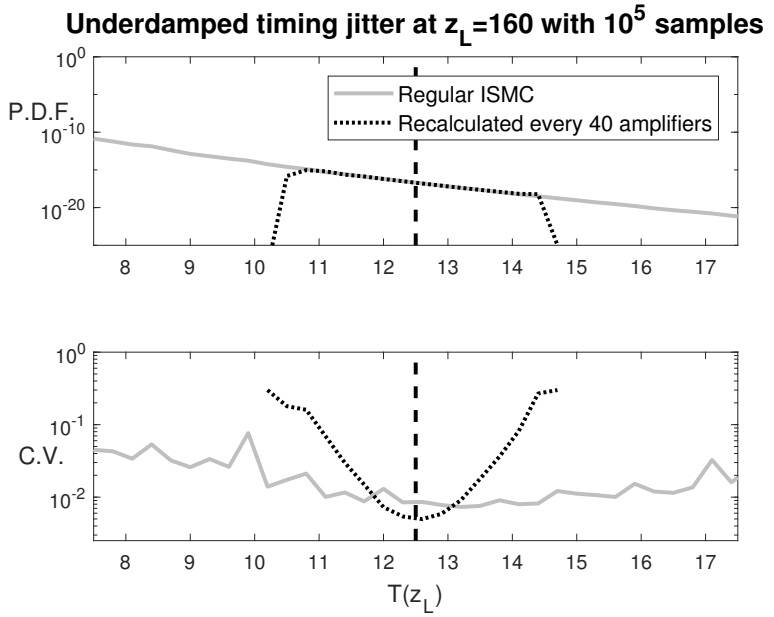

Figure 5.14. ISMC simulations of soliton position in the underdamped regime at $z_L = 160$ comparing tail-only simulations using recalculation every 40 amplifiers with simulations without recalculation. Top: the probability density function of soliton position T after propagation through z_L amplifiers. Bottom: the coefficient of variation reports simulation convergence, showing that convergence is improved by very tight targeting at the bit-slot boundary.

5.4. The Problem of Multiple Paths

As Figures 5.6 and 5.13 indicate, certain propagation distances in the underdamped regime have multiple paths to exit. In order to resolve probabilities in such regions, DMISMC is utilized. The goal of this simulation method is to achieve the improvements in a narrow region of state space that come with DISMC, in a context where multiple paths are unavoidable. We first seek to improve the results for the P.D.F. at distances just past the point at which multiple paths appear, e.g., $z_L = 210$. Note that combining MISMC with dynamic recalculation requires simultaneously recalculating, using the same intermediate position, both the path being currently being used to generate samples *and*

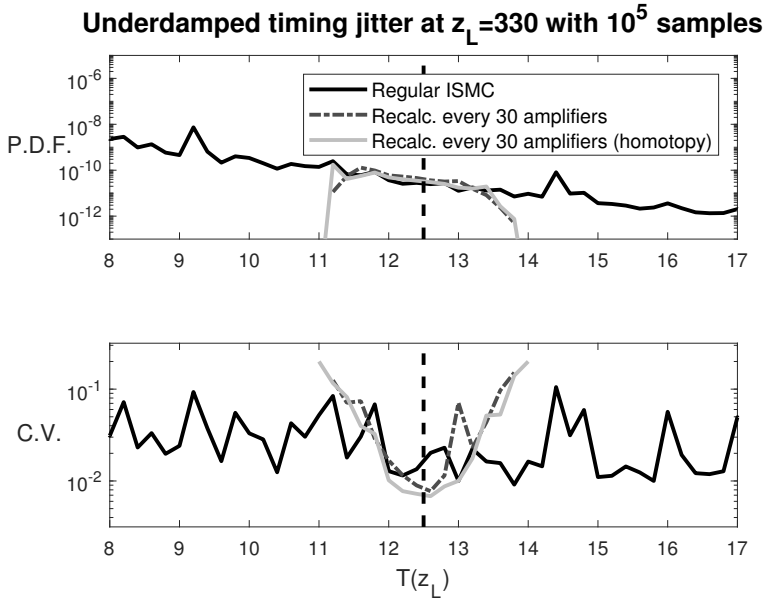

Figure 5.15. Dynamic ISMC computed for $z_L = 330$. Top: Regular ISMC, dynamic ISMC, and homotopy-assisted dynamic ISMC simulations were performed with ten path recalculations per sample for the dynamic simulations. This level of recalculation results in a very narrow spread of samples around the intended target at the bit-slot boundary. Bottom: The C.V. compares the three methods. Dynamic ISMC at this distance results in 50.3% of samples failing due to BVP4C nonconvergence, while the homotopy-assisted method fails in only 3.5% of samples (failed samples were discarded for the purposes of computing the ISMC estimates). This failure rate limits the effectiveness of dynamic ISMC, but the much lower failure rate of the homotopy-assisted method allows the convergence of dynamic ISMC to be significantly improved.

all of the other paths, since dynamically updated versions of the latter are needed for proper implementation of the balance heuristic [**11, 75, 91**]. The results at $z_L = 210$ using one (technically using DISMC), two, or three paths, taken in order from highest to lowest probability, are shown in Fig. 5.16 where each simulation method recalculates paths every fifty-five amplifiers. The results again show that convergence is best when all paths are included, and that using just one path is not sufficient; the one-path simulation still

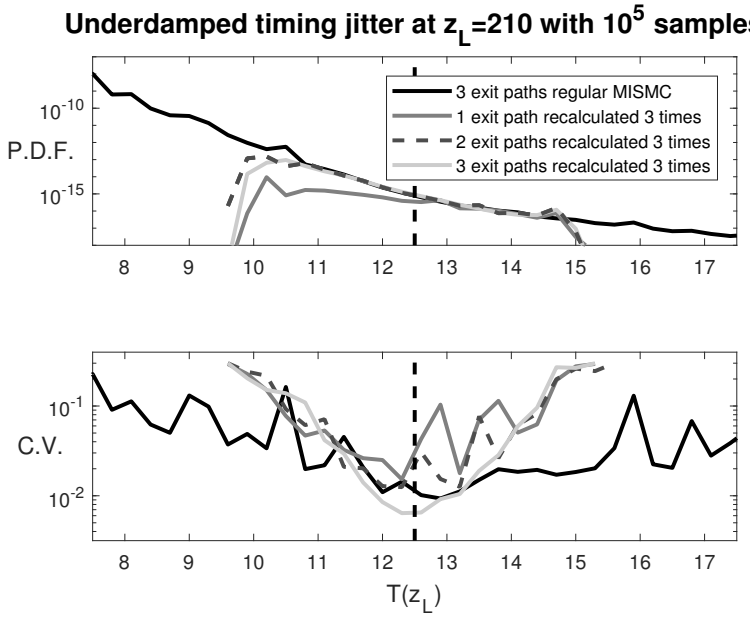

Figure 5.16. Dynamic MISMC with three paths. Dynamic MISMC is implemented starting from each of the three paths from Fig. 5.13 with the three path simulation without recalculation shown in black for comparison. For each dynamic simulation, three path recalculations were performed for each sample. The one-path simulation still converges to a lower probability than the two- or three-path simulations, while only the three-path simulation shows improved convergence relative to non-dynamic MISMC.

converges to a substantially lower probability and the C.V. is markedly worse. The C.V. for the two path simulation is also not improved relative to the non-dynamic simulation (shown in black) while only the three path simulation shows the improved convergence expected with path recalculation.

The multiple-path phenomenon in the laser model, when compared with the two dimensional random walk from Chapter 4, adds significant computational cost while reflecting some similar qualitative behavior. In the simple random walk, the calculation of error paths was a trivial task, whereas here it is an expensive procedure that is subject to failure, especially in the oscillatory underdamped regime in which the multiple-path

phenomenon occurs. This motivates the fact that a significantly lower number of recalculations per sample were attempted in Fig. 5.16 than in Fig. 4.7, for instance. In theory, a higher number of recalculations would increase the simulation yield near the bit-slot boundary, and improve convergence. However, in practice, adding recalculations greatly increases the computational effort of the problem as the boundary value solver can take a large number of iterations to converge. It is also worth noting that in this case, once the remaining distance shrinks sufficiently the three different paths all merge and become identical. This path collapse phenomenon increases the difficulty of recalculation as it creates uncertainty over the number of possible paths when recalculating. In practice, these simulations used the previously computed path as an initial guess for the boundary value solver, and then defaulted to a found path if that failed to converge, but this methodology becomes a bottleneck for the simulations when the number of recalculations increases because failed boundary value solves are computationally intensive. Ad hoc adjustments to the DMISMC procedure, such as avoiding recalculations early on in simulations when it is more likely that there are multiple paths, or using a homotopy method like the one employed in the simulations for Fig. 5.15 can improve such issues in specific circumstances, but there does not appear to be a way to optimize recalculation performance in a general setting without understanding the full landscape of possible paths in the biasing path system.

5.4.1. An exploration of the path landscape

To gain an understanding of the global landscape of error paths, we created a lookup table of solutions using the numerical continuation package AUTO (see [**25**]) to span the

region of (T_0, Ω_0, z_L) phase space where multiple paths were plausible. This region was selected based on the results of previous ISMC simulations and a survey of locations at which BVP4C failed to converge in those simulations. AUTO was used because even though there are numerous other continuation software packages that are more user-friendly, AUTO's scripting interface allowed for exploration of a parameter space with arbitrarily many dimensions. Other continuation packages such as XPP are best-suited for exploration of a one- or two-dimensional parameter space [**34**]. Exploration of phase space was accomplished by numerically continuing no-loop and one-loop solutions in lines in (T_0, Ω_0) space at specified z_L values between $100 \leq z_L \leq 300$, with the found solutions at specified grid values saved into the lookup table. The solutions were then continued in z_L and the (T_0, Ω_0) continuation was repeated at z_L locations divisible by 10 to complete the table.

The path exploration yielded a number of insights about the multiple-path phenomenon. Locations in the lookup table where multiple paths were found to exist are shown in Fig. 5.17. From this, we can see that multiple paths exist for starting locations other than the origin in (T_0, Ω_0) phase space, and they are far more prevalent in the $T_0 > 0$ half-plane. This is due to the onset of the one-loop path: it exists for distances shorter than the origin onset-distance of $z_L = 206$, but only for $T_0 > 0$ and $\Omega_0 < 0$. Here, $\Omega_0 < 0$ suggests that the pulse initially wants to move left, and there is a balance between whether it is more likely for biasing to resist this and aim for a direct exit, or for biasing to use these dynamics to sling-shot around the origin before exiting. A similar balance is found for $T_0 > 0$, $\Omega_0 > 0$, and longer z_L where the starting point in (T_0, Ω_0) phase space recommends a direct exit, but the longer propagation distance suggests oscillating

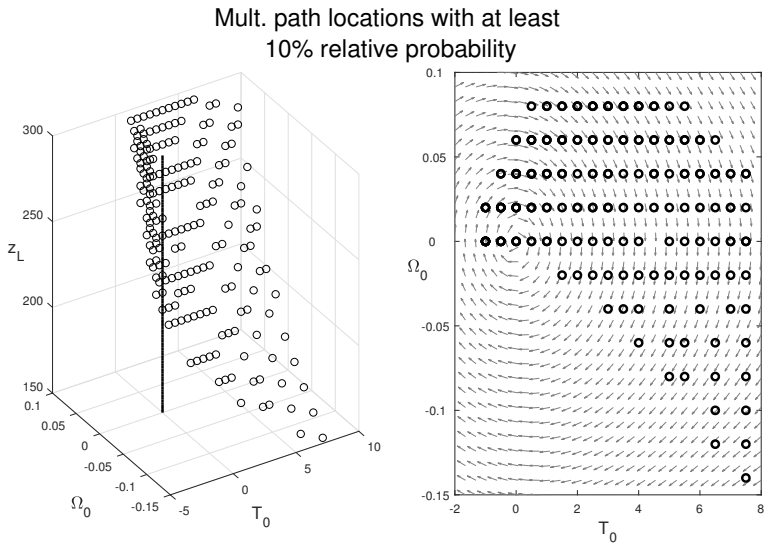

Figure 5.17. Locations found in the lookup table where there were multiple paths with at least 10% relative probability. The locations are shown in (T_0, Ω_0, z_L) phase space in the left panel, with the origin in (T_0, Ω_0) phase space visualized as a vertical line for reference. In the right panel, all multiple-path locations are projected into (T_0, Ω_0) phase space and are overlaid on the mode-locking phase plane to give a sense of the dynamics.

around the origin once. Very few locations in the $T_0 < 0$ half plane have multiple paths as the mode-locking is working with the pulse here. In other words, there is no ambiguity whether or not the pulse should pass the origin before exiting.

The qualitative dynamics ascertained by exploring the path location phase suggest that a path appearance phenomenon may complicate dynamic ISMC simulations. For example, Fig. 5.18 shows a possible trajectory constructed by examining solutions of the exit problem in specific (T, Ω, z_L) regions. In this situation, the starting point has a unique exit trajectory, but as the solution evolves and is pushed away from the optimal path by noise, a second path becomes possible (the dashed curves in the figure). Eventually, the first path disappears and only the second path remains. Performing DISMC for cases of this type presents a new level of difficulty, of course, as multiple importance

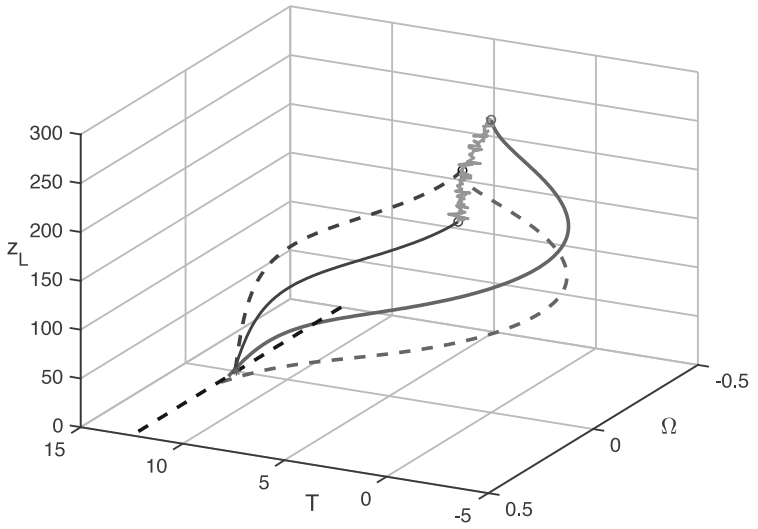

Figure 5.18. A situation where the number of exit paths increases mid-sample. A single exit path, with one oscillation (solid lighter grey), exists at the outset of the sample with $z_L = 330$, but the sample trajectory (lightest grey curve) propagates into a region of (T_0, Ω_0, z_L) phase space where a non-oscillatory path is most likely (solid black). The point at which the two paths are equally likely (dashed curves) is also depicted in between these two points. This situation can be anticipated by considering numerically-computed solutions of the exit problem.

sampling requires enumerating the different paths and keeping track of the likelihood ratios for each of them. In the $z_L < 350$ region in which DISMC simulations were performed in Section 5.3.2, this problem did not manifest as the multiple-path locations were not too numerous. However, for longer distances, we expect this problem to severely curtail DISMC simulation. The reason for this can be seen by viewing Figs. 5.5 and 5.17 together. Multiple-path regions where the competing paths are no-loop or one-loop paths are confined to $T_0 > 0$ while the exit path from the origin is on the other side of $T = 0$. Therefore, we hypothesize that the one-loop/two-loop competition region will be for $T_0 < 0$ as the two-loop path takes the pulse to the right first before oscillating twice. As

the simulation distance increases, such competition regions will become increasingly dense in the phase space around the origin, and it will be inevitable that some amount of samples will deviate from the optimal path and traverse such regions. A general method capable of tracking such situations properly would require combining a continuation/bifurcation method with dynamic MISMC simulations. Such a combined method is beyond the scope of the present thesis, however, and it is hoped that the preceding discussion of the multiple-path phenomenon will provide a sufficient spring-board for future work.

CHAPTER 6

Discussion

In this thesis we have shown that multiple importance sampling techniques based on soliton perturbation theory can be extended to be a useful tool for the study of rare events in mode-locked laser systems, which are of particular interest due to recent technological developments in optical frequency metrology that depend on high-performance laser sources. This rare-event simulation method allows fast and accurate simulation of these rare events and their statistics which are difficult to compute with traditional Monte Carlo methods. This method, which builds on previous work, does depend on the mathematical structure of the equations and being able to use soliton perturbation theory to derive appropriate approximate biasing schemes. However, many laser systems do in fact use soliton or near-soliton pulses so this framework is a reasonable approach.

We have also shown that dynamic importance sampling is a useful addition to multiple importance sampling techniques in situations where system dynamics make the tails of probability distributions especially difficult to sample. In the laser model we considered two parameter regimes where dynamic importance sampling was used to target a saddle point in order to determine the probability of a position slip. In the overdamped regime, computing biasing paths was relatively simple and dynamic importance sampling was straightforward to implement. In the underdamped regime, computing biasing paths was considerably more difficult due to the oscillatory character of the biasing paths. In order to sample the tails, we both had to target the saddle and escape from an oscillatory

well, leading to biasing paths with differing numbers of oscillations coexisting at various propagation lengths. This multiple-path phenomenon motivated the development and implementation of methods blending dynamic importance sampling and multiple importance sampling. Such methods were verified with a toy problem where a random walk was blocked by a wall obstacle, and the ability for the walker to go above or below the wall was the basis for multiple paths.

These methods blending multiple and dynamic importance sampling were then applied to the laser problem in regions where multiple paths have roughly equal probability. This approach is computationally intensive, but it is a tractable resolution to the inherently difficult problem of performing dynamic ISMC when multiple paths exist in phase space. Not all distances in the underdamped regime had multiple paths, however, and at relatively long distances a homotopy method was employed to recompute optimal paths when only a single path was possible. We expect that as the propagation distance increases, however, the increasing number of oscillations in the biasing paths will cause the multiple-path phenomenon to become an ever-present concern, though these longer distances were not explored in this work and such simulations should be the subject of future work. In any case, we anticipate that the methods used here to resolve these difficulties can be a guide to others when rare events arising from multiple paths are possible.

References

[1] Ablowitz, M. J., Ilan, B., and Cundiff, S. T. Noise-induced linewidth in frequency combs. *Opt. Lett. 31*, 12 (2006), 1875–1877.

[2] Ablowitz, M. J., Kaup, D. J., Newell, A. C., and Segur, H. The inverse scattering transform-fourier analysis for nonlinear problems. *Studies in Applied Mathematics 53*, 4 (dec 1974), 249–315.

[3] Ablowitz, M. J., and Segur, H. *Solitons and the inverse scattering transform*, vol. 4 of *SIAM Studies in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa., 1981.

[4] Agrawal, G. *Fiber-Optic Communication Systems*. Wiley Series in Microwave and Optical Engineering. Wiley, 2012.

[5] Agrawal, G. *Nonlinear Fiber Optics*. Optics and Photonics. Elsevier Science, 2013.

[6] Akhmediev, N., and Ankiewicz, A., Eds. *Dissipative Solitons*. Springer Berlin Heidelberg, 2005.

[7] Bauver, M., Forgoston, E., and Billings, L. Computing the optimal path in stochastic dynamical systems. *Chaos 26* (2016), 083101.

[8] Bayliss, A. ESAM 446-1 Course Notes: Numerical solution of partial differential equations, finite difference methods. Northwestern University, Fall 2013.

[9] Benjamin, T. B., and Feir, J. E. The disintegration of wave trains on deep water part 1. theory. *Journal of Fluid Mechanics 27*, 3 (feb 1967), 417–430.

[10] Bergano, N. S. Undersea communication systems. In *Optical Fiber Telecommunications IV-B*. Elsevier, 2002, pp. 154–197.

[11] Biondini, G., Kath, W. L., and Menyuk, C. R. Importance sampling for polarization-mode dispersion: Techniques and applications. *Journal of Lightwave Technology 22*, 4 (2004), 1201.

[12] Blanchet, J., and Lam, H. State-dependent importance sampling for rare-event simulation: An overview and recent advances. *Surveys in Operations Research and Management Science 17* (2012), 38–59.

[13] Bloom, B. J., Nicholson, T. L., Williams, J. R., Campbell, S. L., Bishof, M., Zhang, X., Zhang, W., Bromley, S. L., and Ye, J. An optical lattice clock with accuracy and stability at the 10-18 level. *Nature 506*, 7486 (February 2014), 71–75.

[14] Bucklew, J. A. *Introduction to Rare Event Simulation*. Springer New York, 2004.

[15] Cappé, O., Douc, R., Guillin, A., Marin, J.-M., and Robert, C. P. Adaptive importance sampling in general mixture classes. *Statistics and Computing 18*, 4 (apr 2008), 447–459.

[16] Cappé, O., Guillin, A., Marin, J. M., and Robert, C. P. Population monte carlo. *Journal of Computational and Graphical Statistics 13*, 4 (dec 2004), 907–929.

[17] Coleman, T. F., and Li, Y. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization 6*, 2 (may 1996), 418–445.

[18] Cornuet, J.-M., Marin, J.-M., Mira, A., and Robert, C. P. Adaptive multiple importance sampling. *Scandinavian Journal of Statistics 39*, 4 (Feb. 2012), 798–812.

[19] Cousins, W., and Sapsis, T. P. Reduced-order precursors of rare events in unidirectional nonlinear water waves. *Journal of Fluid Mechanics 790* (feb 2016), 368–388.

[20] Cundiff, S. Soliton dynamics in mode-locked lasers. *Lect. Notes Physics 661* (2005), 183–206. TY - BOOK.

[21] Cundiff, S., and Ye, J. Colloquium: femtosecond optical frequency combs. *Reviews of Modern Physics 75* (2003), 325.

[22] Dematteis, G., Grafke, T., and Vanden-Eijnden, E. Rogue waves and large deviations in deep sea. *Proceedings of the National Academy of Sciences 115*, 5 (jan 2018), 855–860.

[23] Desurvire, E., Bayart, D., Desthieux, B., and Bigo, S. *Erbium-Doped Fiber Amplifiers, Device and System Developments*. Wiley-Interscience, 2002.

[24] Diddams, S. A., Bergquist, J. C., Jefferts, S. R., and Oates, C. W. Standards of time and frequency at the outset of the 21st century. *Science 306*, 5700 (2004), 1318–1324.

[25] Doedel, E. J., et al. *Auto-07p: Continuation and Bifurcation Software for Ordinary Differential Equations*. Concordia Univ., Jan. 2012.

[26] Donovan, G. M., and Kath, W. L. Rare event simulation of the performance of an actively mode-locked fiber laser model. In *Proceedings of the 2007 Conference On Lasers & Electro-Optics/Quantum Electronics and Laser Science* (2007), pp. 1953–1954.

[27] Donovan, G. M., and Kath, W. L. An iterative stochastic method for simulating large deviations and rare events. *SIAM J. Appl. Math. 71*, 3 (2011), 903–924.

[28] Drazin, P. G., and Johnson, R. S. *Solitons*. Cambridge University Press, feb 1989.

[29] Dupuis, P., and Wang, H. Importance sampling, large deviations, and differential games. *Stochastics and Stochastic Reports 76* (2004), 481–508.

[30] Dupuis, P., and Wang, H. Dynamic importance sampling for uniformly recurrent markov chains. *Annals Appl. Prob. 15*, 1A (February 2005), 1–38.

[31] Dürr, D., and Bach, A. The onsager-machlup function as lagrangian for the most probable path of a diffusion process. *Comm. Math. Phys. 60*, 2 (1978), 153–170.

[32] E, W., Ren, W., and Vanden-Eijnden, E. Minimum action method for the study of rare events. *Communications on Pure and Applied Mathematics 57*, 5 (2004), 637–656.

[33] Elvira, V., Martino, L., Luengo, D., and Bugallo, M. F. Efficient multiple importance sampling estimators. *IEEE Signal Processing Letters 22*, 10 (Oct. 2015), 1757–1761.

[34] Ermentrout, B. *Simulating, Analyzing, and Animating Dynamical Systems*. Society for Industrial and Applied Mathematics, Jan. 2002.

[35] Fischer, M., et al. New limits on the drift of fundamental constants from laboratory measurements. *Physical Review Letters 92*, 23 (2004), 230802.

[36] Fogal, S., Biondini, G., and Kath, W. Targeted importance sampling for first-order polarization mode dispersion. In *Lasers and Electro-Optics, 2003. CLEO '03. Conference on* (June 2003), pp. 851–853.

[37] Forgoston, E., and Moore, R. O. A primer on noise-induced transitions in applied dynamical systems. *SIAM Review 60*, 4 (2018), 969–1009.

[38] Freidlin, M., Szücs, J., and Wentzell, A. *Random Perturbations of Dynamical Systems*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2012.

[39] Gardner, C. S., Greene, J. M., Kruskal, M. D., and Miura, R. M. Method for solving the korteweg-deVries equation. *Physical Review Letters 19*, 19 (Nov. 1967), 1095–1097.

[40] Gelfand, I. M., and Fomin, S. V. *Calculus of variations*. Revised English edition translated and edited by Richard A. Silverman. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963.

[41] Gill, J. Lasers: a 40-year perspective. *IEEE Journal of Selected Topics in Quantum Electronics 6* (2000), 1111.

[42] Glasserman, P. *Monte Carlo Methods in Financial Engineering*. Springer New York, 2003.

[43] Glasserman, P., and Wang, Y. Counterexamples in importance sampling for large deviations probabilities. *The Annals of Applied Probability 7*, 3 (Aug. 1997), 731–746.

[44] Gordon, J. P., and Mollenauer, L. F. Phase noise in photonic communications systems using linear amplifiers. *Opt. Lett. 15*, 23 (Dec 1990), 1351–1353.

[45] Gross, E. P. Structure of a quantized vortex in boson systems. *Il Nuovo Cimento (1955-1965) 20*, 3 (May 1961), 454–477.

[46] Hall, K., Moores, J., Rauschenbach, K., Wong, W., Ippen, E., and Haus, H. All-optical storage of a 1.25 kb packet at 10 gb/s. *IEEE Photonics Technology Letters 7*, 9 (1995), 1093–1095.

[47] Hasegawa, A., and Kodama, Y. *Solitons in optical communications*. No. 7. Oxford University Press, USA, 1995.

[48] Hasegawa, A., and Matsumoto, M. *Optical Solitons in Fibers*. Springer Berlin Heidelberg, 2010.

[49] Hasegawa, A., and Tappert, F. Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers. i. anomalous dispersion. *Applied Physics Letters 23*, 3 (Aug. 1973), 142–144.

[50] Hasegawa, A., and Tappert, F. Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers. II. normal dispersion. *Applied Physics Letters 23*, 4 (Aug. 1973), 171–172.

[51] Haus, H. Mode-locking of lasers. *IEEE Journal of Selected Topics in Quantum Electronics 6* (2000), 1173.

[52] Haus, H., and Mecozzi, A. Noise of mode-locked lasers. *IEEE Journal of Quantum Electronics 29* (1993), 983.

[53] He, H. Y., and Owen, A. B. Optimal mixture weights in multiple importance sampling, 2014.

[54] Iannone, E., Matera, F., Mecozzi, A., and Settembre, M. *Nonlinear Optical Communication Networks*. Wiley Series in Microwave and Optical Engineering. John Wiley and Sons, 1998.

[55] Johnson, R. S. *A Modern Introduction to the Mathematical Theory of Water Waves*. Cambridge University Press, 1997.

[56] Kahn, H., and Marshall, A. W. Methods of reducing sample size in monte carlo computations. *Journal of the Operations Research Society of America 1*, 5 (Nov. 1953), 263–278.

[57] Kapteyn, H., Cohen, O., Christov, I., and Murnane, M. Harnessing attosecond science in the quest for coherent x-rays. *Science 317*, 5839 (2007), 775–778.

[58] Kath, W. L. A modified conservation law for the phase of the nonlinear Schrödinger soliton. *Methods and Applications of Analysis 4*, 2 (1997), 141–155.

[59] Kath, W. L. ESAM 448-0 Notes: Importance sampling. Northwestern University, Winter 2014.

[60] Kath, W. L., Mecozzi, A., Kumar, P., and Goedde, C. G. Long-term storage of a soliton bit stream using phase-sensitive amplification: Effects of soliton-soliton

interactions and quantum noise. *Optics Communications 157*, 1-6 (Dec. 1998), 310–326.

[61] Kaup, D. J. A perturbation expansion for the zakharov–shabat inverse scattering transform. *SIAM Journal on Applied Mathematics 31*, 1 (July 1976), 121–133.

[62] King, G., and Zeng, L. Logistic regression in rare events data. *Political Analysis 9*, 2 (2001), 137–163.

[63] Korteweg, D. J., and de Vries, G. On the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary waves. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 39*, 240 (May 1895), 422–443.

[64] Kuhn, H. W., and Tucker, A. W. Nonlinear programming. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (Berkeley, Calif., 1951), University of California Press, pp. 481–492.

[65] Kutz, J. N. Mode-locked soliton lasers. *SIAM Rev. 48*, 4 (2006), 629–678.

[66] Lamb, W. E. Theory of an optical maser. *Phys. Rev. 134* (June 1964), A1429–A1450.

[67] Li, J., and Kath, W. L. Extracting solitons from noisy pulses. *SIAM J. Appl. Math. 72*, 2 (2012), 577–593.

[68] Li, J., and Kath, W. L. A path-based method for simulating large deviations and rare events in nonlinear lightwave systems. *Studies in Applied Mathematics 137*, 2 (Apr. 2016), 159–173.

[69] MathWorks. *MATLAB R2019b Parallel Computing Toolbox User's Guide*, 2019.

[70] Mollenauer, L. F., Stolen, R. H., and Gordon, J. P. Experimental observation of picosecond pulse narrowing and solitons in optical fibers. *Physical Review Letters 45*, 13 (Sept. 1980), 1095–1098.

[71] Moore, R. O. Trade-off between linewidth and slip rate in a mode-locked laser model. *Optics letters 39*, 10 (2014), 3042–3045.

[72] Moore, R. O., Biondini, G., and Kath, W. L. Importance sampling for noise-induced amplitude and timing jitter in soliton transmission systems. In *Nonlinear physics: theory and experiment, II (Gallipoli, 2002)*. World Sci. Publ., River Edge, NJ, 2003, pp. 383–390.

[73] Moore, R. O., Biondini, G., and Kath, W. L. A method to compute statistics of large, noise-induced perturbations of nonlinear Schrödinger solitons. *SIAM Rev. 50*, 3 (2008), 523–549.

[74] Oh, M., and Berger, J. O. Integration of multimodal functions by monte carlo importance sampling. *Journal of the American Statistical Association 88*, 422 (June 1993), 450–456.

[75] Owen, A., and Zhou, Y. Safe and effective importance sampling. *Journal of the American Statistical Association 95*, 449 (Mar. 2000), 135–143.

[76] Owen, A. B. Monte carlo theory, methods and examples. 2013.

[77] Ragone, F., Wouters, J., and Bouchet, F. Computation of extreme heat waves in climate models using a large deviation algorithm. *Proceedings of the National Academy of Sciences 115*, 1 (Dec. 2017), 24–29.

[78] Riecke, H. ESAM 446-2 Notes: Spectral methods for partial differential equations. Northwestern University, Spring 2014.

[79] Runge, P., and Trischitta, P. The SL undersea lightwave system. *Journal of Lightwave Technology 2*, 6 (Dec. 1984), 744–753.

[80] Schwartz, I. B., Forgoston, E., Bianco, S., and Shaw, L. B. Converging towards the optimal path to extinction. *Journal of The Royal Society Interface 8*, 65 (2011), 1699–1707.

[81] Spiller, E. T. *Computational Studies of Rare Events in Optical Transmission Systems*. PhD thesis, Northwestern University, 2005.

[82] Spiller, E. T., and Biondini, G. Importance sampling for dispersion-managed solitons. *SIAM Journal on Applied Dynamical Systems 9*, 2 (Jan. 2010), 432–461.

[83] Spiller, E. T., and Kath, W. L. A method for determining most probable errors in nonlinear lightwave systems. *SIAM Journal on Applied Dynamical Systems 7*, 3 (2008), 868–894.

[84] Süli, E., and Mayers, D. F. *An Introduction to Numerical Analysis*. Cambridge University Press, 2003.

[85] Taha, T. R., and Ablowitz, M. I. Analytical and numerical aspects of certain nonlinear evolution equations. ii. numerical, nonlinear schrödinger equation. *Journal of Computational Physics 55*, 2 (1984), 203–230.

[86] Tchébychef, P. Des valeurs moyennes (traduction du russe, n. de khanikof). *Journal de Mathématiques Pures et Appliquées* (1867), 177–184.

[87] Thorpe, M. J., Balslev-Clausen, D., Kirchner, M. S., and Ye, J. Cavity-enhanced optical frequency comb spectroscopy: application to human breath analysis. *Optics Express 16*, 4 (2008), 2387–2397.

[88] Tonello, A., Wabnitz, S., Gabitov, I., and Indik, R. Importance sampling of gordon-mollenauer soliton phase noise in optical fibers. *IEEE Photonics Technology Letters 18*, 7 (Apr. 2006), 886–888.

[89] Udem, T., Holzwarth, R., and Hänsch, T. W. Optical frequency metrology. *Nature 416* (Mar. 2002).

[90] Vanden-Eijnden, E., and Heymann, M. The geometric minimum action method for computing minimum energy paths. *The Journal of Chemical Physics 128*, 6 (Feb. 2008), 061103.

[91] Veach, E. *Robust Monte Carlo Methods for Light Transport Simulation*. PhD thesis, Stanford University, 1997.

[92] Veach, E., and Guibas, L. J. Optimally combining sampling techniques for monte carlo rendering. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques 95* (1995), ACM Press.

[93] Weideman, J. A. C., and Herbst, B. M. Finite difference methods for an AKNS eigenproblem. *Math. Comput. Simulation 43*, 1 (1997), 77–88. Computation of nonlinear phenomena.

[94] Whitham, G. *Linear and Nonlinear Waves*. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. Wiley, 2011.

[95] Ye, J., and Cundiff, S. *Femtosecond optical frequency comb: Technology, principles, operation & application*. Springer, 2005.

[96] Zabusky, N. J., and Kruskal, M. D. Interaction of "solitons" in a collisionless plasma and the recurrence of initial states. *Physical Review Letters 15*, 6 (Aug. 1965), 240–243.

[97] Zakharov, V., and Ostrovsky, L. Modulation instability: The beginning. *Physica D: Nonlinear Phenomena 238*, 5 (Mar. 2009), 540–548.

[98] Zakharov, V. E. Stability of periodic waves of finite amplitude on the surface of a deep fluid. *Journal of Applied Mechanics and Technical Physics 9*, 2 (1968), 190–194.

[99] Zakharov, V. E., and Shabat, A. B. Exact Theory of Two-dimensional Self-focusing and One-dimensional Self-modulation of Waves in Nonlinear Media. *Soviet Journal of Experimental and Theoretical Physics 34* (1972), 62–69.

APPENDIX A

Computing the Optimal Biasing Direction in General

Here, we derive the solution to the optimal biasing direction problem of Eq. (2.36). Again, this problem is to consider the noise injected at a single amplifier as a perturbation (called $f(z,t) = f_0(t)\delta(z - z_0)$) and minimize the noise subject to a prescribed change in the generic soliton parameter Y of ΔY (the form of this constraint comes from the linearized soliton parameter evolution conditions in Equation 2.35)

$$\min_{f_0} \int_{-\infty}^{\infty} |f_0(t)|^2 \, dt \quad \text{subject to} \tag{A.1a}$$

$$\text{Re} \int_{-\infty}^{\infty} \underline{v}_Y^* e^{-i\Theta} f_0 \, dt = \Delta Y. \tag{A.1b}$$

Using large deviation theory, this problem can be thought of as minimizing the Wentzell-Freidlin least action functional in modified form [**38**]. This problem was also addressed in [**73**], but without the phase sensitivity in the constraint. The solution to this problem is computed by formulating a functional that incorporates the side-constraint with a Lagrange multiplier λ following [**40**]. Therefore, let

$$M = \int_{-\infty}^{\infty} |f_0|^2 \, dt + \lambda \text{Re} \int_{-\infty}^{\infty} \underline{v}_Y^* e^{-i\Theta} f_0 \, dt \tag{A.2}$$

be the constrained functional where we have used the identity that the real part of a quantity is equal to half it plus its conjugate. Note that $M, \lambda, \Theta \in \mathbb{R}$ and $\underline{v}_Y, f_0 \in \mathbb{C}$ so let's split $\underline{v}_Y$ and f_0 into real and imaginary parts via $\underline{v}_Y = v_r + iv_i$ and $f_0 = f_r + if_i$.

Now, M can be rewritten solely in terms of real quantities and we can take the real part to get

$$M = \int_{-\infty}^{\infty} (f_r^2 + f_i^2) + \lambda\big[(v_r f_r + v_i f_i)\cos\Theta + (v_r f_i - v_i f_r)\sin\Theta\big]\,dt. \tag{A.3}$$

Then we take variations to get a system of Euler-Lagrange equations

$$\frac{\delta M}{\delta f_r} = 2f_r + \lambda(v_r\cos\Theta - v_i\sin\Theta) = 0, \tag{A.4a}$$

$$\frac{\delta M}{\delta f_i} = 2f_i + \lambda(v_r\sin\Theta + v_i\cos\Theta) = 0 \tag{A.4b}$$

whose solution is simply $f_0 = -\lambda/2\,\underline{v}_Y\exp(i\Theta)$ after recombining real and imaginary parts. The value for λ is then determined by plugging this value of f_0 into Eq. (A.1b). Doing so gives

$$-\frac{\lambda}{2}\,\mathrm{Re}\int_{-\infty}^{\infty} |\underline{v}_Y|^2\,dt = -\frac{\lambda}{2}||\underline{v}_Y||^2 = \Delta Y \tag{A.5}$$

or $\lambda = -2\Delta Y/||\underline{v}_Y||^2$. Therefore,

$$f_0(t) = \Delta Y\frac{\underline{v}_Y(z_0,t)}{||\underline{v}_Y(z_0,t)||^2}e^{i\Theta(z_0,t)} \tag{A.6}$$

is the optimal direction for the noise to take. In other words, the most probable noise direction is in the direction of the adjoint soliton mode in phase with the soliton.

APPENDIX B

Running Monte Carlo Simulations in Parallel Using MATLAB

Monte Carlo simulations are trivially parallelizable. A Monte Carlo simulation consists of running a process many times and combining process outputs into relevant statistics at the end. This simulation method lends itself to parallel computation as the order of the samples does not matter, so they can be apportioned among a number of workers easily, with each sample taking place on one worker. MATLAB provides parallel architecture for running Monte Carlo simulations in parallel using a parfor loop [**69**]. The iterations of the loop are assigned, in a nondeterministic manner, to each run on one of a number of available workers. The iterates can be said to run in a nondeterministic order as the i^{th} iterate in the loop may run on the k^{th} worker in one execution and run on a different worker in another. The iterates also do not run in order: the last iterate in the loop may run before the first, and so on.

When using a parfor loop, the data structures that pass into the loop and that pass out of the loop are restricted. For example, an array that is written-on during the loop must be "sliced" so that no two iterates in the loop can write on the same index in the array. The trickiest consideration for running MC simulations using a parfor loop is initializing the internal state of the random number generators that MATLAB uses. When MATLAB creates a pool of parallel workers the internal states of the workers are created in a predictable way. Therefore, repeatedly executing the following MATLAB code

```
mypool=parpool(4); % create parpool
parfor i=1:imax
    disp(rand)
end
delete(mypool);
```

generates and displays the same values. This is obviously problematic for Monte Carlo simulations as it undermines statistical independence between simulations. The samples within each simulation (i.e. each time a parallel pool is created) will be statistically independent, but running the same simulation script twice will produce the same results twice. Using the usual expedient of "rng('shuffle')" is also ineffective. Putting the shuffle command above the parfor will have no effect on the states of the random number generators within each worker, and putting the shuffle command inside the parfor loop will cause problems as well. The shuffle command seeds the MATLAB random number generator with the current system time, and two workers may each start iterations at the exact same time, making the results of those two iterates identical.

A workaround used frequently in the code for this thesis, which is not guaranteed to be the simplest or most elegant solution to this issue, is to create a "parallel pool constant" to pass the state of the random number generator of the client into each of the parallel workers. Additionally, the creation of a "random number stream" in the client before so-doing allows each worker to have its own independent state that will be independent over samples and over simulations. MATLAB's documentation presents the passing of such a random number stream as a way to ensure repeatability of the simulations, but it also allows the simulations to be seeded with the state of the random number generator on the client. The base of this code can be executed via the following code.

```
mypool=parpool(4); % create parpool
rng('shuffle') % shuffle the rng on the client
s=rng; % save the resulting state of the rng seed into memory
sc = parallel.pool.Constant(RandStream.create('mrg32k3a','Seed',s.Seed));
% The above creates the random stream as a parallel pool constant called sc
% 'mrg32k3a' is the MATLAB default rng for parallel simulations
% The random stream has been seeded with the seed from the client

parfor i=1:imax
    stream = sc.Value; % access the parpool constant
    % grab the portion of the random stream for this iterate
    stream.Substream=i;
    % set the rng in this worker accordingly
    RandStream.setGlobalStream(stream);
    disp(rand)
end
delete(mypool);
```

Lastly, it should be noted that this random stream method does not impede the ability to make simulations reproducible. A slice-able struct array may be used in the parfoor loop to save the state of the random number generator's state in each worker before the sample is generated, guaranteeing that any sample can be recreated later. Therefore, the struct array must be preallocated before the parfor loop, and the state of the rng should be saved in each sample after the global stream is set.

APPENDIX C

Inequality Constraint in the Wall Walk Model

The treatment of Eq. (4.2) in Section 4.1 was imprecise in that it assumed the optimal path touched the top or bottom of the wall, rather than strictly obeying the inequality condition. The precise way to treat this condition is to define a slack variable and then use the Karush-Kuhn-Tucker conditions [**64**]. Let's treat the positive case $y_0 + \sum_{i=1}^{N_1} \Delta y_i \geq h$ (the negative case is similar). Define the the slackness variable $s = y_0 + \sum_{i=1}^{N_1} \Delta y_i - h$ and then replace the inequality condition in Eq. (4.2) with $y_0 + \sum_{i=1}^{N_1} \Delta y_i - s = h$ and $s \geq 0$. Then add three Lagrange multipliers for each of the equality constraints to form a function

$$Z = \sum_{i=1}^{N} \{(\Delta x_i)^2 + (\Delta y_i)^2\} + \lambda_1 \Big[\sum_{i=1}^{N_1} (\Delta x_i) - (x_B - x_0) \Big] + \lambda_2 \Big[\sum_{i=1}^{N-N_1} (\Delta x_{N_1+i}) - (x_F - x_B) \Big] + \lambda_3 \Big[\sum_{i=1}^{N_1} (\Delta y_i) - (h + s - y_0) \Big]$$

and take derivatives with respect to $\Delta x_i, \Delta y_i$, and the Lagrange multipliers λ_i as before. Since we've introduced an extra variable s and a concomitant inequality condition for s we also have to check the "complementary-slackness" condition [**64**]

$$s \frac{\partial Z}{\partial s} = 0. \tag{C.1}$$

Now $\frac{\partial Z}{\partial s} = -\lambda_3 = -\frac{2(y_0-h-s)}{N_1\sigma^2}$, and since Eq. (C.1) implies that either $s = 0$ and the inequality constraint is binding, or $\frac{\partial Z}{\partial s} = 0$ which implies that $s = y_0 - h$ and therefore

$$\sum_{i=1}^{N_1} \Delta y_i = 0.$$

Furthermore, following the equivalent steps as in Section 4.1 shows that, in fact, $\Delta y_i = 0$ for $i = 1, \ldots, N_1$. Therefore, the main effect of the inequality constraint is to introduce a second type of biasing path in the case when $|y_0| > h$. This biasing path is very intuitive in construction. Since the walker is above (or below) the wall, the walker is not impeded by the wall and only needs to take steps directly towards the target to the right of the wall.