A path-based method for simulating large deviations and rare events in nonlinear lightwave systems

By Jinglai Li and William L. Kath

Errors in nonlinear lightwave systems are often associated with rare, noise-induced large deviations of the signal. We present a method to determine the most probable manner in which such rare events occur by solving a sequence of constrained optimization problems. These results then guide importance-sampled Monte-Carlo simulations to determine the events’ probabilities. The method applies to a general class of intensity-based optical detectors, to arbitrarily shaped and multiple pulses.

1. Introduction

In lightwave systems, additive noise produces large signal deviations that lead to system errors [1, 2, 3], and determining the probabilities associated with such events is essential for evaluating system performance. Because such large deviations can be extremely rare (e.g., \( \leq 10^{-9} \)), standard Monte Carlo (MC) simulations are impractical due to the prohibitively large number of samples needed to produce reliable probability estimates.

A number of methods have been devised for improving simulation efficiency of rare events [4]. In what follows we will focus on importance sampling (IS) [5, 6]. IS generates samples using a biased distribution and then corrects for the biasing using the likelihood ratio [4, 6]. A key issue with applying IS, however, is identifying the most probable regions of sample space that lead to the large deviations of interest, and then finding a biasing distribution that concentrates samples in these regions. Identifying such noise configurations not only improves error probability estimates but also can provide insight mechanisms responsible for the generation of errors.

This work was supported by National Science Foundation grants DMS-0709070 and DMS-1211912, and by National Natural Science Foundation of China grant 11301337.

DOI: 10.1146/((please add article doi))

© YYYY Wiley Periodicals, Inc., A Wiley Company
In soliton-based lightwave systems, the most probable noise configurations leading to large deviations were identified using soliton perturbation theory, both for return-to-zero [7, 8] and differential phase-shift keyed formats [9, 10]. IS has also been used to determine the phase distributions of nonsoliton pulses based using a root-mean-square approximation [11]. Studies further extended the method to dispersion-managed systems by taking advantage of a path-averaged governing equation and a singular perturbation technique [12, 13, 14]. More recently, a method applicable to arbitrarily shaped pulses was developed [15]; here the most probable noise configurations were found using a combination of the singular value decomposition (SVD) and the cross-entropy (CE) method.

Here we pose a constrained optimization problem to identify the most probable noise configurations leading to errors. Compared with the SVD/CE method [15], we avoid the expensive SVD computation by exploiting the mathematical structure of the governing equation, and replace the cross-entropy method with deterministic optimization, significantly improving efficiency. As a result, we are able to extend previous results to determine the large deviations in multiple bit patterns, accounting for pulse interactions.

A related simulation technique is the multicanonical Monte Carlo (MMC) method [16, 17]. MMC produces random samples using a Markov chain Monte Carlo random walk, and uses an iterative procedure to locate the important regions of state space. When the biasing distribution can be determined explicitly, however, as is the case here, IS simulations tend to be much more efficient.

The rest of this paper is organized as follows. Section 2, describes the lightwave system model, and Section 3 presents our method for determining large deviations. Section 4, applies the proposed method to determine transmission errors in lightwave systems and Section 5 provides two numerical examples. Section 6 gives some final remarks.

## 2. Simulation model

Propagation of light in optical fibers is described by the nonlinear Schrödinger equation (NLSE), which in a dimensionless form is [1],

\[
\frac{\partial u}{\partial z} = \frac{i}{2} d(z) \frac{\partial^2 u}{\partial t^2} + i |u|^2 u + \sum_{n=1}^{N_a} s_n(t) \delta(z-z_n),
\]

(1)

Here \( u(t, z) \) is the optical field envelope, \( z \) and \( t \) are dimensionless distance and retarded time [1, 2, 3]. For dispersion-managed (DM) systems, \( d(z) \) is the dispersion map [1] (otherwise, \( d(z) = 1 \)). The last term represents the
random perturbations added at amplifiers due to amplified spontaneous emission (ASE) noise from in-line optical amplifiers [1, 2, 3]. Here $N_a$ is the number of amplifiers, $z_n$ are their locations, and $\delta(z)$ is the Dirac delta distribution. The terms $s_n(t)$ represent i.i.d. Gaussian white noise satisfying $\mathbb{E}[s_n(t)] = 0$ and $\mathbb{E}[s_n(t)s_{n'}(t')] = \sigma^2 \delta(t-t') \delta_{nn'}$ where $\mathbb{E}[\cdot]$ denotes ensemble average, $\delta(t-t')$ is a Dirac delta in same coordinate as the pulse profile, $\delta_{nn'}$ is the Kronecker delta, and $\sigma^2$ is a combination of physical constants and system parameters that determines the noise power [3]. In reality, noise is not white and has a finite bandwidth.

A model detection procedure is illustrated in Fig. 1. At the receiver the signal is filtered and optical power is converted into an electrical voltage $V$ which is then compared to a threshold level $V_D$ to determine whether a “1” bit (if $V > V_D$), or a “0” bit (if $V < V_D$) was sent. Here we assume an optical bandpass filter $\hat{f}(\omega)$ and an integrate-and-dump detector:

$$V = \int_{T_1}^{T_2} |F^{-1}[\hat{f}(\omega)F[u(t, z_{\text{end}})](\omega)](t)|^2 dt ,$$

where $F[\cdot]$ and $F^{-1}[\cdot]$ are the Fourier and inverse transforms. In what follows, we will omit the final distance $z = z_{\text{end}}$. For our purposes, it is more convenient to write Eq. (2) as:

$$V = \int W(t)|(f * u)(t)|^2 dt = \int W(t)\left|\int f(t - \tau)u(\tau)d\tau\right|^2 dt .$$

Here $f * u$ is the time-domain convolution of $f$ and $u$, and $W(t)$ is a window function: $W(t)=1$ for $t \in [T_1, T_2]$, and $W(t) = 0$ otherwise. Eq. (3) can also be written concisely as $V = \|W(f * u)\|_2^2$. Integrals are complete unless limits are explicit. The window function is used to specify a particular bit slot (time interval) of interest when the signal $u(\cdot, z)$ contains a series of pulses representing a bit sequence.

Figure 1. Optical detector schematic showing noisy (solid) and noise-free (dashed) optical signals; $|u|^2$ is the optical power.
Noise can cause the received optical power to vary considerably, so that the detected voltage can be below the threshold value $V_D$. In this case, a transmitted “1” will be detected as a “0”, i.e., an error will occur. A standard MC simulation of such events is straightforward. For each trial, one launches a pulse at $z = 0$, and propagates it to the first amplifier by solving the deterministic part of Eq. (1) numerically. One then adds randomly generated ASE noise to the signal, propagates the signal to next amplifier, and repeats the process. At the end of the transmission line, the optical receiver is applied to obtain the output voltage. One repeats the above process for different noise realizations and uses the voltage statistics to estimate the desired error probability. Because large deviations are infrequent, however, methods that can deal with rare events must be employed.

3. Identifying and simulating rare events with IS

Let $X = (X_1, X_2, \ldots, X_N)$ be an $N$-dimensional random variable and the event of interest be defined by a scalar $g(X) < \rho$, where $\rho$ is a constant. The probability of the event of interest, $P$, is then expressed as

$$P = \int I_\rho(g(x))p(x)dx,$$

where $p(x)$ is the probability density function (PDF) of $X$, and $I_\rho(g(X)) = 1$ for $g(X) < \rho$ and 0 otherwise. An unbiased estimator for $P$ is

$$\hat{P} = \frac{1}{M} \sum_{m=1}^{M} I_\rho(g(X_m)),$$

where the samples $\{X_m\}$ are drawn from the distribution $p(x)$ [4, 6]. If $P$ is very small, however, an unreasonable number of samples are necessary to produce a reliable estimate of $P$. One can, however, write Eqs. (4) as:

$$P = \int I_\rho(g(x))L(x)p^*(x)dx, \quad \hat{P}^* = \frac{1}{M} \sum_{m=1}^{M} I_\rho(g(X^*_m))L(X^*_m),$$

where the samples $X^*_m$ are now drawn from the biasing distribution $p^*(x)$, and where the quantity $L(x) = p(x)/p^*(x)$ is the likelihood ratio [4, 6]. When an appropriate biasing distribution is used, the IS estimator (5) can accurately estimate the probability of the rare event of interest much more efficiently than with straightforward MC methods.

The challenge with implementing IS is to choose a good biasing distribution. For problems where $X = (X_1, X_2, \ldots, X_N)$ are i.i.d. zero-mean Gaussian random variables with $p(x) \propto \exp(-\|x\|_2)$, one choice is to
translate the distribution’s mean to \( x^* \) by letting \( p^*(x) = p(x-x^*) \), where \( x^* \) is the most probable location satisfying the constraint, i.e., solves [5, 18]

\[
\min \|x\|_2, \quad \text{subject to} \quad g(x) < \rho. \tag{6}
\]

When \( g(x) \) is complicated and the dimensionality of \( x \) is high, this optimization problem can be difficult to solve. On the other hand, due to the random sampling nature of the IS method, a reasonable approximation to \( x^* \) can adequately guide the simulation. We thus propose the following adaptive method to obtain an approximate solution to Eq. (6),

1. Choose a sequence \( g(0) = \rho_0 > \rho_1 > \cdots > \rho_{J-1} > \rho_J = \rho \) of constraint values so that \( \epsilon_j = \rho_{j-1} - \rho_j \ll 1 \) for all \( j = 0, ..., J \).
2. Find the solution to \( \min \|x\|_2 \) subject to \( g(x) = \rho_0 \) (it is trivial: \( x_0 = 0 \)); let \( j = 1 \).
3. Compute an (approximate) solution to Eq. (6) using the result of the previous step, \( x_{j-1} \), as the starting point.
4. Stop if \( j = J \); otherwise, let \( j = j+1 \) and go back to step 3.

The main part of the algorithm is to solve the intermediate optimization problem, Eq. (6), in step 3:

\[
\min \|x_j\|_2 = \min \|x_{j-1} + \Delta x_j\|_2, \quad \text{subject to} \quad g(x_{j-1} + \Delta x_j) = \rho_j. \tag{7}
\]

Suppose that we have chosen \( \epsilon_j \) sufficiently small, so that the constraint in Eq. (7) can be approximated by the linear equation \( b_j^T \Delta x_j = -\epsilon_j \), where \( b_j \) is the gradient of \( g(x) \) at \( x_{j-1} \):

\[
b_j = \frac{\partial g}{\partial x} \bigg|_{x=x_{j-1}}.
\]

This approximate problem can be solved analytically,

\[
\Delta x_j = -\frac{\epsilon_j}{b_j^T b_j} b_j + \frac{b_j^T x_{j-1}}{b_j^T b_j} b_j - x_{j-1}, \tag{8}
\]

and therefore

\[
x_j = -\frac{\epsilon_j}{b_j^T b_j} b_j + \frac{b_j^T x_{j-1}}{b_j^T b_j} b_j. \tag{9}
\]

As a result, this adaptive algorithm allows one to obtain an explicit approximation to the solution of Eq. (6). This is much more efficient than solving the full optimization problem numerically. In general, the gradient \( b_j \) can be computed with finite differences, but this can be quite expensive when the dimensionality of \( x \) is large. As will be shown in Section 4, in the present problem we can derive an semi-analytical expression for the gradient by exploiting the mathematical structure of Eq. (1).
4. Application to lightwave systems

We apply the method described in the previous section to lightwave communication systems. In a numerical simulation the noise term $s_n(t)$ at an amplifier is discretized as a set of independent, identically distributed zero-mean normal random variables, one for each real and imaginary part of the signal at the $N_t$ temporal grid points. Let $s_n$ be the discrete counterpart of $s_n(t)$ in vector form and $X^T = [s_1^T, s_2^T, ..., s_{N_t}^T]$ where the superscript stands for matrix transpose. Note that $s_n$ is of dimension $2N_t$ and $X$ is of dimension $2N_tN_a$.

When a pulse is present, i.e., a binary “1” is being transmitted, the event of interest is that the detected voltage $V$ is below a certain threshold value $V_D$. We need to solve the optimization problem [8, 19],

$$\min_{u(t,z)} \int \int \left| \frac{\partial u}{\partial z} - i \frac{d(z)}{2} \frac{\partial^2 u}{\partial t^2} - i |u|^2 u \right|^2 dt \, dz = \min_{\{s_n(t)\}} \sum_{n=1}^{N_a} \|s_n(t)\|_2, \quad (10)$$

subject to $V = \|W(f \ast u)\|_2^2 = V_D$.

The optimization problem Eq.(10) can be solved either for $u(t,z)$, corresponding to the most probable pulse deformation, or $\{s_n(t)\}$, corresponding to the most probable noise configuration, and if either is known the other can be computed from Eq. (1). Here, we will use the method described in Section 3 to solve the problem in terms of $\{s_n\}$.

Specifically, we choose $V_0 = \rho_0 > \rho_1 > ... > \rho_{J-1} > \rho_J = V_D$ (where $V_0$ is the voltage detected in absence of noise), and solve the problem following the steps described in Section 3. Assuming that the translated noise mean at step $j-1$ is $\{s_n^{(j-1)}(t)\}$, the optimization problem (7) becomes

$$\min \sum_{n=1}^{N_a} \|s_n^{(j)}(t)\|_2 = \sum_{n=1}^{N_a} \|s_n^{(j-1)}(t) + \Delta s_n^{(j)}(t)\|_2, \quad (11a)$$

subject to

$$\|W(f \ast u^{(j)})\|_2 = \rho_j, \quad (11b)$$

with

$$\frac{\partial u^{(j)}}{\partial t} = i \frac{d(z)}{2} \frac{\partial^2 u^{(j)}}{\partial t^2} + i |u^{(j)}|^2 u^{(j)} + \sum_{n=1}^{N_a} [s_n^{(j-1)}(t) + \Delta s_n^{(j)}(t)] \delta(z - z_n), \quad (11c)$$

$$u^{(j)}(0,t) = u_0(t),$$

where $u_0(t)$ is the initial condition.
Next we derive a linear approximation of the constraint in Eqs. (11b)–(11c). We let \(u^{(j-1)}(t, z)\) be the solution of Eq. (11c) when \(\Delta s_n^{(j)}(t) = 0\), and we linearize Eq. (11c) around this solution. Since the result is linear, each amplifier’s contribution adds independently, and we obtain

\[
\frac{\partial}{\partial z} \Delta u_n = \mathcal{L}[\Delta u_n], \quad \Delta u_n(t, z_n) = \Delta s_n^{(j)}(t),
\]

for \(n = 1, \ldots, N_a\), where

\[
\mathcal{L}[\Delta u] = i \frac{d^2 \Delta u}{dt^2} + i |u^{(j-1)}|^2 \Delta u + i (u^{(j-1)})^2 \Delta u^*.
\]

For convenience, we introduce an operator \(\Phi_n\) so that \(\Delta u(t, z_{\text{end}}) = \Phi_n \Delta s_n \) describes the linear propagation of \(\Delta s_n\) from position \(z_n\) to the receiver \(z_{\text{end}}\) by Eq. (12). The approximate solution of Eq. (11c) is thus

\[
u'(t, z) = u^{(j-1)}(t, z) + \sum_{n=1}^{N_a} \Delta u_n(t, z).
\]

Substituting Eq. (13) into Eq (11b) and linearizing the result yields

\[
2 \text{Re} \int [W(f * u^{(j-1)}(t, z_{\text{end}}))]^* ([f * \sum_{n=1}^{N_a} \Delta u_n(t, z_{\text{end}})))] dt = -\epsilon_j
\]

with \(\epsilon_j = \rho_j - \rho_{j-1}\). For conciseness, we write Eq. (14) as an inner product:

\[
\langle 2 W F u^{(j-1)}(t), \sum_{n=1}^{N_a} \Delta u_n \rangle = \langle 2 F^\dagger W F u^{(j-1)}, \sum_{n=1}^{N_a} \Delta u_n \rangle = -\epsilon_j,
\]

where \(\langle y(t), v(t) \rangle = \text{Re} \int y^*(t) v(t) \, dt\), \(\dagger\) denotes an adjoint operator (so \(\langle v, F y \rangle = \langle F^\dagger v, y \rangle\)), \(F u = f(t) * u(t)\), and \(F^\dagger = f(-t)^* * u(t)\). We then substitute \(\Delta u(t, z_{\text{end}}) = \Phi_n \Delta s_n^{(j)}\) into Eq (15), obtaining

\[
\sum_{n=1}^{N_a} \langle 2 F^\dagger W F u^{(j-1)}, \Phi_n \Delta s_n^{(j)} \rangle = -\epsilon_j.
\]

Eq. (16) is a linear mapping from \(\{\Delta s_n^{(j)}\}\) to \(\epsilon_j\), and formally we have

\[
\sum_{n=1}^{N_a} \langle 2 \Phi_n^\dagger F^\dagger W F u^{(j-1)}, \Delta s_n^{(j)} \rangle = -\epsilon_j.
\]

An approximation to \(\Phi_n^\dagger\) can be found numerically via the SVD [20], but this is not necessary. Indeed, as shown in the Appendix, when applied to a pulse \(v(t)\), \(\Phi_n^\dagger\) describes the reverse propagation of \(v(t)\) from \(z = z_{\text{end}}\)
to \( z = z_n \). This reverse-propagated pulse \( \Delta u'(t, z_n) \) is given by
\[
\frac{\partial}{\partial z} \Delta u' = -\mathcal{L}^\dagger[\Delta u], \quad \Delta u'(t, z_{\text{end}}) = v(t),
\] (18a)
where
\[
\mathcal{L}^\dagger[\Delta u'] = -\frac{i}{2} \frac{d^2 \Delta u'}{dt^2} - 2i|u^{(j-1)}|^2 \Delta u' + i(u^{(j-1)})^2 \Delta u'^*.
\] (18b)

We thus can write Eq. (16) in a much more compact form,
\[
\sum_{n=1}^{N_a} \langle w_n^{(j)}, \Delta s_n^{(j)} \rangle = -\epsilon_j,
\] (19)
where \( w_n^{(j)} = 2\Phi_n^\dagger \mathcal{F}^\dagger \mathcal{W} \mathcal{F} u^{(j-1)} \). The minimizer of \( \sum_{n=1}^{N_a} \|s_n^{(j)}(t)\|_2 \), subject to the linear constraint (19), then can be found analytically:
\[
s_n^{(j)} = -\frac{\epsilon_j}{\sum_{n=1}^{N_a} \|w_n^{(j)}\|_2} w_n^{(j)} + \frac{\sum_{n=1}^{N_a} \langle w_n^{(j)}, s_n^{(j-1)} \rangle}{\sum_{n=1}^{N_a} \|w_n^{(j)}\|_2} w_n^{(j)}.
\] (20)

Once the iteration has been completed and the mean translation for the most probable noise configuration reaching the voltage threshold \( V_D \) has been found, it is used to guide IS simulations to obtain the associated error probability [8, 9, 12]. The references provide a full description for this portion of the procedure and so we do not repeat these details here.

5. Simulation results

To demonstrate and validate our method, we first apply it to a previously studied system [12]. Here an averaged, nonlocal governing equation, the dispersion-managed NLSE (DMNLSE) [21], was used to guide the IS simulations. This is only an approximation to Eq. (1) with variable coefficients. Here we simulate the system using Eq. (1) directly. The total propagation distance is 4000 km, the average dispersion is 0.15 ps^2/km with a dispersion map strength \( s = 4 \) and dispersion map period \( z_a = 100 \) km. The amplifier spacing equals the dispersion map period. The nonlinear coefficient is 1.7 (W km)^{-1}, the fiber loss is 0.21 dB/km, and the pulses transmitted are DM solitons with 11.8 mW peak power. The noise spontaneous emission factor is 1.5. We use a filtering function for a 10 GHz Gaussian optical filter, \( f(\omega) \propto \exp(-2\ln(2)\omega^2/(2\pi\delta/T_L)^2) \), where \( T_L \) is computational domain width in picoseconds [22]. The system parameters can also be expressed in dimensionless units [12].

To employ our method to recover the detected voltage’s full probability density, we draw samples using several different biasing distributions (i.e.,
with different values of $V_D$), and combine the results using multiple importance sampling [8, 23]. We used seven biasing distributions targeting $V_D = 0.5, 0.8, 0.9, 1, 1.1, 1.2, 1.5$, each with $10,000$ realizations. Fig. 2 compares the normalized voltage distribution found with our method and that obtained previously [12]. There is good agreement between the two results. We also show the coefficient of variation (standard deviation divided by the mean) of the IS simulation [24] in the bottom of the figure. These results show that our method is capable of accurately simulating large, rare pulse deformations without resorting to the averaged equation.

A more challenging example demonstrates the applicability of the method to multiple, arbitrarily-shaped pulses. Specifically, we consider a chirped-return-to-zero (CRZ) system consisting of a dispersion compensating fiber (pre-DCF) followed by 75 dispersion map periods and with a final dispersion compensating fiber (post-DCF) [25]. Each dispersion map consists an amplifier followed by 34 km of $D_+$ fiber and 17.44 km of $D_-$ fiber, and the pre- and post-DCF consist of 51 km and 44 km of $D_+$ fiber respectively, resulting in a total propagation distance of approximately 4000 km. The fiber specifications are given in Table 1. The amplifiers have a gain of 10.82 dB and a spontaneous emission factor of 1.2. We use raised cosine pulses with a chirp parameter $A = -0.46$,

$$u(t, z) = \sqrt{\frac{P_{\text{peak}}}{2}} \left( 1 + \cos(\pi \sin(\frac{\pi t}{T})) \right) \exp(iA\pi \cos(\frac{2\pi t}{T})), $$

where $T=100$ ps is the bit period and $P_{\text{peak}}=1$ mW is the peak optical power. The receiver is a 50 GHz Gaussian filter and an integrate-and-
Table 1

<table>
<thead>
<tr>
<th></th>
<th>Dispersion (ps/(nm-km))</th>
<th>Slope Index Coeff. (m²/W)</th>
<th>Eff. Area (µm²)</th>
<th>Loss (dB/km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D+</td>
<td>20.17</td>
<td>0.062</td>
<td>1.7 × 10⁻²⁰</td>
<td>106.7</td>
</tr>
<tr>
<td>D−</td>
<td>-40.8</td>
<td>-0.124</td>
<td>2.2 × 10⁻²⁰</td>
<td>36.1</td>
</tr>
</tbody>
</table>

Figure 3. The profiles of the optimally deformed signals just before the receiver (after being filtered). Pulses 1–4 are numbered from top to bottom. Noise-free signals (dashed) are shown for comparison.

dump detector within each bit period. The numerical simulations used the periodic bit pattern ‘01110100’ containing all possible 3-bit combinations.

We first compute the most probable pulse deformations. In Fig. 3, we plot the optimally deformed pulses (toward a decreased voltage at the receiver after being filtered). We then employed an IS simulation to reconstruct the PDF of the average power for the four pulses in the bit pattern. Initially all the four pulses are launched with an average power of 0.348 mW. In the simulation for each pulse, we biased the samples toward four different voltage variations: ΔV = 0 mW, -0.05 mW, and -0.10 mW. 10,000 samples were used with each biasing, so a total of 30,000 samples were used to reconstruct the voltage PDF for each pulse. Fig. 4 shows the results. The voltage PDFs for the four pulses are relatively close to one another, but the right-most pulse (i.e, the isolated one) has a much lower probability density in the region where errors usually occur. Thus, the right-most pulse is less susceptible to interactions from the other pulses.
A path-based method for simulating large deviations and rare events

Figure 4. Top: PDF for the average power for all the four pulses in the bit pattern ‘01110100’. Pulses are numbered 1–4 from left to right in the bit pattern. Bottom: the coefficient of variation for the IS simulations.

6. Discussion

We have presented an iterative, path-based method to predict large deviations and determine the probabilities of associated rare events in nonlinear lightwave systems. We first formulated the identification of the most probable noise configurations leading to large deviations as a constrained optimization problem. A straightforward approach with standard numerical optimization techniques may have difficulties since propagation couples signal and noise nonlinearly and because the state space typically has very high dimensionality. By considering a sequence of problems with increasing pulse deformation and exploiting the structure of the equation, however, we are able to solve each subproblem semi-analytically and construct an approximate solution to the overall optimization problem. Errors in the approximation can be controlled by taking sufficiently small steps, but because IS randomly samples the region around the biasing point the method appears relatively tolerant to such errors and provides good results even when relatively large steps are taken.

We considered two examples to demonstrate the effectiveness of the method: a dispersion-managed soliton system and a chirped return-to-zero system. In the first example, we demonstrated good agreement with results obtained previously with an alternate method using the perturbation theory of DM solitons. The second example demonstrates that our method applies to arbitrarily shaped and multiple pulses.

To assess fully the performance of an OFC system, of course, one should also compute the probability that a ‘0’ is mis-detected as a “1”.
While the present method provides a framework for determining large deviations and rare events in the current case, the implementation must be different for zeros due to the absence of a pulse. An extension of the method to the case of zeros will be the subject of future work.

Another possible extension is to the study of rare events in mode-locked lasers; such devices have applications in fields ranging from communications [26] to optical frequency metrology and optical clock technology [27, 28, 29, 30]. In such systems, dispersion and nonlinearity must maintain a very precise balance to achieve mode-locking and produce ultrashort pulses [26]. Additive noise is also present in mode-locked lasers and can cause large deformations leading to loss of mode-locking [31]. We anticipate that the approach presented here, after necessary and appropriate modifications, will also be able to assist with the prediction of large, rare deviations in such systems.

Acknowledgments

We would like to thank Gino Biondini and Graham Donovan for many helpful discussions.

Appendix: Adjoint of the linearized NLSE

Let $\Phi$ be the linear operator characterizing the propagation of a pulse from $z_a$ to $z_b$ via the linearized NLSE. Namely, when applied to a field $v(t)$, it maps $v(t)$ to $\Delta u(t, z_b)$ where $\Delta u(t, z)$ is governed by

$$\frac{\partial}{\partial z} \Delta u = \mathcal{L}[\Delta u], \quad \Delta u(t, z_a) = v(t),$$

(21)

where $\mathcal{L}$ is given by Eq. (12b). Also let $\Phi^\dagger$ be the operator mapping a field $y(t)$ to $\Delta u'(t, z_a)$ via

$$\frac{\partial}{\partial z} \Delta u' = -\mathcal{L}^\dagger[\Delta u'], \quad \Delta u'(t, z_b) = y(t),$$

(22)

where $\mathcal{L}^\dagger$ is given by Eq. (18b). We will drop the superscript $(j-1)$ in Eqs. (12b) and (18b). We will show that $\langle y, \Phi v \rangle = \langle \Phi^\dagger y, v \rangle$. First define

$$\vec{u} = \begin{pmatrix} \text{Re}(u) \\ \text{Im}(u) \end{pmatrix}, \quad \vec{v} = \begin{pmatrix} \text{Re}(v) \\ \text{Im}(v) \end{pmatrix}, \quad \Delta \vec{u} = \begin{pmatrix} \text{Re}(\Delta u) \\ \text{Im}(\Delta u) \end{pmatrix}, \quad \ldots$$

Thus Eq. (21) and Eq. (22) can be recast as

$$\frac{\partial \Delta \vec{u}}{\partial z} = A \Delta \vec{u}, \quad \Delta \vec{u}(t, z_a) = \vec{v}(t),$$

(23)
and
\[
\frac{\partial \Delta \vec{u}'}{\partial z} = -A^T \Delta \vec{u}', \quad \Delta \vec{u}'(t, z_b) = \vec{y}(t),
\] (24)
respectively, where
\[
A = \begin{bmatrix}
- \text{Im}(u^2) & -(d/2)(\partial^2 / \partial t^2) - 2|u|^2 + \text{Re}(u^2) \\
(d/2)(\partial^2 / \partial t^2) + 2|u|^2 + \text{Re}(u^2) & \text{Im}(u^2)
\end{bmatrix}.
\]
The matrix Green’s functions \(G(z, t; z_a, \tau)\) and \(H(z, t; z_b, \tau')\) associated with Eqs. (23) and (24) solve
\[
G(z, t; z > z_a, \tau) = AG(z, t; z_a, \tau) = \delta(t - \tau)I,
\] (25a)
\[
- H(z, t; z < z_b, \tau') = A^T H(z, t; z_b, \tau') = \delta(t - \tau')I,
\] (25b)
where \(I\) is the identity matrix. Then
\[
\Delta \vec{u}(t, z) = \int G(z, t; z_a, \tau) \vec{v}(\tau) \, d\tau,
\] (26)
\[
\Delta \vec{u}'(t, z) = \int H(z, t; z_b, \tau') \vec{y}(\tau') \, d\tau'.
\] (27)
We left multiply Eq. (25a) by \(H^T\) and subtract the transpose of Eq. (25b) right multiplied by \(G\) to obtain
\[
H^T G_x + H^T_x G = H^T A G - (A^T H)^T G.
\] (28)
We will integrate this with respect to \(t\) from \(-\infty\) to \(\infty\) and with \(z\) from \(z_a\) to \(z_b\) (note the left-hand side is a perfect derivative). First consider the \(t\) integral of the right hand side and let \(A = A_1 + A_2\), where
\[
A_1 = \begin{bmatrix}
- \text{Im}(u^2) & -2|u|^2 + \text{Re}(u^2) \\
2|u|^2 + \text{Re}(u^2) & \text{Im}(u^2)
\end{bmatrix}, \quad A_2 = \begin{bmatrix}
0 & -d \frac{\partial^2}{\partial t^2} \\
\frac{d}{2} \frac{\partial^2}{\partial t^2} & 0
\end{bmatrix}.
\] (29)
Immediately, \((A_1^T H)^T = H^T A_1\) since \(A_1\) is just a matrix. Similarly,
\[
\int H^T A_2 G \, dt = \frac{d}{2} \int \begin{bmatrix}
h_{11} & h_{21} \\
h_{12} & h_{11}
\end{bmatrix} \begin{bmatrix}
- \frac{\partial^2 g_{11}}{\partial t^2} & - \frac{\partial^2 g_{12}}{\partial t^2} \\
- \frac{\partial^2 h_{11}}{\partial t^2} & - \frac{\partial^2 h_{12}}{\partial t^2}
\end{bmatrix} \, dt
\]
\[
= \frac{d}{2} \int \begin{bmatrix}
\frac{\partial^2 h_{11}}{\partial t^2} & - \frac{\partial^2 h_{11}}{\partial t^2} \\
- \frac{\partial^2 h_{12}}{\partial t^2} & - \frac{\partial^2 h_{12}}{\partial t^2}
\end{bmatrix} \begin{bmatrix}
g_{11} & g_{12} \\
g_{21} & g_{22}
\end{bmatrix} \, dt = \int (A_2^T H)^T G \, dt.
\] (30)
Therefore, integrating Eq. (28) with respect to both \(t\) and \(z\) and using Eqs. (25) we find
\[
H^T(z_a, \tau; z_b, \tau') = G(z_b, \tau'; z_a, \tau).
\] (31)
Finally we can connect the forward and adjoint propagators:
\[
\langle y, \Phi v \rangle = \int \vec{y}^T(t) \Delta \vec{u}(z_b, t) dt = \int \int \vec{y}^T(t) G(z_b, t; z_a, \tau) \vec{v}(\tau) d\tau dt
\]
\[
= \int \int \vec{v}^T(\tau) G^T(z_b, t; z_a, \tau) \vec{y}(t) d\tau dt = \int \int \vec{v}^T(\tau) H(z_a, \tau; z_b, t) \vec{y}(t) d\tau dt
\]
\[
= \int \vec{v}^T(\tau) \vec{w}(z_a, \tau) d\tau = \int \vec{w}^T(z_a, \tau) \vec{v}(\tau) d\tau = \langle \Phi^\dagger y, v \rangle.
\]

References

16. R. Holzlohrer and C. Menyuk, Use of multicanonical Monte Carlo simulations to obtain accurate bit error rates in optical communications systems, Optics