

AN ITERATIVE STOCHASTIC METHOD FOR SIMULATING LARGE DEVIATIONS AND RARE EVENTS*

GRAHAM M. DONOVAN[†] AND WILLIAM L. KATH[‡]

Abstract. We describe an iterative method capable of determining large deviations responsible for rare events of interest in lightwave systems with additive noise. The method makes use of the singular value decomposition (SVD) to efficiently compute the most important directions in state space, and a stochastic optimization scheme known as the cross-entropy (CE) method to determine the most probable manner in which these large deviations arise. Information from the SVD and CE steps of the method provides a basis for performing importance sampling with Monte Carlo simulation, allowing one to determine the probabilities of the rare events associated with such large deviations. We apply the combined method to investigate some of the mechanisms affecting large amplitude fluctuations in optical systems.

Key words. Monte Carlo simulation, variance reduction, importance sampling, cross-entropy method, singular value decomposition, Gaussian white noise

AMS subject classifications. 78M31, 65C05, 15A18, 60F10

DOI. 10.1137/100789373

1. Introduction. In many systems perturbed by noise it is rare events that are the behaviors of interest, e.g., the formation of a 35 meter rogue ocean wave or severe volcanic hazards [1]. This is also true of many engineered systems; since they are designed with “typical” behaviors in mind, system failures are often associated with deviations that are far from the mean. Because of this, the probabilities of such events can be difficult to predict.

In this paper we present a new method capable of determining large deviations and simulating rare events. We demonstrate the method in the context of nonlinear optical communication systems, but we believe it should be applicable to a wide class of rare-event simulation problems. This technique employs the singular value decomposition (SVD) [2, 3] to reduce the overall effective dimensionality of the system, the cross-entropy (CE) method [4] to locate the most probable regions of state space associated with large deviations, and importance-sampled [5, 6] Monte Carlo simulation to simulate the events and obtain estimates of their associated probabilities. The result is a method capable of studying rare events under general conditions.

To understand the need for this combination of methods, it is easiest to examine them in reverse. Importance sampling (IS) works by biasing the distributions used to select the random variables used in Monte Carlo simulations so that the events of interest occur much more often than they would otherwise [5, 6]. The typical issue with IS is in determining proper biasing distributions; this is particularly difficult when the number of random variables is large. The technique we develop here determines the

*Received by the editors March 19, 2010; accepted for publication (in revised form) March 8, 2011; published electronically June 2, 2011. This work was supported by the National Science Foundation (DMS-0709070).

<http://www.siam.org/journals/siap/71-3/78937.html>

[†]Department of Engineering Sciences and Applied Mathematics, Northwestern University. Current address: Department of Mathematics, University of Auckland, Private Bag 92019, Auckland Mail Centre, Auckland 1142, New Zealand (g.donovan@auckland.ac.nz).

[‡]Department of Engineering Sciences and Applied Mathematics, and Northwestern Institute on Complex Systems, Northwestern University, 2145 Sheridan Rd., Evanston, IL 60208 (kath@northwestern.edu).

biasing distributions using the CE method, a stochastic optimization algorithm [4, 7]. While the CE method is a powerful tool, it cannot easily handle the large number of random variables found in the lightwave systems that we discuss here. For example, in such systems the total number of random variables $N \approx N_a N_f$, where N_a is the number of amplifiers in the system (40–100) and N_f is the number of computational Fourier modes used ($\approx 512, 1024$, or more). The resulting number of random variables, easily in the hundreds of thousands, is simply too large to obtain reasonable results with currently available computing power and the CE method alone.

To simplify the problem we apply the SVD to reduce the effective dimensionality to a much smaller number consisting of only the most important “modes” of the system [8], an approach inspired by approaches to similar problems based upon soliton perturbation theory (SPT) [9, 10, 11, 12, 13, 14]. In previous approaches, the fundamental idea was that the conserved quantities in the system, which are connected to the invariances of the governing equation by Noether’s theorem [15], provide information about the special directions in state space that are most appropriate for biasing the system; the soliton structure allows analytic calculation of these modes. This is not true in general situations, of course, but these modes can still be found numerically by applying the SVD. The important point is that once the number of important directions in state space has been reduced, the CE method is capable of providing optimal biasing distributions for importance-sampled Monte Carlo simulations at a reasonable computational cost.

In this manuscript we focus on the application of this “SVD/CE/IS” method to optical communication systems, rather than a general formulation, but the restrictions on applicability are in fact rather modest, and it should be possible to apply our method to a wide array of problems. The simple constraints are that the system under consideration must feature a linearizable evolution operator, a discrete stochastic component, and rare events arising from the sum of small perturbations. These constraints are discussed in more detail in section 10.

This paper is organized as follows. In section 2 we formulate the optical transmission system problem that we wish to consider. We then present the necessary material on essential elements of this iterative, stochastic method used to simulate large deviations and rare events in this system: IS (section 3), the CE method (section 4), and the SVD (section 5). We then consider in section 6 the linearizations of the optical transmission problem required by the SVD, followed by a discussion of the use of the biasing modes, and an overall summary of the method, in section 7. We then present results of the application of this SVD/CE/IS method to two optical systems problems in section 9, and conclude with discussion of the method and potential further applications (sections 10 and 11).

2. The optical transmission system model. For definiteness, we will consider the propagation of a pulse in a nonlinear optical fiber transmission line with dispersion management [16, 17]. Pulse propagation in this situation is governed by the nonlinear Schrödinger (NLS) equation,

$$(2.1) \quad \frac{\partial U}{\partial z} = \frac{i}{2} d(z) \frac{\partial^2 U}{\partial t^2} + i\gamma |U|^2 U + \sum_{k=1}^{N_a} N_k(t) \delta(z - kz_a),$$

where U is the complex-valued envelope of the optical field, z is the propagation distance, t is retarded time, and fiber loss and periodic amplifier gain have been averaged out [16, 17]. Dispersion management in this case means that the dispersion

coefficient $d(z)$ is a periodic function with period equal to the dispersion map length z_a ; i.e., it is given by the periodic extension of

$$(2.2) \quad d(z) = D_{\text{avg}} + \begin{cases} 4sD_{\text{avg}}, & \left| \frac{z}{z_a} - \frac{1}{2} \right| < \frac{1}{4}, \\ -4sD_{\text{avg}}, & \left| \frac{z}{z_a} - \frac{1}{2} \right| \geq \frac{1}{4}, \end{cases} \quad \frac{z}{z_a} \in [0, 1].$$

Here D_{avg} is the average dispersion, and the parameter s is a measure of the strength of the dispersion map. In addition, in (2.1), γ is the nonlinear coefficient, and $N_k(t)$ is a term representing noise that is added by the amplifiers. Because the noise grows due to gain from subsequent amplifiers, it is known as *amplified spontaneous emission* (ASE) noise [16]. Here, it will be modeled as additive Gaussian white noise added at the equally spaced amplifiers, with a period equal to that of the dispersion map. The noise itself is then zero-mean and delta-correlated,

$$(2.3) \quad \langle N_k(t) \rangle = 0,$$

$$(2.4) \quad \langle N_k(t)N_k^*(t') \rangle = \sigma^2(G_a)\delta(t - t'),$$

although this is technically incorrect since perfectly delta-correlated noise has an infinite bandwidth, which is physically impossible. Typically, however, the noise bandwidth is much larger than that of the pulse, and delta-correlated noise is often used as a shorthand in this situation. In the above, $\sigma^2(G_a)$ is a gain-dependent variance; $\sigma^2(G_a) = \omega_0 n_{\text{sp}}(G_a - 1)^2 / (G_a \log^2 G_a)$, where G_a is the amplifier gain; \hbar is Planck’s constant; ω_0 is the center carrier frequency; and n_{sp} is the spontaneous emission factor.

In an optical transmission line, one specifies the input to the system, which is done by giving the pulse profile $U(0, t)$, and one is interested in the output signal $U(z_L, t)$, where z_L is the final position of the transmission line. As it can be difficult to measure optical fields directly, the optical signal is often converted into an electrical signal with an optical detector. The physics of such devices can be involved, but one relatively simple detector model is an integrate-and-dump detector [18], for which the output voltage is given by

$$(2.5) \quad V = \int_{-T/2}^{T/2} |U_F(z_L, t)|^2 dt.$$

Here $U_F(z_L, t)$ is the pulse envelope after optical filtering using a given transfer function $F(\omega)$ in the Fourier domain. Essentially, an integrate-and-dump detector accumulates all of the output light energy within a given bit slot (of width T , here centered at $t = 0$). It is also common to filter the signal in the electrical domain; (2.5) represents only a very simple electrical filter, and we will not consider more sophisticated electrical filters in the present work.

3. Importance sampling. We begin by discussing the mathematical structure surrounding the calculation of a *single* specific rare event in a system perturbed by noise. While our final interest is in determining the full probability density function (PDF) resulting from a full set of rare events, it is convenient to first discuss single events before describing how to combine their individual contributions.

Monte Carlo methods estimate an expected value using random sampling. Consider the problem of estimating $E[\phi] = \int \phi(\vec{x})f(\vec{x})d\vec{x}$, where $\phi(\vec{x})$ is a known function

and $f(\vec{x})$ is a PDF. A traditional Monte Carlo estimator would be

$$(3.1) \quad \hat{E}_N = \frac{1}{N} \sum_{i=1}^N \phi(\vec{x}_i),$$

where the N random samples \vec{x}_i have been drawn from the distribution f . We are interested in rare events—that is, situations where $E[\phi]$ is small. In such cases, this estimator may require $N > 1/E[\phi]$ samples to provide a good estimate [5, 6], so for rare events the rate of convergence of standard Monte Carlo sampling may be slow.

IS is a variance reduction technique that can provide a dramatic performance increase over traditional Monte Carlo simulations [5, 6]. The idea is to concentrate the random samples in the portion of sample space where these rare events occur. We introduce a *biasing distribution* $f^*(\vec{x})$, so that our estimator becomes

$$(3.2) \quad \hat{E}_N^* = \frac{1}{N} \sum_{i=1}^N \phi(\vec{x}_i) \frac{f(\vec{x}_i)}{f^*(\vec{x}_i)} = \frac{1}{N} \sum_{i=1}^N \phi(\vec{x}_i) w(\vec{x}_i),$$

where

$$(3.3) \quad w(\vec{x}) = \frac{f(\vec{x})}{f^*(\vec{x})}$$

is known as the *likelihood ratio* [5, 6, 19], and now we are sampling to determine the expected value of $\phi(\vec{x})w(\vec{x})$ with samples drawn from $f^*(\vec{x})$. The likelihood ratio can be viewed as a correction factor, appropriately weighting a sample trial *as if it had been drawn from the original distribution*, despite having been drawn from a biased distribution.

If the biasing distribution is chosen properly, so that the majority of samples fall in regions where $\phi(\vec{x})$ is concentrated, a significant reduction in the number of samples required to obtain a good estimate can be realized. The difficulty with IS, however, is in choosing the biasing distribution correctly. An ideal biasing distribution does exist, and it provides an exact estimate with just a single Monte-Carlo sample [5], but to determine this distribution one must know the desired quantity a priori. At the same time, a poor biasing distribution can slow convergence.

It is also useful to consider a real-valued *performance function* $P(\vec{x})$ which depends on a random vector \vec{x} drawn from the PDF $f(\vec{x})$. Suppose we are interested in the rare event where the performance function is greater than some threshold value: $P(\vec{x}) \geq \hat{P}$. Replacing $\phi(\vec{x})$ in (3.1) with an *indicator function* $I_{\{ \cdot \}}$, which is defined to be 1 if the argument is true and 0 otherwise, the Monte Carlo estimator for the probability of this rare event becomes

$$(3.4) \quad \hat{I}_N = \frac{1}{N} \sum_{i=1}^N I_{\{P(\vec{x}_i) \geq \hat{P}\}},$$

where, as before, the \vec{x}_i are drawn from the PDF $f(\vec{x})$. The importance-sampled estimator (3.2) is then

$$(3.5) \quad \hat{I}_N^* = \frac{1}{N} \sum_{i=1}^N I_{\{P(\vec{x}_i) \geq \hat{P}\}} \frac{f(\vec{x}_i)}{f^*(\vec{x}_i)}$$

with the \vec{x}_i drawn from f^* . This is the formulation we will use in the following section.

4. The cross-entropy method. The CE method [7] is a stochastic optimization algorithm which we will use to solve the biasing problem in a reduced-dimensional space. We will present only a simplified version of the CE algorithm necessary for understanding the method developed here; full details may be found in the literature [7, 20, 21].

The CE method constructs an approximation to the ideal biasing distribution by considering a parametric family of such distributions and determining the specific distribution within this family that minimizes the *Kullback–Leibler (KL) distance* [4, 21, 22] between it and the optimal distribution, f_{opt}^* . A stochastic optimization program is posed to find the optimal parameter; although the ideal biasing is not known in closed form, for several classes of probability distributions (including Gaussians) it is possible to estimate the minimum [4]. In this work, f will be a Gaussian distribution, and the biasing distributions are assumed to belong to a family of mean-shifted Gaussian distributions with the same variance as f ; this is one of the cases for which it is possible to estimate the minimum.

In particular, for two functions $g(\vec{x})$ and $h(\vec{x})$ the KL distance is defined as [22]

$$(4.1) \quad \mathcal{D}(g, h) = \text{E}_g \left[\ln \frac{g(\vec{x})}{h(\vec{x})} \right] = \int \ln(g(\vec{x})) g(\vec{x}) d\vec{x} - \int \ln(h(\vec{x})) g(\vec{x}) d\vec{x}.$$

This is also known as the cross-entropy between two probability distributions. (The KL distance is not a true metric, however; for example, it is not symmetric.) If we take $g(\vec{x}) = f_{\text{opt}}^*(\vec{x})$ and $h(\vec{x}) = f^*(\vec{x})$, the first integral on the right-hand-side of (4.1) is fixed, and minimizing the cross-entropy between $f_{\text{opt}}^*(\vec{x})$ and $f^*(\vec{x})$ is equivalent to maximizing $\int \ln(f^*(\vec{x})) f_{\text{opt}}^*(\vec{x}) d\vec{x}$. Noting that the optimal biasing distribution is [5]

$$f_{\text{opt}}^*(\vec{x}) \propto I_{\{P(\vec{x}) > \hat{P}\}} f(\vec{x}),$$

this problem is equivalent to maximizing $\text{E}[I_{\{P(\vec{x}) > \hat{P}\}} \ln(f^*(\vec{x}))]$.

Suppose that, as we will do here, the potential biasing distributions for \vec{x} are selected from a parameterized family $\{f^*(\vec{x}; \vec{v})\}$, where \vec{v} is a vector of parameters, and let the unbiased distribution be denoted by $f^*(\vec{x}; \vec{v}_0) = f(\vec{x})$. In this case one must look for the member of this family that is closest to the optimal distribution. Based on the above discussion, one must then maximize the integral

$$(4.2) \quad D(\vec{v}) = \int I_{\{P(\vec{x}) > \hat{P}\}} \ln(f^*(\vec{x}; \vec{v})) f(\vec{x}) d\vec{x}.$$

This usually must be done numerically, i.e., by using Monte Carlo sampling. Because the optimal biasing distribution is typically far from the unbiased distribution, however, the member of the family that is closest to optimal is also likely to be far from $f(\vec{x})$. Thus, determining the best choice for \vec{v} also becomes a rare event simulation. The problem, of course, is that the region of interest in sample space where rare events occur is generally far from the region in sample space where the unbiased distribution $f(\vec{x})$ is large.

The solution to this problem is to avoid attempting to jump directly to the region where rare events occur, but rather to determine a sequence of intermediate regions that reach the desired region in a series of steps, e.g., regions where the performance function $P(\vec{x})$ is greater than intermediate values \hat{P}_j . In addition, one employs an importance-sampled version of (4.2). Let $D_j(\vec{v})$ be the integral in (4.2) with \hat{P} replaced by \hat{P}_j . Starting with the unbiased distribution, one uses Monte Carlo sampling to minimize the CE distance between the parameterized distribution and the optimal

distribution that reaches \hat{P}_1 . This step, which is done by finding the maximum of $D_1(\vec{v})$ over this first set of samples, will give a parameter vector \vec{v}_1 . One then uses this value to define a biasing distribution, and performs an IS-Monte Carlo simulation with this distribution to minimize the CE distance between the parameterized distribution and the optimal distribution that reaches a second level $\hat{P}_2 > \hat{P}_1$. Of course, since a biasing distribution is being used, each step of the procedure is an importance-sampled Monte Carlo (ISM) simulation of a stochastic optimization program. That is, at step j , one must compute

$$(4.3) \quad \vec{v}_{j+1} = \operatorname{argmax}_{\vec{v}} \hat{D}_j(\vec{v}),$$

where

$$(4.4) \quad \hat{D}_j(\vec{v}) = \frac{1}{M} \sum_{m=1}^M I_{\{\hat{P}(\vec{x}) > \hat{P}_j\}} L(\vec{x}^{(m)}) \ln(f^*(\vec{x}^{(m)}; \vec{v})),$$

$\vec{x}^{(1)}, \dots, \vec{x}^{(M)}$ are i.i.d. samples generated according to $f^*(\vec{x}; \vec{v}_j)$, and $L(\vec{x}^{(m)})$ is the likelihood ratio $f(\vec{x})/f^*(\vec{x}; \vec{v}_j)$. The optimal biasing distribution then can be adaptively determined by performing the following steps [20].

Algorithm 1.

1. Set $j = 0$ and the initial parameter vector \vec{v}_0 .
2. Generate ISM samples according to $f^*(\vec{x}; \vec{v}_j)$ ($j = 0$ is unbiased).
3. Solve (4.3) to find \vec{v}_{j+1} .
4. If the iteration has converged (e.g., the final value of \hat{P} has been reached; see [20] and below), stop; otherwise increase j to $j + 1$ and reiterate from step 2.

Once the iteration has converged, one can then perform ISM simulations using the biasing distribution $f^*(\vec{x}; \vec{v}_{\text{final}})$.

The regions determined by the values \hat{P}_j can also be defined in terms of *sample quantiles* of some quantity of interest [20]. For example, if the goal is to produce large values of voltage V , at each iteration we can define the region as those parts of sample space that produce the largest fraction, e.g., 1%, of all of the randomly generated voltage samples. In this way the iteration systematically moves to larger and larger values of voltage. The iteration is then considered to have converged when the sample quantiles cross some predefined voltage threshold, or, alternatively, when the probability associated with the voltage value is sufficiently small.

A major issue associated with the above algorithm is obviously how to accomplish step 3. Solving (4.3) can be complicated in general. However, if the function $D(\vec{v})$ is convex and differentiable, the solutions of (4.3) can be obtained by solving a system of equations [20]:

$$(4.5) \quad \frac{1}{M} \sum_{m=1}^M I_{\{P(\vec{x}) > \hat{P}_j\}} L(\vec{x}^{(m)}; \vec{v}_0, \vec{v}) \nabla_{\vec{v}} \ln f^*(\vec{x}^{(m)}; \vec{v}) = 0.$$

For many standard distribution families this can be solved analytically. This is the case here, since Gaussian random variables are being used.

An alternative way of updating the biasing distributions is with sample moments [21]. Regardless of which updating method is used, the process is repeated until the target threshold is exceeded with the current biasing distribution. This distribution

then solves the stochastic optimization problem to minimize the KL distance, and thus becomes the biasing distribution used in the importance sampled estimator (3.5) for the rare event of interest, $P(\vec{x}) \geq \hat{P}$.

The above demonstrates how to simulate one particular rare event using ISMC and the CE method. We are interested in simulating full PDFs, however, which often requires taking into account several biasing distributions, since the overall distribution may depend upon the contributions from several regions of sample space. We refer to each of these specific regions as *biasing targets*, and for each individual biasing target a corresponding optimal biasing distribution is found, as above. Each optimal biasing distribution can then be used in ISMC simulations, and the results are combined into a single resultant PDF by the use of the *balance heuristic* [6, 23]. The balance heuristic stipulates that each sample should be combined in the overall statistics with a multiplicative weighting factor,

$$(4.6) \quad w_j(\vec{x}_j^{(m)}) = \frac{f_j^*(\vec{x}_j^{(m)})}{\sum_{k=1}^K f_k^*(\vec{x}_j^{(m)})},$$

where K optimal biasing distributions f_k^* are used (and the sample $\vec{x}_j^{(m)}$ is drawn from the j th distribution f_j^*). Note that samples are weighted according to the relative probability that they are generated from that particular distribution. In the above it is assumed that an equal number of samples are drawn for each biasing target, as is the case in what follows. In this way, the results of simulations using multiple biasing distributions are combined into a single output PDF.

5. The singular value decomposition. As mentioned earlier, the goal is to reduce the effective dimensionality of the problem by finding the most important “modes” or directions in state space and by biasing only in these particular directions. In previous work, it has been shown that this can be done by (1) finding the most probable noise-induced change at each amplifier leading to a given deviation in a measured quantity at the output and (2) finding the most probable combination of these changes across all amplifiers [9, 10, 11, 12, 13, 14]. A key observation is that although the overall deviation at the output can be large, the change induced at each amplifier remains small—it is only when the many small changes combine together in concert that the overall deviation becomes large.

Because the change at each amplifier remains small, the deviation at the output resulting from that change can be determined by linearizing the equation about the unperturbed solution. Once the equation has been linearized, the SVD can be used to efficiently determine the most important modes in the system [8, 24], i.e., those changes that lead to large deviations at the output.

5.1. Numerical implementation of the SVD. Suppose M is an $m \times n$ matrix representing the discretized version of the linear operator mapping the change at an amplifier to the change in the output. Then the SVD is a factorization of the form $M = \mathbf{P}\mathbf{\Sigma}\mathbf{Q}^*$, where

- \mathbf{Q} is $n \times n$ and contains the orthonormal input basis vectors,
- \mathbf{P} is $m \times m$ and contains the orthonormal output basis vectors, and
- $\mathbf{\Sigma}$ is an $m \times n$ matrix with the singular values along the diagonal.

We will also use the facts that the squares of the nonzero singular values of M are equal to the nonzero eigenvalues of M^*M and that the columns of \mathbf{Q} are the eigenvectors of M^*M [3]. Rather than finding all of the elements of the matrix M , we interpret

it as a linear operator and employ an iterative eigendecomposition algorithm to the action M^*M . As a result, it will not be necessary to form the matrix M at all. For this, we use the implicitly restarted Arnoldi method as implemented in the package ARPACK [25].

We are then interested in the largest singular values and their corresponding input basis vectors. For the largest singular values, we refer to these input basis vectors as the SVD-determined modes of the linear operator.

6. Linearized NLS equation. The first part of the operator M arises from the discretized version of the linearized NLS equation. If in (2.1) we let $U = u_0 + \epsilon \Delta u$ and keep only terms of $O(\epsilon)$, we obtain

$$(6.1) \quad \frac{\partial \Delta u}{\partial z} = \frac{i}{2} d(z) \frac{\partial^2 \Delta u}{\partial t^2} + i\gamma(2\Delta u|u_0|^2 + \Delta u^* u_0^2),$$

where the nominal solution u_0 is the noiseless propagation of the initial pulse governed by the NLS equation. We can split this linearized complex equation into real and imaginary parts and deal with a vector system, so that determining the adjoint becomes straightforward. We define $\Delta u = u_1 + iu_2$ and $u_0^2 = \xi + iv$ to obtain the vector evolution equation

$$(6.2) \quad \frac{\partial}{\partial z} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} -\gamma v & -\frac{1}{2}d(z)\frac{\partial^2}{\partial t^2} - 2\gamma|u_0|^2 + \gamma\xi \\ \frac{1}{2}d(z)\frac{\partial^2}{\partial t^2} + 2\gamma|u_0|^2 + \gamma\xi & \gamma v \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}.$$

We will also define the propagation operator \mathcal{L} as the linear operator that maps a change $\vec{u} = [u_1 \ u_2]^T$ from $z = z_c$ to $z = z_L$ via (6.2), where z_c is the current amplifier location.

An alternate representation of the operator \mathcal{L} arises by rewriting (6.2) in more general form as

$$\frac{d\vec{u}}{dz} = A(z, t)\vec{u}.$$

The solution to this linearized problem can be written in terms of the matrix Green's function [26] as $G(z, t; z_0, t_0)$,

$$(6.3) \quad \vec{u}(z, t) = \int G(z, t; z_0, t_0) \vec{u}(z_0, t_0) dt_0,$$

where

$$(6.4) \quad \frac{d}{dz} G(z, t; z_0, t_0) = A(z, t)G(z, t; z_0, t_0)$$

with $G(z_0, t; z_0, t_0) = \mathcal{I}\delta(t - t_0)$ and where \mathcal{I} is the identity matrix.

To obtain the adjoint operator \mathcal{L}^\dagger , we consider the adjoint Green's function. Because the differential operators in the matrix in (6.2) are self-adjoint, computing the adjoint evolution operator is relatively simple:

$$(6.5) \quad -\frac{\partial}{\partial z} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} -\gamma v & \frac{1}{2}d(z)\frac{\partial^2}{\partial t^2} + 2\gamma|u_0|^2 + \gamma\xi \\ -\frac{1}{2}d(z)\frac{\partial^2}{\partial t^2} - 2\gamma|u_0|^2 + \gamma\xi & \gamma v \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}.$$

Then the adjoint propagation operator \mathcal{L}^\dagger is found by solving (6.5) in the reverse direction, from $z = z_L$ to $z = z_c$.

Numerically, these linearized propagators are solved using an integrating factor fourth-order Runge–Kutta method [2], while the solution to the full nonlinear problem is obtained via the split-step Fourier method [27]. Note that it is desirable to have the linear solver z -step size be an integer multiple of the nonlinear solver step size, because at each solution point for the linear solver, the nonlinear solution must be available (ideally, precalculated and stored).

As discussed in section 2, we are interested in determining large deviations in the output voltage detected at the end of the transmission line. As such, we must also linearize the detector in addition to the propagation equation: we need the linearization of the full nonlinear mapping from a change at an amplifier to the output voltage V . Details of this calculation can be found in the appendix.

7. Biasing with the SVD modes. As mentioned in section 2, when loss is compensated by periodic amplification, the signal gain is accompanied by amplified spontaneous emission noise [16], which is modeled as Gaussian white noise added at each amplifier. Let us suppose for simplicity that the amplifiers are equally spaced throughout the system and that the distance between amplifiers is z_a . That is, if there are N_a amplifiers, the system begins at $z = z_0$ and ends at $z = z_0 + N_a z_a$, with the signal amplified at the intermediate points $z = z_0 + j z_a$ for $j = 1, \dots, N_a$. If we are using N_f Fourier modes in our computational scheme, then the simulated system employs $N = N_a N_f$ random variables for the noise components; i.e., at each amplifier we have

$$(7.1) \quad U(z_0 + j z_a^+, t) = U(z_0 + j z_a^-, t) + \mathcal{F}^{-1} \left[\sum_{\omega=-N_f/2}^{N_f/2} \tilde{r}_{\omega,j} e^{i\omega t} \right],$$

where the $\tilde{r}_{\omega,j}$ are the random noise components. Note that these noise components are complex; i.e., both the real and imaginary parts are zero-mean Gaussian random variables. The Fourier domain is usually the most natural one in which to consider adding the random noise, but this is true only in the unbiased case; hence, in what follows, we will revert to the time domain in order to develop the biasing by the SVD modes. The equivalence of the two approaches will be discussed at the end of the section.

As described in section 5, the SVD is applied to the lightwave system to reduce the noise dimensionality. At each amplifier, the SVD determines the input singular modes which produce the largest changes in the output; this results in the most probable way to produce the desired change *from that amplifier forward*. Suppose that we use $n_m \ll N_f$ SVD modes at each amplifier; call these modes $u_{i,j}(t)$ for $i = 1, \dots, n_m$ and $j = 1, \dots, N_a$; i.e., $u_{i,j}(t)$ is the i th mode at the j th amplifier. The next step is to determine the *amount* of each mode that should be added in order to provide the proper biasing for the ISMC simulations.

This biasing problem is posed in the following way: suppose there is only a single SVD mode and at each amplifier location, $z = z_0 + j z_a$; we add to our solution $U(z_0 + j z_a, t)$ a multiple of this SVD mode with the amount determined by a *biasing coefficient* η_j , i.e.,

$$(7.2) \quad U(z_0 + j z_a^+, t) = U(z_0 + j z_a^-, t) + \eta_j u_j(t).$$

The goal is to determine the η_j by the CE method. We do this by solving the constrained optimization problem, i.e., find the coefficients η_j to reach a target voltage

\hat{V} in the most probable way. Because $\eta_j u_j(t)$ can be thought of as a possible realization of a Gaussian white noise source, maximizing the probability of this specific event's actually occurring is then equivalent to minimizing

$$(7.3) \quad \sum_j \int |\eta_j u_j(t)|^2 dt = \sum_j \eta_j^2$$

(here we have assumed that the norm of the SVD mode is unity), subject to the constraint that

$$(7.4) \quad V = \int_{-T}^T |U_F(z_L, t)|^2 dt \geq \hat{V}.$$

The performance function for the CE method, $P(\vec{x})$, is then given for this system as $P = V = V(\vec{\eta})$, where we explicitly write the dependence on the biasing parameters $\vec{\eta} = [\eta_1, \eta_2, \dots, \eta_N]$, and these biasing parameters are equivalent to the general CE method parameter vector \vec{x} . Selecting the target voltage \hat{V} (or set of target voltages) is an important step in building a set of biasing distributions. In general, it may be necessary to use several values of \hat{V} which cover the areas of system state space which one is interested in simulating. If the desired simulation result isn't an entire voltage PDF, but merely the probability that a specific output voltage is exceeded, it may be possible to use only one biasing target and set the value of \hat{V} to correspond to this specific condition.

Because we have connected the biasing coefficients of the SVD modes with the probability of these modes occurring, when determining the biasing distributions the system is noiseless aside from the biasing coefficients; these coefficients are, of course, randomly assigned by the CE method. The performance function therefore depends only upon $\vec{\eta}$, reflecting the significant and necessary reduction in dimensionality. As far as the CE method is concerned, however, this is a black box procedure—the N randomly drawn biasing coefficients determine the output, and the goal is to find the most probable combination of coefficients that produces a desired output voltage.

In the general case where more than one SVD mode is required at each amplifier, one has instead

$$(7.5) \quad U(z_0 + jz_a^+, t) = U(z_0 + jz_a^-, t) + \sum_{i=1}^{n_m} \eta_{i,j} u_{i,j}(t).$$

Here the CE method need control only $n_m N_a$ random variables, a dramatic decrease from the total possible number, $N_f N_a$. This reduction makes the problem computationally feasible, and we are able to generate the appropriate biasing distributions using this formulation. The number of SVD modes required is problem-dependent. Soliton perturbation theory, from which this method evolved, suggests that a maximum of four modes might be required for each pulse when simulating optical systems; physical arguments for specific systems (phase invariance, for example) might reduce this further, although this is not strictly necessary. Alternatively, an ad hoc decision can be made based on the relative magnitude of the computed singular values. Note that computing extra SVD modes involves only modest additional computational cost in the SVD itself, but this also increases the number of random variables under CE method control and thus the number of samples required in the biasing process.

Once the biasing coefficients have been determined and the goal is to perform ISMC simulations, the noise addition is modified. While the reduced problem is

appropriate for generating the biasing distributions, the full simulations require full noise in each Fourier mode for complete accuracy. The simulation is biased toward the rare events of interest by using the SVD modes, multiplied by the now-fixed biasing coefficients, to shift the means of the added Gaussian noise. The addition of biased full-bandwidth noise thus appears as a combination of two terms, i.e.,

$$(7.6) \quad U(z_0 + jz_a^+, t) = U(z_0 + jz_a^-, t) + \sum_{i=1}^{n_m} \eta_{i,j} u_{i,j}(t) + \mathcal{F}^{-1} \left[\sum_{\omega=-N_f/2}^{N_f/2} \tilde{r}_{\omega,j} e^{i\omega t} \right].$$

Here the $\eta_{i,j}$ are the biasing coefficients as determined via the CE method; these no longer change. The $\tilde{r}_{\omega,j}$ are Gaussian white noise coefficients, drawn from distributions with 0 mean. The noise terms of (7.6) can also be rewritten as

$$(7.7) \quad \sum_{i=1}^{n_m} \eta_{i,j} u_{i,j}(t) + \mathcal{F}^{-1} \left[\sum_{\omega=-N_f/2}^{N_f/2} \tilde{r}_{\omega,j} e^{i\omega t} \right] = \mathcal{F}^{-1} \left[\sum_{\omega=-N_f/2}^{N_f/2} \tilde{R}_{\omega,j} e^{i\omega t} \right],$$

where $\tilde{R}_{\omega,j}$ are Gaussian white noise terms in the Fourier domain, mean-shifted by the Fourier transform of the SVD modes multiplied by the biasing coefficients. Mean-shifted normal distributions allow for a particularly simple likelihood ratio calculation, though in principle much more general distributions could easily be used [28]. All that remains, then, is to collect the desired statistics at the end of the transmission line and apply the correction for the biasing via the likelihood ratio.

8. Summary overview of the methodology. A schematic of the overall method is presented in Figure 8.1. Here the dashed lines represent connections or processes which are executed multiple times to collect statistics, due to their stochastic nature. The solid lines represent connections or processes which are performed only once, based upon the aggregation of the stochastic trials.

The method is divided into two parts: a biasing stage and a simulation stage, which occur separately. The output of the biasing stage is required to start the simulation stage. In the biasing stage, given a performance threshold and a small number of modes determined by the SVD, the CE method solves an optimization problem to determine the most probable combinations of these modes, in terms of the biasing coefficients, that can achieve this performance threshold. The CE method is stochastic and iterative, and many realizations of this reduced-dimensionality system are required to obtain the optimal coefficient values.

When the CE method converges, the optimal mode biasing coefficients are interpreted as mean shifts of the original, full-dimensionality noise PDFs. These mean-shifted distributions thus provide the *biasing distributions* required for ISMC simulations of the full system. Here, because the mean shifts are specified in terms of the SVD-determined modes, during the simulation it is necessary to reapply the SVD to determine these mode shapes. Because the simulated statistics are corrected using the likelihood ratios, one obtains realistic probability estimates of the simulated system with much-reduced simulation variance. In particular, accurate tails of the probability distributions associated with quantities of interest can be determined.

As can be seen from Figure 8.1, the components of the method depend little upon the details of the system under consideration. Furthermore, the requirements imposed by the SVD/CE/IS scheme are quite modest; these will be discussed in detail in the context of the specific application.

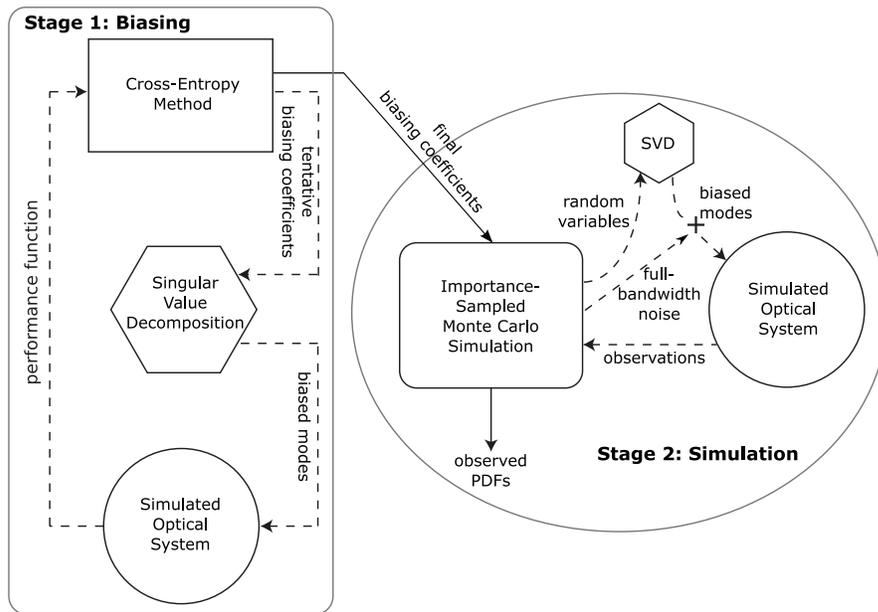


FIG. 8.1. Schematic of the overall method: *Dashed lines represent stochastic elements repeated many times. Solid lines represent elements executed only once. The SVD is employed to reduce the noise dimensionality of the simulated problem to a tractable level for the CE method. The left portion of the diagram represents the biasing stage of the method, and the right portion the simulation stage.*

9. Results.

9.1. Application to a dispersion-managed system. As a demonstration and validation of the SVD/CE/IS method, we will apply it both to a previously studied problem in rare-event simulation and to a problem that has not been solved before. For the first problem, we simulate a sample system with large amplitude pulse distortions in a dispersion-managed communication system [12]. This problem has been studied using a semianalytic method to generate the biasing distributions needed for ISMC simulations, and thus results are available for comparison purposes. Note, however, that these previous results were obtained using an averaged, nonlocal governing equation, the dispersion-managed nonlinear Schrödinger equation (DMNLS) [12]. The DMNLS is an averaged approximation to the NLS equation with variable coefficients. We will apply the SVD/CE/IS method to the full NLS equation with variable coefficients, however, since this equation is more fundamental.

We consider a system with a total propagation distance $z_L = 4000$ km, an average dispersion $D_{avg} = 0.15$ ps²/km, and a dispersion map period $z_a = 100$ km. The nonlinear coefficient is $\gamma = 1.7(W - \text{km})^{-1}$, and the fiber power loss rate is 0.21 dB/km. The noise spontaneous emission factor is $n_{sp} = 1.5$, and the amplifier gain G_a is set to exactly compensate the fiber loss. The initial condition used is a *dispersion managed soliton* (DM soliton) with amplitude $\lambda = 2$ and full-width, half-max power pulsewidth 17 ps computed numerically [12], and the DM map strength is $s = 4$. Simulations are performed in a 100 ps wide computational window using 128 Fourier modes, and a 50 ps wide detector window. In addition, just before the detector a 10 GHz Gaussian optical filter is used (see the appendix). The CE method parameters are 20,000 trials per iteration, with a maximum of 6 iterations, and the best-performing 10% of trials

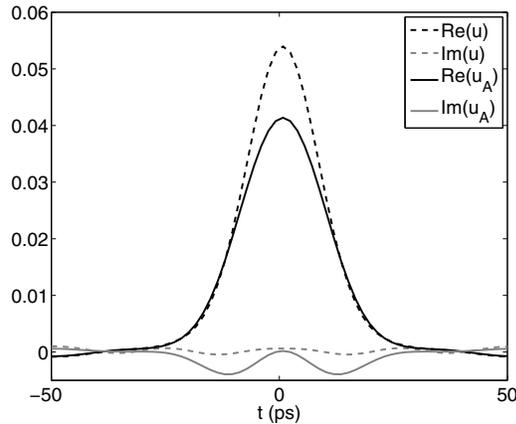


FIG. 9.1. Amplitude mode computed with filter and detection: *The normalized amplitude mode (solid curves) overlaid with the underlying pulse shape (dashed curves) when the mode is extracted from a system with optical filtering and an integrate and dump detector. Both real (black) and imaginary (gray) parts displayed.*

are used to determine the subsequent biasing coefficients (see Algorithm 1, section 4). The subsequent ISMC simulations used 100,000 trials.

Figure 9.1 shows the normalized amplitude mode from a sample trial in the filtered system, along with the (unfiltered) pulse shape at the amplifier where the mode was computed. For illustration purposes both the pulse and the mode have been phase-shifted so that the phase at pulse center is zero; modes and pulses used in the actual computations are not phase-shifted, of course. This perturbation produces the largest change in V when it is propagated to the output. Note that this pulse and mode are from a biasing trial, so full-bandwidth noise is not present but rather only the noise projected in the direction of the SVD mode. In the “pure” soliton regime (infinite domain, no detector), the adjoint amplitude mode is simply the pulse itself (i.e., see [11]). While here the corresponding numerical mode clearly has a similar character, the optimal direction does deviate significantly from the pure mode.

We use the SVD/CE/IS method to compute the amplitude PDF for this system, as described previously [12]. First, using the CE method, we generate the biasing coefficients for one SVD-determined mode, which we have referred to as the amplitude mode. A total of seven biasing targets, corresponding to different specific rare-event outcomes, are used: three biasing distributions which result in progressively larger *increases* in output voltage; three biasing distributions which result in progressively larger *decreases* in output voltage; and one unbiased target resulting in the nominal output voltage. Simulations using the full set of optimal biasing distributions are combined using the balance heuristic. The comparison with previous results is shown in Figure 9.2, and the agreement is seen to be very good. The coefficient of variation (the sample standard deviation divided by the mean) is shown in the lower panel to allow the statistical convergence of the simulations to be assessed. These results show clearly that, even though none of the specific mathematical structure of DM solitons has been employed, the SVD/CE/IS method is capable of simulating rare events in this system with a high degree of accuracy.

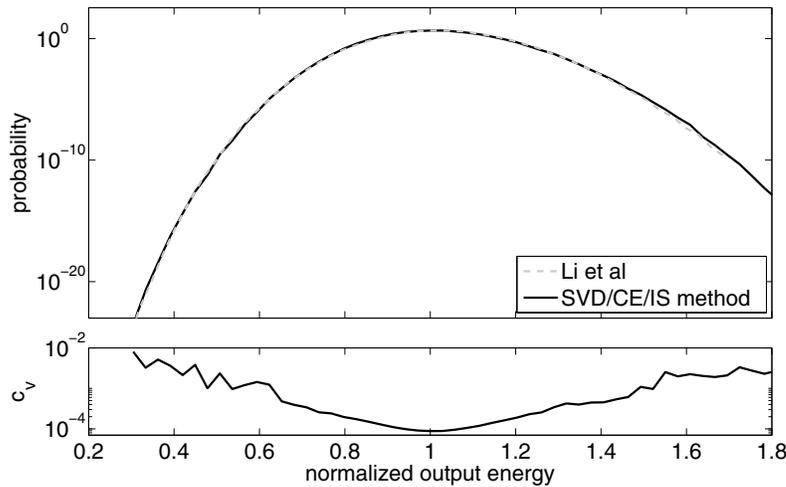


FIG. 9.2. Simulated PDF for DM soliton system: *Top*, the full simulated PDF for the DM soliton system (black, solid) compared with the same PDF generated previously [12] (gray, dashed). *Bottom*, the coefficient of variation of the current simulations.

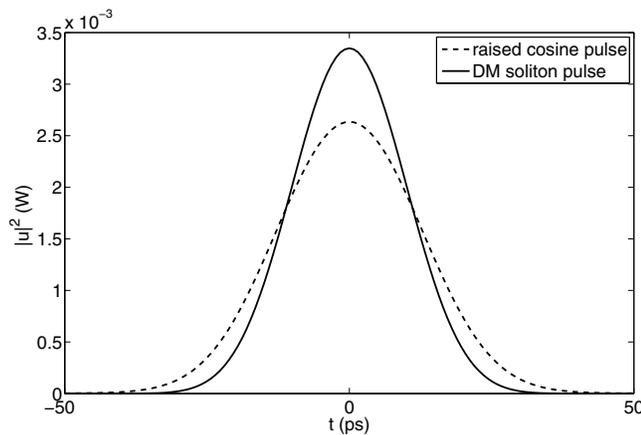


FIG. 9.3. Comparison of pulse shapes: *The two pulse shapes used in the different dispersion-managed systems. The DM soliton shape is the solid curve, and the raised cosine shape is the dashed curve.*

9.2. Beyond analytically tractable problems. As a more practical problem we now consider a 50% duty-cycle, raised cosine initial pulse shape

$$(9.1) \quad U_0 = \frac{1 + \cos(\pi \sin(\pi t/100 \text{ ps}))}{2c_{\text{DM}}}.$$

Here the total pulse power is matched to that of the DM soliton via the normalization factor c_{DM} , numerically determined from the DM soliton. Although we could select any pulse shape, the selected pulse shape is of interest in practical systems (i.e., see [29]). The pulse shapes are compared in Figure 9.3. Note that by eye the difference in pulse shape is not overly dramatic, and there is no difference in pulse power. In

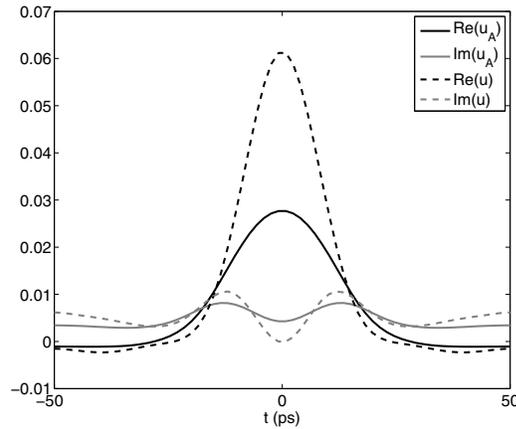


FIG. 9.4. Pulse shape and computed amplitude mode from sample trial: *The pulse shape at the final amplifier (dashed) from the sample trial in Figure 9.5, along with the SVD-computed amplitude mode at that location (solid). Both real and imaginary parts displayed, in black and gray, respectively.*

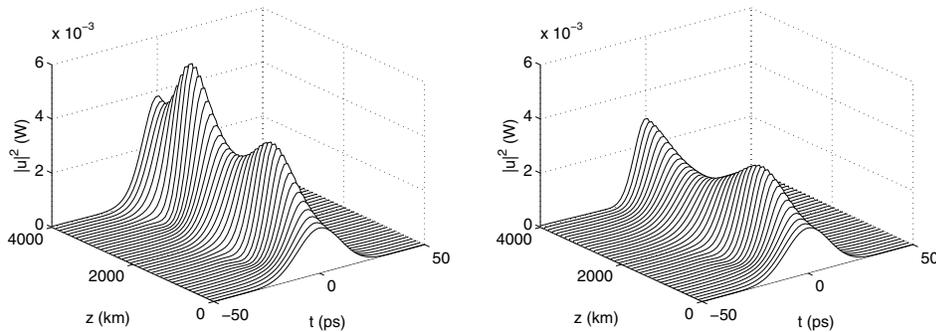


FIG. 9.5. Sample biasing trial for the raised-cosine pulse: *Left panel: a sample Monte Carlo trial from the biasing stage of the method for the raised-cosine initial pulse shape. For this trial, a biasing target exhibiting an increase in pulse energy at the detector is chosen. Right panel: noiseless propagation of the raised-cosine pulse with no biasing or perturbation, for comparison.*

terms of system behavior, however, the change is more noticeable. Without the DM soliton initial condition, the nominal solution is no longer stroboscopically stationary; i.e., it no longer returns to the same pulse shape after each dispersion map period.

We apply the SVD/CE/IS method precisely as before—only the initial condition is changed. In particular, the SVD amplitude mode is computed numerically at each amplifier as before. As an example, Figure 9.4 shows the amplitude mode and the underlying pulse shape at the last amplifier. As before, this mode is the perturbation producing the largest change in output energy for this specific pulse, from this point forward.

When these most probable perturbations are added to the pulse at each amplifier in turn, a specific example exhibiting a large change in output energy is obtained. A single sample trial from a biasing (reduced noise dimensionality) run is shown in the left panel of Figure 9.5. This particular sample trial is biased toward an increase

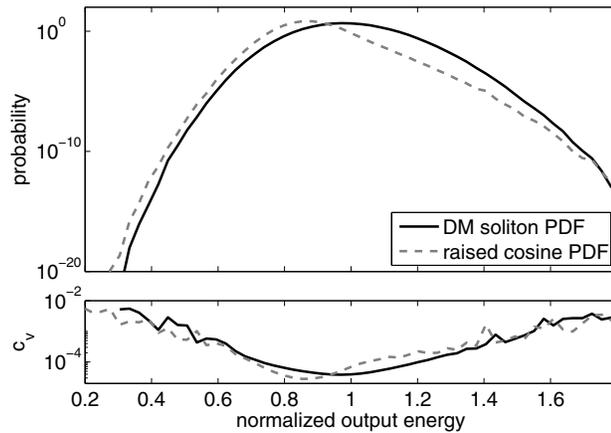


FIG. 9.6. Comparison of amplitude PDFs associated with the two pulse shapes: *Full simulated PDFs obtained from the SVD/CE/IS method applied to the dispersion-managed system with the DM soliton shape (solid) and the raised-cosine pulse shape (dashed). The coefficient of variation for each simulation is given in the lower panel.*

in output detector energy. The noiseless, unbiased propagation of the raised-cosine pulse shape is given for comparison in the right panel of Figure 9.5.

As before, we generate the biasing distributions and then use them to simulate the full output energy PDF for this new initial pulse shape. With the SVD/CE/IS method, adapting to the new, modified system is straightforward, something that is impossible with previous methods based upon the soliton pulses.

The comparison between the simulated PDF for the raised-cosine pulse shape and the previous DM soliton pulse shape is given in Figure 9.6. It is immediately apparent that changing the initial pulse shape has significantly altered the overall system performance. Recall that the input pulse energy has been matched. In this case, one might argue that the two PDFs agree reasonably well in the low-voltage portion, with similar slopes and only a normalization shift between the two curves. As shown, however, the probability of a given output energy differs between the two systems by as much as two orders of magnitude in some energy ranges. This result could not have been inferred a priori. The ability to easily capture the differences in system performance induced by such changes is one of the benefits of the SVD/CE/IS method, in that proposed modifications in system design can be tested quickly and with minimum effort.

9.3. Probing system dynamics. An additional benefit of the SVD/CE/IS method is the ability to probe the underlying system dynamics. This provides a way to improve system design by illuminating the rare-event mechanisms at work. To demonstrate this, we observe that in Figure 9.6 the raised-cosine and DM soliton systems have a particularly large difference in probability at a 30% increase in output energy, and we seek to investigate this difference.

To do this, we examine the optimal biasing distributions for reaching this output energy in each system, and graph three quantities for each system as a function of distance: the normalized energy (computed as if the detector were at this distance), the local biasing strength, and the total biasing strength. These are shown for both systems in Figure 9.7. In the top panel, the normalized output energy is given as

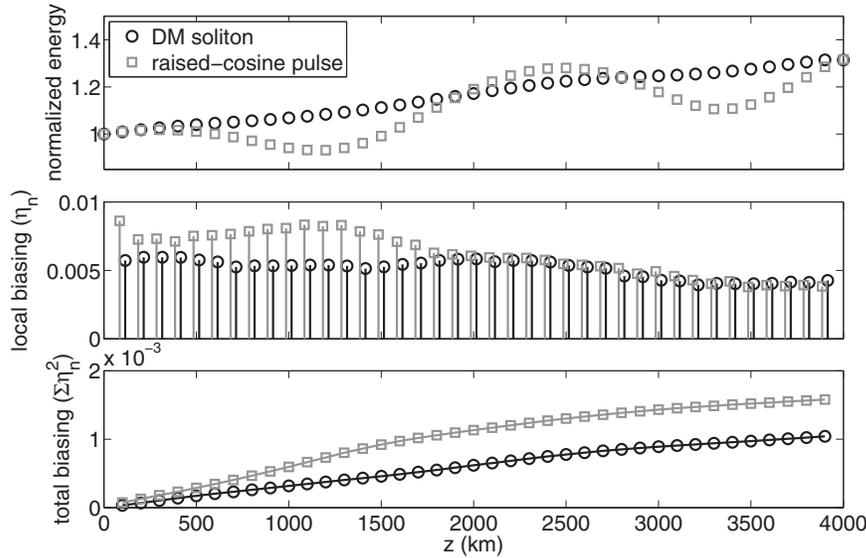


FIG. 9.7. Probing system dynamics: *Top panel:* normalized output energy as a function of propagation distance, observed stroboscopically: raised-cosine pulse (grey squares), DM soliton (black circles). *Middle panel:* local biasing coefficients (η_n) . *Lower panel:* the total biasing strength up to a given location.

a function of propagation distance, observed stroboscopically (once per dispersion map period). The oscillatory route taken by the raised-cosine pulse (grey squares) contrasts with that taken by the DM soliton (black circles), which in the unbiased case is stroboscopically stationary. The middle panel gives the local biasing coefficients (η_n) which optimally target this energy increase. The lower panel gives the total biasing strength up to a given location as the partial sum $\sum_{n=1}^k \eta_n^2$, where $z = nz_a$. At the end of the transmission line, the total amount of biasing required in the raised-cosine case is much greater, reflecting the decreased probability of reaching this state, relative to the DM soliton system.

We observe that the oscillations in the dynamics of the raised-cosine pulse significantly affect its optimal biasing. After being launched, this pulse appears to naturally broaden (and thus the amount of energy in the detection window decreases). The optimal biasing determines that, to increase the amplitude of the pulse significantly, this initial pulse broadening and accompanying amplitude decrease (at $z \sim 1000$ km or so) must be strongly counteracted. After this point, the amount of biasing returns to roughly the same as that for the DM soliton. The DM soliton pulse has no such oscillations, and the local biasing is approximately constant with propagation distance, in agreement with previous results [30]. In general, of course, the precise dynamics are complex, which is exactly why the SVD/CE/IS method is necessary: if the dynamics were immediately obvious, the optimal IS biasing distributions could be determined a priori. Because the SVD/CE/IS method is divided into biasing and simulation stages, however, the optimal biasing coefficients appear explicitly. These coefficients provide information about the manner in which large deviations arise. This ability to probe the underlying dynamics is an additional benefit of the SVD/CE/IS method.

10. Beyond lightwave systems. Up to this point we have formulated and demonstrated the SVD/CE/IS method in the context of lightwave communication

systems, the area of interest that originally inspired this work. The details of the method are not strongly restricted by the context of this particular application, however, and thus should be useful to a much wider range of problems. For example, it certainly should be possible to apply the method with little modification to the complex Ginzburg–Landau equation that governs ultrashort optical pulses in a mode-locked laser. We also expect that this can be applied to any nonlinear evolution operator that can be linearized, e.g., another type of partial differential equation or a system of ordinary differential equations.

With the current formulation, the main restriction is that the stochastic component be added discretely, rather than continuously, as the solution evolves. Of course, applications requiring continuous noise might be accommodated by a discrete noise approximation, and so even this restriction might be relaxed. Thus, it is possible that the method can be applied to a wide class of problems substantially beyond the scope of the lightwave systems considered here.

11. Discussion. In this paper we have formulated the SVD/CE/IS method, a new rare-event simulation method, and demonstrated results in the context of optical communication systems. We first demonstrated good agreement with rare-event simulation results obtained previously by an alternate method, and then showed the applicability of the method to a problem unable to be solved by such earlier methods.

This SVD/CE/IS method has evolved from the earlier, semianalytic methods used in previous studies [9, 11, 12, 14, 31]. The previous methods are highly effective for problems where they can be applied, i.e., in those cases where soliton pulse shapes occur. While solitons have great mathematical appeal, such pulses are not often employed in practice. The SVD/CE/IS method moves beyond soliton pulse shape restrictions, so that rare events can be studied in much broader classes of problems. We have demonstrated that the SVD/CE/IS method is extremely useful for this task.

In terms of computational effort, the method as presented here is still fairly intensive, though not intractably so. Each complete set of simulations took approximately three days to compute on an 8-core Xeon X5355 2.66 GHz workstation. Iterative calculation of the SVD, involving repeated solution of the linearized evolution equations, was the primary computational bottleneck.

Another rare-event simulation technique is the multicanonical Monte Carlo (MMC) method, which has been shown to be an effective method for the study of a number of lightwave systems (i.e., see [32, 33, 34, 35]). MMC is an iterative procedure which, like SVD/CE/IS, converges toward the optimal IS biasing distribution while requiring little a priori knowledge of the system. The principal distinction is that MMC makes no reduction of the dimensionality of the underlying problem and uses the Metropolis algorithm to do the biasing computationally via a directed random walk. Because of this, it is not straightforward to obtain the specific biasing information underlying the simulations, and consequently it is more difficult to obtain insight from the simulations about how rare events can arise. MMC also complicates error estimation due to strong bin-to-bin and sample-to-sample correlations, although a bootstrap technique has been developed to obtain accurate error estimates [32]. SVD/CE/IS simulations do not have the same issues, because of the two-stage approach. For purposes of error analysis of collected statistics, the standard techniques for IS can be used [28]. Recall that in stage 1, the optimal biasing distributions are found, and in stage 2, they are employed in ISMC simulations. As such, if the biasing distributions are in fact suboptimal, the standard methods for assessing the quality of ISMC simulations apply.

We also note that the SVD/CE/IS method could potentially be applied to problems that are quite different from those arising in optical communication systems—the focus on lightwave systems here was merely because these problems motivated its development. The formulation presented here is for a system with additive noise, although it is possible that it could be extended to other types of noise. The use of linearization and the SVD does introduce some limitations, in that each individual perturbation should be small enough for the linearization to be valid. For the systems considered, the perturbations at each noise-addition location are small, and large deviations result from the sum of many small perturbations, so that linearization is appropriate. In a nonlinear system with fewer larger perturbations, however, this may not be the case. In such a case it is possible that the CE method will simply fail to converge to an optimal biasing distribution satisfying the rare-event conditions, in which case the failure would be readily apparent. A more likely scenario, however, is that the biasing solution produced by the CE method would be suboptimal. In this case the problems typically associated with poorly chosen IS distributions would apply: the Monte Carlo simulations would converge slowly and would exhibit abnormally large values of the coefficient of variation in regions of the PDF where the biasing is suboptimal.

Another limitation arises from the computational cost of applying the SVD. The dimensionality reduction provided must be sufficient to justify the cost of the SVD. In the problems studied this reduction is large, e.g., two to three orders of magnitude, and thus the cost of the SVD is well justified. This aspect must be considered when gauging the applicability of the method to other problems. Although it is not a major bottleneck in the problems considered here, computing the adjoint may be a significant computational issue in other applications; checkpointing algorithms [36] may help alleviate this difficulty. In systems of the type studied here, however, the SVD/CE/IS method is clearly a highly effective rare-event simulation technique.

Appendix. Linearization of the detector. For an integrate-and-dump detector at $z = z_L$ given by

$$(A.1) \quad V = \int_{-T/2}^{T/2} |U_F(z_L, t)|^2 dt,$$

we can linearize by $U = u_0 + \epsilon \Delta u$ and obtain

$$(A.2) \quad \Delta V = 2 \operatorname{Re} \int_{-\infty}^{\infty} W(t) [(F * u_0)(t)]^* [(F * \Delta u)(t)] dt,$$

where $F(t) = \mathcal{F}^{-1}[F(\omega)]$ and $W(t)$ is a windowing function which has value 1 in $[-T/2, T/2]$ and 0 otherwise. Splitting into real and imaginary parts, i.e.,

$$(A.3) \quad F * u_0 = U_{F1} + iU_{F2},$$

$$(A.4) \quad F * \Delta u = u_{F1} + iu_{F2},$$

we have

$$(A.5) \quad \Delta V = 2 \int_{-\infty}^{\infty} W \begin{bmatrix} U_{F1} & U_{F2} \end{bmatrix} \begin{bmatrix} u_{F1} \\ u_{F2} \end{bmatrix} dt.$$

Because we wish to employ an iterative method for finding the SVD, however, we also require the adjoint operator M^* so that we can calculate the action M^*M . One way

to obtain the adjoint is to first construct an appropriate inner product. Since we are working with vectors with each component a function of time, we therefore double this equation and think of each element as a constant function of time on the interval $(-T/2, T/2)$:

$$(A.6) \quad \begin{bmatrix} \Delta V \\ \Delta V \end{bmatrix} = 2 \int_{-\infty}^{\infty} W \begin{bmatrix} U_{F1} & U_{F2} \\ U_{F1} & U_{F2} \end{bmatrix} \begin{bmatrix} u_{F1} \\ u_{F2} \end{bmatrix} dt = \mathcal{K}[u].$$

Then, using the inner product

$$(A.7) \quad \langle \vec{v}, \vec{u} \rangle = \int_{-T/2}^{T/2} \vec{v}^T(t') \vec{u}(t') dt',$$

we have

$$(A.8) \quad \langle \vec{v}, \mathcal{K}[\vec{u}] \rangle = \int_{-T/2}^{T/2} \vec{v}^T(t') \left\{ 2 \int_{-\infty}^{\infty} W(t) \begin{bmatrix} U_{F1}(t) & U_{F2}(t) \\ U_{F1}(t) & U_{F2}(t) \end{bmatrix} \begin{bmatrix} u_{F1}(t) \\ u_{F2}(t) \end{bmatrix} dt \right\} dt'.$$

Defining

$$(A.9) \quad X(t) = 2W(t) \begin{bmatrix} U_{F1}(t) & U_{F2}(t) \\ U_{F1}(t) & U_{F2}(t) \end{bmatrix} = 2W(t) \begin{bmatrix} 1 \\ 1 \end{bmatrix} [U_{F1}(t) \quad U_{F2}(t)]$$

and

$$(A.10) \quad \mathbb{F}(t) = \begin{bmatrix} F_1(t) & -F_2(t) \\ F_2(t) & F_1(t) \end{bmatrix},$$

where $F = F_1 + iF_2$, the matrix X represents the filtered, windowed, unperturbed (nonlinear) solution, and \mathbb{F} is the (complex-valued) filter in (real) matrix form. Then we can calculate the adjoint by

$$\begin{aligned} \langle \vec{v}, \mathcal{K}[\vec{u}] \rangle &= \int_{-T/2}^{T/2} \int_{-\infty}^{\infty} \vec{v}^T(t') X(t) \left(\mathbb{F} * \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \right)(t) dt dt' \\ &= \int_{-T/2}^{T/2} \int_{-\infty}^{\infty} \vec{v}^T(t') X(t) \int_{-\infty}^{\infty} \mathbb{F}(t-\tau) \begin{bmatrix} u_1(\tau) \\ u_2(\tau) \end{bmatrix} d\tau dt dt' \\ &= \int_{-\infty}^{\infty} \left\{ \int_{-T/2}^{T/2} \int_{-\infty}^{\infty} \mathbb{F}^T(t-\tau) X(t) dt \vec{v}(t') dt' \right\}^T \begin{bmatrix} u_1(\tau) \\ u_2(\tau) \end{bmatrix} d\tau \\ (A.11) \quad &= \int_{-\infty}^{\infty} \left\{ \int_{-T/2}^{T/2} [\mathbb{F}^T(-t) * X^T(t)] \vec{v}(t') dt' \right\}^T \begin{bmatrix} u_1(\tau) \\ u_2(\tau) \end{bmatrix} d\tau. \end{aligned}$$

Thus we obtain

$$\begin{aligned} \mathcal{K}^\dagger[\vec{v}] &= \int_{-T/2}^{T/2} [\mathbb{F}^T(-t) * X^T(t)] \vec{v}(t') dt' \\ (A.12) \quad &= \mathbb{F}^T(-t) * \left(2W \begin{bmatrix} U_{F1}(t) \\ U_{F1}(t) \end{bmatrix} \right) \int_{-T/2}^{T/2} [1 \ 1] \vec{v}(t') dt'. \end{aligned}$$

This can be simplified further by noting that the filter used in the simulations is real and even; specifically, it is

$$(A.13) \quad F(\omega) = \exp\left(-\frac{2 \ln(2)\omega^2}{(2\pi 5/T_L)^2}\right),$$

where T_L is the length of the computational domain, here expressed in picoseconds [37]. In this case, \mathbb{F} is even and symmetric, and then

$$(A.14) \quad \mathcal{K}^\dagger[\vec{v}] = 2\mathbb{F} * \left(W \begin{bmatrix} U_{F1}(t) \\ U_{F2}(t) \end{bmatrix} \right) \int_{-T/2}^{T/2} [v_1(t') + v_2(t')] dt'.$$

The adjoint of the linearized detection operator thus has a particularly simple form: it is a simple projection of its “input” (i.e., the “output” of the direct linearized problem) that multiplies the windowed and filtered version of the vector $[U_{F1}(t) \ U_{F2}(t)]^T$ (which is the full nonlinear filtered solution at the end of the transmission line). Other detectors can be handled in very much the same way, although in general the nature of the adjoint as a projection which then multiplies the nonlinear solution does not hold. Of course, the full action M^*M is given by $\mathcal{L}^\dagger \mathcal{K}^\dagger \mathcal{K} \mathcal{L}$, where the linear propagators \mathcal{L} and \mathcal{L}^\dagger remain unchanged. Note in particular that this means that the full adjoint operator involves the propagation of the function obtained in (A.14) back to the specific amplifier location using the adjoint linearized equation.

Acknowledgments. The authors would like to thank Tobin Driscoll for useful discussions regarding implementing the SVD for the linearized NLS, and Tito Homem-de-Mello for his insights into the CE method. The authors would also like to thank Jinglai Li and Gino Biondini for a number of helpful comments and suggestions; in particular, they would like to thank Dr. Li for help with the computation of the adjoint detector operator, (A.14).

REFERENCES

- [1] M. J. BAYARRI, J. O. BERGER, E. S. CALDER, K. DALBEY, S. LUNAGOMEZ, A. K. PATRA, E. B. PITMAN, E. T. SPILLER, AND R. L. WOLPERT, *Using statistical and computer models to quantify volcanic hazards*, *Technometrics*, 51 (2009), pp. 402–413.
- [2] L. N. TREFETHEN AND D. BAU, III, *Numerical Linear Algebra*, SIAM, Philadelphia, 1997.
- [3] G. STRANG, *Linear Algebra and Its Applications*, 3rd ed., Harcourt Brace Jonanovich, San Diego, CA, 1988.
- [4] R. Y. RUBINSTEIN AND D. P. KROESE, *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation, and Machine Learning*, Springer-Verlag, New York, 2004.
- [5] G. S. FISHMAN, *Monte Carlo: Concepts, Algorithms and Applications*, Springer-Verlag, New York, 1996.
- [6] A. OWEN AND Y. ZHOU, *Safe and effective importance sampling*, *J. Amer. Statist. Assoc.*, 95 (2000), pp. 135–143.
- [7] R. RUBINSTEIN, *The cross-entropy method for combinatorial and continuous optimization*, *Methodol. Comput. Appl. Probab.*, 1 (1999), pp. 127–190.
- [8] L. N. TREFETHEN, A. E. TREFETHEN, S. C. REDDY, AND T. A. DRISCOLL, *Hydrodynamic stability without eigenvalues*, *Science*, 261 (1993), pp. 578–584.
- [9] R. O. MOORE, G. BIONDINI, AND W. L. KATH, *Importance sampling for noise-induced amplitude and timing jitter in soliton transmission systems*, *Opt. Lett.*, 28 (2003), pp. 105–107.
- [10] R. O. MOORE, T. SCHÄFER, AND C. K. R. T. JONES, *Soliton broadening under random dispersion fluctuations: Importance sampling based on low-dimensional reductions*, *Optics Comm.*, 256 (2005), pp. 439–450.
- [11] R. O. MOORE, G. BIONDINI, AND W. L. KATH, *A method to compute statistics of large, noise-induced perturbations of nonlinear Schrödinger solitons*, *SIAM J. Appl. Math.*, 67 (2007), pp. 1418–1439.

- [12] J. LI, E. SPILLER, AND G. BIONDINI, *Noise-induced perturbations of dispersion-managed solitons*, Phys. Rev. A, 75 (2007), paper 053818.
- [13] E. T. SPILLER, W. L. KATH, R. O. MOORE, AND C. J. MCKINSTRIE, *Computing large signal distortions and bit-error ratios in DPSK transmission systems*, IEEE Photonics Tech. Lett., 17 (2005), pp. 1022–1024.
- [14] E. T. SPILLER AND W. L. KATH, *A method for determining most probable errors in nonlinear lightwave systems*, SIAM J. Appl. Dyn. Syst., 7 (2008), pp. 868–894.
- [15] E. NOETHER, *Invariant variation problems*, Transport Theory Statist. Phys., 1 (1971), pp. 186–207.
- [16] G. P. AGRAWAL, *Nonlinear Fiber Optics*, Academic Press, New York, 1995.
- [17] E. IANNONE, F. MATERA, A. MECOZZI, AND M. SETTEMBRE, *Nonlinear Optical Communication Networks*, Wiley, New York, 1998.
- [18] D. MARCUSE, *Derivation of analytical expressions for the bit-error probability in lightwave systems with optical amplifiers*, J. Lightwave Tech., 8 (1990), pp. 1816–1823.
- [19] J. A. BUCKLEW, *Introduction to Rare Event Simulation*, Springer-Verlag, New York, 2004.
- [20] P. T. DE BOER, D. P. KROESE, S. MANNOR, AND R. Y. RUBINSTEIN, *A tutorial on the cross-entropy method*, Ann. Oper. Res., 134 (2005), pp. 19–67.
- [21] T. HOMEM-DE-MELLO, *A study on the cross-entropy method for rare event probability estimation*, INFORMS J. Comput., 19 (2004), pp. 381–394.
- [22] S. KULLBACK AND R. A. LEIBLER, *On information and sufficiency*, Ann. Math. Stat., 22 (1951), pp. 79–86.
- [23] E. VEACH AND L. J. GUIBAS, *Optimally combining sampling techniques for Monte Carlo rendering*, in Proceedings of SIGGRAPH '95, Los Angeles, CA, 1995, pp. 419–428.
- [24] L. N. TREFETHEN, *Pseudospectra of linear operators*, SIAM Rev., 39 (1997), pp. 383–406.
- [25] R. B. LEHOUCQ, D. C. SORENSEN, AND C. YANG, *ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*, Software Environ. Tools 6, SIAM, Philadelphia, 1998.
- [26] I. STAKGOLD, *Boundary Value Problems of Mathematical Physics*, Classics Appl. Math. 29, Vol. 2, SIAM, Philadelphia, 2000.
- [27] T. R. TAHA AND M. J. ABLOWITZ, *Analytical and numerical aspects of certain nonlinear evolution equations. II. Numerical, nonlinear Schrödinger equation*, J. Comput. Phys., 55 (1984), pp. 203–208.
- [28] G. BIONDINI, W. L. KATH, AND C. R. MENYUK, *Importance sampling for polarization-mode dispersion: Techniques and applications*, J. Lightwave Tech., 22 (2004), pp. 1201–1215.
- [29] V. S. GRIGORYAN, M. SHIN, P. DEVGAN, J. LASRI, AND P. KUMAR, *SOA-based regenerative amplification of phase-noise-degraded DPSK signals: Dynamic analysis and demonstration*, J. Lightwave Tech., 24 (2006), pp. 135–142.
- [30] J. LI AND E. SPILLER, *private communication*, MIT and Marquette University, 2008.
- [31] E. T. SPILLER AND W. L. KATH, *A method for determining most probable errors in nonlinear lightwave systems*, SIAM J. Appl. Dyn. Syst., 7 (2008), pp. 868–894.
- [32] A. O. LIMA, I. T. LIMA, AND C. R. MENYUK, *Error estimation in multicanonical Monte Carlo simulations with applications to polarization-mode-dispersion emulators*, J. Lightwave Tech., 23 (2005), pp. 3781–3789.
- [33] I. NASIEVA, A. KALIAZIN, AND S. K. TURITSYN, *Multicanonical Monte Carlo modelling of BER penalty in transmission systems with optical regeneration*, Optics Comm., 262 (2006), pp. 246–249.
- [34] A. BILENCA, A. DESJARDINS, B. BOUMA, AND G. TEARNEY, *Multicanonical Monte-Carlo simulations of light propagation in biological media*, Optics Express, 13 (2005), pp. 9822–9833.
- [35] R. HOLZLÖHNER AND C. R. MENYUK, *Use of multicanonical Monte Carlo simulations to obtain accurate bit error rates in optical communications systems*, Optics Lett., 28 (2003), pp. 1894–1896.
- [36] A. GRIEWANK AND A. WALTHER, *Algorithm 799: Revolve: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation*, ACM Trans. Math. Software, 26 (2000), pp. 19–45.
- [37] C. K. MADSEN AND J. H. ZHAO, *Optical Filter Design and Analysis: A Signal Processing Approach*, Wiley, New York, 1999.