Transition Matrix Methods

David Yevick (1) and Michael Reimer (2)

1) Department of Physics, University of Waterloo, Waterloo, ON N2L 3G1, Canada, yevick@uwaterloo.ca 2) Department of Physics, University of Waterloo, Waterloo, ON N2L 3G1, Canada, mareimer@sympatico.ca

Abstract: *The transition matrix describes the probabilities of transitions of a system between different sets of states. In this presentation, we demonstrate that techniques similar to the multicanonical method can be employed to find the elements of the transition matrix between pairs of highly unlikely states in an optical communication system. This result is then applied to determine the probability distribution of outage times resulting from polarization mode delay (PMD).*

Introduction

While biased sampling procedures, most notably the multicanonical or importance sampling techniques, are now frequently employed to investigate static system quantities such as the probability distribution functions (pdf) of the differential group delay (DGD) or of the eye opening penalty.[1-4], these methods have not been applied to dynamic system behavior such as the average time spent in an outage. Here we adapt the transition matrix methods of statistical mechanics[5-7] to dynamic communications problems (it should be noted that a transition matrix approach was earlier applied to analyze the error of static PMD calculations in [8]).

Multicanonical Method

Since our transition matrix discussion generalizes our previous work on biased sampling methods, we first review the foundations of the multicanonical method. A randomly fluctuating physical system can generally be described by a vector of N_e system observables, **E**, each of which is affected by N_a stochastically varying parameters **a**. For example, in this paper **E** are suitably defined moments of a pulse that has passed through a series of polarization controllers to which fluctuating voltages specified by **a**, are applied. Typically, the average value of a function $p(\mathbf{E})$ in regions of sample space with an extremely small probability of occurrence must be determined without recourse to an unrealistically large number of samples, $\mathbf{a}_i$.

Accordingly, the multicanonical method biases the sample space. in such a manner that, if $p(\mathbf{E})$ represents the pdf of **E**, the sampling probability after many iterations coincides with the inverse of the pdf and the sampling becomes uniform in the system observables **E**. In the standard implementation, the physically accessible region of **E** is partitioned into N histogram bins centered at $\mathbf{E}_m$ with $m = 1, 2, \ldots N$. Next, the elements of a histogram that will store the unnormalized pdf, $p^0(\mathbf{E}_m)$, and a second histogram that holds temporary values, $H(\mathbf{E}_m)$, are both initialized to unity. The starting value of **E** is computed from a randomly chosen set of system variables $\mathbf{a}^{cur}$. These system variables are then randomly perturbed according to $\mathbf{a}^{new} = \mathbf{a}^{cur} + \delta\mathbf{a}$, leading to new observables $\mathbf{E}^{new}$. To insure that the computation spends more time in states with small values of $p(\mathbf{E})$, this transition is accepted with probability $\min(1, p^0(\mathbf{E}^{cur})/p^0(\mathbf{E}^{new}))$, in which case $\mathbf{a}^{cur}$ is set to $\mathbf{a}^{new}$; otherwise $\mathbf{a}^{cur}$ is unchanged for the following iteration. In either case, the histogram element, $H(\mathbf{E}_m)$, for the new value of $\mathbf{E}^{cur}$ is incremented by unity. After this process is repeated M times, an improved estimate, $p^1(\mathbf{E})$, of $p(\mathbf{E})$ is obtained after the bias in the system variables during the iteration is removed according to $p^1(\mathbf{E}) = cp^0(\mathbf{E})$ $H(\mathbf{E})$. In the following iteration $p^1(\mathbf{E})$ replaces $p^0(\mathbf{E})$, $H(\mathbf{E})$ is reset to unity and the above process is repeated.

Transition Matrices

The multicanonical method can be employed to determine the relative probability for transitions between each pair of system variables even for highly unlikely pairs of states. In particular, for every accepted or rejected transition from a state in the n:th to a state in the m:th histogram bin obtained during a multicanonical calculation, the corresponding element, $\mathcal{T}_{nm}$, of an *unbiased* but *unnormalized transition matrix*, $\mathcal{T}$ is incremented by unity. At any point in the calculation, the rows of $\mathcal{T}_{nm}$ can be normalized to unity, yielding an estimate of the standard *transition matrix T*. We can also construct a *biased multicanonical transition matrix*, *B*, such that B_{nm} gives the likelihood of a transition from the n:th to the m:th state in the multicanonical procedure.. Here T_{nm}, typically at the end of the last iteration, is multiplied by the multicanonical acceptance probability, $\min(1, p^L(\mathbf{E}_n)/p^L(\mathbf{E}_m))$ after which each row of this matrix is normalized to unity. A further matrix *R*, with $R_{nm} = T_{nm}/B_{nm}$, can be constructed that specifies the ratio of the probability of an unbiased transition from state n to state m to the probability of this transition transition probability in the multicanonical method.

An alternative method for biasing transitions is to formulate the acceptance rule directly from an estimate of $\mathcal{T}$ that is dynamically updated as the calculation step.[6] Here, starting from initial values $\mathcal{T}_{nm}$=1, transitions between two states $\mathbf{E}_m^{new}$ and $\mathbf{E}_n^{cur}$ generated by $\mathbf{a}^{new} = \mathbf{a}^{cur} + \delta\mathbf{a}$, are accepted with probability $\min(1, \mathcal{T}_{nm}/\mathcal{T}_{nn})$ while again incrementing

$\mathcal{T}_{nm}$ or $\mathcal{T}_{nn}$ by unity for accepted and rejected transitions, respectively. We have slightly modified this technique by fixing our estimate of the transition matrix after every M transitions and employing this estimate for the next M transitions; since this ensures that for most transitions, the detailed balance condition, $P(\mathbf{E}_n)\ T_{nm} = P(\mathbf{E}_m)\ T_{mn}$ for each n and m, where $P(\mathbf{E}_n)$ denotes the probability of finding the system in the n:th histogram bin, is satisfied. However, we still impose detailed balance at the end of the calculation to avoid unphysical complex eigenvalues of the transition matrix. Since $P(\mathbf{E})$ is identified below with the eigenvector of the transition matrix with unit eigenvalue., we compute the $P^{(0)}(\mathbf{E}_n)$ associated with the transition matrix $T^{(0)}=T$ and iteratively improve T according to $T_{nm}^{(i+1)} = (P^{(i)}(\mathbf{E}_m)T_{mn}^{(i)} + P^{(i)}(\mathbf{E}_n)T_{nm}^{(i)}) / 2P^{(i)}(\mathbf{E}_n)$.

Time-Dependent Problems

Time-dependent systems can be simulated in a similar manner to time-independent systems if the magnitude of the random perturbation, $\delta\mathbf{a}$, is adjusted such that the average change in the observables, $|\mathbf{E}^{new} - \mathbf{E}^{cur}|$, coincides with a measured change over some time interval, Δt.

To sample transitions between physically unlikely states, paths can be composed of randomly generated transitions between states that observe the transition probabilities stored in the biased transition matrix or, equivalently, the multicanonical acceptance rule can be applied in conjunction with the unbiased transition matrix probabilities. In either case, repeated evaluation of the system observables from the system parameters is avoided. Each path from the state m to the state n over the intermediate states 1, 2, ... k must however be assigned the weight $w = R_{m1}R_{12}...R_{kn}$ in the resulting histogram.. Of course, the unbiased transition matrix can be employed directly if only the high probability region of the sample space is of interest.

The time evolution of a system property can also be derived from the eigenvalues and eigenvectors of T. The eigenvector with unit eigenvalue here yields the probability distribution function for finding the system in a given state while the distribution functions associated with the other eigenvectors instead decay with time constants proportional to their eigenvalues as the system tends towards its equilibrium distribution..

Outage Times

To illustrate the above considerations, we compute the pdf of the PMD-induced outage durations in a single mode optical fiber system. Our calculation employs a fiber model consisting of 100 concatenated birefringent fiber segments each of which poseses a DGD of 2.17ps. These are separated by polarization rotators that can be adjusted to generate any desired rotation of the incoming polarization vector $\mathbf{\Omega}_i$ on the Poincaré sphere. For certain mutual orientations of these polarization vectors, the ratio of the emulator DGD to the average DGD of the fiber segment will exceed a certain value associated with a system outage.. The time interval over which the DGD exceeds this value is termed the outage time.

Identifying the vector of system variables $\mathbf{a}$ with the $2l$-dimensional relative angles between the birefringence vectors of successive sections, and the system observables, $\mathbf{E}$, with the DGD, the outage time distribution is normally determined by performing many time steps $\mathbf{a}^{new}=\mathbf{a}^{cur} + \delta\mathbf{a}$, and recording the length of time for which the system remains in the outage region for each outage. However, in our calculations, the Markov chain can be directly modeled by sampling transitions according to the probabilities stored in the unbiased or biased transition matrices (in the latter case, however, the histogram for the time duration of each system path must be incremented by an appropriate weighting factor).

Alternatively, we can take as the initial state of the system a distribution equal to $P(\mathbf{E}_n)$ for n in the in non-outage region and zero elsewhere, corresponding to the average distribution of states that have not yet entered the outage region. The time evolution of this state is then modeled by repeatedly multiplying by the transition matrix. After the k:th step, the sum of the histogram values that fall outside the outage region can be identified with the probability that the system returns to a non-outage state within the time interval $(k\text{-}1)\Delta t<t<k\Delta t$ The values of the system vector in the non-outage region are then set to zero before the $(k+1)$:st time step. The time evolution can also be determined by noting that since the state vector of paths that are restricted to the outage region is repeatedly multiplied by the submatrix formed from the matrix entries for the outage states, the eigenvalues of this submatrix yield the decay rates corresponding to the distributions given by the associated eigenvectors. The desired outage time distributions can therefore be found through projection of the distribution after one time step onto these states.

Numerical Results

To illustrate the above methods, in Fig. 1 we show the pdf of the outage times, for both DGD>$2\tau_{avg}$ and DGD>$3\tau_{avg}$ outage conditions. The circles in this figure refer to an unbiased Markov chain calculation that recorded the outage times observed during 10^9 time steps (fewer than 10,000 samples were recorded in the outage region for the $3\tau_{avg}$. outage condition). The solid line displays the result after constructing an unbiased transition matrix from the accepted and rejected transitions of three $5\text{x}10^6$ sample multicanonical iterations and then generating a biased Markov chain from the the multicanonical acceptance rule together with appropriate weight factors for the paths. Replacing the multicanonical acceptance rule by a rule derived from the ratio of transition matrix elements was found to decrease the sampling of the tail of the pdf for a fixed number of realizations. Finally, successive multiplications of the initial state vector by the unbiased transition matrix obtained after two $2x10^5$ sample multicanonical iterations yielded the dashed line; as was also computed from the eigenvalues and eigenfunctions of the outage state submatrix. Although the biased transition matrix clearly samples a far wider range of states than the standard procedure, some accuracy is necessarily sacrificed since the standard method preserves all information about the system variable correlations in different regions of the system state space.

Finally, Fig. 2 displays the eigenvectors corresponding to the three largest eigenvalues of the transition matrix determined with three $5x10^6$ multicanonical iterations. The insert contains the magnitude of the 50 largest eigenvalues. Note that the eigenvector with unit eigenvalue (solid line) yields the pdf of the DGD for infinitely long fiber samples while the other eigenvectors can be used to compute the DGD at finite fiber lengths.

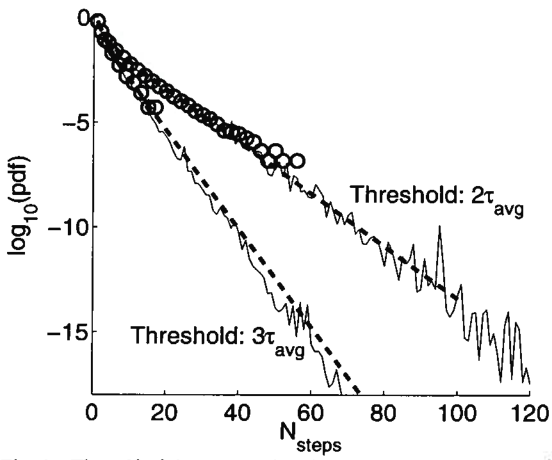

Fig. 1. The pdf of the outage durations associated with an optical fiber emulator for thresholds of $2\tau_{avg}$ and $3\tau_{avg}$. The results of the standard method with 10^9 samples, the multicanonical reweighing method with 3 iterations of $5x10^6$ samples and repeated multiplication of the initial state vector by an unbiased transition matrix obtained after two $2x10^5$ sample multicanonical iterations. are denoted by circles, solid and dashed lines, respectively.

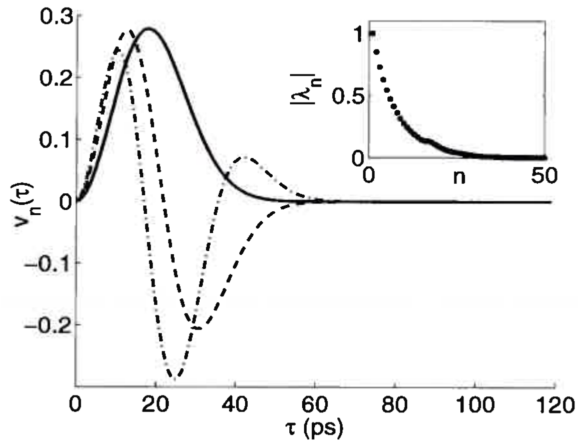

Fig. 2. The first three (solid, dashed and dashed-dotted lines, respectively) transition matrix eigenvector computed after 3 multicanonical iterations of $5x10^6$ samples. The inset contains the magnitude of the first 50 eigenvalues.

Conclusions

That time-dependent effects in communication and other physical systems can be rapidly and accurately modeled, even in very low probability regions by transition matrix methods can significantly impact future communication system simulations. We therefore hope to shortly extend our results to several challenging system calculations.

Acknowledgments

This work was supported in part by the National Sciences and Engineering Research Council of Canada (NSERC), the Canadian Institute for Photonic Innovation (CIPI), Nortel and the Ontario Research and Development Challenge Fund (ORDCF).

References

1. B. Berg and T. Neuhaus:,Phys. Lett. B, V. 267, p. 24, 1991
2. D. Yevick: IEEE Photon. Technol. Lett., vol. 15, pp. 224-226, 2003
3. T. Lu, D. Yevick, B. Hamilton, D. Dumas and M. Reimer, IEEE Photon. Technol. Lett., vol. 17, p. 2583, 2005
4. T. Lu, W. Huang, D. Yevick, M. O'Sullivan and M. Reimer,, J. Opt. Soc. Am. A vol. 22, p. 2804, 2005.
5. G. Smith and A. Bruce, J. Phys. A: Math. Gen. vol. 28, p. 6623 (1995)
6. J-S Wang and R. Swendsen, J. Stat. Phys. vol. 106, p. 245 (2002)
7. M. Fitzgerald, R. Picard and R. Silver, J. Stat. Phys. vol. 98, p. 321 (2000)
8. A. Lima, I. Lima, and C. Menyuk, J. Light. Technol. vol. 23, p. 3781 (2005)