Reliable Extreme-Scale Stochastic Dynamics Simulation based on Generalized Interval Probability – I. Uncertainty Dynamics

Yan Wang
Georgia Institute of Technology, Atlanta, GA30332, yan.wang@me.gatech.edu

1. Introduction

As processor speeds recently reached the frequency limit of 3 gigahertz with the given materials and nanoscale fabrication capabilities, many-core parallel computing architecture has become the only resort to further increase the computational capacities. Some new issues emerge when we design modeling and simulation (M&S) methods and algorithms for extreme-scale scientific computing. Given that a large number of processors with heterogeneous architectures (e.g. CPUs, GPUs, single-precision, double-precision, etc.) are involved in computation, both the resilience of task management (for abnormal disruptions or complete failures of some processors) and the reliability of numerical computation (for non-uniform treatment of round-offs in floating-point computation) are important.

Additionally, uncertainty quantification is particularly important in M&S of multiscale complex systems, such as biological cells, atmospheric turbulence, and materials-product hierarchies, because the traditional modeling by reductionism always provides simplified views without the complete knowledge. In general, two types of uncertainties are recognized. Aleatory uncertainty is inherent randomness in systems, whereas epistemic uncertainty is due to lack of perfect knowledge. Epistemic uncertainty has different sources, such as lack of data, conflicting information from multiple sources, conflicting beliefs among experts' opinions, lack of time for introspection, measurement errors, lack of dependency information, etc. Given the very different natures of the two types of uncertainties, it is important to differentiate and treat them separately in M&S. Neglecting epistemic uncertainty may lead to decisions that are not robust. Mixing epistemic and aleatory uncertainties may increase costs of risk management.

Beyond steady-state or static analysis of uncertainty, the simulation of system dynamics is of particular importance in predicting chemical reactions, liquid and gas transport phenomena, high-energy particle diffusions and irradiations, and others. *There are no stochastic dynamics models that distinguish aleatory and epistemic uncertainties*. Existing stochastic differential equations rely on time-consuming sensitivity analysis (e.g. second-order Monte Carlo, design of experiments) to evaluate the parameter and model uncertainties. Existing imprecise probability representations (e.g. Dempster-Shafer evidence theory, probability bound analysis, possibility theory, and others) are not suitable for computationally intensive dynamics simulation and tend to be overly pessimistic in information fusion.

2. Generalized interval and generalized interval probability

Here, we present a new simulation formalism that distinguishes aleatory and epistemic uncertainties based on *generalized interval probability* [1]. The new form of imprecise probability is based on *generalized interval*, where the probabilistic calculus structure is significantly simplified because of the Kaucher arithmetic. A generalized interval $\mathbf{x} := [\underline{x}, \overline{x}]$ $\underline{x}, \overline{x} \in \mathbb{R}$ is not constrained by $\underline{x} \le \overline{x}$ any more. Therefore, [.2,.1] is also a valid interval and called *improper*, while [.1,.2] is called *proper*.

Generalized interval has better algebraic and semantic properties. Traditional set-based intervals form a *semi-group* because of the lack of invertibility, e.g. [.1,.2]−[.1,.2]=[−.1,.1]≠0. In contrast, generalized intervals form a *group*. [.1,.2]−[.2,.1]=0. The introduction of improper intervals is analogous to the introduction of negative numbers in real analysis, as illustrated in Fig.1. We do not need negative numbers in our

- **classical set-based interval** is defined as
$[\![\underline{x}, \overline{x}]\!] := \ x \in \mathbb{R} \mid \underline{x} \le x \le \overline{x}$
- *Semi-group*: no invertibility
$[\![.1,.2]\!] - [\![.1,.2]\!] = [\![-.1,.1]\!] \ne 0$

- **Generalized interval** is defined as
$\mathbf{x} := [\underline{x}, \overline{x}] \quad \underline{x}, \overline{x} \in \mathbb{R}$
- *Group*
[.1,.2]-dual[.1,.2]
=[.1,.2]-[.2,.1]=[0,0]=0
- The widths of generalized intervals could reduce during calculation
[1,3]+[2,1]=[3,4]
[1,3]+dual[1,3]=4

- Can you manage to calculate with *only zero and positive numbers* to solve all scientific/engineering/financial/… problems?

➢Similar to negative numbers, *improper intervals* in generalized interval system are "negative" intervals that can make our lives much easier!

Fig.1: Computational advantage of generalized interval

daily calculations. However they do bring much better algebraic properties to our number systems with ease of calculation. As a result of introducing improper intervals, the calculus in generalized interval probability is very similar to the one with the traditional precise probability. This resemblance with ease of calculation enables large-scale simulations based on interval probability.

3. Stochastic dynamics simulation under both uncertainties

The time evolution of both aleatory and epistemic uncertainties can be simulated based on generalized interval probability. As two special cases of the differential Chapman-Kolmogorov equation with generalized interval probability, *generalized Fokker-Planck equation* (gFPE) [2] plays a vital role to describe various drift-diffusion processes, including system vibration and dynamics, two-phase flow, high-energy particles irradiation, etc., whereas *interval master equation* (IME) [3] models chemical reactions and long time scale state transitions. The dynamics of both aleatory and epistemic uncertainties can be concisely and efficiently captured.

3.1 Imprecise Markov chain

Interval arithmetic provides rigorous bounds for round-off error and approximation-related model error. In addition, interval probability provides another advantage of relaxing the standard assumption of Markovian property under Gaussian distribution while maintaining computability, since the lower and upper probabilities characterize a collection of different probabilistic estimations as a result of the deviation from the standard assumption. That is, the interval probability provides the effect of non-Markovian properties.

We developed an imprecise Markov chain model based on generalized interval probability. The new model can be intuitively kept track of with its resemblance to the classical Markov chain. It offers much better computational efficiency than other forms of imprecise Markov model.

3.2 Solving gFPE

A path integral algorithm [2] has been developed to solve the generalized Fokker-Planck equation so that the lower and upper probability densities are computed simultaneously. As shown in Fig.2, the time evolution of interval probability can be efficiently computed, where the lower and upper bounds enclose the real-valued distribution. Again, other forms of imprecise probability just cannot provide such computational efficiency.

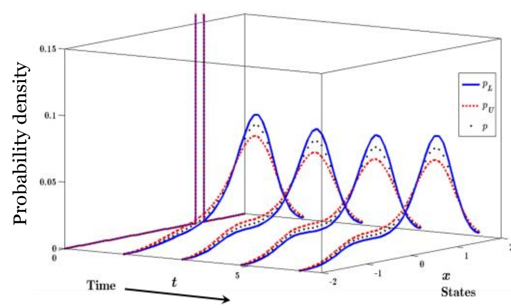

Fig.2: Time evolution of interval probability

3.3 Solving IME

A Krylov subspace projection algorithm [3] has been developed to solve the interval master equation. The approach efficiently saves memory space without the need to store the infinitesimal generator. Additionally, the logic interpretation properties of generalized interval [4,5,6] can be applied to verify the completeness (no under-estimation) and soundness (no over-estimation) of interval bounds. Most importantly, there is no need for additional simulation runs for sensitivity analysis.

The simulation approach by solving the above two equations provides a *holistic* picture of interval probability distribution at any particular time and simulates the evolution of the overall probability densities, which is helpful for us to check the effect of rare events with very small probabilities to occur. Because the probability densities corresponding to the rare events are close to zero, the quantification of numerical errors associated with the floating-point arithmetic operation becomes critical for the accuracy of predictions. Interval arithmetic can provide rigorous error bounds for such small values. By selectively setting the modes of rounding on computers for certain states of interest, the computation of generalized interval probability can bring the benefits of reliable simulation and rigorously verifiable prediction.

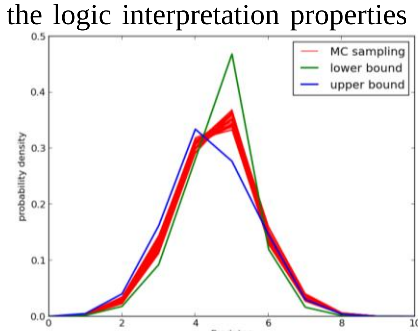

Fig.3: Interval probability vs. Monte Carlo sensitivity analysis

References

[1] Wang Y. (2010) Imprecise probabilities based on generalised intervals for system reliability assessment. *International Journal of Reliability & Safety*, **4**(4): 319-342

[2] Wang Y. (2013) Generalized Fokker-Planck equation with generalized interval probability. *Mechanical Systems & Signal Processing*, **37**(1-2):92-104

[3] Wang Y. (2013) Stochastic dynamics simulation with generalized interval probability. *International Journal of Computer Mathematics* (submitted)

[4] Wang Y. (2008) Semantic tolerance modeling with generalized intervals. *ASME Journal of Mechanical Design*, **130**(8): 081701(1-7)

[5] Wang Y. (2008) Interpretable interval constraint solvers in semantic tolerance analysis. *Computer-Aided Design & Applications*, **5**(5): 654-666

[6] Wang Y. (2008) Closed-loop analysis in semantic tolerance modeling. *ASME Journal of Mechanical Design*, **130**(6): 061701(1-10)