Stochastic Considerations for Dynamic Systems

Moriz Scharpenberg*

Mária Lukáčová-Medviďová†

Technische Universität Hamburg-Harburg, Institute of Numerical Simulation

Schwarzenbergstraße 95, 21073 Hamburg, Germany

Studies and industrial computations frequently need a measure of uncertainty of computed results based on the uncertainty of the underlying data. Often simple gradients are sufficient for an engineer to judge results, but sometimes a more accurate prediction is necessary. A tool for general use is the Monte Carlo method, which gives a prediction for the distribution of results calculated with stochastic data but lacks hard error bounds for many nonlinear problems and may need a lot of computing capacity for higher dimensional systems.

A comparison will be made with a finite volume method and a modified TRAIL approach that uses appropriate Ansatz functions, that are propagated over time. These give alternatives, that may not only be easier to compute in a limited area or for a large number of states but as well provide error bounds. The result is a probability density function at all time steps, from which all interesting quantities, e.g. probabilities for exceeding critical values, can directly be derived.

Even though the alternatives may have some limitations at the current state of development, they may be already interesting for specific purposes. In this paper theoretical background will be given and an example will be presented.

I. Introduction: Extension of Deterministic Models

Many deterministic engineering calculations are based on input data which is still subject to alterations. Even though the process of calculations in early design phases of a product may be close to that of a final one, a lot of data beeing derived from more time consuming methods still cannot be provided and some design descisions may not be finally agreed. If possible, these risks are taken into account by additional scaling factors on the obtained results.

I.A. Stochastic Data

Some data is usually described naturally by a stochastic formulation. Regions of chaotic behaviour, e.g. turbulence in a fluid, are often described in terms of a power spectral density. The term 'naturally' refers to the huge amount of data, which can only be conceived sensibly in a stochastic manner. However, material properties or manufacturing tolerances would normally fit into this category as well, but in many applications their variance is small enough, such that they can be considered deterministic.

But not all of the parameters that are considered deterministic are so from the beginning. Even though for a certain design their value may be considered fixed, since the design changes quickly in early stages, their precise value is not entirely clear. Hence to judge the calculation it would be interesting to see the effect of that uncertainty.

An example may be a specific coefficient which may be hard to compute or measure precisely at a given point of time where an assessment of an engineering design solution is desired. However, a sensible value can be estimated as before with some variance added, depending in size on the engineering knowledge, e.g.

*ms@mds-line.de

†lukacova@tu-harburg.de

experience from previous projects at similar stages or even from an uncertainty analysis for the process of origin.
The salient point is to obtain a sensible probability distribution. In some cases it may make sense just to fit a Gaussian curve through the two side points given by an interval from engineering judgement. But as soon as a property is part of personal interest, e.g. if a value is not to exceed a critical value, the probability density will typically show some additional skewness.

I.B. Uncertainty Analysis

For highly crucial computations there are highly optimized interval methods, e.g. robust analysis, that give hard limits, in which the result is guaranteed to stay within a given range. In some situations, however, a strict boundary is not necessary or even misleading, since very rare parameter combinations may lead to extreme results. This is of importance especially in a design phase, where uncertainties are still large, but the allowed risks are large too. For an engineer only interested in the significance of calculated data, a tendency or feeling for the involved risk is often sufficient.
Depending on the accuracy of the required statements there are appropriate methods available. Just to quickly obtain tendencies a qualitative analysis may just compute simple gradients. These gradients correspond to uncorrelated variances in linearized systems already and can simply be obtained by finite differences or by an automatic differentiation[5] scheme. This approach can even be refined. Using higher order derivatives even lightly noninear effects can be taken into account, but only as long as the system stays sufficiently smooth.
To get a deeper understanding of nonlinear effects, the behaviour of discrete states and the effect of correlation, in a quantitative approach the Monte Carlo analysis (MCA) can be performed. Depending on the number of states and the accuracy needed, computational costs may be high. For nonlinear systems it may not be possible to guarantee hard error bounds for the resulting stochastic distributions.[6]
The advantages of MCA are overwhelming in practice. A calculation framework may be set up entirely independent of the equations in use and a calculation can be parallelized easily. But still, if sufficient structure in the problem can be exploited, other methods, e.g. finite volume (FV) or the Modified TRAIL scheme (MTS), may become more interesting. The existence of hard error bounds allows vindication of safety factors or callibration of MCA. The exploited structure may bring down needed computing power. Such a proven quantitative analysis will be the main topic in the following.

II. A Stochastic Approach

Basically any dynamic system can be described as a system of ordinary differential equations (ODE) of the form

$$\frac{\partial}{\partial t}\tilde{\mathbf{x}} = \tilde{\mathbf{f}}(\tilde{\mathbf{x}}, \mathbf{p}, t) \tag{1}$$

with some states $\tilde{\mathbf{x}}$ that evolve over time t and depend on a vector of parameters $\mathbf{p}$ and a possibly nonlinear vector function $\tilde{\mathbf{f}}$. This equation can futher be simplified by assuming

$$\mathbf{x} = \begin{pmatrix} \tilde{\mathbf{x}} \\ \mathbf{p} \end{pmatrix} \quad \text{and} \quad \mathbf{f} = \begin{pmatrix} \tilde{\mathbf{f}} \\ \mathbf{0} \end{pmatrix} \tag{2}$$

such that equation (1) becomes

$$\frac{\partial}{\partial t}\mathbf{x} = \mathbf{f}(\mathbf{x}, t). \tag{3}$$

II.A. Introduction of Stochastic Data

The most general possible assumption is to define a probability density ρ for the initial states at $t = t_0$ of equation (3), such that

$$\rho(\mathbf{x}, t_0) = \rho_0(\mathbf{x}). \tag{4}$$

If only some parameters in equation (1) are uncertain, for some methods the other parameters and initial states still need to be assumed stochastic as well. Using a very low variance in practice usually does not seem to harm the overall results, but may sometimes disturb numerical schemes.

In some cases a trim calculation needs to be solved prior to time integration, e.g. by reordering equation (3) and solving

$$0 = \frac{\partial}{\partial t}\mathbf{x} - \mathbf{f}(\mathbf{x}, t). \tag{5}$$

for unknown states or state derivatives. It was shown[1] this could be achieved by Taylor expansion around the mean value of the initial distribution, which also allowes the error to be controlled.

II.B. Monte Carlo Analysis

Equation (1) is already enough to start a MCA. Whichever variable is assumed to be stochastic, integration delivers a probability function from the density. Using a waterlevel distribution then for each stochastic variable its inverse probability function returns correctly distributed initial values. For each value a calculation is started and from the results of the combinations of states the density can be calculated numerically.
Many variants for special numerical behaviour or purpose do exist as well as a lot of sophisticated commercial software. The statistical error control seems to work sufficiently for a huge number of applications.

II.C. Fokker Planck Approach

To calculate the effect of (3) on the density distribution (4), a small volume Ω is considered. Stating according to the continuity equation from fluid dynamics that the overall probability density is constant, a change within a piece of volume must be due to in- and outflow of probability normal $\mathbf{n}$ to the surface S

$$\frac{\partial}{\partial t}\int_{\mathbf{x}\in\Omega} \rho(\mathbf{x}, t) dV = \int_{\mathbf{x}\in\partial\Omega} \rho(\mathbf{x}, t)\ \mathbf{f}(\mathbf{x}, t) \cdot \mathbf{n}\, dS \tag{6}$$

By applying the divergence theorem to the right hand side, equation (6) can be transformed into

$$\frac{\partial}{\partial t}\int_{\mathbf{x}\in\Omega} \rho(\mathbf{x}, t) dV = \int_{\mathbf{x}\in\Omega} \mathrm{div}(\rho(\mathbf{x}, t)\ \mathbf{f}(\mathbf{x}, t))\, dV \tag{7}$$

and by stating this has to be true for all volumes, it follows

$$\frac{\partial}{\partial t}\rho(\mathbf{x}, t) = \mathrm{div}(\rho(\mathbf{x}, t)\mathbf{f}(\mathbf{x}, t)) \tag{8}$$

as a general formulation.
This is a form of the Liouville equation for a single system of ODEs or the Fokker[2]-Planck[4] equation with the diffusion term set to zero. It describes exactly how the probability density evolves over time.
Basically a stochastic ordinary differential equation (SODE) is transformed into a deterministic partial differential equation (DPDE), for which computational methods can be constructed. However, it has to be noted that the DPDE has as many dimensions as there are states in the ODE.

II.D. Finite Volume Scheme

Equation (6) gives already a starting point for a solution using a finite volume scheme. The domain is gridded and each gridpoint consists of a piece of volume Ω_i surrounded by borders to pieces of volumes designated to other gridpoints. Figure (1) shows a 2D example, with a rectangular piece of volume at a grid point. The outflow at a gridpoint i is calculated by $\mathbf{f}(b_k)$ and using the area $a(b_k)$ of each linear border b_k separating the volumes Ω_i and Ω_j and a normal vector $\mathbf{n}(b_k)$ the flow is computed by

$$\frac{\partial}{\partial t}\int_{\Omega_i} \rho dV = \sum_j \rho_{\mathrm{upwind}}(b_k) a(b_k) \mathbf{f}(b_k) \mathbf{n}(b_k) \tag{9}$$

where

$$\rho_{\mathrm{upwind}}(b_k) = \begin{cases} \rho_i & \text{if } \mathbf{n}(b_k)\mathbf{f}(b_k) > 0 \\ \rho_j & \text{if } \mathbf{n}(b_k)\mathbf{f}(b_k) \leq 0 \end{cases} \tag{10}$$

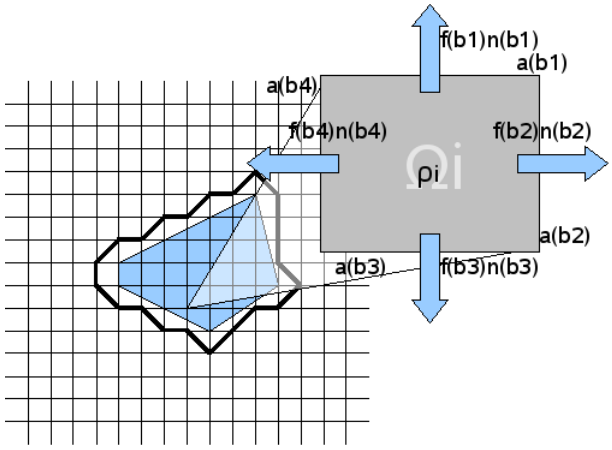

Figure 1. Finite Volume Method: Flow out of and into each grid point is computed. Computing only gridpoints with the possibility of a significant flow allows higher dimensionality for sparse problems.

indicates using the value for the density of the piece of volume, where the flow at the boundary is coming from.
The length of the timestep is calculated from the Courant-Friedrichs-Lewy (CFL) condition

$$\max_i \frac{\Delta t ||\mathbf{f}(\Omega_i)||}{\Delta x} \leq \sigma \tag{11}$$

where for a rectangular grid Δx is the smallest side length, Δt the largest allowed timestep, and σ a factor of safety smaller than one, typically $\sigma \approx 0.8$.
For performance reasons it is essential that only gridpoints in the vicinity of areas are actually computed, where the probability density is sufficiently large, e.g. for a rectangular grid with horizontally and vertically numbered volumes Ω_{ij} larger than some tolerance

$$\mathrm{TOL} > \max|\rho_{n,m}|_{n=[i-1,i+1],m=[j-1,j+1]}. \tag{12}$$

In figure (1) the bold black line indicates the calculation region, while the underlying shape indicates an area with sufficiently large density in the last time step. From step to step only recalculated and adjacent gridpoints have to be checked for modification, since the CFL condition guarantees no further changes.

II.E. Modified TRAIL Scheme

Starting with equation (8) the TRAIL (Trapezoidal Rule for Adaptive Integration of Liouville Dynamics) algorithm is proposed[3] for describing the evolution of probability densities. It adresses the problem of the large number of partial derivatives by using an appropriate set of Ansatz functions, which is possible as long as the probability function stays sufficiently smooth to allow a cost saving algorithm.
The initial density is described by a sum of Gaussian Ansatz functions

$$\rho(\mathbf{x}, t) = \sum_i y_i(t) e^{(\mathbf{x}-\mathbf{x}_i)^T G_i (\mathbf{x}-\mathbf{x}_i)} \tag{13}$$

with the free variables y_i, $\mathbf{x}_i$ and λ_i such that $G_i = \lambda_i I$. Starting with a small number of Ansatz functions optimization routines, e.g. the Gauss-Newton method or the Nelder-Mead scheme, return appropriate values for the free variables and the number of Ansatz functions is increased until a prescribed accuracy is matched. Then an integration is performed using the implicit trapezoidal rule

$$\rho(t+\tau) - \frac{\tau}{2}\dot{\rho}(t+\tau) = \rho(t) - \frac{\tau}{2}\dot{\rho}(t) \tag{14}$$

which provides a local spatial error for each time step by approximating

$$\|\rho(t+\tau) - \frac{\tau}{2}\dot{\rho}(t+\tau) - \rho(t) + \frac{\tau}{2}\dot{\rho}(t)\|_2 \leq \mathrm{TOL}_x \tag{15}$$

that can be controlled. The length of the timestep itself can be estimated by comparing with a step of another order. Values for the free variables of the following time step are obtained by some optimization routine again. If the required tolerance in equation (15) cannot be met, the Ansatz space has to be extended by additional Gaussian functions. If the optimization process breaks down due to a singular Jacobian some Ansatz functions have to be removed.
The whole trouble of the algorithm is the 2-norm in equation (15), which cannot be evaluated exactly. A sensible idea is to introduce a number of sample points $\mathbf{x}_s$ to approximate it. This has the advantage that after linerization the minimization problem becomes an overdetermined system, which can be solved by a QR decomposition from which the global error and the condition of the influence matrix obtained from linearizing the system can directly be read from. If the global error becomes to large, more Gaussians will be added, if the condition becomes too bad, Gaussians will be removed.
The drawback of such an approximation of the 2-norm are frequently strange effect in the optimization, since local maxima generated by overlapping Gaussians may stay unnoticed and gradients happen to show discontinuities during the process. Hence a framework needs to be constructed, that filters inappropriate results and cancels steps if necessary. Using better initial values obtained from linear analysis or good guessing algorithms for the variance controlling parameters of the Gaussians increases the efficiency of the algorithm.
For a lower number of Ansatz functions a maximum optimization seems to work as well. Basically instead of the 2-norm the maximum norm is calculated as an optimization problem

$$\max_{\mathbf{x}}(\rho(\mathbf{x}, t+\tau) - \frac{\tau}{2}\dot{\rho}(\mathbf{x}, t+\tau) - \rho(\mathbf{x}, t) + \frac{\tau}{2}\dot{\rho}(\mathbf{x}, t)) \leq \mathrm{TOL}_x \tag{16}$$

using some reliable global branch and bound method. This approach works very well until the number of variables becomes too large, the global optimization scheme fails more frequently without notice and, hence, the outer minimization optimization fails as well due to wrong gradients.

III. Two State Example

In the following the three approaches are applied to a simple example. A two state model

$$\dot{v} = -g\sin(\theta) - R\frac{v^2}{m} \tag{17}$$

$$\dot{\theta} = A\frac{v}{m} - \frac{g}{v}\cos(\theta) + dAv^2 \tag{18}$$

with gravity constant g, aerodynamic constants A and d, and mass m describes a phugoid motion in terms of the normalized pitch attitude θ and the normalized forward velocity v of a gliding aircraft. A friction constant R is chosen such that a fixed point solution exists and oscillating motion occurs. The theoretical solution of this equation returns multiple fixed points, one of which, oscillation about the center, is practically relevant.
Now assume v and θ to have uncertain initial values and engineering experience gives a guess of their lower and upper value, e.g. approximately $\theta \in [-0.65, 0.60]$ and $v \in [1.5, 2.3]$, from which a worst case scenario can be derived.
The engineering experience, translated in stochastic terms may result in the assumption, that more than 95% of all occurences of the initial values are to be within the given bounds, e.g. larger or smaller parameter values are possible but very rare. Hence the variances σ are $\sigma_v \approx 0.18$ and $\sigma_\theta \approx 0.32$, respectively.

III.A. Monte Carlo Analysis

For MCA $n = 10000$ initial points are chosen, distributed according to the initial probability density and propagated independently. The resulting cloud of points is gridded and the density is visualized. Figure (2) shows the probability density at $t = 0.6$. Depending on the number of points, the density converges quickly to what was expected, but even for a high number of sample points it stays erratic, which visualizes the slow convergence of the statistical error

$$\text{statistical error} \propto \frac{1}{\sqrt{n}} \tag{19}$$

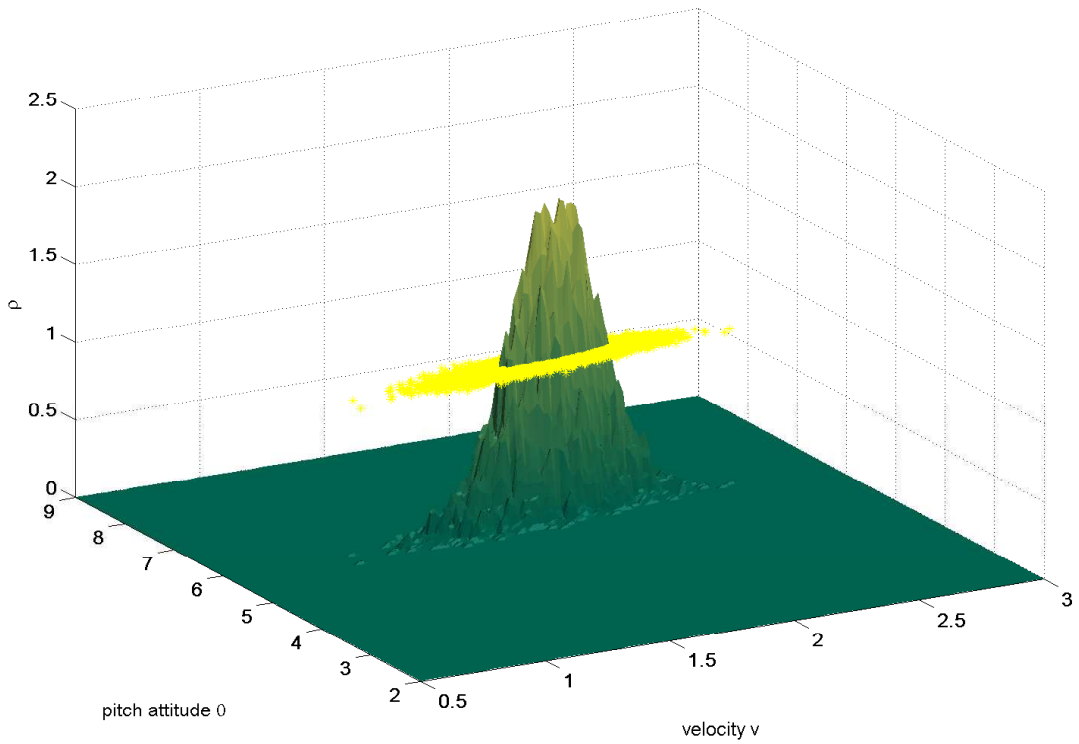

Figure 2. Propagated probability density as 3D view for MCA. The rocky surface give an idea of the error in the estimation of the probability density. However, the density can be seen slightly tilted, which shows the correlation between v and θ. The yellow points show the location of the calculated results.

and indicates large relative errors especially in regions with low probability density. Apart from that the method is implemented entirely independent of the original system, can be parallelized easily and gives quickly a good impression of what is going on.

III.B. Finite Volume Method

The domain is gridded for the FV approach into 50 times 70 small rectangles. To each of them and for each time step equation (9) is solved. From inspection of equations (17) and (18) a maximum gradient for the CFL condition is obtained. Even though the grid is quite coarse the overall result is very good. In 2D the accelleration of the computation due to ommitting unchanged gridpoints is very small due to the overhead costs of the analysis of the gridpoints. However, for a large number of dimensions, this would change dramatically.

III.C. Modified TRAIL Scheme

In a first step the initial density has to be decomposed into a sum of Gaussian Ansatz functions. Similar to equation (15) the sum needs to be close to the target distribution apart from a predefined tolerance

$$\mathrm{TOL}_x \leq \| \sum_i y_i e^{(\mathbf{x}-\mathbf{x}_i)^T G_i (\mathbf{x}-\mathbf{x}_i)} - \rho_0(\mathbf{x}) \|_2 \tag{20}$$

which is approximated the same way as equation (15) is. To stay as stable and efficient as possible, the number of Gaussians is $i = 1$ in a first step and, if the tolerance cannot be met, i is increased in each step. Since only Ansatz function with a single variance in all directions are allowed, in this example eleven functions are needed to obtain the predefined tolerance of $\mathrm{TOL}_x = 10^{-2}$.
Then, using an initial timestep, a future density is calculated and the timestep is checked to not to have

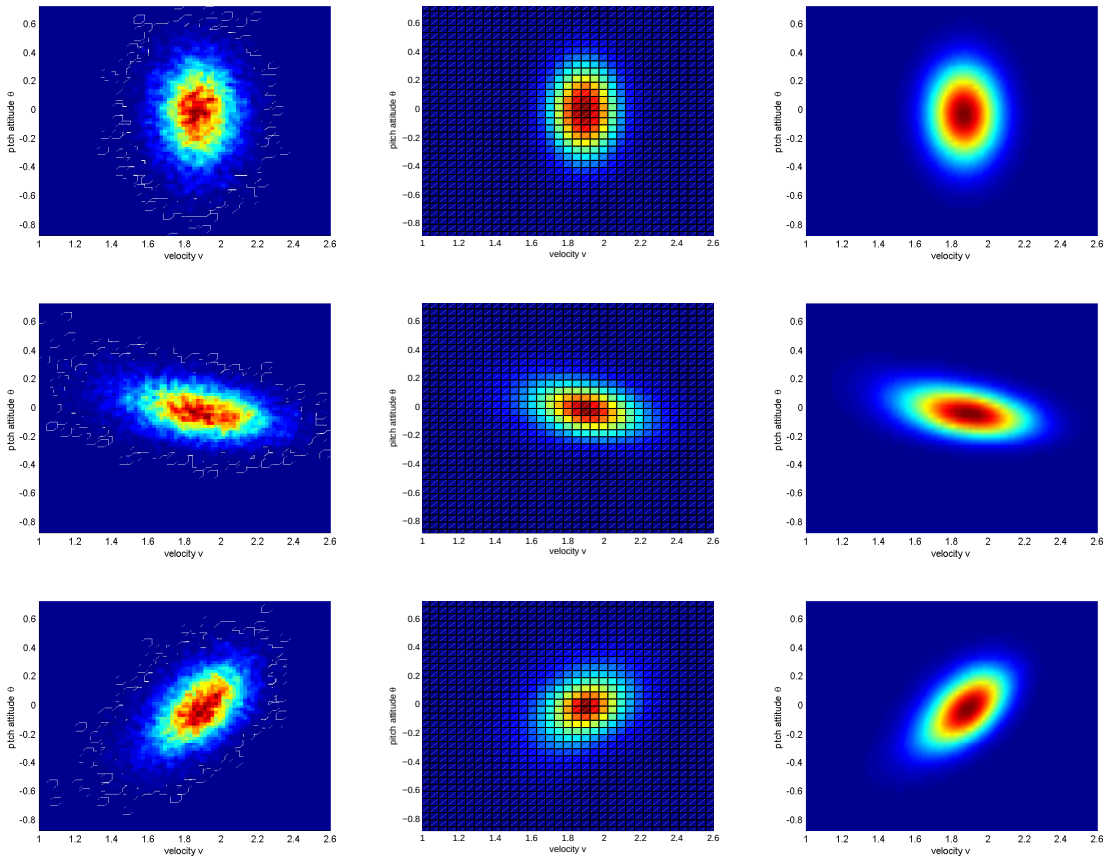

Figure 3. Time series of the probability density for all three methods applied to the example problem: The left column shows the Monte Carlo method, the center column shows the finite volume method and the right column shows the Fokker Planck approach, each at $t = [0.00, 0.15, 0.30]$ All three methods basically lead to the same result, but the different philosophies can be seen: Monte Carlo with its erratic bahaviour, the coarse grid of the finite volume method and the smooth shape of the Fokker Planck approach.

been too large. If the tolerance cannot be met in an intermediate step, the number of Gaussians has to be increased, by adding an additional function to the old density with an amplitude of zero, which can be modified by the optimizer.
Depending on the approximation of the 2-norm this approach may be very fast and gives easily a low error. The most appropriate approximation for it turned out to be a Monte Carlo like estimation of sample points. But if local maxima due to an overlay of Gaussians are not detected by that scheme, a sudden breakdown of the optimizer is possible. Then typically more sample points need to be added at the right places and the optimizations needs to be restarted. With this happening more frequently a bad choice of samplepoints or initial guesses can enlarge the computational time significantly.

III.D. Overall Result

Using the corner points of the given intervals from above as extreme values, at $t = 0.6$ the worst case scenario gives a velocity of $v(0.6) = 2.6$. Now having performed stochastic analysis in a variety of ways, the whole distribution at any point of time is given. Figure 2, for example, visualizes the probability density of the two states at $t = 0.6$ obtained by MCA. θ seems likely to stay in a very limited range compared to the initial value of the variance at that point of time and hence gains certainty which is balanced by greater uncertainty in the velocity. Figure (3) suggests, that during the motion indeed uncertainty is shifted back and forth between the two states. As a result of nonlinearities in the underlying equation, the uncertainty band is also tilted slightly, hence showing a correlation between v and θ.
Integrating the density or using geometrical information from the Gaussian Ansatz functions hard numbers can be obtained. For example using the error controlled MTS reveals that the probability for being above

a velocity threshold of 2.3 is only 2.6% and for being above 2.5 is only 0.15%, maybe tolerable risks at an early stage of design.

IV. The Way Ahead

Typically making use of decreasing computational costs leads to calculations using more realistic and more complex dynamic systems, such that for analysis the effect of more input data becomes ever more complex. At the same time development cycles become shorter and in early design stages some data may not be available, while nevertheless first approximations are requested already. Risk analysis proves to become hence more and more important. Stochastic analysis gives very detailed results on this.
In this paper three different methods were described. The following table gives an overview about the features of each method.

Method	MCA	FV	MTS
Computational Cost	low-high	high	potentially low
Parallelisation	very good	good	possible in parts
Error Control	statistical error control only	yes	yes
Probability Conservation	by construction	yes	within error bounds
Discontinuities	yes	yes	difficult
Code Modifications	not necessary	possibly necessary	possibly necessary

MCA is very common and available in many commercial packages. FV methods are available as well but less common for risk analysis. Especially for low dimensional problems it is robust and allows error control. MTS, making use of the smoothness of many problems, is new and hence not yet available as directly applicable package. But it is promising since it avoids the curse of dimensionality,[3] from which both MCA and FV suffer, and provides hard error bounds as well.
Further improvements of the algorithm will be developed to make it more stable and reliable also for larger equation systems. An appropriate methodology for treating the 2-norm is of importance here. For very few discontinuities extensions may be possible by integrating connected but seperate systems. This will be a topic of future work.

References

[1]B. Christianson , M. Cox, "Automatic Propagation of Uncertainties",Automatic Differentiation: Applications, Theory, and Implementations, Springer, 47-58 (2005)

[2]A. D. Fokker, "Die mittlere Energie rotierender elektrischer Dipole im Strahlungsfeld", Annalen der Physik 43, 810-820 (1914)

[3]I. Horenko and M. Weiser, "Adaptive Integration of Molecular Dynamics", J. Comput. Chem. 24, 1921-1929 (2003)

[4]M. Planck, "Ueber einen Satz der statistichen Dynamik und eine Erweiterung in der Quantumtheorie", Sitzungberichte der Preussischen Akadademie der Wissenschaften, 324-341 (1917)

[5]M. Scharpenberg and M. Lukáčová-Medviďová, "Use of Automatic Differentiation for Sensitivity Analysis of Flight Loads", AST Proceedings, 407-414, Hamburg (2007)

[6]A.Y. Weiße, I. Horenko and W. Huisinga, "Adaptive Approach for Modelling Variability In Pharmacokinetics", CompLife, LNBI 4216, 194-204, Springer-Verlag Berlin Heidelberg (2006)