Efficient Numerical Solution of the Chemical Master Equation in Molecular Biology

Paul Sjöberg[1], Per Lötstedt[1], and Johan Elf[2] **

[1] Division of Scientific Computing, Department of Information Technology, Uppsala University
P.O. Box 337, SE-75105 Uppsala, Sweden.
[2] Department of Cell and Molecular Biology, Uppsala University
P.O. Box 596, SE-75124 Uppsala, Sweden.

Abstract. The chemical master equation (CME) describes the probability for each internal state of the cell or rather the states of a model of the cell. The number of states grows exponentially with the number of chemical species in the model, since each species corresponds to one dimension in the state space.

Two different approaches to handling this difficulty are compared: numerical solution of the *Fokker-Planck approximation* of the CME, by a finite volume method and the *Stochastic Simulation Algorithm* (SSA).

Both methods have their advantages. The convergence rate of SSA is independent of the number of dimensions and the algorithm is rather simple to implement. The Fokker-Planck approach can be more efficient for low-dimensional problems and high accuracy.

1 Stochastic models in molecular biology

Even though, to science, the biological cell is just an extraordinary complex and exquisite system of chemical reactions, the tools usually apt for analyzing chemistry is not always sharp enough to properly describe biochemical dynamics. Some of the assumptions that usually can be made for chemical systems are no longer valid. Biological cells are small and many chemical species are present in very low copy numbers. Due to the inherent discreteness of molecules and the randomness of the encounters between molecules that lead to reactions, the relative variation in copy numbers may become large and it is necessary to account for the stochasticity of the internal state of the cell.

Even though this aspect of *in vivo* biochemistry isn't new [2], the importance of the fluctuations in copy numbers in the cell has been stressed during the last years due to improved experimental techniques as well as an increased effort to model single cell processes [3].

** Present address: Department of Chemistry and Biology, Harvard University, Cambridge, MA 02138, USA.

2 The chemical master equation

The chemical master equation (CME) is used to describe the variation in chemical species [1]. It is assumed that the content of the cell is well mixed, that is, there is a sufficiently long time between reactive collision to ensure that each pair of molecules is equally likely to be the next to collide. The prerequisite for the master equation is that the chemical system can be described as a Markov process, which essentially means that the progress of the system depends only on the present state (chemical composition) of the cell.

Let the molecular species in the model be denoted X_i, $i = 1 \dots N$, where N is the number of components and let x_i be the copy number of molecules of species X_i. The state of the cell is determined by a vector of the copy numbers of each species $\mathbf{x} = (x_1,\ x_2, \dots, x_N)^T$. A reaction is specified

$$\mathbf{x}_r \xrightarrow{w_r(\mathbf{x}_r)} \mathbf{x},$$

where $\mathbf{x}_r$ is the state of the system before the reaction and $w_r(\mathbf{x})$ is the *reaction propensity*, that is the probability for reaction per time unit. Summing over all reactions we now can write the CME:

$$\frac{dp(\mathbf{x},t)}{dt} = \sum_{r=1}^{R} w_r(\mathbf{x}_r)p(\mathbf{x}_r,t) - \sum_{r=1}^{R} w_r(\mathbf{x})p(\mathbf{x},t). \tag{1}$$

The CME state that the change in probability for state $\mathbf{x}$ is simply the probability to reach state $\mathbf{x}$ from any other state $\mathbf{x}_r$ (first sum), minus the probability to leave state $\mathbf{x}$ for any other state (second sum).

The size of the state space grows exponentially with the number of components in the model. There are no analytical solutions for the master equation but for some simple cases. Even low-dimensional problems will be very cumbersome to solve numerically. Two approaches to handling the huge state space are compared here.

The most common strategy is to apply a Monte Carlo method, in this context the *Stochastic Simulation Algorithm* (SSA) [4] is a natural choice. SSA essentially simulates the chemical evolution for an ensemble of model instances by randomly applying the reactions of the system. Data is collected by recording the state of each instance. The simulated data is then used to compute the probability distribution. The algorithm is simple to implement, has low memory demands and a convergence rate that is independent of the dimension of the problem. On the other hand, that convergence rate is slow.

Another approach is to find a way to approximate the CME so that the solution can be represented on a smaller space than the original state space. One alternative is to approximate CME with a partial differential equation, the *Fokker-Planck equation* (FPE) [1]

$$\frac{\partial p(\mathbf{x},t)}{\partial t} = \sum_{r=1}^{R} \left(\sum_{i=1}^{N} n_{ri} \frac{\partial (w_r(\mathbf{x})p(\mathbf{x},t))}{\partial x_i} + \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{n_{ri}n_{rj}}{2} \frac{\partial^2 (w_r(\mathbf{x})p(\mathbf{x},t))}{\partial x_i \partial x_j} \right).$$

The FPE is discretized and solved on a grid that is considerably coarser than the state space [6].

3 Computational efficiency

The convergence rate of solution by the FPE approximation is derived in [7] and SSA have the convergence rate usually associated with Monte Carlo methods [5]. For a certain error ϵ the computational work for the SSA is

$$W_{SSA}(\epsilon) = C_{SSA}\,\epsilon^{-2},$$

while the work for the FPE is

$$W_{FPE}(\epsilon) = C_{FPE}\,\epsilon^{-\left(\frac{N}{r}+\frac{1}{s}\right)},$$

where C_{SSA} and C_{FPE} are independent of ϵ, N is still the dimension of the problem and r and s are the order of accuracy of the space- and time-discretization respectively [7].

Since the work for the FPE approach grows so rapidly with N a parallelization will not achieve so much more than the possibility to treat problems with one additional dimension. For SSA the impact of parallelization is not quite as clear since it depends on how C_{SSA} depends on N. That dependence is in fact very problem dependent, since it essentially is determined by the size of the subspace that is sampled by SSA.

The results for some test problems that all have properties that are troublesome for SSA are shown in Figure 1. The FPE is discretized by a second order finite volume scheme in space. These are steady state solutions which reduce the FPE computational work to $W_{FPE}(\epsilon) = C_{FPE}\,\epsilon^{-\frac{N}{r}}$ [7]. Figure 1 illustrates the SSA insensitiveness of the dimension and how the initially attractive convergence rate of the FPE gets worse with increasing dimension. For low dimension and high accuracy FPE is much more efficient.

Acknowledgments

This work was funded by the Swedish Research Council, the Swedish National Graduate School in Scientific Computing and the Swedish Foundation for Strategic Research.

References

1. van Kampen, N. G.: Stochastic Processes in Physics and Chemistry. Elsevier, Amsterdam, 2nd ed. (1992)
2. McQuarrie, D. A.: Stochastic Approach to Chemical Kinetics. J. Appl. Prob. **4**, (1967) 413–478
3. Raser, J. M., O'Shea, E. K.: Noise in Gene Expression: Origins, Consequences and Control. Science **309** (2005) 2010–2013

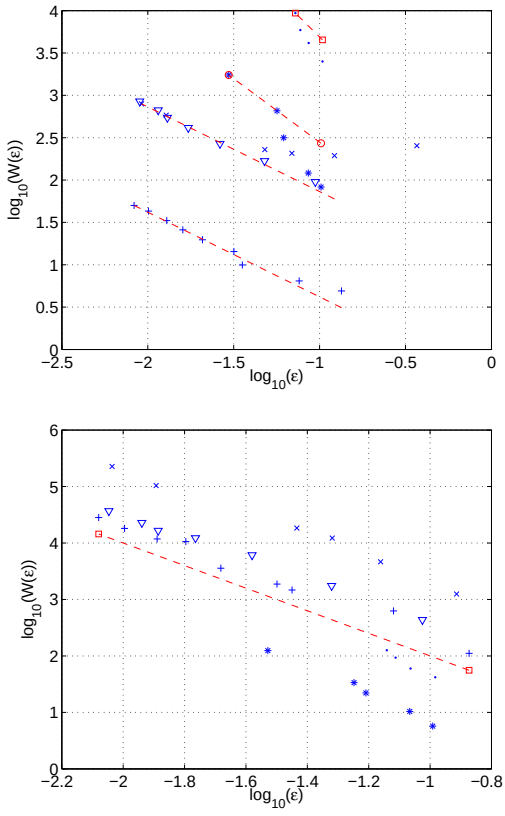

Fig. 1. The computational work using FPE (top) and SSA (bottom) to solve the CME for some different test problems in 2D ($+,\times,\nabla$), 3D ($*$) and 4D ($\cdot$). The dashed lines are reference lines with slopes -1 (no symbol), $-3/2$ ($\circ$) and -2 ($\Box$).

4. Gillespie, D. T.: A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. J. Comput. Phys. **22** (1976) 403–434
5. Heath, M. T.: Scientific Computing: An Introductory Survey. McGraw-Hill, Singapore, (1997)
6. Elf, J., Lötstedt, P., Sjöberg, P.: Problems of high dimension in molecular biology. Proceedings of the 19^{th} GAMM-Seminar, Leipzig, 2003
7. Sjöberg, P., Lötstedt, P., Elf, J.: Fokker-Planck approximation of the master equation in molecular biology. Technical Report 2005-044, Department of Information Technology, Uppsala University, 2005, *submitted for publication*