Jan-Hendrik Prinz
Computational Molecular Biology
Arnimallee 6
FU Berlin
14195 Berlin, Germany
phone: +49 (0)30 838 75775
email: jan-hendrik.prinz@fu-berlin.de

October 13, 2010

Dear Editor,

We have revised our previously submitted draft

Efficient Computation, Sensitivity and Error Analysis of Committor Probabilities for Complex Dynamical Processes
by Jan-Hendrik Prinz, Martin Held, Jeremy C. Smith and Frank Noé.

regarding all aspects of the reviewer. All changed paragraphs have been marked *red* for better readability. In the following we give a detailed explanation of the changes made in this revised version:

- p. 2, A technical point. I think it is possible to define $q(x)$ to be non-zero for $x \in A$ if we think about transitions from $x \in A$ that could find their way to B without returning to A first. I think the definition is just a matter of convenience if the transitions are defined consistently.

  Indeed, the committor theory can also be defined consistently with committor probabilities $q(x) > 0, x \in A$. To be precisely, only states on the boundary of $x \in \partial A$ can have a non-zero committor, henceforth this ambiguity arises from using discrete states, where now every state $x \in A$ can be also on the boundary of A. To be consistent with the picture that all states $x \in A$ are "completely" in A and the idea of a reaction coordinate from $A \to B$ we adopted the case of setting $q(x) \equiv 0, \forall x \in A$. The same is of course true for B.
  I put this into a footnote, to explain our choice.
- p. 3, It would be useful to make a connection with the rate constant expressions derived in terms of committors in Mol. Phys., 104, 1497, 2006 and Int. Rev. Phys. Chem., 25, 237, 2006. The expression in equation (9) is probably equivalent to one of these results.

  The equation in (9) had an error and has been corrected and replaced by a more convenient and easier readable form. The correct expression is almost identical to the rate expressions derived in the papers [Mol. Phys. (2006) vol. 104 (9) pp. 1497-1507] or [Int. Rev. Phys. Chem. (2006) vol. 25 (1&2) pp. 237 - 282] for the steady-state approximations, referred to as k_{AB}^{SS} in these publications. We also assume, that after each jump, the system is in each state (A, B and intermediate) instantaneously in equilibrium, in other words, we assume the Markov property. In addition we only treat the case (for convenience), where A and B consist of only one single state (minimum). According for the different direction of the rate constants and also row-/column-wise definition of the rate matrices, the difference of the rate k_{AB} in expression (9) and k_{AB}^{SS} is by a factor of $\pi(B)$ or P_B^{eq}, which means, that the rate k_{AB} in (9) is the absolute rate of "reactive" trajectories $B \to A$ (or reverse, which depends on the different definition) and k_{AB}^{SS} is the rate of reactive trajectories emerged from B.
- p. 5, If h^A is equivalent to a mean first passage time then this is probably worth mentioning.

  h^A is the hitting-time, which is the minimum number of steps to reach subset A given a particular discrete trajectory (or stochastic process) $X(\omega)$ out of all possible trajectories, that started at $X_0 = x$. If we compute the expectation of the hitting time over all trajectories we get the mean first passage time

  $$\text{mfpt}(x) = \mathbb{E}_X \left(h^A(X) \mid X_0 = x \right)$$

  I added a sentence explaining this to make the distinction clearer and revised the definition of h^A.

- p. 10ff, My experience of iterative methods for calculating the committor probabilities, or eigenvalues of the transition matrix using Lanczos or Arnoldi-type methods, is that insurmountable numerical problems arise when the process of interest becomes slow. Here, "slow" means that the eigenvalue of interest becomes comparable to the precision of the zero eigenvalue of the transition matrix corresponding to the static equilibrium solution. The authors' approach certainly seems more efficient than these other iterative methods, but I wonder sensitive it is to this effect? Is there a clear separation in the eigenspectrum for the examples the authors consider? One particular advantage of the NGT procedure described in reference [34] is that precision is retained, even for very rare events. With the assumed dynamics in the two model systems it would probably be straightforward to compare the performance of the authors' approach with NGT, and this might be a useful avenue for future work. In the mean time, can the authors provide an estimate of how low kT can be compared to the highest overall barrier for their approach?

  We agree with the reviewers opinion, that numerical issues are indeed a big problem in the case of "slow" events. For the used Power Method, the speed of convergence depends on the spectral gap between the first two largest eigenvalues and is thus not optimized for the case of rare events. Thus, the reformulation into an eigenvector problem does not directly help with this issue. However the simplicity allows for a memory efficient implementation with which very large systems can be treated, provided, that the spectral gap is large enough. This is presented in the 3D model, where probability can travel fast enough through the network since the diameter due to the cubicle geometry is small (¡300) and the transition probabilities are not too small. Thus convergence is slow, but assured. In the 2D example, the smaller size allows for a complete error and sensitivity analysis with medium high barriers ($\lambda_2 = 0.999369$). Conclusively for faster and more accurate convergence, especially in the presence of sparse matrices more intelligent methods need to be used, such as Krylov space algorithms or graph based approaches like NGT[36]. The application of the Power Method to very slow processes and its numerical limits might be good way for future investigations.
  I extended the discussion about the power method by a short paragraph explaining its main area of applicability and the need of more advanced methods for dynamics with very slow processes.
- p. 26, "The method is efficently" $\rightarrow$ "The method is efficient"

  Corrected!

Thank you for the helpful and interesting points of discussion!

Kind regards,

Jan-Hendrik Prinz, corresponding author