A Comparative Study of Numerical Algorithms in Calculating Eigenpairs of the Master Equation for Protein Folding Kinetics

Yiming Li
Department of Computational Nanoelectronics, National Nano Device Laboratories
Microelectronics and Information Systems Research Center, National Chiao Tung University
Corresponding address: P.O. BOX 25-178, 1001 Ta-Hsueh Rd., Hsinchu 300, Taiwan; Email: ymli@faculty.nctu.edu.tw

Protein folding has recently been of great interest in both experimental and theoretical points of view [1-4]. In particular, kinetics of protein folding theoretically can be explored in different approaches, such as the mass action models [2], simulation with the all atoms or lattice model [3], and the methods between macroscopic and microscopic models [4]. It is known that a master equation plays a center role, among these approaches, in the computer simulation of the dynamics of protein folding. Deriving from the Liouvillian, a master equation basically describes the time-evolution of a distribution of trajectories (the transition of states). Solution of a master equation: $d\mathbf{P}(t)/dt = A\mathbf{P}(t)$ in general requires calculating a set of eigenvalues λ_N and eigenvector (e.g., the largest nonpositive eigenvalue) for the corresponding matrix A, where $\mathbf{P}(t)$ is the N-dimensional vector of the instantaneous probability of the N conformations. Depending on different free energy, A is a sparse and asymmetric N x N transition (or rate) matrix describing the kinetics of the transitions between these conformations. In particular, people may interest in the largest nonpositive eigenvalue. Because it suggests that some eigenvectors with $\lambda_N << 0$ disappear quickly. On the other hand, any other eigenvectors with $\lambda_N \sim 0$ will determine the slowest dynamical relaxation processes. The size of matrix A grows very quickly with respect to the studied protein length. For examples, a protein with 10 residues has a matrix with N in the order of 10^3; however, for the case of a protein with 20 residues, it leads to a matrix with N in the order of 10^7. To investigate the folding rate for a practical kinetics problem of protein folding (in general it is with more than hundreds of residues), we are encountering a large sparse matrix eigenvalue problem in calculating the desired eigenparis for the master equation.

In this paper, we apply different eigenvaule algorithms to calculate some larger nonpositive eigenvalues and their corresponding eigenvectors of the matrix A coming from the mater equation for a protein folding problem. In terms of the accuracy, stability, and robustness, different algorithms, such as power method, implicitly restarted Arnoldi method, Jacobi-Davidson method, and QR algorithm are compared with respect to different problem size. It is found that the implicitly restarted Arnoldi method is sensitive to the initial guesses and the Jacobi-Davidson method converges slowly for some testing cases. As shown in Fig. 1, for a protein with 9 residues [1], we compute all eigenvalues of the problem using QR method. It is found that the computed largest nonpositive eigenvlaue is equal to -0.04223398667661. QR is the robust algorithm; however, it demands huge storage memory and CPU time when matrix size is large. Therefore, it becomes a drawback if we want to examine a protein with more residues. In our numerical experience, power method is quite sensitive to the initial guesses in comparing with the implicitly restarted Arnoldi method. Shown in Table 1, we have tested the stability of the implicitly restarted Arnoldi method with different initial guesses. We found the implicitly restarted Arnoldi method is sensitive to the initial guesses which complicate the solution of master equation. Not shown here, we have also tested a protein with 15 residues and find a similar stability problem when using the implicitly restarted Arnoldi method.

Different algorithms for solving large sparse matrix eigenvalue problem have been compared and advised for protein folding kinetics. Combining several algorithms and using a biological-based observation, a robust and accurate hybrid method for solving the eigenvalues of mater equation will also be discussed for the kinetics of protein folding.

A full journal publication of this work will be published in the Journal of Computational Electronics.

This work is supported in part by Taiwan NSC under contracts No. NSC – 92 – 2112 – M – 429 - 001 and NSC – 93 – 2752 – E – 009 – 002 - PAE and the grant of Ministry of Economic Affairs, Taiwan under contract No. 92 - EC - 17 - A – 07 - S1 - 0011.

[1] M. Akahoshi *et al*., *Proc. Int. Conf. Biotech. Computing* **5** (1994) 173; U. H. E. Hansmann *et al*., *Computing in Science & Engineering* **5** (2003) 64; C.-D. Huang *et al*., *IEEE Trans. NanoBioscience* **2** (2003) 221; E. S. Keum ., *Proc. IEEE Bioinformatics Conf*. 2003, 615; K. A. Dill *et al*., *Protein Science* **4** (1995) 561.

[2] A. Ikai and C. Tanford, *Nature* **230** (1971) 100; K.A. Dill and H.S. Chan *Nat. Struct. Biol.* **4** (1997) 10; S. W. Englander *Biomol. Struct.* **29** (2000) 213.

[3] Y. Duan and P. A. Kollman, *Science* **282** (1998) 740; 1998; J.-E. Shea and C. L. Brooks III, *Annu. Rev. Phys. Chem*. **52** (2001) 499; V. Daggett, *Acc. Chem. Res*. **35** (2002) 422; S. Banu Ozkan et al., *Biopolymers* **68** (2003) 35; S. Altmeyer *et al*., *Artifcial Life* **10** (2004) 23.

[4] E. Alm and D. Baker, *Proc. Natl. Acad. Sci.* **96** (1999) 11305; E. Alm *et al*., *J. Mol. Biol*. **322** (2002) 463; Micheelsen *et al*., *Europhys. Lett.* **61** (2003) 561.

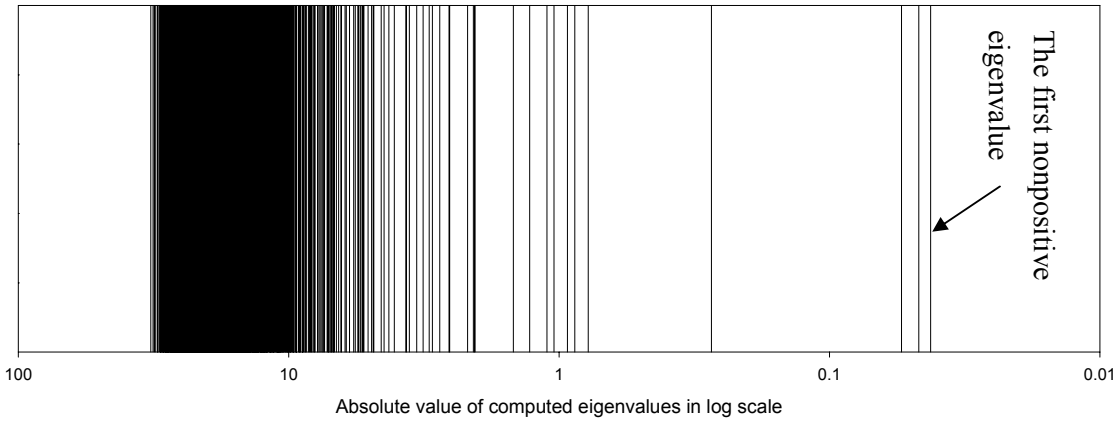

Figure 1: The distribution of all computed eigenvalues using the QR algorithm. The plot is for the absolute value of the computed eigenvalues. The first three largest nonpositive eigenvalues are -0.04223398667661, -0.0467563927012, and -0.05420731932385, respectively. They determine the smallest mode of folding.

Table 1: A stability test of the implicitly restarted Arnoldi method with different initial guesses for the protein with 9 residues. The algorithm is sensitive to initial guesses except the one near the solution -0.042.

Initial guesses	The 1st eigenvalue	The 2nd eigenvalue
1	0	-0.04223398667659
0.04	0	-0.04223398667659
0.0001	0	-0.04223398667664
0	0	0.01827176116506
-0.001	0	-0.04223398667660
-0.04	-0.04223398667659	-0.04675639270119
-1	-1.04365630038126	-0.93086218558657