Dynamics on free energy landscapes via relaxation of large kinetic transition networks

Boris Fačkovec and David J. Wales
Department of Chemistry, University of Cambridge

The grouping problem

Energy landscapes studies are performed in three consecutive steps:
1. **Partitioning** of configuration space into cells corresponding to Markov states (usually basins of local minima, [EL])
2. **Construction of the rate matrix** for the master equation [ME] (most conveniently using transition state theory [TST] with harmonic superposition approximation; traditionally by Markov state modeling [MSM] with some treatment of the embedding problem [EMB], more accurately by discrete relaxation path sampling [DRPS])
3. **Extracting observables** from the rate matrix (commonly [DPS] by regrouping of states [RGRP] and a graph transformation [NGT])

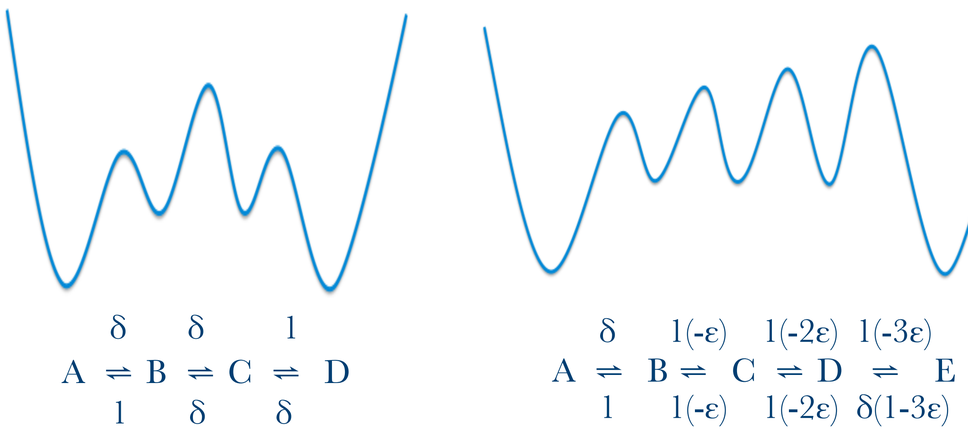
The characteristic properties of the **rate matrices** are
1. **Size**: often of order $>10^5$
2. **Condition number**: very large [SaW]
3. **Sparseness**: the number of relevant transition states is usually about twice the number of the relevant minima

There are two distinct issues of the grouping problem
1. **Calculation problem**: How to efficiently and accurately calculate the rates for the transitions between observables from the rate matrix?
2. **Grouping problem**: How to identify groups of states forming metastable sets, so that a particle stays a long time withing the group before transitioning into another group? This (clustering) problem is also studied in machine learning community [ML].

We propose solutions to both problems
1. **Relaxation approach for extracting observables**, inspired by DRPS
2. A **relaxed recursive regrouping [RRR]**, which is an inproved version of a similar algorithm ("REGROUPFREE" [RGRP]) previously developed in our laboratory.

Breakdown of TST

For pathological cases, flux-based expressions for rates will give nonsensical results



Boundaries between experimentally observed states (defined by a macroscopic quantity, e.g. FRET intensity) rarely coincide exactly with optimal transition states between superbasins. Hence REGROUPFREE will heavily, depend on a regrouping threshold used, making interpretation of results more difficult.

The recursive REGROUPFREE algorithm will also neglect diffusion-like repetitive transitions, which significantly reduce the overall rates.

Relaxation approach

Let us study a system given by a rate matrix of order n with eigenvalues λ_0 .. λ_{n-1}, and eigenvectors ϕ_0 .. ϕ_{n-1}. Using eigenvalue decomposition of the rate matrix the flux expression can be rewritten as (c_i denotes overlap of ϕ_i with state A)

$$k_{BA}^{\mathrm{TST}} = \frac{\partial_t P_A(t)}{P_A(t)}|_{\mathrm{eq}} = -\sum_{i=0}^{n-1} \lambda_i c_i \sum_{a\in A} \phi_{ia}$$

It can be also easily shown that the rate constant based on integration of the **relaxation time** for transition from A in equilibrium to equilibrium between A and B is

$$k_{BA}^{\mathrm{rxn}} = -\frac{(P_B^{\mathrm{eq}})^2}{\sum_{i=1}^{n-1} \frac{c_i}{\lambda_i} \sum_{a\in A} \phi_{ia}}$$

This result is equal to the result previously found by Hummer and Szabo [EXCT].

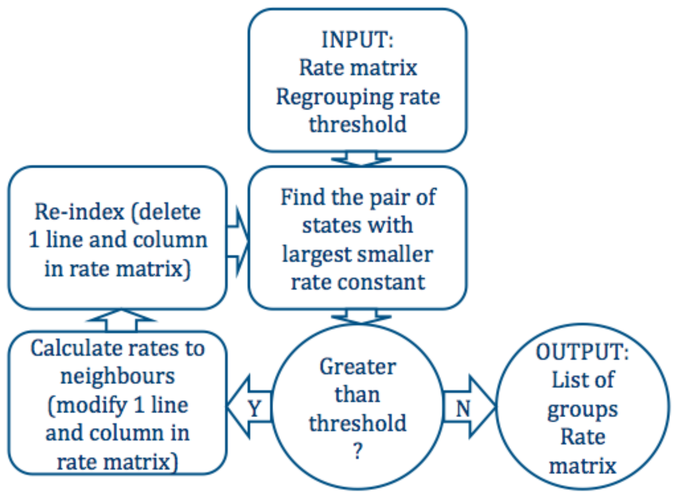
Relaxing algorithm

Since eigenvalue (or singular value) decomposition of rate matrices is very inaccurate due to poor conditioning and expensive due to matrix size, we propose a **numerically stable and cheaper algorithm** for calculating relaxation times. The algorithm is a combination of a linear propagation of population

$$p_i(m\Delta t) = \sum_{j=1}^{n} t_{ij}(\Delta t) p_i[(m-1)\Delta t]$$

(where population in state i at time $m\ \Delta t$, where m is an integer, and $t_{ij}(\Delta t)$ is an element of the transition matrix) and a multiplication of a transition matrix by itself corresponding to doubling the timescale described by the transition matrix. Arbitrarily long timescales can be reliably reached

Relaxed recursive regrouping



Boundary conditions

When extracting a pair of grouped states from the whole matrix, proper boundary conditions must be imposed at the interface between relaxing states and the neighbours:
1. **no-flux boundary conditions**: transitions between relaxing states and neighbours that are not involved in relaxation are disregarded completely = lower bound for the rate.
2. **fast-equilibration boundary conditions**: rates to the neighbours speed up equilibration within relaxing states = upper bound for the rate.

Conclusion

1. We have developed a successful method for calculating rates between grouped states.
2. We have devised a numerically accurate algorithm for relaxation.
3. By imposing different conditions, we can determine the interval for the exact value.
4. We have developed a recursive algorithm scaling almost linearly with size of the studied sparse rate matrix.

References

[DPS] Wales, 2002, Mol. Phys. 100, 3285
[DRPS] Fačkovec et al., 2015, J. Chem. Phys. 143, 044119
[EMB] McGibbon and Pande, 2015, J. Chem. Phys. 143, 034109
[EL] Wales, 2003, *Energy landscapes*, Camb. Uni. Press
[EXCT] Hummer and Szabo, J. Phys. Chem. B 119, 9029
[ME] van Kampen, 1992, Elsevier
[ML] Lafon and Lee, 2006, IEEE Trns. Patt. Mach. Intel
[MSM] Bowman et al., 2014, Springer
[NGT] Wales, 2009, J. Chem. Phys. 130, 204111
[RGRP] Carr and Wales, 2009, PCCP 11, 3341
[RRR] Fačkovec and Wales, in preparation
[SaW] Stevenson and Wales, 2009, J. Chem. Phys. 141, 041104
[TST] Wigner, 1937, J. Chem. Phys. 5, 720

Acknowledgement

Bakala Foundation
Department of Chemistry
King's College, Cambridge

Contact

bf269@cam.ac.uk
dw34@cam.ac.uk
Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom
boris.fackovec.net