Computation of equilibrium densities in metastable dynamical systems by domain decomposition

Susanna Kube and Marcus Weber

Zuse Institute Berlin, Takustraße 7, D-14195 Berlin, Germany

Abstract. Whenever the stationary density of molecular dynamical systems decomposes into almost invariant partial densities, its computation from long-time dynamics simulations is infeasible within the available computer time due to the well-known "trapping problem". In order to avoid this computational difficulty, we suggest a domain decomposition approach that is similar to umbrella sampling methods. In contrast to standard umbrella sampling techniques, our decomposition forms a partition of unity such that the corresponding stationary density can be computed as eigenvector of some mass matrix. This approach has many advantages over traditional approaches used to unbias and recombine the umbrella sampling calculations. The theoretical analysis is illustrated by a two-dimensional example.

Keywords: stationary distribution, metastability, Monte Carlo integration, domain decomposition
PACS: 02.50.Ng, 02.60.Dc, 02.70.Ns, 87.15.hp

INTRODUCTION

Among dynamical systems that possess a unique stationary distribution, there exist *metastable* dynamical systems characterized by the fact that their invariant set can be decomposed into *almost invariant* subsets. Once such a system is within one of these subsets, it stays there for "a long time" before it rapidly switches to another almost invariant subset. In molecular dynamical systems, which are the focus of our applications, the almost invariant subsets are named *metastable conformations*.

Throughout this paper, we consider dynamical systems in phase space $\{(q,p)|\, q \in \Omega \subset \mathbb{R}^d,\, p \in \mathbb{R}^d\}$. Moreover, without loss of generality, we fix the stationary density in position space to be the Boltzmann density

$$\pi(q) = \frac{1}{Z_q} \exp(-\beta V(q)).$$

$\beta = 1/(k_B T)$ denotes the inverse temperature with Boltzmann constant k_B, $V(q)$ refers to the potential energy of the system, and Z_q is the corresponding spatial partition function.

Molecular properties are determined by ensemble averages

$$\langle A \rangle \equiv \int_\Omega A(q)\pi(q)\, dq.$$

One can evaluate such high-dimensional integrals numerically by *importance sampling*,

$$\langle A \rangle \approx \frac{1}{M} \sum_{k=1}^{M} A(q_k), \quad q_k \sim \pi(q),$$

where the *sampling points* q_k are distributed according to the density $\pi(q)$.

The most famous class of algorithms for the generation of sampling points according to a given distribution are *Markov Chain Monte Carlo* (MCMC) methods [1, 2]. MCMC methods are based on constructing a Markov chain $\{q(t_i)\}_{i=1}^M$ that has the desired distribution as its stationary distribution. However, any of the available strategies for sampling from the stationary distribution will, when applied to metastable dynamical systems, suffer from the undesirable fact that they are prone to be "trapped" within conformations.

In this paper, we propose an algorithm that circumvents the trapping problem by realizing a domain decomposition approach in Ω. In other words, the Boltzmann density is split into the weighted sum of several partial densities $\pi_i(q)$,

$$\pi(q) = \sum_{i=1}^{N} w_i \pi_i(q), \tag{1}$$

which can be sampled independently without requiring transitions between conformations. The goal of the partitioning is to separate the conformations such that the partial densities do not inherit metastabilities. This approach is similar to the famous umbrella sampling method by Torrie and Valleau [3].

The difficult task is to recombine the information from single samplings into the overall density, i.e. the computation of coupling factors w_i. Our algorithm reduces the problem of finding the coupling weights to an algebraic eigenvalue problem, for which error estimators are available. The main advantage of our algorithm compared to others, for example the Weighted Histogram Analysis Method (WHAM), is the possibility of a hierarchical refinement of biasing potentials, which leads to a better approximation of the stationary density.

METHODS

Discretization and sampling

The decomposition of high-dimensional spaces is a difficult problem. Regular grids are well-known to be inappropriate, because they suffer from the "curse of dimensionality". In order to overcome this problem, a meshfree discretization can be realized in terms of particles or *nodes* $\{q_i\}_{i=1}^N \in \Omega$ that define a set of basis functions $\phi_i(q) \equiv \phi_i(q_i, q)$. Our algorithm realizes a *decomposition in function space* via *membership functions* ϕ_i:

1. $\sum_{i=1}^N \phi_i(q) = 1, \quad \forall q \in \Omega$ (partition of unity)
2. $\phi_i(q) > 0, \quad \forall q \in \Omega,\ i = 1, \ldots, N$ (positivity)

Several types of basis functions could be used. However, our algorithm only works for "overlapping" basis functions, i.e. for every i there exists an index $j \neq i$ such that $\int_\Omega \phi_i(q)\,\phi_j(q)\,\pi(q)\,dq > 0$. One possible realization are *radial basis functions*, see [4, 5].

The mass matrix S is defined as

$$S(i,j) := \frac{\int_\Omega \phi_i(q)\,\phi_j(q)\,\pi(q)\,dq}{\int_\Omega \phi_i(q)\,\pi(q)\,dq}.$$

It is a row-stochastic matrix, i.e. $\sum_{j=1}^N S(i,j) = 1$, $S(i,j) \geq 0$. The integral can be rewritten in terms of partial densities

$$\pi_i(q) \equiv \frac{\phi_i(q)\pi(q)}{\int_\Omega \phi_i(q)\pi(q)}. \tag{2}$$

The denominator is the statistical weight of basis function $\phi_i(q)$,

$$w_i \equiv \int_\Omega \phi_i(q)\pi(q)\,dq. \tag{3}$$

The vector $\mathbf{w} = [w_1, \ldots, w_N]^\top$ is the left eigenvector of S corresponding to the eigenvalue $\lambda = 1$, $\mathbf{w}^\top S = \mathbf{w}^\top$. It is denoted as *stationary density* of S. According to the Perron-Frobenius Theorem ([6], Thm. 1.4.4.), the eigenvalue $\lambda = 1$ is maximal in modulus among all eigenvalues of S and geometrically simple. Thus, $\mathbf{w}$ is determined uniquely. Moreover, it is positive.

The partial densities $\pi_i(q)$ are Boltzmann densities corresponding to modified potentials $V_i(q)$ [5], i.e. $\pi_i(q) \propto \exp(-\beta V_i(q))$ with $V_i(q) = V(q) - \beta^{-1}\log(\phi_i(q))$. Using these notations results in

$$S(i,j) = \int_\Omega \phi_j(q)\pi_i(q)\,dq. \tag{4}$$

Once the stationary density vector $\mathbf{w}$ of S has been computed, the overall density $\pi(q)$ is obtained by (1).

The computation of matrix entries $S(i,j)$ requires the solution of high-dimensional integrals, which is realized by *Monte Carlo integration*,

$$S(i,j) \approx \frac{1}{n_i}\sum_{k=1}^{n_i} \phi_j(q_k^{(i)}), \quad q_k^{(i)} \sim \pi_i(q).$$

The sampling points $\{q_k^{(i)}\}_{k=1}^{n_i}$ are distributed according to the partial density $\pi_i(q)$. They are generated by *Hybrid Monte Carlo* (HMC), a variant of MCMC [7, 8]. This can be considered as umbrella sampling with modified potential $V_i(q)$.

Error analysis

Actually, we compute just an approximation $\widetilde{S} = S + E$, where E denotes some error matrix. Since the entries of S are computed by Monte Carlo quadrature, it is reasonable to assume random perturbations and to approximate S by the mean $\overline{S}$. For this purpose, we take E from the class of row-wise correlated random matrices with $\mathbb{E}[E(i,j)] = 0$ and $\mathbb{E}[E(i,j)E(k,l)] = \delta_{ik}C_i(j,l)$ for $i,j,k,l = 1,\ldots,N$. The operator $\mathbb{E}$ is the expectation operator. The fact that different matrix rows are uncorrelated is verified by our domain decomposition approach, in which all rows are sampled independently. Following the approach by Singhal and Pande [9], the first order Taylor series expansion for the stationary distribution $\mathbf{w}$ as a function of the matrix entries $s_{ij} \equiv S(i,j)$ is given by

$$\mathbf{w} \doteq \overline{\mathbf{w}} + \sum_{i=1}^{N} G_i^{\mathbf{w}} E(i,:)^\top, \tag{5}$$

where $G_i^{\mathbf{w}}(k,l) = \frac{\partial w_k}{\partial s_{il}}\big|_{\overline{A}}$ with $\overline{A} = \overline{S}^\top - I$. The entries of the sensitivity matrix $G_i^{\mathbf{w}}$ are obtained from the solution of a system of linear equations.

The question remains of how to estimate the mean $\overline{S}$ and the covariance matrices C_k. In order to obtain an indicator whether the sampling chains are rapidly mixing within the basis functions, we run multiple, say c, HMC chains for each partial density $\pi_i(q)$. First, one could test whether the chains sample from the same partial density by computing the Gelman-Rubin convergence factor r, which should decline to one [10]. Moreover, every chain results in a different estimate $\mathbf{s}_i^{(k)}$, $k = 1,\ldots,c$, for row i of S. These rows can be considered as a training set of generalized count vectors, $D = \{n_i^{(1)}\mathbf{s}_i^{(1)},\ldots,n_i^{(c)}\mathbf{s}_i^{(c)}\}$, where $n_i^{(k)}$ denotes the length of chain k. This training set represents the possible outcome of a *Pólya* distribution [11]. The parameters of the distribution can be obtained by maximum-likelihood estimation [12]. As the chain lengths increase, the different estimates for the row will converge to the exact mean. Moreover, if c is large enough, then, by the central limit theorem, the distribution of a row $\mathbf{s}_i = S(i,:)$ converges to a multivariate normal distribution with mean μ_i and covariance matrix C_i. In this case, the stationary density $\mathbf{w}$ will follow a multivariate normal distribution,

$$\mathbf{w} \sim \mathrm{MVN}(\overline{\mathbf{w}}, \sum_{i=1}^{N} G_i^{\mathbf{w}} C_i (G_i^{\mathbf{w}})^\top). \tag{6}$$

Hierarchical refinement

Assume we are given a set of basis functions $\{\phi_1,\ldots,\phi_N\} : \Omega \to [0,1]$ whereof basis function k is selected for hierarchical refinement into s new basis functions. Then, a temporal set of basis functions $\{\widetilde{\phi}_{k1},\ldots,\widetilde{\phi}_{ks}\} : \Omega \to [0,1]$ is constructed such that this basis forms a partition of unity and meets the positivity constraint. Finally, the new basis functions are computed by $\phi_{ki}(q) := \phi_k(q)\widetilde{\phi}_{ki}(q)$, $i = 1,\ldots,s$. In fact, the new basis functions $\{\phi_1,\ldots,\phi_{k-1},\phi_{k+1},\ldots,\phi_N,\phi_{k1},\ldots,\phi_{ks}\}$ form a partition of unity and meet the positivity constraint.

The refinement criterion will be based on the above error estimator. In order to reduce the variance of $\mathbf{w}$, we apply the following greedy algorithm. The basis function with index

$$i = \arg\max_k \|G_k^{\mathbf{w}} C_k (G_k^{\mathbf{w}})^\top\|_2 \tag{7}$$

is the one that contributes most to the variance of $\mathbf{w}$. By adding more sampling points to this basis functions, the variance is expected to reduce. If a predefined maximum number of sampling points is reached for this function, then it will be refined hierarchically.

AN ILLUSTRATIVE EXAMPLE

We consider a two dimensional periodic domain $\Omega = [-\pi,\pi]^2$ and an artificial potential energy

$$V(q) = -5\cos(3q_x) + 0.5q_x^2 - 5\cos(3q_y) + 0.5q_y^2. \tag{8}$$

For $\beta = 0.4$, the corresponding stationary density $\pi(q) \propto \exp(-\beta V(q))$ decomposes into nine well separated peaks, which represent the metastable conformations; see the left picture in Fig. 1. A discretization by radial basis functions

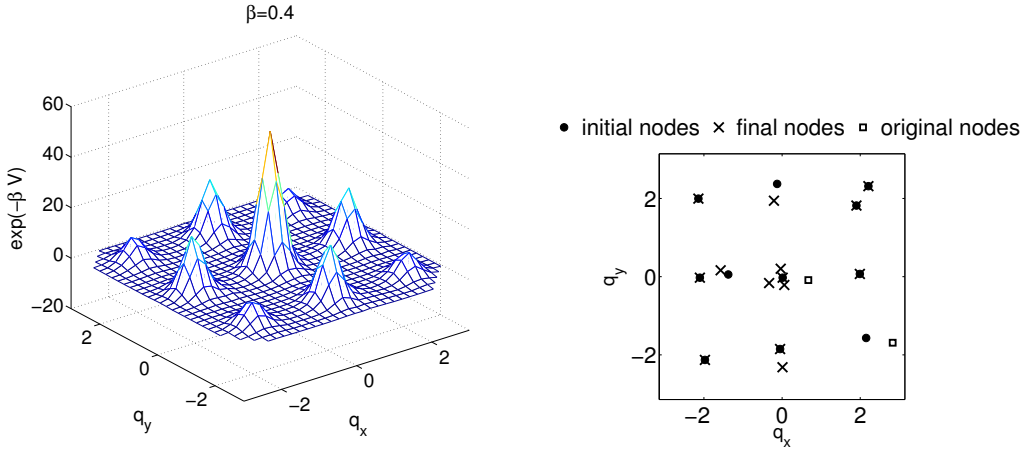

FIGURE 1. Illustrative example: Stationary density for $\beta = 0.4$ (left) and automatic state space decomposition according to Gelman-Rubin convergence (initial nodes) and equilibration of variance contributions (final nodes).

(shape parameter $\alpha = 2$) with two randomly selected nodes was first refined hierarchically to achieve Gelman-Rubin convergence (target convergence factor $r \leq 1.3$; 5 chains in every basis function; refinement whenever the number of sampling points exceeded $5 \cdot 2000$). This resulted in an initial discretization with 11 basis functions. Further adaptive sampling and hierarchical refinement were applied to achieve an equilibration of the values $\sigma_k \equiv \|G_k^{\mathbf{w}} C_k (G_k^{\mathbf{w}})^\top\|_2$ among the basis function until $\sigma_i \leq \frac{3}{2}\text{median}_k(\sigma_k)\,\forall i$ (adding $5 \cdot 500$ points to the function with index from (7); refinement whenever the number of sampling points exceeded $5 \cdot 5000$). The discretization steps are shown on the right hand side of Fig. 1. Finally, one obtains the coupling factors w_i together with entry-wise error distributions (results not shown). They are used to approximate observables, for example the inner energy $U = \int_\Omega V(q)\pi(q)\,dq$. For the sampling with Gelman-Rubin convergence, the value amounts to $U = -2.63$, whereas the value after adaptive sampling and further refinement amounts to $U = -4.77$ (analytic result: $U = -5.03$).

CONCLUSION

The algorithm proposed in the present paper is intended as an alternative method for the computation of stationary densities in molecular dynamical systems. Due to the limited space, a comparison with well-established methods such as the WHAM will be presented elsewhere. Nevertheless, we want to point out that, in contrast to other methods, the possibility of hierarchical refinement makes our algorithm independent of any initial discretization and thus very flexible.

REFERENCES

1. N. Metropolis, A. W. R. ans M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087–1092 (1953).
2. W. K. Hastings, *Biometrika* **57**, 97–109 (1970).
3. G. M. Torrie, and J. P. Valleau, *J.Chem.Phys.* **66**, 1402–1408 (1977).
4. S. Li, and W. K. Liu, *Meshfree Particle Methods*, Springer, 2004.
5. M. Weber, *Meshless Methods in Conformation Dynamics*, Doctoral thesis, Freie Universität Berlin (2006), published by Verlag Dr. Hut, München, `http://www.zib.de/bib/diss/index.de.html`.
6. R. B. Bapat, and T. E. S. Rhagavan, *Nonnegative Matrices and Applications*, Cambridge University Press, 1997.
7. S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, *Phys. Lett. B* **195**, 216–222 (1987).
8. C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard, *J. Comput. Phys.* **151**, 146–168 (1999).
9. N. S. Hinrichs, and V. S. Pande, *J. Chem. Phys.* **126** (2007).
10. A. Gelman, and D. B. Rubin, *Stat. Sci.* **7**, 457–511 (1992).
11. J. E. Mosimann, *Biometrika* **49**, 65–82 (1962).
12. T. P. Minka, Estimating a Dirichlet distribution, Tech. rep., M.I.T. (2000), `http://research.microsoft.com/~minka/papers/dirichlet`.