Uncoupling-Coupling Techniques for Metastable Dynamical Systems[*]

Christof Schütte, Ralf Forster, Eike Meerbach, and Alexander Fischer

Institute of Mathematics II, Free University Berlin, Arnimallee 2–6, 14195 Berlin, Germany
URL: `http://page.mi.fu-berlin.de/~biocomp`

Summary. We shortly review the uncoupling-coupling method, a Markov chain Monte Carlo based approach to compute statistical properties of systems like medium-sized biomolecules. This technique has recently been proposed for the efficient computation of biomolecular conformations. One crucial step of UC is the decomposition of reversible nearly uncoupled Markov chains into rapidly mixing subchains. We show how the underlying scheme of uncoupling-coupling can also be applied to stochastic differential equations where it can be translated into a domain decomposition technique for partial differential equations.

1 Introduction

Application of Markov chain Monte Carlo (MCMC) to biomolecular systems has to tackle the *trapping problem*, i.e., the Markov chain remains for a very long time in one part of the state space before it moves on to another part. Such undesirable behavior of the Markov chain is caused by *metastable sets* (also called *modes* or *conformations*) in the state space, between which transitions are extremely rare. There exists a huge body of literature addressing this notoriously difficult problem (Ferguson et al. [1999], Liu [2001]).

We herein review a novel approach to overcome the trapping problem, the uncoupling-coupling scheme (UC), which has recently been introduced by one of the authors (AF) (Fischer [2003], Fischer et al. [2002]). UC combines statistical reweighting techniques with a *hierarchical decomposition* of the state

space into metastable sets. The key idea is to regard metastable sets as *almost invariant sets* w.r.t. the Markov chain.

It has been shown recently that these metastable sets are strongly connected to the spectral structure of the Markov propagation operator associated with the Markov chain (Schütte et al. [1999]), and that it is even possible for a wide range of problem classes to identify metastable sets by computing dominant eigenvalues of this operator (Deuflhard et al. [2000]). Once n dominant metastable sets are identified, significantly improved convergence properties are achieved by *uncoupling*, i.e., by parallel simulation of n independent chains, each one restricted to one of the metastable sets. Subsequently, information lost in the uncoupling step, i.e., the weighting factors between the n metastable sets, is reconstructed by means of the stationary distribution of an appropriate $(n \times n)$ coupling matrix. We present the uncoupling step in the context of MCMC in Sect. 2, and the underlying uncoupling-coupling scheme in its simplest form in Sect. 3.

In its final section this article is devoted to the demonstration that the UC idea can be translated into a domain decomposition technique for eigenvalue problems of specific partial differential operators. The translation is possible since Markov propagation operators of specific Markov processes (e.g., those governed by stochastic differential equations) are generated by partial differential operators. In such cases the decomposition of state space into metastable sets is strongly connected to the dominant eigenmodes of the generator, and the UC scheme for the propagation operator can be translated into an analogous scheme for the generator. This relation may make uncoupling-coupling techniques as used in biomolecular simulations accessible for research in the direction of domain decomposition.

2 Uncoupling-Coupling Markov Chain Monte Carlo

For biomolecular simulations, MCMC is the method of choice for the task of drawing samples from the canonical distribution. In the presence of strong metastabilities slow convergence can be avoided by uncoupling-coupling, where the state space is decomposed into metastable sets.

Metropolis Algorithm

The Metropolis (or Metropolis-Hastings) algorithm is the most widely used form of MCMC and essentially builds upon Markov chain theory (Brémaud [1999], Liu [2001]).

Suppose that we are interested in a distribution given by a density function $f > 0$ with values in $\Omega \subseteq \mathbb{R}^d$, from which it is practically impossible to draw independent samples (e.g., the canonical distribution of medium-sized biomolecules, where d is in the range of 50 to 500). Usually, f is defined in terms of an *unnormalized density* $\hat{f}$ via

$$f(x) = \frac{\hat{f}(x)}{Z_{\hat{f}}} \quad \text{with} \quad Z_{\hat{f}} = \int_{\Omega} \hat{f}(x)\, dx, \tag{1}$$

where $Z_{\hat{f}}$ denotes the *normalizing constant* of $\hat{f}$. In most applications $\hat{f}$ is the canonical or Boltzmann density $\hat{f} = \exp(-\beta V)$ with inverse temperature β for some potential energy function $V : \Omega \to \mathbb{R}$.

The goal is to obtain expectations of some function g with respect to f, i.e., computing the expectation

$$I_f(g) = \int g(x) f(x)\, dx.$$

The Metropolis algorithm realizes a Markov chain $\mathcal{X} = X^{(1)}, X^{(2)}, X^{(3)}, \ldots$ having f as its invariant density. A sample $\mathbf{x} = (x^{(1)}, \ldots, x^{(n)})$ of $\mathcal{X}$ is obtained by accepting a proposal step $x_{\text{prop}}^{(k+1)}$ with a probability that only depends on the ratio of $f(x_{\text{prop}}^{(k+1)})/f(x^{(k)})$, thereby avoiding a computation of the unknown normalizing constant (which in its integral representation is typically hard to evaluate). The generated (dependent) random sample $\mathbf{x}$ then enables us to estimate the integral I_f by

$$\hat{I}_f(g) = \frac{1}{n} \sum_{k=1}^{n} g(x^{(k)}). \tag{2}$$

The evolution of a Markov chain $\mathcal{X} = (X_k)$ with state space Ω is defined by a *stochastic transition function* $K : \Omega \times \Omega \to \mathbb{R}$, where $K(x, A)$ is the probability density to move from x to the set A in one step (Meyn and Tweedie [1993]). We call f an *invariant density* of the Markov chain given by K, if

$$f(y) = \int_{\Omega} K(x, y) f(x)\, dx \tag{3}$$

holds for all $y \in \Omega$.

In the Metropolis-Hastings algorithm a transition function K which satisfies (3) is realized by first defining an arbitrary but *irreducible* transition kernel $q(x, y)$ together with the *acceptance function*

$$\alpha(x, y) = \begin{cases} \min\left(1, \frac{q(y,x)\, f(y)}{q(x,y)\, f(x)}\right) & \text{for } q(x,y) > 0 \\ 1 & \text{otherwise} \end{cases}. \tag{4}$$

The computation of α requires ratios of the form $f(y)/f(x)$ only, which is feasible even if the normalizing constant $Z_{\hat{f}}$ is unknown.

Based on q and α we define K as the sum of two contributions,

$$K(x, y) = k(x, y) + r(x)\delta(x - y),$$

where the absolutely continuous part k is given by

$$k(x, y) = \begin{cases} q(x, y)\alpha(x, y) & \text{if } x \neq y \\ 0 & \text{otherwise} \end{cases}$$

and the singular component by $r(x) = 1 - \int k(x, y)\, dy$.

With this K one step in the realization of the Markov chain from the state $X_k = x$ consists of: a) propose some y distributed according to $q(x, y)$, b) accept this step by setting $X_{k+1} = y$ with probability $\alpha(x, y)$ or c) reject the proposal leaving $X_{k+1} = x$.

The construction of K guarantees that the associated Markov chain $\mathcal{X}$ is irreducible—provided that q is irreducible—and that for all $x, y \in \Omega$ the detailed balance condition

$$f(x)\, k(x, y) = f(y)\, k(y, x) \tag{5}$$

holds (for details, see e.g.(Tierney [1994])). Due to (5) K is called *reversible* w.r.t. f. If we further assume that $\mathcal{X}$ is aperiodic—which is guaranteed whenever $r > 0$—we can state that f is the unique invariant density of $\mathcal{X}$.

Markov Operator

In the following we want to understand the global behavior of a Markov chain via the eigenmodes of its associated *Markov operator* P. This operator is defined in terms of the transition function K by

$$Pu(y) = \int_\Omega k(x, y)u(x)\, dx + r(y)u(y). \tag{6}$$

P describes the propagation of a phase space density with one step of the Markov chain. One can show that the reversibility of K w.r.t. f implies that its spectrum $\sigma(P)$ is real-valued. More exactly, we have $\sigma(P) \subseteq [-1, 1]$, and the largest eigenvalue is $\lambda_1 = 1$. We have $Pf = f$, and under some ergodicity conditions f is the unique eigenfunction associated with $\lambda_1 = 1$ (up to normalization). Some additional ergodicity typically is sufficient to guarantee that the essential spectrum $\sigma_{ess}(P)$ of P is bounded such that $\sigma_{ess}(P) \subset [-r, r]$ (see Schütte and Huisinga [2003]) and one typically can assume that there are several discrete eigenvalues λ with $|\lambda| > r$. If this is the case then these eigenvalues subsequently are assumed to be ordered due to their absolute value, i.e., such that $\lambda_1 > \lambda_2 \geq \lambda_3 \geq \ldots$.

Discretization

Identification and restricted sampling is carried out by discretizing the operator in essential degrees of freedom. To that end, let for two sets $A, B \subseteq \Omega$ the *transition probability* between A and B within an ensemble distributed w.r.t. the density f and during one step of the Markov chain be given by

$$\kappa(A, B) = \frac{1}{\int_A f(x)\,dx} \int_A \int_B K(x,y) f(x)\,dx\,dy. \tag{7}$$

Then, discretization is done by *coarse graining* with an arbitrary box decomposition of the phase space Ω into m disjoint sets $B_1, \ldots, B_m \subset \Omega$ with $\bigcup B_j = \Omega$. Based on this box decomposition, we introduce the new finite state space $S = \{B_1, \ldots, B_m\}$ and define the transition function $\tilde{K}$ on S via

$$\tilde{K}(B_k, B_l) = \kappa(B_k, B_l). \tag{8}$$

The finite dimensional Markov chain defined by $\tilde{K}$ again is reversible w.r.t. its invariant density $\tilde{f}$ given by $\tilde{f}(B_k) = \int_{B_k} f(x) dx$. Whenever f is unique for K, $\tilde{f}$ is also unique for $\tilde{K}$. Then the phase space is finite and the Markov operator P becomes an $(m \times m)$-transition matrix $\mathbf{P}$ which simply is the stochastic matrix with entries $p_{kl} = \tilde{K}(B_k, B_l) = \kappa(B_k, B_l)$.

Metastability

If λ_2 is close to $\lambda_1 = 1$, we often find that the reason for the undesirably slow convergence is that the Markov chain remains for a long time in a *metastable set* (or *conformation*) of the state space, before it moves on to another one. We will call a set A *metastable* under our Markov chain, if the transition probability from A to itself is close to one, i.e., if $\kappa(A, A) \approx 1$.

For an algorithmic exploitation of metastability the following observation is of importance: If there are n eigenvalues close to $\lambda_1 = 1$ (including λ_1 itself) and a significant spectral gap to all remaining eigenvalues, then there also are n disjoint metastable subsets and vice versa (Meyer [1989], Schütte et al. [2001]). If this is the case, the chain is rapidly mixing *within* the corresponding metastable subsets and the undesirably slow overall convergence results from the rareness of transitions between these metastable sets.

The close connection between a separated cluster of dominant eigenvalues and the existence of metastable subsets has another very important algorithmic consequence: it has been shown that one can *identify* the n metastable subsets only on basis of the *eigenvectors* associated with the n dominant eigenvalues (Schütte et al. [1999, 2001]). This insight leads to a significantly general identification algorithm (Deuflhard et al. [2000]) used for the detection of biomolecular conformations.

Restriction

Assume that we know the n disjoint metastable sets $A_1, \ldots, A_n$ of our Markov chain, and that we now want to sample separately in each A_l, for $l = 1, \ldots, n$. Then, for each l we define a restricted Markov kernel K_l from K on A_l by setting

$$K_l(x,y) = k_l(x,y) + r_l(x)\delta(x - y) \tag{9}$$

with

$$k_l(x,y) = \begin{cases} q(x,y)\alpha(x,y) & \text{if } x \neq y \text{ and } y \in A_l \\ 0 & \text{otherwise} \end{cases}$$

and

$$r_l(x) = 1 - \int k_l(x,y)\, dy.$$

Clearly, detailed balance still holds, so that K_l is again a reversible Markov kernel. Now, let $\hat{f}_l = \mathbf{1}_{A_l}\hat{f}$ be the restricted unnormalized density on A_l, with $\mathbf{1}_A$ denoting the indicator function on A, i.e., $\mathbf{1}_A(x) = 1$ if $x \in A$ and $\mathbf{1}_A(x) = 0$ otherwise. Then, under the assumption, that K_l is irreducible, $f_l = \hat{f}_l / Z_{\hat{f}_l}$ is the unique invariant density of K_l.

We denote by P_l the corresponding propagator of K_l. If we assume that A_l is metastable and that it cannot be subdivided further into two or more almost invariant sets, then we can state the following: The second largest eigenvalue λ_2 of P_l is substantially less than 1, otherwise there would exist a decomposition into two or more metastable subsets. As a consequence, due to $\lambda_2 \ll 1$, the corresponding Markov chain $\mathcal{X}_l$ is rapidly mixing.

For the restricted Markov kernel K_l the detailed balance condition (5) still holds for all $x, y \in A_l$; therefore the density f_l is a scalar multiple of the correct global density f of the unrestricted Markov chain. Thus, we can regain the global density via

$$f = \sum_{l=1}^{n} \xi_l f_l \tag{10}$$

in terms of the local densities f_l. Only the scalar coupling factors ξ_l, $l = 1, \ldots, n$, are unknowns which represent the neglected coupling between the sets A_l. Apparently, the coupling factors need to be ratios of normalizing constants of the form $\xi_l = Z_{\hat{f}_l} / Z_{\hat{f}}$, since then we can reconstruct f from the f_l's:

$$\sum_{l=0}^{n} \xi_l f_l = \sum_{l=0}^{n} \frac{Z_{\hat{f}_l}}{Z_{\hat{f}}} \frac{\hat{f}_k}{Z_{\hat{f}_k}} = \frac{\hat{f}}{Z_{\hat{f}}} = f. \tag{11}$$

Hierarchical Uncoupling-Coupling

Restricted sampling alone does not directly provide the necessary coupling vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)$ and also raises the question of how to decompose the state space. Yet, it is possible to overcome these problems by embedding some Metropolis sampler into a hierarchical annealing structure. For a detailed presentation of this approach we refer to the Uncoupling-coupling Monte Carlo method presented in Fischer [2003], Fischer et al. [2002].

The hierarchical annealing structure is a crucial part for the algorithmic concept in the context of biomolecular simulations with far reaching consequences for the coupling step. However, since it is of lesser importance in the PDE context, we focus on the basic uncoupling-coupling scheme in the following.

3 Basic Uncoupling-Coupling Scheme

We provide some theoretical aspects of uncoupling-coupling, especially the impact of uncoupling on the spectra of restricted Markov operators. Throughout this section we consider a finite state space. Therefore, let $\mathbf{P}$ be an irreducible, aperiodic and reversible stochastic $(m \times m)$-matrix, which might be obtained by a box discretization of a Markov operator P as defined in (8). In this case the state space Ω reduces to $S = \{B_1, \ldots, B_m\}$, the entries p_{ij} of $\mathbf{P}$ are transition probabilities $\kappa(B_k, B_l)$ of the Markov operator between the boxes B_i and B_j, and the global density f becomes a stochastic vector $\boldsymbol{\pi}$.

$\mathbf{P}$ is called a nearly uncoupled Markov chain, if there is a permutation of the state space such that $\mathbf{P}$ becomes diagonal block-dominant, i.e.

$$\mathbf{P} = \widetilde{\mathbf{P}} + \mathbf{E} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{E}_{12} & \cdots & \mathbf{E}_{1n} \\ \mathbf{E}_{21} & \mathbf{P}_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{E}_{n-1,n} \\ \mathbf{E}_{n1} & \cdots & \mathbf{E}_{n,n-1} & \mathbf{P}_{nn} \end{pmatrix}, \tag{12}$$

where each sub-matrix is quadratic and entries in $\mathbf{E}$ are small. In the context of metastability this corresponds to the existence of n metastable sets $S_1, \ldots, S_n$. Computation of an appropriate permutation is by no means trivial, but can be done by the identification algorithm already mentioned in the previous section (Deuflhard et al. [2000]). There are different ways to measure the smallness of $\mathbf{E}$, e.g. by the maximum row sum norm $\| \cdot \|_\infty$ or by some $\boldsymbol{\pi}$-weighted norm.

Reversibility of $\mathbf{P}$ is equivalent to the detailed balance condition

$$\pi_i p_{ij} = \pi_j p_{ji} \tag{13}$$

for all $1 \leq i, j \leq m$. From (13) it easily follows, that $\mathbf{P}$ is self-adjoint w.r.t. the weighted inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\boldsymbol{\pi}} = x_1 y_1 \pi_1 + \ldots + x_m y_m \pi_m. \tag{14}$$

Therefore all eigenvalues of $\mathbf{P}$ are real and contained in the interval $(0, 1]$. Since the diagonal blocks in (12) are nearly stochastic, continuity of the eigenvalues guarantees the existence of n eigenvalues close to 1. We assume that the other eigenvalues are reasonable bounded away from 1, which corresponds to the assumption that the Markov chain is fast-mixing within each metastable set.

The matrix

$$\mathbf{P}^{\text{rest}} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{R}_{nn} \end{pmatrix}, \tag{15}$$

where $\mathbf{R}_{ii} = \mathbf{P}_{ii} + \mathrm{diag}(\mathbf{e}_i - \mathbf{P}_{ii}\mathbf{e}_i)$ and $\mathbf{e}_i = (1, \ldots, 1)$ is a vector of size $|S_i|$, is called the restriction matrix of $\mathbf{P}$ along a given partition. $\mathbf{P}^{\mathrm{rest}}$ is obtained from $\mathbf{P}$ by setting all off-diagonal blocks in $\mathbf{P}$ to zero and adding the sum of the deleted entries of the i-th row to p_{ii}, which is the discrete counterpart of (9).

Since the diagonal blocks of $\mathbf{P}^{\mathrm{rest}}$ are stochastic we can conclude that the dominant eigenvalue 1 is n-fold, while due to continuity all other eigenvalues should be well separated from 1. If the smallness of $\mathbf{E}$ is measured by the ∞-norm, we can state a quantitative bound on this phenomena (Meerbach et al. [2003]):

Theorem 1. *Let $\mathbf{P}$ be a reversible stochastic matrix partitioned according to (12) and $\mathbf{P}^{\mathrm{rest}}$ the restricted matrix, as in (15). Then*

$$\lambda_j(\mathbf{P}^{\mathrm{rest}}) \leq \lambda_j(\mathbf{P}) + 2\|\mathbf{E}\|_\infty \tag{16}$$

holds for each $j = 1, \ldots, m$.

Therefore, by transition from $\mathbf{P}$ to $\mathbf{P}^{\mathrm{rest}}$, we obtain n uncoupled Markov chains restricted to the sets $S_1, \ldots, S_n$, whereby for a metastable decomposition each chain is fast mixing.

The following theorem summarizes some facts about $\mathbf{R}$ and reveals that the behavior of the uncoupled chains is closely related to that of the original chain (Meerbach et al. [2003], Meyer [1989]).

Theorem 2. *Let $\mathbf{P}$ be an irreducible and reversible stochastic matrix partitioned as in (12). Furthermore, let all $\mathbf{P}_{ii}$ be irreducible (substochastic) matrices. Then,*

(a) all $\mathbf{R}_{ii}$ are irreducible,
(b) $\mathbf{R}$ is stochastic with an n-fold dominant eigenvalue 1,
(c) if the (unique) stationary distribution $\boldsymbol{\pi}$ of $\mathbf{P}$ is partitioned according to $\mathbf{P}$,
$$\boldsymbol{\pi} = (\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)}, \ldots, \boldsymbol{\pi}^{(n)}),$$

then for each $i = 1, \ldots, n$ the unique stationary distribution $\mathbf{r}^{(i)}$ of the restriction $\mathbf{R}_{ii}$ is identical to $\xi_i^{-1}\boldsymbol{\pi}^{(i)}$, where $\xi_i = \sum_h \pi_h^{(i)}$ is a constant factor,
(d) the coupling vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ is the unique stationary distribution of the irreducible and stochastic coupling matrix $\mathbf{C} = (c_{ij})$, with

$$c_{ij} := \mathbf{r}^{(i)}\mathbf{P}_{ij}\mathbf{e}. \tag{17}$$

Theorem 2 states that coupling factors, which are needed to reweight the restricted stationary distributions of the $\mathbf{R}_{ii}$'s in order to obtain the stationary distribution of $\mathbf{P}$, can be derived via the coupling matrix $\mathbf{C}$ containing the transition probabilities between the metastable sets.
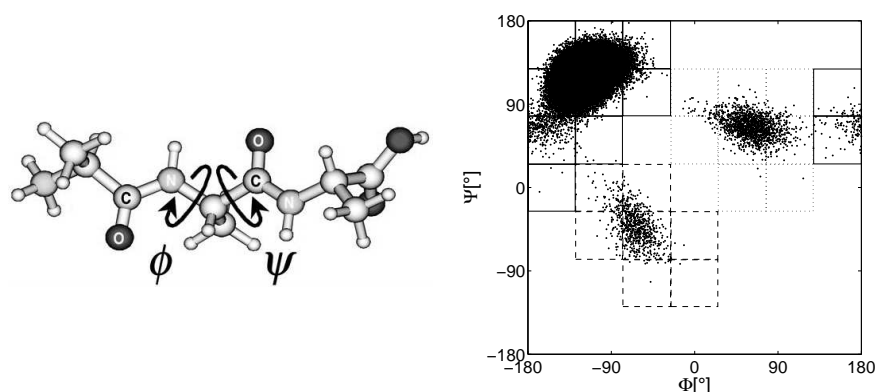
Fig. 1. Left: The trialanine molecule shown in ball-and-stick representation. The overall structure of trialanine is primarily determined by the two torsion angles Φ and Ψ. Right: Plotting Φ versus Ψ results in a so-called Ramachandran plot. The discretization boxes are plotted with different edge lines indicating the different metastable sets they were allocated to.

Theorem 1 is closely related to the theory of stochastic complementation (Meyer [1989]), where an analogous result is stated for the general case of non-reversible matrices. Associated with stochastic complementation are so-called aggregation/disaggregation techniques (Cho and Meyer [1999], Stewart and Wu [1992]), which aim for a given stochastic matrix at a fast computation of the stationary distribution by decomposing the state space into stochastic complements. Yet, due to non-reversibility the setup of stochastic complements is much more intricate than restriction. This is also the reason why for biomolecular simulations, the technique of stochastic complementation does not enable to set up restricted Markov operators for MCMC sampling on a continuous state space. However, for reversible problems, we can utilize aggregation/disaggregation techniques combined with restriction in Section 4 to solve eigenvector problems for discretized differential operators.

Note that, since we do not treat the embedding into a hierarchical annealing structure here, the coupling matrix **C** in Theorem 2, although it shares the same characteristics, is different from the one actually employed in the UC algorithm.

3.1 Trialanine Simulation

As an example how UC is employed in biomolecular simulations we consider trialanine, a small peptide composed of three alanine amino acid residues. Although the continuous state space Ω is high-dimensional, the structural and dynamical properties of trialanine are primarily determined by the two torsion angles Φ and Ψ, as shown in Fig. 1. We herein only illustrate the initial uncoupling step of UC, which starts with a high-temperature MCMC
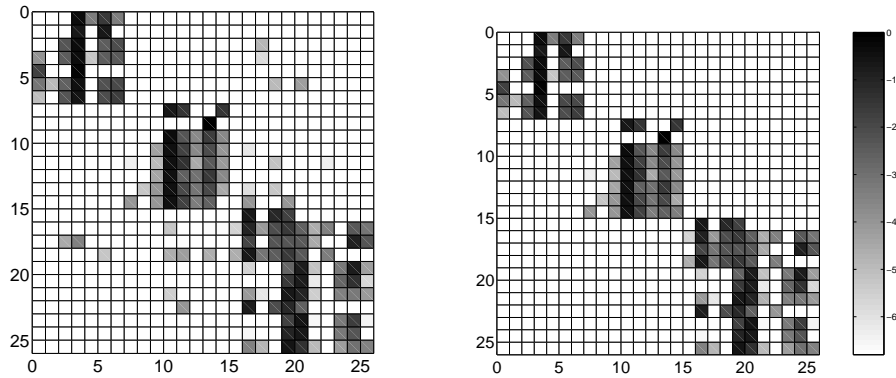
Fig. 2. Left: The permuted transition matrix $\mathbf{P}$ clearly has a block dominant structure. Right: In the resulting restricted matrix $\mathbf{R}$ all off-diagonal entries are set to zero. The intensity of the boxes is chosen due to the logarithmic scale on the far right.

simulation. More precisely, we used the Hybrid Monte Carlo method (Brass et al. [1993]) to sample at a temperature of 650 K and stored the torsion angles for each simulation step. Discretization of each torsion angle domain $\mathcal{D} = (-180°, 180°]$ into 7 equidistant intervals resulted in 26 non-empty boxes $(B_1, \ldots, B_{26})$ in $\mathcal{D}^2$, see Fig. 1. On these boxes we set up a transition matrix $\mathbf{P} = (p_{ij})$ (i.e., the discretized Markov operator), where transition probabilities p_{ij} are obtained by counting the number of transitions between boxes B_i and B_j during simulation.

The first eigenvalues of the resulting (26×26)-transition matrix are

j	1	2	3	4	5
$\lambda_j(\mathbf{P})$	1	0.9952	0.9941	0.5692	0.1425

$\cdots$,

indicating a slow mixing Markov chain with three metastable sets. For the identification of these metastable sets we used the previously mentioned spectral approach (Deuflhard et al. [2000], Weber [2003], Deuflhard and Weber [2003]). In Fig. 1 the identified metastable sets are indicated by different line styles. A corresponding permutation of the transition matrix confirms the computation in that it reveals an obvious block dominant structure, see Fig. 2. Calculating the subdominant eigenvalues of the restrictions $\mathbf{R}_{ii}$ for $i = 1, 2, 3$ results in

$\lambda_2(\mathbf{R}_{11})$	$\lambda_2(\mathbf{R}_{22})$	$\lambda_2(\mathbf{R}_{33})$
0.1376	0.1482	0.5855

,

which shows that this metastable decomposition in fact leads to three fast mixing restricted Markov chains.

4 UC and Domain Decomposition

In this section, we will pursue the goal of connecting the UC scheme for computing canonical distributions with the world of PDEs. In order to do so, we will first introduce a reversible Markov process in continuous time and discuss its biomolecular background and then will focus on its connection to domain decomposition for PDEs.

Molecular Dynamics

In order to specify what we herein consider as molecular dynamics let q denote the position and p the momenta of a single molecular system consisting of N atoms in state $x = (q, p) \in \mathbb{R}^{3N} \times \mathbb{R}^{3N}$. $V = V(q) : \mathbb{R}^{3N} \to \mathbb{R}$ describes the potential energy, which we assume to be differentiable. The statistics of molecular systems in state space is given by the well-known *canonical density* $f = \frac{1}{Z} \exp(-\beta V(x))$. Let μ be the measure induced by f (for a more detailed description of the biophysical background, cf. Schütte and Huisinga [2003]).

There exist several models of molecular dynamics; we will focus on the *Smoluchowski equation*

$$\dot{q} = -\frac{1}{\gamma} \nabla_q V(q) + \frac{\sigma}{\gamma} \dot{W}, \tag{18}$$

which is an approximation of the well-known *Langevin equation* in case of high friction γ, where $\dot{W}$ is given by a standard $3N$-dimensional Brownian motion W. The continuous time Markov process $(X_t)_{t \geq 0}$ defined by (18) leaves the canonical measure μ invariant and is reversible. Furthermore, $(X_t)_{t \geq 0}$ defines an absolutely continuous *stochastic transition function $p(t, x, y)$* that describes the probability that the process if started in x at time $t = 0$ is being found in y at time t (for details see Schütte and Huisinga [2003]).

Markov Operator and Generator

The family of Markov operators $(P_t)_{t \geq 0}$ associated with $(X_t)_{t \geq 0}$ is defined analogously to (6) for every $t > 0$ with $k(x, y) = p(t, x, y)$ and $r \equiv 0$. The family $(P_t)_{t \geq 0}$ forms a strongly continuous semigroup such that the *infinitesimal generator*

$$\mathcal{A}y = \lim_{t \to 0} \frac{P_t y - y}{t}$$

is defined and acts on the domain $\text{dom}(\mathcal{A}) = \{y \in Y : \lim_{t \to 0} (P_t y - y)/t \text{ exists}\}$. In the following we will simply express the relation between P_t and $\mathcal{A}$ by $P_t = \exp(t\mathcal{A})$ (for details see Huisinga et al. [to appear 2004]).

The reversibility of the underlying Markov process $(X_t)_{t \geq 0}$ has the additional implication that all P_t and the generator $\mathcal{A}$ are *self-adjoint* operators on the Hilbert space $L^2(\mu)$ equipped with the scalar product $\langle u, v \rangle_\mu = \int u(x) \bar{v}(x) \mu(dx)$, cf. Schütte and Huisinga [2003].

More insight into the process and into the form of the generator is available if we consider the evolution of a function under the dynamics given by $(X_t)_{t\geq 0}$. This evolution is governed by the *Fokker-Planck equation*

$$\partial_t u = \left(\underbrace{\frac{\sigma^2}{2\gamma^2}\Delta_q + \frac{1}{\gamma}\nabla_q V(q) \cdot \nabla_q + \frac{1}{\gamma}\Delta_q V(q)}_{\mathcal{A}} \right) u \qquad (19)$$

on some suitable subspace of $L^1(\mathrm{d}q)$. In this formula $\mathcal{A}$ is the infinitesimal generator of the semigroup $P_t \colon L^1(\mathrm{d}q) \to L^1(\mathrm{d}q)$. That is, for twice differentiable functions u the generator $\mathcal{A}$ is the elliptic partial differential operator given by the RHS of (19).

Finally, the invariance of canonical density μ under the process $(X_t)_{t\geq 0}$ gives us

$$P_t f = f \qquad \text{and} \qquad \mathcal{A}f = 0. \qquad (20)$$

Therefore, the computation of the canonical density f can be reduced to the computation of the dominant eigenvector of the generator $\mathcal{A}$. If the potential V satisfies some growth and regularity conditions then, both, the spectra of P_t and $\mathcal{A}$ are discrete in $L^2(\mu)$ and satisfy $\sigma(P_t) = \exp(t\sigma(\mathcal{A}))$. Then, metastability can be discussed via the dominant eigenvalues of $\mathcal{A}$ (i.e., those close to the largest one $\lambda_1 = 0$).

Discretization and Uncoupling-Coupling

Discretizing the operator P_t in position and time, one obtains a Markov chain with transition matrix $\mathbf{P}(t)$. Due to the reversibility, one can apply the UC scheme which was presented in section 3. Yet another way is to work with the generator $\mathcal{A}$ instead. In the remaining part of this section, it will be shown to what extent the operator $\mathcal{A}$ is related to P_t and how one can extract the required information from $\mathcal{A}$ in almost the same manner as from P_t. For simplicity in what follows we consider a bounded system: The potential V is smooth and takes infinite values at the boundary and outside of a compact domain Ω with sufficiently smooth boundary $\partial\Omega$.

The *discretization* of $\mathcal{A}$, acting on $L^1(\mathrm{d}q)$, by means of Finite Elements or Finite Differences is well known and produces a large sparse matrix $\mathbf{A}$ with row sum 0 at the interior nodes. Due to the condition on V, we have Dirichlet boundary conditions equal 0 on $\partial\Omega$ and thus at the boundary nodes. If we assume that there is a Perron cluster of eigenvalues of $\mathcal{A}$ close to its largest eigenvalue $\lambda = 0$ then we have the same spectral property for $\mathbf{A}$ (if the discretization grid is fine enough). If the number of nodes is large efficient numerical solution of the eigenvalue problem for $\mathbf{A}$ will therefore have to apply advanced numerical techniques like subspace oriented multigrid solvers (Friese et al. [1999]), appropriate domain decomposition techniques, or suitably preconditioned linear algebra solvers. Alternatively, we can exploit

that the connection between the discretized operators $\mathbf{A}$ and $(\mathbf{P}(t))$ is rather close (Forster [2003]). This allows us to transfer the idea of UC to the generator $\mathbf{A}$. More precisely: decompose the state space into the metastable subsets, restrict $\mathbf{A}$ to these subsets, solve the eigenvector problem locally, and couple the local solutions with weighting factors obtained by some coupling matrix.

The *restriction step* is based on the restricted discretized generator (cp. (15))

$$\mathbf{A}^{\text{rest}} = \begin{pmatrix} \mathbf{A}^{\text{rest}}_{(11)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{\text{rest}}_{(22)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{A}^{\text{rest}}_{(nn)} \end{pmatrix} \tag{21}$$

with $\mathbf{A}^{\text{rest}}_{(ii)} = \mathbf{A}^{(ii)} + \operatorname{diag}(\sum_{j,j\neq i} \mathbf{A}^{(ij)}\mathbf{e}_j)$ and blocks $\mathbf{A}^{(ii)}$ of $\mathbf{A}$ according to a decomposition in metastable subsets S_i. This Neumann-like boundary condition is imposed directly on the discretization level. Neumann boundary conditions make sense since the invariant measure should have local minima at the boundaries between the metastable sets. However, these boundary conditions have no simple continuous analogue. Nonetheless, $\mathbf{A}^{\text{rest}}$ is the requested object, since the local solutions of $\mathbf{A}^{\text{rest}}_{(ii)}\mu^{(i)} = 0$ form—apart from the weighting factors (ξ_i)—the canonical density f. That follows from the fact that the detailed balance condition $\mu_k \mathbf{A}_{kl} = \mu_l \mathbf{A}_{lk}$ (which holds since $\mathcal{A}$ is self-adjoint) implies $\mu^T \mathbf{A}^{\text{rest}} = \mu^T \mathbf{A} = 0$ as follows:

$$(\mu^T \mathbf{A}^{\text{rest}})_j = \mu_j \mathbf{A}^{\text{rest}}_{jj} + \sum_{k\notin S_i} \mu_k \mathbf{A}^{\text{rest}}_{kj} + \sum_{k\in S_i, k\neq j} \mu_k \mathbf{A}^{\text{rest}}_{kj}$$

$$= \mu_j \left(\mathbf{A}_{jj} + \sum_{k\notin S_i} \mathbf{A}_{jk} \right) + \sum_{k\in S_i, k\neq j} \mu_k \mathbf{A}_{kj}$$

$$= \mu_j \mathbf{A}_{jj} + \sum_{k\notin S_i} \mu_k \mathbf{A}_{kj} + \sum_{k\in S_i, k\neq j} \mu_k \mathbf{A}_{kj} = (\mu^T \mathbf{A})_j,$$

where j is a node in S_i and furthermore μ_k and $\mathbf{A}_{kj}$ denote the entries of the large vector $\mu = (\mu^{(1)}, \ldots, \mu^{(n)})$ and the matrix $\mathbf{A}$, respectively. The restricted operator $\mathbf{A}^{\text{rest}}$ is not the generator of $\mathbf{P}(t)^{\text{rest}}$, since $(\mathbf{P}(t)^{\text{rest}})_{t\geq 0}$ does no longer form a semigroup; for the same reason, in general it is $(e^{t\mathbf{A}})^{\text{rest}} \neq e^{t\mathbf{A}^{\text{rest}}}$.

In the *coupling step* the weighting factors can be obtained (Forster [2003]) by the *coupling matrix* $\mathbf{C} = (c_{ij})$ with entries

$$c_{ij} = \langle \mu_i, \mathbf{A}^T \mu_j \rangle_2.$$

The matrix $\mathbf{C}$ arises from a Galerkin discretization of $\mathcal{A}$ on the ansatz space $\mathcal{V} = \operatorname{span}\{\mu_1, \mu_2, \ldots, \mu_n\}$ and inherits the structure of $\mathcal{A}$. The factors $\boldsymbol{\xi} = (\xi_i)$ are the solution of the low-dimensional equation $\mathbf{C}\boldsymbol{\xi} = 0$. Even more

can be achieved: under certain conditions, it is also possible to generate the eigenvectors $v_2, \ldots, v_n$ of the other dominant eigenvalues $\lambda_2, \ldots, \lambda_n$ close to 1 by means of (μ_i) and $\mathbf{C}$. More precisely: the solutions $\nu^{(i)}$ of $(\mathbf{CD})\nu^{(i)} = \hat{\lambda}_i \nu^{(i)}$ with $\mathbf{D} = \mathrm{diag}(\xi_1, \ldots, \xi_n)$ allow to define approximations $\sum_k \nu_k^{(i)} \mu_k$ for the eigenvectors v_i. For a detailed description see Forster [2003].

The efficiency of the entire approach critically depends on the underlying decomposition of state space into metastable subsets. If the dynamics given by (18) is rapidly mixing within each of these metastable subsets then the second eigenvalues of all the diagonal blocks $\mathbf{A}_{(ii)}^{\mathrm{rest}}$ of the discretized restricted generator matrix $\mathbf{A}^{\mathrm{rest}}$ will be separated from the largest eigenvalues $\lambda = 0$ by some significant gap such that iterative eigenvalue solvers can be used to compute the eigenvectors μ_i to $\lambda = 0$ for all blocks $\mathbf{A}_{(ii)}^{\mathrm{rest}}$ efficiently. Thus, in order to construct a fully efficient algorithm one has to integrate annealing strategies and grid refinement into some carefully controlled hierarchical approach. This is still under investigation.

References

A. Brass, B. J. Pendleton, Y. Chen, and B. Robson. Hybrid Monte Carlo simulations theory and initial comparison with molecular dynamics. *Biopolymers*, 33:1307–1315, 1993.

P. Brémaud. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues.* Springer, New York, 1999.

G. E. Cho and C. D. Meyer. Aggregation/disaggregation methods for nearly uncoupled Markov chains. Technical Report NCSU #041600-0400, North Carolina State University, November 1999.

P. Deuflhard, W. Huisinga, A. Fischer, and C. Schütte. Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains. *Lin. Alg. Appl.*, 315:39–59, 2000.

P. Deuflhard and M. Weber. Robust Perron cluster analysis in conformation dynamics. ZIB-Report 03-19, Konrad-Zuse-Zentrum, Berlin, 2003.

D. M. Ferguson, J. I. Siepmann, and D. G. Truhlar, editors. *Monte Carlo Methods in Chemical Physics*, volume 105 of *Advances in Chemical Physics*. Wiley, New York, 1999.

A. Fischer. *An Uncoupling-Coupling Method for Markov Chain Monte Carlo Simulations with an Application to Biomolecules.* PhD thesis, Freie Universität Berlin, 2003.

A. Fischer, C. Schütte, P. Deuflhard, and F. Cordes. Hierarchical uncoupling-coupling of metastable conformations. In T. Schlick and H. H. Gan, editors, *Computational Methods for Macromolecules: Challenges and Applications, Proceedings of the 3rd International Workshop on Algorithms for Macromolecular Modeling, New York, Oct. 12–14, 2000*, volume 24 of *Lecture Notes in Computational Science and Engineering*, Berlin, 2002. Springer.

R. Forster. Ein Algorithmus zur Berechnung invarianter Dichten in metastabilen Systemen. Diploma thesis, Freie Universität Berlin, 2003.

T. Friese, P. Deuflhard, and F. Schmidt. A multigrid method for the complex Helmholtz eigenvalue problem. In C.-H. Lai, P. E. Bjørstad, M. Cross, and O. B. Widlund, editors, *Domain Decomposition Methods in Sciences and Engineering*, DDM-org, pages 18–26, New York, 1999.

W. Huisinga, S. Meyn, and C. Schütte. Phase transitions and metastability in Markovian and molecular systems. *Ann. Appl. Probab.*, to appear 2004.

J. S. Liu. *Monte Carlo Strategies in Scientific Computing.* Springer, New York, 2001.

E. Meerbach, A. Fischer, and C. Schütte. Eigenvalue bounds on restrictions of reversible nearly uncoupled Markov chains. Preprint, 2003.

C. D. Meyer. Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems. *SIAM Rev.*, 31:240–272, 1989.

S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability.* Springer, Berlin, 1993.

C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard. A direct approach to conformational dynamics based on hybrid Monte Carlo. *J. Comput. Phys.*, 151:146–168, 1999.

C. Schütte and W. Huisinga. Biomolecular conformations can be identified as metastable sets of molecular dynamics. In P. G. Ciaret and J.-L. Lions, editors, *Handbook of Numerical Analysis*, volume Computational Chemistry. North-Holland, 2003.

C. Schütte, W. Huisinga, and P. Deuflhard. Transfer operator approach to conformational dynamics in biomolecular systems. In B. Fiedler, editor, *Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems.* Springer, 2001.

W. J. Stewart and W. Wu. Numerical experiments with iteration and aggregation for Markov Chains. *ORSA Journal on Computing*, 4(3):336–350, 1992.

L. Tierney. Markov chains for exploring posterior distributions (with discussion). *Ann. Statist.*, 22:1701–1762, 1994.

M. Weber. Improved Perron cluster analysis. ZIB-Report 03-04, Konrad-Zuse-Zentrum, Berlin, 2003.