Eigenvalue Bounds on Restrictions of Reversible Nearly Uncoupled Markov Chains ⋆

Eike Meerbach * Christof Schütte Alexander Fischer

Abstract

In this paper we analyze decompositions of reversible nearly uncoupled Markov chains into rapidly mixing subchains. We state upper bounds on the 2nd eigenvalue for restriction and stochastic complementation chains of reversible Markov chains, as well as a relation between them. We illustrate the obtained bounds analytically for bunkbed graphs, and furthermore apply them to restricted Markov chains that arise when analyzing conformation dynamics of a small biomolecule.

AMS classification: 15A18; 15A51; 60J10; 60J22

Keywords: Biomolecules; Conformation dynamics; Markov chains; Rapid mixing; Stationary distribution; Stochastic complementation; Uncoupling-coupling

1 Introduction

Markov chains are a popular tool to model the behavior of complex systems like computer networks or biomolecules, and form also the basis of Markov chain Monte Carlo methods like the Metropolis-Hastings sampler.

In many applications so-called nearly uncoupled Markov chains arise that are rapidly mixing within certain parts of the state space while transitions be-

⋆ Supported by the DFG research center "Mathematics for key technologies" (FZT 86) in Berlin.

* Corresponding author. Institut für Mathematik II, Freie Universität Berlin, Arnimallee 2–6, 14195 Berlin, Germany. Tel.: +49-30-838-56965; fax: +49-30-838-75412.

Email addresses: meerbach@math.fu-berlin.de (Eike Meerbach), schuette@math.fu-berlin.de (Christof Schütte), fischer@math.fu-berlin.de (Alexander Fischer).

tween these parts occur rarely. Such behavior is well-known in biomolecular simulations, which, in a broader context than Markov chains, is referred to as metastable or conformation dynamics. The rare transitions between conformations (or metastable sets) pose a big problem, since simulations often get trapped within a conformation. Molecular simulations by Markov chains are typically carried out in a continuous state space, yet by analyzing the outcome, discretization inherits a nearly uncoupled structure to a finite state space.

For finite state spaces a nearly uncoupled Markov chain is often said to consist of weakly coupled subsets containing strongly coupled states. A subset of strongly coupled states then corresponds to a metastable set in the above description of metastable dynamics.

An analytical as well as an algorithmic approach to take advantage of the nearly uncoupled structure is uncoupling, where one decomposes the state space into metastable sets and then defines uncoupled Markov chains on each of its metastable sets. These newly defined Markov chains resemble in many aspects the original chain. For the continuous state space of a biomolecule, uncoupling refers to the process of defining rapidly mixing Markov chains on each conformation.

Uncoupling is often complemented by a coupling step. In an uncoupling-coupling technique restricted (rapidly mixing) Markov chains are coupled together by a $(k \times k)$-coupling matrix. A main characteristic of an uncoupling-coupling technique is that subchains and coupling matrix together still contain all the information to extract the stationary distribution of the original Markov chain.

For general non-reversible Markov chains the concept of uncoupling-coupling has been worked out by Meyer by means of stochastic complementation [**?**,**?**,**?**,**?**]. Stochastic complements are a natural way to define uncoupled Markov chains, which inherit most of the structure of the original chain. Yet, for algorithmic purposes the main drawback of stochastic complements is that they become computationally expensive for larger state spaces.

For reversible Markov chains stochastic complements can be replaced by restriction chains. Restriction chains, in contrast to stochastic complements, do not necessarily inherit irreducibility. Nevertheless, restriction can be a valuable theoretical tool and has been used, e.g., for studying convergence rate analysis [**?**] or improving log-Sobolev inequalities [**?**].

What makes uncoupling by restriction so attractive, is that it can be applied with ease to state spaces of any size, either discrete or continuous. Moreover, if the nearly uncoupled Markov chain under consideration is associated to a Metropolis-Hastings sampler, restriction gives rise to restricted sampling.

Build upon this observation and techniques from conformation dynamics of biomolecules [**?**,**?**,**?**], an uncoupling-coupling technique has been developed by one of the authors, which serves as a generalized Markov chain Monte Carlo method [**?**,**?**].

The efficiency of applying an uncoupling-coupling technique to a nearly uncoupled Markov chain essentially depends on the second eigenvalues of the resulting subchains being bounded far away from 1. The main theorems presented in this paper provide bounds on this spectral gap. The theorems may help to better understand the structure and similarities of uncoupled chains to the original Markov chain and thus shed some light on the usefulness and applicability of uncoupling-coupling techniques. In practice, the eigenvalue bounds are almost sharp in some situations, but fail on others to be close to the actual spectral gap. We discuss both cases in Sect. 4.

Overview. We start Sect. 2 with introducing some basic facts and notation concerning Markov chains in Sect. 2.1, and then focus on nearly uncoupled Markov chains in Sect. 2.2. Next we outline two different uncoupling-coupling schemes, namely stochastic complementation in Sect. 2.3 and Markov chain restriction in Sect. 2.4. In Sect. 3 we present the main theorems on eigenvalue bounds, which are then illustrated in Sect. 4 by examples from graph theory and biomolecular conformation dynamics.

2 Uncoupling-Coupling Schemes

The underlying idea of uncoupling-coupling techniques is to decompose the state space S of a given Markov chain $\mathbf{P}$ into disjoint subsets $S_1, S_2, \ldots, S_k$, and define Markov chains on them whose behavior is related to that of the original Markov chain. The coupling step provides a way to extract information about the global behavior of the original Markov chain, which is achieved by means of a $(k \times k)$-coupling matrix. The coupling matrix together with the k subchains can be regarded as a reduced description of $\mathbf{P}$ that still contains the full information about $\mathbf{P}$'s stationary distribution.

Uncoupling-coupling techniques are encountered in such diverse tasks as improved convergence rate analysis [**?**], the construction of rapidly mixing chains for extended Markov chain Monte Carlo methods [**?**,**?**], or in an approximate way in aggregation-disaggregation techniques [**?**].

2.1 Preliminaries

Throughout this paper we consider finite homogeneous Markov chains (for a formal treatment of terms and well-known facts about Markov chains we refer to [**?**]). Let $\mathbf{P}$ be a stochastic $(n \times n)$-transition matrix associated with a Markov chain over a finite set $S = \{s_1, s_2, \ldots, s_n\}$ of states s_i. The spectral structure of stochastic matrices is characterized in the Perron-Frobenius theory. If $\mathbf{P}$ is stochastic, $\lambda = 1$ is an eigenvalue and its spectrum is contained in the unit circle. If in addition $\mathbf{P}$ is irreducible, the Perron root $\lambda = 1$ is simple and there is a strictly positive left-eigenvector to it. By normalization this yields into a unique stationary distribution $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_n)$, that satisfies

$$\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}, \quad \boldsymbol{\pi} > 0, \quad \text{and} \quad \boldsymbol{\pi}\mathbf{e} = 1,$$

where $\mathbf{e} := (1, 1, \ldots, 1)^T$ is a vector of size n.

The pair $(\mathbf{P}, \boldsymbol{\pi})$ is said to be reversible, if the detailed balance condition

$$\pi_i p_{ij} = \pi_j p_{ji} \tag{1}$$

holds for all $1 \leq i, j \leq n$. A probability vector $\boldsymbol{\pi}$ that satisfies (1) is always a stationary distribution of $\mathbf{P}$, whereas the reverse need not to be true. If (1) holds $\mathbf{P}$ itself as well as its associated Markov chain is called reversible.

A reversible stochastic matrix $\mathbf{P}$ is similar to a symmetric one. More precisely, if $(\mathbf{P}, \boldsymbol{\pi})$ is reversible then

$$\mathbf{P}^{(sym)} := \mathbf{D}\mathbf{P}\mathbf{D}^{-1}, \quad \text{with} \quad \mathbf{D} := \begin{pmatrix} \sqrt{\pi_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\pi_n} \end{pmatrix}, \tag{2}$$

is symmetric. Therefore, all eigenvalues of a reversible stochastic matrix are real, located in the interval $[-1, 1]$, and $\lambda = 1$ is a simple eigenvalue.

2.2 Nearly Uncoupled Markov Chains

An irreducible Markov chain $\mathbf{P}$ is said to be nearly uncoupled, if there exists a decomposition of the state space S into k disjoint subsets $S_1, S_2, \ldots S_k$, such that the subsets are weakly coupled among each other, whereas the states within each subset are strongly coupled. In other words, a realization of $\mathbf{P}$ is slowly mixing on S, but rapidly mixing within each S_i. The subsets S_i are then called metastable w.r.t. $\mathbf{P}$. For a nearly uncoupled Markov chain there exists a permutation of the states s_i such that the transition matrix can be

written in block form

$$\mathbf{P} = \widetilde{\mathbf{P}} + \mathbf{E} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{E}_{12} & \cdots & \mathbf{E}_{1k} \\ \mathbf{E}_{21} & \mathbf{P}_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{E}_{k-1,k} \\ \mathbf{E}_{k1} & \cdots & \mathbf{E}_{k,k-1} & \mathbf{P}_{kk} \end{pmatrix}, \tag{3}$$

where entries in $\mathbf{E}$ are small, and each $\mathbf{P}_{ii}$ possesses good mixing properties. A suitable choice for our purpose to measure the smallness of $\mathbf{E}$ is the infinity norm $\|\mathbf{E}\|_\infty$, which is equivalent to the maximal absolute row sum.

For $\|\mathbf{E}\|_\infty = 0$ the Markov chain is uncoupled into k sets and thus becomes reducible. If we assume all diagonal blocks $\mathbf{P}_{ii}$ being irreducible, $\mathbf{P}$ possesses, according to Frobenius-Perron theory, a k-fold dominant eigenvalue $\lambda = 1$. Furthermore, with each $\mathbf{P}_{ii}$ being rapidly mixing, all other eigenvalues are bounded away from 1 resulting in a large spectral gap.

Now assume that $\|\mathbf{E}\|_\infty = \epsilon$ is small and $\mathbf{P}$ is irreducible. Regarding ϵ as a perturbation parameter, continuity of eigenvalues suggests that the spectra of a regular nearly uncoupled Markov chain must have in addition to the dominant eigenvalue $\lambda = 1$ further $k - 1$ eigenvalues close to 1. Vice versa, such a cluster of k eigenvalues that is separated to the rest of the spectrum by a spectral gap, indicates a nearly uncoupled Markov chain with k metastable sets. A perturbation analysis of this behavior can be found in [**?**,**?**].

It is well known that the subdominant eigenvalue of a regular Markov chain is an indicator of its mixing properties. Thus, a subdominant eigenvalue close to 1 in a nearly uncoupled Markov chain indicates slow mixing, which, e.g., often leads for Metropolis-Hastings algorithms to poor convergence for most expectation values. At this point, uncoupling according to the partition in (3) becomes interesting, in the hope that the uncoupled chains will posses rapid mixing properties.

In practice, the question of whether a given $\mathbf{P}$ is nearly uncoupled or not and how to permute it into block-diagonal form may not been known in advance. For this nontrivial algorithmic task, spectral approaches that first identify a spectral gap (which is related to the number of metastable sets) and then exploit the structure of dominant eigenvectors to identify a suitable permutation has been worked out [**?**,**?**,**?**]. In the following we assume a nearly uncoupled Markov chain to be in block-diagonal form as in (3), whether this being the natural order or a permutation of states after metastable sets has been identified.

Uncoupling by means of stochastic complements can be applied to the whole class of irreducible stochastic matrices, not necessarily restricted to reversible ones. The following definition is proposed by Meyer in [**?**]:

Definition 1 *Let* $\mathbf{P}$ *be an irreducible stochastic* $(n \times n)$*-matrix with a partition*

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \cdots & \mathbf{P}_{1k} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \cdots & \mathbf{P}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{k1} & \mathbf{P}_{k2} & \cdots & \mathbf{P}_{kk} \end{pmatrix} \tag{4}$$

in which all diagonal blocks are square. Denote by $\mathbf{P}_{i'}$, $1 \leq i \leq k$, *the principal block submatrix obtained by deleting the* i*-th block row and the* i*-th block column in* $\mathbf{P}$*. Furthermore let* $\mathbf{P}_{i*}$ *and* $\mathbf{P}_{*i}$ *be the* i*-th block row and* i*-th block column, respectively, in which* $\mathbf{P}_{ii}$ *is deleted, i.e.,*

$$\mathbf{P}_{i*} = (\mathbf{P}_{i1}\ \mathbf{P}_{i2}\ \cdots\ \mathbf{P}_{i,i-1}\ \mathbf{P}_{i,i+1} \cdots \mathbf{P}_{ik})$$

and

$$\mathbf{P}_{*i} = \begin{pmatrix} \mathbf{P}_{1i} \\ \vdots \\ \mathbf{P}_{i-1,i} \\ \mathbf{P}_{i+1,i} \\ \vdots \\ \mathbf{P}_{ki} \end{pmatrix}.$$

Then the inverse of $(\mathbf{I} - \mathbf{P}_{i'})$ *does exist and the matrix*

$$\mathbf{S}_{ii} = \mathbf{P}_{ii} + \mathbf{P}_{i*}(\mathbf{I} - \mathbf{P}_{i'})^{-1}\mathbf{P}_{*i} \tag{5}$$

is called the stochastic complement *of* $\mathbf{P}_{ii}$.

Stochastic complementation provides a neat interpretation: Let $\mathbf{P}$ be partitioned according to a given partition of the state space S, say $S = S_1 \cup S_2 \cup \cdots \cup S_k$. Then it can be shown that stochastic complements of an irreducible stochastic matrix are themselves stochastic and irreducible ([**?**], Theorem 2.3), and the matrix $\mathbf{S}$ defined by (5),

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{S}_{kk} \end{pmatrix},$$

is a (reducible) stochastic $(n \times n)$-matrix. If s_{ij} is an entry within a diagonal block, say S_{ll}, then s_{ij} is determined by the probabilities in $\mathbf{P}$, namely it is the

sum of the one-step transition probability p_{ij} and the probability to leave S_l from i and reenter it in j. Thus, transition probabilities in $\mathbf{S}_{ll}$ can be obtained by keeping track of a realization of the Markov chain associated with $\mathbf{P}$ and masking out every step which is not in S_l.

The following theorem describes the coupling step (see again [**?**] for more details):

Theorem 2 *Let* $\mathbf{P}$ *be an irreducible stochastic matrix partitioned as in Definition 1 with the (unique) stationary distribution* $\boldsymbol{\pi}$ *partitioned accordingly:*

$$\boldsymbol{\pi} = (\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)}, \ldots, \boldsymbol{\pi}^{(k)}).$$

If $\mathbf{s}^{(i)}$ *is the stationary distribution of the stochastic complement* $\mathbf{S}_{ii}$*, then*

$$\mathbf{s}^{(i)} = \frac{\boldsymbol{\pi}^{(i)}}{\boldsymbol{\pi}^{(i)}\mathbf{e}},$$

which is equivalent to

$$\boldsymbol{\pi} = (\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)}, \ldots, \boldsymbol{\pi}^{(k)}) = (\xi_1\mathbf{s}^{(1)}, \xi_2\mathbf{s}^{(2)}, \ldots, \xi_k\mathbf{s}^{(k)}),$$

with $\xi_i := \sum_h \pi_h^{(i)}$*. The scalars* ξ_i *are called the* coupling factors*. The* coupling vector

$$\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_k)$$

is the (unique) stationary distribution of the irreducible and stochastic $(k \times k)$*-*coupling matrix $\mathbf{C} = (c_{ij})$*, whose entries are defined by*

$$c_{ij} = \mathbf{s}^{(i)}\mathbf{P}_{ij}\mathbf{e}. \tag{6}$$

Moreover,

$$\|\mathbf{P} - \mathbf{S}\|_\infty = 2 \max_i \|\mathbf{P}_{i*}\|_\infty$$

holds.

If one is interested in computing the stationary distribution of $\mathbf{P}$, Theorem 2 says that it is sufficient to compute for each $i = 1, \ldots, k$ the restricted stationary distribution $\mathbf{s}^{(i)}$ from $\mathbf{S}_{ii}$ together with its coupling factor ξ_i. Yet, computing $\mathbf{S}_{ii}$ includes inversion of $(\mathbf{I} - \mathbf{P}_{i'})$, which in practice often makes such an approach expensive if not impossible for very large matrices (see [**?**] for ways to reduce the computational cost). This problem is circumvented in so-called inexact aggregation-disaggregation techniques [**?**,**?**] by approximating stochastic complements for the price of computing an approximation of $\boldsymbol{\pi}$. From this point of view, the uncoupling-coupling structure of Theorem 2 is considered as an exact aggregation-disaggregation technique.

2.4 Restriction

For reversible Markov chains stochastic complements can be replaced by restriction chains, which are way easier to compute but still share much of the characteristics of stochastic complements.

Definition 3 *Let* $\mathbf{P}$ *be a stochastic matrix, not necessarily irreducible, partitioned as in (4), and let* $S = S_1 \cup S_2 \cup \cdots \cup S_k$ *be the associated partition of the state space. Then, for each* $i = 1, \ldots, k$*,*

$$\mathbf{R}_{ii} = \mathbf{P}_{ii} + \operatorname{diag}(\mathbf{e}_i - \mathbf{P}_{ii}\mathbf{e}_i)$$

is called the restriction of $\mathbf{P}$ *to the subset* S_i*, where* $\mathbf{e}_i = (1, 1, \ldots, 1)$ *is a vector of size* $|S_i|$*, and*

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{R}_{kk} \end{pmatrix} \tag{7}$$

is called the restriction matrix.

In other words, $\mathbf{R}$ 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}.

The following theorem summarizes in analogy to Theorem 2 some facts about $\mathbf{R}$:

Theorem 4 *Let* $\mathbf{P}$ *be an irreducible and reversible stochastic matrix partitioned as in Definition 1. Furthermore, let all* $\mathbf{P}_{ii}$ *be irreducible (substochastic) matrices. Then,*

(a) all $\mathbf{R}_{ii}$ *are irreducible,*
(b) $\mathbf{R}$ *is stochastic with a* k*-fold dominant eigenvalue* 1*,*
(c) and for each $i = 1, \ldots, k$ *the unique stationary distribution* $\mathbf{r}^{(i)}$ *of the restriction* $\mathbf{R}_{ii}$ *is identical to* $\mathbf{s}^{(i)}$ *of* $\mathbf{S}_{ii}$ *from Theorem 2.*

Proof. Irreducibility is inherited from $\mathbf{P}_{ii}$ to $\mathbf{R}_{ii}$, so (a) holds, and since $\mathbf{R}$ is uncoupled into k blocks, (b) follows. For (c), note that if $r_{ml} \neq p_{ml}$ then, by construction of $\mathbf{R}$, we have $l = m$ or $r_{ml} = r_{lm} = 0$. Therefore the detailed balance condition for $\mathbf{P}$ still holds for $\mathbf{R}$ and $\mathbf{r}^{(i)}$ is obtained by normalization of $\boldsymbol{\pi}^{(i)}$, which is the way $\mathbf{s}^{(i)}$ is defined in Theorem 2. Furthermore $\mathbf{r}^{(i)}$ is unique since $\mathbf{R}_{ii}$ is irreducible. □

Since $\mathbf{r}^{(i)} \equiv \mathbf{s}^{(i)}$, this means that the coupling procedure described in Theorem 2 also applies to restriction chains. Given the coupling vector $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_k)$ (e.g., by means of the coupling matrix $\mathbf{C}$ defined in (6)) we can write the stationary distribution of $\mathbf{P}$ as

$$\boldsymbol{\pi} = (\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)}, \ldots, \boldsymbol{\pi}^{(k)}) = (\xi_1 \mathbf{r}^{(1)}, \xi_2 \mathbf{r}^{(2)}, \ldots, \xi_k \mathbf{r}^{(k)}).$$

Hereby, the straightforward setup of $\mathbf{R}_{ii}$ provides a convenient way to compute the restricted stationary distribution $\mathbf{r}^{(i)}$. Off-diagonal blocks of $\mathbf{P}$ are only needed to compute entries of the coupling matrix $\mathbf{C}$.

Restriction chains arise naturally in the Metropolis-Hastings sampler when the underlying reversible Markov chain is restricted to some region of the state space. For a realization of the underlying Markov chain restriction means to reject all proposal steps that exit from the subset S_i in which the process was started. This small modification directly implements a sampler for the restriction $\mathbf{R}_{ii}$.

This becomes of special interest, if the $\mathbf{R}_{ii}$'s are restrictions of a nearly uncoupled Markov chain to its metastable (or strongly coupled) subsets. Then, the idea is that each $\mathbf{R}_{ii}$ is rapidly mixing, i.e., sampling from the $\mathbf{R}_{ii}$'s may be orders of magnitude faster than for $\mathbf{P}$.

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 a Metropolis-Hastings sampler into a hierarchical annealing structure. For a detailed presentation of this approach we refer to the Uncoupling-coupling Monte Carlo method presented in [**?**,**?**]. In Sect. 4, we analyze for a small biomolecule a nearly uncoupled Monte Carlo Markov chain and illustrate hereby the initial step of Uncoupling-coupling Monte Carlo.

3 Bounds on Subdominant Eigenvalues

Uncoupling, either by stochastic complementation or restriction, is of special interest for a nearly uncoupled Markov chain. From now on we restrict our considerations to reversible matrices, which also allows us to order the real eigenvalues of a reversible stochastic $(n \times n)$-matrix by

$$1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq -1.$$

If necessary for clarification, we also use the notation $\lambda_i(\mathbf{P})$ to denote the i-th eigenvalue of $\mathbf{P}$.

The main goal for uncoupling a nearly uncoupled Markov chain $\mathbf{P}$ is to obtain rapidly mixing subchains $\mathbf{R}_{ii}$ or $\mathbf{S}_{ii}$, for $i = 1, \ldots, k$. Hereby, a reasonable criterion for rapid mixing is that the spectral gap between 1 and λ_2 of each subchain is large. For R and S this means that the spectral gap $1 - \lambda_{k+1}(\mathbf{R})$ and $1 - \lambda_{k+1}(\mathbf{S})$ has to be large, respectively.

For $\mathbf{P}$ being irreducible and nearly uncoupled with k loosely coupled components, $\lambda_k(\mathbf{P}) - \lambda_{k+1}(\mathbf{P})$ has a large spectral gap. By uncoupling, the eigenvalues become $\lambda_i = 1$ for $i = 1, \ldots, k$, and λ_{k+1} may also increase towards 1. Yet, for good metastable decompositions λ_{k+1} should still be bounded away from 1.

Theorem 2 already states that $\|\mathbf{P} - \mathbf{S}\|_\infty = 2 \max_i \|\mathbf{P}_{i*}\|_\infty$ holds. Obviously, the same equality holds for the restricted matrix $\mathbf{R}$, so that, if $\mathbf{P}$ is given in the form stated in (3), we have

$$\|\mathbf{P} - \mathbf{S}\|_\infty = \|\mathbf{P} - \mathbf{R}\|_\infty = 2\|\mathbf{E}\|_\infty.$$

This gives rise to the assumption that eigenvalues of $\mathbf{P}$, $\mathbf{R}$, and $\mathbf{S}$ are indeed close to each other for $\|\mathbf{E}\|_\infty$ being small. We will specify this relationship in the following, especially the impact of uncoupling on the subdominant eigenvalue λ_{k+1}.

Similarity between reversible and symmetric matrices enables us to use Weyls' inequalities ([**?**], III.2, S.62f) to relate subdominant eigenvalues of $\mathbf{R}$ and $\mathbf{S}$ to $\mathbf{P}$.

Theorem 5 (Weyls' inequalities) *Let $\mathbf{A}$ and $\mathbf{B}$ be symmetric $(n \times n)$-matrices with ordered eigenvalues $\lambda_1(\mathbf{A}) \geq \cdots \geq \lambda_n(\mathbf{A})$ and $\lambda_1(\mathbf{B}) \geq \cdots \geq \lambda_n(\mathbf{B})$, respectively. Then, for $j = 1, \ldots, n$,*

$$\lambda_j(\mathbf{A} + \mathbf{B}) \leq \lambda_i(\mathbf{A}) + \lambda_{j-i+1}(\mathbf{B}) \text{ for } i \leq j,$$
$$\lambda_j(\mathbf{A} + \mathbf{B}) \geq \lambda_i(\mathbf{A}) + \lambda_{j-i+n}(\mathbf{B}) \text{ for } i \geq j.$$

If we put $i = j$ in the above inequalities, we immediately obtain

Corollary 6 *For each $j = 1, 2, \ldots, n$,*

$$\lambda_j(\mathbf{A}) + \lambda_n(\mathbf{B}) \leq \lambda_j(\mathbf{A} + \mathbf{B}) \leq \lambda_j(\mathbf{A}) + \lambda_1(\mathbf{B})$$

holds.

This corollary enables us to state bounds on the eigenvalues of $\mathbf{R}$.

Theorem 7 *Let* $\mathbf{P}$ *be a reversible stochastic matrix partitioned according to (3) and* $\mathbf{R}$ *the restricted matrix, as defined in (7). Then*

$$\lambda_j(\mathbf{R}) \leq \lambda_j(\mathbf{P}) + 2\|\mathbf{E}\|_\infty \tag{8}$$

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

Proof. Denote by $\mathbf{E}^{(diag)}$ the diagonal matrix containing the i-th row sum of $\mathbf{E}$ in the i-th diagonal entry, so that

$$\mathbf{R} = \mathbf{P} + \mathbf{E}^{(diag)} - \mathbf{E}.$$

The matrices $\mathbf{P}$ and $\mathbf{R}$ are reversible and symmetrized by the same diagonal matrix $\mathbf{D}$, see (2) and Theorem 4. If we refer to symmetric matrices by the superscript (sym), we have

$$\begin{aligned}\mathbf{R}^{(sym)} &= \mathbf{D}(\mathbf{P} + \mathbf{E}^{(diag)} - \mathbf{E})\mathbf{D}^{-1} \\ &= \mathbf{D}\mathbf{P}\mathbf{D}^{-1} + \mathbf{D}\mathbf{E}^{(diag)}\mathbf{D}^{-1} - \mathbf{D}\mathbf{E}\mathbf{D}^{-1} \\ &= \mathbf{D}^{(sym)} + \mathbf{E}^{(diag)} - \mathbf{E}^{(sym)},\end{aligned}$$

considering that $\mathbf{D}\mathbf{E}^{(diag)}\mathbf{D}^{-1} = \mathbf{E}^{(diag)}$ and that $\mathbf{D}\mathbf{E}\mathbf{D}^{-1}$ must be symmetric to be consistent with the equation. Remind the fact that the spectral radius of a matrix is bounded by every induced matrix norm (like $\|\cdot\|_\infty$ is) and use Corollary 6 to get

$$\begin{aligned}\lambda_1(\mathbf{E}^{(diag)} - \mathbf{E}^{(sym)}) &\leq \lambda_1(\mathbf{E}^{(diag)}) + \lambda_1(-\mathbf{E}^{(sym)}) \\ &= \lambda_1(\mathbf{E}^{(diag)}) + \lambda_1(-\mathbf{E}) \\ &\leq \|\mathbf{E}^{(diag)}\|_\infty + \|-\mathbf{E}\|_\infty = 2\|\mathbf{E}\|_\infty.\end{aligned}$$

Therefore we have

$$\begin{aligned}\lambda_j(\mathbf{R}) &= \lambda_j(\mathbf{R}^{(sym)}) \\ &\leq \lambda_j(\mathbf{P}^{(sym)}) + \lambda_1(\mathbf{E}^{(diag)} - \mathbf{E}^{(sym)}) \\ &= \lambda_j(\mathbf{P}) + \lambda_1(\mathbf{E}^{(diag)} - \mathbf{E}^{(sym)}) \leq \lambda_j(\mathbf{P}) + 2\|\mathbf{E}\|_\infty.\end{aligned}$$

□

We have shown that the subdominant eigenvalue of any diagonal block in $\mathbf{R}$ will be smaller than the k-th eigenvalue of $\mathbf{P}$ plus twice the infinity norm of $\mathbf{E}$. Instead of proving an analogous result for the eigenvalues of $\mathbf{S}$, which would be straightforward, we show a relationship between the eigenvalues of $\mathbf{S}$ and $\mathbf{R}$ and deduce the inequality from there. As a tool we need the well known Geršgorins Theorem ([**?**], VIII.6.3, S. 244):

Theorem 8 (Geršgorins Theorem) *Let* $\mathbf{A}$ *be an* $(n \times n)$*-matrix with entries* $a_{ij} \in \mathsf{C}$ *and define the* Geršgorin discs *by*

$$\mathcal{G}_i = \{z \in \mathsf{C} : |z - a_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|\} \text{ , for } 1 \leq i \leq n.$$

Then all eigenvalues of $\mathbf{A}$ *are contained in* $\bigcup \mathcal{G}_i$*, the union of the Geršgorin discs.*

Proposition 9 *Let* $\mathbf{P} = \widetilde{\mathbf{P}} + \mathbf{E}$ *be an irreducible, stochastic and reversible* $(n \times n)$*-matrix partitioned as in (3). If* $\mathbf{S}$ *is the matrix of the corresponding stochastic complements and* $\mathbf{R}$ *the restriction matrix, then*

$$\lambda_j(\mathbf{S}) \leq \lambda_j(\mathbf{R}) \leq \lambda_j(\mathbf{S}) + 2\|\mathbf{E}\|_\infty$$

holds for all $j = 1, \ldots, n$*.*

Proof. For the proof assume that $\mathbf{S}$ will inherit reversibility from $\mathbf{P}$. We will show this fact afterwards. Set $\mathbf{U} := \mathbf{S} - \widetilde{\mathbf{P}}$ and $\mathbf{V} := \mathbf{R} - \mathbf{S}$, then the entries in $\mathbf{V}$ are

$$v_{ij} = \begin{cases} -u_{ij} \leq 0 & \text{, for } i \neq j, \\ e_i - u_{ii} & \text{, for } i = j, \end{cases}$$

where e_i is the $i-$th row sum of $\mathbf{E}$ and u_{ij} an entry in $\mathbf{U}$. The row sums of $\mathbf{U}$ and $\mathbf{E}$ are equal because both can be converted into a stochastic matrix by adding $\widetilde{\mathbf{P}}$. It follows that $\sum_{j=1}^{n} v_{ij} = 0$ and therefore

$$0 \leq v_{ii} = \sum_{\substack{j=1 \\ j \neq i}}^{n} |v_{ij}|,$$

for all $1 \leq i \leq n$. Applying Geršgorins Theorem shows that an eigenvalue of $\mathbf{V}$ can not be negative. Further we have

$$\|\mathbf{V}\|_\infty = 2 \max_i v_{ii} = 2 \max_i \sum_{\substack{j=1 \\ j \neq i}}^{n} u_{ij} \leq 2\|\mathbf{U}\|_\infty = 2\|\mathbf{E}\|_\infty.$$

Under the assumption that $\mathbf{S}$ is reversible it follows from Theorem 5 that

$$\lambda_k(\mathbf{R}) = \lambda_k(\mathbf{S} + \mathbf{V}) \geq \lambda_k(\mathbf{S}) + \lambda_n(\mathbf{V}) \geq \lambda_k(\mathbf{S})$$

and

$$\lambda_k(\mathbf{R}) = \lambda_k(\mathbf{S} + \mathbf{V}) \leq \lambda_k(\mathbf{S}) + \lambda_1(\mathbf{V}) \leq \lambda_k(\mathbf{S}) + 2\|\mathbf{E}\|_\infty.$$

□

Combining Theorem 7 and Proposition 9

$$\lambda_j(\mathbf{S}) \leq \lambda_j(\mathbf{P}) + 2\|\mathbf{E}\|_\infty \tag{9}$$

follows immediately. Thus the same inequality holds for stochastic complements and restrictions, but stochastic complementation will always be better or equal in lowering the subdominant eigenvalue. On the other hand, if $\|\mathbf{E}\|_\infty$ is small the difference between the spectra of $\mathbf{S}$ and $\mathbf{R}$ will be small too.

In order to complete the proof of Theorem 7, we still have to show that $\mathbf{S}$ inherits reversibility from $\mathbf{P}$. This is done by the following proposition, where we also prove reversibility of the coupling matrix.

Proposition 10 *Let $\mathbf{P}$ be an irreducible and reversible stochastic matrix which is partitioned as in Definition 1, then the following holds:*

(a) Each stochastic complement $\mathbf{S}_{ii}$, $1 \leq i \leq k$, is reversible.
(b) The coupling matrix $\mathbf{C}$ defined in Theorem 2 is reversible.

Proof. (a) As $\mathbf{P}$ is reversible,

$$\mathbf{P}^{(sym)} := \mathbf{D}\mathbf{P}\mathbf{D}^{-1}, \text{ with } \mathbf{D} := \operatorname{diag}(\sqrt{\boldsymbol{\pi}}),$$

is a symmetric matrix. If $\boldsymbol{\pi} = (\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)}, \ldots, \boldsymbol{\pi}^{(k)})$ is partitioned according to $\mathbf{P}$ and $\mathbf{D}_l := \operatorname{diag}(\sqrt{\boldsymbol{\pi}^{(l)}})$ then

$$\mathbf{P}^{(sym)} = \begin{pmatrix} \mathbf{D}_1\mathbf{P}_{11}\mathbf{D}_1^{-1} & \mathbf{D}_1\mathbf{P}_{12}\mathbf{D}_2^{-1} & \cdots & \mathbf{D}_1\mathbf{P}_{1k}\mathbf{D}_k^{-1} \\ \mathbf{D}_2\mathbf{P}_{21}\mathbf{D}_1^{-1} & \mathbf{D}_2\mathbf{P}_{22}\mathbf{D}_2^{-1} & \cdots & \mathbf{D}_2\mathbf{P}_{2k}\mathbf{D}_k^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{D}_k\mathbf{P}_{k1}\mathbf{D}_1^{-1} & \mathbf{D}_k\mathbf{P}_{k2}\mathbf{D}_2^{-1} & \cdots & \mathbf{D}_k\mathbf{P}_{kk}\mathbf{D}_k^{-1} \end{pmatrix}. \tag{10}$$

Let $\mathbf{S}_{ll}$ be an arbitrary stochastic complement of size $(r \times r)$ with stationary distribution $\mathbf{s}^{(l)}$. Reversibility of $\mathbf{S}_{ll}$ is defined via the detailed balance condition (1), which we verify by showing that

$$\widetilde{\mathbf{D}}_l\mathbf{S}_{ll}\widetilde{\mathbf{D}}_l^{-1}, \text{ with } \widetilde{\mathbf{D}}_l = \operatorname{diag}(\sqrt{\mathbf{s}^{(l)}}),$$

is a symmetric matrix.

Theorem 2 states that $\mathbf{s}^{(l)} = \xi_l^{-1}\boldsymbol{\pi}^{(l)}$ with a scalar coupling factor ξ_l, so we have $\widetilde{\mathbf{D}}_l = \xi_l^{-\frac{1}{2}}\mathbf{D}_l$ and therefore

$$\widetilde{\mathbf{D}}_l\mathbf{S}_{ll}\widetilde{\mathbf{D}}_l^{-1} = \mathbf{D}_l\mathbf{S}_{ll}\mathbf{D}_l^{-1} = \mathbf{D}_l\mathbf{P}_{ll}\mathbf{D}_l^{-1} + \mathbf{D}_l\mathbf{P}_{l*}(\mathbf{I} - \mathbf{P}_{l'})^{-1}\mathbf{P}_{*l}\mathbf{D}_l^{-1}.$$

The first term on the right hand side is symmetric, because it is a diagonal block of $\mathbf{P}^{(sym)}$. Thus it remains to show the symmetry of the rightmost addend.
Denote with $\mathbf{D}_{l'}$ the principal submatrix of $\mathbf{D}$ that is produced by deleting

the rows and columns belonging to $\boldsymbol{\pi}^{(l)}$, i.e.,

$$\mathbf{D}_{l'} = \mathrm{diag}(\mathbf{D}_1, \ldots, \mathbf{D}_{l-1}, \mathbf{D}_{l+1}, \ldots, \mathbf{D}_k).$$

Further define $\mathbf{P}_{l*}^{(sym)}$, $\mathbf{P}_{*l}^{(sym)}$, and $\mathbf{P}_{l'}^{(sym)}$ from $\mathbf{P}^{(sym)}$ the same way we defined $\mathbf{P}_{l*}$, $\mathbf{P}_{*l}$, and $\mathbf{P}_{l'}$ from $\mathbf{P}$ in Def. 1. Note that they are submatrices of a symmetric matrix but not symmetric themselves ($\mathbf{P}_{l*}^{(sym)}$ and $\mathbf{P}_{*l}^{(sym)}$ are not even square). Then we have

$$\begin{aligned}
\mathbf{D}_l\mathbf{P}_{l*}(\mathbf{I}-\mathbf{P}_{l'})^{-1}\mathbf{P}_{*l}\mathbf{D}_l^{-1} &= \mathbf{D}_l\mathbf{P}_{l*}\mathbf{D}_{l'}^{-1}\mathbf{D}_{l'}(\mathbf{I}-\mathbf{P}_{l'})^{-1}\mathbf{D}_{l'}^{-1}\mathbf{D}_{l'}\mathbf{P}_{*l}\mathbf{D}_l^{-1} \\
&= \mathbf{P}_{l*}^{(sym)}\mathbf{D}_{l'}(\mathbf{I}-\mathbf{P}_{l'})^{-1}\mathbf{D}_{l'}^{-1}\mathbf{P}_{*l}^{(sym)} \\
&= \mathbf{P}_{l*}^{(sym)}(\mathbf{I}-\mathbf{P}_{l'}^{(sym)})^{-1}\mathbf{P}_{*l}^{(sym)} =: \mathbf{U}.
\end{aligned}$$

Clearly, as the inverse of a symmetric matrix, $\mathbf{V} := (\mathbf{I}-\mathbf{P}_l^{(sym)})^{-1}$ is symmetric.

Let denote the entries of the matrices $\mathbf{U} = (u_{ij})$, $\mathbf{V} = (v_{ij})$, and $\mathbf{P}^{(sym)} = (p_{ij}^{(sym)})$. What remains to show is that $u_{ij} = u_{ji}$ holds for $i, j = 1, \ldots, n$. To that end, let $\mathcal{I}$ be the sorted set containing all row indices of $\mathbf{P}^{(sym)}$, except the ones belonging to the l-th block row. There will be $n-r$ indices in $\mathcal{I} = \{k_1, k_2, \ldots, k_{n-r}\}$ and

$$\begin{aligned}
u_{ij} &= \sum_{s=1}^{n-r} p_{k_ik_s}^{(sym)} \sum_{t=1}^{n-r} v_{st} p_{k_tk_j}^{(sym)} \\
&= \sum_{t=1}^{n-r} p_{k_tk_j}^{(sym)} \sum_{s=1}^{n-r} v_{st} p_{k_ik_s}^{(sym)} \\
&= \sum_{t=1}^{n-r} p_{k_jk_t}^{(sym)} \sum_{s=1}^{n-r} v_{ts} p_{k_sk_i}^{(sym)} = u_{ji}.
\end{aligned}$$

With $\mathbf{U}$ being symmetric, we have actually shown that $\widetilde{\mathbf{D}}_l\mathbf{S}_{ll}\widetilde{\mathbf{D}}_l^{-1}$, as sum of two symmetric matrices, is also symmetric. As already stated, this is equivalent to the reversibility of $\mathbf{S}_{ll}$ w.r.t. $\boldsymbol{\pi}^{(l)}$.

(b) If $\mathbf{P}$ is reversible then $\mathbf{D}^2\mathbf{P}$ is symmetric. For an entry c_{ij} of the coupling matrix $\mathbf{C}$ we have

$$\begin{aligned}
\xi_i c_{ij} &= \xi_i \mathbf{s}^{(i)}\mathbf{P}_{ij}\mathbf{e} = \boldsymbol{\pi}^{(i)}\mathbf{P}_{ij}\mathbf{e} \\
&= \boldsymbol{\pi}^{(i)}\mathbf{D}_i^{-2}\mathbf{D}_i^2\mathbf{P}_{ij}\mathbf{e} \\
&= \mathbf{e}^T\mathbf{D}_i^2\mathbf{P}_{ij}\mathbf{e}.
\end{aligned}$$

$\mathbf{D}_i^2\mathbf{P}_{ij} = \mathbf{D}_j^2\mathbf{P}_{ji}$ follows from the symmetry of $\mathbf{D}^2\mathbf{P}$, so that

$$\begin{aligned}\xi_i c_{ij} &= \mathbf{e}^T \mathbf{D}_j^2 \mathbf{P}_{ji} \mathbf{e} \\ &= \boldsymbol{\pi}^{(j)} \mathbf{P}_{ji} \mathbf{e} = \xi_j c_{ji},\end{aligned}$$

which is the detailed balance condition of $\mathbf{C}$ w.r.t. $\boldsymbol{\xi}$. □

We have shown bounds on subdominant eigenvalues of Markov chains produced by uncoupling. These bounds will provide reasonable information about the spectra of the uncoupled chains, if applied to a nearly uncoupled Markov chain in the sense of section 2.2, that is , if $\|\mathbf{E}\|_\infty$ is small. In that case, uncoupled chains are indeed rapidly mixing. This thread is reversible. Jerrum et al. [**?**] have shown that using the structure of nearly uncoupled Markov chains allows to state upper bounds on the subdominant eigenvalue of a Markov chain, when upper bounds are known for the subdominant eigenvalues of the restriction chains and the coupling matrix. By this means, uncoupling can be used to prove mixing properties for complicated Markov chains, via decomposing it into simpler ones [**?**,**?**].

4 Examples

4.1 Random Walk on Bunkbed Graphs

We first consider an analytically tractable example from graph theory, the class of so-called bunkbed graphs (see, e.g., [**?**]). Given any graph $G = (V, E)$, its associated bunkbed graph $G_2 = (V_2, E_2)$ is defined by

$$V_2 = V \times \{0, 1\}$$

and

$$E_2 = \{\langle (u, i), (v, i)\rangle : \langle (u, v)\rangle \in E, i \in \{0, 1\}\} \cup \{\langle (u, 0), (u, 1)\rangle : u \in V\}.$$

In other words, G_2 is the Cartesian product of the complete graph K_2 and G, which is obtained by placing a copy of G above G and connecting each edge in G with its corresponding edge in the copy. Let $\mathbf{G} = (g_{ij})$ and $\mathbf{P}$ be the adjacency matrices of G and G_2, respectively. Assuming that G is connected, we can then define a weakly coupled random walk on its bunkbed graph by

setting weights on the edges of G by

$$g_{ij} = \begin{cases} \frac{1-\epsilon}{d_*+1} & \text{if } i \text{ is adjacent to j,} \\ 0 & \text{if } i \text{ is not adjacent to j,} \\ (1-\epsilon)(1-\frac{d_i}{d_*+1}) & \text{if } i = j, \end{cases} \tag{11}$$

where $\epsilon \in (0,1)$, d_i is the degree of the vertex i in G and d_* is the maximum degree over the vertices in G; and additionally applying the weight ϵ to all vertices connecting the base graph with its copy. In other words, we construct the adjacency matrix $\mathbf{P}$ on the bunkbed graph by

$$\mathbf{P} = \begin{pmatrix} \mathbf{G} & \epsilon\mathbf{I} \\ \epsilon\mathbf{I} & \mathbf{G} \end{pmatrix}, \tag{12}$$

which by definition is symmetric, stochastic, and irreducible. Now, if we uncouple $\mathbf{P}$ into its two identical blocks in (12), the special structure of $\mathbf{P}$ allows to give analytic expressions for the eigenvalues of $\mathbf{P}$, $\mathbf{S}$, and $\mathbf{R}$ in terms of the eigenvalues of $\mathbf{G}$.

Proposition 11 *Let $\mathbf{P}$ be an irreducible stochastic matrix as given in (12), with $\epsilon > 0$ and an irreducible submatrix $\mathbf{G}$. If λ is a k-fold eigenvalue of $\mathbf{G}$, then:*

(a) $\lambda + \epsilon$, $\lambda - \epsilon$ are k-fold eigenvalues of $\mathbf{P}$,
(b) $\lambda + \epsilon$ is a $2k$-fold eigenvalue of $\mathbf{R}$,
(c) and $\lambda + \epsilon^2(1-\lambda)^{-1}$ is a $2k$-fold eigenvalue of $\mathbf{S}$.

Proof. Statement (b) is obvious, because $\mathbf{R}_{ii} = \mathbf{G} + \epsilon\mathbf{I}, i \in \{1,2\}$, which shifts the eigenvalues of $\mathbf{G}$ by ϵ.

Use (b) to prove (a) by choosing an eigenvector $\mathbf{v}_1 \neq \mathbf{0}$ for $\mathbf{R}_{11}$ to the eigenvalue $\lambda + \epsilon$, so that $\mathbf{R}_{11}\mathbf{v}_1 = (\lambda + \epsilon)\mathbf{v}_1$. From $\mathbf{R}_{11} = \mathbf{R}_{22}$ it follows that $(\mathbf{v}_1,\ \mathbf{v}_1)^T$ and $(\mathbf{v}_1,\ -\mathbf{v}_1)^T$ are eigenvectors of $\mathbf{R}$ to $\lambda + \epsilon$.
Observe that

$$\mathbf{R} = \mathbf{P} - \begin{pmatrix} \mathbf{0} & \epsilon\mathbf{I} \\ \epsilon\mathbf{I} & \mathbf{0} \end{pmatrix} + \begin{pmatrix} \epsilon\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \epsilon\mathbf{I} \end{pmatrix}.$$

Multiplying this equations with $(\mathbf{v}_1,\ \mathbf{v}_1)^T$ and $(\mathbf{v}_1,\ -\mathbf{v}_1)^T$ from the right gives

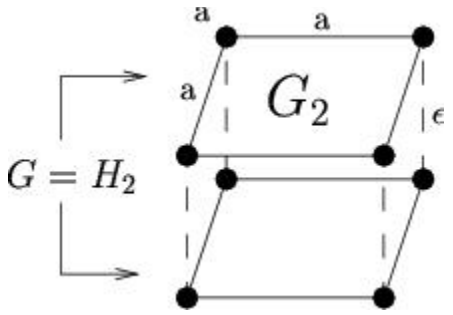
$$\mathbf{P}\begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_1 \end{pmatrix} = (\lambda + \epsilon)\begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_1 \end{pmatrix}$$

and

$$\mathbf{P}\begin{pmatrix} \mathbf{v}_1 \\ -\mathbf{v}_1 \end{pmatrix} = (\lambda - \epsilon)\begin{pmatrix} \mathbf{v}_1 \\ -\mathbf{v}_1 \end{pmatrix},$$

respectively, which is statement (a). Finally

$$\mathbf{S}_{ii} = \mathbf{G} + \epsilon\mathbf{I}(\mathbf{I} - \mathbf{G})^{-1}\epsilon\mathbf{I} = \mathbf{G} + \epsilon^2(\mathbf{I} - \mathbf{G})^{-1},$$



$$\mathbf{P} = \left(\begin{array}{cccc|cccc} a & a & a & 0 & \epsilon & 0 & 0 & 0 \\ a & a & 0 & a & 0 & \epsilon & 0 & 0 \\ a & 0 & a & a & 0 & 0 & \epsilon & 0 \\ 0 & a & a & a & 0 & 0 & 0 & \epsilon \\ \hline \epsilon & 0 & 0 & 0 & a & a & a & 0 \\ 0 & \epsilon & 0 & 0 & a & a & 0 & a \\ 0 & 0 & \epsilon & 0 & a & 0 & a & a \\ 0 & 0 & 0 & \epsilon & 0 & a & a & a \end{array}\right)$$

Fig. 1. Left: G_2 is a bunkbed construction of the 2-dimensional hypercube H_2. Right: if the weights on G_2 are chosen according to (11) the resulting stochastic matrix is $\mathbf{P}$ with $a = (1-\epsilon)/3$.

for $i \in \{1, 2\}$, so if $\mathbf{v}$ is an eigenvector to $\mathbf{G}$ and λ we obtain

$$\mathbf{S}_{ii}\mathbf{v} = \left(\lambda + \frac{\epsilon^2}{1-\lambda}\right)\mathbf{v}. \tag{13}$$

□

As an example take $G = H_d$, the d-dimensional hypercube, $\mathbf{G} = \mathbf{H}_d$ the weighted adjacency matrix as proposed in (11), and $\mathbf{P}$ the weighted adjacency matrix on the bunkbed graph G_2, as given by (12), see Fig. **??**. It is well-known that the eigenvalues of $\mathbf{H}_d$ are $(1-\epsilon)(1-\frac{2(k-1)}{d+1})$ with multiplicity $\binom{d}{k-1}$ for $1 \le k \le d+1$ (e.g. [**?**]). Therefore, due to Proposition 11, we have

$$\begin{aligned} \lambda_1(\mathbf{P}) &= (1-\epsilon) + \epsilon = 1, \\ \lambda_2(\mathbf{P}) &= (1-\epsilon) - \epsilon = 1 - 2\epsilon, \\ \lambda_3(\mathbf{P}) &= (1-\epsilon)\left(1 - \frac{2}{d+1}\right) - 2\epsilon = 1 - 2\frac{1-\epsilon}{d+1}, \end{aligned}$$

as long as $\epsilon < \frac{1}{d+2}$ (otherwise $\lambda_3(\mathbf{P})$ overtakes $\lambda_2(\mathbf{P})$). Uncoupling of $\mathbf{P}$ shifts λ_2 to 1 while λ_3 keeps bounded away from 1 with

$$\lambda_3(\mathbf{S}) \le \lambda_3(\mathbf{R}) \le \lambda_3(\mathbf{P}) + 2\epsilon$$

by Eqs. (8) and (9). Proposition 11 allows to calculate these eigenvalues explicitly, leading to

$$\begin{aligned} \lambda_3(\mathbf{R}) &= \lambda_3(\mathbf{P}), \\ \lambda_3(\mathbf{S}) &= \lambda_3(\mathbf{P}) - \epsilon\left(1 - \frac{\epsilon(d+1)}{2(1-\epsilon) + \epsilon(d+1)}\right). \end{aligned}$$

Thus, the given approximation deviates from the exact results by 2ϵ and $2\epsilon + \epsilon(1 - \frac{\epsilon(d+1)}{2(1-\epsilon)+\epsilon(d+1)})$.

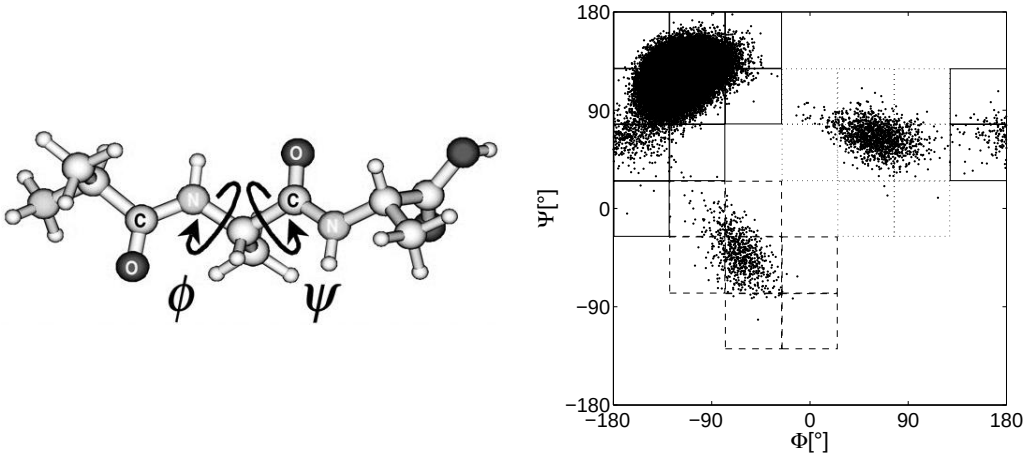


Fig. 2. Left: The trialanine molecule shown in ball-and-stick representation. The overall structure of trialanine is primarily determined by the two torsion angles Φ and Ψ. Right: Plotting Φ versus Ψ results in a so-called Ramachandran plot. The discretization boxes are plotted with different edge lines indicating the different metastable sets they were allocated to.

The analysis can be taken further by showing that $\lambda_2(\mathbf{P}) > \lambda_3(\mathbf{R})+2\epsilon$ is equivalent to $d < \frac{1}{2\epsilon} - \frac{3}{2}$, which is a straightforward calculation. For example, take $\epsilon = 0.01$, then the approximation indicates better subdominant eigenvalues in the diagonal blocks of $\mathbf{R}$ and $\mathbf{S}$ as long as $d < 48$.

4.2 Metastable Sets of Trialanine

Restricted Markov chains can be a useful tool for biomolecular simulations. As an example we consider trialanine, a small peptide composed of three alanine amino acid residues. The structural and dynamical properties of trialanine are primarily determined by the two torsion angles Φ and Ψ as shown in Fig. **??**. Exploration of the high-dimensional continuous state space can be done by means of Uncoupling-coupling Monte Carlo [**?**,**?**], which hierarchically decomposes the state space into metastable sets. We herein only illustrate the initial uncoupling step, which starts with a high-temperature Markov chain Monte Carlo simulation. More precisely, we used the Hybrid Monte Carlo method [**?**], a popular method in this field that combines the benefits of molecular dynamics with the statistical accuracy of Markov chain Monte Carlo. We sampled 10^5 steps at a temperature of 650K 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 in $\mathcal{D}^2$, see Fig. **??**. On these boxes we set up a transition matrix $\mathbf{P} = p_{ij}$, receiving the transition probabilities by counting the number of transitions between them during simulation. Reversibility can be inherited to $\mathbf{P}$ by counting each transition between box i and j as a transition between box j and i too. This approach

is justified, because we can think of $\mathbf{P}$ as of a discretization of a reversible continuous Markov operator governing the dynamics of the molecule [**?**,**?**]. Therefore, if B_{ij} denotes the number of transitions between box i and box j, and B_i the number of data points in box i, the transition probability between box i and box j is given by

$$p_{ij} = \frac{B_{ij} + B_{ji}}{B_i}.$$

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

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

indicating a slow mixing Markov chain with three metastable sets. As already pointed out at the end of Sect. 2.2, identification of metastable sets for given $\mathbf{P}$ is not a trivial task. As dynamical cluster algorithm we used a spectral approach, which analyzes the structure of dominant eigenvectors [**?**,**?**,**?**]. Identified metastable sets are indicated by different line styles in Fig. **??**. A corresponding permutation of the transition matrix confirms the computation in that it reveals an obvious block dominant structure, see Fig. **??**. The maximum row sum over entries outside these blocks is 0.0417, so that we can bound the effect of uncoupling on the subdominant eigenvalue by

$$\lambda_4(\mathbf{R}) \leq \lambda_4(\mathbf{P}) + 2 \cdot 0.0417 = 0.6526.$$

This means, that restriction to the three metastable sets will result in three Markov chains whose subdominant eigenvalues are significantly bounded away from 1. Calculating the subdominant eigenvalues of the restrictions $\mathbf{R}_{ii}$ for $i = 1, \ldots, 3$ shows that $\lambda_4(\mathbf{R}) \leq 0.6526$ is indeed a useful bound:

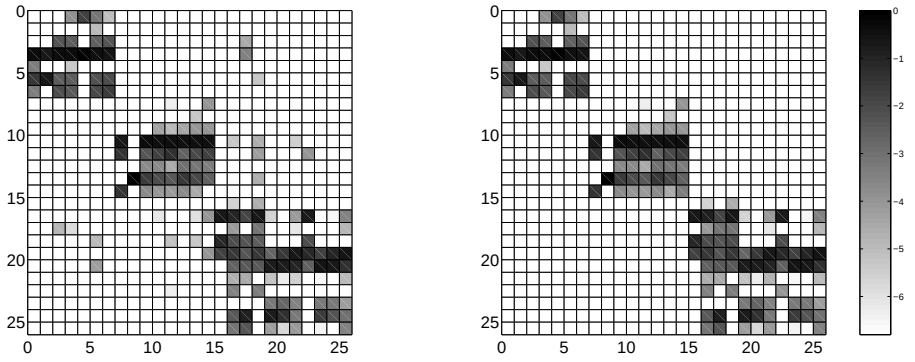


Fig. 3. Left: The permuted transition matrix $\mathbf{P}$ clearly has a block dominant structure. Right: In the resulting restricted matrix $\mathbf{R}$ all off-diagonal entries are set to zero. The intensity of the boxes is chosen due to the logarithmic scale on the far right.

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

Figure **??** illustrates an interesting effect that takes place if we choose a finer discretization of the torsions angles. This might create boxes which cover only a few data points. A transition from one of these boxes to another metastable set induces a large transition probability from the box into the subset. In the permuted transition matrix this leads to large entries outside the diagonal blocks, while the metastable structure is still preserved and uncoupling still lowers the subdominant eigenvalue, although we cannot predict this by Theorem 7.

For a closer consideration of this effect, we construct a (3×3) stochastic matrix. Let $\epsilon > 0$ and

$$\mathbf{P} = \left(\begin{array}{cc|c} \frac{1}{1+\epsilon} & \frac{\epsilon}{1+\epsilon} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \hline 0 & \frac{\epsilon}{1+\epsilon} & \frac{1}{1+\epsilon} \end{array}\right), \tag{14}$$

which is a reversible stochastic matrix with state space $S = \{s_1, s_2, s_3\}$ and stationary distribution

$$\boldsymbol{\pi} = \frac{1}{2+4\epsilon}(1+\epsilon, 2\epsilon, 1+\epsilon). \tag{15}$$

The eigenvalues of $\mathbf{P}$ are given by

$$\lambda_1(\mathbf{P}) = 1, \quad \lambda_2(\mathbf{P}) = \frac{1}{1+\epsilon}, \quad \text{and} \quad \lambda_3(\mathbf{P}) = -\frac{\epsilon}{1+\epsilon}.$$

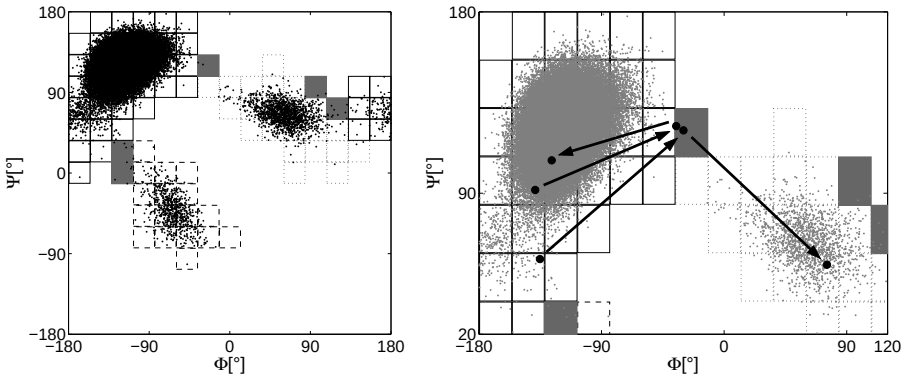


Fig. 4. Left: A refined discretization produces boxes with large transition probabilities to other metastable sets. Here each torsion angle is discretized in 15 intervals. Boxes with a transition probability to another metastable set of 0.25 and larger are marked dark. Right: A close up illustrates the cause of this effect. The arrows connect subsequent sample points in the realization of the Markov chain.

As ϵ tends to zero, $\lambda_2(\mathbf{P})$ tends to 1, thus indicating a slow mixing Markov chain. There is no possible partition that avoids the entry 0.5 outside the diagonal blocks, so that our eigenvalue bounds will provide no further information. Yet, if we restrict along the partition indicated in (**??**) we get

$$\mathbf{R} = \left(\begin{array}{cc|c} \frac{1}{1+\epsilon} & \frac{\epsilon}{1+\epsilon} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \hline 0 & 0 & 1 \end{array}\right),$$

with eigenvalues

$$\lambda_1(\mathbf{R}) = 1, \quad \lambda_2(\mathbf{R}) = 1, \quad \text{and} \quad \lambda_3(\mathbf{R}) = \frac{1-\epsilon}{2(1+\epsilon)}.$$

For ϵ close to zero, $\lambda_3(\mathbf{R})$ is close to 0.5, which shows that the two uncoupled chains are fast mixing. We can understand this effect by examining the underlying dynamics of the given Markov chain. Since the state s_2 makes transitions between s_1 and s_3 likely if it is reached, we call s_2 a transition state. But if ϵ is close to zero, the probability to be in the transition state is close to zero too, as we can see from the stationary distribution (**??**). So metastability is preserved, because the transition state is rarely reached in a realization of the Markov chain.

References

[1] R. Bhatia. *Matrix Analysis.* Springer-Verlag, Berlin-Heidelberg-New York, 1997.

[2] 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.

[3] 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.

[4] 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.

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

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

[7] 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.

[8] O. Häggström. Probability on bunkbed graphs. In K. Eriksson and S. Linusson, editors, *Formal Power Series and Algebraic Combinatorics*, pages 19–27, Linköpings universitet, 2003.

[9] D. Hartfiel and C. D. Meyer. On the structure of stochastic matrices with a subdominant eigenvalue near 1. *Lin. Alg. Appl*, (272):193–203, 1998.

[10] M. Jerrum, J.-B. Son, P. Tetali, and E. Vergoda. Elementary bounds on Poincarè and log-Sobolev constants for decomposable Markov chain. Isaac Newton Institute Preprint, 2003.

[11] N. Madras and D. Randall. Markov chain decomposition for convergence rate analysis. *Annals of Applied Probability*, 12:581–606, 2002.

[12] N. Madras and Z. Zheng. On the swapping algorithm. *Random Structures and Algorithms*, 22(1):66–97, 2003.

[13] R. B. Mattingly. A revised stochastic complementation algorithm for nearly completely decomposable Markov chains. *ORSA Journal on Computing*, 7(2):117–124, 1995.

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

[15] C. D. Meyer. Uncoupling the Perron eigenvector problem. *Lin. Alg. Appl.*, 114/115:69–94, 1989.

[16] B. Mohar. Some applications of Laplace eigenvalues of graphs. In G. Hahn and G. Sabidussi, editors, *Graph Symmetry: Algebraic Methods and Applications*, NATO ASI Ser. C497, pages 225–275. Kluwer, 1997.

[17] C. Schütte. *Conformational Dynamics: Modelling, Theory, Algorithm, and Application to Biomolecules*. Habilitation Thesis, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 1998.

[18] 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.

[19] 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.

[20] E. Seneta. *Non-negative Matrices and Markov Chains*. Springer Series in Statistics. Springer, New York, Heidelberg, Berlin, 1981.

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

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